



**Universidade de Brasília
Departamento de Estatística**

**Estudo de Evasão Acadêmica no curso de Licenciatura em Matemática da
Universidade de Brasília**

Thayane de Souza Soares

Relatório final apresentado para o Departamento de Estatística da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

**Brasília
2022**

Thayane de Souza Soares

**Estudo de Evasão Acadêmica no curso de Licenciatura em Matemática da
Universidade de Brasília**

Orientadora: Maria Teresa Leão Costa

Relatório final apresentado para o Departamento de Estatística da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

**Brasília
2022**

Resumo

A evasão acadêmica é uma problemática a nível mundial e devido as suas, muitas vezes, severas consequências, diversos estudos foram realizados afim de entender tal fenômeno e tentar cessar ou diminuir tais acontecimentos. Nesse presente trabalho o conceito de evasão é atribuído ao aluno que sai de um curso de origem por qualquer motivo que não a formatura. O estudo tem por objetivo verificar a evasão universitária dos discentes dos cursos de Licenciatura em Matemática da Universidade de Brasília, levando em consideração o turno do curso, bem como identificar quais são as características dos estudantes que estão associadas a evasão acadêmica e comparar as informações entre os dois cursos de interesse. Para alcançar o objetivo proposto foi utilizado 710 observações (337 para licenciatura diurno e 373 para o noturno) fornecidas pelo sistema de informação da Universidade de Brasília, com características sociodemográficas e acadêmicas dos estudantes matriculados entre os anos de 2014 a 2019 nos cursos de Licenciatura Diurno em Matemática e Licenciatura Noturno em Matemática. Tais fatores foram estudados por meio do método de Regressão Logística e ao modelar o estudo alguns fatores deram significativos, a quantidade de reprovações de um aluno e o seu índice de rendimento acadêmico foram alguns deles.

Palavras-chaves: evasão, regressão logística, Licenciatura em Matemática.

Lista de Tabelas

| | | |
|----|---|----|
| 1 | Agrupamento Raça/cor | 33 |
| 2 | Agrupamento por forma de ingresso | 34 |
| 3 | Decisão de menção | 34 |
| 4 | Agrupamento por taxa de reprovação | 36 |
| 5 | Disciplinas com os maiores índices de reprovações | 36 |
| 6 | Distribuição dos alunos cotistas por tipo de cotas. | 40 |
| 7 | Período de ingresso dos alunos, Licenciatura em Matemática - UnB, 2014-2019 | 41 |
| 8 | Evasão por semestre de ingresso, Licenciatura em Matemática - UnB, 2014-2019 | 41 |
| 9 | Forma de ingresso dos alunos, Licenciatura em Matemática - UnB, 2014 - 2019 | 42 |
| 10 | Taxa de reprovação geral dos alunos, Licenciatura em Matemática - UnB, 2014 - 2019 | 43 |
| 11 | Análise bivariada por evasão, Licenciatura Diurno em Matemática - UnB, 2014-2019. | 48 |
| 12 | Correlação entre sistema de cotas e tipo de escola do 2 ^o grau, Licenciatura Diurno em Matemática - UnB, 2014 - 2019 | 51 |
| 13 | Modelo proposto através da análise bivariada, Licenciatura Diurno em Matemática - UnB, 2014-2019. | 52 |
| 14 | Estimativas dos parâmetros para as bases de construção, validação e geral. Licenciatura Diurno em Matemática - UnB, 2014-2019. | 52 |
| 15 | Modelo final - Licenciatura Diurno em Matemática - UnB, 2014-2019. | 53 |
| 16 | Modelo para evasão, Licenciatura Diurno da Matemática - UnB, 2014 - 2019 | 53 |
| 17 | Teste Hosmer - Lemeshow, Licenciatura Diurno da Matemática - UnB, 2014 - 2019 | 54 |
| 18 | Matriz de confusão, Licenciatura Diurno da Matemática - UnB, 2014 - 2019 | 55 |
| 19 | Análise bivariada por evasão, Licenciatura Noturno em Matemática - UnB, 2014-2019. | 56 |

| | | |
|----|---|----|
| 20 | Correlação entre cota e tipo de escola do 2 ^o grau, Licenciatura Noturno em Matemática - UnB, 2014 - 2019 | 59 |
| 21 | Modelo proposto através da análise bivariada, Licenciatura Noturno em Matemática - UnB, 2014-2019. | 60 |
| 22 | Estimativas dos parâmetros para as bases de construção, validação e geral, Licenciatura Noturno em Matemática - UnB, 2014-2019. | 61 |
| 23 | Modelo final, Licenciatura Noturno em Matemática - UnB, 2014-2019. | 61 |
| 24 | Modelo para evasão, Licenciatura Noturno da Matemática - UnB, 2014 - 2019 | 61 |
| 25 | Teste Hosmer - Lemeshow para qualidade de modelo | 62 |
| 26 | Matriz de confusão - Licenciatura Noturno de Matemática - UnB, 2014 - 2019 | 63 |

Lista de Figuras

| | | |
|----|--|----|
| 1 | Curva da Função de Regressão Logística | 20 |
| 2 | Função Logística para diferentes valores de β_1 | 20 |
| 3 | Exemplos de gráficos de resíduos. Fonte: Neter et al. Applied Linear Statistical Models (p. 595). | 28 |
| 4 | Distribuição dos Alunos por gênero, Licenciatura em Matemática - UnB, 2014-2019 | 38 |
| 5 | Idade dos alunos, Licenciatura em Matemática - UnB, 2014-2019 | 38 |
| 6 | Raça/cor autodeclarada dos alunos, Licenciatura em Matemática - UnB, 2014 - 2019 | 39 |
| 7 | Tipo de escola do 2 ^o grau dos alunos, Licenciatura em Matemática - UnB, 2014 - 2019 | 40 |
| 8 | Taxa de reprovação por número de semestres cursados, Licenciatura em Matemática - UnB, 2014 - 2019 | 43 |
| 9 | Disciplinas mais reprovadas pelos alunos, Licenciatura em Matemática - UnB, 2014 - 2019 | 44 |
| 10 | Disciplinas com maior atribuição de menção SR, Licenciatura em Matemática - UnB, 2014 - 2019 | 45 |
| 11 | Disciplinas mais cursadas no período verão, Licenciatura em Matemática - UnB, 2014 - 2019 | 46 |
| 12 | Disciplinas mais trancadas, Licenciatura em Matemática - UnB, 2014 - 2019 | 46 |
| 13 | Forma de saída do curso, Licenciatura em Matemática - UnB, 2014 - 2019 | 47 |
| 14 | Correlação entre IRA e Taxa de reprovação, Licenciatura Diurno em Matemática - UnB, 2014 - 2019 | 51 |
| 15 | Resíduos do modelo, Licenciatura Diurno em Matemática - UnB, 2014-2019 | 54 |
| 16 | Curva ROC, Licenciatura Diurno em Matemática - UnB, 2014-2019 | 55 |
| 17 | Correlação entre IRA e Taxa de reprovação, Licenciatura Noturno em Matemática - UnB, 2014 - 2019 | 59 |
| 18 | Resíduos do modelo, Licenciatura Noturno em Matemática - UnB, 2014-2019 | 62 |
| 19 | Curva ROC, Licenciatura Noturno em Matemática - UnB, 2014-2019. | 63 |

Sumário

| | |
|---|----|
| 1 Introdução | 13 |
| 2 Objetivos | 17 |
| 2.1 Objetivo Geral | 17 |
| 2.2 Objetivos Específicos | 17 |
| 3 Referencial Teórico | 19 |
| 3.1 Regressão Logística | 19 |
| 3.2 Estimação dos parâmetros do modelo | 21 |
| 3.3 Regressão Logística Múltipla | 23 |
| 3.4 Inferência no Modelo de Regressão Logística | 24 |
| 3.4.1 Teste da Razão da Verossimilhança | 24 |
| 3.4.2 Teste de Wald | 25 |
| 3.4.3 Intervalos de Confiança | 25 |
| 3.5 Seleção de Modelos | 25 |
| 3.6 Diagnóstico do Modelo | 27 |
| 3.6.1 Resíduos de Pearson | 27 |
| 3.6.2 Resíduos de Pearson Semistudentizados | 28 |
| 3.6.3 Gráficos de Resíduos: | 28 |
| 3.7 Inferência para a resposta média | 30 |
| 3.8 Validação do Modelo | 30 |
| 4 Metodologia | 32 |
| 4.1 Dados | 32 |
| 4.2 Variáveis criadas | 33 |
| 4.3 Método | 37 |
| 5 Resultado | 38 |
| 5.1 Dados Pessoais | 38 |
| 5.2 Vida Acadêmica | 41 |
| 5.3 Licenciatura Diurna em Matemática | 48 |
| 5.3.1 Evasão | 48 |

| | | |
|----------|--|-----------|
| 5.3.2 | Correlação entre variáveis | 50 |
| 5.3.3 | Modelagem | 52 |
| 5.3.4 | Interpretação dos Parâmetros | 53 |
| 5.3.5 | Testes de ajustes e diagnósticos do modelo | 54 |
| 5.3.6 | Curva ROC | 54 |
| 5.3.7 | Matriz de Confusão | 55 |
| 5.4 | Licenciatura Noturno de Matemática | 56 |
| 5.4.1 | Evasão | 56 |
| 5.4.2 | Correlação entre as variáveis | 58 |
| 5.4.3 | Modelagem | 60 |
| 5.4.4 | Interpretação dos Parâmetros | 61 |
| 5.4.5 | Testes de ajustes e diagnósticos do modelo | 62 |
| 5.4.6 | Curva ROC | 62 |
| 5.4.7 | Matriz de Confusão | 63 |
| 6 | Conclusão | 65 |
| 7 | Referências | 69 |

1 Introdução

A evasão escolar é uma preocupação histórica não apenas a nível nacional, como também a nível mundial. A escolaridade, sobretudo infantil, ultrapassa o paradigma de dever, sendo um direito previsto na Constituição Federal Brasileira. Segundo a Lei 12.796 (que altera a Lei nº 9.394 que estabelece as diretrizes e bases da educação nacional - LDB) em seu artigo "Art 4º I - educação básica obrigatória e gratuita dos 4 (quatro) aos 17 (dezessete) anos de idade, organizada da seguinte forma: a) pré-escola; b) ensino fundamental; c) ensino médio".

Muitos são os motivos para o abandono escolar, em particular no ensino médio onde é um evento recorrente. Dentre todos motivos que cercam esse desfecho, pode-se apontar a desigualdade socioeconômica brasileira, de forma que muitos estudantes acabam por abandonar a escola para trabalhar e assim suprir necessidades básicas.

Em contrapartida o ensino superior não é obrigatório e embora esteja cada vez mais acessível a diferentes níveis sociais, ainda assim as desigualdades persistem, significando para muitos a realização de um sonho pessoal e profissional, para outros, além da realização profissional é a maneira encontrada para melhorar de vida no que tange a recursos financeiros. Tais desigualdades podem influenciar em suas trajetórias acadêmicas e até mesmo se vão conseguir concluir ou não o curso.

É certo que o ensino superior, seja por realização pessoal ou por necessidade de engajar em uma carreira promissora, tem uma elevada relevância. Porém, assim como no ensino básico, no ensino superior há uma taxa de evasão assustadora. Segundo Ambiel (2018), "Dados do Ministério da Educação (MEC), que avaliou o período entre 2010 e 2014, indicaram uma taxa de evasão universitária de 11,4% em 2010, que aumentou para 49% em 2014". "Já o Censo da educação superior 2018, elaborado pelo MEC (2018), estudou a trajetória dos alunos universitários que iniciaram seus estudos em 2010 e aponta uma taxa de desistência (evasão) em nada menos que 56,8%" (MUSLINER, 2021).

Devido a alta taxa de evasão universitária, diversos estudos foram feitos a fim de entender as causas desses acontecimentos e agir sobre eles de modo a reduzir tal fenômeno e contribuir para permanência dos estudantes na universidade.

Segundo CASTRO (2014, p.11-12), a teoria de Tinto (1975) é uma dos modelos de evasão mais conhecidos, e embora tenha se mostrado insuficiente para descrever sozinho satisfatoriamente a questão da evasão ou da permanência dos estudantes em seus cursos ainda é muito importante quando se fala em evasão universitária. Ele propõe como aspectos fundamentais para a permanência ou evasão do aluno dois processos: a integração acadêmica e a integração social. A integração acadêmica refere-se ao sentimento de estar integrado ao ambiente da universidade no que diz respeito ao contexto do curso e suas demandas acadêmicas. Inclui a percepção do estudante sobre o seu desempenho acadêmico, a autoestima relacionada a esse desempenho, a percepção de desenvolvimento pessoal, o sentimento de gostar dos conteúdos do curso e uma identificação com suas normas e valores, bem como com o papel de estudante. Já a integração social refere-se ao sentimento de fazer parte de um grupo e de sentir-se bem no ambiente da universidade, o que se manifesta por meio de contatos com colegas de curso (amizades, conversas), participação em atividades de caráter social na universidade (como festas, festivais, esportes, eventos culturais) e também contatos informais com professores e com o pessoal de apoio dos cursos. Além desses dois aspectos, a teoria de Tinto aponta dois outros fatores fundamentais para permanência ou evasão: o compromisso com a instituição (importância de ser aluno de uma determinada universidade); e a meta de concluir a graduação, ou importância de obter o diploma (CABRERA et al., 1992; TINTO, 1975).

É sabido que há múltiplos conceitos de evasão universitária, e no presente estudo, a evasão universitária é dada pela não conclusão do curso escolhido, não importando se concluiu o ensino superior em outra área ou universidade. Com o intuito de promover a permanência dos alunos nos cursos e tendo em consideração a importância do contexto e a multiplicidade de fatores relacionados à evasão, as instituições de ensino superior têm investido cada vez mais em políticas educacionais (COSTA & DIAS, 2015).

A identificação precoce dos estudantes vulneráveis propensos à evasão é crucial para que se obtenha sucesso nas estratégias educacionais. Antecipar esse problema com técnicas preditivas de identificação de fatores de risco, poderia evitar desperdícios financeiros, diminuir taxas de evasão e melhorar a reputação da instituição em relação aos seus pares pois, com a redução da evasão, conseqüentemente obteria maiores índices de formatura, contribuindo assim, para melhora da imagem da instituição (MÁRQUEZ-VERA et al., 2016) (NEILD; BALFANZ; HERZOG, 2007).

Segundo Simas (2012) em média dois em cada dez estudantes, no Brasil, desistem do curso superior. A não conclusão do curso superior expressa uma frustração em inúmeros pontos, o aluno que não concluiu, a família que de certo modo esperava por isso, o estado que não poderá contar com mais essa mão de obra e que no cunho de universidade federal investiu dinheiro naquele tempo cursado, dentre outros fatores. Em Ambiel (2018) "Os motivos para que a pessoa se evada de um curso são diversos, como insatisfação com a escolha profissional, problemas financeiros e de saúde ou insatisfação e dificuldade de

adaptação ao curso escolhido ou ao novo contexto”.

Em Mussliner, B. O. et al (2021, p.42682) foi observado que a evasão é grandemente influenciada pelo rendimento acadêmico do aluno nos primeiros períodos do curso. Por isso, cursos nos quais há muitos casos de reprovação nos períodos iniciais tendem a apresentar elevadas taxas de evasão. Outro ponto interessante trazido pelos autores foi que, especificamente na área de ciências exatas, o baixo desempenho nas provas de ingresso coincide com altas taxas de evasão nesses mesmos cursos. E mais: de acordo com esse trabalho, a criação de cursos oferecidos no turno noturno contribuiu para reduzir as taxas de evasão nos cursos oferecidos anteriormente apenas no turno diurno.

”A Universidade de Brasília (UnB) é uma das mais importantes instituições públicas Brasileiras. Foi avaliado pelo Ministério da Educação, através do Índice Geral de Cursos, como a melhor universidade da região Centro-Oeste”. Mesmo presente no topo de diversos Rankings nacionais e internacionais, a universidade de Brasília também enfrenta problemas com evasão. Segundo o CPA (2012) ”Dentre os resultados obtidos, destaca-se que, devido à expansão universitária, a taxa anual de evasão subiu no ano de 2011 atingindo um patamar relativamente estável desde 2012”. Atualmente a taxa anual de evasão da UnB é melhor que a média das instituições brasileiras de ensino superior, se encontra em torno de 11%, entretanto, isso não significa que a universidade não sofre com tal problemática.

Na matéria intitulada como ”UnB está entre universidades com menores taxas de evasão do país” divulgada por Thais Moura em 20/02/2019 8:29, atualizado 20/02/2019 8:53 no portal Metrópolis, mostra que a Universidade de Brasília tem um dos menores índices de evasão, mas que dentro da universidade há casos preocupantes. São exemplos os cursos de Física com a maior taxa de evasão (72%), seguido em segundo lugar do curso de Matemática com uma taxa de 70% e em terceiro lugar o curso de enfermagem com aproximadamente 70% de evasão também.

Em análise realizada pelo Decanato de Ensino de Graduação – UnB (2016) foi divulgado que dentro do Instituto de Exatas, formado pelos cursos Matemática (bacharel e licenciatura), Ciência da Computação e Estatística a taxa de evasão de discentes do curso de Matemática da Universidade de Brasília, ingressantes no período de 2002 a 2008 foi de 44,2%, sendo que no curso de Matemática diurno essa porcentagem foi de 57,3% e no noturno foi 60,8% considerando o mesmo período.

O quantitativo de formandos em licenciatura em Matemática em universidades públicas ou privadas em todo o Brasil é sempre muito baixo (BITTAR et. al., 2012). Segundo Gatti (1997 apud Bittar 2012) somente 6,2% dos estudantes que começam a graduação em Matemática concluem o curso. É necessário conhecer os motivos da evasão, para que as instituições tomem medidas cabíveis a fim de amenizar tal fenômeno, uma vez que os licenciados não são suficientes para suprir a demanda de docentes na educação

básica.

Nesse contexto, o presente estudo concentra-se em estudar a evasão universitária dos discentes do curso de Licenciatura em Matemática da Universidade de Brasília, levando em consideração o turno do curso, e identificar quais são as características dos estudantes que estão associadas a evasão acadêmica.

2 Objetivos

2.1 Objetivo Geral

Este estudo tem como objetivo geral analisar a evasão dos discentes do curso de Licenciatura em Matemática (diurno e noturno) da Universidade de Brasília (UnB) identificando quais são as características dos estudantes que estão associadas a evasão acadêmica.

2.2 Objetivos Específicos

- Traçar o perfil do aluno do curso de Matemática Licenciatura da Universidade de Brasília;
- Verificar se há correlação entre a evasão e fatores sociodemográficas e acadêmicos dos estudantes;
- Identificar possíveis características associadas entre os que evadiram;
- Averiguar se há diferenças expressivas na evasão entre os cursos Licenciatura Diurno em Matemática e Licenciatura Noturno em Matemática.

3 Referencial Teórico

3.1 Regressão Logística

A Regressão Logística é um método estatístico utilizado para prever valores assumidos por uma variável resposta categórica em função de uma ou mais variáveis explicativas. É também conhecida como modelo logístico, modelo logit, e classificador de máxima entropia. Inicialmente, a Regressão Logística foi muito utilizada na área da saúde, porém atualmente a sua aplicação é mais difundida e ela é aplicada em uma grande diversidade de áreas como a economia, econometria e finanças, dentre outros.

O modelo logístico tem como principal característica uma variável resposta ou dependente que se caracteriza por ser categórica e frequentemente binária. No caso da variável resposta binária temos duas categorias designadas como sucesso ou fracasso. Portanto, a variável resposta, denominada como Y , segue uma distribuição Bernoulli.

$$Y = \begin{cases} 1, & \text{se ocorrer sucesso} \\ 0, & \text{se ocorrer fracasso} \end{cases}$$

e a probabilidade de sucesso é igual a π_i , sendo que $0 \leq \pi \leq 1$. Assim:

$$P(Y = 1) = \pi \quad e \quad P(Y = 0) = 1 - \pi.$$

Considerando uma amostra aleatória de tamanho n temos que Y_i são variáveis aleatórias independentes que seguem distribuição Bernoulli com valores $E[Y_i] = \pi_i$.

O Modelo de Regressão Logística Binária Simples é definido como:

$$Y_i = E[Y_i] + \epsilon_i,$$

e tem por objetivo estimar π_i como uma função da variável explicativa X_i da seguinte forma:

$$E[Y_i] = \pi_i = \frac{e^{(\beta_0 + \beta_1 X_i)}}{1 + e^{(\beta_0 + \beta_1 X_i)}}. \quad (3.1.1)$$

Ao aplicarmos a seguinte transformação, denominada transformação logito temos que:

$$\log(\pi_i) = \ln\left(\frac{\pi_i}{1 - \pi_i}\right) = \beta_0 + \beta_1 X_i, \quad (3.1.2)$$

onde os β_i são coeficientes do modelo, enquanto X_i são os valores da variável explicativa X, com $i = 1, 2, 3, \dots, n$.

As Figuras 1 e 2 apresentam a forma da função de Regressão Logística em geral e como ela se modifica conforme o valor de β_1 . É possível observar:

- o efeito que o coeficiente exerce sobre a chance de um evento ocorrer;
- que coeficientes positivos aumentam a probabilidade de sucesso, enquanto um coeficiente negativo diminui a probabilidade de sucesso.

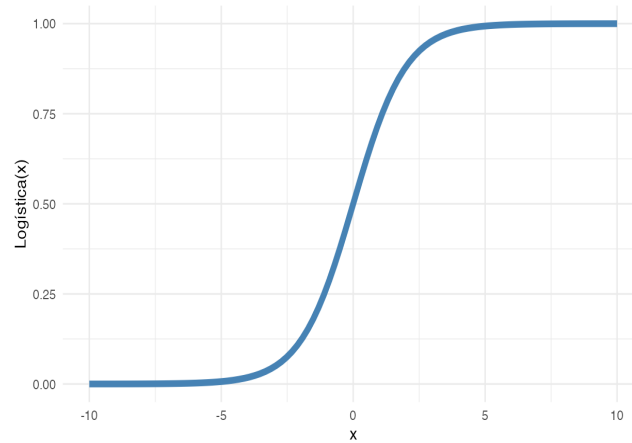


Figura 1: Curva da Função de Regressão Logística

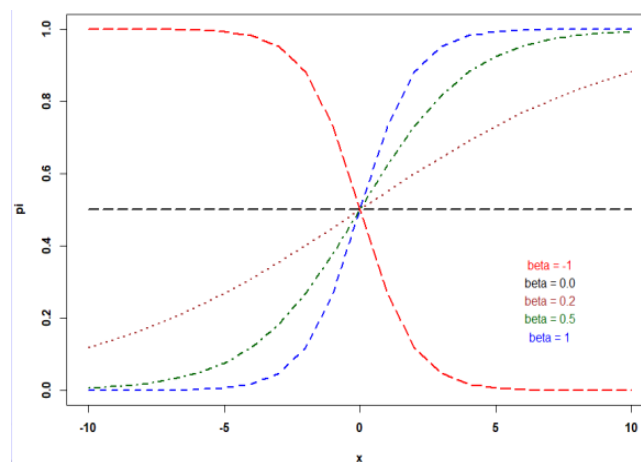


Figura 2: Função Logística para diferentes valores de β_1

Interpretação de $\hat{\beta}_1$

Considerando o valor da função de resposta *logit* ajustada quando $X = X_j$ dado por:

$$\hat{\pi}'(X_j) = \hat{\beta}_0 + \hat{\beta}_1 X_j, \quad (3.1.3)$$

sendo:

$$\hat{\pi}'(X_j) = \ln \left(\frac{\hat{\pi}}{1 - \hat{\pi}} \right). \quad (3.1.4)$$

Considerando também o valor da função de resposta *logit* ajustada em $X = X_j + 1$

$$\hat{\pi}'(X_j + 1) = \hat{\beta}_0 + \hat{\beta}_1(X_j + 1), \quad (3.1.5)$$

têm-se que então a diferença entre os dois valores ajustados é dada por:

$$\hat{\pi}'(X_j + 1) - \hat{\pi}'(X_j) = \hat{\beta}_1. \quad (3.1.6)$$

Sendo assim, de acordo com (3.1.2), $\hat{\pi}'(X_j)$ é o logaritmo da chance (*odds*) estimada quando $X = X_j$, designada por $\ln(odds_1)$. Da mesma forma, $\hat{\pi}'(X_j + 1)$ é o logaritmo da chance (*odds*) estimada quando $X = X_j + 1$, designada por $\ln(odds_2)$. Daí a diferença entre os dois valores de resposta *logit* ajustados podem ser expressos da seguinte forma:

$$\ln(odds_2) - \ln(odds_1) = \ln \left(\frac{odds_2}{odds_1} \right) = \hat{\beta}_1. \quad (3.1.7)$$

Portanto, pode-se observar que a razão de chances estimada é dada por:

$$\hat{OR} = \left(\frac{odds_2}{odds_1} \right) = e^{\hat{\beta}_1}. \quad (3.1.8)$$

3.2 Estimação dos parâmetros do modelo

Afim de estimar os coeficientes β_0 , β_1 do modelo de regressão logística binária, é utilizado o método da máxima verossimilhança. A estimação por método da máxima verossimilhança, permite encontrar os estimadores dos parâmetros do modelo (que são dados por $\hat{\beta}_0$ e $\hat{\beta}_1$), que tem maior probabilidade de replicar o padrão de observações, nos dados da amostra.

Supondo:

$$P(Y_i = y|x_i) = \begin{cases} \pi_i & \text{se } y = 1 \\ 1 - \pi_i, & \text{se } y = 0 \end{cases}$$

onde considera-se Y_i variáveis aleatórias independentes e identicamente distribuídas.

Logo, têm-se que

$$Y_i|X_i \text{ Bernoulli}(\pi_i),$$

e

$$P_{Y_i}(y_i) = \pi_i^{y_i}(1 - \pi_i)^{1-y_i}.$$

A função verossimilhança é dada por:

$$L(\beta; Y_i; X_i) = \prod_{i=1}^n \pi(X_i)^{Y_i} [1 - \pi(X_i)]^{1-Y_i}. \quad (3.2.1)$$

Aplicando (3.1.1) para substituir π_i em (3.2.1)

$$L(\beta; Y_i; X_i) = \prod_{i=1}^n \left(\frac{e^{\beta X_i}}{1 + e^{\beta X_i}} \right)^{Y_i} \left(\frac{1}{1 + e^{\beta X_i}} \right)^{1-Y_i},$$

a função log-verossimilhança será

$$l(\beta) = \sum_{i=1}^n [Y_i \beta X_i - \log(1 + e^{\beta X_i})]. \quad (3.2.2)$$

E por fim, derivando (3.2.2) em relação ao p-ésimo coeficiente e igualando a zero:

$$\frac{\partial l(\beta)}{\partial \beta_p} = \sum_{i=1}^n X_{it}(Y_i - \pi_i) = 0. \quad (3.2.3)$$

Os valores que satisfazem o sistema de equações resultante em (3.2.3) são os estimadores dos coeficientes do modelo. Não há uma fórmula fechada para os valores de que maximizem a função de máxima verossimilhança. Para encontrar estimativas de máxima verossimilhança é necessário utilizar métodos numéricos de solução, como o método de Newton-Raphson.

3.3 Regressão Logística Múltipla

Enquanto na Regressão Logística Simples têm-se a expressão (3.1.1) na Regressão Logística Múltipla está é dada por:

$$E[Y_i] = \pi_i = \frac{e^{(X'_i\beta)}}{1 + e^{(X'_i\beta)}}, \quad (3.3.1)$$

onde:

$$X'\beta = \beta_0 + \beta_1 X_1 + \dots + \beta_{p-1} X_{p-1}, \quad (3.3.2)$$

enquanto, da mesma forma têm-se a extensão de

$$\pi'_i = \ln \left(\frac{\pi}{1 - \pi} \right), \quad (3.3.3)$$

sendo:

$$\pi'_i = X'_i\beta. \quad (3.3.4)$$

Portanto, o modelo de Regressão Linear Múltiplo pode ser enunciado da seguinte forma: Y_i são variáveis aleatórias independentes de Bernoulli com valores $E\{Y_i\} = \pi_i$, onde:

$$E[Y_i] = \pi_i = \frac{e^{(X'_i\beta)}}{1 + e^{(X'_i\beta)}}, \quad (3.3.5)$$

onde X são constantes conhecidas e $E[Y_i]$ é a média condicional de Y_i dados os valores de $X_{i1}, X_{i2}, \dots, X_{i,p-1}$.

A função de Regressão Logística Múltipla é monótona e sigmoïdal em relação a $X'\beta$ e é quase linear quando π está entre 0,2 e 0,8. As variáveis X podem ser variáveis preditoras ou explicativas diferentes, ou algumas podem representar efeitos de curvatura e/ou interação. Além disso, as variáveis preditoras ou explicativas podem ser quantitativas, ou podem ser qualitativas e representadas por variáveis indicadoras. Essa flexibilidade torna o modelo de regressão logística múltipla muito atraente.

Assim como na regressão logística simples, na regressão logística múltipla, também é utilizado o método da máxima verossimilhança para estimar os parâmetros do modelo.

A função de probabilidade para regressão logística múltipla ajustada é dada por:

$$\hat{\pi}_i = \frac{e^{(X'_i\hat{\beta})}}{1 + e^{(X'_i\hat{\beta})}} = \left[1 + e^{-X'_i\hat{\beta}} \right]^{-1}, \quad (3.3.6)$$

onde:

$$X'_i \hat{\beta} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \dots + \hat{\beta}_{(p-1)} X_{(i,p-1)}. \quad (3.3.7)$$

3.4 Inferência no Modelo de Regressão Logística

Após o ajuste do modelo é necessário testar a significância do modelo ajustado bem como a significância dos parâmetros estimados das variáveis explicativas do modelo.

3.4.1 Teste da Razão da Verossimilhança

O Teste da Razão da Verossimilhança é um teste estatístico utilizado para verificar a significância dos coeficientes estimados das variáveis explicativas do modelo. Este teste busca avaliar se todos os parâmetros β_j 's associados às variáveis explicativas são nulos, o que indicaria a não existência de regressão.

As hipóteses a serem testadas são da seguinte forma:

$$H_0) \beta_1 = \beta_2 = \beta_3 = \dots = \beta_p = 0$$

$$H_1) \exists \beta_i \neq \beta_j$$

A medida do Teste da Razão da Verossimilhança testa simultaneamente, se os coeficientes de regressão associados a β são todos nulos com exceção de β_0 .

$$G^2 = -2 \log_e \left[\frac{L(R)}{L(F)} \right] = -2 [\log_e L(R) - \log_e L(F)]. \quad (3.4.1)$$

Observe que se a razão $L(R)/L(F)$ for pequena, então G^2 é grande, indicando que H_1 é a conclusão apropriada. Assim, grandes valores de G^2 levam à conclusão de H_1 . A teoria da grande amostra afirma que quando n é grande, G^2 é distribuído aproximadamente como $\chi^2(p - q)$ quando H_0 vale. Os graus de liberdade correspondem a $df_R - df_F = (n - q) - (n - p) = p - q$. A regra de decisão apropriada, portanto, é:

$$\text{se } G^2 \leq \chi^2(1 - \alpha; p - q) \text{ conclui-se } H_0$$

$$\text{se } G^2 \geq \chi^2(1 - \alpha; p - q) \text{ conclui-se } H_1$$

3.4.2 Teste de Wald

É utilizado na regressão logística para a determinação da significância dos coeficientes do modelo estimado, ele testa se cada coeficiente é significativamente diferente de zero. No caso dos coeficientes logísticos forem estatisticamente significativos, podemos interpretá-los em termos do seu impacto na probabilidade estimada.

As hipóteses a serem testadas são:

$$H_0 : \beta_i = 0 \text{ vs } H_1 : \beta_i \neq 0.$$

E a estatística do teste será dada por:

$$W_i = \frac{\hat{\beta}_i}{\hat{\sigma}(\hat{\beta}_i)}. \quad (3.4.2)$$

Sob hipótese nula, essa estatística possui distribuição assintótica qui-quadrado com 1 grau de liberdade. Para valores maiores que o quantil com acumulada igual a $1 - \alpha$, rejeitamos H_0 .

3.4.3 Intervalos de Confiança

O intervalo de confiança de um determinado β_i com confiança de $100(1 - \alpha)\%$ é definido como:

$$IC[\beta_i; 100(1 - \alpha)\%] = [\hat{\beta}_i \pm Z_{1-(\alpha/2)} \hat{\sigma}(\hat{\beta}_i)], \quad (3.4.3)$$

em que $Z_{1-(\alpha/2)}$ é a quantil da distribuição normal padrão cuja probabilidade acumulada é $1 - (\alpha/2)$.

3.5 Seleção de Modelos

Tem por objetivo identificar um grupo menor de modelos de regressão que são “bons” segundo um critério. Os critérios mais utilizados são:

- AIC_p

O critério de informação de Akaike (AIC) é uma métrica que mensura a qualidade de um modelo estatístico visando também a sua simplicidade.

O valor do modelo é dado da seguinte forma:

$$AIC_p = -\ln L(\hat{\beta}) + 2p, \quad (3.5.1)$$

sendo que, quanto menor o valor do AIC mais adequado o modelo é considerado.

- SBC_p

O critério de Informação Bayesiano - SBC (também conhecido como BIC) mede a qualidade do modelo da seguinte forma:

$$SBC_p = -\ln L(\hat{\beta}) + p \ln(n), \quad (3.5.2)$$

sendo que, quanto menor o valor do SBC mais adequado o modelo é considerado.

- *Pseudo R*²

Mecanismo utilizado para constatar o desempenho do modelo gerado. Encontra o ponto onde adicionar mais variáveis não é útil pois o aumento de R_p^2 é pequeno

O valor do modelo é dado da seguinte forma:

$$R_p^2 = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y}_i)^2}, \quad (3.5.3)$$

sendo

$\sum_{i=1}^n (Y_i - \hat{Y}_i)^2$ = medida da variação total de quando a variável explicativa é considerada.

e

$\sum_{i=1}^n (Y_i - \bar{Y}_i)^2$ = medida da variação total de Y_i .

O *pseudoR*² só tem significado quando comparado a outro *pseudoR*² do mesmo tipo, com os mesmos dados, prevendo o mesmo resultado. Nesta situação, o *pseudoR*² com maior valor indica qual modelo prevê melhor o resultado.

Existem vários procedimentos que podem ser utilizados para seleção de variáveis, com o objetivo de buscar um modelo parcimonioso que descreva o fenômeno, ou seja, com o menor número de variáveis mas que se ajuste bem segundo critérios utilizados.

Eliminação *Backward*: É uma abordagem de regressão passo a passo que começa com um modelo com todas as variáveis explicativas possíveis e a cada etapa verifica segundo critério se existem variáveis a serem eliminadas.

Seleção *Forward*: É uma abordagem de regressão passo a passo que começa com um modelo sem nenhuma variável explicativa, e em cada etapa inclui gradualmente as variáveis com o menor p-valor (inferior ao α fixado) para encontrar um modelo que melhor explica os dados. O processo de seleção para quando não há mais variáveis que atendam ao critério de inclusão no modelo, que consiste em um p-valor menor do que o α fixado.

Regressão *Stepwise*: É uma ferramenta usada nos estágios exploratórios da construção de modelos para identificar um subconjunto útil de preditores. O processo adiciona sistematicamente a variável mais significativa ou remove a variável menos significativa durante cada etapa. O modelo começa sem nenhuma variável explicativa e a cada passo quando as variáveis explicativas (fora do modelo) tem um p-valor menor do que o α fixado esta é incluída no modelo, ao mesmo tempo em que as variáveis que já estão no modelo são ponderadas de acordo com o seu p-valor em cada passo e se esta tiver um p-valor maior que o α fixado ela é retirada do modelo. A Regressão *Stepwise* é uma junção da Seleção de *Forward* e da Eliminação de *Backward*.

3.6 Diagnóstico do Modelo

A análise de resíduos para regressão logística é mais difícil do que para modelos de regressão linear porque as respostas Y_i assumem apenas os valores 0 e 1. Consequentemente, o i -ésimo valor residual, e_i assumirá um de dois valores:

$$e_i = \begin{cases} 1 - \hat{\pi}_i, & \text{se } Y_i = 1 \\ -\hat{\pi}_i, & \text{se } Y_i = 0 \end{cases}$$

Os resíduos não serão normalmente distribuídos e, de fato, sua distribuição sob a suposição de que o modelo ajustado está correto é desconhecida. Gráficos de resíduos contra valores ajustados ou variáveis preditoras geralmente não serão informativos.

3.6.1 Resíduos de Pearson

Os resíduos podem ser mais comparáveis dividindo-se pelo erro padrão estimado de Y_i , a saber, $\sqrt{\hat{\pi}_i(1 - \hat{\pi}_i)}$. Os Resíduos de Pearson resultantes são dados por:

$$r_{Pi} = \frac{Y_i - \hat{\pi}_i}{\sqrt{\hat{\pi}_i(1 - \hat{\pi}_i)}} = \sum \frac{(Y_i - \hat{\pi}_i)^2}{\hat{\pi}_i(1 - \hat{\pi}_i)}, \quad (3.6.1)$$

Os resíduos de Pearson estão diretamente relacionados à estatística de qualidade

de ajuste do qui-quadrado de Pearson.

3.6.2 Resíduos de Pearson Semistudentizados

Os resíduos de Pearson não possuem variância unitária, pois nenhuma previsão foi feita para a variação inerente no valor ajustado $\hat{\pi}_i$. Uma melhor forma de proceder, é dividir os resíduos pelo seu desvio padrão estimado. Esse valor é aproximado por $\sqrt{\hat{\pi}_i(1 - \hat{\pi}_i)(1 - h_{ii})}$, onde h_{ii} é o i -ésimo elemento da matriz diagonal $n \times n$ estimada para regressão logística:

$$H = \hat{W}^{1/2} X (X' \hat{W} X)^{-1} X' \hat{W}^{1/2}. \quad (3.6.2)$$

Assim, \hat{W} é a matriz $n \times n$ com elementos $\hat{\pi}_i(1 - \hat{\pi}_i)$, X é a matriz usual $n \times p$, $\hat{W}^{1/2}$ é uma matriz diagonal com elemento diagonal igual a raiz quadrada daqueles em \hat{W} . Os *resíduos de Pearson studentizados* resultantes são definidos como:

$$r_{SPi} = \frac{r_{Pi}}{\sqrt{1 - h_{ii}}}. \quad (3.6.3)$$

Lembre-se de que, para regressão linear múltipla, a matriz $\hat{Y} = HY$ satisfaz a matriz de expressão $Y = H$. A matriz para a regressão logística é desenvolvida de forma análoga; isso satisfaz aproximadamente a expressão $\hat{\pi}' = HY$, onde $\hat{\pi}'$ é o vetor de preditores lineares ($n \times 1$).

3.6.3 Gráficos de Resíduos:

Na regressão logística utiliza-se os gráficos de resíduos apenas para a detectar a adequação do ajuste do modelo, uma vez que variância não constante está sempre presente.

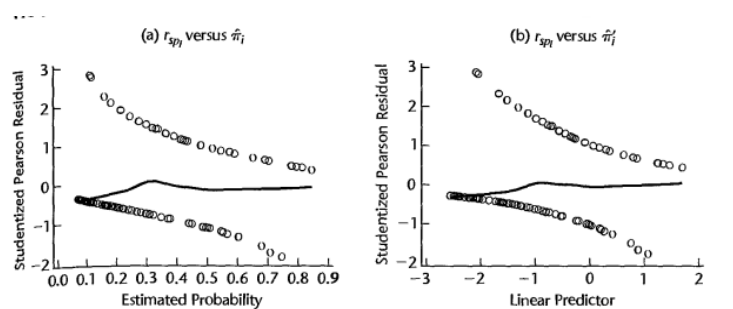


Figura 3: Exemplos de gráficos de resíduos.

Fonte: Neter et al. Applied Linear Statistical Models (p. 595).

Se o modelo de regressão logística estiver correto, uma suavização leve do gráfico dos resíduos em relação á probabilidade estimada ou deve resultar aproximadamente em uma linha horizontal com intercepto no zero. Qualquer desvio significativo desse comportamento sugere que o modelo pode ser inadequado.

3.7 Inferência para a resposta média

A resposta média de interesse para π_h

$$\pi_h = \left[1 + e^{-X'_h \beta} \right]^{-1}. \quad (3.7.1)$$

Onde o estimador é dado por:

$$\hat{\pi}_h = \left[1 + e^{-X'_h \hat{\beta}} \right]^{-1}, \quad (3.7.2)$$

onde \hat{b} é o vetor $[\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_{p-1}]'$ de coeficientes de regressão estimados e X é o vetor $[1, X_{h1}, X_{h2}, \dots, X_{h,p-1}]'$

Os limites de confiança de grandes amostras para a resposta média *logit* (Aproximado $(1 - \alpha)$) é dado da seguinte forma: então obtido da maneira usual:

$$\begin{aligned} L &= \hat{\pi}_h' - z \left(1 - \frac{\alpha}{2} s\{\hat{\pi}_h\} \right) \\ U &= \hat{\pi}_h' + z \left(1 - \frac{\alpha}{2} s\{\hat{\pi}_h\} \right), \end{aligned} \quad (3.7.3)$$

sendo que L e U são, respectivamente, os limites de confiança inferior e superior para n .

A matriz de covariância aproximada estimada dos coeficientes de regressão quando n é grande é dada por:

$$s^2 = \hat{\pi}_h' = X'_h s^2 \{\hat{\beta}\} X_h, \quad (3.7.4)$$

sendo X_h o vetor dos níveis das variáveis X para as quais n deve ser estimado por X_i :

3.8 Validação do Modelo

Refere-se a qualidade (estabilidade e razoabilidade) dos coeficientes de regressão; ao bom poder descritivo (plausível, útil) da parte funcional do modelo.

- Coleta de novos dados (dados independentes) para checar o modelo e seu poder preditivo;
- Comparar os resultados com valores teóricos esperados, resultados empíricos ante-

riores ou resultados simulados;

- Usar uma parte da amostra para checar o modelo e o poder preditivo do mesmo.

Muito frequentemente utiliza-se parte da amostra para fazer a validação do modelo dado a dificuldade de se obter uma nova amostra para este fim. Desta maneira, quando o conjunto de dados é suficientemente grande divide-se a amostra em dois conjuntos:

- Dados de construção do modelo (amostra de treinamento) – usado para desenvolver o modelo;
- Dados de validação ou predição – usado para avaliar a validade e habilidade preditiva do modelo

Assim, serão aplicados métodos para verificar a validade dos dados de validação, que seguirão as etapas:

- Reestimar o modelo escolhido anteriormente usando os novos dados.
- Calibrar a capacidade preditiva do modelo regressivo selecionado.

4 Metodologia

4.1 Dados

Com o intuito de entender como se dá a evasão nos cursos de Matemática (licenciatura diurna e noturna) da Universidade de Brasília, foi realizado um estudo com os dados advindos do Sistema de Informações Acadêmicas de Graduação (SIGRA) e do Sistema Integrado de Gestão de Atividades Acadêmicas (SIGAA), fornecidos pela universidade. A base de dados contém informações sociodemográficas e acadêmicas dos estudantes do Instituto de Ciências Exatas.

Os dados trazem a informação mascarada, portanto sendo impossível identificar o aluno a quem pertence tais dados, e incluem informações sobre seu estado de nascimento, raça/cor autodeclarada ao ingressar na UnB e se é ou não cotista e a qual tipo de cota pertence. Traz informações sobre o IRA do aluno, onde é possível identificar o seu índice de rendimento acadêmico atual ou de quando o aluno formou, ano que esse aluno concluiu o ensino médio, bem como se esse ensino se deu através de escolas públicas ou particulares. Há informação sobre o ingresso do aluno, período que ele ingressou, a forma se foi através de vestibular, PAS, transferências e as demais formas de ingresso possível. Assim como, informações de saída, ou seja, se o aluno saiu ou não do curso, e no caso dos que saíram de que forma saiu, formado, desligado, mudou de curso entre outras. Há também informações de histórico para cada aluno, sendo possível analisar as disciplinas que o aluno cursou, identificar a modalidade de cada uma sendo obrigatória, optativa, modo livre e etc., em que período isso ocorreu, o tanto de créditos de tal disciplina e a menção atribuída ao aluno. Além disso há também informações de quantos créditos o estudante ainda precisa cursar para formar, a sua média em cada semestre e o total de créditos aprovados.

A partir dessa base de dados foi realizado um estudo dos currículos dos cursos de Licenciatura em Matemática afim de entender o intervalo de período onde essas informações podem ser comparadas, sem prejuízo de interpretação. Em seguida sucedeu-se um trabalho de limpeza e preparação do banco, que contou com a identificação de dados duplicados para alguns estudantes, tornando assim necessário analisar caso a caso, visto que os dados são mascarados. Dados organizados, foi realizada a padronização das informações e criação de algumas variáveis, permitindo assim um estudo mais robusto e com maior informação.

O currículo estudado para o curso Licenciatura Diurna em Matemática foi o de código 1325/ ~ 4 disponível no SIGAA, que teve início no ano de 2014, para o curso da Licenciatura Noturna em Matemática o currículo é o de código 1368/ ~ 2, igualmente disponível no SIGAA e com início em 2014. Tendo em vista que no ano de 2020 enfrentou-

se a pandemia popularmente nomeada de COVID 19 que acarretou quarentenas extensas e conseqüentemente repercussão nas rotinas acadêmicas, o estudo leva em consideração os dados até o ano de 2019, onde eventos atípicos não haviam ocorrido.

4.2 Variáveis criadas

- **Idade de ingresso:**

A partir da variável data de nascimento foi possível determinar a idade em anos dos alunos, tendo em vista que o banco de dados dispõe da informação do período de ingresso de cada um. Sendo assim, foi considerado como data de referência o dia 01 de março para os ingressantes no primeiro semestre e 01 de agosto para os ingressantes do segundo semestre. Nos dois casos em que os alunos ingressaram no período de verão foi utilizada como data de referência 1 de janeiro.

- **Raça/cor:**

Tendo em vista que os estudantes podem ou não declarar a sua raça/cor, há uma porcentagem alta, chegando a mais de 40% de dados faltantes para um dos cursos. Além disso, há também raça/cor em que apenas um estudante se autodeclara, como é o caso indígena.

Portanto, com enfoque em análise foi criada uma nova variável Raçaccor que surge através do agrupamento da variável raça/cor inicial, classificada da seguinte forma:

Tabela 1: Agrupamento Raça/cor

| | |
|------------------------------|-------|
| Raça/cor | Negro |
| Branco | Não |
| Amarelo (de origem oriental) | |
| Preta | Sim |
| Pardo | |

- **Modo de Ingresso**

Dada as formas de ingresso dos alunos e a baixa frequência em algumas delas, que é o caso de transferências, dupla diplomação, dupla habilitação, mudança de curso, turno ou habilitação e portador de diploma de curso superior, os dados foram reagrupados da seguinte forma:

Tabela 2: Agrupamento por forma de ingresso

| Forma de ingresso | Classificação |
|---------------------------------------|-------------------------------|
| Programa de Avaliação Seriada | Programa de Avaliação Seriada |
| Sistema de seleção unificada | Sistema de seleção unificada |
| Vestibular | Vestibular |
| Dupla Diplomação | Outros |
| Dupla Habilitação | |
| Mudança de Curso | |
| Mudança de Turno | |
| Mudança de Habilitação | |
| Portador De Diploma De Curso Superior | |
| Transferência Facultativa | |
| Transferência Obrigatória | |

- **Reprovações**

Com o intuito de estudar a evasão de curso, os dados de menção foram agrupados em aprovação ou reprovação, para assim, ter uma melhor ideia de quais disciplinas resultam em um maior número de reprovações. Não ser aprovado em disciplinas pode acarretar em evasão do curso. Essa, pode ser dividida em duas formas, a primeira dada por reprovar uma disciplina obrigatória 3 vezes, e a segunda por exceder o tempo limite máximo de permanência do curso.

Tabela 3: Decisão de menção

| Menção | Decisão |
|--------|-----------|
| SR | Reprovado |
| II | |
| MI | |
| MM | Aprovado |
| MS | |
| SS | |

Dada a preocupação em analisar o andamento acadêmico do estudante e a permanência na universidade, foi criada a variável reprovações que é a quantidade de reprovação do aluno. Importante salientar que o aluno não é medido pela quantidade de aprovação ou não aprovação, porém essas quantidades podem auxiliar em diagnósticos até mesmo para o próprio estudante, além da universidade.

- **Números de semestres cursados**

É dada pela quantidade máxima de semestres que o aluno cursou, para os ativos foi considerado a quantidade de semestres cursados até o segundo semestre do ano de 2019, para os formados a quantidade de semestres que levou até formar e para os

que evadiram o curso a quantidade de semestres quando tal evento ocorreu.

- **Trancou**

Há duas formas de trancamento na Universidade de Brasília, um é o **TR - Trancamento** que é um recurso do qual os alunos podem recorrer, pode ser feito até a metade do semestre, antigamente o aluno perdia nota em seu índice de rendimento acadêmico, porém atualmente isso não afeta mais nota, apenas consta em seu histórico.

A outra forma de trancar uma disciplina é o **TJ - Trancamento justificado** que é quando o aluno tranca a disciplina mas tem justificativa comprobatória e aceita pela universidade, como por exemplo afastamento médico.

A variável Trancou, no presente estudo, leva em consideração apenas a primeira forma de trancamento, que é o não justificado, ou seja é puramente uma decisão do estudante a não continuidade na disciplina. Esta é uma variável dicotômica, que retorna o valor "Sim" para os estudantes que já trancou alguma disciplina durante o curso e "Não" para os que nunca utilizaram dessa saída.

- **Número de Trancamentos**

Assim como analisado se o estudante já trancou alguma disciplina, é de interesse observar quantas vezes isso ocorreu para cada aluno. Ou seja, a variável número de trancamento retorna a quantidade de vezes que o aluno trancou de forma não justificada uma disciplina ao longo da graduação.

- **SR**

A menção SR (Sem rendimento) na Universidade de Brasília é atribuída ao estudante que teve nota zero no resultado de uma disciplina. Dado que é a menor menção possível e que em muitos casos é resultado de um abandono por parte do aluno, visa-se estudar com qual frequência isso ocorre.

- **Número SR**

Assim como para a variável trancou, há também uma variável correspondente se o aluno recebeu alguma menção SR, e a quantidade de vezes que isso ocorreu.

- **Verão**

A Universidade de Brasília dispõe além dos dois semestres regulares um semestre por ano que é denominado como verão. Esse semestre por sua vez é mais curto do que os outros, ocorre sempre no começo do ano, entre janeiro e fevereiro e em muitos casos é uma boa opção para alguns alunos conseguirem a aprovação em alguma matéria mais complicada ou que já tenha reprovado. É uma variável dicotômica,

onde resulta "Sim" nos casos onde o aluno já cursou alguma matéria no verão, e "Não" caso contrário.

- **Número Verão**

A partir do período em que a disciplina foi cursada, é possível saber se ela foi no verão ou não. Diante disto, contou-se quantas vezes o estudante cursou matérias ofertadas no verão.

- **Taxa de Reprovação**

A variável taxa de reprovação foi criada visando dividir em intervalos os níveis de matérias reprovadas pelos alunos, sendo assim possível fazer análises agrupadas. O cálculo para a taxa é dado por:

$$Taxa\ de\ reprova\c{c}\tilde{a}o_i = \frac{\sum\ disciplina\ com\ reprova\c{c}\tilde{a}o\ pelo\ aluno_i}{\sum\ todas\ as\ disciplinas\ cursadas\ pelo\ aluno_i}$$

E agrupadas da seguinte forma:

Tabela 4: Agrupamento por taxa de reprovação

| Taxa de Reprovação | Classificação |
|--------------------|---------------|
| 0 – 0,25 | Baixo |
| 0,25 – 0,50 | Médio - baixo |
| 0,50 – 0,75 | Médio - alto |
| 0,75 – 1 | Alto |

- **Reprovações nas disciplinas mais reprovadas**

Ao analisar as reprovações que mais ocorrem no curso de Licenciatura em Matemática da Universidade de Brasília, foi visto que as disciplinas mais reprovadas são as disciplinas de começo de curso. Além disso, muitas vezes essas disciplinas são não só as mais reprovadas, mas também as que ocorrem com maior frequência as menores menções. Dado as análises descritivas e a importância desse início da vida acadêmica dos estudantes, resolveu-se aprofundar a análises para tais disciplinas:

Tabela 5: Disciplinas com os maiores índices de reprovações

| Disciplina | Semestre do currículo |
|-------------------------------------|-----------------------|
| Cálculo 1 | 1º |
| Introdução à Ciência da Computação | 1º |
| Geometria Analítica para matemática | 1º |
| Organização da Educação Brasileira | 1º |
| Cálculo 2 | 2º |
| Física | 2º |
| Introdução à Álgebra Linear | 2º |

Para cada disciplina foi criado duas variáveis que diz respeito a reprovação, são elas:

- **Reprovou a disciplina**

Resulta sim para quem já reprovou e não caso contrário.

- Quantidade de Reprovações na disciplina

Resulta a quantidade de vezes que o aluno reprovou a disciplina avaliada.

4.3 Método

Inicialmente foi realizado um estudo de análise descritiva para os dados disponibilizados, visando identificar as principais características dos estudantes de Licenciatura em Matemática da Universidade de Brasília. Posteriormente, utilizou-se de técnicas descritivas e inferenciais, afim de verificar a associação da variável resposta "Evasão" e as potenciais variáveis explicativas. Por fim, dado os passos anteriores construídos e analisados, foi construída a modelagem fazendo uso da Regressão Logística.

A Regressão Logística é um modelo estatístico usado para produzir, com base em um conjunto de dados, um modelo que possibilite prever valores assumidos por uma variável categórica (binária ou não) denominada variável resposta ou dependente, com base em uma série de variáveis explicativas quantitativas e/ou categóricas, que são atributos preditores.

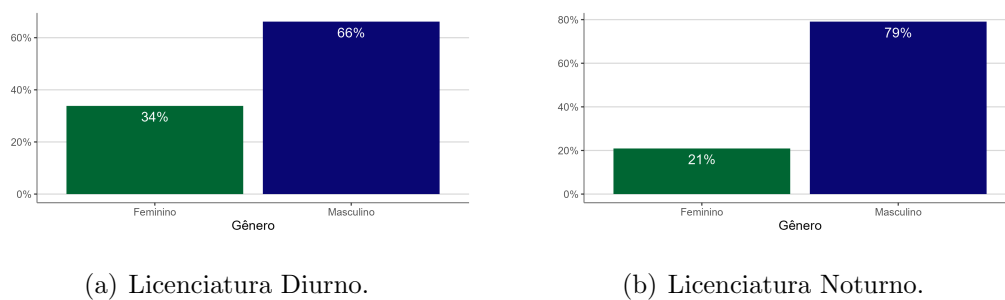
Na limpeza, preparação e análise de dados foi utilizado o software R e eventualmente o software SAS.

5 Resultado

5.1 Dados Pessoais

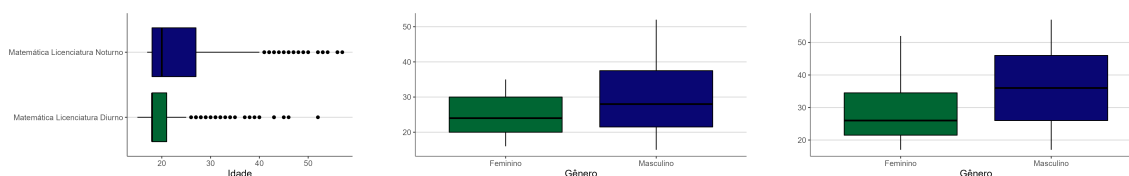
Para melhor estudar a evasão é importante entender tanto as características sociodemográficas dos estudantes, como também sua vida acadêmica.

Figura 4: Distribuição dos Alunos por gênero, Licenciatura em Matemática - UnB, 2014-2019



Ao analisar a quantidade de alunos do curso de Licenciatura em Matemática diurno e noturno da Universidade de Brasília, é possível observar que o curso noturno tem uma quantidade maior de estudantes sendo um total de 373 enquanto o curso diurno tem um número igual a 337 alunos. Em ambos os casos há mais pessoas do gênero masculino, sendo que no curso noturno a taxa de homens é de 79% e no diurno essa taxa é de 66,1%.

Figura 5: Idade dos alunos, Licenciatura em Matemática - UnB, 2014-2019



(a) Licenciatura em Matemática. (b) Licenciatura Diurna. (c) Licenciatura Noturno.

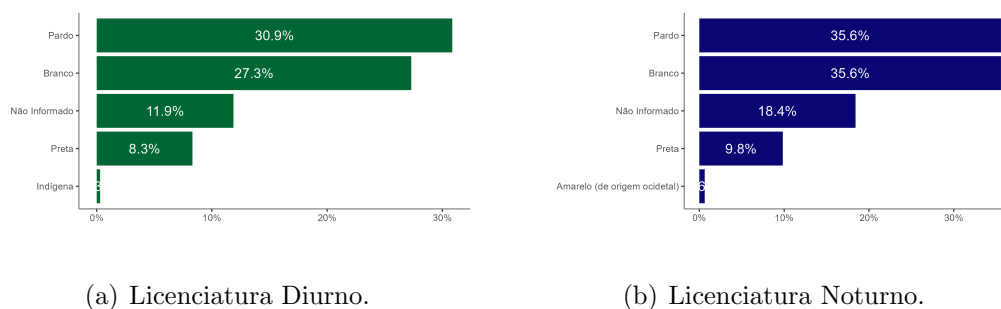
Ao analisar a idade de ingresso dos alunos do curso de Licenciatura em Matemática, pode-se observar que para o curso diurno as idades estão dispostas entre 15 e 52 anos, sendo ambas as idades do gênero masculino, as idades relativas ao gênero feminino estão dispostas entre 16 e 35 anos. A média de idade feminina é de 19,55, a masculina

21.18 e a média geral é de 20.63 anos. Nos 3 casos a mediana é de 18 anos, ou seja, metade dos alunos, nos 3 casos, têm menos de 18 anos de idade.

Enquanto isso, o cenário do curso noturno é idades entre 17 e 57 anos, sendo que a idade mínima de ambos os sexos é 17 anos, o máximo do sexo feminino é 52 anos e o do sexo masculino é de 57 anos. A média de idade feminina é de 21.55, a masculina 24.8 e a média geral é de 24.12 anos. Além disso, metade das mulheres têm até 18 anos, enquanto a metade dos homens têm até 21 anos de idade, a mediana geral é de 20 anos.

Diante disso pode-se notar que no geral os estudantes do curso noturno são mais velhos do que os alunos do curso diurno.

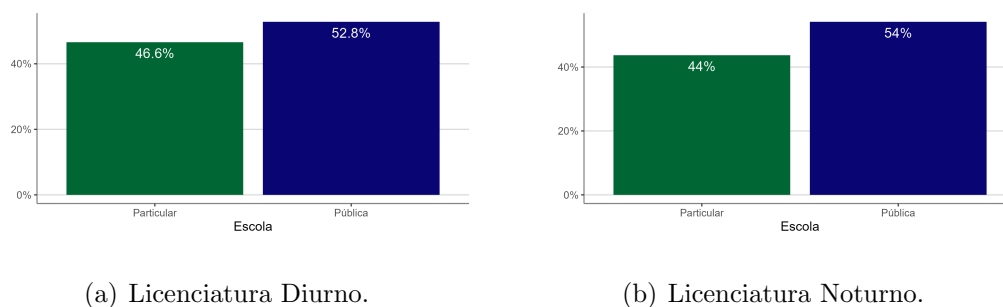
Figura 6: Raça/cor autodeclarada dos alunos, Licenciatura em Matemática - UnB, 2014 - 2019



Ao analisar o gráfico 6(a) que diz respeito a raça/cor autodeclarada dos alunos do curso Licenciatura Diurno em Matemática, é possível observar que a maior parte deles se autodeclararam Pardos ou Brancos, em terceiro lugar têm se o valor da quantidade de alunos que não quiseram informar ou se autodeclarar no que tange a sua raça/cor. Somado o número dos alunos que se autodeclararam Pretos ou Indígenas têm-se um valor inferior a dez por cento do total de alunos, sendo que nenhum aluno se autodeclarou Amarelo (de origem oriental).

Ao observar o gráfico 6(b) a raça/cor dos alunos da Licenciatura Noturno em Matemática, é possível observar que a maior parte dos alunos se autodeclararam Pardos e brancos (sendo inclusive o mesmo número de 112 alunos), em terceiro lugar têm se a quantidade de alunos que não quiseram informar ou se autodeclarar no que tange a sua raça/cor. Somado o número dos alunos que se autodeclararam Pretos ou Amarelos (de origem oriental) têm-se um valor inferior a dez por cento do total de alunos.

Figura 7: Tipo de escola do 2º grau dos alunos, Licenciatura em Matemática - UnB, 2014 - 2019



(a) Licenciatura Diurno.

(b) Licenciatura Noturno.

Dos 337 alunos matriculados no curso de Licenciatura Diurno em Matemática, 52.8% destes advém de escolas de rede pública, não há informação de tipo de escola do 2º grau para 0.6% alunos.

Para o curso Licenciatura Noturno em Matemática, dos 373 alunos matriculados no curso, 54% dos alunos advém de escolas de rede pública, 44% de escolas particulares e 2% dos alunos não informaram o seu tipo de escola do 2º grau.

Em ambos os casos, a maioria dos estudantes cursaram o 2º grau em escolas de rede pública, não sendo uma larga diferença percentual.

Tabela 6: Distribuição dos alunos cotistas por tipo de cotas.

| Cota | Frequência | |
|---------------------------------|------------|---------|
| | Diurno | Noturno |
| Escola Púb., Alta Renda - PPI | 8.9% | 5.3% |
| Escola Púb., Baixa Renda - PPI | 5.3% | 3.5% |
| Escola Púb. | 2.9% | 3.48% |
| Escola Púb., Baixa Renda. | 2.6% | 1.3% |
| Negro | 2.1% | 1.8% |
| Escola Púb., Deficientes - PPI. | 0.2% | 0.3% |

Ao analisar a quantidade de alunos de Licenciatura Diurno em Matemática segundo cotas, pode-se observar que 43,32% dos estudantes são cotistas. Do total de alunos que possuem algum tipo de cota 20,54% deles estão na classe de cotas de escola pública que possuem alta renda e se declaram pretos, pardos ou indígenas, além disso 5% dos ingressantes estão na classe dos cotistas para negros. 56,6% dos estudantes não são cotistas. Sabe-se também que 146 estudantes dos 337 possuem algum tipo de cota, mas não há informação de qual tipo de cota para 71 deles.

Ao analisar a quantidade de alunos de Licenciatura Noturno em Matemática segundo cotas, pode-se observar que 36,7% dos alunos são cotistas. Do total de alunos que possuem algum tipo de cota 14,6% dos ingressantes estão na classe de cotas de escola pública que possuem alta renda e se declaram pretos, pardos ou indígenas, além disso 5,1% dos ingressantes cotistas pertencem a classe de cotas para negros. Além disso não

há informação do tipo de cota para 78 dos 137 alunos cotistas.

5.2 Vida Acadêmica

Tabela 7: Período de ingresso dos alunos, Licenciatura em Matemática - UnB, 2014-2019

| Período de Ingresso | Frequência | |
|---------------------|------------|---------|
| | Diurno | Noturno |
| 2014/1 | 33 | 34 |
| 2014/2 | 25 | 31 |
| 2015/0 | 0 | 1 |
| 2015/1 | 29 | 29 |
| 2015/2 | 20 | 33 |
| 2016/0 | 1 | 0 |
| 2016/1 | 29 | 28 |
| 2016/2 | 25 | 28 |
| 2017/1 | 25 | 37 |
| 2017/2 | 29 | 27 |
| 2018/1 | 31 | 31 |
| 2018/2 | 27 | 32 |
| 2019/1 | 28 | 25 |
| 2019/2 | 35 | 37 |

É possível observar a partir da Tabela 7 a frequência de alunos ingressantes no curso de Licenciatura em Matemática diurno e noturno, entre os anos de 2014 e 2019. O número de ingressantes fica em torno de 20 a 37 alunos por semestre, com exceção de 2015/0 e 2016/0, onde houve um evento atípico que o estudante tem data de início no verão.

Tabela 8: Evasão por semestre de ingresso, Licenciatura em Matemática - UnB, 2014-2019

| Período de Ingresso | Evasão | |
|---------------------|--------|---------|
| | Diurno | Noturno |
| 2014/1 | 81.8% | 88.2% |
| 2014/2 | 80% | 90% |
| 2015/0 | | 100% |
| 2015/1 | 65.5% | 82.8% |
| 2015/2 | 75% | 75.8% |
| 2016/0 | 0% | |
| 2016/1 | 65.5% | 64.3% |
| 2016/2 | 60% | 64.3% |
| 2017/1 | 56% | 67.6% |
| 2017/2 | 41,4% | 40.7% |
| 2018/1 | 65,5% | 41.9% |
| 2018/2 | 38,7% | 28.1% |
| 2019/1 | 3,6% | 4% |
| 2019/2 | 0% | 2,7% |

Ao observar os dados de evasão por semestre nos cursos de Licenciatura em Matemática, diurno e noturno, pode-se perceber que para os ingressantes do ano de 2014 pelo menos 80% dos alunos evadiram o curso, sendo que no curso noturno essa taxa é maior, chegando a 90% dos estudantes.

Considerando o intervalo entre os anos de 2014 até o primeiro semestre do ano de 2017 é notável que mais da metade dos ingressantes evadiram o curso.

Para o segundo semestre de 2017 até o ano de 2018 pelo menos 28% dos alunos já haviam evadido o curso.

Em todos os casos essas taxas são exorbitantes, visto que em nenhum dos casos considerados o aluno teria tempo suficiente para ser desligado por ter excedido o tempo limite de curso (que seria de 8 anos a contar o tempo de ingresso).

Para além do mais, ao considerar a taxa dos ingressantes no ano de 2018 há uma única forma de ser desligado por não cumprir condição obrigatória, que é dada por não ser aprovado em pelo menos quatro disciplinas em dois semestres consecutivos. Sendo assim, é esperado que esse percentual seja mais baixo quanto mais próximo for o semestre de ingresso considerado. Entretanto, isso não significa obrigatoriamente que haverá menos evasões para esses semestres.

Tabela 9: Forma de ingresso dos alunos, Licenciatura em Matemática - UnB, 2014 - 2019

| Forma de Ingresso | Frequência | |
|---------------------------------------|------------|---------|
| | Diurno | Noturno |
| Programa de Avaliação Seriada | 38% | 26.7% |
| Vestibular | 30.5% | 33.8% |
| SISU-Sistema de Seleção Unificada | 21.3% | 22% |
| Portador de Diploma de Curso Superior | 3.2% | 5.9% |
| Mudança de Curso | 2% | 2.1% |
| Mudança de Habilitação | 1.2% | 0.2% |
| Transferência Obrigatória | 0.9% | 5.9% |
| Mudança de Turno | 0.9% | 0% |
| Transferência Facultativa | 0.9% | 0% |
| ENEM - UnB | 0.3% | 0.2% |
| Dupla Diplomação | 0.3% | 0% |
| Dupla Habilitação | 0.3% | 0% |

Diante as formas de ingresso, é possível observar que a maioria dos estudantes ingressam pelo PAS (Programa de Avaliação Seriada), sendo no total 38% dos 337 alunos matriculados no curso Licenciatura Diurno em Matemática e 26.7% dos 373 ingressantes do curso Licenciatura Noturno em Matemática, o Vestibular é responsável pela forma de ingresso de 30.6% dos estudantes do turno diurno e de 33.8% dos 373 alunos do noturno, seguida do SISU que é a forma de ingresso de 21.3% dos estudantes da Licenciatura Diurno de Matemática enquanto na Licenciatura Noturno em Matemática esse valor é de 22%. Somada as demais formas de ingresso têm-se os aproximadamente 10% dos estudantes restantes do turno diurno e 17% dos alunos do noturno.

Foi possível observar que para ambos os cursos há uma semelhança na distribuição da forma de entrada, isso pode ser explicado pelo fato de que as vagas são pré disponibilizadas pela universidade para cada tipo de forma de ingresso, não sendo então possível algo diferente do esperado/estipulado.

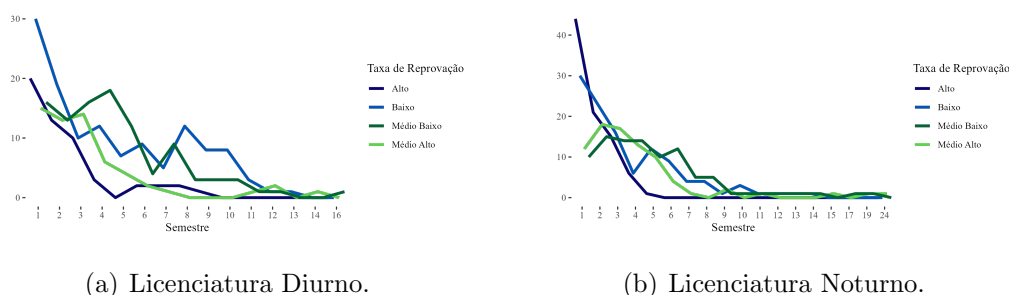
A taxa de reprovação geral dos alunos de Matemática Licenciatura Diurno é de

Tabela 10: Taxa de reprovação geral dos alunos, Licenciatura em Matemática - UnB, 2014 - 2019

| Decisão | Frequência | |
|------------|------------|---------|
| | Diurno | Noturno |
| Aprovação | 66.3% | 58.9% |
| Reprovação | 33.7% | 41.4% |

33.7%. A taxa de reprovação geral dos alunos de Matemática Licenciatura Noturno é de 41.4% dentre disciplinas obrigatórias e não obrigatórias. Observa-se que no geral os alunos do diurno tem maior porcentagem de aprovação.

Figura 8: Taxa de reprovação por número de semestres cursados, Licenciatura em Matemática - UnB, 2014 - 2019



Ao analisar a taxa de reprovação dos alunos de Licenciatura Diurno em Matemática, foi considerado como Baixo o aluno que tem até 25% de reprovações considerando o total de matérias cursadas pelo mesmo, médio baixo o aluno que tem até 50% de reprovações, médio alto os que tem até 75% e alto os alunos que tem mais de 75% de reprovações.

No gráfico acima, os semestres (eixo x) diz respeito ao número de semestres cursado pelo aluno. Podemos notar, conforme Figura 8(a), que o curso de Matemática Licenciatura Diurno no primeiro semestre possui um quantitativo maior de alunos que tem uma taxa de reprovação considerada baixa, ou seja passam em todas ou na maioria das matérias cursadas. Isso também ocorre com o oposto, um quantitativo inferior mas ainda alto de alunos que reprovam a maioria ou todas as matérias. Entretanto, os intervalos denominados média baixa e média alta ficam abaixo dos outros dois intervalos anteriormente mencionados. Logo, no geral tendemos a acreditar que os alunos tendem a ir muito bem ou muito mal no primeiro semestre do curso.

No segundo semestre há uma baixa considerável no quantitativo de alunos com uma taxa alta de reprovações, enquanto isso os que tem uma taxa baixa continua sendo superior e os intervalos de média estão bem próximos, chegando a se cruzarem.

Do terceiro semestre até o sexto a taxa com maior frequência de alunos é a médio baixo, sendo que do sexto em diante as taxas médio baixo e baixo intercalam entre si,

enquanto a alto e médio alto se aproximam, ou até mesmo, atingem a frequência zero!

Logo, nos primeiros semestres é alta o número de aprovações em todas as matérias por um mesmo aluno, de forma semelhante que é alta a taxa de reprovação em todas ou boa parte das matérias por um único aluno também. No que pode ser considerado como meio do curso essa taxa fica mais evidente entre os intervalos médio baixo e baixo, ou seja, as pessoas mais passam do que reprovam suas matérias, assim como no fim do curso.

No curso de Licenciatura Noturno de Matemática, Figura 8(b) é possível observar que no primeiro semestre os alunos têm em grande maioria uma taxa alta de reprovações, sendo que a segunda maior frequência são alunos que têm uma taxa baixa de reprovações.

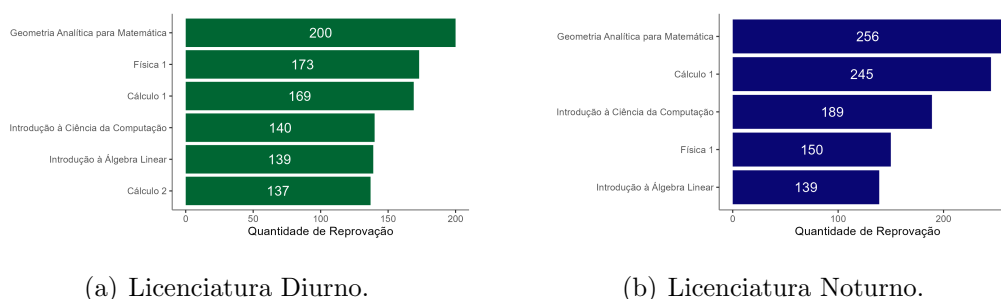
Isso indica que no primeiro, comportamento que se repetiu no segundo e terceiro semestre, os alunos ou são aprovados na maioria das matérias que pegam por semestre, ou reprovam a maior parte das disciplinas que pegam no semestre.

No terceiro semestre todas as curvas de taxa de reprovação se cruzam, e posteriormente há prevalência das taxas médias e baixa.

Além disso, houve um caso curioso, onde um aluno consta como tendo cursado 24 semestres. Isso ocorre porque a vida acadêmica desse aluno foi descontinuada em alguns momentos.

É possível perceber similaridade entre as taxas de reprovação dos alunos pelo número de semestres cursado. Em ambos os cursos considerados Licenciatura Diurno e Noturno os alunos tendem a ter altas ou baixas taxas de reprovações no começo do curso, as taxas média alta e média baixa prevalecem no meio do curso, sendo que nos semestres finais as taxas média baixa e baixa sobressaem novamente.

Figura 9: Disciplinas mais reprovadas pelos alunos, Licenciatura em Matemática - UnB, 2014 - 2019



(a) Licenciatura Diurna.

(b) Licenciatura Noturna.

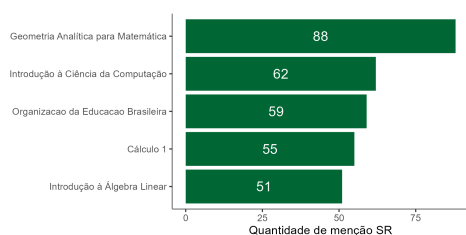
É possível observar diante a Figura 9(a), as seis matérias com maior número de reprovações pelos alunos da Matemática Licenciatura Diurna que é dada pelas disciplinas: Geometria Analítica para Matemática com 200 reprovações, seguida de cálculo 1 e 2, física 2, Introdução a Álgebra Linear e Introdução a Ciência da Computação, todas com

o número de reprovações entre 137 e 173.

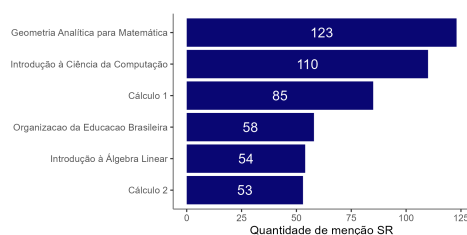
De maneira semelhante, é possível observar por meio da Figura 9(b), as cinco matérias com maior número de reprovações pelos alunos da Matemática Licenciatura Noturno que são: Geometria Analítica para Matemática com mais de 250 reprovações, seguida e Cálculo 1 com 245 reprovações, Introdução a Ciência da Computação, Física 1 e Introdução a Álgebra Linear.

Mais uma vez fica clara a semelhança entre o que ocorre nos dois cursos, uma vez que as disciplinas mais reprovadas no diurno são também em maioria as mais reprovadas no noturno.

Figura 10: Disciplinas com maior atribuição de menção SR, Licenciatura em Matemática - UnB, 2014 - 2019



(a) Licenciatura Diurna.

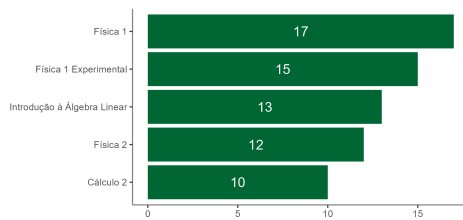


(b) Licenciatura Noturna.

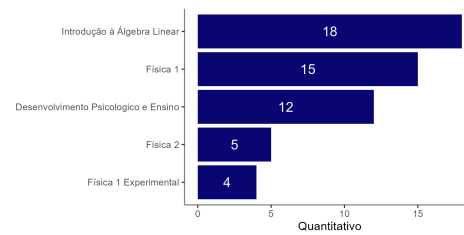
No caso do curso de Licenciatura Diurna, Figura 10(a), com exceção da disciplina Organização da Educação Brasileira todas as outras disciplinas com mais menções SR se repetem as disciplinas mais reprovadas, incluindo Geometria Analítica para Matemática que além de ter o maior número de reprovações também tem o maior número de menções SR.

No curso de Matemática Licenciatura Noturna, Figura 10(b) as disciplinas com mais menções SR são Geometria Analítica para Matemática, seguida de Introdução à Ciência da Computação, ambas com frequência superior a 100. Também é frequente a menção SR nas disciplinas Cálculo 1 e 2, Introdução a Álgebra Linear e Organização da Educação Brasileira.

Figura 11: Disciplinas mais cursadas no período verão, Licenciatura em Matemática - UnB, 2014 - 2019



(a) Licenciatura Diurna.

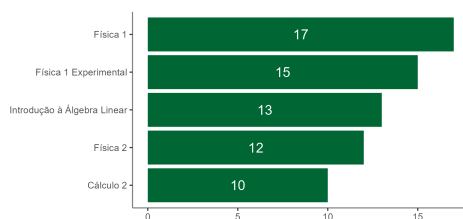


(b) Licenciatura Noturna.

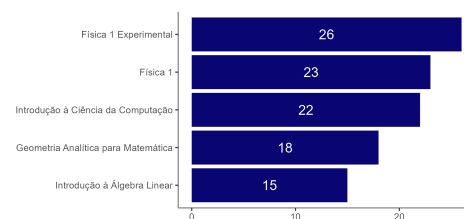
A disciplina mais cursada no verão pelos alunos de licenciatura diurna, Figura 11(a), é Física 1, seguida de Desenvolvimento Psicológico e Ensino. Além disso as disciplinas Introdução a Álgebra Linear e Física 2 estão presentes tanto na lista de disciplinas mais cursadas no verão como também nas disciplinas mais reprovadas. É importante ressaltar que nem todas as disciplinas do curso podem ser ofertadas no período de verão.

A disciplina mais cursada no verão pelos alunos de licenciatura noturna, Figura 11(b), é Introdução à Álgebra Linear, seguido de Física 1, Desenvolvimento Psicológico e Ensino, Física 2 e Física 1 Experimental.

Figura 12: Disciplinas mais trancadas, Licenciatura em Matemática - UnB, 2014 - 2019



(a) Licenciatura Diurna.



(b) Licenciatura Noturna.

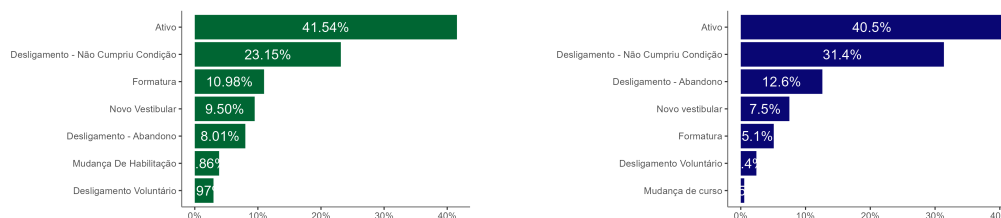
As disciplinas mais trancadas pelos alunos da Licenciatura Diurna em Matemática são Física 1, Física 1 experimental, IAL, Física 2 e Cálculo 2.

Em Licenciatura Noturna em Matemática essas disciplinas são dadas por Física 1 experimental, Física 1, ICC, Geometria Analítica para Matemática e IAL.

É possível notar que em suma os estudantes do noturno utilizam-se mais do recurso de trancar disciplinas. Além disso, que em ambos os casos a 2 matérias mais trancadas coincidem. No curso Licenciatura Noturna em Matemática, os alunos apresentam problemas com a disciplina Introdução a Ciência da Computação, tal problemática não é

tal perceptível no curso de Licenciatura Diurno em Matemática.

Figura 13: Forma de saída do curso, Licenciatura em Matemática - UnB, 2014 - 2019



(a) Licenciatura Diurno.

(b) Licenciatura Noturno.

Pode-se observar que a forma de saída do curso de Licenciatura Diurno em Matemática com maior frequência é o desligamento por não cumprir condição, seguida por formatura e novo vestibular. Sendo que o número de estudantes desligados é 1,75 vezes maior do que o número de estudantes que se formam, quando considerado os estudantes ingressantes entre os anos de 2014 e 2019.

Ao observar o que ocorre no curso de período noturno, pode-se observar que a maior porcentagem dos alunos também evadem o curso por desligamento causado por não cumprir condição. Esses valores em proporção são levemente superior para o curso noturno, sendo que o número de estudantes desligados é superior a 6 vezes o número de estudantes que se formam, quando considerado os estudantes ingressantes do noturno entre os anos de 2014 e 2019.

5.3 Licenciatura Diurna em Matemática

5.3.1 Evasão

Considerando os estudantes ingressantes entre 2014 e 2019, pode-se observar uma taxa de evasão igual a 47,48%, ou seja, praticamente metade dos alunos ingressantes no curso Matemática Licenciatura Diurna não se formam no curso.

Tabela 11: Análise bivariada por evasão, Licenciatura Diurna em Matemática - UnB, 2014-2019.

| Variável | | Evasão | | Estatística | P-valor |
|----------------------------|-------------------------------|--------|-------|-------------|---------------|
| | | Sim | Não | | |
| Gênero | Feminino | 41.2% | 58.8% | 2.333 | 0.1267 |
| | Masculino | 50.7% | 49.3% | | |
| Raça/cor | Negro | 37.1% | 62.9% | 11.282 | 0.0035 |
| | Não Negro | 49.1% | 51% | | |
| Tipo de Escola | Particular | 54.1% | 45.9% | 5.255 | 0.0218 |
| | Pública | 41% | 59% | | |
| Cota | Sim | 40.4% | 59.6% | 4.671 | 0.0306 |
| | Não | 52.9% | 47.1% | | |
| Forma de Ingresso | Programa de Avaliação Seriada | 38.1% | 61.7% | 16.081 | 0.001 |
| | Sistema de Seleção Unificada | 61.6% | 38.4% | | |
| | Vestibular | 54.4% | 45.4% | | |
| | Outras | 30.3% | 69.7% | | |
| Quantidade de Reprovação | Média | 5.21 | 5.44 | -0.3857 | 0.7 |
| Idade de ingresso | Média | 21.7% | 20% | -2.41 | 0.016 |
| IRA | Média | 1.79% | 3.37% | -10.349 | ≤ 0.0001 |
| Reprovação Cálculo 1 | Sim | 47.1% | 52.9% | 0 | 1 |
| | Não | 47.6% | 52.4% | | |
| Reprovação Cálculo 2 | Sim | 35.8% | 64.5% | 0.1539 | 0.6948 |
| | Não | 50.4% | 49.6% | | |
| Reprovação Física | Sim | 46.5% | 53.5% | 0.117 | 0.7323 |
| | Não | 47.7% | 52.3% | | |
| Reprovação Geom. Analítica | Sim | 53.8% | 46.2% | 1.3047 | 0.2533 |
| | Não | 46% | 54% | | |
| Reprovação IAL | Sim | 51.6% | 48.4% | 3.4867 | 0.06186 |
| | Não | 46.5% | 53.5% | | |
| Reprovação ICC | Sim | 50% | 50% | 0.0543 | 0.8157 |
| | Não | 47.3% | 52.7% | | |
| Reprovação Org. Edu. Bras. | Sim | 72.7% | 27.3% | 7.0373 | 0.00798 |
| | Não | 45.7% | 54.3% | | |
| Trancamento | Sim | 74.2% | 25.8% | 33.083 | ≤ 0.0001 |
| | Não | 37.9% | 62.1% | | |
| Verão | Sim | 22.8% | 77.2% | 23.955 | ≤ 0.0001 |
| | Não | 55% | 45% | | |
| SR | Sim | 58% | 42% | 19.195 | ≤ 0.0001 |
| | Não | 33.3% | 66.7% | | |

Ao analisar os alunos de Licenciatura Diurna em Matemática segundo a evasão, é possível notar que o gênero feminino tem maior percentual de não evadidos, quando comparado ao gênero masculino. Entretanto, em ambos os casos a diferença entre os que evadiram e os que não evadiram é menor do que 10%, ou seja, a princípio não consideramos essa uma grande diferença. Ao realizar teste qui quadrado para verificar se há associação entre gênero e evasão, considerando um nível de significância de 5%, descartamos a hipótese de associação. Portanto, acredita-se em primeiro momento que ser do

gênero feminino ou masculino não irá interferir na evasão ou não evasão do curso.

Com relação a raça/cor, pode-se observar que dentre os autodeclarados Negros 62,9% deles evadiram o curso de Licenciatura Diurna em Matemática, enquanto para os autodeclarados Não Negros essa porcentagem é de 51%. Logo, em primeiro momento é clara a diferença entre os que evadiram e não evadiram segundo a sua raça/cor, sendo que para os Negros em sua maioria (uma diferença de mais de 25%) evadiram o curso, enquanto para os não negros essa diferença foi de apenas 2%, ou seja entre os não negros a evasão foi balanceada. Ao testar associação entre evasão e a raça/cor agrupada dos estudantes em questão, podemos a um nível de significância de 5% acreditar que a Raça/cor agrupada tem associação com o fato do estudante evadir ou não o curso.

No que tange ao tipo de escola a qual o aluno cursou o 2º grau de ensino, pode-se observar que entre os alunos que advém da escola de rede particular 54,1% deles evadiram o curso, enquanto, para os alunos da rede pública esse valor é de 41%. Em ambos os casos, a diferença entre os que evadiram ou não evadiram o curso é de menos de 10%. Com o intuito de analisar se há associação entre a evasão e o tipo de escola do 2º grau do estudante, com um grau de significância de 5%, obtivemos um p-valor superior ao valor estipulado ao de significância, logo, acredita-se que não há associação entre esses dois fatores.

Dentre os alunos cotistas e não cotistas, segundo a evasão, observa-se que a maior parte dos cotistas (59,6%) não evadiram o curso, enquanto para os não cotistas a porcentagem dos que evadiram o curso foi de 52,9%. Em ambos os casos a diferença entre os que evadiram e não evadiram não chega a 10%. Ao analisar se há associação entre ser cotista e evadir o curso de Licenciatura Diurna da Matemática, com um nível de significância de 5% há razões para acreditar que estes dois fatores não estão associados.

A forma de ingresso dos alunos foi agrupada em PAS, SISU, Vestibular e Outras, para fim de análise. É possível observar que a maioria dos estudantes que ingressaram pelo PAS e Outras não evadiram do curso, enquanto para o SISU e Vestibular a maior porcentagem destes evadiram. Com exceção da forma de ingresso sendo o Vestibular, que teve uma diferença entre evasão e não evasão de aproximadamente 10%, todas as outras essa diferença entre evadir ou não foi de pelo menos 23%. Ao analisar se há associação entre a forma de ingresso e evasão do curso, com um nível de significância de 5%, acredita-se na associação entre os dois fatores.

A Taxa de Reprovação é uma variável que classifica entre baixo, médio baixo, médio alto e alto as taxas de reprovação de cada aluno, sendo essas taxas dispostas respectivamente da seguinte forma: 0-0,25, 0,25-0,50, 0,50-0,75 e 0,75-1. Conforme fosse de se esperar, a evasão para a taxa de reprovação baixa teve o menor valor entre as quatro, sendo 30% dos que tinham uma taxa baixa de reprovação evadiram o curso, essa taxa de

evasão foi subindo de acordo com a taxa de reprovação, sendo que dentre os alunos com uma taxa alta de reprovação a quantidade de evasão foi de 80%. Ao analisar a associação entre evasão e taxa de reprovação, com um nível de significância de 5%, acredita-se na associação desses dois fatores.

Foram analisadas também as variáveis Trancamento, Verão e SR que indicam respectivamente se o aluno trancou alguma matéria sim ou não, pegou alguma matéria de verão sim ou não e se já ficou com menção SR em alguma matéria sim ou não, para os alunos que já trancou matéria e, ou tirou SR a maior parte dos estudantes evadiram, já para os que fizeram matéria de verão a maior parte não evadiu. Ao testar associação entre evasão e cada uma das três variáveis pode-se observar que, há um nível de 5% de confiança, há indicação de associação nos três casos.

Foi criado uma variável Reprovação para cada uma das 7 matérias mais reprovadas pelos alunos da Licenciatura Diurna de Matemática, sendo elas: Cálculo 1, Cálculo 2, Física, Geometria Analítica, IAL (Introdução à Álgebra Linear), ICC (Introdução à Ciência da Computação) e Organização da Educação Brasileira. Para as disciplinas Geometria Analítica, IAL (Introdução à Álgebra Linear), ICC (Introdução à Ciência da Computação) e Organização da Educação Brasileira, pode-se observar que a porcentagem de evasão é maior para os que já reprovaram tais matérias. Sendo que em ICC essa porcentagem é de exatamente 50%. Ao analisar se há associação entre evasão e cada uma dessas variáveis de reprovação, pode-se observar ao nível de confiança de 5% que apenas Reprovação em Organização da Educação Brasileira tem associação com a evasão. Embora o valor p da reprovação em IAL tenha ficado muito próximo do nível de significância estipulado e podendo assim talvez ser um pouco mais flexível e adentra-la ao modelo, é possível observar que as porcentagens dos que evadem ou não evadem entre os que reprovaram e não reprovaram são muito próximas sendo então algo que pode de fato ser ao acaso, e por esse fator optou-se pela não flexibilização dos 5% no nível de confiança e com isso não ser uma variável de interesse ao criar o modelo.

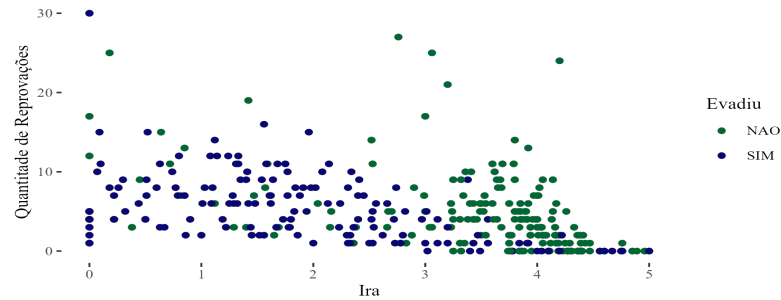
5.3.2 Correlação entre variáveis

Ao analisar as variáveis indicativas para adentrar o modelo, pode-se notar que algumas dessas variáveis podem apresentar correlação entre si. Como o modelo será analisado por meio da Regressão Logística é importante analisar essas variáveis afim de evitar erro por multicolinearidade no modelo.

Tendo em vista que as variáveis de índice de rendimento acadêmico e taxa de reprovação dizem respeito ao rendimento do aluno e que a taxa de reprovação é um dos fatores que influenciam no IRA, será realizado um teste de correlação de Pearson (ρ), (dado que se trata de duas variáveis numéricas), entre elas para testar se estas variáveis

estão ou não correlacionadas de fato.

Figura 14: Correlação entre IRA e Taxa de reprovação, Licenciatura Diurna em Matemática - UnB, 2014 - 2019



Ao realizar o teste é possível observar que as variáveis estão inversamente correlacionadas, ou seja, quanto maior o IRA menor tende a ser a Taxa de reprovações e vice versa. Porém, o valor de r é igual a $-0,403$, ou seja, sabe-se que há uma correlação fraca entre essas duas variáveis.

Portanto, em primeiro momento as duas variáveis farão parte do modelo pois acredita-se que o efeito de uma não irá interferir no efeito da outra.

Cotista x Tipo escola

Dado que alguns tipos de cotas que a Universidade de Brasília dispõe está ligada ao estudante advim de escola pública, é natural que se suspeite que essas variáveis estejam associadas. Para isso, será realizado um teste qui quadrado de Pearson para avaliar a associação entre tipo de escola e cota.

Dessa forma, testou se:

Tabela 12: Correlação entre sistema de cotas e tipo de escola do 2º grau, Licenciatura Diurna em Matemática - UnB, 2014 - 2019

| Variável | Escola | | |
|------------------|---------|------------|-----|
| | Pública | Particular | |
| Sistema de Cotas | Sim | 128 | 17 |
| | Não | 50 | 140 |

Como o teste resulta um p-valor menor que 0.0001 tem-se motivos para rejeitar a hipótese nula do teste, que afirma que não existe associação entre as variáveis.

5.3.3 Modelagem

Diante a tabela 11, é possível observar as variáveis com indicação para entrar no modelo, uma vez que as mesmas possuem associação com a evasão. Um ponto importante é que essas variáveis têm indicação, sendo assim, nem todas permanecerão no modelo, uma vez que estas podem não contribuir significativamente para explicar o evento de interesse, no caso, a evasão dos alunos de Licenciatura Diurna em Matemática da Universidade de Brasília nos anos de 2014-2019.

As variáveis com indicações para entrar no modelo foram: Negro, Tipo de escola, Cota, Forma de ingresso agrupada, Quantidade de reprovação, Reprovou a disciplina Organização de Ensino Brasileiro, Trancou, Verão. Além do mais, ainda que sem indicações expressas via teste, as variáveis gênero e idade serão testadas no modelo.

Tabela 13: Modelo proposto através da análise bivariada, Licenciatura Diurna em Matemática - UnB, 2014-2019.

| Parâmetro | Estimativa | Desvio Padrão | Estatística | P-valor |
|--------------------------------|------------|---------------|-------------|---------|
| Intercepto | 1.9314 | 2.6072 | 0.5488 | 0.4588 |
| IRA | -0.2178 | 0.0928 | 5.5117 | 0.0189 |
| Gênero - Masc | -0.4277 | 0.6526 | 0.4294 | 0.5123 |
| Raça/cor - Não Negro | 0.2329 | 0.6050 | 0.1481 | 0.7003 |
| Ingresso - PAS | 2.6650 | 1.2819 | 4.3218 | 0.0376 |
| Ingresso - SISU | 3.8472 | 1.4248 | 7.2907 | 0.0069 |
| Ingresso - Vestibular | 2.7277 | 1.3070 | 4.3559 | 0.0369 |
| Idade | -0.00214 | 0.0718 | 0.0009 | 0.9763 |
| Reprovação | 0.5933 | 0.8726 | 0.68 | 0.4966 |
| Cotista - Sim | -0.4026 | 0.8353 | 0.2323 | 0.6298 |
| Reprovou Org. Ens. Bras. - Sim | 0.6281 | 0.8291 | 0.5739 | 0.4487 |
| Trancou - Sim | 2.2533 | 0.6690 | 11.3451 | 0.0008 |
| SR - Sim | -0.0705 | 0.7521 | 0.0088 | 0.9254 |
| Verão - Sim | -1.5498 | 0.7406 | 4.3790 | 0.0364 |
| Tipo Escola - Pública | -1.2586 | 0.7649 | 2.7075 | 0.0999 |

Logo, ao observar as estimativas para as variáveis no modelo proposto e analisando seus respectivos p-valor, com um nível de significância de 5%, opta-se pela não continuidade das variáveis Gênero, Negro, Ingresso, Idade, Reprovação, Cotista, Taxa de reprovação, Reprovou Organização do Ensino Brasileiro e Tipo de escola, pois entende-se que estas não estão contribuindo significativamente para analisar o evento de interesse.

A tabela abaixo contém as estimativas para essas variáveis para a amostra de construção, validação e banco geral.

Tabela 14: Estimativas dos parâmetros para as bases de construção, validação e geral. Licenciatura Diurna em Matemática - UnB, 2014-2019.

| Parâmetro | Estimativa-Const. | Estimativa-Valid. | Estimativa-Geral |
|---------------|-------------------|-------------------|------------------|
| Intercepto | 3.6241 | 4.1143 | 3.5483 |
| IRA | -1.1539 | -1.5399 | -1.1789 |
| Reprovação | -0.1718 | -0.0559 | -0.1380 |
| Trancou - Sim | 1.9162 | 2.6783 | 2.0757 |
| Verão - Sim | -1.5382 | -1.1514 | -1.4904 |

Ao usar o método de seleção automático Stepwise, chega-se ao mesmo modelo acima. Pode-se também observar a partir da tabela 14 que as estimativas para as variáveis da amostra de construção, validação e do banco completo ficaram próximas. Com isso, opta-se por seguir com o modelo construído. Tal modelo conta com um AIC igual a 239.542.

Tabela 15: Modelo final - Licenciatura Diurno em Matemática - UnB, 2014-2019.

| Parâmetro | Estimativa | Desvio Padrão | Estatística | P-valor |
|---------------|------------|---------------|-------------|---------------|
| Intercepto | 3.6241 | 0.6401 | 32.0524 | ≤ 0.0001 |
| IRA | -1.1539 | 0.1648 | 49.0565 | ≤ 0.0001 |
| Reprovou | -0.1718 | 0.0471 | 13.2728 | 0.00031 |
| Trancou - Sim | 1.9162 | 0.4308 | 19.7886 | ≤ 0.0001 |
| Verão - Sim | -1.5382 | 0.4693 | 10.744 | 0.0010 |

5.3.4 Interpretação dos Parâmetros

Tabela 16: Modelo para evasão, Licenciatura Diurno da Matemática - UnB, 2014 - 2019

| Parâmetro | Odds Ratio | 2.5 % | 97.5 % |
|---------------|------------|-------|--------|
| IRA | 0.315 | 0.228 | 0.436 |
| Reprovação | 0.842 | 0.768 | 0.9241 |
| Trancou - Sim | 6.795 | 2.921 | 15.808 |
| Verão - Sim | 0.215 | 0.086 | 0.539 |

Para a interpretação dos parâmetros do modelo foi utilizada a razão de chances (Odds Ratio). A estimativa do parâmetro referente ao IRA apresentada na Tabela 16 é negativa, indicando que quanto maior o IRA, menor a taxa de evasão. Sendo assim, sua odds ratio indica que para cada acréscimo de um 1 ponto no IRA, a chance de evasão decresce e é 0.685 menor. De forma semelhante, pode-se verificar também que para cada percentual a mais na taxa de reprovação a chance de evasão é 15.8% menor.

Ao analisar o parâmetro Trancou, observa-se que sua razão de chances é de 6.795, isso significa que um aluno que já tenha trancado qualquer disciplina de forma não justificada tem seis vezes mais chances de evadir o curso do que um aluno que não tenha passado por tal situação.

Ao analisar Verão, é possível observar que os alunos que cursaram matéria no verão tem razão de chances igual a 0.785 vezes menor de evadir do curso, quando comparado ao aluno que nunca pegou nenhuma matéria no verão.

Tabela 17: Teste Hosmer - Lemeshow, Licenciatura Diurno da Matemática - UnB, 2014 - 2019

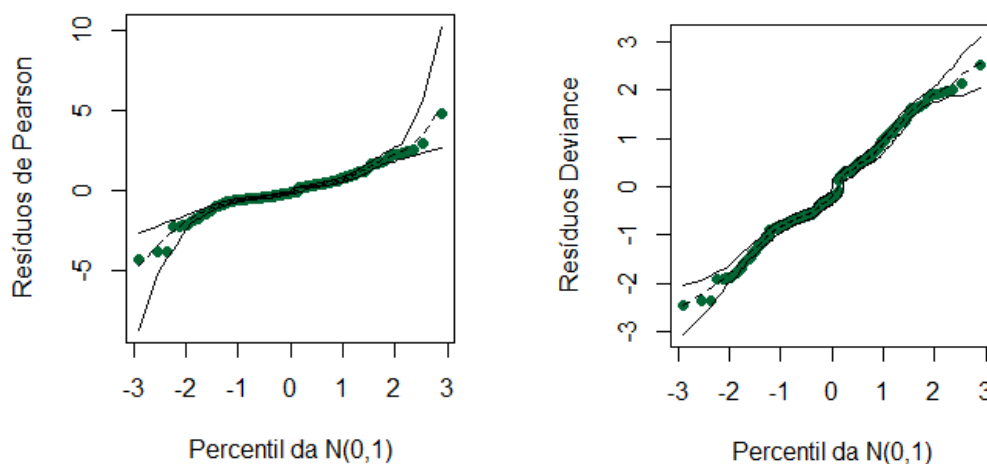
| | Estatística | P-valor |
|-------------------|-------------|---------|
| Hosmer - Lemeshow | 10.407 | 0.2376 |

5.3.5 Testes de ajustes e diagnósticos do modelo

Para testar se o modelo está bem ajustado foi realizado o teste de Hosmer - Lemeshow, e com um nível de significância de 5% não rejeitamos a hipótese de que o modelo está de fato bem ajustado.

Para além disso, os gráficos da Figura 15a e 15b apresentam os resíduos de Pearson e Deviance. É possível observar que os resíduos estão dispersos no interior dos envelopes simulados, sem aparente padrão sistemático dando indício de que o modelo está bem ajustado.

Figura 15: Resíduos do modelo, Licenciatura Diurno em Matemática - UnB, 2014-2019



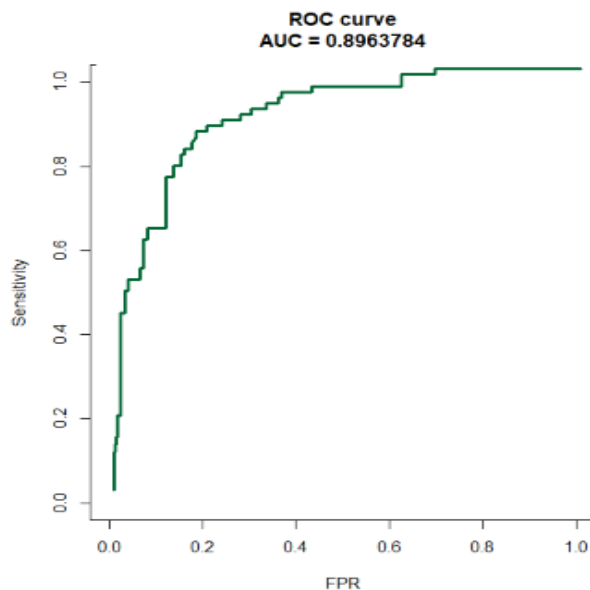
(a) Resíduos de Pearson.

(b) Resíduos Deviance.

5.3.6 Curva ROC

”A curva ROC (Receiver Operating Characteristic curve) é utilizada para avaliar a qualidade do ajuste de um modelo de classificação, ou seja, quão bem o modelo criado está classificando o evento de evasão dos alunos do curso de Licenciatura Diurno em Matemática da Universidade de Brasília entre os anos de 2014 a 2019.

Figura 16: Curva ROC, Licenciatura Diurno em Matemática - UnB, 2014-2019



Diante a curva ROC foi obtido um AUC (Área abaixo da curva ROC) de 0,896, ou seja, o modelo fornece uma habilidade de classificação de fato muito boa.

5.3.7 Matriz de Confusão

Também com o intuito de prever o quanto o modelo está acertando, foi criada uma matriz de confusão. Esta por sua vez, é analisada na base de validação, sendo que a mesma ficou disposta da seguinte forma:

Tabela 18: Matriz de confusão, Licenciatura Diurna da Matemática - UnB, 2014 - 2019

| | Não evadiu | Evadiu |
|------------|------------|--------|
| Não evadiu | 25 | 4 |
| Evadiu | 9 | 29 |

É possível observar através da matriz de confusão que o modelo escolhido erra em um nível muito baixo, sendo que ele diz que nove estudantes evadiram, quando na verdade estes não evadiram, ou seja, acertou em 73,5% dos casos de não evasão. Para os casos em que o modelo afirma o resultado evasão, é possível observar que ele erra em apenas 4 casos, tendo assim uma taxa de acerto de 87,8%.

Logo, para ambos os casos têm-se valores altos de acertos.

Testando diante a matriz de confusão podemos observar uma acurácia de 80,6% o que é muito bom. Ademais, têm-se para a sensibilidade e especificidade os valores de 0,763 e 0,862 respectivamente.

5.4 Licenciatura Noturno de Matemática

5.4.1 Evasão

Ao encontro do que foi observado na forma de saída dos estudantes podemos notar uma alta taxa de evasão dos estudantes do curso de Matemática Licenciatura Noturno, considerando os estudantes ingressantes entre 2014 e 2019, 54% dos alunos evadem do curso, ou seja, menos da metade dos ingressantes se formam no curso de licenciatura em matemática noturno.

Tabela 19: Análise bivariada por evasão, Licenciatura Noturno em Matemática - UnB, 2014-2019.

| Variável | | Evasão | | Estatística | P-valor |
|----------------------------|-------------------------------|--------|--------|-------------|----------------|
| | | Sim | Não | | |
| Gênero | Feminino | 50% | 50% | 0.5688 | 0.4507 |
| | Masculino | 55.6% | 44.4% | | |
| Raça/cor | Negro | 56.6% | 43.4% | 0.6121 | 0.7363 |
| | Não Negro | 51.8% | 48.2% | | |
| Tipo de Escola | Particular | 53.4% | 46.6% | 0 | 1 |
| | Pública | 53.5% | 46.5% | | |
| Cota | Sim | 56.2% | 43.8% | 0.1749 | 0.6757 |
| | Não | 53.4% | 46.6% | | |
| Forma de Ingresso | Programa de Avaliação Seriada | 38.2% | 61.8% | 17.174 | 0.0006 |
| | Sistema de Seleção Unificada | 61.9% | 38.1% | | |
| | Vestibular | 62.7% | 37.3% | | |
| | Outras | 56.6% | 43.4% | | |
| Quantidade de Reprovação | Média | 5.42% | 4.87% | 0.93541 | 0.3507 |
| Idade de ingresso | Média | 24.52% | 23.37% | 4.671 | 0.0306 |
| IRA | Média | 1.16% | 3.169% | -14.996 | ≤ 0.0001 |
| Reprovação Cálculo 1 | Sim | 59.8% | 40.2% | 0.629 | 0.4277 |
| | Não | 54% | 46% | | |
| Reprovação Cálculo 2 | Sim | 50% | 50% | 3.554 | 0.0593 |
| | Não | 31.8% | 68.2% | | |
| Reprovação Física | Sim | 50.9% | 49.1% | 0.0714 | 0.7893 |
| | Não | 47.4% | 52.6% | | |
| Reprovação Geom. Analítica | Sim | 62.5% | 37.5% | 1.3799 | 0.2401 |
| | Não | 54.2% | 45.8% | | |
| Reprovação IAL | Sim | 47.5% | 52.5% | 0.2125 | 0.6448 |
| | Não | 52.3% | 47.7% | | |
| Reprovação ICC | Sim | 75.6% | 24.4% | 7.2479 | 0.007 |
| | Não | 52.5% | 47.5% | | |
| Reprovação Org. Edu. Bras. | Sim | 76.9% | 23.1% | 2.9305 | 0.0869 |
| | Não | 48.7% | 51.3% | | |
| Trancamento | Sim | 88.6% | 11.4% | 66.798 | $3.008e^{-16}$ |
| | Não | 41% | 59% | | |
| Verão | Sim | 31.6% | 68.4% | 13.09 | 0.0002 |
| | Não | 58.5% | 41.5% | | |
| SR | Sim | 65.9% | 34.1% | 36.714 | $1.368e^{-09}$ |
| | Não | 32.3% | 67.7% | | |

Ao analisar os alunos de Licenciatura Noturno em Matemática segundo a evasão, é possível notar que para o gênero feminino a evasão teve um total de 50% das, ou seja, a probabilidade de evadir ou não evadir dado que é do gênero feminino seria dada ao acaso, enquanto as pessoas do gênero masculino em sua maioria evadiu. Entretanto, no caso masculino a diferença entre os que evadiram e os que não evadiram é menor do que 10%, ou seja, a princípio não consideramos essa uma grande diferença. Ao realizar teste qui quadrado para verificar se há associação entre gênero e evasão, considerando um nível de significância de 5%, descartamos a hipótese de associação entre o gênero do aluno e a finalidade evasão do referido curso. Portanto, acredita-se em primeiro momento que ser do gênero feminino ou masculino não irá interferir na evasão ou não evasão de curso.

Para os Negros e não negros analisados, pode-se observar que dentre os autodeclarados Negros 56,6% deles evadiram o curso de Licenciatura Noturno de Matemática, enquanto para os autodeclarados Não Negros essa porcentagem é de 51,8%. Em ambos os casos há uma porcentagem maior de pessoas que evadiram o curso, sendo que a porcentagem de Negros é ligeiramente maior. Ao testar associação entre evasão e a raça/cor agrupada dos estudantes em questão, podemos a um nível de significância de 5% acreditar que a Raça/cor agrupada não tem associação com o estudante evadir ou não o curso.

No que tange ao tipo de escola a qual o aluno cursou o 2º grau de ensino, pode-se observar que entre os alunos que advém da escola de rede particular 53,4% deles evadiram o curso, enquanto, para os alunos da rede pública esse valor é de 53,5%. Em ambos os casos, a diferença entre os que evadiram ou não evadiram o curso é de menos de 10%, além da taxa entre cada tipo de escola ser basicamente a mesma. Com o intuito de analisar se há associação entre a evasão e o tipo de escola do 2º grau do estudante, com um grau de significância de 5%, acredita-se que não há associação entre esses dois fatores.

Dentre os alunos cotistas e não cotistas, segundo a evasão, observa-se que a maior parte dos cotistas (56,2%) evadiram o curso, enquanto para os não cotistas a porcentagem dos que evadiram o curso foi de 53,5%. Em ambos os casos a diferença entre os que evadiram e não evadiram não chega a 10%, além dos cotistas terem uma taxa maior de evasão entre si. Ao analisar se há associação entre ser cotista e evadir o curso de Licenciatura Noturno da Matemática, com um nível de significância de 5% não há indicações que estes dois fatores estejam associados.

A forma de ingresso dos alunos foi agrupada em PAS, SISU, Vestibular e Outras, para fim de análise. É possível observar que a maioria dos estudantes que ingressaram pelo PAS não evadiram o curso, enquanto para o SISU, Vestibular e Outras a maior porcentagem destes evadiram. Com exceção da forma de ingresso sendo o Outras, que teve uma diferença entre evasão e não evasão de 13%, para todas as outras essa diferença entre evadir ou não foi de pelo menos 23%. Ao analisar se há associação entre a forma de ingresso e evasão do curso, com um nível de significância de 5%, acredita-se na associação

dos dois fatores.

Foram analisadas também as variáveis Trancamento, Verão e SR que indicam respectivamente se o aluno trancou alguma matéria sim ou não, pegou alguma matéria de verão sim ou não e se já ficou com menção SR em alguma matéria sim ou não, para os alunos que já trancou matéria e, ou tirou SR a maior parte dos estudantes evadiram, já para os que fizeram matéria de verão a maior parte não evadiu. Ao testar associação entre evasão e cada uma das três variáveis pode-se observar, há um nível de 5% de confiança, que há indicação de associação nos três casos.

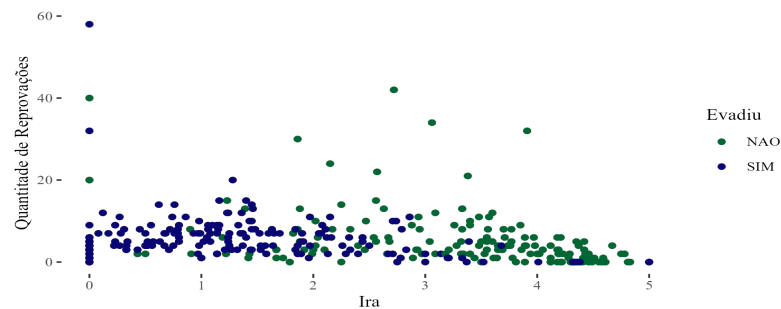
Foi criado uma variável Reprovação para cada uma das 7 matérias mais reprovadas pelos alunos da Licenciatura Diurna de Matemática, sendo elas Cálculo 1, Cálculo 2, Física, Geometria Analítica, IAL (Introdução à Álgebra Linear), ICC (Introdução à Ciência da Computação) e Organização da Educação Brasileira. Com exceção de Cálculo 2 e IAL, nas demais matérias analisadas todos os alunos que já reprovaram as matérias tem maior porcentagem entre os que evadiram o curso, sendo que em Cálculo 2 essa porcentagem é de exatamente 50%. Ao analisar se há associação entre evasão e cada uma dessas variáveis de reprovação, pode-se observar ao nível de confiança de 5% que apenas ICC tem associação com a evasão. Embora o valor p da reprovação em Cálculo 2 tenha ficado muito perto do nível de significância estipulado e podendo assim talvez ser um pouco mais flexível e adentra-la ao modelo, é possível observar que as porcentagens dos que evadem ou não evadem dado os que reprovaram e não reprovaram são muito próximas sendo então algo que pode de fato ser ao acaso, e por esse fator optou-se pela não flexibilização dos 5% no nível de confiança.

5.4.2 Correlação entre as variáveis

Ao analisar as variáveis indicativas para o modelo, pode-se notar que algumas dessas variáveis podem apresentar correlação entre si. Como o modelo será analisado por meio da Regressão Logística é importante analisar essas variáveis afim de evitar erro por multicolinearidade no modelo.

Tendo em vista que as variáveis de índice de rendimento acadêmico e taxa de reprovação dizem respeito ao rendimento do aluno e que a taxa de reprovação é um dos fatores que influenciam no IRA, será realizado um teste de correlação de Pearson (ρ) entre elas, (dado que se trata de duas variáveis numéricas), para testar se estas variáveis estão ou não correlacionadas de fato.

Figura 17: Correlação entre IRA e Taxa de reprovação, Licenciatura Noturno em Matemática - UnB, 2014 - 2019



Ao realizar o teste é possível observar que as variáveis estão inversamente correlacionadas, ou seja, quanto maior o IRA menor tende a ser a Taxa de reprovações e vice versa. Porém, o valor de r é igual a $-0,22$, ou seja, sabe-se que há uma correlação fraca entre essas duas variáveis.

Portanto, em primeiro momento as duas variáveis farão parte do modelo pois acredita-se que o efeito de uma não irá interferir no efeito da outra.

Cotista x Tipo escola

Dado que alguns tipos de cotas que a Universidade de Brasília dispõe é para a escola pública, é natural que se suspeite que essas variáveis estejam associadas. Para isso, será realizado um teste qui quadrado de Pearson para avaliar a associação entre tipo de escola e cota.

Dessa forma, testou se:

Tabela 20: Correlação entre cota e tipo de escola do 2º grau, Licenciatura Noturno em Matemática - UnB, 2014 - 2019

| Variável | Escola | | |
|------------------|---------|------------|-----|
| | Pública | Particular | |
| Sistema de Cotas | Sim | 118 | 17 |
| | Não | 84 | 146 |

Como o teste resulta um p-valor menor que 0.0001 tem-se motivos para rejeitar a hipótese nula do teste, que afirma que as não existe associação entre as variáveis.

5.4.3 Modelagem

Diante a tabela 19 é possível observar todas as variáveis com indicação de entrar no modelo, uma vez que as mesmas possuem associação com a evasão, conforme visto anteriormente na análise bivariada.

Sabe-se que ter indicação para entrar no modelo, ou seja, ter associação ao evento de interesse, não implica na permanência no modelo, uma vez que estas podem não contribuir significativamente para explicar o evento de interesse, no caso, a evasão dos alunos de Licenciatura Noturno de Matemática da Universidade de Brasília nos anos de 2014-2019.

As variáveis com indicações para entrar no modelo, conforme a análise bivariada, foram: forma de ingresso, quantidade de reprovação, IRA, reprovação ICC, trancamento, verão e SR. Além disso, no primeiro momento, foi testado as variáveis gênero e reprovou organização da educação brasileira.

Ao testar tal modelo proposto, pode-se observar que ambas as variáveis (gênero e reprovou Organização da Educação Brasileira) foram não significativas, logo, um novo modelo foi testado apenas com as variáveis com indicação referente a análise bivariada.

Tabela 21: Modelo proposto através da análise bivariada, Licenciatura Noturno em Matemática - UnB, 2014-2019.

| Parâmetro | Estimativa | Desvio Padrão | Estatística | P-valor |
|-----------------------|------------|---------------|-------------|----------|
| Intercepto | 1.9639 | 1.0339 | 3.6078 | 0.0575 |
| IRA | -1.2863 | 0.2386 | 29.0559 | < 0.0001 |
| Ingresso - PAS | 0.9181 | 0.8261 | 1.2349 | 0.2665 |
| Ingresso - SISU | 1.5630 | 0.8555 | 3.3378 | 0.0677 |
| Ingresso - Vestibular | 1.6507 | 0.8277 | 3.9771 | 0.0461 |
| Reprovação | -0.1162 | 0.0569 | 4.1780 | 0.0410 |
| Reprovou ICC - Sim | 1.7200 | 0.6850 | 6.3056 | 0.0120 |
| Trancou - Sim | 3.7397 | 0.7591 | 24.2723 | < 0.0001 |
| SR - Sim | -0.7000 | 0.6916 | 1.0245 | 0.3115 |
| Verão - Sim | -1.3264 | 0.8202 | 2.6155 | 0.1058 |

Logo, ao observar as estimativas para as variáveis no modelo proposto e seus respectivos p-valor, com um nível de significância de 5%, opta-se pela continuidade das variáveis IRA, reprovação, reprovou ICC e trancou.

A tabela abaixo contém as estimativas para as referidas variáveis para a amostra de construção, validação e banco geral.

Tabela 22: Estimativas dos parâmetros para as bases de construção, validação e geral, Licenciatura Noturno em Matemática - UnB, 2014-2019.

| Parâmetro | Estimativa-Const. | Estimativa-Valid. | Estimativa-Geral |
|--------------------|-------------------|-------------------|------------------|
| Intercepto | 2.6372 | 4.3704 | 3.0463 |
| IRA | -1.2060 | -1.3285 | -1.2197 |
| Reprovação | -0.1451 | -0.2910 | -0.1781 |
| Trancou - Sim | 3.2846 | 2.8508 | 3.1889 |
| Reprovou ICC - Sim | 1.2602 | 1.5395 | 1.2480 |

As estimativas para os parâmetros para as amostras de construção, validação e da base de dados geral são relativamente parecidas, com isso, pode-se concluir que este é um modelo válido. Além disso, foi feito também testes com outros modelos que se mostraram significativos, mas com base nos seus respectivos AIC e com testes de máxima verossimilhança o modelo que se mostrou melhor foi de fato o apresentado na tabela 22, sendo assim o modelo final com um AIC de 151.423.

Tabela 23: Modelo final, Licenciatura Noturno em Matemática - UnB, 2014-2019.

| Parâmetro | Estimativa | Desvio Padrão | Estatística | P-valor |
|--------------------|------------|---------------|-------------|----------|
| Intercepto | 2.6372 | 0.5788 | 20.7627 | < 0.0001 |
| IRA | -1.2060 | 0.1823 | 43.7401 | < 0.0001 |
| Reprovação | -0.1451 | 0.0529 | 7.5295 | 0.0061 |
| Trancou - Sim | 3.2846 | 0.6743 | 23.7307 | < 0.0001 |
| Reprovou ICC - Sim | 1.2602 | 0.6403 | 3.8733 | 0.0491 |

5.4.4 Interpretação dos Parâmetros

Tabela 24: Modelo para evasão, Licenciatura Noturno da Matemática - UnB, 2014 - 2019

| Variável Explicativa | Odds Ratio | 2.5 % | 97.5 % |
|----------------------|------------|-------|---------|
| IRA | 0.299 | 0.209 | 0.428 |
| Reprovação | 0.865 | 0.780 | 0.959 |
| Trancou - Sim | 26.69 | 7.121 | 100.100 |
| Reprovou ICC - Sim | 3.526 | 1.005 | 12.370 |

A estimativa do parâmetro referente ao IRA apresentada na Tabela 24 é negativa, indicando que quanto maior o IRA, menor a taxa de evasão. Sendo assim, sua odds ratio indica que para cada acréscimo de um 1 ponto no IRA, a chance de evasão decresce e é 0,701 menor.

Pode-se verificar também que para cada percentual a mais na taxa de reprovação em ICC a chance de evasão é 3 vezes maior. E que para cada percentual a mais na taxa de reprovação a chance de evasão é 13.5% menor.

Além disso, para a variável trancou, é perceptível que a chance de evasão cresce em vinte seis vezes para aqueles alunos que já trancou alguma matéria ao longo do curso.

5.4.5 Testes de ajustes e diagnósticos do modelo

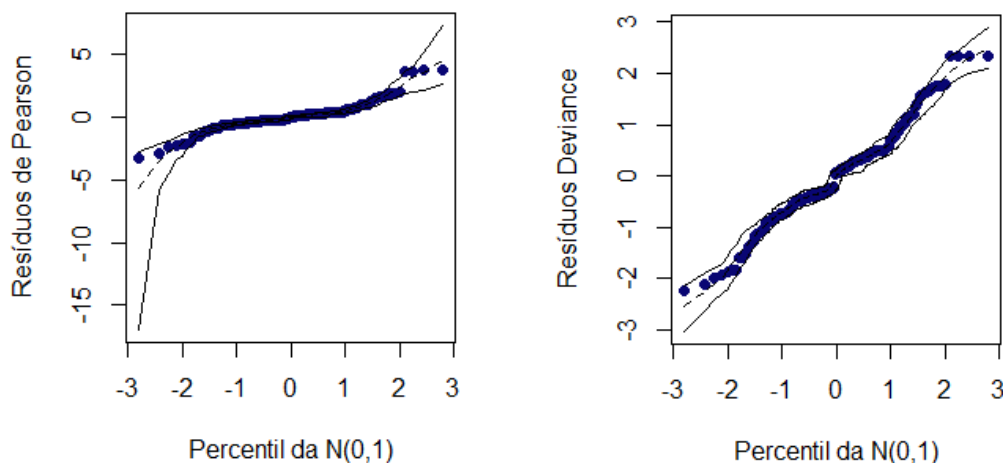
Tabela 25: Teste Hosmer - Lemeshow para qualidade de modelo

| | Estatística | P-valor |
|-------------------|-------------|---------|
| Hosmer - Lemeshow | 9.3042 | 0.3173 |

Para testar se o modelo está bem ajustado foi realizado o teste de Hosmer - Lemeshow, e com um nível de significância de 5% não rejeitamos a hipótese de que o modelo está de fato bem ajustado.

Para além disso, os gráficos da Figura 18a e 18b apresentam os resíduos de Pearson e Deviance. É possível observar que os resíduos estão dispersos no interior dos envelopes simulados, sem aparente padrão sistemático dando indício de que o modelo está bem ajustado.

Figura 18: Resíduos do modelo, Licenciatura Noturno em Matemática - UnB, 2014-2019



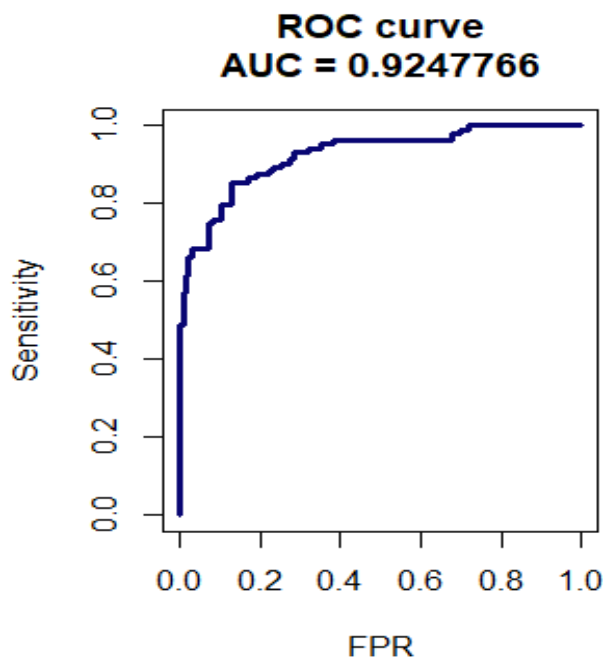
(a) Resíduos de Pearson.

(b) Resíduos Deviance.

5.4.6 Curva ROC

”A curva ROC (Receiver Operating Characteristic curve) é utilizada para avaliar a qualidade do ajuste de um modelo de classificação, ou seja, quão bem o modelo criado está classificando o evento de evasão dos alunos do curso de Licenciatura Noturno em Matemática da Universidade de Brasília entre os anos de 2014 a 2019.

Figura 19: Curva ROC, Licenciatura Noturno em Matemática - UnB, 2014-2019.



Diante a curva ROC foi obtido um AUC (Área abaixo da curva ROC) de 0.9247, ou seja, o modelo fornece uma habilidade de classificação de fato muito boa.

5.4.7 Matriz de Confusão

Também com o intuito de prever o quanto o modelo esta acertando, foi criada uma matriz de confusão. Esta por sua vez, é analisada na base de validação, sendo que a mesma ficou disposta da seguinte forma:

Tabela 26: Matriz de confusão - Licenciatura Noturno de Matemática - UnB, 2014 - 2019

| | Não evadiu | Evadiu |
|------------|------------|--------|
| Não evadiu | 20 | 6 |
| Evadiu | 5 | 49 |

É possível observar através da matriz de confusão que o modelo escolhido erra em um nível baixo, sendo que ele diz que cinco estudantes evadiram, quando na verdade estes não evadiram, ou seja, acertou em 80% dos casos de não evasão. Para os casos em que o modelo afirma o resultado evasão, é possível observar que ele erra em apenas seis casos, e tem uma taxa de acerto de 89,09%.

Testando diante a matriz de confusão podemos observar uma acurácia de 88,89% o que é bom. Ademais, têm-se para a sensibilidade e especificidade os valores de 0.9375

e 0.8182 respectivamente.

6 Conclusão

As motivações para alguém dar início a sua vida acadêmica podem ser diversas, como realizar sonhos, alcançar o sucesso em projetos, alavancar a vida financeira, estar posicionado no mercado de trabalho entre outros. Dado que esse alguém deu início a graduação, pode-se partir do princípio que concluir é sim o seu objetivo. Entretanto, sabe-se que as desistências são grandes, e avaliar o porque isso ocorre é de suma importância.

Foi visto que os cursos de Matemática historicamente sofre muito com o fenômeno da evasão e embora a Universidade de Brasília tenha uma taxa de evasão considerada baixa, os cursos de Licenciatura em Matemática tem índices elevados, notadamente o noturno. Entender o que ocorre e porque ocorre, possibilitando assim entender o que leva esses estudantes a desistirem do curso é o início do que pode mudar esse paradigma.

A história das mulheres no mercado de trabalho teve um início tardio e doloroso e quando esse episódio começou comumente essas estavam de forma majoritária na área da saúde ou educação. É comum até os dias atuais se deparar com muitas pedagogas, entretanto quando o panorama muda para professores sobretudo de ciências exatas é fácil notar que esse é um ambiente com mais homens, embora esse cenário venha cada vez mais se alterando.

Nos cursos de Licenciatura em Matemática diurno e noturno a uma prevalência de pessoas do gênero masculino, sendo inclusive no caso do curso noturno esse percentual chega a ser aproximadamente de 80% homens e apenas 20% mulheres, enquanto no diurno é de 66% homens e 34% mulheres.

No geral, nota-se que as idades dos estudantes do noturno são mais avançadas, desde o ingressante mais novo ao ingressante mais velho, enquanto metade dos estudantes do diurno tem até 18 anos, no noturno esse valor é de 20 anos. Além de que os valores no geral estão mais dispersos.

A raça/cor autodeclarada dos estudantes dos dois curso se dá de forma muito parecida, sendo que em ambos mais pessoas se autodeclararam pardas, branco, não informam e preta respectivamente nessa ordem.

De forma semelhante ocorre com o número de ingressantes cotistas, porém ao se inscrever nos meios de ingresso a Universidade, esses valores para cada cota já são pré-determinados, o que faz com que esses valores de fato muito próximos não fuja do que já era esperado

No que tange a vida acadêmica desses estudantes foi possível notar que com frequência os estudantes do diurno tem uma taxa maior de aprovações em disciplinas já cursadas quando comparado aos estudantes do noturno, embora essa diferença seja apenas

de 7%. Quando essa taxa é observada pelo número de semestres cursados de um aluno, ainda assim essas curvas se parecem, mostrando que no começo ou os alunos vão muito bem ou muito mal, e que no decorrer do curso eles acabam se saindo melhor, tendo menos reprovações. Isso demonstra a importância e o cuidado que se deve ter com as primeiras disciplinas e até mesmo que providências devem ser tomadas. Essa alta reprovação pode indicar um despreparo no ensino médio, ou também a forma das quais algumas disciplinas são abordadas, dentre outros possíveis fatores.

As disciplinas que os estudantes mais reprovam também se coincidem e são todas disciplinas do começo da graduação, sendo portanto disciplinas introdutórias. É preciso que haja um cuidado especial em relação a isso, as disciplinas do começo de curso dão base ao aluno e se isso não ocorre de forma plena pode dificultar o decorrer do curso, tanto em futuras disciplinas como também no desestímulo do estudante frente ao curso.

No geral as disciplinas com maior percentual de reprovação, com mais trancamentos e menções SR, são as mesmas ou pré-requisitos umas das outras. Sendo por exemplo Física 1, Física 2 e Física 1 Experimental. Chama a atenção que a disciplina Geometria Analítica para a Matemática aparece em todos esses casos, outro fator importante é que frente a esses fatores a disciplina Organização da Educação Brasileira aparece em quase todos os casos, essa disciplina é própria da licenciatura e é uma das mais voltadas a educação. Esse fator vai ao encontro do exposto em "Em Mussliner, B. O. et al (2021, p.42682) foi observado que a evasão é grandemente influenciada pelo rendimento acadêmico do aluno nos primeiros períodos do curso. Por isso, cursos nos quais há muitos casos de reprovação nos períodos iniciais tendem a apresentar elevadas taxas de evasão", já citado no começo desse estudo.

Na construção do modelo de regressão logística para o curso de Licenciatura Diurna em Matemática, observou-se que as variáveis IRA, quantidade de reprovações, se um aluno já trancou ou não um ou mais semestre(s) e se o já pegou matéria de verão influenciam significativamente no evento evadir o curso, de tal modo que estes não estão apenas associados mas explicam bem o modelo gerado. Para além disto, pode-se observar que o modelo foi bem ajustado e que teve uma boa acurácia, sensibilidade e especificidade.

De maneira semelhante ocorre para o modelo elaborado para o curso Licenciatura Noturna em Matemática. Novamente aparecem as variáveis IRA, quantidade de reprovação e se o aluno trancou ou não algum semestre, a diferença maior está que nos estudantes do curso noturno variável que diz respeito ao fato do o aluno já ter reprovado ICC também adentra o modelo ajudando a explicá-lo. Nesse caso, o modelo também se mostrou bem ajustado, com um nível bom de acurácia e bons valores para especificidade e sensibilidade.

Sugere-se para futuros estudos:

- Estudar a possibilidade de acesso aos endereços dos estudantes ou regiões administrativas, assim como o meio de transporte utilizado, visando entender se a distância/percurso é um impedimento para a continuidade do curso.
- Analisar aspectos socioeconômicos que possam explicar a descontinuidade.
- Expandir e propagar estudos que fomentem políticas públicas e dê visibilidade a problemática da evasão e as demais difíceis realidades encontradas nas universidades

Esse estudo tem base em um problema real que é a evasão nos cursos de Licenciatura em Matemática da Universidade de Brasília, e por meio dele há interesse em entender melhor a maneira como ocorrem essas evasões, o porque para a partir de então enfrentar um problema conhecido. Não obstante, esse estudo também vêm da necessidade e prazer ao contribuir com a Universidade de Brasília e assim poder propagar cada vez mais o quão importante é lutar por esse espaço e por uma educação cada dia mais digna e acessível.

7 Referências

AMBIEL, Rodolfo Augusto Matteo; BARROS, Leonardo de Oliveira. Relações entre evasão, satisfação com escolha profissional, renda e adaptação de universitários; **Psicologia: Teoria e Prática**, Brasil, vol. 20, núm. 2, p. 254-264, abr. 2018.

BRASIL. Constituição (1988). Constituição da República Federativa do Brasil. Brasília, DF: Senado Federal: Centro Gráfico, 1988.

CASTRO, A.K.S.S; TEIXEIRA, M.A.P. Evasão universitária: modelos teóricos internacionais e o panorama das pesquisas no Brasil; **Psicologia Argumento Artigo**, Curitiba, v. 32, n. 79, p. 9-17, Supl 1., 2014.

COSTA, S. L., & DIAS, S. M. B. (2015). A permanência no ensino superior e as estratégias institucionais de enfrentamento da evasão. **Jornal de Políticas Educacionais**, 9(17-18),51-60. doi:10.5380/jpe.v9i17/18.38650.

MÁRQUEZ-VERA, C. et al. Early dropout prediction using data mining: a case study with high school students. **Expert Systems**, Wiley Online Library, v. 33, n. 1, p. 107–124, 2016.

MEIRELES, Renato Sampaio. **Fatores Causais da Evasão e Retenção Universitária: Uma Análise nos Cursos de Licenciatura da Universidade de Brasília no Período de 2002 até 2018**. Orientador: Profa. Dra. Andrea Felipe Cabello. 2019. Dissertação (Mestrado profissional em Economia Área de Concentração em Gestão Econômica de Finanças Públicas) – Universidade de Brasília, 2019.

MINISTÉRIO DA EDUCAÇÃO (2013). Desenvolvimento, aprimoramento e consolidação de uma educação nacional de qualidade. Brasília: Conselho Nacional de Educação. Recuperado em 21 maio, 2018, de [http:// portal.mec.gov.br/](http://portal.mec.gov.br/).

MINISTÉRIO DA EDUCAÇÃO. Censo da educação superior 2018. Disponível em: http://download.inep.gov.br/educacao_superior/censo_superior/documentos/2019/apresentacao_censo_superior2018.pdf.

MUSSLINER, B. O. et al. O problema da evasão universitária no sistema público de ensino superior: uma proposta de ação com base na atuação de uma equipe multidisciplinar; **Brazilian Journal of Development**, Curitiba, v.7, n.4, p. 42674-42692, abr. 2021.

NETER, J., KUTNER, M., NACHTSHEIM, C. J. e WASSERMAN, W. **Applied linear statistical models**. 5a edição. Illinois: Irwin, 2005.

RODRIGUES Lorrana Thalita Silva; ABREU, Paulo Henrique Oliveira De; COSTA, Thais Laudelino. **A Evasão dos Acadêmicos no Curso de Matemática**. Orientador:

Prof.^a Ma .Kelly Sulâiny Alves Constante. 2014. Dissertação (Especialização em docência universitária X) – Faculdade católica de Anápolis, 2014.

SIMAS, A. As graduações campeãs de desistência. Reportagem do site do jornal A Gazeta do Povo. 2012. Disponível em: <http://www.gazetadopovo.com.br/educacao/vida-na-universidade/ufpr/as-graduacoes-campeas-de-desistencia> 26khijqty1gurtas1veawhyz2. Acesso em: 10 de fevereiro de 2017.

TINTO, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. **Review of Educational Research**, 45, 89-125.