



**Universidade de Brasília
Departamento de Estatística**

**Estudo de fatores associados à conclusão do ensino médio das mulheres no
Distrito Federal, com enfoque na maternidade.**

Letícia Macêdo dos Santos

Trabalho de conclusão de curso apresentado para o Departamento de Estatística da Universidade de Brasília, como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

**Brasília
2024**

Leticia Macêdo dos Santos

Estudo de fatores associados à conclusão do ensino médio das mulheres no Distrito Federal, com enfoque na maternidade.

Orientadora: Profa. Maria Teresa Leão Costa

Trabalho de conclusão de curso apresentado para o Departamento de Estatística da Universidade de Brasília, como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

**Brasília
2024**

Agradecimentos

Em primeiro lugar, agradeço a Deus por me conceder força, sabedoria e perseverança ao longo desta jornada acadêmica. Suas bênçãos foram essenciais para a conclusão deste trabalho.

Gostaria de expressar minha profunda gratidão aos meus pais, Denise e Ricardo, por serem o meu porto seguro e rede de apoio fundamental durante toda a minha trajetória. Seu amor, compreensão e encorajamento foram essenciais para que eu pudesse seguir em frente e concluir meus estudos. Cada gesto e palavra de incentivo foram vitais para a minha perseverança, este trabalho é tanto meu quanto de vocês. Sem vocês, nada disso seria possível!

Ao meu filho, a quem devo gratidão indescritível. Meu pequeno José, sua presença, alegria, amor e energia única trouxeram novo significado à minha vida. Mesmo nos momentos mais desafiadores, seu sorriso e suas pequenas conquistas foram a força que me manteve motivada e determinada a concluir este trabalho. Você é luz! Cada dia ao seu lado é um presente, sou grata por ser sua mãe, e para você, o meu amor eterno.

Ao meu noivo, Andrei, obrigada! Sua paciência, amor, compreensão e apoio durante este período foram admiráveis e essenciais para essa conclusão. Obrigada pelo incentivo e confiança constante, que tornou os desafios mais fáceis e as conquistas mais significativas. Sou verdadeiramente grata por ter você como meu parceiro, não apenas nesta jornada, mas em cada aspecto da minha vida.

Aos meus irmãos, Dyêgo e Lorryne, e à minha Doce Helena, obrigada pela ajuda e amor ao longo de toda a minha vida, sou grata e abençoada por ter vocês!

À minha orientadora, Maria Teresa, minha profunda gratidão. Sua atuação não foi apenas como professora, mas como mentora. Muito obrigada por cada conselho, palavra de encorajamento e por seu exemplo de dedicação e sabedoria.

A todos que, de alguma forma, contribuíram para a realização deste trabalho, meu sincero agradecimento.

Resumo

O estudo propõe investigar fatores que influenciam a conclusão do ensino médio entre mulheres de 14 a 24 anos, com enfoque na maternidade. Por meio de uma análise de regressão logística, o trabalho busca identificar como variáveis de perfis socioeconômicos, composição familiar e saúde, podem impactar a continuidade dos estudos.

A hipótese central é que a maternidade, juntamente com outros fatores desempenham papel significativo na conclusão do ensino médio. Além disso, o estudo examinará o impacto de aspectos como o uso da internet para fins educacionais, a composição do núcleo familiar e possíveis barreiras logísticas enfrentadas pela jovens.

Espera-se que os resultados apresentados contribuam para um entendimento mais profundo dos desafios enfrentados por essas mulheres e, com isso, forneçam subsídios para políticas públicas que promovam maior a equidade educacional e apoio às estudantes que precisam conciliar maternidade e estudos.

Palavras-chave: Regressão Logística, conclusão do ensino médio, mulheres jovens, maternidade, fatores socioeconômicos, barreira educacional, perfil educacional.

Lista de Tabelas

1	Distribuição de Frequência e Percentual por Grupo Etário da PDAD 2021	27
2	Categorias de Renda e Regiões	28
3	Distribuição do nível de escolaridade da PDAD 2021	29
4	Distribuição de mulheres de 14 a 24 anos, segundo raça/cor; Distrito Federal, Brasil - 2021	35
5	Distribuição das mulheres de 14 a 24 anos segundo faixa de renda - Distrito Federal, Brasil - 2021	36
6	Distribuição das mulheres de 14 a 24 anos segundo uso da internet; Distrito Federal, Brasil - 2021	37
7	Distribuição das mulheres de 14 a 24 anos, segundo dificuldades relatadas; Distrito Federal, Brasil - 2021.	39
8	Distribuição de mulheres de 14 a 24 anos, segundo motivos para Buscar Atendimento; Distrito Federal, Brasil - 2021	40
9	Distribuição das mulheres de 14 a 24 anos, segundo as estratégias de Busca de Emprego; Distrito Federal, Brasil - 2021	42
10	Distribuição dos níveis de Escolaridade entre mulheres de 14 a 24 anos da amostra de construção	44
11	Análise bivariada de nível de escolaridade da amostra selecionada da PDAD 2021	45
12	Tabela de Estimativas dos Parâmetros do Modelo	51
13	Tabela de Estimativas dos Parâmetros do Modelo Final	51
14	Porcentagem de concordância	52
15	Estimativas de Razão de Chance (<i>Odds Ratio</i>) com Intervalos de Confiança de 95%	54

Lista de Figuras

1	Curva da função de Regressão Logística	14
2	Curva da função de Regressão Logística para diferentes valores de β_i	14
3	Curva da função de Regressão <i>Logit</i>	15
4	Exemplo gráficos de resíduos	21
5	Exemplo do gráfico da Curva ROC	23
6	Distribuição das mulheres de 14 a 24 anos segundo faixa etária; Distrito Federal, Brasil - 2021	34
7	Distribuição de mulheres de 14 a 24 anos segundo raça cor; Distrito Federal - 2021	34
8	Gráfico de setores da distribuição de mulheres de 14 a 24 anos segundo estado civil; Distrito Federal, Brasil - 2021	35
9	Distribuição das mulheres de 14 a 24 anos segundo dificuldade; Distrito Federal, Brasil - 2021.	38
10	Distribuição das mulheres de 14 a 24 anos, segundo frequência educacional; Distrito Federal, Brasil - 2021.	40
11	Distribuição das mulheres de 14 a 24 anos, segundo modalidade de ensino; Distrito Federal, Brasil - 2021.	41
12	Gráfico de setores da distribuição das mulheres de 14 a 24 anos, segundo filho nascido vivo; Distrito Federal, Brasil - 2021	43
13	Curva ROC do modelo em estudo; Distrito Federal, Brasil - 2021	53
14	Gráfico de dispersão dos resíduos Deviance vs Valores Preditos do modelo em estudo; Distrito Federal, Brasil - 2021	53

Sumário

1 Introdução	9
2 Objetivos	12
2.1 Objetivo Geral	12
2.2 Objetivos Específicos	12
3 Regressão Logística	13
3.1 Modelo de Regressão Logística Múltipla	15
3.2 Estimação dos parâmetros do modelo	16
3.3 Inferência no Modelo de Regressão Múltipla	17
3.3.1 Teste da Razão da Verossimilhança	17
3.3.2 Teste de Wald	18
3.3.3 Intervalo de Confiança	18
3.4 Seleção do Modelo	18
3.5 Diagnóstico do Modelo	19
3.5.1 Resíduos de Pearson	20
3.5.2 Resíduos de Pearson Semistudentizado	20
3.5.3 Gráficos de Resíduos	20
3.6 Inferência sobre a resposta média	21
3.7 Modelo de Regressão Logística Complexa	22
3.8 A Curva ROC	22
4 Metodologia	24
4.1 Banco de dados	24
4.2 Variáveis	24
4.2.1 Variáveis originais	24
4.2.2 Variáveis criadas	25
4.3 Método	31
5 Resultados	32
5.1 Análise Descritiva	32
5.1.1 PDAD 2021	32
5.1.2 Análise Descritiva da amostra do presente estudo	33
5.2 Análise Descritiva Bivariada	43
5.3 Modelagem	49
5.3.1 Avaliação do modelo	52
5.3.2 Interpretação dos parâmetros	54
6 Conclusão	56

Referências 59

1 Introdução

O papel da mulher na história tem se modificado de forma vigorosa no decorrer dos anos. No passado, como destaca Silveira et al. (2019), a mulher era vinculada ao âmbito doméstico, onde era apenas responsável pela procriação e gestão do lar. Entretanto, com as mudanças e reivindicações ao longo da história, o sexo feminino foi ganhando mais espaço e direitos. Apesar da herança histórica e cultural, de ser mãe e esposa, ela passou a ter a possibilidade da decisão a respeito do seu próprio futuro. As mulheres modernas possuem a consciência do seu potencial e demonstram interesse pela auto valorização, incluindo os estudos, fator que antes era de direito apenas dos homens (BARRETO, 2016; GARCIA, 2009).

Embora tardia a inclusão feminina na educação, a presença das mulheres tem sido cada vez mais expressiva. Em dados disponibilizados pelo CENSO, é possível observar que as estudantes representam 51,6% dos alunos no ensino médio, além de representarem maioria nas participações de exames e avaliações de ensino, elas também simbolizam mais de 2/3 dos cargos das diretorias educacionais (FEDERAL, 2023).

Entretanto, apesar da evolução histórica, há um outro contexto que está hodiernamente criando mais visibilidade, nomeado como tragédia silenciosa: a evasão escolar. Com base nos dados do CENSO, o Governo Federal ressalta que no Brasil, 60,3% das pessoas completam a escolaridade básica (educação infantil, ensino fundamental e médio) até os 24 anos, porcentagem que abaixa para 46% ao observar isoladamente os mais pobres. Em contrapartida, se essa porcentagem se igualasse ao Chile (cerca de 93%), o ganho anual na economia seria de R\$135 bilhões. O que reforça o interesse do país de combater essa situação, porém para isso é importante entender a causa do abandono dos estudos onde estão fatores como: desinteresse, necessidade de trabalhar, criminalidade, demanda nas necessidades domésticas e gravidez (QUEIROZ, 2023; SOUSA et al., 2018; SILVA et al., 2020).

Com base em dados do CENSO Escolar, a instituição INESC ressalta que nos anos de 2020 e 2021, o abandono escolar no Ensino Médio no Brasil teve um aumento de 128%, saindo de 165 mil para 377 mil estudantes evadidos no país. No Distrito Federal, o valor absoluto de alunos que desistiram dos estudos nos anos de 2018, 2019 e 2020, foram respectivamente de 3.846, 658 e 1.178, embora observa-se que em 2018 houve o maior número, é possível concluir que a pandemia promoveu um novo aumento na evasão (INESC, 2023).

Conforme afirma Clementino, Cabral e Rodrigues (2020), a sociedade espera que a mulher se anule em prol da criação e constituição da sua família, o que muitas vezes é visto como atitude natural e, em consequência, atrapalha a trajetória acadêmica feminina, já que dificulta conciliar os diversos papéis que exercem, como: mãe, trabalhadora, dona de casa e/ou esposa (LEITE; ALVES, 2022).

O Ministério da Saúde segue a convenção adotada pela Organização Mundial de Saúde (OMS), o qual determina que adolescentes são pessoas na faixa etária de 10 a 19 anos, 11 meses e 29 dias, enquanto que juventude acontece entre os 15 e 24 anos, coincidindo os últimos anos da

adolescência com os primeiros anos da juventude. Ressalta-se que adolescência é a transição da infância para a vida adulta, com alto nível de modificações no corpo e instabilidade emocional, além disso, Carvalho e Matsumoto (2019) menciona que trata-se de um período de aquisição de independência, reconhecimento da identidade, desenvolvimento de valores, relações afetivas e preparação para a carreira profissional, ou seja, é uma transição intensa na vida de um cidadão, transformação na vida física, mental e social.

Em 2020, 14% dos partos realizados no Brasil foram de mães com até 19 anos de idade, a taxa média global de nascimentos vivos de crianças filhas de mães entre 15 e 19 anos é de 46 nascimentos por cada mil meninas, enquanto que no Brasil a estimativa foi de 68,4. Este fato intensifica a preocupação da saúde pública, pois nesta faixa etária a probabilidade de mortes relacionadas a gravidez ou parto é duas vezes maior do que em mães acima desse período, além de, segundo a Organização Mundial de Saúde (OMS), estes filhos possuem maior chance de apresentarem baixo peso ao nascer, uma taxa de mortalidade infantil duas a três vezes maior, um aumento de seis vezes no risco de morte súbita, fora os maiores riscos na mãe ao longo da gravidez, como: anemia, pré-eclâmpsia e aborto (FEDERAL, 2022).

No papel de estudante, a gravidez ao longo da vida acadêmica pode apresentar desafios ainda mais árduos do que o habitual para esta profissão, pois essas mulheres precisam equilibrar as demandas escolares com as responsabilidades da maternidade. Muitas vezes, essas gestações não são planejadas, com isso, assusta, pois, a mudança será repentina e enfática, desde a transformação no cotidiano à adaptação do físico, social e psicológico, necessitando assim de um amparo para minimizar todo o impacto, dado que a maternidade não se resume apenas a gestação (TAUIL, 2019; RABELLO; SILVA; ZORNING, 2010).

A Lei Federal de nº 6.202, de 17 de abril de 1975, é a legislação, no Brasil, que ampara as estudantes gestantes, atribuindo a discente em estado de gestação o regime de exercícios domiciliares, entretanto nem todas as instituições cumprem devidamente o que a lei estabelece. Além da incerteza do cumprimento devido da lei, os desafios permanecem (BRASIL, 1975).

Além disso, como observado por Nogales et al. (2023), a distribuição de mulheres com idades entre 20 e 39 anos segundo o grau de escolaridade apresenta comportamentos distintos quando relacionado com as variáveis condição de maternidade e renda. Verifica-se que mulheres dos grupos de renda baixa e média-baixa apresentam 9,9% e 25,6%, respectivamente, nível superior completo, em contrapartida, 82% e 50,9%, nessa ordem, das mulheres de alta renda e média-alta possuem nível superior completo, ressaltando intensa divergência, possibilitando assim um estudo em conjunto a respeito da relação das variáveis: renda, maternidade e ensino superior.

Embora a maternidade possa ser um momento difícil para todos da família, é necessário apoiar a estudante nessa passagem, seja por parte do parceiro, pais, amigos ou pessoas próximas, dado que a rede de apoio influenciará de forma direta tanto no prosseguimento quanto no abandono dos estudos (URPIA, 2009; OBANDO, 2018).

Brum (2022) afirma que “a maternidade é um desafio e ao mesmo tempo um mundo

de possibilidades”, visto que há excesso de sentimentos, demanda do dever na maternidade e, se ocorrer no período de ensino, terá a complexibilidade da obrigação acadêmica.

Com isso, através dos dados apresentados, é de suma importância uma análise a respeito da conclusão ou não do ensino médio, retratando a influência da maternidade e outros fatores. Destaca-se o crescente número de estudos de caso a respeito deste tema nos últimos anos. Além disso, há um interesse pessoal da autora, visto que ao longo da sua vida acadêmica, ela passou por uma gestação, onde pode vivenciar a dificuldade de conciliar a responsabilidade da maternidade com a demanda educacional.

O presente trabalho, analisará possíveis fatores que influenciam na conclusão do ensino médio das mulheres, com enfoque no impacto do exercício da maternidade, no Distrito Federal. O estudo será realizado com base nos dados da Pesquisa Distrital por Amostra de Domicílios (PDAD).

2 Objetivos

2.1 Objetivo Geral

Investigar os fatores que influenciam na conclusão do ensino médio das mulheres, com enfoque na maternidade, por intermédio de uma regressão logística.

2.2 Objetivos Específicos

- Traçar o perfil das estudantes com idade entre 14 e 24 anos;
- Analisar o nível de escolaridade das mulheres, com enfoque no nível médio;
- Identificar os fatores associados à conclusão do ensino médio, por meio de um modelo de regressão logística.

3 Regressão Logística

Regressão logística é um método estatístico que permite prever valores de uma variável resposta categórica em função de outra(s) variável(s) explicativa(s) independente(s) (AGRESTI, 2007).

Em um modelo de regressão logística, quando a variável resposta de interesse apresentar dois resultados qualitativos possíveis, denomina-se regressão logística binária, porém se a variável dependente possuir mais níveis ou classes, nomea-se regressão logística multinomial.

A variável dicotômica, ou binária, apresenta uma característica e assume dois resultados possíveis "sim" ou "não". No modelo de regressão logística binária a variável resposta, denominada Y , indica duas categorias designadas "sucesso" ou "fracasso", portanto segue uma distribuição de Bernoulli, ou seja:

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

com valor esperado:

$$E[Y_i] = \pi_i.$$

O Modelo de Regressão Logística Binária Simples é definido como:

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

Sendo que deseja-se estimar π_i como uma função da variável explicativa X_i da seguinte forma, dado:

$$E[Y_i] = \pi_i = \frac{\exp[\beta_0 + \beta_1 x_i]}{1 + \exp[\beta_0 + \beta_1 x_i]},$$

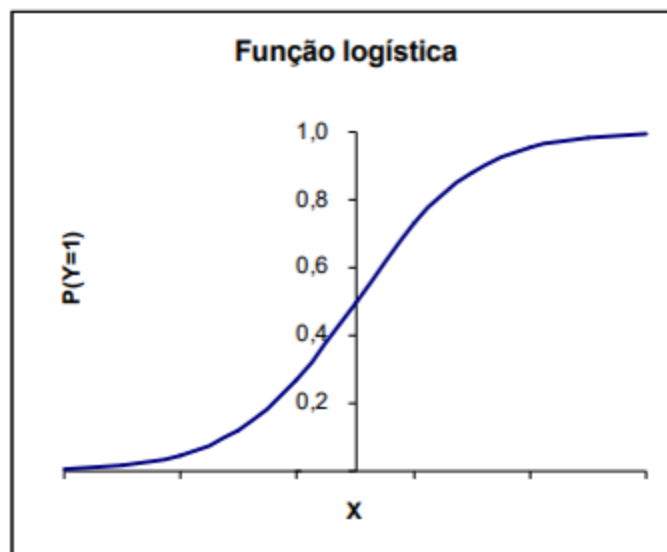
em que X_i são os valores das variáveis explicativas e, β_0 e β_1 são os coeficientes do modelo.

Ao efetuar a transformação logito, encontra-se:

$$\log(\pi) = \ln \left[\frac{\pi}{1 - \pi} \right] = \beta_0 + \beta_1 x_1.$$

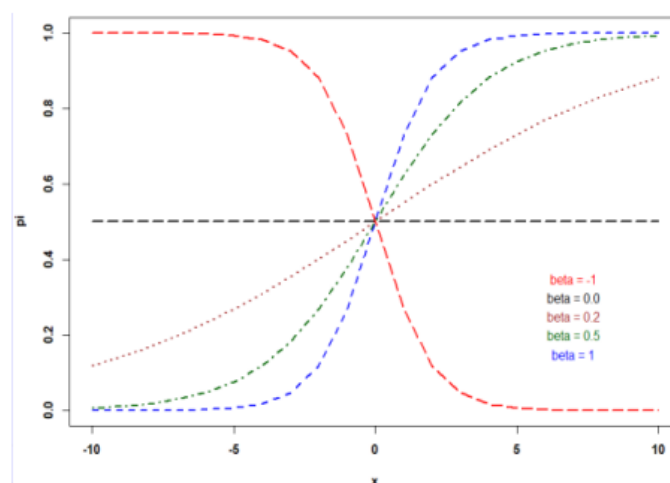
A Figura 1 mostra o comportamento de uma Regressão Logística, onde os valores da variável resposta sempre estão entre 0 e 1, independente das variáveis explicativas.

Figura 1: Curva da função de Regressão Logística



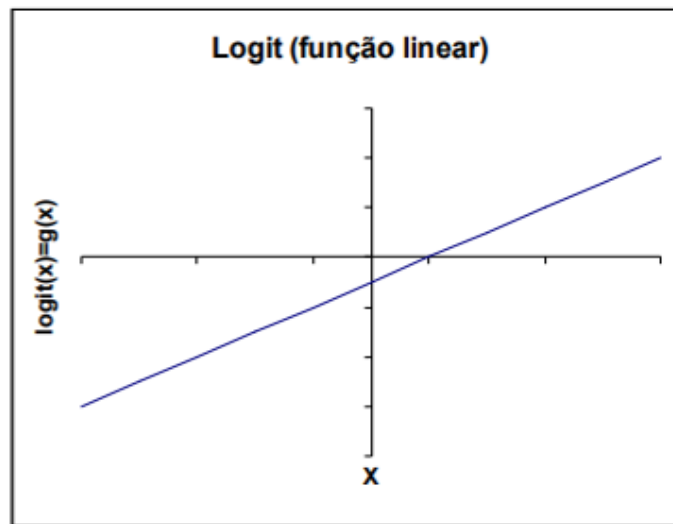
Fonte: Bittencourt (2012)

A Figura 2 mostra o comportamento de uma regressão logística para diferentes valores de β_i , observa-se que o valor do coeficiente causa efeito na chance de um evento ocorrer, dado que se o coeficiente for positivo aumenta a probabilidade de sucesso, enquanto que se for negativo, diminui essa probabilidade.

Figura 2: Curva da função de Regressão Logística para diferentes valores de β_i 

A Figura 3 expõe como a transformação logit modifica o comportamento do gráfico do modelo, possibilitando uma melhor visualização e interpretação mais clara de um estudo.

Figura 3: Curva da função de Regressão Logit



Fonte: Bittencourt (2012)

3.1 Modelo de Regressão Logística Múltipla

O Modelo de Regressão Logística Múltipla é extensão da Regressão Logística Simples, trata-se de quando duas ou mais variáveis independentes são incluídas no modelo para controlar o efeito de diferentes fatores no resultado.

A Regressão Logística Múltipla segue a mesma ideia da simples, onde tem que Y_i são as variáveis aleatórias independentes com distribuição de Bernoulli e possui como valor esperado:

$$E[Y_i] = \pi_i.$$

O Modelo de Regressão Logística Múltipla é definido como:

$$\pi_i = \frac{\exp(X'\beta)}{1 + \exp(X'\beta)},$$

considerando que:

$$X_i'\hat{\beta} = \beta_0 + \beta_1 X_{i1} + \dots + \beta_{p-1} X_{i,p-1},$$

sendo que X_i são os valores conhecidos das variáveis explicativas, os β_i são os coeficientes do modelo e $E[Y_i]$ é a média condicional de Y_i dados os valores de $X_{i,1}, X_{i,2}, \dots, X_{i,p-1}$.

Ao efetuar a transformação logito, encontra-se:

$$\pi'_i = \log(\pi) = \ln \left[\frac{\pi}{1 - \pi} \right],$$

sendo:

$$\pi'_i = X'_i \beta.$$

A função de probabilidade da Regressão Logística Múltipla ajustada é dada por:

$$\hat{\pi}_i = \frac{\exp[X'_i \hat{\beta}]}{1 + \exp[X'_i \hat{\beta}]},$$

sendo que:

$$X'_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \dots + \hat{\beta}_{p-1} X_{i,p-1}.$$

3.2 Estimação dos parâmetros do modelo

Na Regressão Logística, em ambos os modelos (simples e múltiplo), os coeficientes β_i são estimados utilizando o método de máxima verossimilhança, ou seja, busca encontrar os estimadores β que maximizam o logaritmo da função de verossimilhança a partir dos dados da amostra, ou seja, o conjunto de observações.

Supondo que Y_i são variáveis aleatórias independentes e identicamente distribuídas, têm-se que:

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

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

Sabendo que a função de verossimilhança é dada por :

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

e o valor de $\pi_i = \pi(X_i)$ é

$$\pi_i = \frac{\exp(X'_i \beta)}{1 + \exp(X'_i \beta)}.$$

Logo, tem-se:

$$\ln(L[\beta_0, \beta_1, \dots, \beta_p]) =$$

$$\sum_{i=1}^n Y_i (\beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip}) - \sum_{i=1}^n \ln [1 + \exp(\beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip})].$$

Para encontrar as estimativas de máxima verossimilhança é preciso derivar a expressão acima em relação a cada parâmetro do modelo e solucionar o sistema de equações resultantes. Como não há uma fórmula fechada para os valores de β que maximizam a função de máxima verossimilhança é necessário utilizar o método de Newton-Raphson para encontrar as estimativas.

3.3 Inferência no Modelo de Regressão Múltipla

As inferências realizadas no modelo de regressão linear são as mesmas de interesse no Modelo de Regressão Logística: coeficientes de regressão, estimativa de respostas médias e previsão de novas observações.

Com o modelo ajustado, é necessário testar a significância dos parâmetros estimados das respectivas variáveis explicativas utilizadas no modelo.

3.3.1 Teste da Razão da Verossimilhança

O Teste da Razão de Verossimilhança é um teste que compara a qualidade do ajuste, analisa se existe ou não regressão, ou seja, se todos os parâmetros β'_i são nulos, com exceção do β_0 .

As hipóteses a serem testadas são:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$$

$$H_1 : \exists \beta_i \neq \beta_j, \text{ para pelo menos um } i \neq j$$

A estatística do teste é dada por:

$$G^2 = -2(L_0 - L_1) \sim X^2_{(v)gl},$$

em que

- L_0 é o máximo do log da função de verossimilhança sob a H_0 ;
- L_1 é o máximo do log da função de verossimilhança sob a H_1 .

Para uma amostra de tamanho n grande, a estatística G^2 segue uma distribuição X^2 com v graus de liberdade em que v corresponde à diferença no número de parâmetros dos dois modelos L_0 e L_1 .

Dessa forma, a regra de decisão definida é:

$$\text{Se } G^2 \leq X^2(1 - \alpha; v) \text{ não rejeita-se } H_0,$$

$$\text{Se } G^2 > X^2(1 - \alpha; v) \text{ rejeita-se } H_0.$$

3.3.2 Teste de Wald

O Teste de Wald é utilizado para analisar cada coeficiente, de forma isolada, se é diferente de zero, portanto este teste é mais eficiente quando trata-se de um único parâmetro de regressão.

As hipóteses a serem testadas são:

$$\begin{aligned} H_0) \beta_i &= 0 \\ H_1) \beta_i &\neq 0 \end{aligned}$$

A estatística de teste é:

$$z^* = \frac{\widehat{\beta}_i}{s[\widehat{\beta}_i]} .$$

A regra de decisão:

Se $z^* \leq z(1 - \alpha/2)$ conclui-se H_0 ,

Se $z^* > z(1 - \alpha/2)$ conclui-se H_1 .

3.3.3 Intervalo de Confiança

O intervalo de confiança de um parâmetro β_i com confiança $(1-\alpha)$ é calculado como:

$$IC[\beta_i; 100(1 - \alpha)\%] = \widehat{\beta}_i \pm Z_{1-(\alpha/2)} s(\widehat{\beta}_i)$$

em que:

- $Z_{1-(\alpha/2)}$ é o percentil $\left(1 - \frac{\alpha}{2}\right) 100$ da distribuição Normal Padrão;
- $s(\widehat{\beta}_i)$ é a estimativa do erro padrão do estimador $\widehat{\beta}_i$.

3.4 Seleção do Modelo

A seleção de modelos possibilita escolher modelos com dimensionalidade menor, mas assegurando que as informações fornecidas são semelhantes ao modelo total.

Os critérios adaptados para a seleção de modelos da Regressão Logística, com o intuito de analisar a qualidade de um modelo estatístico, são:

- AIC

O critério de informação de Akaike (AIC) é onde é calculado pela expressão:

$$AIC = -\ln L(\hat{\beta}) + 2p.$$

Para análise, quanto menor o valor de AIC, mais adequado é o modelo.

- SBC

O critério de informação Bayesiano é calculado pela fórmula:

$$SBC = -\ln L(\hat{\beta}) + p \ln(n).$$

Para análise, assim como o AIC, quanto menor o valor de SBC, mais adequado é o modelo.

Para efetuar essa redução de dimensionalidade no modelo, existem diversos métodos de seleção de variáveis automáticos frequentemente utilizados. A escolha do método mais adequado deve levar em consideração a complexidade do modelo, sua interpretabilidade dos resultados e a capacidade de generalização dos resultados para novos dados. Há três métodos muito utilizados para determinar quais variáveis independentes devem ser incluídas no modelo, são eles:

- **Eliminação *Backward***: O processo inicia com todas as variáveis independentes incluídas no modelo, depois há a remoção de uma por uma, baseado em critérios estatísticos, como o p-valor do coeficiente de regressão. Ao final do processo, são selecionadas as variáveis independentes que têm p-valores inferiores ao critério de significância, portanto, são consideradas estatisticamente significantes para o modelo.
- **Seleção *Forward***: Diferentemente da Eliminação *Backward*, esta abordagem inicia o processo com nenhuma variável explicativa, a inclusão é efetuada pouco a pouco, considerando o menor p-valor ao comparado com um α já pré fixado.
- **Regressão *Stepwise***: Este procedimento é junção dos dois anteriores. O sistema inicia com nenhuma variável explicativa, mas a cada passo adiciona a variável mais significativa ou retira a menos significativa. Para tal decisão, fixa um α , se a variável apresentar um p-valor menor do que o α pré fixado, ela é incluída no modelo, simultaneamente as variáveis presentes no modelo são analisadas e se apresentarem um p-valor maior do que o α , são retiradas do modelo.

Em todos os métodos, o processo para, quando não há mais variáveis que atendem as condições exigidas para inclusão ou retirada do modelo.

3.5 Diagnóstico do Modelo

Nesta seção aborda-se a análise dos resíduos e a identificação dos casos influentes na regressão logística.

Entretanto, como os resíduos não são normalmente distribuídos e, de fato, sua distribuição sob a suposição de que o modelo ajustados está correto é desconhecida, torna-se a análise

de resíduos para Regressão Logística mais difícil ao comparar com Regressão Linear, visto que a variável resposta Y_i assume apenas os valores 0 e 1. E, conseqüentemente:

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

3.5.1 Resíduos de Pearson

Os Resíduos de Pearson estão diretamente interligados à estatística de qualidade de ajuste qui-quadrado de Pearson. Para isso, os Resíduos de Pearson são dados por:

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

3.5.2 Resíduos de Pearson Semistudentizado

Os Resíduos de Pearson Semistudentizados não possuem variância unitária, dado que não foi considerada a variação inerente no valor ajustado.

Uma forma aconselhada é dividir os resíduos ordinários pelo seu desvio padrão, onde encontra-se o valor aproximado de $\sqrt{\hat{\pi}_i(1 - \hat{\pi}_i)(1 - h_{ii})}$, sendo h_{ii} é o i -ésimo elemento da matriz diagonal $n \times n$ estimada para a regressão logística:

$$H = \widehat{W}^{1/2} X (X' \widehat{W} X)^{-1} X' \widehat{W}^{1/2},$$

sendo que \widehat{W} é a matriz $n \times n$, com elementos $\hat{\pi}_i(1 - \hat{\pi}_i)$. E, $\widehat{W}^{1/2}$ é uma matriz diagonal onde os elementos da diagonal são iguais a raiz quadrada daqueles em \widehat{W} .

Os Resíduos de Pearson Studentizados resultantes são definidos como:

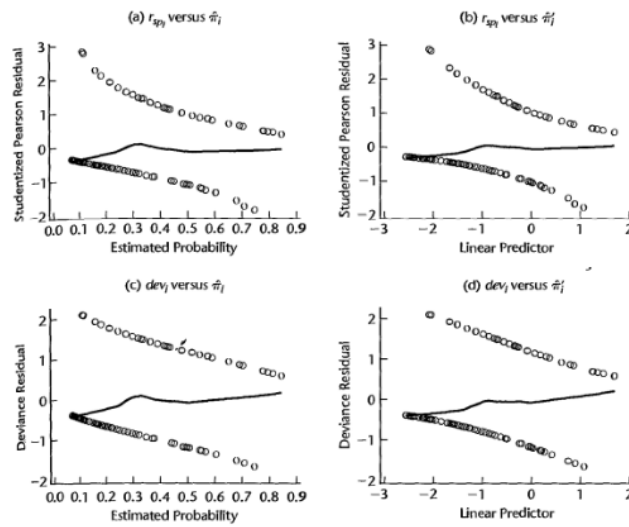
$$r_{SPi} = \frac{r_{pi}}{\sqrt{1 - h_{ii}}}.$$

3.5.3 Gráficos de Resíduos

Na regressão ordinária, os gráficos de resíduos são úteis para diagnosticar a inadequação do modelo, a variância não constante e a presença de valores discrepantes, enquanto que na Regressão Logística o enfoque é utilizar o gráfico de resíduos para detecção de inadequação do modelo.

O comportamento esperado é que ocorra uma leve suavização do gráfico de resíduos em relação a probabilidade ou resultar em uma linha horizontal com intercepto no zero, qualquer comportamento diferente insinua que o modelo não é adequado.

Figura 4: Exemplo gráficos de resíduos



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

3.6 Inferência sobre a resposta média

A resposta média de interesse é denotada pela expressão:

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

O estimador pontual da resposta média é dado por:

$$\widehat{\pi}_h = \left[1 + e^{-X'_h b} \right]^{-1},$$

sendo que b é o vetor de coeficientes de regressão estimados, isto é, $[\widehat{\beta}_0, \widehat{\beta}_1, \dots, \widehat{\beta}_{p-1}]$, e X é o vetor $[1, X_{h,1}, \dots, X_{h,p-1}]'$.

Os limites de confiança de amostra grande para a logística da resposta média são obtidos da maneira usual:

$$L = \widehat{\pi}'_h - z \left(1 - \frac{\alpha}{2} \right) s(\widehat{\pi}'_h),$$

$$U = \widehat{\pi}'_h + z \left(1 - \frac{\alpha}{2} \right) s(\widehat{\pi}'_h).$$

sendo que, L e U representam, respectivamente, os limites de confiança inferior e superior para n . E, $s(\widehat{\pi}'_h)$ é o erro padrão do estimador de $\widehat{\pi}'_h$.

3.7 Modelo de Regressão Logística Complexa

A análise de dados provenientes de pesquisas com amostras complexas, como aquelas que utilizam estratificação ou conglomerados, requer o uso de métodos estatísticos específicos para garantir a validade da inferência. A regressão logística, amplamente utilizada para modelar desfechos binários, pode ser adaptada para essas amostras complexas, mediante o uso de pesos amostrais. Esses pesos ajustam as estimativas para refletir corretamente a estrutura de amostragem, garantindo que os resultados sejam representativos da população-alvo (HOSMER; LEMESHOW; STURDIVANT, 2013). No processo de ajuste da regressão logística, a função de verossimilhança é modificada para incluir os pesos amostrais. A função de verossimilhança aproximada é descrita pela fórmula:

$$\sum_{k=1}^K \sum_{j=1}^{m_k} \sum_{i=1}^{n_{kj}} [w_{kji} \cdot y_{kji} \cdot \ln(\pi(x_{kji})) + w_{kji} \cdot (1 - y_{kji}) \cdot \ln(1 - \pi(x_{kji}))]$$

Nesta fórmula:

- w_{kji} é o peso amostral da observação i na unidade primária j do estrato k ;
- y_{kji} é o valor observado da variável dependente (geralmente 0 ou 1 em regressão logística);
- $\pi(x_{kji})$ é a probabilidade predita pela regressão logística com base nas covariáveis x_{kji}

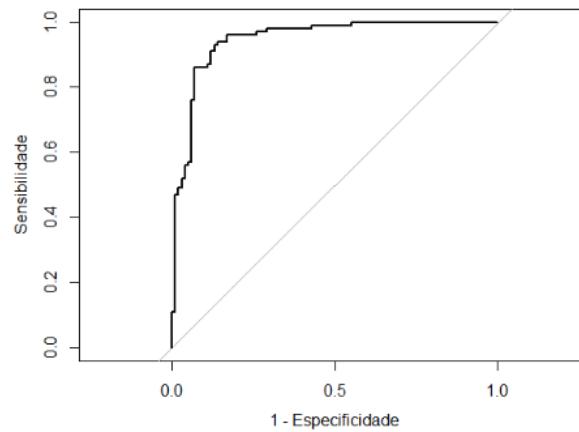
Essa modificação permite a obtenção de estimativas consistentes dos coeficientes de regressão, levando em consideração a complexidade da amostra. A correta interpretação das variâncias e covariâncias dos coeficientes exige a utilização de estimadores robustos e testes como o de Wald ajustado. Esses ajustes são essenciais para garantir a precisão das inferências e evitar viés nos resultados (KORN; GRAUBARD, 1990).

3.8 A Curva ROC

A Curva ROC (Receiver Operating Characteristic) é uma ferramenta amplamente utilizada para avaliar o desempenho de modelos de classificação binária em contextos estatísticos. Ela representa graficamente a relação entre a sensibilidade (taxa de verdadeiros positivos) e especificidade (taxa de falsos positivos), o que facilita a escolha do ponto de corte que melhor equilibra a detecção de verdadeiros positivos e a minimização de falsos positivos.

A Curva de ROC é representada graficamente da seguinte forma:

Figura 5: Exemplo do gráfico da Curva ROC



Fonte: Agresti, An Introduction to Categorical Data Analysis (p.112)

Conforme Agresti (2007), o modelo pode ser interpretado, dado o valor da Área Sob a Curva (AUC) da Curva ROC, da seguinte forma:

- Se $AUC=0.5$: não tem poder discriminatório;
- Se $0.7 \leq AUC < 0.8$: desempenho aceitável;
- Se $0.8 \leq AUC < 0.9$: desempenho excelente;
- Se $AUC \geq 0.9$: desempenho excepcional.

—

4 Metodologia

4.1 Banco de dados

A PDAD 2021 coletou informações detalhadas de uma amostra representativa de domicílios por meio de amostragem aleatória sistemática, abrangendo as 33 regiões administrativas do Distrito Federal e outros locais de interesse da Codeplan, totalizando 43 localidades (Governo do Distrito Federal Secretaria de Estado de Fazenda, Planejamento, Orçamento e Gestão do Distrito Federal - SEFP, 2022).

A coleta de dados foi efetuada entre maio e dezembro de 2021, com a data de referência da pesquisa sendo 1º de julho de 2021, por intermédio de coletor eletrônico (Governo do Distrito Federal Secretaria de Estado de Fazenda, Planejamento, Orçamento e Gestão do Distrito Federal - SEFP, 2022). Essas informações, além de fornecer uma visão detalhada da situação atual da população, a pesquisa oferece subsídios importantes para a formulação de políticas públicas e programas sociais. Ao analisar fatores como acesso a serviços e condições habitacionais, a PDAD 2021 contribui para o desenvolvimento de estratégias que visam melhorar as condições de vida e reduzir desigualdades na região ((IBGE), 2021).

Esta pesquisa coletou dados abrangendo áreas como renda, educação, saúde e condições de moradia que são cruciais para identificar desigualdades e padrões que afetam a qualidade de vida dos residentes (SILVA, 2021a) e oferecem subsídios importantes para a formulação de políticas públicas e programas sociais.

Com base neste banco de dados da PDAD 2021, busca-se analisar os fatores que influenciam a conclusão ou não do ensino médio na população do Distrito Federal. Considerando os objetivos do estudo foram extraídos os dados referentes as mulheres na faixa etária de 14 a 24 anos.

A escolha de analisar as mulheres na faixa etária de 14 a 24 anos visa compreender o impacto da maternidade na conclusão do ensino médio. Esse grupo etário é crítico, pois abrange o período de transição e conclusão do ensino médio, entretanto houve ampliação da faixa devido a problemas de retenção.

4.2 Variáveis

4.2.1 Variáveis originais

A Pesquisa de Desenvolvimento e Acompanhamento de Dados (PDAD) fornece informações abrangentes sobre aspectos socioeconômicos e demográficos, incluindo renda, nível de escolaridade, participação no mercado de trabalho, saúde, e transporte. Os dados permitem avaliar condições econômicas e de vida, como a conclusão de níveis educacionais, o uso de recursos

para educação, o acesso a serviços de saúde e de mobilidade dos indivíduos.

A análise realizada neste estudo se concentrou em variáveis como raça/cor, sexo, renda familiar e maternidade, com o objetivo de identificar padrões e associações que possam revelar como esses fatores afetam a conclusão educacional. Sendo assim, as variáveis consideradas neste estudo são:

- **Dados Pessoais:** Idade, Sexo, Raça/cor, Estado civil, Local de nascimento, Local de moradia (RA);
- **Informações Familiares:** Se mora/reside em companhia do cônjuge, Filho nascido vivo, Número de filhos, Número de filhos residentes no domicílio;
- **Condição de Vida e Saúde:** Condição no domicílio, Dificuldade, Motivo principal de atendimento de saúde, Possui plano de saúde;
- **Educação:** Nível de escolaridade, Nível mais alto de ensino, Turno de ensino, Se frequenta curso;
- **Trabalho e Transporte:** Trabalhou nos últimos 30 dias, Procurou trabalho no último mês/último ano, Quantidade de horas no trabalho, Local do trabalho, Meio de transporte para a unidade de ensino ou para o trabalho, Tempo de deslocamento para unidade de ensino ou para o trabalho;
- **Uso da Internet:** Uso da internet, Motivo do uso da internet.

No contexto da Pesquisa Distrital por Amostra de Domicílios (PDAD) 2021, a variável "peso" desempenha um papel crucial na análise dos dados. O "peso" é um ajuste amostral utilizado para assegurar que as estimativas derivadas dos dados reflitam com precisão a estrutura da população do Distrito Federal. Ele corrige desequilíbrios na amostra que podem surgir devido a diferentes probabilidades de seleção ou taxas de resposta, garantindo a representatividade dos resultados ((CODEPLAN), 2021).

A aplicação adequada do "peso" é essencial para a integridade das análises estatísticas. Sem o ajuste proporcionado por essa variável, as estimativas podem ser enviesadas, comprometendo a qualidade das conclusões. Portanto, o "peso" é vital para garantir a precisão e a relevância dos resultados obtidos a partir da PDAD 2021 (SILVA, 2021b). Assim, a análise descritiva detalhada e a utilização correta da variável "peso" são indispensáveis para assegurar que os dados sejam interpretados corretamente e que as conclusões sejam válidas e aplicáveis.

4.2.2 Variáveis criadas

A criação de novas variáveis em um estudo é uma estratégia importante para enriquecer a análise de dados e tornar os resultados mais robustos e significativos. Segundo Hair et al.

(2014), novas variáveis podem ser derivadas a partir de combinações, transformações ou agrupamentos de variáveis existentes, o que permite uma representação mais fiel e direta do fenômeno em estudo.

Montgomery (2013) destaca que essas novas variáveis podem simplificar a análise e melhorar a interpretação dos resultados, ao permitir que relações complexas entre os dados sejam mais facilmente identificadas (FIELD, 2018). Além disso, Triola (2017) observa que a criação de variáveis pode aumentar a precisão dos modelos estatísticos, revelando padrões que não seriam visíveis nas variáveis originais.

Em síntese, as seguintes variáveis foram criadas a partir da reestruturação ou da combinação de informações das variáveis originais do banco tendo em vista os objetivos do estudo:

- **CNH**

A variável "CNH" indica a posse da Carteira Nacional de Habilitação, sendo categorizada como "Sim" para indivíduos que possuem e "Não" para aqueles que não a possuem. A categorização binária da variável "CNH" simplifica a análise, permitindo que se avalie a relação entre a posse da CNH e outras variáveis de interesse de maneira mais direta. Essa abordagem também facilita a compreensão dos dados e a identificação de padrões e tendências relevantes.

- **Condição**

Nos dados da PDAD, a condição do domicílio é um fator relevante para a análise da conclusão do ensino médio entre mulheres. Para otimizar a interpretação e alinhar a análise com o objetivo do estudo, que é determinar os fatores que influenciam a conclusão do ensino médio, simplificamos as categorias de condição do domicílio. Assim, foram criadas as categorias "pessoa responsável", "filha do responsável" e "filha do responsável e do cônjuge". Essa simplificação permite uma abordagem mais focada e clara, facilitando a análise dos dados e ajudando a identificar como diferentes configurações familiares podem afetar a conclusão do ensino médio. Ao reduzir a complexidade das categorias, buscamos uma compreensão mais precisa das variáveis envolvidas e seu impacto na educação das mulheres.

- **Dificuldade**

A variável "dificuldade" foi criada para consolidar as informações sobre as limitações permanentes nos indivíduos presentes na PDAD 2021 em uma única medida, onde classifica-se as dificuldades com base nas respostas às perguntas sobre problemas auditivos, motores, comunicativos e de manipulação de objetos. A especificação é feita nas categorias "tem dificuldade" e "não tem dificuldade".

- **Estado civil**

Nos dados da PDAD 2021, o estado civil das pessoas foi agrupado em três categorias para melhorar a interpretação e visualização: "solteira", "casada" e "outros". A categoria

”outros” inclui divorciados, separados judicialmente, desquitados, viúvos e aqueles que não souberam informar seu estado civil. Essa simplificação oferece uma visão mais clara e direta sobre o estado civil dos indivíduos, o que contribui para uma análise e interpretação mais eficiente dos dados em relação a outras variáveis.

• Grupo idade

Para aprimorar a análise dos dados relativos à faixa etária dos indivíduos, foi desenvolvida a variável ”grupo_idade”, que categoriza as idades em intervalos específicos: abaixo de 13 anos, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 e acima de 24 anos. Esta categorização visa proporcionar uma visão estruturada e detalhada da distribuição etária, permitindo uma análise mais precisa das características e comportamentos associados a cada faixa etária. A segmentação em grupos etários facilita a identificação de padrões e tendências, contribuindo para uma interpretação mais robusta e informada dos dados.

Tabela 1: Distribuição de Frequência e Percentual por Grupo Etário da PDAD 2021

Grupo de Idade	Frequência	Percentual
0-13 anos	540772	17,96
14 anos	41817	1,39
15 anos	49245	1,64
16 anos	41884	1,39
17 anos	50473	1,68
18 anos	46392	1,54
19 anos	43127	1,43
20 anos	48462	1,61
21 anos	47351	1,57
22 anos	48383	1,61
23 anos	53722	1,78
24 anos	49383	1,64
>24 anos	1949870	64,76

• Local por renda

A variável ”local por renda” foi criada para categorizar diferentes regiões com base na faixa de renda predominante, agrupando-as em quatro categorias principais: Renda Alta, Renda Média-alta, Renda Média-baixa e Renda Baixa. Essa categorização permite uma análise mais detalhada e segmentada da distribuição de renda, ajudando a identificar padrões e diferenças econômicas significativas entre as regiões. Assim, é possível avaliar como a renda afeta outras variáveis de interesse, como o nível de escolaridade e o acesso a serviços, além de facilitar a elaboração de políticas públicas e estratégias de desenvolvimento regional

mais direcionadas e eficazes.

No estudo da Codeplan (2022), a Tabela 2 apresenta a categorização das regiões por faixa de renda. A informação foi adaptada da apresentação disponível online (CODEPLAN, 2022).

Tabela 2: Categorias de Renda e Regiões

Categoria de Renda	Regiões
Renda Alta	Águas Claras, Jardim Botânico, Lago Norte, Lago Sul, Park Way, Plano Piloto, Sudoeste/Octogonal
Renda Média-alta	Arniqueira, Candangolândia, Cruzeiro, Guará, Núcleo Bandeirante, SIA, Sobradinho, Taguatinga, Vicente Pires
Renda Média-baixa	Ceilândia, Gama, Riacho Fundo, Samambaia, Santa Maria, Sobradinho II
Renda Baixa	Brazlândia, Fercal, Itapoã, Paranoá, Planaltina, Recanto das Emas, Riacho Fundo II, São Sebastião, SCIA-Estrutural, Sol Nascente/Pôr do Sol, Varjão

Fonte: Adaptado de Codeplan (2022). *Apresentação PDAD-DF 2021*. Recuperado de https://www.codeplan.df.gov.br/wp-content/uploads/2022/05/Apresentacao_PDAD-DF_2021.pdf

- **Meios de transporte**

A variável meio de transporte foi criada para classificar os diferentes meios de locomoção utilizados pelos alunos para se deslocarem até a escola. Esta variável foi segmentada em três categorias: pessoal, público e privado. Dado que a categoria pessoal inclui a pé e bicicleta, enquanto público refere-se a transportes coletivos fornecidos pelo governo ou instituições públicas, ônibus, ônibus escolar público e metrô. A classe privado engloba meios de transporte como automóvel, motocicleta, ônibus escolar privado e carros de aplicativo.

- **Mora em companhia**

A criação da variável "mora em companhia" é essencial para o estudo dos fatores que influenciam a conclusão do ensino médio, com ênfase na maternidade. Essa variável identifica se o indivíduo reside com um cônjuge ou companheiro, informação relevante para entender o suporte disponível. A presença de um cônjuge ou rede de apoio pode oferecer suporte adicional para mulheres que são mães, ajudando a equilibrar as responsabilidades famili-

ares e acadêmicas. Agrupar as respostas em "Sim" e "Não" permite uma análise clara da influência dessa rede de apoio na capacidade das mães de concluir o ensino médio, fornecendo insights valiosos sobre como o suporte familiar pode impactar o sucesso educacional.

- **Nível de escolaridade**

Embora o banco de dados da PDAD ofereça informações detalhadas sobre escolaridade, a criação da variável "nível de escolaridade", que classifica os indivíduos conforme a conclusão ou não do ensino médio, é essencial para este estudo. Esta variável foi desenvolvida para focar especificamente no marco da conclusão do ensino médio, facilitando a análise do impacto dessa educação sobre outras variáveis socioeconômicas. A utilização dessa variável aprimora a precisão da análise ao simplificar a interpretação dos dados e fornecer uma medida clara sobre o status educacional dos indivíduos. Assim, contribui significativamente para a avaliação dos efeitos da educação no contexto do estudo, alinhando a análise com os objetivos da pesquisa.

Tabela 3: Distribuição do nível de escolaridade da PDAD 2021

Nível de Escolaridade	Frequência	Percentual
Concluiu o ensino médio	1583787	54,5
Não concluiu o ensino médio	1320325	45,5

- **Número de filhos**

O banco de dados inclui informações sobre "quantidade de filhos homens nascidos vivos" e "quantidade de filhas mulheres nascidas vivas". A partir dessas informações, foi criada a variável "número de filhos", que é obtida pela soma das quantidades de filhos e filhas nascidos vivos. Essa abordagem permite uma análise mais consolidada da fecundidade, oferecendo um total agregado de filhos por indivíduo. A nova variável "número de filhos" facilita a interpretação dos dados e proporciona uma visão mais clara e eficiente da distribuição de filhos na amostra.

- **Número de filhos que moram no domicílio**

Nos dados presentes na PDAD, há as informações sobre a "quantidade de filhos homens que moram no domicílio" e a "quantidade de filhas mulheres que moram no domicílio", a partir disso formulou-se a variável "número de filhos residentes", que é obtido pela soma dessas duas variáveis referentes a filhos moradores. A criação e o agrupamento da variável "número de filhos que moram no domicílio" oferecem uma análise mais detalhada e integrada da estrutura familiar, facilitando a interpretação dos dados e fornecendo insights significativos sobre a dinâmica familiar atual.

- **Raça cor**

Para simplificar a análise da diversidade racial e étnica, os indivíduos foram classificados em duas categorias principais: "negro" e "não negro". A categoria "negro" inclui aqueles que se identificam como pardos e pretos, enquanto "não negro" abrange indivíduos identificados como brancos e amarelos. Essa abordagem oferece uma análise mais clara das desigualdades raciais e étnicas, permitindo avaliar o impacto dessas variáveis na conclusão do ensino médio e em outros aspectos do estudo. A simplificação proporciona uma visão mais integrada e compreensiva das diferenças raciais na amostra.

- **Saúde**

A variável 'saúde' foi criada para proporcionar visão detalhada dos diversos aspectos relacionados ao cuidado e à condição de saúde dos indivíduos. Com o objetivo de capturar a diversidade de situações e intervenções relacionadas à saúde, essa variável é subdividida em três categorias principais:

- Problema de saúde: inclui variedade de situações e condições que demandam cuidados médicos específicos, neste tópico são consideradas ocorrências como acidentes, lesões, fraturas e doenças, bem como a continuação de tratamentos para doenças anteriores, problemas odontológicos e processos de reabilitação, que englobam fisioterapia, fonoaudiologia e terapia ocupacional.
- Saúde materna: foca nas necessidades e cuidados específicos durante a gestação, isto é, o pré natal e o parto.
- Prevenção: destina-se a atividades voltadas à antecipação e prevenção de problemas de saúde. Inclui ações como exames complementares de diagnóstico, vacinação, medidas preventivas gerais, solicitação de atestados de saúde e acompanhamento com profissionais de saúde, como psicólogos, nutricionistas ou outros especialistas.

- **Soma dos rendimentos individuais**

As variáveis soma dos rendimentos individuais e soma dos rendimentos individuais corrigidos, representados por "renda_ind" e "renda_ind_r", originalmente em formato categórico, foram convertidas para o formato numérico. Esta transformação é fundamental para a aplicação de técnicas quantitativas e para a obtenção de insights precisos sobre a distribuição e a desigualdade de renda entre os indivíduos. A variável numérica resultante permite uma abordagem mais detalhada e precisa nas análises subsequentes.

- **Turno de ensino**

A variável turno de ensino foi criada para categorizar a jornada educacional com base no horário de frequência das aulas. Esta variável foi dividida em duas categorias: diurno (matutino, vespertino e integral) e noturno.

4.3 Método

No presente trabalho, o objetivo é, através de um modelo de regressão logística, identificar fatores que estão associados à conclusão do ensino médio das alunas do gênero feminino, prevalecendo a variável maternidade.

Neste estudo, a variável de interesse é a conclusão do "Ensino Médio", que é uma variável qualitativa e dicotômica, significando que ela pode assumir apenas dois resultados possíveis: "sim" ou "não".

Além disso, a amostra da PDAD utilizada neste estudo foi selecionada utilizando um plano amostral complexo, conforme mencionado anteriormente, portanto optou-se pelo uso de um modelo de regressão logística binária ajustado para amostragem complexa.

Todas as análises realizadas no desenvolvimento do trabalho foram executadas utilizando o *software* SAS.

5 Resultados

5.1 Análise Descritiva

A análise descritiva de um banco de dados é uma etapa fundamental para compreender as características da amostra e garantir a validade das análises subsequentes. Essa fase inicial fornece visão abrangente das tendências e padrões presentes nos dados. Após essa análise preliminar, realiza-se seleção criteriosa da amostra de interesse, focando em subgrupos específicos que possibilitam aprofundamento analítico mais direcionado. Essa abordagem permite interpretação mais rica e relevante dos resultados, considerando as nuances e particularidades da amostra escolhida.

5.1.1 PDAD 2021

A Pesquisa Distrital por Amostra de Domicílios (PDAD)2021 oferece visão abrangente sobre a população do Distrito Federal, a qual inclui aspectos cruciais como educação, saúde e condições socioeconômicas. Esses dados são essenciais para compreender problemas sociais, como a investigação de fatores que influenciam a conclusão do ensino médio, fornecendo subsídios valiosos para a formulação de políticas públicas eficazes e a implementação de intervenções direcionadas, com o objetivo de melhorar a qualidade de vida e reduzir desigualdades.

A análise da distribuição de renda do Distrito Federal revela disparidade significativa. As rendas de maior poder aquisitivo representam uma porção menor da população, com 16,6% a renda média alta e 20,5% da renda alta. Em contrapartida, a maioria está concentrada nas faixas de renda baixa (25,5%) e média baixa (37,4%). Os dados revelam associação significativa entre renda e nível de escolaridade. Observa-se que 15,2% dos indivíduos com renda alta concluíram o ensino médio, enquanto 10,7% dos de renda baixa o fizeram. Entre os grupos de renda média baixa e média alta, as taxas de conclusão são 17,9% e 10,7%, respectivamente. A não conclusão é mais frequente entre os de renda média baixa(19,4%) e baixa (14,6%), em comparação com o de renda alta (5,6%) e média alta (5,9%). Essa desigualdade reflete diferenças nas condições de vida, que podem impactar diretamente diversos aspectos, incluindo o acesso e a permanência no sistema educacional.

A distribuição por sexo mostra que 52,2% da população é composta por mulheres e 47,8% por homens. Além disso, as mulheres possuem idade média de 35,1 anos ligeiramente superior à dos homens, de 32,9 anos. A mediana de idade segue a mesma tendência, sendo 35 anos para as mulheres e 32 para os homens. Essa diferença na estrutura etária pode influenciar tanto nas dinâmicas familiares quanto o acesso a oportunidades educacionais e profissionais.

A respeito da presença de dificuldade, 93,3% da população não relata problemas significativos, enquanto 6,7% menciona algum grau de dificuldade. Quanto ao estado civil, 46,5% da população é casada, 43,5% solteira e 10% pertence a outras categorias. Na área de saúde, 48,9%

está envolvida em atividades de prevenção, 46,6% lida com problemas de saúde e 1% foca na saúde materna, revelando atenção menor nesse último aspecto, o que pode refletir na necessidade de políticas específicas para grupos mais vulneráveis, como mães jovens.

No banco de dados da PDAD 2021, observa-se que 54,5% da amostra concluiu o ensino médio, enquanto 45,6% não completaram essa etapa educacional. Ao analisar fatores que influenciam a conclusão do ensino médio, com uma razão de chance de 1,038, a idade exerce efeito positivo significativo sobre a conclusão, isto é, à medida que a idade avança, a chance de conclusão aumenta. Os dados ressaltam também que, aqueles que não concluíram essa etapa apresentam idade média tardia: cerca de 24 anos para os homens e 28 para as mulheres, ao comparar com a idade prevista de finalização, 17 a 19 anos. Em contraste, o número de filhos tem efeito negativo, com uma razão de chance de 0,598, indicando que o aumento no número de filhos rediz a probabilidade de conclusão do ensino médio.

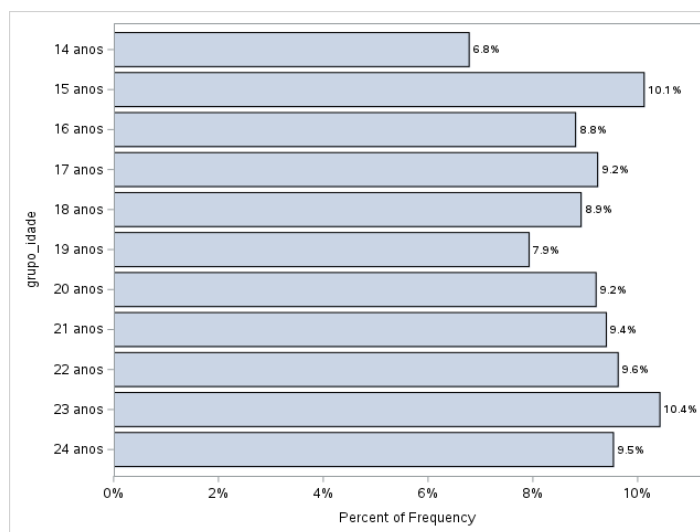
A análise realizada aponta indícios claros de que fatores como idade, filhos e condições socioeconômicas exercem influência significativa na conclusão do ensino médio. Esses resultados sugerem a necessidade de aprofundar a pesquisa sobre como essas e outras variáveis interagem e afetam a trajetória educacional dos indivíduos.

5.1.2 Análise Descritiva da amostra do presente estudo

Dado que o objetivo é compreender o perfil das mulheres em relação à conclusão do ensino médio com foco na maternidade, selecionou-se a faixa etária de 14 a 24 anos por ser um período crucial para a conclusão desse nível de ensino e para a determinação de futuras trajetórias educacionais e profissionais. Com isso, realiza-se uma análise descritiva para explorar as características da amostra e identificar padrões que possam influenciar a conclusão ou não do ensino médio.

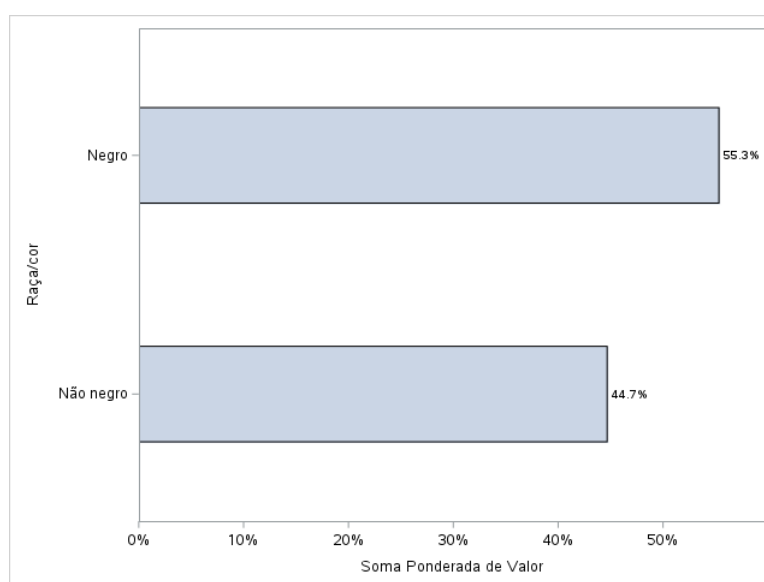
A média de idade da amostra é de 19,12 anos, o que sugere que grande parte dos indivíduos se encontra na fase final do ensino médio ou já completou essa etapa educacional. Esse valor reflete um perfil etário próximo da transição entre o ensino médio e a entrada em níveis de formação superiores ou no mercado de trabalho, destacando a predominância de jovens adultos no conjunto da amostra. A Figura 6 mostra uma distribuição relativamente uniforme entre as idades de 14 e 24 anos, com picos notáveis nas idades de 15 (10,24%) e 23 anos (10,29%), enquanto as menores frequências são observadas aos 16 (7,97%) e 19 anos (7,70%). Associado a isso, a análise da posse de carteira de habilitação (CNH) revela que 68,02% das participantes não possuem CNH. A predominância de jovens sem CNH é consistente com a idade média da amostra, uma vez que muitas ainda estão abaixo da idade mínima para obter a habilitação.

Figura 6: Distribuição das mulheres de 14 a 24 anos segundo faixa etária; Distrito Federal, Brasil - 2021



A composição da amostra quanto à raça/cor mostra que 55,33% dos indivíduos se identificam como "Negro", ou seja, as que se declaram preta ou parda. Por outro lado, 44,67% são classificadas como "Não negro", incluindo as categorias branca e amarela. A Figura 7 ilustra visualmente essa distribuição percentual, destacando a proporção de cada grupo racial na amostra.

Figura 7: Distribuição de mulheres de 14 a 24 anos segundo raça cor; Distrito Federal - 2021



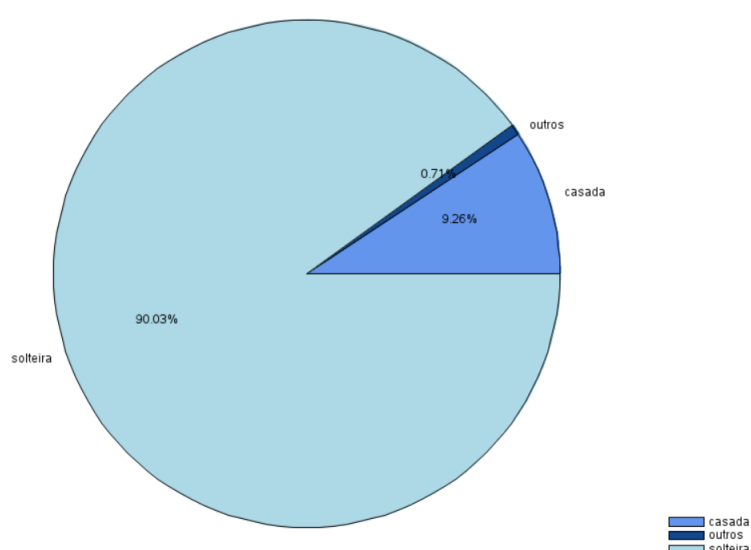
Esta distribuição revela predominância significativa de indivíduos pardos e brancos, como exemplificado na Tabela 4. Esses dados oferecem panorama considerável da diversidade racial presente na amostra, permitindo análise mais relevante sobre as possíveis desigualdades sociais e de acesso a recursos e oportunidades.

Tabela 4: Distribuição de mulheres de 14 a 24 anos, segundo raça/cor; Distrito Federal, Brasil - 2021

Raça/cor	Percentual
Negra	
Parda	45,12
Preta	10,18
Não negra	
Branca	43
Amarela	1,7

A análise dos dados revela que, entre as mulheres de 14 a 24 anos, 90,03% são solteiras, conforme indicado na Figura 8. Esse resultado é esperado, já que muitas priorizam a educação e o desenvolvimento pessoal. Além disso, cerca de 87,9% das participantes não residem com um cônjuge, o que reforça a suposição de que, nesta faixa etária, a prioridade é construir uma base sólida para o futuro. Esse padrão é característico da juventude, na qual o foco está em metas acadêmicas e profissionais, precedendo a formação de vínculos conjugais ou familiares mais duradouros. Os dados revelam que 87,8% dos indivíduos não compartilham a residência com uma segunda pessoa, enquanto 12,2% vivem com alguém.

Figura 8: Gráfico de setores da distribuição de mulheres de 14 a 24 anos segundo estado civil; Distrito Federal, Brasil - 2021



A análise da condição no domicílio revela que a maior parte dos indivíduos na amostra é composta por filhas do responsável e do cônjuge (35,9%), seguidos por aquelas que são filhas apenas do responsável (23,3%), indicando que muitos jovens ainda dependem diretamente de um dos responsáveis pelo domicílio. Netas, irmãs e outros parentes refletem a diversidade de arranjos familiares, representando 27,8% da amostra. Destaca-se que 12,9% são pessoas responsáveis pelo domicílio, sugerindo um núcleo familiar estabelecido em parte da amostra. Essa composição evidencia diferentes dinâmicas familiares e de convivência que podem influenciar a trajetória

educacional das jovens, sobretudo no contexto de conclusão do ensino médio.

Com um total de 259137 mulheres, a amostra revela uma distribuição variada com relação as Regiões Administrativas (RAs) do Distrito Federal. Ceilândia, Samambaia e Taguatinga destacam-se como as regiões de maior concentração de mulheres correspondendo a 11,8%, 8,3% e 6,1%, respectivamente. No entanto, regiões de maior renda, como Plano Piloto e Lago Sul, possuem uma representação menor de jovens. Assim, observa-se que a amostra é composta sobretudo por mulheres residentes em áreas de renda média-baixa e baixa, como apresentado na Tabela 5.

Tabela 5: Distribuição das mulheres de 14 a 24 anos segundo faixa de renda - Distrito Federal, Brasil - 2021

Local por Renda	Frequência	Percentual
Renda Alta	42158	16,3
Renda Baixa	76532	29,5
Renda Média Alta	38448	14,8
Renda Média Baixa	101999	39,4

Fonte: Elaboração dos autores com base em: Codeplan, PDAD – 2021.

A análise do local de nascimento e da mobilidade dos residentes revela padrões significativos. A maioria da população nascida no Distrito Federal (79,9%) mostra uma forte ligação com a região. Em contraste, 20,1% dos residentes nasceram fora do Distrito Federal, com os principais países de origem sendo Brasil, Colômbia e Venezuela. Além disso, uma parte considerável dos indivíduos (78,3%) sempre residiu no Distrito Federal, destacando a estabilidade na região. Entre aqueles que mudaram de local, a principal razão para a mudança foi a busca por trabalho (12,8%), seguida por motivos relacionados à moradia (11%) e acompanhamento de parentes (8,6%). Essa distribuição indica que a mobilidade para o DF é majoritariamente impulsionada por laços familiares e oportunidades de emprego e educação, ressaltando a importância desses fatores na atração de novas moradoras para a região.

Quando considera-se a intenção de formar novo domicílio no DF nos próximos 12 meses, 7,7% das entrevistadas manifestaram essa vontade. As localidades com maior intenção de mudança incluem áreas com grande potencial de desenvolvimento e expansão. Esses dados oferecem visão abrangente sobre a dinâmica populacional no Distrito Federal, refletindo tanto a estabilidade quanto as motivações para a mobilidade na região.

A maioria das residentes acessou a internet nos últimos três meses, cerca de 97,3% relata o uso, sendo 95,9% com acesso diário. Em termos de dispositivos usados, a maioria acessa a internet por meio de telefones celulares ou tablets (96,2%), enquanto 52,5% a utilizam através de microcomputadores. Além disso, 38,5% dos usuários recorrem a outros meios para acessar a internet dentro de casa.

O uso da internet para diferentes fins demonstra ampla aplicação das tecnologias di-

gitais. Como exposto na Tabela 6, educação e cursos são os principais motivos para o uso da internet, com 75,8% dos usuários envolvidos nesse tipo de atividade. Informação e notícias seguem como uma prioridade, com 85,4% dos residentes acessando a internet para essas finalidades. Além disso, 71,3% dos usuários a utilizam para criação e compartilhamento de conteúdos, enquanto 87,1% a utilizam para lazer e cultura. A comunicação é uma função ainda mais predominante, com 90,6% dos residentes empregando a internet para esse propósito. O uso para transações financeiras e comerciais é significativo, com 49,7% dos usuários recorrendo à internet para essas atividades.

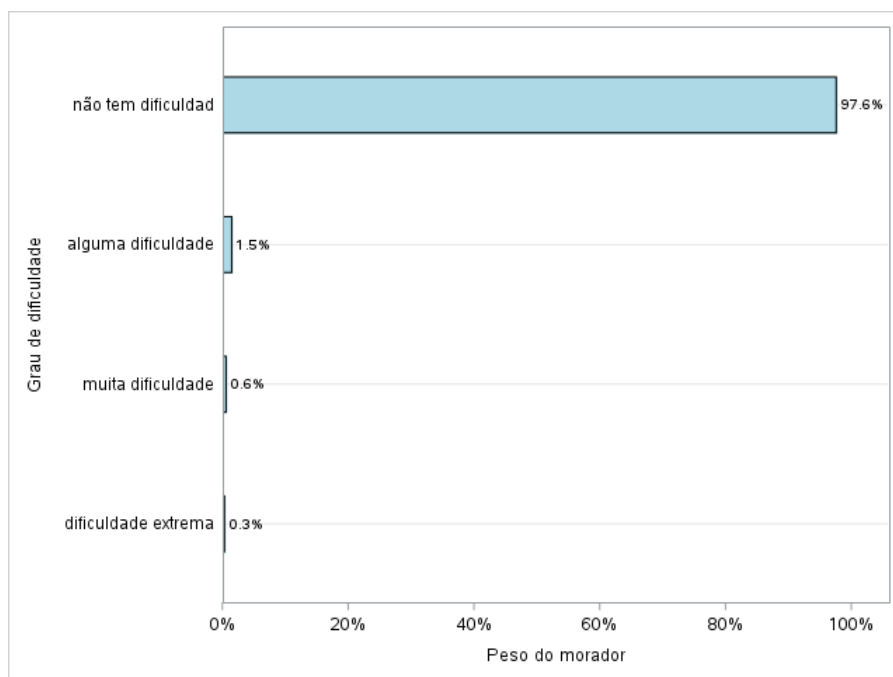
Tabela 6: Distribuição das mulheres de 14 a 24 anos segundo uso da internet; Distrito Federal, Brasil - 2021

Uso da internet	Percentual
Criação/compartilhamento de conteúdos	71,3
Comunicação	90,6
Educação/cursos	75,8
Informações/notícias	85,4
Lazer/cultura	87,1
Trabalho	49,7
Transações financeiras/comerciais	35,7

Esses dados evidenciam a importância da internet no cotidiano das mulheres em estudo, tanto para fins profissionais quanto pessoais, refletindo a alta integração da tecnologia nas atividades diárias.

O diagnóstico das dificuldades enfrentadas pelas mulheres na amostra revela um panorama significativo. A maioria, 98%, não reporta dificuldades severas, indicando que, em geral, os desafios não são de magnitude grave. Entre as mulheres que relataram algum nível de dificuldade, 1,5% enfrentam desafios moderados e 0,9% lidam com limitações severas. A presença dessas dificuldades, mostrada na Figura 9, reflete os desafios diários das mulheres, o que impacta na capacidade de concluir o ensino médio e possivelmente outros aspectos da vida.

Figura 9: Distribuição das mulheres de 14 a 24 anos segundo dificuldade; Distrito Federal, Brasil - 2021.



Como explicitado na Tabela 7, dificuldades graves em visão e audição são raras, com a maioria das participantes relatando ausência de desafios significativos nessas funções. A mobilidade também não apresenta problemas graves para a maioria, que mantém uma boa capacidade de locomoção. Além disso, a comunicação é eficaz para a maioria dos indivíduos, e a habilidade de pegar pequenos objetos está amplamente preservada. As dificuldades permanentes são relativamente raras e, quando presentes, em geral não são de gravidade extrema. A alta porcentagem de indivíduos sem dificuldades permanentes sugere uma prevalência geral de boa saúde e funcionalidade na amostra analisada.

Tabela 7: Distribuição das mulheres de 14 a 24 anos, segundo dificuldades relatadas; Distrito Federal, Brasil - 2021.

Aspecto	Dificuldade Relatada	Percentual
Visão	Incapacidade total	0,2
	Muita dificuldade	1,5
	Sem problemas severos	88,3
Audição	Dificuldades severas	0,3
	Alguma dificuldade	0,7
	Sem problemas auditivos	99
Mobilidade	Grandes dificuldades	0,3
	Alguma dificuldade	0,7
	Sem problemas permanentes	99
Comunicação	Dificuldades permanentes	1,1
	Sem problemas	98,9
Capacidade de pegar pequenos objetos	Dificuldade significativa	0,3
	Alguma dificuldade	0,5
	Sem problemas	99,2

Os dados sobre atendimento em saúde descreve a variedade de opções do primeiro serviço procurado pelas mulheres quando necessitam de cuidados médicos. A maior parte, que corresponde a 41,4%, busca inicialmente o Posto de Saúde ou Unidade Básica de Saúde. A farmácia é a primeira escolha para 15% das mulheres, enquanto 5% recorrem às Unidades de Pronto Atendimento (UPA) e 7% aos Prontos-Socorros de hospitais públicos. Serviços especializados e clínicas privadas também são opções significativas, com 10,6% das mulheres procurando consultórios particulares e 0,9% optando por centros de especialidades do SUS. Outros serviços, como ambulatórios e atendimento no domicílio, representam porcentagens menores, indicando uma preferência por serviços amplamente disponíveis e de fácil acesso.

A tabela mostra que a maioria das consultas é motivada por problemas de saúde (48,9%) e prevenção (45,3%), refletindo a importância dessas áreas para as participantes. As consultas relacionadas à saúde materna representam 3,1%, o que pode ser explicado pelo perfil da amostra, composta por mulheres mais jovens que, em sua maioria, ainda não enfrentaram questões de saúde materna, ou esta baixa porcentagem pode indicar a necessidade de melhorar o acesso e a conscientização sobre esses serviços essenciais.

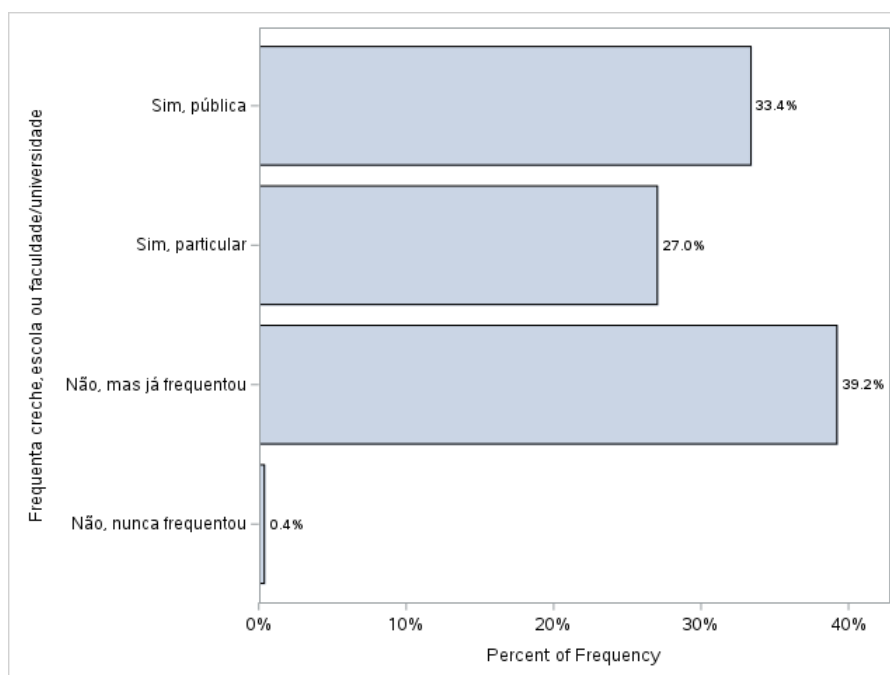
Tabela 8: Distribuição de mulheres de 14 a 24 anos, segundo motivos para Buscar Atendimento; Distrito Federal, Brasil - 2021

Busca de atendimento	Percentual
Saúde materna	3,1
Prevenção	45,3
Problema de saúde	48,9
Outros	2,7

A análise da cobertura de planos de saúde entre as mulheres revela que 25,4% possuem um plano. Dentre essas mulheres, 16,7% têm um plano individual/familiar e 7,5% possuem planos coletivos empresariais. A coparticipação é observada em 13,4% dos casos, enquanto 10,5% não enfrentam coparticipação. Os dados das mulheres com idade entre 14 e 24 anos, da amostra da PDAD de 2021 revelam aspectos importantes sobre a educação das participantes. A grande maioria, 98,6%, possui habilidades básicas de leitura e escrita, demonstrando uma alta taxa de alfabetização. Apenas uma pequena fração, 1,3%, não possui essas habilidades, e 0,2% não informou sua capacidade de leitura e escrita.

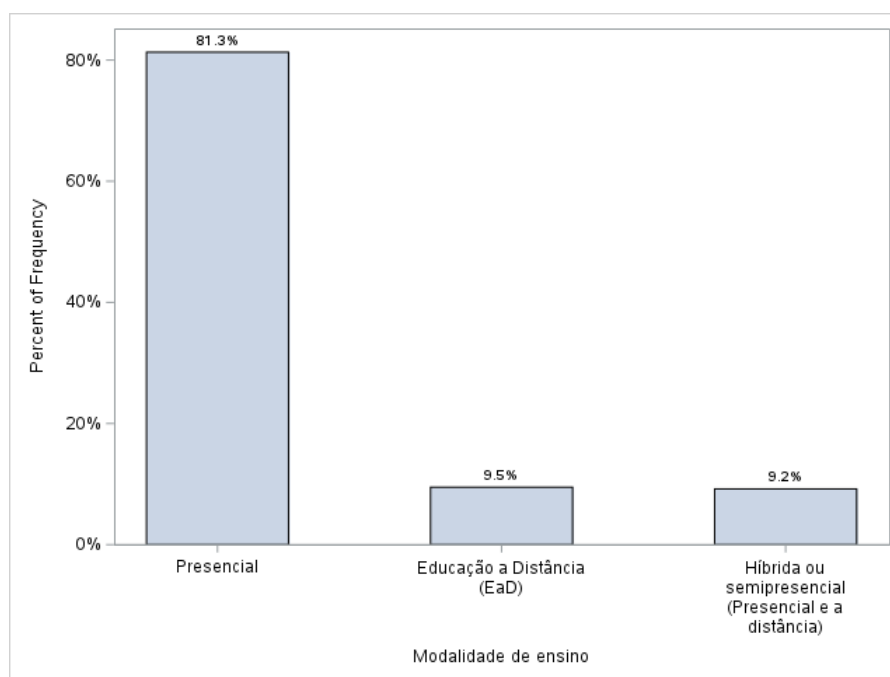
No que diz respeito à frequência escolar, 38,4% das participantes estão matriculadas em escolas públicas, enquanto 20,7% estão em escolas particulares. Um significativo 40,5% já frequentou uma instituição de ensino, mas atualmente não está matriculada, indicando um bom nível de acesso à educação. Apenas 0,5% nunca frequentaram escolas, destacando a ampla cobertura educacional.

Figura 10: Distribuição das mulheres de 14 a 24 anos, segundo frequência educacional; Distrito Federal, Brasil - 2021.



As modalidades de ensino revela que dentre as inscritas, a maioria está em cursos presenciais, com 81,3% das participantes, enquanto 9,5% estão em Educação a Distância (EaD) e 9,2% em modalidades híbridas. Em relação aos turnos de estudo, 83,6% frequentam o turno diurno e 16,4% o noturno.

Figura 11: Distribuição das mulheres de 14 a 24 anos, segundo modalidade de ensino; Distrito Federal, Brasil - 2021.



A análise dos dados revela padrões distintos no uso dos meios de transporte e no tempo de deslocamento, além de informações relevantes sobre o perfil educacional dos indivíduos. O meio de transporte mais utilizado é o ônibus, com 24,5% da amostra optando por esse meio. O transporte escolar público é escolhido por 2% dos indivíduos, enquanto o transporte escolar privado é utilizado por 3,1%. O transporte privado, como táxis ou aplicativos, representa 0,9% dos casos, e o uso de automóveis corresponde a 14% da amostra. Meios de transporte como metrô, motocicleta e bicicleta têm uma participação muito pequena, com 1,7%, 0,2% e 0,2%, respectivamente. Caminhar é uma prática comum para 17,1% dos indivíduos.

Quanto ao tempo de deslocamento, a maior parte dos indivíduos leva até 15 minutos para se deslocar entre o domicílio e a unidade de ensino (24,4%), seguido por 17% que gastam entre 15 e 30 minutos. Menos de 10% dos entrevistados passam mais de 45 minutos no deslocamento, e apenas 0,2% relatam um tempo superior a 2 horas.

No campo educacional, a maioria dos indivíduos está matriculada em cursos de Ensino Médio Regular (19%) ou Ensino Superior/Graduação (17%). Há também uma participação significativa em cursos técnicos do Ensino Médio e de Educação de Jovens e Adultos (EJA). A conclusão de cursos superiores é relativamente baixa, com apenas 1% dos entrevistados completando graduação ou pós-graduação. Em relação aos cursos adicionais, a maioria não está envolvida em cursos preparatórios para concursos (9,4%) e vestibulares (8,9%), e cursos de

línguas estrangeiras têm uma participação de 11,4%.

A análise dos dados sobre a busca de emprego entre mulheres de 14 a 24 anos, obtidos da Pesquisa de Diagnóstico e Análise da Demografia (PDAD) 2021, revela aspectos importantes sobre suas experiências no mercado de trabalho. Entre as participantes, 14,9% procuraram trabalho nos últimos 30 dias, enquanto 84,8% não estiveram em busca de emprego nesse período. Para um intervalo mais longo, 4% das mulheres procuraram emprego nos últimos 12 meses, e 80,7% não buscaram trabalho durante esse período.

A análise das estratégias adotadas pelas mulheres na busca por trabalho mostra uma variedade de abordagens, refletindo a diversidade de métodos utilizados para encontrar oportunidades de emprego. As participantes utilizaram diferentes canais, como buscar diretamente empregadores, responder a anúncios, consultar o Sistema Nacional de Emprego (SINE) e recorrer a agências de emprego. O apoio de redes pessoais, como parentes e amigos, também desempenha um papel significativo. Além disso, uma proporção considerável recorreu a outras estratégias não especificadas, evidenciando uma abordagem flexível e multifacetada na procura de trabalho.

Tabela 9: Distribuição das mulheres de 14 a 24 anos, segundo as estratégias de Busca de Emprego; Distrito Federal, Brasil - 2021

Estratégia	% das participantes
Busca direta a empregadores ou empresas	13,6
Publicação ou resposta a anúncios	14,2
Consulta ao Sistema Nacional de Emprego (SINE)	9
Busca em postos ou agências públicas	9,9
Utilização de agências de emprego privadas	11,8
Auxílio de parentes, amigos ou conhecidos	14,9
Outras medidas não especificadas	9,8

Em relação à aposentadoria, apenas 0,05% das participantes estão aposentadas, e 5,18% recebem pensão. Os valores médios das pensões foram de R\$844,82, enquanto o das aposentadorias foi de R\$1.103,97.

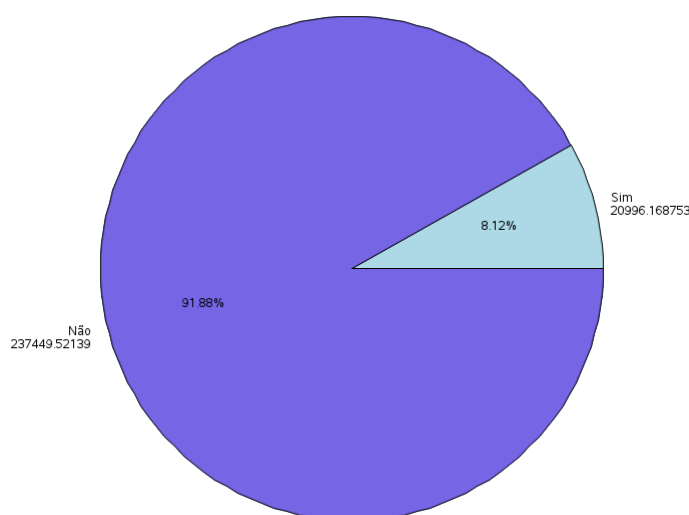
Quanto às atividades profissionais, 23,8% das mulheres trabalharam nos últimos 30 dias. Dentre essas, a maioria há um único emprego (23,3%), enquanto 0,2% possuem dois empregos e 0,1% tinham três ou mais. A duração média no trabalho principal foi de aproximadamente 1,49 anos, ou cerca de 2,42 meses, com uma carga horária semanal média de 35,35 horas. A distribuição geográfica dos locais de trabalho variou amplamente, com a maioria das mulheres empregadas espalhadas por diversas categorias de empresas e setores, o que reflete ampla gama de oportunidades de trabalho.

No que se refere ao transporte utilizado para o deslocamento ao trabalho, observou-se uma variedade de meios de transporte. O ônibus foi o meio mais frequente (12,8%), seguido

por automóvel (4,6%) e transporte privado, como táxis e aplicativos (0,5%). Outros meios de transporte, como metrô, motocicleta, bicicleta e deslocamento a pé, foram utilizados por percentuais menores da amostra, indicando uma diversidade de opções de locomoção para o ambiente laboral.

A análise dos dados sobre a fertilidade das mulheres na amostra revela algumas tendências importantes. Cerca de 8% das participantes relataram ter tido filhos nascidos vivos até junho de 2020, enquanto a vasta maioria, 91,6%, não teve filhos nessa condição. Além disso, o número médio de filhos nascidos vivos é relativamente baixo, com a maioria das mulheres tendo, em média, pouco mais de um filho, sendo que grande parte da amostra tem apenas um filho. Esses resultados ajudam a delinear o perfil reprodutivo predominante na amostra composta de mulheres na faixa de 14 a 24 anos.

Figura 12: Gráfico de setores da distribuição das mulheres de 14 a 24 anos, segundo filho nascido vivo; Distrito Federal, Brasil - 2021



5.2 Análise Descritiva Bivariada

Para garantir a robustez das análises e evitar sobreajuste, a amostra foi dividida em dois subconjuntos: amostra de construção e amostra de validação. A amostra de construção (ou treino) é utilizada para desenvolver e ajustar o modelo. Em contrapartida, a amostra de validação (ou teste) serve para avaliar o desempenho do modelo em novos dados, para garantir que ele não esteja ajustado apenas para o conjunto inicial, mas que tenha a capacidade de generalização. Neste estudo, a amostra treino representa 70% do banco de dados, enquanto a amostra teste corresponde a 30%.

A análise da escolaridade das mulheres com idade entre 14 e 24 anos fornece visão detalhada sobre o perfil educacional da amostra de construção, essencial para compreender as dinâmicas que influenciam a conclusão do ensino médio.

Tabela 10: Distribuição dos níveis de Escolaridade entre mulheres de 14 a 24 anos da amostra de construção

Escolaridade	Percentual (%)
Sem instrução	0,9
Ensino Fundamental Incompleto	17,4
Ensino Fundamental Completo	2,1
Ensino Médio Incompleto	28,1
Ensino Médio Completo	21,3
Ensino Superior Incompleto	21,4
Ensino Superior Completo	6,6
Sem classificação	2,2

A divisão em categorias de "Concluiu" e "Não concluiu" o ensino médio foca no objetivo do estudo e facilita a análise dos determinantes e desafios associados à conclusão dessa etapa educacional. Esta abordagem permite identificar de forma mais clara as barreiras e oportunidades que afetam a conclusão do ensino médio entre as mulheres da amostra. Dentre as participantes, 50,3% concluíram o ensino médio e 49,7% não concluíram o ensino médio.

A variável resposta em análise, nível de escolaridade, é crucial para compreender as dinâmicas educacionais e as desigualdades associadas à conclusão do ensino médio dentro dessa faixa etária específica.

A categoria "concluiu o ensino médio" abrange as mulheres que completaram essa etapa crucial da educação, um fator associado a melhores perspectivas de emprego e maior probabilidade de acesso ao ensino superior (FERREIRA, 2019). É um indicador de sucesso educacional, frequentemente relacionado à mobilidade social ascendente (SILVA; SANTOS, 2020). Em contrapartida, a categoria "não concluiu o ensino médio" inclui aquelas que, por diversos motivos, não conseguiram finalizar essa fase, o que pode estar associado a barreiras econômicas, falta de apoio familiar, ou dificuldades estruturais no sistema educacional (OLIVEIRA, 2018). Estudos mostram que a falta de conclusão do ensino médio está frequentemente ligada à perpetuação da desigualdade social e à menor inserção no mercado de trabalho (PEREIRA; LIMA, 2021).

Para compreender os determinantes da conclusão do ensino médio realiza-se análise bivariada com base na amostra de construção. Esta investigação explorará a relação entre o "nível de escolaridade" e diversas variáveis preditoras, como fatores socioeconômicos e condições familiares, para identificar padrões e influências sobre a conclusão do ensino médio. Para esta análise, as variáveis foram pré selecionadas com base em três critérios principais: o objetivo do estudo, literatura existente e análise pessoal da autora.

Tabela 11: Análise bivariada de nível de escolaridade da amostra selecionada da PDAD 2021

Variáveis	Concluiu o ensino médio		Estatística	P-valor
	Sim	Não		
Idade				
Média	21,3	16,8	Wald 674,2	<,0001
Desvio padrão	12,1	16,05		
Raça/cor				
Negro	46,9%	53,1%	Qui-quadrado 1111,97	<,0001
Não negro	55%	45%		
Estado civil				
Casada	73,3%	26,7%	Qui-quadrado 3945,12	<,0001
Solteira	47,8%	52,2%		
Outros	73,2%	26,8%		
Filho nascido vivo				
Sim	55,1%	44,9%	Qui-quadrado 142,28	<,0001
Não	49,9%	50,1%		
Número de filhos				
Média	1,2	1,3	Teste de Wald 1,4	0,2431
Desvio Padrão	3,7	4,2		
Número de filhos residentes				
Média	1,1	1,2	Teste de Wald 1,3	0,25
Desvio Padrão	3,5	4,1		
Local por renda				
Renda alta	58,3%	41,7%	Qui-quadrado 1120,7	<,0001
Renda baixa	47,5%	52,5%		
Renda média alta	53%	47%		
Renda média baixa	48,1%	51,9%		
Condição				
Pessoa responsável	63,1%	36,9%	Qui-quadrado 2572,48	<,0001
Filho do responsável e do cônjuge	45,4%	54,6%		
Filho somente do responsável	46,7%	53,3%		
Outros	53,9%	46,1%		
Mora em companhia				
Sim	71,4%	28,6%	Qui-quadrado 4052,69	<,0001
Não	47,7%	52,3%		
Dificuldade				
Tem dificuldade	43,4%	56,6%	Qui-quadrado 82,1	<,0001
Não tem dificuldade	50,5%	49,5%		

Tabela 11: Análise bivariada de nível de escolaridade da amostra selecionada da PDAD 2021 (conclusão)

Variáveis	Concluiu o ensino médio		Estatística	P-valor
	Sim	Não		
Saúde			Qui-quadrado	
Saúde materna	61,2%	38,8%	419,3	<,0001
Prevenção	51,3%	48,7%		
Problemas de saúde	49%	51%		
Outros	65,9%	34,1%		
Turno de ensino			Qui-quadrado	
Diurno	21,2%	78,8%	14598,05	<,0001
Noturno	70%	30%		
Utilizou a internet para educação/cursos			Qui-quadrado	
Sim	49,7%	50,3%	433,78	<,0001
Não	55,8%	44,2%		
Meio de transporte			Qui-quadrado	
Pessoal	5,3%	94,7%	12459,65	<,0001
Privado	35,8%	64,2%		
Público	44,2%	55,8%		
Tempo de deslocamento			Qui-quadrado	
Até 15 minutos	14,2%	85,8%	17344,95	<,0001
De 15 a 30 minutos	28,8%	71,2%		
De 30 minutos a 1 hora	54,8%	45,2%		
Acima de 1 hora	75,2%	24,8%		
Trabalhou no último mês			Qui-quadrado	
Sim	79,8%	20,2%	19427,11	<,0001
Não	40,8%	59,2%		
Horas nos afazeres domésticos			Teste de Wald	
Média	11,45	10,4	6,63	0,0101
Desvio Padrão	64,4	62,7		

A análise das mulheres do Distrito Federal, com idades entre 14 e 24 anos, revela relação positiva entre a idade e a conclusão do ensino médio. A média de idade das que concluíram o ensino médio é de 21,3 anos, enquanto as que não o concluíram possuem média de 16,8 anos. Esses resultados indicam que a conclusão do ensino médio tende a ocorrer em idades mais avançadas, o que aponta que as mais jovens ainda se encontram em processo de escolarização ou enfrentam dificuldade para finalizar esta etapa educacional. Além disso, a *Odds Ratio* para idade é de 2,014 com intervalo de confiança de 95% entre 1,911 e 2,124, isto significa que, a cada ano adicional de idade, a chance de uma pessoa ter concluído o ensino médio é aproximadamente duas vezes maior.

Entre as mulheres, a taxa de conclusão do ensino médio é mais alta entre as não negras (55%) do que entre as negras (46,9%). Essa diferença insinua uma disparidade educacional

significativa, sugerindo que as mulheres negras enfrentam maiores desafios para concluir o ensino médio em comparação com as mulheres não negras. Essa variação destaca a necessidade de investigar os fatores que contribuem para essa desigualdade educacional.

A distribuição do estado civil em relação à conclusão do ensino médio mostra que as mulheres solteiras têm a maior proporção entre aquelas que não concluíram o ensino médio (52,2%) e uma proporção considerável entre as que concluíram (47,8%). Em contraste, as mulheres casadas representam uma alta proporção entre as que concluíram o ensino médio (73,3%), mas uma proporção muito menor entre as que não concluíram (26,7%). Essas estatísticas evidenciam uma associação significativa entre estado civil e nível de escolaridade, como demonstrado pelo teste qui-quadrado com valor de χ^2 de 3945,12 e valor p menor que 0,0001.

Os dados revelam forte associação entre ter filhos nascidos vivos até 30 de junho de 2020 e o nível de escolaridade, com diferenças significativas nas proporções entre os grupos. A tabela de contingência indica que 55,07% das pessoas que tem filho concluíram o ensino médio, enquanto 44,9% não concluíram o ensino médio e estão na mesma situação, esta associação é confirmada pela estatística do teste qui-quadrado ($\chi^2 = 142,28$, $p < 0,0001$). No entanto, as análises de regressão logística mostram que o número de filhos e a presença de filhos residentes não tem um impacto estatisticamente significativo na probabilidade de conclusão do ensino médio. Embora, indivíduos com o ensino médio completo tenham, em média, um número ligeiramente menor de filhos (média=1,23) comparado aos que não completaram o ensino médio (média=1,34), essa diferença não se traduz em um impacto significativo na conclusão do ensino médio, segundo os modelos de regressão logística.

A análise da classificação das Regiões Administrativas por renda revela discrepâncias notáveis entre os grupos. A maior proporção de conclusão do ensino médio (58,3%) e a menor de não conclusão (41,7%) estão na de maior poder aquisitivo. Em contraste, as mulheres com renda baixa apresentam a maior proporção de não conclusão (52,5%) e a menor de conclusão (47,5%). As mulheres com renda média alta têm uma taxa de conclusão de 53%, enquanto a de não conclusão é de 47%. Por fim, as mulheres com renda média baixa mostram uma distribuição quase equilibrada, com 48,1% de conclusão e 51,9% de não conclusão. Esses dados indicam que a renda influencia significativamente o nível de escolaridade, com maior taxa de conclusão do ensino médio entre as mulheres de renda mais alta.

O estudo das mulheres que concluíram o ensino médio revela que a proporção de conclusão do ensino médio é maior entre as mulheres cujos responsáveis são apenas pessoas responsáveis pelo domicílio (63,1%), comparado aos outros grupos. As mulheres cujos responsáveis são filhos do responsável e do cônjuge têm uma proporção de conclusão de 45,4%, enquanto as que têm apenas um filho do responsável apresentam 46,8%. Esses resultados sugerem que a condição do responsável influencia a conclusão do ensino médio, com a maior taxa observada entre as mulheres com pessoas responsáveis pelo domicílio.

Embora exista associação estatisticamente significativa entre viver com companhia e o nível de escolaridade, a força dessa associação é relativamente fraca. Isso indica que, apesar da relação observada, viver com companhia não é um fator determinante significativo para a con-

clusão do ensino médio. Outras variáveis ou contextos podem exercer impacto mais substancial no nível de escolaridade. Portanto, a associação identificada pode não ser robusta para fornecer conclusões definitivas sobre como viver com companhia afeta a conclusão do ensino médio.

A análise dos dados mostra que dentre as pessoas não enfrentam dificuldades, 50,5% concluíram o ensino médio, enquanto que 49,5% não. Em contrapartida, as pessoas que possuem dificuldade apresentam maior proporção em não conclusão (56,6%).

A análise mostra associação significativa entre o nível de escolaridade e as categorias de saúde, visto que embora as pessoas apresentam algum problema de saúde, a maioria conclui esta etapa educacional.

Os dados mostram que a proporção de mulheres que concluíram o ensino médio é maior no noturno (70,02%) em comparação com o turno diurno (21,2%). Por outro lado, a proporção de mulheres que não concluíram o ensino médio é significativamente maior no turno diurno (78,9%) do que no turno noturno (30%). Esses resultados indicam que a variável resposta está associada ao turno de ensino.

A internet desempenha papel crucial na educação, dado que fornece recursos e cursos que podem enriquecer a formação acadêmica. No entanto, a análise dos dados mostra que a diferença no uso da internet para educação entre aqueles que concluíram e não concluíram o ensino médio é limitada. Embora a associação seja estatisticamente significativa, a força da relação é fraca, conforme os coeficientes de associação. Isso sugere que, apesar da internet seja recurso valioso, seu impacto na conclusão do ensino médio é menos pronunciado comparado a outros fatores.

A análise referente ao meio de transporte e o tempo de deslocamento, indicam que ambas as variáveis estão associadas ao nível de escolaridade, com a utilização do transporte público e tempos de deslocamento de até 1 hora sendo mais comuns entre aquelas que concluíram o ensino médio. Esses resultados ressaltam a importância de melhorias a acessibilidade ao transporte e reduzir o tempo de deslocamento para apoiar a educação.

O diagnóstico dos dados revela insights significativos entre o nível de escolaridade e a participação no mercado de trabalho, bem como a dedicação a afazeres domésticos. A participação no mercado de trabalho está associada a uma maior taxa de conclusão do ensino médio, como observa-se que ao trabalhar a proporção de conclusão é de 79,8%. Os resultados revelam que a média de horas dedicadas aos afazeres domésticos é sutilmente superior para as mulheres que concluíram o ensino médio (11,45), em comparação com aquelas que não concluíram (10,4). Ainda que a diferença seja pequena, a medida sugere que há tendência que as mulheres com maior nível de escolaridade podem dedicar mais tempo a tarefas domésticas, a cada hora adicional gasta com afazeres domésticos, aumenta cerca de 1% na probabilidade de completar o ensino médio.

Entretanto, apesar de todos os testes executados apresentarem associação entre as variáveis preditoras e o nível de escolaridade, as porcentagens de conclusão e não conclusão estão semelhantes, gerando a suposição de que esta associação é aceita, possivelmente, devido

ao tamanho da amostra.

- Correlação entre as variáveis

A análise da correlação entre as variáveis preditoras é fundamental na construção de modelos estatísticos, como a regressão logística, uma vez que a presença de correlações elevadas entre variáveis pode resultar em multicolinearidade. Este fenômeno ocorre quando duas ou mais variáveis independentes estão altamente correlacionadas, o que prejudica a precisão das estimativas dos coeficientes do modelo.

A multicolinearidade, além de dificultar a interpretação dos efeitos individuais das variáveis, pode aumentar os erros-padrão das estimativas, tornando difícil identificar quais preditores são de fato relevantes. Assim, os resultados do modelo podem se tornar menos confiáveis, com maiores chances de obter conclusões errôneas sobre a significância estatística das variáveis. Por isso, a detecção e o tratamento adequado da multicolinearidade são essenciais para garantir a robustez e a validade das inferências geradas pelo modelo, contribuindo para análises mais precisas e consistentes.

Observou-se forte associação entre filhos nascidos vivos com número de filho e número de filhos residentes, isso ocorre porque a presença de filhos nascidos implica em uma resposta positiva para as outras duas variáveis.

Além disso, a matriz de correlação entre número de filhos e número de filhos residentes apresenta um coeficiente de 0,75047, o que indica correlação positiva e forte entre as variáveis, ou seja, esse valor sugere que a quantidade de filhos residentes aumenta de maneira proporcional ao número de filhos.

A análise também revelou associação forte e significativa de 0,7 entre morar em companhia e estado civil, ou seja, o estado civil é fator importante na determinação da mulher não morar só. Assim como uma relação considerável entre meio de transporte e o tempo de deslocamento.

5.3 Modelagem

A modelagem estatística visa compreender e descrever a relação entre as variáveis, onde testará hipóteses propostas ao longo do trabalho. Com base na análise da Tabela 11, seleciona-se variáveis preditoras, segundo os respectivos p-valores, as quais são consideradas relevantes para a construção do modelo, o qual visa explicar a conclusão ou não do ensino médio das mulheres do Distrito Federal, com idade entre 14 e 24 anos. Tendo em vista que a variável resposta é categórica, a regressão logística é a mais apropriada para identificar padrões e associações entre as variáveis independentes, em função da natureza da variável dependente: nível de escolaridade.

Ao considerar um nível de significância de 25%, a priori todas as variáveis preditoras são incluídas no modelo para permitir a análise abrangente. Entretanto, é importante observar que ao reduzir este α para 5%, valor usual, observa-se que as variáveis "número de filhos" e "número

de filhos residentes” não demonstram associação significativa com o nível de escolaridade. Além disso a análise inicial deve ser complementada pela avaliação da correlação entre as variáveis preditoras, visto que a presença de alta correlação pode indicar multicolinearidade. Portanto, é essencial analisar e, se necessário, remover variáveis altamente correlacionadas.

Para aprimorar o modelo inicial, realiza-se uma análise minuciosa, visando identificar e remover variáveis altamente correlacionadas, com o objetivo de melhorar a precisão e a clareza do modelo. Conforme a observação da semelhança de proporção na conclusão ou não do ensino médio dependendo da variável em análise, opta-se por retirá-las do modelo, exceto quando trata-se do filho nascido vivo, visto que é característica importante de estudo. Além disso, como apresentado na seção ”Correlação entre variáveis” há variáveis que apresentam correlação significativa, o que justifica a realização de testes para também retirá-las. O critério de informação de Akaike (AIC), grandeza que fornece medida do ajuste do modelo, com a exclusão das variáveis ”número de filho”, ”número de filhos residentes”, ”dificuldade”, ”mora em companhia”, ”meio de transporte”, ”uso da internet para educação” e ”horas em afazeres domésticos” resulta em um valor absoluta de 30272,025, uma quantia mais baixa, comparada a outros modelos. Embora inclusão de variáveis decresça o valor do AIC, o seu ganho não é tão significativo, por isso, opta-se por manter o modelo inicial mais enxuto.

Analisar interações entre variáveis é crucial para compreender como múltiplas combinações influenciam os resultados de maneira mais complexa, revelando como o efeito de uma variável pode variar dependendo do nível da outra, dado o objetivo de analisar fatores que influenciam na conclusão do ensino médio das mulheres com idade entre 14 e 24 anos.

O modelo de seleção automática *Stepwise* apresentou um modelo extenso, além de problemas de estimação, o que pode ser explicado pela grande quantidade de parâmetros e a existência de observações em certas combinações de categorias, apesar da grandeza do banco de dados. Por isso, opta-se por efetuar essa seleção manualmente. Adicionou-se individualmente as interações para analisar se deveriam ser incluídas no modelo, em seguida avaliou-se, pelo teste de máxima verossimilhança, a significância de cada interação, dado a hipótese nula de que a interação não melhora o modelo. Quando a hipótese nula foi rejeitada, analisou o Valor do Critério de Informação de Akaike (AIC). Deste modo, a inclusão de cada interação foi, portanto, avaliada com base na melhoria do AIC, fator que determina se a complexidade adicional justifica o ganho no ajuste do modelo, o que traz precisão e relevância do modelo.

Em todas as inclusões de interação deste estudo, rejeitou-se a hipótese nula no teste de máxima verossimilhança, no entanto, a decisão final, baseada no valor do AIC, ressalta que o valor deveria ser baixo, mas também apresentar o número mais reduzido possível de variáveis, dado a complexibilidade e extensão do modelo.

Com essas considerações, a Tabela 12 contém as estimativas das variáveis do modelo resultante do processo de seleção de variáveis para as amostras de construção, validação e total do estudo.

Tabela 12: Tabela de Estimativas dos Parâmetros do Modelo

Parâmetro	Estimativa- Const	Estimativa- Valid	Estimativa - Completa
Intercepto	-24,67	-21,08	-20,92
idade	1,05	0,98	1,02
Estado civil - Casada	2,9	-0,9	-0,38
Estado civil - Outros	-6,79	0,22	-0,37
Filho nascido vivo - Não	1,14	0,9398	1,04
Local por renda - Renda alta	0,7968	0,8691	0,77
Local por renda - Renda baixa	-0,736	-0,8	-0,71
Local por renda - Renda média alta	0,5224	1,07	0,68
Condição - filho do responsável e do cônjuge	0,4917	0,42	0,47
Condição - filho somente do responsável	-0,1799	-0,81	-0,4
Condição - outros	0,0397	0,68	0,24
Saúde - Saúde materna	0,4721	-0,99	0,35
Saúde - Prevenção	-0,21	1,29	-0,04
Saúde - Problema de saúde	-0,33	1,03	-0,24
Turno de ensino - Diurno	-0,34	-0,52	-0,41
Tempo de deslocamento - Acima de 1 hora	1,59	1,91	1,63
Tempo de deslocamento - Entre 15 e 30 min	-0,6924	-0,35	-0,56
Tempo de deslocamento - Entre 30 min e 1 hora	0,4	-0,13	0,32
Trabalhou no último mês - Não	0,192	-0,02	0,19

As estimativas dos parâmetros das amostras de construção, validação e a amostra completa são similares, portanto, pode-se supor que o modelo final é válido para explicar o comportamento dos dados, conforme o objetivo do estudo.

Tabela 13: Tabela de Estimativas dos Parâmetros do Modelo Final

Parâmetro	Estimativa	Erro Padrão	Wald Chi-Square	Valor-p
Intercepto	-20,92	0,2811	5533,8197	<,0001
Idade	1,02	0,00730	19413,4556	<,0001
Estado civil - Casada	-0,38	0,2338	2,6535	0,1033
Estado civil - Outros	-0,37	0,4614	0,6545	0,4185
Filho nascido vivo - Não	1,04	0,0363	818,2259	0,4185
Local por renda - Renda alta	0,77	0,0229	1116,7116	<,0001
Local por renda - Renda baixa	-0,71	0,0237	900,52	<,0001
Local por renda - Renda média alta	0,68	0,59	682,64	<,0001
Condição - filho do responsável e do cônjuge	0,47	0,02	517,67	<,0001
Condição - filho somente do responsável	-0,4	0,024	286,36	<,0001
Condição - outros	0,24	0,023	105,59	j<,0001
Saúde - Saúde materna	0,35	0,08	17,35	<,0001
Saúde - Prevenção	-0,04	0,037	1,26	0,262
Saúde - Problema de saúde	-0,24	0,04	42,32	<,0001
Turno de ensino - Diurno	-0,41	0,0149	737,8	<,0001
Tempo de deslocamento - Acima de 1 hora	1,63	0,03	2543,15	<,0001
Tempo de deslocamento - Entre 15 e 30 min	-0,56	0,021	687,7108	<,0001
Tempo de deslocamento - Entre 30 min e 1 hora	-0,32	0,023	189,68	<,0001
Trabalhou no último mês - Não	0,19	0,0174	121,76	<,0001

5.3.1 Avaliação do modelo

Na avaliação da adequação do modelo de regressão logística ajustado, utilizou-se o teste de Hosmer Lemeshow, que compara as probabilidades previstas com as frequências observadas.

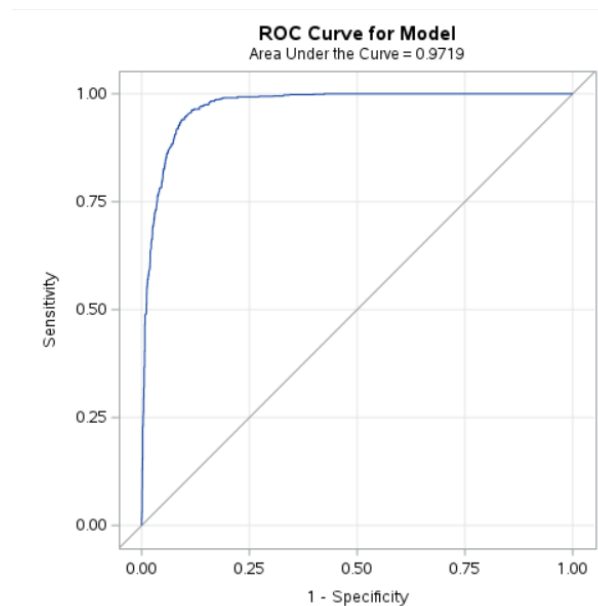
O resultado do teste de Hosmer-Lemeshow apresentou p-valor muito baixo, $<,0001$, apesar da porcentagem de concordância ser acima de 97%. Entretanto, é importante destacar que o tamanho elevado do banco de dados pode influenciar significativamente no resultado do teste devido à sua capacidade de detectar efeitos estatísticos mesmo quando estes são pequenos. Por isso, o teste torna-se altamente sensível a pequenas discrepâncias entre as probabilidades previstas pelo modelo e as observações reais, o que resulta em um valor p baixo. Uma das formas de analisar melhor o modelo para poder tomar uma decisão sobre a sua qualidade é utilizando outras métricas, com a curva de ROC e AUC, para complementar a avaliação do ajuste do modelo, assim como análises dos resíduos, o que verificará se há padrões ou problemas que o teste de Hosmer-Lemeshow não capturou.

Tabela 14: Porcentagem de concordância

	Porcentagem
Concordantes	97,2
Discordantes	2,8

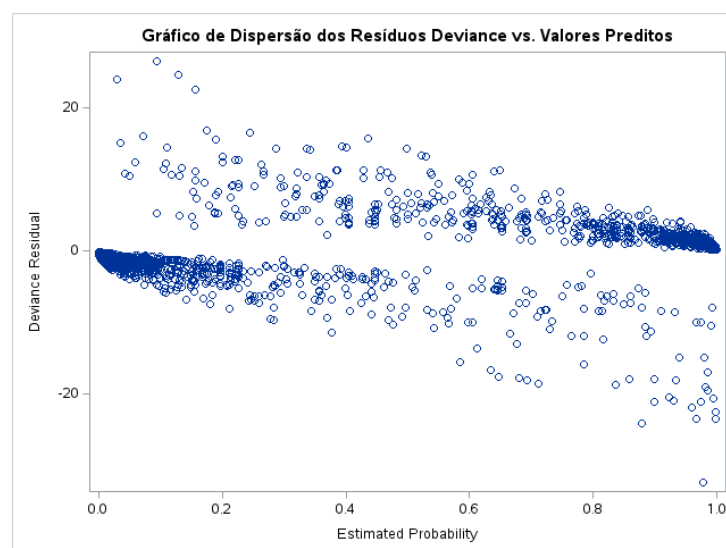
A Curva ROC é uma representação gráfica que avalia a performance de modelos de classificação binária, comparando as taxas de verdadeiros positivos com as taxas de falsos positivos em diferentes pontos de corte. A área sob a curva (AUC) fornece uma medida global da capacidade do modelo em distinguir entre as classes. O valor apresentado na Figura 13 é de 0,97, o que indica que o modelo tem alta capacidade para distinguir entre as duas classes, ou seja, o modelo tem 97% de chance de classificar corretamente uma observação positiva como positiva e uma observação negativa como negativa, quando comparado a uma escolha aleatória.

Figura 13: Curva ROC do modelo em estudo; Distrito Federal, Brasil - 2021



O gráfico de dispersão dos resíduos *deviance* é uma ferramenta essencial na avaliação do ajuste de modelos de regressão, especialmente em modelos de regressão logística. Este gráfico ilustra a relação entre os resíduos *deviance* e os valores ajustados, permitindo a verificação da qualidade do ajuste do modelo. Idealmente, os resíduos *deviance* devem estar distribuídos de maneira uniforme e aleatória ao redor da linha horizontal de zero, sem padrões sistemáticos evidentes. O gráfico apresentado demonstra essa característica, mostrando que os resíduos estão distribuídos de forma aleatória e uniforme, centralizados em torno de zero. Isso indica que o modelo está bem ajustado aos dados e que não há evidências de problemas significativos com a especificação do modelo.

Figura 14: Gráfico de dispersão dos resíduos Deviance vs Valores Preditos do modelo em estudo; Distrito Federal, Brasil - 2021



5.3.2 Interpretação dos parâmetros

Tabela 15: Estimativas de Razão de Chance (*Odds Ratio*) com Intervalos de Confiança de 95%

Efeito	Estimativa Pontual	Limite Inferior (95% IC)	Limite Superior (95% IC)
Idade	2,764	2,724	2,803
Estado civil: Casada vs Solteira	0,321	0,285	0,363
Estado civil: Outros vs Solteira	0,324	0,083	1,257
Filho nascido vivo: Não vs Sim	7,958	6,904	9,173
Local por renda: Renda alta vs Renda média baixa	4,482	4,182	4,803
Local por renda: Renda baixa vs Renda média baixa	1,023	0,957	1,095
Local por renda: Renda média alta vs Renda média baixa	4,100	3,797	4,428
Condição: filho do responsável e do cônjuge vs pessoa responsável pelo domicílio	2,193	1,999	2,404
Condição: filho somente do responsável vs pessoa responsável pelo domicílio	0,914	0,830	1,006
Condição: outros vs pessoa responsável pelo domicílio	1,734	1,576	1,908
Saúde: Saúde materna vs outros	1,506	1,154	1,966
Saúde: Prevenção vs outros	1,022	0,863	1,210
Saúde: Problema de saúde vs outros	0,838	0,708	0,992
Turno de ensino: Diurno vs Noturno	0,445	0,420	0,472
Tempo de deslocamento: Acima de 1 hora vs Até 15 minutos	20,678	18,838	22,698
Tempo de deslocamento: Entre 15 e 30 min vs Até 15 minutos	2,321	2,183	2,468
Tempo de deslocamento: Entre 30 min e 1 hora vs Até 15 minutos	5,560	5,190	5,956
Trabalhou no último mês: Não vs Sim	1,469	1,372	1,573

A análise da *odds ratio* revela informações importante sobre os fatores que influenciam a conclusão do ensino médio. A idade apresenta uma associação positiva significativa, com um *Odds Ratio* de 2,77, sugerindo que, a cada aumento de um ano, a chance de concluir o ensino médio aumenta cerca de 2,7 vezes. A princípio, o estado civil mostra-se fator relevante. Mulheres casadas apresenta probabilidade menor de concluir o ensino médio em comparação às solteiras, com uma *odds* de 0,321, isso sugere que as responsabilidades associadas ao casamento podem interferir no alcance do objetivo educacional, reduzindo cerca de 68% as chances de conclusão. Da mesma forma, pessoas em outros estados civis apresentam menores chances de concluir o ensino médio em comparação com as solteiras, embora o intervalo de confiança para esta categoria seja mais amplo, o que sugere menos precisão nessa estimativa.

Uma variável de impacto no estudo foi filho nascido vivo, com um *odds ratio* de 7,96, indicando que aquelas que responderam que não tem filho, possuem cerca de oito vezes mais chances de concluir o ensino médio do que as que possuem filho. Isso mostra relação forte e positiva com a conclusão, sendo uma das mais expressivas no conjunto de dado, sustentada por um intervalo de confiança estreito o que reforça a precisão dessa estimativa.

As mulheres com renda alta têm 4,48 vezes mais chances de concluir o ensino médio em comparação àquelas com renda média baixa. Essa tendência também é observada para renda média alta, com um *odds ratio* de 4,1, sugerindo que melhores condições econômicas estão claramente associadas a uma maior probabilidade de sucesso educacional. Na análise comparativa entre mulheres de renda baixa e média-baixa, observou-se uma razão de chances (OR) de 1,023. No entanto, como o intervalo de confiança inclui o valor 1, conclui-se que não há diferença estatisticamente significativa entre os grupos, indicando que as chances são equivalentes para ambos.

No que diz respeito à estrutura família, filhas do responsável e do cônjuge têm ligeira-

mente mais chances ($odds=2,19$) de concluir o ensino médio em comparação à responsável pelo domicílio. Contudo, filhas apenas do responsável apresentam chance reduzida, 0,914, o que pode indicar, no contexto, a ausência de um cônjuge na unidade familiar pode ser fator que dificulta o progresso educacional.

Observa-se que a saúde materna ($odds=1,51$) e a prevenção ($odds=1,022$) sugerem uma associação positivo na conclusão do ensino médio. Contudo, o intervalo de confiança para "prevenção" inclui o valor 1, o que indica que o efeito desta variável pode não ser estatisticamente significativo e, portanto, as chances de conclusão podem ser semelhantes entre o grupo "prevenção" e o de comparação ("outros"). Em contraste, ter um problema de saúde reduz essa probabilidade de conclusão ($odds=0,838$). Em relação ao turno de ensino, estudar no turno diurno em vez do noturno diminui as chances de conclusão ($odds=0,445$), o que pode refletir dificuldades de conciliar outras atividades com o estudo durante o dia.

O tempo de deslocamento é fator crítico: mulheres que levam mais de uma hora para a unidade de ensino têm 20,68 vezes mais chances de concluir o ensino médio do que as que levam até 15 minutos. Isso insinua que estudantes que enfrentam deslocamentos longos demonstram maior determinação e perseverança em alcançar seus objetivos educacionais ou a presença de outros fatores de confundimento.

6 Conclusão

O presente estudo visou identificar e compreender fatores, com foco particular na maternidade, que influenciam a conclusão do ensino médio de mulheres com idade entre 14 e 24 anos. O modelo ajustado, apesar de apresentar pequenas discrepâncias devido ao tamanho da amostra, pode ser considerado bom, como evidenciado pela análise dos resíduos e pelo desempenho geral do modelo, com um AUC de 97%.

Entre os fatores analisados, a presença de filhos se destacou como indicador significativo na conclusão do ensino médio, visto que, aquelas que não possuem filhos têm chance quase 8 vezes maior de concluir o ensino médio em comparação com aquelas que não tem. Essa informação reforça que a maternidade precoce muitas vezes interfere na continuidade dos estudos, devido a conciliação de demandas maternas e acadêmicas.

As estimativas revelaram que ser filha do responsável e do cônjuge tem impacto positivo na conclusão, comparado a ser filha apenas do responsável. Esta distinção sugere que a composição familiar, especialmente a presença de um cônjuge, pode oferecer suporte mais significativo, o que pode auxiliar a continuidade dos estudos. Em contraste, ser filha apenas do responsável está associada a maiores dificuldades de concluir o ensino médio, possivelmente devido a uma menor rede de apoio.

A análise mostrou que aspectos como renda e tempo de deslocamento afetam diretamente a conclusão do ensino médio, dado que condições desfavoráveis nessas áreas criam obstáculos adicionais, que podem comprometer a continuidade dos estudos e com isso reforça a necessidade de políticas que abordem essas questões.

O presente estudo deve ser considerado como uma primeira abordagem do tema e apresenta limitações em decorrência da faixa etária analisada. Muitas das mulheres incluídas na amostra estão em idades que, em sua maioria, ainda não concluíram o ensino médio, especialmente aquelas com até 18 anos completos. Essa limitação pode ter impactado as conclusões obtidas. Portanto, recomenda-se a realização de investigações futuras que considerem uma amostra menos ampla e heterogênea, o que possibilitará uma compreensão mais precisa e representativa do fenômeno em questão.

Em síntese, o estudo reforça a importância de uma abordagem abrangente e multifacetada para apoiar a conclusão do ensino médio, pois concluir essa etapa educacional é fundamental para o acesso a oportunidades de emprego melhores e para o desenvolvimento pessoal e profissional. A conclusão do ensino médio possibilita uma carreira mais estável e base mais sólida para a vida adulta.

Embora a conclusão do ensino médio seja essencial e repleta de desafio para todas, as mães enfrentam desafios complementares que tornam essa conquista mais difícil. As responsabilidades maternas, juntamente com as exigências acadêmicas, geram barreiras significativas que podem comprometer seu progresso. Portanto, é crucial oferecer suporte especializado a essas estudantes, incluindo assistência financeira, serviços de creche, programas de apoio acadêmico e

emocional.

No entanto, é essencial reconhecer que, além de mães, muitas outras mulheres, mães e não-mães, enfrentam dificuldades únicas, como a sobrecarga doméstica, pressões familiares e desigualdades socioeconômicas, que podem afetar a capacidade de completar os estudos. Um sistema educacional inclusivo e equitativo, com políticas de suportes específicas, é fundamental para assegurar que todas as mulheres tenham a mesma oportunidade de sucesso e possam concluir essa etapa vital na sua educação.

Referências

- AGRESTI, A. *An Introduction to Categorical Data Analysis*. [S.l.]: John Wiley Sons, 2007. v. 2a ed.
- BARRETO, G. P. A evolução histórica do direito das mulheres. 2016.
- BRASIL. Lei nº 6.202, de 17 de abril de 1975. *Diário Oficial [da] República Federativa do Brasil*, Brasília, DF, 1975. Disponível em: http://www.planalto.gov.br/ccivil_03/_Ato2011-2014/2012/Lei/L12651.htm.
- BRUM, J. d. S. Possibilidades e desafios da maternidade no ensino superior: um estudo de caso do ifes campus itapina. *Instituto Federal do Espírito Santos*, 2022.
- CARVALHO, M. B. de; MATSUMOTO, L. S. Gravidez na adolescência e a evasão escolar. 2019.
- CLEMENTINO, I. T. A.; CABRAL, K. C. de A.; RODRIGUES, F. S. Mulheres, trabalhadoras e mães: desafios para a conclusão do ensino médio na eja em uma escola estadual de fortaleza. *Revista Educação & Ensino*, v. 4, n. 1, 2020.
- CODEPLAN. *Apresentação PDAD-DF 2021*. 2022. Acesso em: [data de acesso]. Disponível em: https://www.codeplan.df.gov.br/wp-content/uploads/2022/05/Apresentacao_PDAD-DF_2021.pdf.
- (CODEPLAN), C. de Planejamento do D. F. Pesquisa distrital por amostra de domicílios (pdad) 2021. *Codeplan*, 2021.
- FEDERAL, G. Evasão escolar no ensino medio atinge meio milhão de jovens por ano, aponta estudo. Governo Federal, 2022. Disponível em: <https://www.gov.br/mdh/pt-br/assuntos/noticias/2022/fevereiro/casos-de-gravidez-na-adolescencia-diminuiram-em-media-18-desde-2019#:~:text=Confira%20o%20hotsite%20do%20Plano%20Nacional%20No%20Brasil%2C,est%C3%A3o%20estimadas%2068%2C4%20gesta%C3%A7%C3%B5es%20nesta%20fase%20da%20vida.>
- FEDERAL, G. Mulheres predominam em estudos, pesquisas e exames. Governo Federal, Brasília, DF, 2023. Disponível em: <https://www.gov.br/inep/pt-br/assuntos/noticias/institucional/mulheres-predominam-em-estudos-pesquisas-e-exames-educacionais>.
- FERREIRA, M. *Educação e Mobilidade Social no Brasil*. [S.l.]: Editora Acadêmica, 2019.
- FIELD, A. *Discovering Statistics Using IBM SPSS Statistics*. 5. ed. Londres: SAGE, 2018.
- GARCIA, L. A mulher e a evolução dos seus direito. 2009.
- Governo do Distrito Federal Secretaria de Estado de Fazenda, Planejamento, Orçamento e Gestão do Distrito Federal - SEFP, C. *Pesquisa Distrital por Amostra de Domicílios (PDAD) 2021*. Brasília, DF: Companhia de Planejamento do Distrito Federal - CODEPLAN, 2022. Disponível em: file:///C:/Users/maced/Downloads/PDAD-DF_2021.pdf.
- HAIR, J. F. et al. *Multivariate Data Analysis*. 7. ed. Harlow: Pearson, 2014.
- HOSMER, D. W.; LEMESHOW, S.; STURDIVANT, R. X. *Applied Logistic Regression*. 3. ed. Hoboken: Wiley, 2013.
- (IBGE), I. B. de Geografia e E. Relatório de dados da pdad 2021. *IBGE*, 2021.

- INESC. Abandono no ensino médio brasileiro duplicou na pandemia. INESC, Brasília, DF, 2023. Disponível em: (https://inesc.org.br/abandono-no-ensino-medio-brasileiro-duplicou-na-pandemia/?gclid=CjwKCAiA-P-rBhBEEiwAQEXhH2_fHyemal31PG0I5M0fTvx0LnToNsQH0A4-o-ACPNl9uWge3zxUSR0CsuUQAvD_BwE).
- KORN, E. L.; GRAUBARD, B. I. Analysis of health surveys. *Wiley Series in Probability and Statistics*, 1990.
- LEITE, A. C. F.; ALVES, F. C. Trabalho, maternidade e permanência no ensino superior. *Universidade Federal do Cariri, Brejo Santo, CE, Brasil*, 2022.
- MONTGOMERY, D. C. *Design and Analysis of Experiments*. 8. ed. Nova Iorque: John Wiley & Sons, 2013.
- NOGALES, A. M. et al. Ter ou não ter filhos? um olhar sobre a maternidade no distrito federal. *Observa DF*, 2023.
- OBANDO, J. M. A linha tênue entre maternidade e evasão escolar. 2018.
- OLIVEIRA, L. Desigualdades educacionais e suas implicações sociais. *Revista de Ciências Sociais*, v. 33, n. 2, p. 123–137, 2018.
- PEREIRA, R.; LIMA, C. Educação e desigualdade: Um estudo sobre a conclusão do ensino médio no Brasil. *Cadernos de Pesquisa em Educação*, v. 30, n. 2, p. 65–80, 2021.
- QUEIROZ, C. Evasão escolar no ensino médio atinge meio milhão de jovens por ano, aponta estudo. 2023.
- RABELLO, A. M. M.; SILVA, A. d. G.; ZORNING, S. Cuidar das mães, que cuidam dos filhos, que cuidam de...: observações e reflexões sobre a gravidez na adolescência. *Primórdios-CPRJ, Rio de Janeiro*, v. 1, n. 1, p. 43–56, 2010.
- SILVA, A. M. da et al. Evasão escolar: causas e estratégias para enfrentamento. Universidade Federal de Minas Gerais, 2020.
- SILVA, J. Análise socioeconômica da população do distrito federal. *Revista de Estudos Sociais*, v. 15, n. 2, p. 120–135, 2021.
- SILVA, J. O impacto do peso amostral nas análises estatísticas. *Revista de Metodologia Estatística*, v. 16, n. 1, p. 22–34, 2021.
- SILVA, J.; SANTOS, A. O impacto da educação na inserção no mercado de trabalho. *Revista Brasileira de Educação*, v. 25, n. 4, p. 345–360, 2020.
- SILVEIRA, P. et al. Ser mulher, mãe e universitária: narrativas de estudantes do curso de pedagogia da UFSC. Florianópolis, SC, 2019.
- SOUSA, C. R. d. O. et al. Fatores preditores da evasão escolar entre adolescentes com experiência de gravidez. *Cadernos Saúde Coletiva, SciELO Brasil*, v. 26, p. 160–169, 2018.
- TAUIL, T. I. Políticas públicas para mães universitárias: um estudo bibliográfico. *Universidade Federal de Ouro Preto*, 2019.
- TRIOLA, M. F. *Essentials of Statistics*. 6. ed. Boston: Pearson, 2017.
- URPIA, A. M. d. O. Tornar-se mãe no contexto acadêmico: narrativas de um self participante. *Universidade Federal da Bahia*, 2009.