

CURSO DE ESPECIALIZAÇÃO EM CIÊNCIA DE DADOS
DISCIPLINA: PROJETO DE CIÊNCIA DE DADOS
GRUPO: ANNE KELLY, JAIR ROCHA, MISAEL SEVERINO

TÍTULO:

1 INTRODUÇÃO

1.1 Contextualização

Segundo a Pesquisa Nacional por Amostra de Domicílios Contínua - PNAD 2019, o Brasil possui 11 milhões de analfabetos (maiores de 15 anos que não sabem ler e escrever). Isso corresponde a 6,6% do total da população do país nesta faixa etária. O Nordeste é a região que apresenta a maior taxa (13,9%), número quatro vezes maior do que nas regiões Sul e Sudeste (ambas com 3,3%)¹. Países vizinhos como Argentina, Uruguai e Chile apresentam taxas bem menores que as do Nordeste e que a média nacional, 2,1%, 2,0% e 1,4%, respectivamente².

O nível de escolarização da população alfabetizada também é algo importante a se considerar quando estamos discutindo o acesso à Educação como direito fundamental para o desenvolvimento da cidadania da população de um território. A proporção de pessoas de 25 anos ou mais que finalizaram a educação básica obrigatória (Ensino Fundamental) é de 48,8%, e apenas 27,4% concluíram o Ensino Médio¹. No Nordeste, três em cada cinco adultos (60,1%) não completaram o ensino médio.

Um olhar mais apurado - mas ainda bastante inicial - sobre os dados do acesso à educação no Brasil, permite-nos constatar que o problema é histórico e tem profundas raízes socioculturais, herança de nossa experiência escravocrata. Para todos os recortes possíveis, negros e pardos são proporcionalmente maioria entre analfabetos e os menos escolarizados. Também constatamos forte correlação entre renda familiar e escolaridade³, o que nos leva a indagar sobre as múltiplas forças que atuam tanto para que os mais pobres encontrem

¹ Disponível em:

<

Acessado em: 01 de julho de 2022.

² Disponível em: <<https://www.cia.gov/the-world-factbook/about/archives/>>. Acessado em: 01 de julho de 2022.

³ Disponível em:

< [Acessado em: 01 de julho de 2022.](https://opiniao.estadao.com.br/noticias/notas-e-informacoes,a-escolaridade-e-a-renda,70002758918#:~:text=A%20experi%C3%Aancia%20internacional%20indica%20uma,trabalho%20e%20possibilita%20maiores%20rendimentos.>)

dificuldades para acessar a educação formal, como a própria ausência da escolarização contribuindo para a não inserção no mercado de trabalho.

Dito isso, passamos a considerar o acesso à educação como meio fundamental para a mobilidade social e construção de uma sociedade mais igualitária. A Constituição Federal de 1988 prevê que a educação básica é um direito do cidadão e um dever do estado, que devem seguir padrões mínimos de qualidade, conforme os artigos 2º e 4º da Lei de Diretrizes e Bases da Educação Nacional – LDBEN - Lei nº 9.394/96 e alterações dadas pela Lei nº 12.796/2013 (BRASIL, 1988; 1996; 2013).

Neste cenário, surgem as políticas para educação de jovens e adultos como forma de contribuir para atenuar a dívida que o estado brasileiro tem com todos os cidadãos que não concluíram a educação básica até os 15 anos ou que não concluíram o ensino médio até os 18 anos. O Exame Nacional para Certificação de Competências de Jovens e Adultos (ENCCEJA) surge como etapa final avaliativa destas políticas de reparação. Através dos dados do ENCCEJA, este trabalho objetiva realizar uma investigação para identificar os perfis dos estudantes participantes deste exame.

1.2 Descrição do Problema

No planejamento e execução das políticas educacionais para jovens e adultos, assim como no processo de ensino e aprendizagem, compreender a pluralidade do corpo estudantil é um desafio a ser considerado. Diz a cartilha “Alunos e Alunas do EJA” da Secretaria de Educação Continuada, Alfabetização e Diversidade (SECAD) do Ministério da Educação (BRASIL, 2006) ser fundamental que os agentes educacionais conheçam este público na sua diversidade, “utilizando metodologias e práticas pedagógicas capazes de respeitar e valorizar suas especificidades” (BRASIL, 2006, p. 4).

Eliane Ribeiro Andrade (2012) e Miguel Arroy (2001), alertam para o tratamento dos alunos da EJA como uma massa sem identidade, resultado do fracasso escolar, repetentes, evadidos, defasados e passíveis de serem acelerados. A alternativa a esta postura seria aquela do compromisso com uma formação humana que passe a entender quem são esses sujeitos e que processos pedagógicos poderiam ser desenvolvidos para dar conta das necessidades e desejos de cada um (ANDRADE, 2012, p. 1).

1.3 Objetivos

Realizar uma investigação exploratória sobre os perfis dos participantes pernambucanos do Exame Nacional para Certificação de Competências de Jovens e Adultos (ENCCEJA) no ano de 2020.

1.4 Justificativa

Dada a amplitude etária e a diversidade sociocultural do povo pernambucano, do Sertão, Agreste, Zona da Mata e Região Metropolitana, podemos supor também em uma grande diversidade dos participantes da EJA. Pensar na escola, sala de aula e nas práticas pedagógicas de uma forma mais flexível, menos uniforme, com novos modelos de avaliação e convivência, parece não apenas interessante, mas também necessário. A inclusão educacional, a partir do reconhecimento dos participantes da política e das suas demandas, facilita o desenvolvimento de projetos de vida. Para isso, diz Andrade:

é essencial que os processos de formação de professores procurem conhecer as diferentes formas de atendimento da EJA, seus sujeitos, cotidianos e de, fundamentalmente, pensar as possibilidades de um dia-a-dia mais promissor para todos aqueles que encontram nessa modalidade educativa, muitas vezes, **a última chance de escolarização** (ANDRADE, p. 1, 2012).

É no intuito de conhecer melhor quem são estes estudantes que o presente trabalho se justifica. Quem são estes sujeitos no estado de Pernambuco? Trabalhadores, desempregados, jovens, idosos, do meio urbano ou rural, sexo, raça, entre outros aspectos mais gerais e específicos serão abordados a partir do que está disponível na base de dados do ENCCEJA. Pretendemos, assim, realizar uma análise que forneça subsídios para uma melhor condução da política de educação para jovens e adultos.

1.5 Escopo Negativo

O presente projeto não pretende realizar uma avaliação da política educacional. De igual forma, não pretende realizar uma avaliação sobre o ENCCEJA. Consideramos que as informações decorrentes da investigação

exploratória sobre os perfis dos estudantes pernambucanos do ENCCEJA 2020, por si, representam uma importante contribuição para compreensão do elemento mais importante desta política, que é o seu público-alvo. Daí, outras análises serão passíveis de serem realizadas por outras pesquisas ou aproveitadas pelo poder público para aprimoramento da política.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Área do negócio

As áreas de negócios abordadas no projeto correspondem a Educação e Direitos Humanos.

2.2 Mineração de dados

O Instituto Nacional de Pesquisas Educacionais Anísio Teixeira - INEP, vinculado ao Ministério da Educação, é o órgão federal responsável, dentre outros assuntos, pelo desenvolvimento de pesquisas estatísticas e indicadores educacionais. No seu portal de acesso à informação e dados abertos, localizamos a base de dados discriminada como “microdados do ENCCEJA 2020”. O processo de mineração dos dados, que buscou encontrar padrões, associações, correlações e novos subconjuntos de dados, se deu sobre a referida base.

O intuito da prospecção desses dados foi o de reunirmos informações suficientes para realizarmos uma profunda análise descritiva dos dados, de modo que esta pudesse ser entendida como a própria descrição dos perfis dos participantes do ENCCEJA.

2.3 Trabalhos relacionados

Neste ano de 2022, o ENCCEJA completa 20 anos de sua criação. Encontramos significativo volume de artigos, monografias, dissertações e teses produzidas ao longo de todo esse período. Trabalhos que utilizaram os microdados do ENEM também foram estudados, quando verificada a proximidade destas pesquisas com os objetivos do nosso trabalho.

Pudemos localizar muitos trabalhos que versavam sobre a relação entre perfis socioeconômicos, desempenho escolar e evasão, como os trabalhos de ARTONI (2022), PINTO (2020). Encontramos pesquisas das áreas da Educação, História, Sociologia e Ciência de Dados. Estes últimos foram particularmente importantes, pois além de fornecer importante base teórica para compreensão do problema, também nos apresentaram sugestões técnicas e metodológicas, como o trabalho de SILVA (2021), por exemplo.

3 MATERIAIS E MÉTODOS

3.1 Descrição da base de dados

A base de dados escolhida tem por título “**Microdados do Exame Nacional para Certificação de Competências de Jovens e Adultos**” e está disponível no portal de acesso à informação e dados abertos do Instituto Nacional de Educação e Pesquisas Educacionais Anísio Teixeira (INEP). Neste portal foram disponibilizados apenas dados, no formato csv, referentes ao ENCCEJA aplicado no Brasil em 2020. No entanto, ao analisar o volume e qualidade dos dados, frente à nossa problemática a ser abordada na pesquisa, concluímos que ainda assim seria bastante interessante seguir com esta base de dados.

Existem duas tabelas principais sobre os dois tipos de aplicação do exame que são:

- A “**MICRODADOS_ENCCEJA_NACIONAL_REGULAR_2020**”, que apresenta dados a respeito do exame feito por jovens sem restrição de liberdade.
- A “**MICRODADOS_ENCCEJA_NACIONAL_PPL_2020**”, que contém dados sobre o exame realizado por pessoas privadas de liberdade (PPL) e jovens sob medidas socioeducativas.

Entretanto, a base de dados é composta por 4 tabelas, onde uma delas possui dados referentes aos participantes regulares, duas sobre os participantes PPL (pessoas privadas de liberdade) e uma sobre as particularidades das provas. Estruturalmente, temos a seguinte divisão das informações entre os arquivos:

Arquivos	Descrição
MICRODADOS_ENCCEJA_NACIONAL_REGULAR_2020	Dados cadastrais, respostas das provas e do questionário do ENCCEJA 2020 para candidatos nacionais sem restrição de liberdade (Regulares).

MICRODADOS_ENCCEJA_NACIONAL_PPL_2020	Dados cadastrais, respostas das provas do ENCCEJA 2020 para candidatos nacionais com restrição de liberdade (PPL's).
MICRODADOS_ENCCEJA_NACIONAL_PPL_QSE_2020	Informações das respostas ao questionário socioeconômico por candidatos nacionais com restrição de liberdade (PPL's).
MICRODADOS_ENCCEJA_ITENS_PROVA_2020	Estrutura geral dos itens das provas

Tabela 1: Descrição dos arquivos da base de dados

3.2 Análise descritiva dos dados

Importamos as 3 (três) bases de dados referentes ao ano de 2020 do ENCCEJA, descritos na tabela 1, no Google Colaboratory, e para todos os arquivos geramos um data frame com a função “head”, para termos uma primeira visualização dos dados. Abaixo, segue resumo das cinco primeiras linhas das três bases importadas.

Base REGULAR

```
[28] df_regular.head(5)
```

	NU_INSCRICAO	NU_ANO	TP_CERTIFICACAO	TP_FAIXA_ETARIA	TP_SEXO	CO_UF_PROVA	SG_UF_PROVA	NO_ENTIDADE_CERTIFICADORA	IN_PROVA_LC	IN_PROVA_MT	...	Q57	Q58	Q59	Q60	Q61	Q62
0	200001318359	2020	2	16	F	43	RS	SECRETARIA DE ESTADO DE EDUCAÇÃO DO RIO GRANDE...	0	1	...	NaN	B	NaN	NaN	NaN	N
1	200002815765	2020	2	12	M	35	SP	SECRETARIA DE ESTADO DE EDUCAÇÃO DO RIO GRANDE...	1	1	...	C	B	NaN	NaN	NaN	N
2	200002310451	2020	2	8	M	53	DF	SECRETARIA DE ESTADO DE EDUCAÇÃO DO DISTRITO F...	1	0	...	NaN	A	A	A	A	B
3	200001895444	2020	2	11	F	13	AM	SECRETARIA DE ESTADO DE EDUCAÇÃO DO AMAZONAS	1	1	...	C	A	B	B	A	B
4	200002181288	2020	2	13	M	35	SP	SECRETARIA DE ESTADO DE EDUCAÇÃO DE SÃO PAULO	1	1	...	NaN	A	A	B	A	B

5 rows × 109 columns

Base PPL

```
[30] df_ppl2020.head(5)
```

	NU_INSCRICAO	NU_ANO	TP_CERTIFICACAO	TP_FAIXA_ETARIA	TP_SEXO	CO_UF_PROVA	SG_UF_PROVA	NO_ENTIDADE_CERTIFICADORA	IN_PROVA_LC	IN_PROVA_MT	...	TX_GABARI
0	200001856888	2020	2	11	M	42	SC	SECRETARIA DE ESTADO DE EDUCAÇÃO DE SANTA CATÁ...	1	1	...	CABBCADDCCADCBBABDBABABDCC
1	200001871210	2020	1	14	M	35	SP	SECRETARIA DE ESTADO DE EDUCAÇÃO DE SÃO PAULO	1	1	...	DABDBACBCDBCADDDBCBDBBDDCD
2	200001856889	2020	1	15	F	35	SP	SECRETARIA DE ESTADO DE EDUCAÇÃO DE SÃO PAULO	1	1	...	
3	200001866885	2020	1	11	M	42	SC	SECRETARIA DE ESTADO DE EDUCAÇÃO DE SANTA CATÁ...	1	1	...	DABDBACBCDBCADDDBCBDBBDDCD
4	200001871211	2020	1	14	M	35	SP	SECRETARIA DE ESTADO DE EDUCAÇÃO DE SÃO PAULO	1	1	...	DABDBACBCDBCADDDBCBDBBDDCD

5 rows × 43 columns

Base PPL_QSE

```
[32] df_pplqse.head(5)
```

	Q001	Q002	Q003	Q004	Q005	Q006	Q007A	Q007B	Q007C	Q007D	...	Q039J	Q039K	Q039L	Q040A	Q040B	Q040C	Q040D	Q040E	Q040F	Q040G
0	A	A	D	A	A	A	B	B	B	B	...	B	A	B	A	A	A	A	B	A	B
1	A	C	A	A	A	B	B	B	A	B	...	B	A	B	A	A	A	A	B	B	B
2	A	NaN	NaN	A	NaN	NaN	A	A	A	B	...	B	B	B	A	B	B	B	B	B	B
3	B	A	A	A	A	B	B	A	A	B	...	B	B	A	A	A	A	B	B	A	B
4	A	NaN	NaN	NaN	NaN	NaN	A	A	A	A	...	A	A	A	A	A	A	A	A	A	A

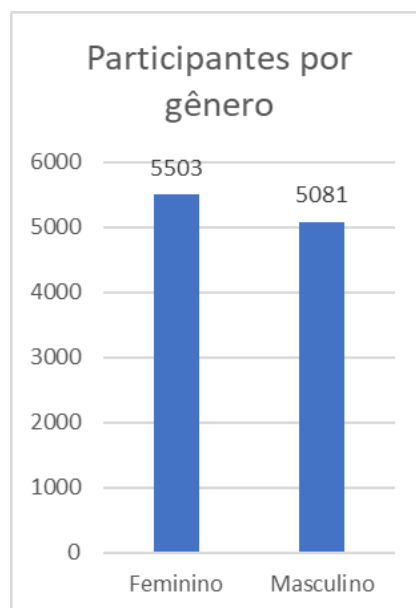
5 rows × 85 columns

Considerando o objetivo de nossa pesquisa, extraímos todas as informações referentes às provas realizadas no Estado de Pernambuco e que foram realizadas por participantes sem restrição de liberdade (base dos regulares) e salvamos em um arquivo “.csv”, para manipulação dos dados por outros softwares. Para tanto, utilizamos os códigos abaixo:

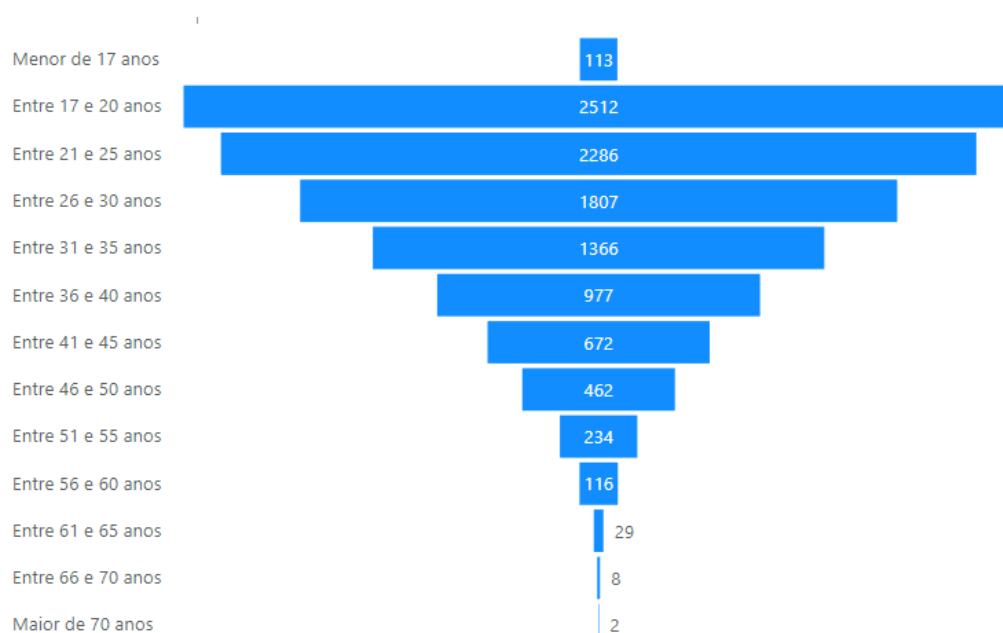
```
[ ] df_regular = pd.read_csv("/content/drive/MyDrive/Projeto CECEJA - TCC/TCC/DADOS/MICRODADOS_ENCCEJA_NACIONAL_REGULAR_2020.csv", encoding='latin-1')  
[ ] pe = df_regular[df_regular['SG_UF_PROVA'] == 'PE']  
[ ] pe.to_csv("PE_Regular")  
[ ] len(pe)  
10584
```

Importamos a base para o Power BI e avançamos na realização da análise descritiva. Também utilizamos o Excel e o RStudio para manipulação dos dados e realização de alguns testes específicos.

Após a extração dos dados, considerando apenas o Estado de Pernambuco, nossa base apresentou 10.584 ocorrências. Observamos um quantitativo de 5503 participantes do sexo feminino e 5081 participantes do sexo masculino. Destacamos que uso do termo “sexo” neste trabalho apenas preserva a informação originalmente apresentada nos questionários respondidos pelos participantes. Antecipamos aqui a sugestão de ajuste para o termo “gênero” nas próximas edições do ENCCEJA, assim como a ampliação das categorias de respostas para que possam abranger de forma mais adequada a diversidade de gêneros hoje existentes.



Também realizamos verificação da distribuição dos participantes por faixa etária.



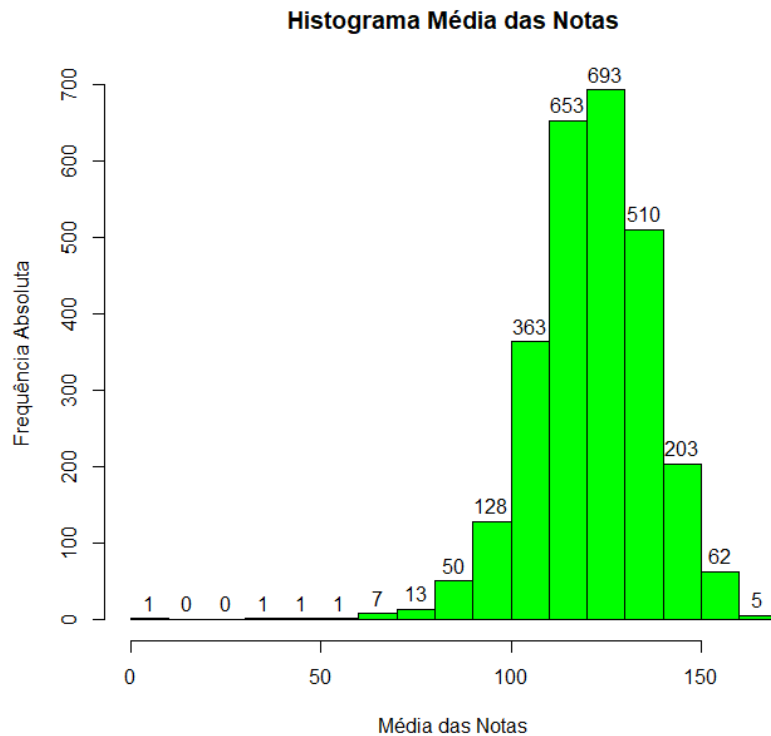
Verificamos os quantitativos de inscrições por cada uma das quatro provas, assim como o quantitativo daqueles que as realizaram ou se ausentaram.



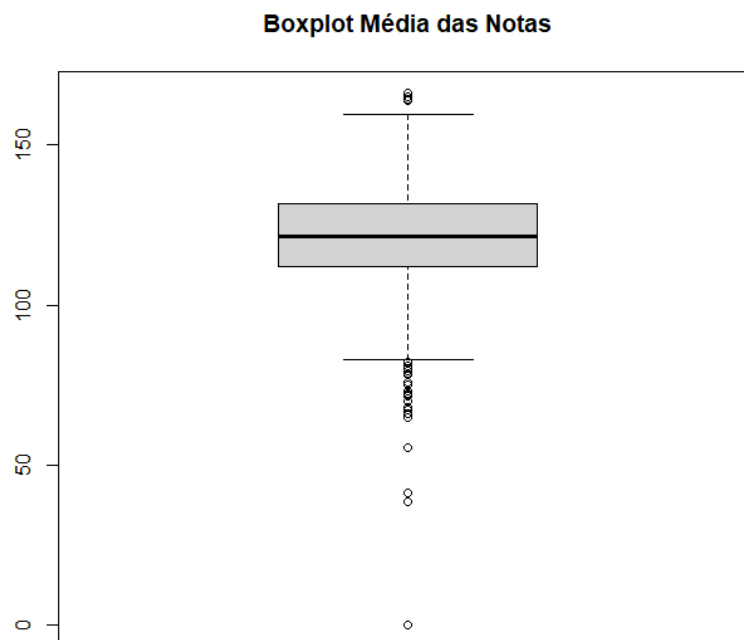
Abaixo, trazemos as informações de média, mediana, moda, desvio padrão e distribuição por quartis das notas de cada uma das provas de todos os participantes.

	LINGUAGENS	MATEMÁTICA	CIÊNCIAS DA NATUREZA	CIÊNCIAS HUMANAS
Média	119	120	137	121
Mediana	120	121	139	122
Moda	120	123	134	117
Desvio padrão	17.28	20.09	14.94	18.69
Mínimo	0	0	0	0
1 Qua	109	107	129	110
2 Qua	120	121	139	122
3 Qua	131	133	147	133
4 Qua	170	182	181	162

Após a importação dos dados para o RStudio, também pudemos realizar outras análises. Abaixo, apresentamos o histograma da média das notas das quatro provas de todos os participantes. É importante esclarecer que o cálculo da média levou em consideração que um único participante pode ter realizado apenas uma, duas, três ou todas as quatro provas.



Também foi possível analisar a média das notas a partir do diagrama boxplot abaixo:



As informações abaixo, também retiradas do R, nos ajudam a realizar uma leitura precisa do boxplot.

```
> summary(base$MEDIA_DAS_NOTAS)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   0.0   112.0   121.5   121.1   131.8   166.2
> |
```

A média das médias de todas as provas corresponde a “121.1”. A menor média ficou em “0.0” e a maior ficou em “166.2”. Podemos verificar outliers, tanto pela presença de notas abaixo do primeiro quartil, como por outras que ficaram acima do quarto quartil.

O desenvolvimento e aprofundamento da análise descritiva representa o próprio objetivo desta pesquisa: realizar uma investigação sobre os perfis dos participantes pernambucanos do ENCCEJA 2020. Por essa razão, as informações extraídas da base, tabelas, gráficos e suas respectivas análises se encontram no capítulo 4, da Análise e Discussão dos Resultados.

3.3 Pré-processamento dos dados

O uso variado das múltiplas ferramentas de manipulação e análise de dados nos permitiu explorar o potencial de cada uma delas. Por outro lado, cada uma também demandou procedimentos específicos para o pré-processamento dos dados.

Podemos considerar que uma primeira etapa do pré-processamento foi realizada no Google Colab, através da linguagem Python e fazendo uso da biblioteca Pandas. Como já apresentado na sessão anterior, criamos outro dataframe filtrando apenas os participantes do ENCCEJA que realizaram provas no Estado de Pernambuco. Esse dataframe foi salvo num arquivo “.csv” e importado para o PowerBI, Excel e para o RStudio.

Para as análises que foram realizadas no Excel, realizamos seguidos filtros e excluímos os dados ausentes para geração das tabelas e gráficos. Também realizamos a conversão de dados categóricos para melhor visualização. Por exemplo, as informações referentes a escolaridade do participante estavam codificadas, conforme tabela abaixo:

A	1ª série do ensino fundamental (antigo primário, 1º grau)
B	2ª série do ensino fundamental (antigo primário, 1º grau)
C	3ª série do ensino fundamental (antigo primário, 1º grau)
D	4ª série do ensino fundamental (antigo primário, 1º grau)
E	5ª série do ensino fundamental (antigo ginásio, 1º grau)
F	6ª série do ensino fundamental (antigo ginásio, 1º grau)
G	7ª série do ensino fundamental (antigo ginásio, 1º grau)
H	8ª série do ensino fundamental (antigo ginásio, 1º grau)
I	1ª série do ensino médio (antigo 2º grau)
J	2ª série do ensino médio (antigo 2º grau)
K	3ª série do ensino médio (antigo 2º grau)

Utilizamos o dicionário de dados como uma matriz de consulta chave-valor e, através de uma função “=procv()”, pudemos criar uma segunda planilha com a tradução de todas as informações. Também através de uma função “=procv()”, conseguimos converter os dados da base em valores numéricos. Essa conversão foi importante para outras análises, como os testes de correlação que serão apresentados no próximo capítulo do artigo. A título de exemplo, apresentamos abaixo o campo “Q05” (coluna G), que se refere à escolaridade do participante e o resultado da função “procv” de cada célula (coluna H), que converteu o dado categórico da escolaridade do participante em um valor numérico:

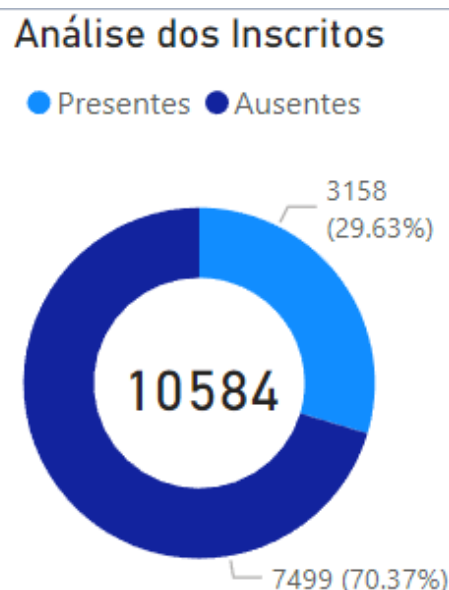
G	H
Q05	Q05_NUM
6ª série do ensino fundamental (antigo ginásio, 1º grau)	6
7ª série do ensino fundamental (antigo ginásio, 1º grau)	7
8ª série do ensino fundamental (antigo ginásio, 1º grau)	8
2ª série do ensino médio (antigo 2º grau)	10
3ª série do ensino médio (antigo 2º grau)	11
8ª série do ensino fundamental (antigo ginásio, 1º grau)	8
2ª série do ensino médio (antigo 2º grau)	10
7ª série do ensino fundamental (antigo ginásio, 1º grau)	7
8ª série do ensino fundamental (antigo ginásio, 1º grau)	8
5ª série do ensino fundamental (antigo ginásio, 1º grau)	5
1ª série do ensino médio (antigo 2º grau)	9
2ª série do ensino médio (antigo 2º grau)	10
1ª série do ensino médio (antigo 2º grau)	9

Os dados processados no Excel foram salvos em um novo arquivo “.csv” e importados para o RStudio. A partir daí, realizamos os ajustes para a correlata leitura dos tipos de dados, conforme trechos dos códigos abaixo:

```
# AJUSTE DOS TIPOS DE DADOS|
str(base)
base$MEDIA_DAS_NOTAS = as.numeric(base$MEDIA_DAS_NOTAS)
base$ESCOLARIDADE_PARTICIPANTE = as.numeric(base$ESCOLARIDADE_PARTICIPANTE)
base$ESCOLARIDADE_PAI = as.numeric(base$ESCOLARIDADE_PAI)
base$ESCOLARIDADE_MAE = as.numeric(base$ESCOLARIDADE_MAE)
base$RENDAMENTO_FAMILIAR = as.numeric(base$RENDAMENTO_FAMILIAR)
base$INCLUSAO_DIGITAL = as.numeric(base$INCLUSAO_DIGITAL)
```

4 ANÁLISE E DISCUSSÃO DOS RESULTADOS

Organizamos a apresentação das análises a partir do tipo de participação dos candidatos do ENCCEJA. Em primeiro lugar, buscamos compreender o público geral que se inscreveu para o processo seletivo, independente de terem ou não realizado a prova. Em segundo lugar, buscamos entender aqueles que realizaram as provas, em comparação com os que deixaram de realizar. Por último, buscamos entender aqueles que foram aprovados, em comparação com os que não foram.



Ao todo, foram 10.584 inscritos no ENCCEJA 2020. Destes, apenas 3.158 realizaram pelo menos uma das 04 provas, enquanto 7.499 não compareceram para nenhuma delas, apesar de terem realizado a inscrição para tanto. Este número corresponde a uma abstenção de 70.37%.

Buscamos informações além da base de dados que explicassem este alto índice de faltas, no entanto, em consulta à literatura, compreendemos que o ano de 2020 não representou uma excepcionalidade, mas um agravamento de uma situação já comum.

No trabalho intitulado “A alta taxa de abstenção (ausência) do participante ao Exame Nacional para Certificação de Competência de Jovens e Adultos”, Rita Lemos Rocha (2020), explica que em edições anteriores (de 2010 a 2018) as abstenções variaram de 34% a 53%. No ano de sua pesquisa (2019), a taxa de abstenção alcançou 62%.

No ENCCEJA de 2020, considerando todo o Brasil, a abstenção foi de 71%⁴, aproximadamente a mesma do Estado de Pernambuco, 70,37%.

Rocha (2020, p. 118) explica que, no intuito de reduzir as abstenções, o INEP disponibilizou um formulário eletrônico para que os inscritos que se ausentaram em 2018 e que iriam participar em 2019, pudessem justificar suas ausências. Foram recebidas 262.392 justificativas. Deste total, 23% não tiveram dispensa do trabalho, conforme tabela abaixo:

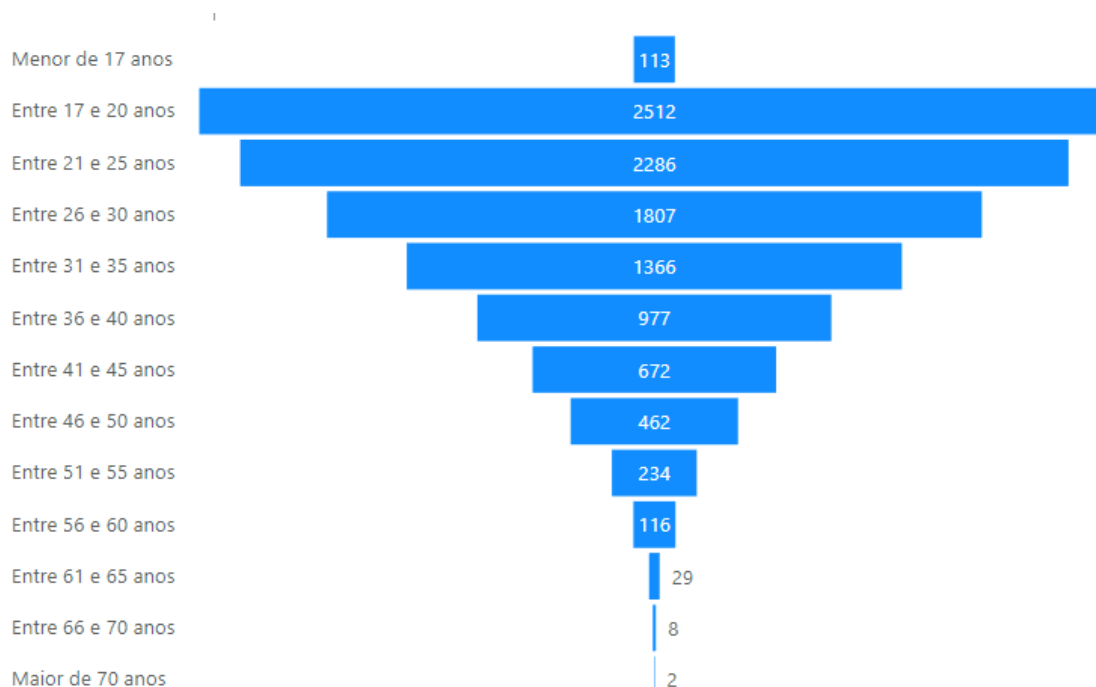
JUSTIFICATIVA APRESENTADA	PERCENTUAL
Não tiveram dispensa do trabalho	23%
Não apresentaram justificativa	21%
Problemas de saúde	16,6%
Maternidade / Paternidade recente	9,7%
Dificuldade com deslocamento	7,7%
Consultas médicas	4,7%
Morte de familiar	4,2%
Outra atividade curricular	3,4%
Viagem de trabalho ou mudança	3,1%
Assalto ou furto no caminho para a prova	2,6%
Acidente de trânsito	2,1%

⁴ Disponível em:

<<https://www.acheconcursos.com.br/governo/resultado-encceja-2020-e-divulgado-pelo-inep-veja-como-consultar-51138#:~:text=Foram%20mais%20de%201%2C6,do%20ensino%20fundamental%20ou%20m%C3%A9dio.>> Acessado em: 01 de julho de 2022.

Cumprimento de pena	1%
Cerimônia de casamento	0,9%

Ainda em relação ao total de inscritos, encontramos a seguinte distribuição por faixa etária:

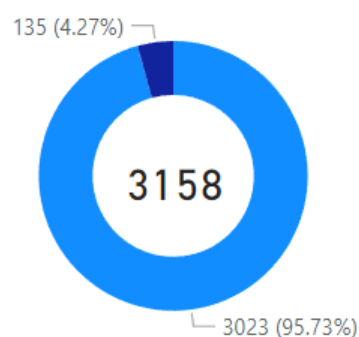


Em geral, o número de inscritos reduz, conforme a idade aumenta. A partir da leitura sobre o tema, foi possível levantar duas hipóteses sobre a questão. Primeiro, quanto mais pessoas se formam mais jovens, menos pessoas se formarão mais velhas. Segundo, que o retorno aos estudos se torna cada vez mais difícil com o avançar da idade.

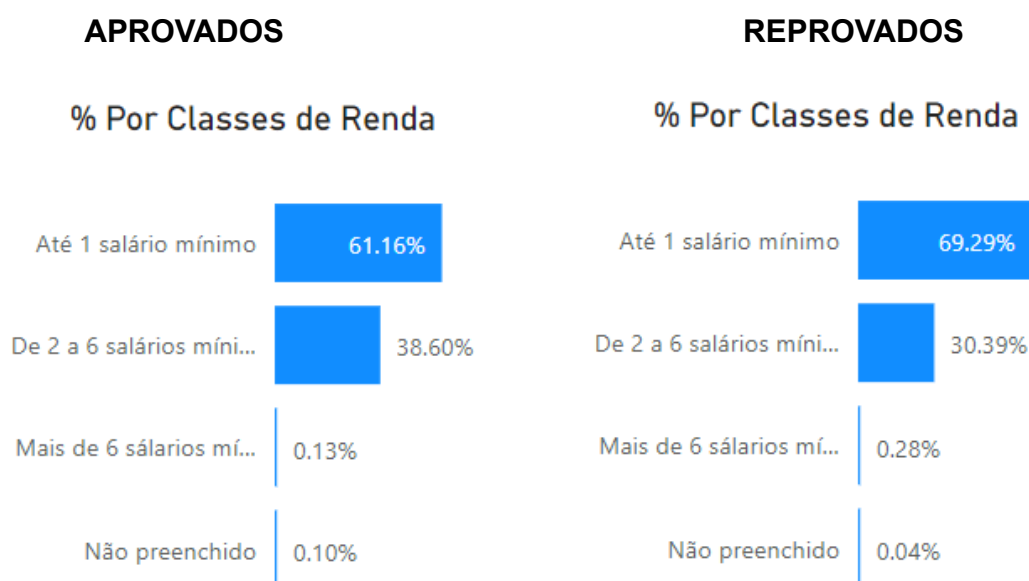
Entre aqueles que realizaram a prova, observamos um alto índice de aprovação, conforme gráfico abaixo:

Análise dos Presentes

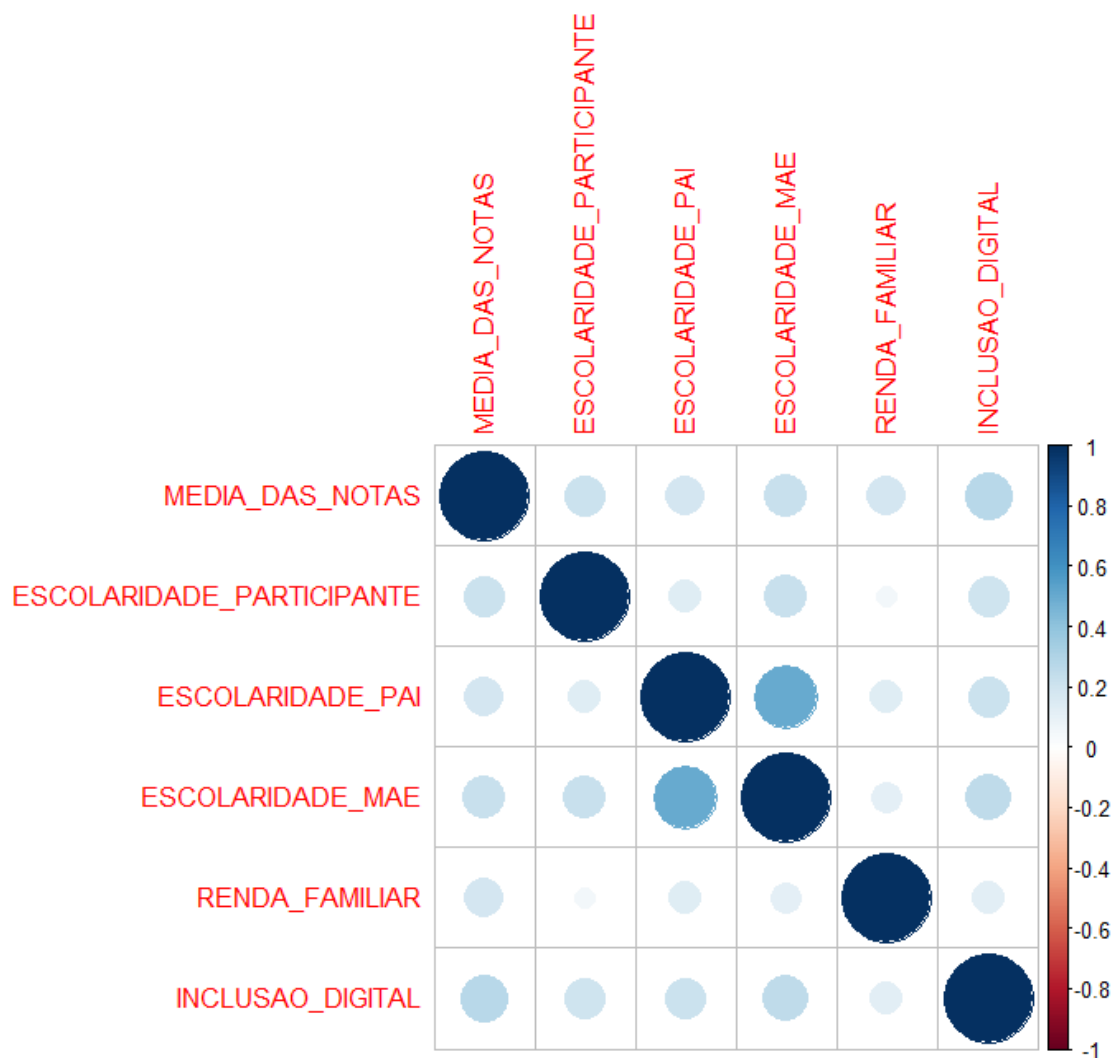
● Aprovados ● Reprovados



Dentre os 3023 aprovados, observamos que 61,16% vivem apenas com até 1 salário mínimo de renda familiar, percentual inferior aos que foram reprovados. Por outro lado, a quantidade percentual de pessoas que recebem dois ou mais salários mínimos aumenta entre os aprovados, em comparação com os reprovados. Estes dados, nos conduzem a supor a influência da renda familiar no desempenho da prova. Abaixo estão dispostos a distribuição percentual dos candidatos aprovados e reprovados por faixa de renda.



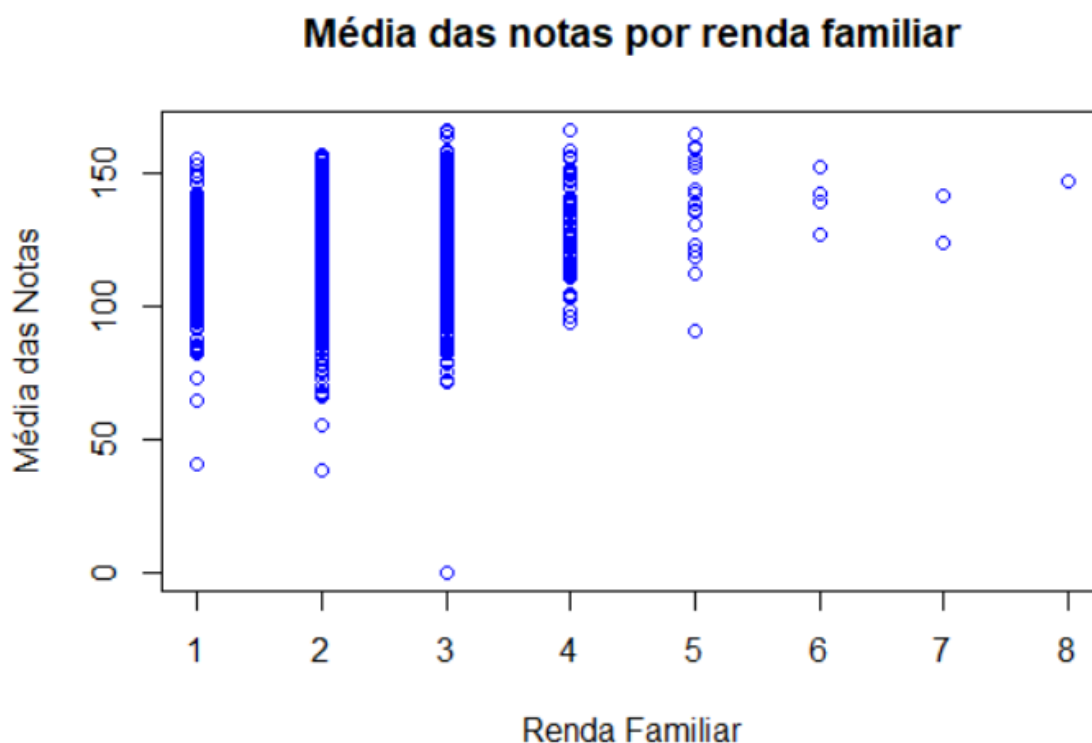
Para verificar uma possível correlação entre renda e desempenho, realizamos a importação de parte da base para o RStudio. Abaixo, apresentamos o gráfico e a tabela de correlações destas duas variáveis, além de outras que também serão aqui comentadas:



	MÉDIA DAS NOTAS	ESCOLARIDADE PARTICIPANTE	ESCOLARIDADE PAI	ESCOLARIDADE MÃE	RENDA FAMILIAR	INCLUSÃO DIGITAL
MÉDIA DAS NOTAS	1.000000	0.210175	0.186622	0.221781	0.183874	0.279799
ESCOLARIDADE PARTICIPANTE	0.210175	1.000000	0.138762	0.220912	0.052701	0.200768
ESCOLARIDADE PAI	0.186622	0.138762	1.000000	0.507553	0.133095	0.214735
ESCOLARIDADE MÃE	0.221781	0.220912	0.507553	1.000000	0.115341	0.255941
RENDA FAMILIAR	0.183874	0.052701	0.133096	0.115341	1.000000	0.127926
INCLUSÃO DIGITAL	0.279799	0.200768	0.214735	0.255941	0.127926	1.000000

A média das notas e renda familiar apresentaram uma correlação positiva muito fraca, de “0.18”. Apesar de resultar em um valor abaixo do esperado, considerando o que indica a literatura sobre a correlação destas duas variáveis,

ao olharmos para faixas de renda mais altas, observamos que as notas mínimas sobem progressivamente. Observe o gráfico de dispersão abaixo:



Cada número do eixo x representa um intervalo de faixa de renda familiar. A partir do número 5, que corresponde a faixa de 6 a 9 salários mínimos, encontramos apenas uma nota abaixo de 100 pontos. Ou seja, considerando esta faixa de renda ou mais, detectamos um único caso de reprovação.

Contudo, quando analisamos a escolaridade da mãe e do pai, podemos identificar uma correlação moderada, de “0,50”. Sugerindo então que casais com o mesmo nível de escolaridade possuem influência moderada no desempenho do participante do exame. É interessante perceber que este fenômeno pode ser encontrado no PNAD (Pesquisa Nacional por Amostra de Domicílios) produzido em 2014, visto que em matéria publicada no G1⁵, em novembro de 2016, a Gerente de Pesquisa do IBGE, Flávia Vinhaes, comenta que “A estrutura familiar parece ter uma importância muito grande em relação tanto ao nível de instrução dos filhos quanto aos índices de alfabetização”. Nesta pesquisa os menores percentuais de filhos sem instrução foi observado no grupo daqueles que moravam com ambos os pais, 10,8%.

⁵ Disponível em:

<

A pesquisa do PNAD também mostrou que nesse ano o índice de alfabetização foi maior dentre aqueles jovens que possuíam 15 anos e que moravam com a mãe. Neste caso, 92,2% para os que moravam com a mãe e com 88,1% para os que moravam com o pai. É interessante perceber que este fenômeno se repete na base de dados do ENCCEJA, onde o desempenho dos filhos que moram com a mãe é também ligeiramente superior aos que moram com o pai. Sendo esses valores “0,20” para mãe e “0,18” para o pai.

Outra correlação interessante, apesar de fraca, pode ser vista em inclusão digital e média de notas, alcançando “0,27”. Em uma matéria publicada em 2015 pelo jornal Gaúcha Zero Hora (GZH)⁶, é apresentado o resultado de um estudo da Organização para a Cooperação e Desenvolvimento Econômico (OCDE), onde é apontado que dispositivos eletrônicos em sala de aula não são suficientes para melhorar o desempenho dos alunos. Os resultados de exames que avaliaram compreensão de escrita, matemática e ciências, em alunos de 15 anos não apresentaram uma melhora significativa. Resultados estes apresentados no Programa para a Avaliação Internacional de Alunos (Pisa). É interessante notar que este mesmo comportamento é percebido nos dados apresentados desse projeto. Segundo o analista de educação da OCDE, “Não é apenas usando equipamentos eletrônicos de maneira intensiva que obtemos sucesso nos testes digitais”.

5 CONCLUSÕES E TRABALHOS FUTUROS

A base de dados do ENCCEJA, em seus mais de 110 (cento e dez) campos, apresentou-nos razoável número de possibilidades de investigação sobre os perfis dos participantes do exame. Percebemos semelhanças entre as razões para o abandono escolar e algumas das principais justificativas das ausências nas provas. Ambas dizem respeito ao atendimento de necessidades urgentes, como o trabalho para geração de renda, o cuidado com a casa e/ou dos filhos.

As abstenções podem ser consideradas como o principal desafio para o exame. Dado o alto índice de aprovação, caberia investigar melhor o desinteresse na realização das provas por grande parte dos inscritos. Assim, são recomendações deste trabalho: compreender as razões das ausências “injustificadas” e oportunizar alternativas para aqueles que apresentarem justificativas. E para futuros trabalhos, sugerimos a realização de estudos georreferenciados sobre os perfis de participantes. Os resultados decorrentes destes estudos, podem, eventualmente, fornecer informações valiosas para

⁶ Disponível em:

<<https://gauchazh.clicrbs.com.br/geral/noticia/2015/09/tecnologia-nao-e-suficiente-para-melhora-r-desempenho-escolar-diz-ocde-4847799.html>>. Acessado em: 01 de julho de 2022.

atuação local e preventiva por parte do poder público no sentido de reduzir as abstenções.

REFERÊNCIAS

ANDRADE, Eliane Ribeiro. Os Sujeitos Educandos na EJA. 2012. Disponível em: <http://www.forumeja.org.br/files/Programa%203_0.pdf>. Acessado em: 01 de julho de 2022.

ARTONI, Carla Baraldi. Relação entre perfil socioeconômico, desempenho escolar e evasão de alunos: escolas do campo e municípios rurais no estado de São Paulo. 2012. Dissertação (Mestrado em Administração de Organizações) - Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2012. doi:10.11606/D.96.2012.tde-23042012-150023. Acessado em: 31 de maio de 2022.

BRASIL. Constituição da República Federativa do Brasil de 1988. Disponível em: http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm> Acessado em: 01 de julho 2022.

_____. Lei nº 9.394, de 20 de dezembro de 1996. Estabelece as Diretrizes e Bases da Educação Nacional.

_____. Lei 12.796 de 04 de abril de 2013. Altera a Lei 9.394 de 20 de dezembro de 1996, que estabelece as diretrizes e bases da educação nacional, para dispor sobre a formação dos profissionais da educação e dá outras providências.

PINTO, Jucinara de Castro Almeida. TENÓRIO, Robinson Moreira. A Influência de fatores socioeconômicos no Desempenho Acadêmico dos Estudantes de Ensino Médio Integrado do IFBA/Campus Barreiras. Disponível em: <http://www.equidade.faced.ufba.br/sites/equidade.oe.faced.ufba.br/files/jucinar_a_pinto_robinson_tenorio_-_a_influencia_dos_fatores_socioeconomicos_no_desempenho.pdf>. Acessado em: 01 de julho de 2022.

ROCHA, Rita Lemos. A alta taxa de abstenção (ausência) do participante ao Exame Nacional para Certificação de Competência de Jovens e Adultos (ENCCEJA). Disponível em: <<http://mestrado.caedufjf.net/wp-content/uploads/2021/02/RITA-LEMOS-ROCHA.pdf>>. Acessado em 01 de julho de 2022.

SILVA, Tayná Arruda Câmara. Uma análise baseada em dados sobre o perfil socioeconômico dos estudantes de ensino médio das escolas públicas do Rio Grande do Norte e seu desempenho escolar. 2021. Disponível em:

<https://repositorio.ufrn.br/bitstream/123456789/43633/1/UmaAnaliseBaseada_Silva_2021.pdf>. Acessado em: 01 de julho de 2022.