

# Análise Exploratória ENEM 2021

## Contexto

O Exame Nacional do Ensino Médio (ENEM) é o principal exame de admissão ao ensino superior no Brasil. É dividido em cinco avaliações: Ciências da Natureza, Ciências Humanas, Linguagens e Códigos, Matemática e Redação.

A edição de 2021 foi marcada por diversos contratempos decorrentes da pandemia de Covid-19 e teve o menor número de inscritos desde 2007.

A base de dados contém campos com dados do participante, dados do exame e respostas ao questionário socioeconômico e pode ser encontrada em:

[https://download.inep.gov.br/microdados/microdados\\_enem\\_2021.zip](https://download.inep.gov.br/microdados/microdados_enem_2021.zip)

## Objetivo

Esta análise tem por fim responder às seguintes perguntas:

- Candidatos provenientes de escolas públicas e privadas tem o mesmo desempenho?
- Quanto maior a renda familiar maior o desempenho?
- A escolaridade dos pais influencia no desempenho dos candidatos?
- Candidatos com acesso à internet tem maior desempenho?

## Metodologia

A viabilização da análise exploratória se deu por meio do Jupyter Notebook em conjunto com as bibliotecas: Numpy e Pandas para manipulação de dados, Matplotlib e Seaborn para visualização de dados.

Também foram utilizados o método QQplot da biblioteca Statsmodel para verificar normalidade de dados e possibilitar a realização do teste de hipóteses. Para comparação de médias de amostras foi utilizado o teste T de Student (contido na biblioteca Scipy),

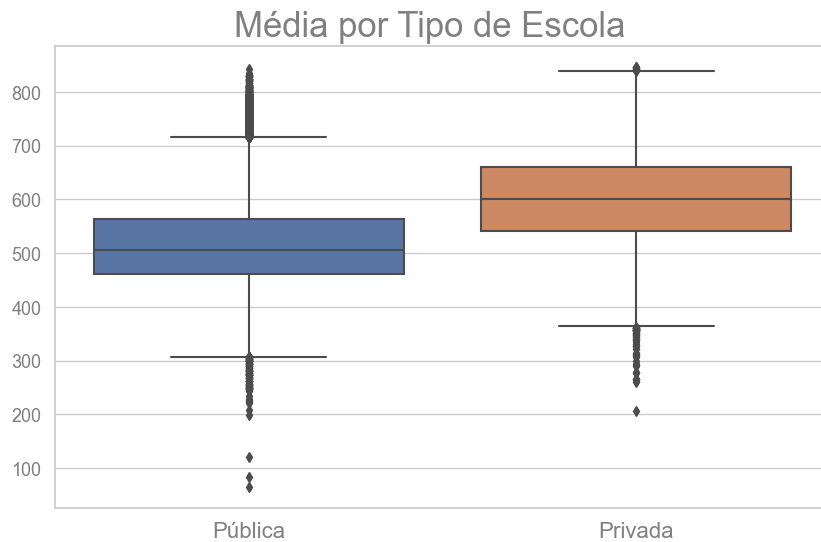
## Resultados

Durante a análise foram realizados testes de hipóteses com a finalidade de responder a cada uma das perguntas listadas no objetivo. Abaixo listamos os resultados dos testes para cada pergunta.

### Candidatos provenientes de escolas públicas e privadas obtêm o mesmo desempenho?

$H_0$	Candidatos provenientes de escolas públicas e privadas obtêm o mesmo desempenho.
$H_A$	Candidatos provenientes de escolas privadas obtêm melhor desempenho que candidatos de escolas públicas.

Logo abaixo temos um boxplot das médias de candidatos separados pelo tipo de escola que frequentaram:



Visualmente se nota a diferença de aproximadamente 85 pontos na média de candidatos provenientes de escolas e privadas.

É possível observar também que a média de candidatos provenientes de escola pública é menor que o média dos 25% piores candidatos de escola privada.

Outro ponto interessante a se destacar é a maior quantidade de outliers entre os candidatos de escolas públicas.

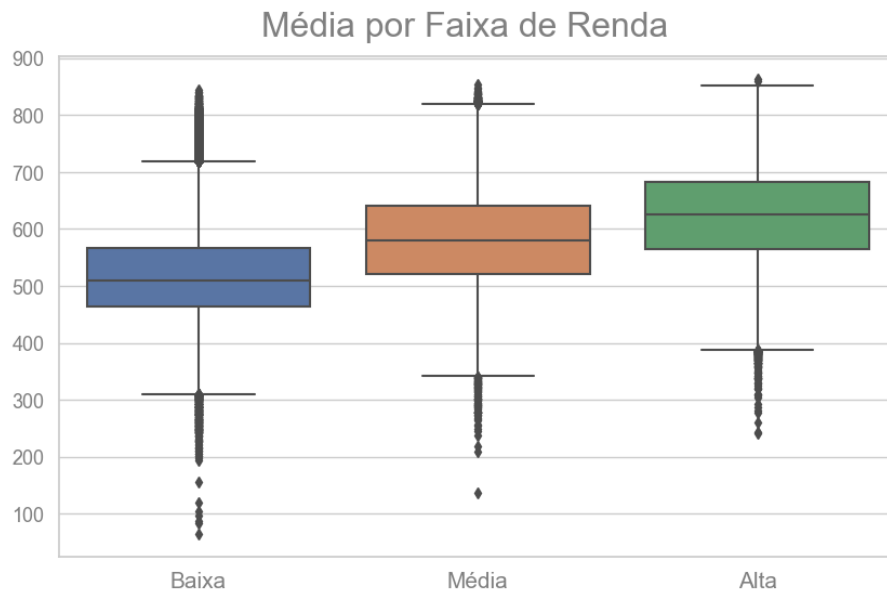
Os resultados do Teste T de Student, abaixo, permitem rejeitar  $H_0$ . Portanto podemos afirmar que candidatos de escolas privadas possuem melhor desempenho que candidatos de escolas públicas.

valor-t	-397,66
valor-p	0

#### Quanto maior a renda familiar melhor o desempenho?

$H_0$	Candidatos de diferentes faixas de renda obtém o mesmo desempenho.
$H_A$	Candidatos com maiores rendas obtém melhores desempenhos.

Abaixo representamos as médias dos candidatos agrupados por suas faixas de renda. Foram considerados renda baixa candidatos cujo renda familiar mensal não superasse R\$3300,00, renda média candidatos entre R\$3300,01 e R\$9900,00 e por fim, renda alta candidatos que somavam mais de R\$9900,01 mensais.



Para aplicação do teste de hipótese foram realizados dois testes independentes, um comparando as amostras de renda baixa e média e um segundo comparando as amostras de renda média e alta. Os resultados estão listados a seguir:

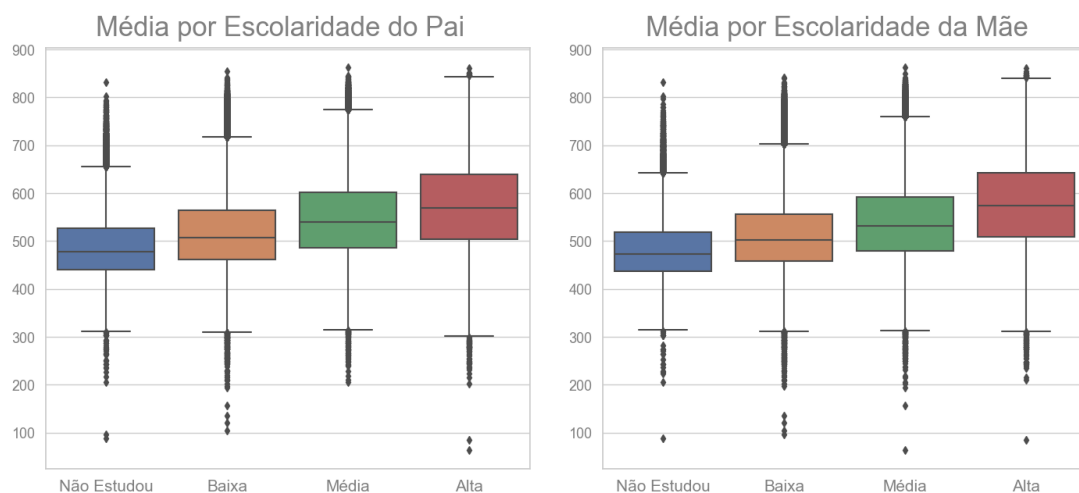
Renda	Baixa x Média	Média x Alta
valor-t	-460,04	-173,47
valor-p	0	0

Os testes são suficientes para rejeitar a hipótese nula e aceitar a hipótese alternativa de que quanto maior a renda familiar de um candidato melhor seu desempenho. Ademais, observando a diferença entre os valores-t obtidos podemos afirmar que a diferença no desempenho entre candidatos de baixa e média renda é maior que entre candidatos de média e alta renda.

#### A escolaridade dos pais influencia no desempenho dos candidatos?

$H_0$	Candidatos cujos pais tem diferentes níveis de escolaridade obtêm desempenhos semelhantes
$H_A$	Quanto maior a escolaridade dos pais do candidato melhor o desempenho

Assim como na pergunta anterior recategorizamos os dados para diminuir o número de categorias para 4: não estudou, escolaridade baixa, escolaridade média e escolaridade alta. Foram considerados escolaridade baixa os pais que cursaram até o final do Ensino Fundamental, os que completaram o Ensino Médio foram considerados de escolaridade média enquanto escolaridade alta são os pais que completaram o Ensino Superior. Abaixo os gráficos de média com as novas categorias e considerando a escolaridade paterna e materna respectivamente.



Assim como na pergunta anterior podemos identificar visualmente uma tendência de crescimento conforme cresce a escolaridade dos pais. Para o teste comparamos cada nível de escolaridade com o seu subsequente (uma vez que há grande diferença de média entre candidatos cujos pais não estudaram e pais com graduação completa, por exemplo). Seguem os resultados.

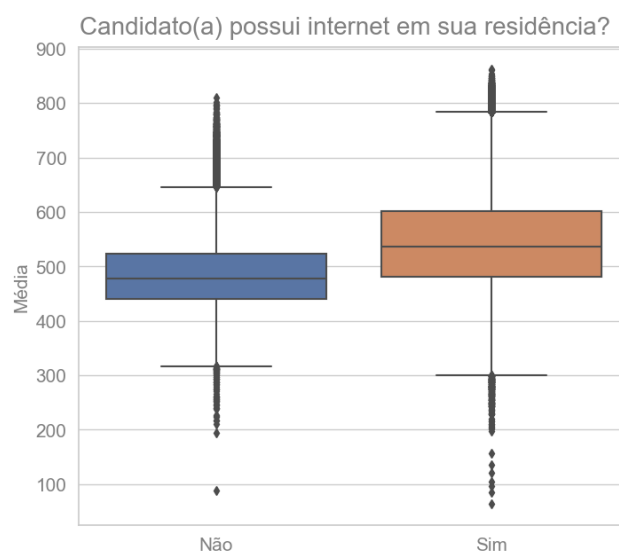
Escolaridade do Pai			
Escolaridade	Não estudou x Baixa	Baixa x Média	Média x Alta
valor-t	-115,57	-234,33	-165,52
valor-p	0	0	0
Escolaridade da Mãe			
Escolaridade	Não estudou x Baixa	Baixa x Média	Média x Alta
valor-t	-95,06	-225,03	-266,37
valor-p	0	0	0

Os valores obtidos permitem rejeitar a hipótese nula e aceitar a hipótese alternativa: quanto maior a escolaridade dos pais melhor o desempenho do candidato.

### Candidatos com acesso à internet tem melhor desempenho?

$H_0$	Candidatos com e sem acesso à internet obtêm desempenhos semelhantes.
$H_A$	Candidatos com acesso à internet obtêm melhor desempenho que candidatos sem acesso.

Por fim, última pergunta refere-se ao acesso à internet em casa. Abaixo apresento um boxplot das médias comparando candidatos com e sem acesso à internet.



É possível observar que a média das notas de candidatos com e sem acesso à internet não chega a 100 pontos, o limite superior, porém evidencia uma diferença superior a 100 pontos. Abaixo os resultados do teste de hipótese realizado.

valor-t	-368,54
valor-p	0

Os valores obtidos permitem rejeitar a hipótese nula e aceitar a hipótese alternativa: candidatos com acesso à internet obtêm melhor desempenho que candidatos sem acesso à internet.

## Conclusão

A análise de dados exploratória dos microdados do ENEM 2021 revelou valiosos insights sobre os fatores que influenciam o desempenho de candidatos. Durante a análise, foram feitas perguntas referentes ao tipo de escola frequentada pelos candidatos, à escolaridade dos pais, à faixa de renda familiar e ao acesso do candidato à internet.

Quanto ao tipo de escola, candidatos provenientes de escolas privadas apresentaram, em média, desempenho superior comparados aos de escolas públicas. Ademais, ficou constatado que quanto maior a renda familiar maior tende a ser o desempenho do candidato; o mesmo se observou da escolaridade dos pais, conforme maior o nível educacional dos pais melhor tende a ser o desempenho do estudante no exame.

O último ponto observado foi o acesso à internet onde foi possível observar que candidatos sem acesso não só tem desempenho médio inferior como também um limite superior muito abaixo daqueles com acesso à internet em casa.

Esta análise pode contribuir para a compreensão de quais condições influenciam o desempenho dos estudantes e podem ser utilizadas para fundamentar políticas públicas educacionais que aspirem à universalização do acesso ao ensino superior como também melhorias na educação básica no Brasil.