

Relatório

Testes de Associação no banco de dados Jovens e Política Online

Jhonatan Alves Passos
Roney Gregory Santos Melo

Outubro
2024

Sumário

1	Objetivos	3
2	Metodologia	3
3	Análise descritiva	6
3.1	Nível de participação online por gênero	6
3.2	Nível de participação online por orientação político-ideológica	7
3.3	Nível de participação online por religião	8
3.4	Opinião x Ofensas	10
3.5	Opinião x Divergentes	11
4	Resultados	12
4.1	Resultado do teste χ^2	14
4.2	Tamanho do efeito	14
5	Análise dos resíduos	15
5.1	Análise post-hoc	16
5.2	Nível de participação online por gênero	19
5.3	Gráfico dos resíduos - Nível de participação online por gênero	20
5.3.1	Variáveis com Significância	22
5.3.2	Variáveis sem Significância	22
5.3.3	Resumo das Associações	23
5.4	Nível de participação online por orientação político-ideológica	23
5.5	Gráfico dos resíduos - Nível de participação online por orientação político-ideológica	24
5.5.1	Variáveis com Significância	25
5.5.2	Variáveis sem Significância	27
5.5.3	Resumo das Associações	27
6	Considerações finais	28
7	Referências	32

Lista de Abreviaturas e Siglas

χ^2 : Qui-Quadrado

H_0 : Hipótese Nula

H_1 : Hipótese Alternativa

$N(0,1)$: Distribuição Normal com média 0 e desvio padrão 1

opi_diverg: Orientação Politico-Ideológica x Nível de disposição para debater com divergentes

opi_ofen: Orientação Politico-Ideológica x Já sofreu ofensas ao expressar opinião política na internet?

ponxgen: Nível de Participação Online x Gênero

ponxopi: Nível de Participação Online x Orientação Politico-Ideológica

ponxreli: Nível de Participação Online x Religião

1 Objetivos

Este relatório tem como objetivo verificar a existência de independência entre as variáveis do estudo ***”Jovens e Política Online”*** (dispostas posteriormente em um tópico específico), testando associações entre elas de forma detalhada. Busca-se compreender se há relações estatisticamente significativas que indicam uma conexão ou padrão de interdependência entre as variáveis analisadas, o que poderá fornecer insights relevantes sobre o comportamento dos jovens em relação à política na internet, e possíveis implicações práticas.

2 Metodologia

Todo o processamento dos dados, incluindo a preparação, limpeza, e análises estatísticas, foi realizado por meio do software R Studio, por intermédio dos pacotes `read_excel()`, `ggplot2()`, `table1()`, `janitor()` e `xtable()`, utilizando suas ferramentas robustas para análise de dados e visualização gráfica. Os scripts desenvolvidos permitiram a criação de gráficos e aplicação de testes estatísticos, garantindo a precisão e a eficiência dos resultados. A editoração e a formatação do relatório foram feitas no Overleaf, utilizando a linguagem LaTeX, o que proporcionou uma apresentação profissional e esteticamente organizada, permitindo um layout consistente e de fácil leitura.

Também foram elaboradas sheets(planilhas) no portal do docs do Google para padronizar tabelas de acordo com as normas da ABNT.

As variáveis de estudo foram:

- Gênero: Variável para auto-declaração de identidade de gênero.
- Religião: Variável para auto-declaração da religião do indivíduo.
- Orientação Politico-Ideológica: Contém classificações de opiniões políticas como, conservador, moderado e progressista. Geradas a partir de uma escala Likert.
- Já sofreu ofensas ao expressar opinião política na internet?: Variável dicotômica, onde as possibilidades eram sim ou não, indicando se o indivíduo sofreu ofensas ou não.
- Nível de disposição para debater com divergentes: atribui níveis de disposição para debates online sobre assuntos que divergem de suas opiniões.

Os cruzamentos realizados foram:

- Nível de Participação Online x Gênero

- Nível de Participação Online x Religião
- Nível de Participação Online x Orientação Politico-Ideológica
- Orientação Politico-Ideológica x Já sofreu ofensas ao expressar opinião política na internet?
- Orientação Politico-Ideológica x Nível de disposição para debater com divergentes

Devido à metodologia empregada para obtenção das informações, identificou-se a necessidade de realizar ajustes em campos do questionário. Com a devida autorização do pesquisadora responsável, foram feitas adaptações nas categorias de gênero e religião, com o objetivo de garantir a consistência dos dados, respeitando sempre as pautas de identidade e expressão de gênero, bem como os aspectos religiosos.

Considerando a possibilidade de campos não preenchidos durante a etapa sócio-demográfica do questionário, o banco de dados inicial apresentou uma amostra de 542 observações. Entretanto, após o pré-processamento, fez-se necessário a exclusão de valores nulos/ausentes, a amostra foi reduzida para algumas variáveis específicas. Para a variável "Nível de participação política online x Gênero", a amostra passou de 542 para 534 observações, devido à exclusão de respostas inadequadas ao contexto da pergunta (como aquelas relacionadas à orientação sexual, conforme detalhado posteriormente). Para a variável Religião, quatro observações foram excluídas devido à ausência de respostas.

Gênero:

O questionário permitia respostas subjetivas e objetivas, o que resultou na inclusão de termos como 'Bi', 'Demi girl', 'Gay', 'Hétero' e 'Hetero', que se referem à orientação sexual e não à identidade de gênero. Além disso, houve uma resposta 'Prefiro não responder', com apenas uma ocorrência, que, devido ao tamanho da amostra, não apresentou representatividade significativa. Essas respostas foram excluídas por não se adequarem ao campo específico de gênero.

Devido à baixa representatividade de pessoas transgênero na amostra e à insuficiência para atender aos pressupostos do teste qui-quadrado, foi necessário agrupar as informações de indivíduos transgêneros e cisgêneros, visando garantir a robustez estatística do estudo, sem desconsiderar as especificidades de cada identidade.

As respostas relacionadas à identidade de gênero, como 'Mulher Cisgênero', 'Mulher Transgênero', 'Mulher' e 'Muie', foram unificadas na categoria 'Mulher', compondo o grupo feminino da variável gênero. Da mesma forma, as respostas 'Homem Cisgênero', 'Homem Transgênero', 'Homem', 'Homem alfa' e 'Homem Masculino' foram unificadas na categoria 'Homem', compondo o grupo masculino.

Indivíduos que se identificaram como 'Não binário' foram mantidos em uma categoria separada, por ser uma identidade de gênero que não se enquadra no binário de gênero.

Religião:

As categorias de religião foram agrupadas da seguinte forma: Católicos, Evangélicos, Sem religião e Outros. Grupos que não alcançaram um valor representativo, como religiões de matriz africana, foram imputados na categoria Outros. Essa redução de dimensionalidade foi necessária para a aplicação do teste qui-quadrado.

Qui-Quadrado, Análise de Resíduos e Post-Hoc:

Realizamos uma análise de associação entre as variáveis utilizando o teste do qui-quadrado de independência de Pearson, onde a hipótese nula (H_0): Não existe diferença entre as frequências obtidas na amostra e as frequências esperadas, dado que a população tem distribuição de probabilidade conhecida. "Para tabelas 2 x 2, um único número, como a razão de chances, pode resumir a associação. Para tabelas I x J, raramente é possível resumir a associação por um único número sem alguma perda de informação.

No entanto, um conjunto de probabilidades de razão de chances ou outro índice resumido podem descrever certas características da associação."Conforme descrito por Agresti (2012) .

O autor destaca que "Como qualquer teste de significância, os testes de independência do qui-quadrado têm limitações de utilidade. Um valor P pequeno indica forte evidência de associação, mas fornece pouca informação sobre a natureza ou a força da associação. Os estatísticos há muito alertam sobre os perigos de confiar apenas nos resultados de testes qui-quadrado em vez de estudar a natureza da associação, por exemplo, (e.g. Berkson 1938; Cochran 1954)".

Para identificar as contribuições individuais das categorias para a associação global, foram calculados os resíduos padronizados e ajustados pela escala z-score. Para controlar a taxa de erro **tipo I** (rejeitar H_0 quando verdadeira, ou seja rejeitar a hipótese de independência quando não há associação) e validar as associações, empregamos testes post-hoc, como o ajuste de Bonferroni por Dunn (1961). As análises foram realizadas utilizando o software R. O teste do qui-quadrado de independência apresentou um p-valor < 0,05, indicando uma associação estatisticamente significativa entre as variáveis. Os testes post-hoc revelaram as categorias que mais contribuíram para essa associação com um p-valor < 0,0033, pós cálculo da correção de Bonferroni.

3 Análise descritiva

Nesta seção, foi realizada uma análise descritiva do banco de dados em estudo, com o objetivo de tornar os resultados mais visíveis e compreensíveis. Essa análise é de extrema importância, pois permite uma compreensão aprofundada das características das variáveis envolvidas, sendo possível obter uma visão geral do comportamento dos dados, proporcionando uma base sólida para as análises subsequentes e facilitando a interpretação dos resultados.

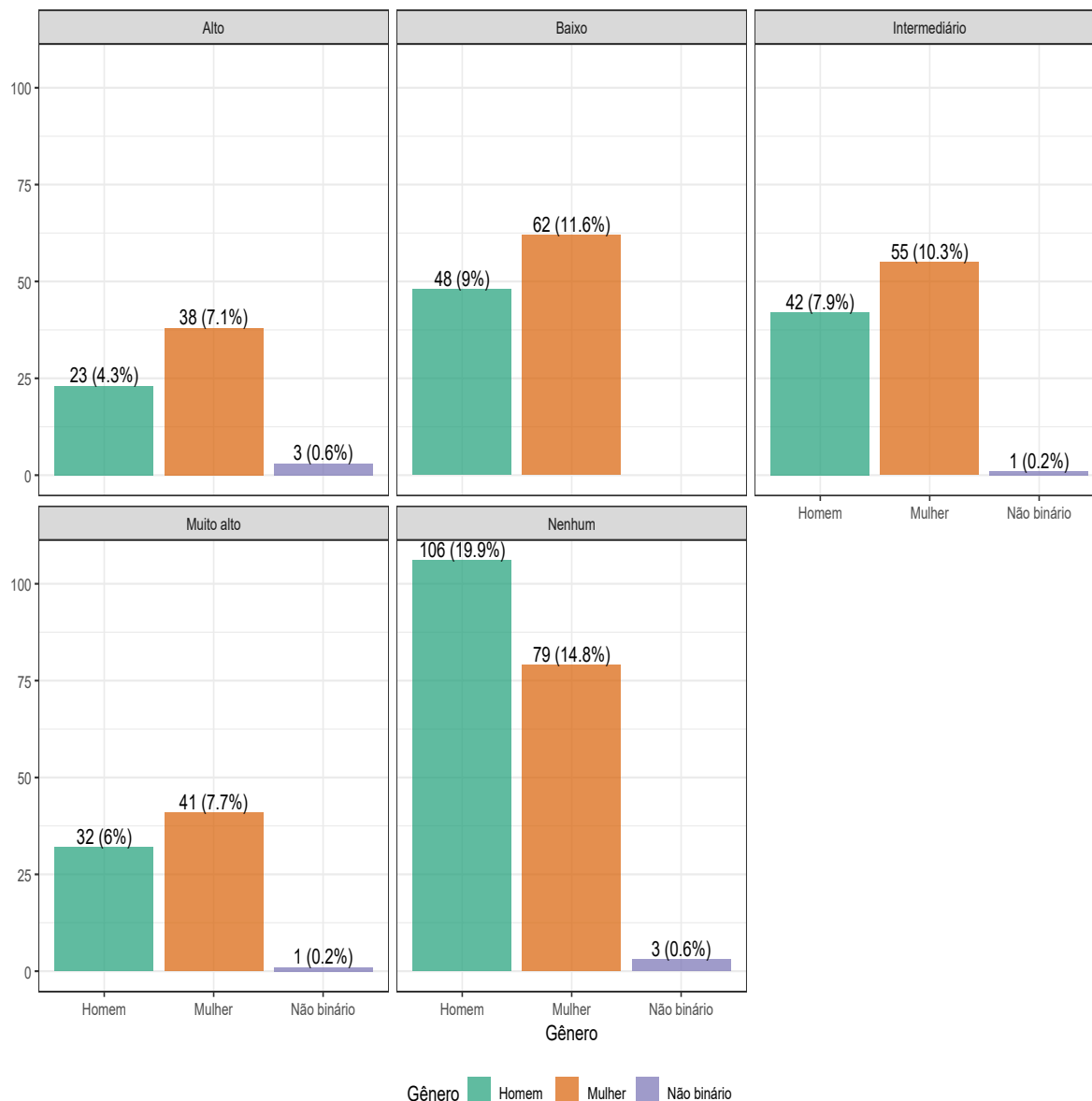
3.1 Nível de participação online por gênero

A amostra é composta por 534 indivíduos, distribuídos em cinco categorias de nível de participação online: "Muito alto", "Alto", "Intermediário", "Baixo" e "Nenhum". O gênero foi agrupado em três grupos: Homem, Mulher e Não Binário.

- O gênero masculino é predominante apenas na categoria de participação "Nenhum", representando 19,9% dos respondentes deste grupo.
- O gênero feminino predomina em todas as outras categorias de participação online, alcançando a maior porcentagem na categoria "Baixo", com 11,6% dos respondentes.
- O gênero não binário teve pouca representatividade em todas as categorias, com uma participação muito baixa, alcançando percentuais entre 0,2% à 0,6%.

Esses resultados sugerem que o gênero feminino possui uma maior participação online em comparação com os outros gêneros, o que fica evidente pelas porcentagens mais altas de mulheres nas categorias que indicam participação ativa ("Baixo", "Intermediário", "Alto" e "Muito alto").

Nível de participação online por gênero (n = 534)



Fonte: Elaboração própria, 2024

3.2 Nível de participação online por orientação político-ideológica

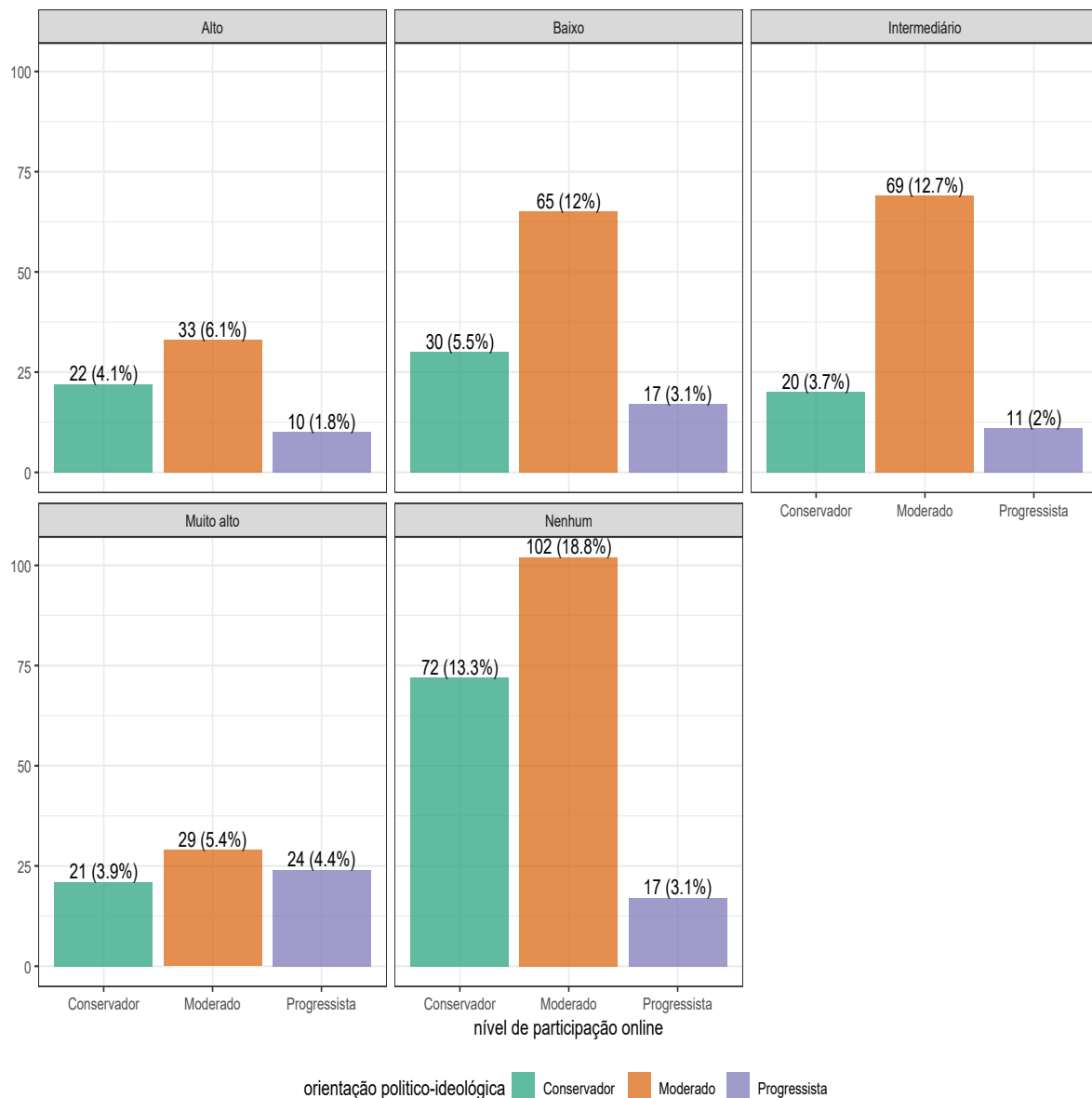
A amostra é composta por 542 indivíduos, distribuídos em cinco categorias de nível de participação online: "Muito alto", "Alto", "Intermediário", "Baixo" e "Nenhum". A orientação político-ideológica foi classificada em 'Conservador', 'Moderado' e 'Progressista'.

- A orientação conservadora não teve predominância nas categorias, tendo a sua maior porcentagem na categoria "Nenhum" (13,3%).
- A orientação moderada predomina em todas as outras categorias de participação on-line, alcançando a maior porcentagem na categoria "Nenhum", com (18,8%).

- A orientação progressista não teve predominância nas categorias, com a sua maior porcentagem na categoria “Muito alto” (4,4%).

Esses resultados sugerem que os indivíduos com orientação moderada possuem uma maior participação on-line em comparação com as outras orientações, o que fica evidente pelas maiores porcentagens em todas as categorias desse item.

Nível de participação online por orientação político-ideológica (n = 542)



Fonte: Elaboração própria, 2024

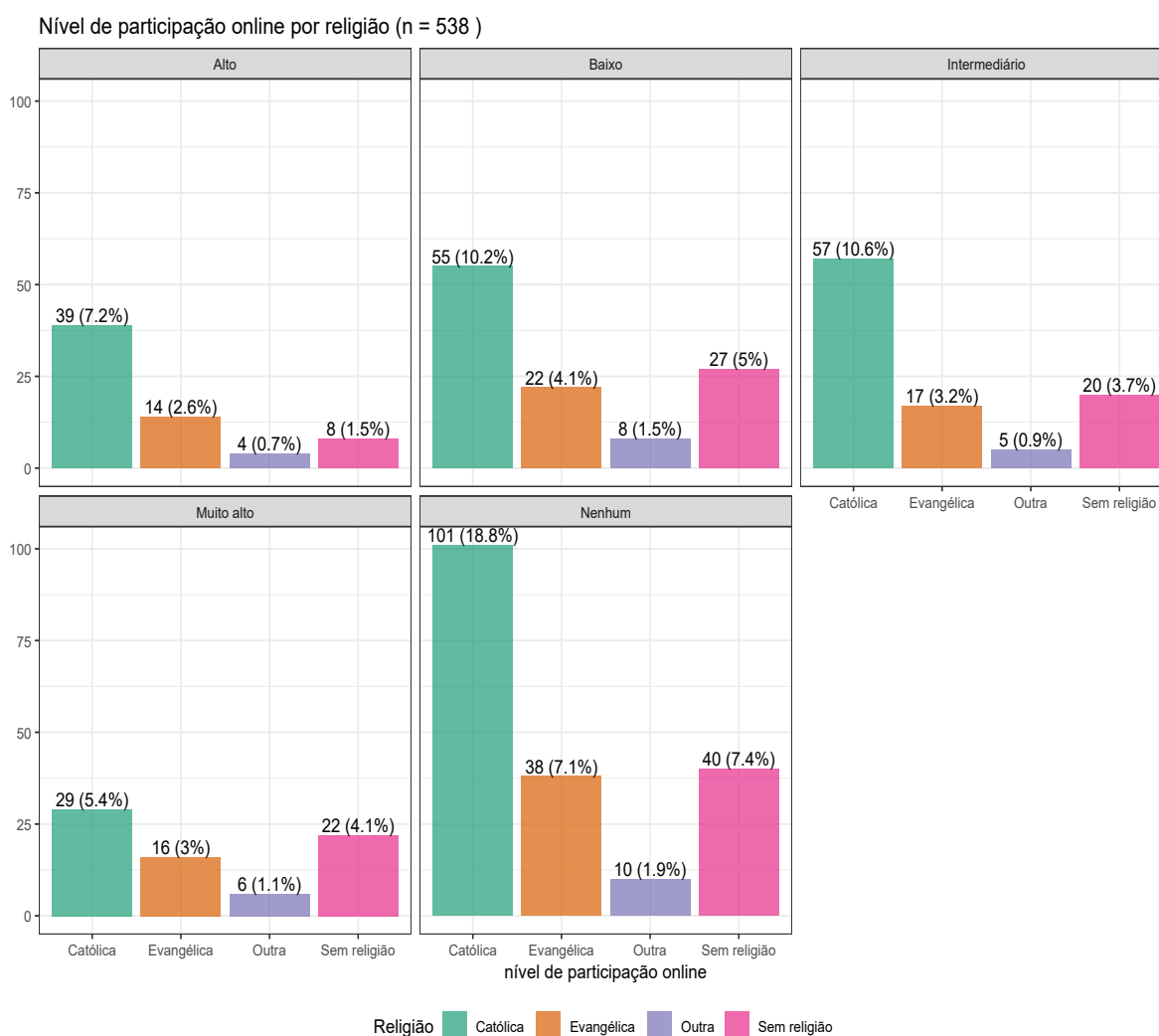
3.3 Nível de participação online por religião

A amostra é composta por 538 indivíduos, distribuídos em cinco categorias de nível de participação online: "Muito alto", "Alto", "Intermediário", "Baixo" e "Nenhum". A

religião foi agrupada em 'Católica', 'Evangélica', 'Outra' e 'Sem religião'.

- A religião católica foi predominante em todas as categorias, tendo a sua maior porcentagem na categoria "Nenhum" (18,8%).
- A religião evangélica não predominou nas outras categorias de participação on-line, alcançando a maior porcentagem na categoria "Nenhum", com (7,1%).
- As outras religiões também não tiveram predominância nas categorias, com a sua maior porcentagem na categoria "Nenhum" (9,3%). Os indivíduos sem religião tiveram destaque na categoria "Nenhum"(7,4%).

Esses resultados sugerem que os indivíduos com a religião católica possuem uma maior participação on-line em comparação com as outras religiões, o que fica evidente pelas maiores porcentagens em todas as categorias desse item.



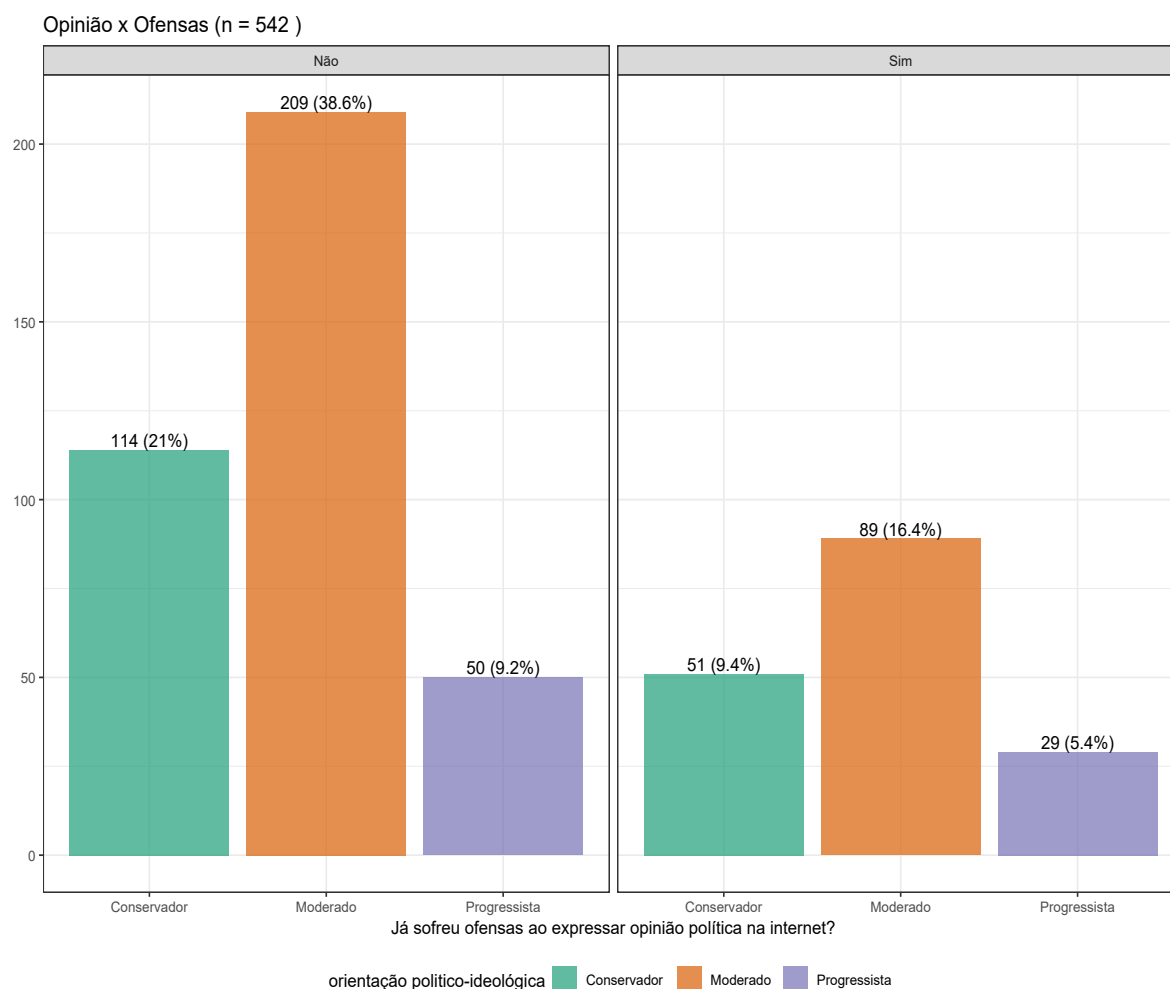
Fonte: Elaboração própria, 2024

3.4 Opinião x Ofensas

A amostra é composta por 542 indivíduos, neste quesito foi perguntado aos indivíduos se “Já sofreu ofensas ao expressar opinião política na internet?”, tendo as suas respostas classificadas em ‘Sim’ e ‘Não’. As orientações políticas foram classificadas em ‘Conservador’, ‘Moderado’ e ‘Progressista’.

- A orientação moderada foi predominante em todas as categorias, tendo a sua maior percentagem na categoria “Não” (38,6%).
- As orientações conservadora e progressista não foram predominantes, com as suas maiores percentagens categoria ‘Não’ com 21% e 9,2% respectivamente.

É possível inferir que os moderados são os menos propensos a ter sofrido crimes políticos online (tanto em termos de percentagem como de números absolutos), enquanto os progressistas apresentam os menores números absolutos e relativos. Os conservadores estão no meio, com uma parcela notável enfrentando crimes online.



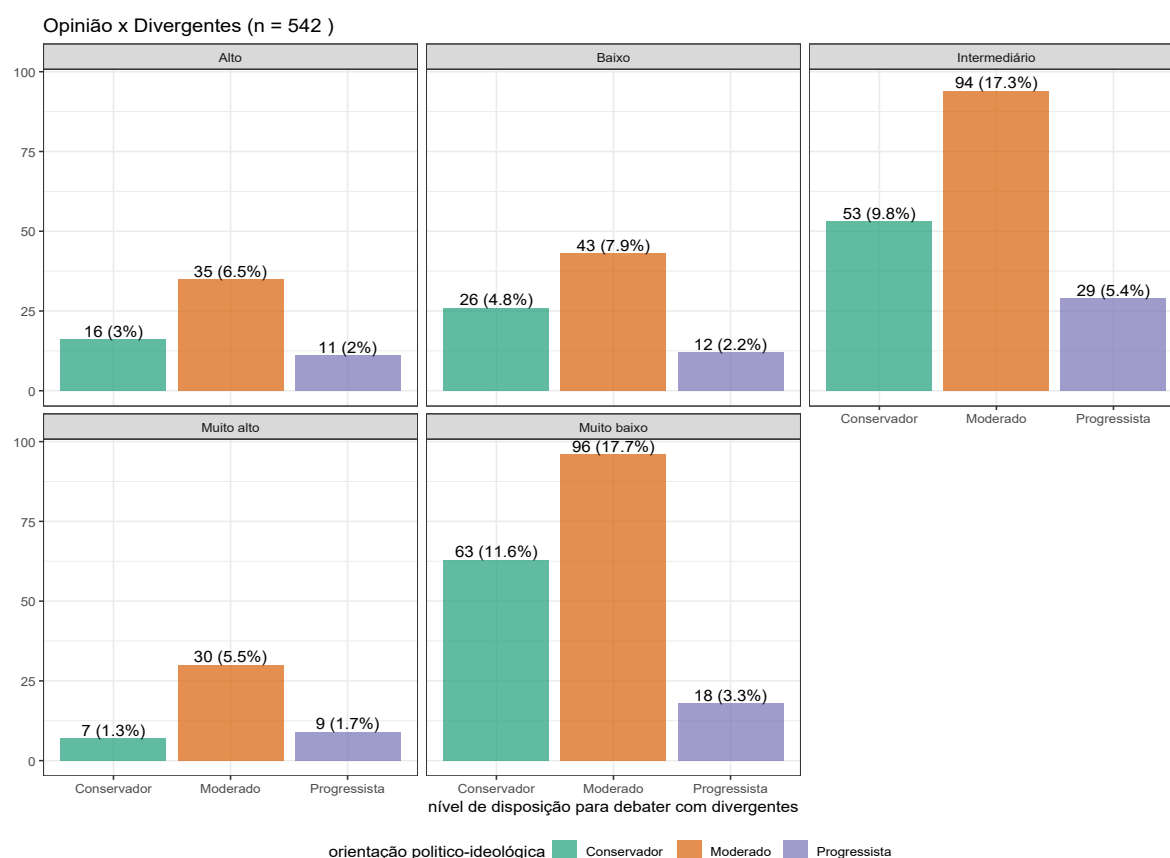
Fonte: Elaboração própria, 2024

3.5 Opinião x Divergentes

A amostra é composta por 542 indivíduos, neste quesito foi perguntado aos indivíduos sobre o “nível de disposição para debater com divergentes”, tendo as suas respostas classificadas em ‘Muito alto’, ‘Alto’, ‘Muito baixo’, ‘Baixo’ e ‘Intermediário’. As orientações políticas foram classificadas em ‘Conservador’, ‘Moderado’ e ‘Progressista’.

- A orientação moderada foi predominante em todas as categorias, tendo a sua maior porcentagem na categoria “Muito baixo” (17,7%).
- As orientações conservadora e progressista não foram predominantes, com as suas maiores porcentagens categoria ‘Muito baixo’ e ‘Intermediário’ com 11,6% e 5,4% respectivamente.

Pode-se inferir que os moderados constituem a maior parte dos indivíduos com diversas disposições para debater com pessoas de opiniões divergentes. Os conservadores têm uma participação relevante nas cinco categorias, com destaque para a disposição muito baixa. Os progressistas têm uma participação menor em quase todas as categorias, com exceção de ‘Muito alto’.



Fonte: Elaboração própria, 2024

4 Resultados

A estatística do qui-quadrado é calculada como a soma dos quadrados das diferenças entre as frequências observadas e esperadas, divididas pelas frequências esperadas. Essa estatística segue aproximadamente uma distribuição qui-quadrado com um determinado número de graus de liberdade, sob a hipótese nula.

O valor esperado (E_{ij}) de cada célula é calculado em uma tabela de contingência, utiliza-se a seguinte fórmula :

$$E_{ij} = \frac{T_i \times T_j}{T_g}$$

Onde,

T_i é o total de observações na linha correspondente.

T_j é o total de observações na coluna correspondente.

T_g é o total geral de observações.

ou seja,

$$E_{ij} = \frac{(\text{total da } i\text{-ésima linha}) \times (\text{total da } j\text{-ésima coluna})}{\text{total geral}}$$

Frequências Observadas (O_i): Estas são as contagens ou frequências reais que foram coletadas e registradas na tabela de contingência. Elas representam os dados brutos, ou seja, o que foi observado na amostra (Agresti, 2012).

Frequências Esperadas (E_i): Estas são as frequências estimadas com base na hipótese nula, que assume que não há associação entre as variáveis. O cálculo do valor esperado visa determinar como seriam as frequências se as variáveis fossem independentes (Fisher Yates, 1963).

A estatística do teste do qui-quadrado (χ^2) é então calculada somando os quadrados das diferenças entre as frequências observadas e esperadas, divididos pelas frequências esperadas para cada célula (Agresti, 2012):

Para uma tabela 2x2 temos;

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Já para uma tabela ixj ($i = \text{Linhas}$, $j = \text{Colunas}$), temos;

$$\chi^2 = \sum_{i=1}^2 \sum_{j=1}^2 \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Este valor é comparado a um valor crítico da distribuição qui-quadrado com graus de liberdade dados por:

Graus de Liberdade

Graus de liberdade são o número de determinações independentes menos o número de parâmetros estatísticos a serem avaliados na população. É um estimador do número de categorias independentes em um teste ou experiência estatística.

$$GL = (i - 1)(j - 1)$$

Onde:

GL é o Grau de liberdade

i é o número de linhas e

j é o número de colunas da tabela de contingência (Agresti, 2012).

Se o valor de χ^2 for maior que o valor crítico, rejeitamos a hipótese nula de independência, indicando que há uma associação significativa entre as variáveis analisadas. Caso contrário, não há evidências estatísticas suficientes para rejeitar a hipótese nula de independência, entre as variáveis comparadas.

Hipóteses a serem testadas:

- H_0 : Não existe associação entre as variáveis (são independentes)
- H_1 : Existe associação (são dependentes)

4.1 Resultado do teste χ^2

Na tabela 1, os resultados dos testes estatísticos para associação entre variáveis são interpretados com base nos p-valores e no nível de significância de 5% (0,05).

Tabela 1: Resultado do teste χ^2

Pares	p-valor	Resultado
ponxgen	0.02639	Há associação entre as variáveis
ponxopi	0.0000143	Há associação entre as variáveis
ponxreli	0.5359	As variáveis são independentes
opi_ofen	0.5039	As variáveis são independentes
opi_diverg	0.224	As variáveis são independentes

Fonte: Elaboração própria, 2024

- ponxgen: p-valor = 0,02639. Como o p-valor é menor que 0,05, há associação entre o Nível de Participação Online e Gênero.
- ponxopi: p-valor = 0,0000143. Também menor que 0,05, indicando associação entre Nível de Participação Online e Orientação Politico-Ideológica.
- ponxreli: p-valor = 0,5359. Como o p-valor é maior que 0,05, o Nível de Participação Online e a Religião são independentes.
- opi_ofen: p-valor = 0,5039. O p-valor é muito maior que 0,05, indicando independência entre orientação política ideológica e ofensas na internet.
- opi_diverg: p-valor = 0,224. Sendo maior que 0,05, logo orientação política ideológica e o nível de disposição são consideradas independentes.

Esses resultados indicam que **"Nível de Participação Online x Gênero"** e **"Nível de Participação Online x Orientação Politico-Ideológica"** possuem associação significativa, enquanto as demais combinações são independentes.

4.2 Tamanho do efeito

O tamanho do efeito será calculado por intermédio do V de Crámer, calculado a partir da estatística do teste qui-quadrado:

$$V = \sqrt{\frac{\chi^2}{N \times \min(i-1, j-1)}}$$

De acordo com Kotrlik et al., 2011 apud Lima, M, 2023, os parâmetros de análise para a mensuração do tamanho do efeito são:

- $0 \leq \phi < 0,10$: irrisório;
- $0,10 \leq \phi < 0,20$: fraco;
- $0,20 \leq \phi < 0,40$: moderado;
- $0,40 \leq \phi < 0,60$: relativamente forte;
- $0,60 \leq \phi < 0,80$: forte;
- $0,80 \leq \phi < 1$: muito forte;
- $\phi = 1$: perfeito.

Tabela 2: Tamanho do efeito

Pares	ϕ	Resultado
ponxgen	0.1276	Fraco
ponxopi	0.1835	Fraco
ponxreli	0.08225	-
opi_ofen	0.05029	-
opi_diverg	0.09899	-

Fonte: Elaboração própria, 2024

Os valores de 0.1276 e 0.1835 correspondem ao tamanho do efeito medido pelo V de Cramér, que avalia a força da associação entre duas variáveis no teste qui-quadrado.

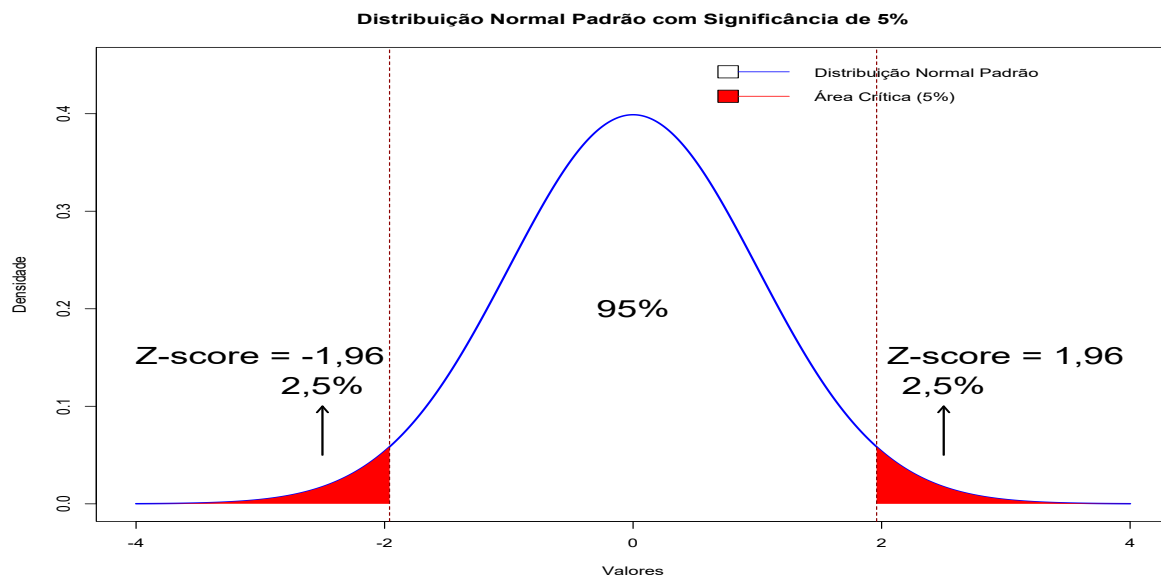
- O valor 0.1276 (para "ponxgen") indica um efeito fraco. Isso significa que há uma associação entre "Nível de Participação Online x Gênero", mas essa associação é de baixa intensidade.
- O valor 0.1835 (para "ponxopi") também representa um efeito fraco entre "Nível de Participação Online x Orientação Politico-Ideológica", com uma associação um pouco mais forte que no caso anterior, mas ainda assim, a relação entre as variáveis é fraca.

5 Análise dos resíduos

Para que o resíduo seja considerado significativo, será realizada uma padronização para seguir a distribuição normal padrão $N(0,1)$. Um resíduo será considerado

significativo se o valor padronizado estiver fora do intervalo $[-1,96; 1,96]$, conforme a tabela da distribuição normal. Esse intervalo corresponde a aproximadamente 95% da área sob a curva da distribuição normal padrão. Portanto, resíduos fora desse intervalo são considerados estatisticamente significativos, indicando que eles são incomuns ou extremos, ao nível de significância de 5%.

A imagem abaixo, exemplifica uma distribuição normal padrão (0,1) com 5% significância.



Fonte: Elaboração própria, 2024

5.1 Análise post-hoc

As análises de post-hoc desempenham um papel fundamental na interpretação de resultados significativos obtidos através de testes qui-quadrado de independência.

Quando um teste identifica uma associação estatística entre variáveis, as análises post-hoc são essenciais para determinar quais grupos ou categorias específicas estão contribuindo para essa associação.

Existem predominantemente duas abordagens para a realização de análises post-hoc após a aplicação do teste qui-quadrado. A primeira, o **Método de Resíduos Padronizados**, calcula os resíduos para cada célula da tabela de contingência. Neste contexto, resíduos que ultrapassam um valor crítico, geralmente maior que 1,96 e menor que -1,96, são considerados indicativos de associações significativas. Em contrapartida, o **Método de Resíduos Ajustados** realiza um ajuste nos resíduos para controlar a taxa de erro Tipo I. Este método aplica um valor crítico ajustado, levando em consideração o número total de células na tabela, com o intuito de mitigar a probabilidade de rejeitar a hipótese nula quando, de fato, ela é verdadeira.

O controle da taxa de erro Tipo I é uma questão premente, particularmente quando múltiplas comparações são realizadas. Os métodos post-hoc, então, são projetados para ajustar os níveis de significância, assegurando que a probabilidade de cometer um erro Tipo I se mantenha em uma faixa aceitável.

Pesquisas sobre a eficácia desses métodos revelam que o método de resíduos ajustados é capaz de manter a taxa de erro sob controle, desde que ajustes adequados sejam implementados. O procedimento de Gardner (MacDonalds e Gardner, 2000) se destaca nesse aspecto, uma vez que demonstrou a capacidade de manter as taxas de erro experimentais próximas aos níveis nominais, configurando-se como uma alternativa recomendada para análises post-hoc.

Portanto, é incontestável que, após a identificação de um resultado significativo em um teste qui-quadrado de independência, as análises post-hoc se tornam imprescindíveis para elucidar quais categorias estão efetivamente associadas. A escolha do método de análise post-hoc deve considerar a necessidade de controle rigoroso da taxa de erro Tipo I, sendo os métodos de resíduos ajustados e o procedimento de Gardner opções eficazes e confiáveis para tal fim. De acordo com o método de Gardner, o cálculo do valor crítico para realizar comparações post hoc em tabelas de contingência (como a tabela 3x5 mencionada no contexto) envolve o ajuste do nível de significância (alfa) para controlar a taxa de erro Tipo I.

O procedimento:

- **Identificação do Número de Categorias:** Deve-se determinar o número total de categorias que foram feitas. Para uma tabela de contingência 3x5, por exemplo, teriam 15 categorias possíveis.
- **Ajuste do Nível de Significância:** Para o modelo de independência, a fórmula é:

$$\alpha_{\text{ajustado}} = \frac{\alpha}{n}$$

Onde,

α = valor de α usado no teste qui-quadrado, comumente é utilizado $\alpha = 0,05$.

n = número de categorias

- **Cálculo do Valor Crítico:** Após determinar o nível de significância ajustado, deve-se usar esse valor para calcular o z-crítico correspondente. O z-crítico pode ser obtido a partir da distribuição qui-quadrado, utilizando o número de graus de liberdade apropriado (que é geralmente $(i - 1)(j - 1)$) para a tabela de contingência).

- **Comparação da Estatística de Teste (Z -crítico):** A estatística de teste calculada para cada comparação é então comparado com o valor crítico ajustado. Se a estatística de teste exceder o valor crítico, a hipótese nula é rejeitada para essa comparação específica. Em outras palavras se a **estatística de teste** (resíduos ajustados) exceder o valor de **z-crítico**, rejeitamos H_0 e podemos afirmar que há associação.

Esse método de ajuste é importante para garantir que a taxa de erro Tipo I permaneça em níveis aceitáveis, especialmente quando múltiplas comparações são realizadas após um teste qui-quadrado significativo.

Para ambos resultados significativos temos o total de categorias de $ixj = 3 \times 15$, Foram obtido um novo p-valor = 0,0033 (alfa corrigido) e valor crítico z-score crítico $\pm 2,9352$.

Tabela 3- Correção de alfa para análise de Post-Hoc.

Correção do alfa = $0,05/(\text{linhas} \times \text{colunas})$				
Linhas	Colunas	Linhas*colunas	alfa corrigido	Z crítico
3	5	15	0,0033	-2,9352

Fonte: Elaboração própria, 2024

Tabela 4- Resultados dos p-valores corrigidos da gênero análise de Post-Hoc.

Nível	Categoria	Resíduo ajustado	valor de p
Alto	Não binário	2,2388	0,025167
Nenhum	Homem	3,2010	0,001369
Nenhum	Mulher	3,2300	0,001238

Fonte: Elaboração própria, 2024

Tabela 5- Resultados dos p-valores corrigidos da Orientação político-ideológica para análise de Post-Hoc.

Nível	Categoria	Resíduo ajustado	valor de p
Muito alto	Progressista	4,6848	0,00000280
Nenhum	Conservador	2,7071	0,006788
Muito alto	Moderado	2,9386	0,003297
Intermediário	Conservador	2,5130	0,011971
Intermediário	Moderado	3,12022	0,001807
Nenhum	Progressista	2,7621	0,005744

Fonte: Elaboração própria, 2024

5.2 Nível de participação online por gênero

De acordo com a **Tabela 1** o **p-valor** de 0.02639 está abaixo do nível de significância 0,05, o que nos permite rejeitar a hipótese nula de independência. Isso significa que há associação estatisticamente significativa entre as variáveis nível de participação online(NPO) e gênero. Em outras palavras, a distribuição de NPO varia de forma significativa conforme o gênero e essa associação não se deve ao acaso.

Com base nos resultados, podemos concluir que existe uma relação significativa entre gênero e nível de participação online. A participação online de indivíduos varia de forma não aleatória entre diferentes gêneros, sugerindo que gênero pode ser um fator importante para entender os padrões de envolvimento político online.

Com base nos resíduos padronizados ajustados e no limite de significância ± 1.96 , vamos analisar os resultados que indicam associação significativa entre as categorias de NPO e gênero.

Tabela 6- Teste de associação qui-quadrado de Pearson e análise de resíduos para Nível de participação e Gênero

Nível de participação Online		Homem	Mulher	Não binário
Muito Alto	<i>O_i</i>	32	41	1
	<i>E_i</i>	34,7828	38,1086	1,1086
	<i>R_i</i>	-0,6983	0,7246	-0,1120
	<i>P-valor</i>			
Alto	<i>O_i</i>	23	38	3
	<i>E_i</i>	30,0824	32,9588	0,9588
	<i>R_i</i>	-1,8907	1,3440	2,2388
	<i>P-valor</i>			0,0252
Intermediário	<i>O_i</i>	42	55	1
	<i>E_i</i>	46,0637	50,4682	1,4682
	<i>R_i</i>	-0,9102	1,01371	-0,43084
	<i>P-valor</i>			
Baixo	<i>O_i</i>	48	62	0
	<i>E_i</i>	51,7041	56,6479	1,6479
	<i>R_i</i>	-0,7941	1,1459	-1,4516
	<i>P-valor</i>			
Nenhum	<i>O_i</i>	106	79	3
	<i>E_i</i>	88,3670	96,8165	2,8165
	<i>R_i</i>	3,2010 *	-3,2300 *	0,1369
	<i>P-valor</i>	0,0014	0,0012	

Fonte: Elaboração própria, 2024

Nota: *O_i*: Valor observado,

E_i: Valor esperado,

R_i: Resíduos ajustados,

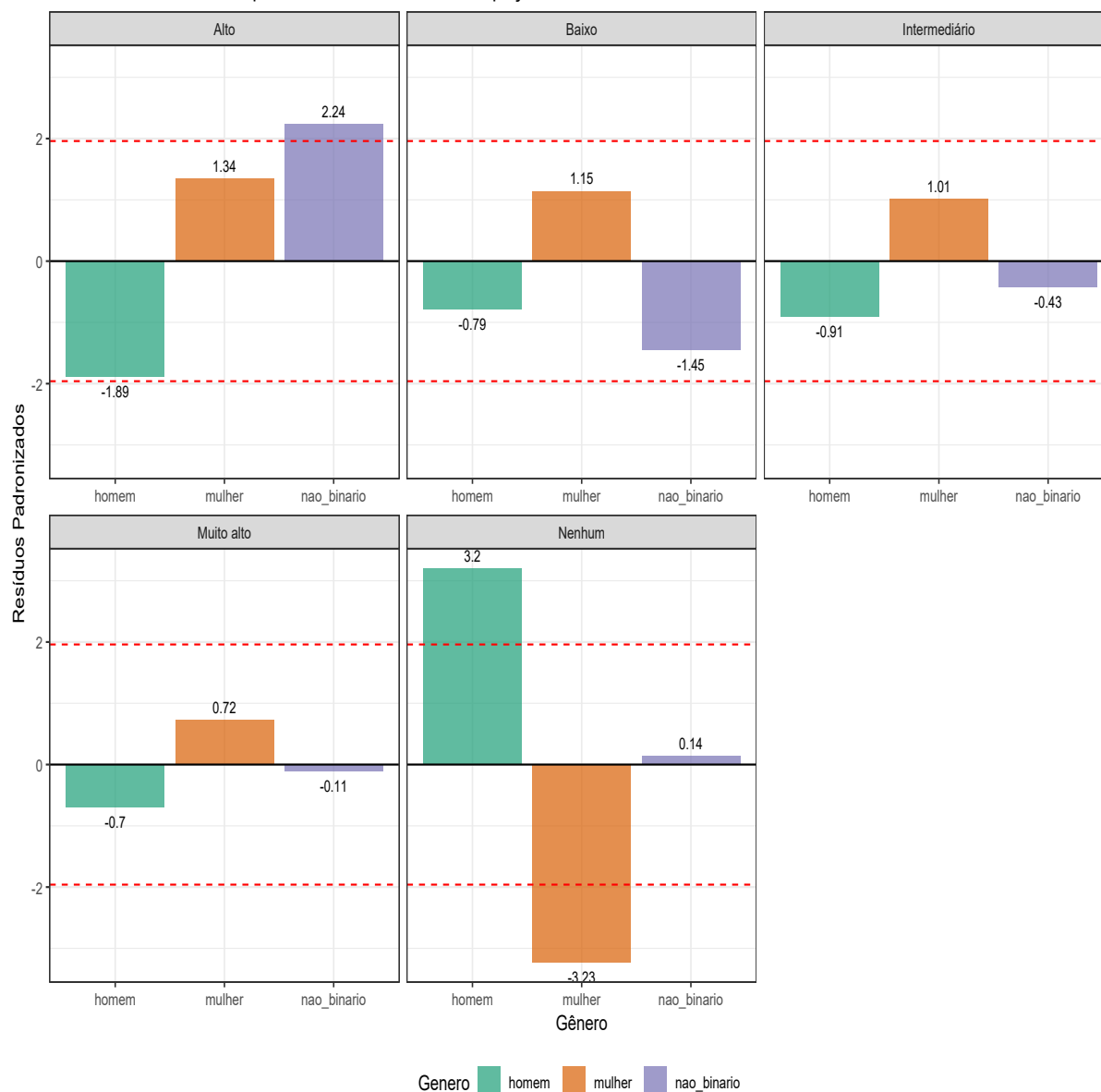
P-Valor: P-Valor Post-Hoc,

*: Resíduos ajustados significantes.

5.3 Gráfico dos resíduos - Nível de participação online por gênero

O gráfico mostra os resíduos padronizados por gênero e nível de participação online, comparando as categorias de gênero (homem, mulher, e não binário) com diferentes níveis de participação (nenhum, baixo, intermediário, alto e muito alto).

Resíduos Padronizados por Gênero e Nível de Participação Online



Fonte: Elaboração própria, 2024

- Valores positivos indicam que há mais observações do que o esperado em determinado grupo. Por exemplo, se um gênero tiver um valor residual alto para um nível de participação específico, isso significa que o número de pessoas desse grupo com esse nível de participação foi maior do que o esperado.
- Valores negativos indicam menos observações do que o esperado. Ou seja, certos grupos tiveram menos pessoas com determinado nível de participação.

5.3.1 Variáveis com Significância

Participação "Alto" para Não Binário:

- O resíduo ajustado é **2,2388**, que está acima do limite de 1.96, indicando que a contagem observada de pessoas não binárias com participação online "Alto"(3) é significativamente maior do que a contagem esperada (0,9588). Isso sugere uma associação significativa e positiva, ou seja, mais indivíduos não binários do que o esperado tendem a ter participação online alta.

Post-hoc: Obteve-se um *p-valor* de **0,0252**, que é maior do que o nível de significância ajustado de **0,0033**. Portanto, não se rejeita a hipótese nula de independência, indicando que não há evidência suficiente para afirmar uma associação significativa entre as variáveis. Em outras palavras, com base nesse teste, não há uma relação estatisticamente significativa entre as variáveis analisadas.

Participação "Nenhum" para Homem e Mulher:

- Para homens, o resíduo ajustado é **3,2010**, acima de 1,96, indicando que o número de homens com "Nenhuma" participação (106) é significativamente maior do que o esperado (88,3670).
- Já para mulheres, o resíduo ajustado é **-3,2300**, abaixo de -1,96, o que indica que o número de mulheres com "Nenhuma" participação (79) é significativamente menor do que o esperado (96,8165).
- Esses resíduos indicam uma associação forte com os homens tendendo a ter participação nula maior do que o esperado, enquanto as mulheres tendem a ter menos participação nula do que o esperado.

Post-hoc: Já durante a avaliação foram retornados os seguintes *p-valores* **0,0014** e **0,0012**, para homem e mulher, respectivamente. Para ambos se rejeita a hipótese nula de independência, indicando que há evidências suficientes para afirmar uma associação significativa entre as variáveis.

5.3.2 Variáveis sem Significância

As demais categorias possuem resíduos padronizados ajustados que não ultrapassam o limite de significância de $\pm 1,96$, o que sugere que não há uma associação significativa entre o gênero e os níveis de participação online "Muito Alto", "Intermediário" e "Baixo" nesses casos.

5.3.3 Resumo das Associações

- Indivíduos não binários tendem a ter uma participação alta maior do que o esperado. Todavia não apresenta evidências significativa de associação
- Homens tendem a ter “nenhuma” participação online maior do que o esperado, enquanto mulheres tendem a ter menos “nenhuma” participação do que o esperado. Apresenta evidências significativas de associação.
- Esses resultados sugerem diferenças claras entre os gêneros em termos de envolvimento em atividades online, com implicações potencialmente importantes para entender como diferentes grupos de gênero se envolvem digitalmente. Apresenta evidências significativas de associação.

5.4 Nível de participação online por orientação político-ideológica

De acordo com a Tabela 1, o **p-valor** de **0.0000143** está abaixo do nível de significância de 0,05, o que nos permite rejeitar a hipótese nula de independência. Isso significa que há associação estatisticamente significativa entre as variáveis **nível de participação online (NPO)** e **orientação político-ideológica**. Em outras palavras, a distribuição de NPO varia de forma significativa conforme a orientação política, e essa associação não se deve ao acaso.

Com base nos resultados, podemos concluir que existe uma relação significativa entre orientação política e nível de participação online. A participação online de indivíduos varia de forma não aleatória entre diferentes orientações políticas, sugerindo que a orientação política pode ser um fator importante para entender os padrões de envolvimento político online.

Com base nos resíduos padronizados ajustados e no limite de significância ± 1.96 , vamos analisar os resultados que indicam associação significativa entre as categorias de NPO e orientação política.

Tabela 7- Teste de associação qui-quadrado de Pearson e análise de resíduos para Nível de participação e orientação político-ideológica.

Nível de participação Online		Conservador	Moderado	Progressista
Muito Alto	<i>O_i</i>	21	29	24
	<i>E_i</i>	22,5277	40,6864	10,7860
	<i>R_i</i>	-0,4153	-2,9386 *	4,6848 *
	<i>P-valor</i>		0,0033	0,0000
Alto	<i>O_i</i>	22	33	10
	<i>E_i</i>	19,7878	35,7380	9,4742
	<i>R_i</i>	0,6356	-0,7276	0,1970
	<i>P-valor</i>			
Intermediário	<i>O_i</i>	20	69	11
	<i>E_i</i>	30,4428	54,9816	14,5757
	<i>R_i</i>	-2,5130	-3,12022 *	-1,12212
	<i>P-valor</i>	0,0120	0,00180	
Baixo	<i>O_i</i>	30	65	17
	<i>E_i</i>	34,0959	61,5793	16,3247
	<i>R_i</i>	-0,9443	0,7294	0,2030
	<i>P-valor</i>			
Nenhum	<i>O_i</i>	72	102	17
	<i>E_i</i>	58,1458	105,0148	27,8395
	<i>R_i</i>	2,7071	-0,5449	-2,7621
	<i>P-valor</i>	0,0068		0,0057

Fonte: Elaboração própria, 2024

Nota: *O_i*: Valor observado,

E_i: Valor esperado,

R_i: Resíduos ajustados,

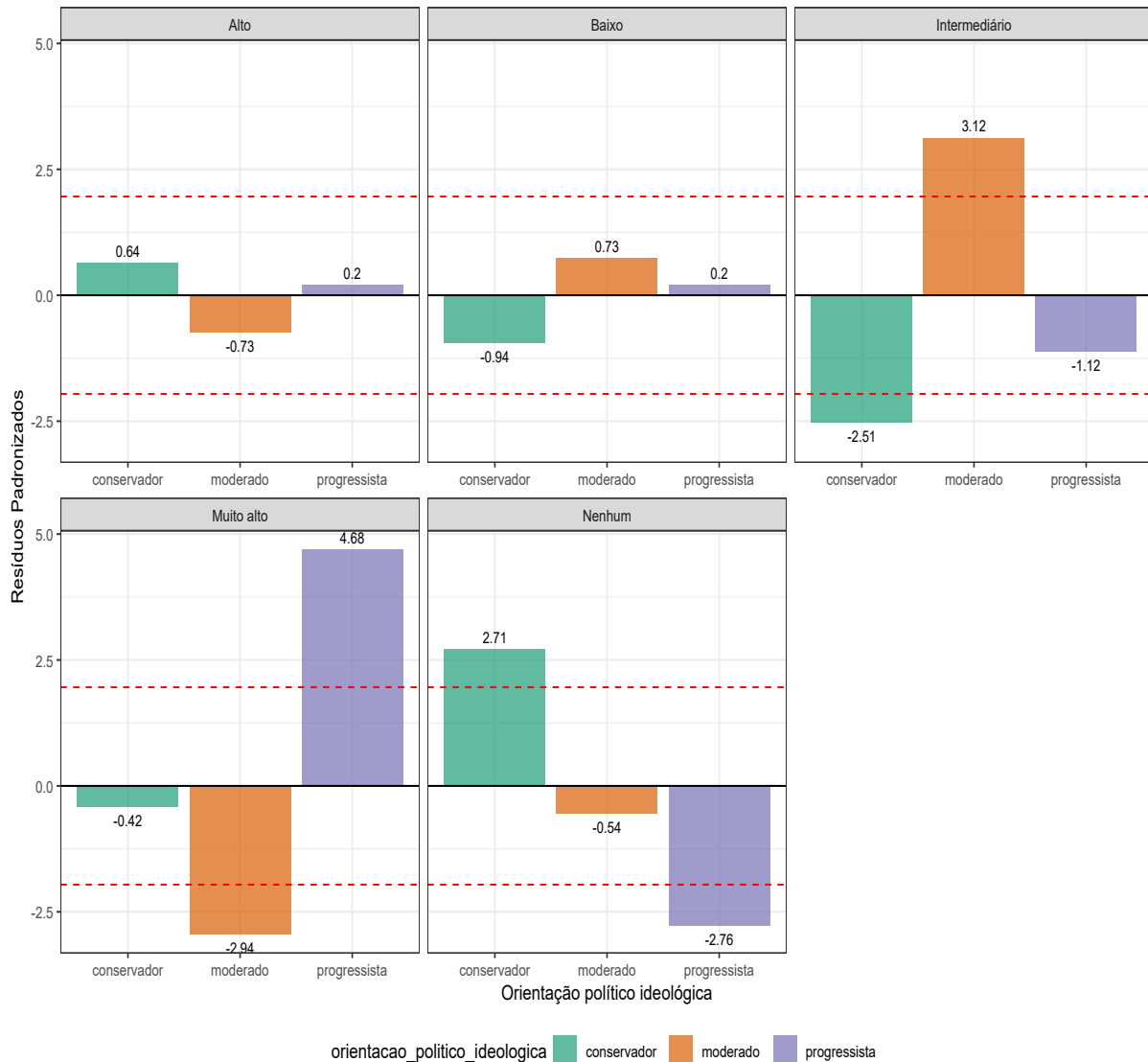
P-Valor: P-Valor Post-Hoc,

*: Resíduos ajustados significantes.

5.5 Gráfico dos resíduos - Nível de participação online por orientação político-ideológica

O gráfico exibe os resíduos padronizados por orientação político-ideológica (conservador, moderado, progressista) e nível de participação online (nenhum, baixo, intermediário, alto e muito alto).

Resíduos Padronizados por Orientação político ideológica e Nível de Participação Online



Fonte: Elaboração própria, 2024

- Valores positivos indicam que há mais observações do que o esperado em determinado grupo. Por exemplo, se uma orientação política tiver um valor residual alto para um nível de participação específico, isso significa que o número de pessoas desse grupo com esse nível de participação foi maior do que o esperado.
- Valores negativos indicam menos observações do que o esperado. Ou seja, certos grupos tiveram menos pessoas com determinado nível de participação.

5.5.1 Variáveis com Significância

Participação "Intermediário" para Conservadores e Moderados:

- O resíduo ajustado é **-2.513** para conservadores, abaixo de -1.96, indicando que

a contagem observada de conservadores com participação "Intermediário"(20) é significativamente menor do que a esperada (30.44).

Post-hoc: Foi retornado o *p-valor* de 0,011971, acima do valor de significância de 0,0033, portanto não se rejeita a hipótese nula de independência, indicando que não há evidência suficiente para afirmar uma associação significativa entre as variáveis.

- Para moderados, o resíduo ajustado é **3.120**, acima de 1.96, indicando que a contagem observada de moderados com participação "Intermediário"(69) é significativamente maior do que a esperada (54.98). Isso sugere uma **associação significativa positiva** para moderados.

Post-hoc: Foi retornado o *p-valor* de 0,001807, abaixo do valor de significância, *p-valor* = 0,0033, portanto se rejeita a hipótese nula de independência, indicando que há evidência suficiente para afirmar uma associação significativa entre as variáveis.

Participação "Muito Alto"para Progressistas e Moderados:

- O resíduo ajustado para progressistas é **4.685**, indicando que a contagem observada de progressistas com participação "Muito Alto"(24) é significativamente maior do que a esperada (10.79).

Post-hoc:Foi retornado o *p-valor* de 0,0000028, abaixo do valor de significância, *p-valor* = 0,0033, portanto se rejeita a hipótese nula de independência, indicando que há evidência suficiente para afirmar uma associação significativa entre as variáveis.

- Para moderados, o resíduo ajustado é **-2.939**, indicando que a contagem observada de moderados com participação "Muito Alto"(29) é significativamente menor do que a esperada (40.69).

Post-hoc:Foi retornado o *p-valor* de 0,003297 está abaixo do valor de significância do *p-valor* 0,0033, portanto se rejeita a hipótese nula de independência, indicando que há evidência suficiente para afirmar uma associação significativa entre as variáveis.

Participação "Nenhum"para Conservadores e Progressistas:

- O resíduo ajustado para conservadores é **2.707**, acima de 1.96, indicando que a contagem observada de conservadores com "Nenhuma"participação (72) é significativamente maior do que a esperada (58.15).
- Para progressistas, o resíduo ajustado é **-2.762**, abaixo de -1.96, indicando que a contagem observada de progressistas com "Nenhuma"participação (17) é significativamente menor do que a esperada (27.84).

Post-hoc: Já durante a avaliação foram retornados os seguintes *p-valor* **0,006788** e **0,005744**, para Conservador e progressista, respectivamente. Para ambos não se rejeita a hipótese nula de independência, indicando que não há evidências suficientes para afirmar uma associação significativa entre as variáveis.

5.5.2 Variáveis sem Significância

As demais categorias de participação online, como "Alto" e "Baixo", possuem resíduos padronizados ajustados que não ultrapassam o limite de significância de ± 1.96 , sugerindo que não há uma associação significativa entre a orientação política e esses níveis de participação online.

5.5.3 Resumo das Associações

- Indivíduos **moderados** tendem a participar mais no nível "Intermediário" e menos no nível "Muito Alto" do que o esperado. Entretanto há associação estatisticamente significativa apenas para Moderado em nível de uso "Muito alto", ou seja há evidências para afirmar que indivíduos moderados participam menos que o valor esperado no nível "Muito alto".
- **Progressistas** têm uma participação maior do que o esperado no nível "Muito Alto" e menor no nível "Nenhum". Todavia apenas participação no "Muito alto" apresenta associação estatisticamente significativa.
- **Conservadores** tendem a ter menos participação "Intermediária" e mais participação "Nenhuma" do que o esperado. Para ambas as associações, não há evidências estatisticamente significativa para afirmar associação, entre Conservador/participação intermediária e Conservador/participação nenhuma

Esses resultados sugerem diferenças claras entre as orientações políticas em termos de envolvimento em atividades online, com implicações importantes para entender como diferentes grupos políticos se engajam digitalmente.

6 Considerações finais

Para analisar a participação online dos jovens na internet, levadas em consideração gênero, orientação-político-ideológica, e a religião, para verificar como essas variáveis os influenciam nesse comportamento online. Para tanto foram utilizadas técnicas estatísticas, como teste qui-quadrado, análise de resíduos e Post-Hoc, conseguindo assim resultados importantes sobre o envolvimento dos jovens em atividades políticas na internet.

O cerne desse relatório é a verificação de associações entre as variáveis do estudo, utilizando-se do teste qui-quadrado de independência, considerando um nível de significância estatística de 5%. Assim foi possível inferir se havia independência ou não entre as variáveis, gerando assim uma base estatística para a compreensão dos resultados.

Gênero e Participação Online

De tal forma que o p-valor foi menor que o nível de significância, foi encontrado associação entre gênero e o nível de participação online, ou seja, essas variáveis são dependentes. As mulheres em comparação aos homens e não binários, tem uma maior participação online nas categorias onde tem a indicação de participação ativa.

Com base nos resíduos padronizados e análise Post Hoc, ficou evidente que os homens tendem a apresentar uma maior participação online nula, enquanto as mulheres em contraste, tem uma participação menor nessa categoria, sugerindo que o gênero feminino de maneira geral, estão mais propensas a participar de atividades online relacionadas a debates políticos, enquanto os homens em uma proporção maior do que o esperado, prefere se abster dos debates.

Analisando os indivíduos não-binários, observa-se que mesmo em menor representatividade, eles alcançaram uma maior participação no nível “Alto” do que o esperado, apontando assim para um possível maior engajamento, inferindo que a internet possa ser um ambiente seguro para a expressão política desse grupo.

Esses resultados reforçam a ideia de que fatores de gênero e orientação política afetam o engajamento online dos jovens.

Orientação Político-Ideológica e Participação Online

Com relação ao nível de participação online e à orientação política revelou que indivíduos com orientação moderada têm uma maior participação em todas as categorias analisadas, com destaque na categoria “Intermediário”, considerando as categorias que incluem algum nível de participação online.

Já os progressistas, alcançaram o maior percentual de indivíduos na categoria de participação “Muito Alta”, enquanto os conservadores na categoria “Nenhum”.

Com base na análise Post-Hoc fica reforçado que os progressistas são mais propensos a um envolvimento com maior intensidade, dado que na categoria “Muito Alto”, foi retornando um quantitativo maior do que o esperado, inferindo assim que essa orientação política está mais inclinada a participar ativamente de debates online.

Assim, esses resultados podem refletir aspectos culturais e sociais sobre como cada grupo se engaja politicamente online, indicando que a orientação política pode influenciar o nível participação em atividades online.

Participação Online e Religião

Analisando a variável religião, fica observado que indivíduos católicos apresentaram uma maior participação em todas as categorias da análise, sendo seguidos pelos que se identificam com outras religiões e os evangélicos. Indicando uma maior disposição dos católicos no engajamento de discussões políticas online.

Opinião X Ofensas

Foi analisada a relação entre a orientação política-ideológica dos indivíduos e o fato de já terem sofrido ofensas ao se expressar politicamente na internet. Observou-se que os indivíduos com orientação moderada teve predominância entre aqueles que assinalaram “Não” como sendo a sua resposta, indicando que são menos propensos a enfrentar ofensas na expressão de suas opiniões. Já os conservadores e progressistas, alcançaram uma menor predominância, com uma parte considerável de indivíduos também não enfrentando ofensas.

Os resultados sugerem que os moderados tem uma participação online menos conflituosa, em contraponto aos conservadores e progressistas que estão mais expostos a situações adversas.

Opinião x Divergentes

Ao questionar sobre o nível de disposição dos indivíduos em debater com pessoas que tem opiniões divergentes, os resultados indicara que a orientação moderada apresentou maior disposição para o debate, devido o seu predomínio em todas as categorias, infere-se que eles em geral são mais aberto ao diálogo, enquanto os progressistas e conservadores demonstraram uma menor disposição para debater, com destaque na categoria “Baixo”. Essa descritiva sugere que os indivíduos moderados podem ter uma visão mais conciliadora ou menos polarizada, facilitando a disposição para o diálogo, enquanto os outros grupos, por serem mais ideologicamente definidos, talvez experimentem uma maior resistência ao interagir com indivíduos com opiniões contrárias.

Associação e tamanho do efeito

Para mensurar o tamanho do efeito, foi utilizado o V de Cramér, o qual mostrou que as associações encontradas possuem relação fraca entre elas, definido assim que as associações são de baixa intensidade, inferindo assim que os outros fatores também podem ter interferência no comportamento dos jovens em relação a participação online, bem como educação, contexto econômico, experiências pessoais e outros.

Análise dos resíduos padronizados

Essa análise foi fundamental para identificar relações entre os grupos, revelando que as associações não eram uniformes e que alguns apresentaram comportamentos distintos do esperado. Observou-se que tanto homens quanto mulheres demonstraram comportamentos inversos em relação ao engajamento em níveis “nenhum”, onde homens apresentam uma maior frequência em nível nulo, podendo levantar a hipótese de que homens tem menos envolvimento online que mulher. Enquanto os moderados exibiram um menor envolvimento online em quase todas as categorias. Além disso, embora os moderados tenham mostrado níveis variados de participação, tanto em contextos de baixo quanto de alto engajamento, a tendência geral aponta para um envolvimento abaixo do esperado, indicando assim que indivíduos moderados tendem a ter menos participação online do que o esperado. Essa diversidade no comportamento dos moderados sugere uma variabilidade significativa dentro desse grupo.

Em contraste, os progressistas apresentaram um alto nível de engajamento, reforçando a ideia de que esses indivíduos veem a internet como um importante espaço para a expressão política e o debate.

Implicações práticas

O estudo em questão oferece uma visão abrangente do comportamento de diferentes grupos no ambiente digital com relação ao seu engajamento político. Observou-se que o engajamento pode ser influenciado por fatores como gênero, identidade, orientação política e religião, o que reforça a ideia de que é necessário criar um ambiente mais inclusivo para promover um debate mais amplo e saudável.

Por fim concluímos que esses resultados são importantes para verificarmos a influência dos fatores supracitados, mas evidenciamos que os resultados aqui expostos correspondem à realidade da amostra estudada, podendo com parcimônia ser ampliada para a população de estudantes do segundo e terceiro anos do Ensino Médio da rede pública do Estado de Sergipe, sugerindo-se considerar outros fatores como contexto econômico e educacionais.

7 Referências

- Agresti, A. (2012). *Categorical Data Analysis*. John Wiley Sons.
- Fisher, R. A., Yates, F. (1963). *Statistical Tables for Biological, Agricultural and Medical Research*. Longman.
- Pearson, K. (1900). X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 50(302), 157–175. <https://doi.org/10.1080/14786440009463897>
- Dunn, O. J. (1961). Multiple Comparisons among Means. *Journal of the American Statistical Association*, 56(293), 52–64. <https://doi.org/10.1080/01621459.1961.10482090>
- Kotrlík, J. W., Williams, H. A., Jabor, M. K. (2011). Reporting and interpreting effect size in quantitative agricultural education research. *Journal of Agricultural Education*, 52(1), 132–142. <https://doi.org/10.5032/jae.2011.01132>
- FCA, Luís. Tamanho do efeito. Bookdown, disponível em: <https://bookdown.org/luisfca/docs/qui-quadrado.html#tamanho-do-efeito>. Acesso em: 19 de setembro de 2024.
- Lima, M. (2023, 7 de agosto). Medidas de associação: Coeficiente phi e V de Cramér. Blog Psicometria Online. <https://www.blog.psicometriaonline.com.br/medidas-de-associacao-coeficiente-phi-e-v-de-cramer/> Acesso em: 19 de setembro de 2024.
- UFPR. Teste do qui-quadrado. Universidade Federal do Paraná, disponível em: http://www.leg.ufpr.br/lib/exe/fetch.php/disciplinas:ce001:teste_do_qui-quadrado.pdf. Acesso em: 19 de setembro de 2024.
- PSICOMETRIA ONLINE. Teste do qui-quadrado de independência no R. YouTube, disponível em: https://www.youtube.com/watch?v=df95rvyb_bk. Acesso em: 19 de setembro de 2024.
- Lima, M. (2022, 14 de novembro). Como fazer o teste qui-quadrado de independência no R? Blog Psicometria Online. <https://www.blog.psicometriaonline.com.br/teste-qui-quadrado-de-independencia-no-r/> Acesso em: 19 de setembro de 2024.

ENRICO, A. Colosimo. Aula 5 e 6: Qui-quadrado e Mantel. UFMG, disponível em: <https://www.est.ufmg.br/enricoc/pdf/avancados_medicina/aula5-6_qui_mantel.pdf>. Acesso em: 19 de setembro de 2024.