

Produtividade acadêmica e gênero na Unesp

Carlos Alberto O. Matos^{1†}, Gabriel A. Pio²

¹Instituto de Ciências e Engenharia, Unesp, Itapeva.

²Faculdade de Engenharia e Ciências, Unesp, Guaratinguetá.

Resumo: A gestão 2021-2024 da Unesp, especialmente por meio da Vice-Reitoria, Pró-Reitorias de Pós-Graduação e de Pesquisa, tem instituído vários prêmios para docentes, pesquisadoras, técnicas administrativas, graduandas, pós-graduandas e pós-doutorandas com o objetivo de reconhecer e valorizar a atuação das mulheres na ciência. A premiação, regra geral, enfatiza como principais critérios a relevância da pesquisa de alto impacto e a produção científica baseada em publicações em periódicos de relevância. Este estudo objetivou avaliar se há um efeito relacionado ao gênero na produtividade científica utilizando o índice H que é um dos principais indicadores de produtividade científica. Dados de 16.753 mulheres e 14.022 homens foram selecionados e processados com mineração de texto. O teste de Wilcoxon rejeitou a H_0 , mas esse resultado pode ter sido consequência do tamanho amostral. Os resultados da Análise Exploratória de Dados e de simulação com técnicas de reamostragem utilizando esse teste e sua versão agnóstica recomendam cautela em relação à decisão de rejeição. Finalmente, dez coeficientes que quantificam o tamanho de efeito, ferramentas amplamente empregadas em Psicologia, Medicina e Ciências Sociais, concluíram que esse efeito, mesmo quando existe, pode ser caracterizado como negligível.

Palavras-chave: índice- h ; mineração de texto; equidade; tamanho de efeito.

Academic productivity and gender at Unesp

Abstract: Unesp's 2021-2024 governance, especially through the Vice-Rector's Office, Dean's Office of Postgraduate Studies and Research, has instituted several awards for teachers, researchers, administrative technicians, undergraduates, postgraduates and post-doctoral students with the objective of recognizing and valuing the role of women in science. The award, as a general rule, emphasizes as its main criteria the relevance of high-impact research and scientific production based on publications in relevant journals. This study aimed to evaluate whether there is a gender-related effect on scientific productivity using the H index, which is one of the main indicators of scientific productivity. Data from 16,753 women and 14,022 men were selected and processed with text mining. The Wilcoxon test rejected H_0 but this result may have been a consequence of the sample size. The results of Exploratory Data Analysis and simulation with resampling techniques using this test and its agnostic version recommend caution in relation to the rejection decision. Finally, ten coefficients that quantify the effect size, tools widely used in Psychology, Medicine and Social Sciences, concluded that this effect, when it exists, can be characterized as negligible.

Keywords: H index; text mining; equity; effect size.

Introdução e Objetivos

A gestão 2021-2024 da Unesp, especialmente por meio da Vice-Reitoria, Pró-Reitoria de Pós-Graduação e Pró-Reitoria de Pesquisa, tem realizado diversas atividades com o objetivo de reconhecer e valorizar a atuação das mulheres na ciência. Entre elas, destacam-se o Edital Conjunto PROPG/PROPE 02/2022 “Prêmio a Mulheres Docentes-Pesquisadoras da Unesp” e o Edital Conjunto PROPG/ProPE 02/2023 “Mulheres que fazem a Unesp” para docentes, pesquisadoras e técnico-administrativas, que consideram a relevância da pesquisa e da produção científica como primeiro critério de premiação.

[†]Autor correspondente: carlos.matos@unesp.br).

Também foram criados o Edital Conjunto PROPG/ProPE 01/2022 “Prêmio às Estudantes e Jovens Pesquisadoras da Unesp” e o Edital Conjunto PROPG/ProPE 01/2023 “Mulheres que fazem a Unesp” para graduandas, pós-graduandas e pós-doutorandas, que priorizam a condução de pesquisa científica de alto impacto como critério de premiação, com base em publicações em periódicos de relevância na área. Além disso, a Portaria 24/01/2023 instituiu o Prêmio “Heleieth Saffioti” para homenagear mulheres que contribuem para o desenvolvimento da Unesp, sejam docentes ou técnicas administrativas, premiando suas relevantes contribuições em ações acadêmicas, administrativas e sociais para a Universidade.

Um dos principais indicadores de produtividade científica é o índice H, que mede o impacto de citação das publicações de um autor. J. E. Hirsch propôs o índice H, “definido como o número de artigos com número de citação maior ou igual a h, como um índice útil para caracterizar a produção científica de um pesquisador” (HIRSCH, 2005).

O índice H é amplamente utilizado na comunidade científica em virtude do fato dele ser um índice representativo da produção intelectual ao longo da vida de um cientista expresso em um único número e de se basear em um cálculo simples, usando os bancos de dados comuns na literatura. Bornmann e Daniel (2007) alertam que esses dois fatores acarretam o perigo de uso indevido do índice, descrevem suas vantagens e desvantagens e resumem os estudos sobre a validade do índice.

O fato deste artigo (HIRSCH, 2005) ter sido citado por 8386 publicações até o presente, atesta a importância e a adequação do referido índice para a avaliação qualitativa da produção científica de um pesquisador.

A mineração de texto é uma prática essencial para lidar com a quantidade de dados textuais disponíveis atualmente, cujo crescimento é exponencial.

Essa técnica envolve a aplicação de algoritmos e ferramentas de análise, para extrair informações a partir de dados não estruturados, que possibilitam identificar padrões, tendências e correlações ocultas nos textos, permitindo a tomada de decisões estratégicas.

A mineração de texto utiliza técnicas de processamento de linguagem natural, aprendizado de máquina e inteligência artificial que possibilitam realizar classificações, agrupamentos, análise de sentimentos, detecção de fraudes e muito mais.

Este estudo utilizou técnicas de mineração de texto e teve como objetivo avaliar se há um efeito relacionado ao gênero na produtividade científica.

Material e Método

A pesquisa do índice H foi realizada na Base Scopus (SCOPUS, 2023). A classificação dos nomes por gênero foi possível graças ao arquivo do Censo de Nomes do IBGE (2010) (NOMES. . . , 2010).

O presente trabalho levantou os seguintes dados de 28242 autores afiliados institucionalmente à UNESP: nome e índice H. Este levantamento reflete a situação encontrada no dia 3 de agosto de 2022. O algoritmo simplificado da mineração de texto utilizada apresentou as seguintes etapas:

1. Seleção do nome conjuntamente com o índice H. O padrão de nome ideal para o processamento foi: **Sobrenome por extenso ou abreviado vírgula espaço Nome por extenso seguido de nomes por extensos ou abreviados.**
2. Remoção dos nomes duplicados (224 linhas referentes a 203 nomes duplicados).
3. Conversão para caracteres maiúsculos.
4. Remoção da acentuação.
5. Remoção de anomalias (números, sinais de pontuação, caracteres UNICODE, etc.)

6. Extração da string após a primeira vírgula.
7. Remoção do primeiro espaço quando existente.

A classificação dos nomes quanto ao gênero foi realizada com base no Censo de Nomes do Brasil que contabiliza 36.764 nomes femininos e 30.197 nomes masculinos. Um algoritmo foi elaborado para evitar o problema de nomes que apareciam classificados tanto como masculinos como femininos. Por exemplo, aparecem 11.694.738 Marias classificadas como nome feminino e 39391 Marias classificadas como nome masculino. Do mesmo modo, João é contabilizado como nome masculino em 2.971.935 vezes e como nome feminino 12.184 vezes. No algoritmo de classificação de gênero, quando um nome aparecia nos dois gêneros, ele foi classificado como pertencendo ao gênero onde ele era mais frequente.

2747 nomes não foram classificados pelo fato de não apresentarem nome por extenso depois da vírgula ou por serem nomes, não redigidos no vernáculo, ausentes do Censo de Nomes do Brasil.

Para o processamento e a análise dos dados foram utilizadas as seguintes bibliotecas do R software versão 3.4.4: `tm`, `plyr`, `tidyverse`, `moments`, `grDevices`, `agnostic`, `effectsize`, `effsize`, `rcompanion`, `DescTools`, `rankFD` e `statspsych` (R Core Team, 2018).

Resultados e Discussão

Após o processamento dos nomes utilizando técnicas de mineração de texto, foram selecionados dados de 14.006 mulheres e 11.275 homens. 2.747 dados não foram passíveis de identificação (NA).

Do mesmo modo que Tukey et al. (1977), consideramos importante a inferência estatística, mas acreditamos que análises gráficas e numéricas podem contribuir para uma melhor compreensão dos dados.

Tabela 1: Índice H - Medidas de posição

	n	Min	Q1	Q2	Média	Q3	Max
Mulheres	14006	0,000	1,000	1,000	2,456	3,000	111,000
Homens	11275	0,000	1,000	1,000	3,175	3,000	105,000
NA	2747	0,000	1,000	1,000	2,780	2,000	112,000

Fonte: Autores.

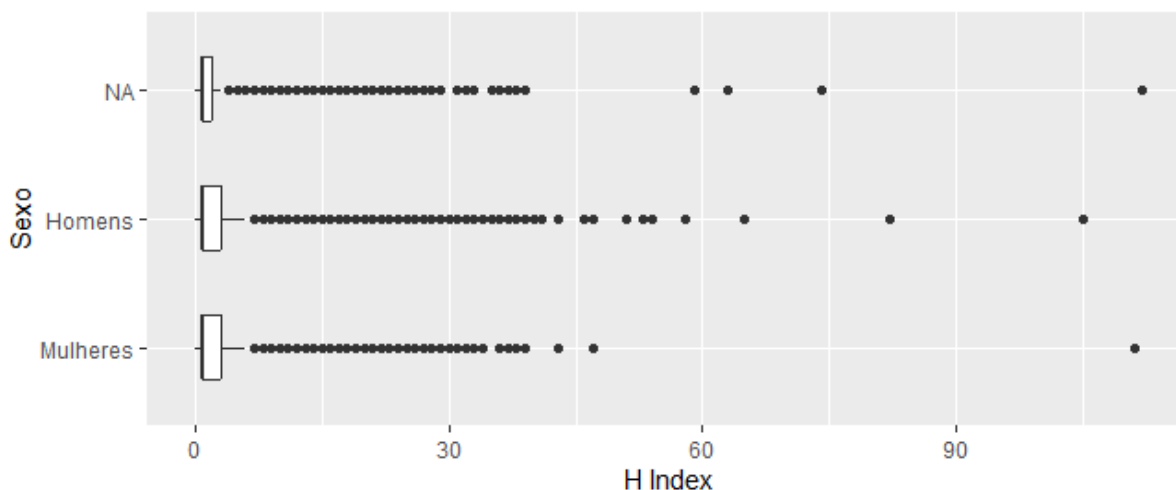
Os valores dos quartis Q1 (1,00), Q2 (1,00) e Q3 (3,00) foram idênticos para ambos os gêneros (Tabela 1). A visualização do boxplot chama a atenção pela quantidade de valores discrepantes (Figura 1).

Valores de índice H superiores a 7 foram categorizados como outliers. 8,32% dos valores de índice H femininos e 13,36% dos valores de índice H masculinos se enquadram nessa categoria.

Os histogramas das distribuições de frequência dos dois gêneros apresentaram alto grau de assimetria: 5,38 para as mulheres e 4,36 para os homens (Figura 2 e Tabela 2). A hipótese nula de ausência de assimetria foi rejeitada ao nível de significância de 5% ($\alpha = 0,05$) pelo teste de D'Agostino. As evidências obtidas na análise exploratória de dados, através dos métodos gráficos e dos resumos numéricos, não são suficientes para confirmar a expectativa acerca da diferença da produtividade acadêmica entre mulheres e homens na Unesp.

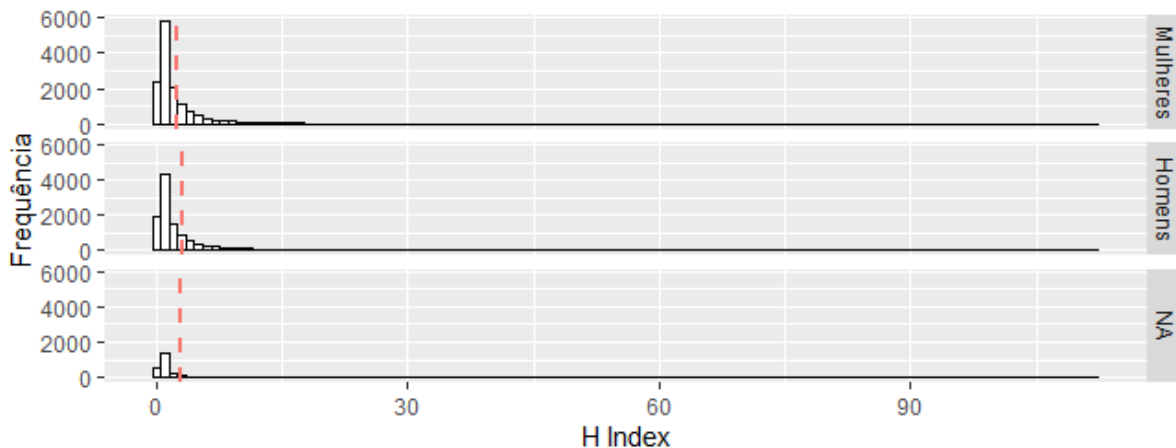
Apesar das semelhanças das medidas de posição quartílicas, o teste de soma de ranks de Wilcoxon rejeitou a hipótese de que as duas amostras pertencem à mesma população ($\alpha = 0,05$), o que evidenciaria a adequação desse tipo de política de equidade de gênero.

Figura 1: Boxplots - Relação entre o índice H e o gênero dos autores(as).



Fonte: Autores.

Figura 2: Histogramas - Relação entre o índice H e o gênero dos autores(as).



Linha tracejada em vermelho = Média.

Fonte: Autores.

Tabela 2: Índice H - Curtose e Assimetria.

	Mulheres	Homens
Curtose	73,9584	38,4254
Assimetria	5,379646	4,361800

Fonte: Autores.

Um alerta em relação a essa conclusão é levantada por Gelman (2009), que afirma ser “um lugar comum entre os estatísticos que um teste qui-quadrado (e, na verdade, qualquer valor P) pode ser visto como uma medida bruta do tamanho da amostra: quando o tamanho da amostra é pequeno, é muito difícil obter uma rejeição (ou seja, um valor P abaixo de 0,05), ao passo que, quando o tamanho da amostra é grande, praticamente qualquer coisa resultará em uma rejeição”.

Na aplicação da Estatística para a Psicologia, Medicina e Ciências Sociais Cohen (1990) e Cohen (1992) reflete sobre o fato de que a hipótese nula (considerada literalmente que é a única maneira de considerá-la num teste formal de hipóteses) ser sempre falsa no mundo real é um fato amplamente compreendido entre os estatísticos. Hipóteses nulas falsas, mesmo em um

grau minúsculo, estão associadas ao caso de amostras grandes o suficiente para produzirem um resultado significativo que levam à sua rejeição.

A Tabela 3 ilustra esse fenômeno para a hipótese de que a proporção de mulheres é igual a de homens em quatro categorias da comunidade universitária. A totalização dos dados de matrícula e formação na Pós-Graduação não está disponível (ANUÁRIO..., 2022).

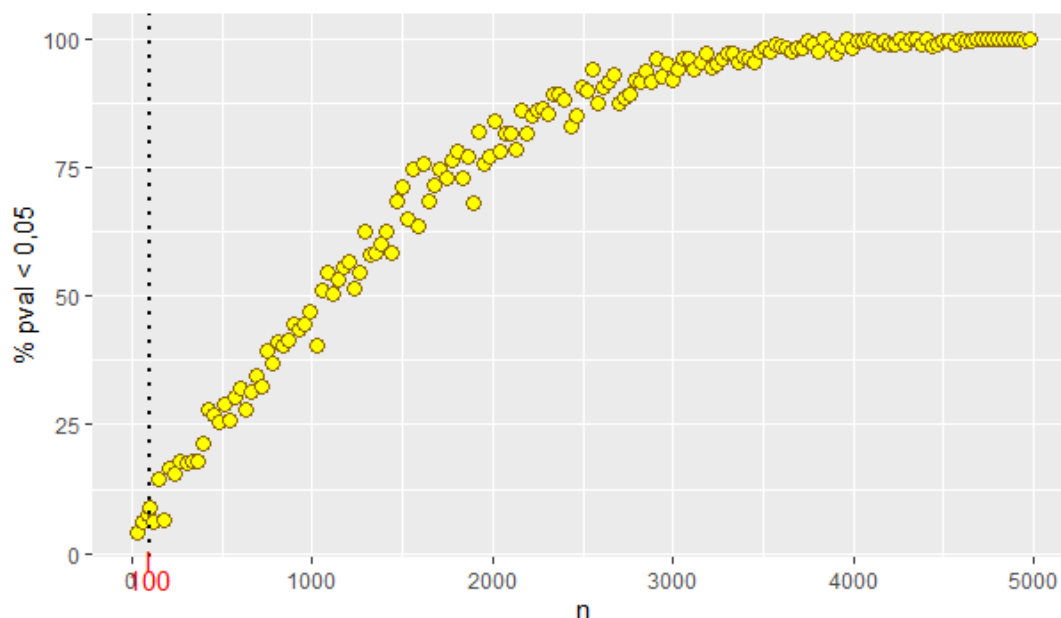
Tabela 3: Resultados do teste binomial exato $H_0 : p = 0,5$,

Sucessos	Tentativas	\hat{p}	Valor P	Sucessos	Tentativas	\hat{p}	Valor P
Alunos matriculados-Graduação 2021				Alunos formados-Graduação 2021			
20733	40792	0,5083	0,0009	2717	4982	0,5454	1,6172E-05
2073	4079	0,5082	0,3014	272	498	0,5462	0,0436
207	408	0,5074	0,8045	27	50	0,5400	0,6718
Docentes ativos 2020				Pessoal técnico administrativo ativo 2020			
1268	3131	0,4050	1,7895E-20	2131	5005	0,4258	7,9857E-20
127	313	0,4058	0,0010	213	500	0,4260	0,0011
13	31	0,4194	0,4731	21	50	0,4200	0,3222

Fonte: Autores.

Considerando que a rejeição da H_0 do teste de soma de ranks de Wilcoxon pode estar associada às amostras de tamanho grande, são apresentados resultados da porcentagem de rejeição de H_0 em função do tamanho da amostra obtidos através de reamostragem (Figura 3). Um detalhamento com amostras de tamanho 30 a 300, apresenta porcentagens de rejeição variando entre 7 e 21% (Tabela 4). Para amostras com $n = 100$, a porcentagem de rejeição da H_0 foi 9,53%.

Figura 3: Resultados da simulação (2000 rodadas) da porcentagem de rejeição de H_0 em função do tamanho da amostra - Teste de Wilcoxon (Fonte: Autor)



Fonte: Autores.

Testes de hipóteses agnósticos são ferramentas de inferência que, ao contrário dos testes usuais, no caso dos dados fornecerem evidências insuficientes para a aceitação ou rejeição uma determinada hipótese, podem permanecer agnósticos, evitando qualquer conclusão. São apresentados resultados da porcentagem de rejeição de H_0 do teste de Wilcoxon agnóstico ($\alpha = \beta = 0,05$) em função do tamanho da amostra ($n = 30 - 4980$) obtidos através de reamostragem com 2000

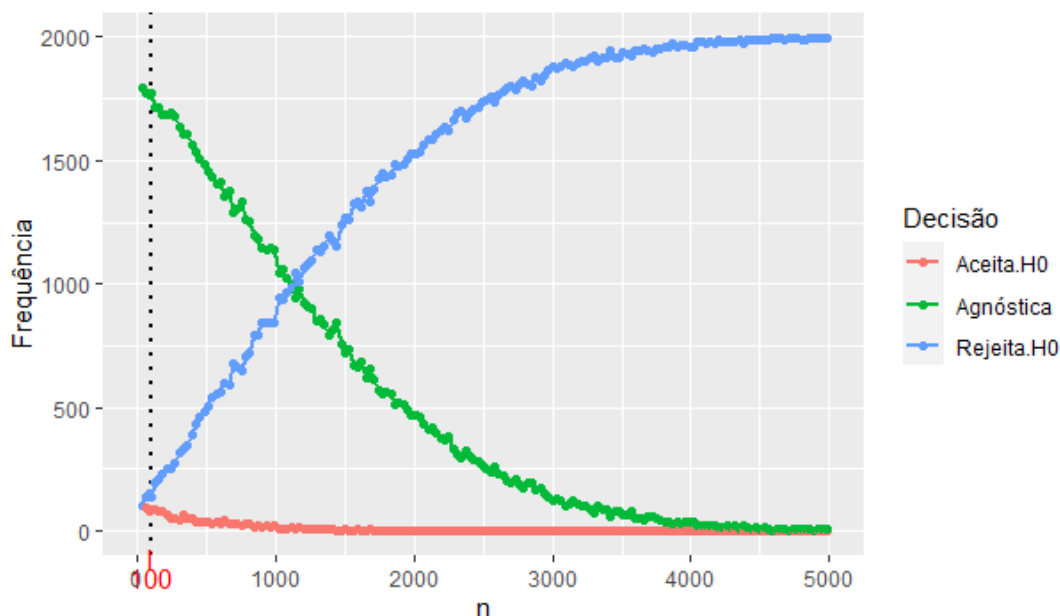
Tabela 4: Resultados da simulação (2000 rodadas) da porcentagem de rejeição de H_0 em função do tamanho da amostra - Teste de Wilcoxon $n= 30 - 300$.

n	pval $\geq 0,05$	pval $< 0,05$	Rejeição de H_0 (%)
30	1877	123	6,55
60	1840	160	8,70
90	1815	185	10,19
100	1826	174	9,53
120	1787	213	11,92
150	1768	232	13,12
180	1730	270	15,61
210	1701	299	17,58
240	1709	291	17,03
270	1684	316	18,76
300	1653	347	20,99

Fonte: Autores.

rodadas (Figura 4). Para amostras com $n = 100$, a porcentagem de rejeição da H_0 diminui para 6,95%. A função `agnostic.wilcox.test` foi adaptada da função `agnostic.t.test` da biblioteca `agnostic` (COSCRATO; IZBICKI; STERN, 2020).

Figura 4: Resultados da simulação (2000 rodadas) da porcentagem de rejeição de H_0 em função do tamanho da amostra - Teste de Wilcoxon Agnóstico.



Fonte: Autores.

Em um estudo quantitativo o valor P informa se existe um efeito, mas não manifesta o tamanho do efeito. A significância estatística (valor P) e a magnitude do efeito são resultados essenciais a serem relatados e interpretados. Um exemplo que é frequentemente citado, foi a recomendação de aspirina para prevenir infarto do miocárdio. Num estudo com mais de 22.000 indivíduos por 5 anos, a associação da aspirina com redução no infarto resultou num valor P altamente significativo do ponto de vista estatístico ($P < 0,00001$), embora o tamanho do efeito tenha sido muito pequeno: uma diferença de risco de apenas 0,77% (o quanto a intervenção altera um risco) com um $r^2 = 0,001$ (proporção de variância compartilhada entre tratados e não tratados). A recomendação do uso de aspirina foi modificada desde então em virtude do risco

de efeitos adversos e da baixa eficácia.(BARTOLUCCI; TENDERA; HOWARD, 2011)

Jacob Cohen (1923-1998) foi um psicólogo e estatístico mais conhecido principalmente por desenvolver vários conceitos e medidas relacionadas ao tamanho de efeito conjuntamente com intervalos de confiança como uma forma de avaliar a significância prática das suas descobertas, além da significância estatística. Seu trabalho contribuiu significativamente para o desenvolvimento da meta-análise.(COHEN, 1994)(COHEN, 2013)

A importância do assunto pode ser dimensionada pelos aproximadamente 27.500.000 resultados da busca "effect size" no rseek.org, pelos 2.320.000 resultados no scholar.google.com, pelos 900 resultados no search.r-project.org e pelos 223 resultados no R-universe.

A literatura é pródiga em relação aos coeficientes de tamanho de efeito. Muitos deles estão relacionados entre si e, do valor de um deles, podem ser calculados os valores de outros. O sinal do efeito indica a direção do efeito e depende de qual amostra é escolhida como a primeira. Se ela for maior em magnitude, e não em tamanho amostral, do que a segunda o sinal do efeito será positivo e, se não for, o sinal será negativo.(BOVA, 2022)

Foram selecionadas algumas estatísticas de tamanho de efeito relacionadas com o teste da soma de ranks de Wilcoxon, que são apresentadas na Tabela 5. Elas, ao contrário dos valores P, não são afetadas pelo tamanho da amostra e dimensionam o grau com que um grupo possui dados com postos mais altos do que um outro grupo.

O d de Cohen d é uma medida de tamanho de efeito padronizado para medir a diferença, em desvios padrões, entre as médias de dois grupos em comparação.(COHEN, 2013)

O r é uma estatística de tamanho de efeito comumente usada para o teste de de Wilcoxon ($r = Z/\sqrt{N}$) que apresenta algumas desvantagens; normalmente não varia totalmente entre -1 e 1 e é afetada pelo tamanho da amostra, fato que se acentua principalmente quando há tamanhos de amostra desiguais entre os grupos.(TOMCZAK; TOMCZAK, 2014)

O coeficiente bisserial de classificação de Glass rg é uma estatística de tamanho de efeito equivalente ao delta de Cliff.(KING; ROSOPA; MINIUM, 2018) O $T - b$ de Kendall é usado eventualmente e varia aproximadamente de -1 a 1.(KENDALL, 1938)

O A de Vargha e Delaney relata a probabilidade de um valor de um grupo ser maior que um valor de um outro grupo. Quando os dois grupos são estocasticamente iguais, seu valor é 0,5. Valores 1 e 0 indicam respectivamente domínio e submissão estocástica completos de um grupo sobre o outro.(VARGHA; DELANEY, 2000)

O Δ de Cliff se relaciona linearmente com o A de Vargha e Delaney. Ele varia de -1 a 1 e a igualdade estocástica dos dois grupos é indicada por 0. Os valores 1 e -1 indicam respectivamente domínio e submissão estocástica completos de um grupo sobre o outro.(VARGHA; DELANEY, 2000) O seu valor absoluto é numericamente igual ao Θ de Freeman.

O Θ de Freeman e o E^2 podem ser empregados no caso de dois grupos, mas são usualmente usados quando há mais de dois grupos, no caso do teste de Kruskal-Wallis.(FREEMAN, 1965)(KING; ROSOPA; MINIUM, 2018)

A razão de chances de probabilidades generalizadas de Agresti calcula as probabilidades de que, no caso de duas observações serem selecionadas aleatoriamente de dois grupos diferentes, o resultado num grupo seja superior ao do outro.(AGRESTI, 1980)

Um coeficiente menos usado é denominado tamanho de efeito em linguagem comum ou probabilidade de superioridade.(MCGRAW; WONG, 1992) (GRISSOM; KIM, 2005) O efeito é convertido numa porcentagem que expressa a probabilidade de uma pessoa retirada aleatoriamente de um grupo possuir uma pontuação maior do que uma pessoa retirada ao acaso do outro grupo. (SANTO; DANIEL, 2017)

Todos os coeficientes acima foram calculados com funções das bibliotecas `effsize`, `rcompanion` e `DescTools`. No cálculo dos intervalos de confiança através de bootstrap, o argumento número de repetições das funções foi padronizado em 100.

A interpretação dos tamanhos dos efeitos varia em função da área do conhecimento. Nas Ciências Sociais, Psicologia e Medicina os critérios mais empregados são os apresentados por

Cohen (2013) que tem 250.126 citações. Outras diretrizes bastante referenciadas são as de Funder e Ozer (2019), Gignac e Szodorai (2016), Lovakov e Agadullina (2021) e Sawilowsky (2009).

Tabela 5: Cálculo de coeficientes de tamanho de efeito

Coeficientes	Efeito Pequeno	Valor	I.C.
d - Cohen	[0,20;0,50)	-0,163 ⁱ	(-0,188;-0,138)
r - Pearson	[0,10;0,30)	-0,043 ⁱ	(-0,058;-0,031)
T - b	[0,10;0,30)	0,038 ⁱ	(0,027;0,049)
Δ - Cliff	[0,11;0,28)	-0,048 ⁱ	(-0,061;-0,034)
rg Coeficiente bisserial de classificação - Glass	[0,11;0,28)	0,026 ⁱ	(0,004;0,046)
A - Vargha e Delaney	(0,34;0,44]	0,476 ⁱ	(0,471;0,482)
	[0,56;0,64)		
Θ - Freeman	[0,11;0,34)	0,048 ⁱ	(0,034;0,061)
E²	[0,01;0,08)	0,009 ⁱ	(0,009;0,013)
Razão de Chance Generalizada - Agresti	[1,44;2,48)	0,885 ⁱ	(0,859;0,919)
Tamanho de efeito em linguagem comum	[0,532;0,597)	0,367 ⁱ	(0,359;0,375)

ⁱEfeito de Tamanho Negligível.

Fonte: Autores.

Outra importante aplicação do tamanho de efeito é no planejamento de uma pesquisa, quando é necessário fazer julgamentos preliminares e informados sobre o tamanho do efeito ou efeitos populacionais desejados, o nível de significância (usualmente 0,05) e o poder de teste (usualmente 0,80) que permitirão a especificação do tamanho da amostra necessário. Quanto menor é o efeito planejado, maior será o tamanho amostral necessário. (BRUNNER; BATHKE; KONIETSCHKE, 2018). Como exemplo, para um efeito médio planejado de 30%, o tamanho de amostra calculado para cada grupo é 59, conforme a recomendação de Noether (1987). Os resultados da reamostragem (2000 rodadas) de amostras com esse tamanho resultaram em valores praticamente idênticos aos apresentados na Tabela 5 que podem ser caracterizados como tamanhos de efeito negligíveis.

Conclusão

Com o objetivo de avaliar se há um efeito relacionado ao gênero na produtividade científica utilizando o índice H, o presente trabalho apresenta o resultado da utilização de diferentes ferramentas estatísticas. A Análise Exploratória de Dados, através dos resumos obtidos de técnicas gráficas e quantitativas, não evidenciou diferenças entre as populações ao contrário do teste não paramétrico de Wilcoxon, que rejeitou a hipótese nula de ausência de diferença na produção intelectual. Para uma amostra de tamanho 100, resultados de reamostragem apresentaram 9,53 % de rejeição da H_0 com o Teste de Wilcoxon. Essa porcentagem cai para 6,95% na versão agnóstica do teste. Dez coeficientes que quantificam o tamanho de efeito, ferramentas amplamente empregadas em Psicologia, Medicina e Ciências Sociais, concluíram que esse efeito, mesmo quando detectado, pode ser caracterizado como negligível. Recomendação importante é a da necessidade de trabalhos de meta-análise em outras áreas do conhecimento para o estabelecimento de critérios de magnitude de tamanho de efeito, que possam servir de referencial para essas áreas.

Agradecimentos

À Equipe de Atendimento do IBGE pela remessa do arquivo Nomes Brasil total fem masc por decada na coluna.zip, que foi essencial para a realização do presente trabalho.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Referências

- AGRESTI, A. Generalized odds ratios for ordinal data. *Biometrics*, JSTOR, p. 59–67, 1980.
- ANUÁRIO Estatístico. Universidade Estadual Paulista Júlio de Mesquita Filho 2022. São Paulo, 2022. v. 1(2001). Disponível em: <https://ape.unesp.br/anuario/flip/>. Acesso em: 7 de janeiro de 2024.
- BARTOLUCCI, A. A.; TENDERA, M.; HOWARD, G. Meta-analysis of multiple primary prevention trials of cardiovascular events using aspirin. *The American journal of cardiology*, Elsevier, v. 107, n. 12, p. 1796–1801, 2011.
- BORNMANN, L.; DANIEL, H.-D. What do we know about the h index? *Journal of the American Society for Information Science and technology*, Wiley Online Library, v. 58, n. 9, p. 1381–1385, 2007.
- BOVA, A. *Nonparametric Methods for Two Independent Samples in R: A Set of Useful Functions (Tutorial)*. 2022. Disponível em: <https://rpubs.com/abova/nptest>.
- BRUNNER, E.; BATHKE, A. C.; KONIETSCHKE, F. *Rank and pseudo-rank procedures for independent observations in factorial designs: Using R and SAS*. [S.l.]: Springer, 2018.
- COHEN, J. Things i have learned (so far). *American Psychologist*, American Psychological Association, Washington, v. 45, n. 12, p. 1304–1312, 1990. Disponível em: <https://doi.org/10.1037/0003-066X.45.12.1304>. Acesso em: 7 de janeiro de 2024.
- COHEN, J. Things i have learned (so far). In: KAZDIN, A. E. (Ed.). *Methodological issues and strategies in clinical research*. Washington: American Psychological Association, 1992. p. 315–333. 765 pp. Disponível em: <https://doi.org/10.1037/10109-028>.
- COHEN, J. The earth is round ($p < .05$). *American psychologist*, American Psychological Association, v. 49, n. 12, p. 997, 1994.
- COHEN, J. *Statistical power analysis for the behavioral sciences*. [S.l.]: Academic press, 2013.
- COSCRATO, V.; IZBICKI, R.; STERN, R. B. Agnostic tests can control the type I and type II errors simultaneously. *Brazilian Journal of Probability and Statistics*, Brazilian Statistical Association, v. 34, n. 2, p. 230–250, 2020.
- FREEMAN, L. C. Elementary applied statistics: for students in behavioral science. (*No Title*), 1965.
- FUNDER, D. C.; OZER, D. J. Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science*, Sage Publications Sage CA: Los Angeles, CA, v. 2, n. 2, p. 156–168, 2019.
- GELMAN, A. *Statistical Modeling, Causal Inference, and Social Science*. 2009. https://statmodeling.stat.columbia.edu/2009/06/18/the_sample_size/. Acesso em: 7 de janeiro de 2024.
- GIGNAC, G. E.; SZODORAI, E. T. Effect size guidelines for individual differences researchers. *Personality and individual differences*, Elsevier, v. 102, p. 74–78, 2016.
- GRISSOM, R. J.; KIM, J. J. *Effect sizes for research: A broad practical approach*. [S.l.]: Lawrence Erlbaum Associates Publishers, 2005.
- HIRSCH, J. E. An index to quantify an individual's scientific research output. *Proceedings of the National academy of Sciences*, National Acad Sciences, v. 102, n. 46, p. 16569–16572, 2005.

- KENDALL, M. G. A new measure of rank correlation. *Biometrika*, JSTOR, v. 30, n. 1/2, p. 81–93, 1938.
- KING, B. M.; ROSOPA, P. J.; MINIUM, E. W. *Statistical reasoning in the behavioral sciences*. [S.l.]: John Wiley & Sons, 2018.
- LOVAKOV, A.; AGADULLINA, E. R. Empirically derived guidelines for effect size interpretation in social psychology. *European Journal of Social Psychology*, Wiley Online Library, v. 51, n. 3, p. 485–504, 2021.
- MCGRRAW, K. O.; WONG, S. P. A common language effect size statistic. *Psychological bulletin*, American Psychological Association, v. 111, n. 2, p. 361, 1992.
- NOETHER, G. E. Sample size determination for some common nonparametric tests. *Journal of the American Statistical Association*, Taylor & Francis, v. 82, n. 398, p. 645–647, 1987.
- NOMES no Brasil. 2010. (<https://censo2010.ibge.gov.br/nomes/#/search>). Acesso em: 06 abr 2022.
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2018. Disponível em: (<https://www.R-project.org/>).
- SANTO, H. E.; DANIEL, F. Calcular e apresentar tamanhos do efeito em trabalhos científicos (1): as limitações do $p < 0,05$ na análise de diferenças de médias de dois grupos (calculating and reporting effect sizes on scientific papers (1): $P < 0.05$ limitations in the analysis of mean differences of two groups). *Revista Portuguesa de Investigação Comportamental e Social*, v. 1, n. 1, p. 3–16, 2017.
- SAWILOWSKY, S. S. New effect size rules of thumb. *Journal of modern applied statistical methods*, v. 8, n. 2, p. 26, 2009.
- SCOPUS. 2023. (<https://www.scopus.com/search/form.uri?display=basic#basic>). Acesso em: 03 ago 2022.
- TOMCZAK, M.; TOMCZAK, E. The need to report effect size estimates revisited. An overview of some recommended measures of effect size. *Akademia Wychowania Fizycznego w Poznaniu*, 2014.
- TUKEY, J. W. et al. *Exploratory data analysis*. Reading, MA: Addison-Wesley, 1977. v. 2.
- VARGHA, A.; DELANEY, H. D. A critique and improvement of the cl common language effect size statistics of mcgraw and wong. *Journal of Educational and Behavioral Statistics*, Sage Publications Sage CA: Los Angeles, CA, v. 25, n. 2, p. 101–132, 2000.