

A Influência do Tamanho da Amostra na Credibilidade de Testes Estatísticos Não-Paramétricos: Uma Abordagem com Simulações de Monte Carlo

The Impact of the Sample Size on the Reliability of Non-Parametric Statistical Tests: A Monte Carlo Simulation Approach

TIAGO MAIA MAGALHÃES¹
GIOVANI PETERSON ALVES MENDES²

DOI: 10.34019/2179-3700.2024.v24.46207

ENVIADO EM: 29/10/2024

APROVADO EM: 14/11/2024

RESUMO

A Estatística é fundamental em diversas pesquisas científicas. Ela desempenha um papel vital ao possibilitar a compreensão de possíveis padrões e relações presentes nos dados, permitindo que pesquisadores tirem conclusões significativas e baseiem suas descobertas em evidências sólidas. A correta aplicação da Estatística garante que as informações coletadas sejam transformadas em conhecimento confiável e útil. No entanto, seu sucesso depende, em grande parte, de um fator-chave, o tamanho da amostra. Assim, o objetivo deste projeto de pesquisa é investigar de que maneira o tamanho da amostra impacta na decisão de dois dos principais testes estatísticos não-paramétricos: os testes de razão de verossimilhanças e de Pearson. Para alcançar esse objetivo, o projeto aplicou o método de Monte Carlo, que permitiu avaliar as taxas de rejeição dos dois testes em diferentes cenários. A variação do tamanho da amostra revelou que, à medida que o tamanho da amostra aumenta, as taxas de rejeição se tornam mais consistentes, aproximando-se de um valor ideal de 5%. Isso sugere que tamanhos de amostra maiores resultam em testes mais confiáveis. A pesquisa também comparou os testes de razão de verossimilhança e de Pearson, concluindo que ambos são equivalentes, uma vez que têm taxas de rejeição semelhantes. Portanto, qualquer um dos testes pode ser escolhido para análises futuras, já que ambos são igualmente eficazes. Em resumo, este projeto evidenciou a importância do tamanho da amostra em testes não-paramétricos. À medida que o tamanho da amostra aumenta, os resultados obtidos desses dois testes podem ser considerados mais confiáveis, tornando-os mais robustos para uma análise. Essas conclusões têm um impacto significativo na pesquisa e na tomada de decisões informadas em diversas áreas.

Palavras-chave: Tamanho da Amostra. Testes Não-Paramétricos. Teste de Razão de Verossimilhanças. Teste de Pearson. Simulações Monte Carlo.

¹ Department of Statistics, Institute of Exact Sciences, Federal University of Juiz de Fora, Juiz de Fora, Brazil, e-mail to tiago.magalhaes@ufjf.br.

² Business school, Federal University of Juiz de Fora, Juiz de Fora, Brazil, e-mail to giovani.peterson@estudante.ufjf.br.

ABSTRACT

Statistics is fundamental in various scientific research. It plays a vital role by enabling the understanding of potential patterns and relationships present in the data, allowing researchers to draw meaningful conclusions and base their findings on solid evidence. The correct application of statistics ensures that the collected information is transformed into reliable and useful knowledge. However, its success largely depends on a key factor: sample size. Thus, the aim of this research project is to investigate how sample size impacts the decision of two of the main non-parametric statistical tests: the likelihood ratio and Pearson's tests. To achieve this objective, the project applied the Monte Carlo method, which allowed for the evaluation of the rejection rates of both tests in different scenarios. The variation in sample size revealed that, as the sample size increases, rejection rates become more consistent, approaching the ideal value of 5%. This suggests that larger sample sizes result in more reliable tests. The research also compared the likelihood ratio and Pearson's tests, concluding that both are equivalent, as they have similar rejection rates. Therefore, either test can be chosen for future analyses, as both are equally effective. In summary, this project highlighted the importance of sample size in non-parametric tests. As sample size increases, the results obtained from these two tests can be considered more reliable, making them more robust for analysis. These conclusions have significant implications for research and informed decision-making in various fields.

Keywords: Sample Size. Non-Parametric Tests. Likelihood Ratio Test. Pearson's Test. Monte Carlo Simulations.

1 INTRODUÇÃO

A Estatística é uma ferramenta essencial em diversas áreas do conhecimento, desempenhando um papel central na análise e interpretação de dados. Desde a Ciência Forense (Longjohn et al., 2022) a área de análise custo-efetividade (Tayar et al., 2018). Em pesquisas científicas, a Estatística possibilita a identificação de padrões, a testagem de hipóteses e a formulação de conclusões baseadas em evidências empíricas. A aplicação adequada de métodos estatísticos é fundamental para garantir que as descobertas obtidas sejam robustas e confiáveis, permitindo que os resultados sejam generalizados e utilizados em decisões informadas. Contudo, o sucesso dessas análises depende de vários fatores, entre os quais o tamanho da amostra ocupa uma posição de destaque.

O tamanho da amostra é um dos principais determinantes da precisão e da credibilidade dos resultados em testes estatísticos. Amostras inadequadamente pequenas podem levar a conclusões errôneas, como a falha em detectar efeitos reais ou a superestimação da variabilidade dos dados. Por outro lado, amostras grandes o suficiente podem garantir que os testes estatísticos sejam mais

sensíveis e específicos, aumentando a probabilidade de detectar padrões significativos. Portanto, compreender como o tamanho da amostra influencia a eficácia dos testes estatísticos é crucial para a validação dos métodos aplicados em pesquisas.

Os testes estatísticos não-paramétricos, como os testes de razão de verossimilhanças (Wilks, 1938) e de Pearson (Pearson, 1900), são amplamente utilizados em situações onde as suposições paramétricas tradicionais, como a normalidade dos dados, não são válidas. Esses testes são especialmente úteis em análises onde as distribuições dos dados são desconhecidas ou difíceis de modelar. No entanto, a eficácia desses testes pode variar consideravelmente com o tamanho da amostra, o que torna a avaliação dessa relação um tema de grande relevância.

Diante da importância do tamanho da amostra para a credibilidade dos resultados obtidos dos testes estatísticos, esta pesquisa tem como objetivo investigar de que maneira a variação do tamanho da amostra afeta a taxa de rejeição dos testes de razão de verossimilhanças e de Pearson. Para isso, utilizou-se simulações Monte Carlo, uma técnica robusta que permite a geração de múltiplos cenários e a avaliação detalhada dos resultados em diferentes condições. A pesquisa visa, assim, fornecer insights sobre a escolha apropriada do tamanho da amostra em análises não-paramétricas.

Este artigo está organizado da seguinte maneira: na próxima seção, apresentamos uma revisão da literatura sobre o papel do tamanho da amostra em testes estatísticos, seguida pela descrição detalhada da metodologia empregada, incluindo as simulações Monte Carlo (Kroese et al, 2014). Em seguida, discutimos os resultados obtidos, comparando o desempenho dos testes de razão de verossimilhanças e de Pearson em diferentes tamanhos de amostra. Por fim, concluímos com uma discussão sobre as implicações dos achados para futuras pesquisas e a aplicação prática dos testes estatísticos não-paramétricos.

2 DESENVOLVIMENTO

Sem perda de generalidade, suponha que nós estamos interessados em observar se existe um padrão nas respostas de dois grupos de respondentes em relação a uma pergunta de duas opções. Dessa forma, nós poderíamos construir uma tabela de contingência e definir a seguinte hipótese nula: “o padrão de resposta dos dois grupos é homogêneo, isto é, igual”. A estatística de Pearson (Pearson, 1900) para testar se a hipótese nula deve ser rejeitada ou não é dada por:

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

enquanto a estatística da razão de verossimilhanças (Wilks, 1938) por

$$\chi_V^2 = \sum_{i=1}^2 \sum_{j=1}^2 \left\{ O_{ij} \times \log \left(\frac{O_{ij}}{E_{ij}} \right) \right\},$$

em que O_{ij} é a frequência observada na linha i e coluna j da tabela, E_{ij} é a frequência esperada na linha i e coluna j da tabela, E_{ij} é calculado pela razão entre total de observações na linha i multiplicado pelo total na coluna j e o total geral de observações e \log é logaritmo de base e . Sob a hipótese nula, as duas estatísticas têm distribuição qui-quadrado com 1 grau de liberdade (12). A hipótese nula será rejeitada, para um dado nível nominal e uma estatística especificada, se a estatística do teste for maior que o quantil $100(1 - \alpha)\%$ de uma distribuição qui-quadrado com 1 grau de liberdade.

3 METODOLOGIA

Para avaliar a influência do tamanho da amostra na credibilidade dos testes de razão de verossimilhanças e de Pearson, empregamos simulações Monte Carlo. Este método é amplamente reconhecido por sua capacidade de gerar múltiplos cenários experimentais, permitindo uma análise robusta das taxas de

rejeição sob diferentes condições de tamanho amostral. As simulações foram realizadas no ambiente de programação R (R Core Team, 2022), utilizando-se scripts customizados para controlar a variação do tamanho da amostra e as características dos dados.

As simulações foram conduzidas variando-se o tamanho da amostra em um intervalo que ia de pequenas amostras ($n = 30$) até amostras consideravelmente grandes ($n = 1.000$), em incrementos específicos. Para cada tamanho amostral, geramos 1.000 replicações, a fim de garantir a estabilidade e a robustez dos resultados. Para cada tamanho de amostra e replicação, foram gerados dois conjuntos dados independentes, seguindo distribuição Bernoulli, com a mesma probabilidade de sucesso e esta igual a 0,5.

Em cada replicação, aplicamos os testes de razão de verossimilhanças e de Pearson para avaliar a homogeneidade dos grupos dentro do conjunto de dados. As taxas de rejeição foram calculadas com base na frequência com que a hipótese nula foi rejeitada em cada cenário experimental, isto é, foi mensurado a proporção de vezes que cada um dos dois testes concluiu que a proporção de sucesso do Grupo I foi diferente do Grupo II, isto é, proporção de vezes em que o teste errou. Este tipo de erro é denominado de Erro do Tipo I (“rejeitar a hipótese nula quando ela é verdadeira”). Estas taxas foram comparadas com o nível de significância padrão de 5%, considerado ideal para testes estatísticos, a fim de verificar a consistência e a credibilidade dos resultados em função do tamanho da amostra.

Além da avaliação individual dos testes, foi realizada uma comparação direta entre o desempenho dos testes de razão de verossimilhanças e de Pearson. Essa comparação focou na equivalência das taxas de rejeição em diferentes tamanhos de amostra, buscando identificar qual dos dois testes apresentou melhor desempenho em determinados cenários. A equivalência foi determinada pela proximidade das taxas de rejeição ao valor de referência, o que indicaria que ambos os testes são igualmente eficazes.

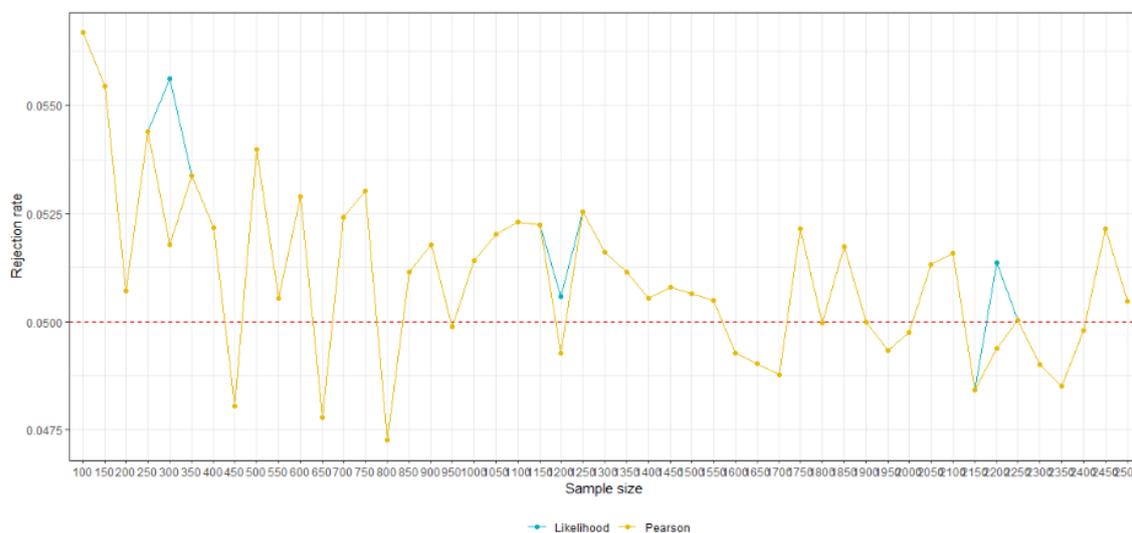
Para facilitar a interpretação dos resultados, gráficos foram gerados a partir das simulações. Estes gráficos ilustram como as taxas de rejeição dos testes variaram conforme o tamanho da amostra, proporcionando uma visualização clara da relação entre o tamanho da amostra e a confiabilidade dos testes. O gráfico foi produzido utilizando-se o pacote ggplot2 (Wickham, 2016), que permitiu uma

análise visual detalhada e uma comunicação eficaz dos achados da pesquisa.

3 RESULTADOS E DISCUSSÕES

A análise dos resultados apresentados no gráfico destacou a relação entre o tamanho da amostra e a taxa coletada dos testes de verossimilhança e Pearson.

Figura 1: Taxas de exclusão em função do tamanho da amostra



Fonte: Elaborado pelos autores (2024)

O gráfico mostra as taxas de exclusão em função do tamanho da amostra, variando de 100 a 2500 observações. A linha vermelha pontilhada representa o nível de significância de 5%, que é o valor de referência ideal para a taxa de exclusão em testes estatísticos.

Observa-se que, em ambos os testes, as taxas de exclusão flutuam em torno da linha de 5%, com variações que indicam a sensibilidade dos testes ao tamanho da amostra. O teste de verossimilhança (indicado pela linha azul) apresenta maior estabilidade em tamanhos de amostra menores, mas sua taxa de inclusão aproxima-se do valor ideal à medida que o tamanho da amostra aumenta. O teste de Pearson (indicado pela linha amarela), por sua vez, mostra uma maior variabilidade em toda a faixa de tamanhos de amostras testadas, embora tenda a convergir para o valor de 5% em amostras maiores.

Essa variabilidade nas taxas de exclusão sugere que ambos os testes são

sensíveis ao tamanho da amostra, com o teste de verossimilhança exibindo um comportamento cuidadoso mais consistente em cenários com amostras pequenas. No entanto, para amostras maiores, as diferenças entre os dois testes tornam-se menos pronunciadas, indicando que ambos os testes podem ser considerados igualmente adequados para tamanhos de amostra acima de 500 observações.

A análise dessas confirma a importância do tamanho da amostra na aplicação de testes estatísticos não-paramétricos. Em contextos onde se trabalha com amostras menores, a escolha do teste pode ter um impacto significativo na taxa de visualização observada, influenciando potencialmente as conclusões do estudo. Por outro lado, para amostras mais robustas, as diferenças entre os testes diminuem, proporcionando maior facilidade na escolha do método a ser utilizado. Essas descobertas corroboram a necessidade de considerar o tamanho da amostra como um fator crítico ao análises planejadas, especialmente em estudos que utilizam testes não paramétricos como os de verossimilhança e Pearson.

REFERÊNCIAS

KROESE, D. P.; BRERETON, T.; TAIMRE, T.; BOTEV, Z. I. **Why the Monte Carlo method is so important today.** *WIREs Computational Statistics*, v. 6, n. 6, p. 386-392, 2014.

LONGJOHN, R.; SMYTH, P.; STERN, H. S. **Likelihood ratios for categorical count data with applications in digital forensics.** *Law, Probability and Risk*, v. 21, n. 2, p. 91-122, 2022.

PEARSON, K. **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 (Series 5)*, v. 50, n. 302, p. 157-175, 1900.

R CORE TEAM. **R: A language and environment for statistical computing.** Vienna: R Foundation for Statistical Computing, 2022.

TAYAR, D. O.; RIBEIRO JR., U.; CECCONELLO, I.; MAGALHÃES, T. M.; SIMÕES, C.; AULER JR., J. O. C. **Propensity score matching comparison of laparoscopic versus open surgery for rectal cancer in a middle-income country: short-term outcomes and cost analysis.** *ClinicoEconomics and Outcomes Research*, v. 10, p. 521-527, 2018.

WICKHAM, H. **ggplot2: Elegant Graphics for Data Analysis.** New York: Springer-Verlag, 2016.

WILKS, S. S. **The large-sample distribution of likelihood ratio for testing composite hypotheses.** *The Annals of Mathematical Statistics*, v. 9, n. 1, p. 60-62, 1938.