ANÁLISE DO EFEITO DO TAMANHO AMOSTRAL NA PRECISÃO DE ESTIMATIVAS DE PROPORÇÃO

GEORGIA LISBOA MAIA[1]; WILLIAN SILVA BARROS[2]; GISELDA MARIA PEREIRA[3]

[1]Universidade Federal de Pelotas – georgialisboamaia@gmail.com
[2]Universidade Federal de Pelotas– wsbarros@hotmaial.com
[3]Universidade Federal de Pelotas– – gmpereira08@gmail.com

1. INTRODUÇÃO

As metodologias de análise quantitativa de diversos fenômenos, sejam eles econômicos, físicos ou sociais, frequentemente, adotam modelos tradicionais buscando uma única resposta precisa, ou uma abordagem conhecida como determinística (ROSS, 1990). Contudo, este modelo acaba muitas vezes por não considerar plenamente dois aspectos: a presença da incerteza e da variabilidade. Desta forma, para tratar essa complexidade é necessário adotar métodos que compreendam a imprevisibilidade permitindo uma análise mais robusta e realista (SALTELLI et al., 2000).

A amostragem estatística, definida como o procedimento utilizado para obter um subconjunto representativo de uma população para inferir suas características (COCHRAN, 1965), exerce papel central na obtenção de resultados confiáveis. Neste sentido, a escolha adequada do tamanho amostral é um passo crucial no planejamento de pesquisas quantitativas, pois amostras pequenas podem gerar estimativas instáveis e pouco representativas. À medida que o tamanho da amostra aumenta, observa-se a estabilização das estimativas em torno de valores centrais, o que confere maior consistência e precisão aos resultados (COCHRAN, 1977; AGRESTI, 2013).

Os processos de amostragem podem ser probabilísticos ou não probabilísticos. Dentre os processos probabilísticos para obtenção de amostras pode-se destacar a amostragem aleatória simples, a amostragem estratificada, a amostragem sistemática e a amostragem por conglomerados, cada um com suas vantagens específicas conforme o contexto da pesquisa (ANDRADE; OGLIARI, 2017; KISH, 1965).

Nesse cenário, a Simulação de Monte Carlo surge como uma ferramenta poderosa para modelar a incerteza, atribuindo distribuições de probabilidade às variáveis incertas e realizando múltiplas iterações para gerar cenários plausíveis. A qualidade da amostragem utilizada na simulação é determinante para a representatividade dos resultados e para a convergência do modelo. Com número suficiente de repetições, a distribuição dos resultados torna-se estável, permitindo conclusões fundamentadas e confiáveis (METROPOLIS; ULAM, 1949; FISHMAN, 1996). Em estudos de amostragem a Simulação de Monte Carlo permite testar diferentes cenários, identificando os possíveis problemas que poderiam ser encontrados em situações reais.

Diante disso, o presente estudo teve como objetivo analisar o impacto do tamanho amostral e da magnitude do erro máximo da estimativa, em amostras aleatórias simples, sobre a estabilidade e confiabilidade das estimativas de proporção populacional, utilizando a Simulação de Monte Carlo.

2. METODOLOGIA

Nesta investigação, utilizou-se uma população finita, dicotômica, com N=100.000 elementos, contendo dois níveis igualmente proporcionais, ou seja, com proporção populacional π=0,5. Esta configuração representa a situação de máxima variabilidade para variáveis binárias, sendo amplamente recomendada para estudos de amostragem por maximizar a variância da estimativa amostral (COCHRAN, 1977).

Foram calculados os tamanhos amostrais (n) para estimação da proporção populacional (π) de 385, 2.401 e 9.604 elementos, para os erros máximos para estimativa de 5%, 2% e 1%, respectivamente, conforme a expressão abaixo: $n = \frac{Z^2_{\alpha/2}.p.(1-P)}{E^2}$ onde $Z_{\alpha/2}$ é o quantil da distribuição normal padrão para o nível de significância desejado, e E é o erro máximo tolerado para a estimativa (AGRESTI, 2013; COCHRAN, 1977).

A partir destes valores base, foram consideradas amostragens em múltiplos e frações do tamanho amostral n, especificamente n/2, n, 2n, 3n, 4n e 5n, para investigar o comportamento da estimativa amostral da proporção. Para cada cenário amostral, foram realizadas 1000 amostragens independentes retiradas da população original, e calculadas as estatísticas descritivas média e desvio padrão da proporção estimada. Posteriormente, as estimativas da proporção obtidas foram testadas pelo teste Z ao nível de 5% de probabilidade, obtendo-se o valor *p* para cada tamanho amostral. Para avaliação da variabilidade entre repetições, este processo foi repetido dez vezes, denominadas repetições R1 a R10, obtendo-se o valor *p* do teste e o desvio padrão do valor *p*, para cada um dos cenários testados.

As simulações computacionais foram implementadas na linguagem R (R CORE TEAM, 2024) e as representações gráficas dos resultados foram obtidas através de planilhas eletrônicas.

3. RESULTADOS E DISCUSSÃO

Os resultados obtidos para o erro tipo I médio, considerando erros máximos admitidos de 5%, 2% e 1% para a estimativa da proporção populacional em diferentes tamanhos amostrais, são apresentados na Figura 1.

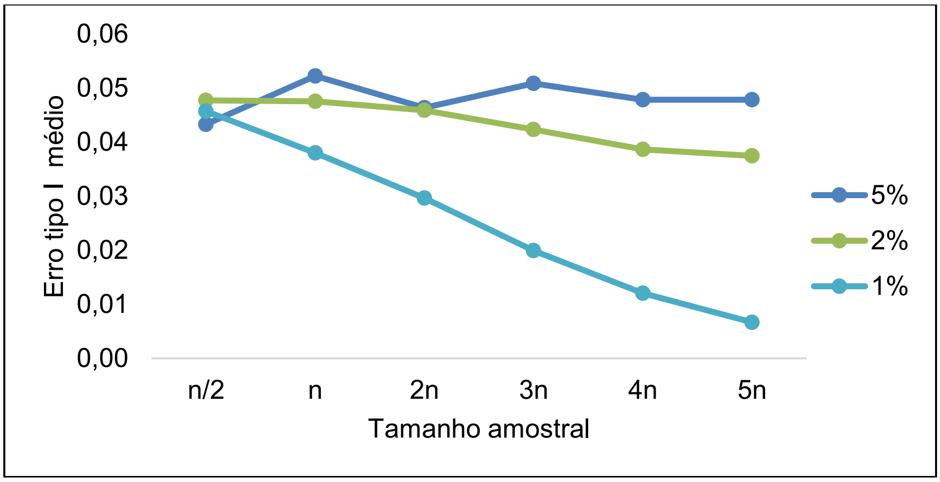

Figura 1- Erro tipo I médio para os testes de hipótese da proporção populacional nos diferentes cenários avaliados.

De forma geral, observou-se que o erro tipo I médio apresentou variação sistemática com o aumento do tamanho da amostra (Figura 1), evidenciando tendência predominante de redução à medida que n aumentou, o que está em consonância com o comportamento esperado de estimadores consistentes (LOHR, 2019). Erro máximo de 5%: o incremento do tamanho amostral até 5n resultou em apenas redução marginal do erro tipo I médio em relação ao tamanho amostral básico, sugerindo que, para níveis de precisão mais modestos, os ganhos adicionais tornam-se decrescentes, aspecto também discutido por COCHRAN (1977) no contexto do princípio de retornos decrescentes em amostragem. Para um erro máximo de 2%: verificou-se um declínio gradual no erro tipo I com o aumento de n, refletindo ganhos moderados de estabilidade. Erro máximo de 1%: houve redução acentuada do erro tipo I, demonstrando que, em cenários de alta exigência de precisão, a ampliação da amostra desempenha papel crucial na diminuição da probabilidade de rejeição incorreta da hipótese nula, o que confirma o relatado por AGRESTI (2013) sobre o impacto do tamanho amostral em testes de hipóteses.

Em relação ao desvio padrão médio do erro tipo I (Figura 2), o aumento de n promoveu redução substancial para erros máximos de 2% e 1%, indicando maior estabilidade das estimativas entre repetições e reforçando o efeito de convergência amostral descrito por KISH (1965). Para o erro máximo de 5%, contudo, o desvio padrão permaneceu praticamente constante, oscilando em torno de 0,02, mesmo com grandes aumentos no tamanho amostral, sugerindo saturação no ganho de precisão para este nível de tolerância ao erro.

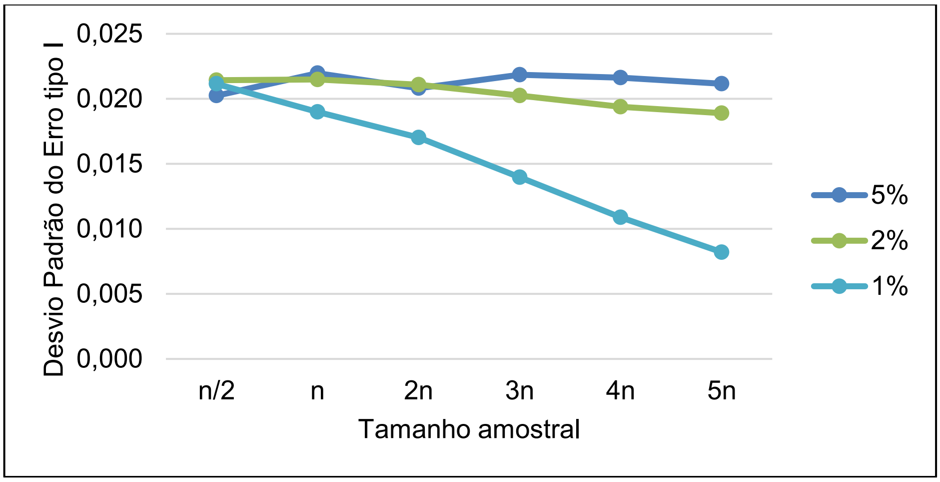

Figura 2- Desvio Padrão do Erro tipo I observado para os testes de hipótese da proporção populacional nos diferentes cenários avaliados.

No que se refere ao desvio padrão médio do erro tipo I (Figura 2), nota-se que à medida que o tamanho da amostra aumenta, o desvio padrão tende a diminuir para os erros máximos de 2% e 1%, indicando um aumento da precisão da estimativa. Para o erro máximo de 5%, contudo, o desvio padrão permaneceu praticamente constante, oscilando em torno de 0,02, mesmo com grandes aumentos no tamanho amostral, sugerindo saturação no ganho de precisão para este nível de tolerância ao erro.

A importância da definição criteriosa do tamanho amostral na estimação de proporções tem sido amplamente reconhecida. Os autores AGRANONIK e HIRAKATA (2011) destacam que o dimensionamento adequado melhora a precisão das estimativas e confere maior confiabilidade às conclusões. Os achados deste estudo corroboram com essa visão, ao mostrarem que, para erros máximos de 5% e 2%, o erro tipo I médio pouco se altera mesmo com aumento substancial da amostra, enquanto para 1% os benefícios são evidentes.

4. CONCLUSÕES

A Simulação de Monte Carlo mostrou-se eficiente para avaliar o efeito do tamanho amostral e do erro máximo na estimação de proporções binárias. Sobretudo, verificou-se que o aumento de n reduz o erro tipo I e sua variabilidade, para alta precisão para o nível de 1%, enquanto para os níveis 2% e 5% os ganhos são mínimos. Esses resultados oferecem suporte prático para definir tamanhos amostrais equilibrando precisão, confiabilidade e custo em pesquisas quantitativas.

5. REFERÊNCIAS BIBLIOGRÁFICAS

AGRANONIK, M.; HIRAKATA, V. N. **Cálculo de tamanho de amostra: proporções.** Revista HCPA. 31(3):382-388, 2011.

AGRESTI, A. **Statistical inference**. 2. ed. New York: CRC Press, 2013.

ANDRADE, D.F., OGLIARI, P. J. **Estatística para as Ciências Agrárias e Biológicas: com noções de experimentação**. 3 ed. Editora da EDUFSC UFSC. 2017.

COCHRAN, W.G. **Sampling techniques**. 3. ed. New York: Wiley, 1977.

COCHRAN, W.G. **Sampling techniques**. 1. ed. New York: Wiley, 1965.

FISHMAN, G.S. **Monte Carlo: concepts, algorithms, and applications**. New York: Springer-Verlag, 1996.

KISH, L. **Survey sampling**. New York: Wiley, 1965.

LOHR, S. L. **Sampling: Design and Analysis**. 2nd ed. Cengage Learning, 2019.

METROPOLIS, N.; ULAM, S. **The Monte Carlo method.** Journal of the American Statistical Association, v. 44, n. 247, p. 335–341, 1949.

R CORE TEAM. **R: A language and environment for statistical computing.** Vienna, Austria: R Foundation for Statistical Computing, 2024.

ROSS, S.M. **Introduction to probability models.** 4. ed. New York: Academic Press, 1990.

SALTELLI, A.; TARANTOLA, S.; CAMPOLONGO, F. **Sensitivity analysis**. New York: Wiley, 2000.