

AVALIAÇÃO DE TRÊS ESTIMADORES DE QUANTIS EM DISTRIBUIÇÕES BETA SIMULADAS VIA MONTE CARLO

CAMILA DE MORAES RAMSON¹; ALEXANDER JOSE DE SENA², GISELDA MARIA PEREIRA³, WILLIAN SILVA BARROS⁴

¹*Universidade Federal de Pelotas – milaramson@gmail.com*

² *Universidade Federal de Pelotas – asena774@gmail.com*

³ *Universidade Federal de Pelotas – gmpereira08@gmail.com*

⁴ *Universidade Federal de Pelotas – willian.barros@ufpel.edu.br*

1. INTRODUÇÃO

A Estatística é definida como uma ciência que dispõem de técnicas para coleta, organização e análise de dados, a fim de obter resultados e prosseguir com o planejamento e tomada de decisões (IGNÁCIO, 2010), integrando a metodologia científica e sendo imprescindível na estruturação de diversas áreas do conhecimento.

De forma complementar, a disseminação de tecnologias e ampliação da capacidade de processamento das ferramentas computacionais, favoreceu a utilização de softwares para a análise de dados, sendo ao mesmo tempo a linguagem R, o maior recurso implementado na área da estatística, nos últimos vinte anos (SMOLSKI et al., 2018).

Uma vez que os dados são coletados, a análise estatística instaura, na maioria dos casos, o cálculo de medidas descritivas. Além desta perspectiva, há ainda as estatísticas inferenciais, cuja análise se dá a respeito de inferências sobre as populações a partir das quais os dados amostrais foram coletados (LARSON, 2006). A determinação dos estimadores implica na contemplação de problemas relacionados ao uso das diferentes fórmulas, aos métodos de estimação e ao processo de amostragem (SILVA et al., 2021).

O emprego de medidas que possibilitam representar de forma resumida um conjunto de dados é imprescindível para sua exploração. A estimativa com base no uso de quantis, como medidas separatrizes de resumo, que dividem o conjunto de dados ordenados em partes, constituem uma ferramenta de interesse para os estatísticos há um longo tempo (KAPPENMAN, 1987). Os quantis podem ser determinados por diferentes métodos e facilitam a compreensão e a interpretação dos dados encontrados a partir das análises.

Dentre as vantagens do estudo dos quantis, destaca-se sua relevância ao resumir quantidades expressivas de elementos em bancos de dados, fornecendo resultados mais precisos, ao passo que, nesta conjuntura, estatísticas como a média e a variância são insuficientemente descritivas e altamente sensíveis a anomalias de dados nas distribuições (GILBERT et al., 2002). Além disso, os quantis também são empregados na construção do Box Plot, que é um tipo de gráfico usado regularmente na pesquisa científica e cuja elaboração é possível por meio de diferentes softwares estatísticos (NETO et al., 2017). Contudo, ainda cabe uma maior exploração a respeito dos estimadores de quantis, observando pontos referentes ao seu comportamento, como a investigação do viés e sua relação com a assimetria da distribuição e tamanho da amostra.

Diante disso, o objetivo deste trabalho foi comparar alguns tipos de estimadores de quantis, em amostras com diferentes tamanhos, obtidas a partir de

populações com distribuições simétrica, assimétrica positiva e assimétrica negativa.

2. METODOLOGIA

Para a realização deste estudo, foram simuladas amostras com tamanhos de 5, 10, 20, 30, 50, 100, 150, 200, 250, 500, 1000 elementos para distribuições Beta simétricas, assimétricas positivas e assimétricas negativas. Para cada uma das amostras foram obtidos os percentis P10, P25, P50, P75, P95, com a utilização dos estimadores tipo um (type 1), tipo dois (type 2) e tipo três (type 3) do *R* (R Core Team, 2023) de acordo com HYNDMAN; FAN (1996), totalizando 99 cenários para cada percentil. Foram realizadas 10.000 simulações de Monte Carlo (RUBINSTEIN; KROESE, 2013) para cada cenário, em que foram obtidos os percentis acima descritos e as respectivas médias e desvios padrões para as estimativas. Os dados foram exportados em arquivo de texto (.txt) e apresentados graficamente, para melhor visualização, com a utilização do *Excel*®.

2. RESULTADOS E DISCUSSÃO

O resumo dos resultados obtidos a partir das simulações para os diferentes cenários testados é apresentado na Figura 1.

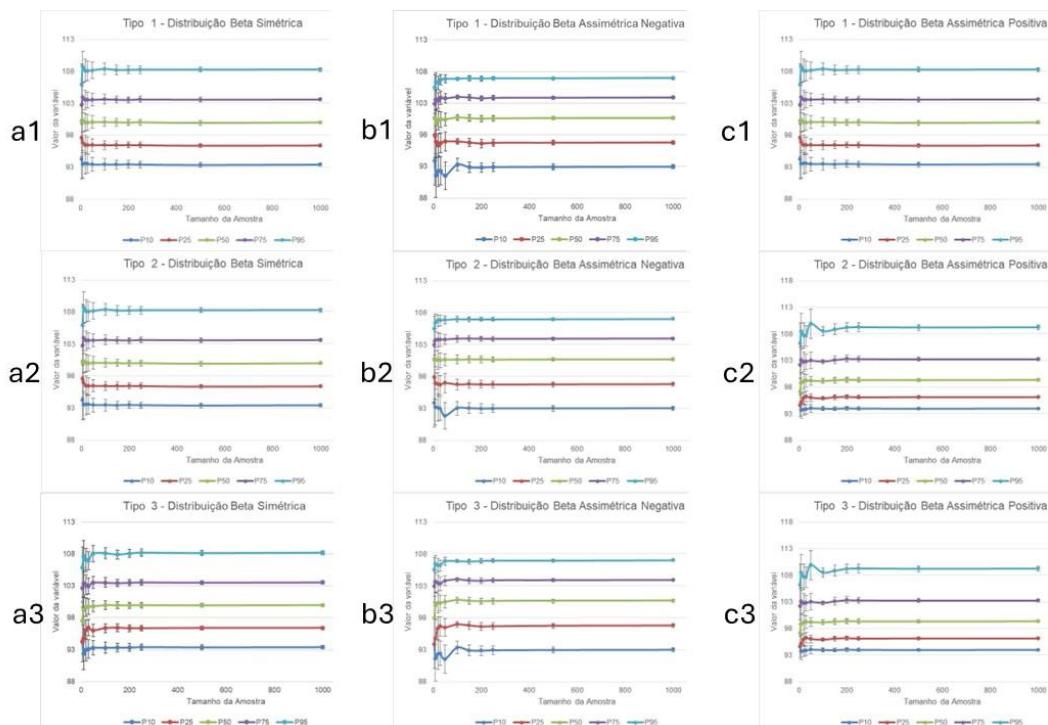


Figura 1- Valores médios e desvios padrões para os quantis obtidos para os diferentes cenários utilizando 10.000 simulações. a1) beta simétrica, utilizando o estimador tipo 1; a2) beta simétrica, utilizando o estimador tipo 2; a3) beta simétrica, utilizando o estimador tipo 3; b1) beta assimétrica negativa, utilizando o estimador tipo 1; b2) beta assimétrica negativa, utilizando o estimador tipo 2; b3) beta assimétrica negativa, utilizando o estimador tipo 3; c1) beta assimétrica positiva, utilizando o estimador tipo 1; c2) beta assimétrica positiva, utilizando o estimador tipo 2; c3) beta assimétrica positiva, utilizando o estimador tipo 3.

Os resultados mostram que as estimativas obtidas para os quantis em estudo, quando a distribuição é simétrica, são semelhantes para os três estimadores testados (Figura 1 - a1, a2 e a3). Este resultado indica que, para distribuições simétricas como a testada, os três estimadores avaliados fornecem valores não tendenciosos e resumem adequadamente o conjunto de dados.

A escolha de um estimador de parâmetro implica na observação de propriedades importantes e inerentes de um bom estimador, como a não tendenciosidade, a fim de evitar que o estimador do parâmetro populacional seja viesado; a acurácia, sendo o estimador consistente quando à medida que aumenta o tamanho da amostra a estimativa se aproxima do parâmetro, e a precisão, cuja base é a variância mínima do estimador (ANDRADE; OGLIARI, 2010). A seleção dos três tipos de estimadores para os percentis do R, utilizados neste trabalho, também se deve ao fato de sua simplicidade e ampla utilização.

Em relação à distribuição Beta assimétrica negativa, observou-se que o estimador tipo 2 (Figura 1 – b2) apresentou maior precisão na estimativa dos quantis indicando um melhor desempenho, especialmente no que se refere a amostras de até 50 elementos. Os estimadores 1 e 3, no entanto, apresentaram maior variabilidade para estimativas (Figura 1– b1 e b3), indicando uma menor precisão dos estimadores para os quantis estudados. Foi observado também que quando a distribuição é assimétrica negativa e as amostras são menores, o percentil 10 apresentou grande variabilidade nas estimativas, especialmente pelos estimadores tipo 1 e 3 (Figura 1– b1 e b3).

Por fim, em relação à distribuição Beta assimétrica positiva, o estimador tipo 1 (Figura 1 – c1) apresentou um desempenho pior que os estimadores tipo 2 (Figura 1 – c2) e tipo 3 (Figura 1 – c3) para os percentis 10, 25, 50 e 75, quando as amostras eram pequenas. Já, para o percentil 95, observou-se que os estimadores tipo 2 e 3 (Figura 1 – c2 e c3) apresentaram menor precisão quando consideradas amostras de até 50 elementos.

Com relação a utilização da distribuição Beta, verificou-se que ela foi adequada para o estudo. De acordo com ALFAIA (2021), esta distribuição possui grande versatilidade de aplicação e frequente uso para modelagem de variáveis aleatórias contínuas.

No que se refere ao impacto do tamanho amostral na obtenção das estimativas dos quantis, observou-se que à medida que este aumenta, é possível obter estimativas mais estáveis (Figura 1). Estes resultados estão de acordo com RODRIGUES (1996) que afirma que conforme o tamanho da amostra se aproxima do tamanho da população, as informações obtidas ganham confiabilidade, o que não pode ser garantido em situações de amostras muito pequenas.

4. CONCLUSÕES

Foram encontrados resultados promissores e elucidativos quanto ao comportamento dos estimadores de quantis para análise dos dados de amostras de diferentes tamanhos, em distribuições Beta, de forma que se adequam e respondem bem a partir de certo tamanho amostral.

Observou-se que maiores variabilidades nas estimativas foram obtidas para distribuições assimétricas e, especialmente, nos casos em que as amostras eram pequenas. Também, surgem novos pontos a serem explorados, como a ampliação das simulações, com a inclusão e comparação dos demais estimadores para

obtenção de percentis, considerando outros modelos de distribuições de probabilidade.

5. REFERÊNCIAS BIBLIOGRÁFICAS

ANDRADE, D. F., OGLIARI, P. J. **Estatística para as Ciências Agrárias e Biológicas: com noções de experimentação.** 2ed. Editora da UFSC. 2010.

ALFAIA, L. M. **A distribuição beta bimodal: propriedades e aplicações.** 2021. 62 f., il. Trabalho de Conclusão de Curso (Bacharelado em Estatística) — Universidade de Brasília.

GILBERT, A. C., KOTIDIS, Y., MUTHUKRISHNAN S., STRAUSS, M. J. 2002. How to summarize the universe: dynamic maintenance of quantiles. **VLDB Endowment.** p. 454–465, 2002.

IGNÁCIO, S. A. Importância da Estatística para o Processo de Conhecimento e Tomada de Decisão. **Nota técnica Ipardes**, nº 6. Curitiba. 2010.

HYNDMAN, R.J., FAN, Y. Sample quantiles in statistical packages. **The American Statistician**, 50 (4), 361–365, 1996.

KAPPENMAN, R. F. Improved distribution quantile estimation. **Communications in Statistics - Simulation and Computation**, v. 16:2, p. 307-320, 1987.

LARSON, M.G. Descriptive statistics and graphical displays. **Circulation**. v.114 p.76-81, 2006.

NETO, J. V., SANTOS, C. B. dos, TORRES, E. M., ESTRELA, C. Boxplot: um recurso gráfico para a análise e interpretação de dados quantitativos. **Rev Odontol Bras Central**; v. 26(76): p. 1-6, 2017.

R CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Viena, Áustria. 2023.

RODRIGUES, E. A. **Um estudo sobre a validação do tamanho da amostra em pesquisa estatística.** 1996. Dissertação. (Mestrado em Engenharia Elétrica) - Universidade Federal de Uberlândia.

RUBINSTEIN, R. Y., KROESE, D. P. **Simulation and the Monte Carlo method.** Hoboken: John Wiley & Sons, 2013.

SILVA, P. L. do N., BIANCHINI, Z. M., DIAS, A. J. R. **Amostragem: Teoria e prática usando R.** Rio de Janeiro: [s.n.], 2021.

SMOLSKI, F. M. da S., BATTISTI, I. E., CHASSOT, T., REIS, D. I., KASZUBOWSKI, E., RIEGER, D. S. Capacitação em análise estatística de dados com o uso do software livre R. **Rev. Ciênc. Ext.** v.14, n.3, p.123-134, 2018.