

AVALIAÇÃO DE ALGUNS MÉTODOS DE ESTIMAÇÃO DE QUANTIS EM DIFERENTES TAMANHOS AMOSTRAIS POR SIMULAÇÃO MONTE CARLO

ALEXANDER JOSÉ DE SENA¹; CAMILA DE MORAES RAMSON², GISELDA MARIA PEREIRA³, WILLIAN SILVA BARROS⁴

¹Universidade Federal de Pelotas – asena774@gmail.com

²Universidade Federal de Pelotas – milaramson@gmail.com

³Universidade Federal de Pelotas – gmpereira08@gmail.com

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

1. INTRODUÇÃO

Análise exploratória de dados é a parte da Estatística que trata da organização, apresentação, resumo e descrição dos dados. Tal subárea se preocupa, em geral, em mostrar o conjunto de dados em elementos visuais, como, por exemplo: tabelas e gráficos, buscando assim, fazer todas as interpretações possíveis visando elucidar os objetivos pesquisados (IGNÁCIO, 2010).

A análise de quantis em diferentes tamanhos amostrais é amplamente utilizada na área de ciências agrárias. Quantis são medidas separatrizes, que dividem o conjunto de dados em 2 partes, onde o Q50 é o mais conhecido e divide o conjunto de dados em 2 partes denominado de mediana, que divide o conjunto de dados em 2 metades, onde, os quantis permitem obter informações sobre a forma, o valor representativo, a dispersão e os *outliers* da distribuição dos dados amostrais (IGNÁCIO, 2010).

A expansão tecnológica e o incremento da capacidade de processamento de dados por computadores contribuíram para a popularização de softwares estatísticos. Dentre essas ferramentas, a linguagem R se destaca como a mais adotada nos últimos vinte anos, conforme apontado por (SMOLSKI, 2018)."

Para alcançar o desafio de buscar o estimador não tendencioso, de variância mínima e consistente as estimativas dos quantis podem ser avaliadas em termos de acurácia e a precisão pelos estudos de simulação. As simulações pelo método de Monte Carlo permitem a geração de um grande número de amostras aleatórias com uma dada distribuição de probabilidade, portanto esse processo possibilita avaliar o comportamento dos estimadores em diferentes tamanhos e distribuições de probabilidades diferentes, fornecendo estimativas que permitem comparar os diferentes estimadores de quantis de acordo com seu viés e precisão em relação ao valor paramétrico.

O objetivo deste trabalho foi comparar a eficiência de três estimadores de quantis do R em amostras aleatórias de diferentes tamanhos, obtidas a partir de populações normalmente distribuídas, com a utilização de simulações de Monte Carlo.

2. METODOLOGIA

Para a realização deste estudo, foram simuladas amostras, obtidas de populações normalmente distribuídas com média 100 unidades e desvio padrão de 5 unidades e tamanhos de 5, 10, 20, 30, 50, 100, 150, 200, 250, 500, 1000 elementos. Para cada uma das amostras foram obtidos os percentis P10, P25, P50, P75, P95, com a utilização dos métodos: tipo um (type 1), tipo dois (type 2) e tipo

três (type 3) conforme HYNDMAN; FAN (1996) do *R* (R Core Team, 2023), totalizando 33 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, a 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®.

3. RESULTADOS E DISCUSSÃO

A análise dos resultados mostrou que o estimador tipo 1 foi eficiente na obtenção de quantis, especialmente para os percentis 25, 50 e 75 (Figura 1).

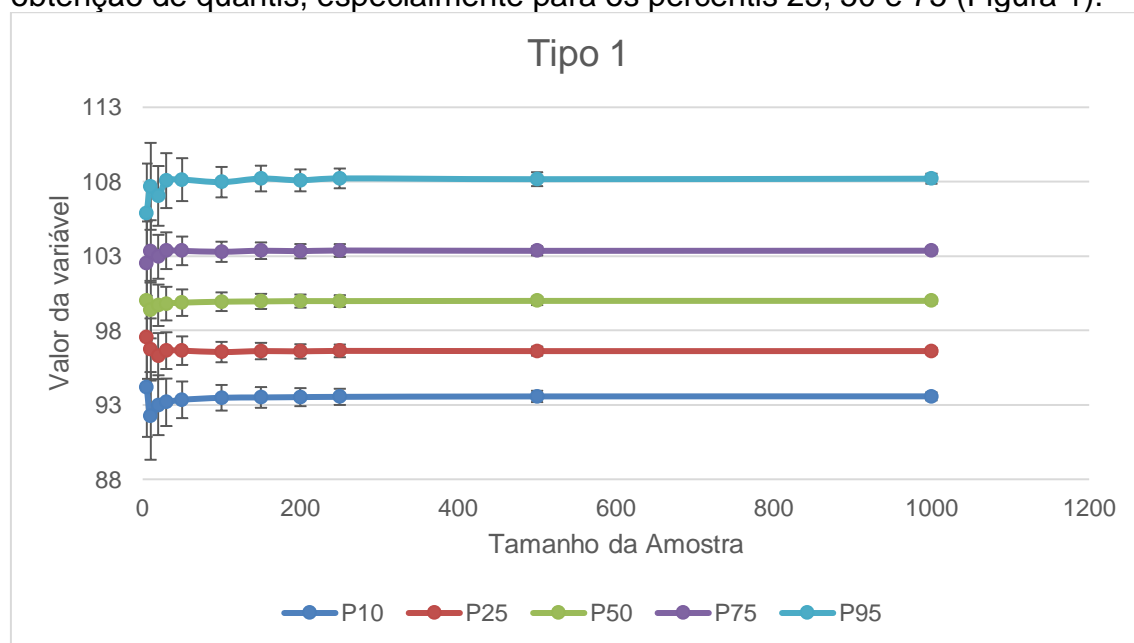


Figura 1- Valores médios e desvios padrões dos quantis P10, P25, P50, P75, P95 obtidos através do estimador tipo 1 para diferentes tamanhos amostrais.

O estimador tipo 2 apresentou estimativas mais precisas que os demais estimadores para os quantis considerados (Figura 2). Este resultado assegura a obtenção de boas estimativas quando as variáveis são normalmente distribuídas, indicando que as propriedades de um bom estimador foram mais bem atendidas.

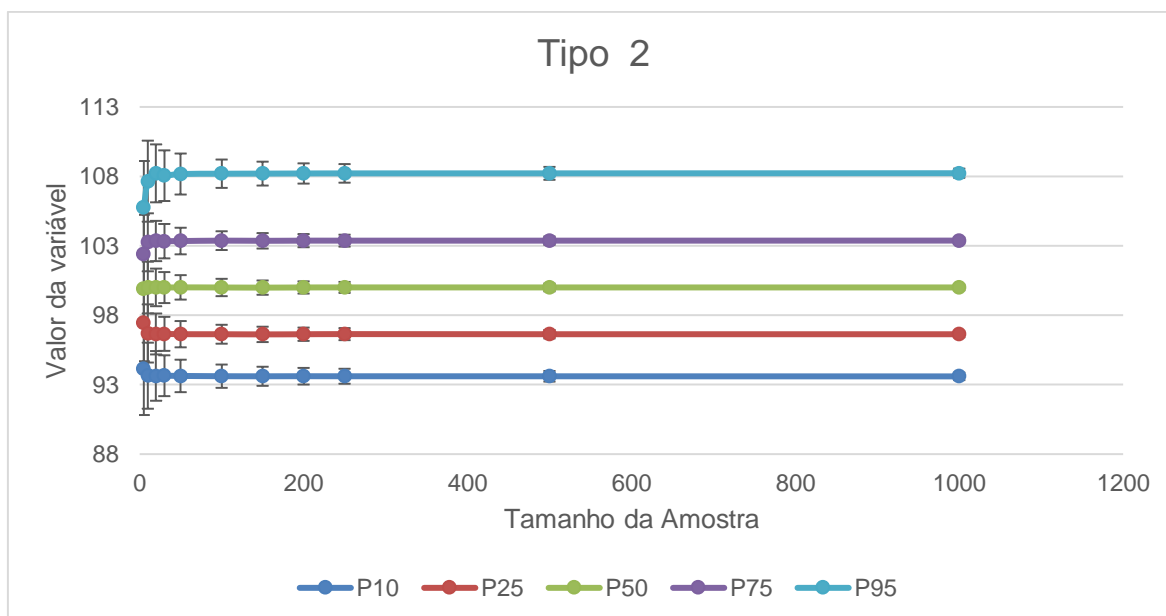


Figura 2- Valores médios e desvios padrões dos quantis P10, P25, P50, P75, P95 obtidos através do estimador tipo 2 para diferentes tamanhos amostrais.

Ao observarmos os resultados obtidos com o estimador tipo 3, evidenciou-se que este produz estimativas menos precisas (Figura 3) que os demais estimadores (Figuras 1 e 2) em amostras de até 50 elementos.

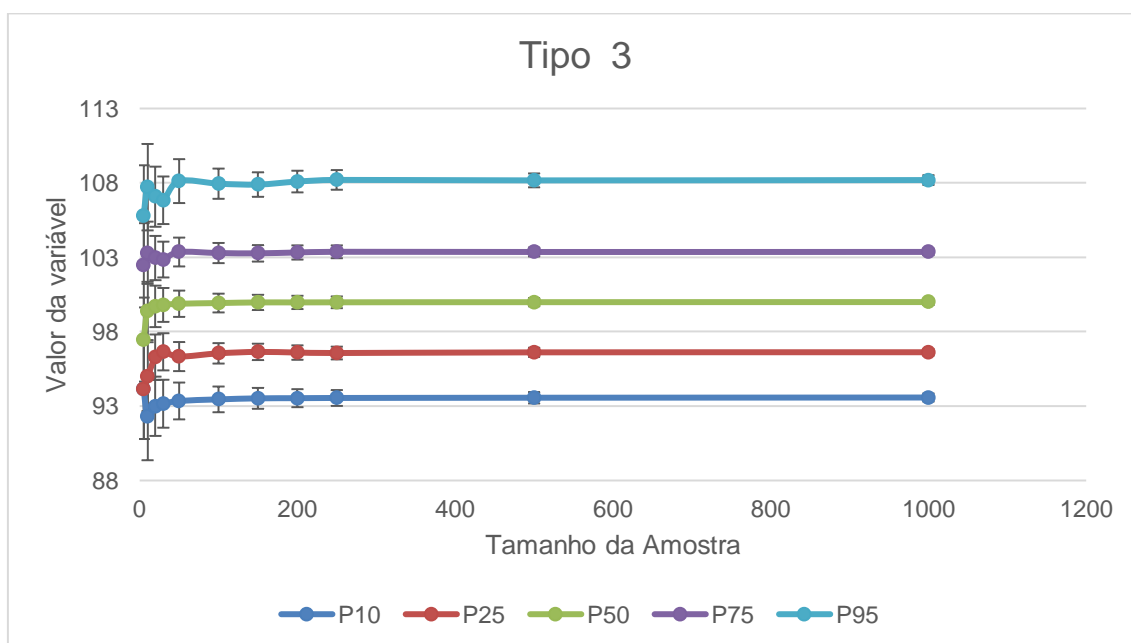


Figura 3- Valores médios e desvios padrões dos quantis P10, P25, P50, P75, P95 obtidos através do estimador tipo 3 para diferentes tamanhos amostrais.

As estimativas dos percentis 10 e 95 apresentam maior variação em relação aos percentis 25, 50 e 75, principalmente, em amostras menores (Figuras 1, 2 e 3).

As simulações de todos os cenários do estudo mostraram a existência de maior variabilidade para estimativas de quantis em amostras menores ($n < 50$), para os três tipos de estimadores avaliados (Figuras 1, 2 e 3). Assim, à medida que o tamanho amostral aumenta as estimativas se aproximam dos valores paramétricos, de acordo com o Teorema Central do Limite (LARSON e FABER (2012)).

4. CONCLUSÕES

O estimador tipo 2 é indicado para obtenção e quantis quando são estudadas variáveis normalmente distribuídas menores que 50 elementos.

Para amostras com tamanhos superiores que 50 elementos os três tipos de estimadores apresentam bom desempenho na obtenção de quantis.

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.

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

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

LARSON, H. J., & FARBER, B. (2012). **Elementary statistics**: Picturing the world.

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

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.