



Universidade de Brasília
IE- Departamento de Estatística
Métodos Estatísticos 2

Juliana Magalhães Rosa

Intervalos de Confiança
Amostra dos Resultados do Saeb

Brasília
2020

Sumário

1 Introdução	2
2 Proporção de alunos do interior	3
3 Proporção de Alunas	5
4 Média das Notas em Língua Portuguesa.....	7
5 Média das Notas em Matemática	9
6 Considerações Finais	11

1 Introdução

A estimação de parâmetros é uma parte importante da inferência estatística. Trata-se do processo de aproximação de medidas populacionais a partir de informações amostrais. Isso pode ser feito por meio de estimadores pontuais, como a média ou a proporção amostral, ou por meio de intervalos de confiança.

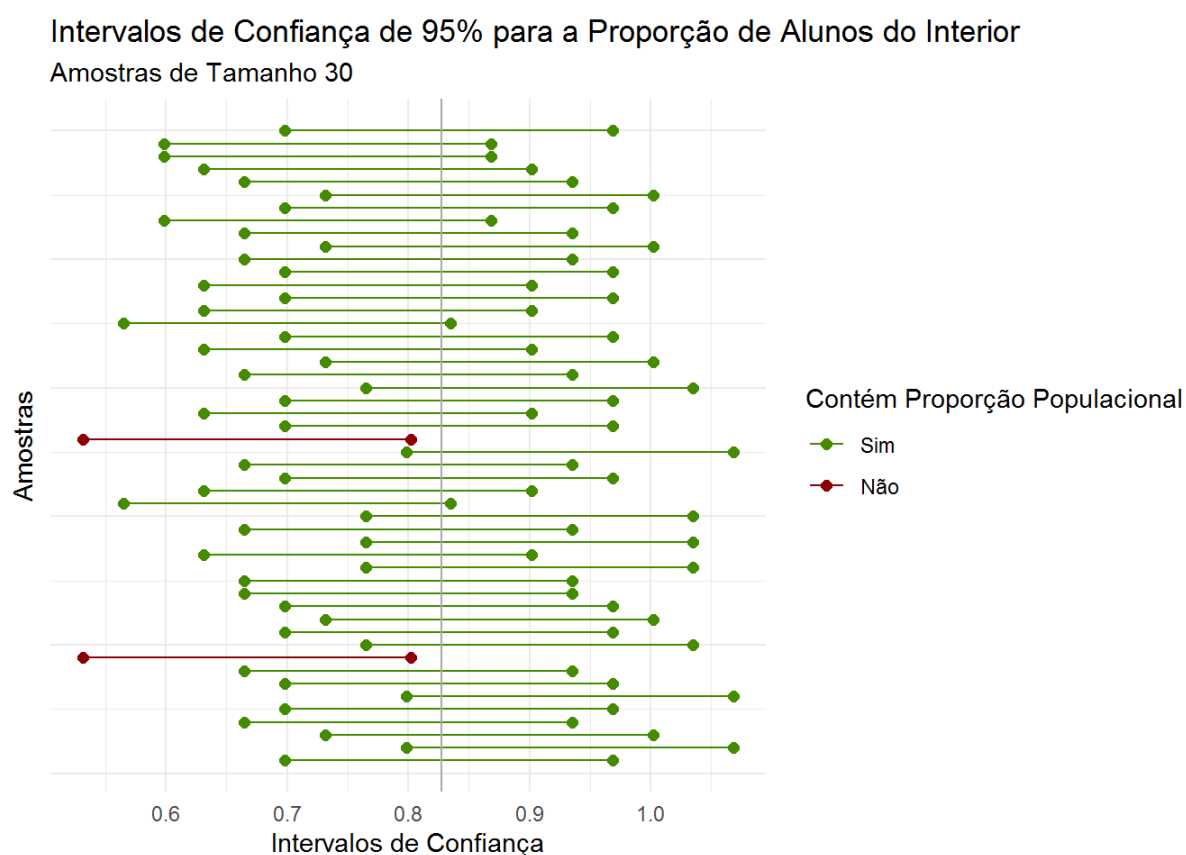
Intervalos de confiança são montados com base em distribuições de probabilidade e a partir de estimadores pontuais. Esse procedimento consiste no estabelecimento de um intervalo de valores (com limite inferior e superior) que estime um parâmetro da população com certo nível de confiança.

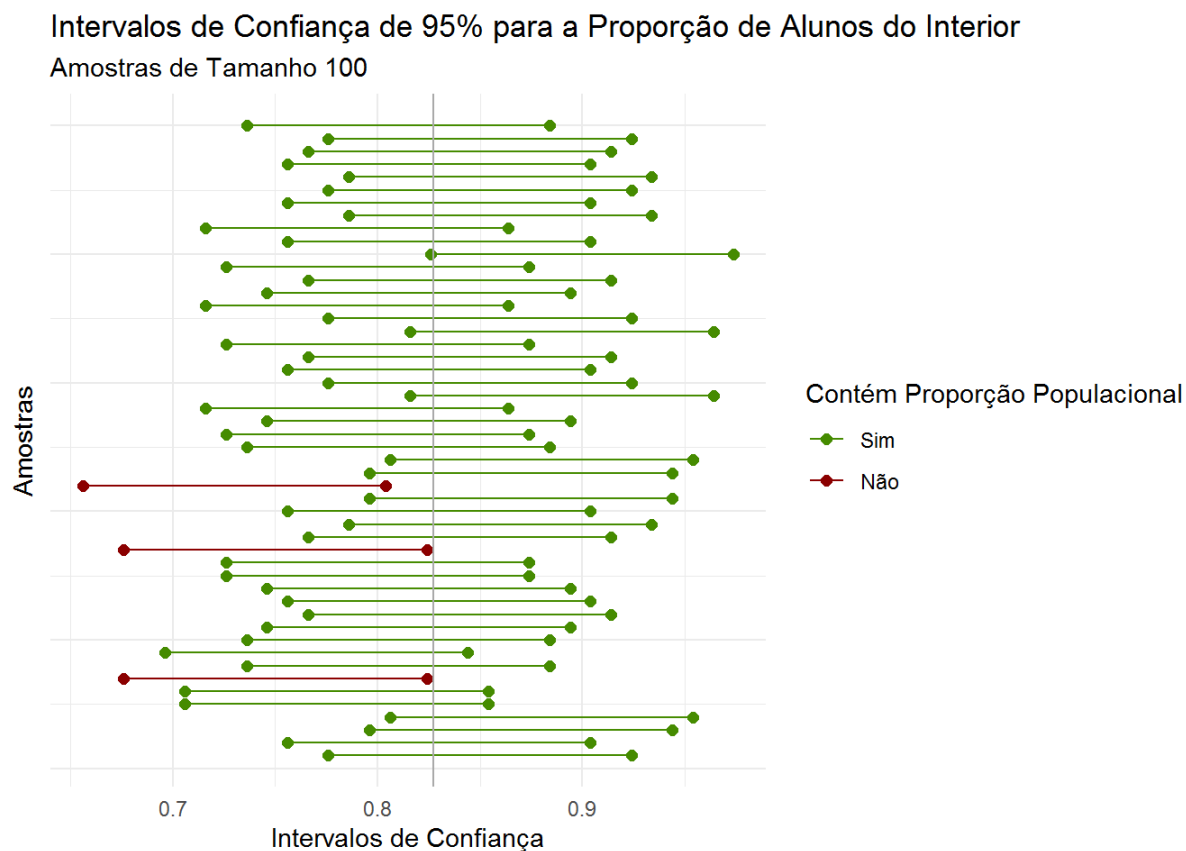
O objetivo deste trabalho é analisar certas variáveis de uma amostra de 2000 alunos do nono ano que participaram do Saeb em 2017, utilizando intervalos de confiança. Esses 2000 estudantes serão a população de estudo e dela serão retiradas 100 amostras para a construção dos intervalos de 95% de confiança com base na distribuição normal.

2 Proporção de Alunos do Interior

A variável categórica AREA indica a área de localização da escola do estudante e assume os valores “capital” e “interior”.

A proporção de alunos do interior será estimada por intervalos de confiança para 50 amostras de tamanho 30 e 50 amostras de tamanho 100.



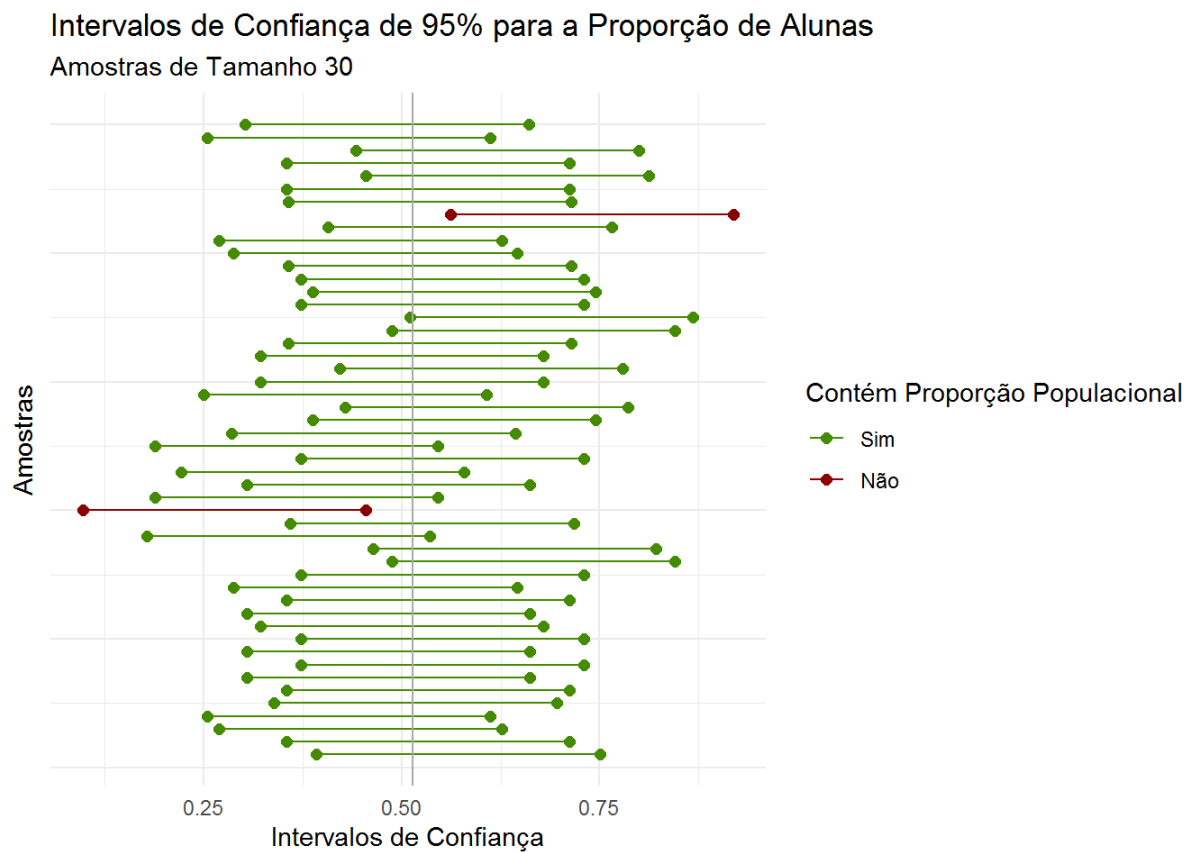


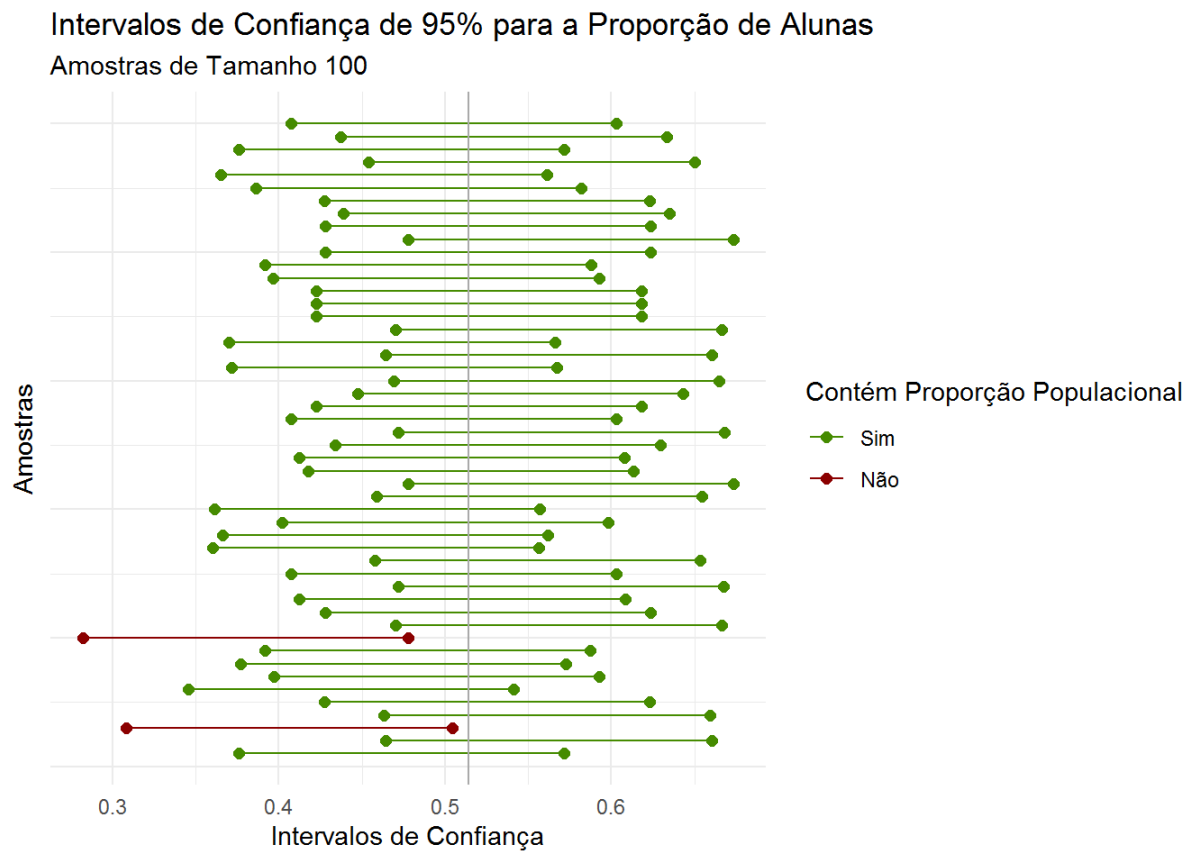
Observando a reta vertical que representa a proporção populacional e as cores que indicam se os intervalos contêm ou não esse parâmetro, percebe-se que são poucos os intervalos que não possuem a proporção real. O nível de confiança de 95% significa que se espera que em torno de 5% dos intervalos não contenham o parâmetro. Considerando que existem 50 amostras, essa porcentagem representa dois ou três intervalos. E, de fato, no primeiro gráfico notam-se dois intervalos vermelhos, e no segundo três. O resto dos estimadores parece, de forma geral, se distribuir em torno do valor populacional de 0,83.

3 Proporção de Alunas

A variável categórica SEXO indica o sexo do estudante e assume os valores “feminino” e “masculino”.

A proporção de estudantes do sexo feminino será estimada por intervalos de confiança para 50 amostras de tamanho 30 e 50 amostras de tamanho 100.



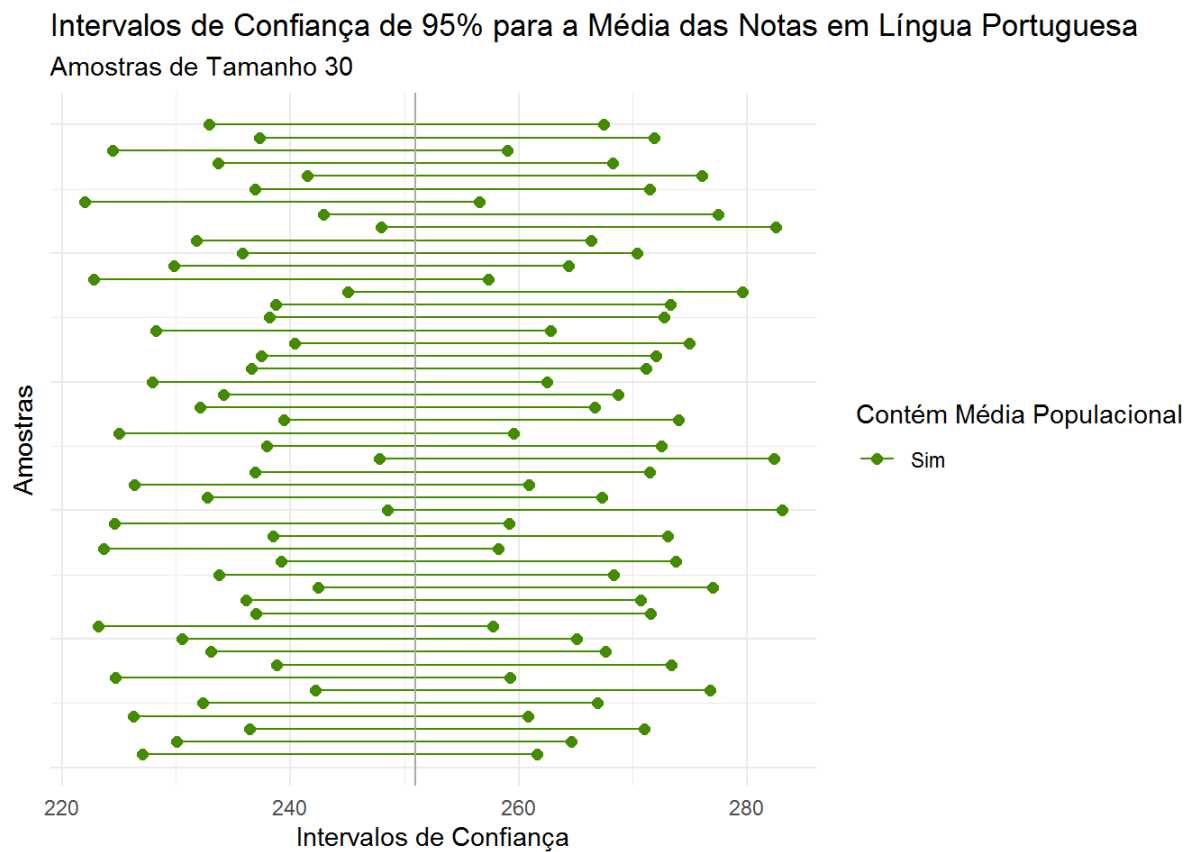


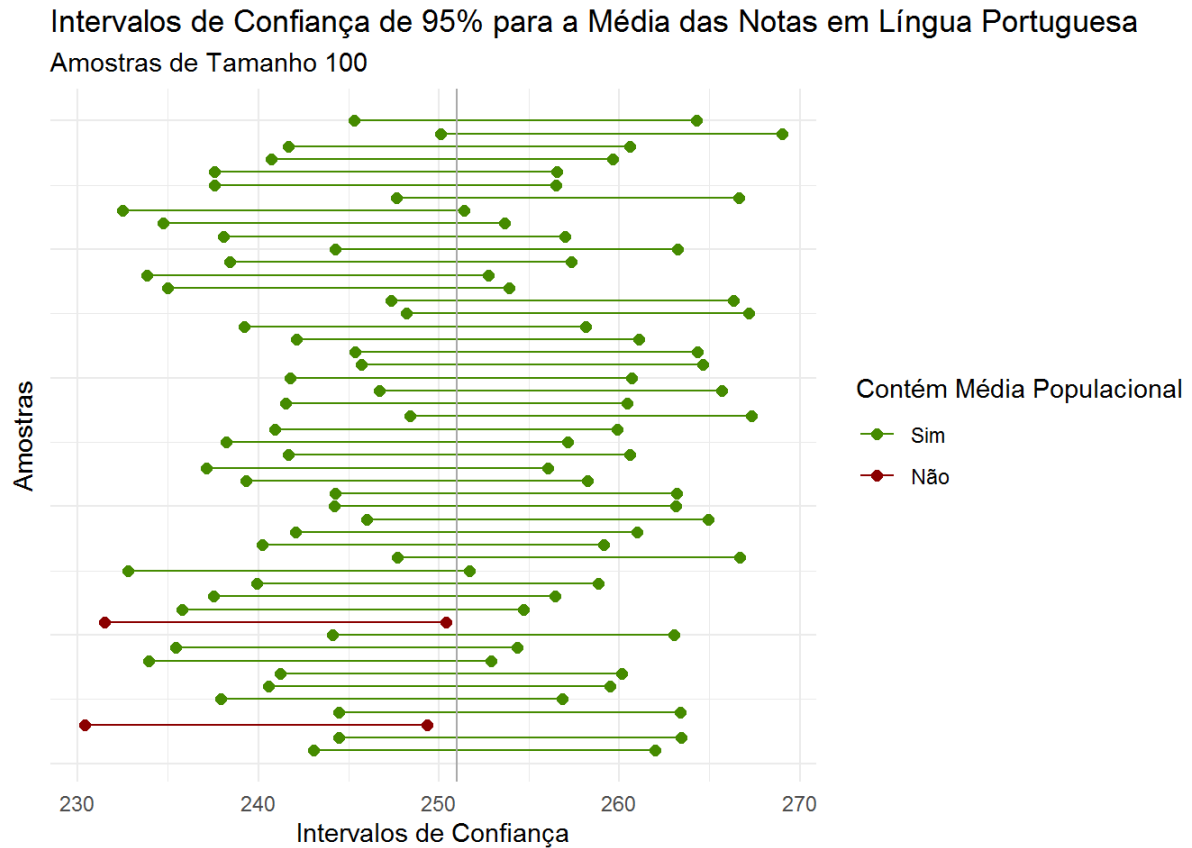
Tanto para as amostras de tamanho 30 como para as de tamanho 100, somente dois dos intervalos construídos não contêm a proporção populacional. Logo, 96% são bons intervalos de estimação, que se distribuem em torno do valor real da proporção de alunas, que é 0,51.

4 Média das Notas em Língua Portuguesa

A variável numérica NOTA_LP indica a proficiência do aluno em Língua Portuguesa, transformada na escala única do Saeb.

A média das notas em língua portuguesa será estimada por intervalos de confiança para 50 amostras de tamanho 30 e 50 amostras de tamanho 100.



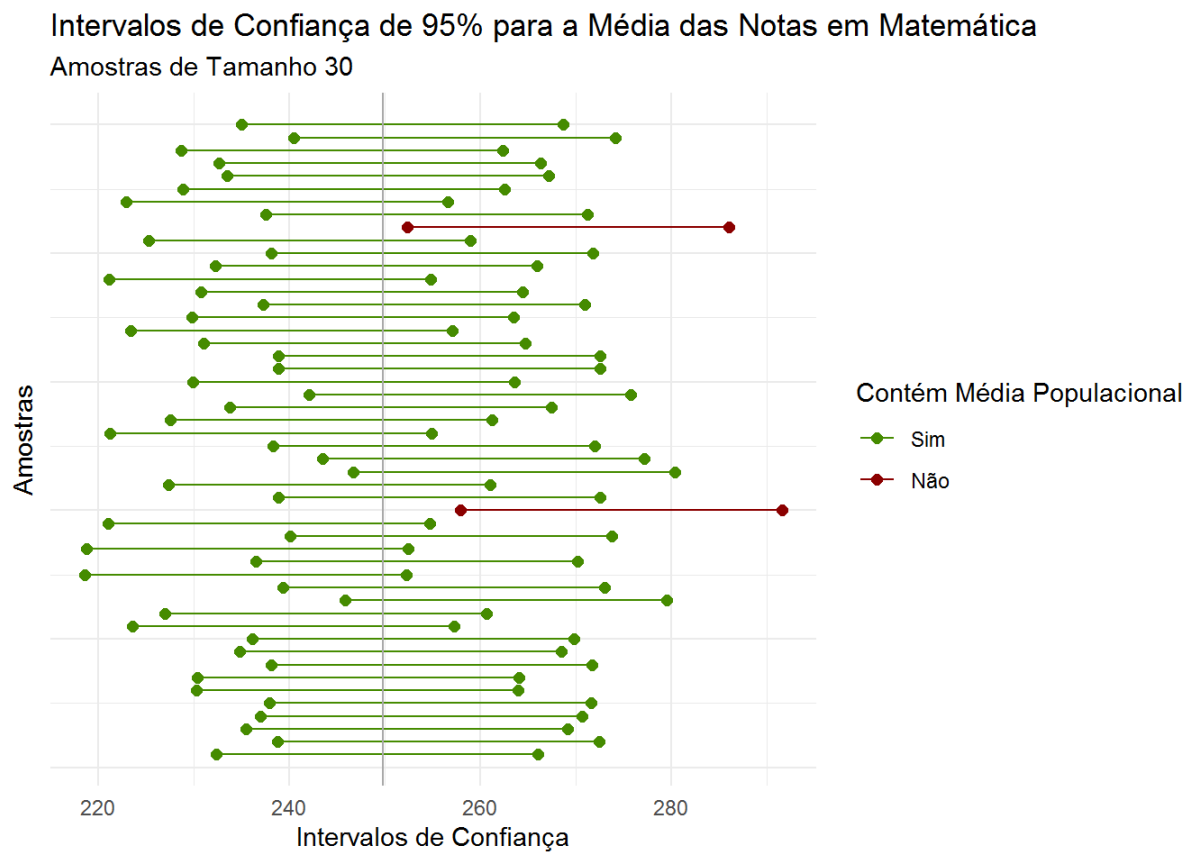


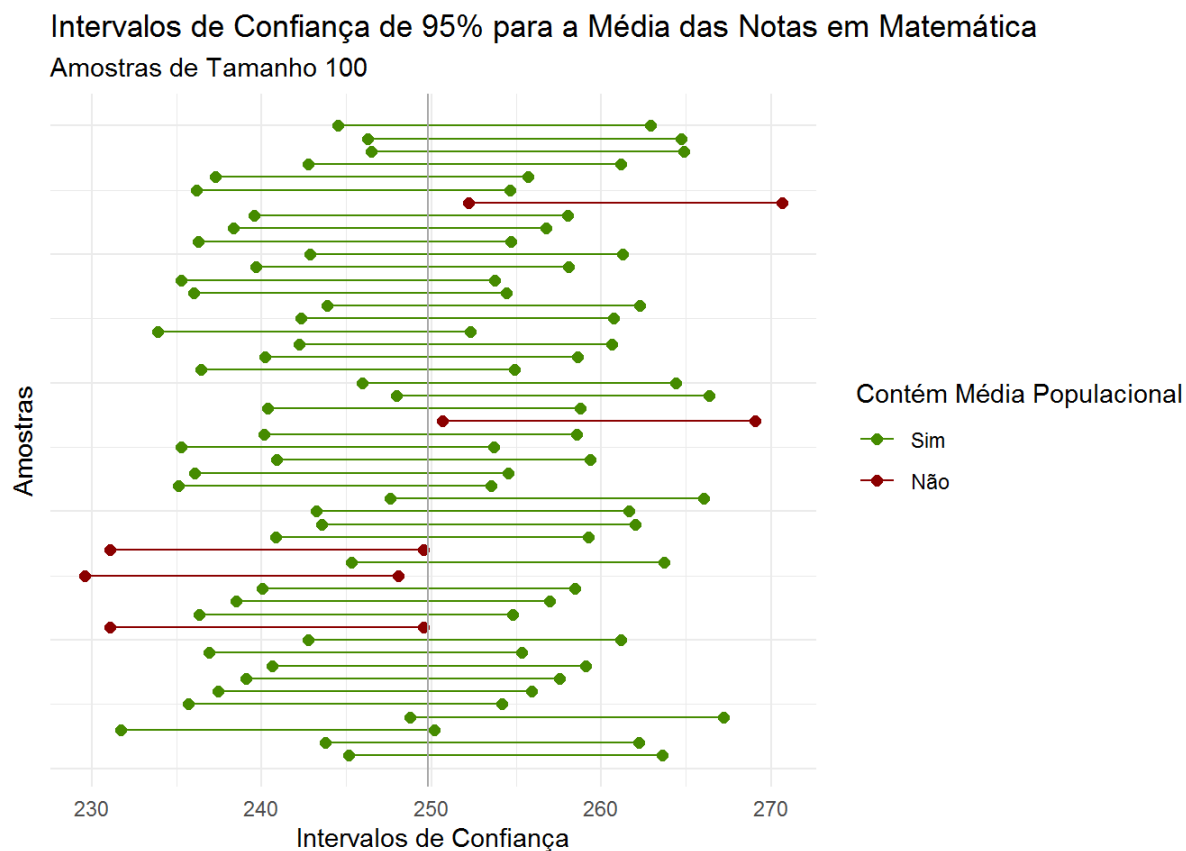
Todos os intervalos montados a partir das amostras de tamanho 30 contêm a média populacional. Para as amostras de tamanho 100, há dois intervalos que não contêm o parâmetro, o que condiz com o esperado para 50 amostras. O valor da média das notas em português na população é de 251,00.

5 Média das Notas em Matemática

A variável numérica NOTA_MT indica a proficiência do aluno em Matemática, transformada na escala única do Saeb.

A média das notas em matemática será estimada por intervalos de confiança para 50 amostras de tamanho 30 e 50 amostras de tamanho 100.





Para as amostras de tamanho 30, existem, novamente, dois intervalos que não contêm a média populacional. Já para as amostras de tamanho 100, cinco dos intervalos formados não incluem o parâmetro, o que é incomum, mas pode acontecer. Nesse segundo caso, em vez dos 5% esperados, há 10% de intervalos que não contêm o valor real da média, que é de 249,80.

6 Considerações Finais

Neste relatório foram apresentados gráficos para os intervalos de confiança de 95% para alguns parâmetros dos dados do Saeb 2017. Essas características populacionais foram a proporção de alunos que estudam no interior, a proporção de estudantes do sexo feminino, a média das notas na prova de língua portuguesa e a média das notas na prova de matemática. Todos os intervalos foram baseados na distribuição normal, e alguns em certos valores da própria população, como o desvio padrão.

Os gráficos mostram os limites inferior e superior de cada um dos 50 intervalos, o valor do parâmetro real e também incluem uma estética de cores para indicar se os intervalos contêm ou não o parâmetro. Tanto as amostras, como os intervalos, e também os próprios gráficos, foram gerados no Software R. O código utilizado pode ser encontrado em anexo.

Em suma, a análise feita ratifica a eficiência de intervalos de confiança como estimadores. Essa forma de estimação permite o controle da precisão e da confiabilidade. Conforme mostram os gráficos, o coeficiente de confiança usado reflete no número de intervalos que de fato contêm o parâmetro de interesse.