

Estudo de Caso 2: Distribuição Amostral e Intervalo de Confiança

Neste estudo de caso, utilizaremos novamente a base de dados IMDB.rds adaptada no último case.

1. Distribuição Amostral.

No último case, você gerou amostras de tamanho K dos dados do arquivo, considerando inicialmente $K=200$. Para realizar essa amostragem, primeiro sorteou K índices, limitados ao tamanho da sua base de dados. Esses índices foram utilizados para selecionar uma amostra aleatória dos dados do arquivo original.

Para essa seção, utilizaremos novamente essa lógica.

Utilize o comando help, referências e fóruns online para gerar/calcular os dados requisitados.

- a. Agora você deve fazer vários sorteios como o do case anterior e, à medida que um sorteio é feito, você deve armazenar os valores de média e variâncias (amostrais) obtidos para a amostra sorteada. Pedem-se $M=450$ sorteios e, portanto, 450 valores de médias e variâncias amostrais extraídos de amostras aleatórias de tamanho K .
- b. A partir dos valores obtidos construa a média das $M=450$ médias amostrais obtidas e compare-as com os valores da população. Construa a média das $M=450$ variâncias amostrais e compare com os valores da variância da população.
- c. Construa o histograma das médias amostrais. A forma do histograma das médias amostrais se assemelha a alguma distribuição de probabilidade conhecida? E das variâncias?
- d. Refaça o exercício considerando diferentes valores para M (de valores baixos como $M=40$, até valores bem elevados, próximos ao tamanho da população como $M=2600$). Discuta o que ocorre com as médias amostrais e com o histograma à medida que o tamanho da amostra cresce.

2. Intervalo de Confiança

- a. Podemos realizar uma estimativa intervalar do retorno médio por meio de um intervalo de confiança. Escreva a expressão analítica do intervalo de confiança para a média populacional em dois casos: I) Desvio padrão populacional conhecido. II) Desvio padrão populacional desconhecido.
- b. Faça o intervalo de confiança para a média de duração de filmes velhos e filmes novos e comente. Considere nível de confiança $\gamma=95\%$ para os dois casos, assumindo desvios padrão populacionais conhecidos e iguais a: $\sigma_{novos} = 25.20$; $\sigma_{velhos} = 27.13$.

Dica: Há funções no R que facilitam a busca do valor tabelado de uma distribuição de probabilidade. Por exemplo, a função `qnorm()`.

- c. Refaça o intervalo de confiança do item b, considerando $\gamma=90\%$ e depois $\gamma=99\%$. Discuta o que ocorre com a amplitude do intervalo ao variarmos o nível de confiança.
- d. Refaça os intervalos de confiança analisados assumindo desvios padrão desconhecidos. Qual a distribuição de probabilidade que você utilizou?
- e. Nos itens b e c qual é o valor da distribuição tabelado que você utilizou em cada caso? Quais foram esses valores no item d? Os valores dos itens b-c em comparação com item d são parecidos. Por que?

3. Entrega no Moodle.

Os cases devem ser enviados no e-disciplinas em um arquivo .pdf com o script do R anexo ao final do próprio PDF, de forma a possibilitar o Ctrl c, Ctrl v do mesmo para efeitos de correção.

Lembre-se de que dissertações e conclusões acerca dos resultados são mais importantes que a própria construção do código em R. Indique todos os resultados da maneira mais expositiva possível.

O prazo de entrega é domingo, 16/04, às 23h59.