

Aplicações do Método Bootstrap

ESTAT0090 – Estatística Computacional

Prof. Dr. Sadraque E. F. Lucena

sadraquelucena@academico.ufs.br

Bootstrap

A grande riqueza em se utilizar o bootstrap é em situações como:

- Você não tem certeza ou uma premissa forte sobre a distribuição normal da população.
- Você gostaria de estimar intervalos de confiança para métricas mais complexas que não possuem distribuição de probabilidade conhecidas ou facilmente definidas, como por exemplo, R^2 de uma regressão ou a mediana.

Exemplo 18.1

Suponha que um fazendeiro gerencia uma plantação de árvores com o objetivo de vendê-las futuramente para a produção de madeira. Sua plantação possui 10.000 árvores, e você está interessado em estimar o diâmetro médio dessas árvores, o que é importante para determinar seu valor de mercado e o momento certo para a colheita.

No entanto, fazer um inventário completo de todas as árvores seria caro e demorado. Então, ele te contrata para usar métodos estatísticos para estimar a média dos diâmetros com base em uma amostra de tamanho 100.

O que se deve fazer nessa situação:

- Estimar o diâmetro médio e obter um intervalo de confiança.
- Vamos obter esse IC pelo método tradicional e usando bootstrap.

Exemplo 18.2

Vamos aproveitar a amostra obtida no Exemplo 18.1 e obter um intervalo de confiança bootstrap para a mediana.

Exemplo 18.3

Considere os dados `datasets::mtcars`. Vamos obter um intervalo de confiança bootstrap de 95% para o R^2 na relação de regressão linear da variável de milhas por galão (mpg) com o peso do carro (wt) e sua cilindrada (disp).

Exemplo 18.4

Agora considere duas áreas distintas em uma fazenda e se deseja comparar a produtividade (diâmetro das árvores) das plantações. Para isso o diâmetro de 30 árvores foi coletado em cada área.

O que deve ser feito nessa situação:

- Hipóteses:
 - H_0 : as produtividades das duas áreas são iguais
 - H_1 : as produtividades das duas áreas são diferentes
- Fazer um teste t.

Teste de comparação de médias bootstrap

- Calcule as médias amostrais \bar{X} e \bar{Y}
- Réplicas bootstrap: Repita B vezes
 - Obtenha amostras artificiais x_1^*, \dots, x_n^* e y_1^*, \dots, y_n^*
 - Calcule a diferença das médias artificiais $\bar{x}^* - \bar{y}^*$
- Calcule o p-valor:
 - Se $H_1 : \mu_x < \mu_y \Rightarrow \text{p-valor} = \frac{1 + \#\{\bar{x}^* - \bar{y}^* < 0\}}{1 + B}$
 - Se $H_1 : \mu_x > \mu_y \Rightarrow \text{p-valor} = \frac{1 + \#\{\bar{x}^* - \bar{y}^* > 0\}}{1 + B}$
 - Se $H_1 : \mu_x \neq \mu_y \Rightarrow \text{p-valor} = \frac{1 + 2 \min[\#\{\bar{x}^* - \bar{y}^* < 0\}, \#\{\bar{x}^* - \bar{y}^* > 0\}]}{1 + B}$

Exemplo 18.5

Em 1882, Simon Newcomb realizou um experimento de medir a velocidade da luz. Ele mediu o tempo que a luz levou para ir de um ponto a outro. Os dados fornecidos confirmam a teoria de que o tempo que a luz leva é de 33,02 milionésimos de segundos? Considere um nível de significância de 5%.

Fim