

## Bootstrap

Estude a função **regress**.

Digite:        `x = [300;800;400;900;1000;100;700;500;800;300;1300;400;600;100;100];`  
                 `y = [245;730;290;925;1120;120;635;590;830;384;1363;390;635;190;101];`  
                 `b = regress(y,x)`

Agora você fará um teste de significância de **b**, o coeficiente da variável independente  $x$ , obtido da análise de regressão de  $y$  por  $x$ . Sua estatística teste será  $b - 0$ , a qual se reduz a **b**.

Digite:        `iterations = 5000;`  
                 `dst = zeros(1,length(iterations));`

Você acaba de definir o número de estatísticas bootstrap que irá calcular para constituir sua  $H_0$ , e criou um vetor para receber estes valores a cada iteração.

Estude as funções **ceil** e **round**.

Estude agora a função **rand**.

Crie agora o loop que calculará os 5000 **b** bootstrap que formarão a  $H_0$ :

Você deve reamostrar os valores de  $x$  e proceder uma nova regressão para obter o **b** bootstrap.

**Dica:** a cada iteração, crie um vetor com os índices (posições) do vetor  $x$  reamostrados e, quando for proceder a nova regressão, indexe seu vetor  $x$  com esse vetor de índices.

Por que usar **ceil** e não **round** ? Faz diferença ?

Achando os valores críticos do intervalo de confiança (IC) de  $H_0$ :

Defina um nível de significância (*alpha*).

Calcule agora a posição, em  $H_0$ , do limite inferior de seu IC. O valor será igual ao número de estatísticas que compõe sua  $H_0$  vezes seu  $\alpha$  dividido por 2. O limite superior será dado pela posição equivalente e simétrica à do limite inferior.

Estude a função **sort**.

Agora ordene em ordem crescente os valores da distribuição de  $H_0$ .

O limite inferior de seu IC está na posição que você calculou acima, na distribuição de  $H_0$  que você ordenou. Já o valor do limite superior está na posição oposta ao limite inferior (número total de posições – posição do limite inferior).

Para visualizar  $H_0$ , digite:

```
hist(dstH0)
title('dstH0 frequency distribution')
ylabel('frequency')
xlabel('b')
clear
```

$b$  é significativo ?

Estude esse algoritmo em detalhes. Compare com o algoritmo final do protocolo 1. Entenda o que foi feito em ambos os casos.

```
a = [8 4 1;0 3 5;0 0 10]
k=length(a);
b(1:k)={repmat([0],k,1)}
for j = 1:k
    for i = 1:k
        b{j}{i}= ones(1,a(i,j))*a(i,j);
    end
end
c=[b{1}{:}]
d=[b{2}{:}]
e=[b{3}{:}]
f=[b{1}{:} b{2}{:} b{3}{:}]
```

## Roteiro para a construção de um algoritmo para o cálculo do intervalo de confiança de $\lambda$

Método: bootstrap das histórias de vida

1. A partir de uma tabela de histórias de vida construa, para cada coluna da tabela, 1 vetor contendo as diferentes histórias de vida que os indivíduos da classe à qual a coluna diz respeito passaram no período estudado. Os vetores devem ter tamanhos equivalentes ao número de indivíduos na classe, e as diferentes histórias de vida devem estar representadas nas mesmas proporções em que aparecem nas tabelas de histórias de vida.
2. Gere a matriz de transição, calcule seu  $\lambda$  dominante e reserve-o.
3. Crie agora a sua  $H_0$ :
  - a. Gere versões bootstrap para os vetores de histórias de vida calculados em 1.
  - b. Com base nos vetores de histórias de vida bootstrap, calcule uma nova matriz de transição e seu respectivo  $\lambda$  dominante. Reserve este  $\lambda$ .
  - c. Repita esse procedimento por pelo menos 2000 vezes
  - d. Ordene a sua  $H_0$  pela ordem crescente de valores
4. Defina um nível de significância
5. Ache os valores dos limites inferior e superior do intervalo de confiança de sua  $H_0$  (método do percentil):
  - a. Determine em sua  $H_0$  as posições dos limites inferior e superior de seu intervalo de confiança. Nelas estarão os valores desejados.