

# INTERVALO DE CONFIANÇA “BOOTSTRAP” COMO FERRAMENTA PARA CLASSIFICAR RAÇAS DO NEMATÓIDE DE CISTO DA SOJA<sup>1</sup>

JOSÉ ERIVALDO PEREIRA<sup>2</sup>, JOÃO FLÁVIO VELOSO SILVA<sup>2</sup>,  
WALDIR PEREIRA DIAS<sup>2</sup> e GERALDO DA SILVA E SOUZA<sup>3</sup>

RESUMO - Estimativas “bootstrap” da média aritmética dos genótipos de soja ‘Pickett’, ‘Peking’, PI88788 e PI90763 e os intervalos de confiança obtidos pela teoria normal e através da distribuição “bootstrap” deste estimador, como o percentil “bootstrap” e o BCa, correção para o viés e aceleração, do parâmetro de diferenciação da cultivar padrão de suscetibilidade Lee são utilizados para classificar raças do nematóide de cisto da soja. Os intervalos de confiança obtidos a partir da distribuição “bootstrap” apresentaram menor amplitude e foram muito similares, dessa forma, o limite inferior do intervalo de confiança percentil “bootstrap” foi tomado como nível de referência nas distribuições “bootstrap” do estimador da média aritmética dos genótipos diferenciadores, permitindo estimar a probabilidade empírica de uma reação positiva ou negativa, e, conseqüentemente, identificar a raça mais provável sob determinado teste.

Termos para indexação: *Heterodera glycines*, genótipos diferenciadores, índice, variância.

## BOOTSTRAP CONFIDENCE INTERVALS AS TOOL TO CLASSIFY RACES OF SOYBEAN'S CYST NEMATODE

ABSTRACT - Bootstrap estimates of arithmetic mean of Pickett, Peking, PI88788 and PI90763 soybean genotypes and the confidence intervals obtained through normal theory and bootstrap distribution of this estimator, like the percentile bootstrap and the BCa (bias-corrected and accelerated) of the differentiation parameter of the standard cultivar Lee for susceptibility, are utilized to classify races of soybean cyst nematode. The confidence intervals obtained from bootstrap distribution showed smaller amplitude and were very similar. Therefore, the lower limit of the bootstrap percentile confidence interval was taken as reference level in the bootstrap distribution of the arithmetic mean estimator of the differential genotypes allowing to estimate the empirical probability of a positive or negative reaction and, consequently, to identify the most probable race under specific test.

Index terms: *Heterodera glycines*, differential genotypes, index, variance.

## INTRODUÇÃO

Populações do nematóide de cisto da soja (NCS), *Heterodera glycines* Ichinohe, podem ser classificadas em 16 raças (Riggs & Schmitt, 1988). Essa classificação é baseada no número médio de fêmeas do

nematóide em quatro genótipos de soja (*Glycine max* (L.) Merr.), quais sejam: ‘Peking’, ‘Pickett’, PI88788 e PI90763, quando comparados com a contagem de fêmeas observada na cultivar padrão de suscetibilidade Lee.

Um genótipo diferenciador tem reação positiva se a porcentagem do número médio de cistos produzidos em relação ao número médio obtido na cultivar padrão de suscetibilidade Lee for igual ou maior que dez (Golden et al., 1970, citados por Riggs & Schmitt, 1988).

A experiência em pesquisa com a variável aleatória número de cistos por volume de solo tem evidenciado a ocorrência de grande aleatoriedade associa-

<sup>1</sup> Aceito para publicação em 14 de abril de 1999.

<sup>2</sup> Eng. Agrôn., M.Sc., Embrapa-Centro Nacional de Pesquisa de Soja (CNPSo), Caixa Postal 231, CEP 86001-970 Londrina, PR. E-mail: erivaldo@cnpso.embrapa.br

<sup>3</sup> Matemático, Ph.D., Embrapa-Secretaria de Administração Estratégica (SEA), SAIN Parque Rural-Final Av. W/3 Norte, Caixa Postal 040315, CEP 70770-901 Brasília, DF. E-mail: geraldo@sede.embrapa.br

da à superdispersão (McSorley, 1982; Ferris, 1984; McCullagh & Nelder, 1989). Riggs et al. (1988) relataram que a mesma população de NCS foi classificada em diferentes raças. Young (1989) observou que o viés na identificação da raça é maior quando são utilizadas menos de quatro repetições e frisou a necessidade de se considerar a variância do índice médio utilizado na diferenciação das raças. O objetivo deste estudo foi estimar e utilizar como critério na determinação de raças do NCS os intervalos de confiança do parâmetro de diferenciação da cultivar-padrão de suscetibilidade Lee obtidos pela teoria normal e via distribuição "bootstrap" (Efron & Tibshirani, 1993) para a média aritmética, visando melhorar a precisão na identificação das raças de NCS.

## MATERIAL E MÉTODOS

A metodologia utilizada consiste em obter a distribuição "bootstrap" para o estimador da média aritmética dos quatro genótipos diferenciadores e do estimador do parâmetro de diferenciação da cultivar padrão de suscetibilidade Lee. O método "bootstrap" foi utilizado na versão não-paramétrica, ou seja, assumindo a distribuição uniforme na geração das amostras "bootstrap". Foram estimados os intervalos de confiança através da teoria normal, e via distribuição "bootstrap" como o percentil "bootstrap", e o BCa.

Dada a distribuição "bootstrap" de um estimador, o intervalo de confiança percentil com coeficiente de confiança igual a  $1-2\alpha$  é definido pelos percentis  $\alpha$  e  $1-2\alpha$  desta distribuição, logo, os valores  $\hat{\theta}^{*(\alpha)}$  e  $\hat{\theta}^{*(1-\alpha)}$  representam os limites do intervalo considerado.

O intervalo BCa é também obtido da distribuição "bootstrap", mas depende das estimativas de  $\hat{a}$  e  $\hat{Z}_0$ , denominados constante de aceleração e correção do viés, respectivamente, estas são dadas por,

$$\hat{a} = \frac{\sum_{i=1}^n (\hat{\theta}_{(.)} - \hat{\theta}_{(i)})^3}{6 \left\{ \sum_{i=1}^n (\hat{\theta}_{(.)} - \hat{\theta}_{(i)})^2 \right\}^{3/2}},$$

onde

$$\hat{\theta}_{(.)} = \sum_{i=1}^n \hat{\theta}_{(i)} / n;$$

$\hat{\theta}_{(.)}$ : é a estimativa Jackknife de  $\theta$ ;

$\hat{\theta}_{(i)}$ : é uma estimativa de  $\theta$  obtida a partir da amostra original sem a  $i$ -ésima observação;  
 $n$ : é o tamanho da amostra original; e

$$\hat{Z}_0 = \Phi^{-1} \left( \frac{\# \{ \hat{\theta}^*(b) < \hat{\theta} \}}{B} \right),$$

onde

$\hat{\theta}$ : estimativa de  $\theta$  da amostra original;

$\hat{\theta}_{(b)}^*$ : é uma estimativa de  $\theta$  para a  $b$ -ésima repetição "bootstrap";

$B$ : é o número de repetições "bootstrap";

$\Phi_{(.)}^{-1}$ : indicando a função inversa da distribuição normal padrão acumulada.

Dessa forma, o intervalo BCa com cobertura igual a  $1-2\alpha$ , é dado por:

$$\text{BCa} = (\hat{\theta}^{*(\alpha_1)}, \hat{\theta}^{*(\alpha_2)}),$$

onde

$$\alpha_1 = \Phi \left( \hat{Z}_0 + \frac{\hat{Z}_0 + Z^{(\alpha)}}{1 - \hat{a}(\hat{Z}_0 + Z^{(\alpha)})} \right);$$

$$\alpha_2 = \Phi \left( \hat{Z}_0 + \frac{\hat{Z}_0 + Z^{(1-\alpha)}}{1 - \hat{a}(\hat{Z}_0 + Z^{(1-\alpha)})} \right);$$

onde

$Z^{(\alpha)}$ : é o  $\alpha$ -ésimo ponto percentil da distribuição normal padronizada;

$\Phi(\cdot)$ : indicando a função de distribuição normal padrão acumulada.

O intervalo de confiança para a média através da

teoria normal foi obtido por,  $\hat{y} \pm t_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}$ ,

onde

$\hat{y}$ : é uma estimativa da média amostral;

$t_{\frac{\alpha}{2}}$ : é o valor tabelado da distribuição t de Student com  $n-1$  graus de liberdade;

$S$ : é uma estimativa do desvio-padrão amostral;

$n$ : é o tamanho da amostra.

Os limites do intervalo de confiança da média aritmética foram multiplicados por 0,10 para se obter o intervalo de confiança do parâmetro de diferenciação da cultivar Lee.

Para caracterizar uma reação positiva de um genótipo diferenciador, assume-se que a média aritmética "bootstrap" deve ser maior ou igual ao limite inferior do intervalo de confiança percentil "bootstrap" para o parâmetro de diferenciação da cultivar padrão de suscetibilidade Lee, ou seja, dez por cento do limite inferior do intervalo de confiança percentil "bootstrap" obtido pela distribuição "bootstrap" do estimador da média aritmética. A probabilidade empírica de um genótipo diferenciador ter uma reação positiva ou negativa pode ser obtida tomando-se esta estimativa do limite inferior do intervalo de confiança percentil "bootstrap" como referência nas distribuições "bootstrap" das médias aritméticas dos genótipos diferenciadores, obtendo-se, assim, a probabilidade da raça de nematóide mais provável para o teste considerado.

## RESULTADOS E DISCUSSÃO

A Tabela 1 apresenta os intervalos de confiança estimados para o parâmetro de diferenciação da cultivar padrão de suscetibilidade Lee nos dois exemplos apresentados. Observa-se que os intervalos obtidos pela distribuição "bootstrap" têm menor amplitude, dessa forma, e por ser mais facilmente obtido, apenas o intervalo percentil "bootstrap" foi considerado na descrição dos dois exemplos dados a seguir:

Os dados observados para o teste realizado em Água Clara, MS, em 17/1/96, onde foram feitas inoculações de 10.872 ovos/vaso, servem de exemplo. As contagens observadas foram:

PI88788  $\Rightarrow$  0, 3, 10, 46, 25;  $n = 5$ ;  
 PI90763  $\Rightarrow$  50, 18, 22, 56, 21, 10;  $n = 6$ ;  
 Peking  $\Rightarrow$  101, 79, 34, 56, 69, 40;  $n = 6$ ;  
 Pickett  $\Rightarrow$  250, 197, 446, 174, 116, 122;  $n = 6$ ;  
 Lee  $\Rightarrow$  285, 186, 436, 269, 322, 467;  $n = 6$ .

Com esses resultados, e utilizando o método de Riggs & Schmitt (1988), observa-se que a estimativa do parâmetro de diferenciação da cultivar padrão de suscetibilidade Lee é igual a 32,75, e as médias dos genótipos diferenciadores 'Pickett', 'Peking', PI88788 e PI90763 são 217,5, 63,2, 20,6 e 29,5, respectivamente, o que sugere a seqüência (+), (+), (-), (-) relacionada ao padrão de suscetibilidade, ou seja, a raça 9 é identificada.

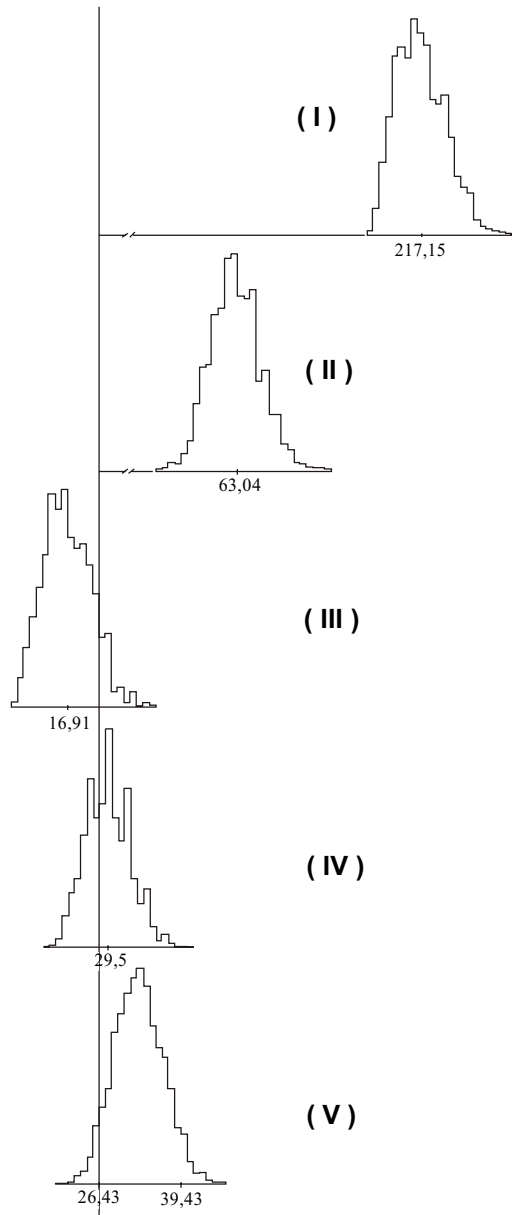
Em relação ao mesmo conjunto de dados, a aplicação do método "bootstrap" com  $B = 5.000$  amostras "bootstrap", as médias aritméticas "bootstrap" estimadas para as diferenciadoras 'Pickett', 'Peking', PI88788 e PI90763 são 217,15, 63,04, 16,91 e 29,50, respectivamente. Além disso, o intervalo de confiança percentil "bootstrap" com coeficiente de confiança igual a 0,90 para o parâmetro de diferenciação da Lee é igual a 26,43 - 39,43; isto sugere a seqüência (+), (+), (-), (+), ou seja, a raça 14 é identificada. Tomando-se agora o limite inferior do intervalo de confiança 26,43 como referência nas distribuições "bootstrap" dos genótipos diferenciadores, observa-se que a probabilidade empírica associada à identificação da raça 14 é de 58,08%. Uma representação gráfica deste procedimento encontra-se na Fig. 1.

A Tabela 2 apresenta a probabilidade empírica de reação positiva ou negativa obtida das distribuições "bootstrap", pode-se observar que as seqüências (+), (+), (-), (-) e (+), (+), (+), identificando as raças 9 e 4, apresentam-se associadas a uma probabilidade de 31,63% e 6,65%, respectivamente.

Nesse exemplo, as amostras originais em cada genótipo diferenciador, e inclusive a da cultivar padrão Lee, apresentaram bastante variação. Essa variabilidade resulta de fontes genéticas e ambientais, e depende do erro de amostragem. Assim, a identificação da raça de NCS associada a uma medida de

**TABELA 1. Intervalos de confiança com coeficiente de confiança igual a 0,90 do parâmetro de diferenciação da cultivar Lee nos dois locais do teste de identificação de raças do nematóide de cisto da soja.**

Intervalos de confiança	Água Clara	Tangará da Serra
Teoria normal	24,00 – 41,49	19,49 – 25,58
Percentil bootstrap	26,43 – 39,43	20,14 – 24,90
BCa	26,58 – 39,58	20,18 – 24,95



**FIG. 1.** Distribuição “bootstrap” do estimador da média aritmética dos genótipos diferenciadores Pickett (I), Peking (II), PI88788 (III), PI90763 (IV) e do estimador do parâmetro de diferenciação da cultivar padrão de suscetibilidade Lee (V). O limite inferior do intervalo de confiança percentil “bootstrap” cujo valor é 26,43 é tomado como referência nas distribuições dos genótipos diferenciadores.

probabilidade constitui um fator importante para o monitoramento de uma população. Entretanto, mesmo quando isso não ocorre, é importante ressaltar que uma estimativa do parâmetro de diferenciação muito próxima das estimativas das médias dos genótipos diferenciadores pode resultar numa identificação viesada da raça, essas dificuldades são corrigidas com o uso do intervalo de confiança. A análise de um segundo exemplo ilustra essa situação. Considerando as contagens observadas no teste de Tangará da Serra, MT, em 15/8/97, onde foram feitas inoculações de 4.000 ovos/vaso, os seguintes resultados foram obtidos:

PI88788  $\Rightarrow$  17, 34, 26, 13, 31, 22, 12;  $n = 7$ ;

PI90763  $\Rightarrow$  0, 0, 0, 0, 0, 0, 0;  $n = 7$ ;

Peking  $\Rightarrow$  0, 0, 0, 0, 0, 0, 0;  $n = 7$ ;

Pickett  $\Rightarrow$  0, 0, 0, 0, 0, 0, 0;  $n = 7$ ;

Lee  $\Rightarrow$  289, 241, 219, 164, 258, 194, 213;  $n = 7$ .

Agora a estimativa do parâmetro de diferenciação da cultivar padrão de suscetibilidade Lee é igual a 22,54 e as médias dos genótipos diferenciadores ‘Pickett’, ‘Peking’, PI88788 e PI90763 são 0, 0, 22,14 e 0, respectivamente; dessa forma, a raça identificada é a 3, embora a média do genótipo PI88788 seja menor mas muito próxima da estimativa pontual do parâmetro de diferenciação.

Ainda nesse exemplo, o método “bootstrap” com  $B = 5.000$  amostras “bootstrap”, as médias aritméticas “bootstrap” estimadas para os genótipos diferenciadores ‘Pickett’, ‘Peking’, PI88788 e PI90763 são 0, 0, 22,09 e 0, respectivamente. O intervalo de confiança percentil “bootstrap” com coeficiente de confiança igual a 0,90 para o parâmetro de diferenciação da cultivar Lee é 20,14 - 24,90, o que indica que as seqüências (-), (-), (+), (-), e (-), (-), (-), (-) são as mais prováveis, ou seja, tomando-se o limite inferior do intervalo de confiança percentil “bootstrap” para a cultivar Lee, cujo valor é 20,14, como nível de referência nas distribuições “bootstrap” dos genótipos diferenciadores, tem-se que a probabilidade do teste está identificando a raça 1 é de 74,26%, e a raça 3 é de 25,74%. Esses resultados são apresentados na Tabela 3.

**TABELA 2. Probabilidade empírica de reação positiva ou negativa e raças mais prováveis no teste de Água Clara, MS.**

Genótipo diferenciador	P(<26,43)	P(≥26,43)	Seqüência mais provável	Segunda seqüência mais provável	Terceira seqüência mais provável
Pickett	0,00	1,00	(+)	(+)	(+)
Peking	0,00	1,00	(+)	(+)	(+)
PI88788	0,8972	0,1028	(-)	(-)	(+)
PI90763	0,3526	0,6474	(+)	(-)	(+)
Raça identificada			14	9	4
Probabilidade da raça identificada			58,08%	31,63%	6,65%

**TABELA 3. Probabilidade empírica de reação positiva ou negativa e raças mais prováveis no teste de Tangará da Serra, MT.**

Genótipo diferenciador	P(<20,14)	P(≥20,14)	Seqüência mais provável	Segunda seqüência mais provável
Pickett	1,00	0,00	(-)	(-)
Peking	1,00	0,00	(-)	(-)
PI88788	0,2574	0,7426	(+)	(-)
PI90763	1,0	0,0	(-)	(-)
Raça identificada			1	3
Probabilidade da raça identificada			74,26%	25,74%

## CONCLUSÃO

1. O intervalo de confiança percentil “bootstrap” é uma estimativa intervalar precisa do parâmetro de diferenciação da cultivar Lee, podendo o limite inferior deste ser utilizado como referência na distribuição “bootstrap” do estimador da média aritmética dos genótipos diferenciadores, obtendo-se, assim, a raça mais provável do NCS para o teste considerado.

## REFERÊNCIAS

- EFRON, B.; TIBSHIRANI, R. **An introduction to the bootstrap**. New York : Chapman & Hall, 1993. 436p.
- FERRIS, H. Nematode damage functions: the problems of experimental and sampling error. **Journal of Nematology**, St. Paul, MN, v.16, n.1, p.1-9, 1984.
- McCULLAGH, P.; NELDER, J.A. **Generalized linear models**. 2.ed. London : Chapman and Hall, 1989. 511p.
- McSORLEY, R. Simulated sampling strategies for nematodes distributed according to a negative binomial model. **Journal of Nematology**, St. Paul, MN, v.14, n.4, p.517-529, 1982.
- RIGGS, R.D.; SCHMITT, D.P. Complete characterization of the race scheme for *Heterodera glycines*. **Journal of Nematology**, St. Paul, MN, v.20, n.3, p.392-395, 1988.
- RIGGS, R.D.; SCHMITT, D.P.; NOEL, G.R. Variability in race tests with *Heterodera glycines*. **Journal of Nematology**, St. Paul, MN, v.20, n.4, p.565-572, 1988.
- YOUNG, L.D. Use of statistics in race determination tests. **Journal of Nematology**, St. Paul, MN, v.21, n.4, p.544-546, 1989.