

Avaliação do desempenho do algoritmo de reamostragem *bootstrap* na verificação da estimação do tamanho ótimo da parcela

Márcio Cláudio Mercês Brito¹

Guido Gustavo Humada-González²

Augusto Ramalho de Moraes³

Juracy Mendes Moreira⁴

1 Introdução

A redução do erro experimental consiste em planejar o experimento visando o controle da variação que ocorre na área experimental. O delineamento adequado depende da cultura, número de tratamentos, tamanho das unidades experimentais e das condições ambientais de cada experimento. A adoção de um tamanho ótimo de parcela é uma das maneiras de se reduzir o erro experimental e, conseqüentemente, maximizar as informações obtidas em um experimento. Para a estimação do tamanho de parcela, vários métodos são empregados e a maioria deles se baseia na utilização de ensaios em branco, ensaios de uniformidade, nos quais, em toda a área experimental é realizado um único tratamento, utilizando-se as práticas de cultivo adequadas para aquele tratamento.

É uma prática comum, na estimação do tamanho ótimo de parcela, o cálculo de coeficientes de variação (CV) de cada tamanho (x) de parcela obtendo-se um conjunto de pontos do tipo (x_i, CV_i) , que são relacionados num sistema de eixos coordenados. Uma curva, então, é traçada utilizando dessas coordenadas resultantes, e o tamanho ótimo de parcela é determinado por alguns métodos convencionais, como por exemplo, o método da máxima curvatura, máxima curvatura modificada e os modelos segmentados de resposta platô.

As metodologias para análises estatísticas, conhecidas como computacionalmente intensivas, vêm sendo cada vez mais utilizadas devido a atual capacidade de processamento das máquinas (“hardware”) e pelo desenvolvimento de programas de computadores (“software”), capazes de processar e armazenar grandes

¹IFBAIANO. e-mail: marcio.brito@gm.ifbaiano.edu.br

²DEX – UFLA. e-mail: gustavohumad@hotmail.com

³DEX – UFLA. e-mail: armorais@dex.edu.br

⁴FAR. e-mail: juracimendesmoreira@yahoo.com.br

conjuntos de dados com qualidade, precisão e alta velocidade. Dentre essas metodologias pode-se destacar o “*bootstrap*”, apresentado por Bradley Efron em um influente artigo publicado no *Annals of Statistics*, em 1979. Vários esquemas diferentes de simulação *bootstrap* têm sido propostos na literatura e muitos deles apresentam bom desempenho em uma ampla variedade de situações.

Neste trabalho, foi utilizado o método do platô de resposta linear para avaliar o desempenho do algoritmo de reamostragem *bootstrap* na verificação da estimação do tamanho ótimo de parcela na cultura do mamoeiro.

2 Material e métodos

Foi conduzido um ensaio de uniformidade com a cultivar Golden, do grupo Solo. O ensaio foi formado por plantas em sacos de polietileno, dispostas em 10 fileiras, com 10 plantas em cada fila. Entre o período do desbaste e do aparecimento da primeira flor funcional, foram feitas avaliações das variáveis: altura de planta, diâmetro de caule, número de folhas, largura e comprimento de folha madura. Foram simulados diversos tamanhos e formas de parcelas, em que cada planta foi considerada como uma unidade básica. Foram simulados diversos tamanhos e formas de parcelas, em que cada planta foi considerada como uma unidade básica. Para cada tamanho e forma de parcela, em cada variável e período de avaliação, foi calculado o coeficiente de variação. Para estimação do tamanho ótimo de parcelas foi utilizado o método do platô de resposta linear. Os dados neste estudo consideram apenas a variável altura de planta.

O método do modelo de regressão linear com resposta platô foi utilizado por Peixoto et al. (2011) como um novo método para o cálculo de tamanho ótimo de parcelas e utiliza o seguinte modelo de regressão

$$CV_i = \begin{cases} \beta_0 + \beta_1 X_i + \varepsilon_i & \text{se } X_i \leq X_c \\ P + \varepsilon_i & \text{se } X_i > X_c \end{cases} \quad (1)$$

em que CV_i é o coeficiente de variação observado experimentalmente entre totais de parcela de tamanho de X_i unidades básicas; X_i é o tamanho da parcela em unidades básicas agrupadas, X_c é o tamanho ótimo de parcelas para o qual o modelo linear se transforma em um platô, em relação à abscissa; P é o coeficiente de variação no ponto correspondente ao platô; ε_i é o erro associado ao $CV_{(x)}$ considerado normalmente e independentemente distribuídos, com média 0 e variância σ_ε^2 constante. As estimativas de β_0, β_1 e P foram obtidas por métodos numéricos.

Com a finalidade de verificar as características dos parâmetros método do platô de resposta linear, utilizou-se a técnica de reamostragem para validar as estimativas, pois, pelo método de *bootstrap* pode-se reexaminar uma análise de regressão, comparando-se os resultados obtidos sob certas circunstâncias assumidas (DRAPER; SMITH, 1998).

Foram obtidas amostras considerando como unidade amostral cada planta. O processo baseou-se na técnica de reamostragem “*bootstrap*” não-paramétrico, conforme Efron e Tibshirani (1993), e seguiu-se o seguinte algoritmo:

- a) os valores observados da variável, coletados em cada planta, formaram a amostra mestre. Foram considerados os tamanhos de parcela formados por 1, 2, 4, 5, 10, 20, 25 e 50 unidades básicas (ou plantas);
- b) foram realizadas 5000 reamostragens com reposição, conforme Xie e Mosjidis (1999), seguindo o exposto no item (i), sendo que foram geradas 5000 novos conjuntos de dados da variável analisada;
- c) para cada conjunto de dados gerado, estimaram-se os parâmetros de interesse β_0, β_1, X_0 e P;
- d) com base nos valores obtidos para os parâmetros em cada uma das 5000 simulações foram calculados a média, a variância, o erro padrão e os intervalos de confiança *bootstrap* Percentil para cada um dos parâmetros, com nível de confiança de 95%. Em adição fez-se a distribuição de frequência e histograma dos valores simulados de cada parâmetro. Além disso, para cada tamanho de ensaio simulado foram obtidas 5000 estimativas de *bootstrap* de β_0, β_1, X_0 e P e determinados os valores mínimo, percentil 2,5%, média, percentil 97,5% e máximo de β_0, β_1, X_0 e P.

Para a realização do processo de reamostragem e obtenção das estatísticas foi utilizado o software livre R (R DEVELOPMENT CORE TEAM, 2013).

3 Resultados e discussão

Pelos resultados apresentados nos histogramas do modelo de regressão linear com resposta platô (MLRP), observou-se a ocorrência de certa assimetria na distribuição por amostragem das estimativas em foco, apresentando uma forma diferente do esperado, nesse caso uma distribuição normal.

A não normalidade dos dados foi confirmada pelos gráficos de probabilidade normal, os quais mostram que os ajustes das estimativas *bootstrap*'s não seguem uma distribuição normal padrão. Pelo valor do teste de aderência de Kolmogorov Smirnov notou-se que a hipótese de normalidade foi rejeitada, uma vez que o valor p calculado foi menor que o nível de significância adotado (0,05) para o teste.

O comportamento das distribuições empíricas dos parâmetros β_0 , β_1 e X_0 são apresentados nas Figuras 1 ((a),(b) e (c)), a distribuição das estimativas do parâmetro coeficiente de regressão β_1 (a) mostra uma certa assimetria à esquerda, enquanto que para o tamanho de parcela X_0 (c) a assimetria é à direita. Pode-se observar que o histograma de frequência do parâmetro β_0 (b) sugere a existência de se ter duas modas.

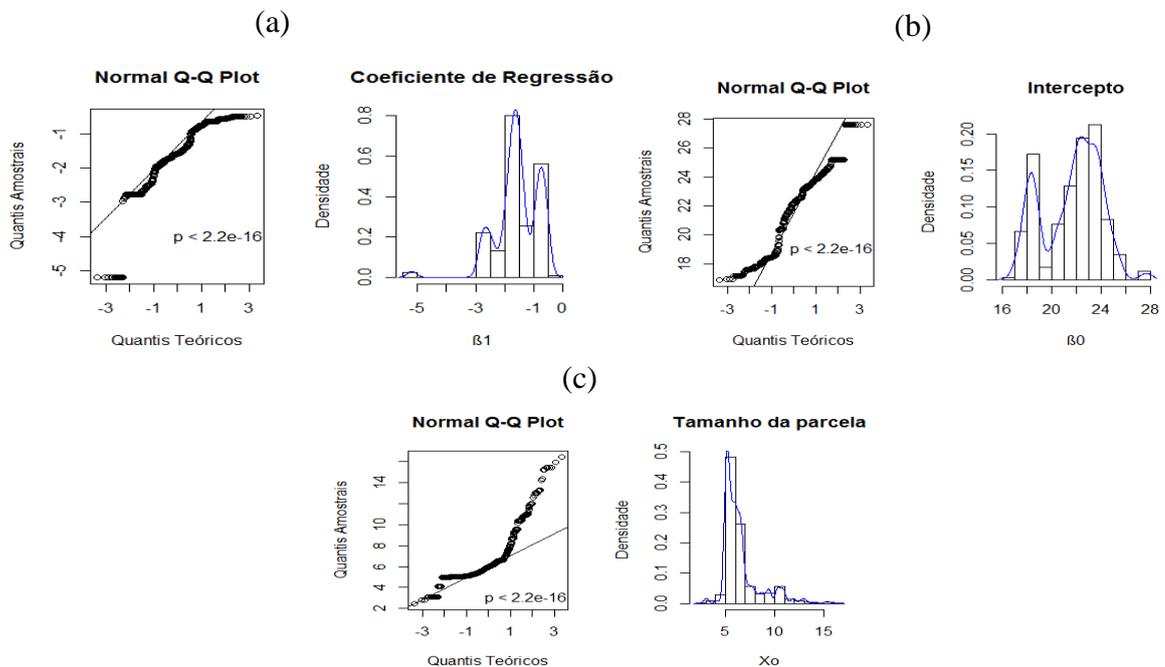


Figura 1 Gráfico de probabilidade normal e histograma do intervalo de confiança para a distribuição *bootstrap*, das 5.000 estimativas dos parâmetros (β_0), (β_1) e (X_0), para a característica altura da planta de mamoeiro, cultivar Golden, obtidas pelo método do modelo de regressão linear com resposta .

Pelos resultados apresentados na Tabela1 verificou-se que os intervalos de confiança calculados, utilizando-se reamostragem *bootstrap*, apresentam uma forte similaridade com os intervalos de confiança assintóticos nas estimativas geradas para o método do modelo de regressão linear com resposta platô.

Segundo Efron e Tibshirani (1993), intervalos *bootstrap* também são aproximados, entretanto oferecem melhor aproximação que os intervalos de confiança padrão.

Tabela 1 Intervalo de confiança assintótico (Ica) e intervalo de confiança *bootstrap* Percentil de 95% (Icb) das 5.000 estimativas dos parâmetros (β_0) e (β_1), do tamanho da parcela (X_0), em unidade básica, obtidas pelo método do modelo de regressão linear com resposta platô (MLRP), para a característica altura do mamoeiro, cultivar Golden, em dez épocas de avaliação

Épocas de Avaliação	Parâmetro					
	β_0		β_1		X_0	
	Ica	Icb*	Ica	Icb*	Ica	Icb*
11/jan	[17.8; 27.2]	[17.4; 25.2]	[-3.1; -0.3]	[-2.8; -0.5]	[3.1; 9.3]	[4.9; 12.8]
19/jan	[16.2; 23.2]	[14.0; 22.2]	[-2.9; -0.8]	[-2.7; -0.5]	[3.8; 7.8]	[4.3; 11.8]
25/jan	[15.2; 21.1]	[12.2; 20.2]	[-2.6; -0.8]	[-2.3; -0.4]	[3.9; 7.6]	[4.9; 12.1]
01/fev	[12.3; 18.2]	[11.7; 19.5]	[-1.3; -0.2]	[-2.3; -0.3]	[4.9; 16.1]	[4.9; 13.8]
16/fev	[22.1; 30.7]	[21.9; 36.2]	[-2.1; -0.5]	[-6.8; -0.8]	[6.1; 16.2]	[3.2; 13.5]
03/mar	[20.1; 29.3]	[20.0; 31.8]	[-2.2; -0.5]	[-4.2; -0.8]	[6.3; 17.9]	[4.2; 15.0]
18/mar	[18.2; 30.1]	[19.3; 35.4]	[-2.6; -0.4]	[-7.7; -0.8]	[5.7; 19.9]	[3.3; 15.7]
02/abr	[19.6; 30.8]	[20.5; 33.0]	[-2.6; -0.5]	[-4.7; -0.9]	[6.2; 19.3]	[4.9; 14.9]
17/abr	[19.7; 30.2]	[20.1; 32.4]	[-2.5; -0.5]	[-4.5; -0.9]	[6.2; 18.3]	[4.9; 14.5]
17/mai	[19.3; 30.1]	[19.7; 35.9]	[-2.5; -0.5]	[-7.7; -0.8]	[5.8; 18.5]	[3.2; 14.6]

*Baseados nas 5000 reamostras

4 Conclusão

Foi possível observar a generalidade de aplicação dessa técnica de estimação através da reamostragem, visto que a mesma pode-se adequar a qualquer situação, sendo seus cálculos rápidos e seus resultados muito eficientes.

Devido à falta de normalidade observada, por prudência, sugere-se que não seja indicada a utilização do intervalo de confiança *bootstrap* Percentil como uma boa alternativa para estimação do tamanho de parcela, até que novos estudos sejam feitos com essa finalidade.

5 Referências

- [1] DRAPER, N. R.; SMITH, H. **Applied regression analysis**. 3th ed. New York: J. Wiley & Sons, 1998. 706 p.
- [2] EFRON, B. (1979). Bootstrap methods: Another look at the jackknife. **Annals of Statistics**, 7:1-26.
- [3] EFRON, B.; TIBSHIRANI, R. J. **An Introduction to the bootstrap**. New York: Chapman & Hall, 1993. 436 p.
- [4] PEIXOTO, A. P. B.; FARIA, G. A.; MORAIS, A. R. Modelos de regressão com platô na estimativa do tamanho de parcelas em experimento de conservação in vitro de maracujazeiro. **Ciência Rural**, Santa Maria, v. 41, n. 11, p. 1907-1913, 2011.
- [5] R DEVELOPMENT CORE TEAM. R: A language and environment for statistical computing., Vienna, Austria,. Disponível em: <http://www.R-project.org>, 2013.
- [6] XIE, C.; MOSJIDIS, J. A. Influence of sampling on precision of genetic correlations in red clover. **Crop Science**, Madison, v. 39, n. 3, p. 863-867, 1999.