

**Reamostragem por Bootstrap para Estimação de Parâmetros de
Dados de Expressão Gênica de Embriões de Aves**

**Claudia Cristina Paro de Paz¹, Alfredo Ribeiro de Freitas², Irineu Umberto Packer³,
Lucia E. Alvares⁴, Jane E. Gabriel⁴, Luiz L. Coutinho⁴**

¹ Instituto de Zootecnia/SP ² CPPSE - Embrapa ³ Departamento de Produção Animal - Esalq - USP

⁴ Laboratório de Biotecnologia Animal - Departamento de Produção Animal - Esalq - USP

Resumo: O objetivo deste trabalho foi calcular intervalos de confiança para média e variância obtidas da expressão dos genes Miogenina e MyoD avaliados em 5 estádios de desenvolvimento (12, 15, 18, 24 e 26) de embriões de aves. Cinco intervalos de confiança foram obtidos por reamostragem com reposição por bootstrap. Em todas as situações estudadas, as médias das estimativas apresentaram viés, os quais foram crescentes com os estádios de desenvolvimento embrionário das aves. O método do percentil com correção de viés acelerado (BCa) apresentou-se mais adequado para aproximar a distribuição da média e da variância.

Palavras Chave: Bootstrap, Intervalo de Confiança, Miogenina, MyoD, RT-PCR Quantitativo.

Introdução

Avanços recentes na área de genética molecular, como estudos de associação entre marcadores e QTL “Quantitative Trait Loci”, manipulação de genes ou seleção de alelos, análises de animais transgênicos e mapeamento genético, têm revelado grande potencial para o melhoramento dos animais domésticos. É possível que o uso destas novas biotecnologias na produção de frangos de corte, adicione melhorias no processo de seleção que atualmente está em torno de 1 a 2% de ganho de peso/ano. A quantificação da expressão de genes Miogenina e MyoD, responsáveis pela miogênese - processo de formação do tecido muscular, é muito importante para o melhoramento. Assim, é de fundamental importância conhecer a verdadeira distribuição de estimadores obtidos do estudo destas características para construir intervalos de confiança e testes de significância exatos. A técnica de bootstrap (EFRON, 1979) utiliza amostragem com reposição obtida por simulação e possibilita obter inferências da população por meio de informações amostrais. O bootstrap “recicla” a informação contida nos dados produzindo uma aproximação da distribuição da estatística sob a suposição de que a amostra em estudo fornece representação adequada da população de interesse (Jeong, Maddala, citado por CRIBARI-NETO e ZARKOS, 2000). Esta técnica é bastante utilizada para estimar a acurácia (compreende o viés - mede se a estatística é consistentemente menor ou maior do que o parâmetro populacional e a precisão - medida de variação na distribuição amostral), o viés de intervalos de confiança e construir testes de hipóteses de estatísticas de interesse, como por exemplo, taxa de crescimento populacional, componentes principais, razão de variáveis (MANLY, 1997) e matriz de covariância em modelos de regressão (CRIBARI-NETO e ZARKOS, 2000), dentre outras. O objetivo deste trabalho foi avaliar por meio da técnica de reamostragem por bootstrap a acurácia de intervalos de confiança da média e variância obtidas da expressão dos genes Miogenina e MyoD avaliados em embriões de duas linhagens de frangos de corte e uma de postura.

Material e Métodos

A formação do tecido muscular depende da ação de genes específicos no desenvolvimento embrionário. Neste estudo foram utilizados dados de embriões de duas linhagens de frangos de

corte (TT=74 e LLC=73) e uma de postura (CC=75), desenvolvidas pelo Centro Nacional de Pesquisa de Suínos e Aves-Embrapa-Concórdia/SC, para medir a expressão dos genes Miogenina e MyoD que são específicos para formação de tecido muscular. Os embriões foram coletados de nos estádios de desenvolvimento 12 (45-49 hs de incubação) 15 (50-55 hs de incubação) 18 (65-69 hs de incubação) 24 (96 hs de incubação) e 26 (156 hs de incubação) e submetidos à extração de RNA total e síntese de cDNA. A quantificação da expressão dos níveis de mRNA de Migenina e MyoD foi realizada por RT-PCR quantitativa e a normalização das amostras foi feita com base na concentração de β -actina. (TSAI e WILTBANK 1996)

Foram utilizadas duas amostras aleatórias como base para reamostragem. A amostra de expressão de Miogenina continha 217 observações, a amostra de expressão de MyoD continha 222 observações. O intervalo de confiança (IC) para média e variância de um parâmetro θ da expressão dos genes Miogenina e MyoD foi calculado por reamostragem com reposição bootstrap (DIXON, 1993; MANLY, 1997). O algoritmo básico de bootstrap é dado por: a) Calcular $\hat{\theta}$ da amostra em estudo; b) obter dos dados, com reposição, amostra aleatória de tamanho n , θ_{1B} , ..., θ_{nB} , para gerar uma distribuição de estimativas bootstrap $\hat{\theta}_B = (\theta_{1B}, \dots, \theta_{nB})$; c) calcular da distribuição de estimativas bootstrap $\hat{\theta}_B = (\theta_{1B}, \dots, \theta_{nB})$; d) repetir os passos (b) e (c) B (número de reamostragens) vezes; e e) usar as B realizações de $\hat{\theta}_B$ para calcular erros-padrão e intervalos de confiança. Foram calculados IC por cinco procedimentos bootstrap considerando-se 2000 repetições ($B=2000$): a) IC padrão sem correção de viés (o IC é dado por $\theta \pm Z_{\alpha/2}\sigma_B$, em o que σ_B é o desvio-padrão das estimativas obtido por reamostragem por bootstrap); b) IC padrão com correção de viés ($\text{viés} = \hat{\theta}_B - \hat{\theta}$); c) IC pelo método de percentil de Efron - reamostragem dos dados originais é usada para gerar a distribuição bootstrap do parâmetro de interesse. O intervalo de confiança $100(1-\alpha)\%$ para um parâmetro θ é dado por dois valores da fronteira da parte central $100(1-\alpha)\%$ desta distribuição, o IC é correto quando o valor observado é a mediana da distribuição; d) Percentil com correção do viés (BC) - Ajusta para a distribuição bootstrap que é viesada, ou seja, sistematicamente baixa ou alta. Os limites dos intervalos corrigidos são obtidos por calcular a distribuição bootstrap e determinar a fração (F) das B repetições que são menores do que o valor observado. Os percentiles apropriados para um IC com 90% de probabilidade são dados por $\Phi(2z_0 - 1,645)$ e $\Phi(2z_0 + 1,645)$ em que z_0 é calculado por $z_0 = \Phi^{-1}(F)$; Φ é a função de distribuição normal cumulativa e Φ^{-1} , a sua inversa; e) Percentil com correção de viés acelerado (BCa) - Faz uma segunda correção que é útil quando alguns pontos são influentes no valor observado, acelera a correção do viés. Estes IC dependem de duas constantes, o z_0 usado para corrigir o viés e o fator de aceleração a , o qual corrige para assimetria. Uma estimativa não-paramétrica de a pode ser obtida por calcular a assimetria da função de influência empírica que mede a sensibilidade da estimativa em relação a cada ponto da amostra original.

Os limites inferior e superior do IC são calculados por $\left(z_0 + \frac{z_0 - 1,645}{1 - a(z_0 - 1,645)}\right)$ e $\left(z_0 + \frac{z_0 + 1,645}{1 - a(z_0 + 1,645)}\right)$, respectivamente, sendo $a \approx \frac{\sum_{i=1}^n (\hat{\theta} - \hat{\theta}_{-i})^3}{6 \left[\sum_{i=1}^n (\hat{\theta} - \hat{\theta}_{-i})^2 \right]^{3/2}}$.

Resultados e Discussão

Os intervalos de confiança por estágio de incubação e genes *Miogenina* e *MyoD*, obtidos com 2000 reamostragens são apresentados na Tabela 1. Em todas as situações estudadas, as médias das estimativas apresentaram viés, os quais foram crescentes com os estádios de desenvolvimento embrionário das aves. Um exemplo da magnitude desses vieses é apresentado na Tabela 2 para o IC padrão com correção de viés. Para a compreensão do cálculo do viés e sua influência na obtenção dos intervalos de confiança é ilustrado um exemplo para a *Miogenina* no estágio 15 do desenvolvimento embrionário. Nesse caso, a estimativa de 2000 reamostragens bootstrap da média foi 1,03762, sendo o viés estimado por $1,03762 - 1,03757 = 0,000053457$. Este valor do viés sugere que a estimativa da amostra original de 1,03757 é superior também ao parâmetro populacional pela quantidade de 0,000053457. Assim, a estimativa com o viés corrigido da média da população original é $1,03757 - 0,000053457 = 1,03751$. Devido ao viés dos estimadores, os intervalos de confiança apresentados na Tabela 1, são obtidos considerando-se o ajuste do viés. Por exemplo, o IC padrão com correção de viés para um parâmetro é dado por $\bar{\theta} \pm Z_{\alpha/2} \sigma_B$, em que σ_B é o desvio-padrão das estimativas obtido por reamostragem por bootstrap. Assim, o IC da média, com 90% de probabilidade, para a *Miogenina* no estágio 15 é calculado por $1,03751 \pm 1,645 \times 0,099603$ ($\sigma_B = 0,099603$), cujos limites são 0,87368 e 1,20135. Com relação ao IC padrão sem correção de viés, o IC é dado por $1,03757 \pm 1,645 \times 0,099603$, cujos limites são 0,87373 e 1,20140. Estes resultados explicam a diferença nos limites do IC pelo método padrão que é devida ao viés. A Tabela 2 apresenta ainda os limites de confiança, em percentagens, com 90% de probabilidade, associado ao método de percentil com correção do viés (BC) e percentil com correção de viés acelerado (BCa). Ambos os métodos implementam o ajuste na distribuição bootstrap viesada para calcular os IC; o método BC considera o fator de correção z_0 , associado ao desvio dos valores da distribuição bootstrap com relação o valor observado, enquanto o método BCa utiliza ambos, z_0 e o fator a , sendo este associado à assimetria dos valores influentes. Observa-se ainda na Tabela 2 que os limites de confiança, em percentagens, obtidos com a distribuição bootstrap ajustada, diferem dos valores nominais de 5% e 95%. Todos os limites de confiança foram viesados para a direita, exceto para BC no estágio 15, indicando que os valores influentes da amostra provocaram assimetria positiva na distribuição bootstrap. Este comportamento dos IC mostrado na Tabela 2 para BC e BCa, foi observado em todos os limites de confiança dos valores da Tabela 1, tanto para *Miogenina* quanto para *MyoD*, indicando que o método do percentil com correção de viés acelerado (BCa) se comportou como mais adequado para aproximar a distribuição da estatística de interesse. Resultados da literatura também comprovam a utilidade da reamostragem bootstrap na construção de intervalos de confiança de estatísticas de interesse. DEMÉTRIO et al. (2000), por exemplo, calcularam erros padrão e intervalos de confiança para os parâmetros índice de fixação e taxa anual de fertilização cruzada de nove locos gênicos de dendezeiro, 97 indivíduos tomados aleatoriamente da população, com o objetivo de construir a

distribuição dos estimadores e comparar os métodos de estimação. Os autores verificaram que as estimativas obtidas por bootstrap apresentaram viés, porém estas técnicas aprimoraram a exploração biométrica dos dados. Segundo CRIBARI-NETO e ZARKOS (2000) na estimativa da matriz de covariância de quadrados mínimos ordinários em modelo de regressão com heterocedasticidade, a correção de erros por bootstrap não funcionou adequadamente, porém estimadores de bootstrap ponderado tenderam a apresentar vícios menores que o estimador de White.

Tabela 1. Intervalos de confiança com 90% de probabilidade para a média e variância, obtidos por cinco critérios de reamostragem por reposição bootstrap, para as expressões dos genes Miogenina e MyoD nos estádios de 12, 15, 18, 24 e 26 de desenvolvimento de embriões de aves

Método	Variável: Miogenina									
	Estatística: Média					Estatística: Variância				
	Estádio 12		Estádio 15		Estádio 18		Estádio 24		Estádio 26	
	LI	LS	LI	LS	LI	LS	LI	LS	LI	LS
Padrão CCV	0,07221	0,10778	0,87368	1,20135	2,94659	3,73598	24,1290	33,7600	75,4290	87,0420
Padrão SCV	0,0722	0,10778	0,87373	1,20140	2,94679	3,73619	24,1310	33,7620	75,4310	87,0440
Efron	0,07274	0,10871	0,87234	1,19971	2,94936	3,74404	24,2930	34,0480	75,4680	87,1220
BC	0,07313	0,10906	0,87158	1,19889	2,95348	3,74826	24,4520	34,4510	75,5400	87,1370
BCa	0,07375	0,10938	0,87730	1,20237	2,96292	3,77043	24,6990	34,8200	75,5400	87,1370
Método	Variável: MyoD									
	Estatística: Média					Estatística: Variância				
	Estádio 12		Estádio 15		Estádio 18		Estádio 24		Estádio 26	
	LI	LS	LI	LS	LI	LS	LI	LS	LI	LS
Padrão CCV	0,002507	0,00464	0,252780	0,500180	1,56629	3,97407	236,131	590,876	393,505	700,162
Padrão SCV	0,002390	0,004530	0,242940	0,490340	1,50985	3,91765	227,624	582,369	370,947	687,604
Efron	0,002240	0,004400	0,230740	0,480260	1,57900	4,01284	220,181	578,225	363,293	683,331
BC	0,002484	0,004600	0,253240	0,497800	1,76892	4,31992	239,199	596,384	397,260	724,639
BCa	0,002528	0,004670	0,263420	0,513990	1,87638	4,65939	254,263	622,358	403,116	735,036
Método	Variável: MyoD									
	Estatística: Média					Estatística: Variância				
	Estádio 12		Estádio 15		Estádio 18		Estádio 24		Estádio 26	
	LI	LS	LI	LS	LI	LS	LI	LS	LI	LS
Padrão CCV	0,50722	0,67620	0,72720	0,94741	3,58534	4,41570	14,6701	18,7490	43,880	54,6890
Padrão SCV	0,50723	0,67620	0,72726	0,94747	3,58547	4,41583	14,6701	18,7490	43,8790	54,6870
Efron	0,51059	0,67818	0,73311	0,94936	3,60812	4,43169	14,8158	18,7950	43,8940	54,8940
BC	0,51206	0,68056	0,74083	0,95775	3,61511	4,44044	14,89420	18,9180	44,0440	54,910
BCa	0,51400	0,68235	0,74605	0,96775	3,62733	4,45894	14,9525	19,0070	44,1930	55,220
Método	Variável: MyoD									
	Estatística: Média					Estatística: Variância				
	Estádio 12		Estádio 15		Estádio 18		Estádio 24		Estádio 26	
	LI	LS	LI	LS	LI	LS	LI	LS	LI	LS
Padrão CCV	0,06513	0,13081	0,07790	0,26550	1,70045	4,30453	51,48420	106,7150	278,0620	720,4170
Padrão SCV	0,06250	0,12817	0,07337	0,26097	1,63637	4,24045	49,94790	105,1780	267,6220	709,6770
Efron	0,06090	0,12665	0,07334	0,26253	1,71847	4,27477	50,28430	104,9130	286,3770	718,7890
BC	0,06644	0,13332	0,08451	0,27767	1,88797	4,59414	53,92020	109,8070	320,6780	779,5870
BCa	0,06874	0,13976	0,09384	0,29821	1,99273	5,05902	55,5610	115,3060	341,230	884,9520

Padrão CCV = Bootstrap Padrão com correção de viés

Padrão SCV = Bootstrap Padrão sem correção de viés

BC = Bootstrap Percentil com correção de viés

BCa = Bootstrap Percentil com correção de viés acelerado

Tabela 2. Média estimada por bootstrap, viés e média corrigida pelo método Padrão com correção de Viés; correção do viés (Z_0), aceleração (a) e limites de confiança em percentagens com 90% de probabilidade associados aos métodos BC e BCa para o gene Miogenina

Estádio	Variável: Miogenina									
	Padrão com Correção de Viés					BC, BCa				
	Estatística Observada	Média Bootstrap	Viés Aproximado	Estatística Corrigida	Correção do Viés (Z_0)	Aceleração	Limite Inferior	Limite Superior	Limite Inferior	Limite Superior
12	0,090000	0,090004	0,00000369	0,089996	0,01943	0,016724	5,4138	95,3881	5,9029	95,9029
15	1,03757	1,03762	0,00005345	1,03751	-0,00564	0,018834	4,8847	94,8826	5,4096	94,8826
18	3,34149	3,34169	0,00020100	3,34129	0,00877	0,02721	5,1835	95,1783	5,9666	95,1783
24	28,9460	28,9480	0,0014600	28,9450	0,055174	0,0377528	6,2452	96,0387	7,4258	96,0387
26	81,2380	81,2400	0,00250099	81,2350	0,006270	-0,000298	5,1307	95,1280	5,1222	95,1280

BC = Bootstrap Percentil com correção de viés

BCa = Bootstrap Percentil com correção de viés acelerado

Conclusões

1. Em todas as situações estudadas, as médias das estimativas bootstrap apresentaram viés, os quais foram crescentes com os estádios de desenvolvimento embrionário das aves.

2. O método do percentil com correção de viés acelerado (BCa) se comportou como mais adequado para aproximar a distribuição da estatística de interesse.

Bibliografia

- CRIBARI-NETO, F., ZARKOS, S.G. Heteroskedasticity-consistent covariance matrix estimation: white's estimator and the Bootstrap. *In: Workshop 2000 - Métodos Computacionais em Estatística*. São Carlos-SP, p.1-19, 2000.
- DEMÉTRIO, C.G.B., VENCOVSKY, R., DIAS, C.T.S. et al. Reamostragem por bootstrap em dados de genética molecular. *In: Workshop 2000 - Métodos Computacionais em Estatística*, São Carlos-SP, 2000.
- DIXON, P.M. The Bootstrap and the Jackknife: Describing the Precision of Ecological Indices. *In: SCHEINER, S.M. GUREVITCH, J. Design and analysis of ecological experiments*. New York: Chapman & Hall, 1993, 445p.
- EFRON, B. Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, v.21, p.460-480, 1979.
- MANLY, B.F.J. *Randomization, bootstrap and Monte Carlo methods in biology - Texts in Statistical Science*. Chapman & Hall, London, 2.ed., 1997, 399p.
- POLLAK, J. Beef cattle breeding - past, present, future. *In: Reunião Anual Sociedade Brasileira de Zootecnia*, 37, 2000, Viçosa. *Palestras...* Viçosa: SBZ, 2000. p.101-111.
- TSAL, S., WILTBANK, M. C. Quantification of mRNA using competitive RT-PCR with standard curve methodology. *Biotechniques*, v.21, p.862-866, 1996.
- VISSCHER, P.M., THOMPSON, R., HALEY, C.S. Confidence intervals in QTL mapping by bootstrapping. *Genetics*, v.143, p.1013-1020, 1996.
- www.de.ufpe.br/cribati/bootstrap.pdf CRIBARI-NETO, F. Bootstrap and applications.