



Estimação da capacidade auxiliada com procedimento *bootstrap* na cadeia produtiva do Leite

Mayra Marques Bandeira, Universidade Federal de Viçosa , mayra.mbandeira@gmail.com

Fernando Luiz Pereira de Oliveira, Universidade Federal de Ouro Preto, fernandoluz@iceb.ufop.br

José Ivo Ribeiro Júnior, Universidade Federal de Viçosa , jivo@ufv.br

Frederico Rodrigues Borges da Cruz, Universidade Federal de Minas Gerais, fcruz@est.ufmg.br

Lupércio França Bessegato, Universidade Federal de Juiz de Fora, lupercio.bessegato@ufjf.edu.br

Resumo: A competitividade do mercado tem levado as empresas e indústrias a elevar a qualidade de seus produtos e serviços. O monitoramento e a avaliação da capacidade do processo tornam-se indispensáveis para que as empresas elevem essa qualidade. Os índices de capacidade são medidas adimensionais que traduzem a capacidade de um processo de atender especificações estabelecidas pelos clientes internos e externos. O objetivo deste trabalho é obter as estimações pontuais de alguns índices de capacidade multivariados propostos e com o auxílio de técnicas de reamostragem estimativas intervalares. Dentre estas técnicas, os testes bootstrap são utilizados de forma a reorganizar o conjunto de dados sem alterar as informações existentes e assim obter reamostras para auxiliar nas estimativas via índices de capacidade. Neste trabalho, utilizando dados reais, foram estimados alguns índices de capacidade em situações em que suposições usuais para obtenção destas estimativas não são válidas fornecendo uma ferramenta adicional para tomada de decisão. Após a realização das reamostragens pelos métodos bootstrap pode-se concluir que o processo não é capaz de atender às especificações estabelecidas pelas estimativas obtidas. Palavras-chave: Controle de Processos; Reamostragem; Índices de Capacidade.

1. Introdução

Segundo Montgomery (2004), o controle da qualidade é composto por ações que garantem a qualidade dos meios que controlam e medem características de um item, processo ou instalação, de acordo com as especificações estabelecidas. Júnior (2013) afirma que qualidade significa atender às especificações, e assim, todos os itens de um mesmo produto ou serviço produzido por um mesmo processo devem apresentar valores da variável resposta localizados dentro dos limites especificados.

Sob outra ótica, Mingoti et al. (2011), sugerem que a qualidade está na capacidade de um produto em atender a determinadas necessidades do cliente. Desse modo, é necessária a utilização de mecanismos de monitoramento do processo e medidas que quantifiquem sua capacidade em atender às exigências dos clientes. Tais medidas podem ser implementadas utilizando-se os índices de capacidade, uma vez que, por meio dos resultados obtidos por eles pode-se analisar a capacidade do processo. A partir da análise de um item ou processo ações podem ser realizadas, de acordo com o resultado obtido, o que leva a melhorias na qualidade caso sejam necessárias. Avaliar a capacidade do processo é importante, uma vez que nos permite quantificar a maneira que o processo de produção está produzindo de acordo com as especificações.



Segundo Oliveira (2007), os índices de capacidade mensuram se o processo sob controle estatístico está atuando conforme as especificações, que podem ser fornecidas pelos clientes ou internamente no processo.

Segundo Pan e Li (2014), os índices de capacidade de processo têm sido muito utilizados na indústria para estimar medidas quantitativas de desempenho de processos que levam à melhoria da qualidade. Em diversas situações, para que estas estimativas sejam confiáveis, algumas suposições precisam ser analisadas. Uma das formas de se obter estimativas da capacidade de processos, quando estas suposições não são atendidas, é através dos métodos de reamostragem. Reorganizar o conjunto de dados observados, segundo Gentle (2009), pode dar uma indicação de como o conjunto de dados é incomum no que diz respeito a uma dada hipótese nula.

Existem muitos procedimentos úteis para a análise dos dados que envolvem o particionamento na amostra original, como o procedimento *bootstrap*. Usando os subgrupos da amostra completa, é possível obter estimativas ou estatísticas de teste sem depender dos pressupostos que levaram à escolha destes. O que torna a técnica de reamostragem *bootstrap* extremamente útil por não necessitarmos de muitas suposições para estimação dos parâmetros das distribuições de interesse.

O método *bootstrap* é um método de reamostragem que, segundo Efron e Tibshirani (1993), é definido como um método de simulação de dados baseado em inferência estatística, que pode ser usado para produzir inferências. Este método consiste na reamostragem com reposição da amostra original. É um procedimento computacional intensivo baseado na técnica da substituição que possibilita estimar a distribuição amostral de estatísticas de interesse. (Filho et al., 2002). Esta técnica de reamostragem é útil quando o cálculo de estimadores for complexo. Uma vantagem é que a reamostragem nos permite alternativas para obter desvios padrões e intervalos de confiança por meio da análise do nosso banco dados.

Segundo Costa (2006), com o uso da reamostragem obtém-se uma distribuição empírica desconsiderando a distribuição normal assumida de uma estatística. A fim de que as aplicações das técnicas de reamostragem resultem em valores confiáveis, é necessário que, a partir de uma amostra mestre, sejam realizados milhares ou até centenas de milhares de reamostragens de tamanhos iguais. De acordo com Davison e Hinkley (1997), para utilização do método *bootstrap* é necessária a realização de um número grande de reamostragens e o cálculo de diversas estatísticas para cada uma destas reamostragens. Com o uso da reamostragem, a distribuição assumida por amostragem é desconsiderada e, gera-se uma distribuição empírica. Sendo assim as estimativas são obtidas nestas distribuições empíricas.

Neste trabalho utilizamos o método *bootstrap* para obter as inferências dos índices de capacidade de processos de produção de leite, retirando centenas de reamostras com reposição a partir da nossa amostra original. Calculamos a média dos índices para cada reamostra e obtivemos as distribuições empíricas dessas reamostras. Através dessa técnica foi possível realizar o que é desejável na prática, a repetição.

Dentre os objetivos deste trabalho destaca-se a obtenção de estimativas intervalares de alguns índices de capacidade multivariados, auxiliada pelo método de reamostragem *bootstrap*. Acredita-se que esta abordagem possa fornecer um auxílio adicional na interpretação da capacidade de processos em setores industriais, em nosso trabalho no setor de produção de leite.



2. Material e métodos

Neste trabalho foram coletados 31 dados referentes à produção de leite de fazendeiros pertencentes à região da Zona da Mata no ano de 2014. Há na região um programa de desenvolvimento que visa promover a evolução acelerada da produção econômica de leite, os dados utilizados neste trabalho foram fornecidos por este programa. As variáveis analisadas foram produção média de leite (L/dia), produção / vacas em lactação (L/dia) e custo total do leite (R\$/L) denominadas por A, B e C, respectivamente. Estas variáveis apresentam limite inferior de especificação (*LIE*) e limite superior de especificação (*LSE*) definido na Tabela 1.

Tabela 1. Limite de especificação definido para as variáveis A, B e C.

Variável	<i>LIE</i>	<i>LSE</i>
A	531,57	4.907,93
B	15,04	28,83
C	0,86	1,13

Na análise de um processo, existem ocasiões em que um grande número de variáveis deve ser analisado simultaneamente, incorporando a correlação existente entre as características de interesse, situação considerada neste trabalho. Segundo Mingoti (2011), é muito comum que a qualidade de um processo seja determinada por mais de uma característica, sendo então desejável avaliar a capacidade do processo por medidas que levem em consideração todas as características simultaneamente. Para Oliveira (2007), existe uma tendência moderna que sugere a utilização dos métodos estatísticos multivariados. Segundo Ferreira (2011), com os métodos multivariados, as análises, descrições e inferências são realizadas com base em respostas simultâneas, valendo-se a correlação entre as variáveis. De modo que a avaliação da capacidade do processo deve ser feita por meio dos índices multivariados.

Os métodos multivariados, segundo Woodall e Montgomery (2014), são fundamentais quando se quer monitorar várias variáveis de qualidade e tirar proveito de todas as relações entre elas, por esse motivo, foram aqui utilizados os índices de capacidade multivariados.

Muitos são os estudos de índices de capacidade para processos multivariados sendo alguns deles: Chen (1994), Bernardo e Irony (1996), Niverthi e Dey (2000), Yeh e Chen (1999), Li e Lin (1996), Wang et. al (2000) e Mingoti e Glória (2005) entre outros. Entretanto, não existe ainda na literatura um consenso sobre qual índice de capacidade multivariado seria o melhor ou mesmo sobre como a capacidade em termos multivariados deveria ser quantificada.

Utilizando o teste Shapiro Wilk e considerando $\alpha = 0,05$, foi testada a normalidade dos dados. Foi realizada uma transformação raiz quadrada nos dados da variável A, denominada A*, buscando normalidade deste. Após a constatação de normalidade dos dados transformados, foram obtidas estimativas pontuais e intervalares dos índices de capacidade multivariados, definidos por:

$$C_{pgeom} = \left(\prod_{i=1}^p C_p(x_i) \right)^{1/p} \quad (1)$$



foram adotados os passos definidos por Ramos e Ho (2003) na realização do método de reamostragem *bootstrap*.

Inicialmente foi considerada uma amostra constituída de n elementos da população em estudo (x_i , $i=1, 2, 3, \dots, n$). Em seguida, através de sorteio com reposição, foram retiradas novas amostras (reamostras) de r valores ($r < n$) a partir da amostra original (x_{ij}^* , $i=1, 2, 3, \dots, r$; $j=1, 2, \dots, B$). A partir daí foi calculada uma estimativa pontual do parâmetro de interesse com base nas observações reamostradas. Este procedimento foi repetido 1000 vezes e ao final, foi possível a obtenção da média amostral, do desvio padrão e do erro quadrático médio amostral das estimativas. Como dito anteriormente, a reamostragem dos dados foi realizada 1000 vezes. Foram estimados, para cada reamostragem, os parâmetros de interesse e assim foi possível obter uma distribuição empírica para as estimativas *bootstrap*, bem como novas estimativas dos parâmetros, como, por exemplo, a média e os intervalos de confiança.

A avaliação da incerteza sobre os valores dos parâmetros obtidos foi realizada por meio dos intervalos de confiança para os parâmetros de interesse pelo método *bootstrap*. Um intervalo de confiança foi definido por limites, tal que para qualquer um α ,

$$P(\theta < \hat{\theta}_\alpha) = \alpha \quad (7)$$

O coeficiente de confiança de um intervalo é definido por

$$\gamma = 1 - (\alpha_1 + \alpha_2), \quad (8)$$

no qual, α_1 e α_2 são as probabilidades de erro pertencentes a cauda esquerda e direita, respectivamente. O desejável é obter intervalos de confiança menores possíveis e nível de confiança maior possível, ou seja, próximo de 100%.

O intervalo $100(1 - \alpha)\%$ de confiança *bootstrap* foi obtido através dos percentis das diferenças dos valores das estatísticas das reamostras em relação ao valor médio da estatística nas reamostras. Então, a diferença entre esses valores, para cada reamostra i , foi obtida da seguinte forma:

$$\text{dif} = \hat{\theta}(b) - \hat{\theta}(\cdot) \quad (9)$$

$$\text{IC}_{\text{boot.percentil}} = [\hat{\theta} - P_{97,5\text{dif}}; \hat{\theta} - P_{2,5\text{dif}}] \quad (10)$$

O método *bootstrap* foi utilizado de forma a reorganizar o conjunto de dados sem alterar as informações existentes, realizamos 1000 reamostras de tamanho 10, 15 e 20.

3. Resultados e discussão

Os índices de capacidade geométrico, $C_{p\text{geom}}$ e $C_{pk\text{geom}}$, obtidos para as variáveis A^* , B e C são iguais a 0,581 e 0,028 respectivamente. Esses valores são bem abaixo de valores tidos como referência na literatura o que indica que o processo não está atendendo as especificações.

Os histogramas permitem visualizar as distribuições empíricas para as 1000 reamostragens pelo método *bootstrap* para os diferentes tamanhos da amostra (n) e podem ser visualizados nas Figura 1 e 2.



Na Tabela 2, são apresentados os valores obtidos para a média dos C_{pgeom} e C_{pkgeom} de acordo com o tamanho da amostra e o erro quadrático médio associado (EQM) a elas, todas realizadas com reamostragem *bootstrap* de tamanho 1.000. Pode-se observar que o EQM diminui quando aumentamos o tamanho (n) da amostra. Os intervalos de confiança para a média dos índices obtidos é apresentado nas tabelas 2 e 3. Utilizamos o método *bootstrap* percentil, com um nível de confiança de 95%.

Tabela 2. Estimativas para os índices C_{pgeom} e C_{pkgeom} para as variáveis A*, B e C pelo método *bootstrap*.

Índices de capacidade	Estimadores	Tamanho amostral		
		10	15	20
C_{pgeom}	Média	0,633	0,616	0,610
	EQM	0,003	0,001	$6,641e^{-5}$
	ICpercentil	(0,622 ; 0,635)	(0,612 ; 0,620)	(0,607 ; 0,614)
C_{pkgeom}	Média	0,010	0,012	0,017
	EQM	0,002	0,001	0,001
	ICpercentil	(0,007 ; 0,012)	(0,010 ; 0,014)	(0,015 ; 0,018)

Os índices de capacidade de Niverthi-Dey, C_{pnd} e C_{pknd} , obtido para as variáveis A*, B e C é igual a 0,417 e -0,239 respectivamente. Foi utilizado o índice global, definido como o valor mínimo dos índices presentes no vetor. Não há valores de referência que possam ser usados para comparação com os valores obtidos por esses índices, porém valores muito baixos nos indicam que o processo não está de acordo com as especificações.

Observam-se na Tabela 3 os valores obtidos para a média dos C_{pnd} e C_{pknd} de acordo com o tamanho da amostra e o EQM associado a elas. Nesta situação, o EQM não diminui muito quando aumentamos o tamanho (n) da amostra, sendo próximos os valores das médias dos índices C_{pnd} obtidos para amostras de tamanho 15 e 20.

Tabela 3. Estimativas para os índices C_{pnd} e C_{pknd} para as variáveis A*, B e C pelo método *bootstrap*.

Índices de capacidade	Estimadores	Tamanho amostral		
		10	15	20
C_{pnd}	Média	0,506	0,473	0,437
	EQM	0,406	0,133	0,084
	ICpercentil	(0,469 ; 0,548)	(0,451 ; 0,497)	(0,419 ; 0,456)
C_{pknd}	Média	-0,279	-0,271	-0,258
	EQM	0,002	0,001	0,001
	ICpercentil	(-0,285 ; -0,273)	(-0,276 ; -0,266)	(-0,262 ; -0,255)

4. Conclusões

Após a realização das reamostragens pelos métodos *bootstrap* pode-se concluir pelas estimativas obtidas que o processo não é capaz de atender às especificações estabelecidas. Isso reforça o fato que medidas devem ser realizadas para identificar o que está afetando o processo a ponto de não atender as especificações sugeridas. Na produção de leite, a variação do tamanho de produtores pode ter ocasionado a não capacidade de atender as especificações.

Como esperado, foi observado que quanto maior o tamanho da amostra (n), mais simétrica é a distribuição empírica, o que permite a realização de inferências considerando



esse fato. Há também uma sensível redução do erro quadrático médio (EQM) quando aumentamos o tamanho (n) da amostra.

Os índices de capacidade geométricos apresentaram valores muito próximos de zero, enquanto valores negativos foram obtidos pelos índices de Niverthi Dey. Dessa forma constatamos o quanto o índice de capacidade de Niverthi Dey penaliza o processo, conforme verificado em outros trabalhos. Em vista disso, sugerimos para trabalhos futuros um estudo que determine valores referência para a proporção de itens fora de especificação para o índice de Niverthi Dey associados com os valores estimados.

Pela natureza dos dados reais, os valores estimados dos índices podem dar uma ideia sobre a capacidade dos processos, muito embora não impliquem que serão observadas as mesmas proporções de não conformidade que seriam esperadas para dados que seguissem uma distribuição normal. Investigações futuras deverão ser conduzidas, envolvendo estas e outras questões correlatas.

5. Agradecimentos

Este trabalho foi parcialmente financiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES e pela Fundação de Amparo à Pesquisa do estado de Minas Gerais - FAPEMIG.

Referências

- BERNARDO, J. M., IRONY, T. X. - A General multivariate bayesian process capability index, *The Statistician*, vol. 45, p. 487-502, 1996.
- CHEN, H. *A multivariate process capability index over a rectangular solid tolerance zone*. *Statistica Sinica*, v.4, n. 2, p.749-758, 1994.
- COSTA, G. G. O. *Um procedimento inferencial para análise fatorial utilizando as técnicas bootstrap jackknife: construção de intervalos de confiança e testes de hipótese*. Tese de Doutorado, Pontifícia Universidade Católica do Rio de Janeiro, 2006.
- DAVISON, A.C.; HINKLEY, D.V. *Bootstrap methods and their application*, New York: Ed. Cambridge, 1997.
- EFRON, B., & TIBSHIRANI, R. J. *An introduction to the bootstrap*, New York: Chapman & Hall, 1993.
- FILHO, M. J. F.; LIBARDI, P. L.; LIER, Q. J., e CORRENTE J. E. *Método convencional e bootstrap para estimar o número de observações na determinação dos parâmetros da função k*. *Revista Brasileira de Ciência do Solo*, v. 26, n.4, p.895-900, 2002.
- FERREIRA, D. F. *Estatística Multivariada*, Lavra,; 2 ed. Editora ULFA, 2011.
- GENTLE, J.E. *Computational Statistics*, New York: Ed. Springer, 2009.
- JUNIOR, J. I. R. *Métodos estatísticos aplicados ao controle da qualidade*, Viçosa, MG: Ed. UFV, 2013.
- JURAN, J. M. *Quality Control handbook*, New York, Mc Graw-Hill, 1974.
- KANE, V. E. Process capability indices, *Journal of Quality Technology*, Milwaukee, vol. 18, n. 1, p. 41-52, 1986.
- LI, Y. E LIN, C. Multivariate C_p Value. *Chinese Journal of Applied Probability and Statistics*, n.12, p. 132-138, 1996.
- MINGOTI, S. A.; OLIVEIRA, F. L. P.; CONCEIÇÃO, M. M. C. *Índices de capacidade para processos multivariados independentes: extensões dos índices de Niverthi e Dey e Mingoti e Glória*. *Revista Produção*, v. 21, n.1, p.94-105, 2011.
- MINGOTI, S. A.; OLIVEIRA, F. L. P. *On Capability Indices for Multivariate Autocorrelated Process*. *Brazilian Journal of Operations & Production Management*, v.8, n. 1, p.133-152. 2011.
- MINGOTI, S. A., GLÓRIA, F. A. A. Comparando os métodos paramétrico e não-paramétrico na determinação do valor crítico do teste estatístico de Hayter & Tsui. *Revista Produção*. 2005.
- MONTGOMERY, D. C. *Introdução ao Controle Estatístico da Qualidade*, São Paulo: Ed. LTC, 2004.



- NIVERTHI, M.; DEY, D. K. *Multivariate process capability: a bayesian perspective*. Communications in Statistics-Simulation and Computation, v. 29, n.2 , p.667-687, 2000.
- OLIVEIRA, F. L. P. *Índices de capacidade para processos multivariados autocorrelacionados*. Dissertação de Mestrado, Departamento de estatística da UFMG, 2007.
- PAN, J.; LI, C. *New capability indices for measuring the performance of a multidimensional machining process*. Expert Systems with Applications, v.41, n.5, p.2409-2414, 2014.
- RAMOS, A.W.; HO, L.L. *Procedimentos inferenciais em índices de capacidade para dados autocorrelacionados via bootstrap*. Revista Produção, v.13, n.3, p.50-62, 2003.
- WANG, K. F., LAWRENCE, P. F, MISKULIN, D. J., SHAHRIARI, H. Comparison of Three Multivariate Process Capability Indices. *Journal of Quality Technology*. v. 32, n. 3, p. 263-275, 2000.
- WOODALL, W. H., MONTGOMERY, D.C. *Some Current Directions in the Theory and Application of Statistical Process Monitoring*. Journal of Quality Technology, v. 46, n. 1, Janeiro 2014.
- YEH, A. B. E CHEN, H. "A Nonparametric Multivariate Process Capability Index", 1999.