

APLICAÇÃO DA TÉCNICA KERNEL A DADOS DE PESAGENS DE BOVINOS¹

Alfredo Ribeiro de Freitas², Luiz Otávio Campos da Silva³, Kepler Euclides Filho³,
José Eduardo dos Reis⁴, Cláudio Massaki Kakuda⁵,
Fábio H. Milanetto Ferreira⁵

¹Financiado pelo CNPq;

² Embrapa Pecuária Sudeste, CP 339, 13560-970 São Carlos-SP;

³ Embrapa Gado de Corte, CP 154, 79002-970 Campo Grande-MS;

^{4,5} Pós-graduação em São Carlos, SP: ⁴ICMC-USP, São Carlos-SP;

⁵Instituto de Física, USP, São Carlos-SP.

Introdução

No estudo da estimação de densidade, que é a construção da estimativa da função densidade dos dados observados de uma amostra, há dois tipos de técnicas: a) paramétrica: quando a distribuição é normal; nesse caso utilizam-se histogramas, além da média e desvio-padrão da amostra, para se obter inferências estatísticas; b) não-paramétrica: apropriada para as distribuições que não se ajustam à normal e apresentam superfícies irregulares. Para se obter inferências estatísticas nesse caso, pode-se usar uma curva de densidade normal *Kernel*, a qual requer um parâmetro de alisamento λ , para determinar o grau de irregularidade na estimativa da função de densidade destas distribuições. Em dados de pesagens de bovinos, é comum a ocorrência de vícios de pesagens dos animais (FREITAS et al., 2000), implicando em distribuições irregulares e multimodais, as quais diferem da distribuição normal. Nestas situações, o método não-paramétrico pode ser mais apropriado do que o paramétrico. O objetivo deste trabalho foi comparar a distribuição paramétrica normal e densidade normal *Kernel* para inferências estatísticas de dados de pesagens de bovinos da raça Gir, pertencentes ao Arquivo Zootécnico Nacional das Raças Zebuínas mantido pela ABCZ-Embrapa Gado de Corte.

Material e Métodos

Foram utilizados dados de nove pesagens: ao nascimento (PN) e oito (P1 a P8), realizadas em intervalos trimestrais, até os dois anos de idade, de 35.715 de bovinos da raça Gir, machos e fêmeas, nascidos de 1976 a 1992. Foram utilizadas duas densidades com base na distribuição normal: a *paramétrica*, com a média e desvio-padrão obtidos da amostra e a *não-paramétrica* ou *Kernel* (HOUGAARD et al., 1989; SAS Institute, 1996; BOWMAN, et al. 1998), que utiliza um parâmetro de alisamento λ , o qual determina o grau de irregularidade na estimativa da função de densidade. O estimador *Kernel* tem a forma:

$$\hat{f}_\lambda(y) = \frac{1}{n\lambda} \left[\sum_{i=1}^n K_0\left(\frac{y - x_i}{\lambda}\right) \right]$$
, em que K_0 é a função *Kernel*; λ é a largura da banda, ou

seja, o parâmetro que controla o grau de alisamento dos dados e x_1, \dots, x_n , variáveis aleatórias contínuas. A função de densidade normal padronizada, comumente usada como função *Kernel*, é $K_0(t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right)$, para $-\infty < t < \infty$. Pode-se selecionar uma

largura de banda para cada estimador *Kernel* por especificar c na fórmula $\lambda = cQn^{-1/5}$, em que Q é o intervalo interquartil (Q₃ - Q₁) da variável y e c uma constante específica. A

discrepância entre o estimador pela função *Kernel* e a densidade verdadeira, $f(x)$, pode ser medida pela curva do quadrado médio do erro integrado aproximado (AMISE), dado por $AMISE(\lambda) = (1/4)\lambda^4 \left(\int t^2 K(t) dt \right)^2 \int_x (f''(x))^2 dx + (1/n\lambda) \int_t K(t)^2 dt$, sendo o parâmetro de alisamento (λ) ótimo aquele que minimiza AMISE. Na expressão acima, pequenos valores de λ e da constante c fornecem estimativas de densidades com superfícies irregulares ou “denteadas”, enquanto que valores grandes de c e de λ estão associados a estimativas de densidade com superfícies mais regulares ou mais lisas (“smoothing”). Foi utilizado o sistema SAS (SAS, 2000) para efetuar as análises.

Resultados e Discussão

Na Tabela 1 são apresentadas as estatísticas associadas as distribuições normal paramétrica e de *Kernel*. Uma vez que λ é o parâmetro de alisamento ótimo, ou seja, aquele que minimiza AMISE, a diferença das estimativas obtidas por estas duas densidades revelam os prejuízos em se considerar que os dados de pesagens se ajustam a uma distribuição normal. Em todas as pesagens, a moda obtida pelo estimador *Kernel*, foi inferior à obtida por meio da função paramétrica normal. Isto mostra que o pico da distribuição, considerando-se a função *Kernel*, situou-se à esquerda daquele obtido pela função paramétrica. Uma explicação para isto é que o estimador *Kernel*, pelo fato de utilizar λ , que determina o grau de irregularidade das distribuições dos dados na estimativa da função de densidade, leva em conta os picos de frequência nas distribuições. Como é sabido, em uma distribuição normal, a moda e a média são iguais e, juntamente com a variância, são parâmetros que determinam esta distribuição. Assim, a discrepância observada na moda, obtida do ajuste de duas densidades, pode ser atribuída à existência de picos de frequências associados à distribuição dos dados, os quais são responsáveis pelo desvio da distribuição com relação à normal paramétrica. Para todas as pesagens, verificaram-se assimetria e curtisse positivas. Verificaram-se, ainda, que os dados não se ajustaram a uma distribuição normal, pelo teste de Kolmogorov-Smirnov, ou seja, rejeitou-se a hipótese de nulidade: “os dados da amostra possuem distribuição normal” ($P < 0,05$).

Conclusões

O estudo dos dados de pesos de bovinos da raça Gir até os dois anos de idade, por meio da densidade normal paramétrica e densidade de *Kernel*, mostrou ser este último mais adequado para se estudar a distribuição dos dados. O estimador *Kernel* considera, no ajustamento, a irregularidade existente nos dados, como os picos de frequências, e, com isso, proporciona resultados mais adequados com relação a distribuição normal paramétrica.

Referências Bibliográficas

- BOWMAN, A. HALL,P.;PRVAN,T. Bandwidth selection for the smoothing of distribution functions. *Biometrika*, v.85,n.4,799-808,1998.
- FREITAS, A, R. de., SILVA, L.O.C., JOSAHKIAN,L.A., ALENCAR, M.M. A qualidade de pesagem de bovinos da raça Gir In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA., 37, 2000, Viçosa, MG. Anais... Viçosa: SBZ, 2000. CD3p
- HOUGAARD, P; PLUM,A; RIBEL,U. *Biometrics*, v.45,n.4,1041-1052,1989.
- SAS Institute 2000. *SAS/INSIGHT User's Guide*. versão 8.2, versão para Windows Cary, NC, USA.

Tabela 1. Estatísticas associadas às funções paramétrica e de Kernel em que Moda_P e Moda_K são as modas obtidas da normal paramétrica e Kernel, respectivamente.

Pesos*	μ	σ	Moda_P	c	λ	Moda_K	AMISE	Prob < F**
PN	25,01	3,16	25,01	0,7852	0,3859	24,90	2,558E-05	0,005
P ₁	59,42	21,09	59,42	0,7852	2,6403	49,64	4,006E-06	0,01
P ₂	108,53	28,49	108,53	0,7852	3,7705	100,06	3,07E-06	0,01
P ₃	151,37	34,47	151,37	0,7852	4,6825	148,57	2,47E-06	0,01
P ₄	182,18	41,70	182,18	0,7852	5,6679	158,65	2,42E-06	0,005
P ₅	210,75	49,94	210,76	0,7852	7,0773	181,72	2,36E-06	0,01
P ₆	238,94	58,72	238,93	0,7852	8,2630	203,06	2,48E-06	0,01
P ₇	263,33	66,67	263,33	0,7852	10,0966	233,80	2,76E-06	0,01
P ₈	291,64	76,47	291,64	0,7852	14,9302	257,01	8,438E-06	0,01

*As medidas de simetria e de curtose foram positivas (P <0,05); ** Teste de Kolmogorov Smirnov.