



## Controle de consumo bovino com intervalos de tolerância via modelos não-paramétricos

Vinícius Basseto Félix<sup>1</sup>, Fabrícia Almeida Garcia<sup>2</sup> e Larissa Bueno Fernandes<sup>3</sup>

<sup>1</sup> Mestrando em Bioestatística - Universidade Estadual de Maringá - elix\_prot@hotmail.com

<sup>2</sup> Consultora de inteligência de mercado em Zoocenia - Gestão Agropecuária

<sup>3</sup> Mestranda em Bioestatística - Universidade Estadual de Maringá

### RESUMO

Este trabalho objetiva a criação de um método que permita tanto um ajuste flexível e rápido, como um tipo de intervalo que atenda ao objetivo de controlar o consumo animal, dado sua ingestão de matéria seca ao longo dos dias de confinamento. Para tal, utilizou-se de um método não-paramétrico e intervalo de tolerância provendo uma solução ágil e com validação biológica de especialistas.

**Palavras chave:** Tolerância; Não-paramétrico, Consumo animal.

## 1 INTRODUÇÃO

Em um contexto de produção bovina, o consumo animal é definido por um nutricionista, sendo que esta ingestão ao longo dos dias de confinamento pode ser apresentada por uma curva de consumo. Porém, há a carência de limites que permitam a indicação de valores que estejam fugindo ao comportamento esperado.

Para compreensão deste consumo é possível obter estimativas pontuais através de um modelo, que são nada mais do que valores singulares. Desta forma é interessante a obtenção de estimativas intervalares, tal que produzam um limite inferior e superior, em que o intervalo obtido é provável de conter a característica de interesse da população (parâmetro). Na Estatística existem algumas formas de intervalos, em que o de confiança é o mais popular, porém o uso de cada um depende do tipo de análise.

- **Intervalos de confiança:**

Este intervalo parte da amostra, e produz limites tal que o valor do parâmetro de interesse esteja provavelmente contido no mesmo. Caso diferentes amostras aleatórias fossem retiradas de uma mesma população cada uma proveria um intervalo de confiança. Dado isto, uma certa proporção destes intervalos conteria o valor verdadeiro do parâmetro, e este percentual é o nível de confiança.

- **Intervalos de predição:**

Após o ajuste de um modelo estatístico pode-se obter intervalos de predição. Duas possibilidades comuns são: Intervalo de confiança da predição: alcance que provavelmente contém o valor médio da variável resposta dado especificações das variáveis explicativas. Diferente do intervalo de confiança aqui define-se uma população especificada pelos valores das variáveis explicativas; Intervalo de predição: alcance que provavelmente contém o valor da variável resposta para uma nova observação.

- **Intervalos de tolerância:**

Por fim, os intervalos de tolerância consistem no alcance dos valores que cobrem  $P\%$  da população, ou seja, contém uma proporção específica da população. Logo, diferente dos demais, além de um nível de confiança  $[100 \times (1 - \alpha)\%]$  é necessário também a especificação de uma proporção.

Porém há um desafio, a maioria destes intervalos baseiam-se em modelos paramétricos, ou seja, possuem pressuposto de seguir uma determinada distribuição de probabilidade. Isto torna-se um problema, pois para o consumo animal é necessário o ajuste de dezenas ou até centenas modelos em tempo real, assim necessita-se de uma solução global, rápida e automática, aliviando a precisão garantida por uma solução paramétrica bem ajustada. Assim, parte-se para uma vertente não-paramétrica.

## 2 METODOLOGIA

### 2.1 Modelo

A modelagem não-paramétrica consiste em inúmeros métodos, alguns dos mais populares são:

- Regressão polinomial local, conhecida como LOWESS (*locally weighted scatterplot smoother*) para regressão simples, ou LOESS para o caso mais geral [1];
- *Splines* [2];
- Regressão não-paramétrica aditiva, ou ainda, modelos *gam* [4].

Neste trabalho será apresentado o uso do método de LOESS, visto que o mesmo apresentou maior flexibilidade de ajuste e melhor resultado que os demais, em testes omitidas deste trabalho.

Esta técnica implica no ajuste de polinômios locais em subconjuntos dos dados, determinados pelo algoritmo do vizinho mais próximo. Tal que se define um parâmetro de suavização, em que este indica a porção de dados a ser utilizados nos ajustes locais. Para estimação utiliza-se do método mínimos quadrados ponderados.

### 2.2 Intervalo de tolerância

Considere uma amostra aleatória  $X$ , tal que  $F_X$ , a função acumulada, é uma função contínua e não-decrescente. Os limites inferior e superior do intervalo de tolerância  $[100 \times (1-\alpha)\%]/[100 \times P\%]$  são dados respectivamente por  $L = x_{(r)}$  e  $U = x_{(s)}$ . Em que  $x_{(j)}$  é o  $j$ -ésimo valor da amostra ordenada ( $x_{(\cdot)}$ ). Vários métodos podem ser utilizados para definição destes limites, alguns deles sendo [7]:

- Método de *Wilks*: método que se baseia na distribuição beta [6];
- Método de *Wald*: semelhante ao método de Wilks para intervalos unilaterais, porém para o caso bilateral produz todos os intervalos possíveis [5];
- Método de *Hahn Meeker*: método que se baseia na distribuição binomial [3].

Neste trabalho utiliza-se o método de *Wilks*.

Para criação de um intervalo para um modelo, utiliza-se da relação não-linear entre a variável resposta  $Y$  dado um conjunto de variáveis explicativas  $X_1, \dots, X_p$ . Pode-se então definir este modelo como

$$Y = f(X_1, \dots, X_p) + \varepsilon, \quad (1)$$

em que  $\varepsilon$  é o termo de erro, assumido com média nula. Após um ajuste, obtém-se

$$y_i = \hat{y}_i + \hat{\varepsilon}_i, \quad (2)$$

em que  $\hat{\varepsilon}_i$  são os resíduos e  $\hat{y}_i$  são os valores ajustados, para  $i = 1, \dots, n$ . Assim, os resíduos podem ser utilizados para construção dos limites inferior e superior de tolerância, que são dados respectivamente por  $L = \hat{y}_i + \hat{\varepsilon}_{(r)}$  e  $U = \hat{y}_i + \hat{\varepsilon}_{(s)}$ , em que se obtém as posições  $r$  e  $s$  via método de *Wilks*.

### 2.3 Dados

Para aplicação, utilizou-se um banco de dados com 5061 observações, com dados de 29/03/2017 a 10/09/2017, em que este apresenta:

- Lote: variável categórica ordinal com 71 níveis (10.020 a 10.588);
- Raça: variável categórica nominal com 5 níveis (Nelore, Composto, Girolando, Redangus e Aberdeenagus);
- Categoria: variável categórica nominal com 2 níveis (Boi Inteiro e Vaca);
- Tipo de ração: variável categórica ordinal com 3 níveis (Adaptação, Crescimento e Terminação);
- Dias de confinamento: variável numérica discreta;
- Índice de Matéria Seca: variável resposta numérica contínua .

### 3 RESULTADOS E DISCUSSÕES

Foram calculados intervalos de tolerância com 95% de confiança e diversos valores de proporção, tal que o valor de proporção foi definido após avaliação de centenas de modelos e testes, de forma que os especialistas da área zootécnica avaliaram aqueles que melhor condiziam com a realidade biológica e a variabilidade presente nos animais.

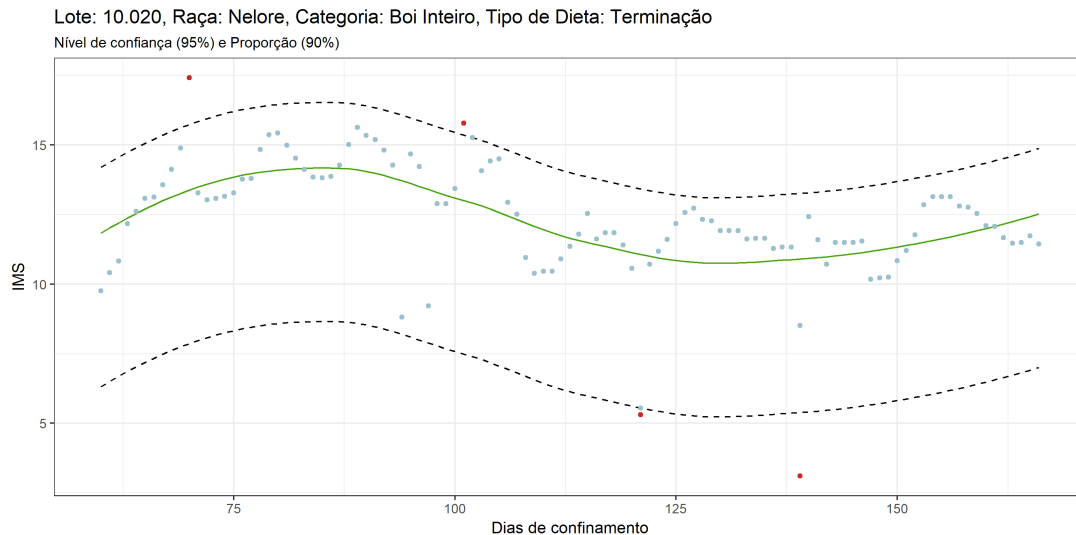


Figura 1: Intervalo de tolerância para combinação especificada, com 95% de confiança e 90% de proporção.

A Figura 1 apresenta o ajuste (linha verde) e o respectivo intervalo de tolerância (linha preta tracejada), em que os pontos vermelhos representam as observações que fugiram os limites calculados.

### 4 CONCLUSÃO

O intervalo de tolerância aliado ao método de ajuste LOESS provê uma solução computacionalmente rápida e global para implementação em um sistema de *Business Intelligence*, sendo flexível o suficiente para adequação das diferentes curvas de consumo, bem como ajustes em tempo real.

### 5 AGRADECIMENTOS

Agradecemos a empresa Gestão Agropecuária por ceder os dados.

### Referências

- [1] COHN, D. A.; GHARAMANI, Z.; JORDAN, M. I. Active learning with statistical models. *Journal of artificial intelligence research*, 1996.
- [2] DE BOOR, C. et al. *A practical guide to splines*. New York: Springer-Verlag, 1978.
- [3] HAHN, G. J.; MEEKER, W. Q. *Statistical intervals: a guide for practitioners*. John Wiley & Sons, 2011.
- [4] LEATHWICK, J. R.; ELITH, J.; HASTIE, T. Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. *Ecological modelling*, v. 199, n. 2, p. 188-196, 2006.
- [5] WALD, A. An extension of Wilk's method for setting tolerance limits. *The Annals of Mathematical Statistics*, v. 14, n. 1, p. 45-55, 1943.
- [6] WILKS, S. S. Determination of sample sizes for setting tolerance limits. *The Annals of Mathematical Statistics*, v. 12, n. 1, p. 91-96, 1941.
- [7] YOUNG, D. S. et al. Tolerance: an R package for estimating tolerance intervals. *Journal of Statistical Software*, v. 36, n. 5, p. 1-39, 2010.