

# NÚMERO DE COLETAS DA PRODUÇÃO DE LEITE NA PREDIÇÃO DE MODELOS PARA ESTIMATIVA DA PRODUÇÃO MÉDIA DIÁRIA DE LEITE DE VACAS NELORE

PAULO SANTANA PACHECO,<sup>1</sup> JOÃO RESTLE,<sup>2</sup> IVAN LUIZ BRONDANI,<sup>3</sup> JOSÉ HENRIQUE SOUZA DA SILVA,<sup>3</sup>  
FABRÍCIA ROCHA CHAVES MIOTTO<sup>4</sup> E JOSÉ NEUMAN MIRANDA NEIVA<sup>5</sup>

1. Pós-doutorando do PPGCA/UFG, bolsista PDJ-CNPq. E-mail: pspacheco@pop.com.br

2. Departamento de Produção Animal/EV/UFG

3. Departamento de Zootecnia/UFSM

4. Doutoranda do PPGCA/EV/UFG, bolsista CNPq

5. Professor UFTO

## RESUMO

Objetivou-se com este experimento avaliar modelos com diferentes números de coletas da produção de leite para estimativa da produção média diária de leite em vacas Nelore até os 217 dias de lactação. As oito avaliações da produção de leite foram realizadas em 29 vacas, através do método direto, aos 14, 42, 70, 98, 126, 154, 182 dias e ao desmame (217 dias). Para seleção dos modelos de regressão linear múltipla, objetivando estimar a produção média diária de leite de acordo com o número de coletas de leite, utilizou-se o procedimento Stepwise do SAS. Procedeu-se à análise residual através dos testes de heterocedasticidade da variância (estatística  $\chi^2$ ), normalidade (estatística W de Shapiro-Wilk), a diagnóstico de observações influentes e a diagnóstico de multicolinearidade. Para a validação dos modelos de regressão selecionados, utilizou-se a estatística

predição da soma de quadrados do erro (PRESS). Os dias da coleta de leite selecionados foram: 98 para 1 coleta; 70 e 98 para 2 coletas; 14, 70 e 98 para 3 coletas; 14, 70, 98 e 154 para 4 coletas; 14, 70, 98, 154 e 182 para 5 coletas; 14, 42, 70, 98, 154 e 182 para 6 coletas e 14, 42, 70, 98, 126, 154 e 182 para 7 coletas. Os  $R^2$  ajustados para modelos com 1; 2; 3; 4; 5; 6 e 7 coletas foram, respectivamente, 0,7900; 0,8718; 0,9064; 0,9596; 0,9698; 0,9875 e 0,9972. Todos os modelos preditos foram considerados satisfatórios na estimativa da produção média diária de leite. Quando houver a possibilidade de executar diversas amostragens ao longo do período de lactação, três coletas em datas estrategicamente escolhidas ao longo do período de lactação são suficientes para predizer com elevada precisão a produção média diária de leite.

**PALAVRAS-CHAVE:** Análise residual, fase de cria, lactação, regressão linear múltipla, seleção de variáveis independentes, técnicas de validação de modelos de regressão.

## ABSTRACT

### NUMBER OF COLLECTIONS OF MILK YIELD ON PREDICTION OF MODELS TO ESTIMATE AVERAGE DAILY MILK YIELD OF NELLORE BEEF COWS

The objective of the experiment was to evaluate models with different number of collections of milk yield to estimate average daily milk yield in Nelore beef cows until 217 days of lactation. The eight evaluations of milk yield were taken from 29 cows, by direct method through

milking, at 14, 42, 70, 98, 126, 154, 182 days and weaning (217 days). For the selection of multiple linear regression models, with objective of estimate the average daily milk yield according to number of milk collects, Stepwise procedure of SAS was used. The data were submitted to

residual analysis by tests of heterocedasticity of variance ( $\chi^2$  statistic), normality (W statistic of Shapiro-Wilk) and diagnosis of outliers (three observations were excluded from original  $n=30$ ), beyond multicollinearity diagnosis. For validation of the selected regression models, the prediction error sum of squares statistic (PRESS), was used. The selected days of milk yield collections were: 98 for 1 collection; 70 and 98 for 2 collections; 14, 70, and 98 for 3 collections; 14, 70, 98 and 154 for 4 collections; 14, 70, 98, 154 and 182 for 5 collections; 14, 42, 70, 98, 154 and

182 days for 6 collections and 14, 42, 70, 98, 126, 154 and 182 days for 7 collections. The adjusted  $R^2$  for models with 1; 2; 3; 4; 5; 6 and 7 collections were, respectively, 0.7900; 0.8718; 0.9064; 0.9596; 0.9698; 0.9875 and 0.9972. All the predicted models were satisfactory to estimate average daily milk yield. In a possibility of to execute various collections of milk yield during lactation period, three milk collections chosen in strategic dates during lactation period were sufficient to estimate with high precision the average daily milk yield.

**KEY WORDS:** Lactation, maternal phase, multiple linear regression, selection of independent variables, residual analysis, techniques of validation of regression models.

## INTRODUÇÃO

Analisando o sistema de cria em bovinos de corte, é incontestável a importância da produção de leite e dos componentes do leite no desenvolvimento dos(as) bezerros(as) até o desmame, conforme demonstrado nos estudos de ALENCAR (1987), RIBEIRO & RESTLE (1991) e RESTLE et al. (2005).

Experimentalmente, são realizadas diversas amostragens da produção de leite em diferentes fases da lactação, objetivando compreender melhor o comportamento da produção e composição do leite ao longo do período de lactação e sua relação com o desempenho das vacas e dos bezerros.

PACHECO et al. (2009) comentaram que em vários estudos relacionados com a produção de leite de vacas de corte, desenvolvidos nas mais variadas situações ambientais e com diversos grupos genéticos, o número ou frequência de amostragens foi determinado previamente, ocasionando diferentes intervalos entre as amostragens. Esses intervalos de amostragens podem ser a cada 21 dias (CERDÓTES et al., 2004), a cada 28 dias (RIBEIRO et al., 1991; RESTLE et al., 2003), a cada 30 dias (HOLLOWAY et al., 1985; ALENCAR et al., 1988; ALBUQUERQUE et al., 1993; BROWN & BROWN, 2002), a cada 50 dias (FREETLY & CUNDIFF, 1998), a cada 90 dias (ALENCAR et al., 1985; ALENCAR, 1987; ALENCAR et al., 1992) ou a cada 100 dias (JOHNSON et al., 2003).

Contudo, existem várias dificuldades enfrentadas na coleta de tais amostras e consequências

desfavoráveis para os animais, como o aumento considerável do estresse em virtude do pouco ou quase inexistente manejo em vacas de corte, se comparadas com vacas de leite, ocasionando dificuldade na realização da ordenha quando se utiliza o método direto; desgaste físico de vacas e bezerros quando são percorridas distâncias razoáveis dos poteiros de pastejo até o centro de manejo e vice-versa; aumento do estresse dos bezerros por causa da separação de suas mães conforme as metodologias empregadas; frequência acentuada das avaliações em rebanhos que utilizam o desmame convencional, necessitando considerável quantidade de tempo e mão de obra para a realização das amostragens.

Objetivou-se, com este estudo, avaliar modelos com diferentes números de coletas da produção de leite para estimativa da produção média diária de leite de vacas Nelore até os 217 dias de lactação.

## MATERIAL E MÉTODOS

O experimento foi desenvolvido no Setor de Bovinocultura de Corte do Departamento de Zootecnia da Universidade Federal de Santa Maria, no município de Santa Maria, RS, localizado na região fisiográfica denominada Depressão Central, a 153 m de altitude, apresentando, segundo classificação de Köppen, clima subtropical úmido (cfa) (MORENO, 1961). A temperatura média anual varia de 14,3 a 25,2 °C, com média das mínimas de 9,7 °C no mês de agosto e média das máximas de 29,9 °C no mês de janeiro, sendo

que geadas podem ocorrer de abril a outubro. A média anual da umidade relativa do ar é de 73% e a precipitação é de 1650,9 mm (RIO GRANDE DO SUL, 1979).

Utilizaram-se 29 vacas da raça Nelore, com idades variando de três a treze anos, tomadas ao acaso do mesmo rebanho experimental e submetidas às mesmas condições de manejo no período anterior ao início do experimento.

No período pós-parto, vacas e bezerros(as) foram submetidos a dois tipos de pastagem: pastagem nativa do parto ao desmame dos bezerros, com lotação de 1,0 vaca mais bezerro(a) ou pastagem cultivada de inverno composta de aveia (*Avena sativa*) mais azevém (*Lolium multiflorum*) do parto até o final do mês de novembro e após em pastagem de capim-papuã (*Brachiaria plantaginea*), até o desmame dos bezerros, com lotação de 1,76 vacas mais bezerros(as).

O período de parição foi de setembro a dezembro, sendo os(as) bezerros(as) obtidos(as) de inseminação artificial (período de 45 dias) ou de monta natural (45 dias seguintes). Durante o período de inseminação, foi utilizado sêmen de oito touros da raça Charolesa e de oito da raça Nelore e, para o repasse em monta natural, quatro touros Charolês e quatro touros Nelore.

As avaliações da produção de leite foram realizadas aos 14, 42, 70, 98, 126, 154, 182 dias e ao desmame. O desmame ocorreu com idade média dos bezerros de 217 dias. O método utilizado para determinar a produção de leite foi o direto, com ordenha manual, segundo metodologia descrita por RIBEIRO et al. (1991) e RESTLE et al. (2003). As vacas eram separadas de seus bezerros pela manhã (um dia antes da ordenha) e soltas em piquete anexo. Por volta das 18 horas, eram recolhidas novamente até o centro de manejo, permanecendo com o bezerro por trinta minutos, tempo suficiente para esgotar o úbere. Após mamarem, os bezerros eram separados novamente das vacas e ficavam contidos no centro de manejo, sendo as vacas soltas em piquete com pasto e água. No dia seguinte, às 6 horas, iniciava-se a ordenha. Para realização da ordenha, a vaca era contida pelo pescoço e a pata traseira esquerda era amarrada com corda, recebendo, então, 3 U.I. de ocitocina

via intramuscular, a fim de estimular a liberação do leite. A ordenha foi realizada, após o úbere ser lavado e massageado, apenas nos dois quartos do lado direito, com duração aproximada de cinco minutos. A produção obtida era multiplicada por dois, para se obter a produção do úbere, sendo também corrigida para 24 horas, pela seguinte fórmula:

$$\text{Produção leite (l/dia)} = \frac{\text{produção [L] dos dois quartos do úbere} * 2 * 60 \text{min} * 24 \text{h}}{\text{tempo [min] entre a última mamada e a ordenha}}$$

Inicialmente, os dados foram ajustados através de um modelo matemático considerando-se os efeitos de sexo do bezerro, a composição racial do(a) bezerro(a) (puros ou cruzados FL Nelore x Charolês), tipo de pastagem, e as covariáveis ordem de parição e idade da vaca (anos), utilizando-se o procedimento GLM (SAS, 1997).

Procedeu-se à análise da matriz de correlação de Pearson entre a produção média diária de leite e a produção de cada dia da coleta, através do procedimento CORR (SAS, 1997).

Para seleção dos modelos de regressão objetivando estimar a produção média diária de leite de acordo com o número de coletas de leite, optou-se pelo procedimento STEPWISE (SAS, 1997), considerando  $P < 0,05$  a probabilidade da estatística F para entrada e permanência de cada variável preditora no modelo.

Procedeu-se à análise residual através dos testes de heterocedasticidade da variância (estatística  $\chi^2$ ) pela opção SPEC do procedimento REG (SAS, 1997) e normalidade (estatística W de Shapiro-Wilk) pela opção NORMAL do procedimento UNIVARIATE (SAS, 1997).

Foi realizado diagnóstico de observações influentes ou *outliers* nas estimativas dos parâmetros das variáveis preditoras, através da análise da estatística do resíduo padronizado de Student (RSTUDENT), obtida pela opção INFLUENCE do procedimento REG (SAS, 1997). Consideraram-se observações com RSTUDENT superior a 2, em valores absolutos, como influentes.

O diagnóstico de multicolinearidade entre os coeficientes de regressão utilizados para estimar a produção média diária de leite, representados pelos dias de coleta de leite, foi realizado conforme su-

gestões de FREUND & LITTEL (1991) e SOUZA (1998), através da análise da matriz de correlação de Pearson, calculada através do procedimento CORR (SAS, 1997), e das medidas fator de inflação da variância (FIV), autovalor ( $\lambda$ ), índice de condição (IC) e proporção da variância associada com cada autovalor. As medidas de multicolinearidade foram obtidas pelas opções VIF e COLLINOINT do procedimento REG (SAS, 1997).

Para a validação dos modelos de regressão múltipla selecionados, utilizou-se a estatística predição da soma de quadrados do erro (PRESS) no cálculo de um  $R^2$  aproximado para validação ou predição do modelo proposto. Optou-se por essa alternativa por considerar observações do próprio conjunto de dados para seu cálculo. A estatística PRESS foi obtida pela opção PRESS do procedimento REG (SAS, 1997).

Informações sobre a teoria e estimação das estatísticas utilizadas na análise residual, nas

observações influentes, na multicolinearidade e na validação de modelos podem ser obtidas em MONTGOMERY & PECK (1982), FREUND & LITTEL (1991), SAS (1997), DRAPER & SMITH (1998) e SOUZA (1998).

## RESULTADOS E DISCUSSÃO

Na Tabela 1 constam as médias para produção média diária de leite ao longo do período de lactação e a matriz contendo os coeficientes de correlação de Pearson entre as diferentes datas de coleta de leite.

Do parto aos 217 dias, a produção média diária de leite foi de 4,4 L/dia, valor superior ao relatado por ALBUQUERQUE et al. (1993), de 3,05 L/dia em 207 dias de lactação, e por CRUZ et al. (1997), de 3,78 L/dia em 238 dias de lactação, ambos os estudos avaliando a raça Nelore.

**TABELA 1.** Médias e desvios-padrão (dp), e coeficientes de correlação de Pearson entre as produções médias diárias de leite em vacas Nelore

		Produção média diária de leite, L (dp) <sup>1</sup>							
	0 - 217 dias	14	42	70	98	126	154	182	217
	4,4	4,7	5,4	5,2	5,0	4,8	3,7	2,8	2,1
	(1,2)	(1,6)	(1,8)	(2,0)	(1,9)	(2,0)	(1,6)	(1,6)	(1,1)
		Coeficientes de correlação de Pearson							
14	0,62**								
42	0,70**	0,65**							
70	0,78**	0,47**	0,48**						
98	0,89**	0,43**	0,59**	0,63**					
126	0,81**	0,23	0,44**	0,55**	0,79**				
154	0,77**	0,20	0,36	0,54**	0,67**	0,82**			
182	0,57**	0,12	0,07	0,35	0,49**	0,43**	0,53**		
217	0,27	-0,10	-0,01	0,15	0,29	0,17	0,18	0,53**	

<sup>1</sup> 14, 42, 70, 98, 126, 154, 182 e 217 são os dias da coleta da produção de leite.

\*\* P<0,01; \* P<0,05.

De acordo com a matriz de correlação de Pearson, as coletas de leite nas diferentes datas correlacionaram-se significativamente (P<0,05) com a produção média diária de toda a lactação (do parto aos 217 dias) dos 14 aos 182 dias, sendo os coeficientes de correlação de elevada magnitude,

variando de 0,57 a 0,89. Além disso, entre as datas de coletas de leite, os coeficientes de correlação de maior magnitude foram verificados com as respectivas datas subsequentes, ou seja, entre o dia 14 com o 42, o dia 42 com o 70, o dia 70 com o 98 e assim por diante. Avaliando a lactação em

vacas Charolês, nas mesmas datas de coleta da produção de leite do presente estudo, PACHECO et al. (2009) verificaram que, dos 14 aos 182 dias, os coeficientes de correlação variaram de 0,68 a 0,85, e a data de amostragem que apresentou maior coeficiente de correlação foi aos 154 dias.

Na Tabela 2 estão apresentadas as estatísticas utilizadas na escolha dos modelos de regressão para estimar a produção média diária de leite em 217 dias de lactação, de acordo com o procedimento Stepwise.

**TABELA 2.** Resumo das estatísticas utilizadas na escolha dos modelos para estimar a produção média diária de leite em vacas Nelore, de acordo com o procedimento Stepwise

Passo (Step)	Dia da coleta da produção de leite incluída no modelo	Número de cole- tas da produção de leite incluídas no modelo	R <sup>2</sup> parcial	R <sup>2</sup>	R <sup>2</sup> ajustado <sup>1</sup>	F	Pr>F
1	98	1	0,7975	0,7975	0,7900	106,3	<,0001
2	70	2	0,0835	0,8810	0,8718	18,3	0,0002
3	14	3	0,0354	0,9164	0,9064	10,6	0,0033
4	154	4	0,0490	0,9654	0,9596	34,0	<,0001
5	182	5	0,0098	0,9752	0,9698	9,1	0,0062
6	42	6	0,0150	0,9901	0,9875	33,4	<,0001
7	126	7	0,0078	0,9979	0,9972	77,8	<,0001
8	217	8	0,0021	1,0000	1,0000	infy	<,0001

<sup>1</sup> R<sup>2</sup> ajustado =  $1 - \frac{(n-1) * (1 - R^2)}{n-p}$ , em que *i* é igual a 1, pois o intercepto está incluído no modelo; *n* é o número de observações utilizadas para estimar o modelo e *p* é o número de parâmetros do modelo, incluindo intercepto.

Neste procedimento, o valor da estatística F parcial é utilizado para inclusão e permanência das variáveis no modelo. As variáveis a serem utilizadas no modelo de regressão são adicionadas uma a uma, considerando o maior valor da estatística F parcial (P<0,05). Para cada nova variável adicionada, o procedimento Stepwise verifica todas as variáveis já incluídas no modelo e exclui a que não apresenta valor F parcial suficientemente significativo (P<0,05) para permanecer no modelo (SAS, 1997).

Em virtude do aumento do R<sup>2</sup> com a inclusão de novos parâmetros no modelo, mesmo que estes não tenham relação com a variável dependente, o R<sup>2</sup> ajustado é indicado em casos de regressão linear múltipla, visando corrigir este problema (SAS, 1997).

Verifica-se que o coeficiente de determinação (R<sup>2</sup>) parcial apresentou maiores incrementos até a quarta coleta de leite incluída no modelo. Considerando o R<sup>2</sup> ajustado, com uma coleta de

leite apenas, realizada no 98° dia de lactação, o modelo estimado explicou 79% da variação da produção média diária de leite até os 217 dias de lactação. Com duas coletas de leite, incluindo o 70° dia de lactação, o modelo estimado explicou 87,18%. Com três coletas, adicionado o 14° dia de lactação, 90,64% e, com quatro coletas de leite, incluindo o 154° dia de lactação, 95,96% da variação na produção média diária de leite foi explicada pelo modelo estimado pelo procedimento Stepwise. PACHECO et al. (2009), trabalhando com vacas Charolês, verificaram incrementos menos expressivos nos valores dos coeficientes de determinação (R<sup>2</sup>) parciais a partir da quarta amostragem de leite incluída no modelo. Com uma (no 154° dia de lactação), duas (70° e 154° dias de lactação) e três (14°, 70° e 154° dias de lactação) amostragens de leite, os autores relataram valores de R<sup>2</sup> ajustado de 0,7191; 0,8711 e 0,9663, respectivamente.

Após o procedimento de escolha dos modelos de regressão para estimar a produção média diária de leite, procedeu-se a diagnósticos em regressão, através de técnicas formais e visuais para detectar desvios das hipóteses clássicas do modelo linear. Este diagnóstico passa por duas etapas: análise residual e análise de multicolinearidade. Conforme SOUZA (1998), esses desvios avaliados na análise residual seriam os resíduos que não apresentam variância constante, com erros de especificação na forma funcional da resposta, com variáveis residuais não independentes, com extremos ocorrendo nos valores das variáveis independentes (observações influentes) e com variáveis residuais que não têm distribuição comum normal.

Na análise residual formal apresentada na Tabela 3, avaliou-se a inconstância da variância dos resíduos pelo teste de heterocedasticidade através da estatística Qui-quadrado ( $\chi^2$ ) e a normalidade dos resíduos pela estatística W de Shapiro-Wilk, sendo este último teste indicado para amostras relativamente pequenas ( $n \leq 2000$ ), conforme SOUZA (1998). Em ambas as análises, as probabilidades das respectivas estatísticas não foram significativas, indicando que os diferentes modelos selecionados de acordo com o número de coletas de leite atendem às pressuposições de resíduos com constância da variância e normalmente distribuídos.

**TABELA 3.** Resumo da análise de heterocedasticidade e de normalidade dos resíduos dos modelos selecionados para estimativa da produção média diária de leite em vacas Nelore, de acordo com o número de coletas da produção de leite

Estatísticas <sup>1</sup>	Número de coletas da produção de leite							
	1	2	3	4	5	6	7	8
$\chi^2$	2,25	4,05	7,47	10,78	19,37	27,75	28,96	-
Pr > $\chi^2$	0,3249	0,5426	0,5879	0,7032	0,4980	0,4238	0,5197	-
W	0,9644	0,9682	0,9675	0,9711	0,9668	0,9701	0,9581	0,9510
Pr < W	0,4197	0,5130	0,4949	0,5893	0,4777	0,5644	0,2947	0,1949

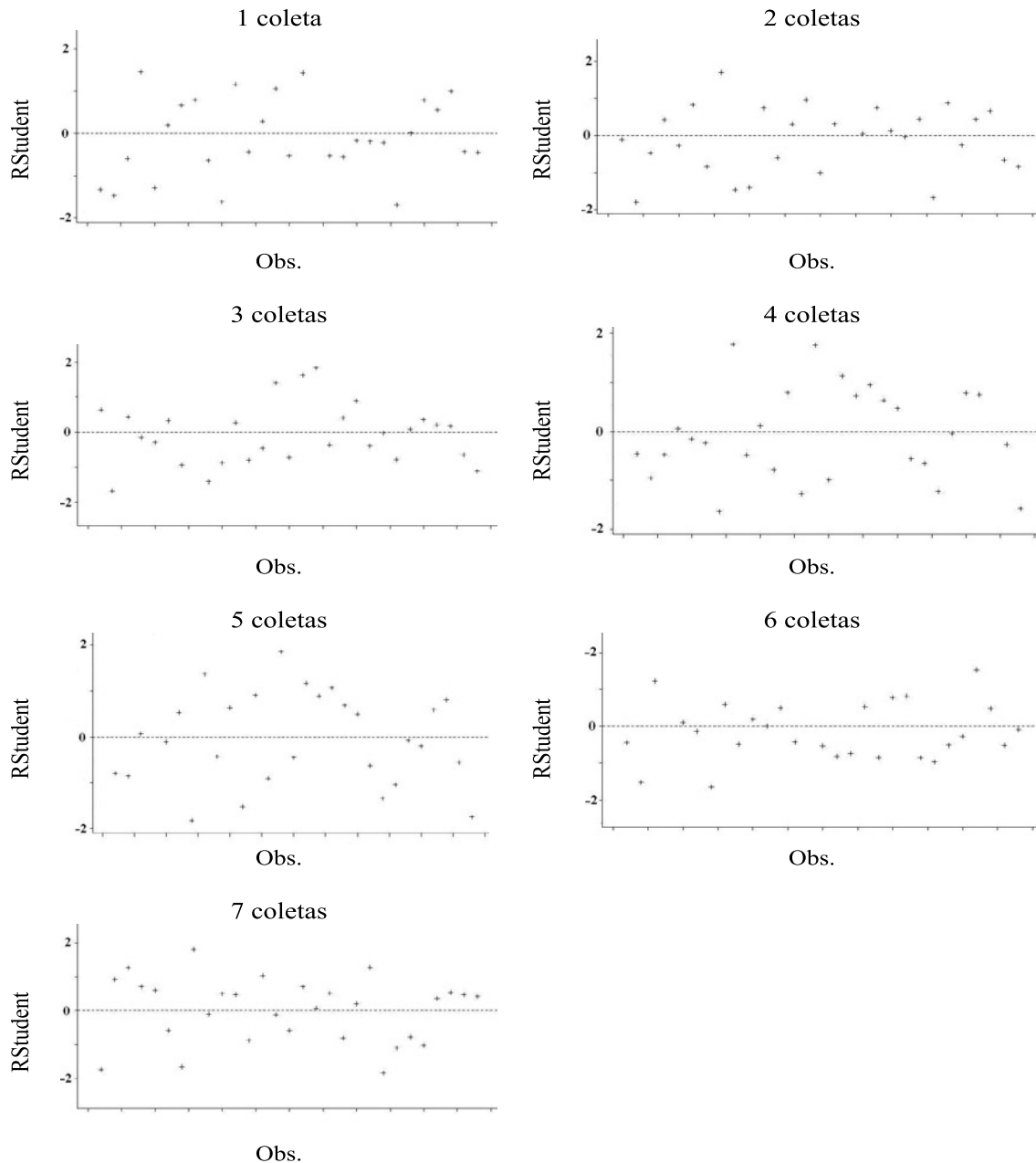
<sup>1</sup>  $\chi^2$  = valor da estatística Qui-quadrado (análise de heterocedasticidade dos resíduos); W = valor da estatística W de Shapiro-Wilk (análise de normalidade dos resíduos).

Na Figura 1 consta a análise residual visual de observações influentes ou *outliers*, para cada modelo selecionado de acordo com o número de coletas de leite, utilizando-se o resíduo padronizado de Student (RStudent) e os valores observados da produção média diária de leite.

Conforme MONTGOMERY & PECK (1982), uma observação influente é uma observação extrema, ou seja, observação atípica do restante dos dados, apresentando elevado desvio-padrão da média. Falhas na coleta dos dados e nos equipamentos de medida são exemplos de eventos que resultam em observações influentes. SOUZA (1998) comenta que, de um modo geral, para modelo linear, todas as estatísticas geradas pelo método de quadrados mínimos podem ser

substancialmente afetadas ou influenciadas por umas poucas observações. Desse modo, é muito importante a identificação de tais observações influentes e a avaliação de seus impactos.

Dentre as várias medidas de avaliação de influência, utilizou-se o resíduo padronizado de Student (RStudent), considerando observações possivelmente influentes aquelas com valores absolutos maiores que 2 para esta estatística (SAS, 1997). De acordo com a Figura 1, esse valor classificatório não foi verificado para nenhuma observação em todos os modelos selecionados de acordo com o número de coletas de leite, não necessitando, portanto, do uso de medidas corretivas ou de exclusão de observações.



**FIGURA 1.** Análise gráfica de observações influentes dos modelos selecionados, considerando resíduo de Student (RStudent) e valores observados da produção média diária de leite (Obs.), de acordo com o número de coletas da produção de leite em vacas Nelore.

Em relação ao diagnóstico de multicolinearidade, uma primeira etapa de avaliação é feita pela análise dos coeficientes de correlação entre as variáveis preditoras dos modelos. Se houver coeficientes elevados, como verificado na Tabela 1, pode ser indicativo de algum grau de confundimento entre essas variáveis preditoras. No

entanto, a severidade da multicolinearidade não deve ser quantificada somente pela magnitude da correlação, devendo ser utilizada como critério norteador e não decisivo no diagnóstico de multicolinearidade.

Diagnóstico de multicolinearidade mais preciso consta na Tabela 4, realizado através das

medidas de fator de inflação da variância (FIV), autovalor ( $\lambda$ ), índice de condição (IC) e proporção da variância associada aos autovalores. O diagnóstico foi ajustado para o intercepto, seguindo a

recomendação de FREUND & LITTEL (1991), visando reduzir de modo substancial os valores dos índices de condição (IC) e desprezar os efeitos da multicolinearidade nas estatísticas associadas.

**TABELA 4.** Diagnóstico de multicolinearidade (intercepto ajustado) <sup>1</sup>

Dia da coleta da produção de leite	$\lambda$	IC	Proporção da variação associada aos autovalores								
			Dia da coleta da produção de leite								
			14	42	70	98	126	154	182	217	
14	4,01	1,00	0,01	0,01	0,02	0,01	0,01	0,01	0,01	0,01	3*10 <sup>-3</sup>
42	1,59	1,59	0,07	0,05	5*10 <sup>-3</sup>	6*10 <sup>-8</sup>	7*10 <sup>-4</sup>	3*10 <sup>-3</sup>	0,06	0,06	0,12
70	0,86	2,16	0,11	0,02	2*10 <sup>-3</sup>	1*10 <sup>-4</sup>	0,04	0,06	0,02	0,02	0,24
98	0,51	2,81	0,09	0,18	0,17	0,02	0,02	3*10 <sup>-3</sup>	0,21	0,21	0,21
126	0,44	3,01	0,10	0,01	0,75	9*10 <sup>-4</sup>	5*10 <sup>-4</sup>	0,01	0,18	0,18	0,06
154	0,25	4,04	0,15	0,54	0,03	0,30	0,06	0,17	0,06	0,06	3*10 <sup>-5</sup>
182	0,22	4,24	0,45	0,15	0,02	0,16	0,01	0,28	0,34	0,34	0,37
217	0,12	5,77	0,01	0,03	0,01	0,50	0,85	0,45	0,12	0,12	1*10 <sup>-3</sup>
FIV			2,12	2,42	1,94	4,34	4,81	3,69	2,22	2,22	1,60

<sup>1</sup>  $\lambda$  = autovalor;

IC = índice de condição =  $IC_i = \sqrt{\frac{\lambda_{\text{máximo}}}{\lambda_i}}$  ; ;

FIV = fator de inflação da variância =  $FIV_i = \frac{1}{(1 - R^2)}$

Conforme SOUZA (1998), a multicolinearidade está intrinsecamente associada ao nível de condicionamento dos dados. Em geral, o mau condicionamento é causado pela escolha do modelo de regressão quando variáveis independentes altamente correlacionadas são incluídas na especificação da superfície de resposta do experimento (MONTGOMERY & PECK, 1982; SOUZA, 1998). A multicolinearidade pode ser um problema e vir a necessitar ação corretiva quando o interesse inferencial concentra-se nas avaliações dos impactos dos coeficientes individuais  $\beta_j$  na resposta (SOUZA, 1998), como é o caso dos modelos que utilizam regressão múltipla.

SOUZA (1998) relata que, em presença de multicolinearidade, as variâncias das estimativas de coeficientes preditores podem crescer exageradamente, mascarando a significância das variáveis utilizadas no modelo e limitando a aplicabilidade e a amplitude de previsão do modelo linear ajustado. Os sintomas de multicolinearidade podem

ser: presença de valores não significativos para as estatísticas de Student associada a cada um dos coeficientes de regressão, apesar de um valor alto para o coeficiente de determinação (SOUZA, 1998); sinais contrários aos esperados para os  $\beta_j$  (RODRÍGUEZ-ALMEIDA et al., 1997); coeficientes de correlação de Pearson elevados entre os coeficientes preditores (RODRÍGUEZ-ALMEIDA et al., 1997).

Conforme apresentado na Tabela 4, as medidas de multicolinearidade apresentaram valores aquém daqueles considerados críticos sugeridos por FREUND & LITTEL (1991), SAS (1997) e SOUZA (1998). De acordo com estes autores, são considerados valores críticos para FIV quando maior que 10, para  $\lambda$  quando menor que 0,01, para IC quando maior que 30 (entre 10 e 30 possível problema) e para proporção da variância associada com os autovalores quando maior que 0,80.

Na Tabela 5 constam os coeficientes de regressão dos modelos selecionados para estima-



tiva da produção média diária de leite em vacas Nelore, de acordo com o número de coletas de leite.

O resumo das estatísticas utilizadas na análise de validação dos modelos selecionados para

estimativa da produção média diária de leite em vacas Nelore, de acordo com o número de coletas de leite, está apresentado na Tabela 6.

**TABELA 5.** Coeficientes de regressão dos modelos selecionados para estimativa da produção média diária de leite em vacas Nelore, de acordo com o número de coletas da produção de leite

Número de coletas da produção de leite	Dia(s) da(s) coleta(s) da produção de leite	Coeficientes de regressão representando os dias de coletas da produção de leite									
		Intercepto	14	42	70	98	126	154	182	217	
1	98	1,436				0,588					
2	70, 98	1,022			0,230	0,433					
3	14, 70, 98	0,595	0,173		0,187	0,399					
4	14, 70, 98, 154	0,361	0,217		0,139	0,276		0,234			
5	14, 70, 98, 154, 182	0,298	0,225		0,136	0,254		0,207	0,096		
6	14, 42, 70, 98, 154, 182	0,155	0,157	0,131	0,131	0,200		0,184	0,136		
7	14, 42, 70, 98, 126, 154, 182	0,100	0,173	0,128	0,127	0,137	0,120	0,115	0,151		
8	14, 42, 70, 98, 126, 154, 182, 217	-8,18*10 <sup>-15</sup>	0,188	0,125	0,125	0,125	0,125	0,125	0,125	0,063	

**TABELA 6.** Resumo das estatísticas utilizadas na análise de validação dos modelos selecionados para estimativa da produção média diária de leite em vacas Nelore, de acordo com o número de coletas da produção de leite

Estatísticas	Número de coletas da produção de leite							
	1	2	3	4	5	6	7	8
GL <sub>Erro</sub> <sup>1</sup>	27	26	25	24	23	22	21	20
SQ <sub>Total</sub> <sup>2</sup>	46,24	46,24	46,24	46,24	46,24	46,24	46,24	46,24
PRESS <sup>3</sup>	10,96	6,45	5,23	2,37	2,09	0,87	0,20	1,19*10 <sup>-26</sup>
S <sub>PRESS</sub> <sup>4</sup>	0,12	0,10	0,09	0,06	0,06	0,04	0,02	5,45*10 <sup>-15</sup>
R <sup>2</sup> <sub>Validação</sub> <sup>5</sup>	0,7630	0,8605	0,8869	0,9487	0,9548	0,9812	0,9957	1,0000

<sup>1</sup> GL<sub>Erro</sub> = grau de liberdade do erro.

<sup>2</sup> SQ<sub>Total</sub> = soma de quadrados total.

<sup>3</sup> PRESS = soma de quadrados predito do erro =  $\sum_{i=1}^n [y_i - \hat{y}_{(i)}]^2$ .

<sup>4</sup> S<sub>PRESS</sub> = desvio-padrão de PRESS =  $\frac{\sqrt{PRESS}}{GL_{Erro}}$ .

<sup>5</sup> R<sup>2</sup><sub>Validação</sub> =  $1 - \frac{PRESS}{SQ_{Total}}$ .

A validação de modelos tem por objetivo verificar se o modelo predito terá sucesso em sua aplicação final, ou seja, na predição de dados. MONTGOMERY & PECK (1982) comentam sobre três procedimentos usados para validar um modelo de regressão, que seriam: a análise dos coeficientes do modelo e valores preditos baseado

em experiência própria ou resultados de outros estudos; coleta de novos dados para avaliar a capacidade de predição do modelo; e o uso dos próprios dados utilizados na predição do modelo. Neste estudo, utilizou-se o último procedimento, que consiste em separar os dados originais em duas partes: uma para construir o modelo de regressão

e a outra para estudar a capacidade de predição desse modelo. A separação dos dados pode ser feita mediante a soma de quadrados predito do erro ou estatística PRESS (ALLEN, 1974), que é então utilizada para computar um  $R^2$  aproximado para validação ou predição do modelo de regressão desenvolvido.

Os valores verificados para os  $R^2$  da validação foram muito próximos dos  $R^2$  ajustados apresentados na Tabela 2. Quanto maior o número de coletas de leite incluídas no modelo (variáveis independentes), mais próximos foram os valores do  $R^2$  da validação (Tabela 6) com o  $R^2$  ajustado (Tabela 2). Na validação de modelos para estimativa da produção de leite em vacas Charolês, PACHECO et al. (2009) verificaram resposta semelhante ao do presente estudo, referente ao uso da estatística PRESS.

## CONCLUSÕES

Todos os modelos preditos foram considerados satisfatórios na estimativa da produção média diária de leite até os 217 dias de lactação em vacas Nelore.

Quando houver a possibilidade de executar diversas amostragens ao longo do período de lactação, três coletas da produção de leite, em datas estrategicamente escolhidas ao longo do período de lactação, são suficientes para predizer com elevada precisão a produção média diária de leite.

## REFERÊNCIAS

- ALBUQUERQUE, L.G.; ELER, J.P.; COSTA, M.J.R.P. Produção de leite e desempenho do bezerro na fase de aleitamento em três raças bovinas de corte. **Revista Brasileira de Zootecnia**, v. 22, n. 5, p. 745-754, 1993.
- ALENCAR, M.M. Efeitos da produção de leite sobre o desenvolvimento de bezerras Canchim. **Revista da Sociedade Brasileira de Zootecnia**, v.16, n.1, p. 1-13, 1987.
- ALENCAR, M.M.; JUNQUEIRA FILHO, A.A.; PARANHOS, N.E. Produção de leite em vacas da raça Canchim. **Revista da Sociedade Brasileira de Zootecnia**, v.14, n. 3, p. 358-366, 1985.
- ALENCAR, M.M.; RIBEIRO, R.P.; VERÍSSIMO, C.J.; DURAM, J.T.; MORO, M.E.G. Efeitos da produção de leite das vacas sobre o desenvolvimento pós-desmama de bezerras da raça Canchim. **Pesquisa Agropecuária Brasileira**, v. 27, n.1, p.105-110, 1992.
- ALENCAR, M.M.; RUZZA, F.J.; PORTO, E.J.S. Desempenho produtivo de fêmeas das raças Canchim e Nelore. III. Produção de leite. **Revista Brasileira de Zootecnia**, v. 17, n. 4, p. 317-328, 1988.
- ALLEN, D.M. The relationship between variable selection and data augmentation and a method of prediction. **Technometrics**, v.16, p. 125-127, 1974.
- BROWN, M.A.; BROWN JR., A.H. Relationship of milk yield and quality to preweaning gain of calves from Angus, Brahman and reciprocal-cross cows on different forage systems. **Journal of Animal Science**, v. 80, p. 2522-2527, 2002.
- CERDÓTES, L.; RESTLE, J.; ALVES FILHO, D.C.; NÖRNBERG, M.F.B.L.; NÖRNBERG, J.L.; HECK, I.; SILVEIRA, M.F. Produção e composição do leite de vacas de quatro grupos genéticos submetidas a dois manejos alimentares no período de lactação. **Revista Brasileira de Zootecnia**, v. 33, n. 3, p. 610-622, 2004.
- CRUZ, G.M.; ALENCAR, M.M.; TULLIO, R.R. Produção e composição do leite de vacas das raças Canchim e Nelore. **Revista Brasileira de Zootecnia**, v. 26, n. 5, p. 887-893, 1997.
- DRAPER, N.R.; SMITH, H. **Applied regression analysis**. 3. ed. New York: Wiley Series in Probability and Statistics. 1998. 705 p.
- FREETLY, H.C.; CUNDIFF, L.V. Reproductive performance, calf growth, and milk production of first-calf heifers sired by seven breeds and raised on different levels of nutrition. **Journal of Animal Science**, v. 76, p. 1513-1522, 1998.
- FREUND, R.J.; LITTEL, R.C. **SAS system for regression**. 2<sup>nd</sup> ed. Cary, North Caroline: SAS Institute Incorporation, 1991.
- HOLLOWAY, J.W.; BUTTS, W.T.; McCURLEY, J.R.; BEAVER, E.E.; PEELER, H.L.; BACKUS, W.L. Breed x nutritional environment interactions for beef female weight and fatness, milk production and calf growth. **Journal of Animal Science**, v. 61, n. 6, p.1354-1363, 1985.
- JOHNSON, C.R.; LALMAN, D.L.; BROWN, M.A.; APPEDDU, L.A.; BUCHANAN, D.S.; WETTEMANN,

- R.P. Influence of milk production potential on forage dry matter intake by multiparous and primiparous Brangus females. **Journal of Animal Science**, v. 81, p.1837-1846, 2003.
- MONTGOMERY, D.C.; PECK, E.A. **Introduction to linear regression analysis**. USA: John Wiley & Sons, Inc./Wiley Series in Probability and Mathematical Statistics, 1982. 504 p.
- MORENO, J.A. **Clima do Rio Grande do Sul**. Porto Alegre: Secretaria da Agricultura, 1961. 42 p.
- PACHECO, P.S.; RESTLE, J.; FREITAS, A.K.; BRONDANI, I.L.; SILVA, J.H.S.; ALVES FILHO, D.C. Número de amostragens da produção de leite para estimativa da produção média diária de leite em vacas Charolês. **Ciência Animal Brasileira**, v. 10, n. 1, p. 147-157, 2009.
- RESTLE, J.; PACHECO, P.S.; MOLETTA, J.L.; BRONDANI, I.L.; CERDÓTES, L. Grupo genético e nível nutricional pós-parto na produção e composição do leite de vacas de corte. **Revista Brasileira de Zootecnia**, v. 32, n. 3, p. 585-597, 2003.
- RESTLE, J.; SENNA, D.B.; PACHECO, P.S.; PADUA, J.T.; VAZ, R.Z.; METZ, P.A.M. Grupo genético e heterose na produção de leite de vacas de corte submetidas a diferentes sistemas de alimentação. **Revista Brasileira de Zootecnia**, v. 34, n. 4, p.1329-1338, 2005.
- RIBEIRO, E.L.A.; RESTLE, J. Desempenho de terneiros Charolês e Aberdeen Angus puros e seus mestiços com Nelore. **Pesquisa Agropecuária Brasileira**, v. 26, n. 8, p.1145-1151, 1991.
- RIBEIRO, E.L.A.; RESTLE, J.; PIRES, C.C. Produção e composição do leite de vacas Charolês e Aberdeen Angus amamentando bezerros puros ou mestiços. **Pesquisa Agropecuária Brasileira**, v. 26, n. 8, p.1267-1273, 1991.
- RIO GRANDE DO SUL. Secretaria da Agricultura. Departamento de Pesquisa. Instituto de Pesquisas Agronômicas. Observações Meteorológicas do Estado do Rio Grande do Sul. **Boletim Técnico**, Porto Alegre, n. 3, 1979. 270 p.
- RODRIGUES-ALMEIDA, F.A.; VAN VLECK, L.D.; GREGORY, K.E. Estimation of direct and maternal breed effects for prediction of expected progeny differences for birth and weaning weights in three multibreed populations. **Journal of Animal Science**, v. 25, p.1203-1212, 1997.
- SAS - STATISTICAL ANALYSIS SYSTEMS. **Sas Institute: User's Guide – Version 6**, Cary, NC: v.2, 1997. 1052p.
- SOUZA, G.S. **Introdução aos modelos de regressão linear e não-linear**. Brasília: Embrapa-SPI/Embrapa-SEA, 1998. 505 p.

---

Protocolado em: 11 jul. 2008. Aceito em: 4 out. 2008.