



**Revista Eletrônica
Paulista de Matemática**

ISSN 2316-9664
Volume 7, dez. 2016
Edição ERMAC

Vitória Castro Santos Barreto
Faculdade de Tecnologia de
Botucatu - FATEC
vitoriabarreto28@gmail.com

Gislaine Cristina Batistela
UNESP – Departamento de
Engenharia de Produção
gislaine@itapeva.unesp.br

Monica Regina Gaiotto
Faculdade de Tecnologia de
Botucatu - FATEC
mgaiotto@fatecbt.edu.br

Danilo Simões
UNESP – Departamento de
Engenharia de Produção
danilo@itapeva.unesp.br

Regressão linear múltipla aplicada ao preço do leite

Multiple linear regression applied to the milk price

Resumo

Os laticínios passaram a remunerar os produtores não só pela quantidade de leite entregue, mas também pela qualidade apresentada pelo produto. As principais variáveis que compõem a qualidade microbiológica do leite e impactam no preço são: estrato sólido total (EST), contagem de célula somática (CCS) e contagem bacteriana total (CBT). O objetivo deste trabalho foi avaliar a relação dessas variáveis com o preço do leite cru pago ao produtor. Para ponderar quais são as que mais impactam no preço do leite, utilizou-se o modelo de regressão linear múltipla. A estimação dos parâmetros foi realizada pelo método dos mínimos quadrados utilizando os *softwares* SAS e R. As discrepâncias entre os dados observados de produtores de um laticínio do Estado de São Paulo e os valores ajustados pelo modelo foram analisadas. Os resultados permitem concluir que o modelo matemático é adequado para estimar o preço do leite recebido pelos produtores. Verificou-se ainda que à medida que o estrato sólido aumenta, o preço do leite sobe. As variáveis quantidade de leite fornecida e contagem bacteriana apresentaram uma variabilidade alta, ao passo que as variáveis estrato sólido, contagem de célula somática e o preço médio do leite pago aos produtores apresentaram uma variação pequena em relação as médias obtidas.

Palavras-chave: Modelo matemático. Contagem bacteriana total. Contagem de célula somática. Estrato de sólido total. Preço do leite.

Abstract

The dairy remunerates the producers not only by the quantity of milk delivered, but also by the quality of the product. The main variables in respect to the microbiological quality and to the impact in the price of milk are: Total Solid Stratum, Somatic Cells and Total of Bacterial. The objective of this work was evaluate the relationship of these variables with the price of raw milk paid to producers. To verify the main variables that impact the milk price, the multiple linear regression model was used. The parameter estimation was carried out by the least-squares method using the SAS and R softwares. The discrepancies between the data observed in respect to producers of a dairy of São Paulo State and the values fitted by the model was analyzed. The results showed that the mathematical model is convenient to estimate the price of milk. It was also observed that when the Solid Stratum increases, the price of milk grows. The variables amount of milk and bacterias showed high variabilities. The variables Solid Stratum, Somatic Cells and the average price of milk paid to producers showed small variations in relation to the averages obtained.

Keywords: Mathematical model. Total bacterial count. Somatic cell count. Total solid stratum. Price and quality of milk.



1 Introdução

O leite é considerado um dos alimentos mais completos da atualidade, por conta das inúmeras substâncias que proporcionam nutrientes importantes e boa qualidade de vida ao consumidor. De acordo com Carvalho et al. (2002), além de seu valor nutricional, o leite está entre os seis produtos mais importantes da agropecuária brasileira e tem grande contribuição social com a geração de empregos.

Em quase todos os países, segundo normas regulamentares, a qualidade do leite é composta, em função das características físico-químicas, sensoriais e microbiológicas (BRITO; BRITO, 1998)

Segundo Nightingale et al. (2008) os laticínios passaram a remunerar os produtores não só pelo volume de leite cru entregue, mas também pela qualidade microbiológica deste. Esta forma de pagamento estimula os produtores a buscarem a melhoria contínua da qualidade do leite cru, seja por meio da genética, nutrição ou manejo.

Basicamente, o leite é uma combinação de elementos sólidos e água. As principais variáveis que compõem a qualidade microbiológica do leite e impactam no preço são: estrato sólido total (EST), contagem de célula somática (CCS) e contagem bacteriana total (CBT). O produto com baixa CBT, baixa CCS e com alta taxa de EST é o melhor remunerado (LACERDA; MOTA; SENA, 2010).

O estrato sólido total consiste na medição dos elementos sólidos do leite, cujos principais componentes são: lipídios (gordura), carboidratos, proteínas, sais minerais e vitaminas. Esses elementos, suas distribuições e interações são determinantes para as propriedades funcionais e a qualidade do leite cru fornecido pelos produtores aos laticínios (VIEIRA; FREITAS, 2006).

Segundo Machado e Pereira (1998), a contagem de células somáticas é um fator que afeta diretamente a qualidade do leite cru, porque tais células são estruturas de defesa do organismo do animal e no momento em que o úbere da vaca sofre uma invasão bacteriana, essas estruturas são “enviadas às glândulas mamárias” a fim de destruir as bactérias. Desta forma, a contagem de células somáticas reflete o estado de saúde do úbere da vaca.

A contagem bacteriana total contempla a contagem do número de bactérias contidas no leite cru, a qual é considerada uma indicação muito importante, pois expressa as condições gerais de higiene da ordenhadeira mecânica, dos utensílios utilizados pelo produtor, bem como a refrigeração do leite cru (ALMEIDA, 2013).

Para o estudo das variáveis que mais impactam no preço do leite cru pago ao produtor, pode-se utilizar a análise de regressão linear múltipla, que é uma metodologia estatística empregada para a previsão de valores, por meio de um modelo matemático que descreva o comportamento de uma variável resposta (ou variável dependente) com base nos valores de uma ou mais variáveis explicativas (ou variáveis independentes).

De acordo com Sassi et al. (2011), esta metodologia descreve a relação matemática existente entre a variável resposta e as variáveis explicativas e pode reduzir o número de variáveis com o mínimo de perda de informação, permitindo detectar padrões de similaridade, associação e correlação entre as variáveis do estudo.

Neste sentido, objetivou-se estudar o relacionamento entre as variáveis microbiológicas tais como, EST, CCS e CBT e o preço do leite cru recebido pelo produtor, por meio da técnica de análise de regressão linear múltipla.

2 Material e métodos

Para a construção do modelo matemático foram consideradas informações provenientes de 31 produtores que integram uma linha de transporte de leite, pertencente a um laticínio do Estado de São Paulo, registradas no período de abril de 2014 a março de 2015.

Inicialmente, realizou-se uma análise estatística descritiva dos dados para compreender o comportamento individual de cada variável em estudo. As variáveis ponderadas para construção do modelo de regressão linear múltipla foram: preço do leite cru recebido pelo produtor (R\$/litro), quantidade de leite cru entregue pelo produtor (qde) (litros); estrato sólido total (EST) (gramas/100ml); contagem de célula somática (CCS) (células/ml) e contagem bacteriana total (CBT) (colônias/ml).

Em seguida, conforme a descrição de Gazola (2002), Paula (2004), Cordeiro e Lima Neto (2006), Montgomery e Runger (2008), Helene (2013) e Morettin e Bussab (2012), estabeleceu-se a construção do modelo matemático de regressão linear múltipla juntamente com a análise de diagnóstico do modelo.

A regressão linear múltipla é uma técnica estatística utilizada quando se deseja analisar o comportamento de uma variável resposta em relação a outras variáveis que são responsáveis pela variabilidade das observações. Destarte, assume-se que há uma relação linear entre a variável resposta e as k variáveis explicativas (CORRAR; PAULO; DIAS FILHO, 2007).

De acordo com Gazola (2002), um dos objetivos da análise de regressão linear múltipla é desenvolver uma equação (ou modelo matemático) que permita ao pesquisador estimar respostas de uma variável, considerando os valores de várias variáveis explicativas.

O modelo matemático de uma regressão linear múltipla com k variáveis explicativas e n observações é expresso por:

$$Y_j = \beta_0 + \beta_1 X_{1j} + \beta_2 X_{2j} + \dots + \beta_k X_{kj} + \varepsilon_j, \quad (1)$$

em que:

Y_j representa a variável resposta, para $j = 1, 2, \dots, n$;

X_{ij} denotam as variáveis explicativas, para $i = 0, 1, \dots, k$;

β_i denotam os parâmetros do modelo (ou coeficientes de regressão) a serem estimados;

ε_j são os erros aleatórios do modelo supostos independentes e normalmente distribuídos de média zero e variância σ^2 .

O parâmetro β_i representa a variação esperada na resposta Y por unidade de variação em X_i , quando todas as outras variáveis explicativas forem mantidas constantes (MONTGOMERY; RUNGER, 2008).

No ajuste de um modelo de regressão linear múltipla, é mais conveniente expressar as operações matemáticas usando a notação matricial. Considerando o modelo expresso em (1), a representação matricial é dada por:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (2)$$

com:

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & X_{11} & X_{21} & \dots & X_{k1} \\ 1 & X_{12} & X_{22} & \dots & X_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1n} & X_{2n} & \dots & X_{kn} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \text{ e } \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

Conforme preconiza Gazola (2002), os coeficientes de regressão podem ser estimados por vários métodos, sendo que um dos mais utilizados é o Método dos Mínimos Quadrados (MMQ). Neste estudo, conforme desenvolvido em Morettin e Bussab (2012), a estimação dos parâmetros do modelo de regressão linear múltipla foi realizada pelo MMQ.

De acordo com Helene (2013), o MMQ fundamenta-se em minimizar o erro quadrático médio das medidas, ou seja, consiste em encontrar estimadores para os parâmetros de forma



que a soma dos quadrados dos desvios entre os valores estimados pelo modelo e os valores observados seja a menor possível.

Um dos critérios utilizados para analisar e comparar os modelos é o coeficiente de determinação múltipla R^2 , que representa a proporção da variação explicada pelo modelo regressão, ou seja, é uma medida de qualidade de ajuste do modelo aos dados sendo calculada por:

$$R^2 = \frac{SQ_R}{SQ_T} = 1 - \frac{SQ_E}{SQ_T}, \quad (3)$$

em que:

SQ_R é a soma de quadrados da regressão;

SQ_E é a soma de quadrados dos erros;

SQ_T é a soma de quadrados total.

O valor de R^2 tem variação entre 0 e 1 e quanto mais próximo de 1 melhor é o ajuste do modelo, ou seja, quanto maior R^2 , mais a variação total de Y é reduzida pelo uso de variáveis explicativas (MONTGOMERY; RUNGER, 2008).

No entendimento de Gazola (2002), na regressão linear múltipla o coeficiente de determinação R^2 tende a aumentar à medida que mais variáveis explicativas são adicionadas ao modelo, independente da variável adicional ser ou não estatisticamente significativa. Este fato leva a um coeficiente que não mede mais a real explicação da variável resposta Y .

De acordo com Montgomery e Runger (2008), existe uma medida que corrige o coeficiente de determinação pela quantidade de variáveis independentes do modelo, denominado coeficiente de determinação ajustado expresso por:

$$R_a^2 = 1 - \left(\frac{n-1}{n-p} \right) (1 - R^2), \quad (4)$$

com $p = k + 1$, isto é, p indica o número de variáveis explicativas mais a constante.

Para avaliar a qualidade do ajuste do modelo verificou-se a possível inadequação das suposições iniciais, a presença de observações mal ajustadas (pontos aberrantes) e influentes utilizando técnicas de diagnóstico (NIGHTINGALE et al. 2008).

Para identificar os pontos de alavanca utilizou-se a matriz de projeção ortogonal de vetores do \mathcal{R}^n no subespaço gerado pelas colunas da matriz \mathbf{X} , denotada pela matriz $\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}$, a qual é uma matriz simétrica e idempotente.

Segundo Cordeiro e Lima Neto (2006), uma característica muito importante da matriz \mathbf{H} é inerente aos elementos da sua diagonal, em que o elemento h_{ii} mede o quão distante a observação y_i , está das demais $(n-1)$ observações no espaço definido pelas variáveis explicativas, isto é, da matriz \mathbf{X} e não envolve as observações em \mathbf{Y} .

O elemento h_{ii} representa uma medida de alavancagem da i -ésima observação. Se $h_{ii} \geq \frac{2p}{n}$, os valores das variáveis explicativas associados a i -ésima observação são atípicos, ou seja, estão distantes do vetor de valores médios das variáveis explicativas. Uma observação com $h_{ii} \geq \frac{2p}{n}$, poderá ter influência na determinação dos coeficientes da regressão (PAULA, 2004; CORDEIRO; LIMA NETO, 2006).

Segundo Montgomery e Runger (2008), um excelente método para detectar observações influentes é a medida da distância desenvolvida por Dennis R. Cook. A distância de Cook é a medida da distância ao quadrado entre a estimativa usual de mínimos quadrados de $\beta(\hat{\beta})$, baseada em todas n observações, e a estimativa obtida quando o j -ésimo ponto for removido, denotado por $\hat{\beta}_{(j)}$. A medida da distância de Cook é definida por:

$$D_j = \frac{(\hat{\beta}_{(j)} - \hat{\beta})' \mathbf{X}' \mathbf{X} (\hat{\beta}_{(j)} - \hat{\beta})}{p \hat{\sigma}^2}, \quad (5)$$



sendo que $\hat{\sigma}^2$ é a estimativa da variância do erro, com $j = 1, 2, \dots, n$.

Desta forma, um grande valor de D_j indica que o j -ésimo ponto exerce influência, ou seja, o j -ésimo ponto é um ponto aberrante (PAULA, 2004).

A análise de resíduos é uma das técnicas de diagnóstico do modelo muito utilizada. O resíduo para a i -ésima observação é obtido por meio da função $r_i = y_i - \hat{y}_i$, que calcula a diferença entre o valor observado (y_i) e o valor ajustado (\hat{y}_i), denominado de resíduo ordinário da variável resposta do modelo. Os resíduos ordinários não são muito informativos, por não apresentar variância constante $\text{Var}(r_i) = \sigma^2(1 - h_{ii})$, pois depende dos valores h_{ii} (PAULA, 2004; CORDEIRO; LIMA NETO, 2006). A solução é comparar os resíduos de forma padronizada, que é obtido pela expressão:

$$r_i^* = \frac{y_i - \hat{y}_i}{\sqrt{\hat{\sigma}^2(1 - h_{ii})}} \quad (6)$$

Caso o modelo de regressão esteja correto o conjunto de resíduos terão a mesma variância e serão adequados para a verificação de normalidade e homocedasticidade (variância constante) dos erros. As observações que possuírem os valores absolutos dos resíduos padronizados maiores que dois poderão ser considerados pontos aberrantes.

Segundo Cordeiro e Lima Neto (2006), como resíduo de cada observação não é independente da variância estimada, não se obtém uma distribuição t -Student, como será esperado. O problema da dependência entre r_i e $\hat{\sigma}^2$ pode ser contornado substituindo $\hat{\sigma}^2$ por $\hat{\sigma}_i^2$, o erro quadrático médio correspondente ao modelo sem a i -ésima observação. O índice (i) indica que a i -ésima observação foi excluída. A expressão do resíduo Studentizado é dada por:

$$t_i = \sqrt{\frac{n - p - 1}{n - p - r_i^{*2}}} r_i^* \quad (7)$$

em que t_i tem distribuição t -Student com $(n - p - 1)$ graus de liberdade.

Os resíduos Studentizado definidos na Equação (7) têm a grande vantagem de serem obtidos diretamente da regressão original com todas as observações.

3 Resultados

A Tabela 1 fornece as medidas descritivas das variáveis utilizadas para modelar o preço do leite pago ao produtor. Observa-se que o preço médio do leite pago aos produtores foi de 1,091 reais por litro com coeficiente de variação de 0,677%, o que indica uma variação muito pequena, com o valor mínimo de 1,076 reais por litro e valor máximo de 1,107 reais por litro.

Tabela 1 - Média, Mediana, Máximo, Mínimo, Desvio Padrão e Coeficiente de Variação das variáveis estudadas.

Medida	Variáveis				
	Preço	Qde	EST	CCS	CBT
Média	1,091	8.971,675	12,114	427.231	164.384
Mediana	1,088	5.182,750	12,111	430.167	176.000
Min.	1,076	594,000	11,574	345.250	68.833
Máx.	1,107	63.912,500	12,572	476.250	415.500
D.P	0,007	11.859,676	0,259	26,685	69,994
C.V. (%)	0,677	132,190	2,141	6,246	42,580

A quantidade de leite cru fornecida pelos produtores ao laticínio foi em média 8.971,675 litros, sendo que o mínimo foi de 594 litros e o máximo foi de 63.912,5. O coeficiente de variação foi de 132,190% indicando variabilidade muito alta.

O estrato sólido total apresentado pelo leite cru fornecido foi em média 12,114 gramas/100ml, com o valor mínimo de 11,574 e o máximo de 12,572. O coeficiente de variação resultou em 2,141% demonstrando uma variação baixa.

A contagem de célula somática do leite cru foi em média 427.231 células/ml, apresentando um coeficiente de variação de 6,246%, que retrata uma variação pequena, com valor mínimo de 68.833 e máximo de 415.500.

A contagem bacteriana total denota em média 164.384 colônias/ml, retratando um valor mínimo de 68.833 e máximo de 415.500. O coeficiente de variação sucedeu em 42,580% que manifesta uma variação alta.

Resultado semelhante foi relatado por Nascimento et al. (2015) que observou o valor médio da variável estrato sólido total em 12,49 gramas/100ml, já a média da contagem de célula somática e contagem bacteriana total apresentam valores discrepantes aos encontrados no presente estudo, apontados em 617.480 células/ml e 934.960 colônias/ml respectivamente.

Na Figura 1 é apresentado o gráfico *boxplot* para as variáveis em estudo. Constata-se que para a variável EST (Figura 1 (c)) não foi detectado valor discrepante (*outlier*) e, visualmente, tem distribuição simétrica, a qual pode ser confirmada pelos valores da média e mediana (Tabela 1).

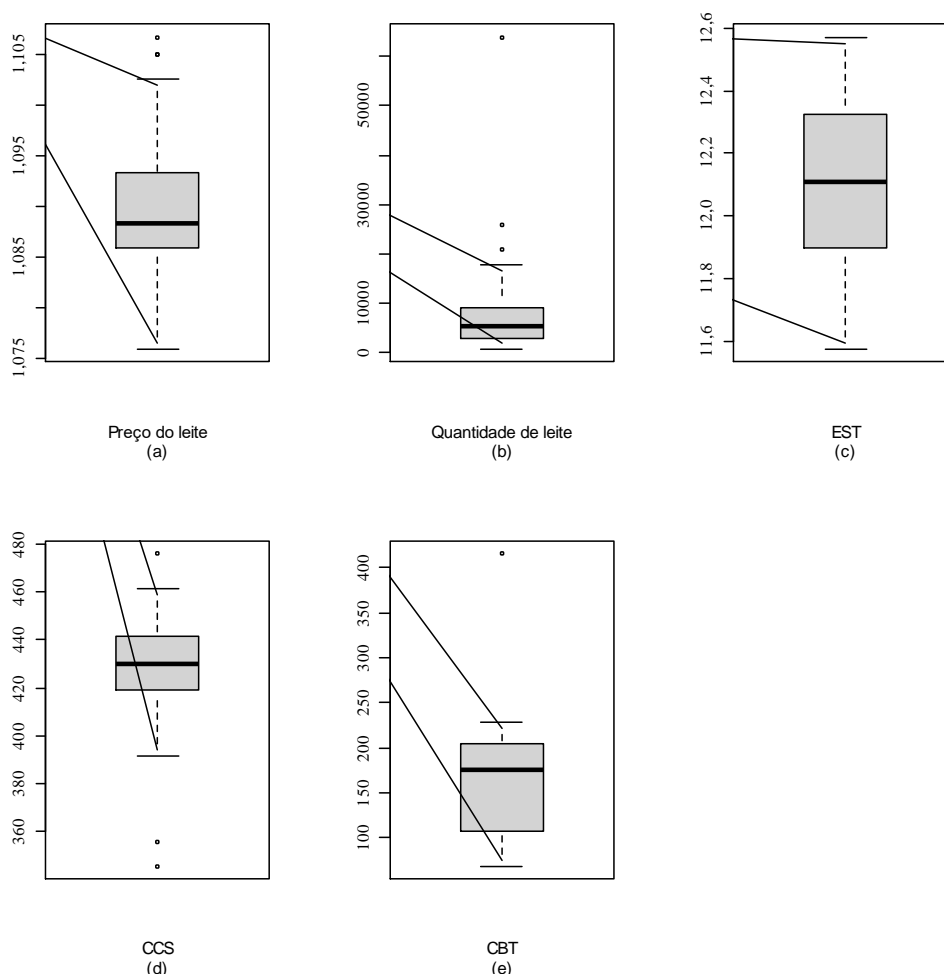


Figura 1 - Gráfico *boxplot* para as variáveis: (a) preço do leite cru pago ao produtor; (b) quantidade de leite cru fornecida pelo produtor; (c) EST; (d) CCS; (e) CBT.



3.1 Construção do modelo

Inicialmente, estuda-se a relação de cada variável explicativa com a variável resposta por meio do coeficiente de correlação linear de Pearson. Considerando o nível de significância de 5%, verifica-se que os valores da correlação evidenciam uma forte relação entre as variáveis e são estatisticamente significativas (valores $p < 0,05$).

Ajusta-se o modelo de regressão múltipla com as variáveis: preço (preço do leite pago ao produtor), qde (quantidade de leite cru entregue pelos produtores), EST (estrato sólido total), CCS (contagem de célula somática) e CBT (contagem bacteriana total), com o interesse de explicar o preço do leite em função das variáveis: quantidade, EST, CCS e CBT.

O modelo inicial proposto é expresso por:

$$P_i = \beta_0 + \beta_1 qde + \beta_2 EST + \beta_3 CCS + \beta_4 CBT + \varepsilon_i, \quad (8)$$

sendo que P_i denota o preço do leite cru pago para o i -ésimo produtor, β_j com $j = 0, 1, \dots, 4$, são os parâmetros do modelo a serem estimados (coeficientes de regressão) e ε_i são os erros aleatórios.

Na Tabela 2 encontram-se as estimativas dos parâmetros do modelo de regressão ajustado.

Tabela 2 - Estimativa dos parâmetros.

Parâmetro	Estimativa	Erro Padrão	Valor p
β_0	1,021	$2,917 \times 10^{-2}$	$< 0,05$
β_1	$8,671 \times 10^{-8}$	$4,647 \times 10^{-8}$	$> 0,05$
β_2	$9,434 \times 10^{-3}$	$2,167 \times 10^{-3}$	$< 0,05$
β_3	$-8,723 \times 10^{-5}$	$2,000 \times 10^{-5}$	$< 0,05$
β_4	$-4,803 \times 10^{-5}$	$8,049 \times 10^{-6}$	$< 0,05$

Pela Tabela 2 verifica-se que as estimativas dos parâmetros β_0 , β_2 , β_3 e β_4 foram estatisticamente significativas ($p < 0,05$) e a estimativa do parâmetro β_1 não foi estatisticamente significativa ($p = 0,073$). Isto significa que a variável quantidade de leite cru entregue pelos produtores foi eliminada do modelo.

O coeficiente de determinação ajustado, R_a^2 , é de 0,882, ou seja, 88,2% da variação no preço do leite pode ser explicada pela variação nas taxas de CBT, CCS e EST.

3.2 Análise de diagnóstico do modelo

Para avaliar a qualidade do modelo de regressão ajustado realiza-se a verificação de possíveis afastamentos das suposições feitas para o modelo, bem como a existência de observações discrepantes com alguma interferência desproporcional ou inferencial nos resultados do ajuste.

A Figura 2 mostra os gráficos de diagnóstico do modelo. Observa-se que os produtores 6, 9 e 13 estão localizados em regiões remotas no subespaço gerado pelas colunas da matriz \mathbf{X} , ou seja, se destacam como pontos de alavanca (Figura 2 (a)) e como pontos influentes (Figura 2 (b)). Pelos gráficos da Figura 2 (b) e (c) nota-se que o produtor 18 é respectivamente um ponto influente e aberrante. Verifica-se, pela Figura 2 (c) e (d), que não existe padrão na relação entre os resíduos e os valores do preço do leite pago ao produtor ajustado, indicando que a suposição de homoscedasticidade não está violada.

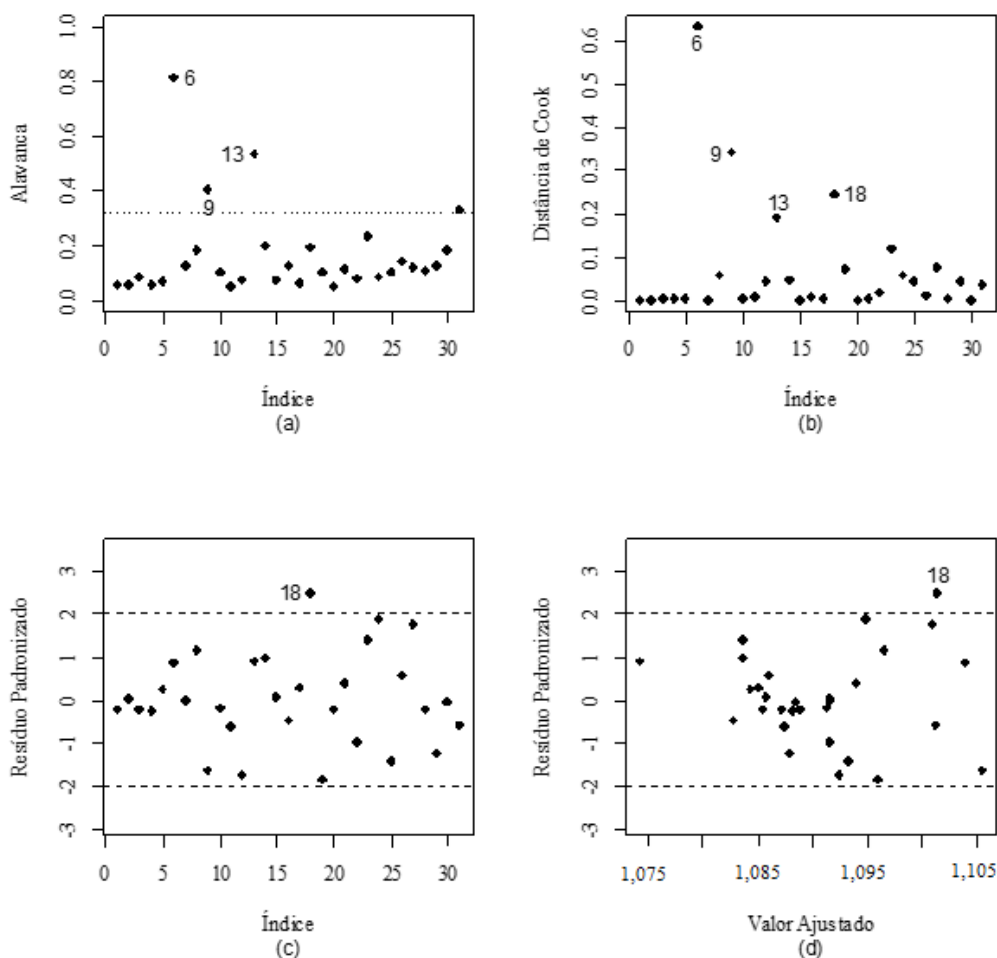


Figura 2 - Gráficos de diagnóstico referentes ao modelo ajustado

Na Figura 3 constata-se, pelo gráfico de envelope, que não há indícios fortes de afastamentos da suposição de normalidade para os erros, ou seja, nota-se uma boa acomodação dos pontos dentro das bandas do envelope gerado para o modelo ajustado.

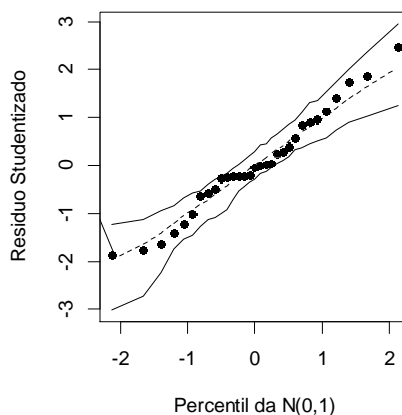


Figura 3 - Gráfico normal de probabilidade.

Desta forma, conclui-se que o modelo de regressão múltipla estimado é apropriado para estimar o preço do leite cru pago aos produtores.



4 Conclusão

Utilizando o modelo de regressão linear múltipla, foi feita uma análise do preço do leite em função da variação nas variáveis: quantidade de leite cru entregue pelos produtores, estrato sólido total (EST), contagem de célula somática (CCS) e contagem bacteriana total (CBT). Sendo assim, as principais conclusões são:

1. O modelo matemático estimado é adequado e gera grande representatividade para estimar o preço do leite cru recebido pelos produtores.
2. As variáveis estrato sólido total, contagem de célula somática e contagem bacteriana total foram incluídas no modelo, enquanto que a quantidade de leite cru entregue pelo produtor foi excluída.
3. Em relação às médias obtidas, observou-se que a variável quantidade de leite fornecida e contagem bacteriana total apresentaram uma variabilidade alta, ao passo que a variável estrato sólido, contagem de célula somática e o preço médio do leite pago aos produtores apresentaram uma variação pequena.
4. Analisando os sinais dos parâmetros, identificou-se que à medida que o estrato sólido total aumenta ocorre maior preço do leite cru, enquanto que um aumento na contagem de célula somática e na contagem bacteriana total acarreta menor preço do leite cru.
5. 88,2% da variação no preço do leite pode ser explicada pela variação nas taxas de CBT, CCS e EST.

5 Referências

ALMEIDA, T. V. **Parâmetros de qualidade do leite cru bovino: contagem bacteriana total e contagem de células somáticas**. Goiânia, 2013. Disponível em: <https://ppgca.evz.ufg.br/up/67/o/2013_Thamara_Venancio_Seminario1corrig.pdf>. Acesso em: 10 jul. 2016.

BRITO, J. R. F.; BRITO, M. A. V. P. Descomplicando as células somáticas. In: BRITO, J. R. F.; DIAS, J. C. (Ed.). **A qualidade do leite**. Juiz de Fora: EMBRAPA, 1998. p. 75-82.

CARVALHO, L. A. et al. **Importância econômica**. Embrapa Gado de Leite: sistema de produção, v. 2, 2002. Disponível em: <<https://sistemasdeproducao.cnptia.embrapa.br/FontesHTML/Leite/LeiteCerrado/importancia.html>>. Acesso em: 11 jul. 2016.

CORDEIRO, G. M.; LIMA NETO, E. A. **Modelos paramétricos**. Recife: Universidade Federal Rural de Pernambuco, 2006.

CORRAR, L. J.; PAULO, E.; DIAS FILHO, J. M. (Coords.) **Análise multivariada: para os cursos de administração, ciências contábeis e economia**. São Paulo: Atlas, 2007.

ESTEVES, E.; SOUSA, C. **Apontamentos de ADPE: regressão linear múltipla**. 2007.

GAZOLA, S. **Construção de um modelo de regressão para avaliação de imóveis**. 2002. 104 f. Dissertação (Mestrado em Engenharia de Produção)– Universidade Federal de Santa Catarina, Florianópolis, 2002.



HELENE, O. **Métodos dos mínimos quadrados com formalismo matricial: guia do usuário**. 2. ed. São Paulo: Livraria da Física, 2013.

LACERDA, L. M.; MOTA, R. A.; SENA, M. J. Contagem de células somáticas, composição e contagem bacteriana total do leite de propriedades leiteiras nos municípios de Miranda do Norte, Itapecurú–Mirim e Santa Rita, Maranhão. **Arquivos do Instituto Biológico, São Paulo**, v. 77, n. 2, p. 209-215, 2010.

MACHADO, P. F.; PEREIRA, A. R. Contagem de células somáticas no leite de rebanhos brasileiros e seus efeitos sobre a qualidade do leite. In: ENCONTRO TECNOLAT DE CONFERÊNCIAS TÉCNICAS, 1., 1998, Campinas. **Anais...** Campinas, 1998.

MONTGOMERY, D. C.; RUNGER, G. C. **Estatística aplicada e probabilidade para engenheiros**. 2. ed. Rio de Janeiro: LTC, 2008.

MORETTIN, P. A.; BUSSAB, W. O. **Estatística básica**. 7. ed. São Paulo: Editora Saraiva, 2012.

NASCIMENTO, G. C. et al. Diagnóstico de situação da qualidade do leite no município de Bambuí, Minas Gerais. In: JORNADA CIENTÍFICA TECNOLOGIA IFMG, 8., 2015, Bambuí. **Anais...** Bambuí: IFMG, 2015.

NIGHTINGALE, C. et al. Influence of variable milk quality premiums on observed milk quality. **Journal of Dairy Science**, Champaign, v. 91, n. 3, p. 1236-1244, 2008.

PAULA, G. A. **Modelos de regressão: com apoio computacional**. São Paulo: IME/USP, 2004.

SASSI, C. P. et al. **Modelo de regressão linear múltipla utilizando os softwares R e Statística: uma aplicação a dados de conservação de frutas**. [2011?]. Disponível em: <http://conteudo.icmc.usp.br/CMS/Arquivos/arquivos_enviados/BIBLIOTECA_113_RT_377.pdf>. Acesso em: 06 jul. 2016.

VIEIRA, L. C.; FREITAS, C. M. K. H. de. Qualidade do leite. In: VEIGA, J. B. de (Ed.). **Criação de gado leiteiro na zona bragantina**. Belém: Embrapa Amazônia Oriental, Belém, 2006. p. 111-116. Disponível em: <<http://ainfo.cnptia.embrapa.br/digital/bitstream/item/126379/1/Cap13.pdf>>. Acesso em: 12 jul. 2016.