

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE MATEMÁTICA
CURSO DE GRADUAÇÃO EM ESTATÍSTICA

Thacyo Euqueres De Villa

**PREDIÇÃO DO CUSTO DE MILHO POR MEIO DE MODELOS DE REGRESSÃO
LINEAR MÚLTIPLA**

Uberlândia – MG
Dezembro/2016

Thacyo Euqueres De Villa

**PREDIÇÃO DO CUSTO DE MILHO POR MEIO DE MODELOS DE REGRESSÃO
LINEAR MÚLTIPLA**

Trabalho de conclusão de curso de graduação
apresentado à Faculdade de Matemática da
Universidade Federal de Uberlândia (UFU) como
requisito parcial para a obtenção do título de Bacharel
em Estatística.

Universidade Federal De Uberlândia
Faculdade De Matemática
Curso De Graduação Em Estatística

Orientador: Prof. Dr. JANSER MOURA PEREIRA

Uberlândia – MG
Dezembro/2016

Thacyo Euqueres De Villa

**PREDIÇÃO DO CUSTO DE MILHO POR MEIO DE MODELOS DE REGRESSÃO
LINEAR MÚLTIPLA**

Trabalho de conclusão de curso de graduação
apresentado à Faculdade de Matemática da
Universidade Federal de Uberlândia (UFU) como
requisito parcial para a obtenção do título de Bacharel
em Estatística.

Aprovada em: 01 de dezembro de 2016.

Prof. Dr. Janser Moura Pereira
Orientador

Prof. Dr. José Waldemar da Silva
Convidado

Profa. Dr. Lúcio Borges de Araújo
Convidado

Uberlândia – MG
Dezembro/2016

"Aos meus pais Jerson e Elaine, meu irmão
Thyago, meus avós, tios e primos pelo carinho,
educação, confiança, amor, incentivo e pelo
exemplo de pessoas que são."

AGRADECIMENTOS

À Universidade Federal de Uberlândia (UFU) pela possibilidade da formação e realização deste trabalho.

Ao meu pai, Jerson, minha mãe, Elaine, e meu irmão, Thyago, que sempre estiveram presentes me apoiando e incentivando durante essa etapa da minha vida.

À minha namorada, Maria Tereza, que compartilhou comigo esse momento e me ajudou bastante dando dicas e apoio moral para o desenvolvimento deste trabalho.

Ao Prof. Janser Moura Pereira pela orientação, apoio, incentivo, ensinamentos e paciência durante as aulas e realização deste e de outros trabalhos da universidade.

Agradeço também aos meus amigos e colegas da universidade que sempre torceram por mim e me apoiaram no decorrer da universidade.

*"E no final o amor que você recebe é igual ao
que você faz"*

Lennon/McCartney

RESUMO

Este trabalho tem como objetivo verificar por meio de modelos de regressão com variáveis dummies se há diferença significativa entre os custos totais de produção do milho nas cidades produtoras, Primavera do Leste – MT e Rio Verde – GO. A pesquisa foi realizada com base nos dados levantados da CONAB (Companhia Nacional de Abastecimento) dos custos de produção do milho no período de 1998 a 2013. Analisou-se 15 (quinze) variáveis, no qual apenas 4 (quatro) foram selecionadas a compor o modelo final. Com base no ajuste do modelo concluiu-se que o município Rio Verde possui em média um custo de R\$ 103,64 a mais do que Primavera do Leste na produção de milho. O modelo de regressão estimado, apresentou ótimo ajuste captando cerca de 97,67% da variação total do custo de produção do milho.

Palavras-chave: *Variável dummy, Custo de produção, Modelos de regressão, Contabilidade de custo.*

ABSTRACT

This study aims to verify through regression models with dummy variables if there is significant difference between the total cost of corn production in the producing cities, Primavera do Leste – MT and Rio Verde – GO. The survey was conducted based on data collected from Conab (Companhia Nacional de abastecimento) of corn production costs from 1998 to 2013. We analyzed 15 (fifteen) variables, in which only 4 (four) were selected to compose the final model. Based on the model fit it was concluded that the Rio Verde municipality has an average cost of R\$ 103,64 more than the Primavera do Leste in corn production. The estimated regression model showed great adjustment capturing approximately 97,67% of the total variation in the cost of maize production.

Keywords: *Dummy, Cost of production, Regression models, Cost account.*

SUMÁRIO

1 INTRODUÇÃO.....	9
2 MATERIAIS E MÉTODOS.....	11
2.1 FORMULAÇÃO MATEMÁTICA.....	11
2.1.1 Regressão Linear Múltipla.....	11
2.1.2 Métodos dos Mínimos Quadrados Ordinários.....	12
2.1.3 Métodos dos Mínimos Quadrados Ponderados.....	15
2.1.4 Métodos dos Mínimos Quadrados Generalizados.....	17
2.1.5 Adequação do modelo.....	20
2.1.6 Multicolinearidade.....	21
2.1.7 Seleção de variáveis.....	22
2.1.7.1 Método “passo atrás” (backward).....	22
2.1.7.1 Método “passo a frente” (forward).....	23
2.1.7.1 Método “passo a passo” (stepwise).....	24
2.1.8 Análise de resíduos.....	24
3 RESULTADOS E DISCUSSÕES.....	27
4 CONCLUSÃO.....	30
REFERÊNCIAS.....	31

1 INTRODUÇÃO

A contabilidade de custos surgiu da necessidade de avaliação dos estoques no setor industrial. Durante muito tempo a Contabilidade de Custos se constituiu como fator limitador para as demonstrações da sua habilidade em assessorar os usuários internos acerca de decisões gerenciais (SANTOS, SCHMIDT E PINHEIRO, 2006). Hoje as mais novas e provavelmente mais importantes tarefas da Contabilidade de Custos contemplam controle e decisão (MARTINS, 2006).

Hansen & Mowen (1997) apontam que a contabilidade de custos é considerada híbrida, ou seja, é aderida de forma significativa tanto pela contabilidade financeira como pela gerencial.

A necessidade de controlar custos, segundo Santos (2009), fez com que a sua apuração se tornasse de grande importância desde o início do capitalismo, pois era devido à contabilidade de custos que o comerciante tinha resposta se estava tendo lucro ou não. Neste contexto, a contabilidade de custos era usada como um instrumento seguro para controlar as variações de custos e de vendas e também para avaliar o crescimento ou o retrocesso do negócio.

A contabilidade de custos tem duas funções relevantes: no auxílio ao controle e na ajuda às tomadas de decisões. No que diz respeito ao controle, sua mais importante missão é fornecer dados para o estabelecimento de padrões, orçamentos e outras formas de previsão e, num estágio imediatamente seguinte, acompanhar o efetivamente acontecido para a comparação com os valores anteriormente definidos (MARTINS 2006).

A análise de regressão ocupa-se do estudo da dependência de uma variável, a variável dependente, em relação a uma ou mais variáveis, as variáveis explicativas, com o objetivo de estimar e/ou prever a média (da população) ou o valor médio da dependente em termos dos valores conhecidos ou fixos (em amostragem repetida) das explicativas. Em algumas situações podem existir variáveis independentes classificadas como qualitativas. Essa informação pode ser incorporada no modelo de regressão por meio das variáveis binárias, conhecidas como variáveis dummy que assumem dois valores distintos, geralmente zero e um (ANDERSON et al., 2007).

Tendo em vista que a contabilidade de custos tem a função de auxiliar no controle e na tomada de decisão, no presente trabalho pretende-se estudar uma relação linear entre o custo total da produção de milho (variável dependente) em função de outras variáveis independentes como o custeio de sementes, custo de mão-de-obra, custo com fertilizantes,

custo com defensivos, entre outros custos das cidades produtoras Primavera do Leste e Rio Verde com a finalidade de encontrar, descrever e prever o comportamento desses custos de forma que possa auxiliar gestores ligados ao setor no planejamento de investimentos. Foram utilizados dados da Companhia Nacional de Abastecimento - CONAB (CONAB, 2016) dos Custos de Contabilidade das cidades produtoras: Primavera do Leste – MT e Rio Verde - GO.

2 MATERIAIS E MÉTODOS

Os dados foram obtidos no site da CONAB referente a 1ª safra de milho por plantio convencional. Foram incluídos no estudo os custos do município de Rio Verde – GO e do município de Primavera do Leste – MT referentes aos anos de 1998 a 2013 totalizando em 32 observações. As variáveis analisadas foram: y representa o custo total de produção (variável dependente); x_1 são os gastos com operação com máquinas; x_2 são os gastos com mão de obra temporária; x_3 são os gastos com mão de obra fixa; x_4 são os gastos com sementes; x_5 são os gastos com fertilizantes; x_6 são os gastos com defensivos; x_7 são os gastos com despesas pós-colheita (Transporte externo; Recepção, limpeza e secagem; PROAGRO e Assistência técnica); x_8 são os gastos com juros; x_9 são os gastos com depreciação de benfeitorias/instalações; x_{10} são os gastos com depreciação de implementos; x_{11} são os gastos com depreciação de máquinas; x_{12} são os gastos com manutenção periódica de máquinas; x_{13} são os gastos com encargos sociais; x_{14} são os gastos com segurança do capital fixo; x_{15} é uma variável dummy que representa o município, se Primavera do Leste = 0 e se Rio Verde = 1.

Todas as análises foram implementadas no freeware R (R Development Core Team, 2016).

2.1 FORMULAÇÃO MATEMÁTICA

2.1.1 Regressão linear múltipla

A análise de regressão múltipla é uma técnica estatística, que pode ser usada para analisar a relação de causa e efeito entre uma única variável dependente e diversas variáveis independentes (HAIR *et. al.*, 2005).

A análise de regressão múltipla tem por objetivo estimar o impacto do incremento de cada variável independente – que se traduz como peso de cada variável independente – sobre a respectiva variação da variável dependente. Os pesos denotam a contribuição relativa das variáveis independentes para a previsão geral e facilitam a interpretação sobre a influência de cada variável explicativa em fazer a previsão. (FÁVERO *et. al.*, 2009; HAIR *et. al.*, 2005).

O modelo de regressão múltipla é dado por (CHARNET, 2008):

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + u_i, \quad (1.1)$$

em que: Y_i é o fenômeno em estudo (variável dependente); β_0 representa o intercepto (constante); $\beta_1, \beta_2, \dots, \beta_k$ são os coeficientes associados a cada variável independente (coeficientes angulares); x_{ki} são as variáveis explicativas (independentes) com $i = 1, 2, \dots, n$; u_i é o termo do erro. O erro u_i , também chamado de resíduo, representa possíveis variáveis que não foram inseridas no modelo, mas que também contribuiriam para a explicação de Y_i , em que (CHARNET, 2008):

$$u_i \stackrel{iid}{\sim} N(0, \sigma^2).$$

Os modelos serão estimados por meio do Método de Mínimos Quadrados. A seguir será apresentado a descrição dos Métodos Mínimos Quadrados Ordinários, Ponderados e Generalizados.

2.1.2 Método dos Mínimos Quadrados Ordinários

Recomendado pela sua precisão, o Método dos Mínimos Quadrados Ordinários (MQO) consiste em determinar os estimadores que minimizam a soma de quadrados dos resíduos (HOFFMANN; VIEIRA, 1998).

Considerando o modelo de regressão linear múltipla com k variáveis independentes na forma matricial tem-se (HOFFMANN; VIEIRA, 1998):

$$y = X\beta + u, \quad (2.1)$$

em que:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & X_{11} & X_{21} & \dots & - \\ 1 & X_{12} & X_{22} & \dots & - \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1n} & X_{2n} & \vdots & - \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} \text{ e } u = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_n \end{bmatrix}$$

Sejam $\hat{\beta}$ e u os vetores dos estimadores dos parâmetros e dos erros (resíduos), respectivamente, isto é,

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} \quad \text{e} \quad u = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_n \end{bmatrix}$$

tem-se:

$$\hat{y} = X\hat{\beta} \quad (2.2)$$

e

$$u = y - X\hat{\beta} = y - \hat{y}$$

em que

$$\hat{y} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \\ \vdots \\ \hat{y}_n \end{bmatrix}$$

A soma dos quadrados dos resíduos matricialmente é dada por:

$$Z = u'u = (y' - \hat{\beta}'X')(y - X\hat{\beta}) = y'y - y'X\hat{\beta} - \hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta}$$

Os produtos das matrizes $y'X\hat{\beta}$ e $\hat{\beta}'X'y$ resultam em um único elemento, isto é um escalar. Além disso, essas matrizes são iguais, pois o transposto de um escalar é o próprio escalar. Então:

$$Z = y'y - 2\hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta} \quad (2.3)$$

A função Z apresenta ponto de mínimo para os valores de β que tornem a diferencial identicamente nula, isto é:

$$\frac{dZ}{d\hat{\beta}} = -2(d\hat{\beta}')X'y + (d\hat{\beta}')X'X\hat{\beta} + \hat{\beta}'X'X(d\hat{\beta}) = 0 \quad (2.4)$$

Como $X'X$ é uma matriz quadrada de ordem p e simétrica, então, pela propriedade reflexiva da transposta tem-se:

$$(d\hat{\beta}')X'X\hat{\beta} = \hat{\beta}'X'X(d\hat{\beta}). \quad (2.5)$$

Sendo assim, a expressão (2.4) pode ser reescrita como:

$$-2(d\hat{\beta}')X'y + 2(d\hat{\beta}')X'X\hat{\beta} = 0 \Leftrightarrow (d\hat{\beta}')(X'X\hat{\beta} - X'y) = 0. \quad (2.6)$$

Portanto, a diferencial de Z será identicamente nula para:

$$X'X\hat{\beta} = X'y \quad (2.7)$$

que é o sistema de equações normais (SEN).

Se $X'X$ é não-singular, existe a matriz inversa $(X'X)^{-1}$. Pré-multiplicando os dois membros da expressão (2.7) por $(X'X)^{-1}$, obtém-se o estimador de β :

$$\hat{\beta} = (X'X)^{-1}X'y \quad (2.8)$$

2.1.3 Método dos Mínimos Quadrados Ponderados

O Método dos Mínimos Quadrados Ponderados (MQP) é recomendado quando há heterogeneidade de variâncias, pois produz estimadores não viesados e de mínima variância (HOFFMANN; VIEIRA, 1998).

Hoffmann e Vieira (1998) consideram o seguinte modelo:

$$y = X\beta + u,$$

admitindo-se que $u \sim N(0, V\sigma^2)$, em que V é uma matriz diagonal positiva definida associada às variâncias de cada u_i , isto é:

$$V = \begin{bmatrix} v_1 & 0 & \cdots & 0 \\ 0 & v_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & v_n \end{bmatrix}$$

Segundo Hoffmann e Vieira (1998), quando os elementos da matriz V fora da diagonal principal são nulos, significa que não há autocorrelação nos resíduos, ou seja, é válida a pressuposição de independência dos resíduos. Segue-se que:

$$E(u_j u_h) = 0 \text{ para } j \neq h.$$

Define-se uma matriz Λ diagonal, de modo que se tenha:

$$\Lambda\Lambda = V^{-1} \text{ e } V = \Lambda^{-1}\Lambda^{-1},$$

em que,

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

com $\lambda_j = 1/\sqrt{v_j}$ para $j = 1, 2, \dots$

Pré-multiplicando ambos os membros de $y = X\beta + u$ por Λ , obtém-se o modelo:

$$\Lambda y = \Lambda X \beta + \Lambda u .$$

Nota-se que o vetor de erros é $\varepsilon = \Lambda u$ e $E(u) = 0$. Então, tem-se que $E(\varepsilon) = 0$. Sabendo que $\Lambda' = \Lambda$, $E(uu') = V\sigma^2$ e $V = \Lambda^{-1}\Lambda^{-1}$, obtém-se:

$$E(\varepsilon\varepsilon') = E(\Lambda uu'\Lambda) = \Lambda V \Lambda \sigma^2 = \Lambda \Lambda^{-1} \Lambda^{-1} \Lambda \sigma^2 = I \sigma^2 .$$

Logo, o modelo $\Lambda y = \Lambda X \beta + \Lambda u$ é homocedástico. Pré-multiplicando o modelo homocedástico por $(\Lambda X)'$, obtém-se o SEN dado por:

$$X'V^{-1}X\hat{\beta} = X'V^{-1}Y .$$

A solução do SEN leva ao estimador:

$$\hat{\beta} = (X'V^{-1}X)^{-1} X'V^{-1}Y .$$

2.1.4 Método dos Mínimos Quadrados Generalizados

De acordo com Hoffmann & Vieira (1998), o Método dos Mínimos Quadrados Generalizados (MQG) deve ser utilizado quando se têm heterogeneidade de variâncias e autocorrelação dos resíduos.

Os autores apresentam o seguinte modelo para ilustrar o problema de autocorrelação:

$$y = X\beta + u ,$$

em que $E(u) = 0$, $E(uu') = V\sigma^2$ e V é uma matriz simétrica, positiva definida de ordem n , que está diretamente associada com a matriz de variâncias e covariâncias dos erros, admitindo-se que os erros são autocorrelacionados na forma de um processo autorregressivo estacionário de primeira ordem $AR(1)$, u_t , da seguinte forma:

$$u_t = \phi u_{t-1} + \varepsilon_t \quad (2.9)$$

em que ϕ é o parâmetro de autocorrelação e ε_t , o ruído branco. Assim,

$$E(\varepsilon_t) = 0, \quad E(\varepsilon_t^2) = \sigma_\varepsilon^2, \quad E(\varepsilon_t \varepsilon_{t-h}) = 0, \text{ se } h \neq 0 \text{ e } -1 \leq \phi \leq 1. \quad (2.10)$$

Nas expressões (2.9) e (2.10), a letra t indica o índice associado às diferentes observações, pois o problema da autocorrelação dos resíduos surge, geralmente, quando se trabalha com séries cronológicas de dados. Então, cada observação corresponde a um certo período de tempo (ano, mês ou semana, geralmente).

A relação $u_t = \phi u_{t-1} + \varepsilon_t$ mostra que o erro da observação relativa a um período está relacionado com o erro da observação anterior. Se $\phi > 0$, diz-se que os erros estão positivamente autocorrelacionados, e se $\phi < 0$, diz-se que há autocorrelação negativa. Para o caso em que $\phi = 0$, pode-se aplicar mínimos quadrados ordinários, ou seja, os erros são independentes.

Utilizando sucessivamente a relação $u_t = \phi u_{t-1} + \varepsilon_t$, obtém-se:

$$\begin{aligned} u_t &= \phi u_{t-1} + \varepsilon_t \\ &= \phi(\phi u_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ &= \phi^2 u_{t-2} + \phi \varepsilon_{t-1} + \varepsilon_t \\ &= \phi^3 u_{t-3} + \phi^2 \varepsilon_{t-2} + \phi \varepsilon_{t-1} + \varepsilon_t \\ &\vdots \\ &= \varepsilon_t + \phi \varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + \dots \end{aligned}$$

Como $V(u_t) = E(u_t^2 - (E(u_t))^2)$ e $E(u_t) = 0$, tem-se $V(u_t) = E(u_t^2)$. Segue-se que:

$$\begin{aligned} E(u_t^2) &= E\left((\varepsilon_t + \phi \varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + \dots)^2\right) \\ &= E(\varepsilon_t^2 + \phi^2 (\varepsilon_{t-1})^2 + \phi^4 (\varepsilon_{t-2})^2 + \dots) \\ &= E(\varepsilon_t^2) + \phi^2 E((\varepsilon_{t-1})^2) + \phi^4 E((\varepsilon_{t-2})^2) + \dots \\ &= \sigma_\varepsilon^2 + \phi^2 \sigma_\varepsilon^2 + \phi^4 \sigma_\varepsilon^2 + \dots \end{aligned} \quad (2.11)$$

Nota-se que, multiplicando $E(u_t^2)$ por ϕ^2 , tem-se:

$$\phi^2 E(u_t^2) = \phi^2 \sigma_\varepsilon^2 + \phi^4 \sigma_\varepsilon^2 + \phi^6 \sigma_\varepsilon^2 + \dots$$

Somando-se membro a membro e termo a termo, $E(u_t^2)$ com $-\phi^2 E(u_t^2)$, obtém-se:

$$\begin{array}{r} E(u_t^2) = \sigma_\varepsilon^2 + \cancel{\phi^2 \sigma_\varepsilon^2} + \cancel{\phi^4 \sigma_\varepsilon^2} + \cancel{\phi^6 \sigma_\varepsilon^2} + \dots \\ -\phi^2 E(u_t^2) = \cancel{-\phi^2 \sigma_\varepsilon^2} - \cancel{\phi^4 \sigma_\varepsilon^2} - \cancel{\phi^6 \sigma_\varepsilon^2} - \dots \\ \hline (1 - \phi^2) E(u_t^2) = \sigma_\varepsilon^2 \end{array}$$

Portanto,

$$E(u_t^2) = \frac{\sigma_\varepsilon^2}{(1 - \phi^2)} = \sigma_u^2. \quad (2.12)$$

Observa-se também que para $h > 0$:

$$\begin{aligned} E(u_t u_{t-h}) &= E\left[\left(\varepsilon_t + \phi \varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + \dots\right) \left(\varepsilon_{t-h} + \phi \varepsilon_{t-h-1} + \phi^2 \varepsilon_{t-h-2} + \dots\right)\right] \\ &= \phi^h \sigma_\varepsilon^2 + \phi^{h+2} \sigma_\varepsilon^2 + \phi^{h+4} \sigma_\varepsilon^2 + \dots \\ &= \phi^h \left(\sigma_\varepsilon^2 + \phi^2 \sigma_\varepsilon^2 + \phi^4 \sigma_\varepsilon^2 + \dots\right) \end{aligned}$$

De acordo com as expressões (2.11) e (2.12), tem-se que:

$$\begin{aligned} E(u_t u_{t-h}) &= \phi^h E(u_t^2) \\ &= \phi^h \frac{\sigma_\varepsilon^2}{1 - \phi^2} \\ &= \phi^h \sigma_u^2 = \text{Cov}_u. \end{aligned}$$

Sendo assim, de maneira análoga ao método dos mínimos quadrados ponderados, encontra-se $\hat{\beta}$.

$$\hat{\beta} = (X'V^{-1}X)^{-1} X'V^{-1}Y,$$

em que

$$V = \frac{1}{1-\phi^2} \begin{bmatrix} 1 & \phi & \phi^2 & \dots & \dots \\ \phi & 1 & \phi & \dots & \dots \\ \phi^2 & \phi & 1 & \dots & \dots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi^{n-1} & \phi^{n-2} & \phi^{n-3} & \dots & \dots \end{bmatrix}.$$

De acordo com Morettin & Toloi (1987), pode-se representar o modelo $AR(2)$, ou seja, se os erros forem autocorrelacionados na forma de um processo autorregressivo estacionário de segunda ordem, da seguinte forma:

$$u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + \varepsilon_t,$$

em que ϕ_1 e ϕ_2 são parâmetros de autocorrelação e ε_t é o ruído branco, tem-se que u_t é estacionário se:

$$\begin{aligned} \phi_1 + \phi_2 &< 1 \\ \phi_2 - \phi_1 &< 1 \\ -1 &< \phi_2 < 1. \end{aligned}$$

Logo, tem-se que:

$$\sigma_u^2 = \frac{\sigma_\varepsilon^2}{1 - \phi_1 \rho_1 - \phi_2 \rho_2},$$

enquanto as funções de autocorrelação são dadas por:

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}, \quad k > 0,$$

com

$$\rho_0 = 1, \quad \rho_1 = \frac{\phi_1}{1 - \phi_2} \quad \text{e} \quad \rho_2 = \phi_2 + \frac{\phi_1^2}{1 - \phi_2}.$$

2.1.5 Adequação do modelo

Após estimar o modelo é necessário verificar se de fato representa o que deseja-se medir, isto é, o modelo expressa a realidade e, conseqüentemente, proporciona uma melhoria nas decisões (CORRAR *et. al.*, 2007). Portanto, avaliou-se: (i) a qualidade do ajuste por meio dos coeficientes de determinação (R^2) e determinação ajustado ($R^2_{ajustado}$); (ii) multicolinearidade entre as variáveis explicativas por meio da análise dos fatores de inflação da variância (*VIF*); (iii) pressuposições de normalidade, independência e homogeneidade de variâncias dos resíduos a partir dos testes de Shapiro-Wilk, Durbin-Watson e Bartlett, respectivamente.

2.1.5.1 Coeficientes de Determinação e Determinação Ajustado

Draper & Smith (1998), define o coeficiente de determinação como um avaliador da qualidade do ajuste do modelo, isto é, representa a proporção da variação total explicada pelo modelo de regressão, como segue:

$$R^2 = \frac{SQReg}{SQTotal} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad 0 \leq R^2 \leq 1$$

Na prática, quanto mais próximo de 1, melhor será o ajuste do modelo. A inclusão de variáveis independentes, mesmo com pouco poder explicativo sobre a variável dependente, aumentará o valor do R^2 . Neste caso, recomenda-se o $R^2_{ajustado}$, como medida alternativa do coeficiente de determinação, que penaliza a inclusão de regressores com baixo poder explicativo (SEBER, 1977).

O coeficiente de determinação ajustado é definido como (DRAPER & SMITH, 1998):

$$R^2_{ajustado} = 1 - \frac{QMErro}{QMTotal} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y})^2 / (n-p)}{\sum_{i=1}^n (y_i - \bar{y})^2 / (n-1)}.$$

em que n é o tamanho da amostra e p é o número de parâmetros do modelo.

2.1.5.2 Multicolinearidade

Outro aspecto importante no ajuste de modelos de regressão linear múltipla é a multicolinearidade. Objetiva-se investigar se há multicolinearidade entre as variáveis explicativas, visto que a forte correlação entre elas acarreta vários efeitos negativos no ajuste do modelo de regressão. A multicolinearidade é um problema comum em regressão linear múltipla, indicando que existe uma relação de linearidade entre as variáveis regressoras, prejudicando assim a estimação dos coeficientes de regressão. O problema de multicolinearidade torna a estimativa dos parâmetros imprecisa, por conta de um alto valor do erro padrão, o que não é conveniente estatisticamente (KUTNER et al., 2004; TAMHANE, DUNLOP, 2000).

A presença de multicolinearidade pode ser observada quando o coeficiente de determinação R_i^2 apresenta um alto valor, mas nenhum dos coeficientes da regressão é estatisticamente significativo. Portanto, uma das formas de detecção é avaliar o Fator de Inflação da Variância (*VIF*). Esse fator mede o grau de associação entre as variáveis explicativas, a partir do coeficiente de determinação do modelo de regressão ajustado apenas entre as variáveis independentes. O Fator de Inflação da Variância é definido como (BERK, 1977):

$$VIF_i = \frac{1}{1 - R_i^2},$$

em que: R_i^2 é o coeficiente de determinação da regressão da variável explicativa X_i sobre as outras variáveis explicativas com $i = 1, 2, \dots, k$, sendo k a quantidade de variáveis explicativas no modelo.

Pode-se observar que, quanto maior o R_i^2 , maior é o valor de *VIF*, indicando alta colinearidade.

Valores de VIF_i maiores que 10 correspondem a um coeficiente de determinação $R_i^2 > 0,90$ (KUTNER et al., 2004; TAMHANE & DUNLOP, 2000). Partindo dessa idéia/princípio de que coeficiente de determinação acima de 90% (entre variáveis explicativas) caracteriza presença de multicolinearidade, então, justifica-se o porquê de considerar valores de VIF_i maiores que 10, pois $VIF_i > 1/(1-0,9)=10$.

2.1.7 Seleção de variáveis

Os métodos de seleção têm sido desenvolvidos para identificar se é necessário a inclusão de todas as variáveis regressoras disponíveis ou incluir apenas um subconjunto destas variáveis. A seguir é apresentado alguns métodos de seleção, nestes métodos as variáveis a compor o modelo são identificadas sequencialmente pela adição ou eliminação de uma variável que exerce grande influência sob a soma de quadrados dos resíduos. O ideal é encontrar critérios que balanceiam as duas idéias seguintes, de tendência oposta: (i) por um lado, para melhorar o processo de previsão, maior número possível de variáveis é desejado, pela melhoria de ajustamento decorrente; (ii) por outro lado, devido aos eventuais custos de obtenção dos x_i 's e às dificuldades de interpretação e de manuseio de um modelo grande, o número de variáveis deve ser tão pequeno quanto possível (DRAPER & SMITH, 1998; CHARNET, 2008).

2.1.7.1 Método “passo atrás” (backward)

Este procedimento caracteriza-se por incorporar, inicialmente, todas as variáveis auxiliares em um modelo de regressão linear múltipla e percorrer etapas, nas quais uma variável por vez pode vir a ser eliminada. Se em uma dada etapa não houver eliminação de alguma variável, o processo é então interrompido e as variáveis restantes definem o modelo final (CHARNET, 2008).

Como critério de eliminação das variáveis, analisa-se as contribuições individuais das variáveis no modelo. A variável de pior desempenho é eliminada, a não ser que esta atenda a um critério mínimo. No julgamento de uma variável, comparamos o modelo completo com o modelo reduzido, pela retirada de tal variável. Observa-se a soma de quadrados de regressão extra assim como a estatística do teste de sua contribuição, dada por:

$$\frac{(SQRe g^c - SQRe g^r)}{\hat{\sigma}^2} \sim F_{(\alpha, 1, n-p)} \quad (2.13)$$

Sendo SQR_{eg}^c e $\hat{\sigma}^2$ calculadas sob o modelo completo e SQR_{eg}^r calculada sob o modelo reduzido. A estatística (2.13) testa a contribuição da variável após a inclusão das demais. A contribuição é significativa se o valor da estatística for maior que um quantil especificado da distribuição F com 1 e $(n - p)$ graus de liberdade, sendo p o número de parâmetros do modelo completo. Assim, se o valor da estatística for menor que esse quantil da distribuição F , a contribuição não é considerada significativa e o modelo reduzido deverá ser preferido. Se observarmos várias variáveis não significantes, apenas uma variável é eliminada em uma etapa (aquela cuja estatística do teste tiver o menor valor). Quando uma variável é eliminada, passamos para a nova etapa cujo modelo completo não contém a variável que foi descartada. Se todas as variáveis são significantes, o processo é concluído, e o modelo completo desta etapa é o modelo final (CHARNET, 2008).

2.1.7.2 Método “passo a frente” (forward)

Este procedimento caracteriza-se por considerar, inicialmente, um modelo de regressão linear simples com a variável independente de maior correlação linear com a variável resposta. Etapas se sucedem, quando uma variável por vez pode vir a ser incorporada. Se em uma etapa não houver uma inclusão, o processo interrompido e as variáveis selecionadas até esta etapa definem o modelo final (CHARNET, 2008).

Numa dada etapa temos um determinado modelo, que denominaremos modelo reduzido da etapa, e o comparamos com modelos em que uma nova variável é acrescentada. Se há um modelo de melhor desempenho que atenda a um critério mínimo exigido, a correspondente variável é incorporada ao elenco de variáveis já selecionadas. Assim, enquanto em uma etapa do procedimento “passo atrás” comparamos vários modelos reduzidos com um único modelo completo, em uma etapa do procedimento “passo a frente”, as comparações são feitas entre vários modelos completos e um único modelo reduzido. No julgamento de uma variável, comparamos o modelo reduzido com o modelo completo que inclui uma variável extra. A estatística (2.13) testa a importância desta inclusão.

Nesse sentido, numa dada etapa deste procedimento, os valores de SQR_{eg}^c e $\hat{\sigma}^2$ são diferentes para cada variável candidata à inclusão e o valor de SQR_{eg}^r permanece o mesmo. Se observarem várias variáveis significantes, ou seja, valores da estatística (2.13) maiores que o quantil da distribuição F com 1 e $(n-p)$ graus de liberdade, sendo p o número de parâmetros do modelo completo, incorpora-se a variável que corresponde ao maior valor da estatística (2.13). Por outro lado, se nenhuma variável é significativa, o processo é concluído, e o modelo reduzido desta etapa é o modelo final (CHARNET, 2008).

2.1.7.3 Método “passo a passo” (stepwise)

Este procedimento é uma generalização do procedimento “passo a frente”, quando após cada etapa de incorporação de uma variável, temos uma etapa em que uma das variáveis já selecionada pode ser descartada. As etapas de eliminação e incorporação de variáveis são efetuadas conforme descrito nos procedimentos anteriores. O procedimento chega ao final quando nenhuma variável é incluída ou descartada (CHARNET, 2008).

Após a decisão, a significância e adequação do modelo ajustado devem ser verificadas e a análise de resíduos deve ser conduzida. Os pressupostos a cerca dos resíduos são apresentados a seguir.

2.1.8 Análise de resíduos

Os resíduos de um modelo de regressão linear têm uma relação muito forte com a qualidade do ajuste, bem como com a confiabilidade dos testes estatísticos sobre os parâmetros do modelo. Nesse sentido, a análise de resíduos tem uma importância fundamental na verificação da qualidade dos ajustes de modelos. Basicamente, essa análise fornece evidências sobre possíveis violações nas suposições do modelo, tais como a de normalidade, homocedasticidade, e quando for o caso ainda fornece indícios de falta de ajuste do modelo proposto (CHARNET, 2008).

A normalidade dos resíduos foi avaliada pelo teste de Shapiro-Wilk, que tem como hipóteses: H_0 - Os resíduos seguem distribuição normal *versus* H_1 - Os resíduos não seguem distribuição normal.

A estatística do teste é:

$$W_c = \left[\sum_{i=1}^{n/2} a_i (e_{n-i+1} - e_i) \right]^2 / \sum_{i=1}^n e_i^2,$$

em que a_i são constantes geradas pelas médias, variâncias e covariâncias das estatísticas de ordem de uma amostra de tamanho n de uma distribuição Normal. A regra de decisão é rejeitar H_0 ao nível de significância α se valor $p < \alpha$ (SHAPIRO, 1965).

A independência dos resíduos foi avaliada pelo teste de Durbin-Watson, que tem como hipóteses: H_0 - Os resíduos do modelo não apresentam autocorrelação serial de ordem 1 *versus* H_1 - Os resíduos do modelo apresentam autocorrelação serial de ordem 1.

A estatística do teste é:

$$d = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2}, \quad 0 \leq d \leq 4.$$

A regra de decisão consiste em (GUJARATI, 2000):

1. Se $d < d_i$ há evidencia de autocorrelação positiva nos resíduos, rejeita-se a hipótese nula H_0 ;
2. Se $d > d_s$ existe evidência de que os erros não são positivamente correlacionados;
3. Se $d_i < d < d_s$, o teste é inconclusivo.
4. Se $(4-d) < d_i$ há evidencia de autocorrelação positiva nos resíduos, rejeita-se a hipótese nula H_0 ;
5. Se $(4-d) > d_s$ existe evidência de que os erros não são negativamente correlacionados;
6. Se $d_i < d < d_s$, o teste é inconclusivo.

em que d_i e d_s são valores críticos da Tabela de Durbin-Watson (GUJARATI, 2000).

A homogeneidade de variância dos resíduos foi avaliada pelo teste de Bartlett, que tem como hipóteses:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2 \text{ versus } H_1 : \sigma_a^2 \neq \sigma_b^2, a \neq b, a, b = 1, 2, \dots, k.$$

A estatística do teste é:

$$\chi_c^2 = \frac{(N-k) \ln(S_p^2) - \sum_{i=1}^k (n_i - 1) \ln(S_i^2)}{1 + \frac{1}{3(k-1)} \left[\left(\sum_{i=1}^k \frac{1}{n_i - 1} \right) - \frac{1}{N-k} \right]},$$

em que:

$$S_p^2 = \frac{\sum_{i=1}^k (n_i - 1) S_i^2}{N + k} \text{ e } S_i^2 = \frac{\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2}{n_i - 1}.$$

Sendo N o tamanho da amostra, k o número de níveis do fator e n_i é a repetição dentro de cada nível do fator.

A regra de decisão por meio da tabela é rejeitar H_0 ao nível de significância α se $\chi^2 > \chi_{(\alpha, k-1)}^2$; E a regra de decisão por meio do valor p é rejeitar H_0 ao nível de significância α se valor $p < \alpha$ (BARTLETT, 1937).

3 RESULTADOS E DISCUSSÕES

Para o ajuste do modelo de regressão linear múltipla considerou-se o custo total como variável dependente e as variáveis independentes foram: operação com máquinas, mão-de-obra temporária, mão-de-obra fixa, sementes, fertilizantes, defensivos, despesas pós-colheita, juros, depreciação de benfeitorias/instalações, depreciação de implementos, depreciação de máquinas, manutenção periódica de máquinas, encargos sociais, segurança do capital fixo e municípios.

Inicialmente, para selecionar as variáveis a compor o modelo de regressão múltipla avaliou-se a multicolinearidade por meio do *VIF*. As variáveis que apresentaram valores do *VIF* maiores que 10 foram excluídas da análise. Em seguida, por meio do critério de "backward" e do teste *F* parcial selecionou-se as variáveis dentre as restantes que apresentaram contribuição significativa para o modelo. Em síntese, as variáveis excluídas do modelo foram: Operações com Máquinas (OperMaquinas), Mão-de-obra temporária (MOBtemp), Defensivos, Despesas pós-colheita (despesaspos), Juros, Depreciação de benfeitorias/instalações (DepBenfInst), Depreciação de implementos (DepImplem), Depreciação de máquinas (DepMaq), Manutenção periódica de máquinas (ManPerMaq), Encargos sociais (EncSoc) e Segurança do capital fixo (SegCapFixo) pois, não foram significativas ao nível de 5% de significância.

Na Tabela 1 são apresentados valores de *VIF* (inferiores a 10) para o modelo final.

Tabela 1 – Valores do *VIF* das variáveis do modelo final.

Variável	<i>VIF</i>
MOBfixa (x_2)	2,8281
Sementes (x_4)	2,2150
Fertilizantes (x_5)	4,4978

Na Tabela 2 são apresentados os resultados do teste *F* parcial referente ao ajuste do modelo final por meio do Método dos Mínimos Quadrados Ordinários.

Tabela 2 – Resultados do teste F parcial do modelo final.

Coefficientes	GL	SQ	QM	<i>F</i>	valor p
MOBfixa (x_2)	1	6745112	6745112	855,23	<0,0000
Sementes (x_4)	1	2665676	2665676	337,99	< 0,0000
Fertilizantes (x_5)	1	799544	799544	101,38	< 0,0000
Município (x_{15})	1	82085	82085	10,41	0,0033
Resíduos	27	212946	7887		

Na Tabela 3 são apresentadas as estatísticas referentes ao ajuste do modelo final para a predição de custo total do milho nos municípios.

Tabela 3 – Estimativas do modelo final.

Coefficientes	Estimativa	Erro Padrão	t	valor p	R²	R² ajustado
Intercepto	143,7251	48,9705	2,935	0,0067	0,9797	0,9767
MOBfixa (x_2)	8,4379	2,2372	3,772	0,0008		
Sementes (x_4)	2,1753	0,2923	7,443	< 0,0000		
Fertilizantes (x_5)	1,9121	0,1842	10,381	< 0,0000		
Município (x_{15})	103,6415	32,1258	3,226	0,0033		

Com base na Tabela 3, ao nível de significância de 5%, todos os parâmetros do modelo estimado são significativos. Portanto, o modelo de regressão múltipla estimado para previsão do custo total da produção de milho é:

$$\hat{y} = 143,7251 + 8,4379x_2 + 2,1753x_4 + 1,9121x_5 + 103,6415x_{15} \quad (3.1)$$

Sendo que: $\hat{y}$ representa o custo estimado da produção total de milho; x_2 gastos com mão-de-obra fixa (MOBfixa); x_4 gastos com sementes (Sementes); x_5 gastos com fertilizantes (Fertilizantes); x_{15} ($x_{15} = 1$, se o município for Rio Verde e $x_{15} = 0$, se o município for Primavera do Leste).

A partir do modelo (3.1) pode-se estimar o valor esperado do custo total da produção dado que o município é Rio Verde ($x_{15} = 1$), tem-se:

$$\begin{aligned} E(\text{Custo} | \text{Rio Verde}) &= \beta_0 + \beta_2x_2 + \beta_4x_4 + \beta_5x_5 + \beta_{15}x_{15} \\ &= \beta_0 + \beta_2x_2 + \beta_4x_4 + \beta_5x_5 + \beta_{15}(1) \\ &= (\beta_0 + \beta_{15}) + \beta_2x_2 + \beta_4x_4 + \beta_5x_5 \end{aligned} \quad (3.2)$$

Analogamente, em relação ao município de Primavera do Leste ($x_{15} = 0$), tem-se:

$$\begin{aligned} E(\text{Custo} | \text{Primavera do Leste}) &= \beta_0 + \beta_2x_2 + \beta_4x_4 + \beta_5x_5 + \beta_{15}x_{15} \\ &= \beta_0 + \beta_2x_2 + \beta_4x_4 + \beta_5x_5 + \beta_{15}(0) \\ &= \beta_0 + \beta_2x_2 + \beta_4x_4 + \beta_5x_5 \end{aligned} \quad (3.3)$$

Ao comparar as equações (3.2) e (3.3), é possível verificar que o custo médio da produção de milho é uma função linear de x_2 (mão-de-obra fixa), x_4 (sementes) e x_5 (fertilizantes) para os municípios Primavera do Leste e Rio Verde. O que difere nas duas curvas é o ponto de intersecção com o eixo y (custo). Em relação ao município de Primavera do Leste o ponto de intersecção com custo é (β_0) , conforme equação (3.3); em relação ao município de Rio Verde é $(\beta_0 + \beta_{15})$, conforme equação (3.2).

A partir da equação de regressão múltipla estimada tem-se que o 143,7251 é a estimativa de β_0 e 103,6415 é a estimativa de β_{15} . Assim, o valor esperado, do custo da produção de milho dado que o município é Rio Verde, é:

$$E(\text{Custo} | \text{Rio Verde}) = (143,7251 + 103,6415) + 8,4379x_2 + 2,1753x_4 + 1,9121x_5 \quad (3.4)$$

Analogamente, em relação ao município Primavera do Leste, tem-se:

$$E(\text{Custo} | \text{Primavera do Leste}) = 143,7251 + 8,4379x_2 + 2,1753x_4 + 1,9121x_5 \quad (3.5)$$

Por meio da análise de regressão foi possível verificar a variação de custo por hectares da produção de milho entre os municípios Primavera do Leste e Rio Verde.

Como β_{15} é positivo, o custo médio da produção de milho na cidade de Rio Verde é maior que o da cidade de Primavera do Leste.

Na Tabela 4 são apresentados os resultados dos testes de normalidade, independência e homogeneidade de variâncias dos resíduos a partir dos testes de Shapiro-Wilk, Durbin-Watson e Bartlett, respectivamente.

Tabela 4 – Resultados dos testes de normalidade, independência e homogeneidade de variâncias dos resíduos.

Testes	Estatística	valor p
Shapiro-Wilk	0,9542	0,1888
Durbin-Watson	1,7759	0,1185
Bartlett	0,1278	0,7207

De acordo com os resultados apresentados na Tabela 4, ao nível de 5% de significância, as pressuposições de normalidade, independência e homogeneidade dos resíduos do modelo estimado foram comprovadas por meio dos testes de Shapiro-Wilk, Durbin Watson e Bartlett, respectivamente.

4 CONCLUSÃO

Neste trabalho foi possível analisar a variação de custo por hectares (R\$/ha) da produção de milho nas cidades Rio Verde - GO e Primavera do Leste - MT. Verificou-se que, além da variável dummy para região ser significativa no modelo, os custos com “mão-de-obra fixa”, “sementes” e “fertilizantes” foram significativos para a predição do custo total de produção. O uso de uma variável dummy para as cidades Primavera do Leste e Rio Verde produz duas equações que podem ser utilizadas para prever e comparar o custo de produção de milho entre os municípios. Os resultados nos mostram que o município de Rio Verde possui em média um custo de R\$ 103,64 a mais do que Primavera do Leste na produção de milho. O modelo de regressão estimado, equação (3.1), apresentou ótimo ajuste captando cerca de 97,67% da variação total do custo de produção do milho.

REFERÊNCIAS

ANDERSON, D. R.; SWEENEY, D. J.; WILLIAMS, T. A. **Estatística aplicada à administração e economia**. 2. Ed. São Paulo: Thomson Learning, 2007.

BARTLETT, M. S. **Properties of sufficiency and statistical tests**. Proceedings of the Royal Society of London, serie A, London, 1937.

BERK, K. N. **Tolerance and condition in regression computations**, Journal of the American Statistical Association, 72 (360), 863-866, 1977.

CHARNET, R. et al. **Análise de modelos de regressão linear: com aplicações**. 2. ed. Campinas, São Paulo: Editora da Unicamp, 2008.

CONAB, Companhia Nacional de Abastecimento. **Custos de produção - Culturas de Verão - Série Histórica**. Disponível em:
<http://www.conab.gov.br/conteudos.php?a=1555&t=2>. Acesso em: 6 de abril de 2016.

CORRAR, S. L.; PAULO, E.; DIAS FILHO, J.M. (Coord.). **Análise multivariada para cursos de administração, ciências contábeis e economia**. São Paulo: Atlas, 2007, 539p.

DRAPER, N.R.; SMITH, H. **Applied regression analysis**. 3. ed. New York: John Wiley e Sons, 1998. 706p.

FÁVERO, L. P.; BELFIORE, P.; SILVA, F. L.; CHAN, B. L. **Análise de dados: modelagem multivariada para tomada de decisões**. Rio de Janeiro: Elsevier, 2009.

GUJARATI, D. N. **Econometria básica**. São Paulo: Makron Books, 2000.

HAIR, J. F.; ANDERSON, R.E.; TATHAM, R. L.; BLACK, W. C.; BABIN, B. J. **Análise multivariada de dados**. 5. Ed. Porto Alegre: Bookman, 2005.

HANSEN, D. R.; MOWEN, M. M. **Cost management**. 2. Ed. Edition, Cincinnati, Ohio: South-Western College Publications, 1997.

HOFFMAN, R.; VIEIRA, S. **Análise de regressão: uma introdução à econometria**. 2. Ed. São Paulo: Câmara Brasileira do Livro, 1998.

KUTNER, M. H. et al. **Applied linear models**. 5th ed. New York: McGraw-Hill Irwin, 2004.

MARTINS, E.; **Contabilidade de custos: Inclui o ABC**. 8. Ed. São Paulo: Atlas, 2006.

MATOS, O.C. **Econometria básica**. São Paulo: Atlas, 124p, 2000.

MORETTIN, P. A.; TOLOI, C. M. C. **Séries temporais**. 2. Ed., Atual: São Paulo, 1987.

SANTOS, J. J. **Contabilidade e Análise de custos:** Modelo contábil, Método de depreciação, ABC – Custeio Baseado em Atividades, análise atualizada de encargos sociais sobre salários. 5. Ed. São Paulo: Atlas, 2009.

SANTOS, J. L.; SCHMIDT, P., & PINHEIRO, P. R. **Fundamentos de Gestão Estratégica de Custos.** São Paulo: Atlas S/A. 2006.

SEBER, G. A. F. **Linear Regression Analysis.** New York: Wiley, 1997.

SHAPIRO, S. S.; WILK, M. B. **An Analysis of Variance Test for Normality.** Biometrika, Vol. 52, pp. 591-611, 1965.

TAMHANE, A.; DUNLOP, D. **Statistics and data analysis: From Elementary to Intermediate.** Prentice Hall, 2000.

R Core Team (2016). **R: A language and environment for statistical computing.** R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.