

# Teste para verificar a igualdade de modelos de regressão e uma aplicação na área médica

Sérgio Ricardo Magalhães<sup>1</sup>, Ednaldo Antônio de Andrade<sup>2</sup>

<sup>1</sup>Prof. e Coordenador do Curso de Sistemas de Informação da Universidade Vale do Rio Verde - UninCor

<sup>2</sup>Prof. da Universidade Federal do Mato Grosso - UFMT

*Resumo – Neste trabalho discutimos os métodos da Identidade de Modelos e o das Variáveis Dummy usados na comparação de modelos de regressão. Considerou-se o ajustamento de  $h$  equações de regressão linear para verificar a igualdade dos modelos de regressão por meio de simulação de dados. Utilizando-se os recursos do Interactive Matrix Language (IML), do sistema SAS®, foram desenvolvidas rotinas apropriadas para a metodologia de comparação de modelos de regressão. Foi realizada uma simulação de dados composta de 100.000 experimentos, considerando diferentes tamanhos de amostras (10, 50 e 100 observações). O desempenho de ambos os métodos mostrou-se praticamente equivalentes com relação aos diferentes tamanhos de amostras. Os resultados de todos os casos simulados pelos métodos foram apresentaram baixos percentuais de Erro Tipo I e Erro Tipo II. Para amostras maiores, a aproximação dada pela estatística  $F$  deve ser a preferida, uma vez que a taxa de Erro tipo I e Erro tipo II foi sempre menor. O Método das Variáveis Dummy foi o mais eficiente para os três tamanhos de amostra, pois, apresentou os menores percentuais de Erro Tipo I e Erro Tipo II.*

*Palavras-chave – modelos de regressão linear, simulação de dados, teste de hipótese*

*Abstract – In this paper we discuss the methods and models of identity of the dummy variables used in the comparison of regression models. It was the adjustment of  $h$  linear regression equations to check the equality of regression models by simulation data. Using the resources system, routines were of the Interactive Matrix Language (IML) of SAS developed for appropriate methodology for the comparison of regression models. We performed a simulation of data consisting of 100,000 experiments, considering different sizes of samples (10, 50 and 100 observations). The performance of both methods proved to be almost equivalent with the different sizes of samples. The results of all cases were simulated by the methods presented low percentage of Type I Error and Type II error. For larger samples, the approximation given by the  $F$  statistic should be preferred, since the rate of error type I and type II error was always lower. The method of dummy variables was the most efficient for the three sizes of sample, therefore, had the lowest percentage of error Type I and Type II error.*

*Keywords – linear regression models, simulation data, test hypothesis*

## I. INTRODUÇÃO

Os modelos de regressão linear têm aplicações nas mais diversas áreas do conhecimento<sup>(1)</sup>. Muitas vezes, um modelo linear é utilizado apenas pela facilidade em descrever o relacionamento aproximado<sup>(2)</sup>.

Em análise de regressão, frequentemente deseja-se saber se um conjunto de  $h$  equações ajustadas são idênticas, ou seja, se o fenômeno em estudo pode ser representado por uma única equação<sup>(3)</sup>.

Em dados da área médica, normalmente a variável resposta  $Y$  e o conjunto de variáveis regressoras  $X_i$ ,  $i=1, 2, \dots, n$  são medidas em um conjunto composto de grupos distintos, visando a comparação de como estes diferem segundo a relação de  $X_i$  e  $Y$ <sup>(4)</sup>. Isto pode ser realizado, a partir da geração de modelos de regressão para cada grupo e, em seguida verificando se as equações correspondentes são paralelas, ou se tem intercepto comum, ou ainda, se são idênticas<sup>(5)</sup>.

Vários autores apresentaram métodos para testar hipóteses relativas à igualdade de modelos lineares<sup>(6, 7, 8, 9 e 10)</sup>, entre outros.

Para se realizar comparações entre equações de regressão, existem diversos métodos. Entre eles, destacam-se Identidade de Modelos<sup>(11)</sup> e o das Variáveis Dummy<sup>(12)</sup> (binárias).

Desde modo, justifica-se o presente estudo que teve por objetivo avaliar os métodos da Identidade de Modelos e das Variáveis Dummy utilizados para a comparação entre equações de regressão lineares, por meio de simulação de dados, bem como verificar se existem divergências entre os métodos em estudo e suas aplicações práticas.

## II. METODOLOGIA

### Modelo estatístico

Consideraram-se, inicialmente o ajustamento dos dados de observação relativos a  $h$  grupos. Para cada um deles, foi ajustado o modelo de regressão linear dado por:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + \varepsilon_i \quad (1)$$

em que:

$y_i$ :  $i$ -ésimo valor da variável resposta,  $i = 1, 2, \dots, N$  observações;

$x_{ki}$ :  $i$ -ésimo valor da  $k$ -ésima variável explicativa,  $k=1, 2, \dots, K$  variáveis;

$\beta_k$ : parâmetros do modelo;

$\varepsilon_i$ : erros aleatórios.

Empregando a notação matricial, o modelo admite a seguinte forma:

$$y = X\beta + \varepsilon \quad (2)$$

em que:

$y$ : vetor de observações, de dimensões  $N \times 1$ , sendo  $N$  o número de observações;

$X$ : matriz das variáveis explicativas, de dimensões  $N \times (K+1)$ , sendo  $K$  o número de variáveis explicativas;

$\beta$ : vetor de parâmetros, de dimensões  $(K+1) \times 1$ , sendo  $(K+1)$  o número de parâmetros;

$\varepsilon$ : vetor de erros aleatórios, de dimensões  $N \times 1$ .

Para a estimação do vetor de parâmetros  $\beta$ , comumente são empregados o método dos quadrados mínimos e o método da máxima verossimilhança, que conduzem aos mesmos estimadores.

De acordo com as pressuposições que os erros podem assumir, existem variações no método de estimação dos quadrados mínimos para o modelo de regressão linear, relativa às diversas formas que a matriz de variâncias e

covariâncias pode assumir. Estas variações são conhecidas como métodos dos quadrados mínimos ordinário, ponderado e generalizado.

No ajuste de um modelo pelo método dos quadrados mínimos ordinários, pressupõe-se que a média dos erros é nula ( $E(\varepsilon_i) = 0$ ); a variância do erro  $\varepsilon_i$ ,  $i=1, 2, \dots, n$  é constante e igual a  $\sigma^2$ ; o erro de uma observação é não correlacionado com o erro de outra observação. Isto é,  $E(\varepsilon_i \varepsilon_j) = 0$ , para  $i \neq j$  e os erros são variáveis aleatórias com distribuição normal<sup>(6)</sup>.

Com base no método dos quadrados mínimos ordinários, estima-se um vetor  $\beta$ , considerando-se como condição que a soma de quadrados dos erros seja mínima. A função quadrática  $Z$ , que representa a soma de quadrados dos erros, é:

$$Z = \varepsilon' \varepsilon = (y - \beta X)' (y - \beta X) \quad (3)$$

Derivando parcialmente em relação a  $\beta$  obtém-se o seguinte sistema de equações normais<sup>(7)</sup>:

$$X'X\hat{\beta} = X'y \quad (4)$$

Como a matriz  $X$  é de posto coluna completo, então  $X'X$  é uma matriz positiva definida e, assim,  $X'X$  é não singular. Portanto, existe a matriz inversa  $(X'X)^{-1}$  e a solução para  $\beta$ , é:

$$\hat{\beta} = (X'X)^{-1} X'y \quad (5)$$

Esta solução única corresponde ao estimador linear não-tendencioso e de variância mínima para  $\beta$ .

### Método da Identidade de Modelos

Trata-se de um teste bastante geral que verifica a igualdade de duas regressões lineares, cujo algoritmo segue os seguintes passos<sup>(11)</sup>:

1. Dadas as seguintes relações lineares:

$$\begin{aligned} y_{1i} &= a_1 + b_1 x_{1i} + e_{1i} & i &= 1, \dots, n_1 \\ y_{2i} &= a_2 + b_2 x_{2i} + e_{2i} & i &= 1, \dots, n_2 \end{aligned} \quad (6)$$

referentes a dois conjuntos de observações.

2. Combinam-se todas as  $n_1 + n_2$  observações e calcula-se a estimativa de quadrados mínimos de  $a$  e  $b$  na regressão combinada  $y = a + bx + e$ . Desta equação obtém-se a soma de quadrados de resíduo ( $S_1$ ) com grau de liberdade igual a  $n_1 + n_2 - p$ , em que  $p$  é o número de parâmetros a ser estimado. Neste caso,  $p = 2$ .
3. Obtém-se a soma de quadrados de resíduo para as duas equações, ou seja,  $S_2$  e  $S_3$ , com os graus de liberdade  $n_1 - p$  e  $n_2 - p$ , respectivamente. Somam-se estas duas somas de quadrados de resíduo, isto é,  $S_4 = S_2 + S_3$  e seus graus de liberdade  $n_1 + n_2 - 2p$ .
4. Obtém-se  $S_5 = S_1 - S_4$ .
5. Calcula-se a estatística  $F$  como:

$$F_c = \frac{S_5/p}{S_4/(n_1 + n_2 - 2p)} \quad (7)$$

com  $p$  e  $n_1 + n_2 - 2p$  graus de liberdade.

Se  $F_c > F$  tabelado, para um determinado nível de significância  $\alpha$ , rejeita-se a hipótese de que os parâmetros  $a$ 's e  $b$ 's são os mesmos para os dois conjuntos de observações.

### Método das Variáveis Dummy

A inclusão de variáveis binárias aditivas ou multiplicativas, permite verificar se duas equações lineares diferem em intercepto, em inclinação, ou ainda em ambos.

Seja a seguinte relação, referente a dois conjuntos de<sup>(12)</sup>:

$$y_i = \alpha_0 + \alpha_1 D + \alpha_2 x_i + \alpha_3 (Dx_i) + e_i \quad (8)$$

$i = 1, \dots, (n_1 + n_2)$

em que:

$D=1$  para observações do primeiro conjunto ( $n_1$  observações)

$D=0$  para observações do segundo conjunto ( $n_2$  observações)

As variáveis binárias foram introduzidas na forma aditiva e multiplicativa. Os coeficientes  $\alpha_1$  e  $\alpha_3$  são diferenças de interceptos e inclinações, respectivamente.

Se  $H_0: \alpha_1=0$  é rejeitada, ou seja,  $\alpha_1$  é significativo, então o valor do intercepto do primeiro conjunto é obtido por  $\alpha_0 + \alpha_1$ , neste caso  $\alpha_0$  é o intercepto do segundo conjunto. Se  $H_0: \alpha_1=0$  não é rejeitada, ou seja,  $\alpha_1$  é não significativo, então  $\alpha_0$  representa o intercepto comum para ambos os conjuntos.

Se  $H_0: \alpha_3=0$  é rejeitada, então o valor da inclinação do primeiro conjunto é obtido por  $\alpha_2 + \alpha_3$ , neste caso  $\alpha_2$  é a inclinação do segundo conjunto. Se  $H_0: \alpha_3=0$  não é rejeitada, então  $\alpha_2$  representa a inclinação comum para ambos os conjuntos.

### Simulação dos métodos

Realizou-se uma simulação de dados composta de 100.000 experimentos, cada qual com 10, 50 e 100 observações.

Para cada experimento, foram gerados modelos de regressão linear simples nos quais os valores das variáveis independentes foram obtidos em um intervalo fechado de 0 a 10, aleatoriamente, pela função RANUNI do sistema SAS<sup>®</sup> (13) que produz uma variável aleatória com distribuição uniforme no nesse intervalo.

Para a geração dos resíduos de cada modelo, foi necessário estimar a variância dos mesmos<sup>(14)</sup>. Fixando-se o coeficiente de determinação  $R^2$  em 90 %, e conhecida a relação

$$R^2 = \frac{\delta_{\text{modelo}}^2}{\delta_{\text{modelo}}^2 + \delta_{\text{erro}}^2}, \text{ em que } \delta_{\text{modelo}}^2 \text{ corresponde à}$$

média dos valores das variáveis dependentes, estimou-se a variância dos resíduos  $\delta_{\text{erro}}^2$  (15).

Estimada a variância dos resíduos  $\delta_{erro}^2$ , geraram-se pela função RANNOR do sistema SAS<sup>®</sup> que produz uma variável aleatória com distribuição normal, os resíduos aleatórios de cada modelo. Estes são, supostamente, independentes e normalmente distribuídos, com média zero e variância comum, isto é,  $\varepsilon_{hi} \sim \text{NID}(0, \delta_{erro}^2)^{(16)}$ .

Com base nos modelos de regressão considerados, e fixando-se os parâmetros de cada modelo para cada uma das situações descritas anteriormente para a comparação dos métodos, foram implementados computacionalmente os métodos da identidade de modelos e das variáveis pelo módulo matricial "*Interactive Matrix Language – IML*" do sistema SAS<sup>®</sup>.

### III. RESULTADOS

Para a avaliação dos métodos, foram considerados quatro casos de regressão linear, representados por (a) o caso mais geral, quando todos os coeficientes são diferentes; (b) regressões paralelas, quando as inclinações são iguais, mas os interceptos são diferentes; (c) regressões concorrentes, quando os interceptos são iguais, mas as inclinações são diferentes e (d) regressões coincidentes, quando todas as retas coincidem.

Os resultados foram analisados com base no procedimento FREQ do módulo BASE, do *Statistical Analysis System* (SAS) e foram determinadas as frequências dos resultados obtidos para os níveis de significância nominal. Estes resultados foram encontrados para os valores do teste F nos modelos para amostras de tamanho 10, 50 e 100, respectivamente.

A avaliação dos métodos da Identidade de Modelos, das Variáveis Dummy e da baseouse no nível nominal de 5 % dos percentuais das taxas de ocorrência do Erro Tipo I, que consiste na rejeição de uma hipótese  $H_0$  tida verdadeira e nos percentuais das taxas de ocorrência do

Erro Tipo II, que consiste na não-rejeição de uma hipótese inicial  $H_0$ , tida como falsa.

A Tabela I ilustra todas as nove situações simuladas utilizando-se os dois métodos em estudo, apresentando uma combinação das frequências do Erro Tipo I e do Erro Tipo II.

**Tabela I**  
DISTRIBUIÇÃO DE FREQUÊNCIAS DE ERRO TIPO I E ERRO TIPO II PARA OS MÉTODOS UTILIZADOS

Casos	M É T O D O S					
	Identidade Modelos (Nº.de Observações)			Variáveis Dummy (Nº.de Observações)		
	10	50	100	10	50	100
a	2811	1126	1219	2415	1001	1003
b	1342	487	219	1083	308	115
c	1312	448	371	1084	487	242
d	1053	1002	18	3101	185	12
Subtotal	6518	3063	1827	7683	1981	1372
Total	11408			9055		

Verifica-se que os métodos da Identidade de Modelos e das Variáveis Dummy sinalizam para resultados bem semelhantes, devido a baixos percentuais de Erro Tipo I e Erro Tipo II.

Pode-se notar que, de modo geral, foram percebidas maiores taxas de combinação de Erros Tipo I e Erros Tipo II nos casos em tamanho da amostra é igual a 10 observações, com uma aparente vantagem para o Método das Identidade de Modelos.

Esperava-se que, com o aumento do número de observações uma redução nas taxas de Erro Tipo I e Tipo II. Fato este que, em geral, ocorreu, mostrando uma melhor eficiência dos métodos para tamanhos de amostras maiores. Por exemplo, para o Método das Variáveis Dummy, verificaram-se menores taxas com tamanho de amostra de 100 observações. Em geral, amostras com 100 observações apresentaram menores taxas de erros, mas estes valores não são bem diferentes dos valores dos outros tamanhos de amostras.

Pôde-se também verificar que para todas as situações estudadas, em todas elas foram observados indícios uma boa precisão para os três métodos estudados, face aos baixos

percentuais de Erro tipo I e Erro tipo II. Contudo, deve-se ressaltar que, para o Método das Variáveis *Dummy*, obteve-se menor probabilidade de ocorrência de Erro Tipo I e de Erro Tipo II.

### III. ILUSTRAÇÃO DO MÉTODO

Julgou-se necessário e adequado, a apresentação de um exemplo numérico para ilustrar os resultados obtidos neste estudo.

Assim, com base nos dados da Tabela 1, foram efetuados cálculos, ilustrando os métodos apresentados. Foram analisados dados coletados no período de 2004 a 2005 provenientes de uma amostra de doadores de sangue do Hemocentro do Hospital Universitário Mário Penna da Universidade Vale do Rio Verde de Belo Horizonte, em pacientes de ambos os sexos.

Para a realização da comparação das metodologias propostas, ajustaram-se retas de regressão da variável pressão sanguínea sistólica versus idade, para uma amostra de 1500 homens e 1500 mulheres, com o objetivo de verificar se estas variáveis tem relação linear semelhante para ambos os sexos.

Nos países industrializados, a pressão arterial média da população aumenta com a idade. Após os 50 anos, a pressão sistólica tende a subir com a idade mais avançada<sup>(17)</sup>, caracterizando em um quadro de Hipertensão Arterial Sistólica<sup>(18)</sup>. Assim, o aumento da pressão sistólica é bem estabelecido como fator de risco cardiovascular<sup>(19)</sup>.

Portanto, nesta aplicação, procurou-se confirmar as teorias que demonstram que a pressão sistólica aumenta continuamente com a idade em ambos sexos<sup>(20)</sup>.

Através do programa de análise estatística SAS®, considerou-se os casos a seguir:

- a) Interceptos diferentes e inclinações iguais;
- b) Interceptos iguais e inclinações

diferentes;

- c) Interceptos e inclinações diferentes;
- d) Interceptos e inclinações iguais.

Explorou-se os dois métodos para a comparação de Modelos de regressão linear, aos quais foram aplicados testes de hipóteses para identificação das situações acima.

Para aplicar o Método da Identidade de Modelos, ajustou-se as retas para cada um dos sexos:

$$\text{Masculino: } \hat{Y}_{ms} = 99,81 + 0,48x$$

$$\text{Feminino: } \hat{Y}_{fen} = 105,14 + 0,37x$$

E, as estimativas dos parâmetros para os dois sexos foram registrados na Tabela II.

**Tabela II**  
ESTIMATIVAS DOS PARÂMETROS PARA OS MODELOS ESTIMADOS DE IDADES *VERSUS* PRESSÃO SISTÓLICA

Grupo	$\hat{\beta}_0$	$\hat{\beta}_1$	$\bar{x}$	$s_x^2$	$s_{Y/X}^2$
Masculino	99,81	0,48	31,08	105,21	328,25
Feminino	105,14	0,37	31,05	115,44	254,81

Para o Método das Variáveis *Dummy*, ajustou-se um modelo de regressão de todo o conjunto e, em seguida, este foi separado, originando um modelo para cada sexo, através da inclusão das variáveis *dummy*.

$$D = \begin{cases} 0, & \text{Se o indivíduo for homem} \\ 1, & \text{Se o indivíduo for mulher} \end{cases} \quad (9)$$

Reta Geral:

$$\hat{Y} = 100,11 + 0,52x + 12,67D - 0,04xD$$

Reta Ajustada - Masculino:

$$\hat{Y}_{ms} = 100,01 + 0,52x \quad (D=0)$$

Reta Ajustada - Feminino:

$$\hat{Y}_{fen} = 113,41 + 0,49x \quad (D=1)$$

Na Tabela III é apresentada a análise de variância do ajuste de retas para esta situação.

**Tabela III**  
ANOVA PELO MÉTODO DAS VARIÁVEIS DUMMY PARA A VARIÁVEL IDADE VERSUS PRESSÃO SISTÓLICA

Fonte de variação	GL	SQ	QM	F
Regressão (x)	1	77071,12	7707,12	20,60
Resíduo	3005	1124445,00	374,19	
Regressão (x,D)	2	926547,00	463273,5	245,81
Resíduo	3004	1428954,00	475,70	
Regressão (x,d,xD)	3	155768,00	51912,67	165,84
Resíduo	3003	926158,00	308,41	

Considerando-se o Método da Identidade de Modelos, procedeu-se a identificação dos casos em que as retas estimadas se enquadravam nos testes do paralelismo e no da igualdade dos parâmetros conforme detalhamento abaixo.

a) Teste do paralelismo:

$$H_0 = \beta_{mas} = \beta_{fem}$$

$$S_{p,Y/X}^2 = 301,25 \text{ e } S_{\beta_{mas} - \beta_{fem}}^2 = 0,04.$$

A estatística de teste foi  $T=0,61$ . Para esta estatística, o valor crítico bilateral dado pelo  $p\text{-value}=2P(T \geq |0,61|) = 0,55$ . Considerando-se nível de significância nominal  $\alpha$  igual a 5%, observou-se que o valor  $p\text{-value} > \alpha$ . Portanto, a hipótese de nulidade não foi rejeitada, ou seja, tiveram evidências amostrais suficientes para que a hipótese de paralelismo não fosse rejeitada.

b) Teste da igualdade de interceptos:

$$H_0 = \beta_{0mas} = \beta_{0fem}$$

$$S_{p,Y/X}^2 = 301,25 \text{ e } S_{\beta_{0mas} - \beta_{0fem}}^2 = 5,01.$$

A estatística de teste foi  $T=-5,61$ . Para esta estatística, o valor crítico bilateral dado pelo  $p\text{-value}=2P(T \geq |-5,09|) \approx 0$ . Portanto, a hipótese de nulidade foi rejeitada para quaisquer níveis de

significância nominal de  $\alpha$ . Houve fortes evidências amostrais de que a hipótese igualdade de interceptos não seja verdadeira.

Em contrapartida, considerando-se o Método das Variáveis *Dummy*, procedeu-se a identificação dos casos em que as retas estimadas se enquadravam nos testes do paralelismo, no da igualdade dos parâmetros e no teste da coincidência, discriminados a seguir.

a) Teste do paralelismo:

$$H_0 = \beta_3 = 0$$

A estatística de teste foi  $F(XD/X,D)=0,52$ . O  $p\text{-value}$  com 1 e 3003 graus de liberdade foi igual a 0,46. Portanto, não foi rejeitada a hipótese de nulidade  $H_0$  para quaisquer valores nominais de  $\alpha$ . Logo, não existiram evidências amostrais para que a hipótese de paralelismo das regressões lineares fosse rejeitada.

b) Teste da igualdade de interceptos:

$$H_0 = \beta_2 = 0$$

A estatística de teste foi  $F(D/X,XD)=253,25$ . O  $p\text{-value}$  com 1 e 3003 graus de liberdade foi aproximadamente igual a zero. Portanto, foi rejeitada a hipótese de nulidade  $H_0$  para quaisquer valores nominais de  $\alpha$  diferentes de zero. Logo, perceberam evidências amostrais para que a hipótese de igualdade de interceptos das equações lineares dos dois sexos não fosse verdadeira.

c) Teste da coincidência:

$$H_0 = \beta_2 = \beta_3 = 0$$

A estatística de teste foi  $F(D/X)=121,68$ . O  $p\text{-value}$  com 2 e 3003 graus de liberdade foi aproximadamente igual a zero. Portanto, foi rejeitada a hipótese de nulidade  $H_0$  para

quaisquer valores nominais de  $\alpha$  diferentes de zero. Logo, não notaram evidências amostrais para que a hipótese de coincidência das regressões lineares estimadas para ambos os sexos.

#### IV. CONCLUSÕES

Verificou-se que para a amostra de dados referentes a pressão sanguínea sistólica e idade, submetidos às duas metodologias do estudo, revelaram que as retas estimadas para o sexo masculino e para o sexo feminino não foram coincidentes. As mesmas foram paralelas, com interceptos diferentes e admitiram a forma  $Y = \beta_0 + \beta_1 x + \varepsilon$ .

Foi possível verificar que a aplicação do Método da Identidade de Modelos foi equivalente ao Método da Variáveis *Dummy*.

Contudo, deve-se ressaltar que para as situações simuladas, para os três tamanhos de amostras, o Método das Variáveis *Dummy*, apresentou-se mais eficiente. Pois, o mesmo apresentou os menores percentuais de Erro Tipo I e Erro Tipo II, sinalizando que as diferenças encontradas são pequenas e portanto, devem ser atribuídas ao acaso.

#### AGRADECIMENTOS

Os autores agradecem ao Hemocentro do Hospital Universitário Mário Penna da Universidade Vale do Rio Verde - UninCor, pela cessão dos dados e também agradecem ao Centro Universitário de Belo Horizonte - UNIBH pela oportunidade de publicação deste estudo.

#### REFERÊNCIAS

[1] ARMITAGE, P.; BERRY, G. **Statistical Methods in Medical Research**. 4 ed. Blackwell, 2004.

[2] HOFFMANN, R.; VIEIRA S. **Análise de regressão: uma introdução à econometria**. 3 ed. São Paulo: HUCITEC, 1998. 379p.

[3] SEBER, G. A. F. **Linear regression analysis**. New York: John Wiley, 1977. 465p.

[4] DAVID C. S; HALL, D. B. A computer program for the regression analysis of ordered categorical repeated measurements. **Computer Methods and Programs in Biomedicine**. v. 51, p.153-169, 2005

[5] COPENHAVER, M. D; HOLLAND, B. S. Computation of the distribution of the maximum Studentized range statistic with application to multiple significance testing of simple effects. **Journal of Statistical Computing and Simulation**. v. 30, p. 1-15, 1998.

[6] CHOW, G. C. Tests of equality between sets of coefficients in two linear regressions. **Econometrica**, Chicago, v. 28, p. 591-605, 1960.

[7] CORDEIRO, G. M.; PAULA, G. A. **Modelos de regressão para análise de dados univariados**. Rio de Janeiro: Instituto de Matemática Pura e Aplicada do CNPq, 1989. 353p.

[8] NETER, J. et al. **Applied linear statistical models**. 4. ed. Chicago: Richard D. Irwin, 1996. 1408 p.

[9] GUJARATI, D. Use of dummy variables in testing for equality between sets of coefficients in two linear regressions: a note. **The American Statistician**, Washington, v. 24, n. 1, p. 50-52, Feb. 1970b.

[10] RATKOWSKY, D. A. **Nonlinear regression modeling: a unified practical approach**. New York : Marcel Dekker, 1983. 276p.

- [11] GRAYBILL, F. A. **Theory and application of the linear model.** Belmont: Duxbury Press, 1976. 704p.
- [12] GUJARATI, D. Use of dummy variables in testing for equality between sets of coefficients in linear regressions: a generalization. **The American Statistician**, Washington, v. 24, n. 5, p. 18-22, Dec. 1970a
- [13] SAS<sup>®</sup> INSTITUTE. **SAS Procedures guide for computers.** 6. ed. Cary N. C.: SAS<sup>®</sup> Institute, 1999. v. 3, 373p.
- [14] VERBEKE, G., MOLENBERGHS, G. **Linear mixed models in practice: a SAS-oriented approach.** Cary: SAS Institute, 1997. 306p.
- [15] LITTELL, R. C., HENRY, P. R., AMMERMAN, C. B. Statistical analysis of repeated measures data using SAS procedures. **J. Anim. Sci.**, v.76, p.1216-31, 1998.
- [16] BROWN, R. L.; DURBIN, J.; EVANS, J.M. Techniques for testing the constancy of regression relationships over time. **J. R.Stat. Soc., B**, v.37, p.149-92, 1975.
- [17] SESSO, H. D; STAMPFER; M. J.; ROSNER, B. et al. Systolic and diastolic blood pressure, pulse pressure, and mean arterial pressure as predictors of cardiovascular disease risk in men. **Hypertension**, v. 36, p. 801-807, 2000.
- [18] FRANKLIN, S.S; KHAN, S. A.; WONG; N.D.; LARSON, M. G., LEVY, D. Is pulse pressure useful in predicting risk for coronary heart disease? **The Framingham Heart Study. Circulation**, v. 100, p 354-360, 1999.
- [19] MILLAR, J.A; LEVER, A. F.; BURKE, V. Pulse pressure as a risk factor for cardiovascular events in the MRC Mild Hypertension trial. **J Hypertens**, v. 17, p. 1065-1072, 1999.
- [20] BENETOS, A.; RUDNICH, A.; SAFAR, M.; GUIZE, L. Pulse pressure and cardiovascular mortality in normotensive and hypertensive subjects. **Hypertension**, v.32 , p. 560-564, 1998.