UNIVERSIDADE FEDERAL DE LAVRAS

AVALIAÇÃO DE MÉTODOS PARA COMPARAÇÃO DE MODELOS DE REGRESSÃO POR SIMULAÇÃO DE DADOS

SÉRGIO RICARDO SILVA MAGALHÃES

2002

SÉRGIO RICARDO SILVA MAGALHÃES

AVALIAÇÃO DE MÉTODOS PARA COMPARAÇÃO DE MODELOS DE REGRESSÃO POR SIMULAÇÃO DE DADOS

Dissertação apresentada à Universidade Federal de Lavras como parte das exigências do Programa de Pós-Graduação em Agronomia, área de concentração em Estatística e Experimentação Agropecuária, para obtenção do título de "Mestre".

Orientador

Ruben Delly Veiga

LAVRAS
MINAS GERAIS - BRASIL
2002

Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da
Biblioteca Central da UFLA

Magalhães, Sérgio Ricardo Silva
Avaliação de métodos para a comparação de modelos de regressão por simulação de dados. -- Lavras : UFLA, 2002.
96 p. : il.

Orientador: Ruben Delly Veiga.
Dissertação (Mestrado) – UFLA.
Bibliografia.

1. Identidade de Modelos. 2. Variáveis Dummy. 3. Análise de Variância. 4. Simulação. I. Universidade Federal de Lavras. II. Título.

CDD-519.536

SÉRGIO RICARDO SILVA MAGALHÃES

AVALIAÇÃO DE MÉTODOS PARA COMPARAÇÃO DE MODELOS DE REGRESSÃO POR SIMULAÇÃO DE DADOS

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Agronomia, área de concentração em Estatística e Experimentação Agropecuária, para obtenção do título de "Mestre".

APROVADA em 27 de fevereiro de 2002.

Prof. Augusto Ramalho de Morais	UFLA
Prof. José Roberto Soares Scolforo	UFLA
Profa. Thelma Sáfadi	UFLA

Prof. Ruben Delly Veiga
UFLA
(Orientador)

LAVRAS
MINAS GERAIS – BRASIL

À memória de meu pai, Ely.

À minha mãe, Maria.

Ao meu irmão, Carlos Eduardo.

DEDICO

AGRADECIMENTOS

A Deus, por mais uma etapa vencida.

À Universidade Federal de Lavras (UFLA), em especial ao Departamento de Ciências Exatas, pela oportunidade concedida para a realização deste curso.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo auxílio financeiro.

Ao professor Ruben Delly Veiga, pela orientação, atenção e amizade.

À professora Thelma Sáfadi, pela co-orientação, contribuições e amizade.

Às secretárias do Departamento de Ciências Exatas da UFLA, Andréa e Ester, pela atenção e amizade.

Aos professores Joel, Eduardo, Thelma, Daniel, Júlio, Luís Henrique, Lucas e Mário, pelos ensinamentos.

Ao amigo Marcelo Cirillo, pela ajuda na elaboração dos programas utilizados neste trabalho.

Aos colegas de turma Ceile, Douglas, Ednaldo, Lívia, Flávio, Marcelo, José Marcelo, Paulo César e Paulo José. Também aos demais colegas de curso: Ermelino, Everton, Iara e Marcos.

A todos parentes e amigos que confiaram e apoiaram, em especial a Flávia, Márcia, Marília, Sandra e Tereza.

SUMÁRIO

RESUMO

MAGALHÃES, Sérgio Ricardo Silva. **Avaliação de métodos para comparação de modelos de regressão por simulação de dados.** Lavras: UFLA, 2002. 96p. (Dissertação - Mestrado em Estatística e Experimentação Agropecuária).*

O presente estudo teve como objetivo avaliar os métodos estatísticos da Identidade de Modelos, das Variáveis Dummy (binárias) e da Análise de Variância, usados para a comparação de modelos de regressão por meio de simulação de dados em computador. 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. Consideraram-se também cinco casos de regressão polinomial quadrática, sendo (a) o caso mais geral, quando todos os coeficientes são diferentes; (b) regressões que possuem o mesmo intercepto; (c) regressões que possuem o mesmo coeficiente relativo ao termo de 1° grau; (d) regressões que possuem o mesmo coeficiente referente ao termo de 2° grau e (e) regressões coincidentes, quando todas as curvas coincidem. Utilizando-se os recursos do módulo Interactive Matrix Language (IML), do Statistical Analysis System (SAS), foram desenvolvidas rotinas apropriadas para a implementação da metodologia de comparação de modelos de regressão, proporcionando uma maior facilidade na sua aplicação. Realizou-se uma simulação de dados composta de 10.000 experimentos, considerando os diferentes tamanhos de amostras (10, 50 e 100 observações), para cada uma das nove situações descritas anteriormente. Os resultados de todas as situações simuladas pelos três métodos foram semelhantes, apresentando baixos percentuais de Erro Tipo I e Erro Tipo II. 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.

---

* Orientador: Ruben Delly Veiga – UFLA, Co-orientadora: Thelma Sáfadi – UFLA

ABSTRACT

MAGALHÃES, Sérgio Ricardo Silva. **Evaluation of methods for comparing regression models by data simulation.** Lavras: UFLA, 2002. 96p. (Dissertation – Master in Statistics and Agricultural Experimentation).*

The present study was intended to evaluate the statistic methods of Models Identity, of the Dummy Variables (binary) and of the Variance Analysis, used for comparing regression models, by means of computer data simulation. Four linear regression cases were considered, stood for (a) the most general case, concurrent regressions, when all the coefficients are different; (b) parallel regressions, when the slopes are equal but the intercepts are different; (c) concurrent regressions, when the intercepts are equal but the slopes are different and (d) coincident regressions when all the straight lines coincide. Five cases of quadratic polynomial regression were also taken into consideration, (a) the most general case, when all the coefficients are different, (b) regressions which possess the same intercept; (c) regressions which possess the same coefficient relative to the 1st degree term; (d) regressions which possess the same coefficient concerning the 2nd degree term and (e) coincident regressions, when all the curves coincide. By utilizing the resources of de modulus of the Interactive Matrix Language (IML) of the Statistical Analysis System (SAS), appropriat routines were developed for implementation of the methodology, providing a greater ease in its application. A data simulation made up of 10.000 experiments, considering the different sample sizes (10, 50 and 100 observations) for each one of the nine situation reported. The results of all the situations simulated through the three methods were similar , presenting a low percent of Type I Error and Type II Error. The Dummy Variable Method proved most efficient for the three sizes of samples, for it presented the lowest percents of Type I Error and Type II Error .

* Adviser: Ruben Delly Veiga – UFLA, Co-adviser: Thelma Sáfadi – UFLA

INTRODUÇÃO

A análise de regressão é uma técnica potencialmente útil na análise de dados, tendo grande aplicação nas mais variadas áreas do conhecimento.

A aplicação da análise de regressão a um conjunto de dados é feita quando pretende-se estudar o comportamento de uma variável dependente em função de uma ou mais variáveis independentes.

Freqüentemente, estuda-se a produção de cultura de grãos de cultivares em função de doses de nitrogênio e estabelece-se que existe uma relação linear ou quadrática. Nestes casos, seria interessante verificar se os coeficientes de regressão entre as cultivares diferem entre si ou não. Isto porque as cultivares com maior coeficiente de regressão apresenta uma melhor resposta à aplicação do nutriente.

Outras aplicações ocorrem quando os dados são provenientes de diferentes grupos, seja pelo local, pela época ou pelo tratamento e a análise de regressão pode ser aplicada separadamente para cada grupo. Surge, então, a necessidade de comparar as equações de regressão, à verificação das semelhanças ou diferenças entre os modelos ou entre determinados coeficientes.

Quando se têm várias equações predizendo valores de uma mesma variável em condições distintas, algumas situações podem ser consideradas: As equações de regressão podem ser consideradas idênticas? Existirá uma equação comum para representar o conjunto? Os coeficientes de regressão dos vários conjuntos são estimadores de um mesmo coeficiente populacional? De que forma diferem as equações?

Análises referentes a essas situações são comuns e de fundamental importância nas áreas de experimentação agropecuária, econometria e biometria florestal. Para realizar comparações entre equações de regressão, existem

diversos métodos. Entre eles, destacam-se Identidade de Modelos, Variáveis Dummy (binárias), Análise de Variância e Comparações Múltiplas.

O presente trabalho teve por objetivo avaliar os métodos da Identidade de Modelos, das Variáveis Dummy (binárias) e da Análise de Variância, utilizados para a comparação entre equações de regressão lineares e quadráticas e/ou de seus coeficientes, empregando a simulação de dados. Pela padronização de rotinas e de teste, pretende-se verificar se existem divergências entre os métodos em estudo e suas aplicações práticas.

2 REFERENCIAL TEÓRICO

2.1 Modelos de regressão

Segundo Draper e Smith (1998), pode-se classificar os modelos de regressão, em relação aos seus parâmetros, em lineares, linearizáveis e não-lineares. Neste trabalho, interessam-nos os modelos lineares ou linearizáveis, com enfoque ao modelos de regressão linear e de regressão quadrática.

Um modelo de regressão linear, conforme Draper e Smith (1998) e Hoffmann e Vieira (1998), pode ser expresso como:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_k x_{ki} + \varepsilon_i$$

em que:

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

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

β_k: parâmetros do modelo;

ε_i: erros aleatórios.

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

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon$$

em que:

$\mathbf{y}$: vetor de observações, de dimensões N x 1, sendo N o número de observações;

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

$\boldsymbol{\beta}$: vetor de parâmetros, de dimensões (K+1) x 1 , sendo (K+1) o número de parâmetros;

$\boldsymbol{\varepsilon}$: vetor de erros aleatórios, de dimensões N x 1.

Para a estimação do vetor de parâmetros $\boldsymbol{\beta}$, 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 as diversas formas que a matriz de variâncias e covarâncias pode assumir. Estas variações são conhecidas como métodos dos quadrados mínimos ordinário, ponderado e generalizado.

Conforme Hoffmann e Vieira (1998), 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 ε_i, $i = 1,2,...,n$ é constante e igual a σ^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.

Com base no método dos quadrados mínimos ordinários, estima-se um vetor $\boldsymbol{\beta}$, considerando-se como condição que a soma de quadrados dos erros seja mínima. Como mostrado por Hoffmann e Vieira (1998), a função quadrática Z, que representa a soma de quadrados dos erros, é:

$$\mathbf{Z} = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = (\mathbf{y} - \boldsymbol{\beta}\mathbf{X})' \, (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

Derivando parcialmente em relação a β obtém-se o seguinte sistema de equações normais, conforme Graybill (1976):

$$\mathbf{X'X\hat{\beta} = X'y}$$

Como a matriz $\mathbf{X}$ é de posto coluna completo, então $\mathbf{X'X}$ é uma matriz positiva definida e, assim, $\mathbf{X'X}$ é não singular. Portanto, existe a matriz inversa $(\mathbf{X'X})^{-1}$ e a solução para β, de acordo com Draper e Smith (1998) e Hoffmann e Vieira (1998), é:

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

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

2.2 Métodos para comparação entre equações de regressão

O estudo de situações, por meio da análise de regressão, em que se faz a comparação entre dois ou mais conjuntos de observações n-dimensionais, tem sido descrito na literatura por Gujarati (1970a), Scolforo (1997), Draper e Smith (1998), Regazzi (1999), entre outros.

Normalmente, preocupa-se primeiramente em estabelecer se os conjuntos de observações, representados por equações de regressão linear, diferem entre si. Se for notada a diferença entre as equações, pode ser interessante avaliar em que ponto diferem, ou seja, quais coeficientes diferem de uma equação para outra.

Em contrapartida, se for notado que as equações não diferem entre si, significa que uma única equação pode ser utilizada para representar todos os conjuntos de observações. Em outras palavras, uma única equação pode ser estimada a partindo-se de todas as observações de todos os conjuntos envolvidos no estudo. Deste modo, pode-se considerar que as diferentes situações em estudo comportam-se da mesma forma. Se isto for verdadeiro, ter-se-á uma equação estimada com melhor precisão e mais confiável, quando comparado à estimação de equações individuais.

Diversos autores apresentaram testes para comparação entre equações de regressão c/ou coeficientes e também a sua utilização prática. Objetivando verificar a igualdade de duas regressões lineares, Chow (1960) sugeriu um teste geral, cujo algoritmo segue os seguintes passos:

1. Dadas as seguintes relações lineares:

$$y_{1i} = a_1 + b_1 x_{1i} + e_{1i} \qquad i = 1,...,n_1$$
$$y_{2i} = a_2 + b_2 x_{2i} + e_{2i} \qquad i = 1,...,n_2$$

   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:

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

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

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

Para Gujarati (1970b), o teste Chow (1960) permite uma avaliação geral da equação, assegurando apenas se duas regressões lineares são iguais ou diferentes. Caso sejam diferentes, não especificam se a diferença é devida a interceptos ou inclinações.

Uma comparação entre coeficientes de regressão, de maneira semelhante à de médias, foi sugerida por Fisher (1970), conduzindo aos mesmos resultados obtidos por Duncan (1970), comparando os coeficientes b_1 e b_2 de duas equações de regressão linear simples, através do teste t.

Brown(1970), para realizar a análise de regressão em H conjuntos de observações (x_{hi}, y_{hi}), considerou aos seguintes modelos de regressão:

$$y_{hi} = a_h + b_h x_{hi} + e_{hi} \qquad \begin{matrix} h = 1,...,H & \text{modelos} \\ i = 1,...,n_h & \text{observações} \end{matrix}$$

para os quais existe interesse em obter um modelo simplificado, em que todos os $b's$ e todos os $a's$ são idênticos. Utilizando regressão linear múltipla, foi realizado o ajustamento das observações, para o modelo reduzido, por meio do método dos quadrados mínimos, deduzindo novas variáveis.

Swamy e Metha (1979) demonstraram que, reunindo dados de duas equações de regressão, é possível obter estimativas mais eficientes do que as estimativas baseadas em cada uma das equações.

Battisti (2001) apresentou um estudo dos métodos estatísticos da Identidade de Modelos, das Variáveis Dummy, da Análise de Variância e da Análise de Agrupamento, usados para a comparação de equações de regressão. O autor realizou uma aplicação em biometria florestal, baseada no modelo volumétrico de Schumacher e Hall[1], abordando este modelo quando aplicado em diferentes estratos. Verifou que os métodos da Identidade de Modelos e das Variáveis Dummy são equivalentes e fornecem resultados mais objetivos.

2.2.1 Identidade de Modelos

Graybill (1976) apresentou um teste para verificar a identidade de H modelos lineares simples, do seguinte modo:

$$\begin{aligned} & y_{1i} = a_1 + b_1 x_{1i} + \varepsilon_{1i} \qquad i = 1,\ldots,n_1 \\ & y_{2i} = a_2 + b_2 x_{2i} + \varepsilon_{2i} \qquad i = 1,\ldots,n_2 \\ & \vdots \\ & y_{Hi} = a_H + b_H x_{Hi} + \varepsilon_{Hi} \qquad i = 1,\ldots,n_H \\ & \sum_{h=1}^{H} n_h = N, \qquad n_h > 2 \quad \text{para todo } h, \qquad \varepsilon_{ij} \sim NID(\varepsilon : 0, \sigma^2) \end{aligned} \tag{1}$$

[1] Procedimento volumétrico tradicional da literatura florestal mundial, que expressa o volume das árvores em função do diâmetro e da altura.

Partindo destes modelos, foram formuladas várias hipóteses e para cada uma apresentou os respectivos testes, a saber:

1. As H equações são paralelas.

Corresponde a testar se as equações possuem inclinações iguais, de acordo com a seguinte hipótese:

$H_0 : \beta_1 = \beta_2 = ... = \beta_H$ (as H linhas são paralelas)

$H_1 : \beta_h \neq \beta_{h'}$ para, pelo menos, um $h \neq h'$ $\quad (h,h' = 1,2,...,H)$

Rejeita-se H_0 se a estatística $W_p \geq F_{\alpha:H-1,N-2H}$, em que:

$$W_P = \frac{\sum_{h=1}^{H}\left[\hat{\beta}_h - \frac{\sum_{j=1}^{H}\hat{\beta}_j b_{jj}}{\sum_{i=1}^{H} b_{ii}}\right]^2 b_{hh}}{(H-1)\hat{\sigma}^2}$$

em que $b_{hh} = \sum_{t=1}^{n_h}\left(x_{ht} - \bar{x}_h\right)^2$.

2. $H_0 : \alpha_1 = \alpha_2 = ... = \alpha_H$ (as H linhas possuem o mesmo intercepto)

$H_1 : \alpha_h \neq \alpha_{h'}$ para, pelo menos, um $h \neq h'$

Rejeita-se H_0 se a estatística $W_I \geq F_{\alpha:H-1,N-2H}$

$$W_I = \frac{\sum_{h=1}^{H}\left[\hat{\alpha}_h - \frac{\sum_{j=1}^{H}\hat{\alpha}_j a_{jj}}{\sum_{i=1}^{H} a_{ii}}\right]^2 a_{hh}}{(H-1)\hat{\sigma}^2}$$

em que $a_{hh} = \frac{n_h \cdot \sum_{t=1}^{n_h}(x_{ht} - \bar{x}_h)^2}{\sum_{S=1}^{n_h} x_{hs}^2}$.

3. $H_0 : \alpha_1 + \beta_1 x_0 = \alpha_2 + \beta_2 x_0 = ... = \alpha_H + \beta_H x_0$ (as H linhas têm intercepto no ponto x_0 conhecido)

   H_1 : pelo menos uma linha não tem interceptos no ponto x_0 conhecido.

   Rejeita-se H_0 se a estatística $W_0 \geq F_{\alpha:H-1,N-2H}$, em que

$$W_O = \frac{\sum_{h=1}^{H}\left[(\hat{\alpha}_h + \hat{\beta}_h x_0) - \frac{\sum_{j=1}^{H}(\hat{\alpha}_j + \hat{\beta}_j x_0)c_{jj}}{\sum_{i=1}^{H} c_{ii}}\right]^2 c_{hh}}{(H-1)\hat{\sigma}^2}$$

em que $c_{hh} = \frac{n_h \cdot \sum_{t=1}^{n_h}(x_{ht} - \bar{x}_h)^2}{\sum_{s=1}^{n_h}(x_{hs} - x_0)^2}$.

Empregando notação matricial, Graybill (1976) derivou um teste para a hipótese em que os H modelos lineares são idênticos. Neste caso, considerou os H seguintes modelos lineares :

$$\begin{aligned}
\mathbf{y}_1 &= \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}_1 \\
\mathbf{y}_2 &= \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}_2 \\
&\vdots \\
\mathbf{y}_H &= \mathbf{X}_H\boldsymbol{\beta}_H + \boldsymbol{\varepsilon}_H
\end{aligned}$$

em que:

$\mathbf{y}_h$: vetor das observações do h-ésimo modelo, de dimensões $n_h \times 1$;

$\mathbf{X}_h$: matriz dos coeficientes do h-ésimo modelo, de dimensões $n_h \times p$;

$\boldsymbol{\beta}_h$: vetor de parâmetros do h-ésimo modelo, de dimensões $p \times 1$;

$\boldsymbol{\varepsilon}_h$: vetor dos erros aleatórios, do h-ésimo modelo, de dimensões $n_h \times 1$.

O modelo completo envolvendo todas as observações de todos conjuntos pode ser escrito como:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

em que:

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_H \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_H \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_H \end{bmatrix} \quad e \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_H \end{bmatrix}.$$

Então, a hipótese de que os H modelos são idênticos foi:

$H_0 : \beta_1 = \beta_2 = ... = \beta_H$ (os H modelos lineares são idênticos)

$H_1 : \beta_h \neq \beta_{h'}$ para, pelo menos, um $h \neq h'$.

Nesta situação, rejeita-se H_0 se a estatística dada por $W \geq F_{\alpha;(H-1)p,N-Hp}$.

em que:

$$W = \left(\frac{\sum_{h=1}^{H} y_h'(X_h X_h^-) y_h - (\sum_{i=1}^{H} y_i' X_i)(\sum_{h=1}^{H} X_h' X_h)^{-1}(\sum_{j=1}^{H} X_j' y_j)}{\sum_{h=1}^{H} y_h' y_h - \sum_{h=1}^{H} y_h' (X_h X_h^-) y_h} \right) \cdot \left(\frac{N - Hp}{(H-1)p} \right)$$

em que:

X^- : matriz inversa de Moore-Penrose;

p : número de parâmetros.

A estatística W segue uma distribuição F (Graybill, 1976), na qual a expressão do numerador representa a diferença entre a soma de quadrados de todos os parâmetros e a soma de quadrados de parâmetros de um modelo reduzido, em que os vetores β_h são considerados iguais.

Regazzi (1993) utilizou esta metodologia, considerando o ajustamento dos dados de observação relativos a H equações de regressão polinomial do segundo grau, empregando a técnica dos polinômios ortogonais. As H equações são dadas por:

$$
\begin{aligned}
y_{1i} &= a_1 + b_1 P_{11i} + c_1 P_{21i} + e_{1i} \\
y_{2i} &= a_2 + b_2 P_{12i} + c_2 P_{22i} + e_{2i} \\
&\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \\
y_{Hi} &= a_H + b_H P_{1Hi} + c_H P_{2Hi} + e_{Hi}
\end{aligned} \qquad (2)
$$

em que:

y_{hi}: i-ésima observação do h-ésimo modelo, sendo i = 1, 2,..., n_h o número de observações e h = 1, 2, ...,H o número de modelos;

a_h, b_h, c_h: parâmetros do h-ésimo modelo;

P_{khi}: polinômio de grau k, correspondente ao i-ésimo valor da variável independente do h-ésimo modelo;

e_{hi}: erro aleatório, associado à i-ésima observação do h-ésimo modelo, sendo $e_{hi} \sim \text{NID}(0, \sigma^2)$;

$\sum_{h=1}^{H} n_h = N$ e $n_h > 3$ para todo h.

O autor considerou as seguintes hipóteses de identidade:

$H_0 : \beta_1 = \beta_2 = ... = \beta_H$ (as H equações são idênticas),

$H_0 : a_1 = a_2 = ... = a_H$ (as H equações têm uma constante de regressão comum),

$H_0 : c_1 = c_2 = ... = c_H$ (as H equações têm os coeficientes de regressão do termo de segundo grau iguais).

O h-ésimo modelo em (2) pode ser escrito na forma matricial como:

$$\mathbf{y}_h = \mathbf{X}_h\,\boldsymbol{\beta}_h + \boldsymbol{\varepsilon}_h \tag{3}$$

em que:

$$\mathbf{y}_h = \begin{bmatrix} Y_{h1} \\ Y_{h2} \\ \vdots \\ Y_{hn_h} \end{bmatrix}_{n_h \times 1}, \quad \mathbf{X}_h = \begin{bmatrix} 1 & P_{1h1} & P_{2h1} \\ 1 & P_{1h2} & P_{2h2} \\ \vdots & \vdots & \vdots \\ 1 & P_{1hn_h} & P_{2hn_h} \end{bmatrix}_{n_h \times p}, \quad \boldsymbol{\beta}_h = \begin{bmatrix} a_h \\ b_h \\ c_h \end{bmatrix}_{p \times 1} \text{ e } \boldsymbol{\varepsilon}_h = \begin{bmatrix} e_{h1} \\ e_{h2} \\ \vdots \\ e_{hn_h} \end{bmatrix}_{n_h \times 1}.$$

Escrevendo esses H modelos na forma do modelo linear geral:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{4}$$

em que:

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_H \end{bmatrix}_{N \times 1}, \quad \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_H \end{bmatrix}_{Hp \times 1}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_H \end{bmatrix}_{N \times 1} \text{ e } \mathbf{X} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{X}_3 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_H \end{bmatrix}_{N \times Hp}.$$

Pelo método dos quadrados mínimos, obteve-se o seguinte sistema de equações normais relativo ao modelo (4):

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y} \tag{5}$$

ou

$$\begin{bmatrix} \mathbf{X}_1'\mathbf{X}_1 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2'\mathbf{X}_2 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{X}_3'\mathbf{X}_3 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_H'\mathbf{X}_H \end{bmatrix} \bullet \begin{bmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2 \\ \hat{\boldsymbol{\beta}}_3 \\ \vdots \\ \hat{\boldsymbol{\beta}}_H \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1'\mathbf{y}_1 \\ \mathbf{X}_2'\mathbf{y}_2 \\ \mathbf{X}_3'\mathbf{y}_3 \\ \vdots \\ \mathbf{X}_H'\mathbf{y}_H \end{bmatrix}$$

e, sendo a matriz $\mathbf{X'X}$ não singular, o estimador do vetor de parâmetros é:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X'X})^{-1}\mathbf{X'y} \qquad (6)$$

Como também a matriz $(\mathbf{X'X})^{-1}$ é bloco diagonal, em que cada bloco é a matriz inversa $(\mathbf{X}_h'\mathbf{X}_h)^{-1}$ de cada modelo, então (6) pode ser escrito do seguinte modo:

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2 \\ \vdots \\ \hat{\boldsymbol{\beta}}_H \end{bmatrix} = \begin{bmatrix} (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y}_1 \\ (\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{y}_2 \\ \vdots \\ (\mathbf{X}_H'\mathbf{X}_H)^{-1}\mathbf{X}_H'\mathbf{y}_H \end{bmatrix}$$

A soma de quadrados de parâmetros relativa ao modelo completo (4) é obtida por:

$$SQPar(c) = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} = \sum_{h=1}^{H} \hat{\boldsymbol{\beta}}_h' \, \mathbf{X}_h' \, \mathbf{y}_h \qquad (7)$$

com Hp graus de liberdade (H modelos, com p parâmetros cada um).

A soma de quadrados total é obtida por:

$$SQTotal(c) = \mathbf{y'y} = \sum_{h=1}^{H} \mathbf{y}_h'\mathbf{y}_h \qquad (8)$$

com N graus de liberdade.

A soma de quadrados de resíduo é obtida pela diferença:

$$\begin{aligned} \text{SQResíduo}(c) &= \mathbf{y'y} - \hat{\boldsymbol{\beta}}'\mathbf{X'y} \\ &= \sum_{h=1}^{H} \mathbf{y}_h' \mathbf{y}_h - \sum_{h=1}^{H} \hat{\boldsymbol{\beta}}_h' \mathbf{X}_h' \mathbf{y}_h \\ &= \sum_{h=1}^{H} (\mathbf{y}_h' \mathbf{y}_h - \hat{\boldsymbol{\beta}}_h' \mathbf{X}_h' \mathbf{y}_h) \end{aligned} \qquad (9)$$

então:

$$\text{SQResíduo}(c) = \sum_{h=1}^{H} \text{SQRes}(h)$$

com N-Hp graus de liberdade.

O esquema da análise de variância relativa ao modelo completo é apresentado na Tabela 1.

TABELA 1 – Esquema da análise de variância relativa ao modelo completo

CV	GL	SQ	QM
Parâmetros (β)	Hp	$\hat{\boldsymbol{\beta}}'\mathbf{X'y}$	
Resíduo (c)	N-Hp	$\mathbf{y'y} - \hat{\boldsymbol{\beta}}'\mathbf{X'y}$	$\frac{\text{SQRes}}{\text{gl}}$
Total	N	$\mathbf{y'y}$	

De acordo com Regazzi (1993), $\frac{SQ\,\text{Res}}{gl} = \hat{\sigma}^2$ é o estimador comum da variância residual. Ele também pode ser obtido pela média ponderada dos estimadores das variâncias residuais de cada modelo.

A seguir são apresentados os testes para as hipóteses, considerados por Regazzi (1993).

O primeiro teste considera a seguinte hipótese de nulidade:

$H_0 : \beta_1 = \beta_2 = ... = \beta_H$ (as H equações são idênticas), isto é, os modelos em (2) reduzem-se à forma:

$$y_{hi} = a + bP_{1hi} + cP_{2hi} + e_{hi} \qquad (10)$$

em que:

y_{hi}, P_{khi} e e_{hi} têm as mesmas especificações dos modelos em (2);

a, b, c : parâmetros comuns.

Empregando a notação matricial, os modelos reduzidos (10) podem ser escritos como:

$$\mathbf{y} = \mathbf{Z}\theta + \boldsymbol{\varepsilon} \qquad (11)$$

em que:

$\mathbf{y}$: vetor dos valores observados da variável resposta, de dimensão (N x 1);

$\boldsymbol{\varepsilon}$: vetor dos erros aleatórios, de dimensão (N x 1);

$\mathbf{Z} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_H \end{bmatrix}_{N \times p}$ em que $\mathbf{X}_h$ com h = 1, 2, ..., H, são iguais às matrizes definidas em (4);

$\theta = \begin{bmatrix} a \\ b \\ c \end{bmatrix}_{p\times 1}$ é o vetor dos parâmetros comuns.

Segundo Graybill (1976), Regazzi (1993) e Draper e Smith (1998), o sistema de equações normais relativo ao modelo reduzido (11), obtido pelo método dos quadrados mínimos, é:

$$\mathbf{Z'Z}\hat{\boldsymbol{\theta}} = \mathbf{Z'y} \qquad (12)$$

como $\mathbf{Z}$ tem posto coluna completo p, então $\mathbf{Z'Z}$ tem dimensão $p \times p$ e não-singular. Portanto, o estimador do vetor dos parâmetros para o modelo reduzido é:

$$\hat{\boldsymbol{\theta}} = (\mathbf{Z'Z})^{-1}\mathbf{Z'y}_{\mathbf{y}} \qquad (13)$$

A matriz $\mathbf{Z'Z}$ é composta pela soma das matrizes $\mathbf{X}_h'\mathbf{X}_h$ de cada modelo, bem como a matriz $\mathbf{Z'y}$. O estimador do vetor dos parâmetros comuns pode ser escrito do seguinte modo:

$$\hat{\boldsymbol{\theta}} = (\sum_{h=1}^{H} \mathbf{X}_h'\mathbf{X}_h)^{-1} \sum_{j=1}^{H} \mathbf{X}_j'\mathbf{y}_j \ .$$

A soma de quadrados de parâmetros relativa ao modelo reduzido é obtida por:

$$\text{SQPar(r1)} = \hat{\boldsymbol{\theta}}'\mathbf{Z'y} \qquad (14)$$

ou

$$\text{SQPar(r1)} = \left(\sum_{j=1}^{H} \mathbf{y}_j' \mathbf{X}_j\right)\left(\sum_{h=1}^{H} \mathbf{X}_h' \mathbf{X}_h\right)\left(\sum_{t=1}^{H} \mathbf{X}_t' \mathbf{y}_t\right)$$

com p graus de liberdade.

A redução devida a H_0 (coeficientes iguais) é obtida pela diferença:

$$\text{Redução } (H_0) = \text{SQPar(c)} - \text{SQPar(r1)} \qquad (15)$$

com (H -1)p graus de liberdade.

Neste caso, o autor testou a seguinte hipótese:

$H_0 : \beta_1 = \beta_2 = \ldots = \beta_H$ (as H equações são idênticas)

$H_1 : \beta_h \neq \beta_{h'}$ para pelo menos um $h \neq h'$

utilizando a estatística F, dada por:

$$F_c = \frac{[\text{SQPar}(c) - \text{SQPar}(r1)]/(H-1)p}{\text{SQRes(c)}/(N - Hp)} \qquad (16)$$

De acordo com Graybill (1976), a estatística (16) apresenta distribuição F central com (H-1)p e (N-Hp) graus de liberdade sob H_0 e normalidade dos erros.

O teste descrito pode ser visualizado na Tabela 2, referente à análise de variância. O critério de decisão considerado foi:

Rejeita-se H_0 se $F_c \geq F_{T[\alpha;(H-1)p,N-Hp]}$, em que $\sum_{h=1}^{H} n_h = N$.

Segundo o autor, a não rejeição de H_0 admite concluir que, a uma significância α, as H equações não diferem entre si. Logo, a equação ajustada com as estimativas dos parâmetros comuns pode ser usada como uma estimativa das H equações envolvidas. São obtidas, dessa forma e nesse caso, estimativas oriundas de amostras maiores, sugerindo que estas são mais confiáveis por apresentarem menores variâncias.

TABELA 2 – Análise de variância relativa ao teste de hipótese $H_0 : \beta_1 = \beta_2 = ... = \beta_H$ (as H equações são idênticas)

CV	GL	SQ	QM	F_c
Parâmetros (β)	(Hp)	$S_1 = \hat{\beta}'\mathbf{X}'\mathbf{y}$		
Parâmetros (θ)	p	$S_2 = \hat{\theta}'\mathbf{Z}'\mathbf{y}$		
Redução (H_0)	(H-1)p	$S_3 = S_1 - S_2$	$V_1 = \frac{S_3}{gl}$	$\frac{V_1}{V_2}$
Resíduo (c)	N-Hp	$S_4 = S_5 - S_1$	$V_2 = \frac{S_4}{gl}$	
Total	N	$S_5 = \mathbf{y}'\mathbf{y}$		

O segundo teste considerado por Regazzi (1993), baseando-se em Graybill (1976), refere-se à seguinte hipótese de nulidade:

$H_0 : a_1 = a_2 = ... = a_H$ (as H equações têm uma constante de regressão comum), isto é, os modelos em (2) reduzem-se à forma:

$$y_{hi} = a + b_h P_{1hi} + c_h P_{2hi} + e_{hi} \qquad (17)$$

em que:

a : parâmetro comum;

$y_{hi}, P_{khi}, b_h, c_h$ e e_{hi} têm as mesmas especificações dos modelos em (2).

A partição de $\boldsymbol{\beta}_h$ e $\mathbf{X}_h$ em (3) é:

$$\boldsymbol{\beta}_h = \begin{bmatrix} a_h \\ \hline \boldsymbol{\delta}_h \end{bmatrix} \qquad \text{e} \qquad \mathbf{X}_h = [\mathbf{u}_h \mid \mathbf{V}_h]$$

em que a_h possui dimensão 1 x 1 e $\boldsymbol{\delta}_h$ possui dimensão (p-1) x 1;

$\mathbf{u}_k$: vetor relativo ao termo constante a, no h-ésimo modelo, de dimensões $n_h \times 1$,

$\mathbf{V}_h$: matriz associada aos termos lineares e quadráticos, no h-ésimo modelo, de dimensões $n_h \times (p-1)$.

Empregando-se a notação matricial, os modelos reduzidos em (17) podem ser escritos como:

$$\mathbf{y} = \mathbf{B}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \qquad (18)$$

em que:

$$\mathbf{y}=\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_H \end{bmatrix}_{N\times 1}, \quad \boldsymbol{\gamma}=\begin{bmatrix} a \\ \boldsymbol{\delta}_1 \\ \boldsymbol{\delta}_2 \\ \vdots \\ \boldsymbol{\delta}_H \end{bmatrix}_{[H(p-1)+1]\times 1}, \quad \boldsymbol{\varepsilon}=\begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_H \end{bmatrix}_{N\times 1} \quad \text{e}$$

$$\mathbf{B}=\begin{bmatrix} \mathbf{u}_1 & \mathbf{V}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{u}_2 & \mathbf{0} & \mathbf{V}_2 & \cdots & \mathbf{0} \\ \mathbf{u}_3 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & & \vdots \\ \mathbf{u}_H & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{V}_H \end{bmatrix}_{N\times[H(p-1)+1]}.$$

O sistema de equações normais relativo ao modelo reduzido (18) é:

$$\mathbf{B'B}\hat{\boldsymbol{\gamma}}=\mathbf{B'y}$$

e o estimador dos parâmetros:

$$\hat{\boldsymbol{\gamma}}=\left(\mathbf{B'B}\right)^{-1}\mathbf{B'y}$$

A soma de quadrados de parâmetros relativa ao modelo reduzido (18) pode ser estimada por:

$$\text{SQPar}(r2)=\hat{\boldsymbol{\gamma}}'\mathbf{B'y}$$

com 1+H(p-1) graus de liberdade.

A redução que H_0 provoca na soma de quadrados de parâmetros do modelo completo é dada por:

$$\text{Redução}(H_0) = \text{SQPar(c)} - \text{SQPar(r2)}$$

com H-1 graus de liberdade.

Para testar a hipótese:

$H_0 : a_1 = a_2 = ... = a_H$ (as H equações têm uma constante de regressão comum)

$H_1 : a_h \neq a_{h'}$, para , pelo menos, um $h \neq h'$,

o autor utilizou a estatística F, dada por:

$$F_c = \frac{[\text{SQPar}(c) - \text{SQPar}(r2)]/(H-1)}{\text{SQRes}(c)/(N-Hp)} \qquad (19)$$

Rejeita-se H_0 se $F_c \geq F_{T[\alpha;(H-1),N-Hp]}$.

Na Tabela 3 é apresentada a análise de variância relativa a este teste.

TABELA 3 – Análise de variância relativa ao teste de hipótese $H_0 : a_1 = a_2 = ... = a_H$ (as H equações têm uma constante de regressão comum)

CV	GL	SQ	QM	F_c
Parâmetros (β)	(Hp)	$S_1 = \hat{\beta}'X'y$		
Parâmetros(γ)	1+H(p-1)	$S_2 = \hat{\gamma}'B'y$		
Redução (H_0)	H-1	$S_3 = S_1 - S_2$	$V_1 = \frac{S_3}{gl}$	$\frac{V_1}{V_2}$
Resíduo (c)	N-Hp	$S_4 = S_5 - S_1$	$V_2 = \frac{S_4}{gl}$	
Total	N	$S_5 = y'y$		

O terceiro teste considerou a seguinte hipótese de nulidade:

$H_0 : c_1 = c_2 = ... = c_H$ (as H equações têm os coeficientes de regressão do termo de segundo grau iguais), isto é, os modelos em (2) reduzem-se à forma:

$$y_{hi} = a_h + b_h P_{1hi} + cP_{2hi} + e_{hi} \qquad (20)$$

em que:

c : parâmetro comum

y_{hi} , P_{khi} , a_h, b_h e e_{hi} têm as mesmas especificações dos modelos em (2);

A partição de β_h e X_h em (3), generalizando para p parâmetros, é:

$$\beta_h = \begin{bmatrix} \alpha_h \\ \psi_h \end{bmatrix}_{p\times 1} \quad \text{e} \quad \mathbf{X}_h = [\mathbf{U}_h \mid \mathbf{V}_h]$$

em que α_h possui dimensão p1 x 1 ($0 < p1 < p$) e ψ_h possui dimensão p2 x 1 ($p2 = p - p1$).

Um caso geral da hipótese H_0 é:

$$H_0 : \psi_1 = \psi_2 = \ldots = \psi_H = \psi$$

Empregando a notação matricial, os modelos reduzidos em (20) podem ser escritos como:

$$\mathbf{y} = \mathbf{W}\xi + \varepsilon \tag{21}$$

em que:

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_H \end{bmatrix}_{N\times 1}, \quad \xi = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_H \\ \psi \end{bmatrix}_{[Hp1+p2]\times 1}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_H \end{bmatrix}_{N\times 1} \quad \text{e}$$

$$\mathbf{W} = \begin{bmatrix} \mathbf{U}_1 & 0 & \cdots & 0 & \mathbf{V}_1 \\ 0 & \mathbf{U}_2 & \cdots & 0 & \mathbf{V}_2 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & \mathbf{U}_H & \mathbf{V}_H \end{bmatrix}_{N\times[Hp1+p2]}.$$

Pelo método dos quadrados mínimos, obtém-se o seguinte sistema de equações normais relativo ao modelo reduzido (21):

$$\mathbf{W'W}\hat{\boldsymbol{\xi}} = \mathbf{W'y}$$

então, o estimador dos parâmetros é:

$$\hat{\boldsymbol{\xi}} = (\mathbf{W'W})^{-1}\mathbf{W'y}\,.$$

A soma de quadrados de parâmetros relativa ao modelo reduzido (21) é dada por:

$$\text{SQPar}(r3) = \hat{\boldsymbol{\xi}}'\mathbf{Wy}$$

com Hp1+p2 graus de liberdade.

A redução que H_0 provoca na soma de quadrados de parâmetros do modelo completo é dada por:

$$\text{Redução}(H_0) = \text{SQPar(c)} - \text{SQPar(r3)}$$

com (H-1)p2 graus de liberdade.

Assim, para testar a hipótese:

$H_0 : \psi_1 = \psi_2 = \ldots = \psi_H$

$H_1 : \psi_h \neq \psi_{h'}$ para, pelo menos, um $h \neq h'$.

em que:

ψ : qualquer coeficiente de interesse a ser comparado, nesse caso, refere-se ao termo quadrático.

Regazzi (1993) utilizou a estatística F, obtida por:

$$F_c = \frac{[\text{SQPar}(c) - \text{SQPar}(r3)]/(H-1)p2}{\text{SQRes}(c)/(N-Hp)}$$

Considerou que rejeita-se H_0 se $F_c \geq F_{T[\alpha;(H-1)p2,N-Hp]}$.

Na Tabela 4 é apresentada a análise de variância relativa a este teste.

TABELA 4 – Análise de variância relativa ao teste de hipótese $H_0 : \psi_1 = \psi_2 = ... = \psi_H$

CV	GL	SQ	QM	F_c
Parâmetros (β)	(Hp)	$S_1 = \hat{\beta}'\mathbf{X}'\mathbf{y}$		
Parâmetros(ξ)	Hp1+p2	$S_2 = \hat{\xi}'\mathbf{W}\mathbf{y}$		
Redução (H_0)	(H-1)p2	$S_3 = S_1 - S_2$	$V_1 = \frac{S_3}{gl}$	$\frac{V_1}{V_2}$
Resíduo (c)	N-Hp	$S_4 = S_5 - S_1$	$V_2 = \frac{S_4}{gl}$	
Total	N	$S_5 = \mathbf{y}'\mathbf{y}$		

Para Regazzi (1993), esse teste é geral, podendo-se aplicá-lo para testar a igualdade de um ou mais coeficientes de regressão. A metodologia adotada por Regazzi (1993) baseando-se em dados relativos à produção de quatro variedades em sete níveis de adubação, sendo considerado o modelo polinomial do segundo grau. O autor concluiu que a identidade de modelos de regressão, ou igualdade de qualquer subconjunto de parâmetros, pode ser verificada pelo teste F.

Em um segundo trabalho, Regazzi (1996), avaliou a identidade de modelos de regressão, considerando o ajustamento de H modelos de regressão no caso da justaposição de r = 2 submodelos polinomiais do primeiro grau e de r = 2 submodelos polinomiais do segundo grau.

Sousa (1989) utilizou essa metodologia na área florestal, estudando a variável peso sob diferentes espaçamentos, envolvendo cinco idades. Encontrou que as variáveis diâmetro, altura e idade, em uma única equação, poderiam estimar o peso do tronco.

Regazzi (1999), apresentou um método para testar as mesmas hipóteses avaliadas por Regazzi (1993), considerando o caso de dados provenientes de delineamentos experimentais (com repetições). Como ilustração, o método foi aplicado a um conjunto de H = quatro equações de regressão polinomial de segundo grau.

2.2.2 Variáveis binárias (*Dummy*)

Muitos autores priorizam a utilização de variáveis binárias, também mencionadas como variáveis *dummy*, indicadoras ou classificatórias, para testar a igualdade de equações ou coeficientes.

Gujarati (1970b) utilizou Variáveis *Dummy*, que são definidas como aquelas que assumem somente dois valores 1 e 0, como uma alternativa para a análise padrão de métodos de análise de variância e do teste de Chow (1960).

O referido autor considerou a seguinte relação, referente a dois conjuntos de dados:

$$y_i = \alpha_0 + \alpha_1 D + \alpha_2 x_i + \alpha_3 (Dx_i) + e_i \qquad i = 1,...,(n_1 + n_2)$$

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 α_1 e α_3 são diferenças de interceptos e inclinações, respectivamente.

Se $H_0 : \alpha_1 = 0$ é rejeitada, ou seja, α_1 é significativo, então, o valor do intercepto do primeiro conjunto é obtido por $\alpha_1 + \alpha_0$. Neste caso, α_0 é o intercepto do segundo conjunto. Se $H_0 : \alpha_1 = 0$ não é rejeitada, ou seja, α_1 é não significativo, então α_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, α_2 é a inclinação do segundo conjunto. Se $H_0 : \alpha_3 = 0$ não é rejeitada, então α_2 representa a inclinação comum para ambos os conjuntos.

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

Gujarati (1970b) notou que este método fornece resultados idênticos aos do teste de Chow (1960). Contudo, indica algumas vantagens para a técnica de

variáveis binárias. Esta técnica indica a(s) fonte(s) de diferença entre as regressões lineares, ou seja, se a diferença é devido a intercepto, ou inclinações, ou ambos. Em uma única regressão obtêm-se todas as informações necessárias, ao passo que o teste Chow é um procedimento de vários estágios.

Num segundo trabalho, Gujarati (1970a) generalizou a técnica de variáveis binárias para os casos com mais que duas regressões lineares e mais que duas variáveis.

Aplicou a técnica utilizando regressão linear múltipla, com duas variáveis independentes e quatro grupos (tratamentos), conforme descrito abaixo:

$$y_{hi} = \beta_{0h} + \beta_{1h}x_{1i} + \beta_{2h}x_{2i} + e_{hi} \qquad h = 1,2,3,4 \qquad i = 1,...,N,$$

o qual foi descrito mais explicitamente da seguinte forma:

$$\begin{aligned}
y_{1i} &= \beta_{01} + \beta_{11}x_{1i} + \beta_{21}x_{2i} + e_{1i} \qquad i = 1,...,n_1 \\
y_{2i} &= \beta_{02} + \beta_{12}x_{1i} + \beta_{22}x_{2i} + e_{2i} \qquad i = 1,...,n_2 \\
y_{3i} &= \beta_{03} + \beta_{13}x_{1i} + \beta_{23}x_{2i} + e_{3i} \qquad i = 1,...,n_3 \\
y_{4i} &= \beta_{04} + \beta_{14}x_{1i} + \beta_{24}x_{2i} + e_{4i} \qquad i = 1,...,n_4
\end{aligned}$$

sendo: $N = n_1 + n_2 + n_3 + n_4$.

De acordo com o autor, estas equações podem diferir de muitos modos, como, por exemplo, $\beta_{01} = \beta_{02} = \beta_{03} = \beta_{04}$, $\beta_{11} = \beta_{12} = \beta_{13} = \beta_{14}$, mas $\beta_{21} \neq \beta_{22} \neq \beta_{23} \neq \beta_{24}$, dentre as muitas outras combinações possíveis.

Uma vez assumido que as equações acima diferem entre si, pode-se definir o seguinte modelo:

$$y_i = \alpha_0 + \alpha_1 D_1 + \alpha_2 D_2 + \alpha_3 D_3 + \alpha_4 x_{1i} + \alpha_5 (D_1 x_{1i}) + \alpha_6 (D_2 x_{1i}) + \alpha_7 (D_3 x_{1i}) + \\ + \alpha_8 x_{2i} + \alpha_9 (D_1 x_{2i}) + \alpha_{10} (D_2 x_{3i}) + \alpha_{11} (D_3 x_{3i}) + e_i \qquad (22)$$

em que:

$D_1 = 1$, se a observação pertence ao segundo grupo
$= 0$, cc.
$D_2 = 1$, se a observação pertence ao terceiro grupo
$= 0$, cc.
$D_3 = 1$, se a observação pertence ao quarto grupo
$= 0$, cc.

Interpretam-se os vários coeficientes da mesma forma descrita por Gujarati (1970b). Como, por exemplo, α_0 é o intercepto para o primeiro grupo e α_1 é a diferença do intercepto para o grupo 2 e, assim, sucessivamente.

Aplicando-se o método dos mínimos quadrados ordinários, obtêm-se as seguintes equações abaixo derivadas da equação (22), assumindo $E(e_i) = 0$, $E(e_i, x_{ij}) = 0$ e $E(e_i, e_{i+k}) = \sigma^2$ para $K = 0$ e zero se $K \neq 0$:

$$\begin{aligned} &\text{grupo 1: } \hat{y} = \hat{a}_0 + \hat{a}_4 x_1 + \hat{a}_8 x_2, \\ &\text{grupo 2: } \hat{y} = (\hat{a}_0 + \hat{a}_1) + (\hat{a}_4 + \hat{a}_5) x_1 + (\hat{a}_8 + \hat{a}_9) x_2, \\ &\text{grupo 3: } \hat{y} = (\hat{a}_0 + \hat{a}_2) + (\hat{a}_4 + \hat{a}_6) x_1 + (\hat{a}_8 + \hat{a}_{10}) x_2, \\ &\text{grupo 4: } \hat{y} = (\hat{a}_0 + \hat{a}_3) + (\hat{a}_4 + \hat{a}_7) x_1 + (\hat{a}_8 + \hat{a}_{11}) x_2. \end{aligned} \qquad (23)$$

De acordo com a significância dos coeficientes estimados, pode-se saber se as regressões lineares são diferentes. Considerando o caso extremo em que pelo teste *t* nenhuma diferença de coeficientes em (22) foi significativa, então a equação relativa ao grupo 1, em (23), fornece a regressão comum para todos os grupos. Neste caso, os grupos não devem ter qualquer efeito sobre a relação da variável dependente Y e preditoras X (Gujarati, 1970a).

O referido autor comentou que a técnica de variáveis binárias é flexível, não sendo necessário diferenciar todos os coeficientes, como em (22). Se, a priori, tem-se a informação de que os interceptos não diferem, então considera-se apenas um intercepto comum para as equações. Salientou também o autor que o número de variáveis binárias é uma a menos que o número de grupos; caso contrário, a matriz **X'X** é singular.

Seber (1977), Draper e Smith (1998), Neter, Wassermann e Kutner (1990) comentaram também sobre o uso de variáveis binárias na regressão.

Segundo Draper e Smith (1998), as variáveis binárias podem assumir quaisquer valores, mas 0 e 1 são mais comumente utilizados. Os autores ilustram a técnica considerando três conjuntos de dados, G, V e W, com o seguinte modelo:

$$Y = \beta_0 + \beta_1 X + \alpha_1 D_1 + \alpha_2 D_2 + e \qquad (24)$$

em que:

$D_1 = 1$, para as observações do conjunto G

$= 0$, cc.

$D_2 = 1$, para as observações do conjunto V

$= 0$, cc.

α_1 e α_2 estimam a diferença nos níveis entre G e W e entre V e W, respectivamente.

Neste caso, considera-se que as três linhas são paralelas, mas possuem interceptos diferentes. Segundo os autores, para se testar a diferença entre os interceptos pode-se utilizar o teste t. Por exemplo, a diferença W-G é estimada por α_1. A estimativa desse coeficiente, dividido pela estimativa de seu respectivo desvio-padrão, obtido tomando-se a raiz quadrada da sua variância ou do termo apropriado da diagonal principal da matriz $(\mathbf{X'X})^{-1}\mathbf{S}^2$, é comparada com o valor crítico da distribuição t, $t_{(n-4,1-\alpha/2)}$ para um teste bilateral, para avaliação da hipótese $H_0 : \alpha_1 = 0$ versus $H_0 : \alpha_1 \neq 0$.

Draper e Smith (1998) abordam termos de interação envolvendo variáveis binárias e ilustram verificação da possibilidade de usar o mesmo modelo ajustado para dois conjuntos de dados, como segue:

$$Y = \beta_0 + \beta_1 X + \beta_{11} X^2 + \alpha_0 D + \alpha_1 XD + \alpha_{11} X^2 D + e \qquad (25)$$

em que D é a variável binária que assume o valor 0 para um conjunto de dados e 1 para o outro. Então, é possível verificar a hipótese de que $H_0 : \alpha_0 = \alpha_1 = \alpha_{11} = 0$.

Se H_0 é rejeitada, conclui-se que os modelos não são iguais. Se H_0 é rejeitada, podem-se verificar subconjuntos de α's. Por exemplo, testar $H_0 : \alpha_1 = \alpha_{11} = 0$. Se H_0 não é rejeitada, conclui-se que os dois conjuntos de dados exibem somente uma diferença nos níveis, mas possuem a mesma inclinação e curvatura.

Mas, se $H_0 : \alpha_1 = \alpha_{11} = 0$ é rejeitada, pode-se testar $H_0 : \alpha_{11} = 0$ versus $H_0 : \alpha_{11} \neq 0$ para verificar se os modelos diferem somente em intercepto e o termo de primeira ordem.

Hoffmann e Vieira (1998) utilizaram a técnica de variáveis binárias para comparar equações de regressão. Comentaram os autores que variáveis binárias podem ser definidas de várias formas e que a escolha da definição, ou da forma mais conveniente, depende das características do problema e das hipóteses que se deseja testar. No entanto, os resultados obtidos são equivalentes.

Também comentaram que o número de variáveis binárias deve ser igual ao número de grupos menos 1.

Scolforo (1997) apresentou a análise de covariância com Variáveis *Dummy*. Considerou três conjuntos de dados de barbatimão[2] de diferentes locais, obtidos de cubagem rigorosa[3], em que uma equação de volume foi estimada para cada conjunto. Este verificou que uma única equação não deve ser utilizada para estimar o volume do barbatimão nas três regiões consideradas.

2.2.3 Análise de Variância

Alguns autores utilizaram a análise de variância seguida de procedimentos para comparações múltiplas no estudo da comparação de modelos de regressão. Segundo Banzatto e Kronka (1995), os testes de comparações múltiplas, ou testes de comparações de médias, servem como um

[2] Árvore da família das leguminosas, subfamília das mimosáceas, cuja casca, adstringente, se usa na medicina e no curtimento de couros (*Stryphodendron barbatiman*).

[3] Procedimento dendrométrico, em que são medidos diâmetros ao longo do fuste com o objetivo de se obter o volume da árvore.

complemento do teste F e são adequados para detectar quais são as diferenças existente entre os tratamentos.

Paula Neto et al. (1983) analisaram sete modelos volumétricos segundo o método de regeneração e idade do plantio ou brotações. Utilizando o teste de Tukey, separaram os modelos em alguns grupos com equações semelhantes. Posteriormente, as árvores-amostra foram agrupadas, obtendo-se as equações específicas para cada grupo.

Herrera (1989) utilizou a análise de variância, seguida da aplicação de testes de médias, para formação de grupos de equações semelhantes. Para isto, utilizou as estimativas de peso seco de todas as árvores, sem casca, pertencentes a 21 condições de espécie, idade e tratamento. Conseguiu formar sete grupos de equações semelhantes e, com o agrupamento dos dados semelhantes em cada grupo, procedeu ao ajuste de novas equações.

Scolforo, Mello e Lima (1994) também utilizaram a análise de variância seguida de teste de aplicação de médias, para verificar a formação de grupos semelhantes, ao ajustar equações de volume para quatro espécies com alto valor de importância para uma floresta semidecídua montana na região de Lavras, MG. Identificaram, em suas análises, que a equação por grupo de espécie foi possível, e que as estimativas de volume propiciadas por estas foram precisas.

Leite e Regazzi (1992) e Silva et al. (1997) também utilizaram o método da análise de variância seguido da técnica de comparações múltiplas para comparar equações de regressão.

2.3 Simulação de dados

Os primeiros indícios de simulação de dados surgiram com a utilização do método de Monte Carlo, por Von Neuman, em 1940, com blindagem de reatores nucleares (Morgan, 1995).

Segundo Naylor et al. (1971), simulação de dados é uma técnica numérica para realizar experiências em um computador digital. Tais experiências envolvem certos tipos de modelos lógicos que descrevem o comportamento de um sistema.

O uso da simulação de dados tem uma grande diversidade de áreas de aplicação, basicamente sob duas linhas de atuação: problemas matemáticos completamente determinísticos, cuja solução é difícil, ou em problemas que envolvem o processo estocástico Monte Carlo, cuja técnica de simulação tem base probabilística ou estocástica.

Estes recursos fornecem dados em situações desejadas ou na ausência de um número suficiente de dados reais, facilitando a repetição do experimento, com rapidez e baixo custo, entre outros fatores.

Mitchell (2000) apresentou rotinas desenvolvidas no sistema computacional SAS®. para comparação de coeficientes de regressão em situações com três ou mais grupos.

3 MATERIAL E MÉTODOS

A metodologia apresentada neste trabalho foi aplicada por meio de um estudo de simulação de dados, com a geração de distribuições comportadas nas suas propriedades. O objetivo foi o de comparar os três métodos estatísticos, ou seja, identidade de modelos, variáveis dummy e análise de variância, que são muito utilizados na comparação de coeficientes e/ou equações de regressão.

Por meio de comparações mais detalhadas entre as metodologias, com o objetivo de padronizar as rotinas de testes e estimativas que são realizadas na prática, pretendeu-se verificar se existem divergências entre os métodos aplicados. Procedeu-se à verficação e comparação dos percentuais de taxas de Erro Tipo I e Erro Tipo II, em todas as situações de regressão linear simples e de regressão linear quadráticas consideradas. Objetivou-se, assim, verificar se estas divergências estão relacionadas relacionadas com as particularidades de cada método e com as propriedades das variáveis estudadas quanto às pressuposições necessárias a cada modelo.

A notação foi apresentada de forma matricial e foram utilizados os recursos do módulo Interactive Matrix Language (IML), do sistema Statistical Analysis System (SAS), para implementação da metodologia, proporcionando uma maior facilidade na sua aplicação.

Para a avaliação dos métodos, foram considerados quatro casos de regressão linear simples e cinco casos de regressão polinomial quadrática, conforme descrito a seguir.

3.1 Regressão linear simples

As situações ilustradas pela Figura 1 foram analisadas para o caso de regressão linear simples.

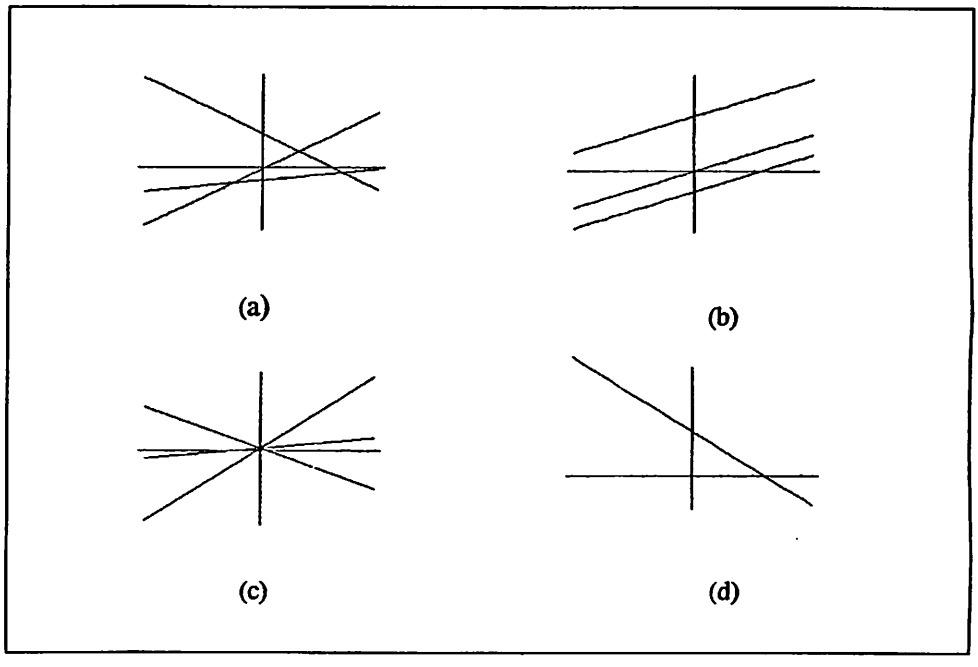

FIGURA 1 – Representação gráfica de algumas situações possíveis de ocorrência de modelos de regressão linear simples, para ilustrar a comparação de equações de regressão.

Na Figura 1, em (a) tem-se o caso mais geral, quando todos os coeficientes são diferentes; em (b) têm-se regressões paralelas, quando as inclinações são iguais, mas os interceptos são diferentes; em (c) têm-se regressões concorrentes, quando os interceptos são iguais, mas as inclinações são diferentes; e em (d) têm-se regressões coincidentes, quando todas as retas são coincidentes.

3.2 Regressão polinomial quadrática

As situações ilustradas pela Figura 2 foram analisadas para o caso de regressão polinomial quadrática.

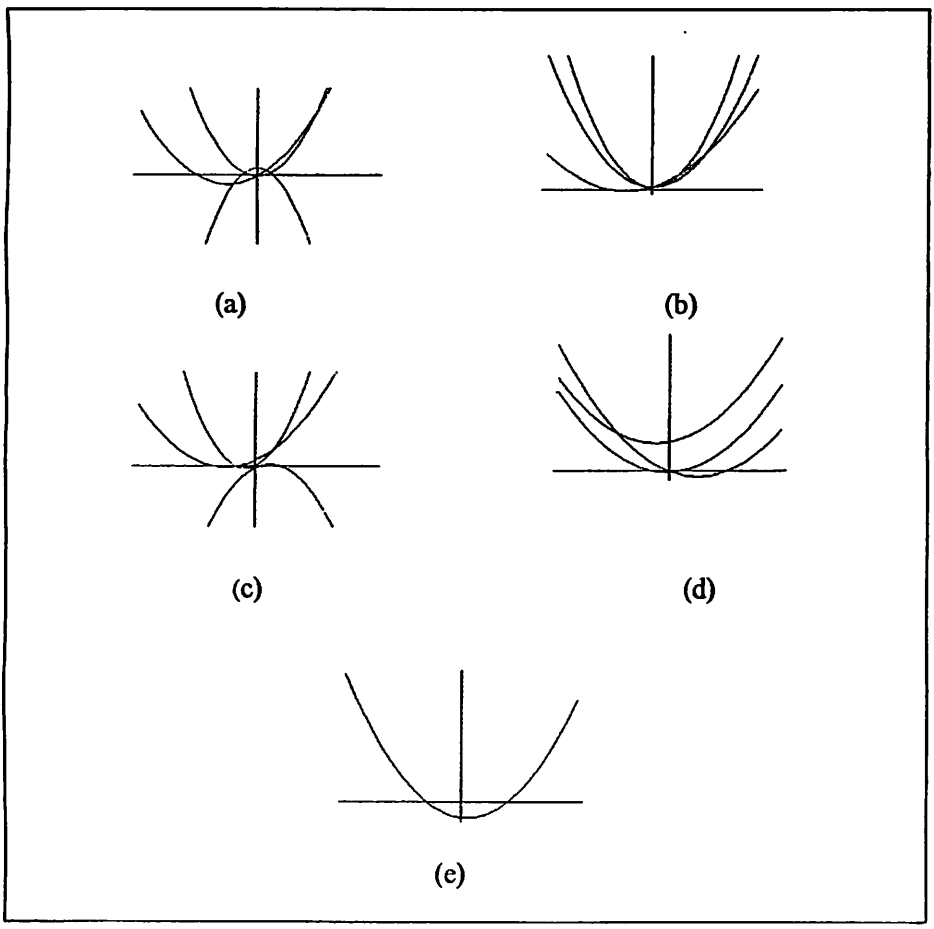

FIGURA 2 – Representação gráfica de algumas situações possíveis de ocorrência de modelos de regressão polinomial quadrática, para ilustrar a comparação de equações de regressão.

Na Figura 2, (a) tem-se o caso mais geral, quando todos os coeficientes são diferentes; em (b) têm-se regressões que possuem o mesmo intercepto; em (c) têm-se regressões que possuem o mesmo coeficiente relativo ao termo de 1° grau; em (d) têm-se regressões que possuem o mesmo coeficiente referente ao termo de 2° grau; em (e) têm-se regressões coincidentes, quando todas as curvas são coincidentes.

3.3 Simulação dos métodos

Dadas as seguintes relações lineares

$$\begin{aligned} y_{1i} &= \beta_{01} + \beta_{11}x_{11i} + \varepsilon_{1i} \\ y_{2i} &= \beta_{02} + \beta_{12}x_{12i} + \varepsilon_{2i} \\ &\vdots \\ y_{hi} &= \beta_{0h} + \beta_{1h}x_{1hi} + \varepsilon_{hi} \qquad \text{em que} \quad h=1,2. \end{aligned}$$

e polinomiais quadráticas

$$\begin{aligned} y_{1i} &= \beta_{01} + \beta_{11}x_{11i} + \beta_{21}x_{21i} + \varepsilon_{1i} \\ y_{2i} &= \beta_{02} + \beta_{12}x_{12i} + \beta_{22}x_{22i} + \varepsilon_{2i} \\ &\vdots \\ y_{hi} &= \beta_{0h} + \beta_{1h}x_{1hi} + \beta_{2h}x_{2hi} + \varepsilon_{hi} \qquad \text{em que } h=1,2. \end{aligned} \tag{26}$$

em que:

y_{hi}: i-ésima observação da variável resposta do h-ésimo modelo, sendo i = 1, 2,..., n_h o número de observações e h = 1,2 o número de modelos;

x_{1hi}, x_{2hi}: i-ésimo valor das variáveis regressoras do h-ésimo modelo;

$\beta_{0h}, \beta_{1h}, \beta_{2h}$: coeficientes do h-ésimo modelo;

ε_{hi}: erro aleatório, associado à i-ésima observação do h-ésimo modelo, sendo supostos independentes e normalmente distribuídos, com média zero e variância comum, isto é,

$$\varepsilon_{hi} \sim \text{NID}(0, \sigma^2), \sum_{h=1}^{H} n_h = N .$$

Realizou-se uma simulação de dados composta de 10.000 experimentos, cada qual com 10, 50 e 100 observações para cada uma das situações ilustradas e descritas pelas Figuras 1 e 2.

Para cada experimento, foram gerados modelos de regressão nos quais os valores das variáveis independentes foram obtidas em um intervalo fechado de 0 a 10, aleatoriamente, pela função RANUNI do sistema SAS®.

Para a geração dos resíduos de cada modelo, foi necessário estimar a variância dos mesmos. Fixando-se o coeficiente de determinação R^2 em 90 %, e conhecida a relação $R^2 = \dfrac{\delta^2_{modelo}}{\delta^2_{modelo} + \delta^2_{erro}}$, em que δ^2_{modelo} corresponde à média dos valores das variáveis dependentes, estimou-se a variância dos resíduos δ^2_{erro}.

Estimada a variância dos resíduos δ^2_{erro}, geraram-se pela função RANNOR do sistema SAS®, os resíduos aleatórios de cada modelo. Estes são, supostamente, independentes e normalmente distribuídos, com média zero e variância comum, isto é, $\varepsilon_{hi} \sim \text{NID}(0, \delta^2_{erro})$.

Com base nos modelos de regressão considerados, e fixando-se os parâmetros de cada modelo para cada uma das situações descritas pelas Figuras 1 e 2 para a comparação dos três métodos, foram implementados

computacionalmente os métodos da identidade de modelos, variáveis dummy e análise de variância, pelo módulo IML do sistema SAS®.

Ressalte-se que, para a implentação computacional do Método da Análise de Variância , foram considerados dois tratamentos conforme a Tabela 5, utilizou-se o Procedimento GLM do sistema SAS®. Os tratamentos para a implementação deste método foram constituídos pelos dois modelos (h=2) e os dados analisados foram os valores preditos pelas equações. Com base nesses dados foi realizada a Análise de Variância, considerando o delineamento inteiramente casualizado.

TABELA 5 – Tratamentos definidos para a metodologia da análise de variância

Tratamentos	Descrição
1	Equações ajustadas ao primeiro modelo
2	Equações ajustadas ao segundo modelo

4 RESULTADOS E DISCUSSÃO

Os resultados foram analisados com base nos procedimento FREQ do módulo BASE, do Statistical Analysis System (SAS).

Para os casos de regressão linear simples e de regressão polinomial quadrática foram determinadas as freqüências 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 Análise de Variância baseou-se 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.

4.1 Regressão linear simples

Para a situação (a), na qual admitiu-se que todos os coeficientes são diferentes, testou-se a hipótese :

$$\begin{cases} H_0 : \beta_1 = \beta_2 \\ H_1 : \beta_1 \neq \beta_2 \quad \text{(as duas equações são diferentes).} \end{cases}$$

Os resultados das frequências apresentados para os 10.000 experimentos simulados por meio dos três métodos utilizados para amostras de tamanho 10, 50 e 100 encontram-se na Tabela 6.

TABELA 6 – Distribuição de freqüências dos níveis de significância para os métodos utilizados nos 10.000 experimentos simulados para a situação de regressão linear simples em que as equações de regressão possuem todos os coeficientes diferentes.

Classes de níveis de significância	MÉTODOS								
	Identidade de Modelos (Nº. de Observações)			Variáveis Dummy (Nº. de Observações)			Análise de Variância (Nº. de Observações)		
	10	50	100	10	50	100	10	50	100
0 ⊢ 2,5	5987	7387	7556	6568	7496	7750	6072	9907	9080
2,5 ⊢ 5,0	3732	2487	2325	3217	2403	2147	3826	54	781
5,0 ⊢ 10	274	121	117	211	98	103	101	39	124
> 10	7	5	2	4	3	0	1	0	15

Verifica-se que, para esta situação, ocorreu uma baixa percentagem do nível de significância acima de 5%, indicando uma boa precisão dos métodos utilizados.

Observou-se uma maior dispersão nos casos em que o tamanho da amostra é menor, ou seja, para amostra de 10 observações, com uma aparente vantagem para o Método da Análise de Variância.

Com o aumento do número de observações, percebeu-se uma maior precisão no Método das Variáveis Dummy. Conforme ilustra a Tabela 6, para amostras de 100 observações, em 1,03 % das simulações seria cometido o Erro Tipo II, ou seja, não seria rejeitada uma hipótese inicial H_0, tida como falsa.

Em geral, o aumento do número de observações não acarretou reduções marcantes na taxa de aceitação de H_0. No Método da Identidade de Modelos, observou-se um maior índice de não rejeição, com 2,81 % para 10 observações e reduzindo-se para 1,19 % para 100 observações.

Para a situação (b), na qual admitiu-se que as duas regressões são paralelas, ou seja, possuem inclinações iguais e interceptos diferentes; testou-se a hipótese :

$$\begin{cases} H_0 : b_1 = b_2 \text{ (as duas equações são paralelas)} \\ H_1 : b_1 \neq b_2. \end{cases}$$

Os resultados apresentados para os 10.000 experimentos simulados por meio dos três métodos para amostras de tamanho 10, 50 e 100 encontram-se na Tabela 7.

TABELA 7 – Distribuição de frequências dos níveis de significância para os métodos utilizados nos 10.000 experimentos simulados para a situação de regressão linear simples em que as equações de regressão são paralelas.

Classes de níveis de significância	MÉTODOS								
	Identidade de Modelos (N°. de Observações)			Variáveis Dummy (N°. de Observações)			Análise de Variância (N°. de Observações)		
	10	50	100	10	50	100	10	50	100
0 ⊢ 2,5	12	3	4	8	1	1	3	8	7
2,5 ⊢ 5,0	120	44	25	95	37	14	44	53	29
5,0 ⊢ 10	4306	4355	2874	3987	4109	1875	4355	3658	521
> 10	5562	5598	7097	5910	5853	8110	5209	6281	9443

Para esta situação percebe-se uma baixa percentagem de ocorrência de níveis de significância abaixo de 5%, ou seja, aqueles que estariam provocando o Erro Tipo I. Este fato indica uma precisão nos três métodos utilizados.

Percebe-se uma maior dispersão para os casos em que o tamanho da amostra é composto de 10 observações, com uma aparente vantagem para o Método da Análise de Variância.

Conforme ilustra a Tabela 7, no Método das Variáveis Dummy, para amostra de 100 observações, somente em 0,15% das simulações seria cometido o Erro Tipo I, ou seja, a rejeição de uma hipótese H_0 tida verdadeira.

De modo geral, com aumento do número de observações, percebeu-se uma maior precisão para todos os métodos.

Para a situação (c), na qual admitiu-se que as duas regressões são concorrentes, ou seja, possuem interceptos iguais, mas inclinações diferentes; testou-se a hipótese :

$$\begin{cases} H_0 : a_1 = a_2 \text{ (as duas equações têm o mesmo intercepto)} \\ H_1 : a_1 \neq a_2 \ . \end{cases}$$

Os resultados apresentados para os 10.000 experimentos simulados por meio dos três métodos utilizados para amostras de tamanho 10, 50 e 100 encontram-se na Tabela 8.

Verifica-se, para esta situação, uma percentagem reduzida de ocorrência de níveis de significância abaixo de 5%, ou seja, aqueles que estariam provocando o Erro Tipo I. Tal fato serve de indicativo para uma precisão dos métodos utilizados.

Observa-se uma maior dispersão nos casos em que o tamanho da amostra é maior, ou seja, para amostra de 100 observações, com uma aparente vantagem para o Método das Variáveis Dummy.

TABELA 8 – Distribuição de frequências dos níveis de significância para os métodos utilizados nos 10.000 experimentos simulados para a situação de regressão linear simples em que as equações de regressão são concorrentes.

Classes de níveis de significância	MÉTODOS								
	Identidade de Modelos (Nº. de Observações)			Variáveis Dummy (Nº. de Observações)			Análise de Variância (Nº. de Observações)		
	10	50	100	10	50	100	10	50	100
0 ⊢ 2,5	11	1	4	8	6	2	1	3	8
2,5 ⊢ 5,0	121	47	27	96	41	21	47	56	182
5,0 ⊢ 10	4306	4385	94	3987	4115	87	352	687	985
> 10	5562	5567	9875	5909	5838	9890	4385	9254	8825

Com o aumento do número de observações, percebeu-se uma maior precisão no Método das Variáveis Dummy. De acordo com a Tabela 8, para amostra de 100 observações, somente em 0,23% simulações seria cometido o Erro Tipo I, ou seja, a rejeição de uma hipótese H_0 tida verdadeira.

E para a situação (d), na qual admitiu-se duas regressões são coincidentes, ou seja, todos os coeficientes são idênticos; testou-se a hipótese :

$$\begin{cases} H_0 : \beta_1 = \beta_2 \text{ (as duas equações são idênticas)} \\ H_1 : \beta_1 \neq \beta_2 \ . \end{cases}$$

Os resultados apresentados para os 10.000 experimentos simulados por meio três métodos utilizados para amostras de tamanho 10, 50 e 100 encontram-se na Tabela 9.

TABELA 9 – Distribuição de frequências dos níveis de significância para os métodos utilizados nos 10.000 experimentos simulados para a situação de regressão linear simples em que as equações de regressão são coincidentes.

Classes de níveis de significância	MÉTODOS								
	Identidade de Modelos (N°. de Observações)			Variáveis Dummy (N°. de Observações)			Análise de Variância (N°. de Observações)		
	10	50	100	10	50	100	10	50	100
0 ⊢ 2,5	2	1	0	4	0	0	0	2	0
2,5 ⊢ 5,0	151	101	1	257	85	0	257	81	8
5,0 ⊢ 10	3258	3826	123	3145	3478	78	3204	2587	1254
> 10	6589	6072	9876	6594	6437	9922	6539	7330	8754

Para esta situação, nota-se uma baixa percentagem de ocorrência de níveis de significância abaixo de 5%, ou seja, aqueles que estariam provocando o Erro Tipo I. Isto indica uma precisão dos métodos utilizados.

Para os casos em que o tamanho da amostra é menor, ou seja, para amostra de 10 observações, percebe-se uma maior dispersão, com uma aparente vantagem para o Método da Identidade de Modelos.

Com o aumento do número de observações, percebeu-se uma maior precisão no Método das Variáveis Dummy. Conforme ilustra a Tabela 9, no Método das Variáveis Dummy para amostra de 50 e 100 observações, em 0,85%

seria cometido o Erro Tipo I, ou seja, a rejeição de uma hipótese H_0 tida verdadeira.

De modo, geral, percebe-se que, com o aumento do tamanho das amostras, ocorreu uma redução significativa dos percentuais de Erro Tipo I.

4.2 Regressão polinomial quadrática

Para a situação (a), na qual admitiu-se que todos os coeficientes são diferentes; testou-se a hipótese :

$$\begin{cases} H_0 : \beta_1 = \beta_2 \\ H_1 : \beta_2 \neq \beta_2 \quad \text{(as duas equações são diferentes).} \end{cases}$$

Os resultados apresentados para os 10.000 experimentos simulados por meio dos três métodos utilizados para amostras de tamanho 10, 50 e 100 encontram-se na
Tabela 10.

Verifica-se, para esta situação, uma baixa percentagem de ocorrência de níveis de significância acima de 5%, ou seja, aqueles que estariam provocando o Erro Tipo II.

Nota-se uma maior dispersão nos casos em que o tamanho da amostra é menor, ou seja, 100 observações, com uma aparente vantagem para o Método da Análise de Variância.

De modo geral, notou-se uma menor variação para os casos em o tamanho da amostra era composto de 50 observações.

Segundo a Tabela 10, para amostras de100 observações, no Método da Análise de Variância, em 1,43 % das simulações seria cometido o Erro Tipo II, ou seja, não seria rejeitada uma hipótese inicial H_0, tida como falsa.

TABELA 10 – Distribuição de frequências dos níveis de significância para os métodos utilizados nos 10.000 experimentos simulados para a situação de regressão polinomial quadrática em que as equações de regressão possuem todos os coeficientes diferentes.

Classes de níveis de significância	MÈTODOS								
	Identidade de Modelos (N°. de Observações)			Variáveis Dummy (N°. de Observações)			Análise de Variância (N°. de Observações)		
	10	50	100	10	50	100	10	50	100
0 ⊢ 2,5	7122	6928	7033	7236	7455	7265	8977	6575	4456
2,5 ⊢ 5,0	2738	2987	1991	2658	2473	1874	515	3058	4780
5,0 ⊢ 10	83	75	976	97	66	861	249	204	65
> 10	12	10	0	9	6	0	259	163	78

Para a situação (b), na qual admitiu-se que as duas regressões têm o mesmo intercepto, testou-se a hipótese :

$$\begin{cases} H_0 : a_1 = a_2 \text{ (as duas equações têm uma constante de regressão comum)} \\ H_1 : a_1 \neq a_2 \ . \end{cases}$$

Os resultados apresentados para os 10.000 experimentos simulados por meio dos três métodos utilizados para amostras de tamanho 10, 50 e 100 encontram-se na
Tabela 11.

Para esta situação, nota-se uma baixa percentagem de ocorrência de níveis de significância abaixo de 5%, ou seja, aqueles que estariam provocando o Erro Tipo I. Este fato serve como um bom indicativo da precisão dos métodos utilizados.

TABELA 11 – Distribuição de frequências dos níveis de significância para os métodos utilizados nos 10.000 experimentos simulados para a situação de regressão polinomial quadrática em que todas as equações de regressão possuem o mesmo intercepto.

Classes de níveis de significância	MÉTODOS								
	Identidade de Modelos (Nº. de Observações)			Variáveis Dummy (Nº. de Observações)			Análise de Variância (Nº. de Observações)		
	10	50	100	10	50	100	10	50	100
0 ⊢ 2,5	9	35	168	7	29	145	35	21	17
2,5 ⊢ 5,0	22	139	299	19	127	251	139	117	101
5,0 ⊢ 10	3267	3135	587	3122	3061	458	840	758	662
> 10	6702	6691	8946	6852	6783	9146	8986	9104	9220

Para amostras de 100 observações, ou seja, para os casos em que o tamanho da amostra é maior, percebe-se uma maior dispersão com uma aparente vantagem para o Método das Variáveis Dummy.

Conforme ilustra a Tabela 11, com aumento do número de observações, percebeu-se uma maior precisão no Método da Análise de Variância. Para amostras de 100 observações, em 1,18 % das simulações seria cometido o Erro Tipo I, a rejeição de uma hipótese H_0 tida verdadeira.

Como os valores paramétricos são iguais, esperava-se uma alta taxa de aceitação de H_0. Este fato pode ser facilmente verificado pela Tabela 11.

Para a situação (c), na qual admitiu-se que as duas regressões possuem o mesmo coeficiente relativo ao termo de 1° grau; testou-se a hipótese :

$$\begin{cases} H_0 : b_1 = b_2 \text{ (as duas equações têm os coeficientes de regressão do termo de primeiro grau iguais)} \\ H_1 : b_1 \neq b_2 \ . \end{cases}$$

Os resultados apresentados para os 10.000 experimentos simulados por meio dos três métodos utilizados para amostras de tamanho 10, 50 e 100 encontram-se na
na Tabela 12.

Verifica-se, para esta situação, uma baixa percentagem de ocorrência de níveis de significância abaixo de 5%, ou seja, aqueles que estariam provocando o Erro Tipo I. Isto indica uma precisão dos métodos utilizados.

Observa-se uma maior dipersão nos casos de tamanho de amostra é maior, ou seja, que é igual a 100 observações, com uma aparente vantagem para o Método das Variáveis Dummy.

Conforme ilustra a Tabela 12, para amostra de 100 observações, somente em 0,21% das simulações seria cometido o Erro Tipo I, ou seja, a rejeição de uma hipótese H_0 tida verdadeira.

TABELA 12 – Distribuição de frequências dos níveis de significância para os métodos utilizados nos 10.000 experimentos simulados para a situação de regressão polinomial quadrática em que todas as equações de regressão possuem o mesmo coeficiente relativo ao termo de 1° grau.

Classes de níveis de significância	MÉTODOS								
	Identidade de Modelos (N°. de Observações)			Variáveis Dummy (N°. de Observações)			Análise de Variância (N°. de Observações)		
	10	50	100	10	50	100	10	50	100
0 ⊢ 2,5	6	7	0	4	4	0	14	31	37
2,5 ⊢ 5,0	33	43	39	29	37	21	152	171	259
5,0 ⊢ 10	3267	3259	364	3157	3087	257	3194	3872	4209
> 10	6694	6691	9597	6810	6872	9922	6640	5926	5495

Para a situação (d), na qual admitiu-se duas regressões possuem o mesmo coeficiente relativo ao termo de 2° grau, testou-se a hipótese :

$$\begin{cases} H_0 : c_1 = c_2 \text{ (as duas equações têm os coeficientes de regressão do termo de segundo grau iguais)} \\ H_1 : c_1 \neq c_2 \ . \end{cases}$$

Os resultados apresentados para os 10.000 experimentos simulados por meio dos três métodos utilizados para amostras de tamanho 10, 50 e 100 encontram-se na
Tabela 13.

Percebe-se, para esta situação, uma baixa percentagem de ocorrência de níveis de significância abaixo de 5%, ou seja, aqueles que estariam provocando o Erro Tipo I. Isto indica uma precisão dos métodos utilizados.

Verifica-se uma maior dispersão nos casos em que o tamanho da amostra é igual a 50 observações, com uma aparente vantagem para o Método das Variáveis Dummy.

Com o aumento do número de observações, nota-se uma maior precisão no Método das Variáveis Dummy. Conforme ilustra a Tabela 13, para amostra de 100 observações, em 1,1% das simulações seria cometido o Erro Tipo I, que é a rejeição de uma hipótese H_0 tida como verdadeira.

TABELA 13 – Distribuição de frequências dos níveis de significância para os métodos utilizados nos 10.000 experimentos simulados para a situação de regressão polinomial quadrática em que todas as equações de regressão possuem o mesmo coeficiente relativo ao termo de 2° grau.

Classes de níveis de significância	MÉTODOS								
	Identidade de Modelos (N°. de Observações)			Variáveis Dummy (N°. de Observações)			Análise de Variância (N°. de Observações)		
	10	50	100	10	50	100	10	50	100
0 ⊢ 2,5	19	1	22	16	0	17	1	5	9
2,5 ⊢ 5,0	77	25	120	5	14	93	25	663	570
5,0 ⊢ 10	3524	3657	547	3364	3291	497	637	3657	3078
> 10	6380	6317	9311	6615	6695	9393	9337	5675	6343

Para esta situação, notou-se, facilmente, que o Método da Ánalise de Variância, diferiu muito em relação aos outros dois.

E para a situação (e), na qual admitiu-se que duas regressões são coincidentes, ou seja, todos os coeficientes são idênticos, testou-se a hipótese :

$$\begin{cases} H_0 : \beta_1 = \beta_2 \text{ (as duas equações são idênticas)} \\ H_1 : \beta_1 \neq \beta_2 \ . \end{cases}$$

Os resultados apresentados para os 10.000 experimentos simulados por meio dos três métodos utilizados para amostras de tamanho 10, 50 e 100 encontram-se na
Tabela 14.

TABELA 14 – Distribuição de frequências dos níveis de significância para os métodos utilizados nos 10.000 experimentos simulados para a situação de regressão polinomial quadrática em que todas as equações de regressão são coincidentes.

Classes de níveis de significância	MÉTODOS								
	Identidade de Modelos (N°. de Observações)			Variáveis Dummy (N°. de Observações)			Análise de Variância (N°. de Observações)		
	10	50	100	10	50	100	10	50	100
0 ⊢ 2,5	4	3	1	2	3	0	2	9	0
2,5 ⊢ 5,0	42	38	41	37	40	18	12	47	69
5,0 ⊢ 10	2674	258	355	2501	1321	214	3658	2796	1567
> 10	7280	9701	9603	7460	8722	9768	6328	7148	8364

Para esta situação, verifica-se uma baixa percentagem de ocorrência de níveis de significância abaixo de 5%, ou seja, aqueles que estariam provocando o Erro Tipo I. Tal fato é um indicativo da uma precisão dos métodos utilizados.

Para os casos em que o tamanho da amostra é igual a 50 observações, percebe-se uma maior dispersão dos níveis de significância, com uma aparente vantagem para o Método da Identidade de Modelos.

Com o aumento do número de observações, percebeu-se uma maior precisão no Método das Variáveis Dummy. De acordo com a Tabela 14, para amostra de 100 observações, em apenas 0,18% das simulações seria cometido o Erro Tipo I, ou seja, a rejeição de uma hipótese H_0 tida verdadeira.

4.3 Considerações finais

A Tabela 15 ilustra todas as nove situações simuladas utilizando-se os três métodos em estudo. Pode-se notar que, de modo geral, foram percebidas maiores taxas de Erro Tipo I e Erro Tipo II nos casos em tamanho da amostra é igual a 50 observações, com uma aparente vantagem para o Método das Variáveis Dummy.

Esperava-se que, com o aumento do número de observações uma redução nas taxas de Erro Tipo I e Tipo II. Mas, este fato, em geral, não ocorreu. Por exemplo, para o Método das Variáveis Dummy, verificaram-se menores taxas com tamanho de amostra de 50 observações. Em geral, amostras com 50 observações apresentaram menores taxas de erros, mas estes valores não são bem diferentes dos valores dos outros tamanhos de amostras. Isto porque seus valores médios foram 1,22 % para amostra de tamanho 10; 1,09 % para amostra de tamanho 50 e 1,84 % para amostra de tamanho 100.

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

TABELA 15 – Distribuição de frequências de Erro Tipo I e Erro Tipo II para os métodos utilizados nos 10.000 experimentos simulados

Casos	MÉTODOS								
	Identidade de Modelos (N°. de Observações)			Variáveis Dummy (N°. de Observações)			Análise de Variância (N°. de Observações)		
	10	50	100	10	50	100	10	50	100
Linear									
a	281	126	119	215	101	103	102	39	139
b	132	47	29	103	38	15	47	61	36
c	132	48	31	104	47	22	48	59	190
d	153	102	1	301	85	0	257	83	8
Subtotal	**698**	**1421**	**1233**	**723**	**271**	**140**	**454**	**242**	**373**
Quadrático									
a	95	85	976	106	72	861	508	367	143
b	31	174	467	26	156	396	174	138	118
c	39	50	39	33	41	21	166	202	296
d	96	26	142	21	14	118	26	668	579
e	46	41	42	39	43	18	14	56	69
Subtotal	**307**	**376**	**1666**	**225**	**326**	**1414**	**888**	**1431**	**1205**
Total	**1005**	**669**	**1846**	**948**	**597**	**1554**	**1342**	**1673**	**1578**
Total Geral	**3520**			**3099**			**4593**		

Sugere-se um estudo bem mais detalhado, no qual deve-se aumentar o número de amostras, com o objetivo de encontrar um tamanho mínimo de amostra que minimize os percentuais de erros. Deve-se também estender a comparação entre os métodos da Identidade de Modelos, das Variáveis Dummy e da Análise de Variância a outros modelos, como, por exemplo, modelos não-lineares e modelos aplicados a algum comportamento biológico.

5 CONCLUSÕES

Os métodos da Identidade de Modelos, das Variáveis Dummy e da Análise de Variância sinalizam para a resultados bem semelhantes, devido a baixos percentuais de Erro Tipo I e Erro Tipo II.

Deve-se ressaltar que para todas as nove 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.

REFERÊNCIAS BIBLIOGRÁFICAS

BANZATTO, D. A.; KRONKA, S. N. **Experimentação agrícola.** Jaboticabal: Funep, 1995. 247p.

BATTISTI, I. D. E. **Comparação entre modelos de regressão com uma aplicação em biometria florestal.** 2001. 79p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária). Universidade Federal de Lavras, Lavras, MG.

BROWN, B. W. Simple comparisons of simultaneous regression lines. **Biometrics**, Washington, v. 26, n. 1, p. 143-144, Mar. 1970.

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

DRAPER, N. R.; SMITH, H. **Applied regression analysis.** 2. ed. New York: John Wiley & Sons, 1998. 709p.

DUNCAN, D. B. Multiple comparison methods for comparing regression coeffcients. **Biometrics**, Washington, v. 26, n. 1, p. 141-143, Mar. 1970.

FISHER, R. A. **Statistical methods for research workers.** 14. ed. New York: Hafner Press, 1970. 362p.

GRAYBILL, F. A. **Theory and application of the linear model.** Belmont: Duxbury Press, 1976. 704p.

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.

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.

HERRERA, M. E. F. **Densidade básica e equação de peso de madeira seca de povoamentos de eucaliptos de acordo com a idade, local, espaçamento e método de regeneração.** 1989. 113p. Dissertação (Mestrado em Ciência Florestal).Universidade Federal de Viçosa, Viçosa, MG.

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

LEITE, H. G.; REGAZZI, A. J. Métodos estatísticos para avaliar a igualdade de equações volumétricas. **Revista Árvore**, Viçosa, v. 16, n. 1, p. 59-71, jan./abr. 1992.

MITCHELL, M. **How can I compare regression coefficients across 3 (or more) groups.** 2000. Disponível em: <http://www.ats.ucla.edu/stat/sas/faq>. Acesso em: 18 set. 2000.

MORGAN, B. J. T. **Elements of simulation.** 6. ed. London: Chapman & Hall, 1995. 351p.

NAYLOR, T. H.; BALINTFY, J. L.; BURDICH, D. S.; CHU, K. **Técnicas de simulação em computadores.** São Paulo: Vozes, 1971. 401p.

NETER, J.; WASSERMAN, W.; KUTNER, M. **Applied linear statistical models.** 3. ed. Burr Ridge, Illinois: Irwin, 1990. 1181p.

PAULA NETO, F.; SOUZA, A. L.; QUINTAES, P. C. G.; SOARES, V. P. Análise de equações volumétricas para *Eucalyptus spp*, segundo o método de regeneração na região de José de Melo – MG. **Revista Árvore,** Viçosa, v. 7, n. 1, p. 56-70, jan./abr. 1983.

REGAZZI, A. J. Teste para verificar a identidade de modelos de regressão. **Pesquisa Agropecuária Brasileira**, Brasília, v. 31, n. 1, p. 1-17, jan. 1996.

REGAZZI, A. J. Teste para verificar a identidade de modelos de regressão e a igualdade de alguns parâmetros num modelo polinomial ortogonal. **Revista Ceres**, Viçosa, v. 40, n. 228, p. 176-195, mar./abr. 1993.

REGAZZI, A. J. Teste para verificar a identidade de modelos de regressão e a igualdade de parâmetros no caso de dados de delineamentos experimentais. **Revista Ceres**, Viçosa, v. 46, n. 266, p. 383-409, jun./ago. 1999.

SAS® INSTITUTE. **SAS Procedures guide for computers.** 6. ed. Cary N. C.: SAS® Institute, 1999. v. 3, 373p.

SCOLFORO, J. R.; MELLO, J. M. de; LIMA, C. S. Obtenção de relações quantitativas para estimativa do volume de fuste em floresta estacional semidecídua montana. **Revista Cerne**, Lavras, v. 1, n. 1, p. 123-134, 1994.

SCOLFORO, J. R. **Técnica de regressão aplicada para estimar: volume, biomassa, relação hipsométrica e múltiplos produtos da madeira.** Lavras: FAEPE, 1997. 292p.

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

SILVA, G. F.; CAMPOS, J. C. C.; SOUZA, A. L. et al. Uso de métodos estatísticos para comparar alternativas de estimação do volume comercial. **Revista Árvore**, Viçosa, v. 21, n. 1, p. 59-70, 1997.

SOUSA, R. N. **Efeito do espaçamento na produção em peso de madeira seca e volume de *Eucalyptus grandis*.** 1989. 86p. Dissertação (Mestrado em Ciência Florestal). Universidade Federal de Viçosa, Viçosa, MG.

SWAMY, P. A. V. B.; MEHTA, J. S. Estimation of common coefficients in two regression evaluations. **Journal of Econometrics**, Lausanne, v. 10, p. 1-14, 1979.

ANEXOS

ANEXO A **Página**

ANEXO B **Página**

ANEXO C **Página**

ANEXO A

Anexo A1 – Estrutura do Programa SAS para o teste de identidade de modelos – Regressão linear simples

```
/* Dissertacao – Teste da identidade modelos – Regressao linear simples*/
 /* 15 de dezembro de 2001*/
 /* Sergio Ricardo Silva Magalhaes e Ruben Delly Veiga*/

options ps=500 ls=76 nodate nonumber;
data teste;
proc iml;

/******* Situacao (A): Mesmo intercepto e mesma inclinação *****/

/* Os dados yobs armazena os yreais dos modelos 1 e 2 */
create yobs var {yreal1,x1,yreal2,x2,aux1,aux0};

/****** Alterar nexp e npares *********/
 npares=10 ; nexp=10000;

do ii=1 to nexp;

/******* Alterar coeficientes a e b e h ***********/
   a={6.33, 6.33};  b={4.78, 4.78};  h=2;

p=nrow(b);
do i=1 to npares;
    x1=ranuni(97)*10 + 1;
    x2=ranuni(89)*10 + 1;
    yob1=a[1,1]+b[1,1]*x1;
    yob2=a[2,1]+b[2,1]*x2;
    yreal1=yob1;
    yreal2=yob2;
    aux1=1;
    aux0=0;
    append var {yreal1,x1,yreal2,x2,aux1,aux0};
end;
end;
run;
quit;

/****** obtencao dos residuos e valores ajustados dos modelos 1 e 2 ********/
proc iml;
/****************** Alterar nexp e npares**************/
nexp=10000; npares=10;

create resival var {e1,e2,yfinal1,yfinal2};
use yobs (keep=yreal1 yreal2);

do ii=1 to nexp;

  read next 10 into yr;
```

```
   yr1=yr[1:10,1:1];
   yr2=yr[1:10,2:2];

   r2=0.9;
   sm1=0;
   sm2=0;

do i=1 to npares;
   m1=yr1[i,1];
   sm1=sm1+m1;
   m2=yr2[i,1];
   sm2=sm2+m2;
end;

   medmod1=sm1/npares;
   medmod2=sm2/npares;
   sigmae1=(medmod1*(1-r2))/r2;
   sigmae2=(medmod2*(1-r2))/r2;

do i=1 to npares;
   e1=rannor(0)*sqrt(sigmae1);
   e2=rannor(0)*sqrt(sigmae2);
   yfinal1=yr1[i,1] + e1;
   yfinal2=yr2[i,1] + e2;
   append var {e1,e2,yfinal1,yfinal2};
end;
end;
run;
quit;

/* *** Situacao (A) : Esquema da Analise de Variância *** */

 data ana1;
 merge yobs resival;

 proc iml;
 create estmocoa var {s1,gl1,s2,gl2,s3,gl3,s4,gl4,s5,gl5,v1a,v2a,fca,nsa};
 /*** Alterar nexp ***/
 nexp=10000;

 use ana1 (keep=yreal1 x1 yreal2 x2 aux1 aux0 e1 e2 yfinal1 yfinal2);

 do i=1 to nexp;
 read next 10 into conj;

  v0=conj[1:10,6:6];
  v1=conj[1:10,5:5];
  vx1=conj[1:10,2:2];
  vx2=conj[1:10,4:4];

   x0=v0||v0;
   x1=v1||vx1;
   x2=v1||vx2;
   x3=x1//x0;
   x4=x0//x2;
   xi=x3||x4;

   yf1=conj[1:10,9:9];
```

```
yf2=conj[1:10,10:10];
yf=yf1//yf2;

z1=v1||vx1;
z2=v1||vx2;

z=z1//z2;

beta=inv(xi`*xi)*xi`*yf;
teta=inv(z`*z)*z`*yf;

/*** Alterar Graus de liberdade ***/
  ha=2;
  pa=2;
  na=20;

s1= beta`*xi`*yf;
gl1=ha*pa;
s2= teta`*z`*yf;
gl2=pa;
s3= s1-s2;
gl3=(ha-1)*pa;
s5=yf`*yf;
s4= s5-s1;
gl4=na-gl1;
gl5=na;

/**** Quadrado Médio ****/

v1a=s3/gl3;
v2a=s4/gl4;
fca=v1a/v2a;

nsa=1-probf(gl1,gl2,fca);

append var {s1,gl1,s2,gl2,s3,gl3,s4,gl4,s5,gl5,v1a,v2a,fca,nsa};
end;
run;
quit;

proc format;
    value fmtnsa
        0    - < 0.0025 = "0%   a 2.5%"
        0.025 - < 0.05 = "2.5% a 5.0%"
        0.05 - < 0.1  = "5.0% a 10.0%"
        0.1  - < 1.0  = "10.0% a 100.0%";
        proc freq data = estmocoa;
    table nsa;
    format nsa fmtnsa. ;

/*  proc print data=estmocoa;
    var nsa;
   run; quit;     */

   proc univariate data=estmocoa plot normal;
    var nsa;
   run; quit:
```

```
/******* Situacao (B) : Mesmo intercepto *****/

/* Os dados yobs armazena os yreais dos modelos 1 e 2 */
 create yobs var {yreal1,x1,yreal2,x2,aux1,aux0};

/****** Alterar nexp e npares *********/
  npares=100; nexp=10000;

  do ii=1 to nexp;

/******* Alterar coeficientes a e b e h ***********/
     a={8.71, 8.71};   b={3.43, 5.97};   h=2;

  p=nrow(b);
  do i=1 to npares;
    x1=ranuni(97)*10 + 1;
    x2=ranuni(89)*10 + 1;
    yob1=a[1,1]+b[1,1]*x1;
    yob2=a[2,1]+b[2,1]*x2;
    yreal1=yob1;
    yreal2=yob2;
    aux1=1;
    aux0=0;
    append var {yreal1,x1,yreal2,x2,aux1,aux0};
  end;
  end;
  run;
  quit;

/****** obtencao dos residuos e valores ajustados dos modelos 1 e 2 ********/
 proc iml;

/****************** Alterar nexp e npares**************/
 nexp=10000;
 npares=100;

 create resival var {e1,e2,yfinal1,yfinal2};
 use yobs (keep=yreal1 yreal2);

 do ii=1 to nexp;

     read next 100 into yr;
     yr1=yr[1:100,1:1];
     yr2=yr[1:100,2:2];

     r2=0.9;
     sm1=0;
     sm2=0;

 do i=1 to npares;
     m1=yr1[i,1];
     sm1=sm1+m1;
     m2=yr2[i,1];
     sm2=sm2+m2;
 end;

     medmod1=sm1/npares;
     medmod2=sm2/npares;
```

```
    sigmae1=(medmod1*(1-r2))/r2;
    sigmae2=(medmod2*(1-r2))/r2;

 do i=1 to npares;
    e1=rannor(0)*sqrt(sigmae1);
    e2=rannor(0)*sqrt(sigmae2);
    yfinal1=yr1[i,1] + e1;
    yfinal2=yr2[i,1] + e2;
    append var {e1,e2,yfinal1,yfinal2};
 end;
 end;
 run;
 quit;

/* *** Situacao (B) : Esquema da Analise de Variância *** */

   data ana2;
   merge yobs resival;
  proc iml;
  create estmocob var {s1b,gl1b,s2b,gl2b,s3b,gl3b,s4b,gl4b,s5b,gl5b,v1b,v2b,fcb,nsb};
  use ana2 (keep=yreal1 x1 yreal2 x2 aux1 aux0 e1 e2 yfinal1 yfinal2);

/*** Alterar nexp ***/
  nexp=10000;

do i=1 to nexp;
  read next 100 into conj;

  v0=conj[1:100,6:6];
  v1=conj[1:100,5:5];
  vx1=conj[1:100,2:2];
  vx2=conj[1:100,4:4];
  yf1=conj[1:100,9:9];
  yf2=conj[1:100,10:10];
  y=yf1//yf2;

  c1=v1//v1;
  c2=vx1//v0;
  c3=v0//vx2;
  b=c1||c2||c3;
/**** Calculo de S1 *****/
    x0=v0||v0;
    x1=v1||vx1;
    x2=v1||vx2;
    x3=x1//x0;
    x4=x0//x2;
    xi=x3||x4;

    beta=inv(xi`*xi)*xi`*y;
    gama=inv(b`*b)*b`*y;

/*** Alterar Graus de liberdade ***/
    hb=2;
    pb=2;
    nb=200;

    s1ba= beta`*xi`*y;
    s1b= s1ba;
```

```
    gl1b=hb*pb;
    s2b= gama`*b`*y;
    gl2b=1+hb*(pb-1);
    s3b= s1b-s2b;
    gl3b=hb-1;
    s5b=y`*y;
    gl4b=nb-gl1b;
    s4b= s5b-s1b;
    gl5b=nb;

/*** Quadrados médios ***/

    v1b=s3b/gl3b;
    v2b=s4b/gl4b;
    fcb=v1b/v2b;

    nsb=1-probf(gl1b,gl2b,fcb);

    append var {s1b,gl1b,s2b,gl2b,s3b,gl3b,s4b,gl4b,s5b,gl5b,v1b,v2b,fcb,nsb};
end;
run;
quit;

proc format;
        value fmtnsb
            0    - < 0.0025 = "0%   a 2.5%"
            0.025 - < 0.05 = "2.5% a 5.0%"
            0.05  - < 0.1  = "5.0% a 10.0%"
            0.1   - < 1.0  = "10.0% a 100.0%";
    proc freq data = estmocob;
        table nsa;
        format nsa fmtnsb. ;

proc print data=estmocob;
    var  nsb;
run;
quit;

proc univariate data=estmocob;
    var nsb;
run;
quit;


/******* Situacao (C): Mesma inclinaçao*** */

/* Os dados yobs armazena os yreais dos modelos  1 e 2 */
  create yobs var {yreal1,x1,yreal2,x2,aux1,aux0};

/****** Alterar nexp e npares *********/
  npares=100;  nexp=500;

do ii=1 to nexp;

/******* Alterar coeficientes a e b e h ***********/
    a={3.21, 17.5};   b={11., 11.3};   h=2;

p=nrow(b);
```

```
do i=1 to npares;
   x1=ranuni(97)*10 + 1;
   x2=ranuni(89)*10 + 1;
   yob1=a[1,1]+b[1,1]*x1;
   yob2=a[2,1]+b[2,1]*x2;
   yreal1=yob1;
   yreal2=yob2;
   aux1=1;
   aux0=0;
   append var {yreal1,x1,yreal2,x2,aux1,aux0};
end;
end;
run;
quit;

/****** obtencao dos residuos e valores ajustados dos modelos 1 e 2 ********/
proc iml;

/****************** Alterar nexp e npares**************/
nexp=10000; npares=100;

create resival var {e1,e2,yfinal1,yfinal2};
use yobs (keep=yreal1 yreal2);

do ii=1 to nexp;

   read next 100 into yr;
   yr1=yr[1:100,1:1];
   yr2=yr[1:100,2:2];

   r2=0.9;
   sm1=0;
   sm2=0;

do i=1 to npares;
   m1=yr1[i,1];
   sm1=sm1+m1;
   m2=yr2[i,1];
   sm2=sm2+m2;
end;

   medmod1=sm1/npares;
   medmod2=sm2/npares;
   sigmae1=(medmod1*(1-r2))/r2;
   sigmae2=(medmod2*(1-r2))/r2;

do i=1 to npares;
   e1=rannor(0)*sqrt(sigmae1);
   e2=rannor(0)*sqrt(sigmae2);
   yfinal1=yr1[i,1] + e1;
   yfinal2=yr2[i,1] + e2;
   append var {e1,e2,yfinal1,yfinal2};
end;
end;
run;
quit;

/* *** Situacao (C) : Esquema da Analise de Variância *** */
```

```
data ana3;
merge yobs resival;

proc iml;
create estmococ var {s1c,gl1c,s2c,gl2c,s3c,gl3c,s4c,gl4c,s5c,gl5c,v1c,v2c,fcc,nsc};
use ana3 (keep=yreal1 x1 yreal2 x2 aux1 aux0 e1 e2 yfinal1 yfinal2);

/*** Alterar nexp ***/
nexp=1000;

  do i=1 to nexp;
    read next 100 into conj;

    v0=conj[1:100,6:6];
    v1=conj[1:100,5:5];
    vx1=conj[1:100,2:2];
    vx2=conj[1:100,4:4];
    yf1=conj[1:100,9:9];
    yf2=conj[1:100,10:10];
    y=yf1//yf2;

  /** Calculo de S1 ***/
    x0=v0||v0;
    x1=v1||vx1;
    x2=v1||vx2;
    x3=x1//x0;
    x4=x0//x2;
    xi=x3||x4;

    beta=inv(xi`*xi)*xi`*y;
    s1bc= beta`*xi`*y;

    c1=v1//v0;
    c2=v0//v1;
    c3=vx1//vx2;
    w=c1||c2||c3;

  cps=inv(w`*w)*w`*y;

  /* Alterar Graus de liberdade */
     hc=2;
     pc=2;
     pc1=1;
     pc2=pc-pc1;
     nc=200;

     s1c= s1bc;
     gl1c=hc*pc;
     s2c= eps`*w`*y;
     gl2c=hc*pc1+pc2;
     s3c= s1c-s2c;
     gl3c=(hc-1)*pc2;
     s5c= y`*y;
     gl5c=nc;
     s4c= s5c-s1c;
     gl4c=gl5c-gl1c;
```

```
/* Quadrado Médios */
  v1c=s3c/gl3c;
  v2c=s4c/gl4c;

  fcc=v1c/v2c;
  nsc=1-probf(gl1c,gl2c,fcc);

  append var {s1c,gl1c,s2c,gl2c,s3c,gl3c,s4c,gl4c,s5c,gl5c,v1c,v2c,fcc,nsc};
end;
run;
quit;

 proc format;
       value fmtnsc
            0    - < 0.0025 = "0%   a 2.5%"
            0.025 - < 0.05  = "2.5% a 5.0%"
            0.05  - < 0.1   = "5.0% a 10.0%"
            0.1   - < 1.0   = "10.0% a 100.0%";
  proc freq data = estmococ;
       table nsc;
       format nsc fmtnsc.;


proc print data=estmococ;
  var  nsc;
run;

proc univariate data=estmococ;
  var nsc;
run;
quit;
```

```
   medmod2=sm2/npares;
   sigmae1=(medmod1*(1-r2))/r2;
   sigmae2=(medmod2*(1-r2))/r2;

do i=1 to npares;
   e1=rannor(0)*sqrt(sigmae1);
   e2=rannor(0)*sqrt(sigmae2);
   yfinal1=yr1[i,1] + e1;
   yfinal2=yr2[i,1] + e2;
   append var {e1,e2,yfinal1,yfinal2};
end;
end;
run;
quit;

/* *** Situacao (B) : Esquema da Analise de Variância *** */

data ana2;
merge yobs resival;

  proc iml;
   create estmocob var {s1b,gl1b,s2b,gl2b,s3b,gl3b,s4b,gl4b,s5b,gl5b,v1b,v2b,fcb,nsb};
   use ana2 (keep=yreal1 x1 yreal2 x2 aux1 aux0 x1q x2q e1 e2 yfinal1 yfinal2);

/*** Alterar nexp ***/
   nexp=10000;

do i=1 to nexp;
   read next 10 into conj;

   v0=conj[1:10,6:6];
   v1=conj[1:10,5:5];
   vx1=conj[1:10,2:2];
   vx2=conj[1:10,4:4];
   vx1q=conj[1:10,7:7];
   vx2q=conj[1:10,8:8];


   yf1=conj[1:10,9:9];
   yf2=conj[1:10,10:10];
   y=yf1//yf2;

   b1=v1//v1;
   b2=vx1//v0;
   b2q=vx1q//v0;
   b3=v0//vx2;
   b3q=v0//vx2q;

   b=b1||b2||b2q||b3||b3q;

  /**** Calculo de S1 *****/

   c1=v1//v0;
   c2=vx1//v0;
   c3=vx1q//v0;
   c4=v0//v1;
   c5=v0//vx2;
   c6=v0//vx2q;
```

```
   xi=c1||c2||c3||c4||c5||c6;

   beta=inv(xi`*xi)*xi`*y;
   gama=inv(b`*b)*b`*y;

/*** Alterar Graus de liberdade ***/
    hb=2;
    pb=3;
    nb=20;

    s1ba= beta`*xi`*y;
    s1b= s1ba;
    gl1b=hb*pb;
    s2b= gama`*b`*y;
    gl2b=1+hb*(pb-1);
    s3b= s1b-s2b;
    gl3b=hb-1;
    s5b=y`*y;
    gl4b=nb-gl1b;
    s4b= s5b-s1b;
    gl5b=nb;

/*** Quadrados médios ***/

    v1b=s3b/gl3b;
    v2b=s4b/gl4b;
    fcb=v1b/v2b;

    nsb=1-probf(gl1b,gl2b,fcb);

    append var {s1b,gl1b,s2b,gl2b,s3b,gl3b,s4b,gl4b,s5b,gl5b,v1b,v2b,fcb,nsb};
end;
run;
quit;

proc format;
        value fmtnsb
            0    - < 0.0025 = "0%   a 2.5%"
            0.025 - < 0.05 = "2.5% a 5.0%"
            0.05  - < 0.1  = "5.0% a 10.0%"
            0.1   - < 1.0  = "10.0% a 100.0%";

 proc freq data = estmocob;
        table nsb;
        format nsb fmtnsb. ;

/*  proc print data=estmocob;
     var nsb;
   run; quit;     */

   proc univariate data=estmocob plot normal;
    var nsb;
   run; quit;

/******* Situacao (C1) : Mesmo coeficiente do 1° grau*** */

  /* Os dados yobs armazena os yreais dos modelos 1 e 2 */
   create yobs var {yreal1,x1,yreal2,x2,aux1,aux0,x1q,x2q};
```

```
/****** Alterar nexp e npares *********/
  npares=10 ; nexp=10000;

  do ii=1 to nexp;

 /******* Alterar coeficientes a e b e h ***********/
     a={6.1, 2.7};   b={4.78, 4.78};  c={9.2, 7.5};  h=2;

 p=nrow(b);
 do i=1 to npares;
     x1=ranuni(97)*10 + 1;
     x2=ranuni(89)*10 + 1;
     x1q=x1**2;
     x2q=x2**2;
     yob1=a[1,1]+b[1,1]*x1 + c[1,1]*x1q;
     yob2=a[2,1]+b[2,1]*x2 + c[2,1]*x2q;
     yreal1=yob1;
     yreal2=yob2;
     aux1=1;
     aux0=0;
     append var {yreal1,x1,yreal2,x2,aux1,aux0,x1q,x2q};
 end;
 end;
 run;
 quit;

/****** obtencao dos residuos e valores ajustados dos modelos 1 e 2 ********/
 proc iml;

/****************** Alterar nexp e npares**************/
 nexp=10000; npares=10;

 create resival var {e1,e2,yfinal1,yfinal2};
 use yobs (keep=yreal1 yreal2);

 do ii=1 to nexp;

   read next 10 into yr;
   yr1=yr[1:10,1:1];
   yr2=yr[1:10,2:2];

   r2=0.9;
   sm1=0;
   sm2=0;

 do i=1 to npares;
   m1=yr1[i,1];
   sm1=sm1+m1;
   m2=yr2[i,1];
   sm2=sm2+m2;
 end;

    medmod1=sm1/npares;
    medmod2=sm2/npares;
    sigmae1=(medmod1*(1-r2))/r2;
    sigmae2=(medmod2*(1-r2))/r2;

 do i=1 to npares;
```

```
    e1=rannor(0)*sqrt(sigmae1);
    e2=rannor(0)*sqrt(sigmae2);
    yfinal1=yr1[i,1] + e1;
    yfinal2=yr2[i,1] + e2;
    append var {e1,e2,yfinal1,yfinal2};
  end;
  end;
  run;
  quit;

/* *** Situacao (C1) : Esquema da Analise de Variância *** */

  data ana3;
  merge yobs resival;

  proc iml;
  create estmococ var {s1c,gl1c,s2c,gl2c,s3c,gl3c,s4c,gl4c,s5c,gl5c,v1c,v2c,fcc,nsc};
  use ana3 (keep=yreal1 x1 yreal2 x2 aux1 aux0 x1q x2q e1 e2 yfinal1 yfinal2);

/*** Alterar nexp ***/
  nexp=10000;

    do i=1 to nexp;
      read next 10 into conj;

      v0=conj[1:10,6:6];
      v1=conj[1:10,5:5];
      vx1=conj[1:10,2:2];
      vx2=conj[1:10,4:4];
      vx1q=conj[1:10,7:7];
      vx2q=conj[1:10,8:8];

      yf1=conj[1:10,9:9];
      yf2=conj[1:10,10:10];
      y=yf1//yf2;

    /** Calculo de S1 ***/

      c1=v1//v0;
      c2=vx1//v0;
      c3=vx1q//v0;
      c4=v0//v1;
      c5=v0//vx2;
      c6=v0//vx2q;
      xi=c1||c2||c3||c4||c5||c6;

      beta=inv(xi`*xi)*xi`*y;
      s1bc= beta`*xi`*y;

      w1=v1//v0;
      w2=v0//v1;
      w3=vx1//vx2;
      w4=vx1q//v0;
      w5=v0//vx2q;

      w=w1||w2||w3||w4||w5;

     eps=inv(w`*w)*w`*y;
```

```
/* Alterar Graus de liberdade */
   hc=2;
   pc=3;
   pc1=1;
   pc2=pc-pc1;
   nc=20;

   s1c= s1bc;
   gl1c=hc*pc;
   s2c= eps`*w`*y;
   gl2c=hc*pc1+pc2;
   s3c= s1c-s2c;
   gl3c=(hc-1)*pc2;
   s5c= y`*y;
   gl5c=nc;
   s4c= s5c-s1c;
   gl4c=gl5c-gl1c;

 /* Quadrado Médios */
   v1c=s3c/gl3c;
   v2c=s4c/gl4c;

   fcc=v1c/v2c;
   nsc=1-probf(gl1c,gl2c,fcc);

   append var {s1c,gl1c,s2c,gl2c,s3c,gl3c,s4c,gl4c,s5c,gl5c,v1c,v2c,fcc,nsc};
 end;
 run;
 quit;


proc format;
        value fmtnsc
            0    - < 0.0025  = "0%   a  2.5%"
            0.025 - < 0.05   = "2.5% a  5.0%"
            0.05  - < 0.1    = "5.0% a 10.0%"
            0.1   - < 1.0    = "10.0% a 100.0%";

 proc freq data = estmococ;
        table nsc;
        format nsc fmtnsc. ;

/*  proc print data=estmococ;
     var nsc;
   run; quit;     */

    proc univariate data=estmococ plot normal;
     var nsc;
   run; quit;


/******* Situacao (C2): Mesmo coeficiente do 2° grau *** */

 /* Os dados yobs armazena os yreais dos modelos  1 e 2 */
   create yobs var {yreal1,x1,yreal2,x2,aux1,aux0,x1q,x2q};

 /****** Alterar nexp e npares *********/
   npares=10 ; nexp=10000;
```

```
 do ii=1 to nexp;

/******* Alterar coeficientes a e b e h ***********/
    a={3.6, 1.78};   b={4.78, 11.41};  c={5.93, 5.93};  h=2;

 p=nrow(b);
 do i=1 to npares;
     x1=ranuni(97)*10 + 1;
     x2=ranuni(89)*10 + 1;
     x1q=x1**2;
     x2q=x2**2;
     yob1=a[1,1]+b[1,1]*x1 + c[1,1]*x1q;
     yob2=a[2,1]+b[2,1]*x2 + c[2,1]*x2q;
     yreal1=yob1;
     yreal2=yob2;
     aux1=1;
     aux0=0;
     append var {yreal1,x1,yreal2,x2,aux1,aux0,x1q,x2q};
 end;
 end;
 run;
 quit;


/****** obtencao dos residuos e valores ajustados dos modelos 1 e 2 ********/
 proc iml;

/****************** Alterar nexp e npares**************/
 nexp=10000; npares=10;

 create resival var {e1,e2,yfinal1,yfinal2};
 use yobs (keep=yreal1 yreal2);

 do ii=1 to nexp;

   read next 10 into yr;
   yr1=yr[1:10,1:1];
   yr2=yr[1:10,2:2];

   r2=0.9;
   sm1=0;
   sm2=0;

 do i=1 to npares;
    m1=yr1[i,1];
    sm1=sm1+m1;
    m2=yr2[i,1];
    sm2=sm2+m2;
 end;

    medmod1=sm1/npares;
    medmod2=sm2/npares;
    sigmae1=(medmod1*(1-r2))/r2;
    sigmae2=(medmod2*(1-r2))/r2;

 do i=1 to npares;
    e1=rannor(0)*sqrt(sigmae1);
```

```
    e2=rannor(0)*sqrt(sigmae2);
    yfinal1=yr1[i,1] + e1;
    yfinal2=yr2[i,1] + e2;
    append var {e1,e2,yfinal1,yfinal2};
  end;
  end;
  run;
  quit;

/* *** Situacao (C2) : Esquema da Analise de Variância *** */

  data ana3;
  merge yobs resival;

  proc iml;
  create estmococ var {s1c,gl1c,s2c,gl2c,s3c,gl3c,s4c,gl4c,s5c,gl5c,v1c,v2c,fcc,nsc};
  use ana3 (keep=yreal1 x1 yreal2 x2 aux1 aux0 x1q x2q e1 e2 yfinal1 yfinal2);

/*** Alterar nexp ***/
  nexp=10000;

   do i=1 to nexp;
     read next 10 into conj;

    v0=conj[1:10,6:6];
    v1=conj[1:10,5:5];
    vx1=conj[1:10,2:2];
    vx2=conj[1:10,4:4];
    vx1q=conj[1:10,7:7];
    vx2q=conj[1:10,8:8];

    yf1=conj[1:10,9:9];
    yf2=conj[1:10,10:10];
    y=yf1//yf2;

   /** Calculo de S1 ***/

    c1=v1//v0;
    c2=vx1//v0;
    c3=vx1q//v0;
    c4=v0//v1;
    c5=v0//vx2;
    c6=v0//vx2q;
    xi=c1||c2||c3||c4||c5||c6;

    beta=inv(xi`*xi)*xi`*y;
    s1bc= beta`*xi`*y;

    w1=v1//v0;
    w2=v0//v1;
    w3=vx1//vx2;
    w3q=vx1q//vx2q;

    w=w1||w2||w3||w3q;

   eps=inv(w`*w)*w`*y;

   /* Alterar Graus de liberdade */
```

```
      hc=2;
      pc=3;
      pc1=1;
      pc2=pc-pc1;
      nc=20;

     s1c= s1bc;
     gl1c=hc*pc;
     s2c= eps`*w`*y;
     gl2c=hc*pc1+pc2;
     s3c= s1c-s2c;
     gl3c=(hc-1)*pc2;
     s5c= y`*y;
     gl5c=nc;
     s4c= s5c-s1c;
     gl4c=gl5c-gl1c;

  /* Quadrado Médios */
    v1c=s3c/gl3c;
    v2c=s4c/gl4c;

    fcc=v1c/v2c;
    nsc=1-probf(gl1c,gl2c,fcc);

    append var {s1c,gl1c,s2c,gl2c,s3c,gl3c,s4c,gl4c,s5c,gl5c,v1c,v2c,fcc,nsc};
  end;
  run;
  quit;

proc format;
        value fmtnsc
            0    - < 0.0025 = "0%   a 2.5%"
            0.025 - < 0.05  = "2.5% a 5.0%"
            0.05  - < 0.1   = "5.0% a 10.0%"
            0.1   - < 1.0   = "10.0% a 100.0%";
proc freq data = estmococ;
          table nsc;
         format nsc fmtnsc. ;

/*  proc print data=estmococ;
     var nsc;
   run; quit;     */

   proc univariate data=estmococ plot normal;
     var nsc;
   run; quit;
```

ANEXO B

Anexo B1 – Estrutura do Programa SAS para o teste das variáveis binárias (dummy) – Regressão linear simples

```
/* Dissertacao – Variaveis dummy – Regressao linear simples*/
 /* 15 de dezembro de 2001*/
 /* Sergio Ricardo Silva Magalhaes e Ruben Delly Veiga*/

options ps=500 ls=76 nodate nonumber;
data teste;

proc iml;

/* obtencao dos dados reais */
  create dadosr var {yr1,xa1,yr2,xa2,aux1,aux0};
  a={6, 6}; b={4.78, 11}; npares=100; exp=10000;
 /*******Alterar a e b de acordo com o teste de interesse ******/

 do j=1 to exp;
 do i=1 to npares;
   xa1=ranuni(97)*10 + 1;
   xa2=ranuni(89)*10 + 1;
   yob1=a[1,1]+b[1,1]*xa1;
   yob2=a[2,1]+b[2,1]*xa2;
   yr1=yob1;
   yr2=yob2;
   aux1=1;
   aux0=0;
   append var {yr1,xa1,yr2,xa2,aux1,aux0};
 end;
 end;
 run;
 quit;

/* *** obtencao dos residuos dos modelos 1 e 2 ****/

 proc iml;

 create resi var {e1,e2};
 use dadosr (keep=yr1 yr2);
 npares=100; exp=10000;

 do j=1 to exp;
   read next 100 into yres;
   yres1=yres[1:100,1:1];
   yres2=yres[1:100,2:2];

   r2=0.9; sm1=0; sm2=0;

 do i=1 to npares;
   m1=yres1[i,1];
   sm1=sm1+m1;
   m2=yres2[i,1];
```

```
    sm2=sm2+m2;
  end;

    medmod1=sm1/npares;
    medmod2=sm2/npares;
    sigmae1=(medmod1*(1-r2))/r2;
    sigmae2=(medmod2*(1-r2))/r2;

  do i=1 to npares;
    e1=rannor(0)*sqrt(sigmae1);
    e2=rannor(0)*sqrt(sigmae2);
    append var {e1,e2};
  end;
  end;
  run;
  quit;

 /*** obtencao dos valores de dx ***/

 proc iml;
 create dx var {edx, j};
 exp=10000; npares=100;
 use dadosr (keep=yr1 xa1 yr2 xa2 aux1 aux0);

 do j=1 to exp;
   read next 100 into auxd;
   d0=auxd[1:100,6:6];
   d1=auxd[1:100,5:5];
   x1=auxd[1:100,2:2];
   x2=auxd[1:100,4:4];

   d=d0//d1;
   x=x1//x2;
   n=nrow(x);

 do i=1 to n;
   dx=d[i,1]*x[i,1];
   edx=dx;
   append var {edx, j};
 end;
 end;
 run;
 quit;

/*** obtencao dos valores ajustados dos modelos 1 e 2 ****/

 data dadosres;
 merge dadosr resi dx;

 proc iml;
 create dadosaj var {yajus,d,x};
 use dadosres (keep=yr1 xa1 yr2 xa2 aux1 aux0 e1 e2) ;

 exp=10000; npares=100;

 do i=1 to exp;
 read next 100 into valy;
   ve1=valy[1:100,7:7];
```

```
  y1=valy[1:100,1:1];
  ve2=valy[1:100,8:8];
  y2=valy[1:100,3:3];
  d0=valy[1:100,6:6];
  d1=valy[1:100,5:5];
  x1=valy[1:100,2:2];
  x2=valy[1:100,4:4];

  d=d0//d1;
  x=x1//x2;

  yajus1=y1 + ve1;
  yajus2=y2 + ve2;
  yajus=yajus1//yajus2;

  append var {yajus,d,x};
end;
run;
quit;

data undados;
merge dadosaj dx;


/**** Verificacao do pvalue para interceptos iguais ****/

proc reg data=undados noprint outest=resula tableout;
by j;
model yajus=d x edx ;
intigual : test  d=0;
run;
quit;

data rfa; set resula;
keep D;

if _TYPE_='PVALUE' THEN PCA=d;
ELSE DELETE;
Run; Quit;

data rfa; set rfa;
if d<0.05 then cta=1;
else cta=0;
run; quit;

proc  means data=rfa;
var cta;
run; quit;

/**** Fim de Interceptos iguais ****/

/**** Verificacao do pvalue para coeficientes iguais ****/

proc reg data=undados noprint outest=resulb tableout;
by j;
model yajus=d x edx ;
cfigual  : test  edx=0;
run;
```

```
quit;
data rfb; set resulb;
keep edx;

if _TYPE_='PVALUE' THEN PCB=edx;
ELSE DELETE;
Run; Quit;

data rfb; set rfb;
if edx<0.05 then ctb=1;
else ctb=0;
run; quit;

proc means data=rfb;
var ctb;
run; quit;

/**** Fim de coeficientes iguais ****/

/**** Verificacao do pvalue para equações iguais ****/

proc reg data=undados noprint outest=resulc tableout;
by j;
model yajus=d x edx ;
eqiguais : test d=0,edx=0;
run;
quit;

data rfc; set resulc;
keep x edx;

if _TYPE_='PVALUE' THEN PC=edx;
ELSE DELETE;
Run; Quit;

data rfc; set rfc;
if edx<0.05 then ctc=1;
else ctc=0;
run; quit;

proc means data=rfc;
var ctc;
run; quit;

/**** Fim de equações iguais ****/
```

Anexo B2 – Estrutura do Programa SAS para o teste das variáveis binárias (dummy) – Regressão polinomial quadrática

```
/* Dissertacao – Variaveis dummy – Regressao polinomial quadrática*/
  /* 15 de dezembro de 2001*/
  /* Sergio Ricardo Silva Magalhaes e Ruben Delly Veiga*/

  options ps=500 ls=76 nodate nonumber;
  data teste;

proc iml;

 /* obtencao dos dados reais */
  create dadosr var {yr1,xa1,yr2,xa2,aux1,aux0,x1q,x2q};
  a={6, 6}; b={4.78, 11}; c={2,5.34}; npares=10; exp=10000;
 /*********Alterar os valores de a, b e c de acordo com o teste a ser feito***********/

  do j=1 to exp;
  do i=1 to npares;
    xa1=ranuni(97)*10 + 1;
    xa2=ranuni(89)*10 + 1;
    x1q=xa1**2;
    x2q=xa2**2;
    yob1=a[1,1]+b[1,1]*xa1 + c[1,1]*x1q;
    yob2=a[2,1]+b[2,1]*xa2 + c[2,1]*x2q;
    yr1=yob1;
    yr2=yob2;
    aux1=1;
    aux0=0;
    append var {yr1,xa1,yr2,xa2,aux1,aux0,x1q,x2q};
  end;
  end;
  run;
  quit;

 /* *** obtencao dos residuos dos modelos 1 e 2 ****/

  proc iml;

  create resi var {e1,e2};
  use dadosr (keep=yr1 yr2);
  npares=10; exp=10000;

  do j=1 to exp;
    read next 10 into yres;
    yres1=yres[1:10,1:1];
    yres2=yres[1:10,2:2];

    r2=0.9; sm1=0; sm2=0;

  do i=1 to npares;
    m1=yres1[i,1];
    sm1=sm1+m1;
    m2=yres2[i,1];
    sm2=sm2+m2;
```

```
 end;
   medmod1=sm1/npares;
   medmod2=sm2/npares;
   sigmae1=(medmod1*(1-r2))/r2;
   sigmae2=(medmod2*(1-r2))/r2;

 do i=1 to npares;
   e1=ranmor(0)*sqrt(sigmae1);
   e2=ranmor(0)*sqrt(sigmae2);
   append var {e1,e2};
 end;
 end;
 run;
 quit;

/*** obtencao dos valores de dx ***/

proc iml;
create dx var {edx,edx2, j};
exp=10000; npares=10;
use dadosr (keep=yr1 xa1 yr2 xa2 aux1 aux0 x1q x2q);

do j=1 to exp;
  read next 10 into auxd;
  d0=auxd[1:10,6:6];
  d1=auxd[1:10,5:5];
  x1=auxd[1:10,2:2];
  x2=auxd[1:10,4:4];
  x1qa=auxd[1:10,7:7];
  x2qa=auxd[1:10,8:8];

  d=d0//d1;
  x=x1//x2;
  xq=x1qa//x2qa;

  n=nrow(x);
do i=1 to n;
  dx=d[i,1]*x[i,1];
  dx2=d[i,1]*xq[i,1];
  edx=dx;
  edx2=dx2;
  append var {edx,edx2, j};
end;
end;
run;
quit;

/*** obtencao dos valores ajustados dos modelos 1 e 2 ****/

data dadosres;
merge dadosr resi dx;

proc iml;
create dadosaj var {yajus,d,x,xq};
use dadosres (keep=yr1 xa1 yr2 xa2 aux1 aux0 x1q x2q e1 e2);

exp=10000; npares=10;
```

```
do i=1 to exp;
read next 10 into valy;
  ve1=valy[1:10,9:9];
  y1=valy[1:10,1:1];
  ve2=valy[1:10,10:10];
  y2=valy[1:10,3:3];
  d0=valy[1:10,6:6];
  d1=valy[1:10,5:5];
  x1=valy[1:10,2:2];
  x2=valy[1:10,4:4];
  x1qa=valy[1:10,7:7];
  x2qa=valy[1:10,8:8];

  d=d0//d1;
  x=x1//x2;
  xq=x1qa//x2qa;
  yajus1=y1 + ve1;
  yajus2=y2 + ve2;
  yajus=yajus1//yajus2;

  append var {yajus,d,x,xq};
end;
run;
quit;

data undados;
merge dadosaj dx;

/**** Verificacao do pvalue para interceptos iguais ****/

proc reg data=undados noprint outest=resula tableout;
by j;
model yajus=d x edx edx2 ;
intigual : test  d=0;
run;
quit;

data rfa; set resula;
keep D;

if _TYPE_='PVALUE' THEN PCA=d;
ELSE DELETE;
Run; Quit;

data rfa; set rfa;
if d<0.05 then cta=1;
else cta=0;
run; quit;

proc  means data=rfa;
var cta;
run; quit;

/**** Fim de Interceptos iguais ****/

/**** Verificacao do pvalue para coeficientes iguais do primeiro grau ****/

proc reg data=undados noprint outest=resulb tableout;
```

```
by j;
model yajus=d x edx edx2 ;
cfigual : test edx=0;
run;
quit;

data rfb; set resulb;
keep edx;

if _TYPE_='PVALUE' THEN PCB=edx;
ELSE DELETE;
Run; Quit;

data rfb; set rfb;
if edx<0.05 then ctb=1;
else ctb=0;
run; quit;

proc means data=rfb;
var ctb;
run; quit;

/**** Fim de coeficientes iguais do primeiro grau ****/

/**** Verificacao do pvalue para coeficientes iguais do segundo grau ****/

proc reg data=undados noprint outest=resulc tableout;
by j;
model yajus=d x edx edx2 ;
cfigual : test edx2=0;
run;
quit;

data rfc; set resulc;
keep edx2;

if _TYPE_='PVALUE' THEN PCC=edx2;
ELSE DELETE;
Run; Quit;

data rfc; set rfc;
if edx2<0.05 then ctc=1;
else ctc=0;
run; quit;

proc means data=rfc;
var ctc;
run; quit;

/**** Fim de coeficientes iguais do segundo grau ****/

/**** Verificacao do pvalue para equaçoes iguais ****/

proc reg data=undados noprint outest=resuld tableout;
by j;
model yajus=d x edx edx2 ;
eqiguais : test d=0,edx=0, edx2=0;
run;
```

```
quit;
data rfd; set resuld;
keep x edx edx2;

if _TYPE_='PVALUE' THEN PC=edx2;
ELSE DELETE;
Run; Quit;

data rfd; set rfd;
if edx2<0.05 then ctd=1;
else ctd=0;
run; quit;

proc means data=rfd;
var ctd;
run; quit;

/**** Fim de equacoes iguais ****/
```

ANEXO C

Anexo C1 – Estrutura do Programa SAS para o teste da análise de variância – Regressão linear simples

```
/* Dissertacao – Analise de Variancia – Regressao linear simples*/
     /* 15 de dezembro de 2001*/
     /* Sergio Ricardo Silva Magalhaes e Ruben Delly Veiga*/

     options ps=500 ls=76 nodate nonumber;
     data teste;

/* Valores de y e x reais ***************************************/
proc iml;
 create yobs var {yreal1,x1,yreal2,x2};
 npares=50 ;     /* Alterar nexp e npares          ****/
 nexp=10000;
 do iexp=1 to nexp;
  a={1, 1};     /* Alterar coeficientes a , b e h     ****/
    b={5, 8};
    h=2;
  p=nrow(b);
  do i=1 to npares;
   x1=ranuni(97)*10 + 1;
   x2=ranuni(89)*10 + 1;
   yreal1=a[1,1]+b[1,1]*x1;
   yreal2=a[2,1]+b[2,1]*x2;
   append var {yreal1,x1,yreal2,x2};
   end;
  end;
run;quit;
/** residuos e valores ajustados dos modelos 1 e 2 *************/
proc iml;
 nexp=10000;       /* Alterar nexp e npares          ****/
 npares=50;
 create resival var {e1,e2,iexp,i,yobs1,yobs2};
 use yobs (keep=yreal1 yreal2);
 do iexp=1 to nexp;
  read next 50 into yr;
  yr1=yr[1:npares,1:1];
  yr2=yr[1:npares,2:2];
  r2=0.9;
  sm1=0;
  sm2=0;
  do i=1 to npares;
   m1=yr1[i,1];
   sm1=sm1+m1;
   m2=yr2[i,1];
   sm2=sm2+m2;
   end;
  medmod1=sm1/npares;
  medmod2=sm2/npares;
  sigmae1=(medmod1*(1-r2))/r2;
  sigmae2=(medmod2*(1-r2))/r2;
```

```
    do i=1 to npares;
      e1=rannor(0)*sqrt(sigmae1);
      e2=rannor(0)*sqrt(sigmae2);
      yobs1=yr1[i,1] + e1;
      yobs2=yr2[i,1] + e2;
      append var {e1,e2,yobs1,yobs2,iexp,i};
      end;
    end;
run;quit;
/** residuos e valores ajustados dos modelos 1 e 2 **********/
proc iml;
 create yf var {y,trat,exp};
 use resival (keep=yobs1 yobs2 iexp i);
 nexp=10000;
 npares=50;
 do iexp= 1 to nexp;
  read next 50 into yp;*print yp;
  y1=yp[1:npares,3];
  y2=yp[1:npares,4];
  y=y1//y2;
     trat1 = j(npares,1,1);
     trat2 = j(npares,1,2);
  trat = trat1//trat2;
     exp = j(2*npares,1,iexp);
  *print y trat exp;
  *print tr;
  append var {y,trat,exp};
 end;
 run;quit;
proc glm data = yf outstat = arqnew noprint;
 class trat;
 model y = trat;
 by exp;
*proc print data=arqnew;
data ultimo (keep = prob);
 set arqnew;
 if _type_ eq "SS1";
proc format;
 value probfmt
       0.0 - < 0.025 = "0 a 2.5%"
      0.025 - < 0.05 = "2.5 a 5.00%"
      0.05 - < 0.5000 = "5.00 a 50.00%"
      0.5000 - < 1.00 = "50.00 a 100.00%";
      proc univariate data = ultimo;
 histogram prob;
proc freq;
 table prob;
 format prob probfmt.;
run; quit;
    *yob1=a[1,1]+b[1,1]*x1;
    *yob2=a[2,1]+b[2,1]*x2;
    *yreal1=yob1;
    *yreal2=yob2;
*proc print data = yobs;
*  title 'Arquivo yobs: variaveis resposta (trat1 e trat2)';
*proc print data = resival;
*  title 'Arquivo resival: residuos e variaveis resposta (trat1 e trat2)';
```

Anexo C2 – Estrutura do Programa SAS para o teste da análise de variância – Regressão polinomial quadrática

```
/* Dissertacao – Analise de Variancia – Regressao linear quadratica*/
     /* 15 de dezembro de 2001*/
     /* Sergio Ricardo Silva Magalhaes e Ruben Delly Veiga*/

     options ps=500 ls=76 nodate nonumber;
     data teste;

/* Valores de y e x reais ************************************/
proc iml;
 create yobs var {yreal1,x1,x1q,yreal2,x2,x2q};
 npares=10 ;     /* Alterar nexp e npares          ****/
 nexp=10000;
 do iexp=1 to nexp;
  a={6, 6};      /* Alterar coeficientes a , b, c e h     ****/
  b={4.78, 11};
  c={2, 5};
    h=2;
  p=nrow(b);
  do i=1 to npares;
   x1=ranuni(97)*10 + 1;
   x2=ranuni(89)*10 + 1;
   x1q=x1**2;
   x2q=x2**2;
   yreal1=a[1,1]+b[1,1]*x1+c[1,1]*x1q;
   yreal2=a[2,1]+b[2,1]*x2+c[2,1]*x2q;
   append var {yreal1,x1,x1q,yreal2,x2,x2q};
   end;
  end;
run;quit;
/** residuos e valores ajustados dos modelos 1 e 2 *************/
proc iml;
 nexp=10000;       /******** Alterar nexp e npares *******/
 npares=10;
 create resival var {e1,e2,iexp,i,yobs1,yobs2};
 use yobs (keep=yreal1 yreal2);
 do iexp=1 to nexp;
  read next 10 into yr;
  yr1=yr[1:npares,1:1];
  yr2=yr[1:npares,2:2];
  r2=0.9;
  sm1=0;
  sm2=0;
  do i=1 to npares;
   m1=yr1[i,1];
   sm1=sm1+m1;
   m2=yr2[i,1];
   sm2=sm2+m2;
   end;
  medmod1=sm1/npares;
  medmod2=sm2/npares;
  sigmae1=(medmod1*(1-r2))/r2;
  sigmae2=(medmod2*(1-r2))/r2;
  do i=1 to npares;
```

```
    e1=rannor(0)*sqrt(sigmae1);
    e2=rannor(0)*sqrt(sigmae2);
    yobs1=yr1[i,1] + e1;
    yobs2=yr2[i,1] + e2;
    append var {e1,e2,yobs1,yobs2,iexp,i};
    end;
   end;
run;quit;
/** residuos e valores ajustados dos modelos 1 e 2 ************/
proc iml;
 create yf var {y,trat,exp};
 use resival (keep=yobs1 yobs2 iexp i);
 nexp=10000;
 npares=10;
 do iexp= 1 to nexp;
  read next 10 into yp;*print yp;
  y1=yp[1:npares,3];
  y2=yp[1:npares,4];
  y=y1//y2;
     trat1 = j(npares,1,1);
     trat2 = j(npares,1,2);
  trat = trat1//trat2;
     exp = j(2*npares,1,iexp);
  *print y trat exp;
  *print tr;
  append var {y,trat,exp};
 end;
 run;quit;
proc glm data = yf outstat = arqnew noprint;
 class trat;
 model y = trat;
 by exp;
*proc print data=arqnew;
data ultimo (keep = prob);
 set arqnew;
 if _type_ eq "SS1";
proc format;
 value probfmt
     0.0 - < 0.025 = "0 a 2.5%"
     0.025 - < 0.05 = "2.5 a 5.00%"
     0.05 - < 0.1000 = "5.00 a 10.00%"
     0.1000 - < 1.00 = "10.00 a 100.00%";
proc univariate data = ultimo;
 histogram prob;
proc freq;
 table prob;
 format prob probfmt.;
run; quit;
```