

**COMPARAÇÃO ENTRE MODELOS DE
REGRESSÃO COM UMA APLICAÇÃO EM
BIOMETRIA FLORESTAL**

IARA DENISE ENDRUWEIT BATTISTI

2001


51629

MFV-36473

80022638 ma

IARA DENISE ENDRUWEIT BATTISTI

**COMPARAÇÃO ENTRE MODELOS DE
REGRESSÃO COM UMA APLICAÇÃO EM
BIOMETRIA FLORESTAL**



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

2001

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

Battisti, Iara Denise Endruweit

Comparação entre modelos de regressão com uma aplicação em biometria
florestal. – Lavras : UFLA, 2001.

86 p. : il.

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

I. Identidade de modelo. 2. Biometria florestal. 3.
Comparação entre modelos de regressão. I. Universidade
Federal de Lavras. II. Título.

CDD-519.537

-634.9285

IARA DENISE ENDRUWEIT BATTISTI


**COMPARAÇÃO ENTRE MODELOS DE REGRESSÃO COM UMA
APLICAÇÃO EM BIOMETRIA FLORESTAL**

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 2 de fevereiro de 2001.

Prof. Augusto Ramalho de Moraes UFLA

Prof. Amauri de Alcmeida Machado UFPel


Prof. Ruben Delly Yciga
UFLA
(Orientador)

LAVRAS
MINAS GERAIS – BRASIL

Aos meus pais Salo e Teresinha;

À minha irmã Lcila Martina;

Ao meu marido Gerson;

DEDICO

AGRADECIMENTOS

A Deus por esta etapa vencida.

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

À Universidade Regional do Noroeste do Estado do Rio Grande do Sul (UNIJUÍ), em especial ao Departamento de Física, Estatística e Matemática (DeFEM), pelo auxílio financeiro.

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

Ao professor Augusto Ramalho de Moraes pelas contribuições e amizade.

Aos professores Joel, Eduardo, Thelma, Daniel, Júlio, Luis Henrique, Lucas pelos ensinamentos.

Ao professor José Roberto Scolforo pelas contribuições. Também a Adriana Assis por fornecer os dados.

À professora Ruth Fricke (UNIJUÍ) pelo incentivo sempre e grande amizade.

À todos colegas da UNIJUÍ, do DeFEM e do LABEST que apoiaram e torceram.

Aos colegas da turma Ana Rita, Andréa, Cristiane, Éverton, Nagib, Teixeira, Moisés, Adriano, Aladir e Alex. Também ao demais colegas de curso: Andréa, Mônica, Ivani, Carlos e Sérgio.

Às amigas da república Cíntia e Rúbia e ao amigo Leonardo.

A todos parentes e amigos que confiaram e apoiaram, em especial ao vô Antônio, vô Gema, vô Hilda e tia Nena.

SUMÁRIO

RESUMO	i
ABSTRACT	ii
1 INTRODUÇÃO.....	1
2 REFERENCIAL TEÓRICO.....	3
2.1 Modelos de regressão	3
2.2 Comparação entre equações de regressão	5
2.2.1 Identidade de modelos	9
2.2.2 Uso de variáveis binárias (<i>dummy</i>).....	32
2.2.3 Comparações múltiplas.....	40
2.2.4 Análise de agrupamento.....	41
3 MATERIAL E MÉTODOS.....	45
3.1 Material	45
3.2 Métodos	48
3.2.1 Modelo de regressão.....	48
3.2.2 Aplicação dos métodos de comparação entre equações de regressão	49
3.2.2.1 Identidade de modelos	50
3.2.2.2 A utilização da técnica de variáveis binárias (<i>dummy</i>).....	55
3.2.2.3 Análise de variância e comparações múltiplas	56
3.2.2.4 A aplicação da análise de agrupamento	57
3.2.3 Softwares estatísticos.....	59
4 RESULTADOS E DISCUSSÃO	60
4.1 Identidade de modelos	60
4.2 Variáveis binárias.....	66
4.3 Análise de variância	69
4.4 Análise de agrupamento.....	70
4.5 Considerações Finais	73
5 CONCLUSÕES	75
REFERÊNCIAS BIBLIOGRÁFICAS.....	76
ANEXOS	80

RESUMO

BATTISTI, Iara Denise Endruweit. COMPARAÇÃO ENTRE MODELOS DE REGRESSÃO COM UMA APLICAÇÃO EM BIOMETRIA FLORESTAL. LAVRAS: UFLA, 2001. 79p. (Dissertação - Mestrado em Estatística e Experimentação Agropecuária).*

O presente estudo teve como objetivo apresentar alguns métodos estatísticos usados para comparação de equações de regressão e uma aplicação na biometria florestal, em que a comparação de equações de regressão é muito importante. Os métodos discutidos foram: análise de variância e comparações múltiplas, identidade de modelos, variáveis binárias e análise de agrupamento. Estes foram aplicados em uma situação na biometria florestal com o objetivo de identificar grupos de equações comuns ou diferenças entre coeficientes de regressão. A partir do modelo volumétrico de Schumacher e Hall, foram consideradas equações para 9 diferentes estratos, distribuídos em 4 grupos, definidos em função das espécies *Eucalyptus grandis* e *Eucalyptus urophylla*; do espaçamento 3,0 x 1,8 m e 3,0 x 2,5 m; da idade de 4 a 7 anos; e do tipo de solo, AQ1, AQ2, Lcm1, Lcm2, LR e Lca. Os resultados indicaram a necessidade de equações independentes por estratos no grupo 1, pelo método de identidade de modelos e variáveis binárias. Os métodos de análise de variância e análise de agrupamento forneceram a possibilidade de uso de uma equação comum para o grupo 1. As equações podem ser consideradas iguais nos grupos 2, 3, 4 pelos diferentes métodos aplicados. O teste de identidade de modelos e a técnica de variáveis binárias são equivalentes, fornecendo resultados mais objetivos.

* Orientador: Ruben Delly Veiga - UFLA

ABSTRACT

BATTISTI, Iara Denise Endrueit. **COMPARISON BETWEEN REGRESSION MODELS WITH AN APPLICATION IN BIOMETRY FOREST.** LAVRAS: UFLA, 2001. 79p. (Dissertation - master in Statistics and Agricultural Experimentation).*

The present study had as main goal to introduce statistical methods found in literatures for comparison of regression equations and an application in biometry forest, in which the regression equation comparisons are very important. The argued methods were: analysis of variance and multiple comparisons, models identity, dummy variables and cluster analysis. These were applied in biometry forest situation with the goal of identifying groups of common equations or differences between regression coefficients. From the volumetric model of Schumacher and Hall, equations were considered for 9 different stratum, distributed in 4 groups, defined regarding the species, *Eucalyptus grandis* and *Eucalyptus urophylla*; with spacing of 3,0 x 1,8 m and 3,0 x 2,5 m; aged from 4 to 7 years; and of soil type, AQ1, AQ2, Lcm1, Lcm2, LR and Lea. The results indicated the necessity for independent equations for stratum in the group 1 and also for model identity methods and dummy variables methods. The analysis of variance and cluster analysis methods supplied the possibility of using a common equation for group 1. While the equations can be considered alike in the groups 2, 3, 4 by the different applied methods. The models identity test and the technique of dummy variables are equivalent, supplying more objective results.

* Adviser: Ruben Delly Veiga - UFLA

1 INTRODUÇÃO

A análise de regressão é uma técnica de análise de dados muito utilizada nos casos em que se deseja estudar a relação entre variáveis quantitativas - uma variável resposta e uma ou mais variáveis explicativas.

Com muita frequência, o estudo é feito em diferentes tratamentos para comparar seus efeitos, podendo ser realizado em diferentes fatores de classificação, como épocas ou locais. Para cada uma dessas situações, a análise de regressão pode ser aplicada separadamente, obtendo-se tantas equações quanto o número de situações distintas.

Neste caso, 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?

Estudos desta natureza são comuns e de fundamental importância na área da experimentação agropecuária, econometria e biometria florestal. Por exemplo:

- num experimento em que foi utilizado fertilizante em determinada cultura, num certo número de variedades, obtém-se a relação entre produção e a quantidade de fertilizante aplicado através de uma equação de regressão. Então, é importante determinar se o aumento na produção por unidade de fertilizante é o mesmo para todas variedades.

- em um estudo sobre a produção em diferentes grupos: (1) proprietários, (2) arrendatários e (3) proprietários e arrendatários. Um interesse prático é comparar estas equações, verificando se são iguais ou em que diferem.

- para um povoamento de *Eucalyptus cloesiana*, ajustou-se uma regressão para o regime de alto fuste e outra para primeira talhadia. Será possível ajustar uma única regressão para as duas situações? (Scolforo, 1997).

São vários os métodos para que se possa diferenciar estas situações, ou seja, identidade das equações de regressão encontradas, igualdade entre coeficientes de regressão, utilização de variáveis binárias, comparações múltiplas e análise de agrupamento.

Deste modo, o objetivo desse estudo foi apresentar e analisar métodos estatísticos de identidade de modelos, variáveis binárias, análise de variância e análise de agrupamento para comparação de equações de regressão e aplicá-los numa área de interesse prático, como a biometria florestal.

2 REFERENCIAL TEÓRICO

2.1 Modelos de regressão

De acordo com Draper e Smith (1981), os modelos de regressão podem ser classificados em lineares, linearizáveis e não-lineares, em relação aos parâmetros. Para o presente estudo interessam os modelos lineares e linearizáveis, lembrando que estes últimos podem ser transformados em lineares através de anamorfose. Algumas considerações sobre estes serão apresentadas abaixo.

Um modelo de regressão linear, segundo Draper e Smith (1981) e Hoffmann e Vieira (1998), pode ser expresso como:

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

em que:

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

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

β_k : parâmetros do modelo;

ε_i : erros aleatórios.

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

$$y = X\beta + \varepsilon$$

em que:

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

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

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

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

Os métodos de estimação do vetor de parâmetros β mais empregados são o método dos quadrados mínimos e o método da máxima verossimilhança, conduzindo aos mesmos estimadores em presença de normalidade.

Existem variações do método de estimação dos quadrados mínimos para o modelo de regressão linear, conforme as pressuposições assumidas quanto aos erros, assim obtendo estimadores não-tendenciosos de variância mínima. Estas variações referem-se aos métodos dos quadrados mínimos ordinário, ponderado e generalizado.

Segundo 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 , $i = 1, 2, \dots, n$ 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.

A partir do método dos quadrados mínimos ordinários encontra-se um vetor β , em que a soma de quadrados dos erros seja mínima. Como apresentada por Hoffmann e Vieira (1998), a função quadrática:

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

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

$$X'X\beta = X'y$$

Sendo a matriz X de posto coluna completo, então $X'X$ é uma matriz positiva definida e, assim, $X'X$ é não singular, logo existe a matriz inversa $(X'X)^{-1}$ e a solução para β de acordo com Draper e Smith (1981) e Hoffmann e Vieira (1998) é:

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

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

Nos itens a seguir, serão descritos detalhadamente os métodos para comparação de equações de regressão encontrados na literatura, para os quais convergem os interesses deste estudo.

2.2 Comparação entre equações de regressão

O estudo de situações, através da análise de regressão, em que se faz a comparação entre dois ou mais conjuntos de observações n -dimensionais, tem sido relatado na literatura, como apresentado por Gujarati (1970a), Draper e Smith (1981), Scolforo (1997), Regazzi (1999), entre outros.

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

Por outro lado, se for verificado que as equações não diferem entre si, significa que uma única equação pode ser utilizada representando todos os conjuntos de observações, ou seja, uma única equação pode ser estimada a partir de todas as observações de todos os conjuntos envolvidos no estudo. Assim, 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 sendo mais confiável, quando comparado à estimação de equações individuais.

Como ilustração, considera-se a regressão linear simples, conforme a Figura 1, representando em (a) o caso mais geral, quando todos os coeficientes são diferentes; (b) regressões paralelas, as inclinações são iguais, mas interceptos são diferentes; (c) regressões concorrentes, interceptos iguais, mas inclinações diferentes; (d) regressões coincidentes, todas as retas são a mesma.

Como já comentado, vários autores apresentaram testes para comparação entre equações de regressão e/ou coeficientes e também a sua utilização prática. A seguir são apresentados alguns dos métodos considerados de maior interesse para o presente estudo.

Para testar a igualdade de duas regressões lineares, Chow (1960) propôs um teste geral, descrevendo os seguintes passos:

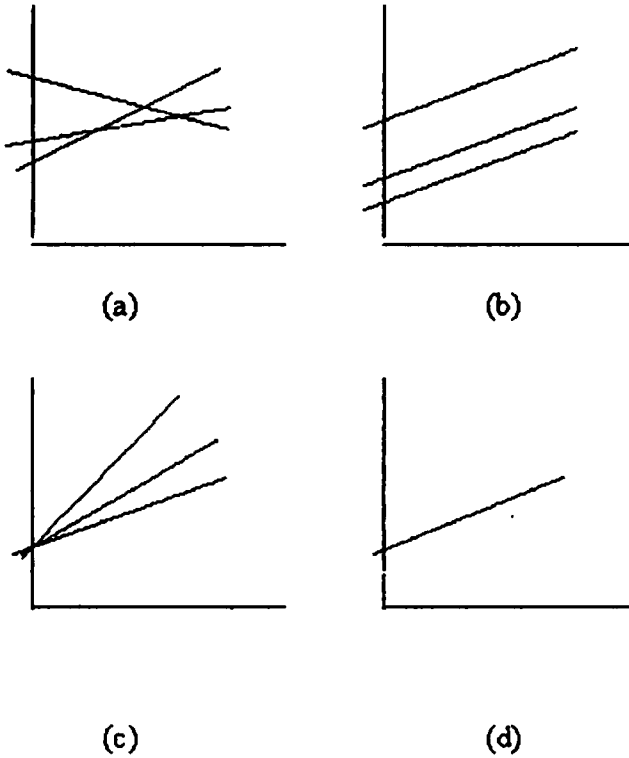


FIGURA 1 - Representação gráfica de algumas situações possíveis de ocorrência de modelos de regressão linear simples, para ilustrar a comparação entre equações de regressão. (Weisberg, 1985)

1. Dadas as seguintes relações lineares:

$$y_{1i} = a_1 + b_1 x_{1i} + e_{1i} \quad i = 1, \dots, n_1$$

$$y_{2i} = a_2 + b_2 x_{2i} + e_{2i} \quad i = 1, \dots, n_2$$

referentes a 2 conjuntos de observações.

2. Combinar todas as $n_1 + n_2$ observações e calcular a estimativa de mínimos quadrados de a e b na regressão combinada $Y = a + bX$. Desta equação obtém-se a soma de quadrados de residuo (S_1) com grau de liberdade igual a $n_1 + n_2 - p$, em que p é o número de parâmetros para a ser estimado. Neste caso, $p = 2$.
3. Obter a soma de quadrados de residuo para as duas equações, ou seja, S_2 e S_3 , com os graus de liberdade $n_1 - p$ e $n_2 - p$, respectivamente. Somar estas duas somas de quadrados de residuo, isto é, $S_4 = S_2 + S_3$ e seus graus de liberdade $n_1 + n_2 - 2p$.
4. Obter $S_5 = S_1 - S_4$.
5. Aplicar o teste F_c 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.

Segundo Gujarati (1970b), o teste de Chow (1960) é geral, afirmando apenas se duas regressões lineares são iguais ou diferentes. Caso sejam diferentes, não especificam se a diferença é devida a interceptos ou inclinações.

Fisher (1970) apresentou a comparação entre coeficientes de regressão de maneira análoga à das médias, levando a resultados idênticos aos obtidos por

Duncan (1970), comparando os coeficientes b_1 e b_2 de duas equações de regressão linear simples, através do teste t .

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

$$y_{hi} = a_h + b_h x_{hi} + e_{hi} \quad h = 1, \dots, H \text{ modelos} \\ i = 1, \dots, n_h \text{ observações}$$

para os quais existe interesse em obter um modelo simplificado, em que todos os b 's e todos os a 's são idênticos. Para isto, deduziu novas variáveis c , utilizando a regressão linear múltipla, fez o ajustamento das observações, para o modelo reduzido, através do método dos quadrados mínimos.

Swamy e Mehta (1979) mostraram que, combinando informações de duas equações de regressão, é possível obter estimativas mais eficientes do que as estimativas baseadas em cada uma das equações.

A homogeneidade de coeficientes de regressão linear foi testada pela estatística F , na análise de covariância, por Snedecor e Cochran (1980) e por Steel, Torrie e Dickey (1997). Estes últimos, utilizando aplicação em delincimento inteiramente casualizado, afirmaram que extensões podem ser feitas para delincimento em blocos casualizados, esquema fatorial e múltiplas covariáveis.

2.2.1 Identidade de modelos

Graybill (1976) apresentou detalhadamente um teste para verificar a identidade de H modelos lineares simples, da seguinte forma:

$$\begin{aligned}
y_{1j} &= \alpha_1 + \beta_1 x_{1j} + \varepsilon_{1j} & j &= 1, \dots, n_1 \\
y_{2j} &= \alpha_2 + \beta_2 x_{2j} + \varepsilon_{2j} & j &= 1, \dots, n_2 \\
&\vdots & & \\
y_{Hj} &= \alpha_H + \beta_H x_{Hj} + \varepsilon_{Hj} & j &= 1, \dots, n_H
\end{aligned} \tag{1}$$

$$\sum_{h=1}^H n_h = N, \quad n_h > 2 \text{ para todo } h, \quad \varepsilon_{ij} \sim NID(\varepsilon : 0, \sigma^2)$$

O autor formulou várias hipóteses e para cada uma apresentou os respectivos testes, a saber:

1. $H_0 : \beta_1 = \beta_2 = \dots = \beta_H$ (as H linhas são paralelas)
 $H_1 : \beta_h \neq \beta_{h'}$ para pelo menos um $h \neq h'$ ($h, h' = 1, 2, \dots, H$)

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

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

em que $b_{hh} = \sum_{i=1}^{n_h} (x_{hi} - \bar{x}_h)^2$

2. $H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_H$ (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 $W_1 \geq F_{\alpha, H-1, N-2H}$

$$W_1 = \frac{\sum_{h=1}^H \left[\hat{\alpha}_h - \frac{\sum_{j=1}^H \hat{\alpha}_j a_{jh}}{\sum_{i=1}^H a_{ii}} \right]^2 a_{hh}}{(H-1)\hat{\sigma}^2}$$

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. $H_0 : \alpha_1 + \beta_1 x_0 = \alpha_2 + \beta_2 x_0 = \dots = \alpha_H + \beta_H x_0$ (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

$$W_0 = \frac{\sum_{h=1}^H \left[(\hat{\alpha}_h + \hat{\beta}_h x_0) - \frac{\sum_{j=1}^H (\hat{\alpha}_j + \hat{\beta}_j x_0) c_{jh}}{\sum_{i=1}^H c_{ii}} \right]^2 c_{hh}}{(H-1)\hat{\sigma}^2}$$

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

O referido autor derivou um teste para a hipótese em que os H modelos lineares são idênticos, utilizando a notação matricial. Neste caso, considerou os H seguintes modelos lineares :

$$\begin{aligned} y_1 &= X_1 \beta_1 + \varepsilon_1 \\ y_2 &= X_2 \beta_2 + \varepsilon_2 \\ &\vdots \\ y_H &= X_H \beta_H + \varepsilon_H \end{aligned}$$

em que:

y_h : vetor das observações do h-ésimo modelo, de dimensões $n_h \times 1$;

X_h : matriz dos coeficientes do h-ésimo modelo, de dimensões $n_h \times p$;

β_h : vetor de parâmetros do h-ésimo modelo, de dimensões $p \times 1$;

ε_h : vetor dos erros alcatórios, do h-ésimo modelo, de dimensões $n_h \times 1$;

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

$$y = X\beta + \varepsilon$$

em que:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_H \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_H \end{bmatrix} \quad X = \begin{bmatrix} X_1 & 0 & \dots & 0 \\ 0 & X_2 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & X_H \end{bmatrix} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_H \end{bmatrix}$$

Então, a seguinte hipótese é testada:

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

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

Rejeita-se H_0 se a estatística $W \geq F_{\alpha(H-1)p, N-Hp}$ sendo $N = \sum_{h=1}^H n_h$,

em que:

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

em que:

X^- : matriz inversa de Moore-Penrose;

p : número de parâmetros.

Representando a diferença entre a soma de quadrados de todos os parâmetros e a soma de quadrados de parâmetros de um modelo reduzido, na qual os vetores β_h são considerados iguais.

Esta metodologia foi utilizada por Regazzi (1993), que considerou o ajustamento dos dados de observação relativos a H equações de regressão polinomial do segundo grau, utilizando a técnica dos polinômios ortogonais. As H equações são dadas por:

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

em que:

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

$$e_{hi} \sim \text{NID}(0, \sigma^2);$$

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

O autor considerou as seguintes hipóteses:

1. $H_0 : \beta_1 = \beta_2 = \dots = \beta_H$ (as H equações são idênticas)
2. $H_0 : a_1 = a_2 = \dots = a_H$ (as H equações têm uma constante de regressão comum)

3. $H_0 : c_1 = c_2 = \dots = c_H$ (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:

$$\mathbf{y}_h = \mathbf{X}_h \boldsymbol{\beta}_h + \boldsymbol{\varepsilon}_h \quad (3)$$

em que:

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

Escrevendo esses H modelos na forma do modelo linear geral:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (4)$$

em que:

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_H \end{bmatrix}_{N \times 1} \quad \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_H \end{bmatrix}_{Hp \times 1} \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_H \end{bmatrix}_{N \times 1}$$

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{X}_3 & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{X}_H \end{bmatrix}_{N \cdot H \times p}$$

O sistema de equações normais relativo ao modelo (4), obtido através do método dos quadrados mínimos, tem a seguinte forma:

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y} \quad (5)$$

ou

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

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

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

e ainda, sendo a matriz $(\mathbf{X}'\mathbf{X})^{-1}$, 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:

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_H \end{bmatrix} = \begin{bmatrix} (X_1' X_1)^{-1} X_1' y_1 \\ (X_2' X_2)^{-1} X_2' y_2 \\ \vdots \\ (X_H' X_H)^{-1} X_H' y_H \end{bmatrix}$$

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

$$\begin{aligned} SQPar(c) &= \hat{\beta}' X' y \\ &= \sum_{h=1}^H \hat{\beta}_h' X_h' y_h \end{aligned} \quad (7)$$

com Hp graus de liberdade (H modelos, com p parâmetros cada um) c , de modo específico para o caso de regressão polinomial de segundo grau, tem-se $p=3$.

A soma de quadrados total é obtida por:

$$\begin{aligned} SQTotal(c) &= y' y \\ &= \sum_{h=1}^H y_h' y_h \end{aligned} \quad (8)$$

com N graus de liberdade.

A soma de quadrados de residuo é obtida pela diferença:

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

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	$\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}$	
Resíduo (c)	N-Hp	$\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}$	$\frac{\text{SQRes}}{\text{gl}}$
Total	N	$\mathbf{y}'\mathbf{y}$	

De acordo com Regazzi (1993), $\frac{SQRes}{gl} = \hat{\sigma}^2$ é o estimador comum da

variância residual, que também pode ser obtido pela média ponderada dos estimadores das variâncias residuais de cada modelo.

O modelo pressupõe homocedasticidade de variâncias entre os H modelos. Um modo prático, segundo Pimentel Gomes (2000), para verificar esta pressuposição, é avaliar se existe uma relação entre variâncias menor que sete. Neste caso, então é possível combinar as variâncias residuais dos H conjuntos ou modelos, obtendo-se uma estimativa comum.

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

O primeiro teste considera a seguinte hipótese de nulidade:

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

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

em que:

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

a, b, c : parâmetros comuns.

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

$$y = Z\theta + \varepsilon \quad (11)$$

em que:

y : vetor dos valores observados da variável resposta, de dimensão $(N \times 1)$;

ε : vetor dos erros aleatórios, de dimensão $(N \times 1)$;

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

em (4).

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

O sistema de equações normais relativo ao modelo reduzido (11), obtido pelo método dos quadrados mínimos, conforme Graybill(1976), Draper e Smith (1981) e Regazzi (1993), é:

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

sendo \mathbf{Z} de posto coluna completo p e, no caso de Regazzi (1993), $p = 3$, então $\mathbf{Z}'\mathbf{Z}$ tem dimensão $p \times p$ e não-singular, logo o estimador do vetor dos parâmetros para o modelo reduzido é:

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

sendo a matriz $Z'Z$ composta pela soma das matrizes $X_h'X_h$ de cada modelo, bem como a matriz $Z'y$. O estimador do vetor dos parâmetros comuns pode ser escrito do seguinte modo:

$$\hat{\theta} = \left(\sum_{h=1}^H X_h'X_h \right)^{-1} \sum_{j=1}^H X_j'y_j$$

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

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

ou

$$\text{SQPar}(r1) = \left(\sum_{j=1}^H y_j'X_j \right) \left(\sum_{h=1}^H X_h'X_h \right) \left(\sum_{t=1}^H X_t'y_t \right)$$

com p graus de liberdade.

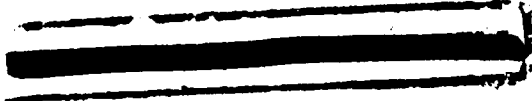
A redução devida a H_0 (coeficientes iguais) é obtida pela diferença:

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

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

Neste caso, o autor testou a seguinte hipótese:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_H \text{ (as H equações são idênticas)}$$


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

utilizando a estatística F, dada por:

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

Segundo 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 foi considerado como:

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

De acordo com o autor, a não rejeição da H_0 permite concluir que, a uma significância α , as H equações não diferem entre si. Portanto, a equação ajustada com as estimativas dos parâmetros comuns pode ser usada como uma estimativa das H equações envolvidas.

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

CV	GL	SQ	QM	F_c
Parâmetros (β)	(Hp)	$S_1 = \hat{\beta}' X' y$		
Parâmetros (θ)	p	$S_2 = \hat{\theta}' Z' y$		
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	$S_5 = y' y$		

O segundo teste considerou a seguinte hipótese de nulidade

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

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

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 β_h e X_h em (3) é:

$$\beta_h = \begin{bmatrix} \alpha_h \\ \delta_h \end{bmatrix} \quad \text{e} \quad X_h = [\mathbf{u}_h \quad \mathbf{V}_h]$$

em que α_h possui dimensão 1×1 e δ_h possui dimensão $(p-1) \times 1$;

\mathbf{u}_h : vetor relativo ao termo constante α , 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 $n_h \times (p-1)$,

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

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

em que:

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_H \end{bmatrix}_{N \times 1} \quad \boldsymbol{\gamma} = \begin{bmatrix} \alpha \\ \delta_1 \\ \delta_2 \\ \vdots \\ \delta_H \end{bmatrix}_{H(p-1)+1 \times 1} \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_H \end{bmatrix}_{N \times 1}$$

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

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

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

e o estimador dos parâmetros:

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

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

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

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

$$H_0: a_1 = a_2 = \dots = a_H \text{ (as } H \text{ equações têm uma constante de regressão comum)}$$

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

o autor utilizou a estatística F, dada por:

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

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

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

TABELA 3 - Análise de variância relativa ao teste de hipótese $H_0: a_1 = a_2 = \dots = a_H$ (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 tipo de comparação estudada por Regazzi (1993) é aquela na qual foi considerado de interesse em avaliar a seguinte hipótese de nulidade.

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

$$y_{hi} = \alpha_h + b_h P_{1hi} + c P_{2hi} + e_{hi} \quad (20)$$

em que:

c : parâmetro comum

$y_{hi}, P_{khi}, \alpha_h, b_h$ 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:

$$\beta_h = \begin{bmatrix} \alpha_h \\ \psi_h \end{bmatrix}_{p \times 1} \quad c \quad X_h = [U_h \quad V_h]$$

em que α_h possui dimensão $p_1 \times 1$ ($0 < p_1 < p$) e ψ_h possui dimensão $p_2 \times 1$ ($p_2 = p - p_1$).

Um caso geral da hipótese H_0 é:

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

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

$$y = W\xi + \varepsilon \quad (21)$$

em que:

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_H \end{bmatrix}_{N \times 1} \quad \boldsymbol{\xi} = \begin{bmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \\ \vdots \\ \boldsymbol{\alpha}_H \\ \boldsymbol{\Psi} \end{bmatrix}_{[Hp_1 + p_2] \times 1} \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_H \end{bmatrix}_{N \times 1}$$

$$\mathbf{W} = \begin{bmatrix} \mathbf{U}_1 & 0 & \cdots & 0 & \mathbf{V}_1 \\ 0 & \mathbf{U}_2 & \cdots & 0 & \mathbf{V}_2 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & \mathbf{U}_H & \mathbf{V}_H \end{bmatrix}_{N \times [Hp_1 + p_2]}$$

O sistema de equações normais relativo ao modelo reduzido (21), obtido pelo método dos quadrados mínimos, é:

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

então, o estimador dos parâmetros é:

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

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

$$\text{SQPar}(r_3) = \hat{\boldsymbol{\xi}}' \mathbf{W}\mathbf{y}$$

com $Hp_1 + p_2$ 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:

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

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

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 $F_c \geq F_{\tau[\alpha; (H-1)p2, N-Hp]}$.

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

Segundo Regazzi (1993), esse teste é geral, podendo-se aplicá-lo para testar a igualdade de um ou mais coeficientes de regressão.

TABELA 4 - Análise de variância relativa ao teste de hipótese

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

CV	GL	SQ	QM	F _c
Parâmetros (β)	(Hp)	$S_1 = \hat{\beta}' X' y$		
Parâmetros (ξ)	Hp1+p2	$S_2 = \hat{\xi}' W y$		
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	$S_5 = y' y$		

A metodologia adotada por Regazzi (1993) foi ilustrada a partir de dados relativos à produção de quatro variedades em sete níveis de adubação, sendo considerado o modelo polinomial do segundo grau. O autor verificou que a identidade de modelos de regressão, ou igualdade de qualquer subconjunto de parâmetros, podem ser verificadas pelo teste F. Também concluiu que a metodologia é geral e pode ser usada em modelos polinomiais de qualquer grau, ortogonal ou não, e também em modelos de regressão múltipla.

O referido autor, 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. As equações usadas para as comparações posteriores foram:

$$\begin{aligned}
y_{1i} &= a_{01} + a_{11}x_{1i} + \lambda_1(x_{1i} - K)Z_i + e_{1i} & i = 1, 2, \dots, n_1 \\
y_{2i} &= a_{02} + a_{12}x_{2i} + \lambda_2(x_{2i} - K)Z_i + e_{2i} & i = 1, 2, \dots, n_2 \\
&\vdots \\
y_{hi} &= a_{0h} + a_{1h}x_{hi} + \lambda_h(x_{hi} - K)Z_i + e_{hi} & i = 1, 2, \dots, n_h
\end{aligned}$$

em que:

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

x_{hi} : i -ésimo valor da variável independente do h -ésimo modelo;

$a_{0h}, a_{1h}, \lambda_h$: parâmetros do h -ésimo modelo;

K : ponto de interseção dos dois submodelos, considerado conhecido e igual para as H equações;

Z_i : variável binária, tal que $Z_i = 0$ para $X_{hi} \leq K$ e $Z_i = 1$ para $X_{hi} > K$;

e_{hi} : erro alcatório, associado à i -ésima observação do h -ésimo modelo, com valores supostos independente e normalmente distribuídos com média zero e variância comum σ^2 , isto é, $e_{hi} \sim NID(0, \sigma^2)$.

Considerou, também, o caso da justaposição de $r = 2$ submodelos polinomiais do segundo grau: neste caso, os H modelos de regressão estudados foram:

$$\begin{aligned}
y_{1i} &= a_{01} + a_{11}x_{1i} + a_{21}x_{1i}^2 + \lambda_1(x_{1i} - K)^2 Z_i + e_{1i} & i = 1, 2, \dots, n_1 \\
y_{2i} &= a_{02} + a_{12}x_{2i} + a_{22}x_{2i}^2 + \lambda_2(x_{2i} - K)^2 Z_i + e_{2i} & i = 1, 2, \dots, n_2 \\
&\vdots \\
y_{hi} &= a_{0h} + a_{1h}x_{hi} + a_{2h}x_{hi}^2 + \lambda_h(x_{hi} - K)^2 Z_i + e_{hi} & i = 1, 2, \dots, n_h
\end{aligned}$$

em que:

$$Z_i = 0, \text{ para } x_{hi} \leq K$$

$$Z_i = 1, \text{ para } x_{hi} > K$$

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

Alguns trabalhos têm utilizado essa metodologia na área florestal. Sousa (1989) estudou 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.

Lopes (1996) aplicou o método para comprovar se uma única equação com estimativas comuns dos parâmetros era suficiente para explicar a estimação do peso da árvore em função das variáveis diâmetro, altura e idade para duas espécies de *Eucalyptus*: *Eucalyptus cloesiana* e *Eucalyptus camaldulensis*, e três idades: 49, 62 e 74 meses. Para isso, utilizou a equação de Schumacher logaritmizada. Porém, em contradição com o autor anterior, concluiu que a estimação do peso nas três idades poderia ser feita por uma equação em função somente do diâmetro e da altura da árvore.

2.2.2 Uso de variáveis binárias (*dummy*)

Alguns autores preferem o uso de variáveis binárias, também referenciando como variáveis *dummy*, indicadoras ou classificatórias, para testar a igualdade de equações ou coeficientes.

Gujarati (1970b) mostrou como variáveis *dummy*, que são definidas como aquelas que assumem somente dois valores 1 e 0, podem ser usadas como uma alternativa para a análise padrão de métodos de análise de variância e de covariância e o teste de Chow (1960), este já descrito anteriormente.

Para isto considerou a seguinte relação, referente a dois conjuntos de dados:

$$y_i = a_0 + a_1 D + a_2 x_i + a_3 (Dx_i) + u_i, \quad i = 1, \dots, (n_1 + n_2)$$

em que:

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

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

As variáveis binárias foram introduzidas na forma aditiva e multiplicativa. Os coeficientes a_1 e a_3 são diferenças de interceptos e inclinações, respectivamente. Se $H_0 : a_1 = 0$ é rejeitada, ou seja, a_1 é significativo, então o valor do intercepto do primeiro conjunto é obtido por $a_1 + a_0$, neste caso a_0 é o intercepto do segundo conjunto. Se $H_0 : a_1 = 0$ não é rejeitada, ou seja, a_1 é não significativo, então a_0 representa o intercepto comum para ambos os conjuntos.

Se $H_0 : a_3 = 0$ é rejeitada, então o valor da inclinação do primeiro conjunto é obtido por $a_2 + a_3$, neste caso a_2 é a inclinação do segundo conjunto. Se $H_0 : a_3 = 0$ não é rejeitada, então a_2 representa a inclinação comum para ambos os conjuntos.

Portanto, incluindo variáveis binárias aditivas ou multiplicativas, é possível verificar se duas equações lineares diferem em intercepto, em inclinação, ou ainda em ambos.

O autor verificou que este método fornece resultados idênticos aos do teste de Chow (1960), mas aponta certa vantagem para a técnica de variáveis binárias. Primeiro, esta técnica indica a(s) fonte(s) de diferença entre as regressões lineares, isto é, se a diferença é devido a intercepto, ou inclinações, ou ambos. Segundo, em apenas uma regressão pode-se obter todas as informações necessárias, ao passo que o teste Chow é um procedimento de vários estágios.

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

A discussão foi realizada utilizando regressão linear múltipla, com duas variáveis independentes e quatro grupos (tratamentos), conforme descrito abaixo:

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

o qual foi descrito mais explicitamente da seguinte forma:

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

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

Segundo 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 combinações possíveis.

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

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

$$+ e, \quad i = 1, \dots, (n_1 + n_2 + n_3 + n_4)$$

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

A interpretação dos vários coeficientes é feita 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 por diante.

As seguintes equações são derivadas da equação (22), pelo método dos mínimos quadrados ordinários, assumindo $E(u_j) = 0$, $E(u_j, x_{ij}) = 0$ e $E(u_j, u_{j,k}) = \sigma^2$ para $K = 0$ e zero se $K \neq 0$:

grupo 1: $\hat{y} = \hat{\alpha}_0 + \hat{\alpha}_4 x_1 + \hat{\alpha}_8 x_2$

grupo 2: $\hat{y} = (\hat{\alpha}_0 + \hat{\alpha}_1) + (\hat{\alpha}_4 + \hat{\alpha}_5)x_1 + (\hat{\alpha}_8 + \hat{\alpha}_9)x_2$ (23)

grupo 3: $\hat{y} = (\hat{\alpha}_0 + \hat{\alpha}_2) + (\hat{\alpha}_4 + \hat{\alpha}_6)x_1 + (\hat{\alpha}_8 + \hat{\alpha}_{10})x_2$

grupo 4: $\hat{y} = (\hat{\alpha}_0 + \hat{\alpha}_3) + (\hat{\alpha}_4 + \hat{\alpha}_7)x_1 + (\hat{\alpha}_8 + \hat{\alpha}_{11})x_2$

Dependendo da 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 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.

Ressaltou, também, que o número de variáveis binárias é uma a menos que o número de grupos; caso contrário, a matriz $X'X$ é singular.

O uso de variáveis binárias também foi discutido na regressão por Seber (1977), Draper e Smith (1981), Neter, Wassermann e Kutner (1990).

Draper e Smith (1981) comentaram que as variáveis binárias podem assumir quaisquer valores, mas 0 e 1 são geralmente melhores e são os mais utilizados. Ilustram a técnica considerando três conjuntos de dados, *G*, *V* e *W*; com o seguinte modelo:

$$Y = \beta_0 + \beta_1 X + \alpha_1 Z_1 + \alpha_2 Z_2 + e$$

em que:

$Z_1 = 1$, para as observações do conjunto G
= 0, cc

$Z_2 = 1$, 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. De acordo com os autores, pode-se utilizar o teste t para testar a diferença entre os interceptos. Por exemplo, a diferença W-G é estimada por α_1 . A estimativa desse coeficiente, dividido pela estimativa de seu respectivo desvio-padrão, que é obtido tomando-se a raiz quadrada da sua variância ou do termo apropriado da diagonal principal da matriz $(X'X)^{-1}S^2$, 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 $H_0 : \alpha_1 \neq 0$.

Um teste alternativo e equivalente, segundo Draper e Smith (1981), é dado por:

$$F = \frac{SQ(\alpha_1 / b_0, b_1, \alpha_2) / 1}{S^2}$$

em que $S^2 = QME$ e $SQ(\alpha_1 / b_0, b_1, \alpha_2)$ é a soma de quadrados relativa ao coeficiente α_1 ajustada para os coeficientes b_0, b_1 e α_2 .

Este é comparado com $F_{(1, n-4, 1-\alpha)}$. O resultado é idêntico para os dois testes, pois o valor de F para 1 grau de liberdade é, teoricamente, o t^2 . O teste

para $H_0 : \alpha_2 = 0$ é realizado da mesma forma, sendo que α_2 representa a diferença entre interceptos dos conjuntos V e W.

De acordo com os autores, a diferença estimada de G-V ($Z_1 - Z_2$) é dada por $a_1 - a_2$, que possui uma variância estimada $V(a_1) + V(a_2) - 2Cov(a_1, a_2)$, e estes três termos são obtidos tomando-se valores adequados na matriz $(X'X)^{-1}S^2$. Se o valor t é não significativo, então não existe diferença significativa entre os interceptos dos dois conjuntos.

Termos de interação envolvendo variáveis binárias são abordados por Draper e Smith (1981). Como ilustração, verificam a possibilidade de usar o mesmo modelo ajustado para dois conjuntos de dados, como segue:

$$Y = \beta_0 + \beta_1 X + \beta_{11} X^2 + \alpha_0 Z + \alpha_1 XZ + \alpha_{11} X^2 Z + e$$

em que Z é a variável binária que assume o valor 0 para um conjunto de dados e 1 para o outro; então é possível verificar que:

1. $H_0 : \alpha_0 = \alpha_1 = \alpha_{11} = 0$. Se H_0 é rejeitada, conclui-se que os modelos não são o mesmo;
2. Se H_0 é rejeitada em (1), 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.
3. Se H_0 em (2) 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.

Os autores comentaram que outras seqüências de testes podem ser usadas, podendo se estender para situações com mais conjuntos de dados e outros modelos com mais variáveis preditoras X_1, X_2, \dots, X_K .

Hoffmann e Vieira (1998) utilizaram a técnica de variáveis binárias para comparar equações de regressão, comentando 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, contudo os resultados obtidos são equivalentes.

Também comentaram que o número de variáveis binárias deve ser igual ao número de grupos menos 1. Caso contrário, a primeira coluna da matriz X será igual à soma das colunas seguintes e, conseqüentemente, a matriz $X'X$ será singular. Apresentam como alternativa para avaliar, por exemplo, o modelo

$$y_i = \alpha + \gamma_1 Z_{1i} + \gamma_2 Z_{2i} + \gamma_3 Z_{3i} + \beta x_i + e_i$$

retirar o parâmetro α e incluir tantas variáveis binárias quanto seja o número de grupos envolvidos, como:

$$y_i = \gamma_1 Z_{1i} + \gamma_2 Z_{2i} + \gamma_3 Z_{3i} + \gamma_4 Z_{4i} + \beta x_i + e_i$$

Scolforo (1997) apresentou a análise de covariância com variáveis *dummy*. Como ilustração considerou três conjuntos de dados de barbatimão¹ de diferentes locais, obtidos a partir de cubagem rigorosa², em que uma equação de

¹ Árvore da família das Leguminosas, subfamília das Mimosáceas, cuja casca, adstringente, se usa na medicina e no curtimento de couros (*Stryphodendron barbatiman*).

² Procedimento dendrométrico, em que são medidos diâmetros ao longo do fuste com o objetivo de se obter o volume da árvore.

volume foi estimada para cada conjunto. O modelo utilizado foi o Spurr, descrito como:

$$V = \beta_0 + \beta_1 D^2 H$$

em que:

V : volume com casca

D : diâmetro (1,30 m do solo)

H : altura

β_0, β_1 : coeficientes do modelo

O autor verificou que uma única equação não deve ser utilizada para estimar o volume do barbatimão nas três regiões consideradas.

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

2.2.3 Comparações múltiplas

Alguns autores utilizaram a técnica das comparações múltiplas no estudo da identidade de modelos de regressão. De acordo com Banzatto e Kronka (1995), os testes de comparações múltiplas, ou testes de comparações de médias, servem como um complemento do teste F, e são adequados para detectar quais são as diferenças existente entre os tratamentos.

Dentre os testes de comparações múltiplas, segundo, Hinkelmann e Kempthorne (1994), Banzatto e Kronka (1995) e Pimentel Gomes (2000) citam-se: teste de Tukey, teste de Duncan, Student Newman Keuls (SNK) e Scott e

Knott. Detalhes sobre estes testes podem ser encontrados em textos tradicionais de estatística experimental, como já mencionados.

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. Após, 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 4 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 propiciada por estas foram precisas.

Leite e Regazzi (1992) e Silva et al. (1997) utilizaram a técnica de comparações múltiplas para comparar equações de regressão.

2.2.4 Análise de agrupamento

Segundo Mardia, Kent e Bibby (1979), a análise de agrupamento tem por finalidade reunir, por algum critério de classificação, as unidades amostrais (indivíduos, objetos, locais, etc) em vários grupos, de tal forma que exista homogeneidade dentro e heterogeneidade entre grupos. Alternativamente, as

técnicas de análise de agrupamento têm por objetivo dividir um grupo original de observações em vários grupos homogêneos, segundo algum critério de similaridade ou dissimilaridade (Cruz, 1990).

Segundo Cruz (1990), no processo de conglomeração existem basicamente duas etapas. A primeira envolve a estimação de uma medida de similaridade (ou dissimilaridade) entre os indivíduos a serem agrupados. A segunda adota uma técnica de agrupamento para a formação dos conglomerados.

De acordo com Bussab, Miazaki e Andrade (1990), nas medidas de similaridade, quanto maior o valor observado, mais parecidos são os objetos, enquanto para medidas de dissimilaridade, quanto maior o valor observado, menos parecidos serão os objetos. Estes resultados podem ser ilustrados graficamente pelo dendograma (Figura 2), com a formação da “árvore” e ramificações.

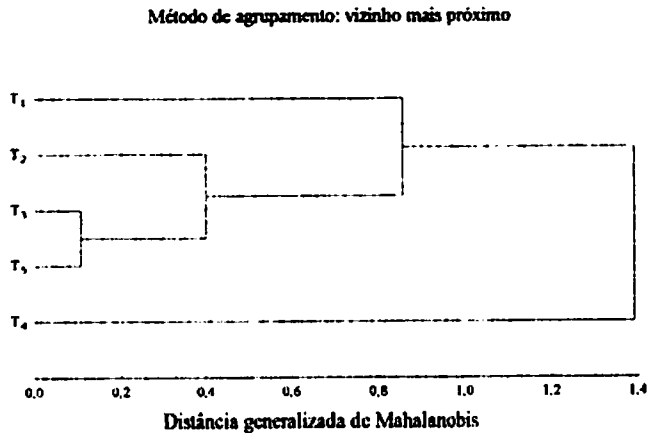


FIGURA 2 - Dendograma resultante de uma análise de agrupamento

Segundo Leite e Regazzi (1992), a medida de dissimilaridade adequada para dados de um modelo de delineamento experimental é a distância generalizada de Mahalanobis, que leva em consideração as correlações entre

características analisadas. Segundo Mardia, Kent e Bibby (1979) e Cruz (1990), a distância generalizada de Mahalanobis entre dois objetos \mathbf{x}_i e $\mathbf{x}_{i'}$ é dada por:

$$D_{ii'}^2 = (\mathbf{x}_i - \mathbf{x}_{i'}) \mathbf{S}^{-1} (\mathbf{x}_i - \mathbf{x}_{i'}) \quad (24)$$

em que:

$$\begin{aligned} \mathbf{x}_i &: [x_{i1}, x_{i2}, \dots, x_{ij}] \\ \mathbf{x}_{i'} &: [x_{i'1}, x_{i'2}, \dots, x_{i'j}] \quad j = 1, 2, \dots, J \text{ variáveis} \end{aligned}$$

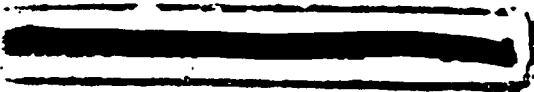
\mathbf{S} : matriz de covariâncias residuais

Bussab, Miazaki e Andrade (1990), apresentando alguns métodos de agrupamentos, definiram que nos métodos hierárquicos, estes utilizados no presente trabalho, os objetos são classificados em grupos em diferentes etapas, de modo hierárquico, produzindo uma árvore de classificação. São subdivididos em aglomerativos e divisivos.

Nos métodos aglomerativos, mais populares, ocorrem fusões sucessivas dos n objetos, sendo obtidos $n-1$, $n-2$, etc grupos, até reunir todos os objetos num único grupo. Já nos métodos divisivos existe um único grupo; e por divisões sucessivas vão sendo obtidos 2,3, etc. grupos.

Os métodos aglomerativos, segundo Bussab, Miazaki e Andrade (1990), podem ser subdivididos em:

1. Método da Centróide;
2. Método das Médias das Distâncias;
3. Método da Ligação Simples ou do Vizinho mais próximo;
4. Método da Ligação Completa ou do Vizinho mais distante.



Para Leite e Regazzi (1992), as técnicas de análise de agrupamento surgem como opções atrativas para comparar equações de regressão, principalmente quando existe mais de uma fonte de comparação e para os casos em que os modelos envolvidos no estudo não apresentam uma mesma relação funcional. Os autores, utilizando a técnica de análise de agrupamento para avaliar a igualdade de equações volumétricas estimadas para *Eucalyptus saligna*, em regime de alto fuste e de primeira talhadia, verificaram a necessidade de se usar equações independentes por regime de corte.

No estudo de Silva et al. (1997), três alternativas de estimação do volume comercial, em árvores de *Eucalyptus grandis*, foram avaliadas utilizando entre outras técnicas a análise de agrupamento.

3 MATERIAL E MÉTODOS

O ajuste de equações volumétricas, baseadas em árvores cubadas rigorosamente é um ponto estratégico para que o volume de cada árvore, contida na parcela, que compõe o inventário florestal possa ser estimado.

De maneira geral estas equações devem ser obtidas para cada estrato, que nada mais é do que a congregação de talhões (unidade mínima de manejo) que possuem mesma idade, espécie, espaçamento, estão implantados num mesmo tipo de solo e sujeitos a mesmos tratamentos silviculturais.

Entretanto é comum que um empreendimento florestal tenha um grande número de estratos, 500, 1000, 2000 ou mais. Assim gerar tantas equações de volume implica num alto custo de cubagem rigorosa.

Para atender ao objetivo deste estudo, fez-se então uso de uma amostra contendo 9 estratos para os quais avaliou-se a possibilidade de agrupamento da base de dados de cubagem rigorosa com o intuito de melhor compreender como deve ser a estratégia de cubagem rigorosa na empresa florestal. Será mesmo que é necessário o ajuste de equação para cada estrato?

3.1 Material

Os dados utilizados neste estudo foram provenientes das propriedades da empresa VCP – Votorantin Celulose e Papel.

Foram usados 9 estratos (Tabela 5), definidos em função das espécies *Eucalyptus grandis* e *Eucalyptus urophylla*; do espaçamento: 3,0 x 1,8 m e 3,0 x 2,5 m; da idade, cuja amplitude variou de 4 a 7 anos; e do tipo de solo: AQ1, AQ2, Lcm1, Lcm2, LR e Lca.

Em cada estrato foram cubadas rigorosamente no mínimo 60 árvores. Estas foram distribuídas nas classes diamétricas com amplitude de classe igual a 5 cm. Em média cada classe diamétrica conteve 10 árvores, o que foi um número suficiente de repetições em cada classe conforme afirmativa de Scolforo e Figueiredo Filho (1997). Para cada árvore foram coletadas as variáveis: diâmetro a 1,30 m do solo (dap) em cm, altura total da árvore em metros, além de se proceder a cubagem rigorosa a partir de seções com 1,40 m de comprimento. O cálculo do volume real da árvore em m³ foi viabilizado através da fórmula de Smalian.

Os estratos disponíveis foram agrupados em quatro grupos, em função de um fator de interesse para aplicação dos testes, do seguinte modo:

- grupo 1: os estratos diferem quanto aos tipos de solos, para a espécie *Eucalyptus grandis*;
- grupo 2: os estratos diferem quanto aos tipos de solos, para a espécie *Eucalyptus urophylla*;
- grupo 3: os estratos diferem quanto à idade, para a espécie *Eucalyptus grandis*;
- grupo 4: os estratos diferem quanto à espécie.

TABELA 5 - Características dos grupos estudados com relação aos estratos, número de árvores por estrato, espécie de *Eucalyptus*, tipos de solos, espaçamento entre plantas e idade após plantio.

Grupo	Estrato	Nº arv.	Espécie *	Solo	Espaçamento (m)	Idade (anos)	Interesse
1	6	60	Eg	Lem2	3,0 x 1,8	7	Avaliar se é possível agrupar árvores de diferentes tipos de solo para <i>E. grandis</i>
	32	59	Eg	AQ1	3,0 x 1,8	7	
	48	60	Eg	LR AQ1	3,0 x 1,8	7	
2	50	58	Eu	AQ2	3,0 x 1,8	6	Avaliar se é possível agrupar árvores de diferentes tipos de solo para <i>E. urophylla</i>
	69	56	Eu	AQ1	3,0 x 1,8	6	
	121	59	Eu	Lem1	3,0 x 1,8	6	
3	4	60	Eg	Lea Lem1	3,0 x 2,5	5	Avaliar se é possível agrupar árvores com diferentes idades
	13	58	Eg	Lea Lem1	3,0 x 2,5	4	
4	47	55	Eg	AQ2	3,0 x 1,8	6	Avaliar se é possível agrupar árvores de diferentes espécies
	50	58	Eu	AQ2	3,0 x 1,8	6	

* Eg - *Eucalyptus grandis* e Eu - *Eucalyptus urophylla*

3.2 Métodos

3.2.1 Modelo de regressão

Para expressar o volume das árvores em função do diâmetro e da altura, utilizou-se o modelo volumétrico de Schumacher e Hall, por ser um dos mais tradicionais da literatura florestal mundial, conforme observado em Loetch, Zöhner e Haller (1964) e utilizado no Brasil com precisão por pesquisadores e na própria operação florestal, conforme Paula Neto et al. (1983), Herrera (1989), Leite e Regazzi (1992), Scolforo (1997) e Assis (1998).

O modelo de Schumacher e Hall tem a seguinte forma:

$$V_i = \beta_0 \cdot D_i^{\beta_1} \cdot H_i^{\beta_2} \cdot e_i$$

em que:

V_i : volume total com casca;

D_i : diâmetro a 1,30m de altura;

H_i : altura total da árvore;

β_i : parâmetros do modelo, sendo $i = 0, 1, 2$;

e_i : erro aleatório.

A linearização do modelo é obtida através da aplicação de propriedades logarítmicas:

$$\ln V_i = \ln \beta_0 + \beta_1 \ln D_i + \beta_2 \ln H_i + \ln e_i$$

e para facilitar a apresentação da metodologia, este é reescrito da seguinte forma:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i,$$

em que:

$$\begin{aligned} y_i &= \ln V_i & \beta_0 &= \ln \beta_0' \\ x_{1i} &= \ln D_i & \varepsilon_i &= \ln e_i \\ x_{2i} &= \ln H_i \end{aligned}$$

3.2.2 Aplicação dos métodos de comparação entre equações de regressão

Para cada um dos quatro grupos descritos anteriormente, foram aplicados os seguintes métodos de comparação entre equações de regressão:

- identidade de modelos;
- variáveis binárias;
- análise de variância;
- análise de agrupamento.

Estes métodos foram descritos detalhadamente no referencial teórico do presente trabalho. Desta forma, julgou-se desnecessário descrevê-los neste item, sendo apresentados os detalhes referentes à aplicação.

Em função da semelhança no procedimento técnico, serão descritos os métodos para o grupo 1. A extensão para os demais é feita de forma simples, sem alterações significativas.

No grupo 1, as comparações de interesse referem-se a comparar os estratos com diferentes tipos de solo para a espécie de *Eucalyptus grandis*. O

interesse aqui foi verificar se as equações ajustadas para os diferentes tipos de solos podem ser expressas por uma mesma equação.

3.2.2.1 Identidade de modelos

Utilizando a metodologia apresentada por Graybill (1976) e Regazzi (1999), os modelos para a amostra 1 são:

$$\begin{aligned} y_{1i} &= \beta_{01} + \beta_{11}x_{11i} + \beta_{21}x_{21i} + \varepsilon_{1i} \\ y_{2i} &= \beta_{02} + \beta_{12}x_{12i} + \beta_{22}x_{22i} + \varepsilon_{2i} \\ y_{3i} &= \beta_{03} + \beta_{13}x_{13i} + \beta_{23}x_{23i} + \varepsilon_{3i} \end{aligned}$$

ou

$$y_{hi} = \beta_{0h} + \beta_{1h}x_{1hi} + \beta_{2h}x_{2hi} + \varepsilon_{hi} \quad (25)$$

em que:

y_{hi} : i -ésima observação da variável resposta do h -ésimo modelo, sendo $i = 1, 2, \dots, n_h$ o número de observações e $h = 1, 2, 3$ o número de modelos, que correspondem aos três estratos;

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,

$$\varepsilon_{hi} \sim \text{NID}(0, \sigma^2)$$

$$\sum_{h=1}^H n_h = N$$

Escrevendo esses H modelos na forma do modelo linear geral:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (26)$$

em que:

\mathbf{y} : vetor dos valores observados da variável resposta ($N \times 1$);

\mathbf{X} : matriz dos valores das variáveis regressoras ($N \times H(K+1)$), sendo $K+1$ o número de parâmetros

$\boldsymbol{\beta}$: vetor dos coeficientes ($H(K+1) \times 1$);

$\boldsymbol{\varepsilon}$: vetor dos erros aleatórios ($N \times 1$), $\boldsymbol{\varepsilon} \sim N(0, \sigma^2 I)$

em que:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix}$$

$$\mathbf{X} = \begin{bmatrix} 1 & X_{111} & X_{211} & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & X_{112} & X_{212} & 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & X_{1160} & X_{2160} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & X_{121} & X_{221} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & X_{122} & X_{222} & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 1 & X_{1256} & X_{2256} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & X_{131} & X_{231} \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & X_{132} & X_{232} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & X_{1358} & X_{2358} \end{bmatrix}$$

As hipóteses consideradas foram as seguintes:

A) Equações idênticas:

$H_0: \beta_1 = \beta_2 = \beta_3$ (as 3 equações são idênticas), isto é, os modelos em (25) reduzem-se à forma:

$$y_{hi} = \beta_0 + \beta_1 x_{1hi} + \beta_2 x_{2hi} + \varepsilon_{hi} \quad (27)$$

em que:

y_{hi}, x_{khi} e ε_{hi} : têm as mesmas especificações dos modelos (25);

$\beta_0, \beta_1, \beta_2$: coeficientes comuns.

Utilizando a notação matricial, os modelos reduzidos (27) podem ser escritos como:

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\theta} + \boldsymbol{\varepsilon} \quad (28)$$

\mathbf{y} : vetor dos valores observados da variável resposta ($N \times 1$)

\mathbf{Z} : matriz dos valores das variáveis regressoras ($N \times (K+1)$)

$\boldsymbol{\theta}$: vetor dos coeficientes comuns ($(K+1) \times 1$)

$\boldsymbol{\varepsilon}$: vetor dos erros aleatórios ($N \times 1$)

Sendo \mathbf{y} e $\boldsymbol{\varepsilon}$ semelhantes ao modelo (26) e :

$$\mathbf{Z} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \mathbf{X}_3 \end{bmatrix} \quad \boldsymbol{\theta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} \text{ é o vetor dos coeficientes comuns.}$$

B) As equações têm o mesmo intercepto:

$H_0: \beta_{01} = \beta_{02} = \beta_{03}$ (as 3 equações têm uma constante de regressão comum), isto é, os modelos em (25) reduzem-se à forma:

$$y_h = \beta_0 + \beta_{1h}x_{1h} + \beta_{2h}x_{2h} + \varepsilon_h \quad (29)$$

em que:

β_0 : coeficiente comum

y_h e ε_h : têm as mesmas especificações dos modelos (25);

Usando a notação matricial, os modelos reduzidos em (29) podem ser escritos como:

$$\mathbf{y} = \mathbf{V}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \quad (30)$$

em que:

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{y}_3 \end{bmatrix} \quad \boldsymbol{\gamma} = \begin{bmatrix} \beta_0 \\ \boldsymbol{\delta}_1 \\ \boldsymbol{\delta}_2 \end{bmatrix} \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \boldsymbol{\varepsilon}_3 \end{bmatrix}$$

$$V = \begin{bmatrix} \mathbf{u}_1 & \mathbf{V}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{u}_2 & \mathbf{0} & \mathbf{V}_2 & \mathbf{0} \\ \mathbf{u}_3 & \mathbf{0} & \mathbf{0} & \mathbf{V}_3 \end{bmatrix}$$

y : vetor dos valores observados da variável resposta ($N \times 1$)

V : matriz dos valores das variáveis regressoras [$N \times (HK+1)$]

γ : vetor dos coeficientes comuns e não-comuns [$(HK+1) \times 1$]

ε : vetor dos erros aleatórios ($N \times 1$)

C) As equações têm um coeficiente de regressão comum:

$H_0 : \beta_{k1} = \beta_{k2} = \beta_{k3}$ (as 3 equações têm um coeficiente de regressão comum), isto é, os modelos em (25) reduzem-se à forma:

Neste caso, considerou-se como igual o terceiro coeficiente, como:

$$y_{hi} = \beta_{0h} + \beta_{1h}x_{1hi} + \beta_3x_{khi} + \varepsilon_{hi} \quad (31)$$

em que:

β_3 : coeficiente comum

y_{hi} e ε_{hi} : têm as mesmas especificações dos modelos em (25);

3.2.2.2 A utilização da técnica de variáveis binárias (*dummy*)

Como uma alternativa para comparar equações de regressão, foi utilizada a metodologia utilizando variáveis binárias, conforme Gujarati (1970a), Draper e Smith (1981) e Scolforo (1997).

Considerou-se que os três modelos eram diferentes entre si. Para testar esta hipótese, o seguinte modelo foi definido para a amostra 1:

$$y_i = \alpha_0 + \alpha_1 Z_1 + \alpha_2 Z_2 + \alpha_3 x_{1i} + \alpha_4 (Z_1 x_{1i}) + \alpha_5 (Z_2 x_{1i}) + \alpha_6 x_{2i} + \alpha_7 (Z_1 x_{2i}) + \alpha_8 (Z_2 x_{2i}) + e_i \quad (32)$$
$$i = 1, \dots, (n_1 + n_2 + n_3)$$

em que:

Z_1 : 1, se as observações pertencem ao grupo 1 e 0, c c

Z_2 : 1, se as observações pertencem ao grupo 2 e 0, c c

α_0 : intercepto para o grupo 3;

α_1, α_2 : diferença de intercepto para o grupo 1 e 2, respectivamente;

α_3 : coeficiente de regressão referente a X_1 , para o grupo 3;

α_4, α_5 : diferença de coeficiente de regressão referente a X_1 , para os grupos 1 e 2, respectivamente;

α_6 : coeficiente de regressão referente a X_2 , para o grupo 3;

α_7, α_8 : diferença de coeficiente de regressão referente a X_2 , para os grupos 1 e 2, respectivamente;

A partir do modelo (32) derivaram-se as seguintes equações, pelo método dos quadrados mínimos ordinários:

$$\begin{aligned}
\text{Grupo 1: } \hat{Y} &= (a_0 + a_1) + (a_3 + a_4)X_1 + (a_6 + a_7)X_2 \\
\text{Grupo 2: } \hat{Y} &= (a_0 + a_2) + (a_3 + a_5)X_1 + (a_6 + a_8)X_2 \\
\text{Grupo 3: } \hat{Y} &= a_0 + a_3X_1 + a_6X_2
\end{aligned}
\tag{33}$$

Dependendo da significância estatística da diferença dos interceptos estimados e coeficientes de regressão estimados, pôde-se saber se os conjuntos de regressões lineares são diferentes.

Aplicou-se o teste *t* para verificar a diferença significativa entre os coeficientes, como segue:

$$\begin{cases} H_0 : a_3 + a_4 = a_3 + a_5 & \text{ou} & a_4 - a_5 = 0 \\ H_1 : a_4 - a_5 \neq 0 \end{cases}$$

$$t = \frac{\lambda' \beta}{\sqrt{\sigma^2 (\lambda' (X' X)^{-1} \lambda)}} \quad n - p \text{ gl}$$

em que:

λ : vetor de contrastes de interesse que avalia uma determinada comparação

3.2.2.3 Análise de variância e comparações múltiplas

Para a utilização da metodologia de comparações múltiplas, apresentada por Leite e Regazzi (1992), foi adotado um modelo em blocos casualizados, sendo considerados como bloco seis classes de diâmetro, com intervalos de 3 cm, excluindo valores considerados discrepantes. Do total de árvores, foram tomadas aleatoriamente 10 árvores em cada uma dessas classes. Os tratamentos definidos são apresentados na Tabela 6.

TABELA 6 - Tratamentos definidos para metodologia de comparações múltiplas para a amostra 1

Tratamentos	Descrição
1	Volumes observados (controle)
2	Equação ajustada aos dados do estrato 6
3	Equação ajustada aos dados do estrato 32
4	Equação ajustada aos dados do estrato 48
5	Equação ajustada aos dados agrupados

Os dados de volumes dos tratamentos 2, 3, 4 e 5 foram obtidos pelas equações estimadas para cada estrato. A partir destes dados foi realizada a análise de variância, sendo que o tratamento 5 refere-se ao modelo reduzido (comum).

3.2.2.4 A aplicação da análise de agrupamento

Outra alternativa utilizada para verificar a igualdade de equações nos diferentes estratos foi a aplicação da metodologia de análise de agrupamento, da forma apresentada por Leite e Regazzi (1992), considerando uma estrutura multivariada de acordo com a Tabela 6, com 5 tratamentos ou casos (equações) e 6 variáveis (classes de diâmetro).

A medida de dissimilaridade utilizada foi a distância generalizada de Mahalanobis e o método de agrupamento hierárquico foi o Método Vizinho Mais Próximo (Ligação Simples).

Para isto, realizou-se uma análise de variância multivariada, considerando a estrutura descrita anteriormente. Dessa análise de variância, obteve-se a matriz soma de quadrados e produtos residual S_1 e, assim, a matriz

$$S = \frac{S_1}{glres}, \text{ de dimensão } 6 \times 6.$$

A distância entre dois tratamentos, x_i e $x_{i'}$, foi obtida por:

$$D_{ii'}^2 = (x_i - x_{i'}) \cdot S^{-1} \cdot (x_i - x_{i'})$$

em que:

$$\begin{aligned} x_i &: [\bar{x}_{i1}, \bar{x}_{i2}, \dots, \bar{x}_{ij}] \\ x_{i'} &: [\bar{x}_{i'1}, \bar{x}_{i'2}, \dots, \bar{x}_{i'j}], \quad i = 1, 2, 3, 4, 5 \text{ tratamentos} \\ & \quad j = 1, 2, 3, 4, 5, 6 \text{ variáveis} \end{aligned}$$

S : matriz de covariâncias residuais

A partir dessa matriz de distância, foi realizada a análise de agrupamento, utilizando o método do Vizinho Mais Próximo. Inicialmente existiam 5 grupos individuais, referentes aos 5 tratamentos, então os grupos mais próximos foram reunidos e calculadas novas distâncias entre grupos, a partir da expressão apresentada por Bussab, Miazaki e Andrade (1990):

$$d(x_i, x_{i'}) = \min\{d(i, i')\} : i \in x_i, \text{ e } j \in x_{i'},$$

formando uma nova matriz de distância, e assim sucessivamente até o último passo em que um único agrupamento contendo os 5 objetos (tratamentos), foi obtido.

3.2.3 Softwares estatísticos

Os sistemas computacionais utilizados para realizar os testes foram os seguintes: SAS - Statistical Analysis System (1999), BIOESTAT (2000) e STATISTICA (1995). Em anexo está disponível uma rotina do sistema computacional SAS, para os métodos de identidade de modelos e variáveis binárias, para comparação dos estratos da amostra 4.

4 RESULTADOS E DISCUSSÃO

4.1 Identidade de modelos

Na Tabela 7, são apresentadas as estimativas dos coeficientes de regressão, coeficientes de determinação e erros padrões residuais para as equações ajustadas aos estratos de cada grupo, assim como aos dados de todos os estratos agrupados. Pelos critérios adotados para verificar o ajuste das equações, observou-se um comportamento adequado dos mesmos, ou seja, coeficientes de determinação (R^2) estão próximos de 100%, erros padrões residuais (S_{y_x}) muito baixos e análise de resíduos sem tendenciosidade.

O resultado da metodologia de identidade de modelos, referente ao teste de hipótese $H_0 : \beta_1 = \beta_2 = \beta_3$, ou seja, as três equações de regressão são idênticas nos grupos 1 e 2 está apresentado na Tabela 8. Assim como o teste de hipótese $H_0 : \beta_1 = \beta_2$, as duas equações de regressão são idênticas nos grupos 3 e 4. Os resultados completos das análises de variância estão apresentados no Anexo A, nas Tabelas 1A, 2A, 3A e 4A.

Com base nos resultados dos testes apresentados na Tabela 8, verifica-se que H_0 foi rejeitada apenas para o primeiro grupo. Neste caso, existem fortes evidências de que há alguma diferença significativa entre as três equações (nível de significância menor que 1%). Assim, a equação referente ao modelo reduzido não pode ser utilizada.

TABELA 7 - Estimativas dos coeficientes de regressão, coeficientes de determinação (R^2) e erros-padrões residuais (S_{yx}) para as equações ajustadas, referentes aos quatro grupos

Equação	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	R^2_{aj} (%)	S_{yx} (ln m ³)
Amostra 1: solo					
Estrato 6	-10,7057	1,9198	1,2144	98,52	0,0743
Estrato 32	-9,7239	2,1490	0,7129	99,02	0,0910
Estrato 48	-10,0733	1,8233	1,1184	99,01	0,0757
Dados agrupados*	-9,9059	2,0762	0,8304	98,92	0,0885
Amostra 2: solo					
Estrato 50	-10,6277	1,7874	1,3133	98,90	0,0840
Estrato 69	-10,5191	1,8842	1,2042	98,78	0,0707
Estrato 121	-10,6913	1,8575	1,2862	98,90	0,0723
Dados agrupados	-10,5504	1,8578	1,2340	98,86	0,0760
Amostra 3: idade					
Estrato 4	-10,4923	1,9359	1,1394	98,99	0,0771
Estrato 13	-10,7960	1,8438	1,3289	98,80	0,0628
Dados agrupados	-10,5410	1,9221	1,1711	98,90	0,0710
Amostra 4: espécie					
Estrato 47	-10,9543	1,7374	1,4557	99,45	0,0561
Estrato 50	-10,6273	1,7854	1,3147	98,94	0,0833
Dados agrupados	-10,6519	1,7882	1,3169	99,21	0,0720

*modelo reduzido (comum)

TABELA 8 - Resumo da análise de variância considerando o teste de hipótese H_0 : as equações são idênticas,

Grupos	Redução (H_0)			
	GL	QM	F	p
1	6	0,0450	6,94	0,0001
2	6	0,0059	1,03	0,4078
3	3	0,0081	1,64	0,1844
4	3	0,0082	1,60	0,1939

Nos grupos 2, 3 e 4, a hipótese H_0 : as equações de regressão são idênticas não foi rejeitada. Isto conduz à possibilidade de utilizar a equação comum como uma estimativa das equações envolvidas em cada amostra.

No caso do grupo 1, julgou-se interessante verificar qual (is) equação (es) diferem entre si; para isso testou-se a igualdade de pares de equações. Este teste está apresentado na Tabela 9.

Observa-se que os testes para os três pares de comparações possíveis foram significativos a 1 % de probabilidade (nas comparações dos estratos 6 e 32; 6 e 48) e 5% de probabilidade (nos estratos 32 e 48), isto é, que todas as equações diferem.

Dado que os modelos são diferentes, aplicou-se o teste para verificar a igualdade de interceptos e dos demais coeficientes de regressão, neste caso os coeficientes relacionados à variável X_1 e os coeficientes relacionados à variável X_2 , como está apresentado na Tabela 10.

TABELA 9 - Análise de variância dos modelos reduzidos, para comparação das equações duas a duas referente aos estratos 6, 32 e 48

CV	GL	SQ	QM	F	p
Modelo Completo	9	736,8515			
M. Reduzido	6	736,7203			
Redução ($H_0 : \beta_1 = \beta_2$)	3	0,1312	0,0437	6,75	0,0003
M. Reduzido	6	736,6252			
Redução ($H_0 : \beta_1 = \beta_3$)	3	0,2263	0,0754	11,64	0,0001
M. Reduzido	6	736,7840			
Redução ($H_0 : \beta_2 = \beta_3$)	3	0,0675	0,0225	3,47	0,0174
Resíduo	165	1,0693	0,0065		
Total	174	737,9208			

Observando a Tabela 10, pode-se verificar que os três coeficientes, testados separadamente para as três equações, diferem entre si, ou seja, os interceptos (β_0) diferem entre si ao nível de 5% e os coeficientes β_1 relativos à variável X_1 (diâmetro a 1,30 m) e os coeficientes β_2 relativos à variável X_2 (altura) diferem entre si, com uma probabilidade menor que 1%.

Na Tabela 11, estão apresentados os testes para igualdade de pares de interceptos (β_0). Observa-se que apenas os interceptos das equações dos estratos 6 e 32 diferem a 1% de probabilidade.

Na Tabela 12, estão apresentados os testes de igualdade de pares de coeficientes β_1 , observando-se que estes podem ser considerados iguais para os estratos 6 e 48.

TABELA 10 - Análise de variância dos modelos reduzidos, para comparação dos coeficientes entre as equações dos estratos 6, 32 e 48

CV	GL	SQ	QM	F	p
Modelo Completo	9	736,8515			
M. Reduzido	7	736,7928			
Redução ($H_0 : \beta_{01} = \beta_{02} = \beta_{03}$)	2	0,00588	0,0294	4,53	0,0122
M. Reduzido	7	736,7895			
Redução ($H_0 : \beta_{11} = \beta_{12} = \beta_{13}$)	2	0,0620	0,0310	4,78	0,0096
M. Reduzido	7	737,9208			
Redução ($H_0 : \beta_{21} = \beta_{22} = \beta_{23}$)	2	0,0808	0,0404	6,23	0,0025
Resíduo	165	1,0693	0,0065		
Total	174	737,9208			

TABELA 11 - Análise de variância dos modelos reduzidos, para comparação de intercepto entre as equações dos estratos 6, 32 e 48

CV	GL	SQ	QM	F	p
Modelo Completo	9	736,8515			
M. Reduzido	8	736,8008			
Redução ($H_0 : \beta_{01} = \beta_{02}$)	1	0,0507	0,0507	7,82	0,0058
M. Reduzido	8	736,8322			
Redução ($H_0 : \beta_{01} = \beta_{03}$)	1	0,0193	0,0193	2,97	0,0864
M. Reduzido	8	736,8329			
Redução ($H_0 : \beta_{02} = \beta_{03}$)	1	0,0186	0,0186	2,87	0,0923
Resíduo	165	1,0693	0,0065		
Total	174	737,9208			

TABELA 12 - Análise de variância dos modelos reduzidos, para comparação do segundo coeficiente entre as equações dos estratos 6, 32 e 48

CV	GL	SQ	QM	F	p
Modelo Completo	9	736,8515			
M. Reduzido	8	736,8255			
Redução ($H_0 : \beta_{11} = \beta_{12}$)	1	0,0260	0,0260	4,02	0,0467
M. Reduzido	8	736,8485			
Redução ($H_0 : \beta_{11} = \beta_{13}$)	1	0,0030	0,0030	0,50	0,4963
M. Reduzido	8	736,8013			
Redução ($H_0 : \beta_{12} = \beta_{13}$)	1	0,0502	0,0502	7,75	0,0060
Resíduo	165	1,0693	0,0065		
Total	174	737,9208			

TABELA 13 - Análise de variância dos modelos reduzidos, para comparação do terceiro coeficiente entre as equações dos estratos 6, 32 e 48

CV	GL	SQ	QM	F	p
Modelo Completo	9	736,8515			
M. Reduzido	8	736,8069			
Redução ($H_0 : \beta_{21} = \beta_{22}$)	1	0,0446	0,0446	6,89	0,0095
M. Reduzido	8	736,8502			
Redução ($H_0 : \beta_{21} = \beta_{23}$)	1	0,0013	0,0013	0,20	0,6578
M. Reduzido	8	736,8002			
Redução ($H_0 : \beta_{22} = \beta_{23}$)	1	0,0514	0,0514	7,92	0,0055
Resíduo	165	1,0693	0,0065		
Total	174	737,9208			

O mesmo resultado é encontrado comparando-se os coeficientes β_2 , isto é, para os estratos 6 e 48 os coeficientes podem ser considerados iguais (Tabela 13).

Percebe-se que os resultados obtidos pela metodologia de identidade de modelos são de fácil interpretação, conforme já comentado por Leite e Regazzi (1992). Deve ser ressaltada, também, a vantagem deste método quanto à possibilidade de comparação de coeficientes específicos dos modelos em estudo. Contudo, ele somente pode ser aplicado quando os modelos de regressão a serem estudados apresentam as mesmas relações funcionais.

4.2 Variáveis binárias

Os resultados da aplicação da técnica de variáveis binárias para comparação de equações são apresentados a seguir.

Assumindo que os três modelos para o grupo 1 diferem entre si, incluíram-se variáveis binárias num modelo completo para testar tal hipótese. A equação estimada a partir do modelo completo para o grupo 1 é a seguinte:

$$\hat{y}_j = -10,0733 - \underset{0,0864}{0,6324} Z_1 + \underset{0,0923}{0,3494} Z_2 + \underset{0,0001}{1,8233} x_{1j} + \underset{0,4963}{0,0965}(Z_1 x_{1j}) + \underset{0,006}{0,3257}(Z_2 x_{1j}) + \underset{0,0001}{1,1184} x_{2j} + \underset{0,6578}{0,0960}(Z_1 x_{2j}) - \underset{0,0055}{0,4054}(Z_2 x_{2j})$$

em que o valor abaixo de cada coeficiente é a probabilidade de significância de cada estimativa (p).

E as equações estimadas para cada estrato são:

Estrato 6: $\hat{y}_j = -10,7057 + 1,9198x_{1j} + 1,2144x_{2j}$

$$\text{Estrato 32: } \hat{y}_j = -9,7239 + 2,1490x_{1j} + 0,7130x_{2j}$$

$$\text{Estrato 48: } \hat{y}_j = -10,0733 + 1,8233x_{1j} + 1,1184x_{2j}$$

Observa-se que alguns coeficientes referentes às variáveis binárias e às variáveis binárias combinadas são significativos ($p < 0,01$), implicando em que os modelos para os três estratos (6, 32 e 48) não são iguais para o grupo 1, isto é, os estratos possuem equações diferentes.

Também realizando comparações de pares de coeficientes, foram obtidos resultados equivalentes ao teste de identidade de modelos, conforme Graybill (1976) e Regazzi (1999).

A equação estimada a partir do modelo completo para o grupo 2 é a seguinte:

$$\hat{y}_j = -10,6913 + \underset{0,8524}{0,0636} Z_1 + \underset{0,6056}{0,1722} Z_2 + \underset{0,0001}{1,8575} x_{1j} - \underset{0,5547}{0,0700}(Z_1 x_{1j}) + \underset{0,8053}{0,0267}(Z_2 x_{1j}) + \underset{0,001}{1,2862} x_{2j} + \underset{0,8927}{0,0270}(Z_1 x_{2j}) - \underset{0,6570}{0,0820}(Z_2 x_{2j})$$

E as equações estimadas para cada estrato são:

$$\text{Estrato 50: } \hat{y}_j = -10,6277 + 1,7875x_{1j} + 1,3132x_{2j}$$

$$\text{Estrato 69: } \hat{y}_j = -10,5191 + 1,8842x_{1j} + 1,2042x_{2j}$$

$$\text{Estrato 121: } \hat{y}_j = -10,6913 + 1,8575x_{1j} + 1,2862x_{2j}$$

A equação estimada a partir do modelo completo para o grupo 3 é a seguinte:

$$\hat{y}_j = -10,7960 + \underset{0,2481}{0,3037} Z_1 + \underset{0,0001}{1,8438} x_{1j} + \underset{0,3586}{0,0921}(Z_1 x_{1j}) +$$

$$\underset{0,0001}{1,3289} X_{2j} - \underset{0,1988}{0,1895}(Z_1 x_{2j})$$

E as equações estimadas para cada estrato são:

Estrato 4: $\hat{y}_j = -10,4923 + 1,9350x_{1j} + 1,1394x_{2j}$

Estrato 13: $\hat{y}_j = -10,7960 + 1,8438x_{1j} + 1,3289x_{2j}$

A equação estimada a partir do modelo completo para o grupo 4 é a seguinte:

$$\hat{y}_j = -10,6273 - \underset{0,3207}{0,3270} Z_1 + \underset{0,0001}{1,7854} x_{1j} - \underset{0,6750}{0,0480}(Z_1 x_{1j}) +$$

$$\underset{0,0001}{1,3147} x_{2j} - \underset{0,4654}{0,1411}(Z_1 x_{2j})$$

E as equações estimadas para cada estrato são:

Estrato 47: $\hat{y}_j = -10,9543 + 1,7374x_{1j} + 1,4558x_{2j}$

Estrato 50: $\hat{y}_j = -10,6273 + 1,7854x_{1j} + 1,3147x_{2j}$

Nestes três últimos casos, pode-se observar que os coeficientes referentes às variáveis binárias e às variáveis binárias combinadas não possuem diferenças significativas, portanto as equações não diferem entre si, conforme apresentado por Gujarati (1970a), Draper e Smith (1981).

Observa-se ainda que os resultados obtidos neste método são iguais aos apresentados pela metodologia de identidade de modelos.

Uma observação importante referente à técnica de variáveis binárias diz respeito à sua vantagem de uma maior facilidade de implementação sob o ponto de vista computacional.

4.3 Análise de variância

Na Tabela 14, estão apresentadas as análises de variâncias, seguindo o modelo de blocos casualizados descrito anteriormente, para os quatro grupos estudados.

Analisando o resultado da Tabela 14 referente à análise de variância dos volumes preditos pelas equações dos estratos de cada amostra, verifica-se que não há diferença significativa entre os tratamentos, isto é, as equações podem ser consideradas pertencentes ao mesmo grupo dentro de cada amostra.

A conclusão obtida para a amostra 1 não está de acordo com o que foi obtido anteriormente no teste de identidade de modelo e utilizando variáveis binárias.

Esta alternativa de teste para comparar equações foi utilizada também por Paula Neto et al. (1983), Herrera (1989), Leite e Regazzi (1992) e Silva et al. (1997). Leite e Regazzi (1992), Scolforo, Mello e Lima. (1994) e Silva et al. (1997) comentaram sobre a dificuldade de interpretação objetiva se existem muitos tratamentos a serem comparados e por este método utilizar apenas as médias nas comparações, não possibilitando inferências a respeito da semelhança das estimativas ao longo das classes de diâmetro.

TABELA 14 - Análises de Variância dos volumes estimados pelas equações dos estratos referentes a cada amostra

CV	GL	SQ	QM	F
Amostra 1: solo				
Tratamento (equação)	4	0,3351	0,0838	1,09 ^{ns}
Classe de diâmetro	5	319,4153	63,8831	**
Resíduo	290	22,2387	0,0767	
Amostra 2: solo				
Tratamento (equação)	4	0,0243	0,0061	< 1 ^{ns}
Classe de diâmetro	4	175,3250	43,8312	**
Resíduo	241	10,9424	0,0454	
Amostra 3: idade				
Tratamento (equação)	3	0,0104	0,0035	< 1 ^{ns}
Classe de diâmetro	4	79,4126	19,8531	**
Resíduo	112	4,5670	0,0408	
Amostra 4: espécie				
Tratamento (equação)	3	0,0205	0,0068	< 1 ^{ns}
Classe de diâmetro	5	136,3882	27,2776	**
Resíduo	135	5,9795	0,0443	

4.4 Análise de agrupamento

As figuras 3, 4, 5 e 6 apresentadas a seguir mostram os resultados da análise de agrupamento. Como ilustração, as matrizes das distâncias de Mahalanobis para cada amostra estão apresentadas no Anexo B.

Uma interpretação visual do agrupamento para a amostra I pode ser feita observando-se o dendograma apresentado na Figura 3.

Para a amostra I (Figura 3), a formação de um único grupo ocorreu a uma distância de Mahalanobis máxima de 1,3. Este resultado mostra que as distâncias entre os tratamentos são pequenas, evidenciando tratamentos semelhantes. Porém um pouco menos para o tratamento 4 referente ao estrato 48.

Estes resultados obtidos por este método para a amostra I indicam que não é necessário utilizar equações independentes por estrato.

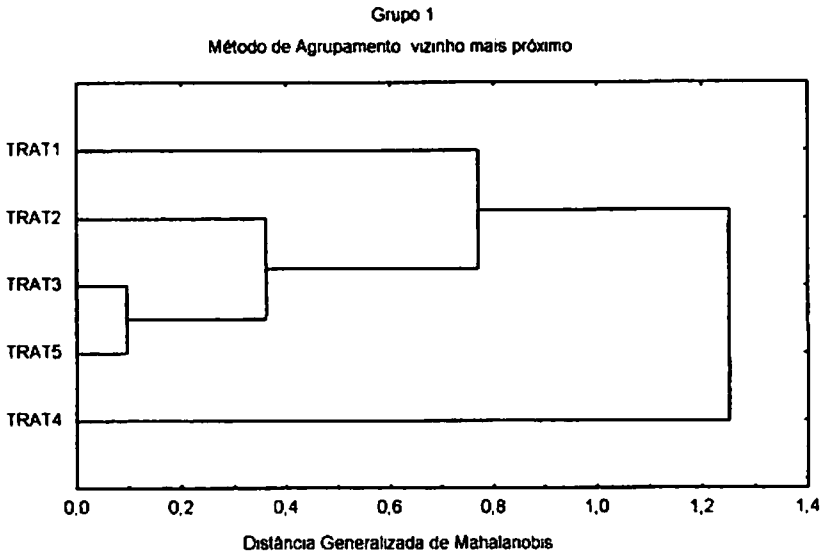


FIGURA 3 - Dendograma resultante da análise de agrupamento para grupo I.

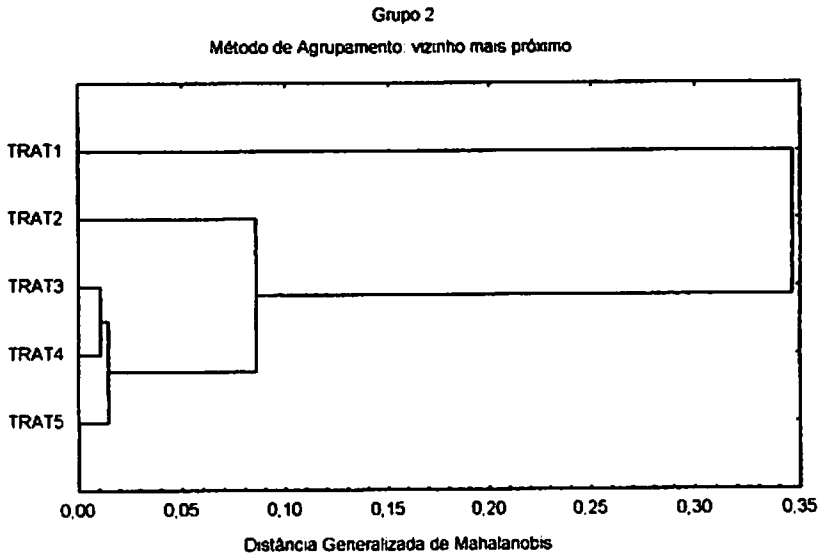


FIGURA 4 - Dendrograma resultante da análise de agrupamento para grupo 2.

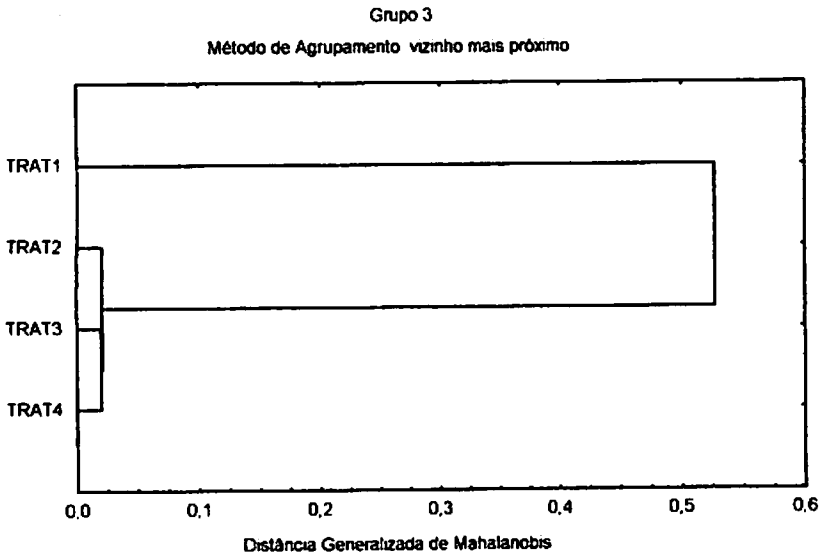


FIGURA 5 - Dendrograma resultante da análise de agrupamento para grupo 3.

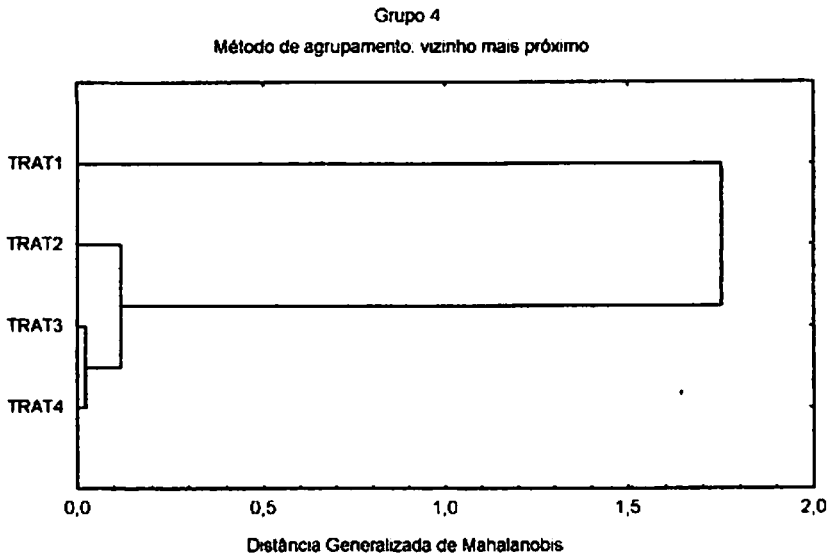


FIGURA 6 - Dendrograma resultante da análise de agrupamento para grupo 4.

Para os demais grupos, o tratamento 1 (volumes observados) está mais distante dos demais grupos (tratamentos) Isto pode estar evidenciando estimativas não tão confiáveis, porém observa-se que a distância em que apenas um grupo é formado não é tão grande.

4.5 Considerações Finais

O teste de identidade de modelos e o uso de variáveis binárias são os mais abordados na literatura, sugerindo uma maior confiabilidade sobre os outros. Porém é possível aplicá-los quando os modelos possuem relações funcionais iguais, o que nem sempre ocorre em situações reais.

Este trabalho teve como objetivo apresentar métodos para comparação entre equações de regressão encontrados em literaturas e uma aplicação. Nos resultados obtidos para a aplicação escolhida, observaram-se algumas discrepâncias entre os métodos aplicados. Algumas destas divergências podem

estar relacionadas com as particularidades de cada método e das propriedades das variáveis estudadas quanto às pressuposições necessárias a cada modelo. Os resultados sinalizam para a necessidade de comparações mais detalhadas entre as metodologias, com objetivo de padronizar as rotinas de testes e estimativas que são realizadas na prática.

Foi ressaltada, neste trabalho, uma aplicação na área florestal como uma evidência de praticidade e economia que os testes de comparações entre equações de volumes de madeira podem trazer. Estas vantagens podem advir, sobretudo, com a possibilidade de tratar desnecessariamente, com uma única equação, amostras, talhões, estratos ou outros fatores que eventualmente estejam sendo utilizados para estratificação.

Sugere-se, então, um estudo de simulação de dados, com possibilidade de geração de distribuições comportadas nas suas propriedades, para que se possa comparar estes métodos, assim como o estudo em outras áreas de aplicação em que possam ser aplicados teste de comparação de coeficientes de regressão.

5 CONCLUSÕES

1. Algumas considerações sobre os métodos:

- O teste de identidade de modelos e a técnica de variáveis binárias levam a resultados semelhantes, sendo que o último possui implementação computacional mais fácil.
- Havendo condições, escolhem-se os testes de identidade de modelos e a técnica de variáveis binárias.
- Comparações múltiplas e análise de agrupamento são alternativas quando os modelos comparados possuem relações funcionais diferentes, porém devem ser utilizados com muita cautela.

2. Os resultados dos métodos aplicados mostraram que:

- Verificou-se a necessidade de equações independentes por estrato no grupo I pelo método de identidade de modelos e variáveis binárias.
- Os métodos de análise de variância e análise de agrupamento forneceram a possibilidade de uso de uma equação comum no grupo I.
- As equações podem ser consideradas iguais nos grupos 2, 3 e 4, pelos diferentes métodos aplicados.

REFERÊNCIAS BIBLIOGRÁFICAS

- ASSIS, A.L. **Acuracidade na estimativa de volumes comerciais de *Eucalyptus grandis* e *Eucalyptus urophylla***. Lavras: UFLA, 1998. 55p. Monografia.
- BANZATTO, D.A.; KRONKA, S.N. **Experimentação agrícola**. Jaboticabal: Funep, 1995. 247p.
- BIOESTAT 2.0: aplicação estatísticas nas áreas das ciências biológicas e médica. Sociedade Civil Mamirauá; Brasília: CNPq, 2000.
- BROWN, B.W. Simple comparisons of simultaneous regression lines. **Biometrics**, Washington, v.26, n.1, p.143-144, Mar. 1970.
- BUSSAB, W.O.; MIAZAKI, E.S.; ANDRADE, D.F. **Introdução à Análise de Agrupamentos**. In: SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA.1990. 105p.
- CHOW, G.C. Tests of equality between sets of coefficients in two linear regressions. **Econometrica**, Chicago, v.28, p.591-605, 1960.
- CRUZ, C.D. **Aplicação de algumas técnicas multivariadas no melhoramento de plantas**. Piracicaba: ESALQ, 1990. 188p. (Tese - Doutorado em Estatística e Experimentação Agronômica).
- DRAPER, N.R.; SMITH, H. **Applied regression analysis**. New York: John Wiley & Sons, 1981. 709p.
- DUNCAN, D.B. Multiple comparison methods for comparing regression coefficients. **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. 704 p.

- 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.** Viçosa: UFV, 1989. 113p. (Dissertação – Mestrado em Ciência Florestal).
- HINKELMAN, K.; KEMPTHORNE, O. **Design and analysis of experiments.** New York: Chichester, 1994. 495p.
- HOFFMANN, R.; VIEIRA S. **Análise de regressão: uma introdução à econometria.** São Paulo: HUCITEC, 1998. 379p.
- X LEITE, H.G.; REGAZZI, A.J. Métodos estatísticos para avaliar a igualdade de equações volumétricas. *Revista Árvore*, Viçosa, v.16, n1, p.59-71, jan./abr. 1992.
- LOETSCH, F.; ZÖHRER, F.; HALLER, K. E. **Forest Inventory.** München: BLV Verlagsgesellschaft, 1964. v.1, 436p.
- LOPES, C.R.G. **Produção de madeira e carvão em plantações de *Eucalyptus camaldulensis* e *Eucalyptus cloeziana*.** Viçosa: UFV, 1996. 78p. (Dissertação – Mestrado em Ciência Florestal).
- MARDIA, K.V.; KENT, J.T.; BIBBY, J.M. **Multivariate analysis.** London: Academic Press, 1979. 518p.
- 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.
- NETER, J.; WASSERMAN, W.; KUTNER, M. **Applied linear statistical models.** 3th 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.

PIMENTEL GOMES, F. *Curso de estatística experimental*. 14.ed. Piracicaba: Nobel, 2000. 467p.

- ✕ 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/STAT[®]: user's guide: Noth Carolina, 1999.

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. ; FIGUEIREDO FILHO, A. *Medição e Volumetria de Árvores Florestais*. Lavras: FAEPE, 1997. 198 p.

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

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

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.

SNEDECOR, G.W.; COCHRAN, W.G. *Statistical methods*. Iowa: The Iowa State University Press, 1980. 507 p.

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

STATISTICA for Windows. Release 5. Copyright Stat Soft. 1984-1995.

STEEL, R.G.D.; TORRIE, J.H.; DICKEY, D.A. Principles and procedures of statistics a biometrical approach. New York: McGraw-Hill, 1997. 666p.

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.

WEISBERG, S. Applied linear regression. New York: John Wiley, 1985. 324p.

ANEXOS

ANEXO A

Página

Tabela 1A - Análise de Variância relativa ao teste de hipótese $H_0 : \beta_1 = \beta_2 = \beta_3$ (as equações dos estratos 6, 32 e 48 são idênticas).....	81
Tabela 2A - Análise de Variância relativa ao teste de hipótese $H_0 : \beta_1 = \beta_2 = \beta_3$ (as equações dos estratos 50, 69 e 121 são idênticas).....	81
Tabela 3A - Análise de Variância relativa ao teste de hipótese $H_0 : \beta_1 = \beta_2$ (as equações dos estratos 4 e 13 são idênticas).....	82
Tabela 4A - Análise de Variância relativa ao teste de hipótese $H_0 : \beta_1 = \beta_2$ (as equações dos estratos 47 e 50 são idênticas).....	82

ANEXO B

Página

Anexo 1B - Matriz de Distância de Mahalanobis para a amostra 1.....	83
Anexo 2B - Matriz de Distância de Mahalanobis para a amostra 2.....	83
Anexo 3B - Matriz de Distância de Mahalanobis para a amostra 3.....	84
Anexo 4B - Matriz de Distância de Mahalanobis para a amostra 4.....	84

ANEXO C

Página

Anexo C1 - Estrutura do Programa SAS para o teste de identidade de modelos e variáveis binárias.....	85
--	----

ANEXO A

Tabela 1A - Análise de Variância relativa ao teste de hipótese $H_0 : \beta_1 = \beta_2 = \beta_3$ (as equações dos estratos 6, 32 e 48 são idênticas).

CV	GL	SQ	QM	F	p
Modelo Completo (β)	9	736,8515			
M. Reduzido (θ)	3	736,5816			
Redução (H_0)	6	0,2699	0,0450	6,94	0,0001
Resíduo	165	1,0693	0,0065		
Total	174	737,9208			

Tabela 2A - Análise de Variância relativa ao teste de hipótese $H_0 : \beta_1 = \beta_2 = \beta_3$ (as equações dos estratos 50, 69 e 121 são idênticas).

CV	GL	SQ	QM	F	P
Modelo Completo (β)	9	1225,2228			
M. Reduzido (θ)	3	1225,1872			
Redução (H_0)	6	0,0356	0,0059	1,03	0,4078
Resíduo	159	0,9166	0,0058		
Total	168	1226,1394			

Tabela 3A - Análise de Variância relativa ao teste de hipótese $H_0 : \beta_1 = \beta_2$ (as equações dos estratos 4 e 13 são idênticas).

CV	GL	SQ	QM	F	P
Modelo Completo (β)	6	668,5129			
M. Reduzido (θ)	3	668,4885			
Redução (H_0)	3	0,0244	0,0081	1,64	0,1844
Resíduo	109	0,5402	0,0050		
Total	115	669,0531			

Tabela 4A - Análise de Variância relativa ao teste de hipótese $H_0 : \beta_1 = \beta_2$ (as equações dos estratos 47 e 50 são idênticas).

CV	GL	SQ	QM	F	P
Modelo Completo (β)	6	698,4013			
M. Reduzido (θ)	3	698,3768			
Redução (H_0)	3	0,0245	0,0082	1,60	0,1939
Resíduo	105	0,5348	0,0051		
Total	111	698,9360			

ANEXO B

Anexo 1B – Matriz de Distância de Mahalanobis para a amostra 1

$$D_1 = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} & \begin{bmatrix} 0 & 1,1977 & 0,9049 & 3,6706 & 0,7705 \\ 1,1977 & 0 & 0,7082 & 1,2518 & 0,3611 \\ 0,9049 & 0,7082 & 0 & 2,1363 & 0,0966 \\ 3,6706 & 1,2518 & 2,1363 & 0 & 1,8304 \\ 0,7705 & 0,3611 & 0,0966 & 1,8304 & 0 \end{bmatrix} \end{matrix}$$

Anexo 2B – Matriz de Distância de Mahalanobis para a amostra 2

$$D_2 = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} & \begin{bmatrix} 0 & 0,3737 & 0,3706 & 0,4863 & 0,3462 \\ 0,3737 & 0 & 0,1698 & 0,2216 & 0,0857 \\ 0,3706 & 0,1698 & 0 & 0,0108 & 0,0144 \\ 0,4863 & 0,2216 & 0,0108 & 0 & 0,0356 \\ 0,3462 & 0,0857 & 0,0144 & 0,0356 & 0 \end{bmatrix} \end{matrix}$$

Anexo 3B – Matriz de Distância de Mahalanobis para a amostra 3

$$D_3 = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{bmatrix} 0 & 0,762 & 0,527 & 0,6947 \\ 0,762 & 0 & 0,0606 & 0,0208 \\ 0,527 & 0,0606 & 0 & 0,0196 \\ 0,6947 & 0,0208 & 0,0196 & 0 \end{bmatrix} \end{matrix}$$

Anexo 4B – Matriz de Distância de Mahalanobis para a amostra 4

$$D_4 = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{bmatrix} 0 & 1,7516 & 2,1498 & 2,1122 \\ 1,7516 & 0 & 0,1171 & 0,1401 \\ 2,1498 & 0,1171 & 0 & 0,0237 \\ 2,1122 & 0,1401 & 0,0237 & 0 \end{bmatrix} \end{matrix}$$

ANEXO C

Anexo C1 – Estrutura do Programa SAS para o teste de identidade de modelos e variáveis binárias

```
/* Dissertacao - Dados Cubagem Rigorosa */  
/* 01 de dezembro de 2000 */  
/* Iara Endruweit Battisti e Ruben Delly Veiga */
```

```
options nodate nonumber;
```

```
data dados;
```

```
input estrato idade espca especie solo parcela arv dap class alt volt lndap lnalt lnvolt z1  
z2;
```

```
w1=z1*lndap;
```

```
w2=z2*lndap;
```

```
u1=z1*lnalt;
```

```
u2=z2*lnalt;
```

```
v1=z1+z2;
```

```
datalines;
```

```
.  
.  
.  
.
```

```
/* variaveis dummy */
```

```
proc reg;
```

```
model lnvolt = z1 lndap w1 lnalt u1;
```

```
output out=ccc p=pred r=res;
```

```
tiiguais: test z1=w1=u1=0; /* teste equacoes iguais */
```

```
tb01b03: test z1=0;
```

```
tb11b13: test w1=0;
```

```
tb21b23: test u1=0;
```

```
run;
```

```
/* identidade de modelos */
```

```
proc iml;
```

```
use dados;
```

```
read all var{'z1' 'z2' 'w1' 'w2' 'u1' 'u2'} into x;
```

```
read all var{'v1' 'lndap' 'lnalt' } into xr; /* teste equacoes iguais */
```

```
read all var{'lnvolt'} into y;
```

```
n=nrow(x);
p=ncol(x);
p1=ncol(xr);
gle=n-p;
glred=p-p1;
xx=x`*`x;
xxr=xr`*`xr;
c=inv(xx);
cr=inv(xxr);
b=c*x`*`y;
br=cr*xr`*`y;
sqt=y`*`y;
sqm=b`*`x`*`y;
sqmr=br`*`xr`*`y;
sqe=sqt-sqm;
qme=sqe/gle;
cd=sqm/sqt;
sqred=sqm-sqmr;
qmred=sqred/glred;
fred=qmred/qme;
print b br cd;
print sqm sqmr sqt;
print gle sqe qme;
print glred sqred qmred fred;
quit;
```

