

TESTE PARA VERIFICAR A IDENTIDADE DE MODELOS DE REGRESSÃO¹

ADAIR JOSÉ REGAZZI²

RESUMO - Neste trabalho foi considerado o ajustamento de H equações de regressão no caso de justaposição de r submodelos polinomiais de grau k , em que os pontos de interseção dos submodelos foram considerados conhecidos. Restrições foram impostas para que os submodelos polinomiais fossem concordantes nos pontos de interseção. O modelo linear para a h -ésima equação é $\underline{Y}_h = X_h \underline{\beta}_h + \underline{\varepsilon}_h$, $h = 1, 2, \dots, H$, em que \underline{Y}_h é um vetor $n_h \times 1$ de realizações de variáveis aleatórias, X_h uma matriz $n_h \times p$ de constantes conhecidas, $\underline{\beta}_h$ um vetor $p \times 1$ de parâmetros desconhecidos e $\underline{\varepsilon}_h$ um vetor $n_h \times 1$ de erros aleatórios suposto NID ($\underline{\varepsilon}_h: \phi, \sigma^2 I$). Na estimação dos parâmetros, utilizou-se o método dos mínimos quadrados. Derivou-se um teste estatístico para a hipótese de que sejam idênticos H modelos de regressão no caso de justaposição de r submodelos polinomiais de grau k . Assim, a hipótese considerada foi: $H_0: \underline{\beta}_1 = \underline{\beta}_2 = \dots = \underline{\beta}_H$ (os H modelos são idênticos) vs. $H_a: \underline{\beta}_i \neq \underline{\beta}_j$, para pelo menos um $i \neq j$ (os H modelos não são todos idênticos). Como ilustração, esse método foi aplicado a um conjunto de $H =$ duas equações de regressão, no caso de justaposição de $r =$ dois submodelos polinomiais do primeiro grau.

Termos para indexação: identidade de modelos, mínimos quadrados, estatística aplicada.

¹ Aceito para publicação em 2 de outubro de 1995.

² Eng. Agr., Dr., Prof. Titular do Dep. de Informática, Universidade Federal de Viçosa, CEP 36571-000 Viçosa, MG. Pesquisador do CNPq.

IDENTITY TEST FOR REGRESSION MODELS

ABSTRACT - It is considered in this paper the adjustment of H regression equations in the case of juxtaposition of r polynomial submodels of k degree. The points of intersection of the submodels are supposed to be known. Appropriate restrictions are imposed in such a way that the polynomial submodels are concordant in the points of intersection. The linear model for the h^{th} equation is $\underline{Y}_h = X_h \underline{\beta}_h + \underline{\varepsilon}_h$, $h = 1, 2, \dots, H$, where \underline{Y}_h is an $n_h \times 1$ vector of observations, X_h is an $n_h \times p$ matrix of known constants, $\underline{\beta}_h$ is an $p \times 1$ vector of unknown parameters and $\underline{\varepsilon}_h$ is an $n_h \times 1$ vector of errors that is distributed NID ($\underline{\varepsilon}_h : \phi, \sigma^2 I$). In the parameters estimation, the Least Square Method was used. A statistic test was derived for the hypothesis that H regression models in the case of juxtaposition of r polynomial submodels of k degree were identical. The hypothesis in consideration is: $H_o: \underline{\beta}_1 = \underline{\beta}_2 = \dots = \underline{\beta}_H$ (H models are identical) vs. $H_a: \underline{\beta}_i \neq \underline{\beta}_j$ for at least one $i \neq j$ (the H models are not all identical). This method is applied to a set of $H =$ two regression equations in the case of juxtaposition of $r =$ two polynomial submodels of first degree.

Index terms: identity of models, least square, applied statistics.

INTRODUÇÃO

Com muita freqüência, estuda-se a relação funcional entre as variáveis Y e X . Um problema que tem aplicação importante é determinar se um conjunto de curvas é idêntico. Por exemplo, ao estudar H diferentes situações experimentais e admitir um modelo linear para cada situação, um pesquisador pode estar interessado em determinar se os H modelos são idênticos. Nesse caso, o interesse está em saber se um conjunto de equações pode ser representado por uma equação comum.

Neter & Wasserman (1974) testaram se duas equações de regressão linear simples eram idênticas, utilizando o teste F . Graybill (1976)

apresentou um método geral para testar a hipótese de igualdade de um conjunto de modelos lineares, empregando o teste F. Regazzi (1993) considerou o ajustamento de H equações de regressão polinomial de grau k, mediante o emprego da técnica dos polinômios ortogonais, em que apresentou, em detalhes, um método para testar as seguintes hipóteses: a) H_0 : as H equações são idênticas; b) H_a : as H equações têm uma constante de regressão comum; c) H_0 : as H equações têm um ou mais coeficientes de regressão iguais. Ele concluiu que o método apresentado é geral e pode ser usado em modelos polinomiais de qualquer grau, ortogonais ou não, e também em modelos de regressão múltipla.

Ao analisar um conjunto de dados de uma variável (Y) qualquer, pode haver uma tendência distinta para cada subconjunto de observações, e não ser satisfatório o ajustamento de um único modelo polinomial. Nesse caso, poderá ser melhor, de acordo com a tendência das observações, ajustar dois ou mais segmentos de polinômios de mesmo grau. O ajustamento de dois ou mais segmentos de polinômios é o que será denominado de submodelos polinomiais. Fuller (1969) e Nogueira (1980) discutiram um método para o ajustamento de dados de observação através da justaposição de funções quadráticas.

Admitindo-se o caso de H modelos de regressão, cada um com r submodelos polinomiais de mesmo grau, pode haver interesse em verificar a identidade dos H modelos, isto é, se o conjunto de H equações pode ser representado por uma única equação comum. Assim, o objetivo do presente trabalho foi apresentar um método geral para o teste de identidade de H modelos de regressão no caso de justaposição de r submodelos polinomiais de grau k, em que os pontos de interseção dos submodelos foram considerados conhecidos.

METODOLOGIA E RESULTADOS

Modelo estatístico completo

Considere-se, inicialmente, o ajustamento de H modelos de regressão no caso de justaposição de $r = 2$ submodelos polinomiais do primeiro grau. As H equações são dadas por:

$$\begin{aligned}
 Y_{1i} &= a_{01} + a_{11}X_{1i} + \lambda_1(X_{1i} - K)Z_i + e_{1i}, & i = 1, 2, \dots, n_1 \\
 Y_{2i} &= a_{02} + a_{12}X_{2i} + \lambda_2(X_{2i} - K)Z_i + e_{2i}, & i = 1, 2, \dots, n_2 \\
 &\vdots & \vdots \\
 &\vdots & \vdots \\
 Y_{Hi} &= a_{0H} + a_{1H}X_{Hi} + \lambda_H(X_{Hi} - K)Z_i + e_{Hi}, & i = 1, 2, \dots, n_H
 \end{aligned}
 \tag{\alpha.1}$$

em que:

- Y_{hi} é a 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} é o i -ésimo valor da variável independente do h -ésimo modelo;
 a_{oh} , a_{1h} e λ_h são os parâmetros do h -ésimo modelo;
 K é o ponto de interseção dos dois submodelos, considerado conhecido e igual para as H equações;
 Z_i é uma variável binária, tal que $Z_i = 0$ para $X \leq K$ e $Z_i = 1$ para $X > K$;
 e_{hi} é o erro aleatório, associado à i -ésima observação do h -ésimo modelo, com valores supostos independentes e normalmente distribuídos com média zero e variância comum σ^2 , isto é, $e_{hi} \sim NID(0, \sigma^2)$;

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

O h -ésimo modelo ajustado pode ser escrito como:

$$\hat{Y}_{hi} = \hat{a}_{oh} + \hat{a}_{1h} X_{hi} + \hat{\lambda}_h (X_{hi} - K) Z_i$$

Assim, para $Z_i = 0$ ($X \leq K$), temos o primeiro submodelo

$$\hat{Y}_{hi} = \hat{a}_{oh} + \hat{a}_{1h} X_{hi}, \quad (\alpha.2)$$

e para $Z_i = 1$ ($X > K$) temos o segundo submodelo

$$\hat{Y}_{hi} = (\hat{a}_{oh} - \hat{\lambda}_h K) + (\hat{a}_{1h} + \hat{\lambda}_h) X_{hi} \quad (\alpha.3)$$

Sejam ($\alpha.2$) e ($\alpha.3$) escritos simplesmente como

$$\hat{Y} = \hat{a}_0 + \hat{a}_1 X, \quad X \leq K$$

e

$$\hat{Y} = \hat{b}_0 + \hat{b}_1 X, \quad X > K, \quad \text{respectivamente.}$$

A restrição para que a justaposição dos dois submodelos seja concordante no ponto de interseção K é dada por:

$$\hat{a}_0 + \hat{a}_1 K = \hat{b}_0 + \hat{b}_1 K.$$

Esta restrição é linear nos parâmetros e reduz o número de parâmetros independentes no modelo de quatro para três. Pode-se obter diretamente as estimativas \hat{a}_0 , \hat{a}_1 , $\hat{\lambda} = \hat{b}_1 - \hat{a}_1$.

Os \hat{b}_1 's são então definidos, em termos dessas três estimativas, por:

$$\hat{b}_0 = \hat{a}_0 - K(\hat{b}_1 - \hat{a}_1) = \hat{a}_0 - \hat{\lambda}K$$

$$\hat{b}_1 = \hat{a}_1 + (\hat{b}_1 - \hat{a}_1) = \hat{a}_1 + \hat{\lambda}$$

As estimativas $\hat{a}_0, \hat{a}_1, \hat{\lambda} = \hat{b}_1 - \hat{a}_1$ são obtidas da equação de regressão

$$\hat{Y} = \hat{a}_0 + \hat{a}_1 X + (\hat{b}_1 - \hat{a}_1) Z$$

em que:

$$\begin{aligned} Z &= 0, & \text{para } X \leq K \\ Z &= (X - K), & \text{para } X > K \end{aligned}$$

A Fig. 1 ilustra o caso dos modelos apresentados em (α.1). Para cada um dos H modelos, a justaposição dos dois submodelos é concordante no ponto de interseção K, isto é, eles se justapõem sem ocorrer saltos.

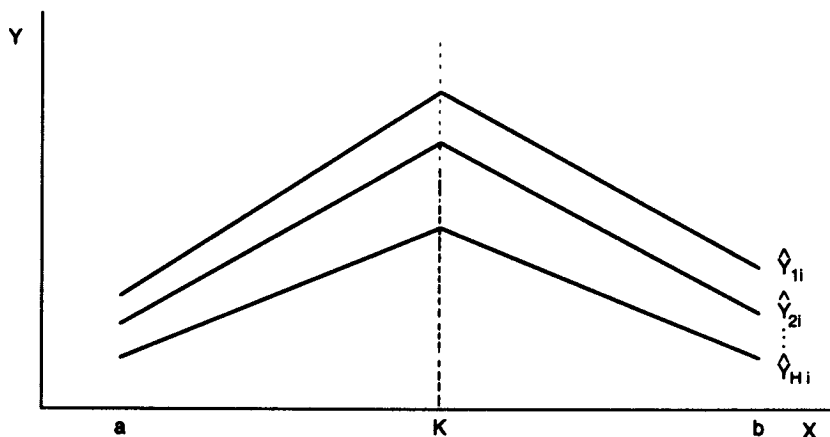


FIG. 1. H modelos ajustados com a justaposição de dois submodelos de regressão linear simples.

Seja agora o ajustamento de um h-ésimo modelo particular, que contém uma seqüência de r submodelos polinomiais de grau k.

Seja o modelo do tipo

$$Y_i = f(X_i) + e_i,$$

em que $f(X_i)$ vem a ser uma seqüência de submodelos do tipo

$$\begin{aligned} f(X_i) &= f_1(X, \underline{\beta}_1), & \text{para } a \leq X \leq X_1 \\ &= f_2(X, \underline{\beta}_2), & \text{para } X_1 < X \leq X_2 \\ &\vdots \\ &= f_r(X, \underline{\beta}_r), & \text{para } X_{r-1} < X \leq b, \end{aligned}$$

e r o número de submodelos e X_1, X_2, \dots, X_{r-1} vem a ser os pontos de interseção dos submodelos, considerados conhecidos.

Tem-se as seguintes restrições:

$$\begin{aligned} \text{i)} \quad & f_t(X_1, \beta_t) = \sum_{j=0}^k \beta_{jt} X^j, \quad t = 1, \dots, r \quad j = 1, \dots, k \quad (\text{grau do} \\ & \text{polinômio}) \\ \text{ii)} \quad & f_t(X_c, \beta_t) = f_{t+1}(X_c, \beta_{t+1}), \quad c = 1, \dots, (r-1) \quad (\alpha.4) \\ \text{iii)} \quad & \frac{d[f_t(X_c, \beta_t)]}{dX} = \frac{d[f_{t+1}(X_c, \beta_{t+1})]}{dX}, \quad k \geq 2 \end{aligned}$$

Os vetores dos parâmetros β_t , ($t = 1, 2, \dots, r$) são restritos, de modo que as curvas e as derivadas primeiras sejam contínuas nos pontos especificados X_c ($c = 1, 2, \dots, (r-1)$). Assim, a justaposição dos submodelos polinomiais serão concordantes nos pontos de interseção e irão se justapor sem formar ângulos, suavemente.

Estas restrições são lineares nos parâmetros e reduzem o número de parâmetros independentes no modelo, os quais são facilmente estimados pelo método dos mínimos quadrados no caso de regressão linear múltipla.

Conforme apresentado por Fuller (1969) e Nogueira (1980), no ajustamento de dados de observação através da justaposição de duas funções polinomiais quadráticas, tem-se:

$$f_1(X) = a_0 + a_1 X + a_2 X^2, \quad \text{para } X \leq K$$

$$f_2(X) = b_0 + b_1 X + b_2 X^2, \quad \text{para } X \geq K$$

As restrições reduzem o número de parâmetros independentes no modelo de seis para quatro: $a_0, a_1, a_2, \lambda = b_2 - a_2$. Os b_i s são então definidos, em termos desses parâmetros, por:

$$b_0 = a_0 + K^2(b_2 - a_2) = a_0 + K^2 \lambda$$

$$b_1 = a_1 - 2K(b_2 - a_2) = a_1 - 2K \lambda$$

$$b_2 = a_2 + (b_2 - a_2) = a_2 + \lambda$$

Estimativas de a_0, a_1, a_2 e $\lambda = (b_2 - a_2)$ podem ser obtidas da equação de regressão

$$Y_i = a_0 + a_1 X_i + a_2 X_i^2 + \lambda (X_i - K)^2 Z_i + e_i,$$

em que:

$$Z_i = 0 \quad \text{para } X \leq K \quad \text{e} \quad Z_i = 1 \quad \text{para } X > K.$$

Admitindo-se então $r = 2$ submodelos polinomiais do segundo grau, os H modelos de regressão podem ser postos da seguinte maneira:

$$\begin{aligned}
 Y_{1i} &= a_{01} + a_{11} X_{1i} + a_{21} X_{1i}^2 + \lambda_1 (X_{1i} - K)^2 Z_i + e_{1i}, & i = 1, 2, \dots, n_1 \\
 Y_{2i} &= a_{02} + a_{12} X_{2i} + a_{22} X_{2i}^2 + \lambda_2 (X_{2i} - K)^2 Z_i + e_{2i}, & i = 1, 2, \dots, n_2 \\
 &\vdots & \vdots \\
 Y_{Hi} &= a_{0H} + a_{1H} X_{Hi} + a_{2H} X_{Hi}^2 + \lambda_H (X_{Hi} - K)^2 Z_i + e_{Hi}, & i = 1, 2, \dots, n_H,
 \end{aligned}
 \tag{\alpha.5}$$

em que:

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

$$Z_i = 1, \text{ para } X > K.$$

A Fig. 2 ilustra o caso dos modelos apresentados em ($\alpha.5$).

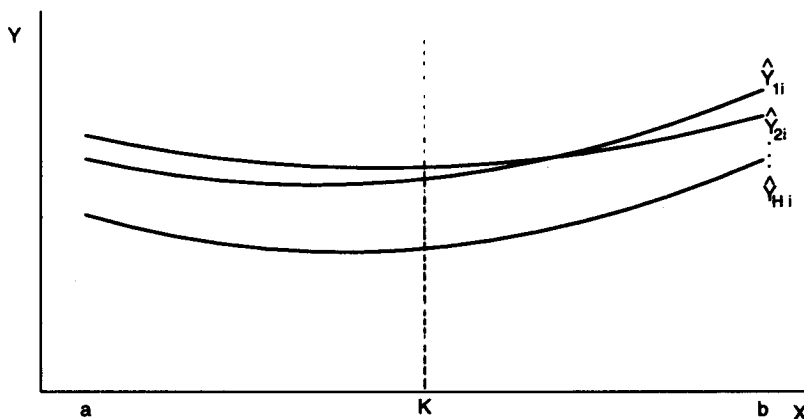


FIG. 2. H modelos ajustados com a justaposição de dois submodelos polinomiais do segundo grau.

No caso de três funções polinomiais quadráticas, para $a \leq X \leq b$, tem-se:

$$f_1(X) = a_0 + a_1 X + a_2 X^2, \text{ para } a \leq X \leq K_1$$

$$f_2(X) = b_0 + b_1 X + b_2 X^2, \text{ para } K_1 < X \leq K_2$$

$$f_3(X) = c_0 + c_1 X + c_2 X^2, \text{ para } K_2 < X \leq b$$

As restrições ($\alpha.4$) reduzem o número de parâmetros independentes no modelo de nove para cinco:

$$a_0, a_1, a_2, \lambda_1 = (b_2 - a_2) \text{ e } \lambda_2 = (c_2 - b_2).$$

Os demais parâmetros são obtidos, conforme demonstrado por Nogueira (1980), e são dados por:

$$b_0 = a_0 + K_1^2 \lambda_1$$

$$b_1 = a_1 - 2K_1 \lambda_1$$

$$b_2 = a_2 + \lambda_1$$

$$c_0 = a_0 + K_1^2 \lambda_1 + K_2^2 \lambda_2$$

$$c_1 = a_1 - 2(K_2 \lambda_2 + K_1 \lambda_1)$$

$$c_2 = b_2 + \lambda_2$$

Estimativas de $a_0, a_1, a_2, \lambda_1 = (b_2 - a_2)$ e $\lambda_2 = (c_2 - b_2)$ podem ser obtidas da equação de regressão

$$Y_i = a_0 + a_1 X_i + a_2 X_i^2 + \lambda_1 (X_i - K_1)^2 Z_{1i} + \lambda_2 (X_i - K_2)^2 Z_{2i} + e_i,$$

em que:

$$Z_{1i} = 0 \quad \text{para} \quad X \leq K_1$$

$$Z_{1i} = 1 \quad \text{para} \quad X > K_1$$

$$Z_{2i} = 0 \quad \text{para} \quad X \leq K_2$$

$$Z_{2i} = 1 \quad \text{para} \quad X > K_2$$

Admitindo-se então $r = 3$ submodelos polinomiais do segundo grau, os H modelos de regressão podem ser postos da seguinte maneira:

$$\begin{aligned} Y_{1i} &= a_{01} + a_{11} X_{1i} + a_{21} X_{1i}^2 + \lambda_{11} (X_{1i} - K_1)^2 Z_{1i} + \lambda_{21} (X_{1i} - K_2)^2 Z_{2i} + e_{2i}, & i = 1, 2, \dots, n_1 \\ Y_{2i} &= a_{02} + a_{12} X_{2i} + a_{22} X_{2i}^2 + \lambda_{12} (X_{2i} - K_1)^2 Z_{1i} + \lambda_{22} (X_{2i} - K_2)^2 Z_{2i} + e_{2i}, & i = 1, 2, \dots, n_2 \\ &\vdots & \vdots \\ Y_{Hi} &= a_{0H} + a_{1H} X_{Hi} + a_{2H} X_{Hi}^2 + \lambda_{1H} (X_{Hi} - K_1)^2 Z_{1i} + \lambda_{2H} (X_{Hi} - K_2)^2 Z_{2i} + e_{Hi}, & i = 1, 2, \dots, n_H \end{aligned} \quad (\alpha.6)$$

em que:

$$Z_{1i} = 0 \quad \text{para} \quad X \leq K_1$$

$$Z_{1i} = 1 \quad \text{para} \quad X > K_1$$

$$Z_{2i} = 0 \quad \text{para} \quad X \leq K_2$$

$$Z_{2i} = 1 \quad \text{para} \quad X > K_2$$

A Fig. 3 ilustra o caso dos modelos apresentados em $(\alpha.6)$.

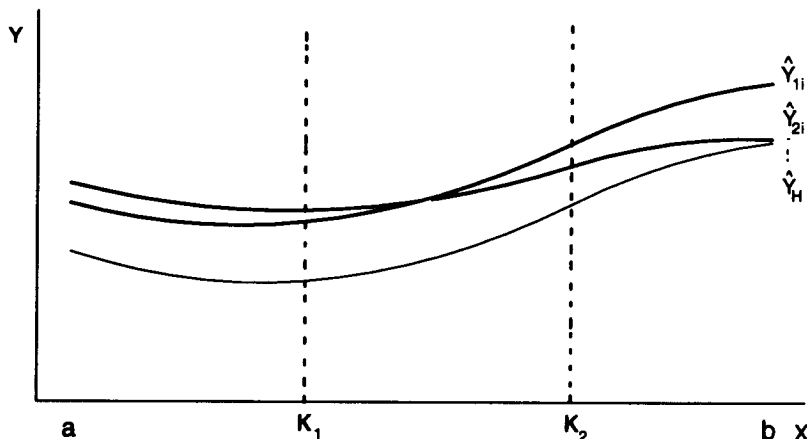


FIG. 3. H modelos ajustados com a justaposição de três submodelos polinomiais do segundo grau.

As considerações anteriores referem-se ao ajustamento dos modelos. A partir de agora será apresentado o teste para a identidade dos H modelos.

A hipótese é a seguinte:

$H_0: \beta_1 = \beta_2 = \dots = \beta_H$ (H modelos são idênticos) vs. $H_a: \beta_i \neq \beta_j$ para pelo menos um $i \neq j$ (os H modelos não são todos idênticos).

O h-ésimo modelo em (α.1), (α.5) ou (α.6) pode ser escrito como

$$Y_h = X_h \beta_h + \varepsilon_h, \quad h = 1, 2, \dots, H \tag{\alpha.7}$$

em que:

Y_h é um vetor $n_h \times 1$ de observações, X_h uma matriz $n_h \times p$ de constantes conhecidas, β_h um vetor $p \times 1$ de parâmetros desconhecidos e ε_h um vetor $n_h \times 1$ de erros aleatórios supostos $NID(\varepsilon_h: \phi, \sigma^2 I)$.

Escrevendo esses H modelos na forma do modelo linear geral $Y = X \beta + \varepsilon$, tem-se que:

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_H \end{bmatrix}, \quad X = \begin{bmatrix} X_1 & \phi & \dots & \phi \\ \phi & X_2 & \dots & \phi \\ \vdots & \vdots & \ddots & \vdots \\ \phi & \phi & \dots & X_H \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_H \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_H \end{bmatrix} \tag{\alpha.8}$$

$N \quad 1 \quad N \quad Hp \quad Hp \quad 1 \quad N \quad 1$

Uma vez que os $\underline{\varepsilon}_h$'s têm distribuição normal e são independentes, tem-se que $\underline{\varepsilon}$ tem distribuição $N(\underline{\varepsilon}; \phi, \sigma^2 I)$.

O sistema de equações normais, obtidos pelo método dos mínimos quadrados, é $X'X\hat{\underline{\beta}} = X'Y$, cuja solução $\hat{\underline{\beta}} = (X'X)^{-1}X'Y$ é o estimador do vetor de parâmetros $\underline{\beta}$.

Assim,

$$\hat{\underline{\beta}} = \begin{bmatrix} \hat{\underline{\beta}}_1 \\ \hat{\underline{\beta}}_2 \\ \cdot \\ \cdot \\ \cdot \\ \hat{\underline{\beta}}_H \end{bmatrix} = \begin{bmatrix} (X'_1 X_1)^{-1} & \phi & \dots & \phi \\ \phi & (X'_2 X_2)^{-1} & \dots & \phi \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \phi & \phi & \dots & (X'_H X_H)^{-1} \end{bmatrix} \begin{bmatrix} X'_1 Y_1 \\ X'_2 Y_2 \\ \cdot \\ \cdot \\ \cdot \\ X'_H Y_H \end{bmatrix}$$

Logo, $\hat{\underline{\beta}}_h = (X'_h X_h)^{-1} X'_h Y_h$, $h = 1, 2, \dots, H$. Assim $\hat{\underline{\beta}}_h$ pode ser obtido diretamente do h-ésimo modelo $Y_h = X_h \underline{\beta}_h + \varepsilon_h$.

Análise de variância relativa ao modelo completo

A soma de quadrados de parâmetros relativa ao modelo completo ($\alpha.8$) é dada por

$$SQPAR(c) = \hat{\underline{\beta}}' X' Y = \sum_{h=1}^H \hat{\underline{\beta}}'_h X'_h Y_h, \quad (\alpha.9)$$

com Hp graus de liberdade.

A soma de quadrados total não-corrigida é dada por

$$SQTOT(c) = Y' Y = \sum_{h=1}^H Y'_h Y_h \quad (\alpha.10)$$

com $N = \sum_{h=1}^H n_h$ graus de liberdade.

A soma de quadrados do resíduo é obtida pela diferença entre ($\alpha.10$) e ($\alpha.9$), em que:

$$SQRES(c) = Y' Y - \hat{\underline{\beta}}' X' Y$$

$$\begin{aligned}
 &= \sum_{h=1}^H (\underline{Y}'_h \underline{Y}_h - \hat{\underline{\beta}}'_h \underline{X}'_h \underline{Y}_h) \\
 &= \sum_{h=1}^H SQRES(h), \tag{\alpha.11}
 \end{aligned}$$

com $N - Hp$ graus de liberdade, e ainda

$SQRES(h) = \underline{Y}'_h \underline{Y}_h - \hat{\underline{\beta}}'_h \underline{X}'_h \underline{Y}_h$ com $n_h - p$ graus de liberdade é a soma de quadrados do resíduo relativa ao h-ésimo modelo, isto é, correspondente à análise de variância de regressão para a h-ésima equação.

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

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

Fonte de variação	GL	SQ	QM*
Parâmetro ($\underline{\beta}$)	Hp	$\hat{\underline{\beta}}' \underline{X}' \underline{Y}$	
Resíduo	$N - Hp$	$\underline{Y}' \underline{Y} - \hat{\underline{\beta}}' \underline{X}' \underline{Y}$	$QMR = \hat{\sigma}^2$
Total	N	$\underline{Y}' \underline{Y}$	

$$* \hat{\sigma}^2 = \frac{\underline{Y}' \underline{Y} - \hat{\underline{\beta}}' \underline{X}' \underline{Y}}{N - Hp} = \frac{\sum_{h=1}^H SQRES(h)}{N - Hp}$$

Na realidade, $\hat{\sigma}^2$ é o estimador comum da variância residual, que pode ser obtido pela média ponderada dos estimadores das variâncias residuais de cada modelo.

Uma vez que o modelo pressupõe homocedasticidade e a violação dessa pressuposição, em casos extremos, pode levar a erros graves nas conclusões, sua verificação poderá ser feita mediante um dos testes para homogeneidade de variâncias, como, por exemplo, os testes de Bartlett ou de Cochran, citados por Li (1964) e Dixon & Massey Junior (1969), respectivamente.

Com base nos trabalhos de Conagin et al. (1990) e Nagai et al. (1991), verifica-se que, dentro de certos limites, a heterogeneidade de variância não é assim um fator tão limitante. Por outro lado, conforme se verifica em Pimentel-Gomes (1990), quando se tem uma relação de variâncias menor que sete, é quase sempre possível combinar as variâncias residuais, obtendo-se uma estimativa comum. Esse fato pode ser estendido para o

problema em questão e, quando esse quociente for além de sete, convirá que se considerem separadamente subgrupos de modelos, em que se tenha, dentro de cada subgrupo, uma razoável homogeneidade de variâncias.

O teste estatístico

Sob a hipótese de nulidade,

$H_0: \beta_1 = \beta_2 = \dots = \beta_H$ (H modelos são idênticos), e tomando-se como exemplo os modelos apresentados em (α.6), estes reduzem-se à forma

$$Y_{hi} = a_0 + a_1 X_{hi} + a_2 X_{hi}^2 + \lambda_1 (X_{hi} - K_1)^2 Z_{1i} + \lambda_2 (X_{hi} - K_2)^2 Z_{2i} + e_{hi} \quad (\alpha.12)$$

em que a_0, a_1, a_2, λ_1 e λ_2 são os p parâmetros comuns, sendo, neste caso, $p = 5$.

Utilizando a notação matricial, os modelos reduzidos (α.12) podem ser escritos como

$$\underline{Y} = Z \underline{\theta} + \underline{\varepsilon} \quad (\alpha.13)$$

\underline{Y} : é o vetor dos valores observados da variável dependente, de dimensão $N \times 1$, igual a (α.8);

$$Z = \begin{bmatrix} X_1 \\ X_2 \\ \cdot \\ \cdot \\ \cdot \\ X_H \end{bmatrix}, \text{ onde } X_h \text{ com } h = 1, 2, \dots, H \text{ é igual a } (\alpha.7);$$

$N \quad p$

$$\underline{\theta} = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \lambda_1 \\ \lambda_2 \end{bmatrix}, \text{ é o vetor } p \times 1 \text{ de parâmetros comuns;}$$

$p \quad 1$

ε : é o vetor de erros aleatórios, de dimensão $N \times 1$ igual a $(\alpha.8)$.

O sistema de equações normais relativo ao modelo reduzido $(\alpha.13)$, obtido pelo método dos mínimos quadrados, é $Z'Z\hat{\theta} = Z'Y$, cuja solução $\hat{\theta} = (Z'Z)^{-1}Z'Y$ é o estimador do vetor de parâmetros comuns θ .

Assim,

$$\hat{\theta} = \left(\sum_{h=1}^H X_h' X_h \right)^{-1} \left(\sum_{j=1}^H X_j' Y_j \right) \quad (\alpha.14)$$

A soma de quadrados de parâmetros relativa ao modelo reduzido $(\alpha.13)$, é dada por:

$$\begin{aligned} SQPAR(r) &= \hat{\theta}' Z' Y \\ &= \left(\sum_{j=1}^H Y_j' X_j \right) \left(\sum_{h=1}^H X_h' X_h \right)^{-1} \left(\sum_{t=1}^H X_t' Y_t \right) \end{aligned} \quad (\alpha.15)$$

com p graus de liberdade.

A soma de quadrados total não-corrigida é dada igualmente por $(\alpha.10)$. Assim,

$$SQTOT(r) = Y'Y = \sum_{h=1}^H Y_h' Y_h \quad (\alpha.16)$$

com N graus de liberdade.

A soma de quadrados do resíduo relativa ao modelo reduzido é obtida pela diferença entre $(\alpha.16)$ e $(\alpha.15)$, em que:

$$\begin{aligned} SQRES(r) &= Y'Y - \hat{\theta}' Z' Y \\ &= \sum_{h=1}^H Y_h' Y_h - \left(\sum_{j=1}^H Y_j' X_j \right) \left(\sum_{h=1}^H X_h' X_h \right)^{-1} \left(\sum_{t=1}^H X_t' Y_t \right) \end{aligned} \quad (\alpha.17)$$

com $N - p$ graus de liberdade.

O teste estatístico para a hipótese $H_0: \beta_1 = \beta_2 = \dots = \beta_H$ (H modelos são idênticos) é baseado na diferença entre a soma de quadrados de parâmetros do modelo completo [SQPAR(c)] e do modelo reduzido [SQPAR(r)], ou seja, na redução que H_0 provoca na soma de quadrados de parâmetros do modelo completo.

Assim, a redução devida à hipótese H_0 , denotada por Redução (H_0) , é obtida pela diferença entre $(\alpha.9)$ e $(\alpha.15)$, ou seja:

$$\begin{aligned} \text{Redução } (H_0) &= SQPAR(c) - SQPAR(r) \\ &= \hat{\beta}' X' Y - \hat{\theta}' Z' Y \\ &= \sum_{h=1}^H \hat{\beta}_h' X_h' Y_h - \hat{\theta}' \sum_{t=1}^H X_t' Y_t \end{aligned} \quad (\alpha.18)$$

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

Para testar a hipótese

$H_0: \beta_1 = \beta_2 = \dots = \beta_H$ (H modelos são idênticos) vs. $H_0: \beta_i \neq \beta_j$ para pelo menos um $i \neq j$ (os H modelos não são todos idênticos), utiliza-se a estatística F , dada por:

$$F_0 = \frac{[SQPAR(c) - SQPAR(r)] / (H - 1)p}{SQRES(c) / (N - Hp)} \quad (\alpha.19)$$

De acordo com Graybill (1976), sob a hipótese de nulidade $H_0: \beta_1 = \beta_2 = \dots = \beta_H$, a estatística $(\alpha.19)$ apresenta distribuição F central, com $(H - 1)p$ e $(N - Hp)$ graus de liberdade.

O teste em questão pode ser facilmente visualizado a partir do esquema da análise de variância apresentado na Tabela 2.

TABELA 2. Esquema da análise de variância relativa ao teste da hipótese $H_0: \beta_1 = \beta_2 = \dots = \beta_H$ (H modelos são idênticos).

Fonte de variação	GL	SQ	QM	F_0
Parâmetro (β)	Hp	$Q_1 = \hat{\beta}' X' Y$		
Parâmetro (θ)	p	$Q_2 = \hat{\theta}' \sum_{i=1}^H X_i' Y_i$		
Redução (H_0)	$(H - 1)p$	$Q_3 = Q_1 - Q_2$	$V_1 = \frac{Q_3}{(H - 1)p}$	V_1/V_2
Resíduo	$N - Hp$	$Q_4 = Q_5 - Q_1$	$V_2 = \frac{Q_4}{N - Hp}$	
Total	N	$Q_5 = Y' Y$		

Assim, rejeita-se H_0 somente se

$$F_0 \geq F_{\alpha; (H-1)p; N-Hp}, \text{ onde } N = \sum_{h=1}^H n_h.$$

A não-rejeição da hipótese H_0 permite concluir que, a uma significância α , os H modelos ou as H equações não diferem significativamente entre si. Assim a equação ajustada com as estimativas dos parâmetros comuns pode ser usada como uma estimativa das H equações envolvidas.

ILUSTRAÇÃO DO MÉTODO

Julgou-se adequada a apresentação de um exemplo numérico para ilustrar os resultados obtidos neste estudo. Assim, com base nos dados da Tabela 3, foram efetuados os cálculos, ilustrando o método apresentado. Os dados, fornecidos pelo Prof. Antônio Bento Mancio, são provenientes de uma pesquisa realizada no Departamento de Zootecnia da Universidade Federal de Viçosa, MG.

TABELA 3. Concentrações médias de progesterona (ng/ml) no soro de novilhas submetidas a três dietas (1-controle, 2-ácido graxo protegido e 3-grão de soja integral), sem e com aplicação de hCG nas amostras nos dias de coleta 6, 9, 12, 15 e 18 após o estro.

Dias após o estro (X)	Progesterona (ng/ml)					
	Dieta 1		Dieta 2		Dieta 3	
	Sem hCG*	Com hCG	Sem hCG	Com hCG	Sem hCG	Com hCG
6	1,57	1,86	1,99	2,55	2,32	3,29
9	3,02	4,42	3,82	5,07	3,81	6,09
12	4,10	5,82	4,91	6,20	5,71	7,74
15	2,85	4,90	3,90	3,13	3,29	5,88
18	1,90	3,12	2,52	1,55	1,78	3,57

* hCG = gonadotropina coriônica humana (hormônio).

Considerando que um dos objetivos foi verificar se dentro de cada dieta existiam diferenças quanto às curvas de resposta da concentração de progesterona em relação aos tratamentos com e sem hormônio hCG, aplicou-se o teste para verificar a identidade de modelos. Assim, dentro de cada dieta, consideraram-se $H =$ dois modelos de regressão, cada um com $r =$ dois submodelos polinomiais do primeiro grau, e o ponto de interseção foi especificado em $K = 12$. Outra hipótese interessante que poderia ser testada, seria a relativa ao teste para verificar se, no tratamento com hCG, as curvas de respostas para as três dietas eram idênticas. O mesmo poderia ser feito para o tratamento sem hCG.

As equações de regressão ajustadas estão apresentadas na Tabela 4.

O resultado do teste para a identidade dos modelos, dentro de cada dieta, está apresentado na Tabela 5.

Com base no teste apresentado na Tabela 5, para as dietas 1 e 3, a hipótese H_0 foi rejeitada ($P < 0,01$), concluindo-se que as duas equações diferem, estatisticamente, a 1% de probabilidade. Assim, dentro de cada uma dessas dietas, existem diferenças quanto à aplicação ou não de hCG. Na dieta 2, a hipótese H_0 não foi rejeitada ($P > 0,05$). Nesse caso, a equação comum pode ser usada como uma estimativa das duas equações envolvidas. Pode-se concluir que, com ou sem a aplicação de hCG, para os animais que receberam a dieta 2, tem-se a mesma curva de resposta.

TABELA 4. Equações de regressão ajustadas à concentração de progesterona (Y), em função do número de dias após o estro (X), nas dietas 1, 2 e 3, com e sem o hormônio hCG.

Dieta	Hormônio (hCG)	Equações ajustadas*	R ² (%)
1	Sem	$\hat{Y}_i = -0,8260 + 0,4113X_i - 0,7899(X_i - 12)Z_i$	98,8
	Com	$\hat{Y}_i = -2,0399 + 0,6790X_i - 1,1581(X_i - 12)Z_i$	96,8
	Comum	$\hat{Y}_i = -1,4329 + 0,5452X_i - 0,9740(X_i - 12)Z_i$	69,4
2	Sem	$\hat{Y}_i = -0,8559 + 0,4937X_i - 0,9114(X_i - 12)Z_i$	98,1
	Com	$\hat{Y}_i = -0,5239 + 0,5591X_i - 1,3809(X_i - 12)Z_i$	94,2
	Comum	$\hat{Y}_i = -0,6899 + 0,5264X_i - 1,1462(X_i - 12)Z_i$	84,5
3	Sem	$\hat{Y}_i = -0,9699 + 0,5409X_i - 1,1886(X_i - 12)Z_i$	98,4
	Com	$\hat{Y}_i = -1,0199 + 0,7490X_i - 1,4747(X_i - 12)Z_i$	98,3
	Comum	$\hat{Y}_i = -0,9949 + 0,6449X_i - 1,3317(X_i - 12)Z_i$	68,8

* $Z_i = 0$ para $X \leq 12$; e $Z_i = 1$ para $X > 12$.

TABELA 5. Resumo das análises de variâncias com o teste de hipótese $H_0: \beta_1 = \beta_2$ (as duas equações são idênticas) vs. $H_0: \beta_1 \neq \beta_2$ (as duas equações não são idênticas), dentro de cada dieta.

Fonte de variação	GL	Quadrados médios		
		Dieta 1	Dieta 2	Dieta 3
Parâmetro (β)	(6)	(21,7325)	(24,3559)	(36,8754)
Parâmetro (θ)	3	41,7303	47,9889	70,4910
Redução (H_0)	3	1,7346**	0,7232n.s.	3,2598**
Resíduo	4	0,0864	0,2330	0,0954

** Significativo, a 1% de probabilidade pelo teste F.

n.s. = não-significativo, a 5% de probabilidade pelo teste F.

CONCLUSÕES

1. A identidade de modelos de regressão no caso de justaposição de submodelos polinomiais pode ser verificada por meio do teste *F*.
2. O método apresentado é geral e pode ser usado em modelos polinomiais de qualquer grau.

REFERÊNCIAS

- CONAGIN, A.; NAGAI, V.; IGUE, T. Poder discriminativo de diferentes testes de comparação de médias. *Revista de Agricultura*, v.65, n.2, p.203-214, 1990.
- DIXON, W.J.; MASSEY JUNIOR, F.J. *Introduction to statistical analysis*. New York: McGraw-Hill, 1969. 638p.
- FULLER, W.A. Grafted polynomials as approximating functions. *Australian Journal of Agricultural Economics*, v.13, p.35-45, 1969.
- GRAYBILL, F.A. *Theory and application of the linear model*. Belmont: Duxbury Press, 1976. 704p.
- LI, J.C.R. *Statistical inference*. Ann Arbor: Edwards Brothers, 1964. v.1, 658p.
- NAGAI, V.; CONAGIN, A.; IGUE, T. Sensibilidade de diferentes testes de homogeneidade das variâncias. *Revista de Agricultura*, v.66, n.1, p.65-76, 1991.
- NETER, J.; WASSERMAN, W. *Applied linear statistical models. Regression, analysis of variance and experimental design*. Homewood: Richard D. Irwin, 1974. 842p.
- NOGUEIRA, M.C.S. *Análise de regressão através da justaposição de submodelos polinomiais*. Piracicaba, SP: ESALQ/USP, 1980. 12p. Seminário.
- PIMENTEL-GOMES, F. *Curso de Estatística Experimental*. 13. ed. Piracicaba, SP: Livraria Nobel, 1990. 468p.
- REGAZZI, A.J. Teste para verificar a identidade de modelos de regressão e a igualdade de alguns parâmetros num modelo polinomial ortogonal. *Revista Ceres*, v.40, n.228, p.176-195, 1993.