

NOTA CIENTÍFICA

ANÁLISE DA ADAPTABILIDADE POR REGRESSÃO SEGMENTADA COM ESTIMAÇÃO DA JUNÇÃO DOS SEGMENTOS¹

JOÃO GILBERTO CORRÊA DA SILVA²

RESUMO - Os métodos que relacionam as respostas individuais de genótipos com o índice de ambiente por meio de equações de regressão lineares simples não são capazes de identificar o genótipo desejável, ou seja, o genótipo responsivo a ambientes favoráveis ou melhorados e que mantenha produtividade razoável em ambientes adversos. O método de regressão linear segmentada proporciona maior flexibilidade para a caracterização dos distintos comportamentos das respostas dos genótipos à variação do ambiente. Entretanto, um aspecto crítico desse método é a fixação arbitrária do índice de ambiente zero como ponto comum de alteração das taxas de respostas dos genótipos. Este artigo apresenta uma generalização do método de regressão linear segmentada que considera o ponto de mudança da taxa de resposta como um parâmetro de adaptabilidade adicional. A aplicação do método é ilustrada por meio de um exemplo.

Termos para indexação: interação genótipo x ambiente, adaptabilidade ao ambiente, estabilidade fenotípica, modelo linear segmentado, modelo segmentado não-linear.

ANALYSIS OF ADAPTATION THROUGH SEGMENTED REGRESSION WITH ESTIMATION OF THE JOIN POINT

ABSTRACT - The methods which relate the individual responses of genotypes to the environment index through simple linear regression equations are not able of identifying the desirable genotype, that is, the genotype which is responsive to favorable or improved environments and maintain reasonable productivity in adverse environments. The segmented linear regression method allows more flexibility for the characterization of the distinct behaviors of the genotype responses to the variation of the environment. However, the arbitrary choice of the null environment index as the change-point of the response rate of the genotypes is a critical aspect of this method. This paper presents a generalization of the segmented linear regression method which considers the change-point as an additional parameter of adaptation. The application of the method is illustrated through an example.

Index terms: genotype x environment interaction, environment adaptation, phenotypic stability, linear segmented model, nonlinear segmented model.

¹ Aceito para publicação em 28 de outubro de 1997.

² Eng. Agr., Ph.D., Prof. Titular, Dep. de Matemática, Estatística e Computação, Inst. de Física e Matemática, Univ. Fed. de Pelotas, Caixa Postal 354, CEP 96010-900 Pelotas, RS. Bolsista do CNPq. E-mail: jgcs@ufpel.tche.br

INTRODUÇÃO

O método de Finlay & Wilkinson (1963) e seus desenvolvimentos ulteriores (Eberhart & Russell, 1966; Perkins & Jinks, 1968a, 1968b; Freeman & Perkins, 1971; Shukla, 1972) têm sido os mais utilizados em estudos de estabilidade fenotípica e adaptabilidade ao ambiente, em pesquisas de melhoramento genético de plantas cultivadas. Este método relaciona as respostas dos genótipos com a produtividade do ambiente, expressa pelo índice de ambiente, por meio de uma equação de regressão linear simples. Dessa forma, ele não é capaz de identificar o genótipo desejável, ou seja, o genótipo responsivo a ambientes favoráveis ou melhorados e que mantenha produtividade razoável em ambientes adversos (Verma et al., 1978; Silva, 1995a).

Silva (1995a, 1995b) expõe os fundamentos e a aplicação de um método de regressão linear segmentada, e o aborda como uma extensão do método de regressão linear simples para proporcionar maior flexibilidade para a caracterização dos distintos comportamentos das respostas dos genótipos à variação do ambiente. A resposta do genótipo na gama de ambientes é representada por um gráfico composto de dois segmentos de reta, conectados no ponto correspondente ao índice de ambiente nulo.

Um aspecto crítico deste método é a escolha arbitrária do índice de ambiente zero para o ponto de alteração da taxa de resposta linear do genótipo ao incremento do índice de ambiente. Em geral, a alteração da taxa de variação pode ser uma característica específica do genótipo, variável com o genótipo.

A utilização das respostas médias dos próprios genótipos no experimento para exprimir a produtividade do ambiente tem conseqüências para as formas das respostas dos genótipos. Ela implica que a resposta média para cada ambiente, global para todos os genótipos, situa-se exatamente sobre uma reta de declividade igual a um. Se cada genótipo mantém um mesmo padrão de comportamento de resposta no intervalo do índice de ambiente, essa propriedade induz a linearidade das respostas individuais dos genótipos. Por outro lado, ela também implica que, se a resposta de um genótipo afasta-se para um lado dessa reta, as respostas dos demais genótipos desviam-se para o lado oposto, de modo que a soma dos desvios mantém-se nula. Esses desvios decorrem do erro experimental, mas também podem resultar de variações sistemáticas dos padrões de comportamento relativo das respostas dos genótipos, como alterações das taxas de resposta lineares no intervalo do índice de ambiente.

Alterações das taxas de resposta podem ocorrer, por exemplo, nas situações em que um genótipo atinge produtividade máxima no intervalo do índice de ambiente, ou é consideravelmente mais sensível a características do ambiente limitantes para a produção, tais como incidência de doenças e escassez ou excesso de água.

Em geral, uma apropriada representação da relação entre as respostas individuais dos genótipos e o índice de ambiente pode demandar uma função polinomial, ou uma função segmentada constituída de segmentos lineares.

Em muitas situações, uma aproximação razoável tem sido lograda por uma função linear simples. É de esperar que a adoção de uma função polinomial quadrática ou uma função segmentada de dois segmentos lineares estenda a aplicabilidade de modelos de regressão baseados no índice de ambiente. Essa última forma funcional é muito conveniente, já que seus parâmetros exprimem importantes propriedades do comportamento das respostas dos genótipos referentes à adaptabilidade ao ambiente.

Este artigo apresenta uma generalização do modelo de regressão linear segmentada (Silva, 1995a) que considera o ponto de mudança da taxa de resposta como um parâmetro de adaptabilidade adicional. O modelo torna-se não-linear. Expõem-se os procedimentos para o estudo da adaptabilidade relativa dos genótipos baseado nesse modelo e se estabelecem as condições para sua adequabilidade. A utilização desse novo modelo é ilustrada pela sua aplicação no estudo da adaptabilidade de cultivares de um experimento com arroz de sequeiro.

MÉTODO

Modelo estatístico

Postula-se que a resposta média observada do i -ésimo genótipo no j -ésimo ambiente é expressa pela seguinte equação:

$$y_{ij} = \begin{cases} b_{01i} + b_{11i} x_j + e_{ij}, & x_j \leq x_{0i} \\ b_{02i} + b_{12i} x_j + e_{ij}, & x_j > x_{0i} \end{cases}$$

$i=1,2,\dots,C$, $j=1,2,\dots,A$, onde x_j é o índice de ambiente, ou seja, o desvio entre a média do ambiente j e a média geral: $x_j = \bar{y}_{.j} - \bar{y}_{..}$; x_{0i} é o ponto de alteração da taxa de resposta linear esperada do genótipo i ; b_{01i} , b_{11i} , b_{02i} e b_{12i} são constantes desconhecidas; e $e_{ij} = \varepsilon_{ij} + d_{ij}$, sendo ε_{ij} o desvio entre as respostas observada e esperada do i -ésimo genótipo no ambiente j , e d_{ij} , o desvio entre o efeito da interação genótipo x ambiente e o componente sistemático postulado para a relação dessa interação com o índice de ambiente. O componente e_{ij} é suposto aleatório, e supõe-se satisfazer as propriedades usuais do componente do erro, ou seja, homogeneidade de variância, ausência de correlação e normalidade.

A correspondente equação para valores esperados, relativa a cada genótipo, é:

$$E(Y_i) = \begin{cases} b_{01i} + b_{11i} x, & x \leq x_{0i} \\ b_{02i} + b_{12i} x, & x > x_{0i} \end{cases}$$

$i = 1,2,\dots,C$. No desenvolvimento que segue, o índice i será omitido, para simplificação das expressões, subentendendo-se sua presença nos símbolos dos parâmetros e da variável resposta, sempre que se referirem a um genótipo particular.

No ponto $x = x_0$ de alteração da taxa de resposta, tem-se:

$$b_{01} + b_{11}x_0 = b_{02} + b_{12}x_0,$$

onde:

$$b_{02} = b_{01} + (b_{11} - b_{12}) x_0$$

$$e$$

$$x_0 = - \frac{b_{02} - b_{01}}{b_{12} - b_{11}}$$

Substituindo essa expressão de b_{02} em função dos demais parâmetros na equação do modelo estatístico, obtém-se:

$$E(Y) = \begin{cases} b_{01} + b_{11}x, & x \leq x_0 \\ b_{01} + b_{11}x + (b_{12} - b_{11})(x - x_0), & x > x_0 \end{cases}$$

que pode ser expressa na forma de uma única equação:

$$E(Y) = b_{01} + b_{11}x + (b_{12} - b_{11})(x - x_0) I(x - x_0), \text{ onde: } I(x - x_0) = \begin{cases} 0, & x \leq x_0 \\ 1, & x > x_0 \end{cases}$$

Segundo esse modelo segmentado, a adaptabilidade de um genótipo é caracterizada pelos seguintes parâmetros: ponto de alteração da taxa de resposta linear, x_0 ; taxas de resposta à esquerda e à direita de

$$x_0, b_{11} \text{ e } b_{12}, \text{ respectivamente; e rendimento médio, } \frac{1}{A} (A_e b_{01} + A_d b_{02}) + \frac{1}{A} (b_{12} - b_{11}) \sum_{x_j > x_0} x_j,$$

onde A_e e A_d são os números de índices de ambiente no experimento à esquerda e à direita de x_0 , respectivamente.

A variância do desvio da regressão ($\sigma_{d_i}^2, i=1,2,\dots,C$) é um parâmetro que deve ser considerado com cautela na avaliação da adaptabilidade. Desvios de regressão significativos podem ser indicação de que o modelo postulado não é completamente apropriado para a caracterização da estabilidade fenotípica e adaptabilidade. As origens de desvios elevados devem ser investigadas pela inspeção das características ambientais e das demais condições e ocorrências durante a condução do experimento (Silva, 1995a).

Essa equação, com parâmetros b_{01}, b_{11}, b_{12} e x_0 (ou b_{01}, b_{11}, b_{02} e b_{12}), é não-linear. Sua representação geométrica é o gráfico constituído de dois segmentos de reta com junção no ponto x_0 . Os significados geométricos dos parâmetros $b_{01}, b_{11}, b_{02}, b_{12}$ e x_0 são indicados na Fig. 1.

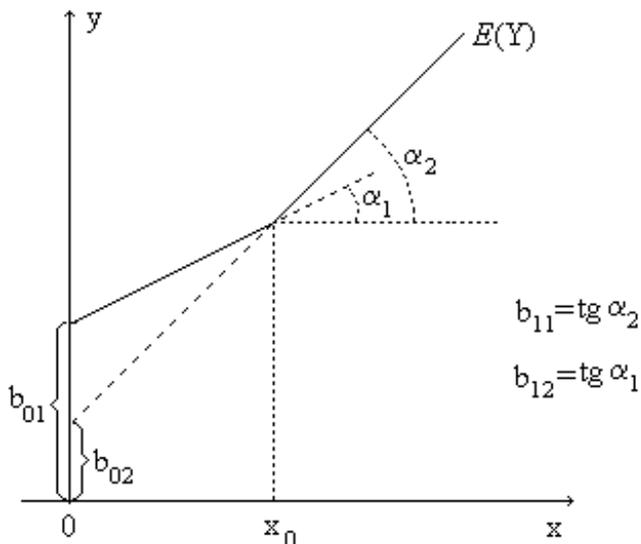


FIG. 1. Gráfico da equação do modelo segmentado com junção no ponto x_0 .

A comparação da adaptabilidade de genótipos baseada na consideração simultânea de quatro parâmetros, pode tornar-se complexa. Entretanto, é de esperar que, muito freqüentemente, os genótipos apresentarão poucos padrões distintos de comportamento de resposta. Na situação mais simples em que as taxas de resposta dos genótipos permanecem constantes, o modelo de regressão segmentada reduz-se ao modelo de regressão linear simples.

Por outro lado, podem ser derivadas conclusões a partir da fixação de um ou mais desses quatro parâmetros. Assim, por exemplo, como no modelo de regressão linear segmentada com junção na origem (Silva, 1995a), entre genótipos com alterações de taxa de resposta no mesmo índice de ambiente x_0 , em geral, é desejável o genótipo com rendimento elevado e com taxas de resposta baixa à esquerda de x_0 e elevada à direita. Entre genótipos de mesmo rendimento médio e mesmas taxas de resposta à esquerda de x_0 e à direita de x_0 com esta taxa maior do que aquela, aqueles com valores maiores de x_0 oferecem menores riscos sob condições adversas de ambiente, mas são menos produtivos sob condições de ambiente favoráveis; os que têm valores menos elevados de x_0 implicam maiores riscos, mas são mais produtivos sob condições favoráveis de ambiente.

Estimação dos parâmetros

A equação desse modelo segmentado não satisfaz as condições de regularidade requeridas pelos métodos de inferência para modelos não-lineares (Gallant, 1974, 1975). As propriedades assintóticas desses métodos (Gauss-Newton, Marquardt e gradiente máximo) requerem que a função de resposta e suas derivadas parciais de primeira ordem sejam contínuas nas variáveis preditoras e nos parâmetros. A continuidade das derivadas parciais não é satisfeita pela equação de regressão segmentada. Conseqüentes dificuldades computacionais decorrem, principalmente com relação à convergência do processo iterativo. Essas dificuldades foram observadas na verificação do comportamento desses métodos quando aplicados ao modelo de regressão segmentada com estimação da junção dos segmentos.

Observa-se, entretanto, que essa equação é não-linear apenas no parâmetro x_0 . A fixação de um valor para esse parâmetro torna o modelo linear nos demais parâmetros b_{01} , b_{11} , b_{12} . De fato, efetuando a translação $z = x - x_0$, obtém-se:

$$E(Y) = b_{01} + b_{11}(z+x_0) + (b_{12}-b_{11}) z I(z)$$

$$= b_{01} + b_{11} x_0 + b_{11} z + (b_{12} - b_{11}) z I(z), \quad I(z) = \begin{cases} 0, & z \leq 0 \\ z, & z > 0 \end{cases}$$

Com a reparametrização: $b_0 = b_{01} + b_{11} x_0$, $b_1 = b_{11}$, $b_2 = b_{12} - b_{11}$, resulta:

$$E(Y) = b_0 + b_1 z + b_2 z I(z),$$

ou seja:

$$E(Y) = b_0 + b_1 x_1 + b_2 x_2,$$

onde: $x_1 = z$, e $x_2 = 0$ se $z \leq 0$, $x_2 = z$ se $z > 0$, que é a equação do modelo de regressão linear múltipla com duas variáveis preditoras x_1 e x_2 , análoga à equação do modelo linear segmentado com junção na origem (Silva, 1995a).

Um procedimento para a estimação dos parâmetros pelo método de quadrados mínimos para modelos lineares é indicado a seguir.

Os estimadores de quadrados mínimos dos parâmetros b_{01} , b_{11} , b_{02} , b_{12} e x_0 relativos a cada genótipo são os valores dessas constantes do modelo segmentado que tornam mínima a soma dos quadrados dos resíduos desse modelo para o genótipo particular. Assim, a estimação desses parâmetros, referente a cada genótipo, pode ser obtida pela busca do ponto x_0 a que corresponde a menor das SQ Resíduo dos ajustamentos dos modelos lineares segmentados que resultam da fixação do ponto de junção x_0 em cada um dos pontos do intervalo entre o menor e o maior dos índices de ambiente no experimento.

Esse processo pode ser executado mediante a seguinte estratégia:

1) Efetua-se o ajustamento da equação de regressão linear separadamente para cada genótipo, com x_0 fixado em cada um dos índices de ambiente determinados pelo experimento, excetuados o primeiro e o último. Identificam-se os índices de ambiente a que correspondem as menores SQ Resíduo. Seleciona-se o subintervalo ou os subintervalos desses índices de ambiente, conforme esses índices sejam contíguos ou não.

Como os parâmetros b_{01} , b_{11} , b_{02} , b_{12} e x_0 têm interpretação geométrica, esses ajustamentos podem ser reduzidos, pela inspeção do correspondente diagrama de dispersão das médias, na medida em que esta representação gráfica indique com clareza um subintervalo para o ponto de alteração da taxa de resposta linear. Nesse caso, pode-se limitar a busca do ponto x_0 a este subintervalo.

2) Subdivide-se o subintervalo, ou o conjunto de subintervalos, selecionado no passo anterior em um novo subconjunto conveniente de subintervalos. Procede-se, então, ao ajustamento do modelo de regressão linear segmentada com x_0 fixado em cada uma das fronteiras desses subintervalos. Identifica-se o subintervalo, ou conjunto de subintervalos, a que correspondem as menores SQ Resíduo.

3) Repete-se o mesmo processo do passo anterior para a identificação de um ou mais subintervalos internos do subintervalo ou dos subintervalos selecionados naquele passo a que correspondem as menores SQ Resíduo.

4) Esse processo de busca é interrompido quando o subintervalo selecionado degenera em um ponto para a precisão dos índices de ambiente adotada, ou as SQ Resíduo determinadas tornam-se iguais para a precisão numérica desejada para a SQ Resíduo. Esse ponto é a estimativa do ponto de junção x_0 . As estimativas dos demais parâmetros $b_{01}, b_{11}, b_{02}, b_{12}$ são:

$$\hat{b}_{01} = \hat{b}_0 - \hat{b}_1 \hat{x}_0, \hat{b}_{11} = \hat{b}_1, \hat{b}_{02} = \hat{b}_0 + (\hat{b}_2 - \hat{b}_1) \hat{x}_0, \hat{b}_{12} = \hat{b}_1 + \hat{b}_2,$$

onde \hat{b}_0, \hat{b}_1 e \hat{b}_2 são as estimativas dos parâmetros b_0, b_1 , e b_2 do modelo linear segmentado com a fixação do ponto de junção em x_0 , determinadas nesse último passo do processo iterativo.

Esse processo, aparentemente tedioso, torna-se exequível com uso de um “pacote” para análise estatística que disponha de procedimento para análise do modelo linear geral, como o PROC GLM do “Statistical Analysis System” (SAS Institute, 1989).

Testes de hipóteses

O modelo de regressão segmentada com estimação do ponto de junção propicia testes de diversas hipóteses de interesse referentes aos parâmetros $b_{01}, b_{11}, b_{02}, b_{12}$ e x_0 que caracterizam a adaptabilidade relativa dos genótipos. Algumas dessas hipóteses são correspondentes e análogas às hipóteses relativas aos parâmetros de adaptabilidade do modelo linear segmentado com junção na origem, descritas por Silva (1995a). Uma hipótese adicional importante diz respeito à homogeneidade da junção dos segmentos para os genótipos. Essas hipóteses e os correspondentes procedimentos para teste são indicados a seguir.

Hipóteses 1

O primeiro conjunto de hipóteses de interesse diz respeito à decisão entre as duas seguintes especificações alternativas de equação para o modelo estatístico, para cada genótipo:

$$\begin{cases} H_{0(i)}: E(Y_i) = b_{0i} + b_{1i}x \\ H_{A(i)}: E(Y_i) = b_{0i} + b_{1i}x + (b_{12i} - b_{11i})(x - x_{0i}) I(x - x_{0i}) \end{cases}$$

$i=1,2,\dots,C$.

Os testes dessas hipóteses são importantes para a decisão referente à adoção do modelo de regressão segmentada em preferência ao modelo de regressão linear simples. Se a hipótese de nulidade H_{0i} é aceita para um genótipo, a relação entre o rendimento do genótipo e o índice de ambiente pode ser adequadamente representada por um modelo de regressão linear simples; caso contrário, ou seja, se a decisão é favorável à hipótese H_{Ai} , o modelo de regressão segmentada revela-se o mais adequado. Se a hipótese de nulidade H_{0i} é aceita para todos os genótipos, procede-se à aplicação do método de regressão linear simples.

Como salientado anteriormente, os métodos usuais de inferência para modelos não-lineares não são aplicáveis ao modelo de regressão segmentada com estimação da junção. Gallant (1977) sugere um procedimento para teste de hipótese com modelos não-lineares aplicável à presente situação. Entretanto, dado que sob cada uma das hipóteses de nulidade o correspondente modelo é linear e que sob as correspondentes hipóteses alternativas os modelos podem ser ajustados pelo procedimento de regressão linear descrito anteriormente, torna-se mais conveniente efetuar os testes das hipóteses em consideração pelas seguintes estatísticas:

$$F_i = \frac{[SQ \text{ Resíduo}(H_{0i}) - SQ \text{ Resíduo}(H_{Ai})] / v_{1i}}{SQ \text{ Resíduo}(H_{Ai}) / v_{2i}}$$

$i=1, 2,\dots,C$, onde $SQ \text{ Resíduo}(H_{0i})$ e $SQ \text{ Resíduo}(H_{Ai})$ são as somas de quadrados dos resíduos do modelo para o i -ésimo genótipo sob as hipóteses de nulidade e alternativa, respectivamente com v_{1i} e v_{2i} graus de liberdade.

Hipóteses 2 e 3

Hipóteses referentes à adequabilidade do modelo de regressão segmentada com estimação da junção (que inclui o modelo de regressão linear simples como um caso particular) e à preferência desse modelo relativamente ao modelo de regressão linear simples, globalmente para todos os genótipos.

Essas hipóteses podem ser testadas por procedimentos análogos aos indicados para as correspondentes hipóteses 1 e 2 do modelo linear segmentado com junção no ponto correspondente ao índice de ambiente zero (Silva, 1995a), com o auxílio da Tabela 1. Nessa Tabela, a SQ Resíduo $R(X_0, X_1, X_2)$ é a soma das somas dos quadrados dos resíduos dos ajustamentos do modelo segmentado separadamente para cada genótipo. O correspondente número de graus de liberdade é $(C-1)(A-4)$, em vez de $C(A-4)$, para levar em

conta a dependência linear dos resíduos implicada pelo ajuste perfeito da equação do modelo para as médias globais dos A ambientes.

TABELA 1. Decomposição da interação genótipo x ambiente para os teste das hipóteses 2 e 3.

Fonte de variação	Graus de liberdade	Soma de quadrados
Genótipo x ambiente	(C-1)(A-1)	
Heter.R(X_0, X_1, X_2)	3(C-1)	SQ Gen. x amb. - SQ Resíduo R(X_0, X_1, X_2)
Heter.R(X_1)	C-1	SQ Gen. x amb.- SQ Resíduo R(X_1)
Heter.R($X_0, X_2 X_1$)	2(C-1)	SQ Heter.R(X_0, X_1, X_2) - SQ Heter.R(X_1)
Resíduo combinado	(C-1)(A-4)	SQ Resíduo R(X_0, X_1, X_2)

Observe-se que a não-significância da Heter.R($X_0, X_2|X_1$) não implica, necessariamente, que o modelo de regressão segmentada não seja melhor que o modelo de regressão linear simples para todos os C genótipos. Por outro lado, a significância da Heter.R($X_0, X_1|X_2$) não implica, necessariamente, que o modelo de regressão segmentada seja mais apropriado para todos os genótipos. Nessas circunstâncias, a escolha entre esses modelos para uma situação particular depende de julgamento do pesquisador. Um meio termo é decidir com base nos testes das hipóteses 1, separadamente para cada genótipo.

Observe-se, ademais, que esse procedimento para os testes das hipóteses 2 e 3 requer a combinação dos quadrados médios dos resíduos das regressões individuais para os C genótipos. Essa combinação é válida somente se a variância dos resíduos da regressão é homogênea para todos os genótipos. Essa homogeneidade de variância pode ser testada por meio de teste apropriado, como os testes de Bartlett, de Hartley, de Cochran ou de Burr-Foster (Kirk, 1968; Anderson & McLean, 1974).

Se a homogeneidade da variância dos resíduos é rejeitada, podem ser de interesse e válidos testes de hipóteses análogas às hipóteses 2 e 3 para subconjuntos de genótipos.

Hipótese 4

Homogeneidade das junções dos segmentos para os C genótipos, ou seja:

$$\begin{cases} H_0: x_{0i} = x_0 \text{ (constante), } i = 1, 2, \dots, C \\ H_A: x_{0i} \neq x_{0i'}, \text{ para pelo menos dois genótipos} \end{cases}$$

O teste desta hipótese indica se deve ser tomado um ponto de junção comum para todos os C genótipos, ou pontos de junção diferenciados para os genótipos. Nesse segundo caso, podem não fazer sentido hipóteses semelhantes às consideradas por Silva (1995a), referentes às taxas de variação dos genótipos nos dois subconjuntos de índices de ambientes demarcados por x_0 .

Esta hipótese refere-se ao modelo de regressão segmentada conjunto para os C genótipos, que, com a fixação da junção x_0 , pode ser posto sob a forma do modelo de regressão linear múltipla:

$$E(Y_i) = b_{0i} + b_{1i}x_1 + b_{2i}x_2, \quad i=1, 2, \dots, C,$$

onde: $x_1=z$, e $x_2=0$ se $z \leq 0$, $x_2 = z$ se $z > 0$ ($z=x-x_0$).

Esta hipótese pode ser testada por uma estatística F semelhante às indicadas para os testes das hipóteses 1. A SQ Resíduo(H_A) é a SQ Resíduo Combinado, ou seja, a soma das somas dos quadrados dos resíduos dos ajustamentos dos modelos segmentados individuais para os C genótipos. A SQ Resíduo(H_0) é a soma dos quadrados dos resíduos do ajustamento do modelo de regressão segmentada conjunto para os C genótipos sob a condição $x_{0i}=x_0$.

A determinação da SQ Resíduo (H_0) requer a busca do ponto x_0 a que corresponde a menor SQ Resíduo dos ajustamentos dos modelos lineares conjuntos para os C genótipos que resultam da fixação do ponto de junção x_0 em cada um dos pontos do intervalo entre o menor e o maior dos índices de ambiente no experimento. Essa busca pode ser procedida pelo mesmo processo descrito para o ajustamento do modelo segmentado separadamente para cada genótipo.

Como observado anteriormente, a combinação de quadrados médios de resíduos requer homogeneidade da correspondente variância estimada. Nessas circunstâncias, esse procedimento para teste da hipótese 4 requer a homogeneidade da variância dos resíduos das regressões individuais para os C genótipos.

Se os testes das hipóteses 1 conduzem à rejeição das correspondentes hipóteses de nulidade para todos os C genótipos e o teste da hipótese 4 à aceitação de ponto de junção comum para todos os genótipos, testes de hipóteses referentes às taxas de resposta nos subespaços de ambiente à esquerda e à direita do

ponto de junção comum podem ser efetuados por procedimentos análogos aos descritos para a análise da adaptabilidade pela regressão linear segmentada com junção no índice de ambiente zero (Silva, 1995a). Se as hipóteses de nulidade do conjunto de hipóteses 1 não são rejeitadas para todos os genótipos, aquelas hipóteses podem ser testadas para um subconjunto dos genótipos para os quais a hipótese de nulidade foi rejeitada.

Para a aplicação desses procedimentos, o modelo deve ser apropriadamente redefinido com a fixação da junção dos segmentos no ponto estimado comum para os C genótipos, pela translação $z = x - e$, e, então, reparametrizado para a forma:

$$E(Y) = b_0 + b_1x_1 + b_2x_2,$$

onde: $x_1 = z$; $x_2 = 0$ se $z \leq 0$, $x_2 = z$ se $z > 0$; $b_0 = b_{01} + b_{11}$; $b_1 = b_{11}$; e $b_2 = b_{12} - b_{11}$.

Observe-se que para esses testes de hipóteses os números de graus de liberdade de resíduos indicados por Silva (1995a) devem ser subtraídos de uma unidade, para levar em conta a estimação do ponto de junção.

Para certos propósitos de estudo de adaptabilidade, pode ser mais conveniente um procedimento para a classificação dos genótipos em grupos homogêneos quanto ao padrão de adaptabilidade ao ambiente. O método de agrupamento sugerido por Silva (1995c) pode ser adaptado para o modelo de regressão segmentada com estimação da junção.

APLICAÇÃO DO MÉTODO

Para ilustração da aplicação do método de regressão segmentada com estimação da junção dos segmentos, considere-se os dados de um experimento do programa de melhoramento de arroz de sequeiro do Instituto Agrônomo utilizados por Silva (1995b, 1995c). Os dados procedem de três anos do experimento em 9 locais do Estado de São Paulo, compreendendo 21 ambientes.

As estimativas dos parâmetros do modelo segmentado com estimação do ponto de junção são apresentadas na Tabela 2. Observe-se que as estimativas das taxas de resposta à esquerda e à direita dos pontos de junção estimados diferem das correspondentes estimativas obtidas pelo ajustamento do modelo linear segmentado com junção no índice de ambiente nulo (Silva, 1995b). Entretanto, sua apreciação revela a nítida caracterização dos mesmos dois grupos de genótipos determinados por este modelo. Os gráficos das linhas segmentadas ajustadas para cada uma das 8 cultivares são apresentados na Fig. 2. As Figs. 3 e 4 mostram os gráficos dessas linhas, juntamente com os correspondentes diagramas das médias ambientais, para uma cultivar de cada um dos dois distintos grupos de cultivares: Pratão Precoce e IAC 5032.

TABELA 2. Estimativas dos parâmetros do modelo de regressão segmentada com estimação da junção dos segmentos, para cada uma das 8 cultivares.

Cultivar	\hat{b}_{01}	\hat{b}_{11}	\hat{b}_{02}	\hat{b}_{12}	\hat{x}_0	S_d^2
1	1.709	0,105	472	2,040	639	237.904
2	1.753	0,243	-829	2,747	1.031	263.424
3	1.943	0,199	819	1,957	639	394.063
4	2.118	1,418	3.134	0,240	863	83.685
5	2.075	1,475	2.952	0,324	762	79.066
6	2.124	1,545	2.852	0,405	639	104.537
7	2.144	1,572	2.938	0,328	639	115.126
8	2.426	1,579	2.924	0,371	412	140.517

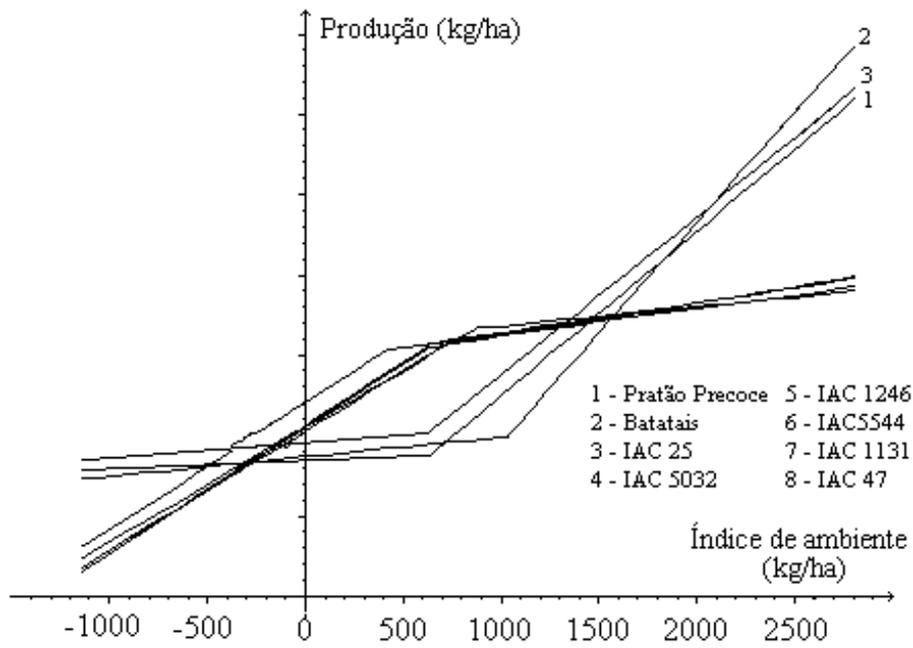


FIG. 2. Representação gráfica da equação do modelo segmentado com estimação da junção para cada uma das oito cultivares.

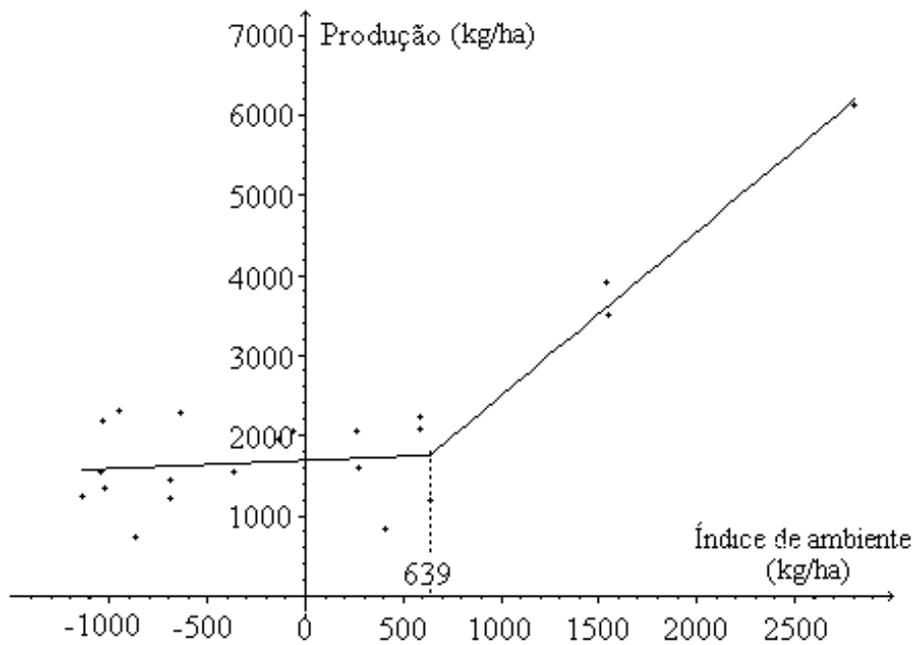


FIG. 3. Representação gráfica da equação do modelo segmentado com estimação da junção, e das médias ambientais observadas, para a cultivar Pratóo Precoce.

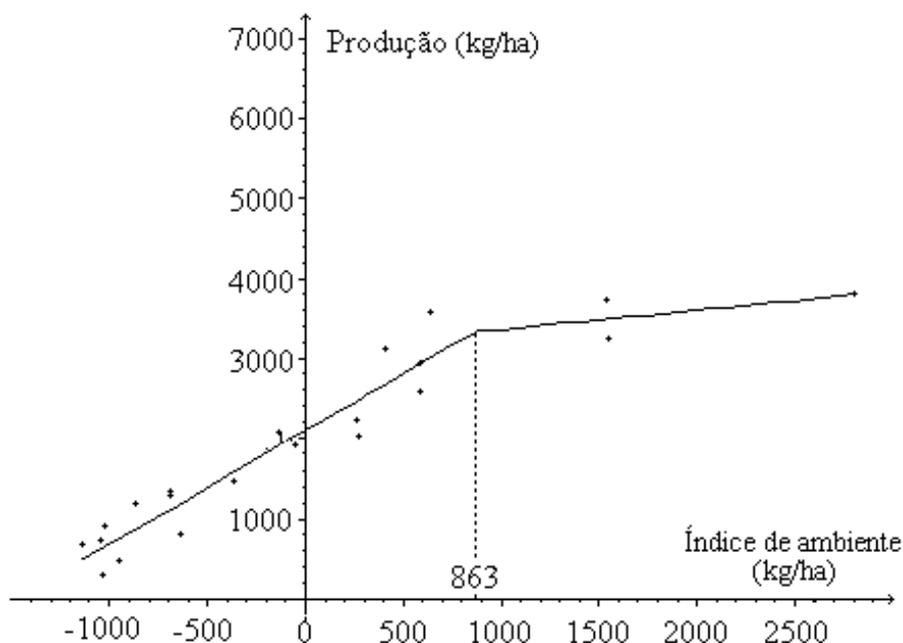


FIG. 4. Representação gráfica da equação do modelo segmentado com estimação da junção, e das médias ambientais observadas, para a cultivar IAC 5032.

Os resultados dos testes das hipóteses referentes à conveniência da adoção do modelo segmentado com estimação da junção dos segmentos relativamente ao modelo de regressão linear simples, para cada cultivar (conjunto de hipóteses 1), estão na Tabela 3. Esses resultados revelam a superioridade do modelo de regressão segmentada para todas as 8 cultivares.

A Tabela 4 mostra as estimativas dos desvios das regressões do modelo linear simples de Finlay & Wilkinson (1963), do modelo polinomial quadrático e do modelo linear segmentado (Silva, 1995a), para cada uma das 8 cultivares. Esses resultados e os da Tabela 2 indicam que, no presente exemplo, o modelo de regressão segmentada com estimação da junção dos segmentos ajustou-se aos dados consideravelmente melhor que o modelo de regressão linear simples e, também, melhor que o modelo de regressão quadrático e o modelo de regressão linear segmentada com junção na origem. Entretanto, a Tabela 3 mostra que os resíduos desses ajustamentos continuam altamente significativos.

TABELA 3. Teste da hipótese referente à especificação do modelo para a análise da adaptabilidade:

$H_0: E(Y_i) = b_{0ii} + b_{1ii}x$, $H_A: E(Y_i) = b_{0ii} + b_{1ii}x + (b_{12i} - b_{11i})(x - x_0) I(x - x_0)$, $I(x - x_0) = 0$ se $x \leq x_0$, $I(x - x_0) = 1$ se $x > x_0$, e testes de significância dos desvios das regressões linear simples e segmentada com estimação da junção, para cada cultivar.

Fonte de variação	GL	SQ	QM	Prob.>F
Pratão Precoce				
Resíduo(H_0)	19	11.778.977	619.946	< 0,0001
Resíduo(H_A)	17	4.434.772	260.869	< 0,0001
Hipótese(H_0)	2	7.344.205	3.672.102	< 0,0001
Batatais				
Resíduo(H_0)	19	13.171.976	693.262	< 0,0001
Resíduo(H_A)	17	4.868.617	286.389	< 0,0001
Hipótese(H_0)	2	8.303.359	4.151.680	< 0,0001
IAC 25				
Resíduo(H_0)	19	13.155.021	692.370	< 0,0001
Resíduo(H_A)	17	7.089.468	417.028	< 0,0001

Hipótese(H ₀)	2	6.065.553	3.032.776	0,0052
IAC 5032				
Resíduo(H ₀)	19	3.979.297	209.437	< 0,0001
Resíduo(H _A)	17	1.813.055	106.650	< 0,0001
Hipótese(H ₀)	2	2.166.242	1.083.121	0,0013
IAC 1246				
Resíduo(H ₀)	19	4.026.163	211.903	< 0,0001
Resíduo(H _A)	17	1.734.522	102.031	< 0,0001
Hipótese(H ₀)	2	2.291.641	1.145.820	0,0008
IAC 5544				
Resíduo(H ₀)	19	4.717.277	248.278	< 0,0001
Resíduo(H _A)	17	2.167.541	127.502	< 0,0001
Hipótese(H ₀)	2	2.549.736	1.274.868	0,0013
IAC 1131				
Resíduo(H ₀)	19	5.379.530	283.133	< 0,0001
Resíduo(H _A)	17	2.347.542	138.091	< 0,0001
Hipótese(H ₀)	2	3.031.988	1.515.994	0,0009
IAC 47				
Resíduo(H ₀)	19	5.664.576	298.136	< 0,0001
Resíduo(H _A)	17	2.779.194	163.482	< 0,0001
Hipótese(H ₀)	2	2.885.382	1.442.691	0,0024
Erro combinado	(882) 557		22.965	

TABELA 4. Estimativas das variâncias dos desvios da regressão referentes aos ajustamentos dos modelos: (1) linear simples (Finlay & Wilkinson, 1963), (2) polinomial quadrático e (3) linear segmentado (Silva, 1995a).

Cultivar	Modelo		
	(1)	(2)	(3)
Pratão precoce	596.981	273.398	329.663
Batatais	670.297	277.706	392.426
IAC 25	669.405	443.719	488.490
IAC 5032	186.472	91.858	121.677
IAC 1246	188.938	86.341	114.332
IAC 5544	225.313	116.047	140.801
IAC 1131	260.168	135.122	167.571
IAC 47	275.171	148.279	156.914

Esses resultados evidenciam que os modelos segmentados permitem mais flexibilidade que o modelo linear simples para a aproximação da relação do rendimento com o índice de ambiente. O modelo segmentado com estimação da junção dos segmentos provê flexibilidade adicional que, de modo geral, deve possibilitar melhor aproximação dessa relação e, portanto, mais apropriada caracterização dos distintos comportamentos das respostas dos genótipos à variação do ambiente.

A Tabela 5 apresenta os resultados dos testes das hipóteses globais para as 8 cultivares referentes à adequabilidade da família de modelos abrangida pelo modelo de regressão segmentada com estimação da junção (que inclui o modelo de regressão linear simples como um caso particular) e à conveniência dessa forma mais flexível de modelo relativamente ao modelo de regressão linear simples (hipóteses 2 e 3). Os resultados evidenciam a adequabilidade daquele novo modelo em sua forma completa e sua superioridade relativamente ao modelo de regressão linear simples, globalmente para as cultivares no experimento.

Os resultados do teste de homogeneidade do ponto de alteração das taxas de resposta para as 8 cultivares, apresentados na Tabela 6, revelam a homogeneidade desse ponto, apesar da considerável amplitude do intervalo dos pontos de junção estimados para as 8 cultivares (Tabela 2). A aceitação da

homogeneidade pode ter decorrido do fato de que 7 dos 8 pontos de junção estimados situaram-se no intervalo compreendido entre dois índices de ambiente consecutivos no experimento bastante distanciados (ou seja, 639 e 1535).

A Tabela 7 apresenta as estimativas dos parâmetros de adaptabilidade segundo o método de regressão segmentada com junção estimada comum para as 8 cultivares, $\hat{x}_0 = 653$.

Os resultados do teste de homogeneidade da taxa de resposta das cultivares à esquerda de $\hat{x}_0 = 653$ e do teste de homogeneidade da taxa de resposta à direita desse ponto mostram heterogeneidade de ambas taxas de resposta (Tabelas 8 e 9). Os resultados das comparações múltiplas das taxas de resposta das cultivares em cada um dos subconjuntos de índices de ambiente demarcados pelo ponto de junção \hat{x}_0 estão na Tabela 7. Esses resultados são similares aos obtidos por Silva (1995b).

TABELA 5. Decomposição da interação cultivar x ambiente segundo o modelo de regressão segmentada com estimação da junção dos segmentos.

Fonte de variação	GL ¹	SQ	QM	Prob.>F
Cultivar x ambiente	140(90)	62.530.916		
Heter.R(X_1, X_2, X_0)	21	35.296.205	1.680.772	< 0,0001
Heter.R(X_1)	7	658.098		
Heter.R($X_2, X_0 X_1$)	14	34.638.107	2.474.150	< 0,0001
Resíduo combinado	119	27.234.711	228.863	< 0,0001
Erro combinado	882(557)		22.965	

¹ Entre parênteses, números de graus de liberdade ajustados para levar em conta a heterogeneidade do erro experimental entre ambientes (Pimentel-Gomes, 1990).

TABELA 6. Teste da hipótese de homogeneidade do ponto de alteração da taxa de resposta para as 8 cultivares: $H_0: x_{0i} = x_0$ (constante), $H_A: x_{0i} \neq x_0, i=1,2,\dots,8$, do modelo de regressão segmentada com estimação da junção dos segmentos.

Fonte de variação	GL	SQ	QM	Prob.>F
Hipótese H_0	6	589.547	98.258	0,8581
Resíduo(H_A)	119	27.234.711	228.863	
Resíduo(H_0)	125	27.824.258		

TABELA 7. Estimativas dos parâmetros de adaptabilidade segundo o método de regressão segmentada com junção estimada $\hat{X}_0 = 653$, para cada uma das 8 cultivares¹.

Cultivar	Média (kg/ha)	\hat{b}_{11}	\hat{b}_{12}	S_d^2
P. Precoce	2.076a	0,108a	2,056a	238.291 **
Batatais	2.087a	0,183a	2,202a	287.997 **
IAC 25	2.276a	0,201a	1,970a	395.019 **
IAC 5032	1.932a	1,434b	0,383b	85.373 **
IAC 1246	1.877a	1,483b	0,396b	79.496 **
IAC 5544	1.908a	1,544b	0,396b	104.556 **
IAC 1131	1.908a	1,570b	0,318b	115.169 **
IAC 47	2.125a	1,477b	0,279b	147.099 **

TABELA 8. Teste da hipótese de igualdade das taxas de resposta dos genótipos à esquerda do ponto de junção estimado $\hat{X}_0 = 653$ para as 8 cultivares do modelo de regressão segmentada com estimação da junção dos segmentos.

Fonte de variação	GL	SQ	QM	Prob.>F
Hipótese H_0	7	25.830.297	3.690.042	<0,0001
Resíduo(H_A)	125	27.824.258	222.594	
Resíduo(H_0)	132	53.654.555		

TABELA 9. Teste da hipótese de igualdade das taxas de resposta dos genótipos à direita do ponto de junção estimado $\hat{X}_0 = 653$ para as 8 cultivares do modelo de regressão segmentada com estimação da junção dos segmentos.

Fonte de variação	GL	SQ	QM	Prob.>F
Hipótese H_0	7	24.402.158	3.486.023	<0,0001
Resíduo(H_A)	125	27.824.258	222.594	
Resíduo(H_0)	132	52.226.416		

REFERÊNCIAS

- ANDERSON, V.L.; McLEAN, R.A. **Design of experiments: a realistic approach**. New York: Marcel Dekker, 1974. 418p.
- EBERHART, S.A.; RUSSELL, W.A. Stability parameters for comparing varieties. **Crop Science**, Madison, v.6, p.36-40, 1966.
- FINLAY, K.W.; WILKINSON, G.N. The analysis of adaptation in a plant-breeding programme. **Australian Journal of Agricultural Research**, Victoria, v.14, p.742-754, 1963.
- FREEMAN, G.H.; PERKINS, J.M. Environmental and genotype-environmental components of variability. VIII. Relations between genotypes grown in different environments and measures of these environments. **Heredity**, Edinburgh, v.27, p.15-23, 1971.
- GALLANT, A.R. **The theory of nonlinear regression as it relates to segmented polynomial regressions with estimated join points**. Raleigh: North Carolina State Univ., 1974. 25p. (Institute of Statistics mimeograph series, 925.)
- GALLANT, A.R. Nonlinear regression. **The American Statistician**, v. 29, n.2, p.73-81, 1975.

- GALLANT, A.R. Testing a nonlinear regression specification: A nonregular case. **Journal of the American Statistical Association**, v.72, n.359, p.523-530, 1977.
- KIRK, R.E. **Experimental design**: procedures for the behavioral sciences. Belmont, CA: Wadsworth Pub. Co., 1968. 577p.
- PERKINS, J.M.; JINKS, J.L. Environmental and genotype-environmental components of variability. III. Multiple lines and crosses. **Heredity**, Edinburgh, v.23, p.339-356, 1968a.
- PERKINS, J.M.; JINKS, J.L. Environmental and genotype-environmental components of variability. IV. Non-linear interactions for multiple inbred lines. **Heredity**, Edinburgh, v.23, p.525-535, 1968b.
- PIMENTEL-GOMES, F. **Curso de Estatística Experimental**. 13.ed. São Paulo: Nobel, 1990. 467p.
- SAS INSTITUTE. **SAS/STAT User's Guide**: version 6. 4.ed. Cary, NC: SAS Institute Inc., 1989. v.2, 846p.
- SHUKLA, G.K. Some statistical aspects of partitioning genotype-environmental components of variability. **Heredity**, Edinburgh, v.29, p.237-245, 1972.
- SILVA, J.G.C. da. Análise da adaptabilidade através de regressão linear segmentada. 1. Fundamentos. **Pesquisa Agropecuária Brasileira**, Brasília, v.30, n.4, p.435-448, abr. 1995a.
- SILVA, J.G.C. da. Análise da adaptabilidade através de regressão linear segmentada. 2. Aplicação. **Pesquisa Agropecuária Brasileira**, Brasília, v.30, n.4, p.449-462, abr. 1995b.
- SILVA, J.G.C. da. Agrupamento de genótipos segundo a adaptabilidade ao ambiente. **Pesquisa Agropecuária Brasileira**, Brasília, v.30, n.9, p.1141-1154, set. 1995c.
- VERMA, M.M.; CHAAL, G.S.; MURTY, B.R. Limitations of conventional regression analysis: a proposed modification. **Theoretical and Applied Genetics**, v.53, p.89-91, 1978.