

ESTUDO DA ESTABILIDADE FENOTÍPICA DE NOVE GENÓTIPOS DE SOJA (*Glycine max L.* *Merrill*) PELO USO DE REGRESSÃO LINEAR SEGMENTADA^{1/}

Carlos Sigueyuki Sediyama ^{2/}
Ney Sussumu Sakiyama ^{2/}

1. INTRODUÇÃO

Na avaliação de genótipos em uma série de ambientes, as diferenças de comportamento causadas pela interação genótipo x ambiente têm sido medidas pelas técnicas de regressão linear (1, 2, 4, 5, 6).

A idéia de dividir os ambientes em dois subgrupos (ambientes desfavoráveis e ambientes favoráveis) foi desenvolvida por VERMA *et alii* (6), que consideraram como genótipo ideal o que tivesse alta capacidade de produção, associada a sensibilidade relativamente baixa em ambientes desfavoráveis e sensibilidade alta em ambientes favoráveis.

O método consiste no ajustamento separado de uma regressão linear simples para um dos dois subgrupos de ambientes. Segundo os pesquisadores, nos métodos que ajustam o coeficiente de regressão, considerando todos os ambientes, o genótipo testado apresenta inclinação única, o que não permite a interpretação de seu «comportamento» nos dois subgrupos, separadamente. Além disso, considerando todos os ambientes, o genótipo ideal poderia ter grande desvio, não se ajustando à linha de regressão, e ser descartado, caso a magnitude dos desvios nos diferentes ambientes não fosse examinada. Entretanto, ressaltaram que, quando se analisa um número pequeno de ambientes em qualquer um dos subgrupos, a inclinação da regressão pode não alcançar significância ou a análise pode ser impraticável.

O presente trabalho teve por finalidade comparar o «comportamento» de nove genótipos de soja, avaliados através da regressão linear simples, com inclinação única para todos os ambientes (1), e da regressão linear segmentada (3).

1/ Aceito para publicação em 14-11-1988.

2/ Departamento de Fitotecnia da UFV. 36570 Viçosa, MG.

2. MATERIAL E MÉTODOS

Neste trabalho foram utilizados 21 ensaios do ano agrícola 1981/82 e 22 ensaios do ano agrícola 1982/83, num total de 43 ensaios, distribuídos em 16 diferentes localidades do Estado de Minas Gerais.

Os experimentos de campo foram delineados em blocos completos casualizados, com 16 tratamentos e três repetições, tendo sido analisados somente os nove genótipos comuns aos dois anos de plantio: Cristalina, Juparaná, Rio Doce, UFV-1, Sucupira, UFV 80-69, UFV 79-56, UFV 79-58 e IAC-2.

Cada unidade experimental foi constituída de quatro fileiras de cinco metros de comprimento, espaçadas de 0,60 m, com densidade populacional em torno de 24 plantas por metro linear. A área útil, de 4,80 m², foi obtida pela colheita de duas fileiras centrais, eliminando-se 0,50 m de cada extremidade.

Inicialmente, realizou-se a análise de variância conjunta dos dados, como esquematizado no Quadro 1.

QUADRO 1 - Esquema da análise de variância conjunta de v genótipos em n ambientes e r blocos casualizados por ambiente

F.V.	g.1.	SQ	QM
Blocos/ambiente	n(r-1)	SQB	QMB
Genótipo	(v-1)	SQG	QMG
Ambiente	(n-1)	SQA	QMA
Genótipo x ambiente	(v-1)(n-1)	SQGA	QMGA
Erro Puro	n(r-1)(v-1)	SQE	QME
Total	nvr-1		

A regressão linear simples foi utilizada para obter os valores de b_1 e de $s^2 d_i$, equivalentes aos parâmetros de estabilidade descritos por EBERHART e RUSSELL (1), para cada genótipo.

A regressão linear segmentada, para cada genótipo, foi obtida, no presente trabalho, de acordo com o seguinte modelo (3):

$$Y_{ij} = B_0 + B_1 X_j + B_2 (X_j - X_p) Z_j + \epsilon_{ij}$$

em que

Y_{ij} representa a produção média, em toneladas/ha, do genótipo i no ambiente j ($i=1,2, \dots, v$) ($j=1,2,\dots,n$);

B_0 representa o intercepto da regressão linear segmentada, estimado por b_0 ;

B_1 representa o coeficiente de regressão do primeiro segmento da regressão linear segmentada (resposta em ambientes desfavoráveis), estimado por b_1 ;

B_2 representa a modificação do B_1 no segmento da regressão linear segmentada, estimado por b_2 ;

$B_1 + B_2$ representa o coeficiente de regressão do segundo segmento da regressão linear segmentada.

- K_j são linear segmentada (resposta em ambientes favoráveis);
 X_p representa a variável independente, dada, no presente trabalho, pela média dos genótipos no ambiente j ($j=1,2,\dots,43$), em toneladas/ha;
- X_p representa o ponto de interseção dos dois segmentos da regressão linear segmentada. No presente trabalho, X_p foi dado pela média de todos os n genótipos em todos os n ambientes, em toneladas/ha;
- Z_j representa um índice cujo valor foi definido por:

$$Z_j = 1 \text{ se } X_j > X_p \text{ e}$$

$$Z_j = 0 \text{ se } X_j \leq X_p,$$

ϵ_{ij} representando a falta de ajuste.

O esquema da análise de variância, para a regressão linear segmentada, é apresentado no Quadro 2.

Para a classificação dos genótipos, com base nos coeficientes de regressão, em relação aos dois subgrupos de ambientes, utilizou-se a sugestão apresentada por VERMA *et alii* (6), conforme o Quadro 3, a qual apresenta nove categorias. As categorias (II) e (III), associadas a alta produtividade média e $s^2d_j = 0$, caracterizam os genótipos ideais.

QUADRO 2 - Esquema da análise de variância da regressão linear segmentada

F.V.	g.1.	SQ	QM
Devida à reg. ($b_1, b_2/b_0$)	2	SQR	QMR
(b_1/b_0)	1	SQX	QMX
($b_2/b_0, b_1$)	1	SQZ	QMZ
Indep. da regressão	(n-3)=40	SQD	QMD
Erro puro	n(r-1)(v-1)=688	SQE	QME

Testes das hipóteses $B_1 = 1,0$ e $(B_1 + B_2) = 1,0$

A hipótese de que as estimativas dos coeficientes de regressão, B_1 e $(B_1 + B_2)$, para cada genótipo, não diferem da unidade foi testada pelo teste t:

$$t = \frac{b_1 - 1,0}{s(b_1)}, \text{ com } n(r-1)(v-1) \text{ g.1. do erro puro.}$$

$$t = \frac{(b_1 + b_2) - 1,0}{s(b_1 + b_2)}, \text{ com } n(r-1)(v-1) \text{ g.1. do erro puro.}$$

$s(b_1)$ e $s(b_1 + b_2)$ foram obtidos da matriz das variâncias e covariâncias de B.

Teste da hipótese $B_2 = 0$.

A contribuição da utilização da regressão linear segmentada, em substituição

QUADRO 3 - Classificação dos genótipos com base nos coeficientes de regressão, em relação aos dois subgrupos de ambientes, segundo VERMA *et alii* (6)

Categoría	Ambiente desfavorável (b_1)	Ambiente favorável ($b_1 + b_2$)
(i)	< 1,0	< 1,0
(ii)		= 1,0
(iii)		> 1,0
(iv)	= 1,0	< 1,0
(v)		= 1,0
(vi)		> 1,0
(vii)	> 1,0	< 1,0
(viii)		= 1,0
(ix)		> 1,0

à regressão linear simples, com inclinação única para todos os ambientes, foi verificada através do teste F (Quadros 1 e 2):

$$F = \frac{rQMZ}{QME}, \text{ com } 1 \text{ e } n(r-1)(v-1) \text{ g.l. do erro puro.}$$

Obtenção e teste do valor de s^2d_i

As estimativas dos desvios da regressão (s^2d_i) foram obtidas do seguinte modo (Quadro 1 e 2):

$$s^2d_i = \frac{SQD}{(n-3)} - \frac{QME}{r}, \text{ sendo } (n-3) \text{ o g.l. independente da regressão.}$$

A hipótese de que $s^2d_i = 0$ foi testada através do teste F, como segue:

$$F = \frac{rQMD}{QME}, \text{ com } n-3 \text{ e } n(r-1)(v-1) \text{ g.l..}$$

3. RESULTADOS E DISCUSSÃO

A análise conjunta preliminar dos 43 ambientes é apresentada no Quadro 4. Observou-se o efeito da interação genótipo x ambiente, pelo teste F, a 1% de probabilidade. Assim, efetuou-se o estudo do «comportamento» de cada genótipo nos 43 ambientes.

Nos Quadros 5 e 6 estão as estimativas dos coeficientes de regressão e dos desvios da regressão, com os respectivos testes estatísticos, para a regressão linear simples (método de EBERHART e RUSSELL (1)) e para a regressão linear segmentada, com ajuste de dois segmentos de reta, respectivamente.

Observou-se que a utilização da regressão linear segmentada não foi eficaz pa-

QUADRO 4 - Análise conjunta preliminar dos nove genótipos em 43 ambientes e três repetições em cada ambiente

F.V.	g.l.	SQ	QM	F
Blocos/ambiente	86	32,20319	0,3744556	3,63 **
Genótipo	8	28,88753	3,6109410	34,97 **
Ambiente	42	307,32130	7,3171730	70,87 **
Genótipo x ambiente	336	66,43492	0,1977230	1,91 **
Erro puro	688	71,03749	0,1032521	
Total	1160			

** Significativo, a 1% de probabilidade.

ra diminuir o desvio de regressão, se comparada com a regressão linear simples.

O valor de b_2 foi diferente de zero, pelo teste de F, a 5% de probabilidade, somente para os genótipos UFV 80-69, UFV 79-56 e UFV 79-58. O genótipo UFV 79-58 (Figura 1) apresentou «comportamento» do tipo (IV), ou seja, $b_1 = 1,0$ e $(b_1 + b_2) < 1,0$, demonstrando ser um genótipo de «comportamento» indesejável para uma faixa de ambientes que inclua o subgrupo de ambientes favoráveis, uma vez que não responde à melhoria de ambiente, dentro desse subgrupo, além de apresentar baixa produtividade média e $s^2 d_i > 0$. O «comportamento» do tipo (IV) só pôde ser detectado com a utilização da regressão linear segmentada. Os genótipos UFV 79-56 e UFV 80-69 apresentaram «comportamento» do tipo (V), o que também não correspondeu ao do genótipo teoricamente ideal, além de terem apresentado baixa produtividade média.

Os demais genótipos podem ser estudados por um ou outro método, uma vez que o valor de b_2 não foi significativo, a 5% de probabilidade.

Os genótipos teoricamente ideais, segundo VERMA *et alii* (6), são os dos tipos (II) e (III), que apresentam $b_1 < 1,0$ e $(b_1 + b_2) \geq 1,0$, associados a altas produtividades médias e $s^2 d_i = 0$. Esses tipos não foram encontrados no presente trabalho.

4. RESUMO E CONCLUSÕES

O modelo de regressão linear segmentada, $Y_{ij} = B_0 + B_1 X_j + B_2(X_j - X_p)Z_j + \epsilon_{ij}$, sendo Y_{ij} a média do i-ésimo genótipo no j-ésimo ambiente; B_0 , o intercepto em Y; B_1 e $(B_1 + B_2)$, as inclinações das duas linhas de regressão: X, a média dos genótipos no j-ésimo ambiente, representando o eixo da variável independente; X_p , a média de todos os genótipos em todos os ambientes, definindo o ponto onde há mudança na inclinação da reta; Z_j , um índice cujo valor foi definido por $Z_j = 1$, se $X_j > X_p$, e $Z_j = 0$, se $X_j \leq X_p$, define parâmetros de estabilidade que podem ser usados para descrever o «comportamento» do i-ésimo genótipo nos ambientes estudados.

Os dados de nove genótipos de soja, testados em 43 ambientes, no Estado de Minas Gerais, indicaram que o uso da regressão linear segmentada permitiu a identificação de genótipos com diferentes «comportamentos» nos ambientes desfavoráveis e favoráveis. Porém, o uso da regressão linear segmentada não reduziu o desvio da regressão, se comparada com a regressão linear simples.

O genótipo UFV 79-58 apresentou «comportamento» do tipo (IV), com baixa

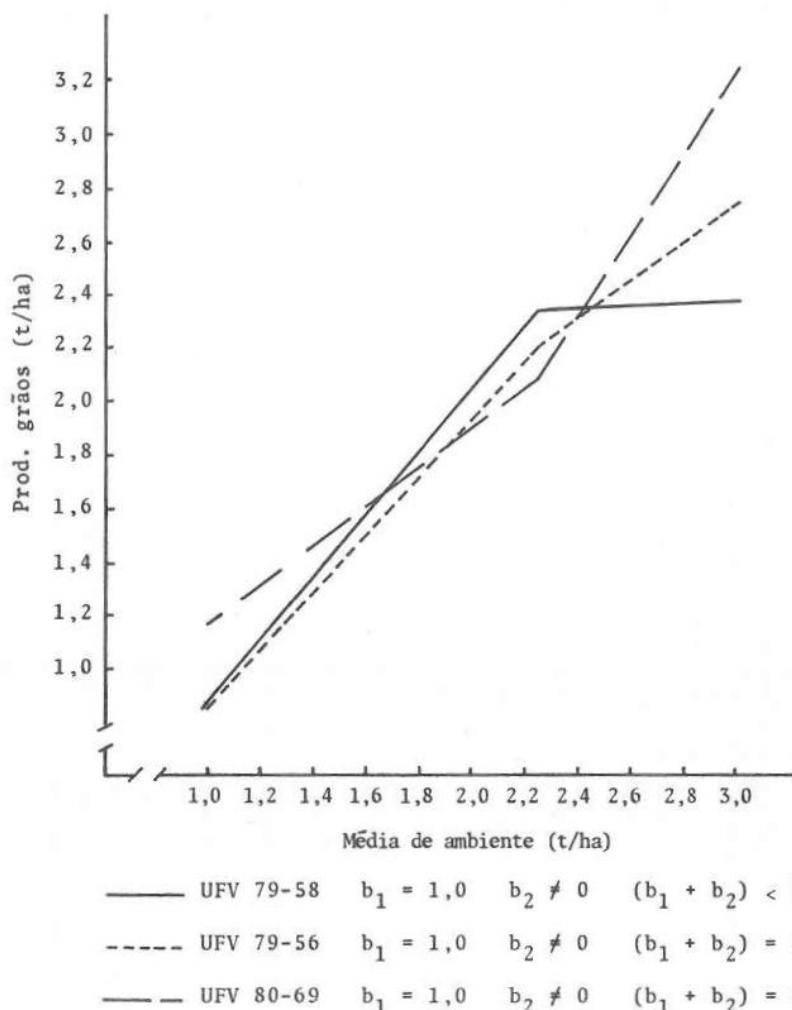
QUADRO 5 - Estimativas das produções de grãos, em Kg/ha, dos coeficientes de regressão (b_i) e dos desvios da regressão ($s^2_{d_i}$), pela regressão linear simples (1)

Genótipo	Produção Kg/ha	b_i	$t(\alpha/2, 688)$	$s^2_{d_i}$	F(41,688)
			Ho: $b_i = 0$	$s^2_{d_i} = 0$	
Cristalina	2544,7	1,2696	2,831**	0,007621	1,308ns
Juparaná	2390,7	1,1577	1,656ns	0,016392	1,476*
Rio Doce	2360,8	1,0584	0,613ns	0,035166	2,022**
UFV-1	2306,9	1,0568	0,596ns	-0,00924	0,732ns
Sucupira	2232,3	1,0443	0,465ns	0,007868	1,228ns
UFV 80-69	2216,2	1,0199	0,209ns	0,009204	1,267ns
UFV 79-56	2115,5	0,8721	-1,343ns	0,003191	1,093ns
UFV 79-58	2104,7	0,6626	-3,542**	0,089340	3,596**
IAC-2	2002,6	0,8587	-1,484ns	-0,003422	0,900ns

** significativo, a 1% de probabilidade, pelo teste de t ou F.
 * significativo, a 5% de probabilidade, pelo teste de t ou F.

QUADRO 6 - Estimativas da produção de grãos, em Kg/ha, dos coeficientes de regressão (b_1 , b_2 e b_1+b_2) e dos desvios da regressão ($s^2 d_i$), pela regressão linear segmentada

Genótipo	Produção Kg/ha	b_{1i}	$t(\alpha/2, 688)$ $Ho: B_{1i}=0$	b_{2i}	$F(1, 688)$ $Ho: B_{2i}=0$	b_1+b_2	$t(\alpha/2, 688)$ $Ho: B_1+B_2=1$	$s^2 d_i$	$F(40, 688)$ $Ho: s^2 d_i=0$
Cristalina	2544,7	1,1868	1,046ns	0,1836	0,895ns	1,3704	1,784ns	0,010977	1,319ns
Juparaná	2390,7	1,0216	0,121ns	0,3019	2,418ns	1,3235	1,558ns	0,015582	1,453*
Rio Doce	2360,8	1,0761	0,426ns	-0,0393	0,041ns	1,0368	0,177ns	0,036870	2,071**
UFV-1	2306,9	1,0064	0,036ns	0,1118	0,331ns	1,1182	0,569ns	-0,008879	0,742ns
Sucupira	2232,3	0,9303	-0,390ns	0,2529	1,697ns	1,1832	0,882ns	0,007465	1,217ns
UFV 80-69	2216,2	0,748	-1,406ns	0,6013	9,593**	1,3501	1,686ns	0,002040	1,059ns
UFV 79-56	2115,5	1,0702	0,393ns	-0,4393	5,121*	0,6309	-1,777ns	-0,000275	0,992ns
UFV 79-58	2104,7	1,1657	0,928ns	-1,1160	33,043**	0,0497	-4,576**	0,064002	2,860**
IAC-2	2002,6	0,7942	-1,152ns	0,1431	0,543ns	0,9373	-0,302ns	-0,003115	0,910ns



produtividade média e $s^2 d_i > 0$, demonstrando ser um genótipo indesejável. Esse «comportamento» do tipo (IV) só pode ser detectado com a utilização da regressão linear segmentada.

Genótipos teoricamente ideais, tipos (II) e (III), não foram encontrados no presente trabalho.

5. SUMMARY

(A STUDY OF FENOTYPE STABILITY OF NINE SOYBEAN GENOTYPES, BY MEANS OF PIECEWISE LINEAR REGRESSION ANALYSIS)

The piecewise linear regression model, $Y_{ij} = B_0 + B_1 X_j + B_2 (X_j - X_p)Z_j + \epsilon_{ij}$, where Y_{ij} is the variety yield at the j th environment, B_0 is the Y intercept, B_1 and $(B_1 + B_2)$ are the slopes of the two regression lines, X_j is the j th environment value represented by the overall mean at a given environment, X_p is the overall mean (point where the slope changes), and Z_j is the indicator variable defined as follows: $Z_j = 1$ if $X_j > X_p$ and $Z_j = 0$ if $X_j \leq X_p$, defines stability parameters that may be used to describe the performance of the i th variety over all environments.

The data from nine varieties tested at forty-three environments indicated that the use of piecewise linear regression allowed identification of varieties with different performances over unfavorable and favorable environments. However, it was not effective to reduce the deviation from regression, as compared to the use of simple linear regression.

The UFV 79-58 variety presented type (iv) performance, with small mean and significant deviation from regression, indicating that it is undesirable. This performance type (iv) could be detected only through piecewise linear regression analysis.

The theoretically ideal varieties, that is, type (ii) and (iii), were not found in this study.

6. LITERATURA CITADA

1. EBERHART, S.A. & RUSSELL, W.A. Stability parameters for comparing varieties. *Crop Sci.*, 6(36-40), 1966.
2. FINLAY, K.W. & WILKINSON, G.N. The analysis of adaptation in a plant-breeding programme. *Aust. J. Agric. Res.*, 14(742-754), 1963.
3. NETTER, J. & WASSERMAN, W. *Applied linear statistical models*. Illinois, Richard D. Irwin, Inc., 1974, 842 p.
4. PERKINS, J.M. & JINKS, J.L. Environmental and genotype-environmental components of variability. III. Multiple lines and crosses. *Heredity*, 23(339-356), 1968.
5. PERKINS, J.M. & JINKS, J.L. Environmental and genotype-environmental components of variability. IV. Non-linear interactions for multiple inbred lines. *Heredity*, 23(525-535), 1968.
6. VERMA, M.M., CHAHAL, G.S. & MURTY, B.R. Limitations of conventional regression analysis: a proposed modification. *Theor. Appl. Genet.*, 53(89-91), 1978.