2-010

Análise de um sistema agroflorestal usando estatística multivariada

Antonio Claudio Almeida de CARVALHO[1], Hilton Thadeu Zarate do COUTO[2], Aderaldo Batista da Silva da Silva GAZEL[3]

[1] Engº Agrônomo, M.Sc. em Estatística e Experimentação Agronômica, pesquisador da Embrapa Amapá a disposição do Instituto de Pesquisas Científicas e Tecnológicas do Estado do Amapá. – IEPA, acacarvalho@uol.com.br . [2] Engº Florestal Ph.D, livre-docente da ESALQ/USP. [3]Engº Agrônomo, M.Sc. em Sistemas Agroflorestais, pesquisador da Embrapa Amapá.

Introdução

Os cultivos consorciados onde são cultivados no mesmo espaço, de forma simultânea ou seqüencial, culturas anuais e/ou fruteiras com espécies florestais, com a presença ou não de animais, receberam a denominação específica de sistemas agroflorestais, ou simplesmente, $SAF's$. Nas regiões tropicais, parece óbvio as vantagens desses sistemas em relação às outras formas de cultivo, principalmente no que tange a sua sustentabilidade a longo prazo. Não obstante, as análises quantitativas dos efeitos diretos das culturas componentes dos $SAF's$, bem como das associações existentes entre essas culturas, são fundamentais para uma melhor compreensão e sistematização desses sistemas de cultivo. Entre os componentes dos $SAF's$, há interações biológicas, econômicas e ambientais, entre outras, o que torna complexo as análises e interpretações dos mesmos.

A forma mais apropriada para análise de um sistema agroflorestal é através do uso das técnicas da estatísticas multivariadas, uma vez que sempre há mais que uma variável de interesse sendo abordada. Não obstante, o tipo de analise estatística mais freqüentemente realizado nos sistemas consorciados é através da transformação do problema multivariado em um problema univariado através da obtenção de uma única variável como função resposta das observações das culturas componentes e, a partir da mesma, aplicam-se os testes estatísticos comumente empregados na análise de dados agronômicos.

Como o objetivo de demonstrar como é feita a aplicação dos testes estatísticos multivariados na comparação dos tratamentos de um experimento envolvendo diferentes arranjos entre as componentes de um sistema agroflorestal, iremos apresentar a seguir, de uma forma resumida, a dedução teórica da análise multivariada com posterior aplicação à análise de dados.

Metodologia

Nos experimentos envolvendo monocultivos, cada parcela fornece resposta apenas de um componente, enquanto que nos $SAF's$, o número de variáveis respostas é igual ou superior ao número de culturas componentes. Suponha um experimento com I tratamentos formados por diferentes $SAF's$, instalado em blocos ao acaso, onde cada tratamento contém K variáveis observadas por parcela.. Numa notação genérica, esse experimento pode ser assim representado:

TABELA 1. Representação genéricas de um ensaio de deferentes $SAF's$ instalado em blocos ao acaso

TRAT.	$Bloco_1$				$Bloco_2$				$\cdots$	$Bloco_J$				TOTAL			
SAF_1	$y_{11}^{(1)}$	$y_{11}^{(2)}$	$\cdots$	$y_{11}^{(K)}$	$y_{12}^{(1)}$	$y_{12}^{(2)}$	$\cdots$	$y_{12}^{(K)}$	$\cdots$	$y_{1J}^{(1)}$	$y_{1J}^{(2)}$	$\cdots$	$y_{12}^{(K)}$	$y_{1\bullet}^{(1)}$	$y_{1\bullet}^{(2)}$	$\cdots$	$y_{1\bullet}^{(K)}$
SAF_2	$y_{21}^{(1)}$	$y_{21}^{(2)}$	$\cdots$	$y_{21}^{(K)}$	$y_{22}^{(1)}$	$y_{22}^{(2)}$	$\cdots$	$y_{22}^{(K)}$	$\cdots$	$y_{2J}^{(1)}$	$y_{2J}^{(2)}$	$\cdots$	$y_{22}^{(K)}$	$y_{2\bullet}^{(1)}$	$y_{2\bullet}^{(1)}$	$\cdots$	$y_{2\bullet}^{(K)}$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
SAF_I	$y_{I1}^{(1)}$	$y_{I1}^{(2)}$	$\cdots$	$y_{I1}^{(K)}$	$y_{I2}^{(1)}$	$y_{I2}^{(2)}$	$\cdots$	$y_{I2}^{(K)}$	$\cdots$	$y_{IJ}^{(1)}$	$y_{IJ}^{(2)}$	$\cdots$	$y_{IJ}^{(K)}$	$y_{I\bullet}^{(1)}$	$y_{I\bullet}^{(2)}$	$\cdots$	$y_{I\bullet}^{(K)}$
TOTAL	$y_{\bullet 1}^{(1)}$	$y_{\bullet 1}^{(2)}$	$\cdots$	$y_{\bullet 1}^{(K)}$	$y_{\bullet 2}^{(1)}$	$y_{\bullet 2}^{(2)}$	$\cdots$	$y_{\bullet 2}^{(K)}$	$\cdots$	$y_{\bullet J}^{(1)}$	$y_{\bullet J}^{(2)}$	$\cdots$	$y_{\bullet 1}^{(K)}$	$y_{\bullet\bullet}^{(1)}$	$y_{\bullet\bullet}^{(2)}$	$\cdots$	$y_{\bullet\bullet}^{(K)}$

$y_{ij}^{(k)}$ é o valor observado na k-ésima variável (k = 1, 2, ⋯, K), no i-ésimo tratamento (i = 1,2,⋯, I) e no j-ésimo bloco (j = 1,2,⋯J)

O modelo matemático de um experimento multivariado, instalado em blocos ao acaso, é dado por:

$$y_{ij} = \mu + \alpha_i + \beta_j + e_{ij}$$

onde :

μ *é um vetor de constantes, representadas pelas médias teóricas dos tratamentos;*

α_i *é o vetor que representa os valores das variáveis no i-ésimo tratamento;*

β_j *é o vetor que representa os valores das variáveis no j-ésimo blocos;*

e_{ij} *é o vetor que representam os erros aleatórios com distribuição normal K-variada.*

As hipóteses a serem testadas no respectivo experimento são:

$$H_0 : \alpha_1 = \alpha_2 = \cdots \alpha_I \quad versus \quad H_A : \alpha_i \neq \alpha_{i'} \quad \Rightarrow \quad para \ \forall \ i \neq i'$$

O quadro de Análise de Variância Multivariada (MANOVA) é dada por:

Fonte de Variação	Grau de Liberdade	Soma de Quadrado e Soma de Produtos
Tratamento	$I-1$	Matriz H
Bloco	$J-1$	Matriz B
Resíduo	$(I-1)(J-1)$	Matriz E
Total	$IJ-1$	Matriz T

De posse das matrizes H, B, E e T, podemos realizar os testes sobre as hipóteses formuladas no experimento. Existem vários critérios para testar as hipóteses H_0 e H_A na análise de variância multivariada, tais como: critério do traço de Hotelling-Lawley, critério do traço de Pillai e pelo critério de Wilks, ente outros. Neste trabalho veremos apenas o critério de Wilks, por sua analogia como a analise de variância univariada. A estatística de Wilks é baseada no teste de razão de verossimilhança dado por: $\lambda = \frac{|E|}{|H+E|}$. Em que $|E|$ representa o determinante da matriz de erros e $|H+E|$ representa o determinante da matriz formada pelas hipóteses a serem testadas (tratamento) acrescida da matriz de erros. A variável λ tem distribuição Wilks e o valor calculado é comparado com valor crítico de $\lambda_{\{k,\ \nu_H,\ \nu_E);\ (\alpha)\}}$ que é encontrado na tabela de Wilks, para $\alpha = 0{,}05$ ou $\alpha = 0{,}01$. Não obstante, o mais comum é aplicar uma transformação na variável λ e utilizar os valores tabelados da própria distribuição F, usada na estatística univariada. Essa transformação é feita como se segue:

TABELA 2. Expressões usadas para transformação da estatística λ em valores da distribuição F

Parâmetros de λ		Expressão para obtenção dos valores do F calculado	Grau de Liberdade da Tabela F	
Nº de Variáveis	g. l. do tratamento		Numerador	Denominador
k	$\nu_H = 1$	$\frac{1-\lambda}{\lambda} \cdot \frac{\nu_E + \nu_H - k}{k}$	k	$(\nu_E + \nu_H - k)$
k	$\nu_H = 2$	$\frac{1-\sqrt{\lambda}}{\sqrt{\lambda}} \cdot \frac{\nu_E + \nu_H - k - 1}{k}$	$2k$	$2(\nu_E + \nu_H - k - 1)$
$k = 1$	ν_H	$\frac{1-\lambda}{\lambda} \cdot \frac{\nu_E}{\nu_H}$	ν_H	ν_E
$k = 2$	ν_H	$\frac{1-\sqrt{\lambda}}{\sqrt{\lambda}} \cdot \frac{\nu_E - 1}{\nu_H}$	$2\nu_H$	$2(\nu_E - 1)$

$\nu_H =$ *número de graus de liberdade de tratamento* e $\nu_E =$ *número de graus de liberdade do erro*

A hipótese H_0 será rejeitada se o valor da estatística λ ou F calculado for igual ou superior ao valor critico tabelado. Rejeitando-se H_0, automaticamente aceita-se H_A, ou seja, existe diferenças entre os tratamentos, quando são analisados todas as variáveis ao mesmo tempo. Dessa forma, o passo seguinte da análise é verificar quais são os tratamentos que diferem entre-se, através da comparação dos contrastes ortogonais. Por questão de espaço, não serão demonstrados aqui a teoria desses contrastes ortogonais.

Resultado e discussão

Demostraremos a seguir, através de um exemplo com dados simulados a realização da analise estatística multivariada de um experimento de avaliação de $SAF's$

TABELA 3. Dados simulados de um experimento de avaliação de $SAF's$, instalado em blocos ao acaso, cujas as variáveis analisadas são as produções anuais de cacau e seringueira (Kg/ha).

Tratamento	BLOCO I		BLOCO II		BLOCO III		BLOCO IV		TOTAL	
	Seringueira	cacau	Seringueira	cacau	Seringueira	cacau	Seringueira	Cacau	Seringueira	Cacau
SAF_1	257,51	1302,83	262,79	1315,63	263,07	1321,58	270,52	392,99	1053,89	5333,03
SAF_2	262,13	1002,12	249,87	1023,40	273,33	1032,25	253,07	959,83	1038,40	4017,60
SAF_3	322,18	1771,26	318,95	1820,37	322,67	1776,27	315,65	705,26	1279,45	7073,16
SAF_4	261,80	941,15	270,79	947,01	264,64	930,91	272,73	876,32	1069,96	3695,39
Total	1103,62	5017,36	1102,40	5106,41	1123,71	5061,01	1111,97	1934,40	4441,70	20119,18

SAF_1 = SERINGUEIRA IAN 717 + CACAU CATONGO
SAF_2 = SERINGUEIRA IAN 717 + CACAU SIAL 70
SAF_3 = SERINGUEIRA RRIM 600 + CACAU CATONGO
SAF_4 = SERINGUEIRA RRIM 600 + CACAU SIAL 70

O primeiro passo para análise de variância multivariada é obter as matrizes H, B, E e T. Abaixo, encontra-se demonstrado como é obtida a matriz T e as demais matrizes, que são obtidas de forma análoga, serão apenas descritas.

$$T = \begin{cases} t_{11} = (257{,}51)^2 + \cdots + (272{,}73)^2 - (4441{,}70)^2 / 16 = 10176{,}42 \\ t_{22} = (1302{,}83)^2 + \cdots + (876{,}32)^2 - (20119{,}18)^2 / 16 = 1786099{,}28 \\ t_{12} = t_{21} = 257{,}51 \times 1302{,}83 + \cdots + 272{,}73 \times 876{,}32 - (4441{,}70 \times 20119{,}18) / 16 = 114842{,}10 \end{cases}$$

$$T = \begin{bmatrix} 10176{,}42 & 114842{,}10 \\ 114842{,}10 & 1786099{,}28 \end{bmatrix}; \quad H = \begin{bmatrix} 9647{,}67 & 113784{,}11 \\ 113784{,}11 & 1768116{,}85 \end{bmatrix}; \quad B = \begin{bmatrix} 72{,}37 & -65{,}73 \\ -65{,}73 & 4024{,}77 \end{bmatrix}; \quad E = \begin{bmatrix} 456{,}35 & 1123{,}72 \\ 1123{,}72 & 13946{,}66 \end{bmatrix}$$

$$\lambda = \frac{|E|}{|H+E|} = \frac{5101868{,}42}{4802201403{,}04} = 0{,}0011 \quad \Rightarrow \quad F = \frac{1-\sqrt{\lambda}}{\sqrt{\lambda}} \cdot \frac{\nu_E - 1}{\nu_H} = \frac{1-\sqrt{0{,}0011}}{\sqrt{0{,}0011}} \cdot \frac{9-1}{3} = 79{,}15$$

Através da tabela da distribuição F, encontramos $F_{\{2\nu_H;\ 2(\nu_E - 1);\ \alpha\}} = F_{\{6;\ 16;\ 0{,}01\}} = 4{,}40$. Dessa forma, como o valor calculado é maior que o valor crítico tabelado, conclui-se que há diferença significativa entre os sistemas agroflorestais testados, ao nível de significância de 1%. Como o teste F foi significativo, devemos prosseguir a análise para saber quais tratamentos diferem entre-se. O teste entre os tratamentos pode ser feito a partir do desdobramento dos graus de liberdade de tratamento., através do qual, podemos testar os $(t-1)$ contrastes ortogonais possíveis. Para o exemplo em estudo, teremos $4-1=3$ contrastes ortogonais a serem testados. Considerando que o interesse seja comparar os dois primeiros $SAF's$ contra os dois últimos, ou seja, os que têm o clone de seringueira IAN 717 contra os que têm o clone de seringueira RRIM 600 e em seguida fazer as comparações entre os dois primeiros e os dois últimos, temos os seguintes contrastes ortogonais:

TABELA 4. Contrastes ortogonais, estimativas dos contrastes e aplicação do teste de Wilks

Contraste	SAF_1 *estimado*		SAF_2 *estimado*		SAF_3 *estimado*		SAF_4 *estimado*		*CONTRASTE estimado*		Significância
	seringueira 1053,89	cacau 53333,03	seringueira 1038,40	cacau 4017,60	Seringueira 1279,45	cacau 7073,16	seringueira 1069,96	cacau 3695,39	Seringueira	Cacau	
W_1	-1		-1		+1		+1		[257,12	1417,92]	**
W_2	-1		+1		0		0		[−15,49	−1315,43]	**
W_3	0		0		-1		+1		[−209,49	−377,77]	**

** indica que o contraste é significativo ao nível de 1% de probabilidade pelo teste Wilks.

Conclusão

Conforme podemos observar, a análise de sistemas agroflorestais através da estatística multivariada é muito eficaz pois os resultados obtidos levam em consideração todas as correlações possíveis entre as variáveis. No exemplo dado acima, verifica-se que tanto para a produção de borracha, quanto para produção de cacau, os SAF_3 e SAF_4 são superiores ao aos SAF_1 e SAF_2. Analisando os dois primeiros Sistemas, verificamos que, em termos de produção de borracha e produção de cacau, o SAF_1 é superior ao SAF_2. De forma análoga, verificamos que o SAF_3 é superior ao SAF_4.