

USO DE MÉTODOS MULTIVARIADOS NA AVALIAÇÃO DO ESTRESSE HÍDRICO EM CULTIVARES DE SOJA

USE OF MULTIVARIATE METHODS IN THE EVALUATION OF WATER STRESS IN SOYBEAN CULTIVARS

OLIVEIRA, M.C.N. DE¹, SANTOS, M.R.G.² NEPOMUCENO, A.L.¹, NEUMAIER, N.¹, FARIAS, J.R.B.¹.

¹Embrapa Soja, Caixa Postal 231, Londrina-Pr, e-mail:mcno@cnpso.embrapa.br,

²UEL-LONDRINA/Embrapa Soja

Resumo -A soja é uma oleaginosa de grande importância para o Brasil e com a população aumentando muitas pesquisas são necessárias para atender a demanda da população mundial. Para garantir produções ótimas é necessário solucionar um dos maiores problemas a médio prazo quanto às questões de temperatura e clima. Dentre os fatores a serem estudados isolados ou em conjunto estão a temperatura, o excesso ou a falta de água, e alta concentração salina nos solos. Nestas pesquisas existe um grande volume dados e para redução dimensional é importante utilizar métodos estatísticos que permitam com poucas variáveis entender o enfoque da pesquisa. Os métodos multivariados são apropriados para melhor interpretar os resultados com poucas variáveis sem perder informação. Desta forma foram utilizados os métodos de agrupamento (k-médias) e o método componente principal. Com estes métodos foi possível obter nove grupos e detectar dentre os materiais convencionais e geneticamente modificados com características de tolerâncias ao déficit hídrico dentro de cada grupo de tratamentos com as variáveis estudadas.

Introdução

A água é de grande importância no desenvolvimento da planta de soja e a sua produtividade dessa cultura está relacionada com vários fatores ambientais. Dentre os fatores a serem estudados isolados ou em conjunto estão a temperatura, o excesso ou a falta de água, e alta concentração salina nos solos (Casagrande et al., 2001; Aragão, 2010). O trabalho terá o enfoque na avaliação de cultivares que suportam o estresse hídrico e as variáveis selecionadas foram às agrônômicas e fisiológicas e aquelas relacionadas com expressão gênica serão comparadas posteriormente. Nestas pesquisas, em geral, tem-se um grande volume de variáveis para que o estudo seja completo, desta forma, sugere-se o método multivariado (Reis, 2001, Mingoti, 2005) para a tomada de decisão na escolha do cultivar que tenha ótimo desenvolvimento nos períodos de ausência de chuvas e se possível maior rendimento. Esse método permite reduzir a dimensão da matriz de dados e com poucas variáveis e, sem perder informação, interpretar o fenômeno em estudo. O objetivo deste trabalho foi agrupar pelo método multivariado de Componentes Principais e Análise de Cluster cultivares de soja dentro de cada nível de disponibilidade hídrica.

Palavras chave: escassez de água, temperatura, componentes principais, agrupamentos

Materiais e métodos

Os dados utilizados neste trabalho fazem parte do projeto “Aprimoramento tecnológico para redução de perdas de soja por seca em sistemas agrícolas sustentáveis” conduzido no ano agrícola 2006/2007. Foram instalados no campo experimental da Fazenda Santa Terezinha, na Embrapa Soja, Londrina, PR. O delineamento utilizado foi o de blocos ao acaso com parcelas subdivididas: em parcelas três níveis de disponibilidade hídrica (estresse no reprodutivo, irrigado e não irrigado) e, em subparcelas 10 cultivares de soja (BR 16, Embrapa 48, BRS 133, BRS 134, BRS 183, BRS 184, BRS 214, BRS 232, BRS245RR, BRS247RR) e quatro blocos. As variáveis estudadas foram: (PFPA) Peso fresco da parte aérea (g), (PSFOL) Peso seco das folhas (g), (PSCP) Pesos seco dos caules + pecíolos (g), (PSVCS) Peso seco das vagens com sementes (g), (PSPA) Peso seco da parte aérea Total (g), (PSS) Peso Seco das Sementes (g), (NS) Número de sementes (g), (PSCS) Peso seco de Cem sementes (g), (RPU) Rendimento em (kg/ha), (FOTO) Taxa Fotossintética ($\text{micromol m}^{-2} \text{s}^{-1}$), (RESTOM)

Resistência Estomática (s.cm⁻¹), (CINT) Concentração Intercelular de CO₂ (ppm), (TRH₂O R₁), Teor relativo de Água (%), (ÓLEO), Teor de óleo (%), (PROT) Teor de Proteína (%), (DIFTEM) Temperatura Ambiente - Temperatura Foliar (°C). O modelo desse delineamento será o que segue:

$$Y_{ijk} = m + \text{Disp Hidrica}_i + \text{Bloco}_j + \text{DispHidr}_i * \text{Bloco} + \text{Cultivares}_{jk} + \text{DispHid} * \text{Culti}_{ik} + \varepsilon_{ijk}$$

com $i = 1, \dots, I; j = 1, \dots, J; k = 1, \dots, K;$

em que,

Y_{ijk} é o efeito da variável resposta da disponibilidade hídrica i , no bloco j no cultivar de soja k ;

m é o efeito da média geral do experimento;

Disp. Hídrica _{i} é o efeito da disponibilidade hídrica i ;

Bloco _{j} é o efeito do bloco j ;

Culti _{k} é o efeito do cultivar de soja k ;

Disp. Hídrica * Cultivar é o efeito da interação disp. Hídrica e cultivar de soja ik ;

ε_{ijk} é o efeito do resíduo aleatório $N \cong (0, \sigma^2)$ associado a cada unidade experimental.

Foram avaliadas todas as pressuposições da Análise de Variância (ANOVA) como: os coeficientes de assimetria e curtose, a normalidade e independência dos erros, a homogeneidade de variâncias dos tratamentos e a aditividade do modelo. Além das análises univariadas foram aplicados também os métodos multivariados de componentes principais e de agrupamento (cluster analysis) para avaliar dentre as variáveis respostas quais são as que mais influenciam no aumento da produtividade quanto aos fatores da disponibilidade hídrica. Os programas estatísticos utilizados foram o SAS-Statistical Analysis System, versão 9.2 (2001), Sanest para DOS (Zonta et al. 1982) e STATISTICA para Windows, (1995). O esquema da Anova e as variâncias complexas para os tratamentos são apresentados (Cochran & Cox, 1957). Na análise de agrupamento utilizou-se um conjunto de técnicas na identificação de padrões de comportamento em bancos de dados através da formação de grupos homogêneos de casos. O método de partição de k -médias foi aplicado para agrupar dados em grupos homogêneos com os dados padronizados. Esse método objetiva minimizar a soma de quadrados de dentro dos grupos, tomando como base a ANOVA. O número de agrupamentos é avaliado pelas somas de quadrados dentro das partições mantendo a homogeneidade interna dentro dos grupos como descrito abaixo. Seja X_1 o vetor médio de observações de cada variável resposta e admitindo-se a formação de k grupos, a partição da soma de quadrados total será:

$$\sum_{J=1}^K \sum_{i \in G_j} (x_{il} - X_1)^2 = \sum_{J=1}^K n_j (\bar{X}_{j1} - X)^2 + \sum_{J=1}^K \sum_{i \in G_j} (X_{il} - X_{j1})^2 \quad \text{e} \quad SQT(1) = SQE(1) + SQD(1)$$

em que,

SQT(1) é a soma de quadrados total da variável 1;

SQE(1) é a soma de quadrados entre grupos da variável 1;

SQD(1) é a soma de quadrados de dentro da variável 1;

G_j é o conjunto que indica os elementos do grupo j ;

n_j é o número de elementos do grupo j ;

\bar{X}_1 é a média da variável 1;

\bar{X}_{j1} é a média da variável 1 no grupo j .

Considerando todas as variáveis a soma de quadrados da partição será: $SQDP = \sum_{i=1}^P SQD(i)$.

A partição é encerrada quando os valores das somas de quadrados dentro de grupos estão próximas permitindo a elaboração do dendograma. A análise de componentes principais é expressa

como combinações lineares das variáveis para k componentes e p variáveis. Maiores detalhes desse método pode ser encontrado em (Reis, 2001; Mingoti, 2005)

Resultados e Discussão

Foram verificadas as pressuposições das análises de variâncias e tais pré-requisitos foram atendidos permitindo a continuidade às análises de variâncias uni e multivariadas (Shapiro & Wilk, 1965; Burr & Foster, 1972, Tukey, 1949, Cochran & Cox, 1957; Parente, 1984). A maioria das variáveis o teste F para os efeitos de disponibilidade hídrica e cultivares apresentaram resposta significativas com exceção do peso seco de caule + pecíolos tanto para os efeitos de disponibilidade hídrica e cultivares, peso seco de caule + pecíolos na colheita para efeito de disponibilidade hídrica e peso fresco da parte aérea para o efeito de cultivares. Os valores de coeficientes de variação variaram de 2,19 a 118,39. Os valores dos coeficientes de assimetria e curtose estão bem próximos de zero, pré-requisito exigido para que os dados tenham distribuição normal. Além disso, foi também verificada se houve correlação entre as variáveis, necessidade primordial para a aplicação de qualquer método multivariado. Os valores dos coeficientes de correlação de Spearman variaram de -0,05 a 0,95.

Análise de agrupamento

Dos 19 possíveis agrupamentos foram selecionados nove grupos pelo método de k-médias e distância euclidiana com ligação simples. Baseando-se nos valores das somas de quadrados de dentro da partição não foi necessário um número maior de grupos para a interpretação das variáveis quanto ao estresse hídrico e as cultivares, pois houve estabilização do total das soma de quadrados (Tabela 1).

Tabela 1. Somas de quadrados de dentro da partição para as variáveis agronômicas e fisiológicas com 19 agrupamentos.

grupos	pfpá R5	Pfpa col	psfol R5	pscp R5	Pscp col	Psvcs R5	Psvcs col	pspa col	pss col	ns col	pscs col	rpu col	Foto R5	restom R5	cint R5	ditemp	trh2o R5	Óleo col	Prot col	TOTAL
1	29,0	12,7	16,1	28,7	27,0	24,0	8,0	10,3	7,4	11,2	25,8	8,3	2,0	1,9	4,1	7,1	7,6	23,8	13,4	268,2
2	28,0	8,5	14,2	26,2	19,2	19,9	5,5	6,7	5,3	6,2	20,1	8,3	1,3	1,9	3,8	6,6	7,5	23,8	12,9	225,9
3	26,5	8,3	14,1	24,7	18,1	15,0	4,8	5,9	4,6	6,2	14,1	7,1	1,3	1,9	3,6	6,6	6,6	23,4	11,3	203,9
4	27,4	7,7	9,5	15,4	21,6	13,8	5,6	7,5	5,3	6,3	14,9	5,2	1,4	1,9	3,7	5,9	6,8	10,7	11,5	182,3
5	15,6	6,0	12,5	22,2	13,1	9,1	3,9	4,8	3,7	5,8	17,9	6,1	1,2	1,4	3,6	5,4	6,0	11,9	10,1	160,1
6	19,9	6,1	9,0	12,0	18,5	11,4	4,0	5,4	4,0	4,7	11,6	5,2	1,0	1,9	3,4	4,9	5,5	10,0	9,9	148,4
7	11,4	5,4	9,0	11,9	18,4	11,3	4,0	5,4	3,9	4,3	11,6	5,2	0,9	1,9	3,4	4,7	5,4	10,0	9,3	137,4
8	14,4	5,2	8,2	11,3	15,5	11,3	4,0	5,3	3,9	3,6	10,3	3,7	0,9	1,8	3,0	4,3	4,5	9,7	8,7	129,7
9	9,6	4,9	7,4	11,1	10,1	7,6	3,7	4,5	3,5	3,5	10,1	3,5	0,9	1,4	3,0	3,4	4,2	8,9	8,1	109,5
10	7,0	4,0	7,9	15,1	9,1	7,0	3,1	3,9	2,7	3,2	9,8	2,3	0,9	1,3	2,1	4,1	2,1	8,2	7,4	101,4
11	7,0	2,1	6,3	14,0	8,5	6,5	2,1	2,8	2,3	1,9	9,8	2,3	0,9	1,2	1,7	4,0	2,1	6,4	6,6	88,6
12	3,9	2,4	5,7	11,4	7,1	6,3	2,2	2,7	2,5	1,9	10,5	2,8	1,0	1,2	1,6	4,3	2,5	6,2	5,6	81,9
13	0,4	2,4	5,7	11,3	7,0	5,2	2,1	2,6	2,2	1,8	10,5	2,5	1,0	1,0	1,6	4,3	2,5	4,1	3,3	71,3
14	0,4	1,7	5,0	10,9	5,3	5,2	1,8	2,2	2,0	1,3	10,5	2,4	1,0	0,7	1,6	4,2	2,5	3,4	3,3	65,0
15	3,8	3,6	6,5	8,0	5,5	6,6	2,1	2,4	1,7	3,4	7,5	1,7	0,3	1,2	1,2	2,8	2,1	6,7	5,0	72,0
16	3,8	3,5	5,4	6,4	7,0	5,4	2,3	2,8	1,7	3,0	5,7	1,5	0,3	1,2	1,2	3,1	2,1	6,7	5,9	69,1
17	0,2	3,4	5,4	6,3	7,0	4,3	2,1	2,7	1,4	2,8	5,7	1,2	0,2	1,0	1,2	3,1	2,1	4,6	3,6	58,5
18	0,1	3,1	5,8	7,1	6,7	3,8	1,9	2,5	1,3	2,8	5,7	1,1	0,1	1,0	1,2	2,9	1,6	4,9	3,1	56,6
19	0,1	3,1	5,7	7,1	6,7	3,8	1,9	2,4	1,3	2,8	5,5	0,8	0,1	1,0	1,0	2,8	1,3	4,8	3,1	55,1

Gráficos do agrupamento

Pelo gráfico entre as ligações e o número de passos é possível identificar o ponto de corte no dendograma a partir do primeiro salto e será o ponto de corte no dendograma entre os valores de 1,5 a 2,0 no eixo em que estão apresentadas as distâncias de ligação (Figs. 1 e 2)

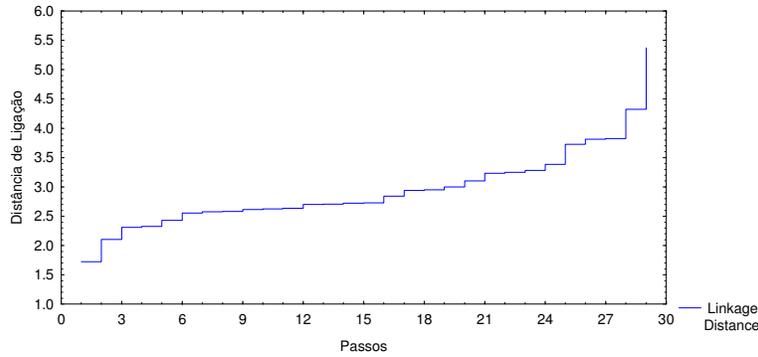


Fig. 1 Gráfico entre as distâncias de ligação e o número de passos

Dendograma

No dendograma realizado pelo método das ligações simples e distância euclidiana foram observados que todos os tratamentos do estresse no estádio reprodutivo estão localizados à direita do gráfico para todos os cultivares e para os tratamentos irrigado e não irrigado à esquerda do gráfico. Avaliando o método das somas de quadrados da partição e o dendograma o número de grupos estão similares.

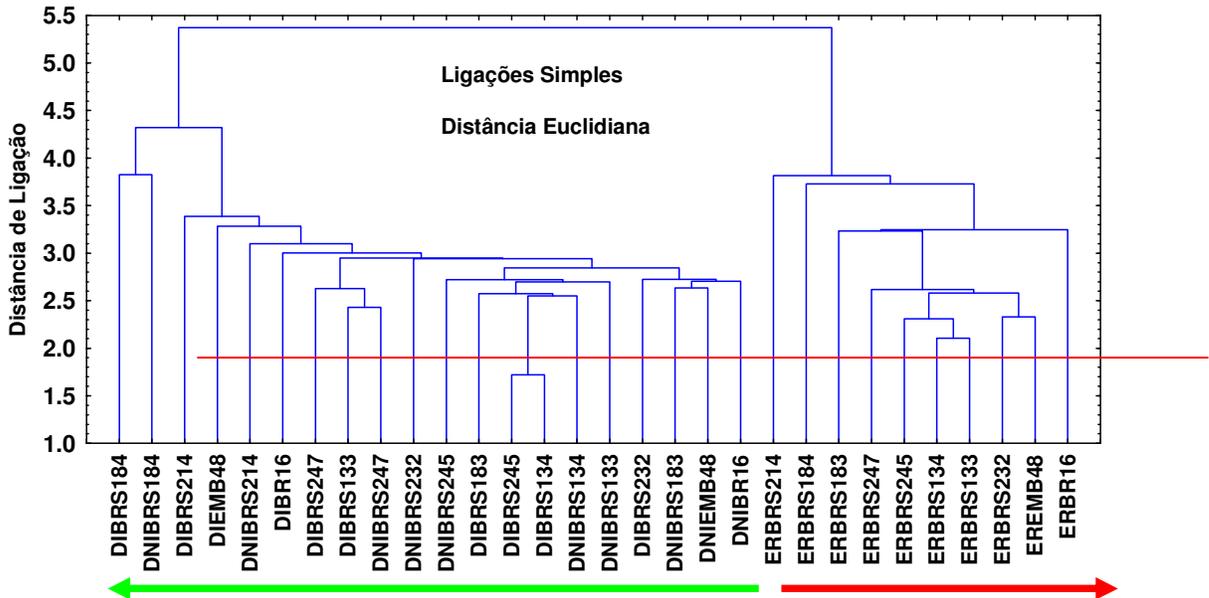


Fig2. Dendograma entre as distâncias de ligação entre os tratamentos pesquisados

Método dos componentes principais

Dentre os 10 componentes avaliados foram selecionados os cinco primeiros que apresentaram autovalores igual ou superior a um e, em consequência maior contribuição da variância explicada pelas variáveis pesquisadas (Tabela 2).

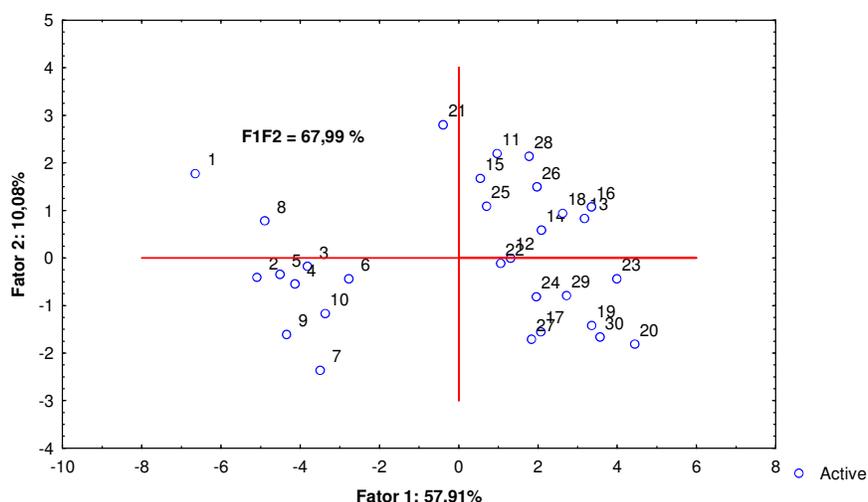
Tabela 2. Valores dos componentes principais considerando as variáveis agrônômicas e fisiológicas da soja.

Componentes	Autovalores	%Total Variância	Autovalor Acumulado	% Acumulado
C1	11,00299	57,91045	11,00299	57,9105
C2	1,91483	10,07807	12,91782	67,9885
C3	1,52734	8,03862	14,44516	76,0271
C4	1,27442	6,70745	15,71957	82,7346
C5	1,09587	5,76775	16,81545	88,5024
C6	0,730638	3,845466	17,546086	92,34782
C7	0,408285	2,148868	17,954371	94,49669
C8	0,328635	1,729657	18,283006	96,22635
C9	0,236961	1,247161	18,519966	97,47351

C10	0,136990	0,720998	18,656956	98,19451
C11	0,124609	0,655836	18,781565	98,85034
C12	0,063166	0,332451	18,844731	99,18279
C13	0,055847	0,293930	18,900577	99,47672
C14	0,047668	0,250884	18,948245	99,72761
C15	0,024124	0,126968	18,972369	99,85458
C16	0,015985	0,084131	18,988354	99,93871
C17	0,008447	0,044459	18,996802	99,98317
C18	0,002998	0,015781	18,999800	99,99895
C19	0,000200	0,001053	19,000000	100,00000

Gráfico entre os componentes principais

Pela Tabela 2 percebe-se que a maior contribuição encontra-se no gráfico entre (F₁F₂) (Fig, 3) com 67,99 %, Considerando os demais fatores F₁F₃ foi 65,95 % F₁F₄ foi de 64,62 % e F₁F₅ foi de 63,48 %, Na Fig, 3 a cultivar BR 16 com o tratamento irrigado ocorreu num grupo distanciado dos demais e bem distante da origem e seu código é 21, Quanto mais distante da origem maior será sua contribuição no estudo, Discussões similares podem ser elaboradas para as demais figuras considerando a legenda dos tratamentos na Tabela 3.



Fig, 3 Gráfico entre o componente 1 e 2

Tabela 3. Contribuição dos tratamentos para avaliação dos cultivares com tolerância ao estresse.

Número da Combinação	Tratamentos	Fator 1	Fator 2	Fator 3	Fator 4	Fator 5
1	ERBR16	13,84499	5,68039	0,45921	1,09437	0,45401
2	DNIEMB48	8,11854	0,29820	0,00067	0,37224	0,48937
3	ERBR133	4,57195	0,05548	1,04278	0,45509	0,00050
4	ERBR134	5,36395	0,54432	4,66433	0,12588	0,00193
5	ERBR183	6,38669	0,21214	8,70357	0,03599	2,35239
6	ERBR184	2,41132	0,34803	0,53557	21,50459	0,02473
7	ERBR214	3,83784	10,06890	13,65280	3,43613	0,37656
8	ERBR232	7,51035	1,10042	0,28385	1,11040	0,25630
9	ERBR245	5,91814	4,65202	1,19143	0,13081	0,60930
10	ERBR247	3,56601	2,47028	0,54638	0,28539	10,2567
11	DNIBR16	0,29272	8,66827	0,73598	0,53879	0,66121
12	DNIEMB48	0,53514	0,00004	2,40834	1,12157	0,54103
13	DNIBR133	3,15239	1,24433	1,76730	0,85411	1,23381
14	DNIBR134	1,36592	0,62196	2,36480	2,75778	0,72814
15	DNIBR183	0,09172	5,03826	0,08263	0,91669	1,27186

16	DNIBRS184	3,51278	2,08335	1,48327	30,10696	3,38061
17	DNIBRS214	1,33794	4,33781	2,33602	1,83677	19,66964
18	DNIBRS232	2,15402	1,58389	14,44796	0,15618	4,94580
19	DNIBRS245	3,52908	3,62993	3,67175	1,98123	5,84066
20	DNIBRS247	6,18344	5,92269	0,90416	0,18027	1,20111
21	DIBR16	0,04845	14,11669	5,11341	0,73258	4,97425
22	DIEMB48	0,35214	0,02315	9,15251	1,47137	17,13402
23	DIBRS133	4,98738	0,34735	0,13433	0,44260	3,64795
24	DIBRS134	1,20567	1,19050	0,24836	4,07070	1,03788
25	DIBRS183	0,15288	2,11967	8,43672	1,73801	5,57275
26	DIBRS184	1,21576	4,02480	2,19098	21,24688	3,11714
27	DIBRS214	1,05929	5,26221	5,39873	0,28238	2,45070
28	DIBRS232	0,99202	8,25215	0,20943	0,56983	5,51791
29	DIBRS245	2,31615	1,12924	0,58633	0,40817	0,71611
30	DIBRS247	3,98530	4,97351	7,24640	0,03623	1,53562

Conclusões

Pode-se concluir que com os métodos multivariados de agrupamento e componentes principais é possível agrupar tratamentos em grupos homogêneos e as variáveis com maiores contribuições na seleção de cultivares tolerantes à seca entre as cultivares convencionais e transgênicas.

Referências Bibliográficas

- ARAGÃO, F.J.L.. Melhoramento do feijão para tolerância à seca via engenharia genética. IN: **Simpósio sobre Tolerância à Deficiência Hídrica em Plantas: Adaptando as Culturas ao Clima do Futuro**. Série documentos. N. 265. 2011. Embrapa Arroz e Feijão Santo Antônio de Goiás, GO, 148 p.
- BURR, I.W. & FOSTER, L.A. **A test for equality of variances**, Mimeo series n^o 282, University of Purdue, West Lafayette, 26p, 1972.
- CASAGRANDE, E.C., FARIAS, J.R.B., NEUMAIER, N., OYA, T., PEDROSO, J., MARTINS, P.K., BRETON, M.C., NEPOMUCENO, A.L. Expressão gênica diferencial durante déficit hídrico em soja. *R. Bras. Fisiol. Veg.*, 13(2): 168-184, 2001.
- MINGOTI, S.A. Análise de dados através de métodos de estatística multivariada, 1^a. edição. 2005. 295 p.
- PARENTE, R.C.P. **Aspectos da análise de resíduos**, 1984, 118f, Dissertação (Mestrado) - ESALQ/USP, Piracicaba.
- REIS, E. Estatística multivariada aplicada. 2^a. edição. Edições Silabo, Lisboa. 2001. 343p.
- SAS INSTITUTE, **SAS user's guide: statistics: version 8,2**, 6, ed, Cary, 2001, SAS Institute, 2000, 291p.
- SHAPIRO, S.S. & WILK, M.B. "An analysis of variance test for normality", *Biometrika* 52:591-611, 1965.
- SCHMITT, J. **Pré-Processamento para a Mineração de Dados: Uso da Análise de Componentes Principais com Escalonamento Ótimo**, 2005, 146p, Dissertação (Mestrado) – Universidade Federal de Santa Catarina, Florianópolis.
- SOUZA, S.R. Classificação de variedades da soja quanto à concentração de isoflavonas utilizando análise estatística multivariada, Londrina, UEL, 2004.
- STATSOFT, Inc. **STATISTICA for windows (Computer program manual)**, Tulsa, Oklahoma, USA, 1995.
- TUKEY, J. **One degree of freedom for non-additivity**, *Biometrics* 5:232-242, 1949.
- ZONTA, E.P., MACHADO, A. A., SILVEIRA JÚNIOR, P. 1982, Sistema de Análise Estatística-SANEST, Registro na SEI No, 066060, UFPEL, Pelotas, 63pp.