

Técnicas de análise exploratória utilizadas em pesquisas de sistemas silvipastoris

Jakelline Furtado de Oliveira¹, Maria Luiza Franceschi Nicodemo², Alfredo Ribeiro de Freitas², Francisco H. Dübbern de Souza², João Carlos Teixeira Mendes³

¹Bacharanda em Estatística da Universidade Federal de São Carlos, São Carlos, CEP 13565-905, bolsista da Embrapa. Endereço eletrônico: jakellinefurtado@hotmail.com

²Pesquisador(a) do CPPSE/Embrapa Pecuária Sudeste. e-mail: mlnicodemo@cppse.embrapa.br.

²Pesquisador(a) do CPPSE/Embrapa Pecuária Sudeste. e-mail: ribeiro@cppse.embrapa.br

²Pesquisador(a) do CPPSE/Embrapa Pecuária Sudeste. e-mail: francisco@cppse.embrapa.br

³Doutorando do Departamento de Engenharia Florestal – ESALQ . e-mail: jctmende@esalq.usp.br.

Resumo: O objetivo deste trabalho foi utilizar técnicas de análises exploratórias em dados de dois experimentos. O primeiro consistiu de quatro métodos de estabelecimento de leguminosas (1: plantio superficial ou sem gradagem; 2: gradagem antes do plantio; 3: gradagem após plantio e 4: gradagem antes e depois do plantio). No segundo foram avaliados o estabelecimento de nove leguminosas forrageiras herbáceas (1: *Lablab purpureum*, 2: *Centrosema acutifolium*, 3: *Clitoria ternatea*, 4: *Pueraria phaseoloides*, 5: *Alysicarpus vaginalis*, 6: *Aeschynomene vilosae*, 7: *Estilosantes Campo Grande* (*Stylosanthes capitata* + *S. macrocephala*), 8: *Calopogonium mucunoides* e 9: *Arachis pintoi*) em dois locais (1: sol e 2: sombra de moderada a intensa de plantações de eucalipto *Eucalyptus grandis* submetido a 40% de corte de desbaste oito anos após seu plantio). Em ambos os experimentos foram analisados a produção de matéria seca, o número de plântulas e o percentual de cobertura vegetal do solo das nove espécies de leguminosas. A análise exploratória mostrou-se eficiente para explorar a variabilidade latente, revelar a dispersão dos dados, detectar padrões de tendências, grau e direção da simetria e principalmente, detectar a presença de outliers.

Palavras-chave: análise descritiva dos dados, diagrama de caixa, leguminosa, sombra, sol, testes de normalidade.

Exploratory data analysis techniques in experiments of Silvopastoral systems

ABSTRACT - The objective was to utilize exploratory data analysis techniques in two experiments. The first one consisted of four methods of herbaceous forage legume seed incorporation into the soil (1: no light disking after broadcast seeding; 2: light disking before; 3: light disking after and 4: sowing, and light disking before and after disking). In the second experiment were evaluated the establishment of nine species (1: *Lablab purpureum*, 2: *Centrosema acutifolium*, 3: *Clitoria ternatea*, 4: *Pueraria phaseoloides*, 5: *Alysicarpus vaginalis*, 6: *Aeschynomene vilosae*, 7: *Estilosantes Campo Grande* (*Stylosanthes capitata* + *S. macrocephala*), 8: *Calopogonium mucunoides* e 9: *Arachis pintoi*) in two local (1: under full sun and 2: under shade promoted by a *Eucalyptus grandis* forest submitted to selective 40% thinning cut eight years after being planted). In both experiments were analyzed the dry matter production, the number of seedlings and vegetal covering percentage. The exploratory data analysis was efficient in order to explore the latent variability of the data, its dispersion and for detecting tendency and outliers.

Key Words: Box-plot, descriptive data analysis, forage legume, shade, sun, tests for normality

PROCI-2009.00353

OLI

2009

SP-PP-2009.00353

Técnicas de análise ...

2009

SP-PP-2009.00353

Técnicas de análise ...

2009

SP-PP-2009.00353



CPPSE-19074-1



CPPSE-19074-1

Introdução

Existe na atualidade grande interesse em pesquisas com sistemas silvipastoris envolvendo a consorciação de bovinos e florestas plantadas, pois estas possibilitam diversificar a produção, maximizar o uso do solo e conseqüentemente, aumentar a renda nas propriedades rurais. Com isto, estas atividades vêm atraindo cada vez mais os pequenos e médios proprietários rurais. Neste consórcio, um dos interesses é o cultivo de leguminosas, com a finalidade de fixar nitrogênio no solo e melhorar as propriedades do solo contribuindo para aumentar a produção. Em um estudo australiano (Congdon & Addison, 2003) foi constatado que de 35 espécies cultivares de leguminosas avaliadas, as de melhor desempenho na sombra também tiveram bom desempenho a pleno sol, indicando que estas espécies poderiam ser utilizadas tanto em plantações florestais jovens quanto em florestas mais fechada.

Em termos de pesquisas, ainda é pouco conhecida a natureza da variabilidade existente entre os diversos fatores ou tratamentos dos experimentos utilizados nos sistemas silvipastoris uma vez que estes envolvem simultaneamente uma diversidade de situações, diferindo dos experimentos tradicionais realizados em agricultura. Assim, o uso da análise exploratória de dados tem grande interesse nas pesquisas com sistemas silvipastoris, uma vez que possibilita investigar as características latentes existentes nos dados, identificar possíveis padrões, tendências, sugerir modelos de análises e ainda contribuir para o planejamento de futuros experimentos (Hartwig & Dearing, 1979; Cleveland, 1994; SAS, 2002-2003; Freitas et al. 2008). Por meio de diagramas de caixa (*box-plots*), por exemplo, pode-se visualizar a dispersão dos dados, detectar padrões de tendências, medidas de locação, grau e direção da simetria, presença de *outliers*, entre outras. O objetivo deste trabalho foi utilizar técnicas de análises exploratórias na análise de dados de plântulas, acúmulo de matéria seca e percentagem de cobertura do solo no estabelecimento de leguminosas em dois experimentos sob a coordenação da Embrapa Pecuária Sudeste, realizados no ano de 2008.

Material e Métodos

Foram realizados dois experimentos na Estação Experimental de Ciências Florestais da Esalq-USP em área experimental ocupada por bosque de *Eucalyptus grandis*., sendo que em ambos foram avaliados o acúmulo de matéria seca, o número de plântulas e o percentual de cobertura vegetal do solo. O primeiro experimento foi instalado em blocos casualizados, cinco repetições e quatro tratamentos referentes à métodos de estabelecimento de leguminosas (1:plantio superficial ou sem gradagem; 2: gradagem antes do plantio; 3:gradagem após plantio e 4:gradagem antes e depois do plantio).

Para a determinação da produção de matéria seca, g/m^2 , (x), em g, aos 90 dias pós-plantio, foram obtidas amostras de plantas, as quais foram cortadas a 5 cm (plantas prostradas) e 10 cm (plantas eretas ou semi eretas) de altura do solo. O número de plântulas (y) foi avaliado em uma área de $0,25m^2$ da parcela e a em percentagem de cobertura do solo (z), foi avaliada subjetivamente.

O segundo experimento foi instalado em delineamento em blocos casualizados, quatro repetições, e 18 tratamentos organizados em esquema fatorial 2×9 (dois locais: sol e sombra e nove espécies: 1:*Lablab purpureum*, 2:*Centrosema acutifolium*, 3:*Clitoria ternatea*, 4:*Pueraria phaseoloides*, 5: *Alysicarpus vaginalis*, 6: *Aeschynomene vilosae*, 7:*Estilosantes Campo Grande*

(*Stylosanthes capitata* + *S. macrocephala*), 8:*Calopogonium mucunoides* e 9:*Arachis pintoï*). Foram utilizadas parcelas de 2,5 m x 5 m, formadas por quatro fileiras de 5 m, espaçadas de 0,5 m, sendo que as duas fileiras centrais constituem a área experimental. As avaliações iniciais foram realizadas quatro, oito e doze semanas após o estabelecimento do experimento.

Para a análise estatística dos dados, foi utilizado o procedimento GLM do SAS (SAS INSTITUTE Inc, 2002/2003), com os dados y e z submetidos à transformação $\sqrt{y + 0,5}$ e $\arcseno(\sqrt{z/100})$, respectivamente.

Para a análise exploratória dos dados foi inicialmente utilizado o diagrama de caixa (*Box-plot*), que permite visualizar a dispersão dos dados em torno da média, o grau e a direção da simetria, a natureza da variância em cada efeito e a presença ou não de *outliers*. Em adição a esta análise visual, foram obtidas as medidas de tendência central, as medidas de dispersão e testes de normalidade por meio da estatística de *Shapiro-Wilks*, para amostra menor ou igual a 2000 (SAS 2002-2003). Todas essas estatísticas foram obtidas considerando-se o módulo *INSIGHT* do SAS (SAS 2002-2003). Para o leitor interessado em conhecer maiores detalhes sobre estas técnicas de análise recomenda-se a leitura do *INSIGHT* e o trabalho de Freitas et al (2008).

Resultados e Discussão

Observando-se a Figura 1, verifica-se que a produção de matéria seca (gráfico superior), apresenta amplitude semelhante entre os dados dos tratamentos 1, 2 e 4, e maior homogeneidade no tratamento 3, porém todos apresentando tendência de viesamento para a direita; no tratamento 3, observa-se uma marcação individual no extremo inferior da caixa, o que pode ser considerado com dado discrepante da amostra (*outlier*). Quanto ao número de plântulas (gráfico central), observa-se resposta crescente dos tratamentos (1 a 4), com maior amplitude no tratamento 4, não se constatando dado influente. Para a cobertura do solo (gráfico inferior), observa-se que os dados dos tratamentos de 1 a 3, são altamente viesados para a direita. O gráfico dos quatro tratamentos, sugere uma preocupação com o atendimento da pressuposição da homogeneidade das variâncias entre tratamentos e também com o ajuste da distribuição normal aos dados, uma vez que se verifica grande afastamento entre média e mediana. Estes são dois requisitos importantes para a análise de variância.

Os resultados quanto à produção de matéria seca e cobertura do solo (Figura 1), obtidos no tratamento 1, pode ser considerado bom, uma vez que as condições para o estabelecimento das sementes na superfície do solo são mais desfavoráveis do que aquelas dos demais tratamentos, conforme resultados de McWilliam e Dowling (1970).

Figura Gráfico do diagrama de caixas para produção de matéria seca (MS), g (superior), número de plântulas (central) e cobertura do solo, % (inferior). Os valores referem-se à média de nove espécies de leguminosas e quatro tratamentos (1:plantio superficial ou sem gradagem; 2: gradagem antes do plantio; 3:gradagem após plantio e 4:gradagem antes e depois do plantio).

Na Figura 2 são apresentados o diagrama de caixa para as nove espécies em relação aos dois locais: sombra (gráfico à esquerda) e sol (gráfico à direita). Para a produção de matéria seca (gráficos superiores), observa-se grande discrepância entre as nove espécies e também entre os dois locais, e como já discutido na Figura 1, este comportamento sugere não atendimento da pressuposição da homogeneidade das variâncias entre espécies dentro de local e também não ajuste da distribuição normal aos dados. Observa-se grande diferença na produção de matéria seca entre as espécies quando se muda de local (sombra e sol); quando expostas ao sol, apenas a espécie 9:*Arachis pintoi* se destacou na produção de matéria seca.

Para a produção de matéria seca, observaram-se marcações individuais com prováveis ocorrências de *outliers* nas espécies 2 e 4 (Figura 2: gráfico superior à esquerda) e 8 (Figura 2: gráfico superior à direita). Houve também diferenças entre as espécies para esta característica, uma vez que os coeficientes de variação oscilaram de 14,5% (3:*Clitoria*) a 189,0% (4:*Pueraria*) na sombra e de 22,4% (2: *Centrosema*) a 107,8% (6: *Aeschynomene*) ao sol.

Quanto ao o número de plântulas (Figura 2), houve variação entre espécies, porém, com comportamento semelhante nos dois locais. Exceto as espécies 2:*Centrosema*, 3:*Clitoria* e 5:*Alysicarpus* (gráfico à esquerda) todas apresentaram marcações individuais; no sol (gráfico à direita), somente as espécies 6:*Aeschynomene*, 7:*Stylosantes Campo Grande* e 9:*Arachis*, apresentaram essas marcações. O número de plântulas também diferiu entre as espécies, uma vez que os coeficientes de variação variaram de 20,9% (8:*Calopogonio*) a 52,5% (7: *Stylosantes Campo Grande*) na sombra e de 16,7% (1:*Lablab purpureus*) a 77,5% (7:*Stylosantes Campo Grande*) ao sol.

Em relação ao percentual de cobertura do solo, observa-se que comportamento em termos de variabilidade entre espécies e local é similar ao observado para o número de plântulas. Conforme constatado para esta variável, o coeficiente de variação também teve grande oscilação, isto é, variou de 27,3% (1:*Lablab purpureus*) a 140,5% (7:*Stylosantes Campo Grande*) na sombra e de 31,4% (1:*Lablab purpureus*) a 133,2% (7:*Stylosantes Campo Grande*) no sol.

Como foi constatado neste trabalho, nas pesquisas com sistemas silvipastoris, deve-se ter grande preocupação com a qualidade dos dados, pois nos experimentos realizados nesta área, é difícil atender os requisitos básicos de uma análise de variância, tais como, normalidade, homogeneidade de variância e independência de erros. A detecção ou não de *outliers*, por exemplo, requer bastante cuidado e necessita ser decidida juntamente com os pesquisadores envolvidos na área. Com este cenário, a aplicação da análise exploratória aos dados se reveste de importância ainda maior. Como já é sabido (Hartwig & Dearing, 1979; Singha & Nocerino 2002), os desvios-padrão e as correspondentes variâncias são sensíveis a dados extremos, falta de normalidade, dados assimétricos e presença de *outliers*. Porém, segundo esses autores, após a eliminação de *outliers*, é possível obter estimativas robustas que correspondam às estimativas clássicas.



Figura 2: Gráfico do diagrama de caixas para acúmulo de matéria seca, g (superior), número de plântulas (central) e cobertura do solo, % (inferior), avaliados de nove espécies em dois tratamentos (sombra: esquerda e sol: direita)

Conclusões

A análise exploratória utilizada em dados oriundos de pesquisas com sistemas silvipastoris, mostrou-se eficiente para explorar a variabilidade latente, revelar a dispersão dos dados, detectar padrões de tendências, grau e direção da simetria e principalmente, detectar a presença de *outliers*.

Literatura citada

- CLEVELAND, W.S. *The elements of graphing data*. New Jersey:AT&T Bell Laboratories; Murray Hill, 1994. 297p.
- FREITAS, A. R.; BARIONI JÚNIOR, W.; FERREIRA, R.P. et al. Técnicas de análises exploratórias em dados de cultivares de alfafa. *Revista Brasileira de Zootecnia*, v.37, n.9, p.1531-1536,2008
- HARTWIG, F.; DEARING, B.E. Exploratory data analysis. In: SULLIVAN, J.L. (Ed.) *Series: quantitative applications in the social science*. Newbury Park: 1979. 83p. (Sage University Paper,16).
- STATISTICAL ANALYSES SYSTEM - SAS. *SAS/INSIGHT User's guide*. versão 9.1.3 - versão para Windows. Cary: SAS Institute, 2002/2003. (CD-ROM).