



TÉCNICAS DE ANÁLISES EXPLORATÓRIAS EM DADOS DE CULTIVARES DE ALFAFA (“MEDICAGO SATIVA” L.)

ALFREDO RIBEIRO DE FREITAS¹; WALDOMIRO BARIONI JÚNIOR², REINALDO DE PAULA FERREIRA³; COSME DAMIÃO CRUZ⁴; JOAQUIM BARTOLOMEU RASSINI⁵, DUARTE VILELA⁶

“1” Pesquisador da Embrapa Pecuária Sudeste, Caixa Postal 339, 13560-970 - São Carlos, SP. Bolsista do CNPq. (ribeiro@cppse.embrapa.br)

“2” Pesquisador da Embrapa Pecuária Sudeste (barioni@cppse.embrapa.br)

“3” Pesquisador da Embrapa Pecuária Sudeste (reinaldo@cppse.embrapa.br)

“4” Professor Titular do Departamento de Biologia da UFV (cdcruz@ufv.br)

“5” Pesquisador da Embrapa Pecuária Sudeste (rassini@cppse.embrapa.br)

“6” Pesquisador da Embrapa Gado de Leite. (vilela@cnppl.embrapa.br)

RESUMO

O objetivo foi utilizar técnicas estatísticas de análises exploratórias (momentos, testes de normalidade e medidas robustas de escalas) em dados da produção de matéria seca (PMS) de 16 cortes de 92 cultivares de alfafa (“Medicago sativa” L.). Exceto nos cortes 1, 3, 7, 10 e 11, os dados de PMS não se comportaram como distribuição normal. Os coeficientes de assimetria foram negativos para os dados de PMS nos cortes 1 a 4, 10 a 12, 14 a 16, enquanto que a curtose foi negativa nos cortes 1, 9, 10, 12 e 16. As três estimativas robustas do desvio-padrão (DAM, S_n e Q_n), diferiram do estimador tradicional (S) em todos cortes, indicando que o uso do S tradicional deva ser substituído por estes estimadores robustos.

PALAVRAS-CHAVE

produção de matéria seca, medidas robustas de escalas, testes de normalidade, testes de locação

EXPLORATORY DATA ANALYSIS TECHNIQUES IN CULTIVARS OF ALFALFA

ABSTRACT

The objective was to utilize exploratory data analysis techniques (moments, tests for normality and robust measures of scale) evaluating dry matter production (DMP) obtained from 16 cuts of 92 accessions of alfalfa (“Medicago sativa” L.). Except cuts 1, 3, 7, 10 and 11, the DMP data do not fit to normal distribution. The skewness coefficients were negative for the DMP data in cuts 1 to 4, 10 to 12, 14 and 16, while kurtosis coefficients were negative in cuts 1, 9, 10, 12 and 16. Three robust estimates of standard deviation (DAM, S_n e Q_n), were different from the traditional standard (S) in all cuts, showing that estimate S should be replaced by these robust estimators in all future statistical analysis.

KEYWORDS

dry matter production, robust measures of scale, tests for normality, location tests

INTRODUÇÃO

Tem sido crescente o uso da alfafa em sistemas intensivos de produção de leite, devido a sua elevada qualidade, produtividade e resposta econômica em relação a outros tipos de alimentos volumosos. Na pesquisa tem sido comum avaliar o comportamento entre cultivares por meio do rendimento de matéria seca realizado em cortes sucessivos no período das águas e da seca. É fundamental realizar um mínimo de medições, de tal forma que a seleção possa ser feita com acurácia e menor custo em termos de tempo e mão-de-obra. Estes dados, no entanto, podem apresentar outliers, correlações absurdas, afastamento da distribuição normal, heterogeneidade de variâncias, grau acentuado de assimetria e curtose, entre outras, que podem omitir resultados importantes e influenciar as inferências obtidas, sendo mais problemáticas em análises multivariadas do que nas univariadas. Com o propósito de conhecer adequadamente a natureza de variação dos dados de produção de matéria seca (PMS) de 16 cortes de 92 cultivares de alfafa ("Medicago sativa" L.) e proceder o refinamento metodológico de modo a obter estimativas mais confiáveis, foram utilizadas técnicas de análises exploratórias (momentos, testes de normalidade e medidas robustas de escalas).

MATERIAL E MÉTODOS

O experimento iniciou-se em junho de 2004 no Campo Experimental da Embrapa Pecuária Sudeste – São Carlos, SP, em um Latossolo Vermelho Amarelo. O preparo do solo consistiu de aplicação de 5,0 t/ha de calcário dolomítico em maio de 2004, seguida de adubação para estabelecimento da cultura, consistindo de aplicação de 80 kg/ha de superfosfato simples, 100 kg/ha de cloreto de potássio e 30 kg/ha de FTE-BR12. Em junho/2005 foi feita a semeadura de 92 acessos de alfafa provenientes do INTA-Argentina, tendo como testemunha a cultivar Crioula. Foram utilizados 20kg de sementes viáveis/ha, previamente inoculadas com estirpes de "Rhizobium meliloti" e espaçamento de 20 cm entre linhas. Para a adubação de cobertura aplicou-se à lanço, após cada corte, 60 kg/ha de K²O, na forma de cloreto de potássio. O delineamento experimental foi o de blocos ao acaso, com duas repetições, sendo as parcelas constituídas de cinco fileiras de cinco metros de comprimento. Considerou-se como bordadura uma fileira de cada lado e 0,50 m de cada extremidade da parcela. Os 16 cortes de avaliação foram realizados quando a maioria dos genótipos apresentavam 10% de plantas em florescimento. As características avaliadas foram: produção de matéria verde e seca; altura da planta; relação caule/folha; aceitação fenotípica, que dá uma idéia do comportamento geral da cultivar (1-excelente; 2-ótimo; 3-bom; 4-regular; 5-ruim); qualidade da forragem (digestibilidade e proteína bruta); incidência de pragas e doenças (0-susceptível; 1-baixa resistência; 2-moderadamente resistente; 3-altamente resistente); e persistência, avaliada através da estimativa da cobertura vegetal das linhas de plantio da área útil de cada parcela, no início e no final da condução do ensaio. No entanto, para o presente estudo foi considerada apenas a produção de matéria seca. Foram utilizadas três técnicas de análises exploratórias do SAS (SAS, 2003) a) momentos: média, desvio-padrão (S), variância, assimetria, curtose e coeficiente de variação; b) testes de normalidade: Teste Shapiro-Wilk (W) e c) medidas robustas de escalas. Foram utilizadas três estatísticas: desvio absoluto da mediana-DAM (para uma distribuição normal $S=1,4826DAM$), S_n e Q_n , para obter um estimador robusto do desvio-padrão. A estatística S é comumente usada para obter inferências de uma população; porém, como ela é sensível à "outliers" é importante o uso de estimadores robustos para substituí-la, principalmente para grandes arquivos de dados coletados em condições de campo, como no presente estudo.

RESULTADOS E DISCUSSÃO

A Tabela 1 apresenta um resumo das técnicas de análises exploratórias utilizando os 16 corte e 92 acessos. Estas técnicas exploratórias são discutidas com o intuito de conhecer o quanto os estimadores são afetados devido às anomalias presentes nos dados. A instabilidade dos dados pode ser visualizada pelos coeficientes de variação que variaram de 17,7% a 51,3%. Com exceção dos dados da produção de matéria seca dos cortes 1, 3, 7, 10 e 11; nos demais, a distribuição dos dados

não se ajustaram à distribuição normal. Pelo teste de Shapiro-Wilks, analisou-se a distribuição dos dados de PMS dos cortes, tendo como ordem de afastamento da distribuição normal, os cortes 3, 7, 1, 10, 11, 16, 6, 12, 15, 4, 5, 14, 13, 9, 8 e 2. No teste de Shapiro-Wilks usa-se a estatística W ($0 < W < 1$) e a probabilidade dessa estatística ($0 < \text{Prob} < 1$); valores próximos de zero, para ambas, W e Prob , indicam que a distribuição dos dados se afasta da curva normal. Observa-se que os coeficientes de assimetria foram negativos nos cortes 1 a 4, 10 a 12, 14 a 16. A assimetria é uma medida da forma da distribuição dos dados com relação a distribuição da curva normal. Os resultados indicam que a cauda desta curva é viesada à esquerda; em uma distribuição normal esperar-se-ia a relação média < mediana < moda. Apesar dos dados, nos cortes, serem amodais, a relação média < mediana não se verificou nos cortes 3, 4, 6, 8, 9, 15 e 16. Da mesma forma, os coeficientes de curtose foram negativos nos cortes 1, 9, 10, 12 e 16; os resultados mostraram que a distribuição é mais achatada do que a normal (Platicúrtica). A curtose é uma medida do grau de achatamento de uma distribuição em relação à curva normal. Segundo COCHRAN e COX (1978), a assimetria, a curtose e a não-normalidade dos dados, afetam as inferências obtidas, a estimação dos efeitos fixos, o uso dos testes t e F e a heterogeneidade da variância do erro, sendo mais problemáticas em análises multivariadas. Os valores dos resultados das medidas robustas de escalas (Desvio Absoluto da Mediana-DAM, S_n e Q_n), de modo geral diferiram do desvio-padrão (S), indicando que o uso do S para eliminar dados extremos de uma amostra, ou seja, média $\pm 3S$; em rotinas de simulação de dados com distribuição normal, testes de hipóteses de médias, intervalos de confiança, entre outras aplicações, deve ser substituído pelos estimadores robusto. Segundo HARTWIG e DEARING (1979), os desvios-padrão e as correspondentes variâncias, são sensíveis à dados extremos, falta de normalidade, dados assimétricos e presença de "outliers". SINGHA e NOCERINO (2002), trabalhando com dados de contaminantes ambientais censurados, obtiveram estimativas confiáveis de parâmetros populacionais de média e desvio-padrão. Segundos eles, após a eliminação de "outliers", é possível obter estimativas robustas em concordância com as correspondentes estimativas clássicas.

CONCLUSÕES

De modo geral, os dados de produção de matéria seca obtidos de 16 cortes de 92 cultivares, não se ajustaram à distribuição normal. Mesmo havendo distribuição normal nos dados de alguns cortes, as três estimativas robustas do desvio-padrão (DAM, S_n e Q_n), diferiram do desvio-padrão clássico (S) em todos os cortes, indicando que o uso de S deve ser substituído por estes estimadores robustos.

REFERÊNCIAS BIBLIOGRÁFICAS

1. COCHRAN, W.G., COX, D.F. 1978. "Desenho experimentales". Mexico: Trillas, 661p.
2. HARTWIG, F., DEARING, B.E. Exploratory data analysis. Series: Quantitative Applications in the Social Science 16 1979. editor: John L. Sullivan. Sage University Paper.83p.
3. SAS Institute 2002/2003. "SAS/INSIGHT User's Guide". versão 9.1.3, versão para Windows Cary, NC, USA.
4. SINGHA, A., NOCERINO, J. Robust estimation of mean and variance using environmental data sets with below detection limit observations. "Chemometrics and Intelligent Laboratory Systems". v.60, n.1-2, 28 January 2002, p.69-86.