



Técnicas de análises exploratórias em dados de cultivares de alfafa

Alfredo Ribeiro de Freitas¹, Waldomiro Barioni Júnior¹, Reinaldo de Paula Ferreira¹, Cosme Damião Cruz², Adônis Moreira³, Duarte Vilela⁴

¹ Embrapa Pecuária Sudeste, Caixa Postal 339, CEP: 13560-970, São Carlos, SP, Brasil.

² Departamento de Biologia Geral da UFV, CEP: 36570-000, Viçosa, MG, Brasil.

³ Embrapa Pecuária Sudeste, Caixa Postal 339, CEP: 13560-970, São Carlos, SP, Brasil. Pesquisador do CNPq.

⁴ Embrapa Gado de Leite, Rua Eugênio do Nascimento, 610 - Dom Bosco, CEP: 36038-330, Juiz de Fora, MG, Brasil.

RESUMO - Objetivou-se utilizar técnicas de análises exploratórias em dados da produção de matéria seca (PMS) de 16 cortes de 92 cultivares de alfafa (*Medicago sativa* L.). A produção de matéria seca variou entre e dentro de cortes, uma vez que os coeficientes de variação oscilaram de 17,7% (corte 2) a 51,7% (corte 9). De modo geral, os dados de produção obtidos nos 16 cortes não se ajustaram à distribuição normal. Os coeficientes de simetria foram negativos nos cortes 1 a 4, 10 a 12, 14 a 16, enquanto os coeficientes de curtose foram negativos nos cortes 1, 9, 10, 12 e 16. As estimativas robustas do desvio-padrão obtidas por quatro métodos diferiram do desvio-padrão clássico (S) em todos os cortes, indicando que a estimativa S deve ser substituída por estimadores robustos.

Palavras-chave: medidas robustas de escalas, produção de matéria seca, testes de locação, testes de normalidade

Exploratory data analysis techniques in cultivars of alfalfa

ABSTRACT - The objective was to utilize exploratory data analysis techniques for evaluating dry matter production (DMP) obtained from 16 cuts of 92 accessions of alfalfa (*Medicago sativa* L.). A significant effect on DMP was observed both within and among cuts; the coefficient of variation ranged from 17.7% (cut 2) to 51.7% (cut 9). The DMP data obtained from 16 cuts of 92 accessions of alfalfa did not fit to the normal distribution. The skewness coefficients were negative for dry matter production data in cuts 1 to 4, 10 to 12, and 14 to 16, while kurtosis coefficients were negative in cuts 1, 9, 10, 12, and 16. Robust estimates of standard deviation obtained from four different methods were different from the traditional standard (S) in all cuts, showing that estimate S should be replaced by these robust estimators in statistical analysis.

Key Words: dry mater production, location tests, robust measures of scale, tests for normality

Introdução

O uso de alfafa em sistemas intensivos de produção de leite tem despertado o interesse de produtores, principalmente no estado de São Paulo, onde a atividade passa por crises e tem desaparecido em determinadas regiões. No entanto, a pecuária leiteira deve ser mantida, não só para atender ao mercado consumidor, mas também como alternativa de diversificação das propriedades rurais do estado. A alfafa, por ser uma forrageira de alta produção de matéria seca (MS) e elevada qualidade nutricional, pode ser uma alternativa na alimentação animal (Vilela, 1998).

Na pesquisa tem sido comum avaliar o comportamento de cultivares por meio do rendimento de MS determinado com cortes sucessivos no período das águas e da seca.

É fundamental realizar um mínimo de medições, de modo que a seleção seja feita com acurácia e menor custo em termos de tempo e mão-de-obra. Esses dados, no entanto, em virtude da variabilidade inerente ao material e de problemas de coleta, podem apresentar *outliers*, correlações absurdas, afastamento da distribuição normal, heterogeneidade de variâncias, grau acentuado de simetria e de curtose, entre outras características que podem ocultar resultados importantes e influenciar as inferências obtidas. Essas anomalias em análises multivariadas são mais problemáticas 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 ao refinamento metodológico de modo

a obter estimativas mais confiáveis, foram utilizadas técnicas de análises exploratórias de dados, que podem ser definidas como o conjunto de procedimentos e de técnicas de manejo de dados com os objetivos de construir gráficos e tabelas sem envolver grande teorização sobre o assunto; fornecer contribuições valiosas para avaliar a qualidade dos dados; gerenciar e monitorar a precisão da coleta dos dados; e calibrar instrumentos de mensuração, entre outros. Em síntese, a análise exploratória investiga características latentes nos dados que indiquem possíveis padrões, tendências ou modelos (Hartwig & Dearing, 1979; Cleveland, 1994; Digby & Kempton, 1996; Gower & Hand, 1996; Pearson, 2001; SAS, 2002-2003) e difere da análise clássica (AC) quanto à sequência de atividades. Na análise clássica, a sequência é: problema => dados => modelo => análises => conclusões; na análise exploratória, é: problema => dados => análises => modelo => conclusões.

Este trabalho foi realizado com os objetivos de: analisar os dados por meio de diagramas de caixa (*box-plots*) e de dispersão (*scatter plot*) para visualizar a dispersão dos dados, detectar padrões de tendências, medidas de locação, grau e direção da simetria, presença de *outliers*, entre outras; comparar os testes de normalidade; obter estimadores robustos da média e do desvio-padrão das características; com base nas análises exploratórias, propor soluções para atender às pressuposições básicas de um modelo misto em estudos longitudinais; e propor alternativas de modelos de análise.

Material e Métodos

Utilizou-se o diagrama de caixa (*Box-plot*) para revelar características importantes, como a dispersão dos dados em torno da média, o grau e a direção da simetria, a existência de heterogeneidade de variâncias e a presença de *outliers* (Cleveland, 1994; Digby & Kempton, 1996; Gower & Hand, 1996). Possíveis *outliers* foram determinados a partir de quatro pontos calculados no gráfico: $L_1 = Q_1 - 1,5(Q_3 - Q_1)$; $L_2 = Q_1 - 3,0(Q_3 - Q_1)$; $U_1 = Q_3 + 1,5(Q_3 - Q_1)$ e $U_2 = Q_3 + 3,0(Q_3 - Q_1)$. Q_1 e Q_3 são os elementos de posição 25 e 75% da amostra ordenada e correspondem, respectivamente, ao primeiro e terceiro quartil. L_1 e U_1 são as delimitações internas; e L_2 e U_2 as delimitações externas. Neste trabalho, os dados menores que L_2 e maiores que U_2 foram considerados dados discrepantes da amostra (*outliers*).

Foram estudados a média aritmética, a mediana, a variância, o desvio-padrão, o erro-padrão da média, o coeficiente de variação e as medidas de simetria e de curtose. A mediana corresponde ao elemento de posição 50% da amostra ordenada. Se a distribuição é simétrica, a média é

igual à mediana; se é assimétrica positiva, a média é maior que a mediana; e se é assimétrica negativa, a média é menor que a mediana.

Cinco testes de normalidade foram utilizados: *Shapiro-Wilks* (S-W) - apropriado para tamanho amostral menor ou igual a 2.000 (Bonett & Seierb, 2002). Testa a hipótese nula de que uma amostra x_1, \dots, x_n pertence a uma população com distribuição normal e utiliza a estatística W ($0 < W < 1$); se W é igual a 1, os dados ajustam perfeitamente à distribuição normal, enquanto valores pequenos de W são evidências de desvios da normalidade; Kolmogorov-Smirnov (K-S) - avalia a discrepância entre a distribuição empírica $F_n(y)$ e a distribuição normal considerada referência $F(y)$, com parâmetros m e s estimados pela média e pelo desvio-padrão amostral; testa a hipótese $H_0: F_n(y) = F(y)$ versus $H_a: F_n(y) \neq F(y)$ e é mais sensível em pontos próximos da mediana da distribuição que nas caudas; Cramer-von Mises (W^2) - é uma alternativa do teste de K-S; Anderson-Darling (A^2) - usado para testar se uma amostra de dados pertence a uma distribuição específica.

Com exceção do *Shapiro-Wilks*, esses testes são apropriados para amostras grandes. Em todos os testes, a hipótese nula a ser testada é que os dados em estudo correspondem a uma amostra aleatória proveniente de uma distribuição normal.

O desvio-padrão amostral tradicional (S) é comumente usado para obter inferências de uma população, pois, juntamente com a média, caracterizam uma distribuição. Como o S é sensível a *outliers*, é importante o uso de estimadores robustos para esta estatística, principalmente para grandes arquivos de dados coletados em condições de campo, cuja qualidade muitas vezes é insatisfatória. Cinco estatísticas foram utilizadas para obter um estimador robusto do desvio-padrão: a) intervalo interquartil, que é a diferença entre os quartis superior e inferior de uma distribuição: $(Q_3 - Q_1)$; b) diferença média de

Gini (G), em que G é dado por $G = \frac{1}{\binom{n}{2}} \sum_{i < j} |y_i - y_j|$;

c) desvio absoluto da mediana (DAM) - DAM = mediana i $|y_i - \text{mediana } j(y_j)|$, em que a mediana $j(y_j)$ é a mediana das n observações e a mediana i é a mediana dos n valores absolutos dos desvios em relação à mediana j . Para uma população normal, o estimador robusto do desvio-padrão pode ser obtido, respectivamente, por $(Q_3 - Q_1)/1,34898$, $\sqrt{\pi G/2}$ e $1,4826 \text{DAM}$. Como a estatística DAM é pouco eficiente para a distribuição normal e inapropriada para

distribuições simétricas, duas alternativas foram estudadas: Sn e Qn.

Resultados e Discussão

Na Figura 1 é apresentado o diagrama de caixa dos 16 cortes de alfafa. A linha horizontal cheia no meio da caixa indica a mediana, que é o elemento de posição 50% ou segundo quartil (Q_2). As partes inferior e superior da caixa correspondem, respectivamente, ao quartil Q_1 e Q_3 . As caixas estreitas (*whiskers*) acima e abaixo da caixa central possuem distância não superior a 1,5 vezes a distância interquartilica ($Q_3 - Q_1$). Marcações individuais nos extremos das caixas estreitas são consideradas dados discrepantes da amostra. Os maiores valores de produção de PMS diferiram entre os cortes e, pela ordem, foram obtidos nos cortes 2, 1, 3, 13, 14 e 15, enquanto os menores foram obtidos nos cortes 9, 8, 7, 6 e 16. Exceto os cortes 1 e 10 (Figura 1A), marcações individuais com prováveis ocorrências de *outliers* são observadas em todos os cortes.

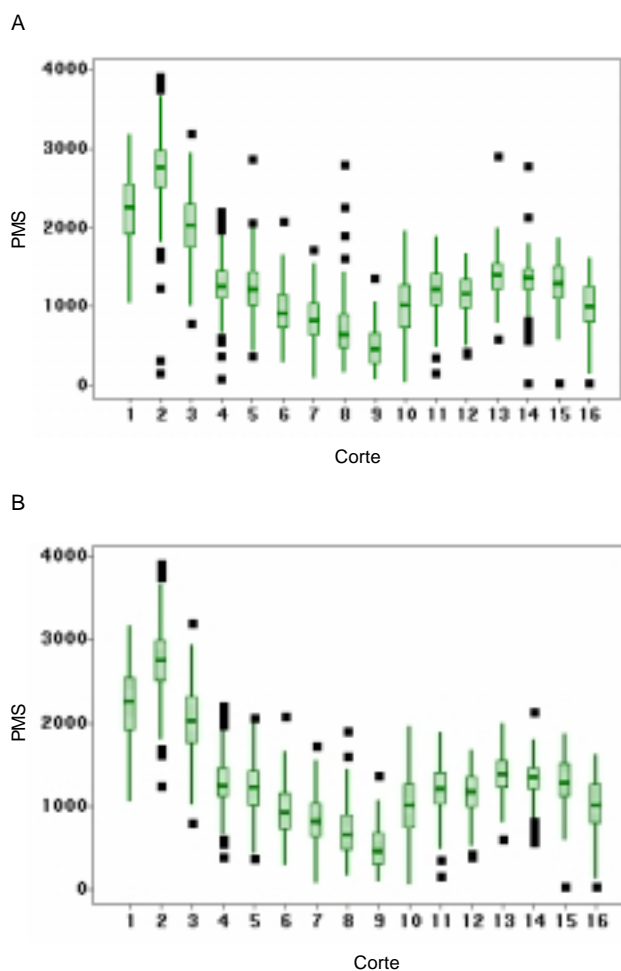


Figura 1 - Produção de matéria seca (PMS), kg/ha, de 16 cortes de 92 cultivares de alfafa, com (A) e sem *outliers* (B).

Na Figura 1 consta o diagrama após a eliminação dos *outliers* conforme delimitações internas (L_1 e U_1) e externas (L_2 e U_2) do diagrama de caixa (Tabela 1). Os critérios são rigorosos para um dado ser considerado discrepante da amostra, isto é, ser menor que L_2 ou maior que U_2 . No caso dos cortes 7, 8, 9 e 10, por exemplo, a produção de MS, kg/ha, deveria ser negativa. Para ser considerado *outlier*, como no corte 1, a PMS, kg/ha, deveria ser menor que 31,43 ou maior que 4.435,09.

A produção de matéria seca diferiu entre e dentro de cortes, uma vez que os coeficientes de variação (Tabela 2) oscilaram de 17,7% (corte 2) a 51,7% (corte 9). Nos cortes 2, 4, 5, 8, 9, 13, 14, observou-se a ocorrência de *outliers*. Nos dados originais, a média foi maior que a mediana nos cortes 3, 4, 6, 7, 8, 9, 15 e 16 e caracterizou-se simetria positiva e menor que a mediana nos cortes 1, 2, 5, 10, 11, 12, 13 e 14, caracterizando-se simetria negativa. A estimativa dessas estatísticas após a retirada dos *outliers* acarretou simetria nos dados, pois houve equivalência entre média e mediana sensível e também redução nos coeficientes de variação.

Os coeficientes de simetria foram negativos nos cortes 1 a 4, 10 a 12, 14 a 16 (Tabela 3). A simetria é uma medida da forma de distribuição dos dados quanto à distribuição da curva normal. Os resultados indicam que a cauda desta curva é viesada à esquerda. Do mesmo modo, os coeficientes de curtose foram negativos nos cortes 1, 9, 10, 12 e 16 e indicam que a distribuição é mais achatada que a normal (platicúrtica). A curtose é uma medida do grau de achata-

Tabela 1 - Valores de delimitações internas (L_1 e U_1) e externas (L_2 e U_2) para detectar dados discrepantes (*outliers*) de produção de matéria seca (PMS), kg/ha, de 16 cortes de 92 cultivares de alfafa

Corte	L_2	L_1	U_1	U_2
1	31,43	975,07	3.491,45	4.435,09
2	1.073,27	1.789,76	3.700,40	4.416,89
3	145,26	955,03	3.114,43	3.924,20
4	103,75	610,00	1.959,97	2.466,21
5	-179,08	423,41	2.030,05	2.632,54
6	-510,80	113,49	1.778,26	2.402,55
7	-868,82	-110,27	1.912,52	2.671,06
8	-770,95	-147,23	1.516,03	2.139,75
9	-828,96	-267,39	1.230,16	1.791,74
10	-852,36	-56,96	2.064,12	2.859,52
11	-145,57	438,77	1.997,01	2.581,35
12	-81,76	455,05	1.886,55	2.423,36
13	237,20	728,66	2.039,22	2.530,68
14	446,00	830,20	1.854,72	2.238,91
15	-65,71	522,98	2.092,80	2.681,48
16	-553,12	126,43	1.938,56	2.618,11

Dados $< L_2$ ou $> U_2$ são considerados discrepantes da amostra (*outliers*). $L_2 < \text{dados} < L_1$ ou $U_1 < \text{dados} < U_2$ são discrepantes da amostra, porém não são *outliers*.

$L_1 = Q_1 - 1,5(Q_3 - Q_1)$; $L_2 = Q_1 - 3,0(Q_3 - Q_1)$.
 $U_1 = Q_3 + 1,5(Q_3 - Q_1)$; $U_2 = Q_3 + 3,0(Q_3 - Q_1)$.
 Q_1 : primeiro quartil; Q_3 : terceiro quartil.

Tabela 2 - Produção de matéria seca (PMS), kg/ha, de 16 cortes de 92 cultivares de alfafa

Corte	Média		Mediana		Erro-padrão da média		Coeficiente de variação	
	Com outlier	Sem outlier	Com outlier	Sem outlier	Com outlier	Sem outlier	Com outlier	Sem outlier
1	2.248,4		2.256,9		31,80		19,2	
2 (2)+	2.730,5	2.757,9	2.757,8	2.762,6	35,69	30,30	17,7	14,8
3	2.041,4		2.035,1		29,89		19,8	
4 (1)	1.266,6	1.273,1	1.250,8	1.250,9	21,79	20,93	23,3	22,2
5 (1)	1.222,1	1.213,1	1.227,02	1.225,6	23,42	21,73	26,0	24,2
6	939,5		921,4		21,19		30,6	
7	825,5		819,4		21,31		35,0	
8 (2)	704,4	684,3	656,3	652,6	26,65	22,73	51,3	44,8
9 (3)	489,5	489,5	463,3	463,3	18,82	18,82	51,7	51,7
10	1.013,1		1.021,8		27,97		37,4	
11	1.210,02		1.217,5		21,19		23,7	
12	1.146,9		1.166,5		19,32		22,8	
13 (1)	1.380,7	1.372,4	1.395,05	1.394,1	19,70	17,95	19,3	17,7
14 (2)	1.331,5	1.330,6	1.357,7	1.357,7	20,13	17,31	20,5	17,5
15	1.289,2		1.284,7		21,39		22,5	
16	1.011,90		1.009,5		23,36		31,3	
Geral	1.304,0	1.303,2	1.212,9	1.212,8	12,00	12,04	50,7	49,4

+ = Número de outliers por corte.

Tabela 3 - Coeficientes de simetria e curtose de dados de produção de matéria seca (PMS), kg/ha, de 16 cortes de 92 cultivares de alfafa

Corte	Simetria		Curtose	
	Com outlier	Sem outlier	Com outlier	Sem outlier
1	-0,13		-0,20	
2 (2)+	-1,55	-0,27	7,09	1,20
3	-0,15		0,16	
4 (1)	-0,16	0,14	1,82	1,04
5 (1)	0,60	-0,13	3,47	0,23
6	0,38		0,48	
7	0,00		0,10	
8 (2)	1,77	0,60	7,18	0,73
9 (3)	0,58	0,58	-0,22	-0,22
10	-0,08		-0,55	
11	-0,41		0,47	
12	-0,43		-0,20	
13 (1)	0,72	-0,30	5,02	0,04
14 (2)	-0,07	-0,51	6,58	1,28
15	-0,41		0,90	
16	-0,30		-0,23	
Geral	0,91	0,92	0,77	

+ = número de outliers por corte.

mento de uma distribuição em relação à curva normal. Segundo Cochran & Cox (1978), a simetria, 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 e são mais problemáticas em análises multivariadas.

Pelo teste de *Shapiro-Wilks*, analisou-se a distribuição dos dados de produção de matéria seca dos cortes, considerando 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 (Tabela 4).

Por meio desse teste, usam-se a estatística W ($0 < W \leq 1$) e sua probabilidade ($0 < Prob \leq 1$); valores próximos de zero, tanto para a estatística quanto para sua probabilidade, indicam que a distribuição dos dados se afasta da curva normal. A retirada dos outliers melhorou a aproximação dos dados à distribuição normal, no entanto, no corte 5, a estatística W passou de 0,0001 para 0,8633 e, no corte 13, de $<0,0001$ para 0,3061, ou seja, a distribuição dos dados foi aproximadamente normal.

Nos dados observados, houve tendência de as estimativas obtidas pelo método tradicional ser inferiores às estimativas robustas nos cortes 1, 9, 10, 12 e 16; superior nos cortes 2, 4, 5, 8, 13 e 14; e semelhante nos demais. Quando os outliers foram eliminados, as estimativas robustas obtidas pela maioria dos métodos diminuíram.

De modo geral, as estimativas robustas do desvio-padrão proporcionadas pelos métodos diferiram das estimadas pelo método tradicional (S), indicando que o uso desse método nas aplicações estatísticas para eliminar dados extremos de uma amostra (média + 3S) e em rotinas para simular dados com distribuição normal nos testes de hipóteses de médias, nos intervalos de confiança, entre outras aplicações, deve ser substituído pelos estimadores robustos. Segundo Hartwig & Dearing (1979), os desvios-padrão e as correspondentes variâncias são sensíveis a dados extremos, falta de normalidade, dados assimétricos e presença de outliers. Singha & Nocerino (2002), em pesquisa com dados de contaminantes ambientais censurados, obtiveram estimativas confiáveis de parâmetros populacionais de média e desvio-padrão. Segundos esses autores, após a eliminação de outliers, é possível obter estimativas robustas que correspondam às estimativas clássicas.

Tabela 4 - Testes de normalidade ajustados a dados de produção de matéria seca (kg/ha) de 16 cortes de 92 cultivares de alfafa

Corte	Shapiro-Wilk		Kolmogorov-Smirnov		Cramer-von Mises		Anderson-Darling	
	Com outlier	Sem outlier	Com outlier	Sem outlier	Com outlier	Sem outlier	Com outlier	Sem outlier
1	0,3683		>0,1500		>0,2500		>0,2500	
2 (2)+	<0,0001	0,0350	<0,0100	0,0877	<0,0050	0,0485	<0,0050	0,0379
3	0,9148		>0,1500		>0,2500		>0,2500	
4 (1)	0,0032	0,0477	<0,0100	0,0383	0,0138	0,0505	0,0069	0,0335
5 (1)	0,0001	0,8633	>0,1500	>0,1500	0,2193	>0,2500	0,1262	>0,2500
6	0,0466		>0,1500		>0,2500		>0,2500	
7	0,8159		>0,1500		>0,2500		>0,2500	
8 (2)	<0,0001	0,0012	<0,0100	0,0782	<0,0050	0,1365	<0,0050	0,0888
9 (3)	<0,0001	0,0000	<0,0100	<0,0100	<0,0050	<0,0050	<0,0050	<0,0050
10	0,3595		>0,1500		>0,2500		>0,2500	
11	0,0921		>0,1500		>0,2500		>0,2500	
12	0,0166		>0,1500		0,1334		0,0667	
13 (1)	<0,0001	0,3061	0,0327	>0,1500	0,1165	0,1709	0,0335	0,1526
14 (2)	<0,0001	0,0008	<0,0100	<0,0100	<0,0050	<0,0050	<0,0050	<0,0050
15	0,0147		>0,1500		>0,2500		>0,2500	
16	0,0666		>0,1500		>0,2500		>0,2500	
Geral	-	-	<0,0100	<0,0100	<0,0050	<0,0050	<0,0050	<0,0050

+ = número de outliers por corte.

- = não calculado pelo SAS, pois a amostra é maior que 2000 dados.

Tabela 5 - Estimativas do desvio-padrão tradicional (S) e robustas obtidas de dados de produção de matéria seca (kg/ha) de 16 cortes de 92 cultivares de alfafa

Corte	S	AI		Gini		DAM		Sn		Qn	
		C/ outlier	S/ outlier	C/ outlier	S/ outlier	C/ outlier	S/ outlier	C/ outlier	S/ outlier	C/ outlier	S/ outlier
1	431,4	466,3		434,3		447,3		456,5		440,5	
2 (2)	484,1	354,1	353,7	437,8	398,1	348,3	340,6	374,4	365,3	389,7	380,9
3	405,4	400,2		404,6		399,7		395,2		406,3	
4 (1)	295,6	250,2	250,1	284,4	275,9	232,9	232,2	268,8	261,4	261,9	259,7
5 (1)	317,7	297,7	292,8	305,6	292,8	297,3	298,9	292,5	289,3	293,2	290,5
6	287,5	308,5		287,2		313,9		283,4		291,8	
7	289,1	285,5		288,9		272,5		276,2		291,5	
8 (2)	361,6	308,2	301,8	333,5	304,6	313,2	310,2	305,8	301,2	311,9	305,8
9 (3)	253,2	277,5	277,5	253,8	253,8	272,9	272,9	264,3	264,3	247,7	247,7
10	379,4	393,1		394,8		396,7		383,1		393,9	
11	287,5	288,8		286,4		288,3		288,2		290,8	
12	262,2	265,3		263,2		266,6		264,3		262,7	
13 (1)	267,2	242,9	241,3	254,9	242,8	246,4	245,0	245,2	243,6	243,9	242,1
14 (2)	273,1	189,9		247,5		185,9		199,6		213,9	
15	290,2	290,9		288,9		286,3		282,0		290,7	
16	316,9	335,8		319,5		339,9		334,3		328,4	
Geral	652,9	497,4	496,2	627,4	624,1	499,3	495,6	534,9	532,0	548,2	545,4

"+" = Número de outliers por corte.

Ver módulo Insight do SAS (SAS, 2002-2003).

AI: amplitude interquartilica: $(Q_3 - Q_1)/1,34898$.

Gini: diferença média de Gini.

DAM = desvio absoluto da mediana.

Conclusões

A aplicação das técnicas de análises exploratórias na avaliação da produção de matéria seca de alfafa indicou grande variabilidade entre os cortes em 92 cultivares, o que sugere boas possibilidades de sucesso no melhoramento genético. Observaram-se valores negativos e positivos para os coeficientes de simetria e de curtose, não ajuste dos dados de produção de

matéria seca à distribuição normal e diferença do desvio-padrão clássico em relação aos estimadores robustos para esta estatística. Portanto, na análise de dados de produção de matéria seca em alfafa, o uso de estatísticas descritivas para a retirada de outliers, transformação de dados e uso de desvio-padrão robusto pode ser uma atividade de rotina anterior à escolha do modelo de análise: problema => dados => análises => modelo => conclusões.

Literatura Citada

- BONETT, D.G.; SEIERB, E. A test of normality with high uniform power. **Computational Statistics & Data Analysis**, v.40, n.3, p.435-445, 2002.
- CLEVELAND, W.S. **The elements of graphing data**. New Jersey: AT&T Bell Laboratories; Murray Hill, 1994. 297p.
- COCHRAN, W.G.; COX, D.F. **Deseno experimentales**. Mexico: Trillas, 1978. 661p.
- DIGBY, P.G.N.; KEMPTON, R.A. **Multivariate analysis of ecological communities**. London: Chapman & Hall, 1996. 206p.
- GOWER, J.C.; HAND, D.J. **Biplots**. London: Chapman & Hall, 1996. 277p.
- 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).
- PEARSON, R.K. Exploring process data. **Journal of Process Control**, v.11, n.2, p.179-194, 2001.
- STATISTICAL ANALYSES SYSTEM - SAS. **SAS/INSIGHT User's guide**. versão 9.1.3 - versão para Windows. Cary: SAS Institute, 2002/2003. (CD-ROM).
- SINGHA, A.; NOCERINOB, 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, p.69-86, 2002.
- VILELA, D. Intensificação da produção de leite. **Estabelecimento e utilização da alfafa**. Juiz de Fora: Embrapa Gado de Leite, 1998. 28p. (Documentos, 26).