

A presença de *Outliers* interfere no Teste F e no teste de comparações múltiplas de médias

CHICARELI, L.S.¹; OLIVEIRA, M.C.N. de²; POLIZEL, A.³; NEPOMUCENO, A.L.²

¹Universidade Estadual de Londrina - UEL; ²Embrapa Soja; Universidade Estadual de Maringá - UEM

Um dos fatores que mais contribuem para o desenvolvimento vegetal é a existência de água. Sendo assim, o estresse hídrico interfere diretamente na produtividade da soja, responsável pelo maior percentual de exportação no mercado mundial. Nesse sentido, tem-se tentado encontrar, por meio de pesquisas e experimentos, cultivares tolerantes ao estresse hídrico comparando com outras sensíveis à ausência de água. Na presença desse estresse vários parâmetros fisiológicos são alterados como: número de estômatos na folha, condutância estomática, teor de clorofila, temperatura da folha, área foliar, entre outros. Objetivamos detectar valores discrepantes (*outliers*) que dificultam a significância da estatística F para os efeitos de tratamentos, bem como diferenças estatísticas entre as médias dos tratamentos. Esses valores podem interferir no atendimento dos pressupostos da análise de variância (Anova). De acordo com Banzatto e Kronka (1989) e Gomes (1990), o teste F é um teste unilateral que tem por finalidade comparar estimativas de variâncias, as hipóteses testadas quando aplicamos o teste F são:

$H_0: \sigma_1^2 = \sigma_2^2$, ou seja, as duas populações possuem variâncias iguais.

$H_a: \sigma_1^2 \neq \sigma_2^2$, ou seja, as duas populações não possuem variâncias iguais.

Os testes de comparações múltiplas servem como complemento para o Teste F e têm por objetivo detectar diferenças entre os tratamentos. O teste de Tukey (Δ) é usado para comparar duas médias e verificar se diferem significativamente entre si e é dado por:
$$\Delta = q \times \frac{s}{\sqrt{r}} = q \times s(\bar{m})$$

em que,

q : amplitude total estudentizada (tabelada em função do número de tratamentos e do grau de liberdade do resíduo);

s : desvio padrão residual;

r : número de repetições das médias

Se $\bar{Y}_a \geq \Delta$ então o contraste é significativo e as médias de Y_a diferem entre si.

Verificou-se que a presença de *outliers* em ensaios agrônômicos afeta os resultados dos testes F e Tukey. Os dados utilizados no trabalho são da área de Ecofisiologia da Embrapa Soja. O delineamento experimental foi o de blocos ao acaso com os tratamentos em arranjo fatorial, contendo quatro blocos com três vasos por tratamento dentro de cada bloco, da seguinte forma: a cultivar BR 16 e a linhagem P58, com dois níveis iniciais de estresse (5% e 15% de Umidade Gravimétrica), totalizando 48 parcelas. A fim de encontrar os possíveis *outliers* e avaliar as pressuposições da análise de variância (ANOVA) foi realizado um diagnóstico exploratório pelo software científico SAS - Statistical Analysis System versão 9.1. Os testes utilizados para verificar a normalidade dos erros, a aditividade do modelo, a homocedasticidade de variância dos tratamentos e a independência dos erros foram respectivamente o de Shapiro & Wilk (1965), Tukey (1949), Burr & Foster (1972) e Parente (1984) (Tabela 1, Fig. 1). O teste de comparação múltipla de médias utilizado foi o de Tukey ($p \leq 0,05$). Observamos que embora haja variabilidade mediana (CV = 14,9 %) com os dados originais para a Temperatura da Folha (°C) ocorreu distribuição assimétrica e curtose leptocúrtica. Conforme Oliveira et al. (1998), valores de assimetria e curtose diferentes de zero indicam a presença

de *outlier* que nesse caso tem valor igual a -27,51 (Fig. 1). O teste de Shapiro & Wilk foi significativo rejeitando a hipótese de normalidade. O teste de não aditividade do modelo apresentou significância pela estatística F e desta forma, o modelo é não-aditivo. As variâncias dos tratamentos foram heterogêneas rejeitando-se a hipótese de nulidade (Tabela 1, Fig. 1).

Tabela 1. Estatísticas descritivas, testes de pressuposições da Anova para dados originais Temperatura da Folha °C.

| ESTATÍSTICAS | | Originais |
|-----------------------------|-----------|-----------|
| Coeficiente de variação (%) | | 14,9309 |
| Coeficiente de assimetria | | -5,1876 |
| Coeficiente de curtose | | 32,3422 |
| Shapiro-Wilk | W | 0,4960 |
| Prob (W) | | <0,0001 |
| Não-Aditividade | F | 29,8004 |
| Prob (F) | | 0,0000 |
| Burr-Foster | calculado | 0,8268 |
| | crítico | 0,3715 |

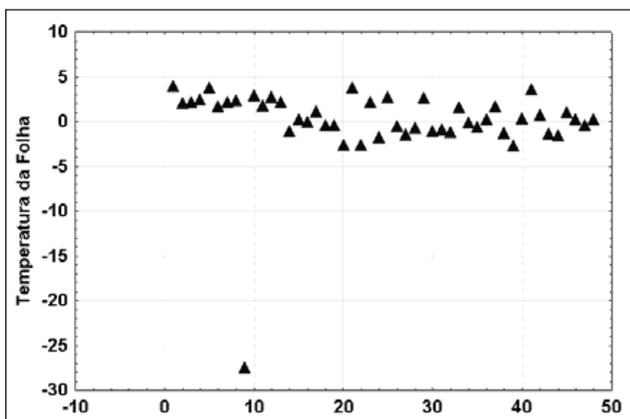


Fig. 1. Gráfico dos Resíduos para a variável Temperatura da Folha.

Em consequência da presença de *outliers* observamos que não houve resposta significativa para a estatística F nos efeitos de tratamentos conforme o modelo estudado. Além disso, houve similaridade nos efeitos de tratamentos para as comparações múltiplas de médias por Tukey a 5 % de probabilidade (Tabelas 1, 2, 3).

Tabela 2. Resultados de Análise de Variância para os dados com *outliers* para a variável Temperatura da Folha °C.

| Causas de variação | Dados Originais | | | |
|--------------------------|-----------------|----------|------|---------|
| | G.L. | QMédio | F | Prob(F) |
| Blocos | 3 | 76,6984 | 3,39 | 0,027 |
| Cultivar | 1 | 34,8502 | 1,54 | 0,2219 |
| Estresse hídrico | 1 | 28,4900 | 1,26 | 0,2686 |
| Cultivar x Estresse | 1 | 19,3548 | 0,85 | 0,3607 |
| Resíduo | 41 | 22,6539 | | |
| Total | 47 | 182,0474 | | |

Tabela 3. Médias para Temperatura da Folha °C na presença de *outliers*.

| Cultivares | Níveis de Estresse | | Médias |
|------------|--------------------|---------|---------|
| | 5% | 15% | |
| BR 16 | 31,16 | 30,89 | 31,02 a |
| P58 | 34,13 | 31,32 | 32,73 a |
| Médias | 32,65 a | 31,11 a | |

Médias seguidas de mesmas letras minúsculas nas linhas e colunas não diferem pelo teste de Tukey com 5 % de probabilidade.

Embora o critério adotado para detectar e retirar *outliers* basei-se em métodos estatísticos é necessária integração entre os pesquisadores para verificar se o mesmo é realmente *outlier* ou é o objeto da pesquisa. Deve-se ter cautela, pois nem sempre valores ruins são discrepantes (Bustos, 1988). Após retirar o valor discrepante obtive-se valores reduzidos para a variabilidade ($CV = 3,17\%$) e todos os pressupostos da Anova foram atendidos (Tabela 4 e Fig. 2). Observamos que os resíduos da Fig. 2 apresentam-se aleatorizados. É importante salientar que se esses procedimentos não são observados corre-se o risco do pesquisador não obter resultados da estatística F significativa para os efeitos de tratamentos e ou recomendar produtos de forma equivocada. Desta forma, esses resultados vêm ao encontro dos obtidos na literatura. Os valores dos coeficientes de assimetria e curtose passaram a ser muito próximos de zero, o modelo é aditivo e as variâncias dos tratamentos homogêneas.

Retirando-se o *outlier* observou-se que os níveis de estresse hídrico foram significativos pelo teste F e os valores médios desses níveis pelo teste de Tukey indicaram que a Temperatura da Folha foi maior na ausência de água apresentando diferença estatística comparando com o nível de 15 % de umidade gravimétrica (Tabelas 5 e 6).

Tabela 4. Valores das estatísticas descritivas, testes de pressuposições da Anova sem η para a variável Temperatura da Folha.

| ESTATÍSTICAS | | <i>S/outlier</i> |
|-----------------------------|-----------|------------------|
| Coeficiente de variação (%) | | 3,1753 |
| Coeficiente de assimetria | | 0,1456 |
| Coeficiente de curtose | | 0,7784 |
| Shapiro-Wilk | W | 0,9804 |
| Prob (W) | | 0,6117 |
| Não-Aditividade | F | 0,0595 |
| Prob (F) | | 0,8084 |
| Burr-Foster | calculado | 0,1183 |
| | crítico | 0,3715 |

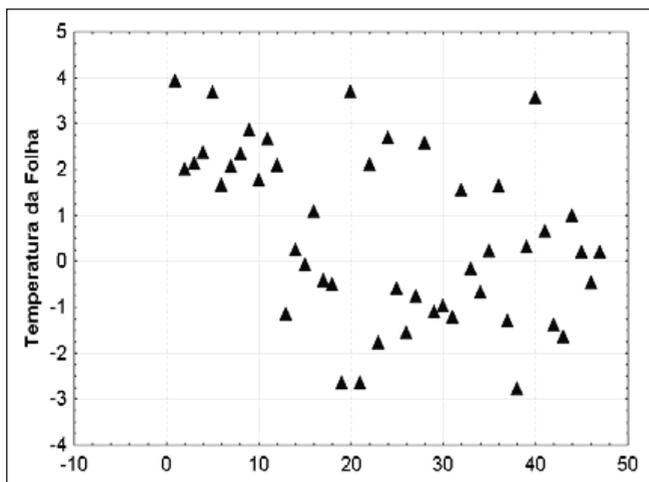


Fig. 2. Gráfico dos Resíduos da variável Temperatura da folha sem *outlier*.

Tabela 5. Resultados de Análise de Variância sem *outliers* para a variável Temperatura da Folha.

| Causas de variação | G.L. | Qmédio | F | Prob(F) |
|---------------------|------|----------|-------|---------|
| Blocos | 3 | 20,5090 | 19,19 | <0,0001 |
| Cultivar | 1 | 2,1773 | 2,04 | 0,1612 |
| Estresse hídrico | 1 | 97,2931 | 91,04 | <0,001 |
| Cultivar x Estresse | 1 | 0,0607 | 0,06 | 0,8128 |
| Resíduo | 40 | 1,0686 | | |
| Total | 46 | 121,1088 | | |

Tabela 6. Médias para Temperatura da Folha °C na ausência de *outliers*.

| Cultivares | Níveis de Estresse | | Médias |
|------------|--------------------|---------|---------|
| | 5% | 15% | |
| BR 16 | 33,93 | 30,89 | 32,37 a |
| P58 | 34,13 | 31,32 | 32,72 a |
| Médias | 34,07 a | 31,10 b | |

Conclui-se que:

- Os *outliers* influenciam nos resultados da estatística F para tratamentos e testes de comparações de médias.
- A ausência de uso do critério apresentado induz o pesquisador a indicar um cultivar e ou produto biológico equivocado.

Referências

BURR, I.W.; FOSTER, L.A. **A test for equality of variances**. West Lafayette: University of Purdue, 1972. 26p. (Mimeo series nº 282).

OLIVEIRA, M.C.N.; CORREA-FERREIRA, B.S. Diagnostico exploratorio na analise de dados entomologicos: efeito do diflubenzurom no consumo alimentar de *Anticarsia gemmatalis* (Hubner). **Anais da Sociedade Entomologica do Brasil**, Londrina, v.27, n.4, p.627-637, dez. 1998.

PARENTE, R.C.P. **Aspectos da análise de resíduos**. 1984. 118f. Dissertação (Mestrado) - ESALQ/USP, Piracicaba.

POLIZEL, A. M. **Avaliações moleculares, morfo-anatômicas e fisiológicas de soja geneticamente modificada com a construção rd29A: DREB1A de Arabidopsis thaliana, visando tolerância à seca.** 2007.125f. Dissertação (Mestrado) – Universidade Estadual de Londrina, Londrina.

SHAPIRO, S.S. & WILK, M.B. An analysis of variance test for normality. **Biometrika**, v.52, p.591-611, 1965.

TUKEY, J.W. One degree of freedom for non-additivity. **Biometrics**, v.5, p.232-242, 1949.