

MÉTODOS ESTATÍSTICOS PARA ANÁLISE ROBUSTA DE DADOS ORIUNDOS DE EXPERIMENTO AGRÍCOLA EM ÁREAS IRRIGADAS

J. A. de Sousa¹, E. A. H. Hoces², H. O. C. Guerra³, R. D. de Lacerda⁴

RESUMO: Geralmente as conclusões dos dados originais de experimentos de campo ou casa de vegetação são extraídas a partir da Análise de variância (ANOVA), entretanto, a utilização desta técnica exige duas pressuposições fundamentais: que os erros sejam distribuídos normalmente e que as variâncias dos diversos tratamentos sejam homogêneas. Entretanto, estas exigências na maioria das vezes são negligenciadas por desconhecimento teórico, fato este que pode levar o pesquisador a conclusões finais equivocadas do seu experimento. Diante do exposto, foi realizada uma análise em dados de um delineamento experimental em blocos inteiramente causalizados com quatro tratamentos de água aplicada ao solo (25, 50, 75 e 100% da ETo) e quatro repetições perfazendo o total de 96 unidades experimentais de abacate comercial (Persea americana Mill.) Cv. Hass, em uma área localizada em Los Molinos, Fazenda no Vale do Peumo na província de Cachapoal, Chile, com coordenadas UTM (299526 E, 6191981 N). Observou-se que o diagnóstico das pressuposições e a transformação Box-Cox viabilizaram uma análise robusta e confiável dos resultados finais do experimento.

PALAVRAS-CHAVE: Homogêneas, análise robusta, abacate comercial

STATISTICAL METHODS FOR ROBUST ANALYSIS OF DATA ORIGINATED FROM AGRICULTURAL EXPERIMENT IN IRRIGATED AREAS

ABSTRACT: Usually, the conclusions originated from field or green house experiments are obtained from Variance Analysis (ANOVA). However, the use of this technique demands two fundamental presuppositions: that the errors must be normally distributed and that the variance of the treatments must be homogeneous. However, most of the time, these exigencies are neglected due to lack of theoretical knowledge, fact that may lead the researcher to equivocated conclusions. Thus, it was made an analysis of an experimental data in entirely random blocks with four soil water treatments (25, 50, 75 e 100% of ETo) and four repetitions summing up 96

¹ Doutor, professor UFCG/CES/UAE. E-mail: jorgeas@ufcg.edu.br

² Ph.D, professor UC/FIA. E-mail: eholzapf@udec.cl

³ Ph.D, professor da UFCG/CTRN/UAEAg. E-mail: hugo_carvalho@hotmail.com

⁴ Doutor, professor do IFRN. E-mail: Rogerio.dantas@ifrn.edu.br

experimental unities of commercial avocado (*Persea americana* Mill.) Cv. Hass, in an area located in Los Molinos farm in the Vale do Peumo, Cachapoal Province, Chile, with UTM coordinates: 299526 E, 6191981 N. It was observed that the presuppositions diagnoses added to Box-Cox transformations proportionated a viable, robust and reliable analysis of the experiment final data.

KEYWORDS: Homogeneous, robust analysis, commercial avocado

INTRODUÇÃO

Cada tipo delineamento experimental é regido por um modelo matemático com restrições e resulta na variável resposta. Para validar os testes de hipótese e outras inferências os modelos matemáticos devem ter suas pressuposições atendidas.

A qualidade de um experimento pode ser avaliada pela magnitude do erro experimental e pelo atendimento das pressuposições do modelo matemático: a) aditividade, ou seja, se existe uma soma de efeitos do modelo; b) os e_{ij} são conjuntamente independentes; c) os e_{ij} apresentam distribuição normal e d) os e_{ij} apresentam variância comum, ou seja, apresenta homocedasticidade de variância (Zimmermann, 2014).

Para verificar se as pressuposições (homogeneidade das variâncias entre os tratamentos e normalidade dos erros) estão sendo satisfeitas, pode-se usar, por exemplo, teste de Lilliefors para verificação da normalidade da distribuição dos erros, teste de Bartlett para verificação da homogeneidade dos erros entre os tratamentos. Além dos diversos testes que podem ser utilizados para diagnosticar estas pressuposições existem ainda os métodos gráficos, onde o pesquisador pode vislumbrar o comportamento dos resíduos de uma maneira mais perceptível e concreta. A verificação destas pressuposições melhora a qualidade da análise dos experimentos, devendo ser aplicado antes de qualquer análise e teste de hipótese que envolva distribuição F (Conagin et al., 1993).

Quando as pressuposições enumeradas não são satisfeitas, uma das alternativas que pode ser utilizada para normaliza os dados e homogeneizar as variâncias é a transformação dos dados por meio da metodologia proposta por Box-Cox (Box e Cox, 1964). Com a nova escala, os dados devem obedecer, aproximadamente, as pressuposições do modelo matemático.

Diante do exposto o objetivo desse trabalho foi realizar o diagnostico de pressuposições do modelo e transformação de Box-Cox, métodos necessários para viabilidade de aplicação e robustez do teste F aplicado amplamente nos experimentos agrícolas.

MATERIAL E MÉTODOS

O estudo foi realizado em um campo comercial de abacate (*Persea Americana Mill*) cv. Hass durante três estações de crescimento consecutivas (2005-2006, 2006-2007 e 2007-2008). O campo está localizado na fazenda "Los Molinos" no Vale do Peumo da Província de Cachapoal, no Chile, coordena a UTM (299526 Este, 6191981 Norte) e altitude de 170 m. A zona tem um clima subúmido mediterrâneo, com temperaturas máximas em janeiro de 32°C, e uma temperatura anual de 14°C, precipitação média anual de 610 mm, concentrada no inverno (junho a setembro). O solo argila arenoso do tipo xerochrept, na superfície, até areia arenosa até a profundidade de 0,9 m; Densidade aparente do solo de 1,20 Mg m⁻³ na superfície a 1,30 Mg m⁻³ a uma profundidade de 0,9 m, capacidade de campo de 0,23 m³ m⁻³ na superfície e 0,21 m³ m⁻³ a uma profundidade de 0,9 m, ponto de murcha permanente de 0,16 m³ m⁻³ na superfície e 0,12 m³ m⁻³ a uma profundidade de 0,9 m, água disponível de 45 a 76 mm, com uma água variável durante a estação, com um mínimo de 1,8 m. Os abacates foram espaçados a 4 m de distância entre plantas com 7 m entre as linhas. Cada linha de plantas foi irrigada usando dois microjet de compensação sem pressão por árvore, localizado a 0,6 m de distância do tronco em ambos os lados na linha, taxa de descarga de 30 Lh⁻¹ cada e diâmetro molhado de 2 m. As taxas de descarga dos emissores foram avaliadas mensalmente em cada estação de irrigação, com valores médios de 30,2 a 29,8 Lh⁻¹ e Coeficiente de Uniformidade Christiane (CUC) de 92,8% nas três estações.

O delineamento experimental utilizado foi em blocos inteiramente casualizado com quatro tratamentos de aplicação de água (25, 50, 75 e 100% do volume teórico de água requerido pela planta, V_{tr}) e quatro repetições. As 16 unidades experimentais foram localizadas em uma área de 0,54 ha. Cada unidade experimental tinha seis plantas, distribuídas em duas filas, com um total de 96 plantas. A irrigação foi aplicada quatro dias por semana (segunda-feira, quarta-feira, sexta-feira e sábado) com base em estimativas diárias da evapotranspiração da cultura.

Para testar o atendimento às pressuposições do modelo matemático e obter o erro experimental foi considerado o modelo matemático do delineamento blocos ao acaso: $y_{ij} = m + t_i + b_j + e_{ij}$, com $i = 4$ tratamentos e, $j = 4$ repetições, onde: m é uma constante associada a todas as observações; t_i é o efeito do i -ésimo tratamento; b_j é o efeito do do j -ésimo bloco; e_{ij} é o erro experimental associado à observação y_{ij} (Storck et al., 2011). Para testar a normalidade e homogeneidade dos erros, realizou-se um diagnóstico gráfico

Para os casos em que ocorreu alguma violação dos pressupostos do modelo matemático, foram realizadas transformações de dados, e os mesmos foram novamente submetidos à análise do atendimento das pressuposições do modelo matemático para verificar se a transformação foi eficiente em adequar a variável ao pressuposto violado. Para transformações da variável, utilizou-se a metodologia de Box e Cox, usada para normalizar os erros e homogeneizar as variações (Box e Cox, 1964). As análises de variância e regressões foram realizadas usando o Software R versão 3.1.0.

RESULTADOS E DISCUSSÃO

Inicialmente foi verificado comportamento diferenciado no atendimento aos pressupostos da normalidade e homogeneidade das variâncias onde aqui ressaltamos os métodos gráficos para variável produção (Draper e Smith, 1998). Percebe-se na figura 1 do gráfico *qxq* que os dados não têm um comportamento normal, Bussab, 2003, afirma que quando os dados têm um comportamento normal, estes devem se dispor ao longo de uma reta. A ausência de normalidade dos erros geralmente causa um aumento no verdadeiro nível de significância do teste de hipótese, resultando em maior probabilidade de erro tipo I do que o especificado pelo nível de significância adotado, além da falta de normalidade ser uma das possíveis causas da heterogeneidade regular dos erros.

Ferreira, Rocha e Mequelino (2012) afirmam que a não normalidade dos erros afeta a eficiência na estimação dos efeitos de tratamentos e implica em perda do poder dos testes e, além disso, há aumento do erro no nível de significância dos testes. Já Cochran (1947) afirma que é possível que a não normalidade seja acompanhada de menor eficiência na estimação dos efeitos e haja perda correspondente de poder no teste F.

A figura 2. Indica claramente que a suposição de homocedasticidade (mesma variância) não está satisfeita, é considerável o comportamento crescente da dispersão dos resíduos ao longo dos dados ajustados. Adicionalmente, vários pesquisadores relatam que a quebra da homogeneidade da variância dos erros pode afetar o desempenho do teste F. Cochran (1947) revela que, se for efetuada a análise de variância quando as variâncias dos erros não forem homogêneas, será quase certo que ocorrerá perda de eficiência na estimação dos efeitos de tratamentos, e haverá, também, uma perda de sensibilidade nos testes de significância: quanto maiores forem as diferenças na variância, maiores serão estas perdas. Scheffée (1959) afirma que o teste F é robusto para a não-normalidade, contudo este não apresenta o mesmo

desempenho sob heterocedasticidade associada à distribuição normal, mostrando-se liberal neste caso. Zhang (2014) complementa destacando que, nesta situação, mesmo para grandes amostras, o teste F continua aceitando H_0 mesmo quando esta hipótese não é verdadeira. Almeida, Elían e Nobre (2008) afirmam que o teste F no contexto da análise de variância com um fator para comparar médias de populações normais independentes apresenta desvios no que tange ao tamanho do teste quando os erros possuem variâncias populacionais diferentes. O teste baseado na estatística F é sensível à falta de homogeneidade de variâncias, pois, sob heterocedasticidade, o tamanho real do teste não coincide com o nível de significância fixado pelo pesquisador. Desta forma, o resultado do teste F fica bastante comprometido quando as variâncias dos erros são heterocedásticas.

O valor da potencia λ que maximizou a função foi de 0,2 (Figura 3), indicando que a melhor transformação para os dados foi a fração:

$$f(y) = \frac{y^{0,2}-1}{0,2}, \lambda \neq 0 \quad (1)$$

Onde: y = dados de produção

A transformação de Box-Cox foi eficiente para dados da produção de abacate, permitindo uma redução na variabilidade (Figura 5), principalmente contribuindo para uma distribuição mais homogenia dos resíduos em função da variável em estudada. Esse resultado também foi obtido por Couto *et al.* (2009) estudando a cultura da abobrinha italiana, com a transformação raiz quarta, evidenciando a eficácia da metodologia de transformação de dados na homogeneidade das variâncias.

CONCLUSÕES

Dessa forma, fica evidente que a transformação adequada é uma possibilidade que permite aproximação de normalização dos dados e estabilização das variâncias ou, pelo menos reduzi a heterocedasticidade, auxiliando nas análises estatísticas e interpretações dos resultados deste tipo de experimento, sem comprometer a precisão experimental.

REFERENCIAS BIBLIOGRAFICAS

ALMEIDA, A.; ELIAN, S.; NOBRE, J. Modificações e alternativas aos testes de Levene e de Brown e Forsythe para igualdade de variâncias e médias. Revista Colombiana de Estatística. (Bogotá), v. 31, n. 2, p. 241-260, 2008.

BOX GEP; COX DR. 1964. An analysis of transformations. *Journal of the Royal Society* 26: 211-252.

BUSSAB, W. O. & MORETTIN, P. A. (2003), *Estatística Básica*, 4a Edição, Atual Editora: São Paulo.

COCHRAN, W.G. Some consequences when the assumptions for the analysis of variance are not satisfied. *Biometrics*, v. 3, n. 1, p. 22-38, 1947.

CONAGIN, A.; NAGQAI, V.; IGUE, T; AMBRÓSIO, L.A. Efeito da falta de normalidade em testes de homogeneidade de variâncias. *Bragantia*, Campinas, 52(2):173-180, 1993.

COUTO MRM; LÚCIO AD; LOPES SJ; CARPES RH. 2009. Transformação de dados em experimentos com abobrinha italiana em ambiente protegido. *Ciência Rural* 39: 1701- 1707

DRAPER, N. E SMITH, H. (1998), *Applied Regression Analysis*. 3rd. John Wiley, New York.

FERREIRA, E. B.; ROCHA, M. C.; MEQUELINO, D. B. Monte Carlo evaluation of the ANOVA's F and Kruskal-Wallis tests under binomial distribution. *Sigmae*, v. 1, n. 1, p. 126 - 139, 2012.

SCHEFFÉ, H. *The analysis of variance*. New Yor: Wiley, 1959. 478p.

STORCK L; GARCIA DC; LOPES SJ; ESTEFANEL V. 2011. *Experimentação vegetal*. Santa Maria: UFSM. 200p

ZHANG, G. A parametric bootstrap approach for one-way ANOVA under unequal variances with unbalanced data. *Communications in Statistics - Simulation and Computation*, p. 37- 41, 2014.

ZIMMERMANN, F. J. P. Efeito de Heterogeneidade de variância e distribuição de probabilidade dos dados sobre o poder e tamanho do teste F. *Revista Agropecuária Brasileira*, v. 22, n. 11, p. 1209-1213, 2014.

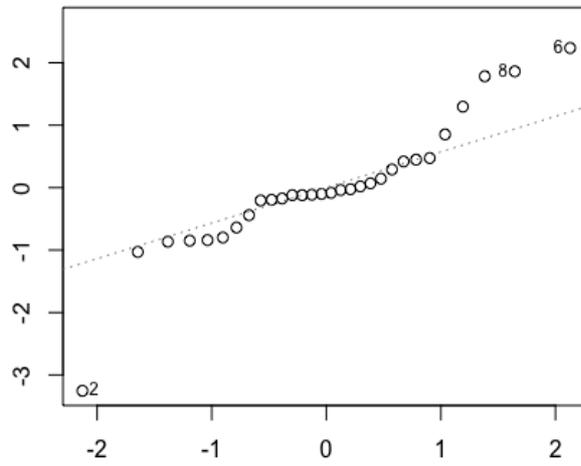


Figura 1. Gráfico qxq (normalidades) para os resíduos do modelo antes da transformação

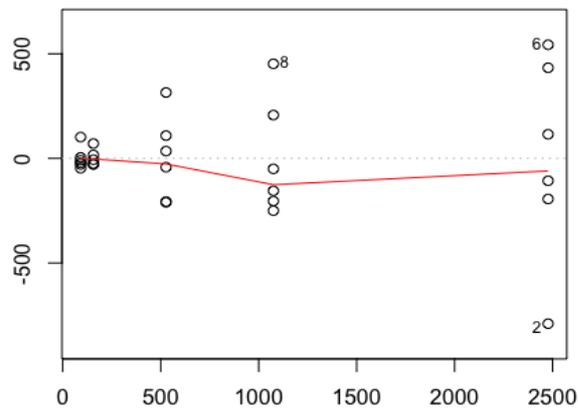


Figura 2. Gráfico dos resíduos contra os valores ajustados antes da transformação

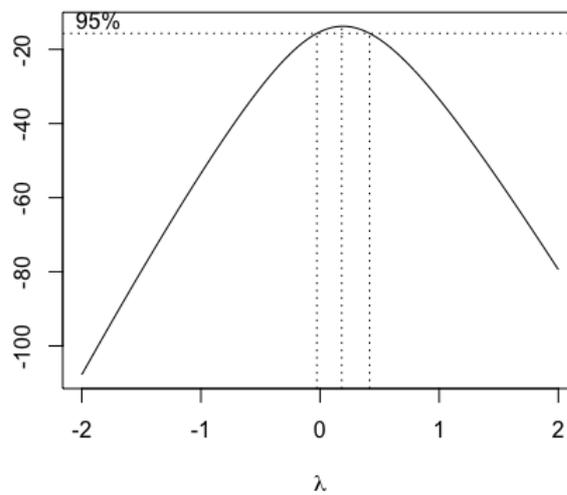


Figura 3. Função de Box-Cox para determinar um melhor lambda para transformação

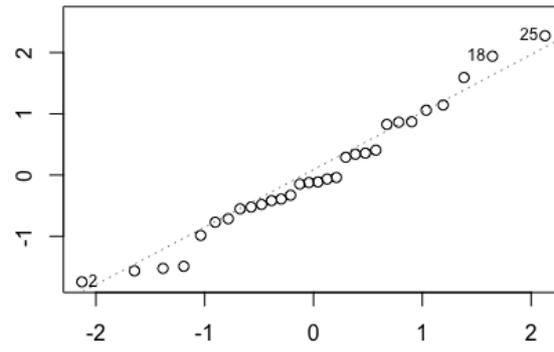


Figura 4. Gráfico *qxq* (normalidades) para os resíduos do modelo após transformação

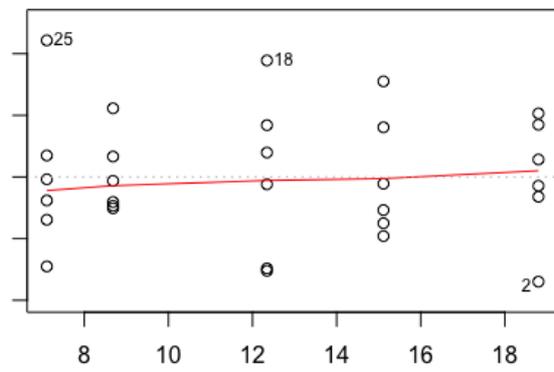


Figura 5. Gráfico dos resíduos contra os valores ajustados após transformação