

Análise da precisão do teste F corrigido em dados longitudinais

Wederson Leandro Ferreira¹, Roberta Araujo Spigolon²

1 Introdução

Dados provenientes de medidas repetidas, obtidos ao longo do tempo, usualmente são analisados estatisticamente pela análise de variância no esquema de parcela subdividida no tempo. Todavia, devido a esta característica inerente aos mesmos à pressuposição de homocedasticidade, ou seja, a igualdade das variâncias dos erros, requerida por esta análise, pode não ser atendida. Isto é, eles podem estar correlacionados pela sua natureza, porque os dados medidos em tempos mais próximos tendem a ser mais correlacionados que os medidos em tempos mais distantes e assim violando a pressuposição de homocedasticidade das variâncias dos erros requerida por esta análise (FERREIRA e MORAIS, 2013).

Sendo assim, outras metodologias de análises devem ser preteridas e uma alternativa possível é optar pela correção dos graus de liberdade do teste F das fontes de variações alocadas na subparcela da análise univariada. Assim, quando se utilizar a análise univariada em esquema de parcelas subdivididas no tempo e a matriz de covariância dos erros intraindivíduos não for homocedástica, uma possível alternativa é ajustar os graus de liberdade do teste F para os fatores presentes na subparcela (MALHEIROS, 2004). Para realizar este ajuste, duas propostas foram apresentadas, respectivamente, por Geisser e Greenhouse (1958) e Huynh e Feldt (1976).

Diante da importância do uso destas correções, objetivou-se analisar a precisão do teste F, via simulação computacional, das respectivas correções propostas, entretanto estipulando os cenários de balanceamento e desbalanceamento dos dados e, ainda, pressupondo erros com distribuição normal.

2 Material e métodos

A análise de dados medidos ao longo do tempo no esquema experimental de parcela subdividida no tempo, pressupõe que a estrutura da matriz de covariância seja homocedástica. Já sob o enfoque dado por Littell et al. (2006) em experimentos utilizando medidas repetidas no tempo, é possível que a estrutura da matriz de covariância se diferencie, impondo uma estrutura de correlação distinta da descrita anteriormente. Uma alternativa para utilizar a matriz de covariância dos erros intraindivíduos, mesmo não homocedástica, é proceder ao ajuste dos graus de liberdade do teste F para os fatores presentes na subparcela.

Autores como Geisser e Greenhouse (1958) e Huynh e Feldt (1976) sugerem algumas correções nos graus de liberdade para o fator tempo e para a interação entre tratamentos e tempos, possibilitando que esses testes possam ser usados, ainda que de forma aproximada, mesmo que a condição de homocedasticidade de variância não seja satisfeita. As correções propostas pelos respectivos autores são:

¹ Docente, IFFarroupilha, *Campus* Alegrete, Alegrete, RS, Brasil. e-mail: wederson.ferreira@iffarroupilha.edu.br.

² Aluna do curso de Engenharia Agrícola, IFFarroupilha/Unipampa, *Campus* Alegrete, Alegrete, RS, Brasil. Bolsista da Fapergs. e-mail: robertaaspigolon@gmail.com.

2.1 Ajuste de Geisser e Greenhouse, dado por

$$\hat{\varepsilon} = \frac{[\text{tr}(\text{CSC}')]^2}{(t-1)\text{tr}(\text{CSC}')^2} \quad (1)$$

ou, ainda, admitindo $A_{qq} = \text{CSC}'$, com $q = t - 1$ contrastes ortogonais normalizados, sobre t medidas repetidas e a_{ii} , com $i = 1, 2, \dots, q$, as raízes latentes de A_{qq} [1].

O fator de correção pode ser alterado por

$$\hat{\varepsilon} = \frac{(\sum_{i=1}^q a_{ii})^2}{(t-1) \sum_{i=1}^q \sum_{j=1}^q a_{ij}^2}. \quad (2)$$

2.2 Ajuste de Huynh e Feldt, dado por

$$\tilde{\varepsilon} = \frac{[N(t-1)\hat{\varepsilon} - 2]}{(t-1)[N - g - (t-1)\hat{\varepsilon}]} \quad (3)$$

em que N é o número total de indivíduos, g é o número de níveis do fator da parcela e t é o número de medidas repetidas.

No estudo de simulação foram realizadas 10.000 simulações no software R, sendo considerado o delineamento inteiramente casualizado com quatro níveis para o fator(tratamento) e cinco para o tempo e, ainda, pressupondo erros normais para o modelo (R CORE TEAM, 2018). Na formulação do modelo matemático para o delineamento considerado, foi utilizado o esquema de parcela subdividida no tempo, com efeito não significativo da hipótese nula para a interação fator x tempo.

3 Resultados e discussão

Pelo estudo realizado, utilizando o nível de significância de 5, 10 e 20%, observou-se os seguintes percentuais de rejeição da hipótese nula de não-significância da interação fator x tempo, tabela 1, pelo teste F, tradicional, pelo teste F corrigido pela correção de Geisser e Greenhouse (GG) e pela correção de Huynh e Feldt (HF), nos seguintes cenários aventados:

Tabela 1. Percentuais de rejeição da hipótese nula de não-significância do teste F da interação fator x tempo.

Distribuição normal			
Dados balanceados (5 repetições)			
Níveis de significâncias (%)			
	5	10	20
Testes utilizados			
F Tradicional	18	27	35
GG	8	15	21
HF	12	15	23
Dados desbalanceados (3, 5 , 7 repetições)			
Níveis de significâncias (%)			
	5	10	20
F Tradicional	24	30	35
GG	15	23	31
HF	21	30	35

Fonte: Os autores.

Pelos resultados, observa-se que a correção de GG impôs maior precisão, menor percentual de rejeição da hipótese nula da interação fator x tempo, no teste F independentemente de balanceamento, ou não, dos dados. Isto, em função do cenário planejado ter simulado não significância desta hipótese nula. Já a correção de HF impôs imprecisão, sobretudo na presença de desbalanceamento dos dados. Sendo assim, sugere-se que pesquisas semelhantes ao cenário proposto ao utilizar a análise tradicional, opte por corrigir o teste F das fontes de variações da subparcela pela correção de GG.

Malheiros (2001) verificou, via estudo de simulação, que o teste F da análise da variância são imprecisos com a correção de HF, em cenários distintos do estudo ora realizado, mas concordantes na indicação da mesma forma de correção, semelhante à indicação de Vieira (2007).

4 Conclusões

Nos cenários simulados, o teste F foi mais preciso utilizando a correção de Geisser e Greenhouse.

Recomenda-se utilizar esta correção em trabalhos práticos, desde que o delineamento possua as diretrizes do simulado computacionalmente.

5 Agradecimentos

Ao instituto Federal Farroupilha, em especial *campus* Alegrete, pelo apoio na condução deste projeto e a FAPERGS pela concessão de bolsas de iniciação científica.

6 Referências bibliográficas

GEISSER, J.; GREENHOUSE, S. W. An extension of Box's results on the use of the F distribution in multivariate analysis. *Annals of Mathematical Statistics*, Ann Arbor, v. 29, p. 855-891, 1958.

FERREIRA, W. L.; MORAIS, A.R. . Análise da influência do café no ganho de peso de animais (ratos) por meio de modelo linear misto. *Revista Brasileira de Biometria*, v. 31, p. 485-500, 2013.

HUYNH, H.; FELDT, L. S. Estimation of the Box correction for degrees of freedom from sample data in the randomized block and split-plot . *Journal of Educational Statistics*, Alexandria, v. 1, n. 1, p. 69-82, 1976.

LITTELL, R. C. et al. SAS mixed models. 2nd ed. Cary: SAS Institute, 2006. 814 p.

MALHEIROS, E. B. Precisão da análise de experimentos com medidas repetidas no tempo usando procedimentos do SAS. *Revista de Matemática e Estatística*, São Paulo, v. 19, n. 2, p. 253-272, 2001.

MALHEIROS, E. B. Precisão de teste F univariados usados em experimentos com medidas repetidas no tempo, quando a condição de esfericidade da matriz de covariâncias não é verificada. *Revista de Matemática e Estatística*, São Paulo, v. 22, n. 2, p. 23-29, 2004.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

VIEIRA, F. T. P. A. et al. Uma abordagem multivariada em Silvicultura com *Leucaena leucocephala* (Lam.) de Wit no Agreste de Pernambuco. *Ciência Florestal*, Santa Maria, v. 17, n. 4, p. 333-342, 2007.