

ANÁLISE DA PRECISÃO DO TESTE F CORRIGIDO EM DADOS LONGITUDINAIS

Roberta Araujo Spigolon¹

Wederson Leandro Ferreira²

Resumo:

: A análise estatística de dados provenientes de medidas repetidas, comumente é realizada pela análise de variância no esquema de parcela subdividida no tempo. Ocorre que geralmente eles podem estar correlacionados por sua natureza, violando a pressuposição de homocedasticidade das variâncias dos erros, requerida por esta análise. Confirmada esta quebra de pressuposição, outras metodologias de análises podem ser utilizadas. Uma das alternativas possíveis é optar pela correção dos graus de liberdade do teste F das fontes de variações alocadas na subparcela da análise. Duas propostas foram apresentadas, respectivamente, por de Geisser e Greenhouse (1958) e Huynh e Feldt (1976). 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. Pelos estudos realizados, observou-se que a correção de GG impôs maior precisão no teste F, independentemente de balanceamento, ou não. Já a correção de HF impôs imprecisão, sobretudo na presença de desbalanceamento dos dados.

Palavras-chave: Análise de variância, distribuição normal, teste de precisão

Modalidade de Participação: Iniciação Científica

ANÁLISE DA PRECISÃO DO TESTE F CORRIGIDO EM DADOS LONGITUDINAIS

¹ Aluno de graduação. robertaaspigolon@gmail.com. Autor principal

² Docente. wederson.ferreira@iffarroupilha.edu.br. Orientador

ANÁLISE DA PRECISÃO DO TESTE F CORRIGIDO EM DADOS LONGITUDINAIS

1 INTRODUÇÃO

A análise estatística de dados provenientes de medidas repetidas, comumente é realizada pela análise de variância no esquema de parcela subdividida no tempo. Ocorre que geralmente estes dados podem estar correlacionados pela natureza destes, violando a pressuposição de homocedasticidade das variâncias dos erros, requerida por esta análise. Confirmada esta quebra de pressuposição, outras metodologias de análises podem ser utilizadas. Uma das alternativas possíveis é optar pela correção dos graus de liberdade do teste F das fontes de variações alocadas na subparcela da análise. 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 METODOLOGIA

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, tornando a distribuição F estatística aproximada.

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, ou seja, que os erros do modelo não possuam distribuição normal com variâncias totalmente iguais.

As correções propostas pelos respectivos autores são:

- 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} . 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) 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.

Pelo exposto, torna-se justificável realizar o estudo ora proposto, pois é carente na literatura indicações das análises estatísticas mais plausíveis a orientar os estatísticos/pesquisados na condução das pesquisas sobre o assunto. Tornando-se assim, necessário realizar estudos que possam subsidiar indicações das formas de análises mais pertinentes ao contexto ora proposto.

No estudo de simulação realizado, foram realizadas 10.000 simulações no software R. Sendo considerado o delineamento inteiramente casualizado, 4 níveis para o fator da parcela e 5 níveis para o tempo e pressupondo erros normais para o modelo. Isto utilizando o esquema de parcela subdividida no tempo para a formulação do modelo teórico matemático e ainda pressupondo, na simulação, não significância da hipótese nula da 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.

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 à correção de GG impôs maior precisão no teste F, independentemente de balanceamento, ou não, dos dados, devido aos valores mais baixos da proporção de rejeição da hipótese nula da não-significância da interação de fator x tempo. Já à correção de HF impôs imprecisão, sobretudo na presença de desbalanceamento dos dados, isto em função dos valores altos da proporção de rejeição da citada hipótese nula, que aproxima-se muito dos percentuais do teste F tradicional.

Portanto, sugere-se que pesquisas semelhantes ao cenário computacional proposto, ao preterir 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 ora estudado, mas concordantes na indicação da mesma forma de correção, semelhante a indicação de Vieira (2007).

4 CONSIDERAÇÕES FINAIS

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

Recomenda-se à correção de Geisser e Greenhouse em trabalhos práticos, desde que o delineamento possua as diretrizes do simulado computacionalmente.

REFERÊNCIAS

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.

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.

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 experimentos Silvicultural com *Leucaena leucocephala* (Lam.) de Wit no Agreste de Pernambuco. **Ciência Florestal**, Santa Maria, v. 17, n. 4, p. 333-342, 2007.