

Ajuste de modelo heteroscedástico em estudo toxicológico usando o SAS

Luzia Pedroso de Oliveira^{1 2}

César Gonçalves de Lima³

Maria Célia Siqueira⁴

Vera Lúcia Scherholz Salgado de Castro⁵

1 Introdução

A principal motivação desse trabalho é apresentar uma aplicação, em estudo toxicológico, de modelo de análise de variância com dois fatores com variâncias residuais heterogêneas, usando os procedimentos MIXED e PLM do SAS.

O experimento foi conduzido no laboratório de ecotoxicologia e biossegurança da EMBRAPA Meio Ambiente e teve como um de seus objetivos avaliar os efeitos de sexo e dose na quantidade de nanopartículas encontradas nos tecidos dos fêos de ratos Wistar que receberam diariamente ao longo de 28 dias, via oral, doses repetidas de um nanocomposto.

O modelo heteroscedástico foi ajustado usando o procedimento MIXED. Os desdobramentos da interação significativa, entre os níveis dos fatores sexo e dose, foram feitos usando o procedimento PLM que possibilitou um maior detalhamento dos mesmos.

A tendência crescente das concentrações médias de nanopartículas nos fêos dos machos em relação às doses do nanocomposto, foi descrita por um modelo de regressão linear ponderada. Nesse caso, ainda que o procedimento MIXED pudesse ser usado, o procedimento REG, por ser mais específico para esse tipo de análise, disponibilizou gráficos mais completos.

2 Material e métodos

No experimento, 5 machos e 5 fêmeas foram atribuídos aleatoriamente para pertencer ao grupo controle ou receber uma das 3 doses do nanocomposto (1, 3 ou 5 µg/ml/100g de peso corpóreo). Ao 29º dia esses ratos foram submetidos a uma autópsia geral e seus órgãos e tecidos foram dissecados e pesados. As concentrações do nanocomposto encontradas nos órgãos e tecidos foram obtidas pela análise ICP-OES -Espectrofotometria de emissão óptica

¹ Doutoranda do PPGEEA ESALQ/USP. E-mail: luziapoliveira@usp.br.

² Agradecimento a CAPES pelo apoio financeiro.

³ ZAB - FZEA/USP.

⁴ Doutoranda do PPGQ UFSCar/EMBRAPA São Carlos.

⁵ EMBRAPA Meio Ambiente.

(Perkin Elmer Optima 7000 DV), sendo, nesse trabalho estudadas as concentrações nos tecidos dos íleos.

O modelo usual de análise de variância com dois fatores assume erros com distribuição normal, independentes e com variâncias homogêneas. A análise gráfica dos resíduos fornece bons indicativos para a verificação dessas suposições.

O fato de não considerar variâncias heterogêneas quando elas estão presentes pode levar a inferências ineficientes e possivelmente enganosas sobre os efeitos fixos do modelo. Littell, et al (2006) apresentam uma extensão do teste de LEVENE para testar homogeneidade de variâncias em experimentos com mais de um fator usando o SAS. Nessa abordagem ajusta-se um modelo de análise de variância incluindo os efeitos principais e interações, considerando como variável resposta, os valores absolutos dos resíduos obtidos do modelo com variâncias homogêneas. No caso da interação ser significativa, um modelo incluindo variâncias diferentes para as combinações dos níveis dos fatores deve ser o ponto de partida. Caso contrário, deve ser verificado se algum efeito principal é significativo e em caso afirmativo deve-se considerar inicialmente variâncias diferentes para seus níveis.

O modelo de análise de variância com dois fatores e variâncias residuais heterogêneas para as combinações dos níveis dos fatores é dado por:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk}; e_{ijk} \sim N(0, \sigma_{ij}^2) \quad (1)$$

sendo, y_{ijk} o valor observado na repetição k do nível i do fator 1 e nível j do fator 2;

μ constante comum a todas as observações; α_i efeito do nível i do fator 1;

β_j efeito do nível j do fator 2; $(\alpha\beta)_{ij}$ efeito da interação entre os níveis dos dois fatores;

e_{ijk} erro aleatório associado à k ésima observação dos níveis i do fator 1 e j do fator 2;

σ_{ij}^2 componente de variância associado aos níveis ij dos fatores.

Simplificações do modelo acima são obtidas considerando, por exemplo, os componentes de variância associados aos níveis de um dos fatores, ou ainda, a combinações dos níveis desse fator.

Nos ajustes dos modelos utilizou-se os procedimentos do SAS. Uma das facilidades desse programa é o recurso `ods graphics` disponível nos procedimentos. Uma vez ativado, exibe painéis com os principais gráficos de resíduos, análise de influência, diagramas de caixa, gráficos de médias, curvas ajustadas etc. O procedimento `PLM`, disponível a partir da versão 9.22, complementa alguns procedimentos, dentre eles o `MIXED`. Por exemplo, possibilita realizar os desdobramentos de interações significativas, exibindo as tabelas de médias com as “famosas” letras, cujo *layout* é preferido, em algumas áreas, para resumir os

resultados dos testes de comparações múltiplas. Os gráficos também são mais completos no PLM, pois incluem os erros padrões e os intervalos de confiança das médias e das curvas ajustadas.

Na busca por um modelo mais parcimonioso para descrever o processo que gerou os dados, geralmente são utilizados os testes da razão de verossimilhanças (TRV), que possibilitam a comparação de modelos encaixados dois a dois, com números diferentes de componentes de variância. De acordo com West, et al (2007), dentre outros, uma vez que os componentes de variância encontram-se no limite do espaço paramétrico torna-se necessário o uso de misturas de χ^2 para obtenção dos níveis descritivos do teste (valores p). É importante destacar que no TRV fornecido pelo PROC MIXED, que compara o modelo ajustado com o modelo com variâncias homogêneas, não é levada em conta essa mistura de χ^2 .

Para testar os efeitos fixos do modelo, Speed (2013), recomenda utilizar juntamente com a opção `method=ml`, a opção `ddfm=KR` na declaração do modelo pois, com essa última, introduz-se um fator de inflação nas variâncias dos efeitos fixos e considera-se conjuntamente a aproximação de Satterthwaite para os graus de liberdade.

Se um dos fatores for quantitativo torna-se necessário descrever a tendência das médias ajustadas em função dos níveis desse fator. Milliken e Johnson (2002), apresentam um exemplo de regressão linear ponderada para descrever o comportamento médio considerando pesos iguais a $1/(\text{erro padrão da média ajustada para o nível } i \text{ do fator quantitativo})^2$.

3 Resultados e discussões

A Figura 1, apresenta o gráfico com os valores observados e médias das concentrações (em ppm) encontradas nos tecidos dos íleos. Já de início tem-se um indicativo de aumento da variabilidade das concentrações com o aumento da dose.

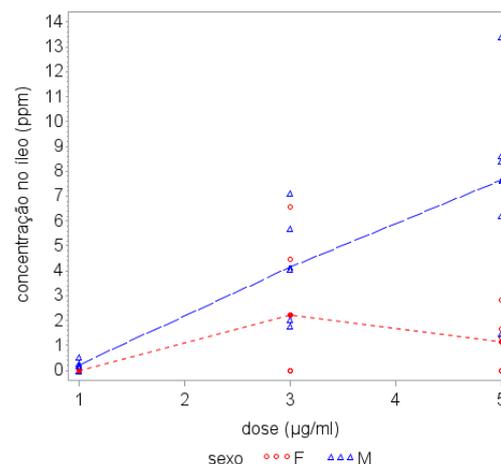


Figura1: Valores observados e médias das concentrações encontradas nos íleos.

A aplicação da extensão de LEVENE no modelo de análise de variância contendo os efeitos fixos de sexo, dose e interação $\text{sexo} \times \text{dose}$ resultou na não rejeição das hipóteses de

variâncias heterogêneas, tanto para as combinações dos níveis dos dois fatores, como também para os níveis do fator dose. Entretanto, o ajuste do primeiro modelo resultou num problema de máxima verossimilhança infinita, mesmo alterando os valores iniciais dos parâmetros. Dessa forma, considerou-se o modelo com variâncias heterogêneas para os níveis de dose, sendo os comandos utilizados, incluindo análise de resíduos e iterativa de influência, apresentados a seguir.

```
ods graphics on;
proc mixed data=concentracoes;
  class sexo dose; model conc_ileo = sexo dose sexo*dose/ ddfm=KR solution
  outpm= res residual influence(effect=rato(dose*sexo) iter=4);
  repeated / group=dose; store sasuser.informacoes; run;
ods graphics off;
```

De acordo com o resultado do TRV a hipótese de variâncias homogêneas para doses foi rejeitada ($p < 0,0001$). Um modelo aninhado ao anterior, foi ajustado considerando apenas duas variâncias residuais diferentes, uma para a dose 1 e outra para as doses 3 e 5 $\mu\text{g/ml}$. Não houve diferença significativa entre esses modelos, de acordo com o TRV ($p = 0,3383$) e, portanto, pelo princípio da parcimônia o modelo escolhido foi esse último.

A interação $\text{sexo} \times \text{dose}$ foi significativa ($p = 0,0037$) e assim sendo, foram feitos os desdobramentos para estudar os efeitos das doses em cada sexo e os efeitos dos sexos em cada uma das doses do nanocomposto. Os comandos utilizados são apresentados a seguir.

```
ods graphics on;
proc PLM restore=sasuser.informacoes; *sendo sasuser.informacoes referente ao
  modelo escolhido e usando method=ml e ddfm=KR;
  lsmeans sexo*dose/cl adjust=tukey plots=all;
  slice sexo*dose / sliceby=sexo cl lines adjust=tukey;
  slice sexo*dose / sliceby=dose cl lines adjust=tukey;
  effectplot/limits; ods output lsmeans= LSM_conc_ileo; run;
ods graphics off;
```

Houve diferenças, a 5% de significância, entre as concentrações médias nos íleos de machos e fêmeas, que receberam as doses 1 ($p = 0,0458$) e 5 $\mu\text{g/ml}$ ($p = 0,0010$).

Não houve diferenças significativas entre as concentrações médias encontradas nos íleos das fêmeas que receberam as diferentes doses do nanocomposto ($p = 0,1373$), já com relação aos machos a concentração média encontrada na dose 1 foi diferente das demais.

Para os machos, a tendência crescente das concentrações médias de nanopartículas em relação às doses do nanocomposto, foi descrita por uma reta com inclinação igual a 1,88, e assim, a um aumento de 1 $\mu\text{g/ml}$ na dose do nanocomposto corresponde um aumento de 1,88 ppm na concentração de nanopartículas encontrada no íleo. A seguir os comandos utilizados:

```
data LSM_conc_ileo; set LSM_conc_ileo; wt=1/(stderr**2); run;
ods graphics on;
proc reg data=LSM_conc_ileo; by sexo; weight wt; model estimate=dose; run;
ods graphics off;
```

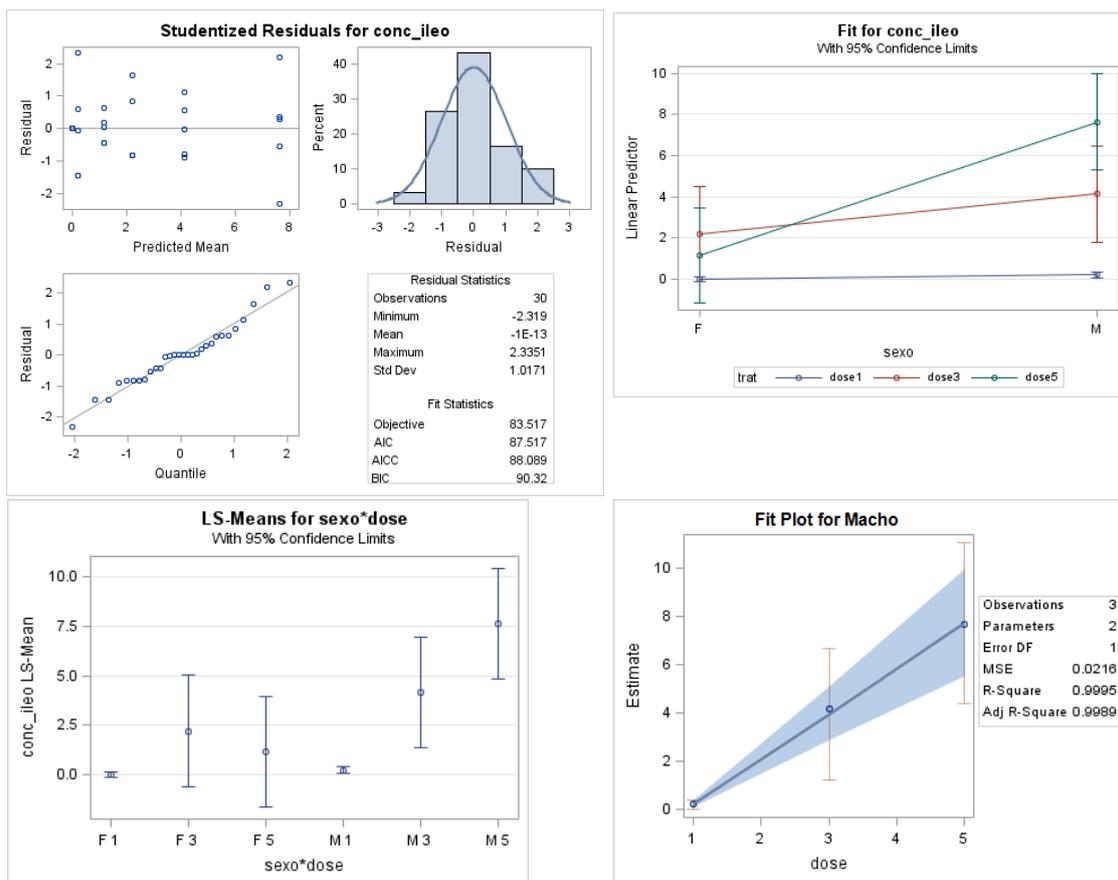


Figura 2 – Principais gráficos obtidos com os comandos descritos ao longo do trabalho

4 Conclusões

O modelo heteroscedástico possibilitou obter um intervalo de confiança mais preciso para a concentração média de nanopartículas nos íleos dos ratos submetidos à dose de 1 $\mu\text{g/ml}$ do nanocomposto, resultando em inferências mais eficientes dos efeitos fixos do modelo.

5 Bibliografia

LITTELL, R. C.; MILLIKEN, G. A.; STROUP, W. W.; WOLFINGER R. D.; SCHABENBERGER, O. **SAS® for mixed models**, Second Edition. Cary, NC: SAS Institute Inc. 2006. 813 p.

MILLIKEN, G. A.; JOHNSON, D. E. **Analysis of messy data**. Volume III: Analysis of covariance. Boca Raton, Chapman & Hall/CRC. 2002. 605 p.

SAS/STAT® 13.1 User's guide. The PLM procedure. SAS Institute Inc. 2013. 6246p.

SPEED, M. **Applied mixed models for professors course notes**. Cary, NC: SAS Institute Inc. 2013.

WEST, B. T.; WELCH, K. B.; GALECKI, A. T. **Linear mixed models: a practical guide using statistical software**. Boca Raton, Chapman & Hall/CRC. 2007. 339p.