

# Análise de experimento fatorial desbalanceado nos ambientes SAS e R

Alfredo José Barreto Luiz<sup>1</sup>

Aline de Holanda Nunes Maia<sup>1</sup>

Ricardo Antônio Almeida Pazianotto<sup>1</sup>

Fernanda Garcia Sampaio<sup>1</sup>

## 1 Introdução

Em experimentos com animais nos quais as variáveis respostas são de difícil obtenção por restrições metodológicas, é comum acontecer que ao final dos procedimentos laboratoriais, os tratamentos possuam número de repetições diferentes. Isso caracteriza uma situação de desbalanceamento, com implicações na análise de variância e testes para avaliação dos efeitos dos tratamentos. Em experimentos fatoriais não balanceados, a matriz X do delineamento experimental possui colunas não ortogonais, o que inviabiliza a estimação dos efeitos principais e as interações de forma independente: para cada efeito a ser estimado é necessário, então, realizar um ajustamento aos demais efeitos presentes no modelo. Em softwares estatísticos como o SAS/STAT<sup>®</sup> do *Statistical Analysis System* [3] e o R [1], existem opções específicas para análise desses experimentos desbalanceados, via cálculo das somas de quadrados apropriadas. Neste trabalho, apresentamos e discutimos um exemplo de análise de dados desbalanceados oriundos de um experimento para avaliar o efeito combinado das concentrações de CO<sub>2</sub> e de cobre sobre parâmetros bioquímicos e defesas antioxidantes do peixe pacu, *Piaractus mesopotamicus*, que é uma das principais espécies cultivadas no Brasil [2].

Vários compostos e mudanças ambientais influenciam a sanidade animal nos sistemas aquícolas. Nas últimas décadas, as implicações na qualidade ambiental relacionadas aos diferentes fatores abióticos, como por exemplo, o dióxido de carbono, e ao uso de produtos químicos, como por exemplo, o sulfato de cobre, têm sido extensivamente estudadas em aquicultura. Entretanto, poucos dados estão disponíveis sobre os efeitos combinados do CO<sub>2</sub> e do cobre sobre a biologia, fisiologia e bioquímica dos peixes tropicais cultivados[2].

Os dados aqui analisados buscavam investigar os efeitos do CO<sub>2</sub> e do CO<sub>2</sub> combinado com sulfato de cobre em pacu, por meio da análise, em laboratório, de diversas características bioquímicas dos indivíduos.

Originalmente, foram obtidas e analisadas 25 variáveis em cada peixe. Algumas das variáveis só podem ser medidas em órgãos dos peixes que são muito pequenos e em determinados indivíduos pode não ser possível a extração de material suficiente para o método de análise. Além disso, algumas análises são extremamente complexas e demoradas, exigindo que as amostras sejam congeladas e analisadas posteriormente, pois os procedimentos laboratoriais não podem ser feitos todos ao mesmo

---

<sup>1</sup> Embrapa Meio Ambiente. e-mail: alfredo@cnpma.embrapa.br

tempo. Também pode ocorrer a morte de peixes antes do prazo final estabelecido para o abate. Essas são algumas das principais justificativas para a não observação de alguns valores em determinados indivíduos participantes do experimento.

## 2 Material e Métodos

Foram utilizados dados de um experimento com *P. mesopotamicus*, conduzido em delineamento experimental inteiramente casualizado, com 10 repetições (peixes). Os tratamentos eram constituídos de combinações dos fatores níveis de CO<sub>2</sub> (meio normocárbico = 0,03% CO<sub>2</sub>  $\cong$  0,2 mmHg ou hipercárbico = 1% CO<sub>2</sub>  $\cong$  6,7 mmHg) e concentrações de cobre (0 ou 0,4 mg Cu<sup>2+</sup> L<sup>-1</sup>).

O padrão de desbalanceamento desse experimento é apresentado na Tabela 1. Algumas combinações de níveis dos fatores apresentaram unidades experimentais (peixes) faltantes (menos de 10 observações) para todas as variáveis-resposta.

Tabela 1. Número de repetições nas combinações de níveis dos fatores para as variáveis estudadas.

Variáveis-resposta	Níveis dos fatores			
	0 mg Cu <sup>2+</sup> L <sup>-1</sup>		0,4 mg Cu <sup>2+</sup> L <sup>-1</sup>	
	0,03 % CO <sub>2</sub>	1 % CO <sub>2</sub>	0,03 % CO <sub>2</sub>	1 % CO <sub>2</sub>
Na/K, RBC, Htc, Hb, CHCM, VCM	10	8	10	10
BpH	9	8	10	10
CuPlasma	7	8	10	3
FOXmv, CATmv	7	8	9	8
GSH-Pxfig, SODmb, GSH-Pxmb	7	7	9	7
SODmv	7	8	8	8
FOXmb	7	6	8	8
FOXfig, CATmb	6	8	9	8
GSH-Pxmv	6	7	8	6
SODfig	5	8	9	8
Glicose, Lactato, Piruvato, Amônia	5	8	7	5
Proteína	5	8	6	5
CATfig	4	7	8	8

Para ilustração dos métodos e discussão do problema de desbalanceamento, apresentaremos a análise detalhada do subconjunto de dados expostos na Tabela 2, referentes à variável resposta Glicose, em peixes da espécie *P. mesopotamicus*.

Tabela 2. Valores da variável Glicose para cada uma das combinações dos níveis dos fatores.

CO <sub>2</sub> \Cu <sup>2+</sup>	0 mg L <sup>-1</sup>	0,4 mg L <sup>-1</sup>
0,03 %	1,19; 1,27; 1,19; 1,27; 1,23	0,62; 1,30; 1,28; 0,87; 1,16; 1,47; 1,52
1 %	0,60; 0,79; 1,22; 0,61; 0,95; 0,68; 0,92; 0,91	1,50; 1,07; 1,05; 1,38; 1,68

No procedimento GLM, no ambiente SAS, o cálculo correto das somas de quadrados (SQ) é obtido com a opção descrita como Erro Tipo III do *MODEL Statement*.

O programa SAS utilizado para análise da variável Glicose obtida num experimento fatorial desbalanceado para avaliar o efeito combinado das concentrações de CO<sub>2</sub> e de cobre sobre parâmetros bioquímicos e defesas antioxidantes do peixe *Piaractus mesopotamicus*, aqui utilizada como exemplo, é apresentado na Figura 1. Os dados foram lidos de uma planilha Excel.

---

```

title 'Análise de Variância de Experimento Fatorial Desbalanceado CO2 x Cobre';
options nonumber nodate pagesize=1000 linesize=80;
Proc import data= um infile=" C:\..."
Dbms=Excel replace;
Getnames=YES;
Run;
* Impressao dos dados;
Proc print data=um;
Var CO2 Cobre Rep Y
Run;
* Análise de variancia;
proc glm data=um;
  class CO2 cobre;
  model y = CO2 cobre CO2*cobre/ss1 ss3;
  lsmeans CO2*cobre / slice = cobre ETYPE=3;
  lsmeans CO2*cobre / slice = CO2 ETYPE=3;
  means CO2*cobre;
  lsmeans CO2*cobre / pdiff ETYPE=3;
run;

```

---

Figura 1. Programa SAS para análise de um experimento fatorial desbalanceado 2x2.

A análise utilizando o pacote LM do R empregou o programa apresentado no Figura 2.

---

```

require(xlsReadWrite)
require(car)
require(reshape)
setwd("C:/...")
dados=read.xls("dadoshc.xls",colClasses=c(rep("factor",2),rep("numeric",26)))
dadosm=melt(dados)
cast(dadosm,CO2+cobre~variable,mean,na.rm=T,margins=c("grand_row"))
cast(dadosm,CO2+cobre~variable,sd,na.rm=T,margins=c("grand_row"))
by(dadosm,dadosm$variable,function(x)summary(glm(value~CO2*cobre,data=x,
  contrasts=list(CO2=contr.sum,cobre=contr.sum))))
by(dadosm,dadosm$variable,function(x)anova(glm(value~CO2*cobre,data=x,
  contrasts=list(CO2=contr.sum,cobre=contr.sum)),test="F"))
by(dadosm,dadosm$variable,function(x)Anova(glm(value~CO2*cobre,data=x,
  contrasts=list(CO2=contr.sum,cobre=contr.sum)),test="F",type="III"))

```

---

Figura 2. Programa R para análise de um experimento fatorial desbalanceado 2x2.

### 3 Resultados e Discussão

Na Tabela 3 são apresentados os resultados da análise de variância para a variável Glicose, mostrando as somas de quadrados dos Tipos I e III, obtidas via procedimento GLM do SAS/STAT.

Tabela 3 – Análise de variância para a variável Glicose no ambiente SAS.

Fonte de variação	GL	SQ	QM	F	Prob>F
Modelo	3	0,96129257	0,32043086	5,40	0,0065
Erro	21	1,24629143	0,05934721		
Total	24	2,20758400			
Erro Tipo I					
CO2	1	0,17992823	0,17992823	3,03	0,0963
Cobre	1	0,31729684	0,31729684	5,35	0,0310
CO2 x cobre	1	0,46406750	0,46406750	7,82	0,0108
Erro Tipo III					
CO2	1	0,08148782	0,08148782	1,37	0,2544
Cobre	1	0,29688889	0,29688889	5,00	0,0363
CO2 x cobre	1	0,46406750	0,46406750	7,82	0,0108

Observa-se que o valor F para o fator CO2 pode ser considerado como significativo a 10% ao utilizarmos a SQ do tipo I, mas é claramente não significativo quando é utilizado a SQ do tipo III. Conforme prevê a teoria, os resultados para a interação entre os fatores não se alteram quando se usam as SQ correspondentes aos erros tipo I ou tipo III. Na Tabela 4, são apresentados os resultados da análise de variância para a variável Glicose, mostrando as somas de quadrados correspondentes aos erros Tipos I e III, obtidas no ambiente R. Ao comparar os resultados obtidos nos ambientes SAS e R, é possível verificar que os valores das SQ são ligeiramente diferentes, mas os valores de F resultam em significâncias muito semelhantes para os fatores.

Tabela 4 – Análise da variância para a variável Glicose no ambiente R.

Fonte de variação	GL	SQ	QM	F	Prob>F
Modelo	3	0,9589	0,3196	5,3412	0,00682
Erro	21	1,2567	0,0598		
Total	24	2,2156			
Erro Tipo I					
CO2	1	0,17617	0,17617	2,9439	0,10092
Cobre	1	0,31824	0,31824	5,3180	0,03140
CO2 x cobre	1	0,46456	0,46456	7,7632	0,01107
Erro Tipo III					
CO2	1	0,07892	0,07892	1,3188	0,26373
Cobre	1	0,29779	0,29779	4,9763	0,03673
CO2 x cobre	1	0,46456	0,46456	7,7632	0,01107

Levando em consideração as análises das 25 variáveis, foi possível comparar o nível de significância do valor F de cada um dos fatores e da interação, calculados considerando os erro tipo I ou III, tanto pelo SAS como pelo R.

Tabela 5. Frequência observada em cada classe de nível de significância dos valores de F correspondentes aos efeitos dos fatores sobre as diferentes variáveis-resposta, calculados pelo R e pelo SAS, utilizando o erro tipo I e tipo III.

Fator	Nível de significância	SAS		R	
		Erro tipo I	Erro Tipo III	Erro tipo I	Erro Tipo III
CO2	1%	10	11	10	11
	5%	3	2	3	2
	10%	1	3	0	3
	ns	11	9	12	9
Cobre	1%	10	10	10	10
	5%	4	5	3	5
	10%	2	2	2	2
	ns	9	8	10	8
CO2xCobre	1%	5	5	5	5
	5%	3	3	3	3
	10%	2	2	2	2
	ns	15	15	15	15

Na Tabela 5, são apresentadas as frequências de variáveis-resposta para em cada nível de significância relativo ao teste F para efeito de fator, para os dois tipos de erro. Pode ser observado que, tanto no SAS quanto no R, a escolha do tipo de erro altera os resultados no que diz respeito ao julgamento da significância dos fatores isolados, mas não da interação. Quanto ao ambiente utilizado, como esperado, não houve diferença quando foi utilizado o erro tipo III, que é o adequado para estes casos.

#### 4 Conclusões

Foram observadas pequenas diferenças nos valores das somas de quadrados para os fatores do modelo entre o R e o SAS, entretanto, os valores calculados de F e, conseqüentemente, seu nível de significância, foram muito semelhantes nos dois ambientes quando se utilizou o erro tipo III. Isto resultou nas mesmas conclusões sobre a existência ou não de influência dos fatores sobre cada uma das variáveis analisadas.

É fundamental, no ambiente R, no módulo *glm* (*glm package*), a especificação da opção *contr.sum* no comando *contrasts*, o que define que a soma de quadrados a ser calculada para cada fator isoladamente terá como base um conjunto de desvios cuja soma é zero. Se essa opção não for especificada, apesar de ser escolhida a opção erro tipo III, os resultados obtidos não serão corretos.

#### 5 Bibliografia

- [1] R DEVELOPMENT CORE TEAM. **R: a language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria, 2010. ISBN 3-900051-07-0, URL <http://www.R-project.org>
- [2] SAMPAIO, F. G.; BOIJINK, C. L.; OBA, E. T.; SANTOS, L. R. B.; KALININ, A. L.; RANTIN, F. T. Antioxidant defenses and biochemical changes in pacu (*Piaractus mesopotamicus*) in response to single and combined copper and hypoxia exposure. **Comparative Biochemistry and Physiology**, v. 147, part C, p. 43-51, 2008.
- [3] SAS INSTITUTE INC. **SAS/STAT® User's Guide**, Version 9.1, v.1-7. SAS Institute Inc., Cary, NC, 2004.