

## APLICAÇÃO DO MÉTODO AID ESTENDIDO NA IDENTIFICAÇÃO DE LINHAGENS DE FRANGOS RESISTENTES A COCCIDIOSIS (*Eimeria tenella*).

Cátia Valéria Presotti - Mestrando em Estatística

Universidade Federal de São Carlos - cvpresotti@yahoo.com.br

Dra. Maria Cecília Mendes Barreto - Departamento de Estatística

Universidade Federal de São Carlos - cbarreto@power.ufscar.br

Giovani Rota Bertani, Carlos A.F. Costa - sac@cnpa.embrapa.br

Embrapa Suínos e Aves, Concórdia SC

Waldomiro Barioni Júnior, Alfredo Ribeiro de Freitas – sac@cppse.embrapa.br

Embrapa pecuária sudeste, São Carlos –SP

**RESUMO:** As técnicas de agrupamentos de médias, entre elas o método de Scott-knott (1974), podem não ser adequadas na aplicação em dados com distribuição diferente da normal. Neste trabalho, apresentaremos uma extensão do algoritmo “Automatic Interaction Detection” (AID), no qual se baseia o método de Scott-Knott, desenvolvido por Barreto (1993), Barreto e Demétrio (1995 e 1998). Essa extensão, usada em modelos lineares generalizados (McCullagh & Nelder, 1989) que adota como medida de homogeneidade de grupos uma estatística baseada na função desvio. Usando os resultados do conjunto de dados provenientes de um delineamento inteiramente casualizado com objetivo de avaliar a taxa de mortalidade das aves, entre as linhagens de frango (CC, CCc e TT) submetidas em dois níveis de infecção: infectados e não infectados, isto é, saber entre as linhagens qual delas é mais resistente à infecção. Através do método AID obtivemos 3 grupos sendo que o mais susceptível a *Eimeria tenella* esta representado pela linhagem TTinf.

**Palavras chave:** AID; Extensão do método AID; Métodos de Agrupamento; Modelos Lineares Generalizados; Scott-Knott.

### 1. INTRODUÇÃO

O método AID (Automatic Interaction Detection) proposto por Morgan e Sonquist como uma da técnica de análise de agrupamentos, consiste em dividir o conjunto de dados sucessivamente em grupos, combinando ou não as classes das variáveis preditoras ou níveis dos fatores, de modo a produzir grupos que se diferenciem ao máximo em relação à média da variável resposta. O processo de subdivisão é aplicado enquanto as divisões obtidas contribuem para explicar a variação dos dados e ao final temos um conjunto de grupos que podem ser identificados por alguns níveis dos fatores ou variáveis preditoras e suas interações mais importantes.

Neste trabalho apresentamos uma extensão do algoritmo AID em modelos lineares generalizados (Barreto,1993; Barreto e Demétrio,1995), baseada no procedimento de Scott-Knott (1974), que adota como medida de homogeneidade de grupos uma estatística baseada na função desvio. Esse procedimento foi aplicado num conjunto de dados provenientes de um delineamento inteiramente casualizado que foi instalado e conduzido na Embrapa Suínos e Aves – Concórdia/SC em 2004, com objetivo de avaliar a taxa de mortalidade das aves, no período de 7 a 25 dias, entre as linhagens de frango (CC, CCc e TT) submetidas em dois níveis de infecção: infectados e não infectados, isto é, saber entre as linhagens qual delas é mais resistente à infecção.

## 2. METODOLOGIA

Ao se procurar grupos homogêneos em um conjunto de dados, uma alternativa seria examinar todas as partições possíveis do conjunto de dados e escolher aquela que é ótima segundo algum critério. O algoritmo AID reduz o número de partições a serem examinadas em ordem crescente do valor esperado de cada tratamento.

Considerando o modelo de um ensaio completamente ao acaso de  $K$  caselas ou tratamentos com  $n_k$  repetições por casela, seja  $D\left(\underset{\sim}{y}, \underset{\sim}{\hat{\mu}}_{1+K}\right)$  a deviance ou desvio do modelo associado. O desvio do mesmo ensaio, considerando todas as médias iguais entre si, é dado por  $D\left(\underset{\sim}{y}, \underset{\sim}{\hat{\mu}}_1\right)$ .

Seja  $D\left(\underset{\sim}{y}, \underset{\sim}{\hat{\mu}}_d\right)$  o desvio do modelo de dois grupos definidos pela partição  $d$  no conjunto de  $K$  tratamentos.

Barreto (1993), Barreto e Demétrio (1995) propõem usar a diferença  $\Delta\left(\hat{\mu}_d\right) = D\left(\underset{\sim}{y}, \underset{\sim}{\hat{\mu}}_1\right) - D\left(\underset{\sim}{y}, \underset{\sim}{\hat{\mu}}_d\right)$ , como uma medida de homogeneidade de dois grupos, indexados pela partição  $d$  uma vez que  $D\left(\underset{\sim}{y}, \underset{\sim}{\hat{\mu}}_d\right)$  é uma medida de discrepância do modelo de dois grupos. Quanto menor o desvio  $D\left(\underset{\sim}{y}, \underset{\sim}{\hat{\mu}}_d\right)$ , mais heterogêneos serão os dois grupos. Desse modo, quanto maior  $\Delta\left(\hat{\mu}_d\right)$ , mais heterogêneos serão os grupos formados pela partição  $d$ .

Para a escolha da divisão mais promissora, a extensão do algoritmo AID (Barreto,1993; Barreto e Demétrio,1995) usa a estatística  $\Delta\left(\hat{\mu}_d\right)$ . Assim, no conjunto de todas as partições possíveis, escolhe-se aquela que produz a maior redução:

$$\Delta\left(\hat{\mu}_{d_0}\right) = \max_d \left\{ \Delta\left(\hat{\mu}_d\right); d \text{ entre as partições possíveis} \right\} \quad (1)$$

Mostra-se que o desvio do modelo de  $G$  grupos,  $D\left(\underset{\sim}{y}; \underset{\sim}{\mu}_{1+G}^{\wedge}\right)$  varia entre o desvio do modelo que corresponde à participação mais fina com  $K$  grupos,  $D\left(\underset{\sim}{y}; \underset{\sim}{\mu}_{1+K}^{\wedge}\right)$ , e o desvio do modelo que corresponde à partição mais rudimentar com um único grupo,  $D\left(\underset{\sim}{y}; \underset{\sim}{\mu}_1^{\wedge}\right)$ .

Uma generalização da estatística  $R^2$  para modelos com  $G$  grupos é dada por:

$$R_s^2(G) = \frac{D\left(\underset{\sim}{y}; \underset{\sim}{\mu}_1^{\wedge}\right) - D\left(\underset{\sim}{y}; \underset{\sim}{\mu}_{1+G}^{\wedge}\right)}{D\left(\underset{\sim}{y}; \underset{\sim}{\mu}_1^{\wedge}\right) - D\left(\underset{\sim}{y}; \underset{\sim}{\mu}_{1+K}^{\wedge}\right)}$$

que pode ser interpretada como a proporção padronizada do desvio explicada na forma de tais modelos.

Assim a extensão do algoritmo AID em modelos lineares generalizados é dada por:

- a) Critérios de seleção:
  - a<sub>1</sub>) Escolher como grupo a ser dividido aquele que produz a maior redução  $\Delta\left(\underset{\sim}{\mu}_d^{\wedge}\right)$ .
  - a<sub>2</sub>) Escolher como grupo a ser dividido aquele que produz no passo seguinte a maior redução  $\Delta\left(\underset{\sim}{\mu}_d^{\wedge}\right)$ .
- b) Critérios de partição ótima:
  - b<sub>1</sub>) Aquela que produz a maior redução de  $\Delta\left(\underset{\sim}{\mu}_d^{\wedge}\right)$ .
  - b<sub>2</sub>) Aquela que produz a redução  $\Delta\left(\underset{\sim}{\mu}_d^{\wedge}\right)$  com o menor nível descritivo.
- c) Regras de parada:
  - c<sub>1</sub>) O número de elementos nos grupos resultantes é menor que algum número pré-fixado.
  - c<sub>2</sub>) O número total de divisões não alcança um certo máximo.
  - c<sub>3</sub>) O nível descritivo do máximo da redução  $\Delta\left(\underset{\sim}{\mu}_d^{\wedge}\right)$  não é maior que um valor pré-fixado.
  - c<sub>4</sub>) O valor de  $R_s^2(G)$  se alterou menos do que um valor pré-fixado em relação à divisão imediatamente anterior.

Tanto na adoção do critério de partição ótima (b<sub>2</sub>) como na adoção da regra de parada (c<sub>3</sub>), é importante o conhecimento da distribuição da estatística  $\Delta\left(\underset{\sim}{\mu}_{d_0}^{\wedge}\right)$ , que é o máximo entre as reduções  $\Delta\left(\underset{\sim}{\mu}_d^{\wedge}\right)$  possíveis.

Para dados com distribuição binomial, Barreto (1993) e Barreto e Demétrio (1995) verificaram, por simulação, que o máximo da redução  $\Delta\left(\underset{\sim}{\mu}_d^{\wedge}\right)$  também apresenta distribuição

aproximadamente proporcional a uma distribuição  $\chi^2$  com  $K/(\pi - 2)$  graus de liberdade e parâmetro de proporcionalidade igual a  $\pi/2(\pi - 2)$ . Esses resultados foram verificados para  $K$  variando de 3 a 10, sendo que a qualidade da aproximação é maior conforme  $K$  aumenta.

### 3. RESULTADOS E DISCUSSÃO

O delineamento experimental foi instalado em um aviário contendo gaiolas metálicas na Estação Experimental de Suruvi, da Embrapa Suínos e Aves, Concórdia, SC no período de novembro a dezembro de 2004. Tendo como objetivo avaliar a resistência genética a coccidiose foi utilizado como modelo a infecção por *Eimeria tenella*. Três linhagens de aves foram estudadas: CC foi selecionada para produção de ovos; CCc é o controle genético utilizando cruzamentos ao acaso e TT foi selecionada para produção de carne. Para cada uma das linhagens as aves foram infectadas com *Eimeria tenella* e um grupo não infectado resultando em 6 tratamentos (CC infectada, CC não infectada; CCc infectada; CCc não infectada; TT infectada e TT não infectada). Os tratamentos tinham entre 6 e 10 repetição, cada repetição era formada por uma gaiola alojando de 10 a 11 aves. A dose infectante foi de 30.000 oocistos de *Eimeria tenella*. A variável resposta do experimento foi o numero de aves mortas no período de 7 a 25 dias. Os dados estão na Tabela\_01.

**Tabela\_01** – Número de aves mortas por gaiola, nos receptivos tratamentos.

Tratamentos	Repetições (gaiolas) <sup>1</sup>									
	I	II	III	IV	V	VI	VII	VIII	IX	X
1. CC <sup>3</sup> inf <sup>2</sup>	2	0	1	4	1	1	1	2	4	3
2. CC não	0	0	0	0	0	0	0	0	.	.
3. CCc inf	1	1	5	4	4	3	1	1	3	.
4. CCc não	0	0	0	0	0	0	1	.	.	.
5. TT inf	2	4	2	6	5	3	5	4	1	3
6. TT não	1	0	1	0	0	0	.	.	.	.

<sup>1</sup> Cada gaiola alojava de 10 a 11 aves; <sup>2</sup> não=não infectado e inf=infectado; <sup>3</sup> Linhas (CC,CCc e TT).

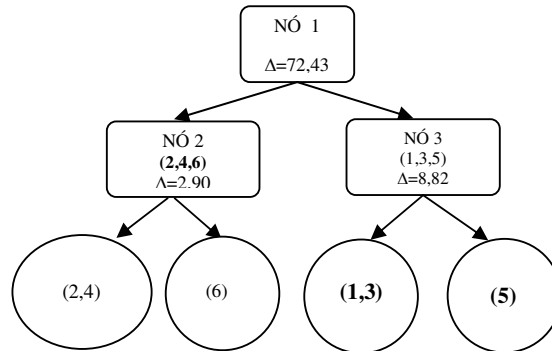
A análise do desvio é de 82,44 correspondentes ao modelo que supõe distribuição binomial ( $r/n$ ) para a variável resposta, taxa de mortalidade de frangos, e a função de ligação logística para o valor esperado.

**Tabela\_02** – Resumo dos resultados do algoritmo de divisão dicotômica.

NÓ	$\Delta_0$	$\frac{2(\pi - 2)\Delta_0}{\pi}$	$\frac{k}{(\pi - 2)}$	Grupos
1	72,43	53,64	5,25	{2,4,6} e {1,3,5}
2	2,90	2,11	6,03	{2,4} e {6}
3	8,82	6,41	6,03	{1,3} e {5}

Na Tabela\_02 estão os principais elementos para a construção de árvores de divisão dicotômica. Considerando como critério de seleção a maior redução  $\Delta(\hat{\mu}_d)$ , como critério de partição ótima a que produz a maior redução e como regra de parada a obtenção de 3 grupos

finais. O algoritmo começa no nó (1) da Tabela\_02. A seguir, examinando os nós (2) e (3), este último é escolhido. Até este passo têm-se 3 grupos. O algoritmo prossegue examinando o nó (2) e para quanto obtém 3 grupos. Os grupos formados são: {2,4,6}, {1,3} e {5}.



**Figura\_01** - Árvore de divisão dicotômica considerando como partição ótima a maior redução e como regra de parada a obtenção de 3 grupos.

Com relação à estatística  $R^2_*$ , aplicação  $(72,43+8,82)/82,39= 98,61\%$  de explicação.

#### 4. CONCLUSÕES

A extensão do algoritmo AID para modelos lineares generalizados aqui apresentadas é indicada na obtenção de grupos homogêneos quando a variável resposta pertence à família exponencial. A medida de homogeneidade de grupos é baseada na função desvio. Foi observado diferente nível de susceptibilidade à infecção por *Eimeria tenella* e não houve uma linhagem totalmente resistente sendo a maior diferença de susceptibilidade observada é a linhagem TT, isto é, que apresenta um maior número de aves infectadas/mortas. A extensão da técnica AID aplicada produziu 3 grupos homogêneos com 98,39% de explicação.

#### REFERÊNCIAS BIBLIOGRÁFICAS

- Barreto, M. C. M; Demétrio, C, G,B.(1995). Uma extensão da técnica AID em modelos lineares generalizados. *Revista Matemática Estatística*, **30**,141-154.
- Barreto, M. C. M.(1993). Uma extensão da técnica AID em modelos lineares generalizados. Tese (Doutoramento) - Escola Superior de Agronomia Luiz de Queiroz, Universidade de São Paulo, 200p.
- McCullagh,P.;Nelder,S.A (1989). *Generalized linear models*. 2 ed. Londres: Chapman and Hall.