

# ANÁLISE DE DADOS DE GASTROENTERITE HEMORRÁGICA CANINA PARA IDENTIFICAR FATORES DE RISCO POR REGRESSÃO LOGÍSTICA<sup>1</sup>

Data analysis of hemorrhagic gastroenteritis to identify risk factors by logistic regression

Paula Roberta Mendes<sup>2</sup>, Joel Augusto Muniz<sup>3</sup>, Hélio Rubens de Carvalho Nunes<sup>4</sup>,  
Lician Vaz de Arruda Silveira Chalita<sup>5</sup>, Luiz Henrique de Aquino<sup>6</sup>

## RESUMO

No presente estudo, ajustou-se um modelo de regressão logística para prever a probabilidade de óbito de cães acometidos por gastroenterite hemorrágica. O modelo Logístico é recomendado para variáveis-resposta dicotômicas em estudo de Coorte. Registraram-se 176 animais censitariamente atendidos com gastroenterite hemorrágica em quatro clínicas veterinárias da cidade de Lavras, sul de Minas Gerais, entre os anos de 1992 e 1999. Após terem sido selecionadas por meio do teste de  $\chi^2$  de Pearson ou teste exato de Fisher, ajustou-se o modelo considerando-se as variáveis sexo, idade, diárias de internação e número de atendimentos. A estimação dos parâmetros foi feita pelo método da máxima verossimilhança. Conclui-se que quando os cães acometidos por gastroenterite hemorrágica são atendidos apenas uma vez, aqueles com idade superior a 6 meses possuem 15,45 vezes mais chances de morrerem ( $P < 0,05$ ) do que aqueles com menos de 6 meses de idade. Quando os animais que apresentam a enfermidade possuem mais de 6 meses de idade, a chance de morrerem, se forem atendidos apenas uma vez, é 20,251 vezes maior ( $P < 0,05$ ) do que se recebessem de 2 a 7 atendimentos.

**Termos para indexação:** Gastroenterite hemorrágica, epidemiologia, modelo logístico, razão de chances, máxima verossimilhança.

## ABSTRACT

This paper presents a study of how to fit a logistic regression model to predict the death probability of dogs with hemorrhagic gastroenteritis. A logistic model is recommended to treat dichotomic variables in Coorte study. Using a census procedure from 1992 to 1999 four veterinary clinic in Lavras, MG, registered 176 infected animals. The variables of the model have been chosen to be sex, age internment days rates and number of clinical treatments by the  $\chi^2$  or Fisher's exact test. The parameters were estimated by the maximum likelihood method. The results showed that if the infected dogs were clinically treated only once then the animals older than six months had their mortality chances 15.45 times ( $P < 0.05$ ) larger than those younger than six months. If the infected animals younger than six months were clinically treated only once then their mortality chances were 20.251 ( $P < 0.05$ ) higher than if they had received two to seven medical treatments.

**Index terms:** Hemorrhagic gastroenteritis, epidemiology, logistic model, maximum likelihood, odds ratio.

(Recebido para publicação em 19 de fevereiro de 2002 e aprovado em 7 de novembro de 2002)

## INTRODUÇÃO

A tríade que guia a Epidemiologia é composta pelo agente, o hospedeiro suscetível e o ambiente. Por outro lado, a Estatística detém sólidos conhecimentos sobre curvas descritivas e interações de fatores, procurando medir o efeito do acaso nos eventos.

Assim, a Estatística pode contribuir para a identificação de fatores de risco para a vida dos animais acometidos por certas enfermidades, enquanto a Epidemiologia, de posse da comprovação desses fatores, pode propor estratégias para aumentar a sobrevivência dos indivíduos acometidos ou da população.

1. Parte do trabalho desenvolvido pela primeira autora como bolsista do PIPIC/CNPq no Departamento de Ciências Exatas/DEX da Universidade Federal de Lavras/UFLA, Caixa Postal 37 – 37200.000 – Lavras, MG.  
2. Médica Veterinária graduada pela UFLA. prmdes@hotmail.com  
3. Professor Titular do DEX/UFLA. Área de Técnica Experimental com Animais, bolsista do CNPq, joamuniz@ufla.br  
4. Acadêmico do Curso de Bacharelado em Estatística pela Universidade Federal de São Carlos, São Carlos, SP. heliorubens@hotmail.com  
5. Professora Assistente, Doutora do Departamento de Bioestatística da Universidade Estadual Paulista - Campus de Botucatu, SP. liciana@ibb.unesp.br  
6. Professor Titular do DEX/UFLA. Área de Técnica Experimental com Animais, lhaquino@ufla.br

Forças importantes da sociedade moderna têm acelerado o reconhecimento dos métodos e perspectivas da Epidemiologia Clínica. Os custos dos serviços, tanto na medicina humana quanto na medicina veterinária, têm aumentado tanto que nem mesmo as sociedades mais afluentes são capazes de pagar todos os serviços de que os pacientes necessitam. Estudos mostram uma grande variação nas práticas clínicas, sem uma correspondente variação nos desfechos delas resultantes, sugerindo que nem todas as práticas comuns e dispendiosas são importantes. Métodos mais rigorosos para avaliar as evidências clínicas estão sendo desenvolvidos e são valorizados por aqueles que tomam decisões clínicas. Essas observações levam ao consenso de que a atenção clínica deveria se embasar nas pesquisas mais sólidas possíveis e ser julgada pelos resultados por ela alcançados a um custo que a sociedade consegue pagar. Além disso, os pacientes individuais devem ser analisados em relação ao grupo maior do qual fazem parte, com o intuito de se fazer previsões sobre eles e auxiliar a decidir quais recursos médico-veterinários trazem melhores benefícios para um número maior de animais (FLETCHER et al., 1996).

Num primeiro passo, a epidemiologia busca reconhecer a existência do problema, definir sua natureza e extensão, bem como determinar-lhe a causalidade. Em outras palavras, procura o diagnóstico da situação existente, para, então, numa segunda etapa, estabelecer o tratamento e/ou as medidas de prevenção. Para isso, seus estudos se dividem em duas etapas: fase descritiva e fase analítica. Na primeira, ocorre a coleta e descrição dos dados ou informações relativas à natureza dos eventos epidemiológicos e/ou características de ocorrência de doenças em populações; na segunda, realizam-se o processamento e a análise das informações obtidas na fase anterior, contemplando também a formulação e o teste de hipóteses, bem como as conclusões (GONSALES et al., 1998).

Um estudo de Coorte é uma forma de pesquisa que visa a verificar se indivíduos, selecionados devido à exposição ao fator de risco, desenvolveram a doença em maior ou menor proporção do que um grupo de indivíduos comparáveis, mas não expostos ao fator de risco. Em muitos estudos de Coorte, os grupos de comparação são obtidos de acordo com o nível de exposição ao fator. Por exemplo, as Coortes do Estudo de Framingham, citadas por Soares e Colosimo (1995) sobre doenças coronárias, foram construídas dividindo-se o grupo acompanhado de acordo com hábitos de fumo, níveis de co-

lesterol. Nesses casos, não há necessidade de um grupo externo de comparação.

Na prática clínica, é freqüente a necessidade de se estudar a relação entre um fator de risco e uma doença na presença de um fator de confundimento. A metodologia de Mantel-Haenszel (1959) permite resolver esse problema se os fatores envolverem variáveis aleatórias discretas. Porém, muitas doenças não são causadas por um único fator, mas por uma combinação de diferentes circunstâncias, chamadas de fatores de exposição ou fatores de risco. No caso de o número de fatores de risco ou confundimento presentes no estudo serem maiores do que um, a interação entre esses fatores deve ser pesquisada (COSTA, 1997).

Soares e Siqueira (1999) enfatizam que comparar dois grupos por meio do resultado observado em uma variável dicotômica é um problema comum na pesquisa médica, aparecendo com freqüência em diversos estudos clínicos. Citam também que o estatístico Pearson, no início do século, propôs medir a discrepância entre os valores observados e esperados das quatro entradas de uma tabela 2 x 2. O teste proposto recebeu o nome de Qui-quadrado devido à distribuição assintótica da estatística do teste.

Há uma limitação teórica na aplicação do teste de Qui-quadrado quando o valor esperado em alguma casela da tabela 2 x 2 é menor que 5, inerente às propriedades estatísticas de distribuições assintóticas ou de grandes amostras. Nesse caso, o teste de  $\chi^2$  de Pearson não deve ser utilizado, pois não garante corretamente o nível de significância de acordo com a distribuição teórica de  $\chi^2$ . Uma alternativa é o teste exato de Fisher que, segundo Soares e Siqueira (1999), se baseia na distribuição exata das contagens das quatro células da tabela em termos de somente uma das células.

A escolha do modelo a ser utilizado é um tópico extremamente importante na análise paramétrica de dados de tempo de vida. A estimação pelo método de máxima verossimilhança só pode ser feita após ter sido definido um modelo probabilístico adequado para os dados (COLOSIMO, 2000). Costa (1997) observou a adequação do modelo Logístico a um conjunto de dados sobre gastroenterite hemorrágica em cães atendidos no Hospital Veterinário da Universidade Estadual de Londrina. O autor ressalta que utilizou tal modelo porque ele permite calcular a razão de chances de dois níveis de uma variável, no caso, morte ou sobrevivência do animal, com a estimação dos parâmetros pelo método

da máxima verossimilhança, usando o software SAS (Statistical Analysis System) e recomendando toda a metodologia utilizada em seu trabalho para ser aplicada a qualquer tipo de estudo de Coorte em animais, desde que a variável-resposta seja dicotômica.

A Gastroenterite Hemorrágica é uma doença que acomete muitos cães jovens, com morbidade e mortalidade elevadas. Os principais sintomas dessa doença são diarreia hemorrágica, vômitos, febre, desidratação, apatia e anorexia. Pode ter etiologia viral, bacteriana ou parasitária. Além disso, sabe-se que fatores importantes considerados hipóteses de associação, descritos na literatura, incluem: idade, raça, porte (peso), estresse ambiental e condições climáticas (COSTA, 1997).

### O Modelo de Regressão Logística

Em um curso de Estatística Médica, a regressão Logística deve ser vista como uma generalização da técnica de Mantel e Haenszel (1959). Deve-se identificar a variável de interesse, que, nesse caso, é dicotômica: o paciente tem ou não tem a doença, o tratamento produziu ou não o efeito considerado. Os fatores de risco e de confundimento, bem como os grupos nos quais os pacientes estão divididos são as variáveis explicativas ou covariáveis (SOARES e COLOSIMO, 1995).

O modelo Logístico é bastante utilizado em Biometria para modelar a probabilidade do desenvolvimento de uma doença, em um período de tempo definido. Sua função assume valores entre 0 e 1 e apresenta uma forma de S, inferindo-se que o risco aumenta rapidamente em certo intervalo de valores intermediários da variável independente e, então, permanece extremamente alto, em torno de 1, para valores altos dessa variável. Em outras palavras, esse modelo é de grande aplicabilidade por considerar a natureza multivariada de uma questão na pesquisa epidemiológica. Assim, esse modelo é muito utilizado para descrever a probabilidade de ocorrência de um evento (COSTA, 1997).

Na presença de  $k$  variáveis independentes categóricas ou contínuas, o modelo Logístico estabelece uma relação entre essas  $k$  variáveis e a probabilidade de ocorrência de sucesso de uma variável dependente dicotômica. Torna-se necessário estabelecer uma relação linear entre as variáveis independentes e uma transformação da variável resposta. Segundo Demétrio (1993), as transformações mais conhecidas são as transformações logística, probito e complemento log-log, que são denominadas de funções de ligação. Assim, o modelo logístico linear amostral é dado por:

$$\text{logit}(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki},$$

$i = 1, \dots, n,$

sendo  $\text{logit}(p_i)$  denominado logaritmo da chance de sucesso com

$$p_i = \frac{\exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki})}{1 + \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki})},$$

em que  $p_i$  denota a probabilidade de sucesso da  $i$ -ésima observação;  $\mathbf{x}'_i = (x_{1i}, x_{2i}, \dots, x_{ki})$  indica o vetor  $k$  dimensional de variáveis relativas à  $i$ -ésima observação;  $\beta' = (\beta_1, \beta_2, \dots, \beta_k)$  refere-se ao vetor de parâmetros a serem estimados e  $\beta_0$  representa o intercepto do modelo. Desse modo, a equação de  $p_i$  pode ser expressa da seguinte forma:

$$p_i = \frac{\exp(\beta_0 + \mathbf{x}'_i \beta)}{1 + \exp(\beta_0 + \mathbf{x}'_i \beta)}.$$

Entre as principais vantagens do uso da transformação logística na probabilidade de sucesso  $p$ , Collet (1991) destaca:

1. A conveniência de seu uso do ponto de vista computacional;
2. A sua interpretação direta em termos do logaritmo da chance de sucesso, sendo útil na análise de dados de estudos epidemiológicos;
3. O fato de modelos com transformação logística, por suas propriedades, serem mais adequados para a análise de dados que foram coletados retrospectivamente, tal como em estudos de caso-controle.

### Cálculo das Chances e das Razão de Chances

Considere uma variável resposta  $Y$

$$P(Y=1/X) = \frac{\exp(\beta_0 + \beta x)}{1 + \exp(\beta_0 + \beta x)},$$

com resposta dicotômica associada à ocorrência ou não de um evento. Considere também uma variável independente  $X$  do tipo binária que define a ocorrência de um fator de risco de uma doença, assumindo o valor 1, se o fator está presente, e o valor zero, se o fator está ausente. Nesse contexto, o logaritmo das probabilidades calculadas sob o modelo logístico, chamado de logaritmo das chances (MCCULLOCH, 2001), é dado pela função

$$g(X) = \log\left(\frac{P(\text{Ocorrência do Evento}/X)}{P(\text{Não Ocorrência do Evento}/X)}\right) = \log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x$$

A razão de chances  $\Psi$  é obtida pelo quociente quando se considera a presença do fator de risco ( $X=1$ ) e sem o fator de risco ( $X=0$ ). O logaritmo da razão de chances é dado por

$$\log(\Psi) = g(X=1) - g(X=0) = \beta_1.$$

Uma vez estimado o parâmetro  $\beta_1$ , a estimativa da razão de chances é obtida fazendo-se  $e^{\beta_1}$ . Um intervalo de confiança para o parâmetro  $\beta_1$  da razão de chances, de acordo com Venzon e Moolgavkar (1988), pode ser construído com base na propriedade de normalidade assintótica do estimador de máxima verossimilhança, utilizando-se a expressão

$$\hat{\beta}_i \pm Z_{\alpha/2} \sqrt{\text{Vâr}(\hat{\beta}_i)}.$$

O ajuste do modelo pode ser feito tanto pelo método dos mínimos quadrados quanto pelo método da máxima verossimilhança (HOSMER e LEMESHOW, 1989). Sempre que possível, a preferência deve ser dada ao método da máxima verossimilhança, pelo fato de esse apresentar para grandes amostras propriedades ótimas, tais como a distribuição normal dos estimadores, a qual permite que esses sejam testados quanto à significância por métodos usuais. A verossimilhança é a probabilidade de se observar os dados obtidos, fixados os valores das variáveis explicativas. Assumindo que as observações vêm de pacientes diferentes e são, portanto, independentes, a verossimilhança é simplesmente o produto da probabilidade de ocorrência de cada resposta observada (SOARES e COLOSIMO, 1995).

Para avaliar a significância de uma variável independente, compara-se o valor do teste da razão de máxima verossimilhança, com e sem a variável em questão. A estatística do teste, de acordo com Mood et al. (1974), é dada por

$$D = -2 \log \Lambda,$$

sendo  $\Lambda$  obtido fazendo-se o quociente da função de verossimilhança  $L$ , calculada considerando-se o modelo reduzido ( $H_0$ ) e o modelo completo ( $H_0 \cup H_1$ ), ou seja

$$\Lambda = \frac{\text{Sup}_{H_0} L}{\text{Sup}_{H_0 \cup H_1} L}.$$

Wilks (1935, 1938), citado por Agresti (2002), mostrou que  $-2\log\Lambda$  possui distribuição assintótica de  $\chi_p^2$ , com  $p$  graus de liberdade. Segundo Collet (1991), a diferença entre as estatísticas do teste de máxima verossimilhança é comparada com o valor de  $\chi^2$  de tabela, com grau de liberdade igual à diferença entre o número de parâmetros no modelo sob  $H_0$  e sob  $H_0 \cup H_1$ . O teste é comumente utilizado para verificar se a inclusão do parâmetro no modelo é significativo ao nível  $\alpha$  de probabilidade.

Ao se realizar a análise estatística de um conjunto de dados, após a estimação dos parâmetros, deve-se avaliar o ajuste do modelo, fazendo-se o diagnóstico da existência de algum ponto discrepante no conjunto de dados. Demétrio (1993, 2001) sugere, entre outras técnicas, a utilização de gráficos de resíduos padronizados de Pearson para se estudar a distribuição dos resíduos, o que permite identificar os pontos discrepantes. Outra alternativa para avaliar a qualidade do ajuste é identificar os pontos influentes, que são obtidos comparando-se os valores amostrais com os valores da matriz de projeção  $H$  de dimensão  $n \times n$ , sendo  $n$  o número de animais. Segundo Costa (1997), o valor-referência da matriz  $H$  é dado pela expressão  $h_{ii} = 2p/n$ , em que  $n$  representa o total de animais estudados. Quando uma observação apresentar valor maior que  $h_{ii}$ , ela será considerada ponto influente. Identificada uma observação influente ou discrepante, a análise é feita com a inclusão ou não dela, avaliando-se as alterações nos valores do teste de máxima verossimilhança, pois, segundo Cordeiro (1986), o valor da estatística do teste da razão de máxima verossimilhança ( $D_p$ ) reduz à medida que melhora o ajuste do modelo e, além disto,  $D_p$  tem distribuição de  $\chi_p^2$  com  $p$  graus de liberdade.

Charnet et al. (1999) destacam que uma ou mais observações discrepantes, quando incorporadas no estudo de um modelo de regressão, podem prejudicar o ajuste, dando uma falsa idéia da qualidade do modelo. Os autores destacam, no entanto, que uma observação discrepante deve ser analisada com muito critério, uma vez que a eliminação dela além de reduzir o conjunto de dados, poderá eliminar informações associadas ao próprio fenômeno em estudo.

Nem sempre a diferença observada entre as respostas em dois grupos, por meio da aplicação do teste  $t$  ou teste Qui-quadrado, pode ser atribuída ao fator usado

para dividir os indivíduos nos dois grupos de comparação. Eventuais diferenças podem ser atribuídas a um outro fator que recebe, nesse caso, a denominação de fator de confusão. Como qualquer realidade clínica é complexa, a possibilidade de existência de fatores de confusão deve ser sempre considerada. De fato, qualquer resposta clínica, como sobrevivência, diminuição da intensidade de uma doença, desaparecimento dos sintomas ou então reaparecimento da doença é influenciada por vários fatores (SOARES e SIQUEIRA, 1999).

Nas clínicas veterinárias, a mortalidade poderia ser menor se os médicos veterinários pudessem identificar fatores de risco que, por acaso ou desconhecimento, passam despercebidos. Em razão do grande número de cães atendidos com Gastroenterite Hemorrágica, decidiu-se associar as variáveis mais comumente observadas em óbitos de animais acometidos por essa enfermidade, a fim de poder estudar seus efeitos. Assim, com o presente estudo tem-se por objetivo construir um modelo para prever a probabilidade de óbito de cães acometidos por gastroenterite hemorrágica, bem como apresentar uma revisão bibliográfica resumida, direcionada aos praticantes de Medicina Veterinária, a respeito do ajuste de modelos de regressão linear Logística.

## MATERIAL E MÉTODOS

Os dados utilizados no presente estudo foram coletados em quatro clínicas veterinárias da cidade de Lavras, Minas Gerais. As fichas de atendimento foram avaliadas previamente e, em face da alta incidência de casos de gastroenterite hemorrágica, decidiu-se aprofundar os estudos sobre essa enfermidade por meio de um estudo retrospectivo de Coorte. Registraram-se 176 animais, com as correspondentes variáveis independentes raça, sexo, idade (meses), pelagem, diárias de internação, data do atendimento e estação do ano correspondente, clínica, método de diagnóstico, número de atendimentos e se o animal veio a óbito ou não.

Fez-se, então, uma triagem das variáveis que compunham o banco de dados, segundo os critérios adotados pela Epidemiologia, utilizando-se a significância de  $\chi^2$  de Pearson ou teste exato de Fisher ao nível  $\alpha = 0,20$ . Assim, após uma análise preliminar das variáveis e com base nos resultados dos testes, decidiu-se que as variáveis independentes a serem ajustadas no modelo seriam: sexo, idade (caracterizando, pela codificação do banco de dados, 1 = cães com menos de 6 meses de idade e 2 = animais acima de 6 meses), diárias de internação (compondo duas categorias: 1 = animais com um ou nenhum dia de internação e 4 =

animais internados por 2 a 7 dias) e número de atendimentos (também dividida em duas categorias: 1 = aqueles atendidos apenas uma vez e 3 = aqueles atendidos de 2 a 7 vezes). No texto, essas variáveis serão denominadas SEXO, IDADE, DIÁRIA e ATENDIMENTO.

Para a análise dos dados, utilizou-se o procedimento logístico (PROC LOGISTIC) do software SAS, conforme sugerido por Allison (1999). Utilizou-se também o procedimento SAS/STAT (SAS INSTITUTE, 1989), de acordo com rotinas apresentadas por Stokes et al. (1995). Foram ajustados quatro modelos, contendo cada um deles um intercepto e uma das quatro variáveis independentes: sexo, idade, diária e atendimento. A seleção do melhor modelo baseou-se na diferença das estatísticas do teste da razão de máxima verossimilhança, que continha apenas as variáveis idade e número de atendimentos. Foi incluído no estudo o efeito da interação entre essas duas variáveis principais no ajuste do modelo. Uma vez que a interação mostrou-se significativa, procurou-se entender o seu comportamento. Foi feito um diagnóstico dos pontos influentes e discrepantes.

## RESULTADOS E DISCUSSÃO

Análises descritivas preliminares das variáveis utilizando-se o Procedimento Univariate (PROC UNIVARIATE) do SAS revelaram que a idade mais freqüente entre os cães atendidos com Gastroenterite Hemorrágica é aquela inferior a 6 meses. Foi observado também que entre os cães atendidos, o mais comum foi receberem apenas um atendimento, permanecendo internado por no máximo um dia.

Com base nos valores do teste da razão de máxima verossimilhança, observou-se que a idade do cão é, significativamente, o fator que melhor explicou o óbito dos cães. Assim, ajustaram-se os modelos considerando as combinações da variável idade com as variáveis atendimento, diária e sexo. Um resumo dos resultados é apresentado na Tabela 1.

**TABELA 1** – Estatísticas *D* relacionadas ao ajuste do modelo Logístico considerando efeitos específicos.

Efeitos ajustados no modelo	D	GL
Idade + Atendimento	189,254	2
Idade + Diária	190,254	2
Idade + Sexo	191,752	2
Idade + Atendimento + Diária	187,130	3
Idade + Atendimento + Sexo	187,725	3

O teste da razão de verossimilhança apresentou uma redução expressiva quando se incluiu o número de atendimentos ao modelo contendo apenas o efeito da idade, ou seja, a inclusão do número de atendimentos no modelo que ajusta somente a idade do animal contribuiu significativamente na explicação do óbito de cães acometidos por Gastroenterite Hemorrágica. Observou-se que a inclusão das variáveis sexo e diária no modelo não acarretou contribuição significativa na explicação do óbito de cães acometidos por gastroenterite hemorrágica. O passo seguinte foi novamente o ajuste de todos os possíveis modelos visando à inclusão de uma nova variável. Um resumo dos resultados é apresentado na Tabela 2.

Após a verificação dos efeitos principais que poderiam ser incluídos no modelo, o próximo passo foi verificar a necessidade da inclusão de um efeito de interação entre a idade do cão e o número de atendimentos. Ajustou-se, então, o modelo dado por:

$$P(\text{Óbito} = 1) = \frac{\exp(\beta_0 + \beta_1 \text{idade} + \beta_2 \text{atend} + \beta_3 \text{idade.atend})}{1 + \exp(\beta_0 + \beta_1 \text{idade} + \beta_2 \text{atend} + \beta_3 \text{idade.atend})}$$

Por meio do teste da razão de verossimilhança, observou-se que o modelo foi significativo ao nível de 0,0021. A interpretação das estatísticas apresentadas na Tabela 2 é semelhante ao teste *F* usado em estudos de regressão linear múltipla, quando se testa a

significância do ajuste de um modelo de regressão (MONTGOMERY e PECK, 1991), ou seja, dado que a hipótese nula foi rejeitada, conclui-se que há ao menos um coeficiente regressor diferente de zero no modelo. Na estimação dos parâmetros, por meio do método da máxima verossimilhança, foram necessárias seis iterações para a convergência do modelo.

Com base na significância do modelo ajustado, obtiveram-se estimativas dos parâmetros, com o intuito de identificar quais os efeitos significativos, dado que os demais se encontram presentes no modelo. Na Tabela 3 são apresentadas as estimativas dos parâmetros, seus erros-padrão, e os valores do teste de Qui-quadrado, com seus respectivos níveis de significância, além das razões de chances.

**TABELA 2** – Resultado do teste de significância do modelo Logístico ajustado a dados de Gastroenterite Hemorrágica em cães.

Teste	Apenas Intercepto	Intercepto e Covariáveis	Qui-Quadrado para Covariáveis
D	197,167	182,452	14,715 com 3 G.L. (P = 0,0021)

**TABELA 3** – Parâmetros estimados, erros-padrão das estimativas, qui-quadrado, significância de qui-quadrado e razão de chances para o modelo Logístico.

Efeito no Modelo	GL	Estimativa	Erro-Padrão	$\chi_c^2$	$P(\chi^2 > \chi_c^2)$	Razão de Chances
Intercepto	1	2,8717	1,7444			.
Idade	1	-3,8358	2,8954	5,7372	0,0166	0,022
Atendimento	1	-1,2344	2,1238	3,1082	0,0779	0,291
Idade*Atendimento	1	1,3692	2,3412	5,0441	0,0247	3,932

Observa-se que o efeito da interação entre idade e atendimento é significativo ao nível  $\alpha = 0,05$ , sugerindo não estudar os efeitos principais, mas sim analisar como a idade do cão interage com o número de atendimentos. Assim, procedeu-se a um desdobramento da interação, ajustando-se um modelo com efeito de idade dentro de cada nível de atendimento e um modelo com a variável atendimento dentro de ca

da nível de idade. Os resultados estão apresentados na Tabela 4.

Observa-se que quando os animais acometidos por Gastroenterite Hemorrágica são atendidos apenas uma vez, aqueles com idade superior a seis meses possuem  $(1,313/0,085 = 15,45)$  vezes mais chances de viverem a óbito do que os com menos de 6 meses de idade. Quando os cães que apresentam a enfermidade possuem

mais de 6 meses de idade, sua chance de virem a óbito, se forem atendidos apenas uma vez, é 20,251 vezes a chance do que se recebessem de 2 a 7 atendimentos.

Pela análise dos resíduos padronizados de Pearson, revelou-se apenas um ponto discrepante, correspondendo à observação 104. As análises foram feitas sem essa observação. Pela matriz de projeção H, pontos influentes foram aqueles com valor  $h_{ii}$  superior a  $(2 \times 4)/165 = 0,04848$ , referindo-se às seguintes observações: 30, 34, 59, 83, 89, 90, 109, 112, 120, 138, 144 e 163. Um exame do banco de dados revelou que tais pontos diferiram entre si em pelo menos uma variável independente ou quanto à variável dependente, ou seja, o óbito. Além disso, vale ressaltar que esses pontos podem ter-se manifestado influentes devido a erros de co-

leta, de relatos dos proprietários dos animais ou de anotação na ficha clínica do animal, ou a variações quanto à condição do animal ao seu primeiro atendimento ou no tratamento, bem como a outros fatores desconhecidos ou não-controláveis.

Com isso, de acordo com as estimativas obtidas, pode-se dizer que quando o animal com Gastroenterite Hemorrágica é atendido apenas uma vez, o modelo ajustado assume a seguinte expressão:

$$P(\text{Óbito} = 1) = \frac{\exp(1,6373 - 2,4666 \cdot \text{idade})}{1 + \exp(1,6373 - 2,4666 \cdot \text{idade})}$$

**TABELA 4** – Parâmetros estimados, erros-padrão das estimativas, qui-quadrado, significância de qui-quadrado e razão de chances para o cada modelo logístico, de acordo com o nível fixado da variável controle.

Variável	Nível	Modelo	GL	Estimativa dos Parâmetros	Erro-padrão	$\chi^2_c$	$P(\chi^2 > \chi^2_c)$	Razão de Chances	Intervalo de Confiança da Razão de Chances (95%)
Atend	Uma vez	Intercepto	1	1,6373	1,1298				
		Idade	1	-2,4666	1,0473	5,5465	0,0185	0,085	[0,011; 0,661]
	Duas a Sete vezes	Intercepto	1	-0,8315	0,8274				
		Idade	1	0,2719	0,6244	0,1897	0,6632	1,313	[0,386; 4,463]
Idade	Até seis meses	Intercepto	1	-0,9641	0,3990				
		Atendimento	1	0,1348	0,1988	0,4601	0,4976	1,144	[0,775; 1,689]
	Mais de seis meses	Intercepto	1	-4,7999	1,5511				
		Atendimento	1	1,5041	0,5763	6,8111	0,0091	20,251	[6,445; 62,661]

Já quando o animal adoecido pela enfermidade em questão possui mais de seis meses de idade, o modelo é descrito por:

$$P(\text{Óbito} = 1) = \frac{\exp(-4,7999 + 1,5041 \cdot \text{atend})}{1 + \exp(-4,7999 + 1,5041 \cdot \text{atend})}$$

## CONCLUSÕES

a) O modelo logístico se prestou a identificar fatores de risco para Gastroenterite Hemorrágica em cães;

b) A variável que mais explica o óbito dos cães acometidos por Gastroenterite Hemorrágica é a idade;

c) Nas clínicas veterinárias de Lavras, quando os cães acometidos por Gastroenterite Hemorrágica são atendidos apenas uma vez, aqueles com idade superior a seis meses possuem 15,45 vezes mais chances de óbito do que os com menos de seis meses.

#### REFERÊNCIAS BIBLIOGRÁFICAS

- AGRESTI, A. **Categorical data analysis**. 2. ed. New Jersey: John Wiley, 2002. 526 p.
- ALLISON, P. D. **Logistic regression using the SAS® System: theory e application**. Cary: SAS Institute, 1999. 304 p.
- CHARNET, R.; FREIRE, C. A. de L.; CHARNET, R. M. R.; BONVINO, B. **Análise de modelos de regressão linear com aplicações**. Campinas: UNICAMP, 1999. 356 p.
- COLLET, D. **Modelling binary data**. London: Chapman & Hall, 1991. 369 p.
- COLOSIMO, E. **Análise de sobrevivência aplicada**. Belo Horizonte: Departamento de Estatística/UFMG, 2000. 109 p. No prelo.
- CORDEIRO, G. M. **Modelos lineares generalizados**. Campinas: SINAPE, 1986. 286 p.
- COSTA, S. C. da. **Regressão logística aplicada na identificação de fatores de risco para doenças em animais domésticos**. 1997. 104 f. Dissertação (Mestrado em Estatística e Experimentação Agronômica) – Escola Superior de Agricultura “Luiz de Queiroz”, Piracicaba, 1997.
- DEMÉTRIO, C. G. B. Modelos lineares generalizados na experimentação agronômica. In: REUNIÃO ANUAL DA REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA, 38.; SIMPÓSIO DE ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO AGRONÔMICA, 5., 1993, Porto Alegre. **Anais...** Porto Alegre: UFRS, 1993. p. 127.
- DEMÉTRIO, C. G. B. Modelos lineares generalizados em experimentação agronômica. In: REUNIÃO ANUAL DA REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA, 46.; SIMPÓSIO DE ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO AGRONÔMICA, 9., 2001, Piracicaba. **Anais...** Piracicaba: ESALQ, 2001. p. 113.
- FLETCHER, R. H.; FLETCHER, S. W.; WAGNER, E. H. Elementos essenciais. In: \_\_\_\_\_. **Epidemiologia clínica**. 3. ed. Porto Alegre: Artes Médicas Sul, 1996. p. 11-28.
- GONSALES, M.; KATZ, G.; ALVES, C. G. P.; CORTÁS, M. C. **Aplicação do método epidemiológico: investigação de caso**. São Paulo: Secretaria de Estado da Saúde, 1998. 28 p.
- HOSMER, D. W.; LEMESHOW, S. **Applied logistic regression**. New York: John Wiley, 1989. 307 p.
- MANTEL, N.; HAENSZEL, W. Statistical aspects of the analysis of data from retrospective studies of disease. **Journal of the National Cancer Institute**, Oxford, v. 22, p. 719-748, 1959.
- MCCULLOCH, C. E. **An introduction to generalized linear mixed models**. In: REUNIÃO ANUAL DA REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA, 46.; SIMPÓSIO DE ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO AGRONÔMICA, 9., 2001, Piracicaba. **Anais...** Piracicaba: ESALQ, 2001. p. 107.
- MONTGOMERY, D. C.; PECK, E. A. **Introduction to linear regression analysis**. 2. ed. New York: John Wiley & Sons, 1991. 527 p.
- MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. **Introduction to the theory of Statistics**. 3. ed. New York: McGraw-Hill, 1974. 564 p.
- SAS INSTITUTE. **SAS/STAT®: user's guide: version 6. 4**. ed. Cary, 1989. 2 v.
- SOARES, J. F.; COLOSIMO, E. Métodos estatísticos na pesquisa clínica. In: REUNIÃO ANUAL DA REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA, 40.; SIMPÓSIO DE ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO AGRONÔMICA, 6., 1995, Ribeirão Preto. **Anais...** Ribeirão Preto: USP; UNESP, 1995. p. 105.
- SOARES, J. F.; SIQUEIRA, A. L. **Introdução à estatística médica**. Belo Horizonte: Departamento de Estatística/UFMG, 1999. 300 p.



STOKES, M. E.; DAVIS, C. S.; KOCH, G. G. **Categorical data analysis using the SAS<sup>®</sup> System.** Cary: SAS Intitute, 1995. 499 p.

VENZON, D. J.; MOOLGAVKAR, S. H. A method for computing profile-likelihood based confidence intervals. **Applied Statistics**, Ann Arbor, v. 37, p. 87-94, 1988.