

## COMPARAÇÃO E APLICAÇÃO DE MODELOS DE REGRESSÃO BINÁRIA NA RETENÇÃO DE CAPACETES DE MOTOCICLISTAS

Ricardo Rasmussen PETTERLE<sup>1</sup>  
Cleiton Aquino de FREITAS<sup>1</sup>  
Ademir Marchi FURTADO<sup>1</sup>  
Fábio Henrique de CARVALHO<sup>2</sup>  
Wagner Hugo BONAT<sup>3</sup>

- RESUMO: O capacete é um equipamento de segurança de uso obrigatório para usuários de motocicletas, composto por um invólucro exterior e um sistema de retenção. Esse sistema mantém o capacete fixo à cabeça, podendo ser utilizado de maneira correta (firmemente apertado) ou incorreta (frouxo ou solto). Mediante o uso incorreto, a ejeção se torna mais provável agravando o traumatismo e a mortalidade dos condutores e passageiros. Nesse contexto, o objetivo deste trabalho é identificar características tanto do capacete quanto do motociclista associadas ao uso incorreto do sistema de retenção. As características avaliadas foram: tipo de capacete, tipo de engate, gênero e posição na moto. Os dados foram coletados na cidade de Curitiba/PR no mês de maio de 2014. A análise dos dados foi feita via modelos lineares generalizados. Como suposição distribucional foi usada a distribuição Bernoulli, uma vez que a variável resposta é binária. Foram comparadas através dos critérios de informação de Akaike e Bayesiano oito funções de ligação: logito, probito, complementar log-log, Cauchy, Aranda-Ordaz, Weibull, Prentice e Stukel. Dentre as funções de ligação avaliadas a Cauchy mostrou um melhor desempenho para o presente conjunto de dados. Os resultados mostraram que o gênero não foi estatisticamente significativo, porém o tipo de capacete, tipo de engate e a posição na moto representam características determinantes para o uso correto do capacete.

<sup>1</sup>Universidade Federal do Paraná - UFPR, Setor de Ciências da Saúde, Curso de Medicina, CEP: 80060-240, Curitiba, PR, Brasil. E-mail: *estatisticoufpr@gmail.com; cleitonaquinodefretas@live.com; ademir.furtado@gmail.com*

<sup>2</sup>Universidade Federal do Paraná - UFPR, Departamento de Cirurgia, CEP: 80060-240, Curitiba, PR, Brasil. E-mail: *fabiohc64@gmail.com*

<sup>3</sup>Universidade Federal do Paraná - UFPR, Departamento de Estatística, Laboratório de Estatística e Geoinformação, CEP: 81531-990, Curitiba, PR, Brasil. E-mail: *wbonat@ufpr.br*

- PALAVRAS-CHAVE: Capacete de motociclista; função de ligação; regressão binária; sistema de retenção.

## 1 Introdução

Segundo a Organização Mundial da Saúde (OMS) no ano de 2010 ocorreram 1,24 milhão de mortes por acidentes de trânsito, sendo a primeira causa de morte entre os indivíduos na faixa etária de 15 – 29 anos no mundo. Segundo estimativa dessa organização, em 2030 poder-se-á atingir a marca de 2,4 milhões de mortes. Por isso a OMS instituiu a década de 2011 a 2020 como a década de ação pela segurança no trânsito (WORLD HEALTH ORGANIZATION, 2013). No Brasil desde 2009, as mortes de motociclistas no trânsito lideram os índices de mortalidade (WAISELFISZ, 2014). Entre os agravantes para a mortalidade está o traumatismo crânio-encefálico que pode ser reduzido com o uso correto do capacete (LIU *et al.*, 2008).

O capacete, no Brasil, é um equipamento de segurança de uso obrigatório em vias públicas para usuários de motocicletas, motonetas, ciclomotores, triciclos e quadriciclos motorizados, segundo a resolução 453 de setembro de 2013 do Conselho Nacional de Trânsito (CONTRAN). Segundo essa resolução, o capacete deve ser afixado à cabeça por uma cinta jugular e um engate. O conjunto dessas duas estruturas, cinta e trava, forma o sistema de retenção do capacete ou retentor.

Quando o retentor está frouxamente ajustado proporciona folga entre a mandíbula e o sistema de retenção. Se o engate não for travado pelo usuário, esse sistema ficará totalmente solto. Nestas duas situações, o capacete não está plenamente otimizado e por não proporcionar a máxima segurança é dito incorretamente afixado (CONRAD *et al.*, 1996; XUEQUN *et al.*, 2011). Mediante o uso incorreto, a ejeção do capacete se torna mais provável, agravando o traumatismo e a mortalidade dos condutores e passageiros (RICHTER *et al.*, 2001). Duas características importantes dos capacetes comercializados no Brasil são o tipo do capacete (fechado, aberto ou modular) e o tipo de engate (serrilhado, não-serrilhado ou duplo-anel). Tais características podem estar associadas com o uso correto ou incorreto do sistema de retenção pelos motociclistas.

O principal objetivo deste artigo é identificar quais características do capacete como também do motociclista estão associadas com o uso correto ou incorreto do sistema de retenção. Identificar tais características é instrumental para ajudar a estabelecer normas e/ou leis que busquem diminuir a mortalidade de motociclistas através do uso correto do capacete. Potenciais características que foram avaliadas: tipo de capacete, tipo de engate, gênero do motociclista e posição na moto. Os dados foram coletados na cidade de Curitiba-PR durante duas semanas no mês de maio de 2014, onde mais de quatro mil motociclistas foram observados.

A modelagem estatística tem por objetivo relacionar uma variável resposta com um conjunto de covariáveis (variáveis explicativas). Para o caso de variável resposta binária (sucesso ou fracasso) o modelo de regressão logística é uma escolha

frequente (ARAUJO *et al.*, 2015; SOUZA *et al.*, 2014). O modelo de regressão logística conduz a interpretações simples além de ser flexível e de fácil utilização do ponto de vista matemático (GIOLO, 2012). Além disso, permite interpretar seus parâmetros em termos de *odds ratio* (razão de chances). O uso da função de ligação logito (canônica) é conveniente uma vez que garante a concavidade da função de variância bem como a existência de estatísticas suficientes para os coeficientes de regressão. Apesar da escolha desta ligação ser uma alternativa conveniente, outras ligações podem proporcionar melhor ajuste.

Neste sentido, a análise dos dados envolve a comparação de oito funções de ligação: logito, probito, complementar log-log, Cauchy, Aranda-Ordaz, Weibull, Prentice e Stukel. As quatro primeiras são funções de ligação que não dependem de parâmetros extras e estão disponíveis em vários softwares estatísticos como por exemplo no *software* R (R CORE TEAM, 2016) através da função `glm`. Por outro lado, as quatro últimas são funções de ligação mais flexíveis, uma vez que dependem de parâmetros extras (SANTOS, 2013).

O artigo está organizado da seguinte forma. A Seção 2 descreve a metodologia de coleta de dados e a metodologia estatística utilizada para análise. A Seção 3 apresenta os principais resultados obtidos. Seção 4 apresenta a discussão e as principais conclusões. O código R e o conjunto de dados estão disponíveis como material suplementar<sup>1</sup>.

## 2 Conjunto de dados

Os dados foram coletados em semáforos na cidade de Curitiba/PR durante a parada de motocicletas no sinal vermelho. Os pesquisadores observaram os motociclistas fronto-lateralmente. Não houve diálogo entre os pesquisadores e os motociclistas. Os horários, dias e os locais escolhidos previamente à coleta de dados foram sorteados entre dois observadores. As coletas ocorreram durante duas semanas no mês de maio de 2014, incluindo dias úteis e finais de semana, das 7h00min às 19h00min, em cinco períodos de duas horas por dia, intervaladas por quatro períodos de 30 minutos.

As coletas foram realizadas nas 9 regionais da cidade e a amostragem respeitou a proporcionalidade populacional da área observada. Cada regional foi avaliada quatro vezes em pontos equidistantes, conforme dados fornecidos pelo Instituto de Pesquisa e Planejamento Urbano de Curitiba ver Tabela 1.

Dois observadores coletaram os dados voluntariamente, sendo que ambos foram previamente treinados durante quatro semanas para reconhecer aos padrões a serem observados e avaliados. Ao final desse treinamento, os dois observadores coletaram 200 amostras oriundas dos mesmos motociclistas a fim de avaliar a concordância das avaliações através do teste de concordância Kappa.

Foram avaliadas cinco características em cada motociclista. Quatro prováveis variáveis explicativas e uma variável resposta. Posição na motocicleta (condutor

---

<sup>1</sup>disponível em: <http://www.leg.ufpr.br/doku.php/publications:papercompanions:rbb20161>

Tabela 1 - População absoluta e percentual de Curitiba por regionais, 2010

Regionais	População	Percentual
Boa Vista	248698	14,05
Portão	243506	13,76
Cajuru	215503	12,17
Matriz	205722	11,74
Boqueirão	197346	12,30
Cidade Industrial	171480	9,69
Pinheirinho	168425	9,52
Santa Felicidade	155794	8,80
Bairro Novo	145433	8,22
Total	1751907	100,00

ou passageiro), gênero (masculino ou feminino), capacete (fechado, modular, aberto, coquinho ou sem capacete) e tipo de engate (serrilhado, não-serrilhado, duplo-anel ou indeterminado) representavam as variáveis explicativas. O modo de uso do retentor, tanto correto (firmemente apertado) quanto incorreto (frouxo ou solto), representa a variável resposta. Os modelos de engates foram divididos conforme o mecanismo de engate: engate serrilhados, não-serrilhados e duplo-anel, respectivamente ver Figura 1. Os modelos de capacete fechado, modular e aberto estão demonstrados na Figura 2.



Figura 1 - Ilustração dos diferentes tipos de engates disponíveis em capacetes comercializados no Brasil.

Optou-se por cruzamento com dois semáforos em cruzamento em “X” ou em “T”. A direção na pista era escolhida conforme a possibilidade de visualização do lado esquerdo do capacete. A Figura 3 ilustra as circunstâncias possíveis durante a observação. Se o observador estivesse na posição A, e um semáforo estivesse indicando parada na via de sentido único, ele poderia avaliar apenas o motociclista 1, mas não o 4, 2 ou 3, pois o 4 estaria com o modelo de trava oposto à capacidade

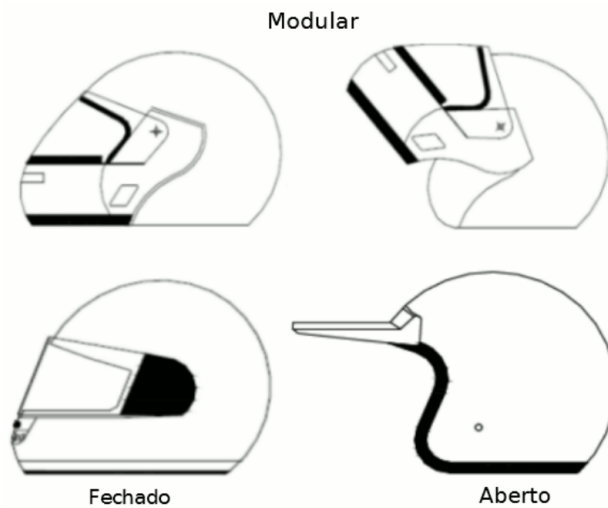


Figura 2 - Ilustração dos diferentes tipos de capacetes comercializados no Brasil.

de observação, o 2 estaria a uma distância grande e em movimento ou iniciando a movimentação e o 3 em movimento. Na Figura 3, a trava está corada em vermelho e o capacete em preto. Assim o observador *A* apenas poderia observar o 1 em plenitude. O observador *B* apenas poderia observar plenamente o 4, o *C* o 2 e o *D* não poderia observar nenhum, devendo mudar de posição. Motociclistas em movimento não foram avaliados.

A Tabela 2 mostra as covariáveis pertencentes ao estudo, sua codificação na base de dados e o total de observações conforme o nível de cada covariável. A partir da Tabela 2, observa-se que a maioria dos motociclistas está alocada predominantemente na categoria de tipo de capacete fechado ( $n = 3361$ ), engate não-serrilhado ( $n = 1719$ ), gênero masculino ( $n = 3655$ ) e condutores ( $n = 3738$ ).

### 3 Metodologia estatística

A modelagem estatística dos dados foi feita através dos modelos lineares generalizados (NELDER e WEDDERBURN, 1972). O software R (R CORE TEAM, 2016), com auxílio do pacote *ResourceSelection* (LELE *et al.*, 2015) foi utilizado para o ajuste dos modelos. Para as funções de ligação que dependem de parâmetros extras fez-se uso das funções implementadas por Santos (2013). O método de máxima verossimilhança foi utilizado para estimação dos parâmetros. Os modelos foram comparados através dos critérios de informação de *Akaike* (AIC), Bayesiano (BIC) e o valor maximizado da função de log-verossimilhança (LL).

Para avaliar a concordância entre os observadores fez-se uso do coeficiente Kappa ( $\kappa$ ) proposto por Cohen (1960) e implementado no pacote *vcd* (MEYER

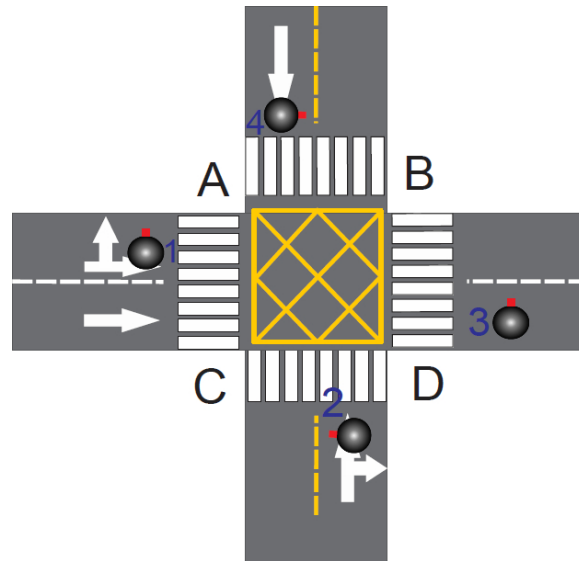


Figura 3 - Posicionamento dos observadores.

Tabela 2 - Análise descritiva das covariáveis tipo de capacete, tipo de engate, gênero e posição na moto

Covariável	Nível	Total	Uso correto	Uso incorreto
Tipo de capacete	0: fechado	3361 (83,8%)	1936 (57,6%)	1425 (42,4%)
	1: aberto	178 (4,4%)	125 (70,2%)	53 (29,8%)
	2: modular	470 (11,8%)	359 (76,4%)	111 (23,6%)
Tipo de engate	0: serrilhado	1189 (29,6%)	952 (80,0%)	237 (20,0%)
	1: não-serrilhado	1719 (42,8%)	780 (45,4%)	939 (54,6%)
	2: Duplo-anel	141 (3,6%)	73 (51,7%)	68 (48,3%)
	3: indeterminado	960 (24,0%)	615 (64,0%)	345 (36,0%)
Gênero	0: masculino	3655 (91,1%)	2238 (61,2%)	1417 (38,8%)
	1: feminino	354 (8,9%)	182 (51,4%)	172 (48,6%)
Posição na moto	0: condutor	3738 (93,2%)	2304 (61,6%)	1434 (38,4%)
	1: passageiro	271 (6,8%)	116 (42,8%)	155 (57,2%)

*et al.*, 2015). Este coeficiente varia entre 0 e 1, podendo ser classificado entre  $\kappa < 0,4$  concordância leve,  $0,4 \leq \kappa \leq 0,8$  concordância moderada,  $0,8 \leq \kappa < 1$  concordância forte e  $\kappa = 1$  concordância perfeita (ARANGO, 2012). A definição do coeficiente Kappa é apresentado na Eq. 1.

$$\kappa = \frac{\Pi_0 - \Pi_e}{1 - \Pi_e}, \quad (1)$$

em que  $\Pi_0$  é a proporção de concordância entre observadores e  $\Pi_e$  é a proporção de não concordância.

Sob a abordagem dos modelos lineares generalizados, considere o seguinte modelo

$$Y_i|x_i, \sim B(\mu_i, n = 1) \\ \mu_i(\boldsymbol{\theta}) = g^{-1}(x_i; \boldsymbol{\theta})$$

em que  $Y_i$  é uma variável aleatória com distribuição binomial com  $n = 1$  (número de realizações) para  $i = 1, \dots, n$ . Tal distribuição é descrita por um parâmetro de locação  $\mu_i(\boldsymbol{\theta})$ , que depende do vetor de parâmetros  $\boldsymbol{\theta}$  composto por dois conjuntos de parâmetros  $\boldsymbol{\beta}_x$  e  $\boldsymbol{\delta}$  que medem o efeito das covariáveis  $x_i$  e controlam o corpo da função de ligação, respectivamente. Para facilitar a apresentação das funções de ligação, denote  $\eta_i = x_i^\top \boldsymbol{\beta}_x$  o preditor linear. Neste estudo, a representação da variável resposta é dada por

$$Y_i = \begin{cases} 1, & \text{se uso incorreto do sistema de retenção} \\ 0, & \text{se uso correto do sistema de retenção.} \end{cases}$$

Assim, a probabilidade de uso incorreto do sistema de retenção dado os valores observados das covariáveis  $x_i$  é denotado por  $E(Y_i|x_i) = P(Y_i = 1|x_i)$ . O objetivo da função  $g(\cdot; \boldsymbol{\theta})$  é linearizar a relação entre os componentes aleatório e sistemático do modelo, permitindo que os valores das covariáveis pertencentes ao intervalo  $(-\infty, \infty)$ , tenham um correspondente no intervalo  $[0, 1]$ . O estimador de máxima verossimilhança de  $\boldsymbol{\theta}$  é obtido pela maximização da função de log-verossimilhança expressa por

$$l(\boldsymbol{\theta}) = \sum_{i=1}^n y_i \log(\mu_i(\boldsymbol{\theta})) + (1 - y_i) \log(1 - \mu_i(\boldsymbol{\theta})).$$

Neste artigo foram consideradas as funções de ligação: logito, probito, complementar log-log, Cauchy, Aranda-Ordaz, Weibull, Prentice e Stukel. As quatro primeiras definem ligações que não dependem de parâmetros extras e as quatro últimas dependem de parâmetros extras que controlam o formato da função de ligação.

A função de ligação logito é a canônica e mais popular para dados binários (HOSMER e LEMESHOW, 2000). As expressões da função de ligação e sua inversa são apresentadas na Eq. 2.

$$g(\mu_i) = \ln \left( \frac{\mu_i}{1 - \mu_i} \right) \quad \text{e} \quad \mu_i = \frac{\exp\{\eta_i\}}{1 + \exp\{\eta_i\}}. \quad (2)$$

Apesar da ligação canônica ser uma alternativa conveniente, outras ligações podem proporcionar melhor ajuste. O modelo probito apresenta forma sigmoidal,

é simétrico em torno de zero e faz uso da função de distribuição da Normal padrão. As expressões da função de ligação e sua inversa são dadas, respectivamente, por

$$g(\mu_i) = \Phi^{-1}(\mu_i) \quad \text{e} \quad \mu_i = \Phi(\eta_i),$$

onde  $\Phi(\cdot)$  denota a função de distribuição da Normal padrão.

A função de ligação Cauchy utiliza a função de distribuição Cauchy. Este modelo também apresenta forma sigmoideal e simetria em torno de zero. As expressões para a função de ligação e sua inversa são dadas, respectivamente, por

$$g(\mu_i) = \tan\left(\pi\left(\mu_i - \frac{1}{2}\right)\right) \quad \text{e} \quad \mu_i = \pi \operatorname{csc}^2(\pi\eta_i),$$

onde  $\tan$  e  $\operatorname{csc}$  denotam as funções tangente e cosecante, respectivamente.

Dentre os modelos que não dependem de parâmetros extras, o único que não é simétrico em torno de zero é o modelo complementar log-log, que usa a distribuição valor mínimo extremo tipo I (GUMBEL, 1935). A definição da ligação e sua inversa são, respectivamente

$$g(\mu_i) = \ln(-\ln(1 - \mu_i)) \quad \text{e} \quad \mu_i = 1 - \exp\{-\exp\{\eta_i\}\}.$$

As funções de ligação logito, proibito e Cauchy apresentam duas propriedades importantes. Quando  $\eta_i = 0$  o valor de  $\mu_i$  é 0,5 e para qualquer número real positivo  $k$  existe simetria, ou seja  $g^{-1}(k) = g^{-1}(-k)$ . A Figura 4(A) apresenta o formato destas funções de ligação, onde é possível identificar tais propriedades. No caso da função de ligação complementar log-log o valor de  $\eta_i = 0$  resulta em  $\mu_i = 0,63$ . Portanto, quando  $\eta_i = 0$  tem-se apenas duas possibilidades para os valores estimados da média,  $\mu_i = 0,5$  e  $\mu_i = 0,63$ . Partindo destas limitações, com o objetivo de tornar o modelo mais flexível neste artigo foram consideradas funções de ligação que dependem de parâmetros extras, permitindo controlar não somente a assimetria, mais também o formato das caudas da função de ligação.

Aranda-Ordaz (1981) propôs um modelo que não se origina de nenhuma distribuição de probabilidade, porém apresenta um parâmetro extra denotado por  $\delta > 0$ . Quando  $\delta = 1$  obtém-se o modelo logístico e na medida em que  $\delta$  tende a zero ( $\delta \rightarrow 0$ ) obtém-se o modelo complementar log-log. Essa ligação é definida da seguinte forma:

$$g(\mu_i) = \ln\left\{\frac{(1 - \mu_i)^{-\delta} - 1}{\delta}\right\},$$

e sua inversa é dada por

$$\mu_i = 1 - (1 + \delta \exp\{\eta_i\})^{-1/\delta}.$$

A Figura 4(B) apresenta o comportamento desta função conforme alguns valores do parâmetro  $\delta$ . Nota-se que quanto menor o valor de  $\delta$  mais rápido a função se aproxima de  $\mu_i = 1$ .



O modelo Weibull proposto por Caron et al. (2009) e Caron (2010) deriva da distribuição Weibull com três parâmetros. Nesse modelo, o parâmetro adicional é representado por  $\delta > 0$ . O modelo complementar log-log é um caso particular do Weibull, obtido na medida em que o parâmetro  $\delta$  tende a infinito ( $\delta \rightarrow \infty$ ). A função de ligação Weibull e sua inversa são representadas da seguinte forma:

$$g(\mu_i) = [-\ln(1 - \mu_i)]^{1/\delta} \quad \text{e} \quad \mu_i = 1 - \exp\{-\eta_i^\delta\}.$$

Conforme aumentam os valores do parâmetro  $\delta$  o formato dessa ligação se altera, chegando mais próxima de  $\mu_i = 1$  (Figura 4)(C). Note que  $\delta$  é positivo, pois atende as mesmas condições da distribuição Weibull.

Prentice (1975) apresentou seu modelo a partir do logaritmo da função de distribuição acumulada da densidade  $\mathcal{F}$  com dois parâmetros,  $\delta_1$  e  $\delta_2$ . A partir dessa ligação, diversos modelos são obtidos como casos particulares conforme os valores de seus dois parâmetros extras. Dessa forma, quando os valores dos parâmetros  $\delta_1$  e  $\delta_2$  são  $(1, 1)$ ,  $(1, \infty)$ ,  $(\infty, 1)$  e  $(\infty, \infty)$  tem-se, respectivamente, as funções de ligação logito, valor mínimo extremo, valor máximo extremo e normal degenerada. Em geral esta função de ligação não apresenta forma fechada, sendo obtida através da densidade da distribuição  $\mathcal{F}$  com parâmetros  $\delta_1$  e  $\delta_2$  dada por:

$$f(y) = \frac{\exp(y\delta_1)(1 + \exp(y))^{-(\delta_1 + \delta_2)}}{\mathcal{B}(\delta_1, \delta_2)},$$

onde  $\mathcal{B}$  representa a função beta. A Figura 4(D) apresenta a forma desta função de ligação de acordo com alguns valores para os parâmetros  $\delta_1$  e  $\delta_2$ .

O modelo proposto por Stukel (1988) é basicamente uma generalização do modelo logito. Antes da função de ligação logito, é aplicada uma função  $h_\delta$  no componente sistemático do modelo permitindo controlar suas caudas, de maneira independente, conforme valores de seus dois parâmetros extras. O modelo logístico é um caso particular, obtido quando ambos os parâmetros são iguais a zero. Outros modelos são obtidos por aproximação como o probito ( $\delta_1 = \delta_2 \approx 0,165$ ), log-log ( $\delta_1 = -0,037$  e  $\delta_2 = 0,62$ ) e complementar log-log ( $\delta_1 = 0,62$  e  $\delta_2 = -0,037$ ). Sua definição é:

$$g(\mu_i) = h_\delta^{-1} \left( \ln \left( \frac{\mu_i}{1 - \mu_i} \right) \right),$$

em que o vetor de parâmetros  $\delta = (\delta_1, \delta_2)$  define sua forma. A inversa da função de ligação é:

$$\mu_i = \frac{\exp\{h_\delta(\eta_i)\}}{1 + \exp\{h_\delta(\eta_i)\}}.$$

A função  $h_\delta$  para  $\eta_i > 0$  ou de maneira semelhante para  $\mu_i > 0,5$  é definida como:

$$h_\delta = \begin{cases} \delta_1^{-1} (\exp(\delta_1 |\eta_i|) - 1) & \text{se } \delta_1 > 0 \\ \eta_i, & \text{se } \delta_1 = 0 \\ -\delta_1^{-1} \log(1 - \delta_1 |\eta_i|) & \text{se } \delta_1 < 0 \end{cases}$$

Para  $\eta_i \leq 0$  ou de maneira equivalente para  $\mu_i \leq 0,5$  a função  $h_\delta$  é dada por:

$$h_\delta = \begin{cases} -\delta_2^{-1} (\exp(\delta_2|\eta_i|) - 1) & \text{se } \delta_2 > 0 \\ \eta_i, & \text{se } \delta_2 = 0 \\ \delta_2^{-1} \log(1 - \delta_2|\eta_i|) & \text{se } \delta_2 < 0 \end{cases}$$

Para valores de  $\delta_1$  e  $\delta_2$  diferentes a função de ligação é assimétrica, porém a simetria é obtida quando  $\delta_1 = \delta_2$ . A Figura 4(E) apresenta a forma desta função de ligação de acordo com alguns valores para os parâmetros  $\delta_1$  e  $\delta_2$ .

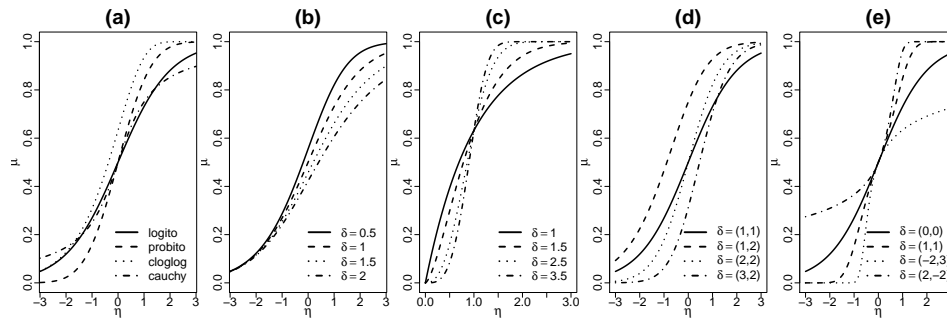


Figura 4 - Funções de ligação logito, probito, complementar log-log e Cauchy (A). Função de ligação Aranda-Ordaz (B), Weibull (C), Prentice (D) e Stukel (E) para diferentes conjuntos de parâmetros.

## 4 Resultados

Nesta seção apresentamos os principais resultados da análise dos dados. Os valores do teste de concordância Kappa e intervalos com 95% de confiança obtidos via bootstrap, referente as 200 avaliações, encontram-se na Tabela 3. Os tipos de engate serrilhado e não-serrilhado apresentam forte concordância e o tipo de engate duplo-anel, concordância perfeita. O modo de uso do capacete nas situações preso e frouxo apontou forte concordância entre os observadores. Finalmente, o modo de uso do capacete avaliado em solto demonstrou concordância perfeita ( $\kappa = 1$ ).

Para o ajuste do modelo de regressão, foram consideradas como covariáveis o tipo de capacete (aberto, fechado ou modular), tipo de engate (serrilhado, não-serrilhado, duplo-anel ou indeterminado), gênero (masculino ou feminino) e posição na moto (condutor ou passageiro). Em uma primeira análise foram ajustados modelos contendo todas as covariáveis e usando as oito funções de ligação consideradas neste artigo. A Tabela 4 apresenta os valores dos critérios de informação de Akaike(AIC) e Bayesiano(BIC), bem como, os valores maximizados da função de log-verossimilhança (LL) e o número de parâmetros (NP) envolvidos nos modelos de acordo com as diferentes funções de ligação.

Tabela 3 - Teste de concordância Kappa e intervalos com 95% de confiança por nível das covariáveis tipo de engate e modo de uso

Covariáveis		Coefficiente Kappa ( $\kappa$ )
Tipo de engate	Serrilhado	0,93(0,87 - 0,98)
	Não-serrilhado	0,96(0,95 - 1,00)
	Duplo-anel	1,00(1,00 - 1,00)
Modo de uso	Preso	0,97(0,93 - 1,00)
	Frouxo	0,93(0,85 - 0,99)
	Solto	1,00(1,00 - 1,00)

Tabela 4 - Critérios de informação de *Akaike* (AIC), Bayesiano (BIC), valor maximizado da função de log-verossimilhança (LL) e número de parâmetros de acordo com diferentes funções de ligação

Função de ligação	NP	LL	AIC	BIC
Logito	7	-2466,64	4947,29	4991,36
Probit	7	-2467,27	4948,54	4992,61
Complementar log-log	7	-2466,27	4946,55	4990,63
Cauchy	7	-2465,04	4944,09	4988,16
Aranda-Ordaz	8	-2466,27	4948,55	4998,92
Weibull	8	-2466,43	4948,86	4999,23
Prentice	9	-2465,04	4948,08	5004,75
Stukel	9	-2464,42	4946,85	5003,52

Os resultados apresentados na Tabela 4 mostram que o modelo Stukel apresentou a maior log-verossimilhança (-2464,42), porém os critérios de informação AIC e BIC mostram valores inferiores aos obtidos pelo modelo usando a função de ligação Cauchy (AIC = 4944,09 e BIC = 4988,16). A diferença em termos de log-verossimilhança entre os modelos Cauchy e Stukel é pequena (0,62), indicando que a inclusão de dois parâmetros extras não resulta em significativa melhora do ajuste. Dessa forma, opta-se pela ligação Cauchy como sendo a que proporciona o melhor balanço entre qualidade de ajuste e simplicidade. Baseado neste modelo a covariável gênero mostrou-se não significativa e foi removida da análise.

A Tabela 5 apresenta as estimativas dos parâmetros, erros-padrão e valores-*p*. O modelo final conta com o efeito de três covariáveis cujas categorias de referência são: capacete fechado, engate serrilhado e condutor, respectivamente.

Os resultados da Tabela 5 mostram que todas as covariáveis apresentam significância. Antes de interpretar os coeficientes associados a cada covariável é interessante apresentar a análise de resíduos para avaliar a qualidade do ajuste. A Figura 5 apresenta os resíduos de Pearson e *deviance* associados ao modelo de regressão binária ajustado usando a função de ligação Cauchy.

De forma geral os resíduos apresentam valores entre -2 e 2 o que indica um

Tabela 5 - Estimativas dos parâmetros, erros-padrão e valores- $p$  do modelo de regressão binária usando função de ligação Cauchy

Parâmetros	Coefficiente	Erro- Padrão	Valor- $p$
$\beta_0$ : constante	-1,2946	0,1044	0,0000
$\beta_1$ : capacete aberto	-0,6620	0,1754	0,0001
$\beta_2$ : capacete modular	-0,6684	0,1287	0,0000
$\beta_3$ : engate não-serrilhado	1,4914	0,1111	0,0000
$\beta_4$ : engate duplo-anel	1,2690	0,1722	0,0000
$\beta_5$ : engate indeterminado	0,8378	0,1199	0,0000
$\beta_6$ : posição na moto passageiro	0,5923	0,1245	0,0000

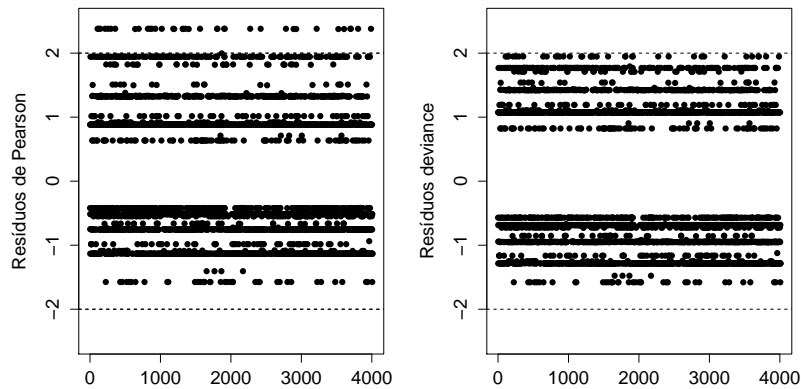


Figura 5 - Resíduos de Pearson e *deviance* associados ao modelo de regressão binária usando função de ligação Cauchy.

bom ajuste do modelo aos dados. Excessão aparece para os resíduos de Pearson que apresenta alguns valores acima de 2, porém tais valores não afetam drasticamente a qualidade do ajuste. Tal resultado, é também confirmado pelo teste the Hosmer-Lemeshow que indicou a não rejeição da hipótese nula de que o modelo apresenta um bom ajuste ( $p - valor = 0,9988$ ).

A Tabela 6 mostra a probabilidade de uso incorreto do sistema de retenção conforme o tipo de capacete, tipo de engate e posição na moto. O capacete fechado, seja ele usado por condutores ou passageiros, apresentou as maiores probabilidades de uso incorreto quando comparado aos capacetes modular e aberto. Os engates não-serrilhado e duplo-anel demonstraram resultados piores em relação ao modelo serrilhado, sugerindo, portanto, o modelo serrilhado como o mais adequado.

A situação mais favorável ao uso correto do sistema de retenção é para

motociclistas com capacete modular, engate serrilhado e condutor da moto. Finalmente, há indicações de que condutores fazem melhor uso do sistema de retenção do que os passageiros.

Tabela 6 - Probabilidades e intervalo com 95% de confiança para o uso incorreto do sistema de retenção, conforme o tipo de capacete, tipo de engate e posição na moto

Tipo de capacete	Tipo de engate	Posição na moto	Probabilidade
Fechado	Não-serrilhado	Passageiro	0,7127(0,6580 - 0,7558)
Fechado	Duplo-anel	Passageiro	0,6641(0,5661 - 0,7372)
Fechado	Não-serrilhado	Condutor	0,5619(0,5360 - 0,5869)
Fechado	Indeterminado	Passageiro	0,5429(0,4646 - 0,6163)
Aberto	Não-serrilhado	Passageiro	0,5402(0,4146 - 0,6549)
Modular	Não-serrilhado	Passageiro	0,5383(0,4303 - 0,6383)
Fechado	Duplo-anel	Condutor	0,4918(0,4084 - 0,5764)
Aberto	Duplo-anel	Passageiro	0,4697(0,3339 - 0,6168)
Modular	Duplo-anel	Passageiro	0,4677(0,3440 - 0,6015)
Fechado	Indeterminado	Condutor	0,3636(0,3326 - 0,3976)
Aberto	Não-serrilhado	Condutor	0,3614(0,2844 - 0,4601)
Modular	Não-serrilhado	Condutor	0,3597(0,3017 - 0,4297)
Aberto	Indeterminado	Passageiro	0,3457(0,2607 - 0,4626)
Modular	Indeterminado	Passageiro	0,3441(0,2701 - 0,4419)
Aberto	Duplo-anel	Condutor	0,3083(0,2324 - 0,4196)
Modular	Duplo-anel	Condutor	0,3069(0,2414 - 0,3978)
Fechado	Serrilhado	Passageiro	0,3051(0,2499 - 0,3778)
Aberto	Indeterminado	Condutor	0,2322(0,1893 - 0,2932)
Modular	Indeterminado	Condutor	0,2313(0,1979 - 0,2748)
Fechado	Serrilhado	Condutor	0,2094(0,1872 - 0,2363)
Aberto	Serrilhado	Passageiro	0,2013(0,1609 - 0,2629)
Modular	Serrilhado	Passageiro	0,2006(0,1652 - 0,2516)
Aberto	Serrilhado	Condutor	0,1504(0,1280 - 0,1813)
Modular	Serrilhado	Condutor	0,1500(0,1320 - 0,1731)

## Discussão

Este artigo identificou características associadas ao uso incorreto do sistema de retenção em motociclistas na cidade de Curitiba/PR. Foram avaliados tanto condições do capacete quanto do motociclista. A análise dos dados envolveu o ajuste e escolha de modelos de regressão para dados binários usando oito funções de ligação, sendo quatro não dependentes de parâmetros extras e quatro dependentes.

A função de ligação Cauchy foi indicada pelos critérios AIC e BIC como a que apresenta o melhor ajuste para o presente conjunto de dados. Uma limitação desta função de ligação quando comparada com a canônica (logito) é que a interpretação

dos coeficientes de regressão não pode ser feita em termos de razão de chances. Neste artigo optou-se pela construção da Tabela 6 que apresenta a probabilidade de uso incorreto do capacete para todas as combinações possíveis entre os níveis das covariáveis consideradas no modelo. Para situações mais gerais, como aquelas que envolvem covariáveis contínuas soluções alternativas devem ser propostas.

Em termos gerais, a escolha da função de ligação e a decisão quanto a sua adequação pode ser complicado em modelos de regressão para dados binários. Isto é devido a ausência de gráficos exploratórios que podem ser usados para avaliar o ajuste do modelo. Sendo assim, a estratégia padrão e adotada neste artigo é ajustar o modelo usando várias funções e comparar o ajuste usando critérios como AIC e BIC. Esta estratégia foi usado em Bonat *et. al.* (2012) no contexto de regressão para dados no intervalo unitário. Em geral os autores concluíram que os resultados finais obtidos pelo modelo em termos da significância dos coeficientes de regressão são pouco sensíveis a escolha da função de ligação. Por outro lado, Santos (2013) comparou diferentes funções de ligação para modelar a proporção de besouros mortos em função do logaritmo natural da dose de inseticida. Esse conjunto de dados tem a característica de ser modelado por ligações assimétricas, como, por exemplo, a ligação complementar log-log. Neste caso, a escolha da função de ligação tem um papel importante sobre os resultados finais obtidos do modelo. Neste sentido, é recomendável usar diferentes funções de ligação como uma estratégia segura de análise.

Para o presente conjunto de dados a escolha da função de ligação não mostrou-se importante, uma vez que todas as ligações consideradas apresentaram resultados próximos em termos de log-verossimilhança maximizada, ver Tabela 4. Porém, em termos computacionais as funções de ligação que dependem de parâmetros extras são mais caras e mais sujeitas a não convergência do algoritmo de maximização numérica.

Quanto a análise dos dados, observou-se que o tipo de capacete, tipo de engate e posição na moto foram características importantes para o uso correto do sistema de retenção. O gênero tanto do condutor como do passageiro não é uma característica importante. Os resultados mostraram que os condutores fazem melhor uso do capacete do tipo modular com engate serrilhado. A situação menos favorável para o uso correto do sistema de retenção se apresentou para os passageiros com o tipo de capacete fechado e o engate não-serrilhado.

Os resultados obtidos através da análise dos dados mostram que apesar de todos os capacetes disponíveis no mercado brasileiro apresentarem especificações técnicas corretas para proteger o motociclista, a probabilidade de uso correto do sistema de retenção depende do tipo de capacete e engate. Embora o estudo não possa explicar o porquê de tais diferenças, elas existem e podem estar associadas tanto à engenharia de construção do produto como a adaptação do motociclista. Os resultados obtidos podem ser usados para identificar tipos de capacetes e engates que devem ser evitados por apresentarem uma alta chance de serem usados incorretamente. Além disso, futuros trabalhos devem buscar entender o que leva o motociclista a usar o capacete incorretamente.

PETTERLE, R. R.; FREITAS, C. A.; FURTADO, A. M.; CARVALHO, F. H.; BONAT, W. H. Comparison and application of binary regression models on the evaluation of helmet retention by motorcyclists. *Rev. Bras. Biom.*, Lavras, v.35, n.2, p.266-282, 2017.

■ **ABSTRACT:** *The helmet is a security equipment of mandatory use for Brazilian motorcyclists. It is composed of an outer wrapper and a retention system. The usefulness of such a system is keeping the helmet on the head. The helmet can be used in the correct or incorrect way. If used incorrectly, the helmet can be more likely ejected, increasing the mortality of the motorcyclists. The main goal of this paper is to identify factors associated with the incorrect use of the helmet. We evaluated the helmet type, coupling type, sex and position on the motorcycle. The data was collected in Curitiba, PR, Brazil. We fitted generalized linear models using the Bernoulli distribution and compared the fit of eight link functions: logit, probit, complement log-log, Cauchy, Aranda-Ordaz, Weibull, Prentice and Studel. For our data, the Cauchy link function presented the best fit. The results showed that gender is not statistically significant, however, helmet type, coupling type and position on the motorcycle are important factors to determine the probability of correct use of the helmet.*

■ **KEYWORDS:** *Motorcycle helmet; retention system; binary regression; link function.*

## Referências

ARANDA-ORDAZ, F. J. On two families of transformations to additivity for binary response data. *Biometrika*, v.68, n.2, p.357–363, 1981.

ARANGO, H. G. *Bioestatística teórica e computacional*. 3 ed. Rio de Janeiro: Guanabara Koogan, 2012.

ARAÚJO, F. C. B.; SOUZA, C. N. P.; RAMOS, E. M. L. S.; BRAGA, R. M. Aspectos associados à recidiva da hanseníase. *Revista Brasileira de Biometria*, São Paulo, v.33, n.1, p.42-50, 2015.

BONAT, W. H.; RIBEIRO Jr, P. J.; ZEVIANI, W. M. Regression models for responses in the unit interval: specification, estimation and comparison. *Revista Brasileira de Biometria*, São Paulo, v.30, n.4, p.415-431, 2012.

CARON, R. *Regressão de dados binários: Distribuição weibull*. Dissertação de Mestrado, Universidade Federal de São Carlos, 2010.

CARON, R.; POLPO, A.; GOGGANS, P. M.; CHAN, C. Y. Binary data regression: Weibull distribution. *Aip Conference Proceedings*, v.1193, n.1, p.187-193, 2009.

CONRAD, P. *et al.* Helmets, injuries and cultural definitions: Motorcycle injury in urban Indonesia. *Accident Analysis and Prevention*, v.28, n.2, p. 193–200, 1996.

COHEN, J. A. coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, v.20, n.1, p.37-46, 1960.

- GIOLO, S. R. *Introdução à análise de dados categóricos*. UFPR. 2012. Disponível em: [http://www.rbras.org.br/rbras57/sites/default/files/SRGIolo\\_57Rbras.pdf](http://www.rbras.org.br/rbras57/sites/default/files/SRGIolo_57Rbras.pdf). Acesso em: 13 Set. 2015.
- GUMBEL, E. J. Les valeurs extrêmes des distributions statistiques. *Annales de l'Institut Henri Poincaré*, v.5, n.2, p.115-158, 1935.
- HOSMER, D. W.; LEMESHOW, S. *Applied logistic regression*, 2.ed., New York: John Wiley e Sons, 2000, 375p.
- LELE, S. R.; KEIM, J. L.; SOLYMOS P. *Resource selection: Resource selection (probability) functions for use-availability data*. R package version 0.2-5, 2015. Disponível em: <http://CRAN.R-project.org/package=ResourceSelection>. Acesso em: 7 Fev. 2016.
- LELE, S. R.; KEIM, J. L.; SOLYMOS P. *Resource selection: Resource selection (probability) functions for use-availability data*. R package version 0.2-5, 2015. Disponível em: <http://CRAN.R-project.org/package=ResourceSelection>. Acesso em: 7 Fev. 2016.
- LIU, B. C. *et al.* Helmets for preventing injury in motorcycle riders. *Cochrane database of systematic reviews* (Online) n.1, 2008.
- MEYER, D.; ZEILEIS A.; HORNIK K. *vcd: Visualizing categorical data*. R, package version 1.4-1, 2015.
- NELDER, J. A. ; WEDDERBURN, W. M. Generalized linear models. *Journal of the Royal Statistical Society. Series A*, v.135, n.3, p.370-384, 1972.
- PRENTICE, R. L. Discrimination among some parametric models. *Biometrika*, v.62, n.3, p.607-614, 1975.
- R CORE TEAM. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing, Austria, 2016
- RICHTER, M. *et al.* Head injury mechanisms in helmet-protected motorcyclists: Prospective multicenter study. *The Journal of Trauma*, v.51, n.5, p.949-58, 2001.
- SANTOS, B. P. *Implementação em R de modelos de regressão binária com ligação paramétrica*. Dissertação de mestrado, Universidade de São Paulo, 2013.
- SOUZA, C. N. P.; LISBOA, J. L. C.; RAMOS, E. M. L. S.; ALMEIDA, S. S.; ARAÚJO, A. D. Fatores contribuintes à ocorrência de mortalidade por doença de Chagas. *Revista Brasileira de Biometria*, São Paulo, v.32, n.4, p.544-552, 2014.
- STUKEL, T. A. Generalized logistic models. *Journal of the American Statistical Association*, v.83, n.402, p.426-431, 1988.
- WASELFISZ, J. J. SECRETARIA NACIONAL DE JUVENTUDE. *Mapa da violência 2014: os jovens do Brasil*. Disponível em: [http://www.mapadaviolencia.org.br/pdf2014/Mapa2014\\_JovensBrasil.pdf](http://www.mapadaviolencia.org.br/pdf2014/Mapa2014_JovensBrasil.pdf). Acesso em: 22 Dez. 2015.
- WORLD HEALTH ORGANIZATION. *Global status report on road safety*, Geneva: World Health Organization, 2013.



XUEQUN, Y.; KE, L.; IVERS, R.; DU, W., SENSERRICK, T. Prevalence rates of helmet use among motorcycle riders in a developed region in China. *Accident Analysis Preview*, v.43, n.1, p.214–219, 2011.

Recebido em 25.02.2016.

Aprovado após revisão em 15.09.2016.