



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE CIÊNCIAS
DEPARTAMENTO DE ESTATÍSTICA E MATEMÁTICA APLICADA
CURSO DE GRADUAÇÃO EM ESTATÍSTICA

LINDEMBERGSON APOLINÁRIO DA SILVA

**USO DOS MODELOS GAMA E NORMAL INVERSO PARA A ANÁLISE DE
DADOS POSITIVOS ASSIMÉTRICOS**

FORTALEZA

2022

LINDEMBERGSON APOLINÁRIO DA SILVA

USO DOS MODELOS GAMA E NORMAL INVERSO PARA A ANÁLISE DE DADOS
POSITIVOS ASSIMÉTRICOS

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Estatística do Centro de Ciências da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de bacharel em Estatística.

Orientadora: Profa. Dra. Jeniffer Johana Duarte Sanchez

FORTALEZA

2022

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

D11u da Silva, Lindembergson Apolinário.

 Uso dos modelos gama e normal inverso para a análise de dados positivos assimétricos /
Lindembergson Apolinário da Silva. – 2022.
 79 f. : il. color.

 Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Centro de Ciências,
Curso de Estatística, Fortaleza, 2022.

 Orientação: Profa. Dra. Jeniffer Johana Duarte Sanchez .

 1. Regressão . 2. Modelos Lineares Generalizados . 3. Dados Positivos Assimétricos . 4. Limite de
 Crédito . I. Título.

CDD 519.5

LINDEMBERGSON APOLINÁRIO DA SILVA

USO DOS MODELOS GAMA E NORMAL INVERSO PARA A ANÁLISE DE DADOS
POSITIVOS ASSIMÉTRICOS

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Estatística do Centro de Ciências da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de bacharel em Estatística.

Aprovada em:

BANCA EXAMINADORA

Profa. Dra. Jeniffer Johana Duarte
Sanchez (Orientadora)
Universidade Federal do Ceará (UFC)

Prof. Dr. Gualberto Segundo Agamez Montalvo
Universidade Federal do Ceará (UFC)

Prof. Dr. José Roberto Silva dos Santos
Universidade Federal do Ceará (UFC)

À minha família e amigos, por acreditarem em mim e por terem me apoiado tanto nessa caminhada. À minha Mãe, essa grande mulher, que sem o seu cuidado, carinho e dedicação eu não teria alcançado nada disso. Você representou em todos esses tempos a confiança e a certeza de que eu nunca estive só. À minha amada Sanara e à nossa pequena Júlia.

AGRADECIMENTOS

Gostaria de agradecer:

A Deus, por me oferecer disposição, força, motivação e por colocar em minha vida pessoas excepcionais e que foram de grande importância, além de me abençoar todos os dias e me dar diversas oportunidades todos os dias.

À minha orientadora, professora Jeniffer Johana Duarte Sanchez por todo o apoio, pela enorme solicitude e por sempre acreditar que eu realizaria este trabalho. Agradeço pela oportunidade de ser seu orientando, pela sua competência e pela sua enorme paciência que teve comigo ao longo da minha monografia.

Ao professor Maurício, meu grande mestre (e acredito que de todos nós), que com seu entusiasmo e amor ao que faz, motivou inúmeros alunos e professores. Ao professor Maurício, só desejo tudo o que há de bom nessa vida. Que Deus o ilumine sempre. Sou eternamente grato por tê-lo conhecido pois foi como um pai para todos nós. Um sincero obrigado por sempre acreditar em mim. Não tenho palavras que possam descrever toda minha gratidão, carinho e amor.

Ao professor Juvêncio Nobre, grande mestre e amigo. O senhor tem um espaço reservado em meu coração. Me ajudou de uma forma surreal em vários momentos difíceis que passei e me ensinou o valor da humildade e honestidade acima de tudo. Você já fez por mim o que muitos familiares meus nunca fizeram e por isso você tem toda a minha admiração, respeito e carinho.

Ao professor Gualberto Agamez, por todo o apoio prestado conosco. É realmente admirável a forma com que se importa com os alunos. Lembro que fazia o possível e o impossível para que tivéssemos uma boa base, para que pudéssemos ser a melhor versão de nós mesmos. Sempre demonstrou muita competência e respeito com todos nós. Você é um exemplo para mim.

À minha mãe, uma mulher batalhadora e que sempre lutou para que nada faltasse pra mim e pra minha irmã. Você é tudo na minha vida e com toda a certeza desse mundo, sem você eu não teria conseguido alcançar tudo o que consegui hoje. Obrigado pelas palavras de conforto, por acreditar sempre em mim, por me apoiar em todas as minhas decisões mesmo não gostando muito de algumas (risos). Me cobrava todo dia para terminar minha monografia e fazia de tudo para que eu pudesse me concentrar e ter o máximo de conforto. Você é admirável e sempre será o meu grande amor. Dedico esse

trabalho à senhora.

À minha namorada Sanara (e se Deus quiser minha futura esposa) por toda a paciência, carinho, amor e companheirismo. Desde que te conheci minha vida mudou de uma forma inexplicável. Graças a você eu conquistei muitas coisas e aprendi tantas outras. Eu gosto muito do seu jeito de resolver as coisas, da forma que consegue me ganhar todos os dias e do jeito que demonstra todo o seu amor por mim. Você me deu a maior alegria que eu poderia ter, a nossa filha, a Julinha, a nossa “jujubinha” que hoje está com 6 meses na barriguinha da mamãe. Nossa pequena Júlia já nos traz tanta alegria. Serão as duas mulheres mais importantes da minha vida. Eu as amo mais do que tudo, não se esqueçam disso.

A todos os meus amigos da UFC, em especial: Yohana, Jamily, Matheus Osterno, Diego Santos, Alysson, Alberto (USP! USP! USP!) e Vitor (Vitão) por todos os momentos bons que vivemos na UFC. Vocês foram essenciais em toda a minha trajetória e por isso só tenho a agradecer.

Aos meus amigos Chrystian Monteiro (Magão do Potira), Lia Menezes e Elton Gomes (Eltão Balada), por todo o carinho e amor e por todos os momentos inesquecíveis, farras, pelos funks escutados nas alturas. Nunca esquecerei do apoio de vocês, de estarem sempre na torcida por mim e por acreditarem em mim. Vocês estarão sempre nas minhas prioridades e no fundo do meu coração. Contem sempre comigo.

Ao meu querido amigo Felipe Soares, por todos os momentos engraçados que passamos juntos e pela longa amizade. Nunca esquecerei de você e do seu apoio. Fico muito orgulhoso das suas conquistas, pelo seu casamento e pelo seu crescimento pessoal e profissional. Se tornou um homem honrado e responsável. Obrigado por tudo.

Aos meus amigos do MetrÓpole, Mariana Dantas, Matheus Sena, Matheus Teixeira, Felipe Neto, Júnior Viana e Gedalva Viana por todo o apoio e respeito. Espero levar suas amizades para a vida toda.

Ao meu querido amigo Ramon Viana, um irmão que a vida me deu. Obrigado por todo o seu amor, companheirismo, lealdade e por sempre acreditar em mim, mesmo quando eu mesmo não acreditava. Nós tivemos muitos momentos felizes na universidade e levamos essa amizade para toda a vida. Você fez coisas por mim que ninguém jamais faria. Você hoje ocupa um lugar imenso no meu coração. Eu sou eternamente grato por você ter entrado em minha vida. Você é um verdadeiro amigo.

Ao meu amigo Bruno Queiroz por todo seu apoio, por me salvar muitas vezes me conduzindo até à universidade quando eu não tinha um tostão no bolso e por todos os momentos de brincadeiras, de estudo e de risos. Agradeço por todos os cafés que tomávamos na cantina da Sâmia, por sempre me motivar e me ajudar. Você tem um coração enorme e por esta razão sou eternamente grato.

Ao grupo de percussão Acadêmicos da Casa Caiada, em especial à nossa querida Catherine Furtado, por conduzir tão bem este grupo e por ter me dado a oportunidade de fazer arte em nossa cidade e conhecer muitas pessoas novas, em especial, Sarah Lima e Sarah Santos. Vocês são mulheres incríveis e desejo toda a felicidade que este mundo possa oferecê-las. Espero que nossa amizade continue sempre forte.

A todos os funcionários do Departamento de Estatística e Matemática Aplicada, em especial aos professores que tive maior contato: Ana Maria, Carlos Diego, Gualberto Agamez, Jeniffer Duarte, João Welliandre, Juvêncio Nobre, Maurício Araújo e Ronald Targino.

E a todos que não mencionei e que me ajudaram (direta ou indiretamente) na realização deste sonho.

“Não há nada que possa me impedir de ser capaz ou ser forte o bastante. Errante é o passo que se limita a ser raso, exceto se for íngreme”

(Guilherme de Sá)

RESUMO

Os modelos lineares generalizados consideram em sua construção a família exponencial de dispersão, a qual tem como membros a distribuição gama e normal inversa. Estas distribuições permitem modelar dados que assumem valores nos reais positivos e com assimetria positiva. A família exponencial possui várias propriedades interessantes que são estendidas a todos os seus membros, o que facilita a obtenção e cálculo de certas quantidades. Neste trabalho é apresentada a metodologia dos Modelos Lineares Generalizados (MLG), desenvolvida por Nelder e Wedderburn (1972), tais como a definição, estimação, funções de ligação, seleção de modelos, análise de resíduos e diagnóstico em MLG's. Também será apresentada uma aplicação utilizando dados reais de limite de crédito, na tentativa de descrever o limite usado médio de clientes utilizando unicamente os modelos gama e normal inverso. Nos ajustes dos modelos e pelas análises de resíduos foi possível observar que para as distribuições gama e normal inversa, utilizando a função de ligação identidade, foram obtidos modelos com bons ajustes após a eliminação de observações destacadas nas análises de resíduos e diagnóstico.

Palavras-chave: Regressão. Modelos Lineares Generalizados. Dados Positivos Assimétricos. Limite de Crédito.

ABSTRACT

Generalized linear models consider in their construction the exponential family of dispersion, which has the gamma and inverse normal distribution as members. These distributions allow modeling data that assume positive real values and with positive asymmetry. The exponential family has several interesting properties that are extended to all its members, which makes it easier to obtain and calculate certain quantities. This work presents the methodology of Generalized Linear Models (GLM), developed by Nelder and Wedderburn (1972), such as definition, estimation, link functions, model selection, residual analysis and diagnosis in GLM's. An application using real credit limit data will also be presented, in an attempt to describe the average used limit of customers using only the gamma and inverse normal models. In the adjustment of the models and by the analysis of residues, it was possible to observe that for the gamma and inverse normal distributions, using the identity link function, models with good adjustments were obtained after eliminating the observations highlighted in the analysis of residues and diagnosis.

Keywords: Regression. Generalized Linear Models. Asymmetric Positive Data. Credit Limit.

LISTA DE FIGURAS

Figura 1 – Comportamento da função de densidade da distribuição gama para alguns valores de μ e ϕ	19
Figura 2 – Função de densidade da distribuição normal inversa para alguns valores dos parâmetros	21
Figura 3 – Boxplot e histograma da variável resposta <i>limite usado médio</i>	47
Figura 4 – Gráficos de dispersão entre algumas covariáveis <i>versus</i> a variável resposta <i>limite usado médio</i>	48
Figura 5 – Gráficos de diagnóstico referentes ao modelo gama com função de ligação identidade ajustado aos dados de limite de crédito médio.	51
Figura 6 – (a) Gráfico de índice <i>versus</i> medida h e (b) gráfico de índices <i>versus</i> l_{max} para a distribuição gama com função de ligação identidade.	52
Figura 7 – Gráfico normal de probabilidades referente ao modelo gama com função de ligação identidade ajustado aos dados de análise de limite de crédito.	52
Figura 8 – Gráfico normal de probabilidades referente ao modelo gama com função de ligação identidade retirando-se todas as observações de forma conjunta ajustado aos dados de análise de limite de crédito médio.	54
Figura 9 – Gráficos de diagnóstico referentes ao modelo normal inverso com função de ligação identidade ajustado aos dados de limite de crédito médio.	57
Figura 10 – (a) Gráfico de índice <i>versus</i> medida h e (b) gráfico de índices <i>versus</i> l_{max} para a distribuição normal inversa com função de ligação identidade.	58
Figura 11 – Gráfico normal de probabilidades referente ao modelo normal inverso com função de ligação identidade ajustado aos dados de análise de limite de crédito médio.	59
Figura 12 – Gráfico normal de probabilidades referente ao modelo normal inverso com função de ligação identidade retirando-se todas as observações de forma conjunta ajustado aos dados de análise de limite de crédito médio.	60

LISTA DE TABELAS

Tabela 1 – Características da distribuição gama	20
Tabela 2 – Características da distribuição normal inversa	21
Tabela 3 – Funções de ligação canônicas	23
Tabela 4 – Medidas descritivas para as variáveis dos dados de análise de limite de crédito.	47
Tabela 5 – Classificação de risco com relação ao <i>behavior score</i>	47
Tabela 6 – Critério de informação de Akaike (AIC) para os modelos finais utilizando a distribuição gama para diferentes funções de ligação.	49
Tabela 7 – Estimativas dos parâmetros referente ao modelo gama com função de ligação identidade ajustado aos dados de análise de limite usado médio.	49
Tabela 8 – Variação percentual e estimativas dos parâmetros utilizando o modelo gama com função de ligação identidade com a exclusão de observações.	53
Tabela 9 – Critério de informação de Akaike (AIC) para os modelos finais utilizando a distribuição normal inversa para diferentes funções de ligação.	56
Tabela 10 – Estimativas dos parâmetros referente ao modelo normal inverso com função de ligação identidade ajustado aos dados de análise de limite de crédito.	56
Tabela 11 – Valores das observações destacadas nos gráficos de diagnóstico para o modelo normal inverso com função de ligação identidade.	58
Tabela 12 – Variação percentual e estimativas dos parâmetros utilizando o modelo normal inverso com função de ligação identidade com a exclusão de observações.	60
Tabela 13 – Descrição de todas as variáveis	77
Tabela 14 – Variáveis excluídas do banco de dados, quantidade zeros, quantidade de valores faltantes e critério de retirada.	78
Tabela 15 – Variáveis preditoras utilizadas para a análise dos dados de limite de crédito médio.	79

SUMÁRIO

1	INTRODUÇÃO	15
2	FAMÍLIA EXPONENCIAL	17
3	MODELOS LINEARES GENERALIZADOS	22
3.1	Estimação	23
3.1.1	<i>Estimação de β</i>	25
3.2	Estimação para o modelo gama	26
3.3	Estimação para o modelo normal inverso	28
3.3.1	<i>Estimação de ϕ</i>	31
3.4	Função de desvio (<i>deviance</i>)	32
3.5	Teste de hipóteses	35
3.6	Seleção de modelos	37
3.7	Técnicas de Diagnóstico em MLGs	37
4	APLICAÇÃO	46
4.1	O conjunto de dados de análise de limite de crédito	46
4.2	Ajuste utilizando o modelo gama	48
4.3	Análise de resíduos e diagnóstico para o modelo gama	50
4.4	Ajuste utilizando o modelo normal inverso	55
4.5	Análise de resíduos e diagnóstico para o modelo normal inverso	56
5	CONCLUSÕES E TRABALHOS FUTUROS	63
	REFERÊNCIAS	64
	APÊNDICES	65
	APÊNDICE A – Função geradora de momentos e cumulantes e propriedades das distribuições gama e normal inversa	65
	APÊNDICE B – Função escore de β	72
	APÊNDICE C – Estimação dos parâmetros	75
	APÊNDICE D – Função digama	76
	APÊNDICE E – Variáveis do banco de dados de análise de limite de crédito	77

1 INTRODUÇÃO

Os modelos de regressão são utilizados para descrever a relação entre uma variável que pode ser denominada de resposta ou dependente, e uma ou mais variáveis explicativas ou preditoras. Um dos objetivos principais da análise de regressão, consiste em estimar ou prever o valor médio da variável resposta condicional aos valores das variáveis preditoras, i.e., dados alguns valores das variáveis explicativas temos um valor estimado para a variável resposta.

O modelo de regressão linear pode ser considerado o modelo mais popular e de simples utilização, sendo um caso particular de uma classe mais geral, que são os Modelos Lineares Generalizados (MLG's), propostos por Nelder e Wedderburn (1972). Estes modelos consistem em flexibilizar a distribuição que será utilizada para a variável resposta, desde que esta pertença à família exponencial de dispersão. Adicionalmente, os MLG's permitem ter uma maior flexibilização para a relação funcional entre a média da variável e o preditor linear dado que é introduzido a função de ligação como aponta Paula (2013).

Diversas distribuições tem sido propostas para modelar dados cuja a variável resposta Y tenha distribuição com suporte positivo e característica assimétrica como por exemplo as distribuições log-normal, Weibull, Pareto e Birnbaum-Saunders como aponta Paula (2013). No entanto, neste trabalho daremos uma atenção especial para as distribuições gama e normal inversa. A principal vantagem de utilizar tais modelos consiste principalmente em evitar a utilização de transformações na variável resposta e não ferir os pressupostos básicos do modelo de regressão linear normal. Conforme Paula (2013) as distribuições gama e normal inversa tem sido bastante utilizadas na análise de tempos de sobrevivência (ou duração) com forte ênfase nas áreas médica e de engenharia e também comuns em áreas do conhecimento como pesca, meteorologia, finanças, seguros e atuária.

Neste trabalho utilizaremos as distribuições gama e normal inverso na tentativa de descrever dados cuja a variável resposta apresente suporte nos números reais positivos e característica de assimetria positiva. Serão ajustados os modelos com base nas distribuições gama e normal inversa para alguns tipos de função de ligação, tais como a ligação identidade, recíproca, recíproca quadrática e logarítmica. As análises serão feitas no software R (R Core Team) (2021) que possui várias rotinas computacionais e modelos implementados para a análise de modelos, bem como ferramentas de visualização

e diagnóstico. Dentre os pacotes utilizados, destacamos o pacote **MASS**, **glm2**. Para a realização dos gráficos de resíduos e envelopes simulados, foram utilizadas rotinas computacionais encontradas em <https://www.ime.usp.br/~giapaula/textoregressao.htm> com adaptações feitas pelo próprio autor.

2 FAMÍLIA EXPONENCIAL

Segundo Cordeiro e Demétrio (2008), uma variável aleatória Y pertence à família exponencial uniparamétrica se sua função densidade de probabilidade (de massa de probabilidade) pode ser expressa como

$$f(y; \theta) = h(y) \exp[\eta(\theta)t(y) - b(\theta)], \quad (2.1)$$

em que $b(\cdot), \eta(\cdot), h(\cdot)$ e $t(\cdot) \in \mathbb{R}$. A família exponencial canônica é obtida através de (2.1) quando $\eta(\theta) = \theta$ e $t(y) = y$. Não existe uma forma única de representar a família exponencial uniparamétrica. Para mais detalhes sobre outras representações desta família, veja Bolfarine e Sandoval (2010) e Demétrio (2001), por exemplo.

Nelder e McCullagh (1989) introduziram na família exponencial canônica definida em (2.1) um novo parâmetro $\phi > 0$ ($\phi^{-1} > 0$) conhecido como parâmetro de *precisão* (dispersão). A incorporação desse parâmetro possibilita ampliar o leque de opções para a distribuição da variável resposta Y . Exemplos de distribuições para a variável resposta são: poisson, binomial, binomial negativa, normal, gama e normal inversa, supondo que essas últimas três possuam um dos parâmetros conhecido.

Com a introdução de ϕ obtemos a família dada por

$$f(y; \theta, \phi) = \exp \left[\phi(\theta y - b(\theta)) + c(y, \phi) \right], \quad (2.2)$$

que é conhecida como a **família exponencial de dispersão**, em que $\phi > 0$, $b(\cdot)$ e $c(\cdot) \in \mathbb{R}$, como apresentada em Paula (2013). Através de (2.2) é possível obter os seguintes resultados

$$\mu = \mathbb{E}[Y] = b'(\theta) \text{ e } \text{Var}[Y] = \phi^{-1}b''(\theta),$$

em que $b''(\theta) = V(\mu) = d\mu/d\theta$ é denotada como a **função de variância**. Segundo Paula (2013), a função de variância desempenha um papel de suma importância na família exponencial pois ela caracteriza a distribuição, ou seja, se conhecermos a função de variância conhecemos a classe de distribuições associadas e vice-versa.

Dado que o foco do trabalho é a utilização dos modelos gama e normal inverso, vamos supor inicialmente que Y tenha distribuição gama com parâmetros $\alpha > 0$ e $\beta > 0$,

em que α é o parâmetro de forma e β o parâmetro de dispersão, com função densidade de probabilidade dada por

$$f(y; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} \exp(-\beta y).$$

Pode-se fazer uma reparametrização considerando $\phi = \alpha$ e $\mu = \alpha/\beta$ de forma que obtemos

$$f(y; \mu, \phi) = \frac{1}{\Gamma(\phi)} \left(\frac{\phi}{\mu}\right)^\phi y^{\phi-1} \exp\left\{-\frac{\phi}{\mu} y\right\}, \quad \phi > 0, \mu > 0, y > 0.$$

Podemos organizar a equação acima de tal forma que

$$f(y; \mu, \phi) = \exp\left\{\phi\left(-\frac{y}{\mu} - \ln(\mu)\right) + (\phi - 1)\ln(y) + \phi\ln(\phi) - \ln(\Gamma(\phi))\right\}.$$

Deste modo, fazendo $\theta = -1/\mu$ têm-se $\mu = -1/\theta$, $b(\theta) = \ln(\mu) = \ln(-1/\theta) = -\ln(-\theta)$ e $c(y, \phi) = (\phi - 1)\ln(y) + \phi\ln(\phi) - \ln(\Gamma(\phi))$ temos

$$f(y; \theta, \phi) = \exp\left\{\phi\left(\theta y + \ln(-\theta)\right) + (\phi - 1)\ln(y) + \phi\ln(\phi) - \ln(\Gamma(\phi))\right\},$$

pertencendo assim à família exponencial canônica de dispersão. É fácil verificar para a distribuição gama que sua função de variância é dada por $V(\mu) = \mu^2$ que pode ser obtida através da segunda derivada de $b(\theta) = -\ln(-\theta)$.

Supondo agora que Y seja uma variável aleatória com f.d.p dada por uma normal inversa de média μ e parâmetro de precisão ϕ denotada por $NI(\mu, \phi)$, dada por:

$$f(y; \mu, \phi) = \frac{\phi^{1/2}}{\sqrt{2\pi y^3}} \exp\left\{-\frac{\phi(y - \mu)^2}{2\mu^2 y}\right\}, \quad y > 0, \mu > 0.$$

Pode-se obter a distribuição normal inversa na família exponencial como (ver Apêndice B) como

$$f(y; \mu, \phi) = \exp\left\{\phi\left(-\frac{y}{2\mu^2} + \frac{1}{\mu}\right) - \frac{1}{2}\left[\log\left(2\pi y^3/\phi\right) + \frac{\phi}{y}\right]\right\}.$$

Fazendo $\theta = -1/2\mu^2$, $b(\theta) = -1/\mu = -(-2\theta)^{1/2}$ e

$$c(y, \phi) = -\frac{1}{2} \log \left(2\pi y^3 / \phi \right) - \frac{\phi}{2y} = \frac{1}{2} \log \left(\phi / 2\pi y^3 \right) - \frac{\phi}{2y},$$

obtemos a forma canônica na família exponencial de dispersão. É simples verificar que a função de variância da distribuição normal inversa é dada por $V(\mu) = \mu^3$ tomando a segunda derivada de $b(\theta) = -(-2\theta)^{1/2}$.

Propriedades das distribuições gama e normal inversa

A seguir serão apresentadas algumas propriedades para as distribuições gama e normal inversa. A Figura 1 (a) apresenta diferentes formas da função de densidade da distribuição gama mantendo $\phi = 0,6$ fixo e variando $\mu = \{2, 3, 4, 5, 6\}$ e a Figura 1 (b) apresenta diferentes formas da função de densidade da distribuição normal inversa para $\mu = 1,5$ fixo, variando $\phi = \{2, 3, 4, 5, 6\}$. Na Figura 1 (a) observamos que para $0 < \phi < 1$ temos que a função densidade da gama possui uma *pole* na origem, isto é, a função densidade assume valor máximo na origem e decresce de forma monótona para $y \rightarrow \infty$. Na Figura 1 (b) observa-se que para $\phi > 1$ a densidade assume valor zero na origem, cresce até o seu ponto de máximo em $y = \mu - \mu/\phi$ e decresce de forma monótona para $y \rightarrow \infty$. Quando $\phi = 1$ temos o caso da distribuição exponencial.

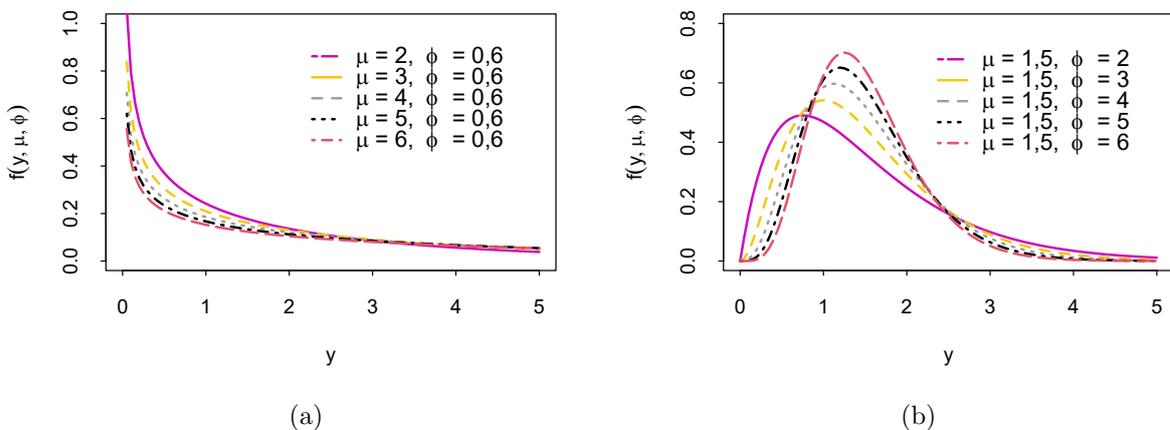


Figura 1 – Comportamento da função de densidade da distribuição gama para alguns valores de μ e ϕ .

Pode-se perceber que à medida em que ϕ aumenta (Figura 1 (b)) a distribuição de Y tende a ficar mais simétrica em torno de sua média. Paula (2013), afirma que

para $\phi \rightarrow \infty$ a distribuição gama converge para uma distribuição normal de média μ e variância $\mu^2\phi^{-1}$. Para mais detalhes sobre propriedades da distribuição gama ver por exemplo Cordeiro e Demétrio (2008), Turkman e Silva (2000) e Paula (2013). A Tabela 1 apresenta as principais características da distribuição gama (para detalhes da obtenção dos resultados ver Apêndice A).

Tabela 1 – Características da distribuição gama	
Características	
Esperança	μ
Variância	μ^2
Moda	$M_o(y) = \mu - \mu/\phi$
Função Geradora de Momentos	$M_Y(t) = \left(1 - \frac{t\mu}{\phi}\right)^{-\phi}, t < \phi/\mu$
Função Geradora de Cumulantes	$\mathcal{K}(t; \mu, \phi) = -\phi \log(1 - \mu t/\phi), t < \phi/\mu$
r -ésimo cumulante	$k_r = (r - 1)! \mu^r \phi^{1-r}$
Assimetria	$A(y) = 2/\sqrt{\phi}$
Curtose	$K(y) = 6/\phi - 3$

Para a distribuição normal inversa temos na Figura 2 o comportamento da função densidade para valores dos seus parâmetros. Na Figura 2 (a) pode-se perceber que à medida que o parâmetro ϕ aumenta, a distribuição normal inversa torna-se mais simétrica em torno da média $\mu = 5$. Nas Figuras 2 (b) e 2 (c) temos os comportamentos da função de densidade com $\phi = 25$ fixo e μ variando e $\mu = 1$ fixo e ϕ variando no intervalo (0,1). Na Figura 2 (d) temos o comportamento da função de densidade considerando fixo $\phi = 15$, variando os valores de μ . Quando $\mu = 1$ e $\phi = 1$ a f.d.p. tem forma padronizada. Para todos os gráficos percebemos que a f.d.p. cresce de forma monótona até atingir seu valor máximo (moda apresentada na Tabela 2) e decresce de forma monótona quando $y \rightarrow \infty$.

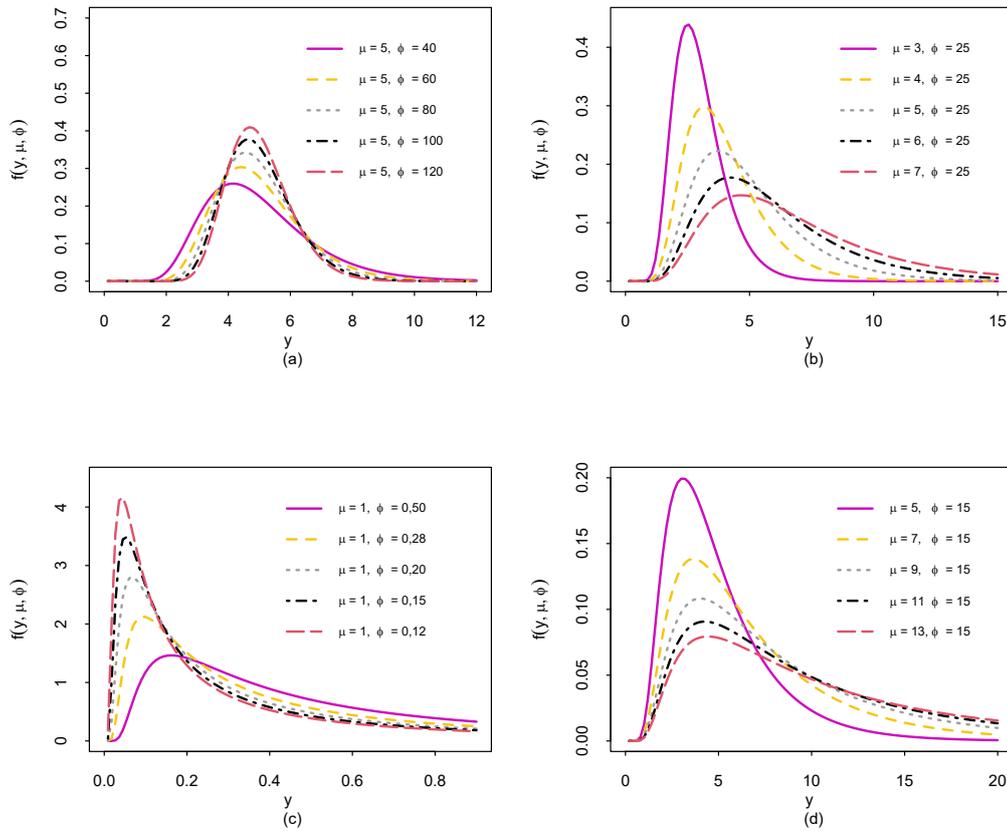


Figura 2 – Função de densidade da distribuição normal inversa para alguns valores dos parâmetros

A Tabela 2 apresenta as principais características da distribuição normal inversa (para detalhes da obtenção dos resultados ver Apêndice A).

Tabela 2 – Características da distribuição normal inversa

Características	
Esperança	μ
Variância	μ^3
Moda	$M_o(y) = \mu \left[\left(1 + \frac{9\mu^2}{4\phi^2} \right)^{1/2} - \frac{3\mu}{2\phi} \right]$
Função Geradora de Momentos	$M_Y(t) = \exp \left[\frac{\mu}{\phi} \left(1 - \sqrt{1 - \frac{2\mu^2 t}{\phi}} \right) \right], t < \phi/2\mu^2$
Função Geradora de Cumulantes	$\mathcal{K}(t; \theta, \phi) = \frac{\phi}{\mu} - \phi \left[\frac{1}{\mu^2} - \frac{2t}{\phi} \right]^{1/2}, t < \phi/2\mu^2$
Assimetria	$A(y) = 3 \left(\frac{\mu}{\phi} \right)^{1/2}$
Curtose	$K(y) = 15\mu/\phi$

3 MODELOS LINEARES GENERALIZADOS

Conforme Cordeiro e Demétrio (2008) os MLGs são construídos através de uma variável resposta univariada, variáveis explanatórias ou explicativas e uma amostra aleatória de n observações independentes de tal modo que:

- i) A variável resposta, conhecida como **componente aleatório** do modelo, tem distribuição pertencente à família exponencial de distribuições definida em (2.2) abrangendo diversas distribuições para variáveis resposta do tipo contagem, proporções ou contínuas;
- ii) As variáveis explicativas entram no modelo na forma de estrutura linear constituindo o **componente sistemático** ou **preditor linear** do modelo, isto é

$$\eta_i = \sum_{k=1}^p x_{ik}\beta_k = \mathbf{x}_i^\top \boldsymbol{\beta} \quad \text{ou} \quad \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta},$$

em que $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)^\top$ é a matriz de especificação do modelo de dimensão $n \times p$ com \mathbf{x}_i^\top para $i = 1, 2, \dots, n$ sendo a i -ésima coluna da matriz especificação, $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^\top$ o vetor de parâmetros de ordem p associado ao modelo e $\boldsymbol{\eta} = (\eta_1, \eta_2, \dots, \eta_n)^\top$ o preditor linear;

- iii) A **função de ligação** que faz a conexão entre os componentes aleatório e sistemático do modelo através de uma relação funcional adequada

$$g(\mu_i) = \eta_i,$$

sendo $g(\cdot)$ uma função monótona diferenciável.

Existem nos MLGs diversos tipos de ligação que podem ser utilizados para modelar a relação entre a variável resposta e seu preditor linear como por exemplo a ligação potência $\eta_i = \mu_i^\lambda$, para $\lambda \in \mathbb{R}$, com casos particulares como sendo a identidade, recíproca e raiz quadrada para $\lambda = 1, -1$ e $1/2$, respectivamente. Neste trabalho, serão consideradas para os modelos gama e normal inverso as funções de ligação identidade ($\eta_i = \mu_i$), recíproca ($\eta_i = 1/\mu_i$), recíproca quadrática ($\eta_i = 1/\mu_i^2$) e logarítmica ($\eta_i = \log(\mu_i)$). É importante ressaltar que nos MLGs, não se define uma distribuição para o erro como no modelo de regressão normal linear, mas sim para a variável resposta Y (Cordeiro e Demétrio, 2008).

Função de ligação canônica

Quando estamos modelando a relação entre a média de Y e o preditor linear, em geral, é interessante que se escolha ligações do tipo $g(\mu_i) = \theta_i = \eta_i$, o preditor linear modela o parâmetro θ_i de forma direta. Em MLGs denotamos estes tipos de ligação como **ligações canônicas** e elas desempenham um papel bastante importante, como por exemplo, a facilidade de interpretação prática dos parâmetros e vantagens referente à simplificação do algoritmo de estimação do vetor de parâmetros $\boldsymbol{\beta}$ como apontam Cordeiro e Demétrio (2008). A Tabela 3 apresenta as ligações canônicas para as distribuições gama e normal inversa, que são as distribuições que possuem as características estudadas neste texto.

Distribuição	Função de ligação canônica
Gama	Recíproca: $\eta = 1/\mu$
Normal Inversa	Recíproca quadrática: $\eta = -1/\mu^2$

A seguir serão apresentados os resultados com relação ao processo de estimação do vetor de parâmetros considerando as funções de ligação mencionadas para os modelos gama e normal inverso.

3.1 Estimação

Seja Y uma variável aleatória unidimensional pertencente à família exponencial canônica de dispersão apresentada em (2.2) e seja Y_1, Y_2, \dots, Y_n uma amostra aleatória de Y com $\mathbf{y} = (y_1, y_2, \dots, y_n)^\top$ o vetor de valores observados da variável resposta. A função de verossimilhança de $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \phi)^\top$ é dada por

$$\begin{aligned} L(\boldsymbol{\theta}, \mathbf{y}) &= \prod_{i=1}^n \exp[\phi(\theta_i y_i - b(\theta_i)) + c(y_i, \phi)] \\ &= \exp \left[\phi \sum_{i=1}^n (y_i \theta_i - b(\theta_i)) + \sum_{i=1}^n c(y_i, \phi) \right], \end{aligned}$$

e sendo assim a função de log-verossimilhança é dada por

$$l(\boldsymbol{\theta}, \mathbf{y}) = \phi \sum_{i=1}^n (y_i \theta_i - b(\theta_i)) + \sum_{i=1}^n c(y_i, \phi). \quad (3.1)$$

Escore e Informação de Fisher de β

Para obter a função escore basta tomar a derivada da função de log-verossimilhança de θ em relação ao parâmetro β . Através de alguns procedimentos algébricos (ver Apêndice B) obtemos a função escore de β como

$$\mathbf{U}_\beta(\theta) = \phi \mathbf{X}^\top \mathbf{W}^{1/2} \mathbf{V}^{-1/2} (\mathbf{y} - \boldsymbol{\mu}), \quad (3.2)$$

em que \mathbf{X} é a matriz de especificação do modelo de dimensão $n \times p$ cujas linhas são dadas por \mathbf{x}_i^\top para $i = 1, 2, \dots, n$, $\mathbf{W} = \text{diag}\{w_1, w_2, \dots, w_n\}$ a matriz diagonal denominada de matriz de pesos, $\mathbf{V} = \text{diag}\{V_1, V_2, \dots, V_n\}$, $\mathbf{y} = (y_1, y_2, \dots, y_n)^\top$ e $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_n)^\top$. Os elementos de \mathbf{W} são

$$w_i = \left(\frac{d\mu_i}{d\eta_i} \right)^2 \frac{1}{V_i},$$

em que $V_i = V(\mu_i)$.

Para obter a informação de Fisher para β temos que

$$-\mathbb{E} \left[\frac{\partial^2 l(\theta, \mathbf{y})}{\partial \beta_j \partial \beta_l} \right] = \phi \sum_{i=1}^n w_i x_{ij} x_{il},$$

com forma matricial dada por

$$\mathbf{K}_{\beta\beta}(\theta) = \phi \mathbf{X}^\top \mathbf{W} \mathbf{X}.$$

Escore e Informação de Fisher de ϕ

Para a função escore de ϕ partimos diretamente da equação (3.1) e obtemos

$$\mathbf{U}_\phi(\theta) = \frac{\partial l(\theta, \mathbf{y})}{\partial \phi} = \sum_{i=1}^n (y_i \theta_i - b(\theta_i)) + \sum_{i=1}^n c'(y_i, \phi). \quad (3.3)$$

Para a matriz de informação de Fisher, tomamos a segunda derivada como sendo

$$\frac{\partial^2 l(\theta, \mathbf{y})}{\partial \phi^2} = \sum_{i=1}^n c''(y_i, \phi),$$

então

$$\mathbf{K}_{\phi\phi}(\boldsymbol{\theta}) = -\mathbb{E}\left[\sum_{i=1}^n c''(y_i, \phi)\right] = \sum_{i=1}^n -\mathbb{E}[c''(Y_i, \phi)]. \quad (3.4)$$

3.1.1 Estimação de β

Para se obter o vetor de parâmetros estimado de um MLG, utiliza-se o método da máxima verossimilhança denotado por $\hat{\beta}$, obtido ao resolver $\mathbf{U}_{\beta}(\boldsymbol{\theta}) = \mathbf{0}$. De forma geral esse sistema de equações não pode ser resolvido de forma analítica e para isso recorrem-se aos métodos iterativos. Um dos métodos mais utilizados para resolver este tipo de sistema de equações é o famoso método de Newton-Rhapson, dado que tem como objetivo estimar as raízes de uma função $f(x)$.

Expandindo $\mathbf{U}_{\beta}(\boldsymbol{\theta})$ até primeira ordem em torno de $\beta^{(0)}$, sendo dada por

$$\mathbf{U}_{\beta} \cong \mathbf{U}_{\beta}^{(0)} + \mathbf{U}'_{\beta}{}^{(0)}(\beta - \beta^{(0)}),$$

em que $\mathbf{U}_{\beta}^{(0)}$ e $\mathbf{U}'_{\beta}{}^{(0)}$ representam \mathbf{U}_{β} e \mathbf{U}'_{β} avaliados em $\beta^{(0)}$, em que \mathbf{U}'_{β} é a derivada da função escore de β em relação ao vetor de parâmetros dada por $-\partial^2 l(\boldsymbol{\theta}, \mathbf{y})/\partial\beta_j\partial\beta_l$. Resolvendo $\mathbf{U}_{\beta} = \mathbf{0}$ temos

$$\beta = \beta^{(0)} + \left[-\mathbf{U}'_{\beta}{}^{-1}\right]^{(0)} \mathbf{U}_{\beta}^{(0)}.$$

De forma geral temos

$$\beta^{(m+1)} = \beta^{(m)} + \left[-\mathbf{U}'_{\beta}{}^{-1}\right]^{(m)} \mathbf{U}_{\beta}^{(m)}.$$

No entanto, a matriz $-\mathbf{U}'_{\beta}$ pode não ser positiva definida, i.e., não possuir inversa usual, e neste caso, a alternativa é substituí-la por $\mathbf{K}_{\beta\beta}$ (matriz de informação esperada de Fisher de β) dada por $\mathbf{K}_{\beta\beta} = \phi\mathbf{X}^{\top}\mathbf{W}\mathbf{X}$. Assim, o processo iterativo escore de Fisher é dado por

$$\beta^{(m+1)} = \beta^{(m)} + \left[\mathbf{K}_{\beta\beta}^{-1}\right]^{(m)} \mathbf{U}_{\beta}^{(m)}. \quad (3.5)$$

sendo $\boldsymbol{\beta}^{(m)}$ e $\boldsymbol{\beta}^{(m+1)}$ os vetores de parâmetros estimados nos passos m e $m + 1$ respectivamente, $\mathbf{U}_{\boldsymbol{\beta}}^{(m)}$ o vetor score avaliado no passo m e $[\mathbf{K}_{\boldsymbol{\beta}\boldsymbol{\beta}}^{-1}]^{(m)}$ é a inversa da matriz de informação esperada de Fisher de $\boldsymbol{\beta}$ avaliada no passo m . Através de algum algebrismo (ver Apêndice C), podemos obter a expressão do estimador de máxima verossimilhança de $\boldsymbol{\beta}$:

$$\boldsymbol{\beta}^{(m+1)} = \left(\mathbf{X}^\top \mathbf{W}^{(m)} \mathbf{X} \right)^{-1} \mathbf{X}^\top \mathbf{W}^{(m)} \mathbf{z}^{(m)}, \quad (3.6)$$

em que $\mathbf{z}^{(m)} = \boldsymbol{\eta}^{(m)} + \left[\mathbf{W}^{-\frac{1}{2}} \right]^{(m)} \left[\mathbf{V}^{-\frac{1}{2}} \right]^{(m)} (\mathbf{y} - \boldsymbol{\mu}^{(m)})$. Existem alguns critérios de parada para o processo iterativo de estimação do vetor de parâmetros, um deles é

$$\sum_{r=1}^p \left(\frac{\beta_r^{(m+1)} - \beta_r^{(m)}}{\beta_r^{(m)}} \right)^2 < \xi,$$

sendo ξ um valor suficientemente pequeno, e β_r representa o r -ésimo parâmetro avaliado nos passos m e $m + 1$ do processo iterativo.

A quantidade \mathbf{z} faz o papel de uma variável modificada, enquanto \mathbf{W} é uma matriz de pesos que se altera a cada passo do processo iterativo. A equação (3.6), segundo Paula (2013), é válida para qualquer MLG. Pode-se notar que em (3.6), para estimar $\boldsymbol{\beta}$, não é necessário conhecer ϕ pois o processo de estimação não depende desse parâmetro. Sobre o número de iterações necessárias para atingir a convergência de (3.6), ocorrerá geralmente para um número finito de passos, independente dos valores iniciais utilizados, como afirma Paula (2013).

3.2 Estimação para o modelo gama

Logarítmica

Para a função de ligação logarítmica temos que $\log(\mu_i) = \eta_i \implies \mu_i = \exp(\eta_i) = \exp(\mathbf{x}_i^\top \boldsymbol{\beta})$. Sabemos que para a distribuição gama a função de variância é dada por $V(\mu_i) = \mu_i^2$, e deste modo temos que o i -ésimo peso será dado por

$$w_i = \left(\frac{d\mu_i}{d\eta_i} \right)^2 \frac{1}{V_i} = \frac{\mu_i^2}{\mu_i^2} = 1.$$

Portanto, a matriz de pesos \mathbf{W} será dada pela matriz identidade de ordem n , $\mathbf{W}^{-\frac{1}{2}} = \mathbf{W}$. Também temos que $\mathbf{V} = \text{diag}\{\exp(\eta_1)^2, \dots, \exp(\eta_n)^2\}$ e $\mathbf{V}^{-\frac{1}{2}} = \text{diag}\{\exp(\eta_1)^{-1}, \dots, \exp(\eta_n)^{-1}\}$. Assim a variável modificada \mathbf{z} será

$$\mathbf{z}^{(m)} = [\log(\boldsymbol{\mu})]^{(m)} + [\mathbf{V}^{-\frac{1}{2}}]^{(m)}(\mathbf{y} - \exp(\boldsymbol{\mu})^{(m)}).$$

O processo iterativo tem forma reduzida dada por

$$\boldsymbol{\beta}^{(m+1)} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{z}^{(m)}, \quad m = 0, 1, 2, \dots,$$

De acordo com Paula (2013), se as colunas de \mathbf{X} forem ortogonais, isto é, $\mathbf{X}^\top \mathbf{X} = \mathbf{I}_p$, em que \mathbf{I}_p é a matriz identidade de ordem p , então temos que

$$\text{Var}[\hat{\beta}_j] = \phi^{-1} \quad \text{e} \quad \text{Cov}[\hat{\beta}_j, \hat{\beta}_l] = 0, \quad (j \neq l),$$

isto é, $\hat{\beta}_j$ e $\hat{\beta}_l$ são assintoticamente independentes e assim a função de ligação logarítmica fornece uma maneira de desenvolver experimentos ortogonais. Ainda segundo Paula (2013) é possível escolher formas convenientes para a matriz \mathbf{X} de modo que $\mathbf{X}^\top \mathbf{X} = \mathbf{I}_p$ e dessa forma obter-se estimativas mutuamente independentes e de variância constante para os coeficientes do preditor linear $\boldsymbol{\eta}$.

Identidade

Para a função de ligação identidade temos que $\mu_i = \eta_i$. Deste modo o i -ésimo peso será dado por

$$w_i = \frac{1}{V_i} = \frac{1}{\mu_i^2} = \eta_i^{-2}.$$

Deste modo as matrizes \mathbf{W} e \mathbf{V} serão dadas respectivamente por $\mathbf{W} = \text{diag}\{\eta_1^{-2}, \eta_2^{-2}, \dots, \eta_n^{-2}\}$ e $\mathbf{V} = \text{diag}\{\eta_1^2, \eta_2^2, \dots, \eta_n^2\}$. Assim temos que

$$\mathbf{W}^{-\frac{1}{2}} = \text{diag}\{\eta_1, \eta_2, \dots, \eta_n\}$$

e

$$\mathbf{V}^{-\frac{1}{2}} = \text{diag}\{\eta_1^{-1}, \eta_2^{-1}, \dots, \eta_n^{-1}\},$$

de tal modo que $\mathbf{W}^{-\frac{1}{2}}\mathbf{V}^{-\frac{1}{2}} = \mathbf{I}_n$. Portanto a variável modificada será $\mathbf{z}^{(m)} = \boldsymbol{\mu}^{(m)} + \mathbf{y} - \boldsymbol{\mu}^{(m)} = \mathbf{y}$. Deste modo o processo iterativo para a estimação do vetor de parâmetros será dado de forma reduzida como sendo

$$\boldsymbol{\beta}^{(m+1)} = (\mathbf{X}^\top \mathbf{W}^{(m)} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}^{(m)} \mathbf{y}.$$

Recíproca

Para a função de ligação recíproca (ligação canônica) temos que $\mu_i = \eta_i^{-1}$. Dado que para o modelo gama $V_i = V(\mu_i) = d\mu/d\eta_i = \mu_i^2$, teremos que $d\mu_i/d\eta_i = -\eta_i^{-2}$ e deste modo

$$\left(\frac{d\mu_i}{d\eta_i} \right)^2 = \eta_i^{-4}.$$

Desta maneira, temos que o i -ésimo peso será dado por

$$w_i = \frac{\eta_i^{-4}}{\eta_i^{-2}} = \eta_i^{-2}.$$

A matriz de pesos será dada por $\mathbf{W} = \text{diag}\{\eta_1^{-2}, \eta_2^{-2}, \dots, \eta_n^{-2}\}$ e $\mathbf{V} = \mathbf{V}$. Teremos que $\mathbf{W}^{-\frac{1}{2}} = \mathbf{V}^{-\frac{1}{2}}$ com $\mathbf{W}^{-\frac{1}{2}} = \text{diag}\{\eta_1^{-1}, \eta_2^{-1}, \dots, \eta_n^{-1}\}$ de tal modo que $\mathbf{W}^{-\frac{1}{2}}\mathbf{V}^{-\frac{1}{2}} = \mathbf{W}$. Portanto, o processo iterativo para a estimação do vetor de parâmetros será dado por (3.6) com variável modificada

$$\mathbf{z}^{(m)} = [\boldsymbol{\mu}^{-1}]^{(m)} + \mathbf{W}^{(m)}(\mathbf{y} - \boldsymbol{\mu}^{(m)}).$$

3.3 Estimação para o modelo normal inverso

Logarítmica

Para a função de ligação logarítmica temos que $\mu_i = \exp(\eta_i)$ e a função de variância é dada por $V(\mu_i) = \mu_i^3$. Deste modo os pesos serão dados por

$$w_i = \left(\frac{d\mu_i}{d\eta_i} \right)^2 \frac{1}{V_i} = \frac{\mu_i^2}{\mu_i^3} = \mu_i^{-1}.$$

Assim, as matrizes \mathbf{W} e \mathbf{V} serão dadas respectivamente por:

$$\mathbf{W} = \text{diag}\{\exp(\eta_1)^{-1}, \exp(\eta_2)^{-1}, \dots, \exp(\eta_n)^{-1}\}$$

e

$$\mathbf{V} = \text{diag}\{(\exp(\eta_1))^3, (\exp(\eta_2))^3, \dots, (\exp(\eta_n))^3\}.$$

Com estes resultados, obtemos a variável modificada

$$\mathbf{z}^{(m)} = \log(\boldsymbol{\mu})^{(m)} + [\mathbf{W}^{-1/2}]^{(m)}[\mathbf{V}^{-1/2}]^{(m)}(\mathbf{y} - \exp(\boldsymbol{\mu})^{(m)}),$$

e segue o processo de estimação do vetor de parâmetros dado em (3.6).

Identidade

Para a função de ligação temos que $\eta_i = \mu_i$ e a função de variância é dada por $V(\mu_i) = \mu_i^3$. Desta maneira os pesos serão dados por

$$w_i = \left(\frac{d\mu_i}{d\eta_i} \right) \frac{1}{V_i} = \frac{1}{\mu_i^3}.$$

Assim, as matrizes \mathbf{W} e \mathbf{V} serão dadas respectivamente por:

$$\mathbf{W} = \text{diag}\{\eta_1^{-3}, \dots, \eta_n^{-3}\}$$

e

$$\mathbf{V} = \text{diag}\{\eta_1^3, \dots, \eta_n^3\}$$

. Com isso temos

$$\mathbf{W}^{-1/2} = \text{diag}\{\eta_1^{3/2}, \dots, \eta_n^{3/2}\}$$

e

$$\mathbf{V}^{-1/2} = \text{diag}\{\eta_1^{-3/2}, \dots, \eta_n^{-3/2}\}$$

de tal modo que $\mathbf{W}^{-1/2}\mathbf{V}^{-1/2} = \mathbf{I}_n$. Assim, temos que a variável modificada será dada por $\mathbf{z}^{(m)} = \mathbf{y}$, seguindo-se com o processo de estimação do vetor de parâmetros do modelo como definido em (3.6).

Recíproca

Para esta função de ligação, temos que $\mu_i = 1/\eta_i$. Então $\eta_i = 1/\mu_i$. Sabendo que $V(\mu_i) = \mu_i^3$ temos que os pesos serão dados por

$$w_i = (-\eta_i^{-2})^2/(\eta_i^{-1})^3 = \eta_i^{-4}/\eta_i^{-3} = \eta_i^{-1} = \mu_i.$$

Deste modo temos que

$$\mathbf{W} = \text{diag}\{\eta_1^{-1}, \eta_2^{-1}, \dots, \eta_n^{-1}\} = \text{diag}\{\mu_1, \mu_2, \dots, \mu_n\}$$

e

$$\mathbf{V} = \text{diag}\{\eta_1^{-3}, \eta_2^{-3}, \dots, \eta_n^{-3}\}.$$

Assim, temos que $\mathbf{W}^{-\frac{1}{2}} = \text{diag}\{\eta_1^{1/2}, \eta_2^{1/2}, \dots, \eta_n^{1/2}\}$ e $\mathbf{V}^{-\frac{1}{2}} = \text{diag}\{\eta_1^{3/2}, \eta_2^{3/2}, \dots, \eta_n^{3/2}\}$. Pode-se definir uma matriz $\mathbf{G} = \mathbf{W}^{-\frac{1}{2}}\mathbf{V}^{-\frac{1}{2}} = \text{diag}\{\eta_1^2, \eta_2^2, \dots, \eta_n^2\}$. Sendo assim, a variável modificada será dada por

$$\mathbf{z}^{(m)} = (\boldsymbol{\mu}^{-1})^{(m)} + \mathbf{G}^{(m)}(\mathbf{y} - \boldsymbol{\mu}^{(m)}),$$

seguindo-se com o processo iterativo para estimação do vetor de parâmetros do modelo como definido em (3.6).

Recíproca quadrática

Para esta função de ligação (ligação canônica) temos que $\eta_i = -1/2\mu_i^2$. Então $\mu_i = (-1/2\eta_i)^{1/2}$. Sabendo que $V(\mu_i) = \mu_i^3$, temos que

$$\left(\frac{d\mu_i}{d\eta_i}\right)^2 = \left\{ \left(-\frac{1}{2}\right)^{3/2} \eta_i^{-3/2} \right\}^2 = \left(-\frac{1}{2}\right)^3 \eta_i^{-3}.$$

Assim, os pesos serão dados por

$$\begin{aligned} w_i &= \left(-\frac{1}{2}\right)^3 \eta_i^{-3} \left(-\frac{1}{2}\right)^{-3/2} \eta_i^{3/2} \\ w_i &= \left(-\frac{1}{2}\right)^{3/2} (\eta_i^{3/2})^{-1} \\ w_i &= \left[\left(-\frac{1}{2\eta_i}\right)^{1/2}\right]^3 = \mu_i^3 = V_i. \end{aligned}$$

Portanto, temos $\mathbf{W} = \text{diag}\{\mu_1^3, \mu_2^3, \dots, \mu_n^3\}$, $\mathbf{V} = \mathbf{W}$ e $\mathbf{W}^{-1/2} = \text{diag}\{\mu_1^{-3/2}, \mu_2^{-3/2}, \dots, \mu_n^{-3/2}\}$.

3.3.1 Estimação de ϕ

O processo para estimar o parâmetro ϕ é realizado igualando a função escore de ϕ definida em (3.3) a zero, obtendo o seguinte resultado:

$$\sum_{i=1}^n c'(y_i, \hat{\phi}) = -\sum_{i=1}^n \phi\{y_i \hat{\theta}_i - b(\hat{\theta}_i)\}.$$

Somando e subtraindo $\sum_{i=1}^n \{y_i \tilde{\theta}_i - b(\tilde{\theta}_i)\}$ no lado direito da igualdade acima, obtemos

$$\sum_{i=1}^n c'(y_i, \hat{\phi}) = \sum_{i=1}^n \phi\{y_i \hat{\theta}_i - b(\hat{\theta}_i)\} + \sum_{i=1}^n \{y_i \tilde{\theta}_i - b(\tilde{\theta}_i)\} - \sum_{i=1}^n \{y_i \tilde{\theta}_i - b(\tilde{\theta}_i)\}.$$

Organizando os termos dos somatórios, reescrevemos a equação acima como

$$\sum_{i=1}^n c'(y_i, \hat{\phi}) = \sum_{i=1}^n \{y_i(\tilde{\theta}_i - \hat{\theta}_i) + b(\hat{\theta}_i) - b(\tilde{\theta}_i)\} - \sum_{i=1}^n \{y_i \tilde{\theta}_i - b(\tilde{\theta}_i)\},$$

Finalmente, multiplicando e dividindo o primeiro somatório do lado direito da igualdade acima por 2, temos

$$\sum_{i=1}^n c'(y_i, \hat{\phi}) = \frac{1}{2} D(\mathbf{y}; \hat{\boldsymbol{\mu}}) - \sum_{i=1}^n \{y_i \tilde{\theta}_i - b(\tilde{\theta}_i)\}.$$

em que $D(\mathbf{y}; \hat{\boldsymbol{\mu}})$ denota o desvio do modelo sob investigação, apresentado em (3.9). Com este resultado, pode-se mostrar que a estimativa de máxima verossimilhança de ϕ para os casos das distribuições normal e normal inversa, é dada por

$$\hat{\phi} = \frac{n}{D(\mathbf{y}; \hat{\boldsymbol{\mu}})},$$

e para o caso da distribuição gama tem-se que

$$2n\{\log(\hat{\phi}) - \psi(\hat{\phi})\} = D(\mathbf{y}; \hat{\boldsymbol{\mu}}),$$

em que $\Psi(\phi) = \Gamma'(\phi)/\Gamma(\phi)$ é a função digama (ver Apêndice D). Paula (2013) apresenta um estimador para ϕ sendo este um estimador consistente (de momentos) que não envolve processos iterativos e que é baseado na estatística de Pearson, válida para modelos pertencentes à família definida em (2.2) dado por

$$\hat{\phi} = \frac{(n-p)}{\sum_{i=1}^n \left\{ \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)} \right\}}, \quad (3.7)$$

em que $V(\hat{\mu}_i)$ é a função de variância estimada sob o modelo que está sendo ajustado.

3.4 Função de desvio (*deviance*)

Segundo Cordeiro e Demétrio (2008), um número muito grande de variáveis explicativas podem levar a um modelo bem ajustado mas com complexidade de interpretação e de outro modo, um modelo com poucas variáveis explicativas pode ser mais atrativo quanto à interpretação porém com ajuste não tão satisfatório. O principal objetivo é conseguir ajustar um modelo parcimonioso, ou seja, com o menor número de parâmetros e que tenha fácil interpretabilidade.

A qualidade do ajuste de um MLG pode ser avaliada através da função desvio que por sua vez mede a distância entre o modelo saturado (número de parâmetros igual ao número de observações, ou $p = n$) e o modelo reduzido (número de parâmetros menor que o número de observações, ou $p < n$), que, segundo Paula (2013), pode ser definida como

$$D^*(\boldsymbol{\mu}, \mathbf{y}) = \phi D(\mathbf{y}, \hat{\boldsymbol{\mu}}) = 2\{l(\mathbf{y}; \mathbf{y}) - l(\hat{\boldsymbol{\mu}}; \mathbf{y})\}, \quad (3.8)$$

em que $l(\mathbf{y}; \mathbf{y})$ é o logaritmo da função de verossimilhança do modelo saturado e $l(\hat{\boldsymbol{\mu}}; \mathbf{y})$ é o estimador de máxima verossimilhança do modelo reduzido. Denotando por $\hat{\theta}_i = \theta_i(\hat{\mu}_i)$

e $\tilde{\theta}_i = \theta_i(\tilde{\mu}_i)$ as estimativas de máxima verossimilhança de θ para o modelo reduzido e saturado respectivamente, então uma forma alternativa para a função desvio pode ser dada por

$$D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2 \sum_{i=1}^n \left[y_i(\tilde{\theta}_i - \hat{\theta}_i) + b(\hat{\theta}_i) - b(\tilde{\theta}_i) \right]. \quad (3.9)$$

Para a distribuição gama, temos que $\tilde{\theta}_i = -1/y_i$, $\hat{\theta}_i = -1/\hat{\mu}_i$, $b(\tilde{\theta}_i) = \log(y_i)$ e $b(\hat{\theta}_i) = \log(\hat{\mu}_i)$ e deste modo temos

$$D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2 \sum_{i=1}^n \left[y_i \left(-\frac{1}{y_i} + \frac{1}{\hat{\mu}_i} \right) + \log(\hat{\mu}_i) - \log(y_i) \right] = 2 \sum_{i=1}^n \left[\log \left(\frac{\hat{\mu}_i}{y_i} \right) + \frac{y_i - \hat{\mu}_i}{\hat{\mu}_i} \right].$$

Para a distribuição normal inversa temos $\tilde{\theta}_i = -1/2y_i$, $\hat{\theta}_i = -1/2\hat{\mu}_i^2$, $b(\tilde{\theta}_i) = -1/y_i$ e $b(\hat{\theta}_i) = -1/\hat{\mu}_i$. Com isso temos

$$\begin{aligned} D(\mathbf{y}; \hat{\boldsymbol{\mu}}) &= 2 \sum_{i=1}^n \left[y_i \left(-\frac{1}{2y_i^2} + \frac{1}{2\hat{\mu}_i^2} \right) - \frac{1}{\hat{\mu}_i} + \frac{1}{y_i} \right] \\ &= 2 \sum_{i=1}^n \left[\left(\frac{1}{y_i} - \frac{1}{2y_i} \right) + \left(\frac{y_i}{2\hat{\mu}_i^2} - \frac{1}{\hat{\mu}_i} \right) \right] \\ &= 2 \sum_{i=1}^n \left[\frac{1}{2y_i} + \frac{y_i - 2\hat{\mu}_i}{2\hat{\mu}_i^2} \right] \\ &= 2 \sum_{i=1}^n \left[\frac{2\hat{\mu}_i^2 + 2y_i^2 - 4y_i\hat{\mu}_i}{4y_i\hat{\mu}_i} \right] \\ &= \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{y_i\hat{\mu}_i^2}. \end{aligned}$$

Cordeiro e Demétrio (2008) destacam que uma maneira de conseguir diminuir o desvio seria aumentar o número de parâmetros, porém, com isto haveria também o aumento da complexidade de interpretação do modelo. Na prática o que se faz é testar um MLG comparando-se $\phi D(\mathbf{y}; \hat{\boldsymbol{\mu}})$ com os percentis de uma χ_{n-p}^2 (distribuição qui-quadrado com $n - p$ graus de liberdade) ou seja, se

$$\phi D(\mathbf{y}; \hat{\boldsymbol{\mu}}) \leq \chi_{n-p; \alpha}^2,$$

isto é, se $\phi D(\mathbf{y}; \hat{\boldsymbol{\mu}})$ for menor ou igual ao quantil $1 - \alpha$ da distribuição χ_{n-p}^2 , considera-se ao nível de $100\alpha\%$ de significância que o modelo proposto está bem ajustado ao dados, ou

se $D(\mathbf{y}; \hat{\boldsymbol{\mu}})$ for próximo do valor esperado $n - p$ da distribuição χ_{n-p}^2 . Pode-se observar que este critério necessita de uma estimativa consistente de ϕ , a qual é dada (3.7)

Resultados assintóticos

Conforme Paula (2013), embora seja de costume comparar os valores observados da função de desvio com os percentis da distribuição qui-quadrado com $n - p$ graus de liberdade, de forma geral, $D(\mathbf{y}; \hat{\boldsymbol{\mu}})$ não possui distribuição assintótica χ_{n-p}^2 . Jorgensen (1987), apresenta a função de desvio nos casos em que o parâmetro de dispersão ϕ é desconhecido como sendo

$$D^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) \sim \chi_{n-p}^2, \quad \text{quando } \phi \rightarrow \infty.$$

Para o caso da distribuição gama, este resultado diz que à medida em que o coeficiente de variação ficar próximo de zero o desvio será bem aproximado por uma distribuição qui-quadrado com $n - p$ graus de liberdade.

Qualidade do ajuste no modelo gama

Para a função de ligação logarítmica pode ser mostrado que para η_i tendo um intercepto, a soma $\sum_{i=1}^n (y_i - \hat{\mu}_i)/\hat{\mu}_i = 0$ e o desvio fica dado apenas por

$$D^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2\phi \sum_{i=1}^n \log(\hat{\mu}_i/y_i).$$

O parâmetro ϕ é desconhecido e deve-se estimá-lo. Um dos métodos de estimação de ϕ é feito através do método de máxima verossimilhança. Supondo que o modelo está bem especificado, para ϕ suficientemente grande, o desvio $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}})$ segue distribuição qui-quadrado com $(n - p)$ graus de liberdade (Paula, 2013). Valores altos para o desvio indicam inadequação do modelo ou falta de ajuste. Se pelo menos uma das observações de y_i for igual a zero, o desvio apresentado será inapropriado para o caso da função de ligação logarítmica no modelo gama. Para contornar esse problema, Nelder e McCullagh (1989) propõem substituir o desvio por

$$D^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2\phi \left\{ C(\mathbf{y}) + \sum_{i=1}^n \log(\hat{\mu}_i) + \sum_{i=1}^n \frac{y_i}{\hat{\mu}_i} \right\},$$

em que $C(\mathbf{y})$ é uma função arbitrária, porém limitada como por exemplo $C(\mathbf{y}) = \sum_{i=1}^n y_i / (1 + y_i)$. Se a parte sistemática do modelo possuir um intercepto, o desvio fica dado por

$$D^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2\phi \left\{ n + C(\mathbf{y}) + \sum_{i=1}^n \log(\hat{\mu}_i) \right\}.$$

Existem na literatura, diversas aproximações para a estimativa de ϕ que podem ser encontrados na série de artigos de Bain e Engelhardt (1975).

Qualidade do ajuste no modelo normal inverso

Para avaliar a qualidade do ajuste utilizamos a função desvio $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) = \phi D(\mathbf{y}; \hat{\boldsymbol{\mu}})$ em que

$$D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{y_i \hat{\mu}_i^2},$$

com $\hat{\mu}_i = g^{-1}(\hat{\eta}_i) = g^{-1}(\mathbf{x}_i^\top \hat{\boldsymbol{\beta}})$. Como ϕ é desconhecido, deve-se estimá-lo por meio do método da máxima verossimilhança, cuja a solução é dada por

$$\hat{\phi} = \frac{n}{D(\mathbf{y}; \hat{\boldsymbol{\mu}})}.$$

Conforme Paula (2013), com a suposição de que o modelo esteja bem especificado, para ϕ grande, o desvio $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}})$ segue distribuição qui-quadrado com $(n - p)$ graus de liberdade. Para avaliar a inadequação do modelo ou falta de ajuste, deve-se observar valores altos do desvio.

3.5 Teste de hipóteses

Hipóteses simples

Os testes de hipóteses são baseados em algumas estatísticas, dentre elas, destacamos a **Razão de Verossimilhanças** e **Teste F**. Paula (2013) apresenta as generalizações para a classe dos MLGs, supondo inicialmente a situação simples do teste de hipóteses:

$$\mathcal{H}_0 : \boldsymbol{\beta} = \boldsymbol{\beta}^0 \text{ contra } \mathcal{H}_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}^0,$$

em que β^0 é um vetor p -dimensional conhecido e ϕ também assumido conhecido.

Teste da razão de verossimilhanças

Para o caso simples, o teste é definido por

$$\xi_{RV} = 2\{L(\hat{\beta}) - L(\beta^0)\}.$$

Pode-se definir esta estatística para os MLGs como a diferença entre as funções desvio

$$\xi_{RV} = \phi\{D(\mathbf{y}; \hat{\boldsymbol{\mu}}^0) - D(\mathbf{y}; \hat{\boldsymbol{\mu}})\},$$

em que $D(\mathbf{y}; \hat{\boldsymbol{\mu}}^0)$ como dado em (3.9) e $\hat{\boldsymbol{\mu}}^0 = g^{-1}(\hat{\boldsymbol{\eta}}^0)$, $\hat{\boldsymbol{\eta}}^0 = \mathbf{X}\beta^0$. Conforme Cordeiro (2016), sob a hipótese nula, a estatística ξ_{RV} tem distribuição assintótica de uma qui-quadrado com p graus de liberdade. Rejeita-se a hipótese nula a um nível $100\alpha\%$, para amostras grandes, se $\xi_{RV} > \chi_{p,1-\alpha}^2$.

Teste F

A estatística F para hipóteses simples tem forma dada por

$$F = \frac{\{D(\mathbf{y}; \boldsymbol{\mu}^0) - D(\mathbf{y}; \hat{\boldsymbol{\mu}})\}/p}{D(\mathbf{y}; \hat{\boldsymbol{\mu}})/(n-p)},$$

que para $\phi \rightarrow \infty$ e sob \mathcal{H}_0 segue uma distribuição $F_{p,(n-p)}$. Segundo Paula (2013) este resultado também vale quando $n \rightarrow \infty$ quando colocarmos no denominador de F uma estimativa consistente de ϕ . O autor também comenta sobre uma interessante propriedade das estatísticas de escore, razão de verossimilhanças e F, que é a invariância em relação a reparametrizações e para estatística F não depender do parâmetro de dispersão ϕ sendo obtida diretamente através das funções de desvio, sendo talvez mais conveniente para usos práticos.

3.6 Seleção de modelos

Critério de Informação de Akaike (AIC)

Paula (2013) discute um critério que já é bem utilizado na Regressão Linear: O *Critério de Informação de Akaike* (1974) ou simplesmente AIC. Este critério é baseado na *função de verossimilhança*, e penaliza, com um fator de correção, a complexidade do modelo. O AIC é dado por

$$\text{AIC} = D^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) + 2p, \quad (3.10)$$

em que p é o número de parâmetros e $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}})$ como dado em (3.8). Um valor pequeno do AIC é considerado como representativo de um melhor ajuste, portanto, na seleção de modelos, aquele com um menor valor de AIC é aquele que melhor se ajusta aos dados.

3.7 Técnicas de Diagnóstico em MLGs

Segundo Cordeiro e Demétrio (2008), os **resíduos** de um modelo são importantes para verificar a distância, ou discrepância, entre os valores observados y_i e os valores ajustados $\hat{\mu}_i$ pelo modelo com o objetivo de analisar a qualidade do ajuste. Portanto, a análise de resíduos avalia *a escolha da distribuição, a função de ligação e o preditor linear*. E ainda, com esta análise, conseguimos identificar observações mal ajustadas ou discrepantes que o modelo não consegue explicar.

Primeiro, podemos definir a *matriz de projeção*, que para os MLG, pode ser compreendida como:

$$\hat{\mathbf{H}} = \hat{\mathbf{W}}^{1/2} \mathbf{X} (\mathbf{X}^T \hat{\mathbf{W}} \mathbf{X})^{-1} \mathbf{X}^T \hat{\mathbf{W}}^{1/2}, \quad (3.11)$$

sendo $\hat{\mathbf{H}}$ uma matriz simétrica e idempotente. Uma propriedade notável da matriz $\hat{\mathbf{H}}$ é

$$\text{tr}(\hat{\mathbf{H}}) = \sum_{i=1}^n \hat{h}_{ii} = p \quad \text{e} \quad 0 \leq \hat{h}_{ii} \leq 1,$$

em que \hat{h}_{ii} é o i -ésimo elemento da diagonal principal da matriz $\hat{\mathbf{H}}$. Com isso, podemos definir os resíduos R_i para os MLG como

$$R_i = h_i(y_i, \hat{\mu}_i). \quad (3.12)$$

em que h_i , segundo Cordeiro e Demétrio (2008), é uma função adequada e de fácil interpretabilidade, geralmente escolhida de modo a estabilizar a variância ou induzir simetria na distribuição amostral de R_i . Para definições mais gerais sobre resíduos, ver por exemplo Cox e Snell (1968). A escolha da função h_i dependerá de qual tipo de anomalia se quer detectar. Cordeiro e Demétrio (2008) citam algumas anomalias mais frequentes, tais como:

- a) uma falsa distribuição populacional para a variável resposta;
- b) uma ou mais observações que não pertencem à distribuição escolhida para a variável resposta;
- c) algumas observações dependentes ou alguma forma de correlação serial;
- d) um parâmetro importante que está sendo omitido do modelo.

Através de uma escolha adequada de h_i constrói-se os seguintes gráficos para a detecção das anomalias citadas:

- i) resíduos ordenados $R_{(i)}$ versus pontos percentuais de alguma distribuição de referência $F(\cdot)$, que podem ser definidos por $F^{-1}[(i - \alpha)/(n - 2\alpha + 1)]$ para $0 \leq \alpha \leq 0,5$;
- ii) R_i versus $\hat{\mu}_i$;
- iii) R_i versus i ;
- iv) R_i versus níveis da variável ou fator correspondente ao parâmetro omitido.

Conforme Paula (2013), uma primeira proposta seria considerar o resíduo ordinário da solução de mínimos quadrados da regressão ponderada de $\hat{\mathbf{z}}$ contra \mathbf{X} , definido por

$$\mathbf{r}^* = \widehat{\mathbf{W}}^{-\frac{1}{2}}(\hat{\mathbf{z}} - \hat{\boldsymbol{\eta}}) = \widehat{\mathbf{V}}^{-\frac{1}{2}}(\mathbf{y} - \hat{\boldsymbol{\mu}}).$$

Podemos entender \hat{r}_{p_i} como sendo o i -ésimo resíduo de Pearson estimado, dado por

$$\hat{r}_{p_i} = \frac{y_i - \hat{\mu}_i}{\sqrt{\widehat{V}_i}}. \quad (3.13)$$

Outro tipo de resíduo utilizado nos MLG's é conhecido como o **resíduo de Pearson Studentizado** dado por

$$t_{S_i} = \frac{(y_i - \hat{\mu}_i)}{\sqrt{\hat{V}_i(1 - \hat{h}_{ii})}},$$

que de acordo com Turkman e Silva (2000), um ponto negativo deste resíduo é que sua distribuição para modelos não normais, é, frequentemente, assimétrica. Paula (2013) comenta que os resíduos mais utilizados em MLGs são definidos a partir de componentes da função desvio, dados por

$$t_{D_i} = \frac{d^*(y_i; \hat{\mu}_i)}{\sqrt{1 - \hat{h}_{ii}}} = \frac{\phi^{1/2}d(y_i; \hat{\mu}_i)}{\sqrt{1 - \hat{h}_{ii}}} \quad (3.14)$$

em que $d(y_i; \hat{\mu}_i) = \pm\sqrt{2}\{y_i(\tilde{\theta}_i - \hat{\theta}_i) + (b(\hat{\theta}_i) - b(\tilde{\theta}_i))\}^{1/2}$. O sinal de $d(y_i; \hat{\mu}_i)$ é o mesmo de $y_i - \hat{\mu}_i$. Willians (1984), através de estudos de simulação, mostra que a distribuição de t_{D_i} está mais próxima da normalidade do que a distribuição dos demais resíduos. O resíduo $d(y_i; \hat{\mu}_i)$ representa uma distância da observação y_i ao valor estimado $\hat{\mu}$, medida na escala do logaritmo da função de verossimilhança (Cordeiro, 2008). Existem algumas vantagens de se utilizar estes resíduos (Cordeiro, 2008):

- i) Não necessitam de funções normalizadoras;
- ii) São de simples computação após o ajuste do MLG;
- iii) São definidos para todas as observações mesmo para observações censuradas (desde que essas contribuam para o logaritmo da verossimilhança).

Os resíduos definidos em (3.14) são chamados de **componentes do desvio estudentizados**.

Para o modelo gama, a expressão para t_{D_i} é dada por

$$t_{D_i} = \pm \frac{\sqrt{2\phi}}{\sqrt{1 - \hat{h}_{ii}}} \{\log(\hat{\mu}_i/y_i) - (y_i - \hat{\mu}_i)/\hat{\mu}_i\}^{1/2},$$

em que $y_i > 0$ e \hat{h}_{ii} é o i -ésimo elemento da diagonal da matriz $\hat{\mathbf{H}}$ com $w_i = (d\mu_i/d\eta_i)^2/\mu_i^2$.

Se η_i possuir um intercepto nulo, t_{D_i} será dado por

$$t_{D_i} = \pm \frac{\sqrt{2\phi}}{\sqrt{1 - \hat{h}_{ii}}} \{\log(\hat{\mu}_i/y_i)\}^{1/2}.$$

Para o modelo normal inverso, t_{D_i} será dado por

$$t_{D_i} = \pm \frac{\sqrt{2\phi}}{\sqrt{1 - \hat{h}_{ii}}} \frac{(y_i - \hat{\mu}_i)}{\hat{\mu}_i \sqrt{y_i}},$$

com $y_i > 0$ e \hat{h}_{ii} é o i -ésimo elemento da diagonal principal de $\widehat{\mathbf{H}}$ com $w_i = (d\mu_i/d\eta_i)^2/\mu_i^3$.

Estatísticas para Diagnóstico

Muito usadas na verificação de pontos atípicos, segundo Cordeiro (2016), as estatísticas para diagnósticos têm sua utilidade quando certas observações não seguem o padrão das outras observações. Assim, podemos utilizar algumas ferramentas de diagnóstico que serão abordadas a seguir:

Pontos de Alavanca

Os *pontos de alavanca* tem o objetivo de avaliar a influência de y_i sobre \hat{y}_i e essa medida é dada por \hat{h}_{ii} , isto é, pelo i -ésimo elemento da diagonal principal da matriz $\widehat{\mathbf{H}}$, segundo Paula (2013). Portanto, como vimos anteriormente, se $\text{tr}(\widehat{\mathbf{H}}) = \sum_{i=1}^n \hat{h}_{ii} = p$ e, ainda, considerando que em média cada valor de h_{ii} deve estar próximo de p/n , então um ponto de alavanca é compreendido como aquele cujo o valor

$$\hat{h}_{ii} > \frac{2p}{n}. \quad (3.15)$$

É pertinente comentar que uma maneira eficiente de visualizar observações consideradas como *pontos de alavanca*, seria utilizar um gráfico de \hat{h}_{ii} contra os valores ajustados \hat{y}_i e traçar um limite $h = 2p/n$ ou $h = 3p/n$, por exemplo. Singer *et al.* (2017) afirmam que esse critério é arbitrário e deve ser encarado com espírito puramente descritivo.

Medida de Influência

Um *ponto influente* é uma observação que afeta significativamente o modelo, segundo Turkman e Silva (2000). Isto é, a modificação dessa observação, ou mesmo sua exclusão, produz efeitos relevantes na estimativas dos parâmetros ou de, especificamente, um parâmetro.

Com isso, podemos utilizar uma medida baseada na verossimilhança que analisa a influência da retirada da i -ésima observação em $\hat{\boldsymbol{\beta}}$, tal que

$$LD_i = 2\{l(\hat{\boldsymbol{\beta}}; \mathbf{y}) - l(\hat{\boldsymbol{\beta}}_{(i)}, \mathbf{y})\},$$

em que $\hat{\boldsymbol{\beta}}_{(i)}$ são as estimativas de máxima verossimilhança do vetor $\hat{\boldsymbol{\beta}}$ sem a observação y_i . Paula (2013), ressalta que não é possível obter uma forma analítica para LD_i e para isso, é utilizada a segunda expansão em série de Taylor em torno de $\hat{\boldsymbol{\beta}}$, obtendo-se

$$LD_i \cong (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^\top \{-\widehat{\mathbf{K}}_{\boldsymbol{\beta}\boldsymbol{\beta}}\}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}),$$

em que $-\widehat{\mathbf{K}}_{\boldsymbol{\beta}\boldsymbol{\beta}}$ é a matriz de informação de Fisher avaliada em $\hat{\boldsymbol{\beta}}$. Substituindo $-\widehat{\mathbf{K}}_{\boldsymbol{\beta}\boldsymbol{\beta}}$ pelo seu respectivo valor esperado e $\boldsymbol{\beta}$ por $\boldsymbol{\beta}_{(i)}$, obtemos

$$LD_i \cong \phi(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)})^\top (\mathbf{X}^\top \widehat{\mathbf{W}} \mathbf{X})(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)}). \quad (3.16)$$

Como seria muito trabalhoso realizar o método iterativo para todas as observações para estimar $\hat{\boldsymbol{\beta}}_{(i)}$, é utilizada apenas o primeiro passo do processo iterativo, tendo $\hat{\boldsymbol{\beta}}$ como valor inicial, como forma de aproximação para $\boldsymbol{\beta}_{(i)}$. Logo,

$$\hat{\boldsymbol{\beta}}_{(i)}^1 = \hat{\boldsymbol{\beta}} - \frac{\hat{r}_{pi} \sqrt{w_i \hat{\phi}^{-1}}}{\sqrt{1 - \hat{h}_{ii}}} (\mathbf{X}^\top \widehat{\mathbf{W}} \mathbf{X})^{-1} \mathbf{x}_i, \quad (3.17)$$

em que $\hat{r}_{pi} = (y_i - \hat{\mu}_i)^2 / V(\hat{\mu}_i)$ são os resíduos de Pearson.

Portanto, substituindo-se (3.17) em (3.16), obtemos

$$LD_i \cong \left\{ \frac{\hat{h}_{ii}}{(1 - \hat{h}_{ii})} \right\} t_{S_i}^2, \quad (3.18)$$

denominada de distância de Cook aproximada. Conforme Paula (2013), a validade da expressão (3.18) tem sido investigada por alguns pesquisadores, constatando que a mesma em geral subestima o verdadeiro valor de LD_i , porém, suficiente para detectar pontos influentes. Para o modelo gama, a distância de Cook aproximada fica dada por:

$$LD_i = \frac{\phi \hat{h}_{ii}}{(1 - \hat{h}_{ii})^2} \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i^2}, \quad (3.19)$$

e para o modelo normal inverso, temos

$$LD_i = \frac{\phi \hat{h}_{ii}}{(1 - \hat{h}_{ii})^2} \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i^3}. \quad (3.20)$$

Influência local

Cook (1986) propôs uma metodologia para avaliar a influência que pequenas mudanças na variabilidade das observações podem causar nas estimativas dos parâmetros. A principal ideia é estudar o comportamento de alguma medida particular de influência segundo perturbações infinitesimais nos dados ou no modelo. Se o interesse for estudar a influência local das observações no ajuste, o método é perturbar as covariáveis ou a variável resposta e utilizar alguma medida adequada para quantificar a influência dessas observações, como aponta Paula (2013).

Considere o logaritmo da função de verossimilhança do modelo sob questão denotado por $L(\boldsymbol{\theta})$ e $\boldsymbol{\theta}$ um vetor r -dimensional. Para os MLGs pode-se ter $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \phi)^\top$ e $r = p + 1$ ou $\boldsymbol{\theta} = \boldsymbol{\beta}$ para ϕ conhecido. Considere também $\boldsymbol{\delta}$ um vetor de dimensão $q \times 1$ de perturbações. Paula (2013) afirma que em geral temos $q = n$ e as perturbações são feitas no logaritmo da verossimilhança tal que este seja $L(\boldsymbol{\theta}, \boldsymbol{\delta})$. Denote também o vetor de não-perturbação $\boldsymbol{\delta}_0$ e assumamos que $L(\boldsymbol{\theta}|\boldsymbol{\delta}_0) = L(\boldsymbol{\theta})$. Verifica-se a influência das perturbações na estimativa de máxima verossimilhança $\hat{\boldsymbol{\theta}}$ a partir de

$$LD(\boldsymbol{\delta}) = 2\{L(\hat{\boldsymbol{\theta}}) - L(\hat{\boldsymbol{\theta}}_\boldsymbol{\delta})\},$$

denotando-se como o afastamento pela verossimilhança, de modo que $\hat{\boldsymbol{\theta}}_\boldsymbol{\delta}$ denota a estimativa de máxima verossimilhança sob o modelo $L(\boldsymbol{\theta}|\boldsymbol{\delta})$.

Em Paula (2013), é discutido que a ideia básica de influência local é estudar o comportamento da função $LD(\boldsymbol{\delta})$ em torno de $\boldsymbol{\delta}_0$. Desta maneira, o que se faz é selecionar uma direção unitária $\boldsymbol{\ell}$ tal que $\|\boldsymbol{\ell}\| = 1$ e estudar o gráfico de $LD(\boldsymbol{\delta}_0 + a\boldsymbol{\ell})$ contra $a \in \mathbb{R}$, conhecido como gráfico de *linha projetada*. Cada uma dessas linhas projetadas será representada por uma curvatura normal $C_\ell(\boldsymbol{\theta})$ em torno de $a = 0$. Paula (2013) sugere a

utilização da direção ℓ_{max} que é a maior curvatura denotada por $C_{\ell_{max}}$ em que a curvatura normal na direção ℓ é dada por

$$C_{\ell}(\boldsymbol{\theta}) = 2|\boldsymbol{\ell}^{\top} \boldsymbol{\Delta}^{\top} (-\mathbf{K}_{\hat{\theta}\hat{\theta}})^{-1} \boldsymbol{\Delta} \boldsymbol{\ell}|,$$

em que $-\mathbf{K}_{\hat{\theta}\hat{\theta}}$ é a matriz de informação observada de Fisher e $\boldsymbol{\Delta}$ é uma matriz de dimensão $r \times q$ tal que seus elementos são $\Delta_{ij} = \partial^2 L(\boldsymbol{\theta}|\boldsymbol{\delta}) / \partial \theta_i \partial \delta_j$, para $i = 1, \dots, r$ e $j = 1, \dots, q$. O máximo de $\boldsymbol{\ell}^{\top} \mathbf{B} \boldsymbol{\ell}$, sendo $\mathbf{B} = \boldsymbol{\Delta}^{\top} (-\mathbf{K}_{\hat{\theta}\hat{\theta}})^{-1} \boldsymbol{\Delta}$, será o maior autovalor em absoluto de \mathbf{B} . Assim, $C_{\ell_{max}}$ será o maior autovalor da matriz \mathbf{B} e ℓ_{max} seu maior autovetor associado. Convém fazer o gráfico $|\ell_{max}|$ contra as ordens das observações e assim verificar quais pontos exercem maior influência na vizinhança de $LD(\delta_0)$.

Estendendo para os MLG's, supondo ϕ conhecido, podemos denotar a função de verossimilhança como $L(\boldsymbol{\beta}|\boldsymbol{\delta})$. A ideia que se segue, é perturbar a verossimilhança considerando $L(\boldsymbol{\beta}|\boldsymbol{\delta})$ de tal forma que

$$L(\boldsymbol{\beta}, \boldsymbol{\delta}) = \sum_{i=1}^n \delta_i L_i(\boldsymbol{\beta}), \quad 0 \leq \delta_i \leq 1,$$

então a matriz $\boldsymbol{\Delta}$ assume a forma $\boldsymbol{\Delta} = \sqrt{\phi} \mathbf{X}^{\top} \widehat{\mathbf{W}}^{\frac{1}{2}} \mathbf{D}(\hat{\mathbf{r}}_p)$ em que $\mathbf{D}(\hat{\mathbf{r}}_p) = \text{diag}\{\hat{r}_{P_1}, \hat{r}_{P_2}, \dots, \hat{r}_{P_n}\}$ e $\hat{r}_{P_i} = \sqrt{\phi}(y_i - \hat{\mu}_i) / \sqrt{V_i}$ é o i -ésimo resíduo de Pearson estimado. Assim, estendendo para os MLGs, a curvatura normal na direção ℓ , substituindo $-\mathbf{K}_{\beta\beta}$ por $\phi(\mathbf{X}^{\top} \widehat{\mathbf{W}} \mathbf{X})$, tem-se que

$$C_{\ell}(\boldsymbol{\theta}) = 2|\boldsymbol{\ell}^{\top} \mathbf{D}(\hat{\mathbf{r}}_p) \widehat{\mathbf{H}} \mathbf{D}(\hat{\mathbf{r}}_p) \boldsymbol{\ell}|.$$

em que $\widehat{\mathbf{H}}$ é a matriz de projeção definida em (3.11).

Conforme Paula (2013), se o interesse é calcular a curvatura normal na direção de ℓ da i -ésima observação, então pode-se avaliar o gráfico de índices de $C_i = 2\hat{h}_{ii}\hat{r}_{P_i}^2$. Para avaliar a influência local das observações nas estimativas dos parâmetros utiliza-se ℓ_{max} como o autovetor correspondente ao maior autovalor da matriz

$$\mathbf{B} = \mathbf{D}(\hat{\mathbf{r}}_p) \widehat{\mathbf{H}} \mathbf{D}(\hat{\mathbf{r}}_p).$$

Do mesmo modo, se o interesse for detectar observações influentes na estimativa de um coeficiente particular associado à uma variável \mathbf{X}_p , o vetor $\boldsymbol{\ell}_{\max}$ fica dado por

$$\boldsymbol{\ell}_{\max}^\top = \left(\frac{v_1 \hat{r}_{p1}}{\sqrt{C_{\ell_{\max}}}}, \dots, \frac{v_n \hat{r}_{pn}}{\sqrt{C_{\ell_{\max}}}} \right),$$

em que v_1, v_2, \dots, v_n são obtidos através da regressão linear de \mathbf{X}_i contra as colunas de \mathbf{X}_j para $1 \leq i < j \leq p$, com a matriz de pesos $\widehat{\mathbf{V}}$, ou seja,

$$\mathbf{v} = \widehat{\mathbf{V}}^{\frac{1}{2}} \mathbf{X}_i - \widehat{\mathbf{V}}^{\frac{1}{2}} \mathbf{X}_j (\mathbf{X}_j^\top \widehat{\mathbf{H}} \mathbf{X}_j)^{-1} \mathbf{X}_j^\top \widehat{\mathbf{V}} \mathbf{X}_i.$$

Técnicas gráficas

Paula (2013) e Cordeiro e Demétrio (2008) apresentam as mais importantes técnicas gráficas para métodos de diagnóstico em MLGs, dentre elas:

a) **Resíduos versus alguma função dos valores ajustados**

Recomenda-se utilizar o gráfico de algum tipo de resíduo estudentizado (t_{S_i} ou t_{D_i}) *versus* $\hat{\eta}_i$, ou *versus* os valores ajustados transformados de tal forma que se tenha variância constante para a distribuição em uso. Desta forma, pode-se utilizar $\log(\hat{\mu}_i)$ para a distribuição gama e $-2\hat{\mu}_i^{-1/2}$ para a distribuição normal inversa e, conforme Cordeiro e Demétrio (2008), um padrão nulo desse gráfico nos daria um indicativo de uma distribuição dos resíduos em torno de zero com amplitude constante. Um padrão com desvios sistemáticos indicaria algum tipo de curvatura ou mudanças sistemáticas na amplitude com o valor ajustado.

b) **Resíduos contra as variáveis explicativas**

Este gráfico é utilizado para mostrar se existe alguma relação sistemática entre os resíduos e uma variável já incluída no modelo. Um padrão nulo para esse gráfico é uma distribuição aleatória de média zero com amplitude constante.

c) **Gráfico de índices**

Este tipo de gráfico revela observações com resíduo, h_{ii} , distância de Cook modificada, grandes.

d) **Gráfico normal de probabilidades para t_{D_i} com envelope simulado**

Utilizado para detectar afastamentos de normalidade e informa sobre a existência de pontos discrepantes ou sobre a falta de homogeneidade de variâncias. Esse método

foi proposto por Atkinson (1985) que sugeriu a utilização de bandas de confiança através de simulações, o que se denomina por *envelope simulado*. A ideia consiste em gerar resíduos que tenham média zero e matriz de variância-covariância $(\mathbf{I}_n - \widehat{\mathbf{H}})$ sendo \mathbf{I}_n a matriz identidade de ordem n e $\widehat{\mathbf{H}}$ a matriz de projeção definida para os MLGs. Os passos para a construção desse método podem ser encontrados em Paula (2013).

e) **Gráfico de \widehat{z}_i contra $\widehat{\eta}_i$**

Verifica a adequabilidade da função de ligação (uma tendência linear indica uma adequação da função de ligação);

f) **Gráficos de \mathbf{LD}_i , \mathbf{C}_i ou $|\ell_{max}|$ contra a ordem das observações**

São utilizados para a detecção de pontos influentes.

4 APLICAÇÃO

4.1 O conjunto de dados de análise de limite de crédito

Na análise de limite de crédito, diversos fatores influenciam no valor do limite que um determinado cliente pode ter, baseados em diversas variáveis relativas às compras. Para traçar o limite adequado para cada perfil de cliente é necessário por exemplo, determinar o limite usado a fim de conceder melhores propostas de crédito. O banco de dados de estudo refere-se à atividade de compra de clientes na empresa UI-Credit, uma administradora de cartões que faz concessões de crédito para clientes e parceiros através dos seus cartões Private Label (PL).

Uma amostra aleatória simples de 100 observações do ano de 2021 foi analisada e extraída para fins de estudo. É importante salientar que o banco de dados de análise de limite de crédito possui um total de 61 variáveis (incluindo a variável resposta) das quais 60 são variáveis explicativas. Nos dados coletados, algumas variáveis possuem informações desatualizadas, possuem excessos de zeros (maior que 50% das observações), são combinações lineares de outras variáveis explicativas (algumas delas trazendo mesma a mesma informação), ou possuindo valores faltantes (*missing values*). No Apêndice E, na Tabela 13 é apresentada uma descrição de cada variável considerada para a aplicação. Na Tabela 14 são apresentadas as variáveis que foram excluídas com base em alguns critérios. Na Tabela 15 são apresentadas as variáveis restantes (25 variáveis) que serão utilizadas para o ajuste dos modelos. Pelos gráficos da Figura 3, pode-se perceber que a variável resposta possui assimetria positiva, mostrando que é razoável supor as distribuições assimétricas gama e normal inversa para a modelagem desta. Para o ajuste destes modelos foram usados os códigos em R disponibilizados por Paula (2013).

Na Tabela 4, temos as principais medidas descritivas das variáveis utilizadas. Pela Tabela 4, notamos algumas características importantes. Os clientes que possuem o menor valor de *limite usado médio*, tem valores dados por 34,60 e 5.921,66 reais. O mínimo e o máximo de *behavior score* são 43 e 98,55, representando clientes de médio e baixo risco, respectivamente. Se olharmos para a variável *idade conta*, notamos que o cliente mais novo tem *idade conta* igual a 2 meses e o cliente mais velho, 149 meses. A variável *rank vendas* possui valor mínimo de -710,64 e máximo de 6515,81. Pode-se notar que para a variável resposta, temos o valor 2,5678 de assimetria, confirmando assimetria positiva.

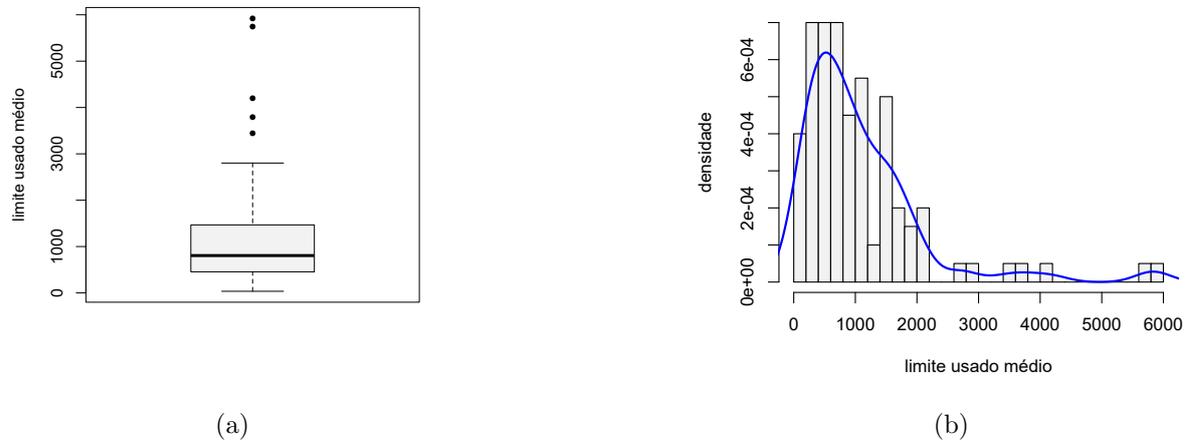


Figura 3 – Boxplot e histograma da variável resposta *limite usado médio*.

Tabela 4 – Medidas descritivas para as variáveis dos dados de análise de limite de crédito.

Variáveis	Mínimo	Máximo	Média	Mediana	D.Padrão	Assimetria	Curtose
LUM	34,60	5921,66	1087,98	805,26	1031,92	2,5678	8,5389
BS	43,00	98,55	84,29	88,04	12,51	-1,4256	1,7014
VLA	300,00	9884,00	2414,21	1833,70	1896,92	1,8483	3,9171
VLN	435,00	13650,00	3198,41	2583,00	2385,61	2,0382	4,7814
IC	2,00	149,00	44,53	35,00	34,60	1,1535	0,8023
TA	12,00	40,00	25,08	12,00	13,76	0,1397	-1,9790
MedS4M	45,93	98,22	84,04	86,37	10,53	-1,1655	1,5124
CF	7,00	80,00	40,29	35,50	21,58	0,2712	-1,2375
MedVF	64,81	1402,51	320,77	250,67	223,66	2,2350	6,4325
MaxVF	128,65	2991,73	800,45	630,74	550,16	1,4772	2,5769
MaxCV	17,97	1014,89	168,48	92,48	181,57	2,2236	5,6456
VCV	21,48	3407,38	629,40	350,89	695,83	1,6724	2,6190
QCV	1,00	52,00	10,05	7,00	9,79	1,8473	3,9028
LC	300,00	9884,00	2502,97	1977,50	2012,97	1,8927	3,8714
QIFOP	1,00	12,00	3,01	2,00	2,09	1,4672	2,8365
QCT	2,00	58,00	16,25	12,00	12,25	1,4720	2,0360
R3M	-10,19	495,97	63,61	43,26	80,31	3,4356	13,3264
R6M	-3,59	1070,37	152,60	92,17	167,03	2,6975	9,8364
R9M	29,32	1168,89	218,35	154,67	197,01	2,2509	6,2492
PMaxCP	49,99	2000,00	362,46	210,48	367,39	2,2450	5,5501
PVCP	49,99	2946,92	467,41	254,81	485,12	2,3457	7,1632
PQCP	1,00	11,00	1,78	1,00	1,53	3,2575	14,1213
MDA	0,00	31,00	3,74	2,00	5,24	2,66	8,87
MCP	0,00	2284,82	515,14	311,98	495,37	1,52	2,20
VCP	0,00	11399,27	1583,16	1023,04	1834,85	2,59	8,96
RKV	-710,64	6515,81	976,56	572,06	1360,10	1,6623	3,4526

Uma regra que será útil na avaliação das características dos clientes com relação ao *behavior score* são as classificações de faixas de risco, apresentadas na Tabela 5.

Tabela 5 – Classificação de risco com relação ao *behavior score*.

Faixas de Behavior	Risco
[0 – 35)	Alto Risco
[35 – 70)	Médio Risco
≥ 70	Baixo Risco

Também é interessante observar a relação marginal entre as variáveis explicativas e a variável resposta. Na Figura 4 são apresentados os gráficos de dispersão entre algumas variáveis. Pelos gráficos da Figura 4 podemos notar que embora a variável *idade conta* não pareça ter um comportamento bem definido com relação ao *limite usado médio*, observamos que quanto mais “velho” um cliente é, maior o *limite usado médio*. Uma relação mais definida aparece quando a *média do valor de faturas* aumenta, nota-se um aumento no *limite usado médio*. Também temos que a medida em que a *quantidade de compras total* aumenta, o *limite usado médio* parece aumentar. Interpretações semelhantes são observadas quando analisamos o *limite combinado*, *valor de compras à vista* e *valor do limite antigo*.

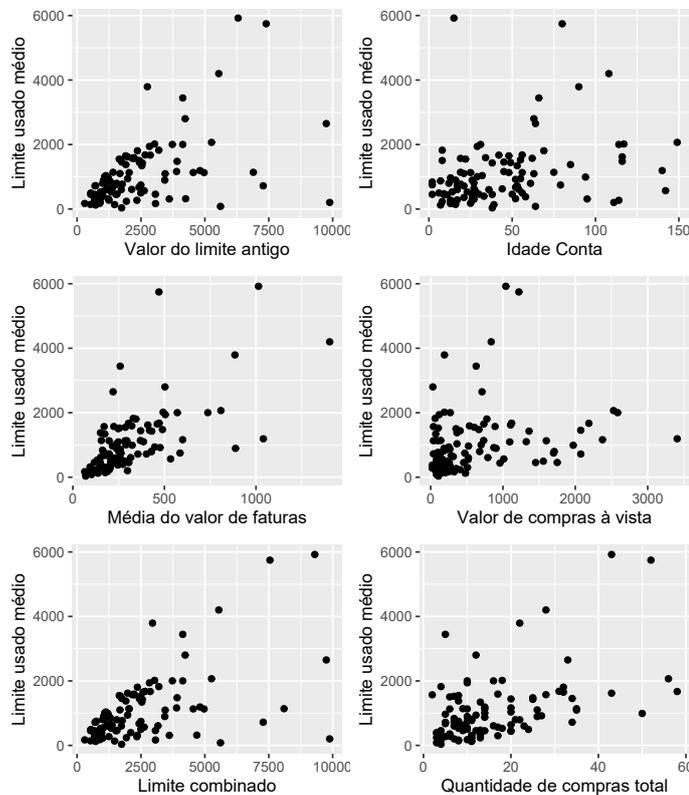


Figura 4 – Gráficos de dispersão entre algumas covariáveis *versus* a variável resposta *limite usado médio*.

4.2 Ajuste utilizando o modelo gama

Para a variável resposta consideramos $Y \sim Gama(\mu, \phi)$, denotando a distribuição gama, de parâmetros μ e ϕ . Para analisarmos o ajuste dos modelos, empregou-se o critério de informação de Akaike (AIC) para a seleção das variáveis que irão compor os

modelos relativos às suas respectivas funções de ligação. Para compor o modelo, foram selecionadas as variáveis que fossem significativas ao nível de 5%. Na Tabela 6, são apresentados os modelos após a seleção de variáveis e o respectivo AIC relativo às suas funções de ligação. Para simplificar a interpretação e visualização, denotaremos para o modelo gama com função de ligação identidade, $(\beta_0, \dots, \beta_7)$ como os parâmetros que acompanham cada variável apresentada na Tabela 6.

Tabela 6 – Critério de informação de Akaike (AIC) para os modelos finais utilizando a distribuição gama para diferentes funções de ligação.

Função de ligação	Modelo	AIC
identidade $\eta = \mu$	$\eta = \beta_0 + \beta_1 \text{VLA} + \beta_2 \text{IC} + \beta_3 \text{MedVF} + \beta_4 \text{VCV} + \beta_5 \text{LC} + \beta_6 \text{QCT} + \beta_7 \text{RKV}$	1415,8
recíproca $\eta = 1/\mu$	$\eta = \beta_0 + \beta_1 \text{BS} + \beta_2 \text{VLN} + \beta_3 \text{IC} + \beta_4 \text{CF} + \beta_5 \text{MedVF} + \beta_6 \text{VCV} + \beta_7 \text{PmaxCP} + \beta_8 \text{RKV}$	1522
logarítmica $\eta = \log(\mu)$	$\eta = \beta_0 + \beta_1 \text{BS} + \beta_2 \text{VLN} + \beta_3 \text{CF} + \beta_4 \text{MedVF} + \beta_5 \text{VCV} + \beta_6 \text{RKV}$	1489,9

Na Tabela 7 são apresentadas as estimativas dos parâmetros, erros-padrão das estimativas e os respectivos p -valores associados aos parâmetros para o modelo gama com função de ligação identidade. O motivo de serem apresentados apenas os resultados do modelo gama com função de ligação identidade se dá pelos ajustes feitos através das técnicas de diagnóstico e pode-se constatar que os modelos gama com funções de ligação logarítmica e inversa não se ajustaram tão bem quando comparado ao modelo gama com função de ligação identidade. Por este motivo, prosseguiremos uma análise somente para o modelo gama com função de ligação identidade.

Tabela 7 – Estimativas dos parâmetros referente ao modelo gama com função de ligação identidade ajustado aos dados de análise de limite usado médio.

Coefficientes	Estimativa	Erro Padrão	p -valor
Intercepto	1049,249	98.537	< 0,01
VLA	0,193	0,089	0,03305
IC	-2,117	0,862	0,01589
MedVF	-1,531	0,381	< 0,01
VCV	-0,452	0,075	< 0,01
LC	-0,188	0,088	0,03608
QCT	-7,299	4,244	0,08886
RKV	1,063	0,084	< 0,01

4.3 Análise de resíduos e diagnóstico para o modelo gama

Como mencionado na Seção 4.2, apresentaremos a análise de resíduos e diagnóstico apenas para o modelo gama com função de ligação identidade, dado que para as funções de ligação inversa e logarítmica o ajuste não foi superior aos ajustes utilizando a função de ligação identidade. Deste modo, serão omitidos os resultados para as funções de ligação inversa e logarítmica e prosseguiremos somente com a função de ligação identidade.

A Figura 5 apresenta os gráficos de diagnóstico referente ao modelo gama com função de ligação identidade. Podemos observar que temos pontos com alto grau de alavancagem no gráfico de índices *versus* a distância de Cook. Os pontos são #22, #88, #82, #67, #79 e #50 com destaque para a observação #22. A observação #22 representa um cliente que tem baixo valor de *limite usado médio*, sendo este valor igual a 122,25 reais, com idade conta de 7 meses, isto é, um cliente novo se comparado ao cliente mais velho (149 meses), com valor de *behavior score* de 67,5 (médio risco) possuindo um valor baixo para a variável de *rank vendas*. A observação #88 representa um cliente com valor de *limite usado médio* igual a 34,60 que pode ser um valor muito baixo se comparado com o valor máximo de *limite usado médio*, com *idade conta* de 38 semanas, com valor de *behavior score* 67 (médio risco) e valor de *rank vendas* -710,64, que é o menor valor para essa variável, indicando que esse cliente é bastante ruim em relação à vendas. A observação #82 representa um cliente com valor de *limite usado médio* de 3.444,1 reais, com idade conta de 66 meses, isto é, um cliente mais velho em relação à média e à mediana de *idade conta* com *behavior score* de 86,28 (baixo risco), com valor de *rank vendas* de 891,7. A observação #67 representa um cliente com valor relativamente alto de *limite usado médio*, com 19 meses, ou seja, um cliente mais novo com relação aos demais clientes, com *behavior score* de 95 (baixo risco) e valor de *rank vendas* -110,4. A observação #79 representa um cliente com *limite usado médio* de 451,4, *idade conta* de 38 semanas, *behavior score* de 81,96 (baixo risco) e valor de *rank vendas* -55,33. Por fim, a observação #50 representa um cliente com valor de *limite usado médio* de 1378, *idade conta* de 85 meses, *behavior score* de 94 (baixo risco) e *rank vendas* de 239,3.

Podemos também perceber através da Figura 5 (d) indícios de que a função de ligação apresenta uma tendência em torno da reta nula com exceção de alguns pontos. A linha tracejada representa a reta nula com inclinação de 45 graus e a linha cheia representa uma reta suavizada através do método *lowess* robusto para ajudar na detecção de tendência.

Já na Figura 5 (c) podemos detectar pontos aberrantes (outliers). Notamos que os pontos #67 e #82 estão fora dos limites -2 e 2, o que pode ser um indicativo de que o modelo não se ajusta de forma adequada aos dados.

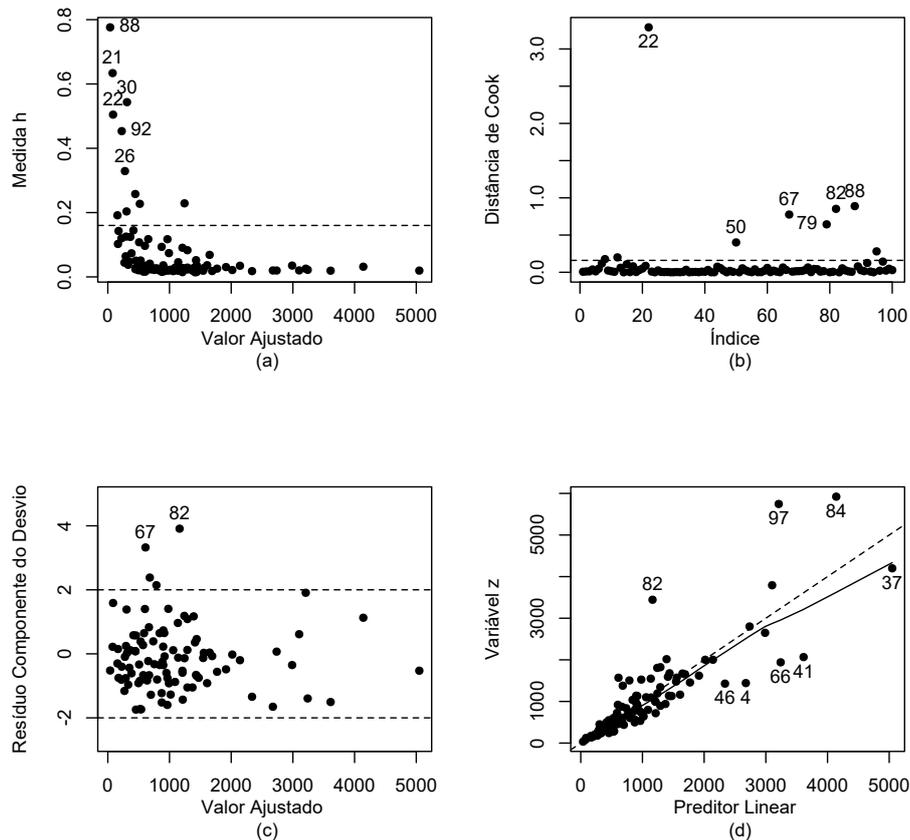


Figura 5 – Gráficos de diagnóstico referentes ao modelo gama com função de ligação identidade ajustado aos dados de limite de crédito médio.

A Figura 6 (a) e a Figura 6 (b) apresentam os gráficos de índices *versus* medida \hat{h} e índices *versus* l_{max} , respectivamente. Podemos perceber alguns pontos que são localmente influentes com destaque para as observações #82 e #67, em conjunto com alguns pontos já destacados e mencionados anteriormente.

O gráfico normal de probabilidades apresentado na Figura 7 apresenta indícios de que a distribuição gama com função de ligação identidade não seja adequada para explicar o limite usado médio, no entanto, devemos retirar as observações destacadas como potencialmente influentes nos gráficos de diagnóstico e reavaliar as estimativas dos parâmetros com a retirada das observações.

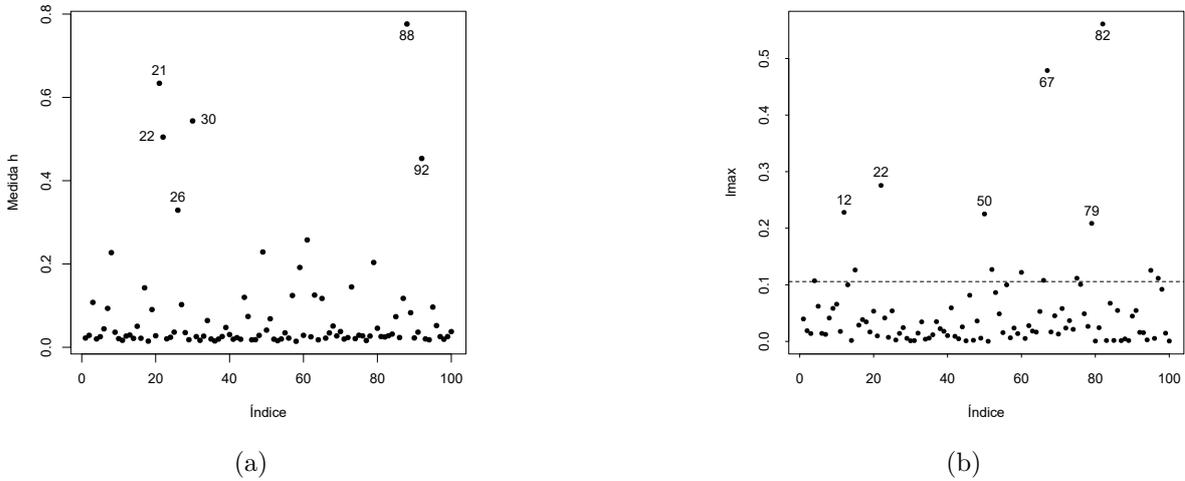


Figura 6 – (a) Gráfico de índice *versus* medida h e (b) gráfico de índices *versus* l_{max} para a distribuição gama com função de ligação identidade.

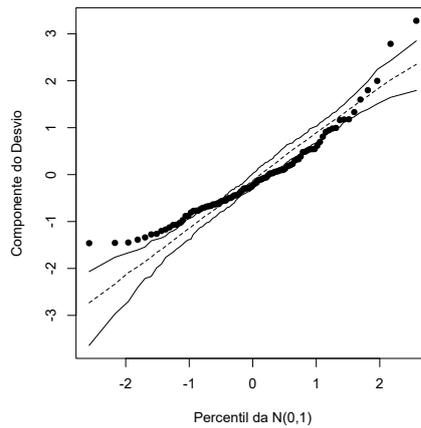


Figura 7 – Gráfico normal de probabilidades referente ao modelo gama com função de ligação identidade ajustado aos dados de análise de limite de crédito.

Na Tabela 8, apresentamos a variação percentual e estimativa dos parâmetros do modelo gama utilizando a função de ligação identidade ao excluir algumas observações conjuntamente. A variação percentual das estimativas é dada por

$$\mathbf{VP} = \left[\frac{\widehat{\beta}_{\mathbf{e}} - \widehat{\beta}_{\mathbf{c}}}{\widehat{\beta}_{\mathbf{c}}} \right] \times 100,$$

em que $\widehat{\beta}_{\mathbf{c}}$ é a estimativa do vetor de parâmetros do modelo quando se considera todas as observações e $\widehat{\beta}_{\mathbf{e}}$ é o vetor de parâmetros estimados com a retirada das observações, em que $\mathbf{e} = (e_1, e_2, \dots, e_n)^\top$ é o vetor de observações excluídas, e_i , $i = 1, \dots, n$ denotando a

i -ésima exclusão.

Para a retirada de observações, foram avaliadas as estimativas e os gráficos de diagnóstico para subconjuntos retirados de um a um, dois a dois, três a três elementos e assim por diante e por fim retirando-se todas as observações de forma conjunta. Na Tabela 8, foram apresentadas apenas os subconjuntos que após a retirada dos mesmos, produziram uma redução substancial do AIC, bem como a redução do desvio do modelo.

Tabela 8 – Variação percentual e estimativas dos parâmetros utilizando o modelo gama com função de ligação identidade com a exclusão de observações.

	Parâmetros estimados							
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_7$
Modelo completo	1049,249	0,193	-2,117	-1,531	-0,452	-0,188	7,299	1,063
	Estimativas após a retirada							
#{22}	1074,508	0,189	-1,511	-1,631	-0,480	-0,188	-8,866	1,103
#{22,82}	1023,674	0,178	-1,625	-1,504	-0,475	-0,176	-6,984	1,052
#{22,67,82}	970,847	0,159	-1,216	-1,485	-0,446	-0,159	-5,091	1,012
#{22,67,82,88}	970,401	0,159	-1,230	-1,484	-0,447	-0,159	-5,070	1,012
#{22,67,82,88,79,50}	963,743	0,153	-1,834	-1,425	-0,436	-0,146	-5,866	1,003
	p -valor							
Todas	< 0,01	0,3305	0,0159	< 0,01	< 0,01	0,0361	0,0889	< 0,01
#{22}	< 0,01	0,0318	0,1062	< 0,01	< 0,01	0,0313	0,0409	< 0,01
#{22,82}	< 0,01	0,0254	0,0482	< 0,01	< 0,01	0,0258	0,0719	< 0,01
#{22,67,82}	< 0,01	0,0254	0,1012	< 0,01	< 0,01	0,0255	0,1449	< 0,01
#{22,67,82,88}	< 0,01	0,0272	0,1293	< 0,01	< 0,01	0,0264	0,1597	< 0,01
#{22,67,82,88,79,50}	< 0,01	0,0255	0,0120	< 0,01	< 0,01	0,0312	0,1033	< 0,01
	Variação Percentual (%)							
#{22}	2,41	-1,81	-28,63	6,56	6,15	0,27	21,48	3,75
#{22,82}	-2,44	-7,72	-23,23	-1,75	4,92	-6,30	-4,31	-1,11
#{22,67,82}	-7,47	-17,50	-42,56	-2,96	-1,32	-15,47	-30,25	-4,83
#{22,67,82,88}	-7,51	-17,40	-41,91	-3,05	-1,20	-15,48	-30,54	-4,79
#{22,67,82,88,79,50}	-8,15	-20,63	-13,39	-6,88	-3,52	-22,15	-19,63	-5,68

Analisando os modelos após a retirada de observações, constatou-se que o melhor modelo obtido é o que exclui as observações #22 #67,#82,#88,#79 e #50 de forma conjunta, levando ao valor de desvio $D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 7,58$. Esse valor representa uma diminuição de 36,64% em relação ao desvio do modelo com todas as observações (11,96). O valor do AIC com a retirada de todas as observações foi de 1299,29, que representa uma redução em torno de 8,2% em relação ao AIC do modelo completo (1415,78).

Podemos observar também que embora as variações percentuais nas estimativas dos parâmetros β_1 , β_2 , β_5 e β_6 sejam grandes, não houve mudanças inferenciais significativas e as variáveis continuam significativas marginalmente ao nível de 5% com exceção da variável *quantidade de compras totais*, porém, se mantém estatisticamente significativa ao nível de

10%. A estimativa (erro padrão) do parâmetro de dispersão após a retirada das observações foi $\hat{\phi} = 12,56 (1,81)$ levando ao valor do desvio escalonado $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 12,56 \times 7,58 = 95,20$ com 86 graus de liberdade, conduzindo ao valor $p = 0,232$ indicando uma boa qualidade de ajuste.

A Figura 8 mostra o gráfico normal de probabilidades para o modelo gama com função de ligação identidade ajustado após a retirada de todas as observações. Pode-se perceber que o gráfico normal de probabilidades não apresenta indícios que a distribuição gama com função de ligação identidade seja inadequada para explicar o limite usado médio.

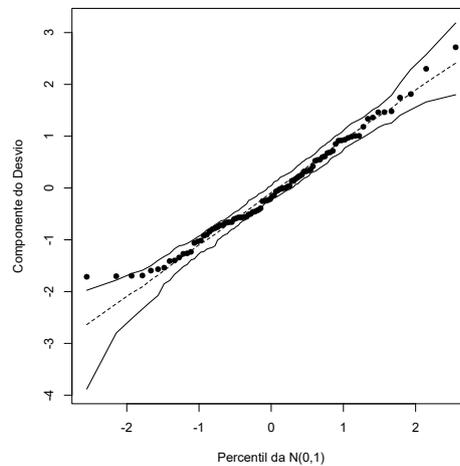


Figura 8 – Gráfico normal de probabilidades referente ao modelo gama com função de ligação identidade retirando-se todas as observações de forma conjunta ajustado aos dados de análise de limite de crédito médio.

Interpretação dos parâmetros

Para a interpretação dos parâmetros, consideramos o modelo com a retirada de todas as observações, dado por:

$$\begin{aligned} \hat{\mu}_i = & 963,743 + 0,153 \text{ VLA} - 1,834 \text{ IC} - 1,425 \text{ MedVF} - 0,436 \text{ VCV} \\ & - 0,146 \text{ LC} - 5,866 \text{ QCT} + 1,003 \text{ RKV}. \end{aligned}$$

- $\hat{\beta}_0$: representa a valor esperado do *limite usado médio* considerando fixas e iguais a zero as variáveis restantes do modelo, que neste caso, espera-se uma variação de R\$ 963,743 reais no limite usado médio;

- $\hat{\beta}_1$: quando o *limite antigo* aumenta ou tem o acréscimo de 1 real, o *limite usado médio* aumenta 0,153 em média, mantendo-se fixas as demais covariáveis;
- $\hat{\beta}_2$: representa uma variação esperada de R\$ -1,834 reais no *limite usado médio* a cada mês acrescido na *idade de conta* mantendo-se fixas as demais covariáveis;
- $\hat{\beta}_3$: representa uma variação esperada de R\$ -1,425 reais no *limite usado médio* quando aumenta-se 1 real na *média do valor de fatura*, mantendo-se fixas as demais covariáveis.
- $\hat{\beta}_5$: representa a variação esperado do valor de *limite usado médio* em R\$ -0,146 reais a quando o *limite combinado* é acrescido de 1 real, mantendo-se fixas as demais covariáveis.
- $\hat{\beta}_6$: representa a variação esperada no *limite usado médio* de R\$ -5,866 reais à medida em que a *quantidade de compras* é acrescida de uma unidade, mantendo-se fixa as demais covariáveis.
- $\hat{\beta}_7$: variação esperada no *limite usado médio* de R\$ 1,003 real a cada unidade acrescida no valor de *rank vendas*, mantendo-se fixas as demais covariáveis;

Podemos observar que alguns parâmetros estimados parecem não ter contribuição adequada para explicar a variável resposta se interpretarmos de forma isolada, como por exemplo as variáveis *média do valor de fatura*, *valor de compras à vista*, *limite combinado* e *quantidade de compras totais*, que acompanham os parâmetros β_3 , β_4 , β_5 e β_6 , respectivamente. O motivo se dá pelo fato de que o *limite usado médio* tende a aumentar com o aumento da *média do valor de faturas*, como pode ser evidenciado através do gráfico de dispersão na Figura 4. O mesmo ocorre com a variável *valor de compras à vista*, pois a medida em que o *valor de compras à vista* aumenta espera-se um aumento no *limite usado médio*. Comportamento bem parecido ocorre com as variáveis *limite combinado* e *quantidade de compras totais*. Porém, pode ser que faça sentido dadas as estimativas obtidas e suas interpretações se analisarmos na presença das outras variáveis explicativas, ou seja, se olharmos o efeito de determinada estimativa em conjunto com as demais.

4.4 Ajuste utilizando o modelo normal inverso

Para a variável resposta consideramos $Y \sim NI(\mu, \phi)$ denotando a distribuição normal inversa de média μ e parâmetro de dispersão ϕ . Na Tabela 9, são apresentados os modelos após a seleção de variáveis e o respectivo AIC relativo às suas funções de ligação.

Tabela 9 – Critério de informação de Akaike (AIC) para os modelos finais utilizando a distribuição normal inversa para diferentes funções de ligação.

Função de ligação	Modelo	AIC
identidade $\eta = \mu$	$\eta = \beta_0 + \beta_1\text{VLA} + \beta_2\text{IC} + \beta_3\text{TA} + \beta_4\text{MedVF} + \beta_5\text{VCV} + \beta_6\text{LC} + \beta_7\text{RKV}$	1416,9
logarítmica $\eta = \log(\mu)$	$\eta = \beta_0 + \beta_1\text{IC} + \beta_2\text{TA} + \beta_3\text{CF} + \beta_4\text{MedVF} + \beta_5\text{MaxVF} + \beta_6\text{MaxCV} + \beta_7\text{VCV} + \beta_8\text{QCV} + \beta_9\text{QIFOP} + \beta_{10}\text{QCT} + \beta_{11}\text{R3M} + \beta_{12}\text{R6M} + \beta_{13}\text{R9M} + \beta_{14}\text{PVCP} + \beta_{15}\text{PQCP} + \beta_{16}\text{RKV}$	1506,4

Para as funções de ligação inversa e recíproca quadrática, não houve convergência para nenhum dos modelos na geração dos gráficos de envelope e outros gráficos de diagnóstico, mesmo após retirar variáveis com excesso de zeros e variáveis que poderiam causar multicolinearidade.

Foram analisados para os dois modelos da Tabela 9 os gráficos de diagnóstico e envelopes simulados, bem como as estimativas dos parâmetros após a retirada de observações influentes e pontos de alavanca. Constatou-se através dessas análises que o modelo normal inverso com função de ligação identidade se ajustou melhor aos dados quando comparado ao modelo com função de ligação logarítmica. Deste modo, utilizaremos o modelo normal inverso com função de ligação identidade para a realização dos ajustes.

Para o modelo considerado, temos na Tabela 10 as estimativas dos parâmetros, erros-padrão e p -valor associado.

Tabela 10 – Estimativas dos parâmetros referente ao modelo normal inverso com função de ligação identidade ajustado aos dados de análise de limite de crédito.

Coefficientes	Estimativa	Erro Padrão	p-valor
Intercepto	1005,403	90,926	< 0,01
VLA	0,203	0,068	< 0,01
IC	-1,579	0,705	0,0276
TA	-0,831	0,804	0,0303
MedVF	-1,731	0,338	< 0,01
VCV	-0,448	0,061	< 0,01
LC	-0,195	0,068	< 0,01
RKV	1,016	0,077	< 0,01

4.5 Análise de resíduos e diagnóstico para o modelo normal inverso

Na Figura 9 (a) de valores ajustados *versus* medida \hat{h} podemos notar alguns pontos que exercem grande influência em relação aos seus respectivos valores ajustados.

Na Figura 9 (b) notamos que a observação #88 é uma observação influente. Na Figura 9 (c) notamos 2 possíveis pontos aberrantes e na Figura 9 (d) notamos a presença de uma tendência linear, indicando que a função de ligação identidade parece ser adequada.

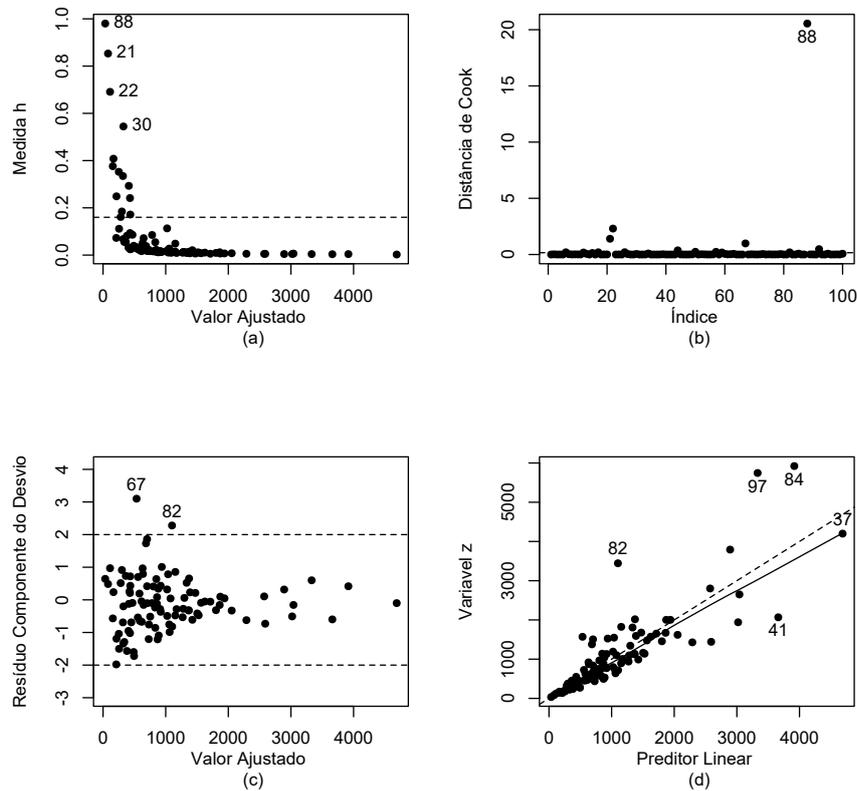


Figura 9 – Gráficos de diagnóstico referentes ao modelo normal inverso com função de ligação identidade ajustado aos dados de limite de crédito médio.

Na Figura 10 (a) temos o gráfico de índices *versus* medida \hat{h} destacando pontos de alavanca, que são os pontos #88 (ponto influente e de alto grau de alavancagem), #21, #22, #30, #27, #59, #92, #26 e #8. Na Figura 10 (b) notamos alguns pontos que são localmente influentes, que sob pequenas perturbações nos dados ou no modelo mudam substancialmente as estimativas dos parâmetros. Os pontos são #67, #12, #44, #18, #6, #13, #17, #82 e #100.

Para esta aplicação, foi analisada a retirada do ponto #88 e os demais pontos influentes localmente, e deste modo foram obtidos bons resultados após a retirada de todas as observações em conjunto. Para resumir as características dessas observações, na Tabela 11 temos os valores observados da variável resposta e as demais variáveis explicativas que compõem o modelo.

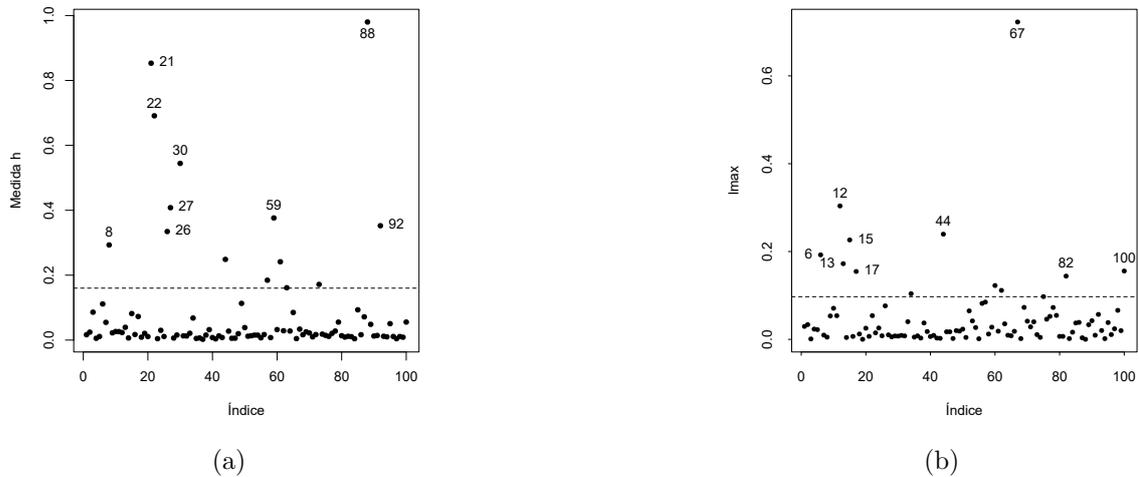


Figura 10 – (a) Gráfico de índice *versus* medida h e (b) gráfico de índices *versus* l_{max} para a distribuição normal inversa com função de ligação identidade.

Da Tabela 11, percebemos que grande parte dos clientes possuem *tempo de alteração* de 40 meses, com exceção da observação #82, que possui *tempo de alteração* de 12 meses, porém, com maior tempo de idade de conta, com valor de 66 meses e maior valor de *limite usado médio*, com valor de 3.444,11, o que pode ser considerado bem alto se comparado com as outras observações. Podemos notar que a observação #88, que foi bastante influente possui o menor valor de *rank vendas* dentre todos os clientes na amostra observada, pode ser que este fato explique a razão dessa observação ser tão destacada no gráfico de índices *versus* distância de Cook.

Tabela 11 – Valores das observações destacadas nos gráficos de diagnóstico para o modelo normal inverso com função de ligação identidade.

Observação	Valores							
	LUM	VLA	IC	TA	MedVF	VCV	LC	RKV
#6	178,32	889	25	40	189,38	263,23	889	-234,30
#12	1504,84	2436	8	40	247,78	349,86	2436	304,60
#13	282,43	870	17	40	181,73	21,48	870	-137,20
#17	136,00	1469	39	40	75,36	157,20	1469	-501,71
#44	168,34	3069	13	40	114,94	325,59	3069	-409,54
#67	1569,91	2242	19	40	170,31	38,61	2242	-110,37
#82	3444,11	4135	66	12	258,30	628,91	4135	891,73
#88	34,60	1743	38	40	69,67	108,08	1743	-710,64
#100	234,55	700	8	40	133,90	407,73	700	-208,29

Observando a Figura 11, notamos que o gráfico normal de probabilidades não apresenta indícios que o modelo normal inverso com função de ligação identidade seja adequado para explicar o limite usado médio, no entanto convém fazer a retirada das observações já mencionadas e reavaliar as estimativas.

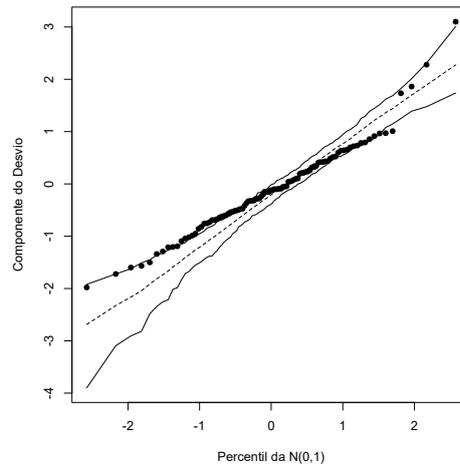


Figura 11 – Gráfico normal de probabilidades referente ao modelo normal inverso com função de ligação identidade ajustado aos dados de análise de limite de crédito médio.

Na Tabela 12, são apresentados os resultados das estimativas dos parâmetros, erros-padrão e p -valor com a eliminação das observações discutidas. A eliminação foi feita através de subconjuntos de 2 a 2 elementos, de 3 a 3 elementos e assim por diante até a eliminação de todas as observações conjuntamente. No entanto, a Tabela 12 apresenta somente os subconjuntos que levaram uma maior redução do AIC e do desvio do modelo. Após uma análise considerado a eliminação das observações, constatou-se que o modelo que melhor ajustou-se aos dados de análise de limite de crédito foi aquele considerando a retirada conjunta de todas as observações, que correspondem às observações destacadas na Tabela 11. Com a retirada conjunta de tais observações, obteve-se um valor de $AIC = 1268,75$ em relação ao valor de $AIC = 1416,864$ do modelo com todas as observações, representando uma redução de 10,45%. O desvio do modelo retirando todas as observações conjuntamente foi $D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 0,00899$ em relação ao desvio do modelo completo que foi de 0,01643, que representa uma redução de 45,22% no desvio. A estimativa do parâmetro de dispersão para este modelo foi $\hat{\phi} = 9219,56$ e o desvio escalonado foi $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 9219,56 \times 0,00899 = 82,98$ que conduz ao valor $p = 0,51$ indicando um bom ajuste do modelo.

Pode-se perceber na Tabela 12 que após a retirada de todas as observações de forma conjunta, as variações percentuais nas estimativas dos parâmetros em relação às estimativas obtidas com o modelo completo não foram grandes, ou seja, não chegaram a reduzir mais do que 10% nas suas estimativas com exceção dos parâmetros β_3 à variável *tempo de alteração* (TA) resultando em uma diminuição de 22,55%. Também não tivemos

alteração dos sinais que acompanham as variáveis logo após a retirada e as variáveis continuaram marginalmente significativas ao nível de 5%.

Tabela 12 – Variação percentual e estimativas dos parâmetros utilizando o modelo normal inverso com função de ligação identidade com a exclusão de observações.

	Parâmetros estimados							
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_7$
Modelo completo	1005,403	0,203	-1,579	-0,831	-1,731	-0,448	-0,195	1,016
	Estimativas após a retirada							
#{88}	1025,058	0,216	-1,809	-1,902	-1,682	-0,468	-0,200	1,039
#{67,82}	927,474	0,162	-0,974	-1,155	-1,567	-0,393	-0,157	0,937
#{67,82,17}	958,436	0,171	-0,914	-1,201	-1,688	-0,398	-0,167	0,967
#{67,12,44,82}	914,962	0,151	-0,789	-1,319	-1,564	-0,377	-0,147	0,921
#{6,12,13,17,44,67,82,100}	960,931	0,184	-1,532	-0,644	-1,634	-0,414	-0,177	0,979
	p-valor							
#{88}	< 0,01	0,0028	0,0138	0,0608	< 0,01	< 0,01	0,0045	< 0,01
#{67,82}	< 0,01	0,0037	0,0744	0,0574	< 0,01	< 0,01	0,0046	< 0,01
#{67,82,17}	< 0,01	0,0015	0,0890	0,0436	< 0,01	< 0,01	0,0019	< 0,01
#{67,12,44,82}	< 0,01	0,0035	0,1814	0,0269	< 0,01	< 0,01	0,0041	< 0,01
#{6,12,13,17,44,67,82,100}	< 0,01	< 0,01	< 0,01	0,0270	< 0,01	< 0,01	< 0,01	< 0,01
	Variação Percentual (%)							
#{88}	1,95	6,52	14,54	128,80	-2,85	4,57	2,47	2,20
#{67,82}	7,75	-20,32	-38,32	38,91	-9,52	-12,32	-19,44	-7,79
#{67,82,17}	-4,67	-15,66	-42,09	44,46	-2,47	-11,14	-14,41	-4,83
#{67,12,44,82}	-9,00	-25,66	-50,03	58,64	-9,67	-15,91	-24,63	-9,37
#{6,12,13,17,44,67,82,100}	-4,42	-9,28	-2,99	-22,55	-5,62	-7,46	-9,40	-3,65

A Figura 12 apresenta o gráfico normal de probabilidades para o modelo normal inverso com função de ligação identidade após a retirada de todas as observações de forma conjunta. Pode-se perceber que os pontos estão dentro das bandas de confiança indicando que o modelo normal inverso com função de ligação identidade, após a retirada de tais observações é adequado para explicar o *limite usado médio*.

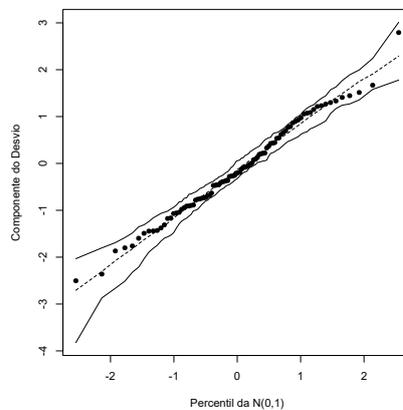


Figura 12 – Gráfico normal de probabilidades referente ao modelo normal inverso com função de ligação identidade retirando-se todas as observações de forma conjunta ajustado aos dados de análise de limite de crédito médio.

Interpretação dos parâmetros

Para a interpretação dos parâmetros, consideramos o modelo com a retirada de todas as observações, dado por:

$$\widehat{\mu}_i = 960,931 + 0,184 \text{VLA} - 1,532 \text{IC} - 0,644 \text{TA} - 1,634 \text{MedVF} \\ - 0,414 \text{VCV} - 0,177 \text{LC} - 0,979 \text{RKV}.$$

- $\widehat{\beta}_0$: representa o valor esperado do *limite usado médio* considerando fixas e iguais a zero todas as variáveis consideradas no modelo, neste caso, espera-se um *limite usado médio* de R\$ 960,931 reais;
- $\widehat{\beta}_1$: representa um aumento no valor esperado do *limite usado médio* de R\$ 0,184 reais ao acréscimo de R\$ 1 real no *valor do limite antigo*, mantendo-se fixas as demais covariáveis;
- $\widehat{\beta}_2$: representa uma variação esperada de R\$ -1,532 reais no *limite usado médio* a cada mês acrescido na *idade de conta* mantendo-se fixas as demais covariáveis;
- $\widehat{\beta}_3$: representa uma variação esperada de R\$ -0,644 reais no *limite usado médio* quando aumenta-se 1 mês no *tempo de alteração do limite* mantendo-se fixas as demais covariáveis;
- $\widehat{\beta}_4$: representa a variação esperada do valor de *limite usado médio* em R\$ -1,634 reais quando aumenta-se 1 real na *média do valor de fatura*, mantendo fixas as demais covariáveis.
- $\widehat{\beta}_5$: representa a variação esperada do valor de *limite usado médio* em R\$ -0,404 reais a quando o *valor de compras à vista* é acrescido em 1 real, mantendo-se fixas as demais covariáveis.
- $\widehat{\beta}_6$: representa a variação esperada no *limite usado médio* de R\$ -0,177 reais à medida em que o *limite combinado* é acrescido de uma 1 real, mantendo-se fixas as demais covariáveis.
- $\widehat{\beta}_7$: variação esperada no *limite usado médio* de aproximadamente R\$ 1 real a cada unidade acrescida no valor de rank vendas, mantendo-se fixas as demais covariáveis;

De forma semelhante ao que podemos observar no modelo gama com função de ligação identidade, alguns parâmetros parecem não ter contribuição adequada para explicar a variável resposta se analisarmos de forma isolada, como por exemplo as variáveis

média do valor de faturas, valor de compras à vista e limite combinado, que acompanham os variáveis β_4 , β_5 e β_6 , respectivamente, para o modelo normal inverso com função de ligação identidade. No entanto, se analisarmos de forma isolada a estimativa na presença das demais variáveis explicativas pode ser que a contribuição seja adequada para explicar a variável resposta.

5 CONCLUSÕES E TRABALHOS FUTUROS

Os modelos lineares generalizados foram considerados para modelar o limite de crédito médio utilizado, em particular os modelos gama e normal inverso, dado que a variável resposta possuía suporte nos números reais positivos e apresentava característica de assimetria. Para o ajuste dos modelos, foram consideradas as funções de ligação identidade, recíproca, recíproca quadrática e logarítmica, obtendo-se resultados satisfatórios após a eliminação de algumas observações de forma conjunta. A função de ligação identidade foi a mais adequada para explicar o limite usado médio de clientes para a concessão de crédito com base nos resultados obtidos através da análise de diagnóstico.

É importante salientar a importância dos modelos gama e normal inverso, que são pertencente a classe dos MLG's frente aos modelos normais lineares, uma vez que a suposição de normalidade dos erros e homogeneidade de variâncias precisavam ser satisfeitas. Notamos que essa metodologia flexibiliza o uso de outras distribuições como as utilizadas no texto que explicam muito bem o fenômeno de estudo. Sendo assim, o uso da metodologia apresentada neste texto contribuiu para estudos envolvendo limite e concessão de crédito. O *software* R foi de grande importância para as análises apresentadas neste trabalho, uma vez que possui rotinas computacionais próprias para a análise dos MLG's, oferecendo forte apoio computacional com destaque para as rotinas computacionais oferecidas por Paula (2013).

Para trabalhos futuros, outras distribuições que poderiam ser utilizadas para o ajuste de dados cuja a variável resposta apresenta suporte nos números reais positivos e tenha característica de assimetria, são as distribuições Weibull, Log-normal, Pareto, Birnbaum-Saunders e Beta Prime. Outra proposta seria elaborar um plano amostral mais bem especificado, dado que a amostra utilizada foi de todo o ano de 2021 não levando em conta as informações e as características próprias de cada mês, nem fatores externos como a pandemia, que afetou bastante o comportamento de compras de todos os clientes. Neste caso, seria interessante selecionar uma amostra considerando tais fatores ou criando estratos para se ter uma amostra que seja mais representativa da população.

REFERÊNCIAS

- BAIN, L. J.; ENGELHARDT, M. A two-moment chi-square approximation for the statistic $\log(\bar{x}/\tilde{x})$. **Journal of the American Statistical Association**, v. 70, n. 352, p. 948–950, 1975.
- BOLFARINE, H.; SANDOVAL, M. **Introdução à Inferência Estatística**. 2. ed. Rio de Janeiro: Sociedade Brasileira de Matemática (SBM), 2010.
- CORDEIRO, D. B. **Ajuste de Modelos Lineares Generalizados Para Dados Positivos Assimétricos (manuscrito)**. Paraíba, 2016. 70 p.
- CORDEIRO, G. M.; DEMÉTRIO, C. G. B. **Modelos Lineares Generalizados e Extensões**. Piracicaba: [s.n.], 2008.
- COX, D. R.; SNELL, E. J. A general definition of residual (with discussion). **Journal of the Royal Statistical Society**, v. 30, n. 2, p. 248–175, 1968.
- DEMÉTRIO, C. G. B. **Modelos Lineares Generalizados em Experimentação Agronômica**. Piracicaba: [s.n.], 2001.
- JORGENSEN, B. Exponential dispersion models (with discussion). **Journal of the Royal Statistical Society**, v. 49, p. 127–162, 1987.
- NELDER, J. A.; MCCULLAGH, P. **Generalized Linear Models**. 2. ed. New York: Chapman and Hall, 1989.
- NELDER, J. A.; WEDDERBURN, R. W. M. Generalized linear models. **Journal of the Royal Statistical Society**, v. 135, No. 3, p. 370–384, 1972.
- PAULA, G. A. **Modelos de regressão com apoio computacional**. São Paulo: Instituto de Matemática e Estatística (IME), 2013.
- SINGER, J.; NOBRE, J.; ROCHA, F. **Análise de Dados Longitudinais: Versão parcial preliminar**. São Paulo: [s.n.], 2017.
- TURKMAN, M. A. A.; SILVA, G. L. **Modelos lineares generalizados: da teoria à prática**. Lisboa: [s.n.], 2000.
- WILLIAMS, D. Residuals in generalized linear models: Versão parcial preliminar. **Proceedings of the 12th. International Biometrics Conference**, Tokyo, p. 59–68, 1984.

APÊNDICE A – FUNÇÃO GERADORA DE MOMENTOS E CUMULANTES E PROPRIEDADES DAS DISTRIBUIÇÕES GAMA E NORMAL INVERSA

Função geradora de momentos e cumulantes

Seja Y uma variável aleatória com função de densidade como dada em (2.2), sua função geradora de momentos (f.g.m.) é dada por

$$M_Y(t; \theta, \phi) = \mathbb{E}[e^{tY}] = \exp \left\{ \phi [b(\phi^{-1}t + \theta) - b(\theta)] \right\}. \quad (\text{A.1})$$

A demonstração de (A.1) pode ser encontrada em Cordeiro e Demétrio (2008). A f.g.c. é dada por $\varphi(t; \theta, \phi) = \log [M_Y(t; \theta, \phi)] = \phi [b(\phi^{-1}t + \theta) - b(\theta)]$, sendo esta o logaritmo natural da f.g.m. Deste modo, o r -ésimo cumulante pode ser obtido derivando sucessivas vezes em relação a t , obtendo-se a expressão geral

$$k_r = \phi^{1-r} b^{(r)}(\phi^{-1}t + \theta),$$

em que $b^{(r)}(\cdot)$ indica a r -ésima derivada de $b(\cdot)$ em relação a t . Para $t = 0$, o r -ésimo cumulante da família exponencial de dispersão é dado por

$$k_r = \phi^{1-r} b^{(r)}(\theta). \quad (\text{A.2})$$

A partir de (A.2) obtemos $k_1 = \mathbb{E}[Y] = \mu = b^{(1)}(\theta)$ e $k_2 = \text{Var}[Y] = \phi^{-1} b^{(2)}(\theta) = \phi^{-1} V(\mu)$.

Propriedades da distribuição gama

i) **Esperança:** Temos para a distribuição gama que

$$\mathbb{E}[Y] = b'(\theta) = \frac{d}{d\theta} [-\ln(-\theta)] = -\frac{-1}{-\theta} = -\frac{1}{\theta}.$$

Mas $\theta = -1/\mu$, logo,

$$\mathbb{E}[Y] = b'(\theta) = \frac{1}{1/\mu} = \mu.$$

ii) **Variância:** Obtendo $b''(\theta)$ temos

$$b''(\theta) = \frac{d}{d\theta}(-\theta^{-1}) = \frac{1}{\theta^2} = \frac{1}{1/\mu^2} = \mu^2.$$

Deste modo, obtemos que a variância de Y será dada por

$$\text{Var}[Y] = b''(\theta)\phi^{-1} = \frac{\mu^2}{\phi}.$$

iii) **Moda:** Seja Y reescrita como

$$f(y; \mu, \phi) = \exp \left\{ \phi \left(-\frac{y}{\mu} - \ln(\mu) \right) + (\phi - 1) \ln(y) + \phi \ln(\phi) - \ln(\Gamma(\phi)) \right\}.$$

Sabe-se que ponto que maximiza $f(y; \mu, \phi)$ é o mesmo que maximiza $h(y; \mu, \phi) = \ln(f(y; \mu, \phi))$, e deste modo temos

$$h(y; \mu, \phi) = \phi \left(-\frac{y}{\mu} - \ln(\mu) \right) + (\phi - 1) \ln(y) + \phi \ln(\phi) - \ln(\Gamma(\phi)),$$

que tomando a derivada em relação à y obtemos

$$\frac{\partial}{\partial y} h(y; \mu, \phi) = \frac{\phi - 1}{y} - \frac{\phi}{\mu}$$

Igualando esta derivada à zero temos e resolvendo para y obtemos

$$y = \mu - \frac{\mu}{\phi},$$

que é o ponto que maximiza $f(y; \mu, \phi)$.

iv) **Função geradora de momentos:**

Utilizando o resultado apresentado para funções geradoras, temos para a distribuição gama que

$$b(\phi^{-1}t + \theta) = -\ln \left[-(\phi^{-1}t - 1/\mu) \right],$$

e deste modo temos que a f.g.m pode ser expressa como

$$M_Y(t; \mu, \phi) = \exp \left\{ \phi \left[-\ln \left(\frac{1}{\mu} - \frac{t}{\phi} \right) - \ln(\mu) \right] \right\}$$

$$\begin{aligned}
&= \exp \left\{ -\phi \left[\ln \left(\frac{1}{\mu} - \frac{t}{\phi} \right) + \ln(\mu) \right] \right\} \\
&= \exp \left\{ -\phi \ln \left[\left(\frac{1}{\mu} - \frac{t}{\phi} \right) \mu \right] \right\} \\
&= \exp \left\{ \ln \left[\left(1 - \frac{t\mu}{\phi} \right) \right]^{-\phi} \right\} \\
&= \left(1 - \frac{t\mu}{\phi} \right)^{-\phi}, \quad t < \phi/\mu.
\end{aligned}$$

v) **Função geradora de cumulantes:** A função geradora de cumulantes para a distribuição gama é dada por

$$\mathcal{K}(t; \mu, \phi) = \log \left[\left(1 - \frac{\mu t}{\phi} \right)^{-\phi} \right] = -\phi \log(1 - \mu t/\phi), \quad t < \phi/\mu.$$

Derivando em relação à t e avaliando em $t = 0$ obtemos a esperança da variável aleatória Y :

$$k_1 = \frac{\partial}{\partial t} \mathcal{K}(t; \mu, \phi) = -\phi \left(\frac{-\mu/\phi}{1 - \mu t/\phi} \right) = (\mu\phi)(\phi - \mu t)^{-1} \Big|_{t=0} = \mu.$$

Para obter a variância, temos

$$k_2 = \frac{\partial^2}{\partial t^2} \mathcal{K}(t; \mu, \phi) = \mu^2 \phi (\phi - \mu t)^{-2} \Big|_{t=0} = \frac{\mu^2}{\phi}.$$

O terceiro momento central é dado pelo cumulante de ordem 3, isto é

$$k_3 = \frac{\partial^3}{\partial t^3} \mathcal{K}(t; \mu, \phi) = 2\phi\mu^3 (\phi - \mu t)^{-3} \Big|_{t=0} = \frac{2\mu^3}{\phi^2}.$$

O quarto momento central é dado pelo cumulante de ordem 4, isto é

$$k_4 = \frac{\partial^4}{\partial t^4} \mathcal{K}(t; \mu, \phi) = 6\phi\mu^4 (\phi - \mu t)^{-4} \Big|_{t=0} = \frac{6\mu^4}{\phi^3}.$$

De forma geral, o r -ésimo cumulante é dado por

$$k_r = \frac{\partial^{(r)}}{\partial t^{(r)}} = (r-1)! \mu^r \phi (\phi - \mu t)^{-r}.$$

Avaliando k_r no ponto $t = 0$, obtemos a expressão geral

$$k_r \Big|_{t=0} = \frac{(r-1)! \mu^r}{\phi^{r-1}} \text{ ou } (r-1)! \mu^r \phi^{1-r}.$$

vi) **Assimetria:** A assimetria da distribuição gama tal como foi definida, é dada por

$$A(y) = \frac{k_3}{k_2^{3/2}} = \frac{\left(\frac{2\mu^3}{\phi^2}\right)}{\left(\frac{\mu^2}{\phi}\right)^{3/2}} = \frac{2}{\sqrt{\phi}}.$$

vii) **Curtose:** A curtose é dada por

$$K(y) = \frac{k_4}{k_2^2} - 3 = \frac{\left(\frac{6\mu^4}{\phi^3}\right)}{\left(\frac{\mu^2}{\phi}\right)^2} - 3 = \frac{6}{\phi} - 3.$$

Propriedades da distribuição normal inversa

i) **Esperança:** Temos

$$\mathbb{E}[Y] = b'(\theta) = (-2)(-1/2)(-2\theta)^{-1/2} = (-2\theta)^{-1/2},$$

mas $\theta = -1/2\mu^2$ e portanto

$$\mathbb{E}[Y] = \left[-2 \left(-\frac{1}{2\mu^2} \right) \right]^{-1/2} = (\mu^{-2})^{-1/2} = \mu.$$

ii) **Variância:** Primeiro obtemos $b''(\theta)$:

$$b''(\theta) = (-2\theta)^{-3/2} = \left[-2 \left(-\frac{1}{2\mu^2} \right) \right]^{-3/2} = (\mu^{-2})^{-3/2} = \mu^3.$$

Deste modo temos que a variância de Y será dada por

$$\text{Var}[Y] = \phi^{-1}b''(\theta) = \frac{\mu^3}{\phi},$$

sendo $V(\mu) = \mu^3$ a função de variância associada à função de distribuição normal inversa.

iv) **Moda:**

Seja Y reescrita na família exponencial de dispersão. Deste modo temos que

$$\begin{aligned} f(y; \mu, \phi) &= \exp \left\{ \phi \left(-\frac{y}{2\mu^2} + \frac{1}{\mu} \right) - \frac{1}{2} \left[\log(2\pi y^3/\phi) + \frac{\phi}{y} \right] \right\} \\ &= \exp \left\{ -\frac{\phi y}{2\mu^2} + \frac{\phi}{\mu} - \frac{1}{2} \log(2\pi) - \frac{3}{2} \log(y) + \frac{1}{2} \log(\phi) - \frac{\phi}{2y} \right\}. \end{aligned}$$

Sabe-se que ponto que maximiza $f(y; \mu, \phi)$ é o mesmo que maximiza $h(y; \mu, \phi) = \ln(f(y; \mu, \phi))$, e deste modo temos

$$h(y; \mu, \phi) = -\frac{\phi y}{2\mu^2} + \frac{\phi}{\mu} - \frac{1}{2} \log(2\pi) - \frac{3}{2} \log(y) + \frac{1}{2} \log(\phi) - \frac{\phi}{2y},$$

que tomando a derivada em relação à y obtemos

$$\frac{\partial}{\partial y} h(y; \mu, \phi) = -\frac{\phi}{2\mu^2} - \frac{3}{2y} + \frac{\phi}{2y^2}$$

Igualando esta derivada à zero temos

$$\begin{aligned} \implies -\frac{\phi}{2\mu^2} - \frac{3}{2y} + \frac{\phi}{2y^2} &= 0 \\ \implies -\frac{\phi}{\mu^2} - \frac{3}{y} + \frac{\phi}{y^2} &= 0 \\ \implies -\phi y^2 - 3\mu^2 y + \phi \mu^2 &= 0 \end{aligned}$$

Basta agora resolver esta equação de segundo grau:

$$\begin{aligned} y &= \frac{3\mu^2 \pm \sqrt{9\mu^4 + 4\phi^2\mu^2}}{-2\phi} \\ &= \frac{3\mu^2 \pm \sqrt{\mu^2(9\mu^2 + 4\phi^2)}}{-2\phi} \end{aligned}$$

$$\begin{aligned}
&= -\frac{3\mu^2}{2\phi} \pm \mu \sqrt{\frac{9\mu^2 + 4\phi^2}{4\phi^2}} \\
&= \mu \left[\left(1 + \frac{9\mu^2}{4\phi^2}\right)^{1/2} - \frac{3\mu}{2\phi} \right].
\end{aligned}$$

Para que o máximo exista, basta que $\mu > 0$ e $\phi > 0$, o que já é garantido pelo próprio espaço paramétrico da distribuição normal inversa.

v) **Função geradora de momentos:**

Para a distribuição normal inversa, obtemos anteriormente que $\theta = -1/2\mu^2$, $b(\theta) = -1/\mu = -(-2\theta)^{1/2}$, e deste modo

$$b(\phi^{-1}t + \theta) = - \left[-2 \left(\frac{t}{\phi} - \frac{1}{2\mu^2} \right) \right]^{1/2} = - \left[\frac{1}{\mu^2} - \frac{2t}{\phi} \right]^{1/2},$$

e dessa forma temos que

$$M_Y(t; \mu, \phi) = \exp \left\{ \phi \left(\frac{1}{\mu} - \left[\frac{1}{\mu^2} - \frac{2t}{\phi} \right]^{1/2} \right) \right\} = \exp \left\{ \frac{\phi}{\mu} - \phi \left[\frac{1}{\mu^2} - \frac{2t}{\phi} \right]^{1/2} \right\}.$$

vi) **Função geradora de cumulantes:**

Podemos obter a função geradora de cumulantes como sendo

$$\mathcal{K}(t; \mu, \phi) = \log(M_Y(t; \theta, \phi)) = \frac{\phi}{\mu} - \phi \left[\frac{1}{\mu^2} - \frac{2t}{\phi} \right]^{1/2}.$$

Derivando em relação à t e avaliando em $t = 0$ obtemos a esperança da variável aleatória Y :

$$k_1 = \frac{\partial}{\partial t} \mathcal{K}(t; \mu, \phi) = -\frac{\phi}{2} \left[\frac{1}{\mu^2} - \frac{2t}{\phi} \right]^{-1/2} \left(-\frac{2}{\phi} \right) \Big|_{t=0} = (\mu^{-2})^{-1/2} = \mu.$$

Para obter a variância temos que

$$k_2 = \frac{\partial^2}{\partial t^2} \mathcal{K}(t; \mu, \phi) = -\frac{1}{2} \left[\frac{1}{\mu^2} - \frac{2t}{\phi} \right]^{-3/2} \left(-\frac{2}{\phi} \right) \Big|_{t=0} = \frac{(\mu^{-2})^{-3/2}}{\phi} = \frac{\mu^3}{\phi}.$$

O terceiro momento central é dado pelo cumulante de ordem 3, isto é

$$k_3 = \frac{\partial^3}{\partial t^3} \mathcal{K}(t; \mu, \phi) = -\frac{3}{2\phi} \left[\frac{1}{\mu^2} - \frac{2t}{\phi} \right]^{-5/2} \left(-\frac{2}{\phi} \right) = \frac{3}{\phi^2} \left[\frac{1}{\mu^2} - \frac{2t}{\phi} \right]^{-5/2} \Big|_{t=0} = \frac{3}{\phi^2} \mu^5.$$

O quarto momento central é dado pelo cumulante de ordem 4, isto é

$$k_4 = \frac{\partial^4}{\partial t^4} \mathcal{K}(t; \mu, \phi) = \left(-\frac{5}{2}\right) \frac{3}{\phi^2} \left[\frac{1}{\mu} - \frac{2t}{\phi}\right]^{-7/2} \left(-\frac{2}{\phi}\right) \Big|_{t=0} = \frac{15\mu^7}{\phi^3}.$$

vii) **Assimetria:**

A assimetria de uma função densidade Y é dada por

$$A(y) = \frac{k_3}{k_2^{3/2}}.$$

Deste modo obtemos que

$$A(y) = \left(\frac{3\mu^5}{\phi^2}\right) / \left(\frac{\mu^3}{\phi}\right)^{3/2} = 3 \left(\frac{\mu}{\phi}\right)^{1/2}.$$

viii) **Curtose:**

Para a distribuição normal inversa temos que a curtose denotada por $K(y)$ é dada por

$$K(y) = \frac{k_4}{k_2^2} - 3 = \frac{15\mu^7}{\phi^3} \frac{\phi^2}{\mu^6} - 3 = \frac{15\mu}{\phi} - 3.$$

APÊNDICE B – FUNÇÃO ESCORE DE β

Tomando a derivada da função escore em relação ao j -ésimo parâmetro, temos

$$\frac{\partial l(\boldsymbol{\theta}, \mathbf{y})}{\partial \beta_j} = \phi \sum_{i=1}^n \left(y_i \frac{\partial \theta_i}{\partial \beta_j} - \frac{\partial b(\theta_i)}{\partial \beta_j} \right).$$

Note por exemplo que θ_i depende μ_i que depende de η_i que por sua vez depende de β_j da seguinte forma: $\eta_i = \mathbf{x}_i^\top \boldsymbol{\beta}$ sendo η_i o nosso preditor linear que por sua vez relaciona-se com a média μ_i através de uma função de ligação $g(\mu_i) = \eta_i$ e existe uma função que relaciona $\theta_i = \mu_i$ para a família exponencial de dispersão. Assim temos

$$\frac{\partial l(\boldsymbol{\theta}, \mathbf{y})}{\partial \beta_j} = \phi \sum_{i=1}^n \left(y_i \frac{d\theta_i}{d\mu_i} \frac{d\mu_i}{d\eta_i} \frac{d\eta_i}{d\beta_j} - \frac{db(\theta_i)}{d\theta_i} \frac{d\theta_i}{d\mu_i} \frac{d\mu_i}{d\eta_i} \frac{d\eta_i}{d\beta_j} \right).$$

Sabemos que

$$V = V(\mu) = \frac{d\mu}{d\theta} \implies V^{-1} = \frac{d\theta}{d\mu},$$

então temos

$$\frac{\partial l(\boldsymbol{\theta}, \mathbf{y})}{\partial \beta_j} = \phi \sum_{i=1}^n \left(y_i V_i^{-1} \frac{d\mu_i}{d\eta_i} \frac{d\eta_i}{d\beta_j} - \frac{db(\theta_i)}{d\theta_i} V_i^{-1} \frac{d\mu_i}{d\eta_i} \frac{d\eta_i}{d\beta_j} \right).$$

No entanto, também podemos perceber que

$$\frac{d\eta_i}{d\beta_j} = \frac{d(\mathbf{x}_i^\top \boldsymbol{\beta})}{d\beta_j} = \frac{d}{d\beta_j} \left(\sum_{j=1}^p x_{ij} \beta_j \right) = \frac{d}{d\beta_j} (x_{i1} \beta_1 + x_{i2} \beta_2 + \dots + x_{ij} \beta_j + \dots + x_{ip} \beta_p) = x_{ij}.$$

e

$$\frac{db(\theta_i)}{d\theta_i} = \mu_i.$$

Simplificando obtemos

$$\frac{\partial l(\boldsymbol{\theta}, \mathbf{y})}{\partial \beta_j} = \phi \sum_{i=1}^n \left(y_i V_i^{-1} \frac{d\mu_i}{d\eta_i} x_{ij} - \mu_i V_i^{-1} \frac{d\mu_i}{d\eta_i} x_{ij} \right) = \phi \sum_{i=1}^n \left(V_i^{-1} \frac{d\mu_i}{d\eta_i} (y_i - \mu_i) x_{ij} \right).$$

Vamos reorganizar o termo

$$\frac{d\mu_i}{d\eta_i} \frac{1}{V_i} = \sqrt{\left(\frac{d\mu_i}{d\eta_i}\right)^2 \frac{1}{V_i^2}} = \sqrt{\frac{\left(\frac{d\mu_i}{d\eta_i}\right)^2 \frac{1}{V_i}}{V_i}},$$

e definir

$$w_i = \left(\frac{d\mu_i}{d\eta_i}\right)^2 \frac{1}{V_i}.$$

Logo, temos que

$$\frac{\partial l(\boldsymbol{\theta}, \mathbf{y})}{\partial \beta_j} = \phi \sum_{i=1}^n \left(\sqrt{\frac{w_i}{V_i}} (y_i - \mu_i) x_{ij} \right).$$

Finalmente, podemos obter uma forma matricial para a função escore como sendo

$$\mathbf{U}_\beta(\boldsymbol{\theta}) = \phi \mathbf{X}^\top \mathbf{W}^{1/2} \mathbf{V}^{-1/2} (\mathbf{y} - \boldsymbol{\mu}).$$

Para obter a matriz de informação de Fisher tomamos a segunda derivada da função

$$\begin{aligned} \frac{\partial l(\boldsymbol{\theta}, \mathbf{y})}{\partial \beta_j \partial \beta_l} &= \phi \sum_{i=1}^n (y_i - \mu_i) \frac{d^2 \theta_i}{d\mu_i^2} \left(\frac{d\mu_i}{d\eta_i} \right)^2 x_{ij} x_{il} + \phi \sum_{i=1}^n (y_i - \mu_i) \frac{d^2 \theta_i}{d\mu_i} \frac{d^2 \mu_i}{d\eta_i^2} x_{ij} x_{il} \\ &\quad - \phi \sum_{i=1}^n \frac{d\theta_i}{d\mu_i} \left(\frac{d\mu_i}{d\eta_i} \right)^2 x_{ij} x_{il}, \end{aligned}$$

de modo que

$$\begin{aligned} \mathbb{E} \left[\frac{\partial l(\boldsymbol{\theta}, \mathbf{y})}{\partial \beta_j \partial \beta_l} \right] &= -\phi \sum_{i=1}^n \frac{d\theta_i}{d\mu_i} \left(\frac{d\mu_i}{d\eta_i} \right)^2 x_{ij} x_{il} \\ &= -\phi \sum_{i=1}^n \left(\frac{d\theta_i}{d\eta_i} \right) \frac{1}{V_i} x_{ij} x_{il} \\ &= -\phi \sum_{i=1}^n w_i x_{ij} x_{il}. \end{aligned}$$

Portanto, podemos escrever a informação de Fisher para $\boldsymbol{\beta}$ na forma matricial como sendo

$$\mathbf{K}_{\beta\beta}(\boldsymbol{\theta}) = \mathbb{E} \left[- \frac{\partial^2 l(\boldsymbol{\theta}, \mathbf{y})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} \right] = \boldsymbol{\phi} \mathbf{X}^\top \mathbf{W} \mathbf{X}.$$

APÊNDICE C – ESTIMAÇÃO DOS PARÂMETROS

Podemos reescrever a Equação 3.5 multiplicando ambos os lados da equação por $\mathbf{K}_{\beta\beta}^{(m)}$:

$$\begin{aligned}\mathbf{K}_{\beta\beta}^{(m)}\boldsymbol{\beta}^{(m+1)} &= \mathbf{K}_{\beta\beta}^{(m)}\boldsymbol{\beta}^{(m)} + \mathbf{K}_{\beta\beta}^{(m)}\left[\mathbf{K}_{\beta\beta}^{-1}\right]^{(m)}\mathbf{U}_{\beta}^{(m)} \\ \mathbf{K}_{\beta\beta}^{(m)}\boldsymbol{\beta}^{(m+1)} &= \mathbf{K}_{\beta\beta}^{(m)}\boldsymbol{\beta}^{(m)} + \mathbf{U}_{\beta}^{(m)},\end{aligned}$$

no entanto, sabemos que $\mathbf{K}_{\beta\beta} = \phi\mathbf{X}^{\top}\mathbf{W}\mathbf{X}$ e $\mathbf{U}_{\beta} = \phi\mathbf{X}^{\top}\mathbf{W}^{\frac{1}{2}}\mathbf{V}^{-\frac{1}{2}}(\mathbf{y} - \boldsymbol{\mu})$, e deste modo temos

$$\begin{aligned}\phi\mathbf{X}^{\top}\mathbf{W}^{(m)}\mathbf{X}\boldsymbol{\beta}^{(m+1)} &= \phi\mathbf{X}^{\top}\mathbf{W}^{(m)}\mathbf{X}\boldsymbol{\beta}^{(m)} + \phi\mathbf{X}^{\top}\left[\mathbf{W}^{\frac{1}{2}}\right]^{(m)}\left[\mathbf{V}^{-\frac{1}{2}}\right]^{(m)}(\mathbf{y} - \boldsymbol{\mu}^{(m)}) \\ \phi\mathbf{X}^{\top}\mathbf{W}^{(m)}\mathbf{X}\boldsymbol{\beta}^{(m+1)} &= \phi\mathbf{X}^{\top}\mathbf{W}^{(m)}\underbrace{\mathbf{X}\boldsymbol{\beta}^{(m)}}_{\boldsymbol{\eta}^{(m)}} + \phi\mathbf{X}^{\top}\underbrace{\left[\mathbf{W}^{\frac{1}{2}}\right]^{(m)}\left[\mathbf{W}^{\frac{1}{2}}\right]^{(m)}}_{\mathbf{W}^{(m)}}\left[\mathbf{W}^{-\frac{1}{2}}\right]^{(m)}\left[\mathbf{V}^{-\frac{1}{2}}\right]^{(m)}(\mathbf{y} - \boldsymbol{\mu}^{(m)}) \\ \mathbf{X}^{\top}\mathbf{W}^{(m)}\mathbf{X}\boldsymbol{\beta}^{(m+1)} &= \mathbf{X}^{\top}\mathbf{W}^{(m)}\boldsymbol{\eta}^{(m)} + \mathbf{X}^{\top}\mathbf{W}^{(m)}\left[\mathbf{W}^{-\frac{1}{2}}\right]^{(m)}\left[\mathbf{V}^{-\frac{1}{2}}\right]^{(m)}(\mathbf{y} - \boldsymbol{\mu}^{(m)}) \\ \mathbf{X}^{\top}\mathbf{W}^{(m)}\mathbf{X}\boldsymbol{\beta}^{(m+1)} &= \mathbf{X}^{\top}\mathbf{W}^{(m)}\left(\boldsymbol{\eta}^{(m)} + \left[\mathbf{W}^{-\frac{1}{2}}\right]^{(m)}\left[\mathbf{V}^{-\frac{1}{2}}\right]^{(m)}(\mathbf{y} - \boldsymbol{\mu}^{(m)})\right).\end{aligned}$$

Denotaremos $\mathbf{z}^{(m)} = \boldsymbol{\eta}^{(m)} + \left[\mathbf{W}^{-\frac{1}{2}}\right]^{(m)}\left[\mathbf{V}^{-\frac{1}{2}}\right]^{(m)}(\mathbf{y} - \boldsymbol{\mu}^{(m)})$ e sendo assim temos

$$\begin{aligned}\mathbf{X}^{\top}\mathbf{W}^{(m)}\mathbf{X}\boldsymbol{\beta}^{(m+1)} &= \mathbf{X}^{\top}\mathbf{W}^{(m)}\mathbf{z}^{(m)} \\ \boldsymbol{\beta}^{(m+1)} &= \left(\mathbf{X}^{\top}\mathbf{W}^{(m)}\mathbf{X}\right)^{-1}\mathbf{X}^{\top}\mathbf{W}^{(m)}\mathbf{z}^{(m)},\end{aligned}$$

demonstrando assim o resultado procurado.

APÊNDICE D – FUNÇÃO DIGAMA

Sabemos que a função gama é definida por

$$\Gamma(\alpha) = \int_0^{\infty} u^{\alpha-1} e^{-u} du.$$

Derivando $\Gamma(\alpha)$ em relação a α obtemos

$$\Gamma'(\alpha) = \int_0^{\infty} \ln(u) u^{\alpha-1} e^{-u} du.$$

A função digama é definida por

$$\Psi(\alpha) = (\ln[\Gamma(\alpha)])' = \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}.$$

Pode-se notar que $\Psi'(1) = \Gamma'(1)$ pois $\Gamma(1) = 1$. Pode-se achar derivadas de ordens superiores de ordem r como por exemplo

$$\Gamma^r(\alpha) = \int_0^{\infty} [\ln(u)]^r u^{\alpha-1} e^{-u} du.$$

As derivadas de ordem r da função $\Gamma(\alpha)$ são realizadas através de $\Psi^r(\alpha)$. Sabe-se que $\Gamma^{(1)}(\alpha) = \Psi(\alpha)\Gamma(\alpha)$. Derivando a segunda vez utilizando a regra do produto, obtemos

$$\Gamma^{(2)}(\alpha) = \Psi^{(1)}(\alpha)\Gamma(\alpha) + \Psi(\alpha)\Gamma^{(1)}(\alpha).$$

APÊNDICE E – VARIÁVEIS DO BANCO DE DADOS DE ANÁLISE DE LIMITE DE CRÉDITO

Tabela 13 – Descrição de todas as variáveis

Índice	Variável	Descrição
y	Limite usado médio	Média nos últimos quatro meses do saldo total (variável resposta)
x_1	Behavior Score	Pontuação de comportamento de mercado de um cliente (limitada entre 0 e 100)
x_2	Valor do limite antigo	Valor de limite antes do aumento;
x_3	Valor do limite novo	Valor do limite pós aumento;
x_4	Aumento	Diferença entre o valor do limite novo e o valor do limite antigo
x_5	Aumento Percentual	Razão entre o aumento e o valor do limite antigo
x_6	Idade de conta	Idade em meses do cliente;
x_7	Tempo de Alteração	Tempo (em meses) até a ocorrência de aumento do limite;
x_8	Média de score 4 meses	Média dos últimos quatro meses dos valores de score;
x_9	Diferença de score	Diferença entre o valor de score mais recente e a média do valor de score nos últimos 4 meses
x_{10}	Contagem de fatura	Quantidade de faturas de um determinado cliente;
x_{11}	Média do valor de fatura	Valor médio das faturas de um determinado cliente;
x_{12}	Máximo do valor de fatura	Valor máximo dentre todas as faturas de um determinado cliente;
x_{13}	Máximo do valor de compras à vista	Valor máximo dentre as compras à vista de um determinado cliente;
x_{14}	Valor de compras à vista	Valor de compra à vista de um determinado cliente;
x_{15}	Quantidade de compras à vista	Quantidade de compras à vista realizada por um determinado cliente;
x_{16}	Limite combinado	Soma dos saldos totais de um determinado cliente; o saldo total é o valor extratado em uma determinada compra mais o valor de parcelas à vencer;
x_{17}	Quantidade de operações em instituições financeiras	Quantidade de instituições em que determinado cliente possui vínculo com cartão de crédito;
x_{18}	Tempo total em meses	Tempo total de um determinado cliente no SCR (Sistema de Informações de Crédito do Banco Central)
x_{19}	Valor de compras total	Total do valor de compras de um determinado cliente
x_{20}	Quantidade de compras total	Soma da quantidade de compras à vista e parceladas;
x_{21}	Percentual do limite usado médio	Razão entre o limite usado médio e o valor do limite antigo
x_{22}	Rentabilidade em 3 meses	Quantia em reais rentabilizada por um determinado cliente num período de três meses;
x_{23}	Rentabilidade em 6 meses	Quantia em reais rentabilizada por um determinado cliente num período de três meses;
x_{24}	Rentabilidade em 9 meses	Quantia em reais rentabilizada por um determinado cliente num período de três meses;
x_{25}	Máximo do valor de compras parcelado pós aumento	Valor máximo de todas as compras parceladas de um determinado cliente pós aumento de limite;
x_{26}	Valor de compras parcelado pós aumento	Valor da compra parcelada pós aumento de limite;
x_{27}	Quantidade de compras parceladas pós aumento	Quantidade de compras parceladas pós aumento de limite de um determinado cliente
x_{28}	Valor de compras total pós aumento	Valor total de compras pós aumento de um determinado cliente
x_{29}	Quantidade de compras total pós aumento	Quantidade de compras total pós aumento de um determinado cliente
x_{30}	Compras à vista relativo	Razão entre o valor de compras à vista e o valor do limite antigo
x_{31}	Máximo do valor de compras à vista relativo	Razão entre o máximo do valor da compra e o valor do limite antigo
x_{32}	Valor de fatura relativo	Razão entre a média do valor da fatura e o valor do limite antigo
x_{33}	Máximo do valor de fatura relativo	Razão entre o valor máximo de fatura e o valor do limite antigo
x_{34}	Rank vendas	Pontuação do quanto determinado cliente é ou não bom em vendas/compras; quanto maior for o valor de rank vendas melhor será determinado cliente em relação à vendas e quanto menor for o valor de rank vendas, pior será determinado cliente em vendas. Esta variável possui contradomínio pertencente aos números reais.
x_{35}	Tempo de tombamento	Tempo em meses a partir da data de tombamento (cartão de loja vira Mastercard)
x_{36}	Créditos à vencer	Valor total das parcelas à vencer, trazidas a valor presente.
x_{37}	Créditos a vencer (até 30 dias)	Valor total das parcelas à vencer até 30 dias, trazidas a valor presente.
x_{38}	Créditos a vencer (31 a 60 dias)	Valor total das parcelas à vencer de 31 até 60 dias, trazidas a valor presente.
x_{39}	Créditos a vencer (61 a 90 dias)	Valor total das parcelas à vencer de 61 até 90 dias, trazidas a valor presente.
x_{40}	Créditos a vencer (91 a 180 dias)	Valor total das parcelas à vencer de 91 até 180 dias, trazidas a valor presente.
x_{41}	Créditos a vencer (181 a 360 dias)	Valor total das parcelas à vencer de 181 até 360 dias, trazidas a valor presente.
x_{42}	Limite de crédito com vencimento até 360 dias (percentual)	Percentual do limite de crédito à vencer até 360 dias.
x_{43}	Limite de crédito com vencimento acima de 360 dias (percentual)	Percentual do limite de crédito à vencer acima de 360 dias.
x_{44}	Máximo de dias em atraso	Máximo dos dias de atraso de um determinado cliente
x_{45}	Máximo de compras à vista pós aumento	Máximo do valor de compras à vista
x_{46}	Valor de compras à vista pós aumento	Valor de compras à vista pós aumento
x_{47}	Quantidade de compras à vista pós aumento	Quantidade de compras à vista pós aumento
x_{48}	Valor total do limite de crédito (percentual)	Percentual do valor total de limite de crédito
x_{49}	Contagem de rotativos	Contagem de clientes que pagaram mais que valor mínimo da fatura e menos que o valor integral da fatura.
x_{50}	Percentual de rotativos	Razão entre a contagem de rotativos e a contagem de faturas
x_{51}	Máximo de compras parcelado	Valor máximo de compras parceladas de um determinado cliente
x_{52}	Valor de compras parcelado	Valor de compras parceladas de um determinado cliente
x_{53}	Quantidade de compras parcelado	Quantidade de compras parceladas de um determinado cliente
x_{54}	Percentual de compras parceladas	Percentual de compras parceladas de um determinado cliente
x_{55}	Compras parceladas relativa	Razão entre o valor mais recente de compras parceladas e o valor do limite antigo
x_{56}	Máximo do valor de compras parceladas relativa	Razão entre o máximo de compras parceladas e o valor do limite antigo
x_{57}	Máximo do valor rolado (valor de fatura em atraso)	Valor máximo das faturas em atraso de um determinado cliente
x_{58}	Mínimo do valor rolado (valor de fatura em atraso)	Valor mínimo das faturas em atraso de um determinado cliente
x_{59}	Percentual de FX1 (faturas em atraso)	Razão entre a contagem de faturas em atraso e a contagem de faturas
x_{60}	Quantidade de compras online	Quantidade de compras feitas de forma online

Tabela 14 – Variáveis excluídas do banco de dados, quantidade zeros, quantidade de valores faltantes e critério de retirada.

Índice	Variável	Quantidade de zeros	Quantidade de NA (valor ausente)	Critério de retirada
x_4	Aumento	0	0	Combinação linear de outras variáveis ($x_3 - x_2$)
x_5	Aumento Percentual	0	0	Combinação linear de outras variáveis (x_4/x_2)
x_9	Diferença de score	0	0	Combinação linear de outras variáveis ($x_1 - x_2$)
x_{18}	Tempo total em meses	0	0	Variável desatualizada no período de coleta
x_{19}	Valor de compras total	0	0	Combinação linear de outras variáveis ($x_{14} + x_{52}$)
x_{21}	Percentual do limite usado médio	0	0	Combinação linear de outras variáveis (y/x_2)
x_{28}	Valor de compras total pós aumento	0	0	Combinação linear de outras variáveis ($x_{26} + x_{46}$)
x_{30}	Compras à vista relativo	0	0	Combinação linear de outras variáveis (x_{14}/x_2)
x_{31}	Máximo do valor de compras à vista relativo	0	0	Combinação linear de outras variáveis (x_{13}/x_2)
x_{32}	Valor de fatura relativo	0	0	Combinação linear de outras variáveis (x_{11}/x_2)
x_{33}	Máximo do valor de fatura relativo	0	0	Combinação linear de outras variáveis (x_{12}/x_2)
x_{35}	Tempo de tombamento	0	10	Valores faltantes
x_{36}	Créditos à vencer	4	0	Variável desatualizada no período de coleta
x_{37}	Créditos a vencer (até 30 dias)	4	0	Variável desatualizada no período de coleta
x_{38}	Créditos a vencer (31 a 60 dias)	10	0	Variável desatualizada no período de coleta
x_{39}	Créditos a vencer (61 a 90 dias)	14	0	Variável desatualizada no período de coleta
x_{40}	Créditos a vencer (91 a 180 dias)	15	0	Variável desatualizada no período de coleta
x_{41}	Créditos a vencer (181 a 360 dias)	37	0	Variável desatualizada no período de coleta
x_{42}	Limite de crédito com vencimento até 360 dias (percentual)	37	0	Variável desatualizada no período de coleta
x_{43}	Limite de crédito com vencimento acima de 360 dias (percentual)	34	0	Variável desatualizada no período de coleta
x_{45}	Máximo de compras à vista pós aumento	0	32	Valores faltantes
x_{46}	Valor de compras à vista pós aumento	0	32	Valores faltantes
x_{47}	Quantidade de compras à vista pós aumento	0	32	Valores faltantes
x_{48}	Valor total do limite de crédito (percentual)	12	0	Variável desatualizada no período de coleta
x_{49}	Contagem de rotativos	0	46	Valores faltantes
x_{50}	Percentual de rotativos	0	46	Valores faltantes
x_{53}	Quantidade de compras parcelado	6	0	Incluída na variável x_{20}
x_{54}	Percentual de compras parceladas	6	0	Combinação linear de outras variáveis (x_{53}/x_{20})
x_{55}	Compras parceladas relativa	6	0	Combinação linear de outras variáveis (x_{52}/x_2)
x_{56}	Máximo do valor de compras parceladas relativa	6	0	Combinação linear de outras variáveis (x_{51}/x_2)
x_{57}	Máximo do valor rolado (valor de fatura em atraso)	73	0	Excesso de zeros
x_{58}	Mínimo do valor rolado (valor de fatura em atraso)	74	0	Excesso de zeros
x_{59}	Percentual de FX1 (faturas em atraso)	73	0	Excesso de zeros
x_{60}	Quantidade de compras online	87	0	Excesso de zeros

Tabela 15 – Variáveis preditoras utilizadas para a análise dos dados de limite de crédito médio.

Variável	Abreviação
Behavior Score	BS
Valor do limite antigo	VLA
Valor do limite novo	VLN
Idade de conta	IC
Tempo de Alteração	TA
Média de score 4 meses	MVE4
Contagem de fatura	CF
Média do valor de fatura	MVF
Valor máximo de fatura	VMF
Máximo do valor de compras à vista	MCV
Valor de compras à vista	VCV
Quantidade de compras à vista	QCV
Limite combinado	LC
Quantidade de operações em instituições financeiras	QIFOP
Quantidade de compras total	QCT
Rentabilidade em 3 meses	R3M
Rentabilidade em 6 meses	R6M
Rentabilidade em 9 meses	R9M
Máximo do valor de compras parcelado pós aumento	MCPos
Valor de compras parcelado pós aumento	VCPos
Quantidade de compras total pós aumento	QCPos
Máximo de dias em atraso	MDA
Máximo de compras parceladas	MCP
Valor de compras parceladas	VCP
Rank vendas	RKV