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

CAROLINE GOMES DUARTE

USO DAS DISTRIBUIÇÕES POISSON, POISSON-GAMA, POISSON-INVERSA GAUSSIANA E POISSON-LINDLEY GENERALIZADA PARA DADOS DE CONTAGEM

FORTALEZA
2021

CAROLINE GOMES DUARTE

USO DAS DISTRIBUIÇÕES POISSON, POISSON-GAMA, POISSON-INVERSA GAUSSIANA E POISSON-LINDLEY GENERALIZADA PARA DADOS DE CONTAGEM

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. Sílvia Maria de Freitas

FORTALEZA

2021

CAROLINE GOMES DUARTE

USO DAS DISTRIBUIÇÕES POISSON, POISSON-GAMA, POISSON-INVERSA GAUSSIANA E POISSON-LINDLEY GENERALIZADA PARA DADOS DE CONTAGEM

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

Prof$^{\underline{a}}$. Dra. Sílvia Maria de Freitas (Orientadora)
Universidade Federal do Ceará (UFC)

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

Prof. Dr. Juvêncio Santos Nobre
Universidade Federal do Ceará (UFC)

À Deus.

À minha família, Luciana, Atanael e Raphael.

AGRADECIMENTOS

Primeiramente agradeço a Deus, por moldar a pessoa que sou hoje e me conceder a oportunidade de entrar para o curso de Estatística de UFC.

Aos meus pais, Luciana e Atanael, por sempre acreditarem no meu potencial e por fazerem o possível e o impossível para me ajudar durante toda a minha vida. Ao meu irmão Raphael, o qual foi o responsável pelo meu ingresso no curso, e me motivou ao longo do mesmo. Aos meus familiares, em especial minha vó, Áurea Maria, que sempre me incentivou e apoiou.

À minha orientadora, Dra. Sílvia Maria, por me acolher tão bem como minha co-orientadora de pesquisa do PET, por me conceder a oportunidade de ser bolsista de Iniciação Científica e que com sua sabedoria e experiência me ensinou e auxiliou durante todas minhas bolsas.

Agradeço também ao Prof. Júlio Barros, pelas experiências e ensinamentos durante o período que participei do PET, que me ajudaram a crescer como pessoa e estudante.

À todos professores do curso de Estatística, em especial ao Prof. Dr. Maurício Mota, por ter me ensinado em várias disciplinas e pela colaboração, em especial na minha composição da monografia.

Aos componentes da minha banca, professores Juvêncio Nobre e Gualberto Agamez, por serem excelentes professores, pelo apoio e por sempre estarem a disposição para ensinar.

Aos meus amigos que cultivei durante o curso, Luan, Daniel, Keyliane e Matheus por me ajudarem durante todo o período de minha graduação e pelos momentos de desespero e alegria compartilhados. À minha colega de curso, Yohana, que foi a primeira pessoa a me acolher durante meu percurso e me incentivar a entrar para o PET.

Agradeço meus amigos do desporto que me receberam desde o primeiro ano de graduação, me proporcionando ótimos momentos durante toda a minha permanência no time de basquete feminino da UFC, seja através de vitórias ou derrotas.

"Eu posso aceitar falhas, mas não posso aceitar não tentar."

(Michael Jordan)

RESUMO

Os Modelos Lineares Generalizados (MLGs) surgiram como uma alternativa aos Modelos Lineares Clássicos quando a variável resposta não segue uma distribuição Normal. Caso o espaço paramétrico da variável aleatória assuma valores discretos e não negativos, tem-se um estudo para dados de contagem. A análise comum para estes dados é através de uma distribuição de Poisson, Binomial ou Binomial Negativa, via MLG. Entretanto, um dos cuidados que se deve ter ao fazer a análise de dados discretos, é com a superdispersão. Esse termo é utilizado, na área da Estatística, quando estamos trabalhando com dados de contagem, em que a média da variável resposta é maior que a variância, segundo o modelo proposto. Dessa forma, a utilização de um modelo com base apenas na distribuição Poisson, que tem como suposição a equidispersão, isto é, média igual à variância, poderia ser bem sensível e não muito útil, na presença de superdispersão. Uma alternativa para dados com essa característica é o uso das distribuições compostas, através dos modelos em dois estágios, ou hierárquicos, como uma forma para modelar essa superdispersão. A metodologia dos modelos em dois estágios associa, uma distribuição à resposta condicionada a sua média e, posteriormente, uma distribuição ao parâmetro de média, de forma que, incondicionalmente, se tem uma distribuição composta para a variável resposta. Neste trabalho é utilizada a distribuição clássica de Poisson, para dados de contagem, e as distribuições Gama, Inversa Gaussiana e Lindley Generalizada para o parâmetro de média da Poisson, que implicam nas distribuições compostas Poisson-Gama, Poisson-Inversa Gaussiana e a Poisson-Lindley Generalizada. Assim, o objetivo principal deste trabalho é apresentar esses modelos hierárquicos, ou modelos em dois estágios, que permitem a modelagem de dados de contagem com superdispersão. Também foram abordados alguns tipos de resíduos da estrutura dos MLGs (Pearson, Componente da *deviance* e Quantílicos), adaptados para as distribuições compostas utilizadas na monografia. A performance de cada abordagem foi realizada usando-se conjuntos de dados reais descritos na literatura.
Palavras-chave: Dados de Contagem, Superdispersão, Distribuição Poisson, Distribuição Compostas.

ABSTRACT

The Generalized Linear Models (GLM) is an alternative to Normal Models when the response variable does not come from a normal distribution. When the parametric space of the the variable of interest assumes discrete and non-negative values, we study the Count Models. The common approach when analyzing these data is through a Poisson distribution. However, one of the precautions to be taken when analyzing discrete data is overdispersion. This term is used, in the area of Statistics, when the average of the variance of the response variable exceeds the nominal variance predicted by the proposed model. Thus, the use of a model based on the Poisson distribution, which assumes equidispersion, would be unfounded in the presence of overdispersion. An alternative to this problem is to use mixed distributions, through models in two stages, or hierarchical, as a way to accommodate this overdispersion. The methodology of the two-stage models associates a distribution to the response conditioned to its average and, later, a distribution to the average parameter, so that, unconditionally, there is a compound distribution for the response variable. In this work, the classical Poisson distribution is used for counting data, and the Gama, Inverse Gaussian and Generalized Lindley distributions for the Poisson mean parameter, thus generating the Poisson-Gama, Poisson-Inverse Gaussian and Generalized Poisson-Lindley. So, the main objective of this work is to present these hierarchical models, or models in two stages, that allow the modeling of count data with overdispersion. Also addressed were some types of residues in the MLG structure (Pearson, deviance component, randomized quantile), adapted for the composite distributions used in the monograph. The performance of each approach was performed using real data sets described in the literature.
Keywords: Counting Data, Overdispersion, Poisson Distribution, Compound Distribution.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

MLGs	Modelos Lineares Generalizados
MLG	Modelo Linear Generalizado
FE	Família Exponencial
EMV	Estimação de Máxima Verossimilhança
GAMLSS	Modelos Aditivos Generalizados para Posição, Escala e Forma
PG	Poisson-Gama
PIG	Poisson-Inversa Gaussiana
PLG	Poisson-Lindley Generalizada

SUMÁRIO

1 INTRODUÇÃO

Em muitas áreas é frequente deparar-se com a investigação de características, feitas em unidades experimentais, que apresentem, por exemplo, o número de defeitos que podem aparecer em um processo de produção (Engenharia); o número de sinistros associados a uma carteira de seguros (Atuária); etc. Dados deste tipo são denominados como dados discretos, pois são expressos em termos de contagens associados a uma característica de interesse (BICKEL; DOKSUM, 1977), não sendo então possível assumir uma distribuição Normal para a modelagem dos mesmos.

De forma geral, dados de contagem podem ser modelados através da distribuição Poisson, muitas vezes associada a modelos de regressão, que são casos particulares da classe dos Modelos Lineares Generalizados (MLGs), difundido por Nelder e Wedderburn (1972), cuja estimação dos parâmetros do modelo é feita utilizando-se a metodologia de MLG (MCCULLAGH; NELDER, 1989).

Ao se utilizar o modelo probabilístico Poisson, acabamos por assumir que a média e variância são iguais. Entretanto, a restrição de modelar média igual à variância faz com que tal modelo tenha uma forte limitação de aplicação em diversas áreas. Visto que, na prática, é bastante comum que dados de contagem apresentem uma variância superior à variância nominal da distribuição Poisson assumida pelo modelo. Tal comportamento é denominado de superdispersão.

Uma forma de contemplar essa superdispersão é utilizar os modelos para dados inflacionados de zeros, conhecidos como ZIP - Poisson Inflacionada de Zeros e ZINB - Binomial-Negativa Inflacionada de Zeros (LAMBERT, 1992; RIDOUT, HINDE e DEMÉTRIO, 1998; HILBE, 2014). Os modelos em dois estágios também podem ser utilizados como outra forma alternativa para modelar a superdispersão (HINDE; DEMETRIO, 1998).

Segundo Hinde e Demétrio (1998) e, conforme utilizado em Mendes (2017), a metodologia dos modelos em dois estágios associa, em um primeiro estágio, uma distribuição à resposta condicionada à sua média ($Y|\mu \sim f_{y|\mu}$) e, em segundo estágio, associa uma distribuição ao parâmetro de média, ($\mu \sim g(\mu)$), de modo que, incondicionalmente, se tem uma distribuição composta ($Y \sim h_y$) para a variável resposta Y.

Neste trabalho é utilizada a distribuição clássica de Poisson (BICKEL; DOKSUM, 1977), para dados de contagem, e as distribuições Gama, Inversa Gaussiana e Lindley

Generalizada para o parâmetro de média da Poisson, que implicam nas distribuições compostas Poisson-Gama (GREENWOOD; YULE, 1920), Poisson-Inversa Gaussiana (HOLLA, 1967) e a Poisson-Lindley Generalizada (WONGRIN; BODHISUWAN, 2016). Também foram abordados alguns tipos de resíduos da estrutura dos MLGs (Pearson e Componente da *deviance*, adaptados para as distribuições compostas utilizadas na monografia.

A proposta do trabalho é fazer um estudo dos modelos através da modelagem com o uso das distribuições compostas citadas acima, para dados de contagem inflacionados de zeros. Será estudada a performance de cada abordagem e o impacto dessas modelagens na qualidade dos ajustes usando-se conjuntos de dados reais. Para o desenvolvimento dos algoritmos necessários para o uso dos métodos acima descritos, foi-se utilizado o software estatístico R (Rstudio Team, 2020).

Essa monografia esta subdividida de forma que possa ser um auxilio no estudo desse tema da seguinte forma: no Capítulo 2 são apresentados os Modelos Lineares Generalizados e os resíduos que serão utilizados, para posteriormente apresentar, nos Capítulo 3 a 6, as distribuições compostas, com base na Poisson, e as formas de seus respectivos resíduos. No Capítulo 7 são ajustados os modelos e apresentado os resíduos e, no Capítulo 8 as considerações finais.

2 MODELOS LINEARES GENERALIZADOS

Em modelos de regressão linear, a distribuição Normal tem um papel fundamental, visto que se assume que a fonte de variação e a variável resposta seguem uma distribuição Normal, permitindo que o procedimento de inferência exatas destes modelos, como testes e intervalos de confiança, sejam realizados (MYERS *et al.*, 2010). Entretanto, muitas vezes na prática, algumas pressuposições como aditividade do componente sistemático do modelo, normalidade, variância constante ou até mesmo independência das variáveis respostas, não são atendidas, fazendo com que um modelo linear clássico não seja apropriado (COSTA, 2003). Dessa forma, o Modelo Linear Generalizado (MLG) nos permite ajustar um modelo de regressão onde a distribuição considerada não precisa ser normal, podendo seguir qualquer distribuição da classe de distribuições chamada família exponencial.

O termo modelo linear generalizado foi introduzido por Nelder e Wedderburn (1972), e se refere a uma classe de modelos que servem como alternativa à distribuição Normal do modelo linear clássico. Essa modelagem consiste em aplicar uma transformação na média dos dados, isto é, uma transformação na esperança da variável resposta. Sendo a ideia básica, estimar os parâmetros de um modelo linear utilizando o método de máxima verossimilhança baseado na distribuição dos dados originais (COSTA, 2003).

2.1 Família exponencial de distribuições

Um importante conceito subjacente ao MLG é a família exponencial de distribuições. Nelder e Wendderburn (1972) introduzem um parâmetro ϕ na família exponencial linear canônica no qual este parâmetro possibilita englobar novas distribuições para a variável resposta (Y). Denotada como da Família Exponencial (FE), ou também como da família exponencial de dispersão, sua função de probabilidade pode ser definida na forma

$$f(y;\theta,\phi) = \exp\left\{\frac{1}{a(\phi)}\Big[y\theta - b(\theta)\Big] + c(y;\phi)\right\}\ , \tag{2.1}$$

sendo $a(\cdot)$, $b(\cdot)$ e $c(\cdot)$ funções conhecidas, θ o parâmetro canônico (natural) e, $\phi > 0$, um parâmetro de dispersão, considerado conhecido (AGRESTI, 2002).

Assumindo que $f(y;\theta,\phi)$ tem a forma de uma FE, temos como propriedade, conforme Bickel e Doksum (2001)

$$\mathbb{E}\left[\frac{\partial l(\theta,\phi;y)}{\delta\theta}\right]=0$$

e

$$\mathbb{E}\left[-\frac{\partial^2 l(\theta,\phi;y)}{\partial\theta^2}\right]=\mathbb{E}\left[\left(\frac{\partial l(\theta,\phi;y)}{\delta\theta}\right)^2\right].$$

Através de (2.1) o valor esperado e variância de Y podem ser obtidos, respectivamente, como

$$\begin{aligned}\mathbb{E}(Y) &= b'(\theta)=\mu\\ \mathrm{Var}(Y) &= a(\phi)b''(\theta)=a(\phi)V(\mu),\end{aligned}$$

em que $V=V(\mu)=\frac{\delta\mu}{\delta\theta}$ é a função de variância e depende unicamente da média μ. Assim, o parâmetro natural θ pode ser expresso na forma

$$\theta=\int V^{-1}(\mu)\;d\mu=q(\mu)\;,$$

em que $q(\mu)$ é uma função conhecida da média μ e $V^{-1}(\mu)$ é a função inversa da função de variância.

2.2 Definição do modelo

Os MLGs envolvem uma variável resposta univariada, variáveis explanatórias e uma amostra aleatória de n observações independentes (CORDEIRO; DEMÉTRIO, 2008), sendo que:

I. O **componente aleatório** do modelo é caracterizado por um conjunto de variáveis respostas independentes $Y_1,\ldots,Y_n$ que seguem uma distribuição proveniente da família exponencial de dispersão, com médias $\mu_1,\ldots,\mu_n$, que englobam distribuições tanto para dados contínuos como, Normal, Normal Inversa e Gama, como para dados de contagens, como a Poisson, Binomial e Binomial Negativa. Tendo sua função densidade de probabilidade conforme descrita em (2.1).

II. As variáveis explicativas ou covariáveis, entram na forma de uma estrutura linear, ou seja, uma soma linear de seus efeitos, gerando o chamado **componente sistemático** do modelo, dando origem a um vetor de preditores lineares, isto é

$$\eta_i = \sum_{j=1}^{p} x_{ij}\beta_j = \boldsymbol{x}_i^\top \boldsymbol{\beta} \ , \ \text{para } i = 1, \ldots, n \ ,$$

também escrito como

$$\boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta}$$

em que $\boldsymbol{\eta}$, preditor linear, é um vetor de dimensões $n \times 1$, $\boldsymbol{X}$ é a matriz de especificação do modelo, com dimensão $n \times p$ e $\boldsymbol{\beta}$ é um vetor de p parâmetros desconhecidos.

III. A ligação entre os componentes aleatórios e sistemáticos é feita através de uma **função de ligação**. Essa ligação é feita através de uma função $g(\cdot)$ conhecida, que relaciona a média μ_i ao preditor linear, isto é

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

podendo ser expressa como

$$\mu_i = g^{-1}(\eta_i) = g^{-1}(\boldsymbol{x}_i^\top \boldsymbol{\beta})$$

em que $g(\cdot)$ é uma função monótona e duplamente diferenciável. Um exemplo é a função de ligação logarítmica para os modelos log-lineares ou a função de ligação logit para modelos logísticos.

2.3 Estimação por máxima verossimilhança

Considerando $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \phi)^\top$ e uma amostra aleatória independente $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)^\top$, com $y_i \sim FE(\boldsymbol{\theta})$, para se obter as estimativas dos parâmetros $\boldsymbol{\beta}$ e ϕ do modelo, pode-se recorrer ao método de máxima verossimilhança, no qual determinam-se os valores de $\boldsymbol{\beta}$ e ϕ que maximizam o logaritmo da função de verossimilhança.

Dada a função densidade de probabilidade (2.1) e considerando o parâmetro de dispersão ϕ conhecido, a função de verossimilhança de $\boldsymbol{\theta}$ é dada por

$$L(\boldsymbol{\theta}; \mathbf{y}) = \prod_{i=1}^{n} \left[\exp\left\{ \frac{1}{a_i(\phi)} \Big(y_i\theta_i - b(\theta_i) \Big) + c(y_i; \phi) \right\} \right] .$$

Portanto, o logaritmo da função de verossimilhança é dada como

$$\begin{aligned} l(\boldsymbol{\theta}) &= \log[L(\boldsymbol{\theta}; \mathbf{y})] \\ &= \sum_{i=1}^{n} \log\left[\exp\left\{\frac{1}{a_i(\phi)}\Big(y_i\theta_i - b(\theta_i)\Big) + c(y_i;\phi)\right\}\right] \\ &= \sum_{i=1}^{n}\left[\frac{1}{a_i(\phi)}\Big(y_i\theta_i - b(\theta_i)\Big) + c(y_i;\phi)\right], \end{aligned} \tag{2.2}$$

em que, como citado, $\theta_i = q(\mu_i)$, $\mu_i = g^{-1}(\eta_i)$ e $\eta_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}$.

Para obter as estimativas dos vetores de médias $\boldsymbol{\mu}$ e do vetor de parâmetros $\boldsymbol{\theta}$, é necessário estimar primeiramente o vetor $\boldsymbol{\beta}$. Para isto, se faz uso da expressão (2.2) e da regra da cadeia, para obtenção do vetor escore formado pelas derivadas parciais de primeira ordem do logaritmo da função de verossimilhança em relação a $\boldsymbol{\beta}$. Assim

$$\begin{aligned} U_{\beta_j}(\boldsymbol{\theta}) = \frac{\partial l(\boldsymbol{\theta})}{\partial \beta_j} &= \frac{\partial}{\partial \beta_j}\sum_{i=1}^{n}\left[\frac{1}{a_i(\phi)}\Big(y_i\theta_i - b(\theta_i)\Big) + c(y_i;\phi)\right] \\ &= \frac{\partial}{\partial \beta_j}\left[\sum_{i=1}^{n}\frac{1}{a_i(\phi)}\Big(y_i\theta_i - b(\theta_i)\Big) + \sum_{i=1}^{n} c(y_i;\phi)\right] \\ &= \sum_{i=1}^{n}\frac{1}{a_i(\phi)}\frac{\partial}{\partial \beta_j}\Big(y_i\theta_i - b(\theta_i)\Big) \\ &= \sum_{i=1}^{n}\frac{1}{a_i(\phi)}\left(y_i\frac{\partial \theta_i}{\partial \beta_j} - \frac{\partial b(\theta_i)}{\partial \beta_j}\right), \end{aligned}$$

sendo θ_i uma função de μ_i, μ_i uma função inversa de η_i e η_i o i-ésimo elemento do vetor de preditores lineares, além disso, sabendo que $V_i = V(\mu_i)$ e $\mu_i = b'(\theta_i)$, então

$$\begin{aligned} \frac{\delta l(\boldsymbol{\theta})}{\partial \beta_j} &= \sum_{i=1}^{n}\frac{1}{a_i(\phi)}\left(y_i\frac{\partial \theta_i}{\partial \mu_i}\frac{\partial \mu_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial \beta_j} - \frac{\partial b(\theta_i)}{\partial \theta_i}\frac{\partial \theta_i}{\partial \mu_i}\frac{\partial \mu_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial \beta_j}\right) \\ &= \sum_{i=1}^{n}\frac{1}{a_i(\phi)}\left(y_i V_i^{-1}\frac{\partial \mu_i}{\partial \eta_i}x_{ij} - \mu_i V_i^{-1}\frac{\partial \mu_i}{\partial \eta_i}x_{ij}\right) \\ &= \sum_{i=1}^{n}\frac{1}{a_i(\phi)}\left(\frac{\partial \mu_i}{\partial \eta_i}V_i^{-1}(y_i - \mu_i)x_{ij}\right). \end{aligned}$$

Além disso, considera-se uma matriz diagonal $n \times n$, $\mathbf{W}$, de pesos, em que o i-ésimo elemento é definido como

$$w_i = \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 \frac{1}{V_i}, \tag{2.3}$$

assim, pode-se escrever o vetor escore em relação a $\boldsymbol{\beta}$, considerando a matriz de pesos, na forma

$$\begin{aligned} U_{\beta_j}(\boldsymbol{\theta}) &= \sum_{i=1}^{n}\frac{1}{a_i(\phi)}\left(\sqrt{\left(\frac{\partial \mu_i}{\partial \eta_i}\frac{1}{V_i}\right)^2}(y_i - \mu_i)x_{ij}\right) \\ &= \sum_{i=1}^{n}\frac{1}{a_i(\phi)}\left(\sqrt{\frac{w_i}{V_i}}(y_i - \mu_i)x_{ij}\right), \end{aligned} \tag{2.4}$$

em que $j = 1, 2, \ldots, p$, sendo p o número de parâmetros de regressão do modelo.

Para a função escore de ϕ seguindo o mesmo procedimento adotado anteriormente, tem-se

$$\begin{aligned} U_\phi(\boldsymbol{\theta}) &= \frac{\partial l(\boldsymbol{\theta})}{\partial \phi} \\ &= \frac{\partial}{\partial \phi} \sum_{i=1}^{n} \left[\frac{1}{a_i(\phi)} \bigg(y_i\theta_i - b(\theta_i) \bigg) + c(y_i;\phi) \right], \end{aligned} \tag{2.5}$$

que não possui forma mais simplificada, visto que varia em relação a qual distribuição se está utilizando. Assim, função escore dos parâmetros pode ser expressa como

$$\mathbf{U}_\theta = (\mathbf{U}_\beta^\top, \mathbf{U}_\phi^\top). \tag{2.6}$$

Assim como na função escore que utiliza-se a função de verossimilhança, a informação de Fisher é definida utilizando a segunda derivada do logaritmo da função de verossimilhança. A matriz de informação de Fisher para o parâmetro β é definida como na forma

$$\begin{aligned} K_\beta(\boldsymbol{\theta}) = -\mathbb{E}\left[\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \beta_j \partial \beta_l}\right] &= \mathbb{E}\left[\frac{\partial l(\boldsymbol{\theta})}{\partial \beta_j}\frac{\partial l(\boldsymbol{\theta})}{\partial \beta_l}\right] \\ &= \sum_{i=1}^{n} \mathbb{E}\left[\frac{1}{a_i(\phi)}\frac{\partial \mu_i}{\partial \eta_i} V_i^{-1}(y_i-\mu_i)x_{ij} \times \frac{1}{a_i(\phi)}\frac{\partial \mu_i}{\partial \eta_i} V_i^{-1}(y_i-\mu_i)x_{il}\right] \\ &= \sum_{i=1}^{n} \left(\frac{1}{a_i(\phi)}\right)^2 V_i^{-2} \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 \mathbb{E}(y_i-\mu_i)^2 x_{ij}x_{il} \ . \end{aligned}$$

Considerando que $\mathrm{Var}(Y) = \mathbb{E}(Y-\mu)^2$ e utilizando dos elementos da matriz $\mathbf{W}$, (2.3), temos

$$\begin{aligned} K_\beta(\boldsymbol{\theta}) &= \sum_{i=1}^{n} \left(\frac{1}{a_i(\phi)}\right)^2 V_i^{-1}V_i^{-1} \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 \mathrm{Var}(y_i)\ x_{ij}x_{il} \\ &= \sum_{i=1}^{n} \left(\frac{1}{a_i(\phi)}\right)^2 V_i^{-1} w_i \,\mathrm{Var}(y_i)\ x_{ij}x_{il} \\ &= \sum_{i=1}^{n} \frac{1}{a_i(\phi)^2}\frac{1}{V_i} w_i \,\mathrm{Var}(y_i)\ x_{ij}x_{il} \ . \end{aligned}$$

Além disso, como definido anteriormente, $\mathrm{Var}(y_i) = a_i(\phi)V_i$, portanto

$$\begin{aligned} K_\beta(\boldsymbol{\theta}) &= \sum_{i=1}^{n} \frac{1}{a_i(\phi)^2}\frac{1}{V_i} w_i a_i(\phi) V_i\ x_{ij}x_{il} \\ &= \sum_{i=1}^{n} \frac{1}{a_i(\phi)} w_i\ x_{ij}x_{il} \end{aligned}$$

O procedimento da obtenção da informação de Fisher de ϕ é feita de maneira similar, em que se tem

$$
\begin{aligned}
K_{\phi}(\boldsymbol{\theta}) &= -\mathbb{E}\left[\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \phi}\right] \\
&= -\mathbb{E}\left\{\frac{\partial}{\partial \phi}\left[\frac{\partial}{\partial \phi}\sum_{i=1}^{n}\left[\frac{1}{a_i(\phi)}\Big(y_i\theta_i - b(\theta_i)\Big) + c(y_i;\phi)\right]\right]\right\} \\
&= -\sum_{i=1}^{n}\mathbb{E}\left\{\frac{\partial}{\partial \phi}\left[\Big(y_i\theta_i - b(\theta_i)\Big)\frac{\partial}{\partial \phi}\frac{1}{a_i(\phi)} + \frac{\partial}{\partial \phi}c(y_i;\phi)\right]\right\} \\
&= -\sum_{i=1}^{n}\mathbb{E}\left\{\Big(y_i\theta_i - b(\theta_i)\Big)\frac{\partial^2}{\partial \phi^2}\left[\frac{1}{a_i(\phi)}\right] + \frac{\partial^2}{\partial \phi^2}\left[c(y_i;\phi)\right]\right\} \\
&= -\sum_{i=1}^{n}\mathbb{E}\left\{\Big(y_i\theta_i - b(\theta_i)\Big)\frac{\partial^2}{\partial \phi^2}\left[\frac{1}{a_i(\phi)}\right] + \frac{\partial^2}{\partial \phi^2}\left[c(y_i;\phi)\right]\right\}
\end{aligned}
$$

A Estimação de Máxima Verossimilhança (EMV) do vetor de parâmetros $\boldsymbol{\beta}$ é obtida igualando-se a equação de U_{β_j}, em (2.4), a zero para todos os valores de j. Da mesma forma, para a estimação de ϕ, iguala-se U_{ϕ}, (2.5), a zero. Em geral, as equações são não-lineares em $\boldsymbol{\beta}$ e ϕ, de forma que se faz necessário o uso de métodos iterativos para a solução da EMV, como por exemplo, escore de Fisher e Newton-Raphson, sendo ambos baseados na aproximação de Taylor da função escore.

O Método interativo Newton-Raphson é bastante útil quando as derivas parciais de segunda ordem são facilmente avaliadas. Entretanto, utiliza-se no caso dos MLGs o método escore de Fisher que coincide com o método de Newton-Raphson ao utilizar as funções de ligação canônica (CORDEIRO; DEMÉTRIO, 2008). A diferença dos dois métodos é que no Newton-Raphson faz-se uso da matriz Hessiana observada e no método escore de Fisher, utiliza-se a matriz Hessiana esperada. Além disso, quando se recorre ao método escore de Fisher, o inverso da matriz de informação (ou seja, o negativo do inverso da matriz Hessiana esperada) também é a covariância assintótica de $\hat{\boldsymbol{\beta}}$, assim, o método escore de Fisher acaba por fornece também este resultado.

2.4 Análise de resíduos

A análise de diagnóstico é uma etapa fundamental no ajuste de modelos de regressão, pois ela verifica possíveis afastamentos das suposições feitas para o modelo, bem como a existência de observações influentes que causam alguma interferência nos resultados do ajuste. Tal etapa começa com uma análise de resíduos para detectar a presença de pontos aberrantes e avaliar a adequação da distribuição proposta para a variável resposta (PAULA, 2013).

No caso dos MLGs, essa análise também se baseia na validação da adequação da parte sistemática (preditor linear) do modelo e a adequação da função de ligação utilizada. Boa parte dos métodos de análise de resíduos são semelhantes aos procedimentos usados para o modelo clássico de regressão, com algumas adaptações. Entretanto, deve-se usar com cautela, visto que alguns resultados dependem fortemente das propriedades do modelo proposto.

Alguns trabalhos citados na literatura apresentam extensões da análise de resíduos para os modelos lineares generalizados. Podemos citar, por exemplo, Pregibon (1981), Cox e Snell (1968), Williams (1984) e Fahrmeir e Tutz (1994). Para mais detalhes, veja (PAULA, 2013).

2.4.1 Tipos de resíduos

Um resíduo, geralmente denotador por r_i, é uma medida que expressa a distância entre uma observação (y_i) e o seu valor ajustado ($\hat{\mu}_i$):

$$r_i = d_i(y_i, \hat{\mu}_i),$$

sendo d_i uma função adequada de fácil interpretação usualmente escolhida para estabilizar a variância ou induzir simetria na distribuição amostral de r_i. Garantindo comparabilidade dos resíduos e detecção de resíduos discrepantes (COX; SNELL, 1968).

A seguir, apresentam-se alguns tipos de resíduos mais comumente utilizados ao se tratar dos MLGs, conforme apresentado em Cordeiro e Demétrio (2008), McCullagh e Nelder (1989) e Paula (2013).

a) **Resíduo ordinário**: Mensura a diferença do valor observado para o valor ajustado de uma dada observação

$$r_i = y_i - \hat{\mu}_i, \tag{2.7}$$

sendo $\hat{\mu}_i$ o valor estimado pelo modelo adotado. De acordo com Cordeiro e Lima (2006), por não apresentarem variância constante, os resíduos ordinários não são muito informativos no diagnóstico de MLGs. Dessa forma, procura-se comparar resíduos de forma padronizada.

b) **Resíduo de Pearson**: São componentes da estatística χ^2 de Pearson generalizada e sua forma mais simples é dada por

$$r_i^P = \frac{y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}}. \tag{2.8}$$

Conforme Cordeiro e Demétrio (2008), os resíduos de Pearson têm como desvantagem o fato de que sua distribuição é, em geral, fortemente assimétrica para modelos não-normais.

c) **Resíduo de Pearson padronizado**: O resíduo de Pearson têm uma versão padronizada, sendo definido como

$$r_i^{P^*} = \frac{y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)(1 - \hat{h}_{ii})}}, \tag{2.9}$$

em que h_{ii} é o i-ésimo elemento da diagonal da matriz de projeção $\mathbf{H}$

$$\mathbf{H} = \mathbf{W}^{\frac{1}{2}}\mathbf{X}(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}\mathbf{W}^{\frac{1}{2}}, \tag{2.10}$$

sendo $\mathbf{W}$ dado por (2.3). A matriz $\mathbf{H}$ mede a influência em unidades estudentizadas de $\mathbf{y}$ sobre $\hat{\boldsymbol{\mu}}$, tendo como propriedades que $\text{tr}(\mathbf{H}) = p$ e $0 \leq h_{ii} \leq 1$. Assim, além de consideramos a função de variância do modelo, como no caso dos resíduos de Pearson, agora levamos em conta a medida de *leverage* estimada, $\hat{h}_{ii}$.

d) **Resíduo componente da *deviance***: O resíduo componente da *deviance* é definido pela contribuição de cada observação para a *deviance*, ou desvio, do modelo, sendo uma medida de distância de y_i em relação a $\hat{\mu}_i$ na escala do logaritmo da verossimilhança (log-verossimilhança).

Proposto por Nelder e Wedderburn (1972), o desvio de um MLG é definido como se segue. Considere uma amostra aleatória $Y_1, \ldots, Y_n$ pertencente à família exponencial, $Y_i \sim FE(\mu_i, \phi)$, com função de log-verossimilhança conforme (2.2), assim

$$l(\boldsymbol{\mu}; \mathbf{y}) = \sum_{i=1}^{n} l(\mu_i, y_i).$$

Considerando um modelo saturado, ou seja, um modelo onde o número de parâmetros de regressão é igual ao número de observações, $p = n$, tem-se que este modelo atribui toda a variação dos dados ao componente sistemático, ajustando-se perfeitamente e reproduzindo os próprios dados, entretanto, tal modelo é de difícil interpretação. A função log-verossimilhança usada para estimar este modelo é descrita como

$$l(\mathbf{y};\mathbf{y}) = \sum_{i=1}^{n} l(y_i, y_i).$$

A medida desvio é expressa como sendo a diferença entre os máximos da função de log-verossimilhança do modelo saturado e do modelo sob pesquisa, isto é,

$$D(\mathbf{y};\hat{\boldsymbol{\mu}}) = 2\left[l(\mathbf{y};\mathbf{y}) - l(\hat{\boldsymbol{\mu}};\mathbf{y})\right] = \sum_{i=1}^{n} d_i. \tag{2.11}$$

Portanto, o resíduo componente da *deviance* fica definido por

$$r_i^D = sinal(y_i - \hat{\mu}_i)\sqrt{d_i}, \tag{2.12}$$

em que $sinal(\alpha) = -1$, se $\alpha < 0$ e $sinal(\alpha) = 1$, se $\alpha > 0$. Um valor grande para r_i^D indica que a i-ésima observação é mal ajustada pelo modelo.

e) **Resíduo componente da *deviance* padronizado**: Assim como o resíduo de Pearson que possui sua forma padronizada, o resíduo componente *deviance* possui uma forma padronizada que leva em conta a medida de *leverage*, sendo definido como

$$r_i^{D*} = \frac{r_i^D}{\sqrt{(1-\hat{h}_{ii})}}. \tag{2.13}$$

f) **Resíduo quantílico aleatorizado**: Sugerido por Dunn e Smyth (1996), os resíduos quantílicos aleatorizados se baseiam no teorema da inversa da função de distribuição acumulada e tem como proposta apresentar uma distribuição Normal, independentemente da distribuição da variável resposta. O resíduo quantílico é definido por

$$r_i^q = \begin{cases} \Phi^{-1}[\mathrm{F}(y_i;\hat{\mu}_i,\phi)], & \text{se F é contínua} \\ \Phi^{-1}[\mathrm{F}(y_i^-;\hat{\mu}_i,\phi) - u_i\mathrm{p}(y_i;\hat{\mu}_i,\phi)], & \text{se F é discreta} \end{cases} \tag{2.14}$$

sendo $\mathrm{F}(y_i;\hat{\mu}_i,\phi)$ a função de distribuição acumulada de Y_i, $\mathrm{p}(y_i;\hat{\mu}_i,\phi)$ a função de probabilidade, $\mathrm{F}(y_i^-;\hat{\mu}_i,\phi) = \lim_{y\to y_i^-} \mathrm{F}(y;\hat{\mu}_i,\phi)$, u_i uma variável aleatória uniforme no intervalo $(0,1]$ e $\Phi(\cdot)$ a função de distribuição acumulada de uma normal padrão.

2.4.2 Gráficos para os resíduos

Comumente, dentre os principais tipos de gráficos para diagnosticar o ajuste do modelo com base nos resíduos, destacam-se:

a) **Gráfico de índices**: Utilizado para localizar observações com altos valores de resíduos. Pode ser útil para identificar possíveis *outliers*, ou seja, observações que se destacam da tendência geral dos dados.

b) **Resíduos *versus* valores ajustados**: Verifica a constância da variância para a distribuição em uso (MCCULLAGH; NELDER, 1989). Se espera que o gráfico possua um comportamento dos resíduos em torno de zero com amplitude constante, em que os desvios sistemáticos podem ter algum tipo de amplitude ou curvatura diferente com o valor ajustado. Ou seja, para um modelo bem ajustado, se espera que os resíduos estejam distribuídos aleatoriamente em torno de zero e com variância constante.

c) **Quantis-Quantis com envelopes simulados**: Os envelopes, no caso dos MLGs com distribuições diferentes da normal, são construídos com os resíduos sendo gerados a partir do modelo ajustado (PIERCE; SCHAFER, 1986). Weisberg (2005) cita que o gráfico normal de probabilidade se destaca por dois aspectos: a identificação de valores que se destacam no conjunto das observações e a identificação da distribuição original dos dados (MARCIANO, 2009).

d) **Variável ajustada *versus* preditor linear**: Usado para avaliar se a função de ligação do modelo está adequada. Uma tendência linear no gráfico indicaria que a escolha da função de ligação é adequada.

e) **Worm-plot**: É uma ferramenta de diagnóstico para visualização de quão bem um modelo estatístico se ajusta aos dados, comparar com o ajuste de outros modelos e encontrar locais em que o ajuste pode ser ser melhorado. Se os dados forem normais, os pontos devem estar próximos da curva vermelha e com poucas oscilações (BUUREN, 2007). No caso dos MLGs, como o worm-plot testa a normalidade dos dados, se é utilizados os resíduos quantílicos. É um gráfico muito utilizado para a classe de Modelos Aditivos Generalizados para Posição, Escala e Forma (GAMLSS).

2.5 Seleção do Modelo

A escolha de um modelo apropriado é um tópico extremamente importante na análise de dados (BOZDOGAN, 1987). Na prática, podemos nos deparar com um ou mais modelos, no qual sempre se tem o interesse de utilizar o modelo mais parcimonioso, ou seja, um modelo que explique bem o comportamento da variável resposta e que envolva o mínimo de parâmetros possíveis. Nessa linha, diversos critérios de seleção de modelos são apresentados na literatura. Critérios, tais como, C_p de Mallows (MALLOWS, 1973), Regressão Stepwise (HOCKING, 1976), Critério de Informação de Akaike (AKAIKE, 1975), Critério de Informação Bayesiano (SCHWARZ, 1978) e Critério de Informação Generalizado (KONISHI; KITAGAWA, 2008).Além disso, pode-se utilizar de ferramentas de diagnósticos para seleção do modelo.

Proposto por Akaike (1973), o Critério de Informação de Akaike (AIC) é uma medida relativa da qualidade de ajuste de um modelo estatístico estimado, baseando-se na minimização da distância de Kullback-Leibler (K-L). Assim, o critério AIC avalia a qualidade do ajuste estimado pelo método da máxima verossimilhança. Sua forma geral é dada por

$$AIC = -2l(\widehat{\boldsymbol{\beta}}; \mathbf{y}) + 2p, \tag{2.15}$$

sendo $l(\boldsymbol{\beta}; \mathbf{y})$ a função de log-verossimilhança maximizada para o modelo estimado e p o número de parâmetros. Seleciona-se o modelo que retorna o menor valor AIC.

Outro critério de avaliação do modelo, que tem como base a estimação de máxima verossimilhança, é o Critério de Informação Bayesiano (BIC). Também chamado de Critério de Schwarz, esse método foi introduzido por Schwarz (1978), sendo um critério consistente, ele é definido em termos da probabilidade a posteriori. Tem-se como forma geral

$$BIC = -2l(\widehat{\boldsymbol{\beta}}; \mathbf{y}) + p \ \ln n, \tag{2.16}$$

sendo p o número de parâmetros, $l(\widehat{\boldsymbol{\beta}}; \mathbf{y})$ é a função de log-verossimilhança maximizada para o modelo estimado e n o número de observações da amostra.

Como visto pelas forma funcionais, ambos os critérios fundamentam-se na função de verossimilhança impondo, entretanto, penalizações diferentes. Além disso, ambos servem para comparar modelos encaixados ou aninhados (BURNHAM; ANDERSON, 2002).

2.6 Superdispersão para dados de contagem com excesso de zeros

Um dos cuidados que se deve ter ao fazer a análise de dados de contagem é a possibilidade da existência de superdispersão. Esse termo é usado quando a razão entre variância e a média é maior do que um, ou seja, a média observada nos dados excede à variância nominal estipulada pelo modelo, que é assumida pela distribuição Poisson. Isso pode ocorrer em diversas áreas, como por exemplo na Epidemiologia, Psicologia, Agronomia e dentre outras. McCullagh e Nelder (1989) citam que, na prática, não é incomum obter superdispersão, a dispersão nominal que é a exceção.

Uma atenção maior indicada em situações com a superdispersão é que a não consideração desta, na análise dos dados, pode levar a sub ou superestimação dos erros-padrão das estimativas dos parâmetros do modelo em questão, causando assim uma estimação incorreta destes e, como consequência, podendo levar a uma tomada de decisão errônea (HINDE; DEMETRIO, 1998). Dessa forma, alguns modelos foram apresentados na literatura como algumas alternativas ao modelo Poisson. Hinde e Demétrio (1998), por exemplo, mostram uma revisão de modelos superdispersos para dados de contagem. Outras discussões podem ser encontradas em Breslow (1984), Lawless (1987) e Lindsey (1995).

A superdispersão, em dados de contagem, pode ocorrer por várias razões, uma delas é quando existe excesso de zeros no conjunto de dados, ou seja, a incidência de zeros é maior do que a esperada pela distribuição do modelo base, implicando em uma variabilidade maior do que o esperado. No caso da Poisson, isto implica em uma variação maior do que a média postulado do modelo. Dessa forma, utilizar um modelo com base em uma distribuição de Poisson não seria adequado, visto que esta possui a propriedade de equidispersão.

Assim, algumas alternativas são propostas na literatura para contornar o problema da inflação de zeros no conjunto. Entre eles, podemos citar a utilização de distribuições alternativas (por exemplo, a distribuição Binomial Negativa), como mostrado por Breslow (1984); abordar uma forma mais geral para a função de variância, conforme

McCullagh e Nelder (1989), por exemplo; assumir distribuições compostas para a variável resposta, através dos modelos hierárquicos, conforme descrito por Hinde e Demétrio (1998); e, incorporar um efeito aleatório ao preditor linear, segundo Costa (2003) .

Nesta monografia são utilizados, como alternativa, os modelos hierárquicos, também denotados como modelos em dois estágios. Sendo abordado da seguinte forma: em primeiro estágio, associa-se uma distribuição à variável resposta condicionada à sua média ($Y|\mu \sim f_{y|\mu}$) e, em segundo estágio, associa-se uma distribuição à média ($\mu \sim g(\mu)$), de modo que, incondicionalmente, tem-se uma distribuição composta ($Y \sim h_y$) para a variável resposta Y.

De forma específica, são abordados os modelos Poisson, Poisson-Gama, Poisson-Inversa Gaussiana e Poisson-Lindley Generalizada, todos derivados de misturas da distribuição de Poisson com as distribuições Gama, Inversa Gaussiana e a Lindley Generalizada, para modelar o parâmetro de média (λ) da Poisson. Dessa forma se tem:

$$\begin{cases} Y_i|\theta_i \sim \text{Pois}(\lambda_i) \\ \lambda_i \sim \text{Distribuíções citadas.} \end{cases}$$

Esses modelos hierárquicos tem uma similaridade, na forma da composição dos modelos propostos, com o método da análise Bayesiana. Visto que na inferência bayesiana, assume-se uma distribuição à variável resposta condicionada ao seus parâmetros e uma distribuição a priori para estes parâmetros (PAULINO *et al.*, 2018).

3 DISTRIBUIÇÃO POISSON

Proposta por Siméon-Denis Poisson (1937), a distribuição de Poisson representa a distribuição de probabilidade do número de ocorrências de um evento em um determinado intervalo contínuo (de tempo, volume, área, etc) dado o número médio de vezes que o evento ocorre nesse intervalo, sendo eventos independentes entre si.

Associada às variáveis aleatórias discretas, ela é, comumente, a distribuição inicial utilizada para analisar dados de contagem, cujo suporte é o conjunto dos números inteiros não negativos. Essa distribuição tem como característica principal a pressuposição de equidispersão, em outras palavras, assume que a variância dos dados é igual à média.

Considerando a variável aleatória Y sendo o número de ocorrências do experimento em um intervalo δ_t, os possíveis resultados desse experimento seguem uma distribuição de Poisson, $Y \sim \text{Pois}(\lambda)$, com função de probabilidade dada por:

$$p(y;\lambda) = \frac{e^{-\lambda}\lambda^y}{y!} \qquad (3.1)$$

com $y = 0, 1, 2, \ldots$, sendo $\lambda > 0$ o parâmetro da distribuição.

A Figura 1 apresenta o comportamento da função de probabilidade da distribuição Poisson assumindo diferentes valores para λ.

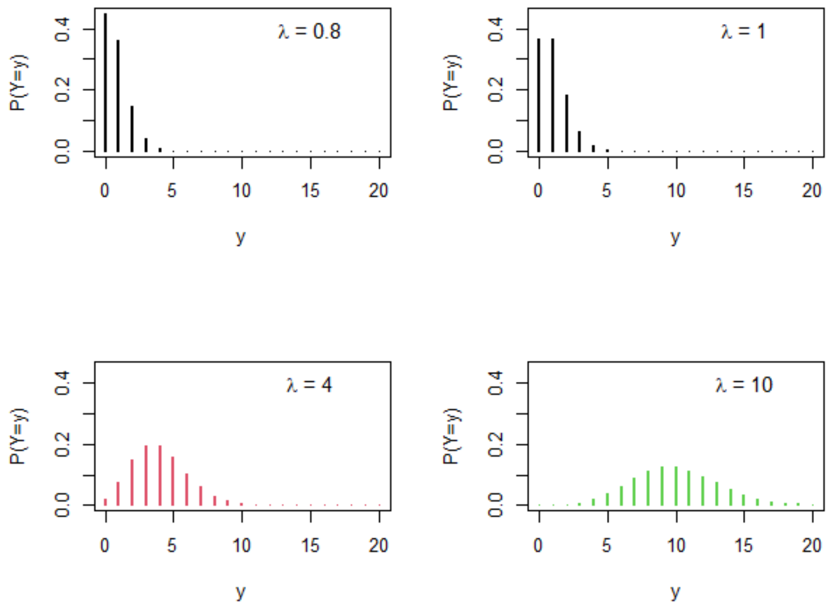


Figura 1 – Função de probabilidade da distribuição Poisson para diferentes valores de λ

3.1 Propriedades

Propriedade 3.1.1 *Se* $Y \sim \text{Pois}(\lambda)$*, então o valor esperado,* $\text{E}(Y)$*, e a variância,* $\text{Var}(Y)$*, é expresso pelo parâmetro da distribuição,* λ*.*

Prova: Por se tratar de dados discretos, o valor esperado de Y é encontrado como

$$\text{E}(Y) = \sum_{y=0}^{\infty} y \frac{e^{-\lambda}\lambda^y}{y!} = e^{-\lambda}\lambda \sum_{y=1}^{\infty} \frac{\lambda^{y-1}}{(y-1)!},$$

fazendo $x = y - 1$, obtem-se

$$\text{E}(Y) = e^{-\lambda}\lambda \sum_{y=0}^{\infty} \frac{\lambda^x}{x!} = \lambda e^{-\lambda} e^{\lambda} = \lambda. \tag{3.2}$$

O segundo momento em relação a origem é obtido por

$$\begin{aligned}
\text{E}(Y^2) &= \sum_{y=0}^{\infty} y^2 \frac{e^{-\lambda}\lambda^y}{y!} = e^{\lambda}\lambda \sum_{y=1}^{\infty} y \frac{1}{(y-1)!}\lambda^{y-1} \\
&= e^{\lambda}\lambda \sum_{y=1}^{\infty} (y-1+1) \frac{1}{(y-1)}\lambda^{y-1} \\
&= e^{\lambda}\lambda \left[\sum_{y=1}^{\infty} (y-1) \frac{1}{(y-1)!}\lambda^{y-1} + \sum_{y=1}^{\infty} \frac{1}{(y-1)!}\lambda^{y-1} \right] \\
&= e^{\lambda}\lambda \left[\sum_{x=0}^{\infty} x \frac{1}{x!}\lambda^{x} + \sum_{x=0}^{\infty} \frac{1}{x!}\lambda^{x} \right] \\
&= \lambda(\lambda - e^{-\lambda}e^{\lambda}) = \lambda^2 + \lambda.
\end{aligned}$$

Assim, conforme os resultados de $E(Y)$ e $E(Y^2)$, a variância associada a Y, é dada

$$\begin{aligned}
\text{Var}(Y) &= \text{E}(Y^2) - \text{E}^2(Y) \\
&= \lambda^2 + \lambda - \lambda^2 = \lambda.
\end{aligned} \tag{3.3}$$

Conforme comentado, a distribuição de Poisson tem a característica de possuir o valor esperado e variância iguais, o que pode ser visualizado conforme os resultados obtidos a partir da propriedade (3.1.1).

3.2 Modelo Poisson

O modelo Poisson possui um papel importante na análise de dados de contagem. Ele possui algumas características como: proporcionar uma descrição dos dados experimentais cuja variância é proporcional à média, ser deduzido teoricamente de princípios elementares com um número mínimo de restrições e, determinar o número de eventos em um intervalo, dado que ocorreram independentemente e aleatoriamente (CORDEIRO; DEMÉTRIO, 2008).

Considere $Y_1, \ldots, Y_n$ variáveis respostas condicionalmente independentes e um vetor de covariáveis $\boldsymbol{x}_i^\top = (x_{i1}, x_{i2}, \ldots, x_{ip})$, $i = 1, \ldots, n$. Assumindo que a distribuição condicional $Y_i|\boldsymbol{x}_i^\top$ segue uma Pois(μ_i), com $\mu_i > 0$, isto é, $Y_i|\boldsymbol{x}_i^\top \sim$ Pois(μ_i), a função de probabilidade de $Y_i|\boldsymbol{x}_i^\top$, conforme (3.1) é dada na forma

$$f(y_i|\boldsymbol{x}_i^\top) = \frac{e^{-\mu_i}(\mu_i)^{y_i}}{y_i!} \tag{3.4}$$

em que $y_i = 0, 1, \ldots$ Sabendo que (3.4) pertence à FE, com uma função de ligação canônica sendo a logarítmica para a média, então a relação do preditor linear com μ_i é definido por

$$\eta_i = \log(\mu_i) = \boldsymbol{x}_i^\top \boldsymbol{\beta} \ , \tag{3.5}$$

sendo $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ o vetor de parâmetros, de dimensão $p \times 1$.

Dessa forma, sabendo que (3.4) pode ser escrita da forma (2.1) e, conforme (3.2) e (3.3), temos como média e variância do modelo, respectivamente

$$\mathrm{E}(Y_i|\boldsymbol{x}_i^\top) = \mu_i = e^{\boldsymbol{x}_i^\top \boldsymbol{\beta}} \tag{3.6}$$

e

$$\mathrm{Var}(Y_i|\boldsymbol{x}_i^\top) = \mu_i = e^{\boldsymbol{x}_i^\top \boldsymbol{\beta}}. \tag{3.7}$$

3.3 Resíduos

Utilizando alguns resíduos abordados na seção (2.4.1), a seguir são apresentados alguns destes adaptados para a forma do modelo Poisson. Levando em consideração que esta distribuição pertencente à família exponencial de distribuição, tem-se como parâmetro de dispersão $a(\phi) = 1$. Dessa forma, a função de variância, $V(\hat{\mu}_i)$, fica escrita apenas na forma da própria variância do modelo, Var$(\hat{y}_i)$.

I. **Resíduos de Pearson**: Utilizando a forma geral (2.8), e a variância para a Poisson (3.7), o resíduo de Pearson fica então especificado como

$$r_i^P = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\mu}_i}}. \tag{3.8}$$

II. **Resíduos de Pearson padronizado**: Para a composição dos resíduos de Pearson padronizado, é necessário utilizarmos a matriz de pesos $\mathbf{W}$. Considerando, por exemplo, uma função de ligação logarítmica, $\eta_i = \log(\mu_i)$, o i-ésimo elemento da diagonal de $\mathbf{W}$, (2.3), é dado por

$$\begin{aligned} w_i &= \left[\left(\frac{\partial \eta_i}{\partial \mu_i}\right)^2 V(\mu_i)\right]^{-1} = \left[\left(\frac{\partial \log(\mu_i)}{\partial \mu_i}\right)^2 V(y_i)\right]^{-1} \\ &= \left[\left(\frac{1}{\mu_i}\right)^2 \mu_i\right]^{-1} = \left[\frac{1}{\mu_i^2}\ \mu_i\right]^{-1} = \left(\frac{1}{\mu_i}\right)^{-1} = \mu_i\ . \end{aligned} \tag{3.9}$$

Assim, utilizando (2.9), o resíduo padronizado é descrito como

$$r_i^P = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\mu}_i(1 - \hat{h}_{ii})}}, \tag{3.10}$$

em que $\hat{h}_{ii}$ é o i-ésimo elemento da diagonal da matriz de projeção $\widehat{\mathbf{H}}$, (2.10), com a matriz de pesos sendo $\widehat{\mathbf{W}} = \operatorname{diag}(\hat{\mu}_1, \ldots, \hat{\mu}_n)$.

III. **Resíduo componente da *deviance***: Utilizando a função de probabilidade definida em (3.4), podemos considerar a função log-verossimilhança do modelo Poisson como sendo

$$l(\boldsymbol{\mu}) = \sum_{i=1}^{n} \log(f(y_i|\boldsymbol{x}_i^\top)) = -\sum_{i=1}^{n} \mu_i + y_i \sum_{i=1}^{n} \log(\mu_i) - \sum_{i=1}^{n} \log y_i!, \tag{3.11}$$

e, pela definição em (2.11) para os desvios, obtemos

$$\begin{aligned} d_i &= 2\Big[-y_i + y_i \log(y_i) - \log y_i! - \Big(-\hat{\mu}_i + y_i \log(\hat{\mu}_i) - \log y_i!\Big)\Big] \\ &= 2\Big[-y_i + y_i \log(y_i) + \hat{\mu}_i - y_i \log(\hat{\mu}_i)\Big] \\ &= 2\left[y_i \log\left(\frac{y_i}{\hat{\mu}_i}\right) - (y_i - \hat{\mu}_i)\right], \end{aligned}$$

para $y_i > 0$. Caso $y_i = 0$, a substituição no logaritmo seria indefinido, $\log(0)$, assim, obtemos o desvio, para este caso como

$$\begin{aligned} f(y_i = 0|\boldsymbol{x}_i^\top) &= \frac{e^{-\mu_i}(\mu_i)^0}{0!} = \frac{e^{-\mu_i} \times 1}{1} = \exp(-\mu_i)\ , \\ \Rightarrow d_i &= 2\Big[l(y_i; y_i) - l(\hat{\mu}_i; y_i)\Big] \\ &= 2\Big[\log\Big(\exp\{-y_i\}\Big) - \log\Big(\exp\{-\hat{\mu}_i\}\Big)\Big] \\ &= 2\Big[\log\Big(\exp\{0\}\Big) - \Big(-\hat{\mu}_i\Big)\Big] = 2\hat{\mu}_i \end{aligned}$$

Portanto, conforme (2.12), temos como resíduo componente da *deviance* para a Poisson

$$r_i^D = \begin{cases} sinal(y_i - \hat{\mu}_i)\sqrt{2}\left[y_i \log\left(\frac{y_i}{\hat{\mu}_i}\right) - (y_i - \hat{\mu}_i)\right]^{\frac{1}{2}} & \text{se } y_i > 0 \\ sinal(y_i - \hat{\mu}_i)\sqrt{2}\Big[\hat{\mu}_i\Big]^{\frac{1}{2}} & \text{se } y_i = 0 \end{cases} \tag{3.12}$$

IV. **Resíduo componente da *deviance* padronizado**: Para utilizar os resíduos componente da *deviance* de forma padronizada, basta dividirmos os resíduos (3.12) por $(1 - \hat{h}_{ii})^{1/2}$, onde $w_i = \hat{\mu}_i$, citado anteriormente em (3.9).

4 DISTRIBUIÇÃO POISSON-GAMA

Na literatura também é conhecida como distribuição de Pascal, quando r é um natural positivo, ou Binomial Negativa, quando r é um real positivo, esta distribuição apresenta formas especiais que foram surgidas por Pascal e Fermat (DEMÉTRIO; CORDEIRO, 2008).

Esta distribuição pode representar o número necessário de ensaios até se obter um número fixo de sucessos, dado que r é um número inteiro. Em 1907, Gosset usou a distribuição Binomial Negativa para modelar dados de contagem no lugar da distribuição de Poisson, visto que ela apresenta uma maior flexibilidade por não possuir a propriedade de equidispersão.

Esta distribuição também foi obtida por Greenwood e Yule (1920) como consequência da suposição de modelos a propensão de acidentes. Os autores consideraram que o número de acidentes seguia uma distribuição de Poisson, com parâmetro λ, em que λ variava de acordo com uma distribuição Gama, de parâmetros α e β.

Assim, considerando Y uma variável aleatória com distribuição Gama, com parâmetros α e β, é possível expressar a função de probabilidade de $Y \sim \text{Gama}(\alpha, \beta)$ como

$$g(y; \alpha, \beta) = \frac{\left(\frac{\beta}{\alpha}\right)^{\beta}}{\Gamma(\beta)} y^{\beta-1} \exp\left\{-\frac{y\beta}{\alpha}\right\}, \quad y > 0 \tag{4.1}$$

em que $\alpha, \beta > 0$. A esperança e variância são dadas, respectivamente, por

$$\begin{aligned} \mathrm{E}(Y) &= \alpha \\ \mathrm{Var}(Y) &= \frac{\alpha^2}{\beta}. \end{aligned} \tag{4.2}$$

Assumindo que a variável aleatória condicional de Y dado o parâmetro λ segue uma distribuição Poisson, $Y|\lambda \sim \text{Pois}(\lambda)$, e que o parâmetro em si, λ, segue uma distribuição gama, $\lambda \sim \text{Gama}(\alpha, \beta)$, a função de probabilidade marginal de Y pode ser obtida utilizando a condicional e a distribuição do parâmetro, ou seja

$$f(y) = \int_{\lambda} f(Y|\lambda) g(\lambda) d\lambda.$$

Substituindo-se pelas respectivas funções de probabilidades denotadas em (4.1) e (3.1), obtém-se a marginal

$$
\begin{aligned}
f(y) &= \int_0^\infty \frac{e^{-\lambda}\lambda^y}{y!} \frac{\left(\frac{\beta}{\alpha}\right)^\beta}{\Gamma(\beta)} \lambda^{\beta-1} \exp\left\{-\frac{\beta\lambda}{\alpha}\right\} d\lambda = \frac{1}{y!} \frac{\left(\frac{\beta}{\alpha}\right)^\beta}{\Gamma(\beta)} \int_0^\infty e^{-\lambda}\lambda^y\lambda^{\beta-1} \exp\left\{-\frac{\beta\lambda}{\alpha}\right\} d\lambda \\
&= \frac{1}{y!\Gamma(\beta)} \left(\frac{\beta}{\alpha}\right)^\beta \int_0^\infty e^{-\lambda}\lambda^{y+\beta-1} e^{-\left(\frac{\beta\lambda}{\alpha}\right)} d\lambda = \frac{1}{y!\Gamma(\beta)} \left(\frac{\beta}{\alpha}\right)^\beta \int_0^\infty \lambda^{y+\beta-1} e^{-\lambda\left(\frac{\beta}{\alpha}+1\right)} d\lambda,
\end{aligned}
$$

e dado que $\Gamma(\gamma) = \int_0^\infty t^{\gamma-1}e^{-t}dt$, então tem-se que

$$
\begin{aligned}
f(y) &= \frac{1}{y!\Gamma(\beta)} \left(\frac{\beta}{\alpha}\right)^\beta \frac{\Gamma(y+\beta)}{\left(\frac{\beta}{\alpha}+1\right)^{y+\beta}} \\
&= \frac{\Gamma(y+\beta)}{\Gamma(y+1)\Gamma(\beta)} \left(\frac{\beta}{\alpha}\right)^\beta \left(\frac{\alpha}{\beta+\alpha}\right)^\beta \left(\frac{\alpha}{\beta+\alpha}\right)^y.
\end{aligned}
$$

Portanto

$$
f(y;\alpha,\beta) = \frac{\Gamma(y+\beta)}{\Gamma(y+1)\Gamma(\beta)} \left(\frac{\beta}{\alpha+\beta}\right)^\beta \left(\frac{\alpha}{\alpha+\beta}\right)^y \tag{4.3}
$$

com $y = 0, 1, \ldots,$ α e β positivos. Assim, Y segue uma distribuição Poisson-Gama, $Y \sim \mathrm{PG}(\alpha, \beta)$, ou binomial negativa.

Considerando uma reestruturação da função de probabilidade (4.3), sendo

$$
p = \frac{\alpha}{\alpha+\beta} \quad \text{e} \quad 1-p = \frac{\beta}{\alpha+\beta}
$$

se obtem a $\mathrm{PG}(\alpha, \beta)$ na forma

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

Na Figura 2 é apresentado o comportamento da distribuição Poisson-Gama (PG) assumindo $\beta = 10$ e diferentes valores para p.

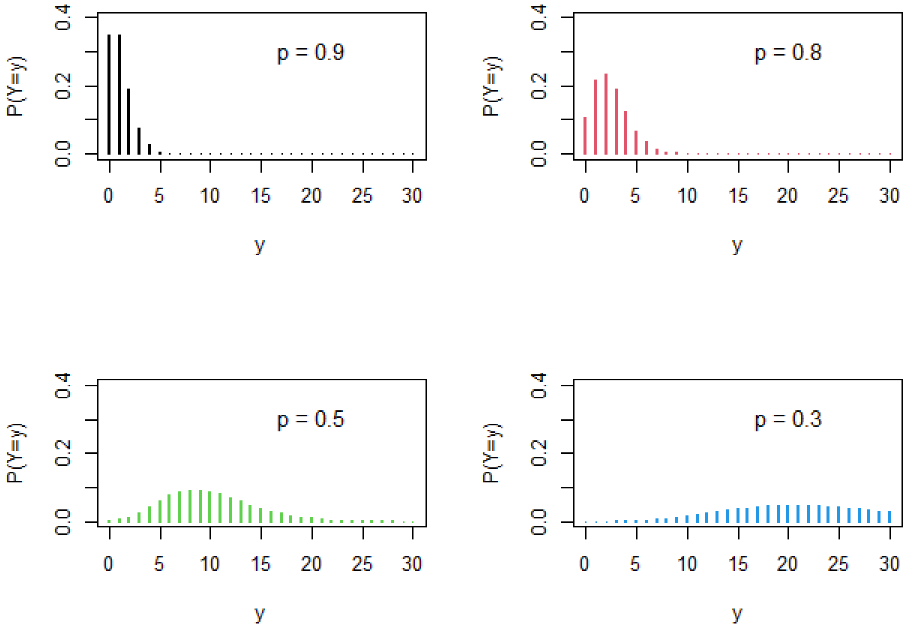


Figura 2 – Função de probabilidade da distribuição PG para diferentes valores de p

4.1 Propriedades

Propriedade 4.1.1 *Se $Y \sim PG(\alpha, \beta)$, então o valor esperado e variância são dados por α e $\alpha + \alpha^2/\beta$, respectivamente.*

Prova: Dado que a distribuição PG é uma distribuição composta, é possível encontrar o valor esperado de Y a partir da propriedade da esperança condicional, levando em consideração que $\lambda \sim \text{Gama}(\alpha, \beta)$ com média e variância dados conforme (4.2)

$$\mathrm{E}(Y) = \mathrm{E}[\mathrm{E}(Y|\lambda)] = \mathrm{E}(\lambda) = \alpha,$$

e a variância de Y pode ser escrita através da propriedade da variância e esperança condicionais

$$\begin{aligned}\mathrm{Var}(Y) &= \mathrm{E}[\mathrm{Var}(Y|\lambda)] + \mathrm{Var}[\mathrm{E}(Y|\lambda)] = \mathrm{E}[\lambda] + \mathrm{Var}(\lambda) \\ &= \mathrm{E}(\lambda) + \mathrm{Var}(\lambda) = \alpha + \frac{\alpha^2}{\beta}.\end{aligned}$$

Observa-se que a variância de Y, escrita em função do valor esperado, cresce mais rapidamente com relação à média do que para a distribuição Poisson. Isso permite uma maior flexibilidade entre média e variância. Sendo esse um dos modelos adequados para estudar a superdispersão para dados de contagem.

4.2 Modelo Poisson-Gama

Sejam $Y_1, \dots, Y_n$ variáveis condicionalmente independentes, dado um vetor de variáveis explicativas $\boldsymbol{x}_i^\top = (x_{i1}, x_{i2}, \dots, x_{ip})$, $i = 1, \dots, n$. Considerando que a distribuição condicional $Y_i|\boldsymbol{x}_i^\top$ segue uma PG(μ_i, ϕ), a função de probabilidade de $Y_i|\boldsymbol{x}_i^\top$, conforme (4.3), é definida por

$$f(y_i|\boldsymbol{x}_i^\top) = \frac{\Gamma(y_i+\phi)}{\Gamma(y_i+1)\Gamma(\phi)} \left(\frac{\phi}{\mu_i+\phi}\right)^{\phi} \left(\frac{\mu_i}{\mu_i+\phi}\right)^{y_i}, \tag{4.4}$$

com $y_i = 0, 1, \dots$ e $\mu_i, \phi > 0$. A função de probabilidade definida (4.4) pertence a família exponencial uniparamétrica, sendo a Binomial Negativa (ou PG) um caso particular dos modelos lineares generalizados.

É comum, assim como no modelo Poisson, adotar como função de ligação a funções logarítmica, $g(\mu_i) = log(\mu_i)$; raiz quadrada, $g(\mu_i) = (\mu_i)^{1/2}$, e identidade, $g(\mu_i) = \mu_i$, sendo esta não tão usada pois pode retornar um problema de estimação com restrição, visto que $\mu_i > 0$. Entretanto, utiliza-se preferencialmente a função logarítmica devido à possibilidade de comparação de forma direta com a regressão Poisson, além de uma fácil interpretação das estimativas dos parâmetros do modelo.

Tendo em vista que a distribuição incondicional da variável resposta, dado um conjunto de variáveis regressora, segue uma $PG(\mu_i, \phi)$, a esperança e variância do modelo podem ser expressos, utilizando-se uma função de ligação logarítmica, respectivamente por:

$$\mathrm{E}(Y_i|x_i^T) = \mu_i = e^{\boldsymbol{x}_i^\top \boldsymbol{\beta}} \tag{4.5}$$

e

$$\mathrm{Var}(Y|x_i^T) = \frac{\mu_i^2}{\phi} + \mu_i = \mu_i\left(\frac{\mu_i}{\phi} + 1\right). \tag{4.6}$$

4.3 Resíduos

Utilizando-se alguns resíduos abordados na seção (2.4.1), a seguir são apresentados alguns destes adaptados para a forma do modelo Poisson-Gama.

Sendo uma distribuição também pertencente à família exponencial de distribuição, seu parâmetro de dispersão $a(\phi)$ é 1. Dessa forma, a função de variância, $V(\hat{\mu}_i)$, assim como no caso da Poisson, fica expressa na forma da variância do modelo, $\mathrm{Var}(\hat{y}_i)$.

I. **Resíduos de Pearson**: Utilizando a variância para a PG (4.6), o resíduo de Pearson fica descrito como

$$r_i^P = \frac{y_i - \hat{\mu}_i}{\sqrt{\frac{\hat{\mu}_i^2}{\phi} + \hat{\mu}_i}} = \frac{\sqrt{\phi}(y_i - \hat{\mu}_i)}{\sqrt{\hat{\mu}_i^2 + \hat{\mu}_i \phi}}. \tag{4.7}$$

II. **Resíduos de Pearson padronizado**: Para a composição dos resíduos de Pearson padronizado é necessário utilizarmos a matriz de pesos **W**. Considerando a função de ligação logarítmica, o i-ésimo elemento de **W**, é composto como

$$\begin{aligned} w_i &= \left[\left(\frac{\partial \eta_i}{\partial \mu_i}\right)^2 V(\mu_i)\right]^{-1} = \left[\left(\frac{\partial \log(\mu_i)}{\partial \mu_i}\right)^2 V(y_i)\right]^{-1} \\ &= \left[\left(\frac{1}{\mu_i}\right)^2 \mu_i \left(\frac{\mu_i}{\phi} + 1\right)\right]^{-1} = \left[\frac{1}{\mu_i}\left(\frac{\mu_i}{\phi} + 1\right)\right]^{-1} \\ &= \left[\frac{1}{\phi} + \frac{1}{\mu_i}\right]^{-1} = \frac{\phi \mu_i}{\mu_i + \phi}. \end{aligned} \tag{4.8}$$

Assim, o resíduo padronizado é descrito como

$$r_i^P = \frac{y_i - \hat{\mu}_i}{\sqrt{\left(\frac{\hat{\mu}_i^2}{\phi} + \hat{\mu}_i\right)(1 - \hat{h}_{ii})}}, \tag{4.9}$$

em que $\hat{h}_{ii}$, utiliza o i-ésimo elemento da matriz peso **W** dado por (4.8).

III. **Resíduo componente da *deviance***: Utilizando a função de probabilidade definida em (4.4), a função log-verossimilhança do modelo Poisson-Gama fica definida por:

$$\begin{aligned} l(\boldsymbol{\mu}) &= \sum_{i=1}^{n} \log(f(y_i|\boldsymbol{x}_i^\top)) \\ &= \sum_{i=1}^{n} \left[\log\left(\frac{\Gamma(y_i + \phi)}{\Gamma(y_i + 1)\Gamma(\phi)}\right) + \phi \log\left(\frac{\phi}{\mu_i + \phi}\right) + y_i \log\left(\frac{\mu_i}{\mu_i + \phi}\right)\right] \\ &= \sum_{i=1}^{n} \left[\log\left(\frac{\Gamma(y_i + \phi)}{\Gamma(y_i + 1)\Gamma(\phi)}\right) + \phi \log \phi + y_i \log \mu_i - (\phi + y_i)\log(\mu_i + \phi)\right] \end{aligned}$$

Utilizando a definição para os desvios, obtemos

$$\begin{aligned}
d_i &= 2\left[\log\left(\frac{\Gamma(y_i+\phi)}{\Gamma(y_i+1)\Gamma(\phi)}\right) + \phi\log(\phi) + y_i\log y_i - (\phi+y_i)\log(y_i+\phi)\right] \\
&- \left[\log\left(\frac{\Gamma(y_i+\phi)}{\Gamma(y_i+1)\Gamma(\phi)}\right) + \phi\log(\phi) + y_i\log\hat{\mu}_i - (\theta+y_i)\log(\hat{\mu}_i+\phi)\right] \\
&= 2\Big[y_i\log y_i - (\phi+y_i)\log(y_i+\phi) - y_i\log\hat{\mu}_i + (\phi+y_i)\log(\hat{\mu}_i+\phi)\Big] \\
&= 2\left[y_i\log\left(\frac{y_i}{\hat{\mu}_i}\right) + \phi\log\left(\frac{\hat{\mu}_i+\phi}{y_i+\phi}\right) + y_i\log\left(\frac{\hat{\mu}_i+\phi}{y_i+\phi}\right)\right] \\
&= 2\left[\phi\log\left(\frac{\hat{\mu}_i+\phi}{y_i+\phi}\right) + y_i\log\left(\frac{y_i}{\hat{\mu}_i}\frac{(\hat{\mu}_i+\phi)}{(y_i+\phi)}\right)\right],
\end{aligned}$$

para $y_i > 0$. Caso $y_i = 0$ se obtem o desvio como sendo

$$\begin{aligned}
f(y_i = 0|\boldsymbol{x}_i^\top) &= \sum_{i=1}^{n}\left[\phi\log\phi - \phi\log(\mu_i+\phi)\right] \\
\Rightarrow d_i &= 2\Big[l(y_i;y_i) - l(\hat{\mu}_i;y_i)\Big] \\
&= 2\Big[\phi\log\phi - \phi\log(y_i+\phi) - \Big(\phi\log\phi - \phi\log(\hat{\mu}_i+\phi)\Big)\Big] \\
&= 2\Big[-\phi\log(\phi) + \phi\log(\hat{\mu}_i+\phi)\Big] \\
&= 2\phi\log\left(\frac{\hat{\mu}_i+\phi}{\phi}\right).
\end{aligned}$$

Portanto, o resíduo componente da *deviance* para a Poisson-Gamaé dado por:

$$r_i^D = \begin{cases} sinal(y_i-\hat{\mu}_i)\sqrt{2}\left[\phi\log\left(\frac{\hat{\mu}_i+\phi}{y_i+\phi}\right) + y_i\log\left(\frac{y_i}{\hat{\mu}_i}\frac{(\hat{\mu}_i+\phi)}{(y_i+\phi)}\right)\right]^{\frac{1}{2}} & \text{, se } y_i > 0 \\ sinal(y_i-\hat{\mu}_i)\sqrt{2}\left[\phi\log\left(\frac{\hat{\mu}_i+\phi}{\phi}\right)\right]^{\frac{1}{2}} & \text{, se } y_i = 0 \end{cases} \tag{4.10}$$

IV. **Resíduo componente da *deviance* padronizado**: Para utilizar os resíduos componente da *deviance* de forma padronizada, basta dividirmos os resíduos (4.10) juntamente por $(1-\hat{h}_{ii})^{1/2}$, onde w_i é dado conforme (4.8).

5 DISTRIBUIÇÃO POISSON-INVERSA GAUSSIANA

A distribuição Poisson-Inversa Gaussiana (PIG) foi proposta (HOLLA, 1967) como uma alternativa para a distribuição de Poisson para casos com superdispersão, de forma que ela consiste de uma mistura da distribuição Poisson e da distribuição Inversa Gaussiana.

Em alguns estudos de seguro e medicamento, a distribuição PIG foi proposta como uma boa alternativa pala modelar dados com superdispersão ou de cauda longa (se a probabilidade individual se torna pequena apenas após um certo número grande de y, dado a distribuição de probabilidade $P(y)$) do que a Poisson-Gama. Dessa forma, o modelo de regressão PIG é adequado para analisar a relação entre dados com variável resposta de contagem associada à covariáveis até mesmo em casos de superdipersão (PUTRI *et al.*, 2020).

Holla (1965) realizou um estudo acerca da distribuição Poisson-Inversa Gaussiana tanto para o caso univariado como para o caso multivariado.

Seja X uma variável aleatória com distribuição Inversa Gaussiana, ou Normal Inversa, de parâmetros μ e σ. Podemos expressar a função de probabilidade de $X \sim \mathrm{IG}(\mu, \sigma)$ como

$$g(x; \mu, \sigma) = \sqrt{\frac{\sigma}{2\pi x^3}} \exp\left\{-\frac{\sigma(x-\mu)^2}{2\mu^2 x}\right\}, \tag{5.1}$$

em que $x > 0$ e $\mu > 0$, sendo a média. A esperança e variância da PIG são dadas, respectivamente

$$\begin{aligned} \mathrm{E}(X) &= \mu \\ \mathrm{Var}(X) &= \frac{\mu^3}{\sigma}. \end{aligned} \tag{5.2}$$

Assumindo que uma variável aleatória $Y|\lambda \sim \mathrm{Pois}(\lambda)$, dada em (3.1), e $\lambda \sim \mathrm{IG}(\mu, \sigma)$, expressa em (5.1), então, marginalmente, a função de probabilidade marginal de Y, é dada por

$$\begin{aligned} f(y) &= \int_0^\infty p(y;\lambda) g(\lambda; \mu, \sigma) \mathrm{d}\lambda \\ &= \int_0^\infty \frac{e^{-\lambda}\lambda^y}{y!} \left(\frac{\sigma}{2\pi\lambda^3}\right)^{1/2} \exp\left\{\frac{\sigma(\lambda-\mu)^2}{2\mu^2\lambda}\right\} \mathrm{d}\lambda \\ &= \left(\frac{2\sigma}{\pi}\right)^{\frac{1}{2}} \frac{1}{y!}\, e^{\sigma/\mu} \left[\frac{\sigma}{2\left(1+\frac{\sigma}{2\mu^2}\right)}\right]^{\frac{1}{2}\left(y-\frac{1}{2}\right)} K_{y-1/2}\left[\sqrt{2\sigma\left(1+\frac{\sigma}{2\mu^2}\right)}\right], \end{aligned} \tag{5.3}$$

com $y = 0, 1, \ldots$ e $\mu, \sigma > 0$. Além disso, K_α representa a função de Bessel, obtida segundo a solução de uma equação diferencial de segunda ordem (HOLLA, 1967). Portanto, temos que (5.3) representa a função de probabilidade de de $Y \sim PIG(\mu, \sigma)$.

A Figura 3 apresenta o comportamento da distribuição Poisson-Inversa Gaussiana assumindo diferentes valores para μ e σ.

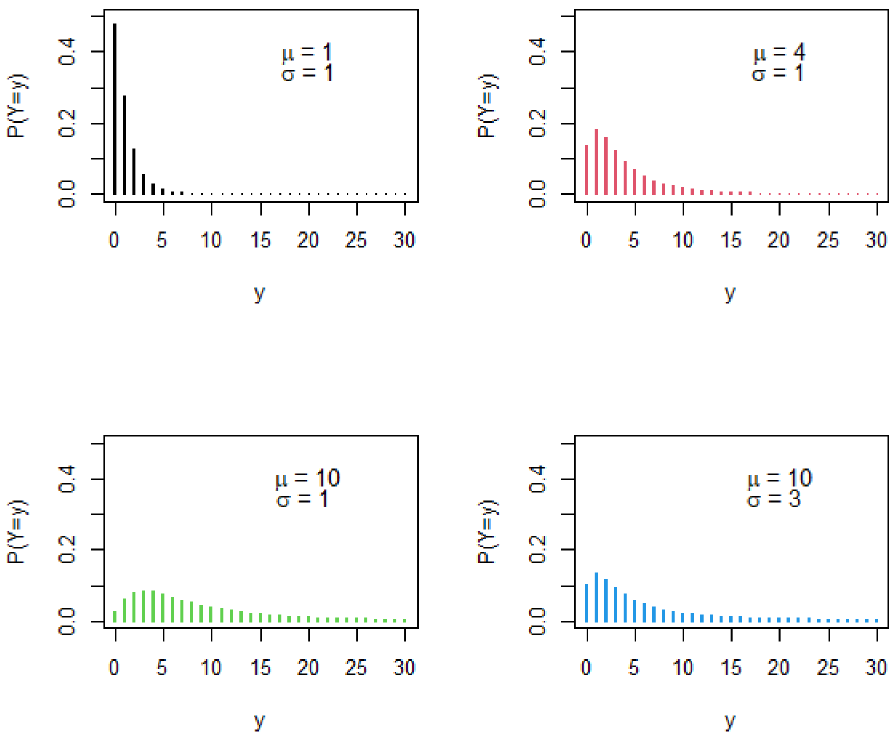


Figura 3 – Função de probabilidade da distribuição PIG para diferentes valores de μ e σ

5.1 Propriedades

Propriedade 5.1.1 *Considere $Y \sim PIG(\mu, \sigma)$, então o valor esperado e variância são dados por μ e $\mu + \mu^3/\sigma$, respectivamente.*

Prova: Dado que a distribuição PIG é uma distribuição composta, o valor esperado de Y vai ser expresso partindo da propriedade da esperança condicional. Da mesma forma, e levando em consideração que $\lambda \sim \mathrm{IG}(\mu, \sigma)$, a variância de Y pode ser escrita a partir da esperança e variância condicionais de $(Y|\lambda)$, da forma seguinte:

$$\mathrm{E}(Y) = \mathrm{E}[\mathrm{E}(Y|\lambda)] = \mathrm{E}(\lambda) = \mu$$

e

$$\begin{aligned}\mathrm{Var}(Y) &= \mathrm{E}[\mathrm{Var}(Y|\lambda)] + \mathrm{Var}[\mathrm{E}(Y|\lambda)] \\ &= \mathrm{E}(\lambda) + \mathrm{Var}(\lambda) = \mu + \frac{\mu^3}{\sigma} = \mu\left(\frac{\mu^2}{\sigma} + 1\right).\end{aligned}$$

5.2 Modelo Poisson-Inversa Gaussiana

Considere $Y_1, \ldots, Y_n$ variáveis condicionalmente independentes, dado um vetor de variáveis explicativas $\boldsymbol{x}_i^\top$. Assumindo que a distribuição condicional $Y_i|\boldsymbol{x}_i^\top$ segue uma PIG(μ_i, σ), a função de probabilidade, conforme (5.3), é definida por

$$f(y_i|\boldsymbol{x}_i^\top) = \left(\frac{2\sigma}{\pi}\right)^{\frac{1}{2}} \frac{1}{y_i!}\, e^{\sigma/\mu_i} \left[\frac{\sigma}{2\left(1+\frac{\sigma}{2\mu_i^2}\right)}\right]^{\frac{1}{2}\left(y_i - \frac{1}{2}\right)} K_{y_i - 1/2}\left[\sqrt{2\sigma\left(1 + \frac{\sigma}{2\mu_i^2}\right)}\right] \qquad (5.4)$$

em que $y_i = 0, 1, \ldots$ e $\mu_i, \sigma > 0$.

A esperança e variância da distribuição incondicional da resposta, associada ao modelo, pode ser expressa, utilizando uma função de ligação logarítmica, respectivamente

$$\mathrm{E}(Y_i|\boldsymbol{x}_i^\top) = \mu_i = e^{\boldsymbol{x}_i^\top \boldsymbol{\beta}} \qquad (5.5)$$

e

$$\mathrm{Var}(Y_i|\boldsymbol{x}_i^\top) = \frac{\mu_i^3}{\sigma} + \mu_i = \mu_i\left(\frac{\mu_i^2}{\sigma} + 1\right) \;. \qquad (5.6)$$

5.3 Resíduos

Para a obtenção dos resíduos, é utilizada uma adequação da função de variância, $\mathrm{Var}(\hat{\mu}_i)$, onde se utiliza a própria variância do modelo, conforme (5.6).

I. **Resíduos de Pearson**: Com base na forma geral (2.8), e a variância para a PIG (5.6), o resíduo de Pearson fica escrito como

$$r_i^P = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\mu}_i\left(\frac{\hat{\mu}_i^2}{\sigma} + 1\right)}} = \frac{\sqrt{\sigma}(y_i - \hat{\mu}_i)}{\sqrt{\hat{\mu}_i(\hat{\mu}_i^2 + \sigma)}} \qquad (5.7)$$

II. **Resíduos de Pearson padronizado**: Considerando a função de ligação logarítmica, $\eta_i = \log(\mu_i)$, o i-ésimo elemento de $\mathbf{W}$, em (2.3), é composto na forma seguinte:

$$\begin{aligned} w_i &= \left[\left(\frac{\partial \eta_i}{\partial \mu_i}\right)^2 V(\mu_i)\right]^{-1} = \left[\left(\frac{\partial \log(\mu_i)}{\partial \mu_i}\right)^2 V(y_i)\right]^{-1} \\ &= \left[\left(\frac{1}{\mu_i}\right)^2 \mu_i \left(\frac{\mu_i^2}{\sigma} + 1\right)\right]^{-1} = \left[\frac{\mu_i}{\sigma} + \frac{1}{\mu_i}\right]^{-1} \\ &= \frac{\mu_i \sigma}{\mu_i^2 + \sigma}. \end{aligned} \tag{5.8}$$

Assim, o resíduo padronizado é descrito como

$$r_i^P = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\mu}_i(1 - \hat{h}_{ii})}}, \tag{5.9}$$

com a matriz $\widehat{\mathbf{H}}$, utilizando o i-ésimo elemento da matriz pesos dado por (5.8).

III. **Resíduo componente da *deviance***: Considerando a função de probabilidade (5.4), a função de log-verossimilhança para o modelo PIG pode ser escrita na forma abaixo:

$$\begin{aligned} l(\boldsymbol{\mu}, \sigma) &= \sum_{i=1}^{n} \log\left(f(y_i; \mu_i, \sigma)\right) \\ &= \sum_{i=1}^{n} \left\{ \log\left(\frac{2\sigma}{\pi}\right)^{1/2} + (-\log y_i!) + \log e^{\sigma/\mu_i} + \log\left[\frac{\sigma}{2\left(1 + \frac{\sigma}{2\mu_i^2}\right)}\right]^{\frac{1}{2}(y_i - 1/2)} \right. \\ &\quad \left. + \log \mathrm{K}_{y_i - 1/2}\left[\sqrt{2\sigma\left(1 + \frac{\sigma}{2\mu_i^2}\right)}\right]\right\} \\ &= \sum_{i=1}^{n} \left\{ \frac{1}{2}\log\left(\frac{2\sigma}{\pi}\right) - \log y_i! + \frac{\sigma}{\mu_i} + \frac{1}{2}\left(y_i - \frac{1}{2}\right)\log\sigma \right. \\ &\quad \left. - \frac{1}{2}\left(y_i - \frac{1}{2}\right)\log\left[2\left(1 + \frac{\sigma}{2\mu_i^2}\right)\right] + \log \mathrm{K}_{y_i - 1/2}\left[\sqrt{2\sigma\left(1 + \frac{\sigma}{2\mu_i^2}\right)}\right]\right\}. \end{aligned}$$

O desvio será dado de forma a considerar a log-verossimilhança do modelo estimado ($\hat{\mu}_i = \hat{\mu}_i$) e do modelo saturado ($\tilde{\mu}_i = y_i$), definido como:

$$\begin{aligned}
d_i &= 2[l(y_i, y_i) - l(\hat{\mu}_i, y_i)] \\
&= 2\left[\frac{\sigma}{y_i} - \frac{1}{2}\left(y_i - \frac{1}{2}\right)\log\left[2\left(1 + \frac{\sigma}{2y_i^2}\right)\right] + \log \mathrm{K}_{y_i - 1/2}\left[\sqrt{2\sigma\left(1 + \frac{\sigma}{2y_i^2}\right)}\right]\right. \\
&\quad - \left.\left(\frac{\sigma}{\hat{\mu}_i} - \frac{1}{2}\left(y_i - \frac{1}{2}\right)\log\left[2\left(1 + \frac{\sigma}{2\hat{\mu}_i^2}\right)\right] + \log \mathrm{K}_{y_i - 1/2}\left[\sqrt{2\sigma\left(1 + \frac{\sigma}{2\hat{\mu}_i^2}\right)}\right]\right)\right] \\
&= 2\left\{\sigma\left(\frac{1}{y_i} + \frac{1}{\hat{\mu}_i}\right) + \left(\frac{y_i}{2} - \frac{1}{4}\right)\left(\log\left[2\left(1 + \frac{\sigma}{2\hat{\mu}_i^2}\right)\right] - \log\left[2\left(1 + \frac{\sigma}{2y_i^2}\right)\right]\right)\right. \\
&\quad \left. + \mathrm{K}_{y_i - 1/2}(y_i) - \mathrm{K}_{y_i - 1/2}(\hat{\mu}_i)\right\},
\end{aligned} \tag{5.10}$$

em que $\mathrm{K}_{y_i-1/2}(\alpha)$ é dado por

$$\mathrm{K}_{y_i-1/2}(\alpha) = \mathrm{K}_{y_i-1/2}\left[\sqrt{2\sigma\left(1 + \frac{\sigma}{2\alpha^2}\right)}\right].$$

Portanto, temos como resíduo *deviance* para a Poisson-Inversa Gaussiana

$$r_i^D = sinal(y_i - \hat{\mu}_i)\sqrt{d_i}, \tag{5.11}$$

em que d_i é expresso por (5.10).

IV. **Resíduo componente da *deviance* padronizado**: Para utilizar os resíduos componente da *deviance* de forma padronizada, basta dividirmos os resíduos (5.11) por $(1 - \hat{h}_{ii})^{1/2}$, onde $\hat{h}_{ii}$ é o elemento da diagonal da matriz de projeção $\widehat{\mathbf{H}}$, em (2.10), sendo w_i definido em (5.8).

6 DISTRIBUIÇÃO POISSON-LINDLEY GENERALIZADA

A distribuição Lindley foi introduzida por Lindley (1958), para estudos realizados sobre vida útil. Segundo Ghitany *et al.* (2011), esta distribuição pode ser especialmente útil para modelagens em estudos de mortalidade.

Assim como o caso da distribuição PG, que pode ser expressa por meio de uma mistura entre as distribuições Poisson e Gama, alguns autores sugeriram ramificações dessa natureza em relação a distribuição Lindley. Podemos citar duas, a Poisson-Lindley introduzida por Sankaran (1970) e a Poisson-Lindley Generalizada apresentada por Wongrin e Bodhisuwan (2016). Esta última sendo uma mistura entre a distribuição Poisson e a distribuição Lindley Generalizada apresentada por Elbatal *et al.* (2013).

A distribuição Lindley Generalizada é uma distribuição para dados de natureza contínua, especificada por três parâmetros (α, β e θ) e, dado sua maior flexibilidade, comparada à Lindley, ela é bastante usada como alternativa para dados de tempo de vida. Conforme Elbatal et al (2013), sua função densidade de probabilidade é apresentada da seguinte forma:

$$g(x;\alpha,\beta,\theta) = \frac{1}{\theta+1}\left(\frac{\theta^{\alpha+1}x^{\alpha-1}}{\Gamma(\alpha)} + \frac{\theta^{\beta}x^{\beta-1}}{\Gamma(\beta)}\right)e^{-\theta x} \tag{6.1}$$

em que $x > 0$ e $\alpha,\beta,\theta > 0$, denotada por $X \sim \mathrm{NGL}(X;\alpha,\beta,\theta)$, intitulada como "New Generalized Lindley" por Wongrin e Bodhisuwan (2016).

Conforme Elbatal *et al.* (2013), sua esperança e variância são expressas por

$$\mathrm{E}(X) = \frac{\alpha\theta + \beta(\alpha+1)}{(\theta+\beta)\theta} \tag{6.2}$$

e

$$\mathrm{Var}(X) = \frac{(\alpha+1)[\alpha\theta + \beta(\alpha+2)]}{(\theta+\beta)\theta^2}\ . \tag{6.3}$$

Como comentado, a Poisson-Lindley Generalizada é uma mistura entre a distribuição (3.1) e (6.1). Assim, seja $Y|\lambda$ uma variável aleatória seguindo uma distribuição de Poisson com parâmetro λ. Se $\lambda \sim \mathrm{NGL}(\alpha,\beta,\theta)$, então Y, marginalmente, tem uma distribuição Poisson-Lindley Generalizada (PLG). A função de probabilidade de Y pode ser encontrada utilizando-se a definição da marginal, dada por:

$$
\begin{aligned}
f(y) &= \int_0^\infty p(y|\lambda) g(\lambda;\alpha,\beta;\theta)\ \mathrm{d}\lambda \\
&= \int_0^\infty \frac{e^{-\lambda}\lambda^y}{y!}\frac{1}{\theta+1}\left(\frac{\theta^{\alpha+1}\lambda^{\alpha-1}}{\Gamma(\alpha)}+\frac{\theta^{\beta}\lambda^{\beta-1}}{\Gamma(\beta)}\right) e^{-\theta\lambda}\ \mathrm{d}\lambda \\
&= \frac{1}{y!(\theta+1)}\int_0^\infty e^{-\lambda}\lambda^y\left(\frac{\theta^{\alpha+1}\lambda^{\alpha-1}}{\Gamma(\alpha)}+\frac{\theta^{\beta}\lambda^{\beta-1}}{\Gamma(\beta)}\right) e^{-\theta\lambda}\mathrm{d}\lambda \\
&= \frac{1}{y!(\theta+1)}\int_0^\infty e^{-\lambda(\theta+1)}\lambda^y\left(\frac{\theta^{\alpha+1}\lambda^{\alpha-1}}{\Gamma(\alpha)}+\frac{\theta^{\beta}\lambda^{\beta-1}}{\Gamma(\beta)}\right)\mathrm{d}\lambda \\
&= \frac{1}{y!(\theta+1)}\left(\int_0^\infty e^{-\lambda(\theta+1)}\lambda^y\frac{\theta^{\alpha+1}\lambda^{\alpha-1}}{\Gamma(\alpha)}\mathrm{d}\lambda+\int_0^\infty e^{-\lambda(\theta+1)}\lambda^y\frac{\theta^{\beta}\lambda^{\beta-1}}{\Gamma(\beta)}\mathrm{d}\lambda\right) \\
&= \frac{1}{y!(\theta+1)}\left(\frac{\theta^{\alpha+1}}{\Gamma(\alpha)}\int_0^\infty e^{-\lambda(\theta+1)}\lambda^{y+\alpha-1}\mathrm{d}\lambda+\frac{\theta^{\beta}}{\Gamma(\beta)}\int_0^\infty e^{-\lambda(\theta+1)}\lambda^{y+\beta-1}\mathrm{d}\lambda\right).
\end{aligned}
$$

Utilizando-se da integral do núcleo de uma Gama(α,β), com a parametrização mais comumente conhecida,

$$\int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}\mathrm{d}x = \frac{\Gamma(\alpha)}{\beta^\alpha},$$

tem-se,

$$
\begin{aligned}
f(y) &= \frac{1}{y!(\theta+1)}\left(\frac{\theta^{\alpha+1}}{\Gamma(\alpha)}\frac{\Gamma(y+\alpha)}{(\theta+1)^{y+\alpha}}+\frac{\theta^{\beta}}{\Gamma(\beta)}\frac{\Gamma(y+\beta)}{(\theta+1)^{y+\beta}}\right) \\
&= \frac{1}{y!(\theta+1)}\left(\frac{1}{(\theta+1)^y}\frac{\theta^{\alpha+1}\Gamma(y+\alpha)}{(\theta+1)^\alpha\Gamma(\alpha)}+\frac{1}{(\theta+1)^y}\frac{\theta^{\beta}\Gamma(y+\beta)}{(\theta+1)^\beta\Gamma(\beta)}\right).
\end{aligned}
$$

Assim, tem-se que Y segue uma distribuição Poisson-Lindley Generalizada de parâmetros α, β e θ, positivos, $Y \sim \mathrm{PLG}(\alpha,\beta,\theta)$, com função de probabilidade

$$f(y;\alpha,\beta,\theta) = \frac{1}{y!(\theta+1)^{y+1}}\left[\left(\frac{\theta}{\theta+1}\right)^\alpha\frac{\theta\Gamma(y+\alpha)}{\Gamma(\alpha)}+\left(\frac{\theta}{\theta+1}\right)^\beta\frac{\Gamma(y+\beta)}{\Gamma(\beta)}\right], \tag{6.4}$$

com $y = 0, 1, \ldots$ e $\alpha,\beta,\theta > 0$.

6.1 Propriedades

Propriedade 6.1.1 *Seja $Y \sim \mathrm{PLG}(\alpha,\beta,\theta)$, o k-ésimo momento fatorial pode ser denotado por*

$$\mathrm{E}[(Y)_k] = \frac{1}{\theta+1}\left(\frac{\Gamma(k+\alpha)}{\Gamma(\alpha)\theta^{k-1}}+\frac{\Gamma(k+\beta)}{\Gamma(\beta)\theta^k}\right). \tag{6.5}$$

Prova: O k-ésimo momento fatorial pode ser encontrado de forma que

$$E[(Y)_k] = \int_0^\infty \sum_{y=0}^\infty y^k\ \frac{e^{-\lambda}\lambda^y}{y!}\ \lambda^k\ g(\lambda)\ \mathrm{d}\lambda$$

Entretanto, $\sum_{y=0}^{\infty} y^k \frac{e^{-\lambda}\lambda^y}{y!}$ corresponde ao k-ésimo momentonão central de uma Poisson. Assim, obtém-se

$$E[(Y)_k] = \int_0^{\infty} \lambda^k g(\lambda)\mathrm{d}\lambda \ .$$

Substituindo pela função de probabilidade (6.4) e utilizando a propriedade da função gama, obtemos

$$\begin{aligned}
E[(Y)_k] &= \int_0^{\infty} \lambda^k \frac{1}{\theta+1}\left(\frac{\theta^{\alpha+1}\lambda^{\alpha-1}}{\Gamma(\alpha)} + \frac{\theta^{\beta}\lambda^{\beta-1}}{\Gamma(\beta)}\right) e^{-\theta\lambda}\mathrm{d}\lambda \\
&= \frac{1}{\theta+1}\left(\int_0^{\infty} \lambda^k \frac{\theta^{\alpha+1}\lambda^{\alpha-1}}{\Gamma(\alpha)}\mathrm{d}\lambda + \int_0^{\infty} \lambda^k \frac{\theta^{\beta}\lambda^{\beta-1}}{\Gamma(\beta)} e^{-\theta\lambda}\mathrm{d}\lambda\right) \\
&= \frac{1}{\theta+1}\left(\frac{\theta^{\alpha+1}}{\Gamma(\alpha)}\int_0^{\infty} \lambda^{k+\alpha-1}e^{-\theta\lambda}\mathrm{d}\lambda + \frac{\theta^{\beta}}{\Gamma(\beta)}\int_0^{\infty} \lambda^{k+\beta-1}e^{-\theta\lambda}\mathrm{d}\lambda\right) \\
&= \frac{1}{\theta+1}\left(\frac{\theta^{\alpha+1}}{\Gamma(\alpha)}\frac{\Gamma(k+\alpha)}{\theta^{k+\alpha}} + \frac{\theta^{\beta}}{\Gamma(\beta)}\frac{\Gamma(k+\beta)}{\theta^{k+\beta}}\right) \\
&= \frac{1}{\theta+1}\left(\frac{\theta^{\alpha}\theta}{\Gamma(\alpha)}\frac{\Gamma(k+\alpha)}{\theta^{k}\theta^{\alpha}} + \frac{\theta^{\beta}}{\Gamma(\beta)}\frac{\Gamma(k+\beta)}{\theta^{k}\theta^{\beta}}\right) \\
&= \frac{1}{\theta+1}\left(\frac{\Gamma(k+\alpha)}{\Gamma(\alpha)\theta^{k-1}} + \frac{\Gamma(k+\beta)}{\Gamma(\beta)\theta^{k}}\right).
\end{aligned}$$

Propriedade 6.1.2 *Supondo $Y \sim \mathrm{PLG}(\alpha, \beta, \theta)$, o valor esperado de Y e variância são*

$$\mathrm{E}(Y) = \frac{\theta\alpha + \beta}{\theta(\theta+1)} \tag{6.6}$$

$$\mathrm{Var}(Y) = \frac{\beta[\theta(-2\alpha+\theta+2)+1] + \alpha\theta[\alpha+(\theta+1)^2] + \beta^2\theta}{\theta^2(\theta+1)^2} \ . \tag{6.7}$$

Prova: Defini-se o r-ésimo momento não central em termos do momento fatorial

$$\mathrm{E}(Y^r) = \sum_{k=0}^{r} \left\{ {n \atop k} \right\} \mathrm{E}[(Y)_k]$$

em que

$$\left\{ {n \atop k} \right\} = \frac{1}{k!} \sum_{i=0}^{k} (-1)^i \binom{k}{i} (k-i)^n \ .$$

denotada como Número de Stirling de segundo tipo. Para encontrar o primeiro e segundo momento não central, substituí-se $r = 1$ e $r = 2$, obtendo, respectivamente

$$\begin{aligned}
\mathrm{E}(Y) &= \sum_{k=0}^{1} \left\{ {1 \atop k} \right\} \mathrm{E}[(Y)_k] \\
&= \left\{ {1 \atop 0} \right\} \mathrm{E}[(Y)_0] + \left\{ {1 \atop 1} \right\} \mathrm{E}[(Y)_1] \\
&= \mathrm{E}[(Y)_1]
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}(Y^2) &= \sum_{k=0}^{2} \left\{ {2 \atop k} \right\} \mathrm{E}[(Y)_k] \\
&= \left\{ {2 \atop 0} \right\} \mathrm{E}[(Y)_0] + \left\{ {2 \atop 1} \right\} \mathrm{E}[(Y)_1] + \left\{ {2 \atop 2} \right\} \mathrm{E}[(Y)_1] \\
&= \mathrm{E}[(Y)_1] + \mathrm{E}[(Y)_2]
\end{aligned}$$

Substituindo a equação (6.5) para $k = 1$ e $k = 2$, tem-se

$$\begin{aligned}
\mathrm{E}[(Y)_1] &= \frac{1}{\theta+1}\left[\frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} + \frac{\Gamma(\beta+1)}{\theta\Gamma(\beta)}\right] \\
&= \frac{1}{\theta+1}\left[\frac{\alpha!}{(\alpha-1)!} + \frac{\beta!}{\theta(\beta-1)!}\right] \\
&= \frac{1}{\theta+1}\left[\alpha + \frac{\beta}{\theta}\right] = \frac{\alpha}{\theta+1} + \frac{\beta}{\theta(\theta+1)} = \frac{\theta\alpha+\beta}{\theta(\theta+1)} \ .
\end{aligned}$$

$$\begin{aligned}
E[(Y)_2] &= \frac{1}{\theta+1}\left[\frac{\Gamma(\alpha+2)}{\Gamma(\alpha)\theta} + \frac{\Gamma(\beta+2)}{\Gamma(\beta)\theta^2}\right] \\
&= \frac{1}{\theta+1}\left[\frac{(\alpha+1)\Gamma(\alpha+1)}{\Gamma(\alpha)\theta} + \frac{(\beta+1)\Gamma(\beta+1)}{\Gamma(\beta)\theta^2}\right] \\
&= \frac{1}{\theta+1}\left[\frac{(\alpha+1)\alpha!}{\theta(\alpha-1)!} + \frac{(\beta+1)\beta!}{\theta^2(\beta-1)!}\right] \\
&= \frac{\theta(\alpha+1)\alpha + (\beta+1)\beta}{\theta^2(\theta+1)} \ .
\end{aligned}$$

Assim, para o valor esperado de Y, com base no momento fatorial

$$\mathrm{E}(Y) = \frac{\theta\alpha+\beta}{\theta(\theta+1)} \ .$$

e para o 2º momento não central

$$\begin{aligned}
\mathrm{E}(Y^2) &= \frac{\theta\alpha+\beta}{\theta(\theta+1)} + \frac{\theta(\alpha+1)\alpha + (\beta+1)\beta}{\theta^2(\theta+1)} \\
&= \frac{\theta^2\alpha + \theta\beta + \theta\alpha^2 + \theta\alpha + \beta^2 + \beta}{\theta^2(\theta+1)}
\end{aligned}$$

Portanto, sabendo que $\mathrm{Var}(Y) = E(Y^2) - [E(Y)]^2$, a variância de Y é dada como

$$\begin{aligned}
\mathrm{Var}(Y) &= E(Y^2) - [E(Y)]^2 \\
&= \frac{\beta[\theta(-2\alpha+\theta+2)+1] + \alpha\theta[\alpha+(\theta+1)^2] + \beta^2\theta}{\theta^2(\theta+1)^2} \ .
\end{aligned}$$

Um método comum para estimarmos os parâmetros de uma distribuição é pela maximização da função de log-verossimilhança. Dado que $Y_i \sim \mathrm{PLG}(\alpha, \beta, \theta)$ a função de log-verossimilhança, conforme Wongrin e Bodhisuwan (2016), é dada por

$$
\begin{aligned}
\mathrm{l}(y_i;\alpha,\beta,\theta) &= \log\prod_{i=1}^{n} f(y_i;\alpha,\beta,\theta)\\
&= \sum_{i=1}^{n}\ln\left\{\frac{1}{y_i!(\theta+1)^{y_i+1}}\left[\left(\frac{\theta}{\theta+1}\right)^{\alpha}\frac{\theta\Gamma(y_i+\alpha)}{\Gamma(\alpha)}+\left(\frac{\theta}{\theta+1}\right)^{\beta}\frac{\Gamma(y_i+\beta)}{\Gamma(\beta)}\right]\right\}\\
&= \sum_{i=1}^{n}\ln\Big(\theta^{\alpha+1}(\theta+1)^{\beta}\Gamma(\beta)\Gamma(y_i+\alpha)+\theta^{\beta}(\theta+1)^{\alpha}\Gamma(\alpha)\Gamma(y_i+\beta)\Big)-\sum_{i=1}^{n}\ln y_i!\\
&\quad -\left(\sum_{i=1}^{n}y_i+n\right)\ln(\theta+1)-n\ln\Gamma(\alpha)-n\ln\Gamma(\beta)-n(\alpha+\beta)\ln(\theta+1).
\end{aligned} \tag{6.8}
$$

Realizando a primeira derivada em relação a cada parâmetro, obteremos as expressões dos componentes do vetor escore. Os estimadores de máxima verossimilhança (EMV) podem ser obtidos igualando cada função escore a zero e resolvendo o sistema de equações. Entretanto, as funções escores obtidas são não lineares e não tem uma solução analítica. Como consequência, a estimação de máxima verossimilhança deve ser obtida considerando métodos de aproximação numérica.

Propriedade 6.1.3 *De acordo com Wongrin e Bodhisuwan (2016), a distribuição Poisson-Lindley (Sankaran, 1970) pode ser obtida a partir da distribuição $PLG(\alpha,\beta,\theta)$, definida em 6.4, considerando $\alpha = 1$ e $\beta = 2$. Assim, $Y \sim PL(\theta)$, cuja função de probabilidade é dada por*

$$
f(y;\theta) = \frac{\theta^2(y+\theta+2)}{(\theta+1)^{y+3}}\ ,
$$

em que $y = 0, 1, 2, \ldots$ e $\theta > 0$.

Seu valor esperado e variância são expressos, respectivamente, como

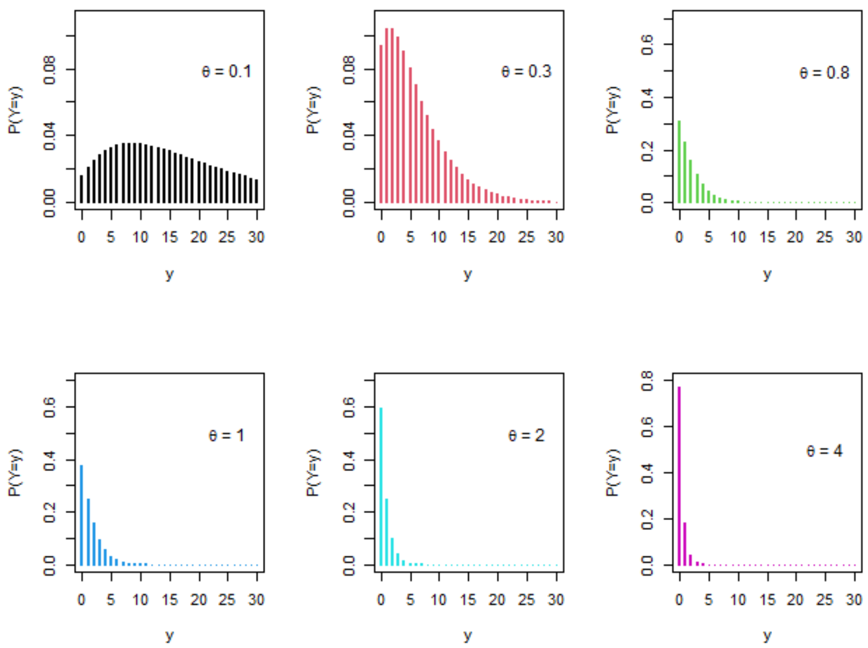
$$
\mathrm{E}(Y) = \frac{\theta+2}{\theta(\theta+1)}
$$

e

$$
\mathrm{Var}(Y) = \frac{\theta^3+4\theta^2+6\theta+2}{\theta^2(\theta+1)^2}\ .
$$

A Figura 4 abaixo, apresenta o comportamento da distribuição Poisson-Lindley para diferentes valores para θ.

Figura 4 – Função de probabilidade da distribuição PL para diferentes valores de θ



6.2 Distribuição Poisson-Lindley Generalizada com dois parâmetros

A distribuição Poisson-Lindley Generalizada como visto, possui três parâmetros. Entretanto, assim como a Poisson-Lindley, que é uma simplificação utilizando apenas um parâmetro, θ, a Poisson-Lindley Generalizada pode ser reduzida a dois parâmetros, o que facilita a parte computacional e prática para a aplicação desta distribuição.

Primeiramente, consideramos a distribuição Lindley generalizada (LG) apresentada por Mahmoudi e Zekerzadeh (2010), onde eles focaram na redução de parâmetros da distribuição Lindley generalizada expressa por Zakerzadeh e Dolati (2009). Considerando os parâmetros α e θ, positivos, e $\beta = 1$, obtemos a expressão da função de probabilidade da distribuição Lindley com dois parâmetros como sendo

$$g(y; \alpha, \beta = 1, \theta) = \frac{1}{\theta + 1} \frac{\theta^{\alpha+1} y^{\alpha-1}(\alpha + y)}{\Gamma(\alpha + 1)} e^{-\theta y} \ , \tag{6.9}$$

em que $y > 0$ e $\alpha, \theta > 0$, denotada como $Y \sim \mathrm{GL}(Y; \alpha, \beta = 1, \theta)$. O valor esperado e a variância são denotados, respectivamente, por:

$$\mathrm{E}(Y) = \frac{\alpha(\theta + 1) + 1}{\theta(\theta + 1)} \ . \tag{6.10}$$

$$\mathrm{Var}(Y) = \frac{(\alpha + 1)[\alpha(\theta + 1) + 2]}{\theta^2(\theta + 1)} - \left[\frac{\alpha(\theta + 1) + 1}{\theta(\theta + 1)}\right]^2 \ . \tag{6.11}$$

A função de probabilidade marginal de Y, pode ser expressa por meio da integral da função de probabilidade da Poisson (3.1) juntamente com a GL (6.9), onde $Y|\lambda \sim Pois(\lambda)$ e $\lambda \sim GL(\alpha, \beta = 1, \theta)$, sendo especificada por:

$$\begin{aligned}
f(y;\alpha,\beta=1,\theta) &= \int_0^\infty p(y;\lambda)g(y;\alpha,\beta=1,\theta)\mathrm{d}\lambda \\
&= \int_0^\infty \frac{e^{-\lambda}\lambda^y}{y!}\frac{1}{\theta+1}\frac{\theta^{\alpha+1}\lambda^{\alpha-1}(\alpha+\lambda)}{\Gamma(\alpha+1)}e^{-\theta\lambda}\mathrm{d}\lambda \\
&= \frac{\theta^{\alpha+1}}{y!(\theta+1)\Gamma(\alpha+1)}\int_0^\infty e^{-\lambda}\lambda^y\lambda^{\alpha-1}(\alpha+\lambda)e^{-\theta\lambda}\mathrm{d}\lambda \\
&= \frac{\theta^{\alpha+1}}{y!(\theta+1)\Gamma(\alpha+1)}\int_0^\infty (\alpha+\lambda)e^{-\lambda(\theta+1)}\lambda^{y+\alpha-1}\mathrm{d}\lambda \\
&= \frac{\theta^{\alpha+1}}{y!(\theta+1)\Gamma(\alpha+1)}\left[\frac{\alpha\Gamma(y+\alpha)}{(\theta+1)^{y+\alpha}}+\frac{\Gamma(y+\alpha+1)}{(\theta+1)^{y+\alpha+1}}\right].
\end{aligned}$$

Portanto, temos que a marginal de Y segue uma distribuição Poisson-Lindley Generalizada com parâmetros α e θ, ou seja, $Y \sim PLG(\alpha, \beta = 1, \theta)$, com função de probabilidade

$$f(y;\alpha,\theta) = \frac{\theta^{\alpha+1}\Gamma(y+\alpha)[\alpha(\theta+1)+(y+\alpha)]}{y!\Gamma(\alpha+1)(\theta+1)^{y+\alpha+2}}\ , \tag{6.12}$$

com $y = 0, 1, \ldots$ e $\alpha, \theta > 0$.

A Figura 5 apresenta o comportamento da distribuição Poisson-Lindley generalizada com diferentes valores de α e θ.

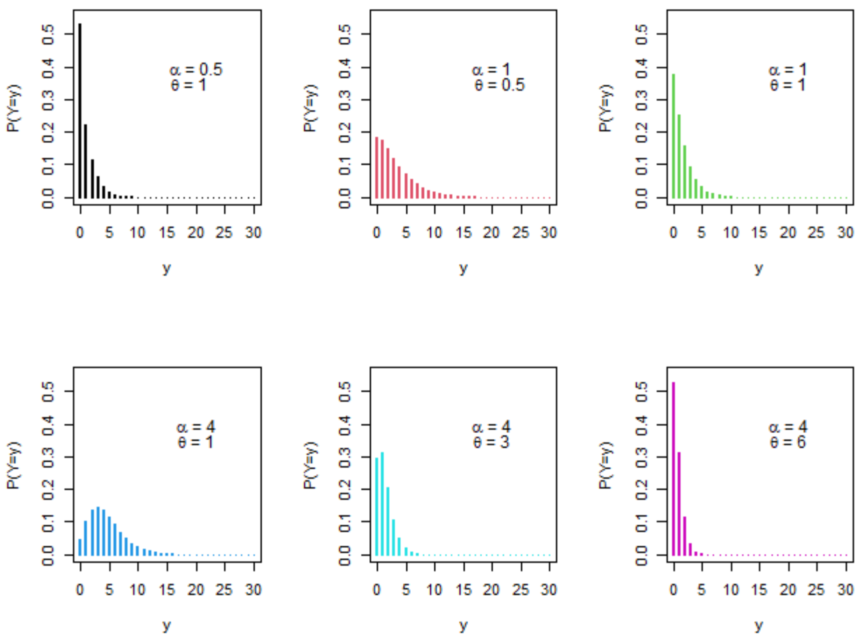


Figura 5 – Função de probabilidade da distribuição PLG para diferentes valores de α e θ

6.2.1 Propriedades

Propriedade 6.2.1 *Suponha que $Y \sim PLG(\alpha, \beta = 1, \theta)$, o valor esperado e a variância podem ser expressos como*

$$\mathrm{E}(Y) = \frac{\alpha(\theta+1)+1}{\theta(\theta+1)} \ . \tag{6.13}$$

e

$$\mathrm{Var}(Y) = \frac{(\alpha+1)[\alpha(\theta+1)+2]}{\theta^2(\theta+1)} - \left[\frac{\alpha(\theta+1)+1}{\theta(\theta+1)}\right]^2 + \frac{\alpha(\theta+1)+1}{\theta(\theta+1)} \ . \tag{6.14}$$

Podemos obter as expressões, da esperaça e variância da $PLG(\alpha, \beta = 1, \theta)$ utilizando as propriedades de esperança e variância condicionais, considerando que $\lambda \sim GL(\alpha, \beta = 1, \theta)$

$$\mathrm{E}(Y) = \mathrm{E}[\mathrm{E}(Y|\lambda)] = \mathrm{E}(\lambda) = \frac{\alpha(\theta+1)+1}{\theta(\theta+1)} \ .$$

e

$$\begin{aligned}\mathrm{Var}(Y) &= \mathrm{Var}[\mathrm{E}(Y|\lambda)] + \mathrm{E}[\mathrm{Var}(Y|\lambda)] = \mathrm{Var}(\lambda) + \mathrm{E}(\lambda)\\ &= \frac{(\alpha+1)[\alpha(\theta+1)+2]}{\theta^2(\theta+1)} - \left[\frac{\alpha(\theta+1)+1}{\theta(\theta+1)}\right]^2 + \frac{\alpha(\theta+1)+1}{\theta(\theta+1)} \ .\end{aligned}$$

Propriedade 6.2.2 *Considere que $Y \sim PLG(\alpha, \beta = 1, \theta)$, a função log-verossimilhança a ser maximizada, é dada por*

$$\begin{aligned}l(y;\alpha,\theta) &= \sum_{i=1}^{n}\left\{\log\left[\frac{\theta^{\alpha+1}\Gamma(y_i+\alpha)[\alpha(\theta+1)+(y_i+\alpha)]}{y_i!\Gamma(\alpha+1)(\theta+1)^{y_i+\alpha+2}}\right]\right\}\\ &= \sum_{i=1}^{n}\left\{\log\left[\theta^{\alpha+1}\Gamma(y_i+\alpha)\Big[\alpha(\theta+1)+(y_i+\alpha)\Big]\right] - \log\Big[y_i!\Gamma(\alpha+1)(\theta+1)^{y_i+\alpha+2}\Big]\right\}\\ &= \sum_{i=1}^{n}\Big\{\log\theta^{\alpha+1} + \log\Gamma(y_i+\alpha) + \log[\alpha(\theta+1)+(y_i+\alpha)] - \log y_i! - \log\Gamma(\alpha+1)\\ &\quad - \log(\theta+1)^{y_i+\alpha+1}\Big\}\\ &= \sum_{i=1}^{n}\Big\{\log\Gamma(y_i+\alpha) - \log y_i! - \log\Gamma(\alpha+1) + (\alpha+1)\log\theta - (y_i+\alpha+2)\log(\theta+1)\\ &\quad + \log\Big[\alpha(\theta+1)+(y_i+\alpha)\Big]\Big\}.\end{aligned} \tag{6.15}$$

Propriedade 6.2.3 *A função densidade acumulada de $Y \sim PLG(\alpha, \beta = 1, \theta)$ pode ser escrita em função de uma mistura de duas binomiais negativas, denotado como:*

$$F(y;\alpha,\beta=1,\theta)=\frac{\theta}{1+\theta}\ G_1(y)+\frac{1}{1+\theta}\ G_2(y)\ , \tag{6.16}$$

em que $G_1(y)$ e $G_2(y)$ são função acumuladas de uma binomial negativa de parâmetros $r=\alpha$ e $r=\alpha+1$, respectivamente, e $p=\frac{\theta}{1+\theta}$.

Seja a função de probabilidade dada conforme (6.12), podemos expressá-la como:

$$\frac{\theta^{\alpha+1}\alpha\Gamma(y+\alpha)}{y!(\theta+1)^{y+\alpha+1}\Gamma(\alpha+1)}+\frac{\theta^{\alpha+1}(\alpha+y)\Gamma(y+\alpha)}{y!(\theta+1)^{y+\alpha+2}\Gamma(\alpha+1)}\ .$$

Considere uma binomial negativa com $r>0$ e $0<p<1$, ou seja, $BN(r,p)$. A sua função de probabilidade por ser expressa como:

$$P(y)=\frac{\Gamma(y+r)}{y!\Gamma(r)}p^r(1-p)^y\ .$$

Se admitirmos que $r=\alpha$ e $p=\frac{\theta}{1+\theta}$, obtém-se:

$$\begin{aligned}g_1(y)&=\frac{\Gamma(y+\alpha)}{y!\Gamma(\alpha)}\left(\frac{\theta}{1+\theta}\right)^{\alpha}\left(\frac{1}{1+\theta}\right)^{y}\\&=\frac{\alpha\Gamma(y+\alpha)}{y!\Gamma(\alpha+1)}\frac{\theta^{\alpha}}{(1+\theta)^{\alpha+y}}\ .\end{aligned} \tag{6.17}$$

Admitindo que $r=\alpha+1$ e $p=\frac{\theta}{1+\theta}$, obtém-se:

$$\begin{aligned}g_2(y)&=\frac{\Gamma(y+\alpha+1)}{y!\Gamma(\alpha+1)}\left(\frac{\theta}{1+\theta}\right)^{\alpha+1}\left(\frac{1}{1+\theta}\right)^{y}\\&=\frac{(\alpha+y)\Gamma(y+\alpha)}{y!\Gamma(\alpha+1)}\frac{\theta^{\alpha+1}}{(1+\theta)^{\alpha+1+y}}\ .\end{aligned} \tag{6.18}$$

Assim, utilizando as parametrizações (6.17) e (6.18) e, multiplicando cada uma por $\frac{\theta}{1+\theta}$ e $\frac{1}{1+\theta}$, respectivamente, encontramos a função de probabilidade da PLG:

$$\begin{aligned}p(y;\alpha,\beta=1,\theta)&=\frac{\theta}{1+\theta}\ g_1(y)+\frac{1}{1+\theta}\ g_2(y)\\&=\frac{\theta}{(1+\theta)}\frac{\alpha\Gamma(y+\alpha)}{y!\Gamma(\alpha+1)}\frac{\theta^{\alpha}}{(1+\theta)^{\alpha+y}}+\frac{1}{(1+\theta)}\frac{(\alpha+y)\Gamma(y+\alpha)}{y!\Gamma(\alpha+1)}\frac{\theta^{\alpha+1}}{(1+\theta)^{\alpha+1+y}}\\&=\frac{\alpha\Gamma(y+\alpha)}{y!\Gamma(\alpha+1)}\frac{\theta^{\alpha+1}}{(1+\theta)^{\alpha+y+1}}+\frac{(\alpha+y)\Gamma(y+\alpha)}{y!\Gamma(\alpha+1)}\frac{\theta^{\alpha+1}}{(1+\theta)^{\alpha+2+y}}\\&=\frac{\theta^{\alpha+1}\alpha\Gamma(y+\alpha)}{y!\Gamma(\alpha+1)(1+\theta)^{\alpha+y+1}}+\frac{\theta^{\alpha+1}(\alpha+y)\Gamma(y+\alpha)}{y!\Gamma(\alpha+1)(1+\theta)^{\alpha+y+2}}\ ,\ y=0,1,\ldots\end{aligned}$$

Essa é uma propriedade importante, visto que na utilização de resíduos quantílicos tem-se a necessidade da geração de números aleatórios. E poder reescrever a PLG com base em uma distribuição conhecida, simplifica computacionalmente.

6.2.2 Modelo Poisson-Lindley Generalizada com dois parâmetros

Seja $Y_1, \ldots, Y_n$ as variáveis respostas dado um conjunto de covariáveis $\boldsymbol{x}_i^\top$. Assumindo que a distribuição condicional de Y_i dado $\boldsymbol{x}_i^\top$ segue uma distribuição PLG$(\alpha, \beta = 1, \theta)$ com parâmetros α e $\theta > 0$, temos como valor esperado do modelo, conforme (6.13)

$$\begin{aligned} \mathrm{E}(Y_i|\boldsymbol{x}_i^\top) &= \mathrm{E}\left\{\mathrm{E}\left[(Y_i|\boldsymbol{x}_i^\top)|\lambda_i\right]\right\} \\ &= \frac{\alpha_i(\theta+1)+1}{\theta(\theta+1)} \\ &= \mu_i \ . \end{aligned} \tag{6.19}$$

Os componentes da variância para o modelo PLG com dois parâmetros pode ser composta na seguinte forma, considerando (6.19) e (6.14)

$$\begin{aligned} \mathrm{Var}(Y_i|\boldsymbol{x}_i^\top) &= \mathrm{E}\left\{\mathrm{Var}\left[(Y_i|\boldsymbol{x}_i^\top)|\lambda_i\right]\right\} + \mathrm{Var}\left\{\mathrm{E}\left[(Y_i|\boldsymbol{x}_i^\top)|\lambda_i\right]\right\} \\ &= \mu_i + \frac{(\alpha_i+1)[\alpha_i(\theta+1)+2]}{\theta^2(\theta+1)} - \mu_i^2 \ . \end{aligned} \tag{6.20}$$

Assim, como os modelos anteriormente citados, no modelo PLG a função de ligação logarítmica é adotada para fazer a ligação entre a média da variável resposta e as covariáveis. Reparametrizando a média da distribuição PLG com $\mu_i = e^{\boldsymbol{x}_i^\top \boldsymbol{\beta}}$ temos:

$$\begin{aligned} \frac{\alpha_i(\theta+1)+1}{\theta(\theta+1)} = \mu_i &\Rightarrow\ \alpha_i(\theta+1)+1 = \mu_i\theta(\theta+1) \\ &\Rightarrow\ \alpha_i(\theta+1) = \mu_i\theta(\theta+1) - 1 \\ &\Rightarrow \alpha_i = \frac{\mu_i\theta(\theta+1)-1}{\theta+1} \ , \end{aligned}$$

portanto,

$$\alpha_i = \frac{e^{\boldsymbol{x}_i^\top \boldsymbol{\beta}}\theta(\theta+1)-1}{\theta+1} \ . \tag{6.21}$$

Portanto, a função de probabilidade de $Y_i|\boldsymbol{x}_i^\top \sim PLG(\alpha, \theta)$ pode ser expressa na forma de um modelo linear por substituir α_i na função de probabilidade

$$\begin{aligned} f(Y_i|\boldsymbol{x}_i^\top) = \frac{\Gamma\left(y_i + \frac{e^{\boldsymbol{x}_i^\top \boldsymbol{\beta}}\theta(\theta+1)-1}{\theta+1}\right)}{y_i!\Gamma\left(\frac{e^{\boldsymbol{x}_i^\top \boldsymbol{\beta}}\theta(\theta+1)-1}{\theta+1}+1\right)} \times \frac{\theta^{\frac{e^{\boldsymbol{x}_i^\top \boldsymbol{\beta}}\theta(\theta+1)-1}{\theta+1}+1}}{(\theta+1)^{y_i+\frac{e^{\boldsymbol{x}_i^\top \boldsymbol{\beta}}\theta(\theta+1)-1}{\theta+1}+2}} \times \Bigg[e^{\boldsymbol{x}_i^\top \boldsymbol{\beta}}\theta(\theta+1) - 1 \\ + y_i + \frac{e^{\boldsymbol{x}_i^\top \boldsymbol{\beta}}\theta(\theta+1)-1}{\theta+1}\Bigg]. \end{aligned} \tag{6.22}$$

6.2.3 *Resíduos*

Para o modelo PLG, a função de variância para os resíduos também será utilizada em função da variância do modelo.

I. **Resíduos de Pearson**: Sendo (6.20) a variância para a PLG, o resíduo de Pearson é expresso como:

$$r_i^P = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\mu}_i + \frac{(\hat{\alpha}_i+1)[\hat{\alpha}_i(\theta+1)+2]}{\theta^2(\theta+1)} - \hat{\mu}_i^2}}\ . \tag{6.23}$$

II. **Resíduos de Pearson padronizado**: Novamente considerando a função de ligação logarítmica, o i-ésimo elemento de $\mathbf{W}$, é composto como

$$\begin{aligned} w_i &= \left[\left(\frac{\partial \eta_i}{\partial \mu_i}\right)^2 V(\mu_i)\right]^{-1} = \left[\left(\frac{\partial \log(\mu_i)}{\partial \mu_i}\right)^2 V(y_i)\right]^{-1} \\ &= \left[\left(\frac{1}{\mu_i}\right)^2 \left(\mu_i + \frac{(\alpha_i+1)[\alpha_i(\theta+1)+2]}{\theta^2(\theta+1)} - \mu_i^2\right)\right]^{-1} \\ &= \left[\frac{1}{\mu_i} + \frac{(\alpha_i+1)[\alpha_i(\theta+1)+2]}{\mu_i^2\theta^2(\theta+1)} - 1\right]^{-1} \\ &= \left[\frac{\mu_i\theta^2(1+\theta) + (\alpha_i+1)[\alpha_i(\theta+1)+2] - \mu_i^2\theta^2(1+\theta)}{\mu_i^2\theta^2(\theta+1)}\right]^{-1} \\ &= \frac{\mu_i^2\theta^2(\theta+1)}{\mu_i\theta^2(\theta+1)(1-\mu_i) + (\alpha+1)[\alpha_i(\theta+1)+2]}\ . \end{aligned} \tag{6.24}$$

Assim, o resíduo padronizado é descrito como

$$r_i^P = \frac{y_i - \hat{\mu}_i}{\sqrt{\left(\hat{\mu}_i + \frac{(\hat{\alpha}_i+1)[\hat{\alpha}_i(\theta+1)+2]}{\theta^2(\theta+1)} - \hat{\mu}_i^2\right)(1-\hat{h}_{ii})}}\ . \tag{6.25}$$

III. **Resíduo componente da *deviance***: Seja $\Theta = (\boldsymbol{\beta}^\top, \theta)^\top$ um vetor de parâmetros. A função de log-verossimilhança para o modelo pode ser escrita da seguinte forma, considerando a substituição de α_i, dado por 6.21

$$
\begin{aligned}
l(\Theta) = \sum_{i=0}^{n} \Bigg\{ & \log\left[\Gamma\left(y_i + \frac{\mu_i\theta(\theta+1)-1}{\theta+1}\right)\right] - \log y_i! - \log\left[\Gamma\left(1 + \frac{\mu_i\theta(\theta+1)-1}{\theta+1}\right)\right] \\
& + \left(\frac{\mu_i\theta(\theta+1)-1}{\theta+1} + 1\right)\log\theta - \left(y_i + \frac{\mu_i\theta(\theta+1)-1}{\theta+1} + 2\right)\log(\theta+1) \\
& + \log\left[\mu_i\theta(\theta+1) - 1 + y_i + \frac{\mu_i\theta(\theta+1)-1}{\theta+1}\right]\Bigg\}, \quad \text{com} \quad \mu_i = \exp(\boldsymbol{x}_i^\top\boldsymbol{\beta}).
\end{aligned}
$$

O desvio será dado de forma a considerar a log-verossimilhança do modelo estimado e do modelo saturado ($\tilde{\mu}_i = y_i$), de forma que:

$$d_i = 2[l(y_i, y_i) - l(\hat{\mu}_i, y_i)].$$

IV. **Resíduo componente da *deviance* padronizado**: Para utilizar os resíduos componente da *deviance* de forma padronizada, basta dividirmos os resíduos *deviance* juntamente por $\sqrt{(1-\hat{h}_{ii})}$, sendo w_i dado conforme (6.24).

7 APLICAÇÕES

7.1 Aplicação I - Abelhas

Os dados considerados foram retirados de Marciano (2009) e Mendes (2017). Um estudo foi realizado na área de apicultura, com o intuito de verificar o número de abelhas que polinizam determinada espécie de planta no decorrer do tempo. Para isto, realizou-se quatro coletas em um intervalo de tempo variável segundo a hora do dia. Os horários das coletas forma: 4,5,6,8,10,12,14,16 e 18 horas, tendo no total 36 observações. Os dados foram ajustado considerando a variável resposta o número de abelhas coletando pólen.

Na Tabela 1 está apresentado um resumo geral do comportamento descritivo do número de abelhas coletando polens nos intervalos considerados.

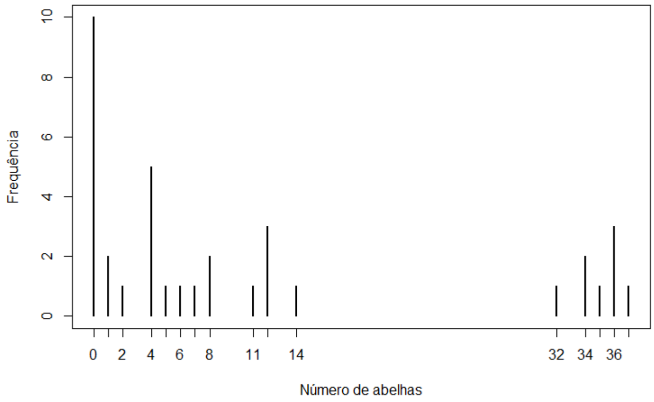
Tabela 1 – Medidas descritivas para o número de abelhas coletando polens

	Mín	1ª Q.	Mediana	Média	3ª Q.	Máx	Variância
Número de abelhas	0,0	0,0	4,5	11,1	12,5	37,0	184,8

Através desta Tabela 1 é possível verificar que o número mínimo de abelhas observadas corresponde a zero, algo que ocorre até o primeiro quartil, ou seja, pelo menos 25% dos dados. Além disso, tem-se uma quantidade máxima de 37 abelhas, elevando assim a variabilidade dos dados. Ao comparar a variância com a média apresentada dos dados, vemos que a variância é bem superior à média amostral ($s^2 = 184,8 > \bar{x} = 11,1$), caracterizando um conjunto de dados com superdispersão, influenciado pelo excesso de zeros.

No Figura 6 obtemos a visualização gráfica da frequência do número de abelhas. É possível observar que assim como na tabela descritiva, é perceptível que existe uma grande frequência de valores baixos e alguns valores muito altos.

Figura 6 – Número de abelhas coletando pólen segundo o tempo, em horas, observado



Por meio do Figura 7 obtemos a comparação da quantidade dos valores observados em relação aos valores estimados para cada uma das distribuições. Além disso, na Tabela 2 são apresentados os valores do valor-p em relação ao teste Anderson-Darling (AD) para cada distribuição estimada. Observe que as distribuições Poisson-Gama e Poisson-Lindley Generalizada obtiveram os maiores p-valores, indicando uma melhor qualidade de ajuste através dessas distribuições.

Figura 7 – Número de abelhas coletando pólen segundo o tempo, em horas

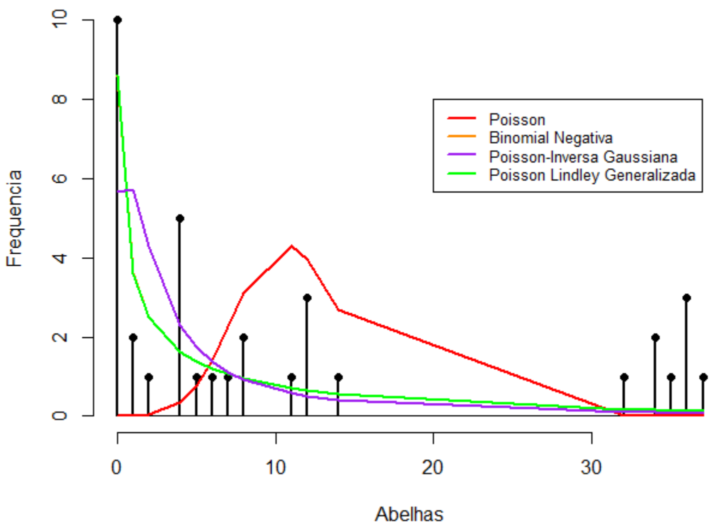


Tabela 2 – Teste AD em relação aos valores estimados para a frequência de abelhas

	P	PG	PIG	PLG
Valor-p	<0,001	0,6095	0,0895	0,6063

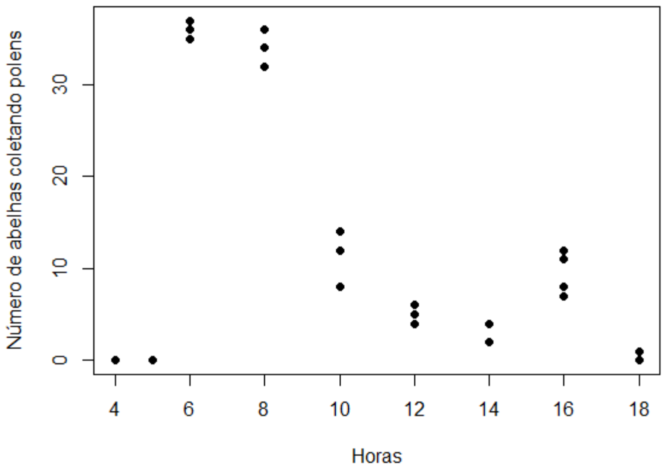
7.1.1 Modelagem dos dados

Na Figura 8, observamos o comportamento da variável resposta, número de abelhas coletando pólen, dado a covariável tempo. Conforme a figura temos que o comportamento não segue uma tendência linear nem quadrática. Assim, segundo o modelo definido por Marciano (2009), ajustou-se um modelo cúbico para o conjunto de dados. Como função de ligação, será utilizada a logarítmica,

$$log(\mu_i) = \beta_0 + \beta_1 t_i + \beta_2 t_i^2 + \beta_3 t_i^3$$

sendo t a hora observada, com $i = 1, \ldots, 9$, que são os nove diferentes tempos considerados no experimento.

Figura 8 – Número de abelhas coletando pólen segundo o tempo, em horas, observado



Afim de obter o melhor ajuste, usou-se os modelos Poisson, Poisson-Gama, Poisson-Inversa Gaussiana e Poisson-Lindley Generalizada. Utilizou-se o *software* R com as seguintes funções *glm*(), *glm.nb*(), *gamlss*() e *optim*() para a modelagem dos dados, respectivamente.

Como o modelo Poisson-Lindley Generalizada não possui função pronta no *software* R para estimação, foi utilizada a função *optim* atráves do método *SANN*, que consiste em utilizarmos a verossimilhança para a estimação dos parâmetros do modelo (ver detalhes em Béslisle, 1992). Como as estimativas variam de acordo com os valores iniciais propostos, escolheu-se recorrer como valores iniciais para o vetor de β's as estimativas dadas pelos modelos Poisson, Poisson-Gama e Poisson-Inversa Gaussiana e, para o parâmetro θ, avaliou-se uma grade de parâmetros possíveis conforme feito por Mendes (2017).

As Tabelas 3, 4 e 5 apresentam os valores das estimativas para o modelo PLG considerando as estimativas de β_i de cada modelo como valores iniciais e uma grade de valores para o parâmetro θ. As linhas destacadas em negrito correspondem ao vetor de estimativas que produziu melhor ajuste, segundo os critérios AIC e BIC, para cada base de modelo.

Tabela 3 – Estimativas dos parâmetros do modelo da PLG considerando como valores iniciais os coeficientes do ajuste dado um modelo Poisson

$\theta_{inicial}$	$\hat{\theta}$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	AIC	BIC	LogVerossimilhança
1,5000	**1,5000**	**-11,1202**	**4,4863**	**-0,4293**	**0,0122**	**281,7012**	**288,0352**	**136,8506**
1,6000	1,6000	-11,1202	4,4863	-0,4293	0,0122	284,2893	290,6234	138,1446
1,7000	1,7000	-11,1202	4,4863	-0,4293	0,0122	286,6998	293,0339	139,3499
1,8000	1,8000	-11,1202	4,4863	-0,4293	0,0122	288,9505	295,2846	140,4753
1,9000	1,9000	-11,1202	4,4863	-0,4293	0,0122	291,0570	297,3911	141,5285
2,0000	2,0000	-11,1202	4,4863	-0,4293	0,0122	293,0329	299,3670	142,5165
2,1000	2,1000	-11,1202	4,4863	-0,4293	0,0122	294,8901	301,2242	143,4451
2,2000	2,2000	-11,1202	4,4863	-0,4293	0,0122	296,6392	302,9733	144,3196
2,3000	2,3000	-11,1202	4,4863	-0,4293	0,0122	298,2894	304,6235	145,1447
2,4000	2,4000	-11,1202	4,4863	-0,4293	0,0122	299,8490	306,1831	145,9245
2,5000	2,5000	-11,1202	4,4863	-0,4293	0,0122	301,3254	307,6595	146,6627
2,6000	2,6000	-11,1202	4,4863	-0,4293	0,0122	302,7252	309,0592	147,3626
2,7000	2,7000	-11,1202	4,4863	-0,4293	0,0122	304,0542	310,3883	148,0271
2,8000	2,8000	-11,1202	4,4863	-0,4293	0,0122	305,3178	311,6519	148,6589
2,9000	2,9000	-11,1202	4,4863	-0,4293	0,0122	306,5208	312,8548	149,2604

Tabela 4 – Estimativas dos parâmetros do modelo da PLG considerando como valores iniciais os coeficientes do ajuste dado um modelo Poisson-Gama

$\theta_{inicial}$	$\hat{\theta}$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	AIC	BIC	LogVerossimilhança
1,0000	**0,9976**	**-14,5834**	**5,2567**	**-0,4761**	**0,0126**	**269,7263**	**276,0603**	**130,8631**
1,1000	1,2911	-14,6479	5,2113	-0,4759	0,0131	291,0732	297,4073	141,5366
1,2000	1,2000	-14,6581	5,1268	-0,4515	0,0120	319,8205	326,1545	155,9102
1,3000	1,3000	-14,6581	5,1268	-0,4515	0,0120	325,4338	331,7679	158,7169
1,4000	1,4000	-14,6581	5,1268	-0,4515	0,0120	330,6041	336,9382	161,3021
1,5000	1,5000	-14,6581	5,1268	-0,4515	0,0120	335,3823	341,7164	163,6912
1,6000	1,4779	-14,6220	5,1488	-0,4531	0,0116	306,6840	313,0181	149,3420
1,7000	1,9082	-14,5924	5,2606	-0,4720	0,0126	315,6770	322,0111	153,8385
1,8000	1,6345	-14,8664	5,5476	-0,5157	0,0143	305,8952	312,2293	148,9476
1,9000	1,7578	-14,6214	4,9993	-0,4317	0,0113	351,0942	357,4283	171,5471
2,0000	2,1397	-14,5359	5,1481	-0,4594	0,0124	352,9966	359,3307	172,4983
2,1000	2,1000	-14,6581	5,1268	-0,4515	0,0120	357,8703	364,2044	174,9351
2,2000	2,4031	-14,8654	5,3677	-0,4925	0,0137	325,9426	332,2767	158,9713
2,3000	2,4341	-14,5723	5,2011	-0,4735	0,0131	330,5663	336,9004	161,2832
2,4000	2,5010	-14,5796	5,0359	-0,4355	0,0110	354,0897	360,4238	173,0448
2,5000	2,5000	-14,6581	5,1268	-0,4515	0,0120	368,7682	375,1023	180,3841
2,6000	2,6000	-14,6581	5,1268	-0,4515	0,0120	371,1322	377,4663	181,5661

Tabela 5 – Estimativas dos parâmetros do modelo da PLG considerando como valores iniciais os coeficientes do ajuste dado um modelo Poisson-Inversa Gaussiana

$\theta_{inicial}$	$\hat{\theta}$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	AIC	BIC	LogVerossimilhança
1,6300	1,6300	-18,0367	5,9499	-0,5023	0,0128	503,5537	509,8878	247,7768
1,6400	1,4884	-18,2734	6,4516	-0,5842	0,0155	316,6324	322,9665	154,3162
1,6500	1,8477	-18,1542	6,3106	-0,5744	0,0159	332,0304	338,3645	162,0152
1,6600	1,6600	-18,0367	5,9499	-0,5023	0,0128	505,6960	512,0300	248,8480
1,6700	1,7943	-17,9512	6,0778	-0,5363	0,0142	335,9223	342,2563	163,9611
1,6800	1,6800	-18,0367	5,9499	-0,5023	0,0128	507,0972	513,4313	249,5486
1,6900	**1,4783**	**-18,0333**	**6,2719**	**-0,5797**	**0,0161**	**305,3850**	**311,7190**	**148,6925**
1,7000	1,7000	-18,0367	5,9499	-0,5023	0,0128	508,4776	514,8117	250,2388
1,7100	1,6625	-17,8798	6,0923	-0,5400	0,0143	323,2781	329,6122	157,6391
1,7200	1,6428	-18,1498	6,2241	-0,5563	0,0150	324,7084	331,0424	158,3542
1,7300	1,6117	-18,0657	6,0839	-0,5359	0,0139	329,4586	335,7926	160,7293
1,7400	1,6817	-18,0103	6,1696	-0,5601	0,0153	316,4487	322,7828	154,2244

A Figura 9 retorna uma visualização gráfica referente ao comportamento do ajuste para cada um dos modelos. Além disso, na Tabela 6, estão apresentados os critérios Akaike (AIC) e Bayesiano (BIC) para cada um dos ajustes. Como esperado, o modelo Poisson é o que possui maior valor nos dois critérios, sendo o modelo que apresenta o pior ajuste. Sendo o modelo que menos está se adequando ao conjunto de dados.

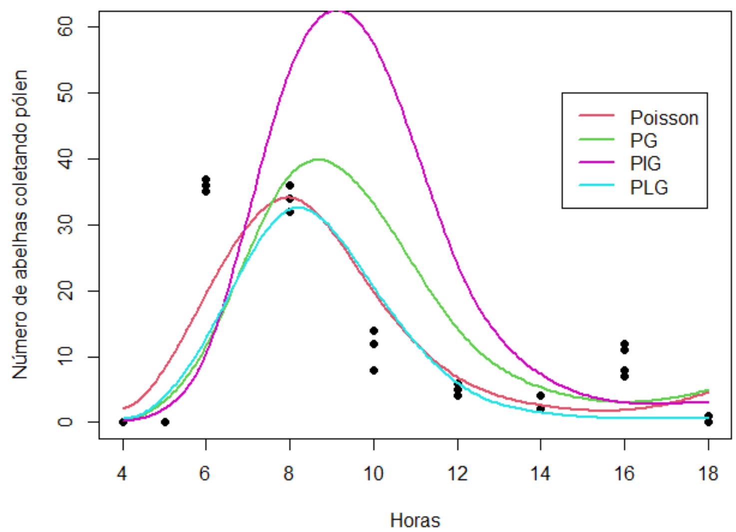


Figura 9 – Modelos do 3º grau ajustados e valores observados para o número de abelhas

Tabela 6 – Critérios AIC e BIC para cada modelo ajustado ao conjunto das abelhas

Critérios	Poisson	PG	PIG	PLG
AIC	356,80	227,78	222,87	260,58
BIC	363,14	235,70	230,79	266,92

Os modelos PG e PIG são os que possuem os menores valores de AIC e BIC.

Entretanto, comparada com a Poisson, a PLG não possui uma medida de qualidade ruim, sendo um modelo que está se adequando melhor aos dados do que a Poisson. É importante notar que apesar do modelo PIG possuir os menores valores nos critérios, ele aparenta superestimar alguns valores dos dados.

A Tabela 7 mostra as estimativas e os erros padrão dos parâmetros de cada modelo ajustado, em que cada β_i está relacionado as horas do modelo cúbico. Nota-se que as estimativas de β_1, β_2 e β_3 não diferenciam muito em relação ao modelo utilizado, observando-se mais diferença ao se tratar do intercepto. Além disso, não há mudança de sinais nas estimativas.

Tabela 7 – Estimativas e erros padrão dos ajustes para o número de abelhas que polinizaram

Parâmetros	Poisson	PG	PIG	PLG
Intercepto	$-11,120$ *** (1,119)	$-14,658$ *** (3,501)	$-18,037$ ** (6,558)	$-14,896$ *** (0,005)
β_1	$4,486$ *** (0,360)	$5,127$ *** (1,127)	$5,950$ ** (2,050)	$5,506$ *** (0,015)
β_2	$-0,429$ *** (0,036)	$-0,451$ *** (0,109)	$-0,502$ * (0,193)	$-0,513$ *** (0,003)
β_3	$0,012$ *** (0,001)	$0,012$ *** (0,003)	$0,012$ * (0,006)	$0,014$ *** (<0,001)
θ				0,996
σ/ϕ		0,835	0,898	

A Tabela 8 apresenta os valores preditos em comparação com os valores observados, com as respectivas variâncias associadas a cada modelo. Além disso, contém os resíduos de Pearson e *Deviance*, ambos padronizados. O modelo PIG contempla a maior variabilidade que o modelo Poisson, sendo seguido pelos modelos PG e PGL. Ao verificar a análise dos valores dos resíduos, os ajustes com PG, PIG e PLG aparentam ser os que estão distribuídos mais próximos de zero.

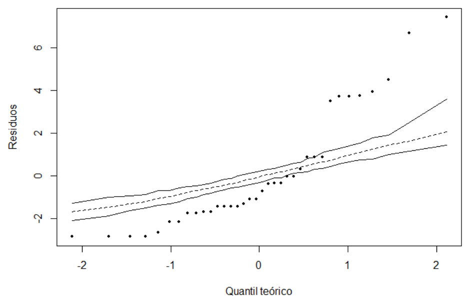
Tabela 8 – Valores observados e preditos com suas respectivas variâncias e resíduos para cada modelo ajustado

	Poisson				Poisson-Gama				Poisson-Inversa Gaussiana				Poisson-Lindley generalizada			
y_{obs}	$\hat{y}$	Var($\hat{y}$)	r_i^{P*}	r_i^{D*}	$\hat{y}$	Var($\hat{y}$)	r_i^{P*}	r_i^{D*}	$\hat{y}$	Var($\hat{y}$)	r_i^{P*}	r_i^{D*}	$\hat{y}$	Var($\hat{y}$)	r_i^{P*}	r_i^{D*}
0	2,087	2,087	-1,512	-2,138	0,546	0,904	-0,617	-0,984	0,234	0,248	-0,523	-1,022	0,695	1,642	-0,620	-1,052
0	2,087	2,087	-1,512	-2,138	0,546	0,904	-0,617	-0,984	0,234	0,248	-0,523	-1,022	0,695	1,642	-0,620	-1,052
0	2,087	2,087	-1,512	-2,138	0,546	0,904	-0,617	-0,984	0,234	0,248	-0,523	-1,022	0,695	1,642	-0,620	-1,052
0	2,087	2,087	-1,512	-2,138	0,546	0,904	-0,617	-0,984	0,234	0,248	-0,523	-1,022	0,695	1,642	-0,620	-1,052
0	8,177	8,177	-3,021	-4,272	3,298	16,313	-0,859	-1,720	2,132	12,922	-0,640	-1,765	3,971	8,203	-1,436	-2,404
0	8,177	8,177	-3,021	-4,272	3,298	16,313	-0,859	-1,720	2,132	12,922	-0,640	-1,765	3,971	8,203	-1,436	-2,404
0	8,177	8,177	-3,021	-4,272	3,298	16,313	-0,859	-1,720	2,132	12,922	-0,640	-1,765	3,971	8,203	-1,436	-2,404
0	8,177	8,177	-3,021	-4,272	3,298	16,313	-0,859	-1,720	2,132	12,922	-0,640	-1,765	3,971	8,203	-1,436	-2,404
36	19,558	19,558	3,930	3,513	11,580	172,097	1,940	1,296	10,466	1286,815	0,742	1,296	12,806	25,894	4,676	3,556
37	19,558	19,558	4,169	3,706	11,580	172,097	2,020	1,335	10,466	1286,815	0,771	1,335	12,806	25,894	4,877	3,672
35	19,558	19,558	3,691	3,318	11,580	172,097	1,861	1,257	10,466	1286,815	0,713	1,257	12,806	25,894	4,474	3,438
36	19,558	19,558	3,930	3,513	11,580	172,097	1,940	1,296	10,466	1286,815	0,742	1,296	12,806	25,894	4,676	3,556
32	34,100	34,100	-0,392	-0,396	37,549	1725,170	-0,141	-0,149	53,414	169725,159	-0,053	-0,143	32,371	65,071	-0,047	-0,062
36	34,100	34,100	0,355	0,352	37,549	1725,170	-0,039	-0,040	53,414	169725,159	-0,043	-0,038	32,371	65,071	0,457	0,505
34	34,100	34,100	-0,019	-0,019	37,549	1725,170	-0,090	-0,093	53,414	169725,159	-0,048	-0,090	32,371	65,071	0,205	0,259
34	34,100	34,100	-0,019	-0,019	37,549	1725,170	-0,090	-0,093	53,414	169725,159	-0,048	-0,090	32,371	65,071	0,205	0,259
14	19,843	19,843	-1,387	-1,465	33,205	1352,874	-0,546	-0,709	57,480	211500,933	-0,096	-0,688	20,588	41,478	-1,043	-1,056
8	19,843	19,843	-2,811	-3,199	33,205	1352,874	-0,716	-1,077	57,480	211500,933	-0,109	-1,045	20,588	41,478	-1,993	-2,351
12	19,843	19,843	-1,862	-2,011	33,205	1352,874	-0,602	-0,817	57,480	211500,933	-0,100	-0,792	20,588	41,478	-1,360	-1,453
12	19,843	19,843	-1,862	-2,011	33,205	1352,874	-0,602	-0,817	57,480	211500,933	-0,100	-0,792	20,588	41,478	-1,360	-1,453
4	6,912	6,912	-1,164	-1,265	14,277	258,262	-0,667	-0,958	24,137	15680,956	-0,164	-0,938	6,047	12,359	-0,600	-0,461
5	6,912	6,912	-0,765	-0,805	14,277	258,262	-0,603	-0,818	24,137	15680,956	-0,156	-0,801	6,047	12,359	-0,307	-0,060
4	6,912	6,912	-1,164	-1,265	14,277	258,262	-0,667	-0,958	24,137	15680,956	-0,164	-0,938	6,047	12,359	-0,600	-0,461
6	6,912	6,912	-0,365	-0,373	14,277	258,262	-0,538	-0,695	24,137	15680,956	-0,148	-0,681	6,047	12,359	-0,014	-0,066
2	2,585	2,585	-0,376	-0,392	5,322	39,225	-0,561	-0,734	7,320	443,954	-0,266	-0,733	1,505	3,265	0,289	0,552
4	2,585	2,585	0,910	0,842	5,322	39,225	-0,223	-0,243	7,320	443,954	-0,166	-0,243	1,505	3,265	1,456	1,410
4	2,585	2,585	0,910	0,842	5,322	39,225	-0,223	-0,243	7,320	443,954	-0,166	-0,243	1,505	3,265	1,456	1,410
4	2,585	2,585	0,910	0,842	5,322	39,225	-0,223	-0,243	7,320	443,954	-0,166	-0,243	1,505	3,265	1,456	1,410
12	1,861	1,861	7,616	5,067	3,067	14,322	2,473	1,546	2,967	32,035	1,726	1,596	0,583	1,418	10,559	4,109
7	1,861	1,861	3,860	2,947	3,067	14,322	1,089	0,831	2,967	32,035	0,770	0,858	0,583	1,418	5,935	3,033
11	1,861	1,861	6,865	4,675	3,067	14,322	2,196	1,420	2,967	32,035	1,535	1,465	0,583	1,418	9,634	3,914
8	1,861	1,861	4,611	3,407	3,067	14,322	1,366	0,994	2,967	32,035	0,962	1,026	0,583	1,418	6,860	3,272
1	4,627	4,627	-1,923	-2,335	4,869	33,250	-0,762	-1,134	2,974	32,254	-0,397	-1,141	0,644	1,540	0,329	0,905
1	4,627	4,627	-1,923	-2,335	4,869	33,250	-0,762	-1,134	2,974	32,254	-0,397	-1,141	0,644	1,540	0,329	0,905
0	4,627	4,627	-2,454	-3,470	4,869	33,250	-0,959	-2,035	2,974	32,254	-0,598	-2,047	0,644	1,540	-0,595	-1,008
0	4,627	4,627	-2,454	-3,470	4,869	33,250	-0,959	-2,035	2,974	32,254	-0,598	-2,047	0,644	1,540	-0,595	-1,008

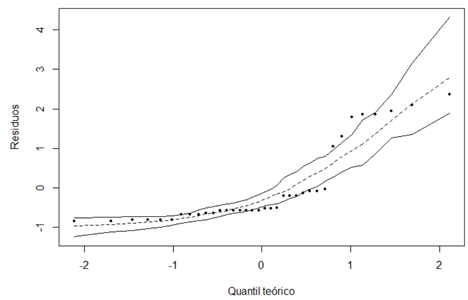
Os gráficos apresentados na Figura 10 são referentes aos resíduos de Pearson padronizados para cada modelo ajustado. Como citado, os envelopes para modelos que não tem distribuição normal, são gerados a partir do modelo ajustado. Assim, utilizando como base Atkinson (1981), que sugeriu o uso de uma espécie de banda para a flutuação dos pontos, foram construídos os resíduos simulados para o modelo PIG e PLG. Para o modelo Poisson e PG, utilizou-se a função *hnp*() do R.

Apesar de ser esperado que haja alguns pontos fora da banda de confiança, ao se utilizar o ajuste do Modelo Poisson, mais de 95% das observações estão fora região de confiança, sendo notável, portanto, que tal modelo não se ajusta às respectivas observações. Em contraste, temos que para os demais modelos, aproximadamente 30% das observações estão fora das bandas de confiança. Nota-se que os valores residuais elevados não estão dentro do envelope para a PLG, mas para a PG e PIG a grande maioria se encontra contido.

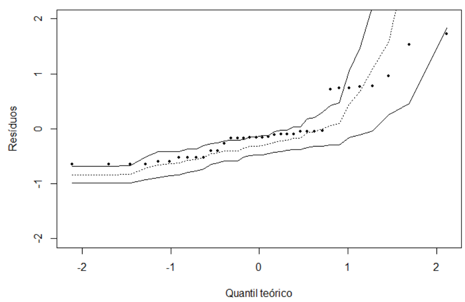
Figura 10 – Envelope simulado para os resíduos de Pearson padronizado dos modelos ajustados



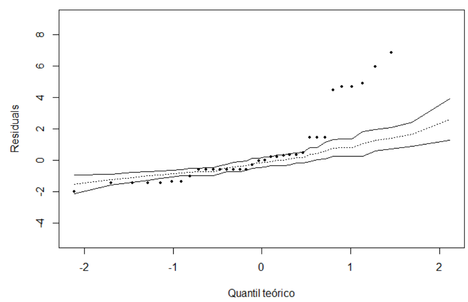
(a) Poisson



(b) Poisson-Gama



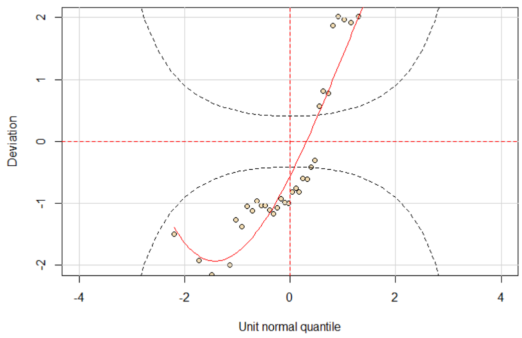
(c) Poisson-Inversa Gaussiana



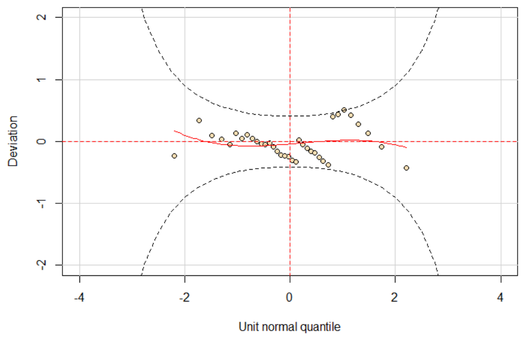
(d) Poisson-Lindley Generalizada

No Figura 11 é apresentado o *worm plot* para os resíduos. Este gráfico consiste em utilizar os resíduos quantílicos. Caso o modelo esteja bem ajustado, se espera que os resíduos estejam distribuídos sobre a reta que passa em zero. Assim, anulamos completamente a possibilidade de bom ajuste da Poisson pelos resíduos quantílicos. Apesar de não estarem tão distribuídos sobre a linha horizontal, o gráfico mostra que a PG e PIG aparentam estar bem ajustadas.

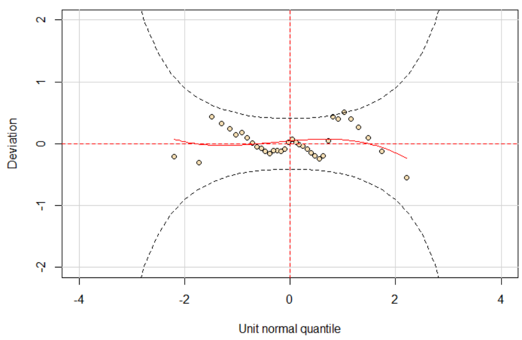
Figura 11 – Envelope simulado para os resíduos quantílicos dos modelos ajustados



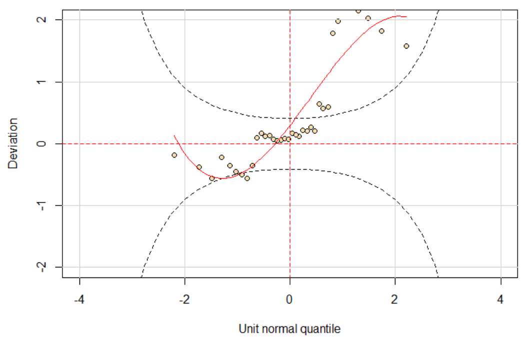
(a) Poisson



(b) Poisson-Gama



(c) Poisson-Inversa Gaussiana



(d) Poisson-Lindley Generalizada

Pelo excesso de zeros apresentado nos dados, e levando em consideração tanto a quantidade dos critérios AIC e BIC, além dos resíduos, o modelo que melhor se ajusta aos dados foi a PIG, sendo seguida pela PG. Entretanto, vimos que a PLG realmente apresenta uma melhoria em comparação ao modelo Poisson.

7.2 Aplicação II - Vítimas de Homicídio

O banco de dados foi retirado do livro *Categorical Data Analysis* de Alan Agresti (2002). O estudo consiste em uma pesquisa com 1308 indivíduos que responderam a seguinte pergunta "Com o passar dos 12 meses, quantas pessoas você conhece pessoalmente que foram vítimas de homicídio?"Os dados foram divididos entre pessoas que se identificaram tendo como raça branca ou negra.

A Tabela 9, mostra como estão distribuídos os dados em função da raça do indivíduo que participou da pesquisa.

Tabela 9 – Número de vítimas conhecidas de homicídios no ano passado, por raça

	Raça	
Resposta	Negro	Branco
0	119	1070
1	16	60
2	12	14
3	7	4
4	3	0
5	2	0
6	0	1
Média	0,522	0,092
Variância	1,150	0,155

Através da Tabela 9 é possível visualizar a questão da superdispersão dos dados, visto que em ambos os grupos, a variância dos dados é maior do que a média. A evidência disso pode ser refletida no fato da quantidade de respostas zero que o conjunto de dados possui. A Tabela 10 apresenta a frequência estimada de vítimas conhecidas para cada uma das distribuições.

Tabela 10 – Frequência estimada para as distribuições Poisson, PG, PIG e PLG

	Poisson		Poisson-Gama		PIG		PLG	
Resposta	Negros	Brancos	Negros	Brancos	Negros	Brancos	Negros	Brancos
0	94,34	1047,74	117,94	1104,50	116,13	1069,66	124,59	1070,02
1	49,25	96,66	22,00	38,12	26,10	62,50	28,45	60,04
2	12,85	4,46	9,12	5,25	8,25	11,18	5,17	13,45
3	2,24	0,14	4,47	0,90	3,58	3,38	0,69	3,76
4	0,29	0,00	2,37	0,17	1,86	1,27	0,08	1,16
5	0,03	0,00	1,30	0,03	1,07	0,53	0,01	0,38
6	0,00	0,00	0,74	0,01	0,66	0,24	0,00	0,13

Observa-se que ao estimar o número de pessoas da raça branca, a PLG chegou mais próximo, enquanto ao estimar a frequência da raça negra a PG se aproxima mais. Apesar de todas as distribuições conseguirem estimar bem as frequências de cada grupo, tem-se que a distribuição Poisson é que menos se assemelha, quando comparada com as demais.

7.2.1 Modelagem dos dados

O autor Agresti (2002) sugere ajustar um modelo de primeira ordem, visto que a variável resposta está em função apenas de uma covariável dicotômica, com a função de ligação logarítmica. Assim, como estamos trabalhando com uma mesma função de ligação para todos os modelo, a equação para o ajuste é dado como

$$\log(\mu_{ij}) = \beta_0 + \beta_1 x_{ij} \tag{7.1}$$

em que μ_{ij} corresponde ao i-ésimo sujeito da j-ésima raça, com $i = 1, \ldots, 1308$ e $j = 1, 2$.

Propõe-se inicialmente um modelo Poisson, no qual a variável resposta, o número de pessoas assassinadas que o i-ésimo entrevistado da j-ésima raça conhecia pessoalmente, denotado por Y_{ij}, segue uma distribuição de Poisson, isto é, $Y_{ij} \sim Pois(\mu_i)$. O ajuste apresentou medida *Deviance* igual a 844.71 com 1306 graus de liberdade, sendo um indicativo de presença de superdispersão, provavelmente pelo excesso de zeros que observamos nos dados.

Em seguida, ajustou-se um o modelo Poisson-Gama e Poisson-Inversa Guassiana para o conjunto de dados. O ajuste do modelo Poisson-Gama retornou uma *Deviance* igual a 412.60 com 1306 graus de liberdade, que é praticamente a metade da medida apresentado pelo ajuste da Poisson.

Outra alternativa é utilizar o modelo PLG proposto, em que a variável resposta segue uma distribuição Poisson-Lindley generalizada. Como esta distribuição não possui um função para o modelo pronta no R, fio utilizado a função *optim*() atraveés do método SANN. Como valores iniciais para as estimativas de β, utilizou-se os parâmetros das distribuições Poisson, PG e PIG. Para o parâmetro θ, avaliou-se uma grade de valores possíveis.

As Tabelas 11, 12 e 13 apresentam, respectivamente, as estimativas obtidas utilizando uma grade de valores para θ e as estimativas dos outros modelos. A linha destacada em ambas as tabelas apresenta as estimativas que levaram aos menores valores dos critérios AIC e BIC.

Tabela 11 – Estimativas dos parâmetros do modelo da PLG considerando como valores iniciais os coeficientes do modelo Poisson

$\theta_{inicial}$	$\hat{\theta}$	$\hat{\beta}_0$	$\hat{\beta}_1$	AIC	BIC	LogVerossimilhança
2,1000	2,7024	-2,3023	1,4657	1013,0158	1023,3683	504,5079
2,2000	2,6805	-2,2841	1,4319	1013,0159	1023,3684	504,5080
2,4000	2,6829	-2,2857	1,4120	1013,0205	1023,3730	504,5102
2,5000	2,6903	-2,2938	1,4312	1012,9800	1023,3325	504,4900
2,6000	2,6538	-2,2663	1,4170	1013,0399	1023,3924	504,5199
2,7000	2,6964	-2,2982	1,4400	1012,9830	1023,3355	504,4915
2,8000	**2,6775**	**-2,2871**	**1,4298**	**1012,9502**	**1023,3028**	**504,4751**
2,9000	2,6697	-2,2784	1,4179	1012,9953	1023,3478	504,4976

Tabela 12 – Estimativas dos parâmetros do modelo da PLG considerando como valores iniciais os coeficientes do modelo Poisson-Gama

$\theta_{inicial}$	$\hat{\theta}$	$\hat{\beta}_0$	$\hat{\beta}_1$	AIC	BIC	LogVerossimilhança
2,0000	2,6884	-2,2905	1,4739	1013,0880	1023,4405	504,5440
2,1000	2,6311	-2,2564	1,3993	1012,9962	1023,3487	504,4981
2,3000	2,6862	-2,2888	1,4248	1013,0028	1023,3553	504,5014
2,4000	2,6957	-2,2969	1,4230	1012,9990	1023,3515	504,4995
2,5000	2,6757	-2,2828	1,4137	1012,9907	1023,3432	504,4953
2,6000	**2,6584**	**-2,2743**	**1,4207**	**1012,9576**	**1023,3102**	**504,4788**
2,7000	2,7059	-2,2952	1,4517	1013,1058	1023,4583	504,5529
2,8000	2,6469	-2,2664	1,4021	1012,9729	1023,3254	504,4864
2,9000	2,6823	-2,2899	1,4621	1013,0053	1023,3578	504,5027

Tabela 13 – Estimativas dos parâmetros do modelo da PLG considerando como valores iniciais os coeficientes do modelo Poisson-Inversa Gaussiana

$\theta_{inicial}$	$\hat{\theta}$	$\hat{\beta}_0$	$\hat{\beta}_1$	AIC	BIC	LogVerossimilhança
2,0000	2,7154	-2,3110	1,4505	1013,0188	1023,3713	504,5094
2,1000	2,6492	-2,2661	1,4069	1012,9967	1023,3492	504,4984
2,2000	2,7148	-2,3111	1,4560	1013,0149	1023,3675	504,5075
2,3000	2,7312	-2,2928	1,4519	1013,3563	1023,7088	504,6782
2,4000	2,6615	-2,2726	1,4040	1013,0081	1023,3606	504,5041
2,6000	**2,6731**	**-2,2833**	**1,4453**	**1012,9825**	**1023,3350**	**504,4912**
2,7000	2,6635	-2,2756	1,3820	1013,0271	1023,3796	504,5135
2,8000	2,6518	-2,2688	1,3943	1012,9850	1023,3375	504,4925
2,9000	2,7019	-2,3005	1,4582	1013,0189	1023,3714	504,5095

Levando em conta os três melhores modelos PLG selecionados, escolheu-se como modelo final para a PLG o que apresentasse menor medida de qualidade de ajustamento, o AIC e BIC. Com isso, na Tabela 14 têm-se as estimativas, erros padrão associado a cada um dos parâmetros dos modelos Poisson, PG, PIG e PLG.

Tabela 14 – Estimativas e erros padrão dos ajustes para o número de vitimas conhecidas de homicídios

Parâmetros	Poisson	PG	PIG	PLG
Intercepto	$-2,3832$ *** (0,0971)	$-2,3832$ *** (0,1172)	$-2,3943$ *** (0,1221)	$-2,2871$ *** (0,0484)
β_1	$1,7331$ *** (0,1466)	$1,7331$ *** (0,2385)	$1,8287$ *** (0,2519)	$1,4298$ *** (0,1389)
θ				2,6775
σ/ϕ		0,2023	1,8662	

Como estamos trabalhando com covariáveis dicotômicas, assumimos que o β_1 está relacionado a raça negra tendo como base a raça branca, ou seja, estamos verificado se existe diferença entre eles. Como em todos os modelos, β_1 deu significativo, isso implica que existe uma diferença no número de casos que o individuo sendo negro teve um conhecido assassinado com um individuo que se denomina na raça branca.

A Tabela 15, retorna um resumo dos critérios Akaike (AIC) e Bayesiano (BIC) para cada um dos ajustes.

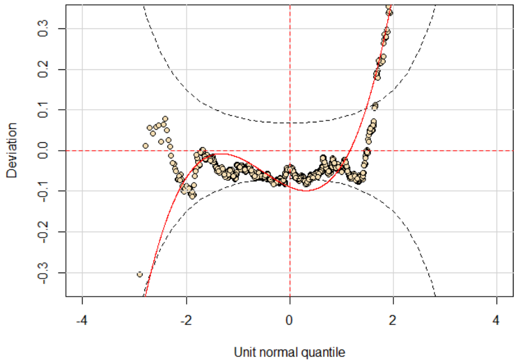
Tabela 15 – Critérios AIC e BIC para cada modelo ajustado ao conjunto das vítimas de homicidios

Critérios	Poisson	PG	PIG	PLG
AIC	1121,99	1001,80	1006,54	1012,95
BIC	1132,34	1017,32	1022,07	1023,30

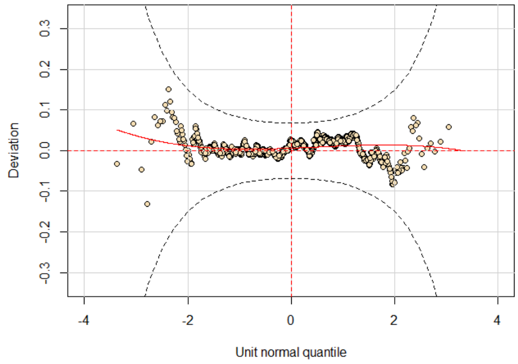
Através do gráfico *worm plot* pode-se averiguar quão bem um modelo se ajusta aos dados, além de permitir uma comparação com diferentes modelos. Essa análise é feita com os resíduos quantílicos de cada modelo. A Figura 12 ilustra o comportamento dos resíduos quantílicos para cada modelo que foi ajustado.

O esperado é que os resíduos estejam distribuídos em torno da linha horizontal. Nota-se que os resíduos do modelo Poisson é que menos está se adequando. Visto que a maioria de seus valores estão abaixo da linha horizontal, e em algum momento, seus resíduos tendem a subir verticalmente. Por outro lado, o modelo Poisson-Gama aparenta ser o mais

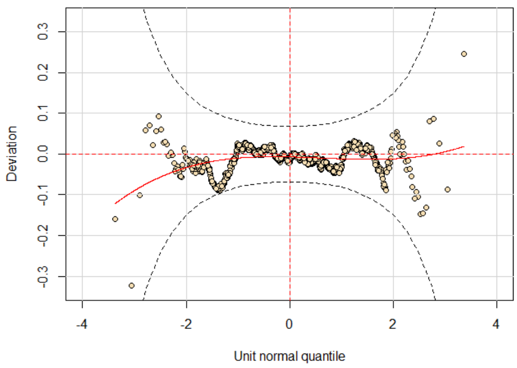
Figura 12 – Worm plot para os resíduos quantílicos dos modelos ajustados



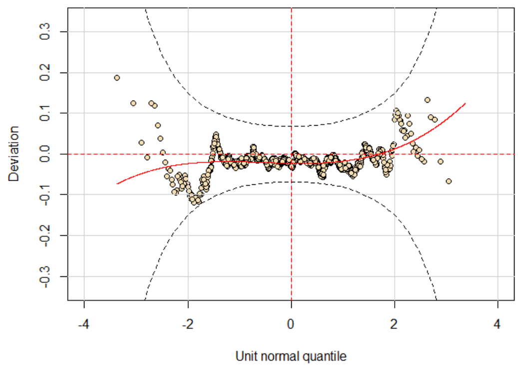
(a) Poisson



(b) Poisson-Gama



(c) Poisson-Inversa Gaussiana



(d) Poisson-Lindley Generalizada

adequado, visto que os resíduos estão mais comportados em torno da linha horizontal. Os resíduos da PLG e PIG não aparentam estar mal comportados, em comparação com os resíduos da Poisson.

Assim, considerando as medidas de qualidade e o comportamento dos resíduos quantílicos, têm-se que a PG, PIG e PLG possuem uma melhor alternativa para dados de contagem com excesso de zeros, quando comparadas com a Poisson, sendo a Poisson-Gama a que melhor está se adequando ao conjunto de dados, neste caso.

8 CONSIDERAÇÕES FINAIS

O propósito deste trabalho foi apresentar alternativas para a análise de dados de contagem com excesso de zero, por meio de distribuições compostas para os modelos em dois estágios, que tem como base a distribuição de Poisson. Comparou-se os resultados obtidos das distribuições PG, PIG e PLG, de forma a verificar se realmente há uma melhora nos ajustes quando comparada com a distribuição de Poisson. Além disso, realizou-se uma análise dos resíduos para cada distribuição, tendo como foco implementar um código no *software* R, de forma que retornasse os resíduos e os gráficos, respectivamente, para a distribuição PLG, visto que esta não possui implementação pronta.

Nas aplicações verificamos que o modelo Poisson-Gama e Poisson-Inversa Gaussiana apresentaram melhores valores nos critérios AIC e BIC, além de um bom comportamento quando analisado os resíduos. Entretanto, a PLG também apresentou um ótimo valor nos critérios de informação e comportamente quando comparada com o ajuste de Poisson.

Assim, apesar do modelo PLG não ser totalmente estável, por depender de valores iniciais para estimação de seus parâmetros, e variar de acordo com o valor inicial, ele acaba sendo um modelo simples quando comparado com algumas outras alternativas de misturas para dados com excesso de zeros, podendo servir como alternativa para a distribuição de Poisson quando trabalhado com dados com superdispersão.

Em pesquisas realizadas sobre o tema, notou-se a escassez de estudos na área computacional e nas técnicas de diagnósticos para a Poisson-Lindley Generalizada. Assim, alguns temas que podem ser desenvolvidos em trabalhos futuros seria complementar a parte de diagnóstico com a elaboração de técnicas para detecção de *outliers* e pontos influentes, com sua implementação no *software* R. Além disso, existe a simplificação da PLG para um parâmetro, chamada Poisson-Lindley, que poderia ser uma alternativa na estimação para dados de contagem com excesso de zeros. Entretanto, não se encontra na literatura nenhuma exemplo da sua forma de modelo e estimação computacional.

REFERÊNCIAS

AGRESTI, A. **Categorical Data Analysis (2nd Edition)**. [*S. l.*: *s. n.*], 2002. v. 33.

AKAIKE, H. A new look at the statistical model identification. **Automatic Control, IEEE Transactions on**, v. 19, p. 716 – 723, 01 1975.

ATKINSON, A. C. **Plots, transformations, and regression : an introduction to graphical methods of diagnostic regression analysis**. Oxford New York: Clarendon Press Oxford University Press, 1985. ISBN 0198533594.

BICKEL, P.; DOKSUM, K. **Mathematical Statistics**. [*S. l.*: *s. n.*], 1977.

BICKEL, P.; DOKSUM, K. **Mathematical Statistics: Basic Ideas and Selected Topics 2ed**. [*S. l.*]: Prentice Hall, 2001. (Mathematical Statistics: Basic Ideas and Selected Topics, v. 1).

BOZDOGAN, H. Model selection and akaike's information criterion (aic): The general theory and its analytical extensions. **Psychometrika**, v. 52, p. 345–370, 02 1987.

BRESLOW, N. Extra-poisson variation in log-linear models. **Applied Statistics**, v. 33, 01 1984.

BURNHAM, K.; ANDERSON, D. **Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach**. [*S. l.*: *s. n.*], 2002. v. 67.

BUUREN, S. Worm plot to diagnose fit in quantile regression. **Statistical Modelling**, v. 7, 12 2007.

BéLISLE, C. Convergence theorems for a class of simulated annealing algorithms on d. **Journal of Applied Probability**, v. 29, p. 885–895, 12 1992.

CORDEIRO, G. M.; DEMETRIO, C. G. B. **Modelos lineares generalizados**. Universidade estadual de Campinas. Dep. de estatística Campinas: [*S. n.*], 1986.

CORDEIRO, G. M.; DEMÉTRIO, C. G. B. **Modelos lineares generalizados e extensões**. [*S. l.*]: Departamento de Estística e Informática, Recife, PE. Departamento de Ciências Exatas, USP, 2008.

COSTA, S. C. **Modelos Lineares Generalizados Mistos para Dados Longitudinais**. Tese (Doutorado em Agronomia) – Escola Superior de Agricultura, USP, 2003.

COX, D.; SNELL, E. A general definition of residuals. **Journal of the Royal Statistical Society. Series B**, v. 30, 07 1968.

DUNN, P. K.; SMYTH, G. K. Randomized quantile residuals. **Journal of Computational and Graphical Statistics**, [American Statistical Association, Taylor Francis, Ltd., Institute of Mathematical Statistics, Interface Foundation of America], v. 5, n. 3, p. 236–244, 1996. ISSN 10618600.

ELBATAL, I.; MEROVCI, F.; ELGARHY, M. A new generalized lindley distribution. **Mathematical Theory and Modeling**, v. 3, p. 30–47, 11 2013.

EMILIANO, P. C. **Fundamentos e Aplicações dos critérios de informação: akaike e bayesiano.** Dissertação (Mestrado em Estatística e Experimentação Agropecuária) – Universidade Federal de Lavras, Minas Gerais, 2009.

FAHRMEIR, L.; TUTZ, G. Multivariate statistical modelling based on generalized linear models. 2nd ed. **Journal of the American Statistical Association**, v. 91, 06 1996.

GHITANY, M.; ALQALLAF, F.; AL-MUTAIRI, D.; HUSAIN, H. A two-parameter weighted lindley distribution and its applications to survival data. **Mathematics and Computers in Simulation**, v. 81, p. 1190–1201, 02 2011.

GREENWOOD, M.; YULE, G. U. An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. **Journal of the Royal Statistical Society**, [Wiley, Royal Statistical Society], v. 83, n. 2, p. 255–279, 1920. ISSN 09528385.

HILBE, J. M. **Modeling Count Data**. [*S. l.*]: Cambridge University Press, 2014.

HINDE, J.; DEMETRIO, C. Overdispersion: Models and estimation. **Computational Statistics Data Analysis**, v. 27, p. 151–170, 02 1998.

HOCKING, R. R. A biometrics invited paper. the analysis and selection of variables in linear regression. **Biometrics**, v. 32, n. 1, p. 1, 1976.

HOLLA, M. On a poisson-inverse gaussian distribution. **Metrika**, v. 11, p. 115–121, 02 1967.

KONISHI, S.; KITAGAWA, G. **Information criteria and statistical modeling**. [*S. l.*]: Springer, 2008.

LAMBERT, D. Zero-inflated poisson regression, with an application to defects in manufacturing. **Technometrics**, [Taylor Francis, Ltd., American Statistical Association, American Society for Quality], v. 34, n. 1, p. 1–14, 1992. ISSN 00401706.

LAWLESS, J. Negative binomial and mixed poisson regression. **Canadian Journal of Statistics**, v. 15, p. 209 – 225, 09 1987.

LINDLEY, D. Fiducial distributions and bayes' theorem. **Journal of the Royal Statistical Society. Series B**, v. 20, 01 1958.

LINDSEY, J. **Modelling frequency and count data**. Oxford: Clarendon Press, 1995. ISBN 0198523319.

MAHMOUDI, E.; ZAKERZADEH, H. Generalized poisson–lindley distribution. **Communications in Statistics—Theory and Methods**, v. 39, p. 1785–1798, 06 2010.

MALLOWS, C. L. Some Comments on CP. **Technometrics**, v. 15, n. 4, p. 661, 1973.

MARCIANO, F. W. P. **Principais tipos de resíduos utilizados na análise de diagnóstico em MLG com aplicações para os modelos: Poisson, ZIP e ZINB.** [*S. l.*], 2009.

MCCULLAGH, J. A.; NELDER, P. **Generalized linear models**. Boca Raton London New York: Chapman and Hall, 1989. ISBN 978-0412317606.

MENDES, A. M. F. **Modelo Poisson-Lindley Generalizada para dados de contagem com superdispersão**. Dissertação (Monografia (Graduação em Estatística)) – Departamento de Estatística e Matemática Aplicada, UFC, Fortaleza, 2017.

MYERS, R.; MONTGOMERY, D.; VINING, G.; ROBINSON, T. Generalized linear models: With applications in engineering and the sciences: Second edition. 01 2010.

NELDER, J. A.; WEDDERBURN, R. W. M. Generalized linear models. **Journal of the Royal Statistical Society. Series A (General)**, [Royal Statistical Society, Wiley], v. 135, n. 3, p. 370–384, 1972. ISSN 00359238.

PAULA, G. A. **Modelos de regressão: com apoio computacional**. [*S. l.*: *s. n.*], 2013.

PAULINO, C. D.; TURKMAN, M. A. A.; MURTEIRA, B.; SILVA, G. L. **Estatística Bayesiana**. 2a. ed. Lisboa: Fundação Calouste Gulbenkian, 2018.

PIERCE, D. A.; SCHAFER, D. W. Residuals in generalized linear models. **Journal of the American Statistical Association**, [American Statistical Association, Taylor Francis, Ltd.], v. 81, n. 396, p. 977–986, 1986. ISSN 01621459.

POISSON, S. Probabilité des jugements en matière criminelle et en matière civile, précédées des règles générales du calcul des probabilitiés. **Paris, France: Bachelier**, 1837.

PREGIBON, D. Logistic regression diagnostics. **The Annals of Statistics**, Institute of Mathematical Statistics, v. 9, n. 4, p. 705–724, 1981. ISSN 00905364.

PUTRI, G.; NURROHMAH, s.; FITHRIANI, I. Comparing poisson-inverse gaussian model and negative binomial model on case study: horseshoe crabs data. **Journal of Physics: Conference Series**, v. 1442, p. 012028, 01 2020.

Rstudio Team. **RStudio: Integrated Development Environment for R**. Boston, MA, 2020. Disponível em: http://www.rstudio.com/.

SANKARAN, M. 275. note: The discrete poisson-lindley distribution. **Biometrics**, v. 26, p. 145, 03 1970.

SCHWARZ, G. Estimating the dimension of a model. **The Annals of Statistics**, Institute of Mathematical Statistics, v. 6, n. 2, p. 461–464, 1978. ISSN 00905364.

WEISBERG, S. **Applied linear regression**. Hoboken, N.J: Wiley-Interscience, 2005. ISBN 0471663794.

WILLIAMS, D. Generalized linear model diagnostics using the deviance and single case deletions. **Journal of the Royal Statistical Society. Series C. Applied Statistics**, v. 36, 01 1987.

WONGRIN, W.; BODHISUWAN, W. The poisson-generalised lindley distribution and its applications. v. 38, p. 645–656, 01 2016.

WONGRIN, W.; BODHISUWAN, W. Generalized poisson–lindley linear model for count data. **Journal of Applied Statistics**, v. 44, p. 2659–2671, 11 2017.

ZAKERZADEH, H.; DOLATI, A. Generalized lindley distribution. **Journal of Mathematical Extension**, v. 3, p. 13–25, 06 2009.

APÊNDICE A – CÓDIGO UTILIZADO NAS APLICAÇÕES

```
######## Log-verossimilhan a PLG
L=function(y,x,par){
  k<-length(par)
  theta<-par[1]
  beta<-as.matrix(par[2:k])
  ee = function(aux){
    sol = TRUE
    for(i in 1:length(aux)){
      if(sol & aux[i]){
      } else{
        sol = FALSE
      }
    }
    return(sol)
  }
  if(theta>0){
    al=(exp(x%*%beta)*theta*(theta+1)-1)/(theta+1)
    if(ee(al>0) & max(al)<100){
      A=sum(log(gamma(y+al))-(log(factorial(y))+log(gamma(1+al)))+
             (al+1)*log(theta)-(y+al+2)*log(theta+1)+log((al*(theta+1)+y+al)))
      return(A)
    }else{
      return(-Inf)
    }
  }else{
    return(-Inf)
  }
}
```

```

# -logverossimilhanca
Lest<-function(par){
  return(-L(y,x,par))
}

########  Dados
abelhas = read.table(choose.files(), header = TRUE) #Banco de dados
abelhas = data.frame(abelhas)

colnames(abelhas) = c("Horas", "y")  #Titulo das colunas
attach(abelhas)
head(abelhas)

table(y)    #Frequencia
summary(y)    #Descritiva da distribuicao dos dados
var(y)     #Variabilidade

plot(table(y), type = "h", ylab = "Frequencia", xlab = "N mero de abelhas")
plot(y ~ Horas, data = abelhas, pch = 16, xlab = "Horas", ylab = "Numero de abelhas coletando polens")



####### Modelos ##########

############################### Poisson
modelo.po = glm(y ~ Horas + I(Horas^2) + I(Horas^3), family = "poisson"(link = "log"))
summary(modelo.po)
```

```
coef.po = coefficients(modelo.po)

################################ Poisson-Gama
library(MASS)

modelo.pg = glm.nb(y ~ Horas + I(Horas^2) + I(Horas^3),
   link = log)
summary(modelo.pg)
coef.pg = coefficients(modelo.pg)

################################ Poisson-Inversa Gaussiana
library(gamlss)

modelo.pig = gamlss(y ~ Horas + I(Horas^2) + I(Horas^3),
   data = abelhas, family = PIG(mu.link = "log"), trace =
   FALSE)   #Global Deviance: -2*loglik
summary(modelo.pig)

coef.pig = coefficients(modelo.pig)

################################ Poisson-Lindley
   Generalizada
x = model.matrix(modelo.pg)

# ---------- Poisson
i = 1.5
j = 1
n = length(y)
p = ncol(x)

matplg.po = matrix(0, ncol = 9, length(seq(1.5,3,by=0.1))
   -1)
```

```
colnames(matplg.po) = c("Valor inicial", "Theta Estimado", "Beta_0", "Beta_1", "Beta_2", "Beta_3", "AIC", "BIC", "LogVerossimilhanca")

while(i <= 3){
  e3 = optim(c(i,coef.po),Lest,method = "SANN")
  predito = exp(x%*%e3$par[2:5])
  AICplg = -2*L(y,x,e3$par)+2*p
  BICplg = -2*L(y,x,e3$par)+log(n)*p
  Logvero = Lest(e3$par)

  matplg.po[j,1] = i
  matplg.po[j,2:6] = unlist(e3$par)
  matplg.po[j,7] = AICplg
  matplg.po[j,8] = BICplg
  matplg.po[j,9] = Logvero

  j = j+1
  i = i+0.1
}

# ---------- Poisson-Gama
i = 1
j = 1
n = length(y)
p = ncol(x)

matplg.pg = matrix(0, ncol = 9, length(seq(1,3,by=0.1))-4)
colnames(matplg.pg) = c("Valor inicial", "Theta Estimado", "Beta_0", "Beta_1", "Beta_2", "Beta_3","AIC", "BIC", "LogVerossimilhanca")

```

```
while(i<=2.7){
  e3=optim(c(i,coef.pg),Lest,method="SANN")
  predito=exp(x%*%e3$par[2:5])
  AICplg=-2*L(y,x,e3$par)+2*p
  BICplg=-2*L(y,x,e3$par)+log(n)*p
  Logvero = Lest(e3$par)

  matplg.pg[j,1] = i
  matplg.pg[j,2:6] = unlist(e3$par)
  matplg.pg[j,7] = AICplg
  matplg.pg[j,8] = BICplg
  matplg.pg[j,9] = Logvero

  j = j+1
  i = i+0.1
}

# ---------- Poisson-Inversa Guassiana
i = 1.6
j = 1
n = length(y)
p = ncol(x)

matplg.pig = matrix(0, ncol = 9, length(seq(1.6,1.8,by
   =0.01))-1)
colnames(matplg.pig) = c("Valor inicial", "Theta Estimado",
    "Beta_0", "Beta_1", "Beta_2", "Beta_3", "AIC", "BIC", "
   LogVerossimilhanca")

while(i <= 1.8){
  e3 = optim(c(i,coef.pig),Lest,method = "SANN")
  predito = exp(x%*%e3$par[2:5])
```

```
    AICplg = -2*L(y,x,e3$par)+2*p
    BICplg = -2*L(y,x,e3$par)+log(n)*p
    Logvero = Lest(e3$par)

    matplg.pig[j,1] = i
    matplg.pig[j,2:6] = unlist(e3$par)
    matplg.pig[j,7] = AICplg
    matplg.pig[j,8] = BICplg
    matplg.pig[j,9] = Logvero

    j = j+1
    i = i+0.01
}

# ---------- Modelo Poisson-Lindley Generalizada
#Selecionando com melhor AIC/BIC
plg.po = matplg.po[which.min(apply(cbind(matplg.po[,7],
    matplg.po[,8]), 1, sum)),]
plg.pg = matplg.pg[which.min(apply(cbind(matplg.pg[,7],
    matplg.pg[,8]), 1, sum)),]
plg.pig = matplg.pig[which.min(apply(cbind(matplg.pig[,7],
    matplg.pig[,8]), 1, sum)),]

m.plg = matrix(c(plg.po,plg.pg,plg.pig), nrow = 3, byrow =
    TRUE)
colnames(m.plg) = c("Valor inicial", "Theta Estimado", "
    Beta_0", "Beta_1", "Beta_2", "Beta_3", "AIC", "BIC", "
    LogVerossimilhanca")

final.plg = m.plg[which.min(apply(cbind(m.plg[,7],m.plg
    [,8]), 1, sum)),]

```

```
coef.plg = c(unlist(final.plg[2:6])) #Estimativas
predito.plg = exp(x%*%coef.plg[2:5])   #Valores estimados

dseg = function(y,x,par){
  k = length(par)
  theta = par[1]
  beta = par[2:k]
  al = (exp(x%*%beta)*theta*(theta+1)-1)/(theta+1)
  vdc = exp(x%*%beta)
  res = sum(theta*vdc*(vdc*theta*trigamma(y+al)+digamma(y+
     al)+vdc*theta*trigamma(1+al)
                     +digamma(1+al)-log(theta/(theta+1))+((
                        theta+1)*(theta+2)*(y*(theta+1)-(
                        theta+2)))/
                       ((vdc*theta*(theta+1)-1)*(theta+2)+y
                          *(theta+1))^2))*(t(x)%*%x)
  return(res)
}

## sabemos que hat(beta) ~ N(beta,invertida)
n = 36
inform = (1/n)*dseg(y,x,coef.plg);inform # matriz de
   informacao
invertida = ginv(inform);invertida # inverso da matrz de
   informacao
epb0 = sqrt(invertida[1,1]);epb0 # erro padrao beta0
epb1 = sqrt(invertida[2,2]);epb1 # erro padrao beta1
epb2 = sqrt(invertida[3,3]);epb2 # erro padrao beta2
epb3 = sqrt(invertida[4,4]);epb3 # erro padrao beta3

t0cal = coef.plg[2]/epb0;t0cal # estatistica t calculada
t1cal = coef.plg[3]/epb1;t1cal
```

```
t2cal = coef.plg[4]/epb2;t2cal
t3cal = coef.plg[5]/epb3;t3cal
p0 = pt(abs(t0cal),n-1,lower.tail = F);p0 # valor-p associado
p1 = pt(abs(t1cal),n-1,lower.tail = F);p1
p2 = pt(abs(t2cal),n-1,lower.tail = F);p2
p3 = pt(abs(t3cal),n-1,lower.tail = F);p3

####### Gr fico
plot(y ~ Horas, abelhas, pch=16, ylim = c(0,60), xlab="Horas", ylab="Numero de abelhas coletando polen")
lines(spline(Horas, fitted.values(modelo.po), n=201), pch = 16, col = 2, lwd = 2)
lines(spline(Horas, fitted.values(modelo.pg), n=201), pch = 16, col = 3, lwd = 2)
lines(spline(Horas, fitted.values(modelo.pig), n=201), pch = 16, col = 6, lwd = 2)
lines(spline(Horas, predito.plg, n = 201), pch = 16, col = 5,lwd = 2)
legend(14.5, 50, c("Poisson","PG","PIG","PLG"), lty = c(1,1,1), lwd = c(2,2,2), col = c(2,3,6,5))

####### AIC/BIC
AIC.geral = data.frame(cbind(modelo.po$aic, modelo.pg$aic, modelo.pig$aic, final.pl[7], final.plg[7]))
colnames(AIC.geral) = c("Poisson", "PG", "PIG", "PL", "PLG")

BIC.geral = data.frame(cbind(BIC(modelo.po), BIC(modelo.pg), BIC(modelo.pig), final.pl[8], final.plg[8]))
colnames(BIC.geral) = c("Poisson", "PG", "PIG", "PL", "PLG")
```

```
AICBIC.geral = rbind(AIC.geral, BIC.geral)

####### Valores Estimados e Variancia
# ---------- Poisson
predito.po = exp(x%*%coef.po)
var.po = exp(x%*%coef.po)

# ---------- Poisson-Gama
predito.pg = exp(x%*%coef.pg)
var.pg = ((exp(x%*%coef.pg)^2)/modelo.pg$theta) + exp(x%*%coef.pg)

# ---------- Poisson-Inversa Gaussiana
predito.pig = exp(x%*%coef.pig)
var.pig = ((exp(x%*%coef.pig)^3)/modelo.pig$sigma.coefficients) + exp(x%*%coef.pig)

# ---------- Poisson Lindley Generalizada
predito.plg
alpha = (exp(x%*%coef.plg[2:5])*coef.plg[1]*(coef.plg[1]+1)-1)/(coef.plg[1]+1)
var.plg = exp(x%*%coef.plg[2:5]) + (((alpha+1)*(alpha*(coef.plg[1]+1)+2))/(coef.plg[1]^2*(coef.plg[1]+1))) - exp(x%*%coef.plg[2:5])^2


########## Residuos de pearson padronizado
library(gdata)
library(expm)
n = length(y)

```

```
# ---------- Poisson
w.po = diag(unmatrix(exp(x%*%coef.po)), n) #Matriz de pesos
#modelo.po$weights = diag(w.po)
h.po = sqrtm(w.po)%*%x%*%solve(t(x)%*%w.po%*%x)%*%t(x)%*%
   sqrtm(w.po)    #Matriz de projecao
h.po = diag(h.po)    #diagonal (leverage)

pearsonPond.po =  (y - predito.po)/sqrt(var.po*(1-h.po))
#rstandard(modelo.po, type = 'pearson')

# ---------- Poisson-Gama
w.pg = diag(unmatrix((modelo.pg$theta*exp(x%*%coef.pg))/(
   exp(x%*%coef.pg)+modelo.pg$theta)), n)
#modelo.pg$weights = diag(w.pg)
h.pg = sqrtm(w.pg)%*%x%*%solve(t(x)%*%w.pg%*%x)%*%t(x)%*%
   sqrtm(w.pg)
h.pg = diag(h.pg)

pearsonPond.pg = (y - predito.pg)/sqrt(var.pg*(1-h.pg))
#rstandard(modelo.pg, type = 'pearson')

# ---------- Poisson-Inversa Gaussiana
w.pig = diag(unmatrix((exp(x%*%coef.pig)*modelo.pig$sigma.
   coefficients)/((exp(x%*%coef.pig)^2)+modelo.pig$sigma.
   coefficients)), n)

h.pig = sqrtm(w.pig)%*%x%*%solve(t(x)%*%w.pig%*%x)%*%t(x)
   %*%sqrtm(w.pig)
h.pig = diag(h.pig)

pearsonPond.pig = (y - predito.pig)/sqrt(var.pig*(1-h.pig))

```

```
# ---------- Poisson-Lindley Generalizada
alpha
mi = predito.plg
theta = coef.plg[1]

w.plg = diag(unmatrix((mi*(theta^2)*(theta+1)*(1-mi) + (
   alpha+1)*(alpha*(theta+1)+2))/((mi^2)*(theta^2)*(theta
   +1))))
h.plg = sqrtm(w.plg)%*%x%*%solve(t(x)%*%w.plg%*%x)%*%t(x)
   %*%sqrtm(w.plg)
h.plg = diag(h.plg)

pearsonPond.plg = (y - predito.plg)/sqrt(var.plg*(1-h.plg))

######### Residuos componente da deviance padronizado
n = length(y)
x = model.matrix(modelo.pg)

# ---------- Poisson
logl.po = c()
for(i in 1:n){
  logl.po[i] = -exp(x[i,]%*%coef.po) + y[i]*(x[i,]%*%coef.
     po) - log(factorial(y[i]))       #"Normal"
  i = i+1
}

logl2.po = c()
for(i in 1:n){
  logl2.po[i] = y[i]*try(log(y[i]), T) - y[i] - log(
     factorial(y[i]))    #Saturado
  i = i+1
}
```

```

i=0
devi.po = c()
for(i in 1:n){
  if(i < 9){
    devi.po[i] = -2*(logl.po[i])
  }else{
    if(i > 34){
      devi.po[i] = -2*(logl.po[i])
    }else{
      devi.po[i] = -2*(logl.po[i]-logl2.po[i])
    }
  }
  i = i+1
}

dev.po = c(sign(y-predito.po)*sqrt(devi.po))  #residuos
   deviance
#residuals(modelo.po, "deviance")

devPadro.po = dev.po/sqrt(1-h.po)  #deviance padronizado
#rstandard(modelo.po,type='deviance')

# ---------- Poisson-Gama
k = modelo.pg$theta

logl.pg = log(gamma(k+y)/(gamma(y+1)*gamma(k))) + y*log(exp
   (x%*%coef.pg)) + k*log(k) - (y+k)*log(exp(x%*%coef.pg)+k
   )   #Verossimilhanca

logl2.pg = log(gamma(k+y)/(gamma(y+1)*gamma(k))) + y*log(y)
    + k*log(k) - (y+k)*log(y+k) #Verossimilhanca Saturada
```

```

i=0
devi.pg = c()
for(i in 1:n){
  if(i < 9){
    devi.pg[i] = -2*(logl.pg[i])
  }else{
    if(i > 34){
      devi.pg[i] = -2*(logl.pg[i])
    }else{
      devi.pg[i] = -2*(logl.pg[i]-logl2.pg[i])
    }
  }
  i = i+1
}

dev.pg = c(sign(y-predito.pg)*sqrt(devi.pg))
#residuals(modelo.pg, type = "deviance")

devPadro.pg = dev.pg/sqrt(1-h.pg)
#rstandard(modelo.pg,type='deviance')


# ---------- Poisson-Inversa Gaussiana
library(actuar)
library(Bessel)

phi = modelo.pig$sigma.coefficients

logl.pig = (phi/exp(x%*%coef.pig)) + 1/2*(log(2*phi)-log(
    exp(x%*%coef.pig))) + 1/2*(y-1/2)*(log(phi)-log(2*(1+phi
    /(2*exp(x%*%coef.pig)^2)))) + log(besselK(sqrt(2*phi*(1+
```

```
    phi/(2*exp(x%*%coef.pig)^2))),y-1/2)) - log(factorial(y)
    )

logl2.pig = (phi/y) + 1/2*(log(2*phi)-log(y)) + 1/2*(y-1/2)
    *(log(phi)-log(2*(1+phi/(2*y^2)))) + log(besselK(sqrt(2*
    phi*(1+phi/(2*y^2))),y-1/2)) - log(factorial(y))

i=0
devi.pig = c()
for(i in 1:n){
  if(i < 9){
    devi.pig[i] = -2*(logl.pig[i])
  }else{
    if(i > 34){
      devi.pig[i] = -2*(logl.pig[i])
    }else{
      devi.pig[i] = -2*(logl.pig[i]-logl2.pig[i])
    }
  }
  i = i+1
}

dev.pig = c(sign(y-predito.pig)*sqrt(abs(devi.pig)))

devPadro.pig = dev.pg/sqrt(1-h.pig)

# ---------- Poisson-Lindley Generalizada
predito.plg
theta = coef.plg[1]

al = (predito.plg*theta*(theta+1)-1)/(theta+1)
```

```
logl.plg = log(gamma(y+al)) - (log(factorial(y)) + log(
    gamma(1+al))) + (al+1)*log(theta) -
  (y+al+2)*log(theta+1) + log((al*(theta+1) + y + al))

al2 = (y*theta*(theta+1)-1)/(theta+1)
logl2.plg = log(gamma(y+al2)) - (log(factorial(y)) + log(
    gamma(1+al2))) + (al2+1)*log(theta) -
  (y+al2+2)*log(theta+1) + log((al2*(theta+1) + y + al2))

i=0
devi.plg = c()
for(i in 1:n){
  if(i < 9){
    devi.plg[i] = -2*(logl.plg[i])
  }else{
    if(i > 34){
      devi.plg[i] = -2*(logl.plg[i])
    }else{
      devi.plg[i] = -2*(logl.plg[i]-logl2.plg[i])
    }
  }
  i = i+1
}

dev.plg = c(sign(y-predito.plg)*sqrt(abs(devi.plg)))

devPadro.plg = dev.plg/sqrt(1-h.plg)

######## Envelope simulados (pearson padronizado)
library(gdata)
library(expm)

```

```
# ---------- Poisson
n = length(y)

observacoes = matrix(0, ncol = 36, nrow = 99)
residuos = matrix(0, ncol = 99, nrow = 36)

#gerando valores de y assumindo que o modelo esta correto
set.seed(123)
for(i in 1:99){
  observacoes[i,] = rpois(36, fitted(modelo.po))
  i = i + 1
}

#residuos para os valores de y gerado
for(i in 1:99){
  a = observacoes[i,]

  modelo = glm(a ~ Horas + I(Horas^2) + I(Horas^3), family
     = "poisson"(link = "log"))

  predito.po = exp(x%*%coefficients(modelo))
  var.po = exp(x%*%coefficients(modelo))

  w.po = diag(unmatrix(exp(x%*%coefficients(modelo))), n)
  h.po = sqrtm(w.po)%*%x%*%solve(t(x)%*%w.po%*%x)%*%t(x)%*%
     sqrtm(w.po)
  h.po = diag(h.po)

  pearsonPond.po =  (a - predito.po)/sqrt(var.po*(1-h.po))

  residuos[,i] = sort(pearsonPond.po)

```

```
  i = i+1
}

#selecionando o menor e o maior de cada linha
minEmax = matrix(0, ncol = 2, nrow = 36)
for(i in 1:36){
  minEmax[i,1] = min(residuos[i,])
  minEmax[i,2] = max(residuos[i,])

  i = i + 1
}

n = 36
i = 1:36
plot(qnorm((i-3/8)/(n+1/4)), minEmax[,1], type = "l", ylim
   = c(-3,7.5), ylab = "Residuos", xlab = "Quantil teorico
   ")
lines(qnorm((i-3/8)/(n+1/4)), minEmax[,2], type = "l")
lines(qnorm((i-3/8)/(n+1/4)), rowMeans(minEmax), lty = 3) #
   plot(minEmax[,2], type = "l", ylim = c(-3,4.5), add = T)

points(qnorm((i-3/8)/(n+1/4)), sort(rstandard(modelo.po,
   type ="pearson")), cex = 0.5, pch = 16)

library(hnp)
hnp(modelo.po, resid.type = "pearson", halfnormal = F, pch
   = 16, xlab = "Quantil teorico", ylab = "Residuos", print
    = T)

# --------- Poisson-Gama
hnp(modelo.pg, resid.type = "pearson", halfnormal = F, pch
   = 16, xlab = "Quantil te rico", ylab = "Residuos",
```

```
    print = T)

# ---------- Poisson-Lindley Generalizada
observacoes = matrix(0, ncol = 36, nrow = 99)
residuos = matrix(0, ncol = 99, nrow = 36)

#gerando valores de y considerando a funcao acumulada da
    PLG
set.seed(123)
for(i in 1:99){
  p = (coef.plg[1])/(1+coef.plg[1])
  alfa = (exp(x%*%coef.plg[2:5])*coef.plg[1]*(coef.plg
      [1]+1)-1)/(coef.plg[1]+1)
  observacoes[i,] = coef.plg[1]/(1+coef.plg[1])*rnbinom(36,
       alfa, p) + 1/(1+coef.plg[1])*rnbinom(36, alfa + 1, p)

  i = i + 1
}

for(k in 1:99){
  a = observacoes[k,]

  i = 1
  j = 1
  n = length(a)
  p = ncol(x)

  matplg.pg = matrix(0, ncol = 9, length(seq(1,3,by=0.1))
      -4)
  colnames(matplg.pg) = c("Valor inicial", "Theta Estimado
      ", "Beta_0", "Beta_1", "Beta_2", "Beta_3", "AIC", "BIC
      ", "LogVerossimilhanca")
```

```

while(i<=2.7){
  e3=optim(c(i,coefficients(modelo.pg)),Lest,method="SANN
     ")
  AICplg =-2*L(a,x,e3$par)+2*p
  BICplg =-2*L(a,x,e3$par)+log(n)*p
  Logvero = Lest(e3$par)

  matplg.pg[j,1] = i
  matplg.pg[j,2:6] = unlist(e3$par)
  matplg.pg[j,7] = AICplg
  matplg.pg[j,8] = BICplg
  matplg.pg[j,9] = Logvero

  j = j+1
  i = i+0.1
}

plg.pg = matplg.pg[which.min(apply(cbind(matplg.pg[,7],
   matplg.pg[,8]), 1, sum)),]
final.plg = plg.pg

coef.plg.obs = c(unlist(final.plg[2:6])) #Estimativas
predito.plg.obs = exp(x%*%coef.plg.obs[2:5])   #Valores
   estimados

alpha.obs = (exp(x%*%coef.plg.obs[2:5])*coef.plg.obs[1]*(
   coef.plg.obs[1]+1)-1)/(coef.plg.obs[1]+1)
var.plg.obs = exp(x%*%coef.plg.obs[2:5]) + (((alpha.obs
   +1)*(alpha.obs*(coef.plg.obs[1]+1)+2))/(coef.plg.obs
   [1]^2*(coef.plg.obs[1]+1))) - exp(x%*%coef.plg.obs
   [2:5])^2
```

```


  mi = predito.plg.obs
  theta = coef.plg.obs[1]

  w.plg = diag(unmatrix((mi*(theta^2)*(theta+1)*(1-mi) + (
     alpha.obs+1)*(alpha.obs*(theta+1)+2))/((mi^2)*(theta
     ^2)*(theta+1))))
  h.plg = sqrtm(w.plg)%*%x%*%solve(t(x)%*%w.plg%*%x)%*%t(x)
     %*%sqrtm(w.plg)
  h.plg = diag(h.plg)

  pearsonPond.plg.obs = (a - predito.plg.obs)/sqrt(var.plg.
     obs*(1-h.plg))

  residuos[,k] = sort(pearsonPond.plg.obs)

  k = k+1
}

minEmax = matrix(0, ncol = 2, nrow = 36)
for(i in 1:36){
  minEmax[i,1] = min(residuos[i,])
  minEmax[i,2] = max(residuos[i,])

  i = i + 1
}

n = 36
i = 1:36
plot(qnorm((i-3/8)/(n+1/4)), minEmax[,1], type = "l", ylim
   = c(-5,9), ylab = "Residuos", xlab = "Quantil teorico")
```

```
lines(qnorm((i-3/8)/(n+1/4)), minEmax[,2], type = "l")
lines(qnorm((i-3/8)/(n+1/4)), rowMeans(minEmax), lty = 3) #
   plot(minEmax[,2], type = "l", ylim = c(-3,4.5), add = T)
points(qnorm((i-3/8)/(n+1/4)), sort(pearsonPond.plg), cex =
    0.5, pch = 16)

# ---------- Poisson-Inversa Gaussiana
library(gamlss)
library(actuar)

n = length(y)

modelo.pig = gamlss(y ~ Horas + I(Horas^2) + I(Horas^3),
   data = abelhas, family = PIG(mu.link = "log"), trace =
   FALSE)   #Global Deviance: -2*loglik

x = model.matrix(modelo.pig)

observacoes = matrix(0, ncol = 36, nrow = 19)
residuos = matrix(0, ncol = 19, nrow = 36)

set.seed(51)
for(i in 1:19){
  observacoes[i,] = rpoisinvgauss(36, fitted(modelo.pig),
     modelo.pig$sigma.coefficients)
  i = i + 1
}

for(i in 1:19){
  a = observacoes[i,]

```

```
  modelo = gamlss(a ~ Horas + I(Horas^2) + I(Horas^3), data
      = abelhas, family = PIG(mu.link = "log"), trace =
     FALSE)

  predito.pig.obs = exp(x%*%coefficients(modelo))
  var.pig.obs = ((exp(x%*%coefficients(modelo))^3)/
     modelo$sigma.coefficients) + exp(x%*%coefficients(
     modelo))

  w.pig = diag(unmatrix((exp(x%*%coefficients(modelo))*
     modelo$sigma.coefficients)/((exp(x%*%coefficients(
     modelo))^2)+modelo$sigma.coefficients)), n)

  h.pig = sqrtm(w.pig)%*%x%*%solve(t(x)%*%w.pig%*%x)%*%t(x)
     %*%sqrtm(w.pig)
  h.pig = diag(h.pig)

  pearsonPond.pig.obs = (a - predito.pig.obs)/sqrt(var.pig.
     obs*(1-h.pig))

  residuos[,i] = sort(pearsonPond.pig.obs)

  i = i+1
}

minEmax = matrix(0, ncol = 2, nrow = 36)
for(i in 1:36){
  minEmax[i,1] = min(residuos[i,])
  minEmax[i,2] = max(residuos[i,])

  i = i + 1
}
```

```
n = 36
i = 1:36
plot(qnorm((i-3/8)/(n+1/4)), minEmax[,1], type = "l", ylim = c(-2,2), ylab = "Residuos", xlab = "Quantil teorico")
lines(qnorm((i-3/8)/(n+1/4)), minEmax[,2], type = "l")
lines(qnorm((i-3/8)/(n+1/4)), rowMeans(minEmax), lty = 3) # plot(minEmax[,2], type = "l", ylim = c(-3,4.5), add = T)

points(qnorm((i-3/8)/(n+1/4)), sort(pearsonPond.pig), cex = 0.5, pch = 16)

############# Gr fico - Residuo quantilico
library(statmod)

# ---------- Poisson
Acumu.po = ppois(y-1, predito.po)
Prob.po = dpois(y, predito.po)
quantilico.po = qnorm(Acumu.po + runif(36)*Prob.po)
wp(resid = quantilico.po)
#wp(resid = qresiduals(modelo.po))

# ---------- Poisson-Gama
Acumu.pg = pnbinom(y-1, size = modelo.pg$theta, prob = modelo.pg$theta/(modelo.pg$theta+predito.pg))
Prob.pg = dnbinom(y, size = modelo.pg$theta, prob = modelo.pg$theta/(modelo.pg$theta+predito.pg))
quantilico.pg = qnorm(Acumu.pg + runif(36)*Prob.pg)
wp(resid = quantilico.pg)
#wp(resid = qresiduals(modelo.pg))

# ---------- Poisson-Inversa Gaussiana
```

```
wp(gamlss(y ~ Horas + I(Horas^2) + I(Horas^3), data =
   abelhas, family = PIG(mu.link = "log"), trace = FALSE))

# ---------- Poisson-Lindley Generalizada
p = (coef.plg[1])/(1+coef.plg[1])
Acumu.plg = (coef.plg[1]/(1+coef.plg[1]))*pnbinom(y-1,
   alpha, p) + (1/(1+coef.plg[1]))*pnbinom(y-1, alpha + 1,
   p)
Prob.plg = (coef.plg[1]/(1+coef.plg[1]))*dnbinom(y, alpha,
   p) + (1/(1+coef.plg[1]))*dnbinom(y, alpha + 1, p)
quantilico.plg = qnorm(Acumu.plg + runif(36)*Prob.plg)
wp(resid = quantilico.plg)





############# Frequencia
abelhas2 = abelhas
abelhas = abelhas$y

media = mean(abelhas)
variancia = var(abelhas)

table.abelhas = as.numeric(paste(data.frame(table(abelhas))
   [,1]))
n = length(table.abelhas) #Number of abelhas
N = length(abelhas) #Total number of abelhas


# -------------------- Poisson ----------------------
lambE = mean(abelhas)
```

```
E = matrix(0, ncol = 2, nrow = n)
E[,1] = table.abelhas
E[,2] = round(dpois(table.abelhas, lambE)*N, 3)   #round: casas decimais

P = data.frame(E)
colnames(P) = c("X","Freq"); P

# ---------------------Binomial Negativa ---------------------
library(MASS)
rE = fitdistr(abelhas,"Negative Binomial", method = "Nelder-Mead")[[1]][1]
rE = array(as.numeric(unlist(rE)))
pE = N*rE/(N*rE+sum(abelhas))

E = matrix(0, ncol = 2, nrow = n)
E[,1] = table.abelhas
E[,2] = round(dnbinom(table.abelhas, rE, pE)*N,3)

BN = data.frame(E)
colnames(BN) = c("X","Freq"); BN

# --------------------- Poisson-Inversa Gaussiana ---------------------
library(actuar)
library(Bessel)

ll = function(x,par){
  x = abelhas
  phi = par[1]
```

```
  R = sum((phi/media)  + 1/2*(log(2*phi)-log(media)) +
     1/2*(x-1/2)*(log(phi)-log(2*(1+phi/(2*media^2)))) +
     log(besselK(sqrt(2*phi*(1+phi/(2*media^2))),x-1/2)) -
     log(factorial(x)))
  -R
}

mu.start = c(0.1) #c(0.01)
out2 <- nlm(ll, mu.start,x = abelhas)

phi = array(as.numeric(unlist(out2)))[2]
dpoisinvgauss(table.abelhas, media, phi, log = FALSE)*N

E = matrix(0, ncol = 2, nrow = n)
E[,1] = table.abelhas
E[,2] = round(dpoisinvgauss(table.abelhas, media, phi, log
   = FALSE)*N,3)

PIG = data.frame(E)
colnames(PIG) = c("X","Freq"); PIG

# ------- Poisson Lindley Generalizada ----------
logvero = function(x,par){
  x = abelhas
  n = N
  a = par[1]
  b = par[2]
  t = par[3]
  R = suppressWarnings(sum(log(t^(a+1)*(t+1)^b*gamma(b)*
     gamma(x+a)+t^b*(t+1)^a*gamma(a)*gamma(x+b))) - sum(log
     (factorial(x))) - (sum(x)+n)*log(t+1)
```

```
                        - n*log(gamma(a)) - n*log(gamma(b))
                           - n*(a+b)*log(t+1))
  -R
}
mu.start = c(1,1,0.1) #c(1,0.01,0.01) #c(1,0.1,0.1)
out <- nlm(logvero, mu.start, x = abelhas)

y = table.abelhas+1
R = 0
for (x in y){
  a = array(as.numeric(unlist(out)))[2]
  b = array(as.numeric(unlist(out)))[3]
  t = array(as.numeric(unlist(out)))[4]
  R[x] = (1/(factorial(x-1)*(t+1)^((x-1)+1)))*(((t/(t+1))^a
     )*((t*gamma((x-1)+a))/gamma(a))+((t/(t+1))^b)*(gamma((
     x-1)+b)/gamma(b)))*N
}

R[is.na(R)] = 0
R = R[R != 0]

E = matrix(0, ncol = 2, nrow = n)
E[,1] = table.abelhas
E[,2] = round(R,3)

PGL = data.frame(E)
colnames(PGL) = c("X","Freq"); PGL

# ------------ Geral ----------------
Obs = as.numeric(paste(data.frame(table(abelhas))[,2]))
Estimados.Geral = cbind(Abelhas=table.abelhas,Obs.= Obs,
   Poisson=P[,2],B.Negativa=BN[,2],PI.Gaussiana=PIG[,2],P.L
```

```
.Generalizada=PGL[,2]); Estimados.Geral
# ----------- Grafico ------------
abelhas
x = table.abelhas
Y = data.frame(table(abelhas))[,2]

hist(abelhas, breaks = 1000, ylab = "Frequencia", xlab = "Abelhas")
points(x, Y, pch = 16)

# -- Poisson
lines(x, P[,2], col = "red", lwd = 2)
# -- Binomial Negativa
lines(x, BN[,2], col = "darkorange", lwd = 2)
# -- Poisson Lindley Generalizada
lines(x,PGL[,2], col = "green", lwd = 2)
# -- Poisson-Inverso Gaussiano
lines(x,PIG[,2], col = "purple", lwd = 2)

legend(20, 8, legend=c("Poisson", "Binomial Negativa", "Poisson-Inversa Gaussiana", "Poisson Lindley Generalizada"),
       col=c("red", "darkorange", "purple" , "green"), lty=1, cex=0.8, lwd = 2)

# ------------ Anderson Darling Teste ------------
X = c(abelhas)
l = tail(table(X), n=1)
x_eval = 0:37

# -- Poisson
```

```
ppois_stepfun <- stepfun(x = x_eval, y = c(0, ppois(q = x_eval, lambda = lambE)))
T.P = dgof::cvm.test(x = X, y = ppois_stepfun, type = "A2")

# -- Poisson-Gama
pnbinom_stepfun <- stepfun(x = x_eval, y = c(0, pnbinom(q = x_eval, size = rE, prob = pE)))
T.Bn = dgof::cvm.test(x = X, y = pnbinom_stepfun, type = "A2")

# -- Poisson-Inversa Gaussiana
ppinvgaus_stepfun <- stepfun(x = x_eval, y = c(0, ppoisinvgauss(q = x_eval, media, phi)))
T.PIG = dgof::cvm.test(x = X, y = ppinvgaus_stepfun, type = "A2")

# -- Poisson Lindley Generalizada
f = as.vector("numeric")
for(i in 1:length(x_eval)){
  y = i-1
  f[i] = (1/(factorial(y)*((t+1)^(y+1))))*(((t/(t+1))^a)*(t*gamma(y+a)/gamma(a))+((t/(t+1))^b)*(1/gamma(b))*gamma(y+b))
}
cumsum(f)
pplindgener_stepfun <- stepfun(x = x_eval, y = c(0,cumsum(f)))
T.Pgl = dgof::cvm.test(x = X, y = pplindgener_stepfun, type = "A2")

Valor.p = cbind(P = array(as.numeric(unlist(T.P$p.value))), PG = array(as.numeric(unlist(T.Bn$p.value))), PIG =
```

```
array(as.numeric(unlist(T.PIG$p.value))),
                 PL = array(as.numeric(unlist(T.Pl$p.value))
                    ), PGL = array(as.numeric(unlist(T.Pgl$p
                    .value)))); Valor.p
```