



Universidade Federal da Paraíba
Centro de Informática
Programa de Pós-Graduação em Modelagem Matemática e Computacional

REPARAMETRIZAÇÃO DA DISTRIBUIÇÃO POISSON INFLACIONADA EM
ZERO: ESTIMAÇÃO E APLICAÇÃO

Irineu Barbosa da Silva Neto

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Modelagem Matemática e Computacional, da Universidade Federal da Paraíba, UFPB, como parte dos requisitos necessários à obtenção do título de Mestre em Modelagem Matemática e Computacional.

Orientadores: Tatiene Correia de Souza
Tarciana Liberal Pereira

João Pessoa
Agosto de 2023

Ata da Sessão Pública de Defesa de Dissertação de Mestrado de **IRINEU BARBOSA DA SILVA NETO**, candidato ao título de Mestre em Modelagem Matemática e Computacional, realizada no dia 29 de agosto de 2023.

1 Aos vinte e nove dias do mês de agosto do ano de dois mil e vinte e três, às 8h, via
2 videoconferência, reuniram-se os membros da Banca Examinadora constituída para julgar o
3 Trabalho Final do discente IRINEU BARBOSA DA SILVA NETO, vinculado a Universidade
4 Federal da Paraíba sob a matrícula nº 20211025927, candidato ao grau de Mestre em
5 “*Modelagem Matemática e Computacional*”, na linha de pesquisa “*Modelagem*
6 *Probabilística*”, do Programa de Pós-Graduação em Modelagem Matemática e
7 Computacional. A comissão examinadora foi composta pelos professores Tatiene Correia de
8 Souza, Orientadora e Presidente da Banca; Tarciana Liberal Pereira de Araujo, Orientadora;
9 Pedro Rafael Diniz Marinho, Examinador Interno ao Programa; e Hemílio Fernandes Campos
10 Coelho, Examinador Externo ao Programa. Dando início aos trabalhos, a Presidente da Banca
11 cumprimentou os presentes, comunicou a finalidade da reunião e passou a palavra ao
12 candidato para que fizesse, oralmente, a exposição do trabalho de dissertação intitulado
13 “*Reparametrização da distribuição poisson inflacionada em zero: estimação e aplicação*”.
14 Concluída a exposição, o candidato foi arguido pela Banca Examinadora, que emitiu o
15 seguinte parecer: “**aprovado**”. Do ocorrido, eu, Gean Paulo P. M. de Barros, secretário do
16 Programa de Pós-Graduação em Modelagem Matemática e Computacional (PPGMMC),
17 lavrei a presente ata, que vai assinada por mim e pelos membros da Banca Examinadora.

João Pessoa, 29 de agosto de 2023.


Gean Paulo Pereira Maurício de Barros
Secretário do PPGMMC
SIAPE 2326476


Prof^a. Dr^a. Tatiene Correia de Souza
Orientadora (PPGMMC)


Prof^a. Dr^a. Tarciana Liberal Pereira de Araujo
Orientadora (PPGMMC)


Prof. Dr. Pedro Rafael Diniz Marinho
Examinador Interno ao Programa (PPGMMC)

Prof. Dr. Hemílio Fernandes Campos Coelho
Examinador Externo ao Programa (UFPB)

Documento assinado digitalmente
 TATIENE CORREIA DE SOUZA
Data: 29/08/2023 12:06:52-0300
Verifique em <https://validar.iti.gov.br>

Documento assinado digitalmente
 TARCIANA LIBERAL PEREIRA DE ARAUJO
Data: 01/09/2023 09:14:01-0300
Verifique em <https://validar.iti.gov.br>

Documento assinado digitalmente
 PEDRO RAFAEL DINIZ MARINHO
Data: 29/08/2023 13:12:47-0300
Verifique em <https://validar.iti.gov.br>

Documento assinado digitalmente
 HEMILIO FERNANDES CAMPOS COELHO
Data: 31/08/2023 18:10:20-0300
Verifique em <https://validar.iti.gov.br>

Catálogo na publicação
Seção de Catalogação e Classificação

S586r Silva Neto, Irineu Barbosa da.

Reparametrização da distribuição Poisson inflacionada em zero : estimação e aplicação / Irineu Barbosa da Silva Neto. - João Pessoa, 2023.
76 f. : il.

Orientação: Tatiene Correia de Souza.
Coorientação: Tarciana Liberal Pereira.
Dissertação (Mestrado) - UFPB/CI.

1. Matemática computacional - Modelagem. 2. Estimação de máxima verossimilhança. 3. Método dos momentos. 4. Poisson inflacionada. 5. ZIPM. I. Souza, Tatiene Correia de. II. Pereira, Tarciana Liberal. III. Título.

UFPB/BC

CDU 519.6(043)

REPARAMETRIZAÇÃO DA DISTRIBUIÇÃO POISSON INFLACIONADA EM
ZERO: ESTIMAÇÃO E APLICAÇÃO

Irineu Barbosa da Silva Neto

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO PROGRAMA DE
PÓS-GRADUAÇÃO EM MODELAGEM MATEMÁTICA E COMPUTACIONAL
(PPGMMC) DO CENTRO DE INFORMÁTICA DA UNIVERSIDADE FEDERAL DA
PARAÍBA COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO
DO GRAU DE MESTRE EM CIÊNCIAS EM MODELAGEM MATEMÁTICA E
COMPUTACIONAL.

Examinada por:

Prof. Tatiene Correia de Souza, D.Sc.

Prof. Tarciana Liberal Pereira, D.Sc.

Prof. Hemílio Fernandes Campos Coêlho, D.Sc.

Prof. Pedro Rafael Diniz Marinho, D.Sc.

JOÃO PESSOA, PB – BRASIL
AGOSTO DE 2023

*Aos meus pais, Irinaldo Barbosa
e Regina Cely, e a minha esposa,
Gezy Kristina.*

Agradecimentos

Gostaria de agradecer, primeiramente a Deus por me fazer acreditar e permitir a conclusão desse trabalho.

As minhas orientadoras, Tatiene e Tarciana, por todo apoio durante essa jornada e pelos ensinamentos que foram de grande relevância para o conclusão deste trabalho.

A minha esposa, Gezy Kristina, pelo amor e por nunca me deixar desistir.

Aos meus filhos, João Neto e Benício, por me motivarem a cada dia.

Aos meus pais, Irinaldo e Regina, pela educação que me deram e por sempre estarem ao meu lado.

Aos meus irmãos, Fabiana, Fagner, Irís e Jéssica, pela força e por terem me apoiado durante toda esta etapa da minha vida.

Ao meu amigo José Railton que sempre me deu apoio a continuar.

Aos meus colegas, Carlos, David e Eduarda, que foram meus companheiros de estudo no PPGMMC.

Aos professores do Programa em Pós-graduação em Modelagem Matemática e Computacional da UFPB.

Ao governo do estado da Paraíba que me proporcionou apoio via licença para capacitação.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), pela concessão de bolsa de estudo.

Aos membros da banca examinadora, pelas sugestões e correções.

Resumo da Dissertação apresentada ao PPGMMC/CI/UFPB como parte dos requisitos necessários para a obtenção do grau de Mestre em Modelagem Matemática e Computacional (M.Sc.)

REPARAMETRIZAÇÃO DA DISTRIBUIÇÃO POISSON INFLACIONADA EM ZERO: ESTIMAÇÃO E APLICAÇÃO

Irineu Barbosa da Silva Neto

Agosto/2023

Orientadores: Tatiene Correia de Souza
Tarciana Liberal Pereira

Programa: Modelagem Matemática e Computacional

O presente trabalho aborda a estimação da média da distribuição Poisson inflacionada em zero (ZIP) por meio dos métodos de estimação de máxima verossimilhança (EMV) e de momentos. Foi proposta uma nova parametrização para a Poisson inflacionada, denominada de ZIPM, que introduz os parâmetros de média (μ) e de inflação (δ). O objetivo principal foi investigar a eficácia da estimação para média da distribuição Poisson inflacionada em zero e a adequação da nova parametrização proposta.

Os resultados demonstraram que tanto o EMV quanto o método dos momentos produziram estimativas satisfatórias para a média da distribuição ZIPM. Não foram encontradas diferenças significativas entre os métodos em termos da média estimada, erro quadrático médio (EQM) e viés relativo. Além disso, a nova parametrização ZIPM mostrou-se promissora, permitindo um controle refinado da média e da inflação.

No entanto, é importante ressaltar que certos erros de especificação podem ser cometidos ao utilizar a ZIPM e os métodos de estimação. A adequação do modelo e a validade das estimativas dependem do atendimento dos pressupostos estatísticos subjacentes e da aderência dos dados à distribuição Poisson.

Em conclusão, este trabalho contribui para o campo da estatística ao propor a parametrização ZIPM e ao explorar os métodos de estimação para a média da distribuição ZIP. Os resultados encorajadores indicam a utilidade da nova parametrização e fornecem informações valiosas para pesquisadores que lidam com dados inflacionados em zero. No entanto, é necessário cautela ao interpretar os resultados, considerando as possíveis limitações do modelo e os erros de especificação associados.

Palavras-chave: Estimação de Máxima Verossimilhança, Momentos, Poisson inflacionada, ZIPM, Reparametrização.

Abstract of Dissertation presented to PPGMMC/CI/UFPB as a partial fulfillment of the requirements for the degree of Master of Mathematical and Computational Modeling (M.Sc.)

REPARAMETRIZAÇÃO DA DISTRIBUIÇÃO POISSON INFLACIONADA EM ZERO: ESTIMAÇÃO E APLICAÇÃO

Irineu Barbosa da Silva Neto

August/2023

Advisors: Tatiene Correia de Souza
Tarciana Liberal Pereira

Program: Computational Mathematical Modelling

The present work addresses the estimation of the mean of the Zero-Inflated Poisson (ZIP) distribution using the methods of Maximum Likelihood Estimation (MLE) and Moments. A new parameterization for the Zero-Inflated Poisson, named ZIPM, was proposed, introducing the mean (μ) and inflation (δ) parameters. The main objective was to investigate the estimation effectiveness for the mean of the Zero-Inflated Poisson distribution and the suitability of the proposed new parameterization.

The results demonstrated that both MLE and the method of moments produced satisfactory estimates for the mean of the ZIPM distribution. No significant differences were found between the methods in terms of estimated mean, mean squared error (MSE), and relative bias. Furthermore, the new ZIPM parameterization showed promise, allowing for refined control of the mean and inflation.

However, it is important to note that certain specification errors can occur when using ZIPM and estimation methods. Model adequacy and the validity of estimates depend on meeting underlying statistical assumptions and the data adhering to the Poisson distribution.

In conclusion, this work contributes to the field of statistics by proposing the ZIPM parameterization and exploring estimation methods for the mean of the ZIP distribution. The encouraging results indicate the utility of the new parameterization and provide valuable insights for researchers dealing with zero-inflated data. Nevertheless, caution is necessary when interpreting results, considering potential model limitations and associated specification errors.

Keywords: Maximum Likelihood Estimation, Moments, Zero-Inflated Poisson, ZIPM, Reparameterization.

Sumário

1	Introdução	1
2	Distribuição Poisson	3
2.1	Função de probabilidade da distribuição Poisson	3
2.2	Estimação via máxima verossimilhança	4
2.3	Estimação via método dos momentos	5
2.4	Estimação via máxima verossimilhança para o parâmetro da distribuição Poisson	6
3	Distribuição Poisson inflacionada em zero	7
3.1	Teste de superdispersão	8
3.2	Uma nova parametrização para distribuição ZIP	9
3.3	Estimação dos parâmetros da ZIPM via máxima verossimilhança	14
3.4	Intervalos de confiança ZIPM	18
3.5	Estimação dos parâmetros da ZIPM via método dos momentos	18
4	Avaliação numérica	21
4.1	Estimação pontual - máxima verossimilhança	21
4.1.1	Avaliação sem cometer erro de especificação	21
4.1.2	Avaliação com erro de especificação: gerando ZIPM e estimando Poisson	25
4.2	Estimação Intervalar - máxima verossimilhança	28
4.2.1	Avaliação sem cometer erro de especificação	28
4.2.2	Avaliação com erro de especificação	32
4.3	Estimação pontual - Método dos momentos	36
4.3.1	Estimação pontual - Método dos momentos	36
5	Aplicações	39
5.1	Dados de movimento fetal	39
5.2	Beneficiários do fundo de pensão	41
6	Conclusões	43

Referências Bibliográficas	44
Apêndice A - Códigos	46

Capítulo 1

Introdução

A modelagem estatística desempenha um papel fundamental na análise e interpretação de dados em diversas áreas do conhecimento. Uma das distribuições mais utilizadas nesse contexto é a distribuição de Poisson, que descreve a ocorrência de eventos em um intervalo de tempo ou espaço fixo. Larsen e Marx (2005) utilizaram como exemplo a distribuição Poisson para prever o número de emissões de uma fonte radioativa, bem como o número de guerras iniciadas em um determinado ano. A distribuição de Poisson é adequada para modelar dados equidispersos, uma vez que a variância de uma variável que segue essa distribuição é igual a média. Dessa forma, na presença de uma grande quantidade de zeros a distribuição pode não ser mais adequada uma vez que a contagem excessiva de zeros pode causar a superdispersão dos dados. Uma solução é considerar um parâmetro extra para lidar com a contagens excessivas de zeros subestimadas pela distribuição Poisson padrão. Tal distribuição é chamada de distribuição Poisson Inflacionada em Zero (ZIP - Zero Inflated Poisson).

A distribuição Poisson inflacionada em zero foi apresentada por Lambert (1992) em uma aplicação sobre o número de defeitos em um processo industrial. Ele observou que quando um equipamento de fabricação está devidamente alinhado, defeitos praticamente não acontecem, porém quando está desalinhado, os defeitos ocorrem de acordo com uma distribuição de Poisson. Depois disso, vários estudos têm demonstrado a importância da modelagem estatística de dados inflacionados através da distribuição Poisson. Por exemplo, Broek (1995) utilizou a distribuição Poisson inflacionada em zeros para analisar números de homens infectados pelo HIV. Yang et al. (2017) utilizaram a distribuição ZIP em um estudo sobre inquiridos de saúde.

Um ponto que merece atenção em relação a distribuição ZIP é o fato da mesma não ser expressa em termos da média da distribuição inflacionada, uma vez que o parâmetro λ , que indexa a distribuição, representa a média da distribuição Poisson. Nesta dissertação, buscando auxiliar no estudo de variáveis dessa natureza, propomos uma nova parametrização para a distribuição de Poisson inflacionada em zero, em termos da média da distribuição inflacionada, semelhante ao que foi feito por Bayes e Valdivieso (2016)

para distribuição beta inflacionada. Essa nova parametrização da distribuição ZIP é mais atrativa para uso em modelos de regressão e gráficos de controle uma vez que um dos parâmetros que indexa a distribuição é a média da variável considerando o excesso de zeros.

A estimação dos parâmetros da distribuição proposta foi realizada através dos métodos da máxima verossimilhança e dos momentos. Por meio de simulação de Monte Carlo avaliamos numericamente o desempenho do estimador proposto para a média, de forma pontual e intervalar considerando cenários com e sem erros de especificação. Os resultados foram analisados em termos da média estimada, erro quadrático médio (EQM), viés relativo (VR) e taxas de convergência.

Como a distribuição Poisson permite valores iguais a zero no seu domínio, uma etapa essencial antes de utilizar a distribuição ZIP é verificar se os dados indicam a presença de um inflacionamento em zero. Dessa forma, é prudente analisar até que ponto o modelo que segue distribuição Poisson padrão suporta o excesso de zeros. Nesse contexto, Broek (1995) apresentou um teste para avaliarmos se uma amostra com determinadas quantidade de valores nulos pode ser ajustada por uma distribuição inflacionada. Foram realizados testes de qualidade dos dados, como o teste de inflação de Broek (1995) e o teste de convergência .

Este trabalho está organizado como segue. No Capítulo 2, apresentamos uma revisão teórica dos conceitos fundamentais sobre a distribuição Poisson e os estimadores de máxima verossimilhança. No capítulo 3, apresentamos a distribuição de Poisson inflacionada em Zero . Abordamos o teste de superdispersão, utilizado para verificar a presença de inflação nos dados. Em seguida, propomos uma nova parametrização, denominada ZIPM (Zero Inflated Poisson with mean μ) e realizamos a estimação para média da distribuição ZIPM através do métodos da máxima verossimilhança e também pelo método dos momentos. No Capítulo 4, realizamos uma avaliação numérica dos estimadores propostos para a média da distribuição ZIPM. Apresentamos resultados de estimação pontual e intervalar e exploramos situações de erro de especificação, gerando dados ZIP e estimando-os por meio da distribuição de Poisson. No Capítulo 5, aplicamos a nova parametrização a dois conjuntos de dados reais para exemplificar o uso da distribuição proposta. Por fim, no Capítulo 6, concluímos o trabalho, resumindo os principais resultados, destacando as contribuições e limitações do estudo e sugerindo possíveis direções para pesquisas futuras.

Capítulo 2

Distribuição Poisson

A distribuição Poisson foi introduzida por Siméon Denis Poisson em 1837 (POISSON, 1837). De acordo com Hilbe (2014), quando lidamos com uma variável aleatória Y que segue distribuição Poisson, podemos observar as seguintes características:

1. É uma distribuição discreta com um único parâmetro, a média λ , que corresponde ao número esperado de vezes que um evento ocorre por unidade de tempo, área ou volume.
2. Os valores esperados de uma variável aleatória que segue distribuição Poisson são inteiros não negativos.
3. A probabilidade de valores iguais a zero é inversamente proporcional a média λ .
4. A média e a variância da variável aleatória Y que segue distribuição Poisson são iguais ao parâmetro λ , ou seja, $E(Y) = \text{Var}(Y) = \lambda$.

2.1 Função de probabilidade da distribuição Poisson

Considere a variável aleatória Y que segue uma distribuição de Poisson com parâmetro λ ($\lambda > 0$). A função de probabilidade é dada por :

$$P(Y) = \frac{e^{-\lambda} \lambda^y}{y!}, y = 0, 1, 2, 3, \dots, \quad (2.1)$$

em que λ é o número médio de ocorrências no intervalo de tempo ou espaço e escrevemos $Y \sim \mathcal{P}(\lambda)$. O valor esperado e a variância da variável aleatória Y são iguais a λ , ou seja, $E(Y) = \text{Var}(Y) = \lambda$.

A distribuição Poisson é bastante utilizada para descrever situações que envolvem, por exemplo, o número de chamadas que chegam a uma central telefônica, o número de partículas α emitidas por uma fonte radioativa ou o número de pessoas que chegam a

determinada fila, sempre em um intervalo de tempo fixado (BOLFARINE; SANDOVAL, 2001).

2.2 Estimação via máxima verossimilhança

Nesta seção apresentamos o método da máxima verossimilhança e os fundamentos que lhes são característicos. Esse método estima os parâmetros de uma distribuição maximizando a probabilidade de uma amostra aleatória ser observada. O procedimento para obter EMV envolve a construção da função de verossimilhança, que é uma função dos parâmetros desconhecidos. Essa função é obtida multiplicando as funções de probabilidade correspondentes a cada observação independente dos dados (LARSEN; MARX, 2005). A seguir temos a definição para a função de verossimilhança.

Seja Y_1, Y_2, \dots, Y_n uma amostra aleatória de uma população com função de probabilidade $f(y; \theta)$, onde θ é um parâmetro desconhecido. A função de verossimilhança, $L(\theta)$, é definida por

$$L(\theta) = \prod_{i=1}^n f(y_i; \theta)$$

Note que a função de verossimilhança de uma amostra aleatória Y_1, Y_2, \dots, Y_n é a densidade conjunta das variáveis aleatórias Y_1, Y_2, \dots, Y_n .

O método da máxima verossimilhança seleciona o valor de $\hat{\theta}$ mais provável de ter produzido as observações Y_1, Y_2, \dots, Y_n , isto é, os estimadores de máxima verossimilhança são os estimadores que atribuem à amostra selecionada a maior probabilidade de ser observada. A maximização da função de verossimilhança é realizada através de técnicas matemáticas. O valor de θ que maximiza a função de verossimilhança é chamado de estimador de máxima verossimilhança (EMV) e é denotado por $\hat{\theta}$.

Para determinar o estimador de máxima verossimilhança procedemos da seguinte maneira

- Obtemos uma amostra aleatória Y_1, Y_2, \dots, Y_n da população com função de probabilidade $f(y; \theta)$;
- Encontramos a função de verossimilhança $L(\theta)$ da amostra aleatória Y_1, Y_2, \dots, Y_n ;
- Determinamos a expressão para θ que maximiza $L(\theta)$. Na maioria dos casos é mais conveniente utilizarmos a função $\ln L(\theta)$ para obter o EMV;
- Substituímos θ por $\hat{\theta}$ na expressão para o estimador de máxima verossimilhança de θ e encontramos o estimador $\hat{\theta}$ para θ .

Nem sempre o estimador de máxima verossimilhança é uma função explícita da amostra aleatória, nesse caso precisamos recorrer à otimizações numéricas para encontramos o estimador.

O estimador de máxima verossimilhança é um método poderoso para estimar parâmetros desconhecidos. Ao maximizar a verossimilhança dos dados observados, ele fornece estimativas consistentes, eficientes e assintoticamente normais. O EMV desempenha um papel fundamental na inferência estatística e é amplamente utilizado em diversas áreas de pesquisa

2.3 Estimação via método dos momentos

O método dos momentos pode ser utilizado para estimar os parâmetros de uma distribuição de probabilidade e consiste em igualar os momentos populacionais aos seus respectivos momentos amostrais.

Os momentos populacionais são características importantes que descrevem a distribuição teórica dos dados. Eles fornecem informações essenciais sobre o comportamento estatístico da população. Os momentos amostrais são as estimativas calculadas a partir dos dados observados em uma amostra. Ao igualar os momentos populacionais aos momentos amostrais, estamos procurando encontrar os valores dos parâmetros que tornam a distribuição teórica consistente com os dados observados. De acordo Sahoo (2015) o método dos momentos é um dos métodos clássicos para estimar parâmetros e a motivação vem do fato de que os momentos amostrais são em certo sentido, estimativas para os momentos populacionais.

Nesta seção apresentamos o método dos momentos e o passo a passo dessa técnica de estimação de parâmetros. Começamos com a definição dos momentos amostrais e populacionais, respectivamente, e em seguida a descrição do método.

Seja Y_1, Y_2, \dots, Y_n uma amostra aleatória de tamanho n . O r -ésimo momento amostral e o r -ésimo momento populacional, $r \geq 1$, são dados respectivamente, por

$$m_r = \frac{1}{n} \sum_{i=1}^n Y_i^r$$

e

$$\mu_r = E[Y^r].$$

O método dos momentos permite encontrarmos os estimadores para $\theta = (\theta_1, \theta_2, \dots, \theta_k)$ resolvendo-se as equações

$$m_r = \mu_r,$$

$r = 1, 2, \dots, k$.

O método dos momentos apresenta uma abordagem útil para a estimação dos parâmetros de uma distribuição e na maioria das vezes oferece estimadores consistentes. Entretanto, é importante ressaltar que, esse método está sujeito a certas limitações. Em

situações em que a amostra é pequena os momentos amostrais têm estimativas imprecisas, a estimação dos parâmetros pode ser afetada e gerar resultados menos confiáveis.

2.4 Estimação via máxima verossimilhança para o parâmetro da distribuição Poisson

Seja Y_1, Y_2, \dots, Y_n uma amostra aleatória obtida a partir de uma população que segue distribuição Poisson com parâmetro λ . Considerando o fato que temos variáveis aleatórias independentes e identicamente distribuídas, então a função de verossimilhança $L(\lambda)$ é dada por:

$$L(\lambda) = \prod_{i=1}^n \frac{e^{-\lambda} \lambda^{y_i}}{y_i!} = \frac{e^{-n\lambda} \lambda^{\sum_{i=1}^n y_i}}{\prod_{i=1}^n y_i!}, \quad i = 1, \dots, n.$$

Assim, a função de log-verossimilhança $l(\lambda)$ é expressa por:

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

A primeira derivada de $l(\lambda)$ em relação a λ é

$$\frac{\partial l}{\partial \lambda} = -n + \frac{1}{\lambda} \sum_{i=1}^n y_i. \quad (2.2)$$

Como estamos interessados em obter um estimador que maximize a log verossimilhança, devemos verificar se a segunda derivada de $l(\lambda)$ é negativa, de fato

$$\frac{\partial l}{\partial \lambda^2} = -\frac{\sum_{i=1}^n y_i}{\lambda^2} < 0.$$

Como resultado, igualando a equação 2.2 a zero e isolando o parâmetro λ obtemos o EMV $\hat{\lambda}$,

$$\begin{aligned} -n + \frac{1}{\hat{\lambda}} \sum_{i=1}^n y_i &= 0 \\ \Rightarrow \hat{\lambda} &= \frac{\sum_{i=1}^n y_i}{n} = \bar{y} \end{aligned}$$

Portanto, \bar{y} , a média amostral, é o estimador de máxima verossimilhança para λ .

Capítulo 3

Distribuição Poisson inflacionada em zero

Uma variável aleatória Y que segue distribuição Poisson, $Y \sim \mathcal{P}(\lambda)$, se caracteriza pela equidispersão, ou seja, $E(Y) = \text{Var}(Y)$ (KUMAR; RAMACHANDRAN, 2020). Entretanto, ao modelar dados reais, a equidispersão nem sempre é satisfeita. Sendo assim, geralmente é necessário utilizar uma distribuição que contemple qualquer sub ou superdispersão que esteja nos dados (HILBE, 2014).

Conjuntos de dados onde a ocorrência de zeros é maior do que a esperada pela distribuição de probabilidade são ditos inflacionados em zero. A distribuição de Poisson inflacionada em zero (ZIP) fornece uma abordagem adequada para ajustar esse inflacionamento (LAMBERT, 1992). As distribuições inflacionadas geralmente fornecem um bom ajuste aplicada a conjuntos de dados positivamente distorcidos quando comparadas as distribuições padrões (TLHALOGANYANG; THAGA, 2020). Dessa forma, essas distribuições vêm sendo bastante utilizadas na literatura, pois vários conjuntos de dados são caracterizados pela superdispersão causada por esse excesso de zeros. Por exemplo, o número de pedidos de ambulância para doenças relacionadas ao calor (BASSIL et al., 2011), o número de células somáticas no leite (SERGIO, 2012) ou ainda o número de aberrações cromossômicas em função da dose de radiação (OLIVEIRA et al., 2016).

A distribuição de Poisson inflacionada em zero (ZIP) tem um parâmetro adicional, chamado parâmetro de inflação (δ) que lida com o excesso de zeros subestimado pela distribuição Poisson padrão. Segundo Lambert (1992), uma variável aleatória segue distribuição de Poisson inflacionada em zero se sua distribuição de probabilidade é dada por:

$$P(Y) = \begin{cases} \delta + (1 - \delta) \cdot e^{-\lambda}, & y = 0 \\ (1 - \delta) \cdot \frac{e^{-\lambda} \lambda^y}{y!}, & y = 1, 2, 3, \dots \end{cases}, \quad 0 \leq \delta < 1 \quad \text{e} \quad \lambda > 0. \quad (3.1)$$

Usamos a notação $Y \sim \text{ZIP}(\delta, \lambda)$ para indicar que a variável aleatória Y tem distribuição de Poisson inflacionada em zero com parâmetros δ e λ . O valor esperado e a

variância da variável aleatória Y que segue distribuição Poisson inflacionada em zero são dados por:

$$E(Y) = (1 - \delta)\lambda. \quad (3.2)$$

e

$$\text{Var}(Y) = \lambda(1 - \delta)(1 + \delta\lambda). \quad (3.3)$$

De fato,

$$E(Y) = \sum_{y=0}^{\infty} y \cdot P(Y = y) = 0 \cdot P(Y = 0) + \sum_{y=1}^{\infty} y \left[(1 - \delta) \cdot \frac{e^{-\lambda} \cdot \lambda^y}{y!} \right]$$

$$E(Y) = (1 - \delta) \cdot \sum_{y=1}^{\infty} y \frac{e^{-\lambda} \cdot \lambda^y}{y!}$$

$$\Rightarrow E(Y) = (1 - \delta)\lambda.$$

Além disso, temos que

$$E(Y^2) = \sum_{y=0}^{\infty} y^2 \cdot P(Y = y) = 0 \cdot P(Y = 0) + \sum_{y=1}^{\infty} y^2 \left[(1 - \delta) \cdot \frac{e^{-\lambda} \cdot \lambda^y}{y!} \right]$$

$$E(Y^2) = (1 - \delta) \cdot \sum_{y=1}^{\infty} y^2 \frac{e^{-\lambda} \cdot \lambda^y}{y!}$$

$$E(Y^2) = (1 - \delta)(\lambda + \lambda^2).$$

Logo,

$$\text{Var}(Y) = E(Y^2) - [E(Y)]^2 = (1 - \delta)(\lambda + \lambda^2) - (1 - \delta)^2\lambda^2$$

$$\text{Var}(Y) = \lambda \cdot (1 - \delta) \cdot (1 + \delta\lambda).$$

De acordo 3.2 e 3.3, $\text{Var}(Y) > E(Y)$ para todo λ . Essa superdispersão é gerada pelo excesso de zeros na distribuição ZIP. Tem-se ainda que, quanto maior for a proporção de zeros maior será a variância da variável aleatória e a medida que o valor de δ se aproxima de zero a distribuição ZIP se aproxima da distribuição Poisson padrão.

3.1 Teste de superdispersão

As distribuições discretas inflacionadas em zero são usadas para modelar dados de contagem que possuem mais zeros do que o esperado pela distribuição que estamos usando na modelagem (RIDOUT et al., 1998). Sendo assim, uma vez que a distribuição Poisson permite valores iguais a zero no seu domínio, e assumindo que os dados possam apresentar excesso de zeros, é pertinente verificar até que ponto o modelo com distribuição Poisson padrão pode ser utilizada para ajustar os dados.

Broek (1995) desenvolveu um teste para averiguar se o número de zeros presentes nos dados é compatível com o esperado pela distribuição de Poisson. O teste consiste em analisar se em um conjunto de dados de contagem o parâmetro de inflação em zero, δ , é igual a zero. Em caso afirmativo temos que a superdispersão não está presente, e a variável aleatória Y segue distribuição Poisson (TLHALOGANYANG; THAGA, 2020).

Diante o exposto, usamos as hipóteses $H_0 : \delta = 0$ versus $H_1 : \delta > 0$ para verificar a inflação de zero, onde a rejeição de H_0 indica a presença de inflação nos dados. A estatística de teste desenvolvida por Van den Broek, e apresentada a seguir, segue distribuição qui-quadrado com 1 grau de liberdade e é expressa por:

$$S_1 = \frac{(n_0 - np_0)^2}{np_0(1 - p_0) - n\bar{y}p_0^2}$$

em que n : número total de observações, n_0 : número de zeros observados, \bar{y} : média das observações, $p_0 = e^{-\hat{\lambda}}$, onde $\hat{\lambda}$ é o estimador do parâmetro da distribuição de Poisson.

De acordo com Broek (1995), uma média razoavelmente alta e poucos zeros causam problemas com a distribuição aproximadamente χ^2 , a menos que o tamanho da amostra seja grande. No entanto, na prática, isso não é um problema, pois com uma média grande, a presença de zeros é evidente (JANSKUL; HINDE, 2002).

3.2 Uma nova parametrização para distribuição ZIP

Nesta seção apresentamos uma nova parametrização para distribuição Poisson inflacionada em zero com média μ (ZIPM) e obtivemos o estimador de máxima verossimilhança e intervalos de confiança para média (μ) de uma variável aleatória Y que segue distribuição ZIPM.

A função de probabilidade da distribuição Poisson inflacionada em zero, tem como um dos seus parâmetros a média da distribuição Poisson (λ) ao invés da média da distribuição ZIP apresentada em (3.1). Em termos de interpretação, esse fato pode ser apontado como um ponto fraco dessa distribuição uma vez que λ é a média da distribuição Poisson que engloba uma parte dos valores iguais a zero. Adicionalmente, para modelos de regressão e cartas de controle é mais atrativo que um dos parâmetros seja a média da distribuição inflacionada que contempla todos os valores de Y e não uma média condicional. Propomos nessa seção uma reparametrização, segundo a apresentada por Bayes e Valdivieso (2016), na qual estão presentes dois parâmetros, sendo um dos parâmetros a média da distribuição inflacionada (e não a média condicional ao número de zeros permitido pela distribuição de Poisson) e o outro chamado de parâmetro de inflação, δ . Dessa forma, de (3.2), tem-se que

$$E(Y) = (1 - \delta)\lambda = \mu.$$

$$\Rightarrow \lambda = \frac{\mu}{(1 - \delta)}.$$

Portanto, sob a nova parametrização a função de probabilidade da variável aleatória Y que segue distribuição Poisson inflacionada em zero, $Y \sim \text{ZIPM}(\delta, \mu)$, é dada por :

$$P(Y = y) = \begin{cases} \delta + (1 - \delta) \cdot e^{-\frac{\mu}{(1-\delta)}}, & y = 0 \\ (1 - \delta) \cdot \frac{e^{-\frac{\mu}{(1-\delta)}} \cdot (\frac{\mu}{1-\delta})^y}{y!}, & y = 1, 2, 3, \dots \end{cases}$$

em que $\mu > 0$ é a média da distribuição Poisson inflacionada.

Note que o parâmetro δ , nomeado aqui de parâmetro de inflação, não é a proporção de zeros na amostra, uma vez que $P(Y = 0) = \delta + (1 - \delta)e^{-\frac{\mu}{1-\delta}}$.

Além disso, tem-se que

$$E(Y) = \mu$$

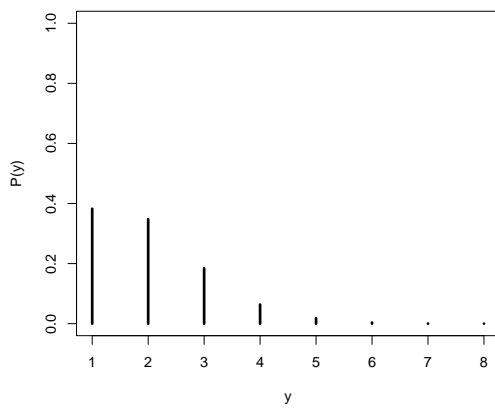
e

$$\text{Var}(Y) = \mu \left(\frac{\delta\mu - \delta + 1}{1 - \delta} \right).$$

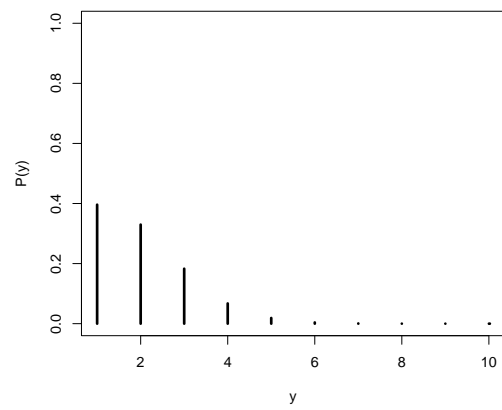
A seguir apresentamos algumas representações gráficas da função de probabilidade da distribuição ZIPM para os valores dos parâmetros usados na simulação. A Figura 3.1 representa a função de probabilidade com média $\mu = 1$ e parâmetro de inflação $\delta = 0.05, 0.10, 0.15, 0.20$. A Figura 3.2 representa a função de probabilidade com média $\mu = 1.4$ e parâmetro de inflação $\delta = 0.05, 0.10, 0.15, 0.20$. A Figura 3.3 representa a função de probabilidade com média $\mu = 1.8$ e parâmetro de inflação $\delta = 0.04, 0.08, 0.12, 0.16$. Os gráficos ilustram o comportamento do inflacionamento na distribuição ZIPM. O parâmetro de inflacionamento (δ) controla o pico da curva em cada cenário. Notamos que o pico da curva diminui à medida que o δ aumenta e que os gráficos apresentam comportamento assimétricos.

De modo geral, os gráficos nos ajudam a identificar o padrão de inflacionamento presente na distribuição ZIPM. Se houver uma maior concentração nas barras com probabilidade zero, isso pode indicar um maior inflacionamento na distribuição, com um número relativamente maior de zeros extras. Por exemplo, considere a Figura (3.3). Observamos que as barras com probabilidade zero aumentam progressivamente à medida que incrementamos o valor do parâmetro δ de 0.04 para 0.08, 0.12 e 0.16.

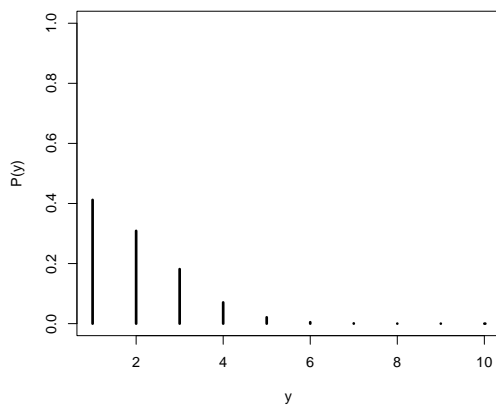
Figura 3.1: Função de probabilidade da distribuição ZIPM, com parâmetro de inflação: $\delta = 0.05, 0.10, 0.15, 0.20$ e $\mu = 1$



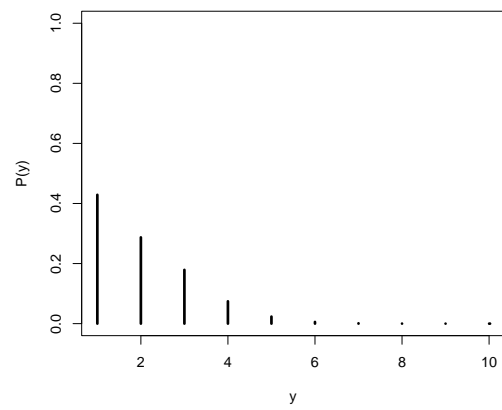
(a) $\delta = 0.05$ e $\mu = 1$



(b) $\delta = 0.10$ e $\mu = 1$

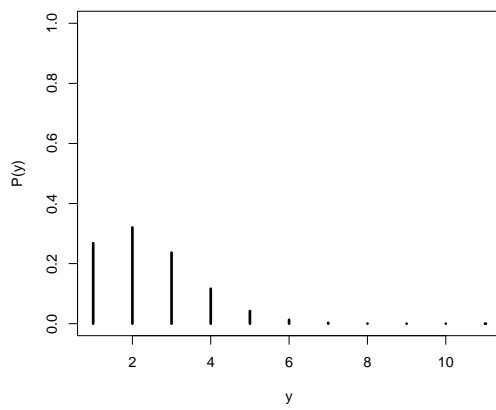


(c) $\delta = 0.15$ e $\mu = 1$

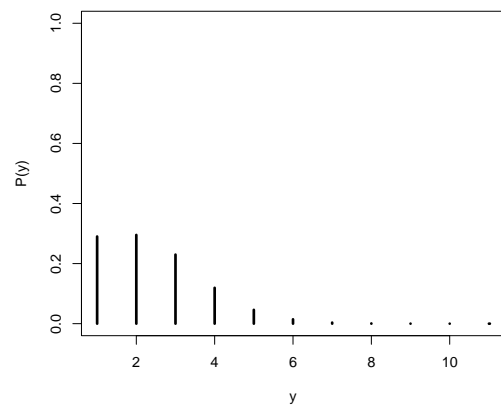


(d) $\delta = 0.20$ e $\mu = 1$

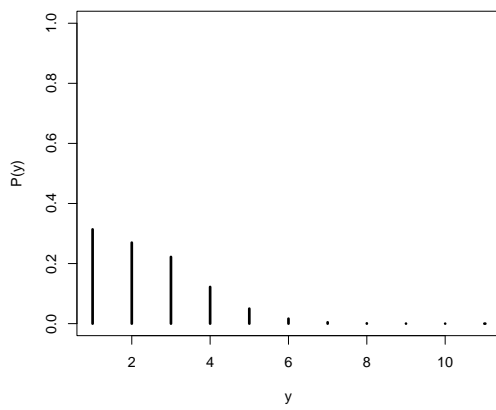
Figura 3.2: Função de probabilidade da distribuição ZIPM, com parâmetro de inflação: $\delta = 0.05, 0.10, 0.15, 0.20$ e $\mu = 1.4$



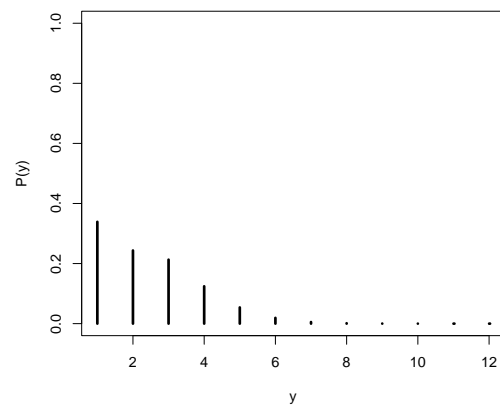
(a) $\delta = 0.05$ e $\mu = 1.4$



(b) $\delta = 0.10$ e $\mu = 1.4$

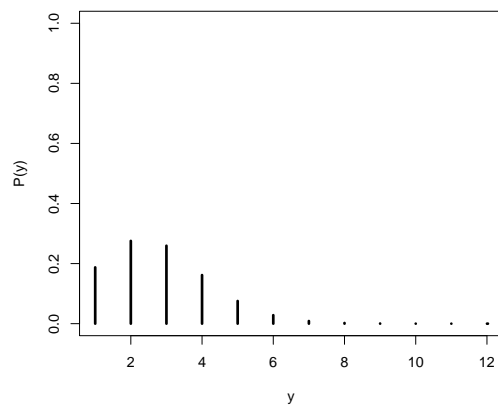


(c) $\delta = 0.15$ e $\mu = 1.4$

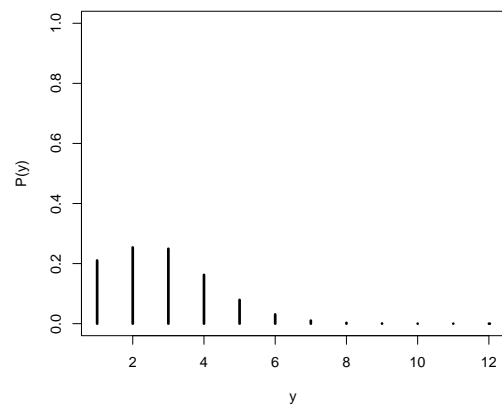


(d) $\delta = 0.20$ e $\mu = 1.4$

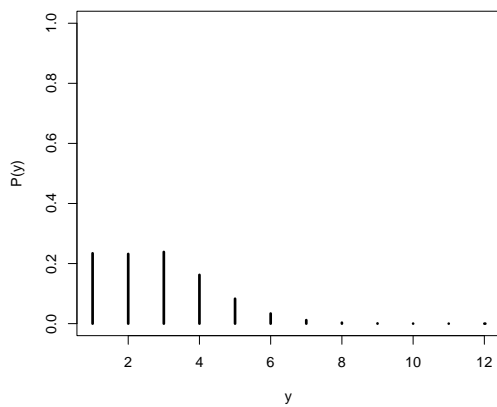
Figura 3.3: Função de probabilidade da distribuição ZIPM, com parâmetro de inflação: $\delta = 0.04, 0.08, 0.12, 0.16$ e $\mu = 1.8$



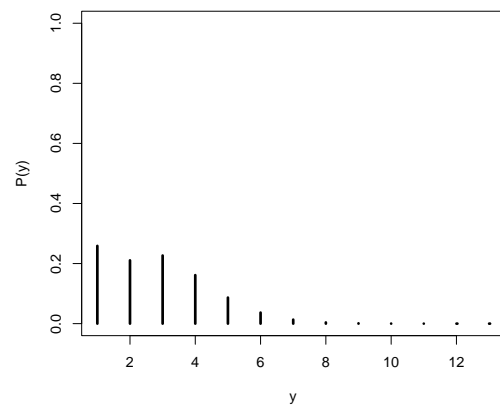
(a) $\delta = 0.04$ e $\mu = 1.8$



(b) $\delta = 0.08$ e $\mu = 1.8$



(c) $\delta = 0.12$ e $\mu = 1.8$



(d) $\delta = 0.16$ e $\mu = 1.8$

3.3 Estimação dos parâmetros da ZIPM via máxima verossimilhança

Seja Y_1, Y_2, \dots, Y_n uma amostra aleatória da variável aleatória Y que segue distribuição ZIPM(δ, μ), com n_0 observações iguais a zero e $n - n_0$ observações diferentes de zero. Dessa forma, a função de verossimilhança para $\theta = (\delta, \mu)$ dada a amostra aleatória será

$$L(\theta) = \prod_{i=1}^{n_0} (\delta + (1 - \delta) \cdot e^{-\frac{\mu}{1-\delta}}) \prod_{i=n_0+1}^n \left[(1 - \delta) \cdot \frac{e^{-\frac{\mu}{1-\delta}} \cdot \left(\frac{\mu}{1-\delta}\right)^{y_i}}{y_i!} \right].$$

Logo, a função de log-verossimilhança, $l(\theta)$, é expressa por

$$l(\theta) = \sum_{i=1}^{n_0} \ln[\delta + (1 - \delta) \cdot e^{-\frac{\mu}{1-\delta}}] + \sum_{i=n_0+1}^n \ln \left[(1 - \delta) \cdot \frac{e^{-\frac{\mu}{1-\delta}} \cdot \left(\frac{\mu}{1-\delta}\right)^{y_i}}{y_i!} \right]$$

ou ainda,

$$l(\theta) = \sum_{i=1}^{n_0} \ln[\delta + (1 - \delta) \cdot e^{-\frac{\mu}{1-\delta}}] + \sum_{i=n_0+1}^n \left[\ln(1 - \delta) - \frac{\mu}{(1 - \delta)} + y_i [\ln \mu - \ln(1 - \delta)] - \ln y_i! \right].$$

A função escore (também conhecida como gradiente) é um vetor de derivadas parciais da função de verossimilhança em relação aos parâmetros do modelo. Para o caso da distribuição Poisson inflacionada em zero com média μ , a função escore é dada por

$$U(\theta) = \left(\frac{\partial l(\theta)}{\partial \delta}, \frac{\partial l(\theta)}{\partial \mu} \right),$$

em que,

$$\begin{aligned} \frac{\partial l(\theta)}{\partial \delta} &= n_0 \cdot \frac{1 - \frac{\mu \cdot e^{-\frac{\mu}{1-\delta}}}{1-\delta} - e^{-\frac{\mu}{1-\delta}}}{\delta + (1 - \delta) \cdot e^{-\frac{\mu}{1-\delta}}} \\ &\quad + \sum_{i=n_0+1}^n \left[\frac{-1}{(1 - \delta)} - \frac{\mu}{(1 - \delta)^2} + \frac{y_i}{1 - \delta} \right] \\ \Rightarrow \frac{\partial l(\theta)}{\partial \delta} &= n_0 \cdot \frac{1 - \frac{\mu \cdot e^{-\frac{\mu}{1-\delta}}}{1-\delta} - e^{-\frac{\mu}{1-\delta}}}{\delta + (1 - \delta) \cdot e^{-\frac{\mu}{1-\delta}}} + \sum_{i=n_0+1}^n \frac{y_i - \delta \cdot y_i - \mu + \delta - 1}{(1 - \delta)^2} \end{aligned}$$

e

$$\begin{aligned}\frac{\partial l(\theta)}{\partial \mu} &= -n_0 \cdot \frac{e^{-\frac{\mu}{1-\delta}}}{\delta + (1-\delta) \cdot e^{-\frac{\mu}{1-\delta}}} + \sum_{i=n_0+1}^n -\frac{1}{1-\delta} + y_i \cdot \left(\frac{1}{\mu}\right) \\ \Rightarrow \frac{\partial l(\theta)}{\partial \mu} &= -n_0 \cdot \frac{e^{-\frac{\mu}{1-\delta}}}{\delta + (1-\delta) \cdot e^{-\frac{\mu}{1-\delta}}} + \sum_{i=n_0+1}^n \frac{y_i - \delta y_i - \mu}{\mu \cdot (1-\delta)}.\end{aligned}$$

Os estimadores de máxima verossimilhança são determinados numericamente maximizando a função de log-verossimilhança através de um método de otimização não linear, pois não podem ser obtidos de forma fechada. A matriz de informação de Fisher é obtida a partir do negativo dos valores esperados das derivadas de segunda ordem da função de log-verossimilhança. Tem-se que

$$\begin{aligned}\frac{\partial^2 l(\theta)}{\partial \delta^2} &= \frac{n_0 \left[\frac{\mu(\delta+\mu-1)e^{-\frac{\mu}{1-\delta}}}{(1-\delta)^3} + \frac{\mu e^{-\frac{\mu}{1-\delta}}}{(1-\delta)^2} \right] \left[\delta + (1-\delta)e^{-\frac{\mu}{1-\delta}} \right]}{\left[\delta + (1-\delta) \cdot e^{-\frac{\mu}{1-\delta}} \right]^2} \\ &\quad - \frac{n_0 \left[1 - \frac{\mu e^{-\frac{\mu}{1-\delta}}}{(1-\delta)} - e^{-\frac{\mu}{1-\delta}} \right] \left[1 - \frac{\mu e^{-\frac{\mu}{1-\delta}}}{(1-\delta)} - e^{-\frac{\mu}{1-\delta}} \right]}{\left[\delta + (1-\delta) \cdot e^{-\frac{\mu}{1-\delta}} \right]^2} \\ &\quad + \sum_{i=n_0+1}^n \left[\frac{y_i - \delta \cdot y_i + \delta - 2\mu - 1}{(1-\delta)^3} \right] \\ \Rightarrow \frac{\partial^2 l(\theta)}{\partial \delta^2} &= n_0 \cdot \frac{\mu^2 \cdot e^{-\frac{\mu}{1-\delta}}}{(1-\delta)^3 \cdot \left[(1-\delta) \cdot e^{-\frac{\mu}{1-\delta}} + \delta \right]} - n_0 \cdot \left[\frac{1 - \frac{\mu \cdot e^{-\frac{\mu}{1-\delta}}}{1-\delta} - e^{-\frac{\mu}{1-\delta}}}{\delta + (1-\delta) \cdot e^{-\frac{\mu}{1-\delta}}} \right]^2 \\ &\quad + \sum_{i=n_0+1}^n \left[\frac{y_i - \delta \cdot y_i + \delta - 2\mu - 1}{(1-\delta)^3} \right] \\ \frac{\partial^2 l(\theta)}{\partial \mu^2} &= \frac{-n_0 \left[\frac{-e^{-\frac{\mu}{1-\delta}}}{1-\delta} \cdot \left(\delta + (1-\delta)e^{-\frac{\mu}{1-\delta}} \right) - e^{-\frac{\mu}{1-\delta}} \cdot \left(\frac{-e^{-\frac{\mu}{1-\delta}}}{1-\delta} + \frac{\delta \cdot e^{-\frac{\mu}{1-\delta}}}{1-\delta} \right) \right]}{\left(\delta + (1-\delta)e^{-\frac{\mu}{1-\delta}} \right)^2} \\ &\quad - \sum_{i=n_0+1}^n \frac{y_i}{\mu^2} \\ \Rightarrow \frac{\partial^2 l(\theta)}{\partial \mu^2} &= n_0 \cdot \frac{\delta \cdot e^{-\frac{\mu}{1-\delta}}}{(1-\delta) \cdot \left[\delta + (1-\delta) \cdot e^{-\frac{\mu}{1-\delta}} \right]^2} - \sum_{i=n_0+1}^n \frac{y_i}{\mu^2}\end{aligned}$$

$$\frac{\partial^2 l(\theta)}{\partial \delta \partial \mu} = \frac{\partial^2 l(\theta)}{\partial \mu \partial \delta} = n_0 \cdot \frac{e^{-\frac{\mu}{1-\delta}} \cdot \left(-\delta^2 e^{-\frac{\mu}{1-\delta}} + 2\delta \cdot e^{-\frac{\mu}{1-\delta}} - e^{-\frac{\mu}{1-\delta}} + \delta^2 + \mu \cdot \delta - 2\delta + 1 \right)}{(1-\delta)^2 \cdot \left[\delta + (1-\delta)e^{-\frac{\mu}{1-\delta}} \right]^2} - \sum_{i=n_0+1}^n \frac{1}{(1-\delta)^2}$$

Para obter a matriz da informação de Fisher, multiplicamos cada derivada de segunda ordem por -1 e depois calculamos as suas respectivas esperanças. Dessa forma, a matriz da informação de Fisher para distribuição Poisson inflacionada (ZIPM) é:

$$M(\theta) = \begin{pmatrix} M_{\delta\delta} & M_{\delta\mu} \\ M_{\mu\delta} & M_{\mu\mu} \end{pmatrix},$$

em que,

$$\begin{aligned} M_{\delta\delta} &= E \left\{ -\frac{\partial^2 l(\theta)}{\partial \delta^2} \right\} = \\ &= E \left\{ -n_0 \cdot \frac{\mu^2 \cdot e^{-\frac{\mu}{1-\delta}}}{(1-\delta)^3 \cdot [(1-\delta) \cdot e^{-\frac{\mu}{1-\delta}} + \delta]} + n_0 \left[\frac{1 - \frac{\mu \cdot e^{-\frac{\mu}{1-\delta}}}{1-\delta} - e^{-\frac{\mu}{1-\delta}}}{\delta + (1-\delta) \cdot e^{-\frac{\mu}{1-\delta}}} \right]^2 \right. \\ &\quad \left. - \sum_{i=n_0+1}^n \left[\frac{y_i - \delta \cdot y_i + \delta - 2\mu - 1}{(1-\delta)^3} \right] \right\} \end{aligned}$$

$$\begin{aligned} \Rightarrow M_{\delta\delta} &= E \left\{ -\frac{\partial^2 l(\theta)}{\partial \delta^2} \right\} = -n_0 \cdot E \left\{ \frac{\mu^2 \cdot e^{-\frac{\mu}{1-\delta}}}{(1-\delta)^3 \cdot [(1-\delta) \cdot e^{-\frac{\mu}{1-\delta}} + \delta]} \right\} \\ &\quad + n_0 \cdot E \left\{ \left[\frac{1 - \frac{\mu \cdot e^{-\frac{\mu}{1-\delta}}}{1-\delta} - e^{-\frac{\mu}{1-\delta}}}{\delta + (1-\delta) \cdot e^{-\frac{\mu}{1-\delta}}} \right]^2 \right\} \\ &\quad - E \left\{ \sum_{i=n_0+1}^n \left[\frac{y_i - \delta \cdot y_i + \delta - 2\mu - 1}{(1-\delta)^3} \right] \right\} \end{aligned}$$

$$\begin{aligned} \Rightarrow M_{\delta\delta} &= E \left\{ -\frac{\partial^2 l(\theta)}{\partial \delta^2} \right\} = -n_0 \cdot \frac{\mu^2 \cdot e^{-\frac{\mu}{1-\delta}}}{(1-\delta)^3 \cdot [(1-\delta)e^{-\frac{\mu}{1-\delta}} + \delta]} \\ &\quad + n_0 \cdot \left[\frac{1 - \frac{\mu \cdot e^{-\frac{\mu}{1-\delta}}}{1-\delta} - e^{-\frac{\mu}{1-\delta}}}{\delta + (1-\delta) \cdot e^{-\frac{\mu}{1-\delta}}} \right]^2 - (n - n_0) \cdot \frac{\delta - \mu - \delta \cdot \mu - 1}{(1-\delta)^3} \end{aligned}$$

$$M_{\mu\mu} = E \left\{ -\frac{\partial^2 l(\theta)}{\partial \mu^2} \right\} = E \left\{ -n_0 \cdot \frac{\delta \cdot e^{-\frac{\mu}{1-\delta}}}{(1-\delta) \cdot \left[\delta + (1-\delta) \cdot e^{-\frac{\mu}{1-\delta}} \right]^2} + \sum_{i=n_0+1}^n \frac{y_i}{\mu^2} \right\}$$

$$\Rightarrow M_{\mu\mu} = E \left\{ -\frac{\partial^2 l(\theta)}{\partial \mu^2} \right\} = -n_0 \cdot E \left\{ \frac{\delta \cdot e^{-\frac{\mu}{1-\delta}}}{(1-\delta) \cdot \left[\delta + (1-\delta) \cdot e^{-\frac{\mu}{1-\delta}} \right]^2} \right\} \\ + E \left\{ \sum_{i=n_0+1}^n \frac{y_i}{\mu^2} \right\}$$

$$\Rightarrow M_{\mu\mu} = E \left\{ -\frac{\partial^2 l(\theta)}{\partial \mu^2} \right\} = -n_0 \cdot \frac{\delta \cdot e^{-\frac{\mu}{1-\delta}}}{(1-\delta) \cdot \left[\delta + (1-\delta) e^{-\frac{\mu}{1-\delta}} \right]^2} + (n - n_0) \cdot \frac{1}{\mu}$$

$$M_{\delta\mu} = E \left\{ -\frac{\partial^2 l(\theta)}{\partial \delta \partial \mu} \right\} \\ = E \left\{ -n_0 \cdot \frac{e^{-\frac{\mu}{1-\delta}} \cdot \left(-\delta^2 e^{-\frac{\mu}{1-\delta}} + 2\delta \cdot e^{-\frac{\mu}{1-\delta}} - e^{-\frac{\mu}{1-\delta}} + \delta^2 + \mu \cdot \delta - 2\delta + 1 \right)}{(1-\delta)^2 \cdot \left[\delta + (1-\delta) e^{-\frac{\mu}{1-\delta}} \right]^2} \right. \\ \left. + \sum_{i=n_0+1}^n \frac{1}{(1-\delta)^2} \right\}$$

$$M_{\delta\mu} = E \left\{ -\frac{\partial^2 l(\theta)}{\partial \delta \partial \mu} \right\} \\ = E \left\{ -n_0 \cdot \frac{e^{-\frac{\mu}{1-\delta}} \cdot \left(-\delta^2 e^{-\frac{\mu}{1-\delta}} + 2\delta \cdot e^{-\frac{\mu}{1-\delta}} - e^{-\frac{\mu}{1-\delta}} + \delta^2 + \mu \cdot \delta - 2\delta + 1 \right)}{(1-\delta)^2 \cdot \left[\delta + (1-\delta) e^{-\frac{\mu}{1-\delta}} \right]^2} \right\} \\ + E \left\{ \sum_{i=n_0+1}^n \frac{1}{(1-\delta)^2} \right\}$$

$$M_{\delta\mu} = E \left\{ -\frac{\partial^2 l(\theta)}{\partial \delta \partial \mu} \right\} \\ = -n_0 \cdot \frac{e^{-\frac{\mu}{1-\delta}} \cdot \left(-\delta^2 \cdot e^{-\frac{\mu}{1-\delta}} + 2\delta \cdot e^{-\frac{\mu}{1-\delta}} - e^{-\frac{\mu}{1-\delta}} + \delta^2 + \mu\delta - 2\delta + 1 \right)}{(1-\delta)^2 \cdot \left[\delta + (1-\delta) e^{-\frac{\mu}{1-\delta}} \right]^2} \\ + \frac{n - n_0}{(1-\delta)^2}.$$

Além disso, como

$$\frac{\partial^2 l(\theta)}{\partial \delta \partial \mu} = \frac{\partial^2 l(\theta)}{\partial \mu \partial \delta},$$

temos que

$$M_{\mu\delta} = E \left\{ -\frac{\partial^2 l(\theta)}{\partial \mu \partial \delta} \right\} = E \left\{ -\frac{\partial^2 l(\theta)}{\partial \delta \partial \mu} \right\} = M_{\delta\mu}.$$

3.4 Intervalos de confiança ZIPM

De acordo com Larsen e Marx (2005), as estimativas pontuais não fornecem nenhuma indicação de sua precisão inerente. Por sua vez, o intervalo de confiança (IC) é uma estimativa estatística que nos permite analisar a probabilidade do verdadeiro valor do parâmetro pertencer a um intervalo. Do ponto de vista informativo o IC é mais completo do que gerar somente uma estimacão pontual. Essa estimativa tem como base uma amostra dos dados disponveis.

Uma medida comum de incerteza associada a um intervalo de confiana é o nvel de confiana que indica a probabilidade de que o intervalo de confiana contenha o verdadeiro valor do parâmetro desconhecido (PINHEIRO et al., 2012). A taxa de cobertura refere-se à porcentagem de intervalos de confiana que capturam o parâmetro populacional verdadeiro, enquanto a taxa de no cobertura é a porcentagem de intervalos que no o capturam.

Adicionalmente, de acordo com Pinheiro et al. (2012) quanto maior for a amplitude do intervalo de confiana, maior será a incerteza associada à estimativa. Um intervalo de confiana mais amplo indica uma estimativa menos precisa, pois abrange uma variedade maior de valores possveis para o parâmetro. Por outro lado, um intervalo de confiana mais estreito indica maior preciso, pois restringe a faixa de valores possveis.

A obteno da matriz de informao de Fisher nos permite obter intervalos de confiana e realizar testes de hipteses sobre os parâmetros do modelo. Sendo $\hat{\theta} = (\hat{\delta}, \hat{\mu})$ o estimador de máxima verossimilhana de $\theta = (\delta, \mu)$, temos pelo Teorema Central do Limite (TCL), que a distribuio de $\hat{\theta}$ converge para uma distribuio assintoticamente Normal, ou seja, $\hat{\theta} \stackrel{a}{\sim} N(\theta, M(\theta)^{-1})$ em que $M(\theta)$ é a matriz da informao de Fisher. Dessa forma, os intervalos de confiana assintticos para os parâmetros da distribuio ZIPM so dados por $\hat{\delta} \pm z_{1-\frac{\alpha}{2}} EP(\hat{\delta})$ e $\hat{\mu} \pm z_{1-\frac{\alpha}{2}} EP(\hat{\mu})$, ao nvel de significncia $(1 - \alpha)$ e EP é o erro padro do estimador.

3.5 Estimaco dos parâmetros da ZIPM via método dos momentos

No contexto da distribuio ZIPM, os momentos populacionais esto relacionados à taxa média de ocorrncia dos eventos (μ) e ao parâmetro de inflaco (δ).

Seja Y_1, Y_2, \dots, Y_n uma amostra aleatória de uma variável aleatória que segue distribuição ZIPM(δ, μ), com n_0 observações iguais a zero e $n - n_0$ observações diferentes de zero.

Para aplicar o método dos momentos calculamos os momentos amostrais correspondentes e em seguida igualamos os momentos populacionais aos seus respectivos momentos amostrais. O primeiro momento amostral é dado pela média amostral, ou seja, a soma de todos os valores observados dividida pelo tamanho da amostra. Temos

$$m_1 = \frac{1}{n} \sum_{i=1}^n Y_i = \bar{Y}.$$

Além disso, o segundo momento amostral é dado por

$$m_2 = \frac{1}{n} \sum_{i=1}^n Y_i^2.$$

Por sua vez, o primeiro e segundo momentos populacionais são dados, respectivamente, por

$$\mu_1 = E(Y) = \mu$$

e

$$\mu_2 = E(Y^2) \tag{3.4}$$

Como $\text{Var}(Y) = E(Y)^2 - [E(Y)]^2$, temos de 3.4 que

$$\mu_2 = E(Y^2) = \text{Var}(Y) + [E(Y)]^2.$$

Entretanto, $\text{Var}(Y) = \mu \left(\frac{\delta\mu - \delta + 1}{1 - \delta} \right)$ e $E(Y) = \mu$, logo

$$\mu_2 = \mu \left(\frac{\delta\mu - \delta + 1}{1 - \delta} \right) + \mu^2.$$

Igualando o primeiro momento amostral ao primeiro momento populacional obtemos

$$\hat{\mu} = \bar{Y}.$$

Além disso, de $m_2 = \mu_2$ temos a seguinte equação

$$\frac{1}{n} \sum_{i=1}^n Y_i^2 = \hat{\mu} \left(\frac{\hat{\delta}\hat{\mu} - \hat{\delta} + 1}{1 - \hat{\delta}} \right) + \hat{\mu}^2.$$

ou ainda

$$\frac{1}{n} \sum_{i=1}^n Y_i^2 = \hat{\mu} \left[\left(\frac{\hat{\delta}\hat{\mu} - \hat{\delta} + 1}{1 - \hat{\delta}} \right) + \hat{\mu} \right].$$

logo

$$\frac{1}{n} \sum_{i=1}^n Y_i^2 = \hat{\mu} \left(\frac{\hat{\mu} - \hat{\delta} + 1}{1 - \hat{\delta}} \right). \quad (3.5)$$

Como $\hat{\mu} = \bar{Y}$, temos de (3.5) que

$$\frac{1}{n} \sum_{i=1}^n Y_i^2 = \bar{Y} \left(\frac{\bar{Y} - \hat{\delta} + 1}{1 - \hat{\delta}} \right) \quad (3.6)$$

ou ainda

$$\frac{\sum_{i=1}^n Y_i^2}{\sum_{i=1}^n Y_i} = \frac{\bar{Y} - \hat{\delta} + 1}{1 - \hat{\delta}}$$

Portanto, isolando $\hat{\delta}$ na equação (3.6) temos

$$\hat{\delta} = \frac{\bar{Y}}{1 - \frac{\sum_{i=1}^n Y_i^2}{\sum_{i=1}^n Y_i}} + 1$$

Capítulo 4

Avaliação numérica

Uma vez que o interesse da nova parametrização proposta é a média da distribuição inflacionada, abordamos nesse capítulo a avaliação numérica do estimador de máxima verossimilhança para média da distribuição Poisson inflacionada em zero considerando a nova parametrização.

Os resultados que são apresentados neste capítulo são oriundos de simulações de Monte Carlo com 10.000 réplicas e amostras de tamanhos 50, 100, 150 e 200. Todas as simulações foram realizadas com o *software* estatístico R (R Core Team, 2022). A maximização da função de log-verossimilhança foi realizada usando o método quasi-Newton BFGS com primeiras derivadas. Para os estudos da simulação, as amostras da distribuição ZIPM foram geradas usando a função *rzip* que está incluída no pacote *extraDistr* (Frank E. Harrell Jr., 2021), em que o valor de λ foi calculado a partir da média (μ) e do parâmetro de inflação (δ).

Após gerar as réplicas da distribuição ZIPM, aplicamos o teste de inflacionamento de Broek (1995), para verificar se as réplicas apresentavam a característica de excesso de zeros esperada na distribuição ZIPM, assegurando que o parâmetro estimado fosse coerente com a natureza da distribuição ZIPM, evitando estimativas inválidas.

4.1 Estimação pontual - máxima verossimilhança

4.1.1 Avaliação sem cometer erro de especificação

Para a avaliação do estimador de máxima verossimilhança Consideramos como critérios de avaliação a média, o erro quadrático médio (EQM) e o viés relativo (VR) do estimador de máxima verossimilhança de μ .

Seguindo a sugestão dada Broek (1995), Jansakul e Hinde (2002) ressaltam que, quando a contagem média é alta e há poucos zeros, a aproximação χ^2 pode ser imprecisa se o tamanho da amostra, (n), não for grande o suficiente. Como medida preventiva, Jansakul e Hinde (2002) inicialmente utilizam o parâmetro λ com valores na faixa de 0.5

a 2.0, a fim de garantir a precisão. Os valores para o média, μ , e para o parâmetro de inflação, δ , foram escolhidos de acordo Jansakul e Hinde (2002).

Na Tabela 4.1 encontram-se os resultados das medidas utilizadas na avaliação dos os estimadores pontuais da distribuição ZIPM com média $\mu = 1$ e $\delta = (5\%, 10\%, 15\%, 20\%)$. Notamos que as médias dos estimadores se aproximam dos verdadeiros valores à medida que o tamanho da amostra aumenta. Tem-se também que o valor do EQM se reduz com o aumento do tamanho da amostra, considerando todos os valores de δ . Além disso, observamos que as estimativas do viés relativos dos estimadores, em módulo, foram consideravelmente pequenas e diminuíram à medida que o tamanho da amostra aumenta. Para exemplificar considere $\delta = 0.15$. Quando $n = 50$, observamos que a média, EQM e VR foram 1.0183, 0.0271 e 1.8250, enquanto para $n = 200$, temos os valores 1.0019, 0.0070 e 0.1933, respectivamente.

Na Tabela 4.2 encontram-se os resultados para o estimador pontual da média com $\mu = 1.4$ e $\delta = (5\%, 10\%, 15\%, 20\%)$. Observa-se, assim, como nos cenários da tabela 4.1, que as médias dos estimadores se aproximam dos verdadeiros valores à medida que o tamanho da amostra aumenta. Tem-se também que o valor do EQM se reduz com o aumento do tamanho da amostra. Além disso, observamos que as estimativas do viés relativos dos estimadores foram consideravelmente pequenas e diminuíram a medida que o tamanho da amostra aumenta. Por exemplo, considerando $\delta = 0.20$, para $n = 50$, observamos que a média, EQM e VR foram 1.3993, 0.0702 e -0.0501 , já para $n = 200$, os valores foram de 1.4004, 0.0108 e 0.0280, respectivamente. Não verificamos diferenças consideráveis entre todos os cenários observados, no que se refere ao aumento do valor referente ao parâmetro de inflação δ .

Na Tabela 4.3 encontram-se os resultados para o estimador pontual da média $\mu = 1.8$ e $\delta = (4\%, 8\%, 12\%, 16\%)$. Observa-se, assim, como nos cenários anteriores, que as médias dos estimadores se aproximam dos verdadeiros valores à medida que o tamanho da amostra aumenta. Tem-se também que o valor do EQM se reduz com o aumento do tamanho da amostra. Além disso, observamos que as estimativas do viés relativos dos estimadores foram consideravelmente pequenas e diminuíram a medida que o tamanho da amostra aumenta. Para exemplificar considerando $\delta = 0.12$, para $n = 50$, observamos que a média, EQM e VR foram 1.8138, 0.0524 e 0.7685, já para $n = 200$, os valores foram 1.8090, 0.0134 e 0.4967, respectivamente.

Dessa forma, os resultados apresentados mostram que o estimador de máxima verossimilhança para o parâmetro μ da distribuição poisson inflacionada em zero, apresenta um comportamento assintótico esperado, evidenciando bom desempenho.

Tabela 4.1: Média, EQM e VR do EMV da média da ZIPM (μ) com $\delta = (0.05, 0.10, 0.15, 0.20)$ e $\mu = 1$

(δ, μ)	Medida	$n = 50$	$n = 100$	$n = 150$	$n = 200$
(0.05; 1.0)	Média	1.0230	1.0158	1.0119	1.0099
	EQM	0.0256	0.0124	0.0084	0.0062
	VR	2.3044	1.5802	1.1947	0.9947
(0.10; 1.0)	Média	1.0224	1.0109	1.0066	1.0052
	EQM	0.0270	0.0131	0.0087	0.0066
	VR	2.2437	1.0927	0.6638	0.5192
(0.15; 1.0)	Média	1.0183	1.0097	1.0048	1.0019
	EQM	0.0271	0.0142	0.0093	0.0070
	VR	1.8250	0.9682	0.4786	0.1933
(0.20; 1.0)	Média	1.0117	1.0026	0.9991	0.9980
	EQM	0.0289	0.0148	0.0100	0.0073
	VR	1.1689	0.2572	-0.0859	-0.1982

Tabela 4.2: Média, EQM e VR do EMV da média da ZIPM (μ) com $\delta = (0.05, 0.10, 0.15, 0.20)$ e $\mu = 1.4$

(δ, μ)	Medida	$n = 50$	$n = 100$	$n = 150$	$n = 200$
(0.05; 1.4)	Média	1.4048	1.4027	1.4012	1.4014
	EQM	0.0633	0.0381	0.0192	0.0095
	VR	0.3457	0.1913	0.0885	0.0976
(0.10; 1.4)	Média	1.4020	1.4017	1.4007	1.4003
	EQM	0.0660	0.0406	0.0206	0.0102
	VR	0.1432	0.1198	0.0465	0.0226
(0.15; 1.4)	Média	1.3993	1.3998	1.4029	1.4004
	EQM	0.0702	0.0434	0.0219	0.0108
	VR	-0.0501	-0.0160	0.2036	0.0280
(0.20; 1.4)	Média	1.3993	1.3998	1.4029	1.4004
	EQM	0.0702	0.0434	0.0219	0.0108
	VR	-0.0501	-0.0160	0.2036	0.0280

Tabela 4.3: Média, EQM e VR do EMV da média da ZIPM (μ) com $\delta = (0.04, 0.08, 0.12, 0.16)$ e $\mu = 1.8$

(δ, μ)	Medida	$n = 50$	$n = 100$	$n = 150$	$n = 200$
(0.04; 1.8)	Média	1.8176	1.8088	1.8044	1.8041
	EQM	0.0443	0.0221	0.0151	0.0111
	VR	0.9789	0.4906	0.2456	0.2268
(0.08; 1.8)	Média	1.8110	1.8057	1.8041	1.8049
	EQM	0.0480	0.0247	0.0163	0.0121
	VR	0.6126	0.3185	0.2300	0.2746
(0.12; 1.8)	Média	1.8138	1.8053	1.8077	1.8090
	EQM	0.0524	0.0267	0.0180	0.0134
	VR	0.7685	0.2944	0.4250	0.4967
(0.16; 1.8)	Média	1.8134	1.8129	1.8176	1.8200
	EQM	0.0558	0.0294	0.0197	0.0147
	VR	0.7418	0.7147	0.9763	1.1107

4.1.2 Avaliação com erro de especificação: gerando ZIPM e estimando Poisson

A estimação da média na distribuição ZIP desempenha um papel fundamental na análise de dados com excesso de zeros. Nesta seção cometemos um erro de estimação, gerando dados inflacionados e estimando a média como sendo proveniente de uma distribuição de Poisson. Essa interpretação equivocada resultou em consequências significativas, afetando tanto a precisão das estimativas quanto a interpretação dos resultados conforme abordamos a seguir.

As Tabelas 4.4, 4.5 e 4.6 trazem os resultados da estimação com erro de especificação cometido, considerando os mesmos cenários das tabelas 4.1, 4.2 e 4.3. Como o parâmetro a ser avaliado é o da Poisson, temos nas Tabelas 4.4, 4.5 e 4.6 o valor de λ para cada valor de μ e δ . De modo geral, percebemos que em todos os cenários a média do estimador esteve distante do valor do parâmetro a medida que aumentamos o valor do n . Observamos uma redução do EQM com o aumento da amostra, no entanto, mesmo com essa redução, o VR relativo permaneceu alto, indicando que ainda há uma considerável variação nos resultados estimados em relação à verdadeira média. Esse alto VR relativo pode ser uma consequência direta da inadequação da distribuição de Poisson para modelar os dados inflacionados e reforça a necessidade de adotar abordagens mais apropriadas, como a distribuição ZIPM. Por exemplo, para $\mu = 1$, $\delta = 0.20$, $\lambda = 1.25$ e $n = 200$ a média, EQM e VR do estimador λ foram 0.9966, 0.0706 e -20.2759 , respectivamente.

Ao assumir erroneamente a distribuição de Poisson para os dados inflacionados, ocorreu uma distorção nas estimativas da média. Isso pode ser atribuído à natureza distintiva das distribuições ZIPM e Poisson, uma vez distribuição ZIPM modela a probabilidade de zeros excessivos e a média de todos os valores.

Essas observações têm implicações importantes para a interpretação dos resultados e conclusões obtidos. É fundamental reconhecer as limitações da suposição de uma distribuição de Poisson e considerar cuidadosamente a escolha da distribuição mais adequada para modelar os dados inflacionados. Explorar métodos estatísticos mais avançados, como a utilização de modelos de regressão ZIP baseado na distribuição ZIPM proposta, pode ajudar a obter estimativas mais precisas, proporcionando uma análise mais robusta e confiável.

Tabela 4.4: Média, EQM e VR da média da distribuição Poisson (λ) com $\delta = (0.05, 0.10, 0.15, 0.20)$ e $\lambda = (1.05, 1.11, 1.18, 1.25)$ - Erro de especificação.

(δ, μ)	λ	Medida	EMV	$n = 50$	$n = 100$	$n = 150$	$n = 200$
(0.05; 1.0)	1.05	Média	$\hat{\lambda}$	1.0025	1.0016	1.0012	1.0015
		EQM	$\hat{\lambda}$	0.0238	0.0131	0.0095	0.0078
		VR	$\hat{\lambda}$	-4.7574	-4.8472	-4.8847	-4.8597
(0.10; 1.0)	1.11	Média	$\hat{\lambda}$	1.0054	1.0009	0.9996	1.0003
		EQM	$\hat{\lambda}$	0.0340	0.0234	0.0199	0.0180
		VR	$\hat{\lambda}$	-9.5156	-9.9185	-10.0325	-9.9730
(0.15; 1.0)	1.18	Média	$\hat{\lambda}$	1.0042	1.0016	1.0006	0.9993
		EQM	$\hat{\lambda}$	0.0530	0.0427	0.0390	0.0375
		VR	$\hat{\lambda}$	-14.6398	-14.8585	-14.9516	-15.0584
(0.20; 1.0)	1.25	Média	$\hat{\lambda}$	1.0009	0.9979	0.9962	0.9966
		EQM	$\hat{\lambda}$	0.0871	0.0764	0.0730	0.0706
		VR	$\hat{\lambda}$	-19.9364	-20.1686	-20.2979	-20.2759

Tabela 4.5: Média, EQM e VR da média da distribuição Poisson (λ) com $\delta = (0.05, 0.10, 0.15, 0.20)$ e $\lambda = (1.47, 1.56, 1.65, 1.75)$ - Erro de especificação.

(δ, μ)	λ	Medida	EMV	$n = 50$	$n = 100$	$n = 150$	$n = 200$
(0.05; 1.4)	1.47	Média	$\hat{\lambda}$	1.4013	1.4018	1.4005	1.4008
		EQM	$\hat{\lambda}$	0.0363	0.0201	0.0156	0.0129
		VR	$\hat{\lambda}$	-4.9140	-4.8782	-4.9647	-4.9462
(0.10; 1.4)	1.56	Média	$\hat{\lambda}$	1.4025	1.4015	1.4018	1.4011
		EQM	$\hat{\lambda}$	0.0566	0.0401	0.0345	0.0321
		VR	$\hat{\lambda}$	-9.8388	-9.9032	-9.8819	-9.9247
(0.15; 1.4)	1.65	Média	$\hat{\lambda}$	1.4039	1.4022	1.4008	1.4020
		EQM	$\hat{\lambda}$	0.0951	0.0779	0.0728	0.0691
		VR	$\hat{\lambda}$	-14.7649	-14.8672	-14.9524	-14.8807
(0.20; 1.4)	1.75	Média	$\hat{\lambda}$	1.3972	1.4019	1.4015	1.4003
		EQM	$\hat{\lambda}$	0.1631	0.1409	0.1347	0.1320
		VR	$\hat{\lambda}$	-20.1614	-19.8924	-19.9161	-19.9836

Tabela 4.6: Média, EQM e VR da média da distribuição Poisson (λ) com $\delta = (0.04, 0.08, 0.12, 0.16)$ e $\lambda = (1.88, 1.96, 2.05, 2.14)$ - Erro de especificação.

(δ, μ)	λ	Medida	EMV	$n = 50$	$n = 100$	$n = 150$	$n = 200$
(0.04; 1.8)	1.88	Média	$\hat{\lambda}$	1.7998	1.7991	1.7978	1.7990
		EQM	$\hat{\lambda}$	0.0435	0.0246	0.0188	0.0152
		VR	$\hat{\lambda}$	-4.0101	-4.0491	-4.1196	-4.0529
(0.08; 1.8)	1.96	Média	$\hat{\lambda}$	1.7997	1.8007	1.8016	1.8030
		EQM	$\hat{\lambda}$	0.0661	0.0450	0.0381	0.0339
		VR	$\hat{\lambda}$	-8.0157	-7.9648	-7.9202	-7.8462
(0.12; 1.8)	2.05	Média	$\hat{\lambda}$	1.8063	1.8029	1.8070	1.8081
		EQM	$\hat{\lambda}$	0.1021	0.0818	0.0724	0.0677
		VR	$\hat{\lambda}$	-11.6934	-11.8572	-11.6607	-11.6029
(0.16; 1.8)	2.14	Média	$\hat{\lambda}$	1.8091	1.8114	1.8172	1.8198
		EQM	$\hat{\lambda}$	0.1592	0.1350	0.1228	0.1167
		VR	$\hat{\lambda}$	-15.5739	-15.4681	-15.1964	-15.0773

4.2 Estimação Intervalar - máxima verossimilhança

4.2.1 Avaliação sem cometer erro de especificação

Nesta seção apresentamos os resultados das simulações de Monte Carlo para a avaliação da taxa (%) de cobertura e não cobertura dos IC assintóticos da distribuição ZIPM relacionados à distribuição. Para tal avaliação, consideramos níveis de confiança de 95% e de 99%.

Conforme Chen et al. (2003), quando uma população contém muitos valores zeros e o tamanho da amostra é pequeno, os intervalos de confiança baseados na aproximação normal para a média populacional podem ter baixas taxas de cobertura. Segundo Dietz e Böhning (2000), a normalidade assintótica de uma estimativa de máxima verossimilhança pode ser afetada em situações onde o tamanho da amostra é pequeno. Eles mostraram que para uma distribuição ZIP, quando $\lambda = 3$ e $\lambda = 5$, os ICs tem uma boa taxa de cobertura para a maioria dos tamanhos de amostra ($n = 30, 60, 100, 200, 500$). Quando λ tem um valor baixo, como 1, e o tamanho da amostra é pequeno, os ICs apresentam resultados inferiores. Além disso, para um tamanho amostral pequeno n , quando δ aumenta, a taxa de cobertura dos ICs tende a diminuir.

A Tabela 4.7 apresenta os resultados das taxas de coberturas e não coberturas dos intervalos de confiança para a média da distribuição ZIPM, considerando $\mu = 1$ e $\delta = (0.05, 0.10, 0.15, 0.20)$. Com base nos resultados percebemos que as taxas de coberturas empíricas convergem lentamente para as taxas nominais à medida que o tamanho da amostra aumenta. Percebe-se que as taxas de coberturas empíricas ficaram acima das taxas nominais em todos os cenários e que com o nível de significância de 99% os resultados foram melhores. Por exemplo, com o nível de significância de 99%, $\mu = 1$ e $\delta = 0.20$, tivemos a taxa de cobertura de 99.16% e as taxas de não cobertura de (0.25%) à esquerda e de (0.59%) à direita. Adicionalmente, as taxas de cobertura para média μ com o nível de significância de 95% ficaram relativamente menores para $\delta = 0.20$ quando comparadas com os cenários em que $\delta = 0.05$, $\delta = 0.12$ e $\delta = 0.15$.

A Tabela 4.8 apresenta os resultados das taxas de coberturas e não coberturas de intervalos de confiança para a média da distribuição ZIPM, considerando $\mu = 1.4$ e $\delta = (0.05, 0.10, 0.15, 0.20)$. Com base nos resultados percebemos que as taxas de coberturas empíricas se aproximam das taxas nominais à medida que o tamanho da amostra aumenta. Adicionalmente, as taxas de cobertura para média μ ficaram relativamente maiores para $\delta = 0.05$ quando comparadas com os cenários em que $\delta = 0.10$, $\delta = 0.15$ e $\delta = 0.20$. As maiores taxas de não cobertura foram observadas quando consideramos $n = 50$ e $\delta = 0.05$ ao nível de significância de 95%, sendo 1.36 à esquerda e 1.03 à direita.

Tabela 4.7: Taxa de cobertura (95% e 99%) e de não cobertura dos intervalos de confiança para média da distribuição Poisson inflacionada em zero com média $\mu = 1$ e $\delta = (0.05, 0.10, 0.15, 0.20)$.

δ	$1 - \alpha$	$n = 50$	$n = 100$	$n = 150$	$n = 200$
0.05	95%	98.83 (0.64; 0.53)	98.86 (0.68; 0.46)	98.62 (0.90; 0.48)	98.23 (1.00; 0.77)
0.10	95%	98.59 (0.76; 0.65)	98.46 (0.93; 0.61)	98.50 (0.78; 0.72)	98.30 (0.89; 0.81)
0.15	95%	98.75 (0.61; 0.64)	98.42 (0.94; 0.64)	98.01 (1.01; 0.98)	97.87 (0.97; 1.16)
0.20	95%	98.28 (0.77; 0.95)	98.27 (0.81; 0.92)	97.71 (1.05; 1.24)	97.45 (0.91; 1.64)
0.05	99%	99.87 (0.09; 0.04)	99.93 (0.06; 0.01)	99.89 (0.08; 0.03)	99.90 (0.04; 0.06)
0.10	99%	99.89 (0.07; 0.04)	99.92 (0.07; 0.01)	99.85 (0.07; 0.08)	99.77 (0.07; 0.16)
0.15	99%	99.87 (0.06; 0.07)	99.83 (0.09; 0.08)	99.71 (0.10; 0.19)	99.85 (0.09; 0.06)
0.20	99%	99.80 (0.06; 0.14)	99.81 (0.11; 0.08)	99.82 (0.11; 0.07)	99.16 (0.25; 0.59)

Tabela 4.8: Taxa de cobertura (95% e 99%) e de não cobertura dos intervalos de confiança para média da distribuição Poisson inflacionada em zero com média $\mu = 1.4$ e $\delta = (0.06, 0.10, 0.12)$.

δ	$1 - \alpha$	$n = 50$	$n = 100$	$n = 150$	$n = 200$
0.05	95%	97.61 (1.36; 1.03)	97.19 (1.58; 1.23)	96.79 (1.66; 1.55)	96.72 (1.62; 1.66)
0.10	95%	97.20 (1.36; 1.44)	96.74 (1.53; 1.73)	96.84 (1.50; 1.66)	96.00 (1.81; 2.19)
0.15	95%	96.67 (1.61; 1.72)	96.06 (1.77; 2.17)	95.57 (1.91; 2.52)	95.65 (1.98; 2.37)
0.20	95%	96.45 (1.5; 2.05)	95.62 (1.87; 2.51)	95.35 (2.16; 2.49)	94.78 (2.27; 2.95)
0.05	99%	99.64 (0.13; 0.23)	99.71 (0.14; 0.15)	99.48 (0.18; 0.34)	99.45 (0.18; 0.37)
0.10	99%	99.60 (0.13; 0.27)	99.52 (0.18; 0.30)	99.51 (0.18; 0.31)	99.20 (0.35; 0.45)
0.15	99%	99.49 (0.24; 0.27)	99.36 (0.30; 0.34)	99.20 (0.35; 0.45)	99.25 (0.30; 0.45)
0.20	99%	99.51 (0.18; 0.31)	99.20 (0.25; 0.55)	99.16 (0.27; 0.57)	98.95 (0.32; 0.73)

A Tabela 4.9 apresenta os resultados das taxas de coberturas e não coberturas de intervalos de confiança para a média da distribuição ZIPM, considerando $\mu = 1.8$ e $\delta = (0.04, 0.08, 0.12, 0.16)$. Diante os resultados obtidos, observamos que as taxas de cobertura empíricas tendem a se aproximar das taxas nominais à medida que o tamanho da amostra aumenta. Adicionalmente, as taxas de cobertura para média μ ficaram relativamente maiores para $\delta = 0.04$ quando comparadas com os cenários em que $\delta = 0.08$, $\delta = 0.12$ e $\delta = 0.16$. As maiores taxas de não cobertura foram observadas quando consideramos $n = 50$ e $\delta = 0.04$ ao nível de significância de 95%, sendo 1.46 à esquerda e 1.85 à direita.

Analisando as taxas de cobertura dos Intervalos de Confiança (IC) notamos melhores resultados para $\mu = 1.4$ e $\mu = 1.8$, quando comparadas com $\mu = 1$, para todos tamanhos de amostras considerados. Essa melhoria sugere que o IC proporciona uma estimativa mais confiável e acurada para valores maiores da média.

Tabela 4.9: Taxa de cobertura (95% e 99%) e de não cobertura dos intervalos de confiança para média da distribuição Poisson inflacionada em zero com média $\mu = 1.8$ e $\delta = (0.04, 0.08, 0.12, 0.16)$.

δ	$1 - \alpha$	$n = 50$	$n = 100$	$n = 150$	$n = 200$
0.04	95%	96.69	96.22	95.88	95.98
		(1.46; 1.85)	(1.83; 1.95)	(1.96; 2.16)	(1.98; 2.04)
0.08	95%	96.21	95.16	95.51	95.20
		(2.06; 3.45)	(2.08; 2.76)	(2.07; 2.42)	(2.32; 2.48)
0.12	95%	95.22	94.72	94.64	94.46
		(2.05; 2.73)	(2.38; 2.90)	(2.61; 2.75)	(2.82; 2.72)
0.16	95%	94.96	94.01	93.71	93.63
		(2.15; 2.89)	(2.98; 3.01)	(3.52; 2.77)	(3.82; 2.55)
0.04	99%	99.46	99.52	99.29	99.27
		(0.20; 0.34)	(0.21; 0.27)	(0.22; 0.49)	(0.37; 0.36)
0.08	99%	99.35	99.16	99.08	99.09
		(0.13; 0.52)	(0.35; 0.49)	(0.34; 0.58)	(0.38; 0.53)
0.12	99%	99.04	98.91	98.79	98.79
		(0.29; 0.67)	(0.38; 0.71)	(0.44; 0.77)	(0.48; 0.73)
0.16	99%	98.95	98.75	98.62	98.48
		(0.33; 0.72)	(0.55; 0.70)	(0.75; 0.63)	(0.93; 0.59)

4.2.2 Avaliação com erro de especificação

Nesta seção apresentamos os resultados das simulações de Monte Carlo da distribuição Poisson para a avaliação da taxa (%) de cobertura e não cobertura dos IC assintóticos relacionados à distribuição cometendo o erro de especificação quando utilizamos dados inflacionados na estimação. Para tal avaliação, consideramos níveis de confiança de 95% e de 99%. Os cenários considerados foram os mesmos das tabelas 4.4, 4.5 e 4.6, sendo divididos nas tabelas 4.10, 4.11 e 4.12.

De modo geral, as taxas de cobertura dos ICs diminuem consideravelmente à medida que o tamanho da amostra aumentou. Percebe-se também que a medida que o parâmetro de inflação aumenta, e conseqüentemente o parâmetro λ , a taxa de não cobertura à direita também aumenta significativamente. Por exemplo, na tabela 4.10, para $\lambda = 1.05$, $\delta = 0.05$ e $n = 200$, a taxa de cobertura, ao nível de confiança de 95%, ficou em 87.94%, 0.24% à esquerda e 11.82% à direita. Já quando o $\lambda = 1.25$, $\delta = 0.20$ e $n = 200$, a taxa de cobertura, ao nível de confiança de 95%, ficou em 9.05%, 0.00% à esquerda e 90.49% à direita.

A baixa taxa de cobertura do IC indica que, na maioria dos casos, o intervalo estimado não inclui a verdadeira média da distribuição. Essa falta de precisão no IC pode levar a conclusões errôneas e a tomada de decisões equivocadas com base nos resultados obtidos. Esses resultados têm implicações importantes para a interpretação dos resultados da pesquisa e destacam a necessidade da utilização da distribuição ZIPM para abordagens mais adequadas na estimação intervalar.

Tabela 4.10: Taxa de cobertura (95% e 99%) e de não cobertura dos intervalos de confiança para média da distribuição Poisson inflacionada em zero com média $\lambda = (1.05, 1.11, 1.18, 1.25)$ e $\delta = (0, 05, 0.10, 0.15, 0.20)$. - Erro de especificação.

μ	δ	λ	$1 - \alpha$	$n = 50$	$n = 100$	$n = 150$	$n = 200$
1.0	0.05	1.05	95%	89.96 (0.78; 9.26)	90.57 (0.39; 9.04)	88.01 (0.32; 11.67)	87.94 (0.24; 11.82)
1.0	0.10	1.11	95%	84.79 (0.26; 14.95)	76.01 (0.13; 23.86)	68.63 (0.06; 31.31)	64.31 (0.04; 35.65)
1.0	0.15	1.18	95%	72.51 (0.09; 27.40)	55.20 (0.02; 44.78)	42.09 (0.00; 57.91)	31.00 (0.00; 69.00)
1.0	0.20	1.25	95%	57.34 (0.02; 42.64)	33.78 (0.00; 66.22)	17.74 (0.00; 82.53)	9.05 (0.00; 90.49)
1.0	0.05	1.05	99%	97.07 (0.07; 2.86)	96.64 (0.03; 3.33)	96.01 (0.01; 3.89)	95.15 (0.02; 4.83)
1.0	0.10	1.11	99%	92.73 (0.04; 7.23)	88.52 (0.01; 11.47)	84.91 (0.00; 15.09)	81.53 (0.00; 18.47)
1.0	0.15	1.18	99%	84.74 (0.01; 15.25)	75.44 (0.00; 24.56)	62.89 (0.00; 37.11)	52.59 (0.00; 47.41)
1.0	0.20	1.25	99%	71.72 (0.00; 28.28)	50.12 (0.00; 49.88)	32.59 (0.00; 67.41)	20.35 (0.00; 79.65)

Tabela 4.11: Taxa de cobertura (95% e 99%) e de não cobertura dos intervalos de confiança para média da distribuição Poisson inflacionada em zero com média $\lambda = (1.47, 1.56, 1.65, 1.75)$ e $\delta = (0, 05, 0.10, 0.15, 0.20)$. - Erro de especificação.

μ	δ	λ	$1 - \alpha$	$n = 50$	$n = 100$	$n = 150$	$n = 200$
1.4	0.05	1.47	95%	89.89 (0.72; 9.39)	88.34 (0.44; 11.22)	86.21 (0.25; 13.54)	84.49 (0.19; 15.32)
1.4	0.10	1.56	95%	79.61 (0.22; 20.17)	72.38 (0.07; 27.55)	61.75 (0.00; 38.25)	53.93 (0.03; 46.04)
1.4	0.15	1.65	95%	63.85 (0.05; 36.15)	45.84 (0.01; 54.15)	30.51 (0.01; 69.48)	21.67 (0.00; 78.33)
1.4	0.20	1.75	95%	46.68 (0.00; 53.32)	23.70 (0.00; 76.30)	10.20 (0.00; 89.80)	4.06 (0.00; 95.94)
1.4	0.05	1.47	99%	96.36 (0.03; 3.61)	95.92 (0.04; 4.44)	94.93 (0.01; 5.06)	95.15 (0.02; 4.83)
1.4	0.10	1.56	99%	89.89 (0.02; 10.09)	85.86 (0.00; 14.14)	79.27 (0.00; 20.73)	74.01 (0.02; 25.97)
1.4	0.15	1.65	99%	79.25 (0.00; 20.75)	66.47 (0.00; 33.53)	51.07 (0.00; 48.93)	39.29 (0.00; 60.71)
1.4	0.20	1.75	99%	62.23 (0.00; 37.77)	38.11 (0.00; 61.89)	22.22 (0.00; 77.78)	9.90 (0.00; 90.10)

Tabela 4.12: Taxa de cobertura (95% e 99%) e de não cobertura dos intervalos de confiança para média da distribuição Poisson inflacionada em zero com média $\lambda = (1.88, 1.96, 2.05, 2.14)$ e $\delta = (0, 04, 0.08, 0.12, 0.16)$. - Erro de especificação.

μ	δ	λ	$1 - \alpha$	$n = 50$	$n = 100$	$n = 150$	$n = 200$
1.8	0.04	1.88	95%	91.17 (0.79; 8.04)	89.35 (0.40; 10.25)	86.94 (0.33; 12.73)	86.09 (0.31; 13.60)
1.8	0.08	1.96	95%	82.19 (0.24; 17.57)	74.19 (0.07; 25.06)	68.83 (0.06; 31.11)	61.74 (0.03; 38.23)
1.8	0.12	2.05	95%	71.01 (0.09; 28.9)	54.54 (0.01; 45.45)	42.47 (0.01; 57.52)	32.19 (0.00; 67.81)
1.8	0.16	2.14	95%	57.31 (0.02; 42.67)	33.99 (0.00; 66.01)	29.59 (0.00; 79.41)	12.89 (0.00; 87.11)
1.8	0.04	1.88	99%	97.28 (0.09; 2.63)	96.45 (0.02; 3.53)	94.99 (0.01; 5.00)	94.58 (0.02; 5.40)
1.8	0.08	1.96	99%	92.12 (0.03; 7.78)	88.79 (0.01; 11.2)	84.14 (0.00; 15.58)	80.06 (0.01; 19.93)
1.8	0.12	2.05	99%	84.98 (0.01; 15.01)	73.66 (0.00; 26.34)	63.10 (0.00; 36.90)	51.34 (0.00; 48.66)
1.8	0.16	2.14	99%	73.80 (0.00; 26.20)	54.13 (0.00; 45.87)	38.48 (0.00; 61.52)	25.82 (0.00; 74.18)

4.3 Estimação pontual - Método dos momentos

Apresentamos nesta seção a avaliação do estimador para μ pelo método dos momentos à partir das simulações de Monte Carlo. Focamos na avaliação das estimativas pontuais, calculando a média, o erro quadrático médio (EQM) e o viés relativo (VR) do estimador para a média μ pelo método dos momentos.

Os resultados que são apresentados neste capítulo são oriundos de simulações considerando 10.000 réplicas de Monte Carlo e amostras de tamanhos 50, 100, 150 e 200. Todas as simulações foram realizadas com o *software* estatístico R.

4.3.1 Estimação pontual - Método dos momentos

. Na Tabela 4.13 encontram-se os resultados das medidas utilizadas na avaliação dos estimadores pontuais da distribuição ZIPM com média $\mu = 1$, considerando δ igual a 5%, 10%, 15% e 20%. Observa-se que as médias dos estimadores se aproximam dos verdadeiros valores à medida que o tamanho da amostra aumenta. Tem-se também que o valor do EQM se reduz com o aumento do tamanho da amostra, considerando todos os valores do δ . Além disso, observamos que as estimativas do viés relativos dos estimadores foram consideravelmente pequenas e diminuíram à medida que o tamanho da amostra aumenta. Para exemplificar considere $\delta = 0.20$. Quando $n = 50$, observamos que a média, EQM e VR foram 0.9976, 0.0254 e -0.2312 e para $n = 200$, temos 0.9959, 0.0064 e -0.4091 , respectivamente.

Na Tabela 4.14 encontra-se os resultados para os estimadores pontuais da distribuição ZIPM com média $\mu = 1.4$ e δ igual a 5%, 10%, 15% e 20%. Na Tabela 4.15 encontra-se os resultados para os estimadores pontuais da distribuição ZIPM com média $\mu = 1.8$ e δ igual a 4%, 8%, 12% e 16%. Aumentamos o valor da média da distribuição ZIPM com intuito de verificar se existe alguma mudança na avaliação dos cenários presentes na Tabela 4.13. Para ambas as médias, $\mu = 1.4$ e $\mu = 1.8$, e para todos valores de δ considerados, observamos que as médias dos estimadores se aproximam dos verdadeiros valores à medida que o tamanho da amostra aumenta. Tem-se também que o valor do EQM se reduz com o aumento do tamanho da amostra. Além disso, observamos que as estimativas do viés relativos dos estimadores foram consideravelmente pequenas e diminuíram à medida que o tamanho da amostra aumenta. Não verificamos diferenças consideráveis entre todos os cenários observados, no que se refere ao aumento do valor referente ao parâmetro de inflação δ .

Portanto, os resultados apresentados mostram que o estimador obtido pelo método dos momentos para a média da distribuição poisson inflacionada em zero com parâmetros μ e δ , apresenta um comportamento consistente e não viesado.

Com base na estimação pontual da média da distribuição ZIPM utilizando os métodos de estimação de máxima verossimilhança (EMV) e dos momentos, pode-se con-

cluír que ambos os estimadores produziram resultados satisfatórios em relação à média estimada, erro quadrático médio (EQM) e viés relativo (VR). Não foram observadas diferenças significativas entre os resultados obtidos pelos dois métodos. Isso sugere que tanto o EMV quanto o método dos momentos são abordagens viáveis e eficazes para estimar a média da distribuição ZIPM. Portanto, a escolha entre esses métodos pode depender de outros fatores, como a facilidade de implementação ou a disponibilidade de informações adicionais sobre a distribuição estudada.

Tabela 4.13: Média, EQM e VR do EMV da média da ZIPM (μ) com $\delta = (0.05, 0.10, 0.15, 0.20)$ e $\mu = 1$.

(δ, μ)	Medida	$n = 50$	$n = 100$	$n = 150$	$n = 200$
(0.05; 1.0)	Média	1.0019	1.0003	0.9999	1.0005
	EQM	0.0218	0.0106	0.0070	0.0052
	VR	0.1916	0.0304	-0.0072	0.0453
(0.10; 1.0)	Média	1.0050	0.9995	0.9998	1.0001
	EQM	0.0233	0.0114	0.0075	0.0057
	VR	0.4950	-0.0503	-0.0176	0.0086
(0.15; 1.0)	Média	1.0021	0.9999	0.9998	0.9983
	EQM	0.0240	0.0124	0.0080	0.0061
	VR	0.2180	-0.0059	-0.0149	-0.1692
(0.20; 1.0)	Média	0.9976	0.9962	0.9991	0.9959
	EQM	0.0254	0.0128	0.0087	0.0064
	VR	-0.2312	-0.3765	-0.4857	-0.4091

Tabela 4.14: Média, EQM e VR do EMV da média da ZIPM (μ) com $\delta = (0.05, 0.10, 0.15, 0.20)$ e $\mu = 1.4$.

(δ, μ)	Medida	$n = 50$	$n = 100$	$n = 150$	$n = 200$
(0.05; 1.4)	Média	1.3981	1.3999	1.3996	1.3993
	EQM	0.0302	0.0151	0.0101	0.0075
	VR	-0.1331	-0.0077	-0.0241	-0.0483
(0.10; 1.4)	Média	1.3988	1.3997	1.4009	1.3994
	EQM	0.0330	0.0167	0.0109	0.0083
	VR	-0.0891	-0.0206	0.0612	-0.0407
(0.15; 1.4)	Média	1.3977	1.4000	1.3997	1.4008
	EQM	0.0352	0.0180	0.0121	0.0091
	VR	-0.1646	-0.0021	-0.0181	0.0586
(0.20; 1.4)	Média	1.3943	1.3997	1.4001	1.3997
	EQM	0.0389	0.0198	0.0133	0.0098
	VR	-0.4055	-0.0199	0.0099	-0.0185

Tabela 4.15: Média, EQM e VR do EMV da média da ZIPM (μ) com $\delta = (0.04, 0.08, 0.12, 0.16)$ e $\mu = 1.8$.

(δ, μ)	Medida	$n = 50$	$n = 100$	$n = 150$	$n = 200$
(0.04; 1.8)	Média	1.7954	1.7967	1.7982	1.7984
	EQM	0.0383	0.0190	0.0129	0.0094
	VR	-0.2569	-0.1841	0.1024	-0.0869
(0.08; 1.8)	Média	1.7955	1.7971	1.7999	1.8017
	EQM	0.0421	0.0205	0.0141	0.0103
	VR	-0.2520	-0.1619	-0.0049	0.0964
(0.12; 1.8)	Média	1.7989	1.7995	1.8038	1.8066
	EQM	0.0447	0.0227	0.0153	0.0113
	VR	-0.0593	-0.0272	0.2109	0.3673
(0.16; 1.8)	Média	1.7998	1.8071	1.8143	1.8179
	EQM	0.0486	0.0250	0.0166	0.9970
	VR	-0.0098	0.3936	0.7944	-0.1982

Capítulo 5

Aplicações

Neste capítulo, abordaremos duas aplicações a dados reais utilizando a distribuição Poisson inflacionada em zero, com média μ , a fim de destacar sua relevância como uma alternativa à dados de contagem inflacionados em zero. A primeira variável utilizada é o registro do número de movimentos respiratórios e corporais em fetos de cordeiros nos últimos dois terços da gestação. Os dados foram utilizados por Leroux et al. (1994) EM em um estudo sobre distribuições discretas. A segunda variável apresenta o número de viúvas e a quantidade de filhos que tinham direito ao benefício, os dados foram do relatório anual do fundo de pensão S.P.P. para o ano de 1952 e citados por Kennedy e Gentle (1980).

5.1 Dados de movimento fetal

Em um estudo sobre os movimentos respiratórios e corporais em fetos de cordeiros, realizado com o objetivo de investigar possíveis mudanças na quantidade e padrão de atividade fetal durante os últimos dois terços da gestação, foram registrados os números de movimentos do feto por meio de ultrassom (LEROUX et al., 1994). Tais alterações na atividade podem ser atribuídas a fatores físicos, como a redução no volume de fluido amniótico e a presença de espaço vazio dentro do útero, além do desenvolvimento do sistema nervoso central.

Neste estudo, foram analisadas uma sequência específica de contagens de números de movimentos em 240 intervalos consecutivos de 5 segundos. As análises aqui discutidas fornecem uma descrição do padrão de atividade fetal em um ponto fixo da gestação. No conjunto de dados do estudo verificamos que cerca de 76% das observações que representam o número de movimentos são iguais a zero. Dessa forma, a distribuição Poisson não seria adequada para modelar esses dados, por isso a importância de utilizarmos a distribuição ZIPM.

A Tabela 5.1 apresenta a distribuição de probabilidade do número de movimentos onde observamos uma grande proporção de valores iguais a 0. Uma maneira mais confiável de escolher entre a distribuição de Poisson e a distribuição ZIPM é aplicar o teste de

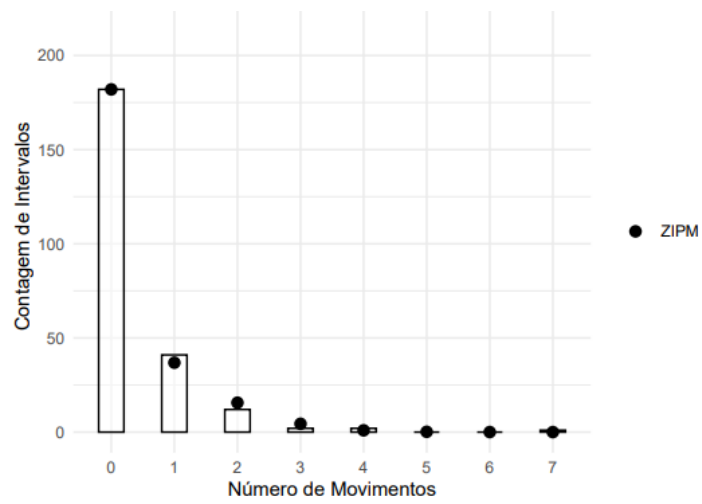
pontuação de Broek (1995). Com base no teste de pontuação, temos que o p -valor = 9.86×10^{-7} , ou seja, H_0 deve ser rejeitada e concluímos que os dados não seguem uma distribuição de Poisson, mas sim uma distribuição ZIPM.

Tabela 5.1: Distribuição de frequência do número de Movimentos

Número de Movimentos	0	1	2	3	4	5	6	7
Frequência	182	41	12	2	2	0	0	1

Na Figura 5.1 apresentamos o histograma da distribuição referente aos dados dos cordeiros fetais e a representação gráfica da ZIPM. Observamos que os pontos da distribuição ZIPM e as barras do histograma estão bem alinhados, o que indica uma boa correspondência entre a distribuição teórica e os dados observados. Logo, podemos concluir que a distribuição ZIPM se ajusta bem aos dados observados e é uma escolha adequada para modelar a distribuição dos valores em questão.

Figura 5.1: Histograma e função de probabilidade estimada para o número de movimentos



A estimativa de máxima verossimilhança para média da distribuição ZIPM com esses dados é $\hat{\mu} = 0.3584$. Por sua vez, a estimativa pelo método dos momentos para média da distribuição ZIPM com esses dados é $\hat{\mu} = 0.3583$.

Apesar dos resultados indicarem uma rejeição da hipótese nula (H_0) com base no teste de pontuação de Broek (1995), é importante considerar a possibilidade de um erro de especificação no modelo utilizado para a estimação da média. Caso tivéssemos cometido o erro de especificação e a média tivesse sido estimada erroneamente pela distribuição Poisson, teríamos obtido um valor de $\hat{\mu} = 0.8473$, em contraste com as estimativas corretas da distribuição ZIPM que resultaram em $\hat{\mu} = 0.3584$ pelo método de máxima verossimilhança e $\hat{\mu} = 0.3583$ pelo método dos momentos. Essa diferença considerável ressalta a importância de utilizar o modelo apropriado na análise estatística.

5.2 Beneficiários do fundo de pensão

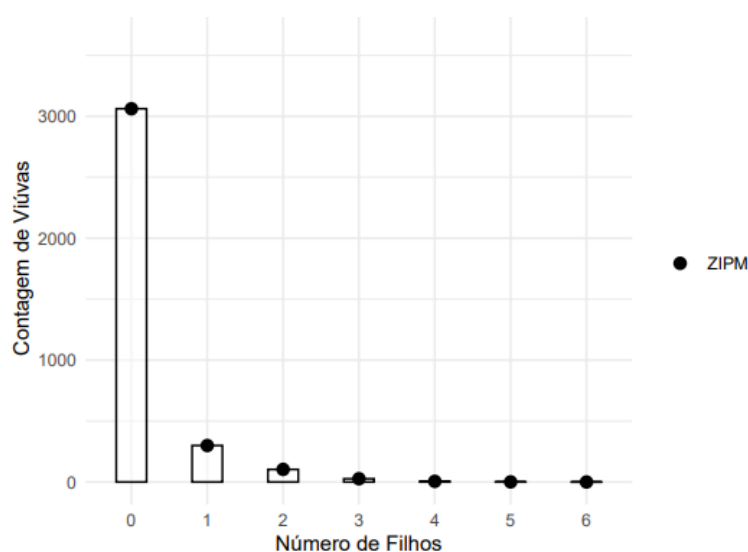
Analisamos nesta aplicação os dados do relatório anual do fundo de pensão S.P.P. para o ano de 1952, conforme citado por Kennedy e Gentle (1980). O relatório informa que aproximadamente 4.075 viúvas receberam pensões do fundo. A Tabela 5.2 apresenta a distribuição de frequência do número de filhos por viúva que tinha direito a apoio do fundo. Observamos a predominância de viúvas sem filhos, cerca de 75%, indicando que os dados podem está inflacionados. Assim, a utilização da distribuição Poisson não seria adequada para modelar esses dados, sendo necessário a utilização a distribuição ZIPM. Aqui aplicamos novamente o modelo objeto de nosso estudo, a distribuição ZIPM, a fim de determinar os possíveis estimadores.

Tabela 5.2: Distribuição de frequência do número de filhos por viúvas que tinham direito a apoio do fundo

Número de filhos	0	1	2	3	4	5	6
Frequência	3062	587	284	103	33	4	2

Na Figura 5.2 apresentamos o histograma da distribuição dos dados e a representação gráfica da ZIPM. Observamos que os pontos da distribuição ZIPM e as barras do histograma estão bem alinhados, o que indica uma boa correspondência entre a distribuição teórica e os dados observados. A semelhança entre o formato do histograma e do gráfico ZIPM sugere que essa distribuição é uma escolha adequada para modelar os dados em questão.

Figura 5.2: Histograma e função de probabilidade estimada para o número de filhos



Com base no teste de Broek (1995), temos que o p -valor $< 2,401757 \times 10^{-142}$, ou seja, H_0 deve ser rejeitada e portanto os dados não seguem uma distribuição de Poisson,

mas sim uma distribuição ZIPM. A estimativa de máxima verossimilhança para média da distribuição ZIPM com esses dados é $\hat{\mu} = 0.3995$. Além disso, a estimativa pelo método dos momentos para média da distribuição ZIPM com esses dados é $\hat{\mu} = 0.3995$. Caso tivesse sido cometido um erro de especificação e a média tivesse sido estimada erroneamente pela distribuição Poisson, teríamos obtido um valor de $\hat{\mu} = 1.0377$, em contraste com as estimativas corretas da distribuição ZIPM que resultaram em $\hat{\mu} = 0.3995$ tanto pelo método de máxima verossimilhança quanto pelo método dos momentos.

Capítulo 6

Conclusões

Com base na nova parametrização proposta para a Poisson inflacionada, denominada de ZIPM, que introduz os parâmetros de média μ e de inflação δ , esta dissertação representa uma contribuição significativa para o campo da estatística. A introdução da ZIPM abre caminho para uma modelagem mais flexível e precisa de dados com excesso de zeros, fornecendo uma abordagem mais robusta para a análise desses tipos de dados.

Ao incorporar os parâmetros μ e δ na parametrização da Poisson inflacionada em zero, a ZIPM nos permite um controle mais refinado da média e da inflação, possibilitando uma descrição mais precisa do fenômeno em estudo. Essa nova parametrização amplia o leque de ferramentas estatísticas disponíveis e fornece uma opção adicional para os pesquisadores lidarem com dados inflacionados em zero.

Os resultados obtidos nesta dissertação, tanto por meio do método de Estimação de Máxima Verossimilhança (EMV) quanto pelo método dos Momentos, validaram a eficácia da ZIPM na estimação da média μ . Os estimadores obtidos demonstraram boa qualidade em termos de média estimada, erro quadrático Médio (EQM) e viés relativo (VR). Além disso, foram construídos intervalos de confiança para estimativas da média μ , permitindo uma avaliação mais robusta.

Entretanto, é importante reconhecer que, apesar dos resultados encorajadores, a ZIPM e os métodos de estimação utilizados podem apresentar erros de especificação em determinados cenários. A qualidade das estimativas e dos intervalos de confiança pode ser afetada por pressupostos não atendidos, como dados não inflacionados ou vice-versa. É essencial estar ciente dessas possíveis limitações e considerar cuidadosamente a adequação do modelo ZIPM e a robustez das estimativas em cada contexto específico de aplicação.

Portanto, a introdução da nova parametrização ZIPM representa uma contribuição valiosa para a área de estatística e oferece uma alternativa promissora para a modelagem de dados inflacionados em zero. A ZIPM pode ser aplicada em diversas áreas de estudo, auxiliando pesquisadores na análise e interpretação de conjuntos de dados complexos com excesso de zeros.

Referências Bibliográficas

BASSIL, K. L.; COLE, D. C.; MOINEDDIN, R.; LOU, W.; CRAIG, A. M.; SCHWARTZ, B.; REA, E. The relationship between temperature and ambulance response calls for heat-related illness in toronto, ontario, 2005. **Journal of Epidemiology & Community Health**, BMJ Publishing Group Ltd, v. 65, n. 9, p. 829–831, 2011.

BAYES, C. L.; VALDIVIESO, L. A beta inflated mean regression model for fractional response variables. **Journal of Applied Statistics**, Taylor Francis, v. 43, p. 1814–1830, 02 2016.

BOLFARINE, H.; SANDOVAL, M. C. **Introdução à inferência estatística**. [S.l.]: SBM, 2001. v. 2.

BROEK, J. Van den. A score test for zero inflation in a poisson distribution. **Biometrics**, v. 51, p. 738–743, 07 1995.

CHEN, J.; CHEN, S.-Y.; RAO, J. Empirical likelihood confidence intervals for the mean of a population containing many zero values. **Canadian Journal of Statistics**, v. 31, p. 53 – 68, 03 2003.

DIETZ, E.; BÖHNING, D. On estimation of the poisson parameter in zero-modified poisson models. **Computational Statistics & Data Analysis**, v. 34, p. 441–459, 02 2000.

Frank E. Harrell Jr. **extraDistr: Additional Univariate and Multivariate Distributions**. [S.l.], 2021. R package version 1.9.1. Disponível em: <<https://cran.r-project.org/package=extraDistr>>.

HILBE, J. M. **Modeling count data**. 1. ed. New York: Cambridge University Press, 2014.

JANSAKUL, N.; HINDE, J. P. Score tests for zero-inflated poisson models. **Computational Statistics & Data Analysis**, v. 40, n. 1, p. 75–96, 2002. Disponível em: <<https://EconPapers.repec.org/RePEc:eee:csdana:v:40:y:2002:i:1:p:75-96>>.

KENNEDY, W.; GENTLE, J. **Statistical computing**. [S.l.]: Marcel Dekker., 1980.

KUMAR, C. S.; RAMACHANDRAN, R. On some aspects of a zero-inflated overdispersed model and its applications. **Journal of Applied Statistics**, Taylor Francis, v. 47, n. 3, p. 506–523, 2020.

LAMBERT, D. Zero-inflated poisson regression, with an application to defects in manufacturing. **Technometrics**, Taylor Francis, v. 34, n. 1, p. 1–14, 1992. Disponível em: <<https://www.tandfonline.com/doi/abs/10.1080/00401706.1992.10485228>>.

LARSEN, R. J.; MARX, M. L. **An introduction to mathematical statistics**. [S.l.]: Prentice Hall, 2005.

LEROUX, B.; DOUGLAS, J. B.; PUTERMAN, M. L. Empirical fitting of discrete distributions. **Biometrics**, [Wiley, International Biometric Society], v. 50, n. 2, p. 576–579, 1994. ISSN 0006341X, 15410420. Disponível em: <<http://www.jstor.org/stable/2533403>>.

OLIVEIRA, M.; EINBECK, J.; HIGUERAS, M.; AINSBURY, E.; PUIG, P.; ROTHKAMM, K. Zero-inflated regression models for radiation-induced chromosome aberration data : a comparative study. **Biometrical journal.**, Wiley-VCH, v. 58, n. 2, p. 259–279, 2016. Disponível em: <<http://dro.dur.ac.uk/17160/>>.

PINHEIRO, J. I. D.; CARVAJAL, S. S. R.; CUNHA, S. B. d.; GOMES, G. C. **PROBABILIDADE E ESTATÍSTICA: Quantificando a incerteza**. São Paulo - SP: Elsevier, 2012.

POISSON, S. D. **Recherches sur la probabilité des jugements en matière criminelle et en matière civile**. Paris: Bachelier, Imprimeur - Libraire, 1837.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2022. Disponível em: <<https://www.R-project.org/>>.

RIDOUT, M.; DEMETRIO, C.; HINLE, J. Models for count data with many zeros. international biometric conference. **Cape Town**, v. 13, p. 1–13, 01 1998.

SAHOO, P. **Probability and Mathematical Statistics**. 1e. ed. Louisville, USA: University of Louisville, 2015. 416 p.

SERGIO, V. **Utilização das Distribuições Inflacionadas de Zeros no Monitoramento da Qualidade do Leite**. 43 f. Monografia (graduação) — Universidade Federal de Juiz de Fora, Minas Gerais, 2012.

TLHALOGANYANG, B.; THAGA, K. Are zero inflated distributions compulsory in the presence of zero inflation. 10 2020.

YANG, S.; HARLOW, L.; PUGGIONI, G.; REDDING, C. A comparison of different methods of zero-inflated data analysis and an application in health surveys. **Journal of Modern Applied Statistical Methods**, v. 16, p. 518–543, 05 2017.

Apêndice A

Códigos

Apresentamos aqui os códigos utilizados para realizar as simulações do capítulo 4 desta dissertação.

```
#FUNCAO DE PROBABILIDADE DA DISTRIBUICAO POISSON INFLACIONADA (ZIPM.R)
```

```
library(Formula)
library(LaplacesDemon)
library(ggplot2)
library(extraDistr)
library(gamlss)

dzip<- function(y, delta, mu){
  y<- as.vector(y)
  n<- length(y) # sample size
  n0<- sum(y==0) # number of zeros

  n0 <- sum(y == 0)
  n <- length(y)
  prop <- n0/n
  n1 <- n - n0
  y1 <- y[y > 0] # valores de y>0

  c1 = 1-delta
  c2 = mu/c1

  f = ifelse((y==0),delta + c1*exp(-c2), c1*dpois(y1,c2))
  sum(f)
}
```

```

#MV

ZIPfit <- function(y){
  n <- length(y) # sample size
  n0 <- sum(y == 0) # number of zero

  loglik<-function(pa){
    delta <-pa[1]
    mu <-pa[2]

    #delta = ifelse(n0==0,0,delta)

    n0 <- sum(y == 0)
    n <- length(y)
    prop <- n0/n
    n1 <- n - n0
    y1 <- y[y > 0]
    c1 = (1-delta)
    c2 = mu/c1

    l =
    ifelse((y==0), log(delta + c1*exp(-c2)), log(c1)
    + dpois(y1,c2, log=TRUE))

    sum(l)
  }

  escore <- function(pa){
    delta <-pa[1]
    mu <- pa[2]

    n0 <- sum(y == 0)
    n <- length(y)
    prop <- n0/n
    n1 <- n - n0
    y1 <- y[y > 0]

    c1 = 1-delta
    c2 = mu/c1
    a <- mu*exp(-c2)
    b <- exp(-c2)
  }
}

```

```

#DERIVADA EM RELACAO A DELTA
  Up1=ifelse(y==0,(1-a/(1-delta)-b)/(delta+(1-delta)*b),0)
  Up2=ifelse(y>0,(y1-delta*y1-mu+delta-1)/(1-delta)^2,0)
  Up = sum(Up1,Up2)
#DERIVADA EM RELACAO A MU
  Umu1=ifelse(y==0,-b/(delta+(1-delta)*b),0)
  Umu2=ifelse(y>0,(y1-delta*y1-mu)/(mu*(1-delta)),0)
  Umu=sum(Umu1,Umu2)
  c(Up,Umu)
}

y0<- y
if(any(y==0)) y0<-y0[-which(y==0)]

#Initial values

n<- length(y) # sample size
n0<- sum(y==0)
lambda_hat = mean(y0)
mu.hat = mean(y)
delta_hat <- 1 - mu.hat/lambda_hat
delta0 <- ((n0/n) - exp(-lambda_hat))/(1-exp(-lambda_hat))
mu_0 = (1-delta0)*lambda_hat
ini <- c(delta0,mu_0)

#OTIMIZACAO

opt <-optim(ini,loglik,escore, method = "BFGS",
control = list(fnscale=-1), hessian = TRUE)
if (any(c(mu, delta) < 0)) return(NA)

if (opt$convergence !=0) print("FUNCTION DID NOT CONVERGE!")

#Return

z <- c()
z$mv <- opt$pa
z$conv <- opt$convergence

```

```

z$loglik <- opt$value
z$vcov <- opt$hessian
return(z)
}

#MATRIZ DA INFORMACAO DE FISHER (Function.Fisher.R)

fisher <- function(y){
  n <- length(y) # sample size
  n0 <- sum(y == 0) # nu
  delta =fit$mv[1]
  mu =fit$mv[2]

  #####Fisher#####

  k = exp(-(mu.hat/(1-delta.hat)))
  k1= delta.hat + (1-delta.hat)* k
  k11= (delta.hat + (1-delta.hat)* k)^2
  k2 = (1-delta.hat)^3
  k3 = (1-delta.hat)^2
  k4 = 1 - (mu.hat*k)/(1-delta.hat)- k
  k5=(k4)^2

  IF_delta_delta2 =
  ifelse(y>0, (mu.hat+delta.hat*mu.hat-delta.hat+1)/(k2),
  -(k*(mu.hat)^2)/((k2)*(delta.hat + (1-delta.hat)*k)) + (k5)/(k11))
  s.IF_delta_delta2 = sum(IF_delta_delta2)

  IF_mu_mu2 =
  ifelse(y>0, 1/mu.hat,-(delta.hat*k)/((1-delta.hat)*(k1)^2))
  s.IF_mu_mu2 = sum(IF_mu_mu2)

  IF_mu_delta = ifelse(y>0, 1/(k3), -k*(-k*(delta.hat)^2 + 2*delta.hat*k - k
  + delta.hat^2 - 2*delta.hat + delta.hat*mu.hat+ 1)/((k3)*(k1)^2))

  s.IF_mu_delta=sum(IF_mu_delta)
  s.IF_delta_mu= s.IF_mu_delta

  IF.v1 =
  (matrix(c(s.IF_delta_delta2,s.IF_delta_mu,
  s.IF_delta_mu,s.IF_mu_mu2),2,2))

```

```

inv.IF.v1 = solve(IF.v1)

ep.delta=sqrt(inv.IF.v1[1,1])
ep.mu = sqrt(inv.IF.v1 [2,2])

return(c(ep.delta, ep.mu))
}
#Teste de Vandenbroek (zero.test.R)

zero.test <- function(y) {

if(is.table(y)) { # expand to vector of values
  if(length(dim(y)) > 1) stop ("x must be a 1-way table")
  y <- rep(as.numeric(names(x)), unname(c(x)))
}
lambda <- mean(y)
p0_tilde <- exp(-lambda)
n0 <- sum(1*(!(y >0)))
n <- length(y)
numerator <- (n0 - n*p0_tilde)^2
denominator <- n*p0_tilde*(1-p0_tilde) - n*lambda*(p0_tilde^2)
stat <- numerator/denominator
pvalue <- pchisq(stat,df=1, ncp=0, lower.tail=FALSE)
result <- list(statistic=stat, df=1, prob=pvalue)
#cat(paste("Score test for zero inflation\n\n",
#          "\tChi-square =", round(stat,5), "\n",
#          #          "\tdf = 1\n",
#          #          "\tpvalue:", format.pval(pvalue), "\n"))
#invisible(result)
  return(pvalue)
}

#https://rdr.io/cran/vcdExtra/src/R/zero.test.R

#SCRIPT DA SIMULACAO DE MONTE CARLO

rm(list = ls(all=TRUE))
source("ZIPM.R")
source("Function.Fisher.R")
source("zero.test.R")

```

```

library(Formula)
library(LaplacesDemon)
library(ggplot2)
library(extraDistr)
library(gamlss)
library(vcdExtra)

set.seed(2023) #fixar uma semente
R= 20000

#CENARIOS
delta = 0.05
mu = 1
lambda <- mu/(1-delta)
pa = c(delta,mu)
var.Y=mu*((delta*mu-delta+1)/(1-delta))
var.Y
prob.zero = delta+(1-delta)*exp(-mu/(1-delta))
prob.zero

#VALORES TABELADOS DA DISTRIBUICAO NORMAL

vt5 <- qnorm(0.975)
vt1 <- qnorm(0.995)

Cobdelta <- rep(0,R)
TbSdelta <- rep(0,R)
TbIdelta <- rep(0,R)

Cobmu <- rep(0,R)
TbSmu<- rep(0,R)
TbImu<- rep(0,R)

var.y=rep(0,R)

results <- c()
results1 <- c()

contpois=0
contpois.NIP=0

```

```

pvalue = rep(0,R)
coef.mu = rep(0,R)
coef.delta=rep(0,R)
teste=0

index = rep(0,R)

vn = 50

# INICIO DO LOOP
for (i in 1:R)
{
  y <- rzip(n=vn,lambda = lambda,delta)

  fit = ZIPfit(y)

  n<- length(y) # sample size
  n0<- sum(y == 0) # number of zeros

  delta.hat = fit$mv[1]
  mu.hat = fit$mv[2]

  if ((zero.test(y)<=0.05)|| (fit$conv !=0)|| (delta.hat<=0)|| (mu.hat<=0))
  {
    index[i] = i
    contpois=contpois+1
    next
  }

  coef.delta[i] = delta.hat
  coef.mu[i]=mu.hat

  var.y[i]= mu.hat*((delta.hat*mu.hat-delta.hat+1)/(1-delta.hat))

  #####

  ep.delta=fisher(y)[1]
  ep.mu = fisher(y)[2]

```



```

#delta
coefdelta = fit$mv[1]
ICinfdelta = coefdelta - vt5*ep.delta
ICsupdelta = coefdelta + vt5*ep.delta

Cobdelta[i] = ifelse(((delta>ICinfdelta)&& (delta<ICsupdelta) ),1,0)
TbSdelta[i] = ifelse((delta>ICsupdelta),1,0)
TbIdelta[i] = ifelse((delta<ICinfdelta),1,0)

#COBERTURA mu
coefmu = fit$mv[2]
ICinfmua = coefmu - vt5*ep.mu
ICsupmu = coefmu + vt5*ep.mu
Cobmu[i] = ifelse(((mu>ICinfmua)&&(mu<ICsupmu)),1,0)
TbSmu[i] = ifelse((mu>ICsupmu),1,0)
TbImu[i] = ifelse((mu<ICinfmua),1,0)

} # fim do loop

# NAO POIZ
(contpois/R)*100
index[index != 0]

NR=10000
if ((R - (length(index[index != 0]))) < NR) {
  print("Voce precisa aumentar o valor de R")
}

coef.mu.f = coef.mu[-index]
length(coef.mu.f)
summary(coef.mu.f)

coef.delta.f=coef.delta[-index]
length(coef.delta.f)
summary(coef.delta.f)
#NAO CONVERGENCIA E NAO POIZ
contpois
prob.zero

```

```

##### Considerando as 10000 replicas
NR=10000

Cobmu.f= Cobmu[-index]
summary(Cobmu.f)

TC.mu.f = sum(Cobmu.f[1:NR])/NR
TC.mu.f=(TC.mu.f)*100
TC.mu.f

Cobdelta.f=Cobdelta[-index]
summary(Cobdelta.f)
TC.delta.f = sum(Cobdelta.f[1:NR])/NR
TC.delta.f=(TC.delta.f)*100
TC.delta.f

mean.mu.hat = mean(coef.mu.f[1:NR])
vies.mu.hat = mean.mu.hat - mu
VR.mu.hat= 100*(vies.mu.hat/mu)
var.mu.hat =mean(var(coef.mu.f[1:NR]))
eqm.mu.hat = var.mu.hat + ((vies.mu.hat )^2)

mean.mu.hat
vies.mu.hat
var.mu.hat
eqm.mu.hat
VR.mu.hat

Cobmu.f= Cobmu[-index]
summary(Cobmu.f)

TC.mu.f = sum(Cobmu.f[1:NR])/NR
TC.mu.f=(TC.mu.f)*100

TbSmu.f=TbSmu[-index]
TbSmu.f=na.omit(TbSmu.f)
tnc_S_mu.f = sum(TbSmu.f[1:NR])/NR
tnc_S_mu.f=(tnc_S_mu.f)*100

TbImu.f=TbImu[-index]

```

```
TbImu.f=na.omit(TbImu.f)
tnc_I_mu.f = sum(TbImu.f[1:NR])/NR
tnc_I_mu.f=(tnc_I_mu.f)*100
```

```
Cobdelta.f=Cobdelta[-index]
summary(Cobdelta.f)
Cobdelta.f=na.omit(Cobdelta.f)
tam.delta.f = length(Cobdelta.f)
TC.delta.f = sum(Cobdelta.f[1:NR])/NR
TC.delta.f=(TC.delta.f)*100
summary(Cobdelta.f)
```

```
TbSdelta.f=TbSdelta[-index]
TbSdelta.f=na.omit(TbSdelta.f)
tnc_S_delta.f = sum(TbSdelta.f[1:NR])/NR
tnc_S_delta.f=(tnc_S_delta.f)*100
```

```
TbIdelta.f=TbIdelta[-index]
TbIdelta.f=na.omit(TbIdelta.f)
tnc_I_delta.f = sum(TbIdelta.f[1:NR])/NR
tnc_I_delta.f= (tnc_I_delta.f)*100
```

```
Cobdelta.sem=(Cobdelta)
tam.delta.sem = length(Cobdelta.sem)
TC.delta.sem = sum(Cobdelta.sem)/(tam.delta.sem-contpois.NIP)
(TC.delta.sem)*100
```

```
var.y.est = mean(var.y)
var.y.est
summary(var.y.est )
var.y.est
```

```
Cobdelta.sem=na.omit(Cobdelta)
tam.delta.sem = length(Cobdelta.sem)
TC.delta.sem = sum(Cobdelta.sem)/(tam.delta.sem-contpois.NIP)
(TC.delta.sem)*100
```

```
Cobmu.sem=na.omit(Cobmu)
tam.mu.sem = length(Cobmu.sem)
```

```

TC.mu.sem = sum(Cobmu.sem)/(tam.mu.sem -contpois.NIP)
(TC.mu.sem)*100

mu
delta
n

#-----
#>>>>>>>!!!FIM DO LOOP!!!<<<<<<<<<<=====
#-----

# Taxa de balanceamento - 5%

tnc_S_delta <- (sum(TbSdelta)/R)*100
tnc_I_delta <- (sum(TbIdelta)/R)*100
tnc_S_mu <- (sum(TbSmu)/R)*100
tnc_I_mu <- (sum(TbImu)/R)*100

tnc <- c(tnc_S_delta.f, tnc_I_delta.f, tnc_S_mu.f, tnc_I_mu.f)

tdelta <- (sum(Cobdelta)/R)*100
tmu <- (sum(Cobmu)/R)*100

results1 = cbind(TC.delta.f, tnc_I_delta.f, tnc_S_delta.f)
results2 = cbind(TC.mu.f, tnc_I_mu.f, tnc_S_mu.f)

rownames(results1)=c("delta")
rownames(results2)=c("mu")
results=rbind(results1,results2)
colnames(results)=c("Taxa.Cob", "TaxaN.Cob.I", "TaxaN.Cob.S")

results

m.Y = mu
dp.Y = sqrt(var.Y)
cv.Y = dp.Y/m.Y

```

```

m.Y
dp.Y
cv.Y*100

#METODO DOS MOMENTOS

rm(list = ls(all=TRUE))
source("zero.test.R")

library(Formula)
library(LaplacesDemon)
library(ggplot2)
library(extraDistr)
library(gamlss)
set.seed(2023)

R = 25000
n= 50
pvalue= rep(0,R)
delta = 0.5
mu = 1
lambda=mu/(1-delta)
pa = c(delta,mu)
coef_results<-c()
contpois=0

pvalue = rep(0,R)
coef.mu = rep(0,R)
coef.delta=rep(0,R)
index = rep(0,R)
teste=0
nc=0
ni=0

for (i in 1:R){

  #Poisson inflated
  y = rzip(n,lambda,delta)

  mu.hat <- mean(y)
  delta.hat <- mean(y)/(1-(sum(y^2)/sum(y))) + 1

```

```

    if ((zero.test(y)<=0.05)|| (delta.hat<=0)|| (mu.hat<=0)) {
      index[i] = i
      contpois=contpois+1
      next
    }

    coef.delta[i] = delta.hat
    coef.mu[i]=mu.hat
  }# fim do loop

# NAO POIZ
(contpois/R)*100
index[index != 0]

NR=10000
if ((R - (length(index[index != 0]))) < NR) {

  print("Voce precisa aumentar o valor de R")
}

coef.mu.f = coef.mu[-index]
length(coef.mu.f)
summary(coef.mu.f)

coef.delta.f=coef.delta[-index]
length(coef.delta.f)
summary(coef.delta.f)
#NAO CONVERGENCIA E NAO POIZ

contpois

##### Considerando as 10000 replicas
NR=10000

mean.mu.hat = mean(coef.mu.f[1:NR])
vies.mu.hat = mean.mu.hat - mu
VR.mu.hat= 100*(vies.mu.hat/mu)
var.mu.hat =mean(var(coef.mu.f[1:NR]))
eqm.mu.hat = var.mu.hat + ((vies.mu.hat )^2)

```

```

mean.mu.hat
vies.mu.hat
var.mu.hat
eqm.mu.hat
VR.mu.hat

#MÁXIMA VEROSSIMILHANÇA DA DISTRIBUIÇÃO POISSON (Poisson_01.R)

library(Formula)
library(LaplacesDemon)
library(ggplot2)
library(extraDistr)
library(gamlss)

#MV
ZIPfit1 <- function(y){
  n <- length(y) # sample size
  n0 <- sum(y == 0) # number of zero

  loglik<-function(pa){
    lambda <-pa[1]
    fat = factorial(y)
    n = length(y)
    log.L = (-n*lambda + sum(y)*log(lambda)-sum(log(fat)))
  }

  escore <- function(pa){
    lambda <-pa[1]

    #delta = ifelse(n0==0,0,delta)
    -n + (1/lambda)*sum(y)
  }

  #Initial values
  n<- length(y) # sample size
  n0<- sum(y==0)
  lambda0 = mean(y)
  ini <- c(lambda0)

```

```

#Otimização

opt<-
optim(ini,loglik,escore,method = "BFGS",
control = list(fnscale=-1), hessian = TRUE)

if (opt$convergence !=0) print("FUNCTION DID NOT CONVERGE!")

#Return

z <- c()
z$mv <- opt$pa
z$conv <- opt$convergence
z$loglik <- opt$value
z$vcov <- opt$hessian
return(z)
}

#ERRO DE ESPECIFICACAO
rm(list = ls(all=TRUE))
source("Poisson_01.R")
source("zero.test.R")
source("ZIPM.R")

library(Formula)
library(LaplacesDemon)
library(ggplot2)
library(extraDistr)
library(gamlss)
library(vcdExtra)

set.seed(2023) #fixar uma semente
R= 20000
#CENARIO

delta = 0.05
mu = 1
lambda <- mu/(1-delta)
pa = c(delta,mu)
var.Y=mu*((delta*mu-delta+1)/(1-delta))

```



```

var.Y
prob.zero = delta+(1-delta)*exp(-mu/(1-delta))
prob.zero

#VALORES TABELADOS DA DISTRIBUICAO NORMAL

vt5 <- qnorm(0.975)
vt1 <- qnorm(0.995)

Coblambda <- rep(0,R)
TbSlambda <- rep(0,R)
TbIlambda <- rep(0,R)

var.y=rep(0,R)

results <- c()
results1 <- c()

contpois=0
contpois.NIP=0

pvalue = rep(0,R)
coef.lambda = rep(0,R)
teste=0

index = rep(0,R)

vn = 200

# inicio do loop
for (i in 1:R)
{
  y <- rzip(n=vn,lambda = lambda,delta)

  fit = ZIPfit(y)
  fit1 = ZIPfit1(y)

  n<- length(y) # sample size
  n0<- sum(y == 0) # number of zeros

```

```

#AQUI - \hat{\lambda}=mean(y)
lambda.hat = mean(y)
delta.hat = fit$mv[1]
mu.hat = fit$mv[2]

if ((zero.test(y)<=0.05)|| (fit$conv !=0)|| (delta.hat<=0)|| (mu.hat<=0))
{
  index[i] = i
  contpois=contpois+1
  next
}

coef.lambda[i]=lambda.hat

#####INFORMACAO DE FISHER#####
IF.v1 = n/((lambda.hat))
inv.IF.v1 = solve(IF.v1)

ep.lambda=sqrt(inv.IF.v1)

#COBERTURA LAMBDA
coeflambda = lambda.hat
ICinflambda = coeflambda - ep.lambda*vt5
ICsuplambda = coeflambda + ep.lambda*vt5

Coblambda[i] = ifelse(lambda>ICinflambda & lambda<ICsuplambda,1,0)
TbSlambda[i] = ifelse(lambda>ICsuplambda,1,0)
TbIlambda[i] = ifelse(lambda<ICinflambda,1,0)

}
# fim do loop

# NAO POIZ
(contpois/R)*100
index[index != 0]

NR=10000
if ((R - (length(index[index != 0]))) < NR) {

```

```

    print("Voc? precisa aumentar o valor de R")
}

coef.lambda.f = coef.lambda[-index]
length(coef.lambda.f)
summary(coef.lambda.f)

contpois
prob.zero

##### Considerando as 10000 replicas
NR=10000

Coblambda.f= Coblambda[-index]
summary(Coblambda.f)

TC.lambda.f = sum(Coblambda.f[1:NR])/NR
TC.lambda.f=(TC.lambda.f)*100
TC.lambda.f

mean.lambda.hat = mean(coef.lambda.f[1:NR])
vies.lambda.hat = mean.lambda.hat - lambda
VR.lambda.hat= 100*(vies.lambda.hat/lambda)
var.lambda.hat =mean(var(coef.lambda.f[1:NR]))
eqm.lambda.hat = var.lambda.hat + ((vies.lambda.hat )^2)

mean.lambda.hat
vies.lambda.hat
var.lambda.hat
eqm.lambda.hat
VR.lambda.hat

Coblambda.f= Coblambda[-index]
summary(Coblambda.f)

TC.lambda.f = sum(Coblambda.f[1:NR])/NR
TC.lambda.f=(TC.lambda.f)*100

TbSlambda.f=TbSlambda[-index]

```

```

TbSlambda.f=na.omit(TbSlambda.f)
tnc_S_lambda.f = sum(TbSlambda.f[1:NR])/NR
tnc_S_lambda.f=(tnc_S_lambda.f)*100

TbIlambda.f=TbIlambda[-index]
TbIlambda.f=na.omit(TbIlambda.f)
tnc_I_lambda.f = sum(TbIlambda.f[1:NR])/NR
tnc_I_lambda.f=(tnc_I_lambda.f)*100

var.y.est = mean(var.y)
var.y.est
summary(var.y.est )
var.y.est

Coblambda.sem=na.omit(Coblambda)
tam.lambda.sem = length(Coblambda.sem)
TC.lambda.sem = sum(Coblambda.sem)/(tam.lambda.sem -contpois.NIP)
(TC.lambda.sem)*100

lambda
delta
n

#-----
#>>>>>>>>>!!!FIM DO LOOP!!!<<<<<<<<<<<=====
#-----

# Taxa de balanceamento - 5%

tnc_S_lambda <- (sum(TbSlambda)/R)*100
tnc_I_lambda <- (sum(TbIlambda)/R)*100
tnc <- c(tnc_S_lambda.f, tnc_I_lambda.f)

#Taxas de cobertura - 95%

tmu <- (sum(Coblambda)/R)*100

#Taxa <- c( (sum(Cobdelta)/R)*100, (sum(Cobmu)/R)*100)

results2 = cbind(TC.lambda.f, tnc_I_lambda.f, tnc_S_lambda.f)

```

```
rownames(results2)=c("lambda")
results=rbind(results2)
colnames(results)=c("Taxa.Cob","TaxaN.Cob.I","TaxaN.Cob.S")
# taxa cobertura e nao cobertura
results
```