

---

Modelos para Dados de Contagem: um  
estudo sobre o número de ovos do mosquito  
*Aedes aegypti*

*Camila Macedo Lima Nagamine*

---

# **Modelos para Dados de Contagem: um estudo sobre o número de ovos do mosquito *Aedes aegypti***

**Camila Macedo Lima Nagamine**

Orientadora: Cecília Candolo  
Co-Orientadora: Maria Sílvia de Assis Moura

Dissertação apresentada ao Departamento de Pós-Graduação em Estatística da Universidade Federal de São Carlos - UFSCar, como parte dos requisitos para obtenção do título de Mestre em Estatística.

**UFSCar - São Carlos**  
**Setembro/2007**

**Ficha catalográfica elaborada pelo DePT da  
Biblioteca Comunitária da UFSCar**

N147md

Nagamine, Camila Macedo Lima.

Modelos para dados de Contagem : um estudo sobre o número de ovos do mosquito *Aedes aegypti* / Camila Macedo Lima Nagamine. -- São Carlos : UFSCar, 2007. 72 f.

Dissertação (Mestrado) -- Universidade Federal de São Carlos, 2007.

1. Dados de contagem. 2. Distribuição de Poisson. 3. Distribuição binomial negativa. 4. Modelos inflacionados de zeros. I. Título.

CDD: 519.536 (20ª)

# Dedicatória

---

---

---

A André, pelo apoio e incentivo em todas as etapas deste trabalho. Pela compreensão, carinho, por ser meu alicerce e porto seguro nas horas mais difíceis.

---

# Agradecimentos

---

---

À Deus, pela fé e coragem ao longo desse trabalho e principalmente por ter concebido a graça da realização de um sonho.

À minha orientadora, Profa. Dr<sup>a</sup> Cecília Candolo pela orientação e apoio.

À minha Co-orientadora, Profa. Dr<sup>a</sup> Maria Sílvia Assis Moura pela orientação, incentivo e sincera amizade que construímos no decorrer deste trabalho.

Aos meus pais, Severino e Lúcia pelo apoio e carinho.

Às minhas irmãs, Michela, Jamille e Priscilla pela força e carinho.

À minha amiga Milena, por me motivar sempre com dedicação, confiança e carinho em todos os momentos.

Às colegas, Camila Bassetto e Flávia Pedroso pela força e amizade.

À funcionária, Luíza Maria da Silva, pelo carinho e alegria ao longo dessa caminhada.

Ao Prof. Dr<sup>o</sup>. Francisco Chiaravalloti e Angelita Barbosa, da Faculdade de Medicina de São José do Rio Preto, pela ajuda e fornecimento dos dados.

Enfim, agradeço a todos familiares e amigos que torceram por mim e que também são responsáveis por essa conquista.

# Resumo

---

---

**N**este trabalho são considerados quatro tipos de modelos para dados de contagens: o Modelo Poisson, o Modelo Binomial Negativo e variações destes modelos para contagens inflacionadas de zeros (ZIP e ZINB).

Informações do número de ovos postos pelo mosquito *Aedes aegypti* na cidade de São José do Rio Preto, SP, foram analisadas neste trabalho através de modelos para dados de contagens. O número de ovos do mosquito *Aedes aegypti* é primeiramente estudado sem o uso de covariáveis e através de alguns testes, foi verificado que modelos inflacionados de zeros (ZIP e ZINB) são mais adequados para analisar os dados em questão.

O modelo ZIP mostrou-se inadequado, devido à superdispersão gerada pela variabilidade dos dados. Nesse caso, o modelo ZINB, mostrou-se mais flexível ao conjunto de dados como é mostrado graficamente e em seguida foi modelado para três períodos distintos, considerando a variável resposta número de ovos do mosquito *Aedes aegypti*.

**Palavras-chaves:** Dados de Contagem, Modelos Inflacionados de Zeros, Poisson, Binomial Negativo, *Aedes aegypti*, Excesso de zeros.

# Abstract

---

---

In this work we considered four types of models for count data: the Poisson model, the Negative Binomial Model and variations of these models for count inflated of zeros (ZIP and ZINB).

Information on the number of put eggs for the mosquito aedes in the city of São José of Rio Preto, SP, was analyzed through models for counting data. The number of eggs of the mosquito aedes is studied firstly without the covariable and through some tests, it was verified that models zero inflated (ZIP and ZINB) are more adapted to analyze the data under study.

The ZIP model was shown inadequate, due to overdispersion generated by the variability of the data. The ZINB model, was shown more flexible to the data as it is shown graphically.

# Sumário

---

---

<b>Dedicatória</b>	<b>i</b>
<b>Agradecimentos</b>	<b>ii</b>
<b>Resumo</b>	<b>iii</b>
<b>Abstract</b>	<b>iv</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Contextualização . . . . .	1
1.2 Objetivo . . . . .	2
1.3 Apresentação dos Capítulos . . . . .	2
<b>2 Revisão de Literatura</b>	<b>3</b>
2.1 A Distribuição Poisson . . . . .	3
2.2 A Distribuição Binomial Negativa . . . . .	4
2.3 Modelos Inflacionados de Zeros . . . . .	6
2.4 Testes de Hipóteses . . . . .	7
2.4.1 Teste da Razão de Verossimilhança ( $RV$ ) . . . . .	8
2.4.2 Teste de Wald . . . . .	8
2.4.3 Teste Escore . . . . .	8
2.5 Intervalos de Confiança . . . . .	10
2.6 Medidas de Qualidade de Ajuste do Modelo . . . . .	11
2.6.1 Deviance . . . . .	11
2.6.2 AIC . . . . .	12
2.6.3 BIC . . . . .	12



<b>3</b>	<b>Metodologia</b>	<b>13</b>
3.1	Modelo Poisson Inflacionado de Zeros (ZIP)	13
3.1.1	Estimador de Máxima Verossimilhança	14
3.2	Regressão ZIP	15
3.2.1	Estimação para o Modelo de Regressão ZIP	16
3.2.2	Matriz de Informação Observada	17
3.2.3	Teste Escore ZIP	18
3.3	Modelo Binomial Negativo Inflacionado de Zeros (ZINB)	20
3.3.1	Estimador de Máxima Verossimilhança	21
3.4	Regressão ZINB	23
3.4.1	Estimação para o modelo de Regressão ZINB	24
3.4.2	Matriz de Informação Observada	25
3.4.3	Teste Escore ZINB	29
3.5	Procedimento de Ajuste dos Modelos	30
<b>4</b>	<b>Aplicação</b>	<b>33</b>
4.1	A Dengue	33
4.2	O Aedes aegypti	34
4.3	A Ovitrapa	35
4.4	Dados	35
4.5	Área Trabalhada	36
4.6	Modelo sem Covariáveis	36
4.6.1	Estimativas dos Parâmetros	36
4.6.2	Intervalos de Confiança	38
4.6.3	Resultados do Teste de Wald	40
4.6.4	Resultado do Teste Escore ZIP	42
4.6.5	Resultado do Teste Escore ZINB	42
4.6.6	Análise Gráfica	43
4.7	Modelo com Covariáveis	57
4.7.1	Descrição dos Modelos	57
4.7.2	Resultados e Discussões	61
<b>5</b>	<b>Conclusões</b>	<b>64</b>
<b>6</b>	<b>Perspectivas Futuras</b>	<b>66</b>
<b>7</b>	<b>Referências Bibliográficas</b>	<b>67</b>

<b>A</b>	<b>Apêndice</b>	<b>70</b>
A.1	EMV para os modelos ZIP e ZINB . . . . .	70
A.2	Logaritmo da função de verossimilhança do modelo de regressão ZIP . . . . .	72
A.3	Logaritmo da função de verossimilhança do modelo de regressão ZINB . . . . .	72
<b>B</b>	<b>Apêndice</b>	<b>73</b>
B.1	Programa - EMV ZIP . . . . .	73
B.2	Programa - EMV ZINB . . . . .	74
B.3	Algoritmo IC Assintótico . . . . .	75

---

# Introdução

---

## 1.1 Contextualização

A distribuição Poisson é uma base padrão para o ajuste de dados de contagens, porém, quando a variância observada é muito maior do que as médias observadas, ocorre o fenômeno da superdispersão e, nesse caso, o Modelo Binomial Negativo pode ser mais apropriado, por possibilitar que se modele a variância de modo mais adequado. Para os casos em que existe uma incidência muito alta de zeros em relação aos valores esperados para essa distribuição, Modelos Inflacionados de Zeros são utilizados.

O excesso de zeros em dados de contagem é uma ocorrência de certa forma, comum em experimentos nas mais diversas áreas, que pode ser causada por uma combinação dos chamados zeros estruturais e zeros amostrais. No primeiro caso, os zeros estruturais ocorrem independentemente da distribuição discreta de probabilidade. No último caso, os zeros amostrais estão relacionados com a ocorrência de zeros segundo o modelo adotado.

O modelo Poisson Inflacionado de Zeros (ZIP), sem covariáveis, é discutido por vários autores, entre eles, Xie e Goh (2001), Cameron e Trivedi (1998) e Broek (1995), mas, é Lambert (1992) quem primeiro apresenta este modelo associado ao uso de covariáveis em uma aplicação sobre o número de defeitos em um processo industrial. Em seguida, outros autores como Ridout *et al.* (1998), Famoye e Singh (2006) e Cheung (2002) ajustam o modelo ZIP com covariáveis, a dados provenientes de contagem aplicado a diversas áreas de pesquisa.

O modelo Binomial Negativo Inflacionado de Zeros (ZINB) é também uma alternativa na modelagem de dados de contagens, visto que, em algumas situações, esse modelo pode ser mais

apropriado. Neste contexto, Ridout *et al.* (2001) propuseram um teste escore para testar o modelo ZIP contra uma alternativa de que o modelo correto é o ZINB. Yau *et al.* (2003) mostram um estudo de dados de contagens com super-dispersão e zeros extras, usando uma mistura do modelo de Regressão Binomial Negativo Inflacionado de Zeros.

## 1.2 Objetivo

O objetivo deste trabalho foi fazer um levantamento de modelos para dados de contagens com excessos de zeros, particularmente o modelo Poisson Inflacionado de Zeros e Binomial Negativo Inflacionado de Zeros.

Foi desenvolvida a parte analítica para obtenção de estimadores de máxima verossimilhança dos parâmetros dos modelos ZIP e ZINB, e de suas covariâncias, porque, na literatura estudada não foi encontrada uma descrição detalhada desta metodologia.

Em seguida foi estudada uma aplicação dos modelos acima citados para determinar variáveis que influenciam na postura de ovos do *Aedes aegypti*, isto é, para modelar a quantidade de ovos em função das variáveis ambientais: pluviosidade, temperatura e umidade. Os dados coletados fazem parte de um Projeto de Pesquisa do Departamento de Epidemiologia e Saúde Coletiva da Faculdade de Medicina de São José do Rio Preto.

## 1.3 Apresentação dos Capítulos

O Capítulo 2 traz uma revisão de literatura, envolvendo alguns tópicos necessários para o desenvolvimento do trabalho em questão. No Capítulo 3, a metodologia envolvendo modelos para dados de contagens com excessos de zeros é apresentada e a seguir, no Capítulo 4, são apresentados os resultados de estudos aplicados aos modelos ZIP e ZINB, com e sem covariáveis, relativos ao número de ovos postos pelo mosquito *Aedes aegypti* citado anteriormente. Por fim, são apresentadas a conclusão e algumas perspectivas futuras deste trabalho.

---

# Revisão de Literatura

---

Dados de contagem em geral, podem ser considerados como provenientes de uma distribuição Poisson, pois conta eventos em um intervalo de tempo e nesses casos, é comum fazer uma transformação da variância a fim de validar as pressuposições básicas para o ajuste do modelo a ser analisado. Nesse capítulo é visto uma revisão dos modelos para dados de contagem: Poisson, Binomial Negativo e Inflacionados de Zeros.

## 2.1 A Distribuição Poisson

O nome da distribuição Poisson está associado ao matemático francês S. Denis Poisson (1781-1840), que apresentou um livro onde trata da aplicação da teoria das probabilidades na justiça comum. A distribuição Poisson de parâmetro  $\lambda$  é uma distribuição discreta de probabilidade. A variável aleatória  $Y$  é o número de ocorrências de um evento em um intervalo de tempo especificado e o parâmetro  $\lambda$  corresponde ao número médio de eventos ocorrendo neste intervalo. O intervalo pode ser de tempo, distância, área, volume ou outra unidade pertinente.

Meyer (1984) e James (1981), dentre outros autores, mostram que a distribuição Poisson pode ser obtida através do Processo de Poisson, onde são feitas algumas hipóteses, chamadas *hipóteses de Poisson*, e uma equação diferencial é calculada, cuja solução, obtida por recorrência, gera a expressão (2.1.1).

Segundo Jonhson e Kotz (1969), uma variável aleatória  $Y$  tem distribuição Poisson com parâmetro  $\lambda$  se sua função de probabilidade é dada por:

$$f(y, \lambda) = Pr(Y = y) = \frac{e^{-\lambda} \lambda^y}{y!}, \quad y = 0, 1, \dots, \quad (2.1.1)$$

em que  $\lambda > 0$  é o único parâmetro definindo a distribuição.

A média e variância da distribuição Poisson podem ser obtidas por

$$E(Y) = \sum_{y=0}^{\infty} y \frac{e^{-\lambda} \lambda^y}{y!} = e^{-\lambda} \sum_{y=1}^{\infty} \frac{\lambda^y}{(y-1)!} = \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = \lambda, \quad (2.1.2)$$

e, de maneira análoga, pode-se calcular  $E(Y^2)$  e então obter a variância de  $Y$  através de

$$Var(Y) = E(Y^2) - [E(Y)]^2 = \lambda. \quad (2.1.3)$$

### • Características da distribuição Poisson

Serão apresentadas quatro características da distribuição Poisson que são importantes para entender modelos de regressão para contagens:

- 1) O parâmetro  $\lambda$  é a média da distribuição.
- 2) O parâmetro  $\lambda$  também é a variância da distribuição. Em dados reais, muitas variáveis de contagens têm uma variância maior do que a média, o que é chamado de superdispersão.
- 3) Quando  $\lambda$  aumenta, a probabilidade de contagens de zero diminui.
- 4) Quando  $\lambda$  aumenta, a distribuição Poisson aproxima-se de uma distribuição Normal.

## 2.2 A Distribuição Binomial Negativa

A Distribuição Binomial Negativa é também conhecida como distribuição de Pascal ou distribuição de Polya.

Considere um experimento estatístico com dois resultados possíveis: sucesso ou fracasso, em que sucesso ocorre com probabilidade  $p$  e fracasso ou falha ocorre com probabilidade  $q = 1 - p$ . Se o experimento é repetido indefinidamente e os ensaios são independentes, então a variável aleatória  $Y^*$  é o número de ensaios em que ocorre o  $k$ -ésimo sucesso, e tem uma distribuição Binomial Negativa com parâmetros  $k$  e  $p$ .

A distribuição de probabilidades de  $Y^*$  é dada por,

$$f(y^*, k, p) = \binom{y^* - 1}{k - 1} p^k (1 - p)^{y^* - k}, \quad y^* = k, k + 1, k + 2, \dots \quad (2.2.1)$$

A distribuição Binomial Negativa pode ser escrita de diversas maneiras dependendo da parametrização utilizada. Será utilizada nas seções seguintes a distribuição Binomial Negativa na notação de Nelder e Wedderburn (1972), onde é feita uma reparametrização na definição acima,

tal que  $p = \frac{k}{k + \lambda}$ ,  $0 < p \leq 1$ , o parâmetro  $k$  é o parâmetro de dispersão e é assumido que,  $k \geq 0$ . Tomando  $y = y^* - k$  temos a distribuição Binomial Negativa com parâmetros  $k$  e  $\lambda$  com função de densidade dada por

$$f(y, k, \lambda) = \frac{\Gamma(y + k)}{\Gamma(k)y!} \left(\frac{k}{k + \lambda}\right)^k \left(\frac{\lambda}{k + \lambda}\right)^y, \quad y = 0, 1, 2, \dots \quad (2.2.2)$$

Seja  $Y$  uma variável aleatória com distribuição Binomial Negativa com parâmetros  $k$  e  $\lambda$ ,  $y = 0, 1, 2, \dots$ . A média de  $Y$  é dada por

$$\begin{aligned} E(Y) &= \sum_{y=0}^{\infty} y \frac{\Gamma(y + k)}{\Gamma(k)y!} \left(\frac{k}{k + \lambda}\right)^k \frac{\lambda^y}{(k + \lambda)^y} = \left(\frac{k}{k + \lambda}\right)^k \frac{1}{\Gamma(k)} \sum_{y=0}^{\infty} y \frac{\Gamma(y + k)}{y!} \frac{\lambda^y}{(k + \lambda)^y} \\ &= \left(\frac{k}{k + \lambda}\right)^k \frac{1}{\Gamma(k)} \Gamma(1 + k) \lambda \left(\frac{k}{k + \lambda}\right)^{-k} k^{-1} = \lambda. \end{aligned} \quad (2.2.3)$$

De maneira análoga pode-se calcular a  $E(Y^2) = \frac{\lambda(k\lambda + k + \lambda)}{k}$  e, assim obter a variância que é dada por

$$Var(Y) = \frac{\lambda(k + \lambda)}{k} = \lambda + \frac{\lambda^2}{k}. \quad (2.2.4)$$

Nota-se que a variância da Binomial Negativa tem um termo adicional  $\frac{\lambda^2}{k}$ , comparativamente com a variância da Poisson, que, em muitos casos, ajuda a ajustar melhor um conjunto de dados onde existe superdispersão.

#### • Modelo Binomial Negativo como mistura de Poisson

Segundo Jonhson e Kotz (1969), suponha que têm-se uma mistura de distribuições Poisson, tal que o valor esperado,  $\lambda$ , das distribuições Poisson varia de acordo com uma distribuição Gamma com função de densidade de probabilidade dada por

$$P(\lambda) = [\beta^\alpha \Gamma(\alpha)]^{-1} \lambda^{\alpha-1} \exp\left[-\frac{\lambda}{\beta}\right], \quad (2.2.5)$$

em que,  $\lambda > 0$ ,  $\alpha > 0$  e  $\beta > 0$ .

De forma que,

$$\begin{aligned}
 Pr(X = r) &= [\beta^\alpha \Gamma(\alpha)]^{-1} \int_0^\infty \lambda^{\alpha-1} \exp\left[-\frac{\lambda}{\beta}\right] \left[\lambda^r \exp\left[-\frac{\lambda}{\beta}\right]\right] d\lambda \\
 &= [\beta^\alpha \Gamma(\alpha)]^{-1} \int_0^\infty \lambda^{\alpha+r-1} \exp[-\lambda(\beta^{-1} + 1)] d\lambda \\
 &= \binom{\alpha + r - 1}{\alpha - 1} \left(\frac{\beta}{\beta + 1}\right)^{-1} \left(\frac{1}{\beta + 1}\right)^\alpha.
 \end{aligned} \tag{2.2.6}$$

Então,  $X$  tem uma distribuição Binomial Negativa com parâmetros  $(\alpha, \beta)$ .

## 2.3 Modelos Inflacionados de Zeros

Os modelos Inflacionados de Zeros têm sido usados para a modelagem de dados nas mais diversas áreas, tais como Epidemiologia, Sociologia, Medicina, Odontologia, Agricultura, Econometria dentre outras.

Lambert (1992) foi a pioneira no uso de modelos de regressão para contagens com distribuição inflacionada de zeros. A autora propôs uma mistura finita das distribuições Bernoulli e Poisson para modelar o excesso de zeros para dados de contagem envolvendo defeitos de fabricação de placas de circuito impresso.

Um dos modelos mais utilizados em dados de contagens com excessos de zeros é o modelo Poisson Inflacionado de Zeros (ZIP). Neste contexto, Ridout *et al.* (1998) apresentam uma revisão de literatura e discutem uma metodologia geral para modelar dados de contagem inflacionados de zeros, com ênfase em aplicações na Agricultura e Horticultura, Broek (1995) apresenta um a estatística Escore para testar se o modelo ZIP se ajusta melhor a um conjunto de dados que uma distribuição Poisson, Xie e Goh (2003) fazem um breve resumo de aplicações envolvendo o modelo ZIP, autores como Jung *et al.* (2005), Khoshgoftaar e Szabo (2001) e Cheung (2002) também fazem uso do modelo ZIP nas mais diversas aplicações.

O modelo Binomial Negativo Inflacionado de Zeros (ZINB) também é uma boa alternativa para dados de contagens, visto que, nos casos de dados de contagens com muitos zeros, a superdispersão pode acarretar maiores problemas, e, esse modelo pode ser mais adequado do que o modelo ZIP. Na literatura, existem trabalhos recentes sobre o modelo ZINB. Dentre eles, Lewsey e Thomson (2004) mostram um estudo comparativo dos modelos ZINB e ZIP, e concluem que em um estudo relativo a dentes com cáries, o modelo ZINB resulta em um melhor ajuste do que o modelo ZIP; Yau *et al.* (2003) ajustam um modelo de regressão ZINB para analisar dados referentes à recuperação de pacientes que submeteram-se a uma cirurgia no fígado, Ridout *et al.* (2001) abordam neste contexto um Teste Escore do modelo ZIP contra uma alternativa ZINB.

Uma outra distribuição para dados de contagens com excesso de zeros, definida de maneira similar aos modelos ZIP e ZINB é a Poisson Generalizada Inflacionada de Zeros, que pode ser



encontrada em Gupta e Tripathi (1996). Essa distribuição é adequada quando o modelo ZIP não consegue explicar a superdispersão e é uma boa alternativa neste tipo de situação, assim como o modelo ZINB. Recentemente, Famoye e Singh (2006) utilizaram essa distribuição para explicar dados referentes a violência doméstica. Entretanto, neste trabalho não será abordada essa distribuição.

Jonhson e Kotz (1969) discutem uma maneira simples de modificar uma distribuição discreta para modelar os zeros extras, através de uma mistura finita de distribuições. Uma proporção de zeros,  $\omega$ ,  $0 \leq \omega < 1$ , é adicionado à uma fração da proporção de zeros da distribuição discreta original,  $f(0)$ , enquanto diminui em proporções restantes dos valores  $y$  diferentes de zero através de

$$\begin{cases} P(Y_i = 0) = \omega + (1 - \omega) f(0) \\ P(Y_i = y_i) = (1 - \omega) f(y_i), y_i > 0. \end{cases} \quad (2.3.1)$$

A restrição  $0 < \omega < 1$  pode ser abandonada desde que

$$\omega \geq -\frac{f(0)}{[1 - f(0)]},$$

com igualdade para truncamento à esquerda. Do ponto de vista prático,  $\omega < 0$  esta associado à situações onde o número de zeros é menor do que o esperado.

A função densidade (2.3.1) com distribuição Poisson é dada por

$$f(y, \lambda, \omega) = \begin{cases} \omega + (1 - \omega) e^{-\lambda}, & y = 0 \\ (1 - \omega) e^{-\lambda} \frac{\lambda^y}{y!}, & y > 0 \end{cases}. \quad (2.3.2)$$

Se  $Y$  tem distribuição Binomial Negativa sua função de densidade é dada por

$$f(y, \lambda, \omega) = \begin{cases} \omega + (1 - \omega) \left(\frac{k}{k+\lambda}\right)^k, & y = 0 \\ (1 - \omega) \frac{\Gamma(y+k)}{\Gamma(k)y!} \left(\frac{k}{k+\lambda}\right)^k \left(\frac{\lambda}{k+\lambda}\right)^y, & y = 1, 2, \dots, \end{cases} \quad (2.3.3)$$

sendo que  $k \geq 0$  é o parâmetro de dispersão e  $\lambda$  é a média da distribuição Binomial Negativa com parâmetros  $k$  e  $\lambda$ .

No Capítulo seguinte os modelos ZIP e ZINB serão vistos mais detalhadamente.

## 2.4 Testes de Hipóteses

Nesta seção apresentamos três estatísticas usadas para testar hipóteses de restrições lineares sobre os parâmetros estimados por máxima verossimilhança. Este texto foi baseado no trabalho de Paula (2004), supondo inicialmente a seguinte situação de hipóteses simples:  $H_0 : \beta = \beta_0$  contra

$H_1 : \beta \neq \beta_0$ , em que  $\beta$  é o parâmetro do modelo e  $\beta_0$  é um valor especificado para  $\beta$ , supostamente conhecido.

### 2.4.1 Teste da Razão de Verossimilhança ( $RV$ )

Testes da razão de verossimilhança seguem a forma usual de comparar os valores maximizados do logaritmo da função de verossimilhança sem restrição ( $l(\hat{\beta}, y)$ ) e sob a hipótese nula  $H_0$  ( $l(\beta_0, y)$ ). Se a diferença é grande, então  $H_0$  é rejeitada.

O teste da razão de verossimilhança, no caso de hipóteses simples, é usualmente definido por

$$RV = 2[l(\hat{\beta}, y) - l(\beta_0, y)]. \quad (2.4.1)$$

Para amostras grandes, rejeita-se  $H_0$  a um nível de significância  $\alpha$  se,

$$RV > \chi_{\nu, 1-\alpha}^2, \quad (2.4.2)$$

em que  $(\nu, 1 - \alpha)$  é o número de graus de liberdade e  $\nu$  é a dimensão de  $\beta$ .

### 2.4.2 Teste de Wald

O teste de Wald é baseado na distribuição normal assintótica de  $\hat{\beta}$  visto que,

$$\hat{\beta} \sim \mathcal{AN}[\beta, I^{-1}].$$

em que,  $\mathcal{AN}$  é a notação para indicar que a distribuição é assintoticamente normal e  $I$  é a matriz de informação de Fisher.

Assim, a estatística para esse teste é dada por

$$W = (\hat{\beta} - \beta_0)\hat{I}(\hat{\beta} - \beta_0), \quad (2.4.3)$$

em que,  $\hat{I}$  denota a matriz de informação de Fisher avaliada em  $\hat{\beta}$ .

Para grandes amostras, rejeita-se  $H_0 : \beta = \beta_0$  se,

$$W > \chi_{(\nu, 1-\alpha)}^2, \quad (2.4.4)$$

em que  $(\nu, 1 - \alpha)$  é o número de graus de liberdade e  $\nu$  é a dimensão de  $\beta$ .

### 2.4.3 Teste Escore

A teoria de Testes Escores, que, são geralmente mais simples que Testes da Razão de Verossimilhança porque a estatística de teste requer que os parâmetros de estimação sejam obtidos

somente sob a hipótese nula. Por esta razão, Testes Escore têm sido propostos frequentemente no contexto de Modelos Lineares Generalizados para testar hipóteses relativas à comparação de modelos que incorporam superdispersão, excesso de zeros, além de adequacidade da função de ligação e outros, Smyth (2003).

A função escore é definida por

$$U(\beta) = \frac{\partial l(\beta; y)}{\partial \beta}. \quad (2.4.5)$$

Dessa forma, o teste escore, também conhecido como teste de Rao, é definido para  $U(\hat{\beta}) = 0$  e, a estatística do teste escore é calculada da seguinte forma

$$S = U(\hat{\beta}_0)^T [I_0]^{-1} U(\hat{\beta}_0), \quad (2.4.6)$$

em que  $I_0$  é a matriz de informação de Fisher avaliada em  $\beta_0$ , ou seja, sob a hipótese nula.

Para grandes amostras, rejeita-se  $H_0$  a um nível de significância  $\alpha$  se,

$$S > \chi_{\nu, 1-\alpha}^2, \quad (2.4.7)$$

em que  $(\nu, 1 - \alpha)$  é o número de graus de liberdade e  $\nu$  é a dimensão de  $\beta$ .

Nesse contexto, Broek (1995) apresenta um Teste Escore para testar o modelo Poisson contra o modelo Poisson Inflacionado de Zeros (ZIP). Sob a hipótese nula  $H_0$ : a distribuição é Poisson a estatística escore, para o caso sem covariáveis é dada por

$$S_1 = \frac{(n_0 - np_0)^2}{np_0(1 - p_0) - n\bar{y}p_0^2}, \quad (2.4.8)$$

onde  $n$  é o número total de observações,  $n_0$  é o número de zeros nas observações,  $\bar{y}$  é a média das observações, e  $p_0 = e^{-\hat{\lambda}}$ , em que  $\hat{\lambda}$  é a estimativa do parâmetro Poisson sob a hipótese nula. Esta estatística terá uma distribuição assintótica  $\chi^2$  com um grau de liberdade sob a hipótese nula. Na Seção 3.1.2 será apresentado o desenvolvimento de um teste escore ZIP, baseado neste de Broek (1995), utilizando uma parametrização diferente para a função de densidade do modelo Poisson Inflacionado de Zeros.

Outro Teste Escore é proposto por Ridout *et al.* (2001), para testar o modelo Poisson Inflacionado de Zeros (ZIP) contra a alternativa Binomial Negativo Inflacionado de Zeros (ZINB) para o caso com covariáveis.

Considerando o modelo ZINB dado por

$$f(y, \omega, \alpha, k) = \begin{cases} (1 - \omega) + \omega \left( \frac{\alpha^{-1}}{\alpha^{-1} + \lambda} \right)^{\alpha^{-1}}, & y = 0 \\ \omega \frac{\Gamma(y + \alpha^{-1})}{\Gamma(\alpha^{-1})y!} \left( \frac{\alpha^{-1}}{\alpha^{-1} + \lambda} \right)^{\alpha^{-1}} \left( \frac{\lambda}{\alpha^{-1} + \lambda} \right)^y, & y = 1, 2, \dots, \end{cases} \quad (2.4.9)$$

em que  $\alpha \geq 0$  é um parâmetro de dispersão que é assumido não depender de covariáveis. Essa distribuição sob a hipótese nula  $H_0 : \lim_{\alpha \rightarrow 0}$  se reduz ao modelo Poisson Inflacionado de Zeros.

Sob  $H_0$ : a distribuição é ZIP, e a estatística do Teste Escore é dada por

$$S_2 = \frac{1}{2} \sum_{i=1}^n \hat{\lambda}_i^{c-1} \left\{ [(y_i - \hat{\lambda}_i)^2 - y_i] - I_{(y_i=0)} \hat{\lambda}_i^2 \hat{\omega}_i / \hat{p}_{0,i} \right\} \sqrt{\hat{J}^{\alpha\alpha}}, \quad (2.4.10)$$

onde  $i = 1, 2, \dots, n$ ,  $c = 0$ ,  $\hat{p}_{0,i} = Pr(Y_i = 0)$ , e  $I_{(y_i=0)}$  é uma variável indicadora que assume valor 1 quando  $y_i = 0$  e 0 caso contrário, no modelo ZIP, que é a distribuição sob a hipótese nula. O  $\hat{J}^{\alpha\alpha}$  é o elemento da inversa da matriz de informação que para  $\alpha$ , corresponde à estimativa de máxima verossimilhança sob a hipótese nula  $H_0 : \lim_{\alpha \rightarrow 0}$ .

Assintoticamente, sob a hipótese nula, a estatística  $S_2$  tem uma distribuição normal padrão. Na Seção 3.3.2 é apresentado o Teste Escore ZINB, considerando o caso sem covariáveis.

## 2.5 Intervalos de Confiança

Neste trabalho utilizaremos a metodologia da teoria assintótica para calcular intervalos de confiança. Os intervalos de confiança são baseados na distribuição amostral do estimador pontual.

Intervalos de confiança para os parâmetros do modelo Poisson Inflacionado de Zeros, por exemplo, podem ser obtidos através dos estimadores de máxima verossimilhança e suas variâncias estimadas. Considerando o vetor de parâmetros  $\beta = (\lambda, \omega)$ ,  $l(\beta, y)$  o logaritmo da função de verossimilhança e  $\hat{\beta}$  o estimador de máxima verossimilhança para  $\beta$ , então,  $\hat{\beta}$  tem uma distribuição assintótica normal, Cox e Hinkley (1974), dada por

$$\hat{\beta} = (\hat{\lambda}, \hat{\omega}) \sim \mathcal{AN}(\beta, I^{-1}),$$

onde  $I(\beta)$  é a matriz de informação de Fisher. Na prática, ao invés de se trabalhar com o inverso da informação de Fisher, trabalha-se com o inverso da matriz de informação observada, Welsh (1996), que pode ser escrita como

$$-J(\beta) = \begin{pmatrix} \frac{\partial^2 l(y_i, \lambda, \omega)}{\partial \lambda^2} & \frac{\partial^2 l(y_i, \lambda, \omega)}{\partial \lambda \partial \omega} \\ \frac{\partial^2 l(y_i, \lambda, \omega)}{\partial \omega \partial \lambda} & \frac{\partial^2 l(y_i, \lambda, \omega)}{\partial \omega^2} \end{pmatrix}. \quad (2.5.1)$$

A inversa da matriz de informação observada é definida como

$$J^{-1}(\beta) = \begin{pmatrix} \hat{V}ar(\hat{\lambda}) & \hat{C}ov(\hat{\lambda}, \hat{\omega}) \\ \hat{C}ov(\hat{\lambda}, \hat{\omega}) & \hat{V}ar(\hat{\omega}) \end{pmatrix}. \quad (2.5.2)$$

Assim, intervalos de confiança com coeficientes de confiança  $100(1 - \alpha)\%$  para os parâmetros  $\lambda$  e  $\omega$  podem ser obtidos diretamente a partir da normalidade assintótica dos estimadores de máxima verossimilhança e são dados por

$$IC(\lambda; 100(1 - \alpha)\%) = \hat{\lambda} \pm z_{\frac{\alpha}{2}} \sqrt{\hat{V}ar(\hat{\lambda})}, \quad (2.5.3)$$

$$IC(\omega; 100(1 - \alpha)\%) = \hat{\omega} \pm z_{\frac{\alpha}{2}} \sqrt{\hat{V}ar(\hat{\omega})}, \quad (2.5.4)$$

em que  $z_{\frac{\alpha}{2}}$  é o quantil  $\frac{\alpha}{2}$  da distribuição normal padrão e,  $\alpha$  é o nível de confiança.

## 2.6 Medidas de Qualidade de Ajuste do Modelo

Muitas medidas escalares tem sido desenvolvidas para resumir a qualidade global do ajuste em modelos de regressão para dados de contagens, ou variáveis dependentes categóricas. Uma medida escalar pode ser útil para comparar modelos e em última instância selecionar um modelo final.

As informações resultantes de medidas de ajuste são informações parciais, pois, devem ser avaliadas dentro do contexto da teoria que motiva a análise, pesquisa anterior, e os parâmetros calculados do modelo sendo considerado.

A seguir algumas medidas de ajuste são apresentadas, seguindo a teoria apresentada em Long e Freese (2001), e que, posteriormente, no Capítulo 4, serão utilizadas para analisar alguns modelos.

### 2.6.1 Deviance

A deviance compara um determinado modelo com um modelo que tem um parâmetro para cada observação, de forma que, o modelo reproduz os dados observados perfeitamente. A deviance denotada por Long e Freese (2001) é calculada através da expressão:

$$D = -2 \ln \mathcal{L}(M_k), \quad (2.6.1)$$

onde  $\mathcal{L}(M_k)$  é a função de verossimilhança do modelo  $M_k$  e o número de graus de liberdade são iguais a  $N - k$ ,  $k$  é o número de parâmetros no modelo e  $N$  o número total de observações. Modelo com menor valor de Deviance é considerado o modelo mais adequado.

## 2.6.2 AIC

O critério de informação de informação Akaike (1973) conhecido como AIC é definido como

$$AIC = -2 \ln \mathcal{L}(M_k) + 2k, \quad (2.6.2)$$

onde  $\mathcal{L}(M_k)$  é a função de verossimilhança maximizada do modelo e  $k$  é o número de parâmetros no modelo em questão. O modelo com menor AIC é considerado o melhor modelo ajustado.

Na notação de Long e Freese (2001), esta medida é calculada através da expressão:

$$AIC = \frac{-2 \ln \mathcal{L}(M_k) + 2k}{N}, \quad (2.6.3)$$

sendo  $N$  o número de observações.

## 2.6.3 BIC

O critério de informação Bayesiano (BIC) é dado pela expressão

$$BIC = -2 \ln \mathcal{L}(M_k) + k(\ln N), \quad (2.6.4)$$

e foi proposto por Schwarz (1978).

Long e Freese (2001) consideram o modelo  $M_k$  com Deviance igual a  $D(M_k)$  e a medida de ajuste BIC é escrita na forma

$$BIC_k = D(M_k) - gl_k \ln N, \quad (2.6.5)$$

onde  $gl_k$  são os graus de liberdade associados com a deviance,  $N$  é o número total de observações e  $k$  é o número de parâmetros no modelo. O modelo com menor valor BIC é considerado o melhor modelo ajustado.

## Metodologia

---

---

Neste Capítulo será apresentado o desenvolvimento para a obtenção dos estimadores dos modelos, considerando o método de máxima verossimilhança, os testes escores e a construção das matrizes de informação para os modelos ZIP e ZINB, visto que esta parte analítica não está explícita nos trabalhos vistos na literatura estudada.

### 3.1 Modelo Poisson Inflacionado de Zeros (ZIP)

A partir da definição vista na Seção 2.3 e, segundo Xie e Goh (2001), o modelo Poisson Inflacionado de Zeros considera que alguns zeros, os zeros estruturais, ocorrem com probabilidade  $1 - \omega$  e os zeros amostrais, com probabilidade  $\omega$  e denotam o modelo Poisson Inflacionado de Zeros como sendo

$$f(y, \omega, \lambda) = \begin{cases} (1 - \omega) + \omega e^{-\lambda}, & y = 0 \\ \omega e^{-\lambda} \frac{\lambda^y}{y!}, & y > 0 \end{cases} . \quad (3.1.1)$$

Sendo que,  $1 - \omega$  é a proporção de zeros que não é explicada pelo modelo em questão, e para o parâmetro  $\omega$  existe a restrição de que  $0 < \omega < 1$ .

A partir da densidade do modelo Poisson Inflacionado de Zeros, calcula-se a média para essa distribuição,

$$\begin{aligned} E(Y) &= \sum_{y=0}^{\infty} y f(y, \omega, \lambda) = \sum_{y=1}^{\infty} y \omega \frac{e^{-\lambda} \lambda^y}{y!} \\ &= \omega e^{-\lambda} \sum_{y=1}^{\infty} y \frac{\lambda^y}{y!} = \omega e^{-\lambda} \lambda e^{\lambda} = \omega \lambda. \end{aligned} \quad (3.1.2)$$

Para obter a variância, é necessário calcular a  $E(Y^2)$

$$\begin{aligned} E(Y^2) &= \sum_{y=0}^{\infty} y^2 f(y, \omega, \lambda) = \sum_{y=1}^{\infty} y^2 \omega e^{-\lambda} \frac{\lambda^y}{y!} \\ &= \omega e^{-\lambda} \sum_{y=1}^{\infty} y^2 \frac{\lambda^y}{y!} = \omega e^{-\lambda} \lambda e^{\lambda} (\lambda + 1) \\ &= \omega (\lambda^2 + \lambda) = \omega \lambda (\lambda + 1), \end{aligned} \quad (3.1.3)$$

e assim, obtém-se a variância, que é dada por

$$\begin{aligned} Var(Y) &= E(Y^2) - [E(Y)]^2 = [\omega (\lambda^2 + \lambda)] - [\lambda \omega]^2 \\ &= \omega \lambda^2 + \omega \lambda - \omega^2 \lambda^2 \\ &= \omega \lambda + \omega \lambda [\lambda - \omega \lambda] \\ &= \omega \lambda [1 + \lambda(1 - \omega)]. \end{aligned} \quad (3.1.4)$$

Se  $\omega = 0$ , a  $E(Y) = 0$  e  $Var(Y) = 0$ . Para  $\omega = 1$ ,  $E(Y) = \lambda$ , que é corresponde à média do modelo Poisson e  $Var(Y) = \lambda$ , que corresponde à variância do modelo Poisson. Ou seja, quando  $\omega = 1$ , o modelo ZIP se reduz ao modelo Poisson padrão.

### 3.1.1 Estimador de Máxima Verossimilhança

A obtenção dos estimadores de  $\omega$  e  $\lambda$ , para o modelo Poisson Inflacionado de Zeros, pode ser feita através do método de máxima verossimilhança.

Seja  $Y_1, Y_2, \dots, Y_n$  uma amostra de tamanho  $n$ , com  $n_0$  observações iguais a zero, e  $n_y$  observações diferentes de zero para  $y = 1, 2, \dots$ , tais que,  $n_1$  são as observações iguais a 1,  $n_2$  são as observações iguais a 2 e assim por diante, de forma que  $n = n_0 + \sum_{y=1}^{\infty} n_y$ .



O logaritmo da função de verossimilhança para o modelo ZIP é definido da seguinte forma

$$\begin{aligned}
 l(y_1, y_2, \dots, y_n, \omega, \lambda) &= \log[\mathcal{L}(y_1, y_2, \dots, y_n, \lambda, \omega)] \\
 &= \log \left[ \prod_{i=1}^n f(y_i, \lambda, \omega) \right] \\
 &= \log \left\{ [(1 - \omega) + \omega e^{-\lambda}]^{n_0} \prod_{y=1}^{\infty} \left( \omega e^{-\lambda} \frac{\lambda^y}{y!} \right)^{n_y} \right\} \\
 &= n_0 \log(1 - \omega + \omega e^{-\lambda}) + \sum_{y=1}^{\infty} n_y \log(\omega e^{-\lambda} \frac{\lambda^y}{y!}). \quad (3.1.5)
 \end{aligned}$$

As derivadas do logaritmo da função de verossimilhança com respeito a  $\omega$  e  $\lambda$ , são respectivamente,

$$\frac{\partial l(y, \omega, \lambda)}{\partial \omega} = \frac{-n_0(1 - e^{-\lambda})}{(1 - \omega + \omega e^{-\lambda})} + \sum_{y=1}^{\infty} \frac{n_y}{\omega}, \quad (3.1.6)$$

$$\frac{\partial l(y, \omega, \lambda)}{\partial \lambda} = -\frac{n_0 \omega e^{-\lambda}}{(1 - \omega + \omega e^{-\lambda})} + \sum_{y=1}^{\infty} n_y \left( \frac{y - \lambda}{\lambda} \right). \quad (3.1.7)$$

Os estimadores de máxima verossimilhança de  $\omega$  e  $\lambda$ , de acordo com o **Apêndice A1** são, respectivamente,

$$\hat{\omega} = \frac{1 - \frac{n_0}{n}}{1 - e^{\hat{\lambda}}}, \quad (3.1.8)$$

$$\hat{\lambda} = \frac{\bar{y}}{\hat{\omega}}, \quad (3.1.9)$$

onde  $\bar{y} = \sum_{i=1}^n \frac{y_i}{n}$ .

## 3.2 Regressão ZIP

O modelo de Regressão Poisson Inflacionado de Zeros é proveniente do modelo Poisson Inflacionado de Zeros de tal forma que o parâmetro  $\lambda_i$  pode ser escrito em função de covariáveis. Neste modelo é assumido que a variável resposta  $Y_i$ ,  $i = 1, 2, \dots, n$ , é uma contagem e tem distribuição Poisson Inflacionada de Zeros.

O modelo de regressão ZIP, modela separadamente o parâmetro  $\omega_i$  e o parâmetro  $\lambda_i$ , sendo que o parâmetro  $\omega_i$  pode ser modelado através de modelo logístico e assim, na expressão (3.1.1), foi feita uma reparametrização em relação ao parâmetro  $\omega_i$ , onde  $\omega_i = \frac{1}{1 + \theta_i}$  e, então a distribuição de probabilidade para a variável resposta do modelo de Regressão ZIP é dado por

$$f(y_i, \theta_i, \lambda_i) = \begin{cases} \frac{\theta_i}{1 + \theta_i} + \frac{1}{1 + \theta_i} e^{-\lambda_i} & y_i = 0 \\ \frac{1}{1 + \theta_i} \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} & y_i > 0, \end{cases} \quad (3.2.1)$$

onde  $-f(0) \leq \theta_i < \infty$ .

A função de verossimilhança para  $f(y_i, \theta_i, \lambda_i)$  é dada por

$$\mathcal{L}(y_i, \theta_i, \lambda_i) = \prod_{i=1}^n \left\{ \left[ \frac{1}{1 + \theta_i} (\theta_i + e^{-\lambda_i}) \right] I_{y_i=0} \times \left[ \frac{1}{1 + \theta_i} \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} \right] I_{y_i>0} \right\}, \quad (3.2.2)$$

e seu logaritmo é dado por

$$l(y_i, \theta_i, \lambda_i) = - \sum_{i=1}^n \log(1 + \theta_i) + \sum_{i:y_i=0} \log(\theta_i + e^{-\lambda_i}) - \sum_{i:y_i>0} \log(y_i!) + \sum_{i:y_i>0} [y_i \log(\lambda_i) - \lambda_i]. \quad (3.2.3)$$

Considere  $Y_i$ ,  $i = 1, 2, \dots, n$  uma variável resposta com distribuição ZIP dada por (3.2.1), onde, o interesse é associar a variável resposta  $Y_i$  às covariáveis, utilizando a função de ligação logarítmica para  $\lambda_i$  e  $\theta_i$ , representadas respectivamente por

$$\log(\lambda_i) = \mathbf{X}\beta \Rightarrow \lambda_i = \exp(\mathbf{X}\beta) \quad (3.2.4)$$

e

$$\log(\theta_i) = \mathbf{Z}\gamma \Rightarrow \theta_i = \exp(\mathbf{Z}\gamma), \quad (3.2.5)$$

onde,  $\beta$  e  $\gamma$  correspondem aos vetores dos coeficientes de regressão para as matrizes de covariáveis  $\mathbf{X}$  e  $\mathbf{Z}$ , respectivamente, sendo que  $\mathbf{X}$  são as covariáveis relacionadas ao modelo Poisson, e,  $\mathbf{Z}$  são as covariáveis relacionadas ao excesso de zeros explicado pelo parâmetro  $\theta$ .

Segundo Lambert (1992), as covariáveis que envolvem os zeros amostrais ( $\mathbf{X}$ ) podem ou não, ser as mesmas covariáveis que afetam a probabilidade da parte que envolve os zeros estruturais ( $\mathbf{Z}$ ).

### 3.2.1 Estimação para o Modelo de Regressão ZIP

Para o modelo de regressão ZIP, o logaritmo da função de verossimilhança é dado por

$$\begin{aligned} l(y_i, \beta, \gamma; \mathbf{X}, \mathbf{Z}) &= \sum_{i:y_i=0} \log \left[ e^{\mathbf{z}'_i \gamma} + \exp(-e^{\mathbf{x}'_i \beta}) \right] + \sum_{i:y_i>0} \left[ y_i \mathbf{x}'_i \beta - e^{\mathbf{x}'_i \beta} \right] \\ &\quad - \sum_{i:y_i>0} \log(y_i!) - \sum_{i=1}^n \log(1 + e^{\mathbf{z}'_i \gamma}), \end{aligned} \quad (3.2.6)$$

onde  $\mathbf{X}$  e  $\mathbf{Z}$  são as matrizes das covariáveis,  $\mathbf{x}'_i$  e  $\mathbf{z}'_i$ , os vetores das covariáveis. Os estimadores de máxima verossimilhança são obtidos maximizando a função dada por (3.2.6) em relação a

$\beta$  e  $\gamma$ , ou de forma equivalente, determinando o máximo absoluto do logaritmo da função de verossimilhança, uma vez que a função logarítmica é estritamente crescente.

As primeiras derivadas de (3.2.6) com respeito a  $\beta$  e  $\gamma$  são respectivamente,

$$\frac{\partial l(y_i, \beta, \gamma)}{\partial \beta_j} = - \sum_{i:y_i=0} \left[ \mathbf{x}'_{ij} \frac{e^{\mathbf{x}'_i \beta} \exp(-e^{\mathbf{x}'_i \beta})}{\exp(-e^{\mathbf{x}'_i \beta}) + e^{\mathbf{z}'_i \gamma}} \right] + \sum_{i:y_i>0} \left[ \mathbf{x}'_{ij} y_i - \mathbf{x}'_{ij} e^{\mathbf{x}'_i \beta} \right] \quad (3.2.7)$$

e

$$\frac{\partial l(y_i, \beta, \gamma)}{\partial \gamma_r} = \sum_{i:y_i=0} \left[ \mathbf{z}'_{ir} \frac{e^{\mathbf{z}'_i \gamma}}{\exp(-e^{\mathbf{x}'_i \beta}) + e^{\mathbf{z}'_i \gamma}} \right] - \sum_{i=1}^n \left[ \mathbf{z}'_{ir} \frac{e^{\mathbf{z}'_i \gamma}}{1 + e^{\mathbf{z}'_i \gamma}} \right]. \quad (3.2.8)$$

Sendo os componentes do vetores  $j = 1, \dots, p$  e  $r = 1, \dots, p^*$ . Os estimadores de máxima verossimilhança para  $\beta$  e  $\gamma$  podem ser obtidos através de um método iterativo tipo Newton-Raphson.

### 3.2.2 Matriz de Informação Observada

Um dos objetivos nos modelos de regressão é fazer inferências, testes de hipóteses e construir intervalos de confiança sobre o modelo de regressão. Aqui será considerada a matriz de informação observada que é dada por

$$J = - \sum_{i=1}^n \frac{\partial^2 l(y_i, \beta; X, Z)}{\partial \beta^2} \Big|_{\beta=\hat{\beta}}. \quad (3.2.9)$$

As segundas derivadas de (3.2.6) com respeito a  $\beta_j$  e  $\gamma_r$  são respectivamente<sup>1</sup>,

$$\begin{aligned} \frac{\partial^2 l(y_i, \beta, \gamma)}{\partial \beta_j^2} = & - \sum_{i:y_i>0} \left[ e^{\mathbf{x}'_i \beta} (\mathbf{x}'_{ij})^2 \right] \\ & - \sum_{i:y_i=0} \left[ (\mathbf{x}'_{ij})^2 \frac{(e^{\mathbf{x}'_i \beta - 2e^{\mathbf{x}'_i \beta}})}{(\exp(-e^{\mathbf{x}'_i \beta}) + e^{\mathbf{z}'_i \gamma})^2} + (\mathbf{x}'_{ij})^2 \frac{(e^{\mathbf{x}'_i \beta - e^{\mathbf{x}'_i \beta} + \mathbf{z}'_i \gamma} - e^{2\mathbf{x}'_i \beta - e^{\mathbf{x}'_i \beta} + \mathbf{z}'_i \gamma})}{(\exp(-e^{\mathbf{x}'_i \beta}) + e^{\mathbf{z}'_i \gamma})^2} \right], \end{aligned} \quad (3.2.10)$$

e

$$\begin{aligned} \frac{\partial^2 l(y_i, \beta, \gamma)}{\partial \gamma_r^2} = & \sum_{i:y_i=0} \left[ (\mathbf{z}'_{ir})^2 \frac{e^{\mathbf{z}'_i \gamma}}{\exp(-e^{\mathbf{x}'_i \beta}) + e^{\mathbf{z}'_i \gamma}} - (\mathbf{z}'_{ir})^2 \frac{(e^{\mathbf{z}'_i \gamma})^2}{(\exp(-e^{\mathbf{x}'_i \beta}) + e^{\mathbf{z}'_i \gamma})^2} \right] \\ & + \sum_{i=1}^n \left[ -(\mathbf{z}'_{ir})^2 \frac{e^{\mathbf{z}'_i \gamma}}{1 + e^{\mathbf{z}'_i \gamma}} + (\mathbf{z}'_{ir})^2 \frac{(e^{\mathbf{z}'_i \gamma})^2}{(1 + e^{\mathbf{z}'_i \gamma})^2} \right]. \end{aligned} \quad (3.2.11)$$

<sup>1</sup>As derivadas segundas e as derivadas cruzadas foram calculadas utilizando o software Maple

As derivadas cruzadas em relação a  $\beta$  e  $\gamma$  são simétricas, dessa forma

$$\frac{\partial^2 l(y_i, \beta, \gamma)}{\partial \gamma \partial \beta} = \frac{\partial^2 l(y_i, \beta, \gamma)}{\partial \beta \partial \gamma}$$

tal que,

$$\frac{\partial^2 l(y_i, \beta, \gamma)}{\partial \beta_j \partial \gamma_r} = \sum_{i: y_i=0} \left[ \mathbf{x}'_{ij} \mathbf{z}'_{ir} \frac{e^{\mathbf{x}'_i \beta} \exp(-e^{\mathbf{x}'_i \beta}) e^{\mathbf{z}'_i \gamma}}{(\exp(-e^{\mathbf{x}'_i \beta}) + e^{\mathbf{z}'_i \gamma})^2} \right]. \quad (3.2.12)$$

Isso resulta em uma matriz de informação observada dada por:

$$J = \begin{pmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{pmatrix}, \quad (3.2.13)$$

de forma que os elementos que compõe essa matriz são dados por,

$$J_{11} = \frac{\partial^2 l(y_i, \beta, \gamma)}{\partial \beta_j^2}, \quad J_{22} = \frac{\partial^2 l(y_i, \beta, \gamma)}{\partial \gamma_r^2}, \quad J_{12} = J_{21} = \frac{\partial^2 l(y_i, \beta, \gamma)}{\partial \beta_j \partial \gamma_r}.$$

A matriz inversa de variância-covariância  $J^{-1}$  é dada por

$$J^{-1} = \begin{pmatrix} \text{Var}(\hat{\beta}) & \text{cov}(\hat{\beta}, \hat{\gamma}) \\ \text{cov}(\hat{\beta}, \hat{\gamma}) & \text{Var}(\hat{\gamma}) \end{pmatrix}, \quad (3.2.14)$$

Assintoticamente, intervalos de confiança podem ser construídos para os parâmetros  $\beta$  e  $\gamma$ .

### 3.2.3 Teste Escore ZIP

Nesta Seção é apresentado um Teste Escore para testar situações onde o número de zeros é muito grande para uma distribuição Poisson ajustar bem os dados, como é o caso dos dados que serão analisados referente ao número de ovos do mosquito *Aedes aegypti*.

O objetivo do Teste Escore ZIP é verificar se há alguma evidência nos dados para apoiar a suposição de que a distribuição Poisson Inflacionada de Zeros é realmente mais apropriada do que a distribuição Poisson.

O Teste Escore ZIP foi baseado no trabalho de Broek (1995) mostrado na Seção 2.4.3, onde a parametrização utilizada nesta seção é diferente da utilizada por Broek.

Considere a função de densidade para o modelo ZIP dada por

$$f(y_i, \theta, \lambda) = \begin{cases} \frac{\theta}{1+\theta} + \frac{1}{1+\theta} e^{-\lambda} & y_i = 0 \\ \frac{1}{1+\theta} \frac{e^{-\lambda} \lambda^{y_i}}{y_i!} & y_i > 0. \end{cases} \quad (3.2.15)$$

Essa distribuição sob a hipótese nula  $H_0 : \theta = 0$  se reduz ao modelo Poisson padrão.

O logaritmo da função de verossimilhança para o modelo ZIP é dado por

$$l(y_i, \theta, \lambda) = - \sum_{i=1}^n \log(1 + \theta) + \sum_{i:y_i=0} \log(\theta + e^{-\lambda}) + \sum_{i:y_i>0} [y_i \log(\lambda) - \lambda - \log(y_i!)] . \quad (3.2.16)$$

As primeiras derivadas de  $l(\theta, \lambda, y_i)$  com respeito aos parâmetros  $\theta$  e  $\lambda$  são respectivamente,

$$\frac{\partial l(y_i, \theta, \lambda)}{\partial \theta} = -\frac{n}{1 + \theta} + \frac{n_0}{\theta + e^{-\lambda}}, \quad (3.2.17)$$

$$\frac{\partial l(y_i, \theta, \lambda)}{\partial \lambda} = -\frac{n_0 e^{-\lambda}}{\theta + e^{-\lambda}} + \frac{n\bar{y}}{\lambda} - n + n_0. \quad (3.2.18)$$

Sob a hipótese nula  $H_0 : \theta = 0$ , com  $\hat{\lambda}$  e  $\hat{\theta}$  os estimadores de máxima verossimilhança, temos que o vetor escore é dado por

$$U(\hat{\lambda}, 0) = \left[ 0, -n + \frac{n_0}{e^{-\hat{\lambda}}} \right]. \quad (3.2.19)$$

A matriz de informação observada, obtida através das derivadas de segunda ordem é dada por

$$J = \begin{pmatrix} \frac{n}{(1 + \hat{\theta})^2} - \frac{n_0}{(\hat{\theta} + e^{-\hat{\lambda}})^2} & \frac{n_0 e^{-\hat{\lambda}}}{(\hat{\theta} + e^{-\hat{\lambda}})^2} \\ \frac{n_0 e^{-\hat{\lambda}}}{(\hat{\theta} + e^{-\hat{\lambda}})^2} & \frac{n_0 e^{-\hat{\lambda}}}{\hat{\theta} + e^{-\hat{\lambda}}} - \frac{n_0 (e^{-\hat{\lambda}})^2}{(\hat{\theta} + e^{-\hat{\lambda}})^2} - \frac{n\bar{y}}{\hat{\lambda}^2} \end{pmatrix}, \quad (3.2.20)$$

Sob  $H_0 : \theta = 0$ , essa matriz se reduz a

$$J = \begin{pmatrix} -n + \frac{n_0}{(e^{-\hat{\lambda}})^2} & -\frac{n_0}{e^{\hat{\lambda}}} \\ -\frac{n_0}{e^{\hat{\lambda}}} & \frac{n\bar{y}}{\hat{\lambda}^2} \end{pmatrix}, \quad (3.2.21)$$

A estatística do Teste Escore para testar o modelo ZIP é escrita da seguinte forma

$$S_1 = U^T(\hat{\lambda}, 0) J^{-1}[2, 2] U(\hat{\lambda}, 0), \quad (3.2.22)$$

onde  $J^{-1}[2, 2]$  é o elemento superior à esquerda da inversa da matriz de informação calculada no estimador de máxima verossimilhança sob  $H_0$  e é dado por

$$J^{-1}[2, 2] = \frac{[n(e^{-\hat{\lambda}})^2 - n_0]\hat{\lambda}^2}{n^2\bar{y}(e^{-\hat{\lambda}})^2 - n\bar{y}n_0 + n_0^2\hat{\lambda}^2}. \quad (3.2.23)$$

Assim, a estatística escore para testar o modelo ZIP no caso sem covariáveis é obtida através da expressão dada por

$$S_1 = \frac{\left(-n + \frac{n_0}{e^{-\hat{\lambda}}}\right)^2 \left(n^2\bar{y}(e^{-\hat{\lambda}})^2 - n\bar{y}n_0 + n_0^2\hat{\lambda}^2\right)}{n\bar{y}(e^{-\hat{\lambda}})^2}, \quad (3.2.24)$$

onde  $\bar{y}$  é a média amostral,  $n_0$  o número de zeros nas observações e  $n$  o número total de observações.

Sob a hipótese nula  $H_0 : \theta = 0$  a estatística  $S_1$  terá uma distribuição assintótica  $\chi^2$  com um grau de liberdade.

### 3.3 Modelo Binomial Negativo Inflacionado de Zeros (ZINB)

Seguindo a notação usada por Yau *et al.* (2003), o modelo Binomial Negativo Inflacionado de Zeros pode ser escrito como sendo

$$f(y, \omega, \lambda, k) = \begin{cases} (1 - \omega) + \omega \left(\frac{k}{k + \lambda}\right)^k, & y = 0 \\ \omega \frac{\Gamma(y + k)}{\Gamma(k)y!} \left(\frac{k}{k + \lambda}\right)^k \left(\frac{\lambda}{k + \lambda}\right)^y, & y = 1, 2, \dots, \end{cases} \quad (3.3.1)$$

sendo que  $\lambda$  é a média da distribuição Binomial Negativa com parâmetros  $(k, \lambda)$ ,  $k$  é o parâmetro de dispersão que é assumido não depender de covariáveis,  $k \geq 0$  e  $1 - \omega$  é a proporção de zeros que não é explicada pelo modelo em questão, e para o parâmetro  $\omega$  existe a restrição de que  $0 < \omega < 1$ .

A partir da densidade do modelo Binomial Negativo Inflacionado de Zeros, calcula-se a média para essa distribuição,

$$\begin{aligned} E(Y) &= \sum_{y=0}^{\infty} yf(y, \omega, \lambda, k) = \sum_{y=1}^{\infty} y\omega \frac{\Gamma(y + k)}{\Gamma(k)y!} \left(\frac{k}{k + \lambda}\right)^k \left(\frac{\lambda}{k + \lambda}\right)^y \\ &= \omega \frac{1}{\Gamma(k)} \left(\frac{k}{k + \lambda}\right)^k \sum_{y=1}^{\infty} y \frac{\Gamma(y + k)}{y!} \frac{\lambda^y}{(k + \lambda)^y} \\ &= \omega \frac{1}{\Gamma(k)} \left(\frac{k}{k + \lambda}\right)^k \Gamma(1 + k) \lambda \left(\frac{k}{k + \lambda}\right)^{-k} k^{-1} \\ &= \omega \lambda. \end{aligned} \quad (3.3.2)$$

Para obter a variância, é necessário calcular a  $E(Y^2)$

$$\begin{aligned}
 E(Y^2) &= \sum_{y=0}^{\infty} y^2 f(y, \omega, \lambda, k) \\
 &= \omega \frac{1}{\Gamma(k)} \left( \frac{k}{k+\lambda} \right)^k \sum_{y=1}^{\infty} y^2 \frac{\Gamma(y+k)}{y!} \frac{\lambda^y}{(k+\lambda)^y} \\
 &= \frac{\omega (k + (1+k)\lambda) \lambda}{k} \\
 &= \frac{\omega \lambda (k + \lambda + \lambda k)}{k}
 \end{aligned} \tag{3.3.3}$$

e, assim, obtém-se a variância dada por

$$\begin{aligned}
 Var(Y) &= E(Y^2) - [E(Y)]^2 = \frac{\omega \lambda (k + \lambda + \lambda k)}{k} - [\omega \lambda]^2 \\
 &= \frac{\omega \lambda (k + \lambda + \lambda k)}{k} - \omega^2 \lambda^2 \\
 &= \omega \lambda \left[ \lambda(1 - \omega) + 1 + \frac{\lambda}{k} \right].
 \end{aligned} \tag{3.3.4}$$

Se o parâmetro  $\omega = 0$ , a  $E(Y) = 0$  e a  $Var(Y) = 0$ . Para  $\omega = 1$ ,  $E(Y) = \lambda$ , que corresponde à média do modelo Binomial Negativo e  $Var(Y) = \lambda + \frac{\lambda^2}{k}$ , que corresponde à variância do modelo Binomial Negativo. Ou seja, quando  $\omega = 1$ , o modelo ZINB se reduz ao modelo Binomial Negativo com parâmetros  $(k, \lambda)$ .

### 3.3.1 Estimador de Máxima Verossimilhança

O estimador de máxima verossimilhança para o parâmetro  $\omega$  é obtido através da derivada do logaritmo da função de verossimilhança com respeito a  $\omega$ . Seguindo a notação usada no modelo Poisson Inflacionado de Zeros, para uma amostra de tamanho  $n$  com  $n_0$  observações iguais a zeros e  $n_y$  observações diferentes de zero para  $y = 1, 2, \dots$ , o logaritmo da função de verossimilhança para o modelo Binomial Negativo Inflacionado de Zeros é dado por

$$l(y, \omega, \alpha, k) = n_0 \log \left( (1 - \omega) + \omega \left( \frac{k}{k+\lambda} \right)^k \right) + \sum_{y=1}^{\infty} n_y \log \left( \omega \frac{\Gamma(y+k)}{\Gamma(k)y!} \left( \frac{k}{k+\lambda} \right)^k \frac{\lambda^y}{(k+\lambda)^y} \right). \tag{3.3.5}$$

As derivadas do logaritmo da função de verossimilhança com respeito aos parâmetros  $\omega$ ,  $\lambda$  e  $k$  são respectivamente

$$\frac{\partial l(y, \omega, \lambda, k)}{\partial \omega} = -\frac{n_0 \left(1 - \left(\frac{\hat{k}}{\hat{k} + \hat{\lambda}}\right)^{\hat{k}}\right)}{1 - \hat{\omega} + \hat{\omega} \left(\frac{\hat{k}}{\hat{k} + \hat{\lambda}}\right)^{\hat{k}}} - \sum_{y=1}^{\infty} \frac{n_y}{\hat{\omega}}, \quad (3.3.6)$$

$$\frac{\partial l(y, \omega, \lambda, k)}{\partial \lambda} = -\frac{n_0 \hat{\omega} \hat{k} \left(\frac{\hat{k}}{\hat{k} + \hat{\lambda}}\right)^{\hat{k}}}{(\hat{k} + \hat{\lambda}) \left(1 - \hat{\omega} + \hat{\omega} \left(\frac{\hat{k}}{\hat{k} + \hat{\lambda}}\right)^{\hat{k}}\right)} - \sum_{y=1}^{\infty} \frac{n_y \hat{k}}{\hat{\lambda} + \hat{k}} + \sum_{y=1}^{\infty} \frac{n_y y \hat{k}}{(\hat{\lambda} + \hat{k}) \hat{\lambda}}, \quad (3.3.7)$$

$$\begin{aligned} \frac{\partial l(y, \omega, \lambda, k)}{\partial k} &= \frac{n_0 \hat{\omega} \left(\frac{\hat{k}}{\hat{k} + \hat{\lambda}}\right)^{\hat{k}} \left((\hat{k} + \hat{\lambda}) \ln \left(\frac{\hat{k}}{\hat{k} + \hat{\lambda}}\right) + \hat{\lambda}\right)}{(\hat{k} + \hat{\lambda}) \left(1 - \hat{\omega} + \hat{\omega} \left(\frac{\hat{k}}{\hat{k} + \hat{\lambda}}\right)^{\hat{k}}\right)} + \sum_{y=1}^{\infty} n_y \Psi(y + \hat{k}) - \sum_{y=1}^{\infty} \frac{n_y y}{\hat{k} + \hat{\lambda}} \\ &\quad - \sum_{y=1}^{\infty} n_y \Psi(\hat{k}) + \sum_{y=1}^{\infty} n_y \ln \left(\frac{\hat{k}}{\hat{k} + \hat{\lambda}}\right) + \sum_{y=1}^{\infty} \frac{n_y \hat{\lambda}}{\hat{k} + \hat{\lambda}} \end{aligned} \quad (3.3.8)$$

onde, a função  $\Psi(\cdot)$  é a função *Digamma*, dada por  $\Psi(\cdot) = \frac{\partial \ln[\Gamma(\cdot)]}{\partial(\cdot)}$ .



As estimativas dos parâmetros do modelo Binomial Negativo Inflacionado de Zeros podem ser obtidas iterativamente usando o seguinte sistema de equações

$$\left\{ \begin{array}{l} \hat{\omega} = \frac{1 - \frac{n_0}{n}}{1 - \left(\frac{\hat{k}}{\hat{k} + \hat{\lambda}}\right)^{\hat{k}}} \\ \frac{n_0 \hat{\omega} \hat{k} \left(\frac{\hat{k}}{\hat{k} + \hat{\lambda}}\right)^{\hat{k}}}{(\hat{k} + \hat{\lambda}) \left(1 - \hat{\omega} + \hat{\omega} \left(\frac{\hat{k}}{\hat{k} + \hat{\lambda}}\right)^{\hat{k}}\right)} = \frac{(n - n_0)\hat{k}}{\hat{\lambda} + \hat{k}} - \frac{n\bar{y} \hat{k}}{(\hat{\lambda} + \hat{k})\hat{\lambda}} \\ \frac{n_0 \hat{\omega} \left(\frac{\hat{k}}{\hat{k} + \hat{\lambda}}\right)^{\hat{k}} \left((\hat{k} + \hat{\lambda}) \ln \left(\frac{\hat{k}}{\hat{k} + \hat{\lambda}}\right) + \hat{\lambda}\right)}{(\hat{k} + \hat{\lambda}) \left(1 - \hat{\omega} + \hat{\omega} \left(\frac{\hat{k}}{\hat{k} + \hat{\lambda}}\right)^{\hat{k}}\right)} = \sum_{y=1}^{\infty} n_y \Psi(y + \hat{k}) - \frac{n\bar{y}}{\hat{k} + \hat{\lambda}} \\ -(n - n_0)\Psi(\hat{k}) + (n - n_0) \ln \left(\frac{\hat{k}}{\hat{k} + \hat{\lambda}}\right) + \frac{(n - n_0)\hat{\lambda}}{\hat{k} + \hat{\lambda}} \end{array} \right. \quad (3.3.9)$$

### 3.4 Regressão ZINB

Lewsey e Thomson (2004) sugerem substituir o modelo Poisson pelo modelo Binomial Negativa num ajuste de dados de experiência com cáries. Nesta seção o foco é o estudo do modelo de Regressão ZINB.

Considere então a expressão dada por (2.4.9), e a mesma reparametrização usada no modelo de Regressão ZIP em que  $\omega_i = \frac{1}{1 + \theta_i}$ , a função de densidade para o modelo de regressão ZINB é dado por

$$f(y_i, \theta_i, \lambda_i, k) = \begin{cases} \frac{\theta_i}{1 + \theta_i} + \frac{1}{1 + \theta_i} \left(\frac{k}{k + \lambda_i}\right)^k, & y_i = 0 \\ \frac{1}{1 + \theta_i} \frac{\Gamma(y_i + k)}{\Gamma(k) y_i!} \left(\frac{k}{k + \lambda_i}\right)^k \left(\frac{\lambda_i}{k + \lambda_i}\right)^{y_i} & y_i = 1, 2, \dots \end{cases} \quad (3.4.1)$$

em que  $-f(0) \leq \theta_i < \infty$ .

A função de verossimilhança para  $f(y_i; \theta_i, \lambda_i, k)$  é dada por

$$\mathcal{L}(y_i, \theta_i, \lambda_i, k) = \prod_{i=1}^n \left\{ \left[ \frac{1}{1 + \theta_i} \left( \theta_i + \left( \frac{k}{k + \lambda_i} \right)^k \right) \right] I_{y_i=0} \times \left[ \frac{1}{1 + \theta_i} \frac{\Gamma(y_i + k)}{\Gamma(k) y_i!} \left( \frac{k}{k + \lambda_i} \right)^k \left( \frac{\lambda_i}{k + \lambda_i} \right)^{y_i} \right] I_{y_i>0} \right\} \quad (3.4.2)$$

O logaritmo da função de verossimilhança para o modelo de regressão ZINB é dado por

$$\begin{aligned} l(y_i, \theta_i, \lambda_i, k) = & - \sum_{i=1}^n \log(1 + \theta_i) + \sum_{i:y_i=0} \log \left[ \theta_i + \left( \frac{k}{k + \lambda_i} \right)^k \right] \\ & + \sum_{i:y_i>0} \left[ k \log \left( \frac{k}{k + \lambda_i} \right) + y_i \log \left( \frac{\lambda_i}{k + \lambda_i} \right) \right] - \sum_{i:y_i>0} [\log \Gamma(y_i + 1) + \log \Gamma(k) - \log \Gamma(y_i + k)]. \end{aligned} \quad (3.4.3)$$

Considere  $Y_i$ ,  $i = 1, 2, \dots, n$  uma variável resposta com distribuição ZINB dada por (3.4.1), onde, o interesse é associar a variável resposta  $Y_i$  às covariáveis, utilizando a função de ligação logarítmica para  $\lambda_i$  e  $\theta_i$ , segundo Lambert (1992) podem ser representadas respectivamente por

$$\log(\lambda_i) = \mathbf{X}\beta \Rightarrow \lambda_i = \exp(\mathbf{X}\beta) \quad (3.4.4)$$

e

$$\log(\theta_i) = \mathbf{Z}\gamma \Rightarrow \theta_i = \exp(\mathbf{Z}\gamma), \quad (3.4.5)$$

onde,  $\beta$  e  $\gamma$  correspondem aos vetores dos coeficientes de regressão para as matrizes de covariáveis  $\mathbf{X}$  e  $\mathbf{Z}$ , respectivamente. Sendo que  $\mathbf{X}$  são as covariáveis relacionadas ao modelo Binomial Negativo, e,  $\mathbf{Z}$  são as covariáveis relacionadas ao excesso de zeros explicado pelo parâmetro  $\theta$ .

### 3.4.1 Estimação para o modelo de Regressão ZINB

Para o modelo de regressão ZINB, o logaritmo da função de verossimilhança é dado por

$$\begin{aligned} l(y_i, \beta, \gamma, k, \mathbf{X}, \mathbf{Z}) = & - \sum_{i=1}^n \log(1 + e^{z_i' \gamma}) + \sum_{i:y_i=0} \left[ k \log \left( \frac{k}{k + e^{x_i' \beta}} \right) + y_i \log \left( \frac{e^{x_i' \beta}}{k + e^{x_i' \beta}} \right) \right] \\ & + \sum_{i:y_i=0} \log \left[ e^{z_i' \gamma} + \left( \frac{k}{k + e^{x_i' \beta}} \right)^k \right] + \sum_{i:y_i>0} [-\log \Gamma(y_i + 1) - \log \Gamma(k) + \log \Gamma(y_i + k)], \end{aligned} \quad (3.4.6)$$

onde  $\mathbf{X}$  e  $\mathbf{Z}$  são as matrizes das covariáveis,  $x_i'$  e  $z_i'$  são os vetores das covariáveis.

Os estimadores, são obtidos maximizando a função dada por (3.4.6) em relação a  $\beta$  e  $\gamma$ , ou de forma equivalente, determinando o máximo absoluto do logaritmo da função de verossimilhança, uma vez que a função logarítmica é estritamente crescente.

As primeiras derivadas de (3.4.6) com respeito a  $\beta$  e  $\gamma$  são respectivamente<sup>2</sup>,

$$\frac{\partial l(y_i, \beta, \gamma, k)}{\partial \beta_j} = - \sum_{i: y_i=0} \left[ \left( \frac{k}{k + e^{\mathbf{x}'_i \beta}} \right)^k k \mathbf{x}'_{ij} e^{\mathbf{x}'_i \beta} (k + e^{\mathbf{x}'_i \beta})^{-1} \left( \left( \frac{k}{k + e^{\mathbf{x}'_i \beta}} \right)^k + e^{\mathbf{z}'_i \gamma} \right)^{-1} \right] \\ \sum_{i: y_i > 0} \left[ -\mathbf{x}'_{ij} \frac{k e^{\mathbf{x}'_i \beta}}{k + e^{\mathbf{x}'_i \beta}} + y_i \mathbf{x}'_{ij} \left( \frac{e^{\mathbf{x}'_i \beta}}{k + e^{\mathbf{x}'_i \beta}} - \mathbf{x}'_{ij} \frac{(e^{\mathbf{x}'_i \beta})^2}{(k + e^{\mathbf{x}'_i \beta})^2} \right) (k + e^{\mathbf{x}'_i \beta}) (e^{\mathbf{x}'_i \beta})^{-1} \right], \quad (3.4.7)$$

$$\frac{\partial l(y_i, \beta, \gamma, k)}{\partial \gamma_r} = \sum_{i: y_i=0} \left[ \mathbf{z}'_{ir} e^{\mathbf{z}'_i \gamma} \left( \left( \frac{k}{k + e^{\mathbf{x}'_i \beta}} \right)^k + e^{\mathbf{z}'_i \gamma} \right)^{-1} \right] - \sum_{i=1}^n \left[ \mathbf{z}'_{ir} \frac{e^{\mathbf{z}'_i \gamma}}{1 + e^{\mathbf{z}'_i \gamma}} \right] \quad (3.4.8)$$

e

$$\frac{\partial l(y_i, \beta, \gamma, k)}{\partial k} = \sum_{i: y_i=0} \left[ \frac{\left( \frac{k}{k + e^{\mathbf{x}'_i \beta}} \right)^k \left( \ln \left( \frac{k}{k + e^{\mathbf{x}'_i \beta}} \right) + \left( (k + e^{\mathbf{x}'_i \beta})^{-1} - \frac{k}{(k + e^{\mathbf{x}'_i \beta})^2} \right) (k + e^{\mathbf{x}'_i \beta}) \right)}{\left( \frac{k}{k + e^{\mathbf{x}'_i \beta}} \right)^k + e^{\mathbf{z}'_i \gamma}} \right] \\ + \sum_{i: y_i > 0} \left[ \ln \left( \frac{k}{k + e^{\mathbf{x}'_i \beta}} \right) + \left( (k + e^{\mathbf{x}'_i \beta})^{-1} - \frac{k}{(k + e^{\mathbf{x}'_i \beta})^2} \right) (k + e^{\mathbf{x}'_i \beta}) \right. \\ \left. - \frac{y_i}{k + e^{\mathbf{x}'_i \beta}} - \frac{\frac{d}{dk} \Gamma(k)}{\Gamma(k)} + \frac{D(\Gamma)(y_i + k)}{\Gamma(y_i + k)} \right], \quad (3.4.9)$$

onde  $D$ , representa a derivada da função Gamma ( $\Gamma$ ) em relação a  $y_i + k$ .

Através das primeiras derivadas, os estimadores de máxima verossimilhança para  $\beta$ ,  $\gamma$  e  $k$  podem ser obtidos através de um método iterativo tipo Newton-Raphson.

### 3.4.2 Matriz de Informação Observada

Com o objetivo de se obter os intervalos de confiança e fazer testes de hipóteses para o modelo de Regressão ZINB foram calculadas as segundas derivadas com respeito aos parâmetros  $\beta$ ,  $\gamma$  e  $k$  e também as derivadas cruzadas entre estes parâmetros para a construção da matriz de informação observada.

<sup>2</sup>As derivadas de primeira, segunda ordem e as derivadas cruzadas do modelo ZINB foram calculadas utilizando o software Maple.

As segundas derivadas de (3.4.6) com respeito a  $\beta$ ,  $\gamma$  e  $k$  são respectivamente dadas por,

$$\frac{\partial^2 l(y_i, \beta, \gamma, k)}{\partial \beta_j^2} = - \sum_{i: y_i > 0} \left[ (\mathbf{x}'_{ij})^2 \frac{k e^{\mathbf{x}'_i \beta} (y_i + k)}{(k + e^{\mathbf{x}'_i \beta})^2} \right] \quad (3.4.10)$$

$$- \sum_{i: y_i = 0} \left[ (\mathbf{x}'_{ij})^2 \frac{k^2 e^{\mathbf{x}'_i \beta} \left( - \left( \frac{k}{k + e^{\mathbf{x}'_i \beta}} \right)^k e^{\mathbf{x}'_i \beta + \mathbf{z}'_i \gamma} + \left( \frac{k}{k + e^{\mathbf{x}'_i \beta}} \right)^{2k} + \left( \frac{k}{k + e^{\mathbf{x}'_i \beta}} \right)^k e^{\mathbf{z}'_i \gamma} \right)}{(k + e^{\mathbf{x}'_i \beta})^2 \left( \left( \frac{k}{k + e^{\mathbf{x}'_i \beta}} \right)^k + e^{\mathbf{z}'_i \gamma} \right)^2} \right],$$

$$\frac{\partial^2 l(y_i, \beta, \gamma, k)}{\partial \gamma_r^2} = \sum_{i: y_i = 0} \left[ (\mathbf{z}'_{ir})^2 \frac{e^{\mathbf{z}'_i \gamma}}{\left( \frac{k}{k + e^{\mathbf{x}'_i \beta}} \right)^k + e^{\mathbf{z}'_i \gamma}} - (\mathbf{z}'_{ir})^2 \frac{(e^{\mathbf{z}'_i \gamma})^2}{\left( \left( \frac{k}{k + e^{\mathbf{x}'_i \beta}} \right)^k + e^{\mathbf{z}'_i \gamma} \right)^2} \right] \quad (3.4.11)$$

$$+ \sum_{i=1}^n \left[ -(\mathbf{z}'_{ir})^2 \frac{e^{\mathbf{z}'_i \gamma}}{1 + e^{\mathbf{z}'_i \gamma}} + (\mathbf{z}'_{ir})^2 \frac{(e^{\mathbf{z}'_i \gamma})^2}{(1 + e^{\mathbf{z}'_i \gamma})^2} \right]$$

e,

$$\begin{aligned}
& \frac{\partial^2 l(y_i, \beta, \gamma, k)}{\partial k^2} = \\
& \sum_{i: y_i=0} \left[ k^3 \left( \frac{k}{k + e^{\mathbf{x}'_i \beta}} \right)^k \left( \ln \left( \frac{k}{k + e^{\mathbf{x}'_i \beta}} \right) \right)^2 e^{\mathbf{z}'_i \gamma} + 2 k^2 \left( \frac{k}{k + e^{\mathbf{x}'_i \beta}} \right)^k \left( \ln \left( \frac{k}{k + e^{\mathbf{x}'_i \beta}} \right) \right)^2 e^{\mathbf{x}'_i \beta + \mathbf{z}'_i \gamma} \right. \\
& + 2 k^2 \left( \frac{k}{k + e^{\mathbf{x}'_i \beta}} \right)^k \ln \left( \frac{k}{k + e^{\mathbf{x}'_i \beta}} \right) e^{\mathbf{x}'_i \beta + \mathbf{z}'_i \gamma} + k \left( \frac{k}{k + e^{\mathbf{x}'_i \beta}} \right)^k \left( \ln \left( \frac{k}{k + e^{\mathbf{x}'_i \beta}} \right) \right)^2 e^{2 \mathbf{x}'_i \beta + \mathbf{z}'_i \gamma} \\
& + 2 k \left( \frac{k}{k + e^{\mathbf{x}'_i \beta}} \right)^k \ln \left( \frac{k}{k + e^{\mathbf{x}'_i \beta}} \right) e^{2 \mathbf{x}'_i \beta + \mathbf{z}'_i \gamma} + k \left( \frac{k}{k + e^{\mathbf{x}'_i \beta}} \right)^k e^{2 \mathbf{x}'_i \beta + \mathbf{z}'_i \gamma} \\
& \left. + e^{2 \mathbf{x}'_i \beta} \left( \frac{k}{k + e^{\mathbf{x}'_i \beta}} \right)^{2k} + \frac{e^{2 \mathbf{x}'_i \beta + \mathbf{z}'_i \gamma} \left( \frac{k}{k + e^{\mathbf{x}'_i \beta}} \right)^k}{k (k + e^{\mathbf{x}'_i \beta})^2 \left( \left( \frac{k}{k + e^{\mathbf{x}'_i \beta}} \right)^k + e^{\mathbf{z}'_i \gamma} \right)^2} \right] \\
& + \sum_{i: y_i > 0} \left[ \frac{\left( \frac{1}{(k + e^{\mathbf{x}'_i \beta})} - \frac{k}{(k + e^{\mathbf{x}'_i \beta})^2} \right) (k + e^{\mathbf{x}'_i \beta})}{k} + \left( -\frac{2}{(k + e^{\mathbf{x}'_i \beta})^2} + \frac{2k}{(k + e^{\mathbf{x}'_i \beta})^3} \right) (k + e^{\mathbf{x}'_i \beta}) \right. \\
& + \frac{1}{(k + e^{\mathbf{x}'_i \beta})} - \frac{k}{(k + e^{\mathbf{x}'_i \beta})^2} + \frac{y_i}{(k + e^{\mathbf{x}'_i \beta})^2} - \frac{\frac{d^2}{dk^2} \Gamma(k)}{\Gamma(k)} + \frac{\left( \frac{d}{dk} \Gamma(k) \right)^2}{(\Gamma(k))^2} \\
& \left. + \frac{(D^{(2)}(\Gamma)(y_i + k))}{\Gamma(y_i + k)} - \frac{(D(\Gamma)(y_i + k))^2}{(\Gamma(y_i + k))^2} \right]. \tag{3.4.12}
\end{aligned}$$

As derivadas cruzadas em relação a  $\beta_j$  e  $\gamma_r$  são simétricas, dessa forma

$$\frac{\partial^2 l(y_i, \beta, \gamma, k)}{\partial \beta_j \partial \gamma_r} = \frac{\partial^2 l(y_i, \beta, \gamma, k)}{\partial \gamma_r \partial \beta_j}$$

tal que,

$$\frac{\partial^2 l(y_i, \beta, \gamma, k)}{\partial \beta_j \partial \gamma_r} = \sum_{i: y_i=0} \left[ \mathbf{x}'_{ij} \mathbf{z}'_{ir} \frac{e^{\mathbf{z}'_i \gamma} \left( \frac{k}{k + e^{\mathbf{x}'_i \beta}} \right)^k k e^{\mathbf{x}'_i \beta}}{\left( \left( \frac{k}{k + e^{\mathbf{x}'_i \beta}} \right)^k + e^{\mathbf{z}'_i \gamma} \right)^2 (k + e^{\mathbf{x}'_i \beta})} \right]. \tag{3.4.13}$$

As derivadas cruzadas em relação a  $\beta_j$  e  $k$  são dadas por

$$\begin{aligned} \frac{\partial^2 l(y_i, \beta, \gamma, k)}{\partial \beta_j \partial k} &= \sum_{i:y_i > 0} \left[ \mathbf{x}'_{ij} \frac{e^{\mathbf{x}'_i \beta} (-e^{\mathbf{x}'_i \beta} + y_i)}{(k + e^{\mathbf{x}'_i \beta})^2} \right] \\ - \sum_{i:y_i = 0} \left[ \mathbf{x}'_{ij} \frac{\left( \left( \frac{k}{k + e^{\mathbf{x}'_i \beta}} \right)^k k^2 \ln \left( \frac{k}{k + e^{\mathbf{x}'_i \beta}} \right) e^{\mathbf{z}'_i \gamma} + \left( \frac{k}{k + e^{\mathbf{x}'_i \beta}} \right)^k k \ln \left( \frac{k}{k + e^{\mathbf{x}'_i \beta}} \right) e^{\mathbf{x}'_i \beta + \mathbf{z}'_i \gamma} \right) e^{\mathbf{x}'_i \beta}}{(k + e^{\mathbf{x}'_i \beta})^2 \left( \left( \frac{k}{k + e^{\mathbf{x}'_i \beta}} \right)^k + e^{\mathbf{z}'_i \gamma} \right)^2} \right. \\ &\quad \left. + \mathbf{x}'_{ij} \frac{\left( \left( \frac{k}{k + e^{\mathbf{x}'_i \beta}} \right)^k k e^{\mathbf{x}'_i \beta + \mathbf{z}'_i \gamma} + e^{\mathbf{x}'_i \beta} \left( \frac{k}{k + e^{\mathbf{x}'_i \beta}} \right)^{2k} + \left( \frac{k}{k + e^{\mathbf{x}'_i \beta}} \right)^k e^{\mathbf{x}'_i \beta + \mathbf{z}'_i \gamma} \right) e^{\mathbf{x}'_i \beta}}{(k + e^{\mathbf{x}'_i \beta})^2 \left( \left( \frac{k}{k + e^{\mathbf{x}'_i \beta}} \right)^k + e^{\mathbf{z}'_i \gamma} \right)^2} \right] \end{aligned} \quad (3.4.14)$$

e,

$$\begin{aligned} \frac{\partial^2 l(y_i, \beta, \gamma, k)}{\partial k \partial \beta_j} &= \sum_{i:y_i > 0} \left[ \mathbf{x}'_{ij} \frac{e^{2\mathbf{x}'_i \beta} (-1 + y_i e^{-\mathbf{x}'_i \beta})}{(k + e^{\mathbf{x}'_i \beta})^2} \right] \\ - \sum_{i:y_i = 0} \left[ \mathbf{x}'_{ij} \frac{\left( \left( \frac{k}{k + e^{\mathbf{x}'_i \beta}} \right)^k k^2 \ln \left( \frac{k}{k + e^{\mathbf{x}'_i \beta}} \right) e^{\mathbf{z}'_i \gamma} + \left( \frac{k}{k + e^{\mathbf{x}'_i \beta}} \right)^k k \ln \left( \frac{k}{k + e^{\mathbf{x}'_i \beta}} \right) e^{\mathbf{x}'_i \beta + \mathbf{z}'_i \gamma} \right) e^{\mathbf{x}'_i \beta}}{(k + e^{\mathbf{x}'_i \beta})^2 \left( \left( \frac{k}{k + e^{\mathbf{x}'_i \beta}} \right)^k + e^{\mathbf{z}'_i \gamma} \right)^2} \right. \\ &\quad \left. + \mathbf{x}'_{ij} \frac{\left( \left( \frac{k}{k + e^{\mathbf{x}'_i \beta}} \right)^k k e^{\mathbf{x}'_i \beta + \mathbf{z}'_i \gamma} + e^{\mathbf{x}'_i \beta} \left( \frac{k}{k + e^{\mathbf{x}'_i \beta}} \right)^{2k} + \left( \frac{k}{k + e^{\mathbf{x}'_i \beta}} \right)^k e^{\mathbf{x}'_i \beta + \mathbf{z}'_i \gamma} \right) e^{\mathbf{x}'_i \beta}}{(k + e^{\mathbf{x}'_i \beta})^2 \left( \left( \frac{k}{k + e^{\mathbf{x}'_i \beta}} \right)^k + e^{\mathbf{z}'_i \gamma} \right)^2} \right]. \end{aligned} \quad (3.4.15)$$

As derivadas cruzadas em relação a  $k$  e  $\gamma_r$  são simétricas, dessa forma

$$\frac{\partial^2 l(y_i, \beta, \gamma, k)}{\partial k \partial \gamma_r} = \frac{\partial^2 l(y_i, \beta, \gamma, k)}{\partial \gamma_r \partial k}$$

tal que,

$$\frac{\partial^2 l(y_i, \beta, \gamma, k)}{\partial k \partial \gamma_r} = - \sum_{i:y_i=0} \left[ \frac{z'_{ir} e^{z'_i \gamma} \left( k \ln \left( \frac{k}{k + e^{x'_i \beta}} \right) + \ln \left( \frac{k}{k + e^{x'_i \beta}} \right) e^{x'_i \beta} + e^{x'_i \beta} \right) \left( \frac{k}{k + e^{x'_i \beta}} \right)^k}{\left( \left( \frac{k}{k + e^{x'_i \beta}} \right)^k + e^{z'_i \gamma} \right)^2 (k + e^{x'_i \beta})} \right]. \quad (3.4.16)$$

Isso resulta em uma matriz de informação observada dada por:

$$- J = \begin{pmatrix} J_{11} & J_{12} & J_{13} \\ J_{21} & J_{22} & J_{23} \\ J_{31} & J_{32} & J_{33} \end{pmatrix}, \quad (3.4.17)$$

onde,

$$J_{11} = \frac{\partial^2 l(y_i, \beta, \gamma, k)}{\partial \beta_j^2}, \quad J_{22} = \frac{\partial^2 l(y_i, \beta, \gamma, k)}{\partial \gamma_r^2}, \quad J_{33} = \frac{\partial^2 l(y_i, \beta, \gamma, k)}{\partial k^2}, \quad J_{12} = J_{21} = \frac{\partial^2 l(y_i, \beta, \gamma, k)}{\partial \beta_j \partial \gamma_r},$$

$$J_{13} = \frac{\partial^2 l(y_i, \beta, \gamma, k)}{\partial \beta_j \partial k}, \quad J_{31} = \frac{\partial^2 l(y_i, \beta, \gamma, k)}{\partial k \partial \beta_j}, \quad J_{23} = J_{32} = \frac{\partial^2 l(y_i, \beta, \gamma, k)}{\partial k \partial \gamma_r},$$

sendo  $J$  uma matriz de dimensão  $(p + p^* + 1) \times (p + p^* + 1)$ ,  $j = 1, \dots, p$  e  $r = 1, \dots, p^*$ .

A matriz inversa de variância-covariância  $J^{-1}$  é dada por

$$J^{-1} = \begin{pmatrix} \hat{V}ar(\hat{\beta}) & \hat{c}ov(\hat{\beta}, \hat{\gamma}) & \hat{c}ov(\hat{\beta}, \hat{k}) \\ \hat{c}ov(\hat{\beta}, \hat{\gamma}) & \hat{V}ar(\hat{\gamma}) & \hat{c}ov(\hat{k}, \hat{\gamma}) \\ \hat{c}ov(\hat{\beta}, \hat{k}) & \hat{c}ov(\hat{k}, \hat{\gamma}) & \hat{V}ar(\hat{k}) \end{pmatrix}. \quad (3.4.18)$$

Assintoticamente, intervalos de confiança podem ser construídos para os parâmetros  $\beta$ ,  $k$  e  $\gamma$ .

### 3.4.3 Teste Escore ZINB

Nesta Seção é apresentado o Teste Escore para o modelo ZINB contra o modelo ZIP, com o objetivo de verificar a suposição de que o modelo ZINB é mais apropriado para ajustar o número de ovos do que o modelo ZIP, como foi citado na Seção 2.4.3, proposto por Ridout *et al.* (2001).

Fazendo uma reparametrização no modelo ZINB dado em (3.3.1), em que  $k = \alpha^{-1}$  e a proporção de zeros é  $\omega$ , temos

$$f(y, \omega, \alpha, \lambda) = \begin{cases} (1 - \omega) + \omega \left( \frac{\alpha^{-1}}{\alpha^{-1} + \lambda} \right)^{\alpha^{-1}}, & y = 0 \\ \omega \frac{\Gamma(y + \alpha^{-1})}{\Gamma(\alpha^{-1}) y!} \left( \frac{\alpha^{-1}}{\alpha^{-1} + \lambda} \right)^{\alpha^{-1}} \left( \frac{\lambda}{\alpha^{-1} + \lambda} \right)^y, & y = 1, 2, \dots, \end{cases} \quad (3.4.19)$$

em que  $\alpha \geq 0$  é um parâmetro de dispersão que é assumido não depender de covariáveis. Essa distribuição sob a hipótese nula  $H_0 : \lim_{\alpha \rightarrow 0} \alpha$  se reduz ao modelo Poisson Inflacionado de Zeros. O Teste Escore para testar a hipótese nula  $H_0 : \lim_{\alpha \rightarrow 0} \alpha$ , tem a vantagem de ajustar os dados para uma distribuição ZIP, que é a distribuição sob a hipótese nula, não havendo então, a necessidade do ajuste para o modelo ZINB.

Sendo assim, para o caso sem covariáveis, a estatística escore é dada por

$$S_2 = \frac{\sum_{i=1}^n ((y_i - \hat{\lambda})^2 - y_i) - n\hat{\lambda}^2\hat{\omega}}{\hat{\lambda} \sqrt{n(1 - \hat{\omega}) \left( 2 - \frac{\hat{\lambda}^2}{e^{\hat{\lambda}} - 1 - \hat{\lambda}} \right)}}. \quad (3.4.20)$$

em que  $\hat{\lambda}$  e  $\hat{\omega}$  são os estimadores dos parâmetros obtidos por máxima verossimilhança sob  $H_0$  e  $n$  o número total de observações. Assintoticamente, sob  $H_0$ , a estatística  $S_2$  tem uma distribuição Normal padrão.

### 3.5 Procedimento de Ajuste dos Modelos

Nesta Seção é apresentado o procedimento de ajuste segundo Long e Freese (2001) implementado no programa Stata e que será utilizado para a modelagem do número de ovos no Capítulo seguinte, visto que, com a parametrização utilizada nas seções anteriores, não foi possível ajustar os modelos ZIP e ZINB em outros *Softwares Estatísticos*.

Nesta parametrização, a função de verossimilhança para o modelo ZINB é dada por

$$\mathcal{L}(y_i, \omega_i, \lambda_i, \alpha) = \prod_{i=1}^n \left\{ \left[ \omega_i + (1 - \omega_i) \left( \frac{\alpha^{-1}}{\alpha^{-1} + \lambda_i} \right)^{\alpha^{-1}} \right]_{I_{y_i=0}} \times \left[ (1 - \omega_i) \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(\alpha^{-1})y_i!} \left( \frac{\alpha^{-1}}{\alpha^{-1} + \lambda_i} \right)^{\alpha^{-1}} \left( \frac{\lambda_i}{\alpha^{-1} + \lambda_i} \right)^{y_i} \right]_{I_{y_i>0}} \right\}, \quad (3.5.1)$$

que é particionada em duas funções de verossimilhança: uma em relação ao parâmetro  $\omega_i$  e outra em relação aos parâmetros  $\alpha$  e  $\lambda_i$ , ou seja,

$$Y_i = \begin{cases} 0, & \text{com probabilidade } \omega_i \\ BN(\alpha, \lambda), & \text{com probabilidade } (1 - \omega_i). \end{cases} \quad (3.5.2)$$



Dessa maneira, seja  $T_i$  uma variável indicadora tal que,

$$T_i = \begin{cases} 1, & \text{se } Y_i = 0 \text{ (zero estrutural)} \\ 0, & \text{se } Y_i \sim BN(\alpha, \lambda) \text{ (zero amostral)}. \end{cases} \quad (3.5.3)$$

O processo de ajuste é desenvolvido em três passos vistos a seguir:

### **Passo 1) Grupo da proporção de zeros**

O modelo de regressão ZINB relaciona o parâmetro  $\omega_i$  às covariáveis, através da função logito dada por

$$\text{logito}(\omega_i) = \mathbf{z}'_i \boldsymbol{\gamma}, \quad (3.5.4)$$

em que,  $\boldsymbol{\gamma}$  corresponde ao vetor dos coeficientes do modelo de regressão ZINB e  $\mathbf{z}'_i$  corresponde ao vetor de covariáveis relacionado ao parâmetro  $\omega$ . Sendo assim, a parte inflacionada é modelada por

$$\omega_i = \frac{\exp(\gamma_0 + \gamma_1 z_1 + \gamma_2 z_2 + \dots + \gamma_n z_n)}{1 + \exp(\gamma_0 + \gamma_1 z_1 + \gamma_2 z_2 + \dots + \gamma_n z_n)}. \quad (3.5.5)$$

### **Passo 2) Grupo da distribuição Binomial Negativa**

Para as observações que não são sempre nulas, a probabilidade de cada contagem (inclusive de zeros) é determinada por uma distribuição Binomial Negativa. Da mesma forma, o modelo de regressão ZINB relaciona o parâmetro  $\lambda_i$  às covariáveis usando a função logarítmica dada por

$$\log(\lambda_i) = \mathbf{x}'_i \boldsymbol{\beta}, \quad (3.5.6)$$

A probabilidade de uma contagem (inclusive de zeros) é determinada por

$$Pr(y_i | \mathbf{x}_i, T_i = 0) = \frac{\Gamma(y_i + \alpha^{-1})}{y_i! \Gamma(\alpha^{-1})} \left( \frac{\alpha^{-1}}{\alpha^{-1} + \lambda_i} \right)^{\alpha^{-1}} \left( \frac{\lambda_i}{\alpha^{-1} + \lambda_i} \right)^{y_i}, \quad (3.5.7)$$

em que,  $\lambda_i = \exp(\mathbf{x}'_i \boldsymbol{\beta})$ .

### **Passo 3) Mistura: Modelo Binomial Negativo Inflacionado de Zeros**

A proporção de zeros ( $\omega_i$ ) em cada grupo é definida por

$$\begin{aligned} Pr(T_i = 1) &= \omega_i \\ Pr(T_i = 0) &= 1 - \omega_i, \end{aligned} \quad (3.5.8)$$

e as probabilidades de zeros dentro de cada grupo são

$$\begin{aligned} Pr(y_i = 0|T_i = 1, \mathbf{x}_i, \mathbf{z}_i) &= 1 \\ Pr(y_i = 0|T_i = 0, \mathbf{x}_i, \mathbf{z}_i) &= \text{resultado do modelo regressão Binomial Negativo.} \end{aligned} \quad (3.5.9)$$

Então, a probabilidade de uma contagem ser zero é dada por

$$Pr(y_i = 0|\mathbf{x}_i, \mathbf{z}_i) = \omega_i + [(1 - \omega_i) \times Pr(y_i = 0|\mathbf{x}_i, T_i = 0)]. \quad (3.5.10)$$

Para outros resultados diferente de zeros, têm-se

$$Pr(y_i = m|\mathbf{x}_i, \mathbf{z}_i) = (1 - \omega_i) \times Pr(y_i = m|\mathbf{x}_i, T_i = 0). \quad (3.5.11)$$

---

## Aplicação

---

### 4.1 A Dengue

A dengue constitui um grave problema de saúde pública no Brasil e na maioria dos países tropicais, onde as condições climáticas favorecem a proliferação do *Aedes aegypti* o principal vetor da doença.

A dengue é transmitida ao homem por algumas espécies de mosquitos do gênero *Aedes*, sendo o *Aedes aegypti* a mais importante. Além do homem, não se conhece nenhum outro hospedeiro de importância significativa que atue como reservatório.

Pode-se contrair a dengue clássica e a hemorrágica. Geralmente, quando contaminada pela primeira vez, a pessoa contrai a dengue clássica, em uma segunda contaminação, existe um risco muito maior de se contrair a dengue hemorrágica, que é muito mais grave e pode levar à morte. Avaliando-se a série histórica de incidências de dengue no Brasil, verifica-se uma tendência ascendente com pico em 2002 e com um significativo aumento da forma hemorrágica da doença.

A dengue e especialmente a sua manifestação mais grave, a Dengue Hemorrágica são consideradas doenças infecciosas que estão em expansão no mundo e oferecem agravos significativos à saúde da população.

O mosquito *Aedes aegypti* é a principal espécie responsável pela transmissão da dengue. Somente as fêmeas se alimentam de sangue, as quais possuem hábitos alimentares diurnos e utilizam-se preferencialmente de depósitos artificiais de água limpa para colocar os seus ovos, que por sua vez, têm uma alta capacidade de resistir à dissecação, mantendo-se viáveis na ausência de água por até 450 dias, Tauil (2002).

Apesar da dengue ser muito pesquisada, não está disponível uma vacina preventiva eficaz. No momento, a única medida preventiva é o controle do vetor de transmissão.

## 4.2 O *Aedes aegypti*

O *Aedes aegypti* é um vetor oriundo do continente africano, trazido juntamente com os escravos. No Brasil o mosquito vetor foi erradicado pela primeira vez em 1958, e em seguida em 1967.

Em 1976, o *Aedes aegypti* reentrou no Brasil pelo litoral paranaense, porém a infestação definitiva do território nacional não seguiu a lógica da propagação pelas fronteiras imediatas do norte do País. Foi a partir dos portos de Salvador e do Rio de Janeiro, e não por terra, através de fronteiras com regiões infestadas.

Com o relaxamento da vigilância entomológica ocorrido no final da década de setenta e início dos anos oitenta, o vetor foi reintroduzido instalando-se definitivamente no País, Chiaravalloti (2004).

O *Aedes aegypti* é um mosquito com hábitos domiciliares que se prolifera facilmente em criadouros, ou seja, recipientes com água limpa, pouca matéria orgânica e expostos em locais sombreados, podendo manter dessa forma a densidade populacional. Essas condições são favorecidas pela presença de chuva. É na estação chuvosa que a população deste mosquito atinge níveis elevados e ocorrem epidemias de dengue. Essa associação do mosquito com o ser humano apresenta grande importância, pois a fêmea infectada, ao realizar a hematofagia em vários hospedeiros dissemina a doença, Osanai et al (1983).

O *Aedes aegypti* é uma espécie domiciliada, cuja convivência com o homem é favorecida pela utilização de recipientes artificiais ao desenvolvimento de suas formas lavriárias. A estratégia de controle desses recipientes é a principal arma de controle e continuação da doença.

O mosquito tem mostrado uma grande capacidade de adaptação a diferentes situações ambientais. Adultos já foram encontrados em altitudes elevadas e larvas em água poluída. Assim, a dengue apresenta um ciclo epidemiológico urbano e seus principais elementos são o homem - o hospedeiro do vírus e, o *Aedes aegypti*- o vetor de transmissão.

Como medida de vigilância e controle para o *Aedes aegypti* têm-se indicado as armadilhas de oviposição (ovitrampa). Os dados coletados através destas armadilhas são bem sucedidos quando usados para monitorar o impacto de vários tipos de medidas de controle envolvendo a redução da fonte de mosquitos, uma vez que os dados provenientes destas armadilhas são esperados para observar a tendência na abundância de adultos.

### 4.3 A Ovitampa

A ovitampa é uma armadilha constituída por um pote preto fosco, com abertura de 5 *cm* de diâmetro por 12 *cm* de profundidade, sem tampa, com uma palheta de madeira compensada, de 2 *cm* × 12,5 *cm*, com uma face rugosa voltada para cima para facilitar a aderência do ovo colocado pela fêmea do mosquito *Aedes aegypti*. Essa palheta é presa verticalmente por um clip, no interior da armadilha. No pote são colocadas 200 *ml* de água de torneira ou uma infusão de capim. As fêmeas são atraídas pela cor preta (por isso elas colocam seus ovos em vasos de planta) e fazem a oviposição na palheta.

Essa palheta com os ovos permite monitorar o *Aedes* porque indica a presença e o nível de infestação do mosquito numa determinada área. As paletas das ovitampas permanecem expostas por uma semana. As palhetas retiradas de cada armadilha são acondicionadas individualmente em sacos plásticos, encaminhadas ao laboratório acomodadas em uma placa de isopor e colocadas para secar em temperatura ambiente. Após a secagem, contam-se os ovos presentes nas palhetas com o auxílio de uma lupa.

O uso da ovitampa, como método de vigilância entomológica é considerado mais econômico e operacionalmente viável que os levantamentos de índices de infestação larvária. Estas armadilhas podem produzir melhores medidas de risco por estarem mais próximas das fêmeas adultas, e detectar precocemente novas infestações, o que as diferem das outras estratégias usadas hoje.

Verificou-se que o melhor local para a colocação de ovitampas, tanto em termos de positividade como o número de ovos coletados, é no peridomicílio não coberto, que o número médio de ovos está altamente correlacionado com a positividade, e é de 99,6% a probabilidade de uma palheta ser positiva para pelo menos um ovo de *Aedes aegypti*, Barbosa (2006).

### 4.4 Dados

Os dados analisados neste trabalho foram gentilmente cedidos pelo professor Dr<sup>o</sup> Francisco Chiaravalotti Neto, da Faculdade de Medicina de São José do Rio Preto, e estão descritos em detalhes por Barbosa (2006). A cidade de São José do Rio Preto está localizada na região noroeste do Estado de São Paulo, apresenta uma área de 575 *km*<sup>2</sup> e uma população de 406.826 habitantes estimada para o ano de 2005.

São José do Rio Preto está a 489 *m* acima do mar, possui clima tropical, temperatura média anual de 25,4° e pluviosidade por volta de 200 *mm* no período de outubro a março.

A área selecionada para a obtenção dos dados faz parte de um bairro de classe média chamado Boa Vista, considerado um dos mais antigos da cidade de São José do Rio Preto.

Os dados foram obtidos pela coleta de ovos postos pelo mosquito *Aedes aegypti* em armadilhas de oviposição (ovitampa) em 100 casas dispostas numa área de 50 quadras. Juntamente foram obtidas medidas climáticas correspondentes ao período de colocação e retirada das armadilhas. As

informações sobre a temperatura máxima, média e mínima, foram fornecidas pelo Aeroporto de São José do Rio Preto, a umidade máxima, média e mínima, pela Divisão Regional Agrícola de São José do Rio Preto da Secretaria da Agricultura do Estado de São Paulo - DIRA, e os dados pluviométricos foram medidos por uma pessoa treinada, que realizou a observação todos os dias as 6 : 00 horas da manhã. A coleta foi feita quinzenalmente durante um ano, perfazendo um total de 2.600 observações, sendo que os dados foram obtidos em ciclos quinzenais, com a colocação em uma semana e retirada na seguinte.

## 4.5 Área Trabalhada

As quadras selecionadas para a realização das atividades foram as mais homogêneas possíveis e sem a presença de pontos estratégicos (locais com grande concentração de recipientes, como borracharias, ferros velhos e etc). Para a instalação de armadilhas de oviposição (ovitrampa) foi selecionada uma área com 50 quadras com aproximadamente 1.700 casas térreas. Em cada quadra, duas casas foram escolhidas para a instalação das armadilhas. Para que as armadilhas apresentassem distribuição uniforme na área, as casas foram selecionadas em faces opostas de cada uma das quadras, ora nas faces norte e sul, ora nas faces leste e oeste.

O local selecionado para a instalação das armadilhas foi o peridomicílio ao relento, de preferência próximo à plantas, a uma altura que variou de 0.5 m a 1 m, sendo de fácil acesso para uma possível retirada, mesmo quando o morador não estivesse em casa.

## 4.6 Modelo sem Covariáveis

Nesta Seção, apresenta-se um estudo sem covariáveis, no qual é considerado apenas a variável resposta  $Y$ , número de ovos postos pelo mosquito *Aedes aegypti*, e, faz-se uma breve discussão sobre os resultados obtidos em relação aos dois tipos de modelos utilizados nesta dissertação.

### 4.6.1 Estimativas dos Parâmetros

Considerando os modelos ZIP e ZINB dados pelas expressões (3.1.1) e (2.4.9) e os estimadores de máxima verossimilhança dados pelas equações (3.1.8), (3.1.9) e (3.3.9), correspondentes aos modelos ZIP e ZINB, respectivamente, foram calculados os valores das estimativas para esses dois modelos, considerando  $Y_i$ , o número de ovos do mosquito *Aedes aegypti* para as 26 quinzenas. O processo iterativo para o cálculo das estimativas dos parâmetros  $\hat{\omega}$ ,  $\hat{\lambda}$  e  $\hat{k}$  teve como valores iniciais as estimativas encontradas através do métodos dos momentos, o programa encontra-se no **Apêndice B**.

Na Tabela 4.1 os valores de  $n$  correspondem ao total de observações, os valores de  $n_0$  equivalem aos números de zeros nas observações,  $\bar{Y}$  é a média amostral do número de ovos e  $S^2$  é a variância amostral do número de ovos para cada quinzena. Encontram-se também as médias e variâncias dos modelos ZIP e ZINB, bem como as estimativas dos parâmetros destes modelos para cada uma das 26 quinzenas.

**Tabela 4.1:** Estimativas para os parâmetros dos modelos ZIP e ZINB

Quin.	$\bar{Y}$	$S^2$	$n$	$n_0$	ZIP				ZINB				
					$\hat{\omega}$	$\hat{\lambda}$	$E(Y)$	$Var(Y)$	$E(Y)$	$Var(Y)$	$\hat{\omega}$	$\hat{\lambda}$	$\hat{k}$
1	7,54	363,20	96	66	0,31	24,13	7,54	132,02	7,53	131,33	0,33	22,83	0,88
2	17,13	857,77	100	42	0,58	29,53	17,13	229,55	17,02	213,39	0,63	27,02	0,65
3	4,41	95,51	96	55	0,42	10,32	4,41	30,46	4,33	37,26	0,42	10,32	0,62
4	5,34	175,08	99	70	0,29	18,22	5,32	73,63	5,20	70,27	0,32	16,26	0,69
5	12,63	725,43	96	62	0,35	35,66	12,62	301,77	12,60	305,70	0,36	35,00	1,16
6	21,59	1282,47	97	39	0,59	36,10	21,55	336,54	21,31	334,89	0,61	34,98	0,95
7	25,27	877,59	97	16	0,83	30,26	25,26	154,31	25,10	142,84	0,88	28,53	0,79
8	33,99	2037,09	99	22	0,77	43,70	33,95	371,85	33,10	357,12	0,80	42,02	0,82
9	30,79	1203,93	96	31	0,67	45,47	30,78	487,59	30,73	496,06	0,68	45,20	1,48
10	23,43	1054,31	97	24	0,75	31,13	23,41	205,04	23,36	185,33	0,81	28,85	0,69
11	33,88	1696,19	99	16	0,83	40,41	33,86	263,95	33,49	233,63	0,88	38,06	0,71
12	33,25	1784,75	97	14	0,85	38,85	33,22	220,33	33,17	208,27	0,89	37,28	0,85
13	21,79	1485,97	97	27	0,72	30,19	21,76	205,53	21,56	134,33	0,89	24,23	0,39
14	33,02	1581,08	99	16	0,83	39,38	33,00	251,50	32,67	243,56	0,86	37,99	0,88
15	42,54	2872,87	100	12	0,88	48,34	42,53	289,30	42,18	243,56	0,92	45,85	0,70
16	23,36	1512,27	99	31	0,68	34,00	23,33	274,66	23,12	322,43	0,68	34,00	0,48
17	16,85	846,65	100	35	0,65	25,92	16,84	169,69	16,50	188,48	0,66	25,00	0,52
18	5,07	151,37	100	60	0,40	12,67	5,07	43,59	4,86	49,52	0,40	12,16	0,53
19	18,28	1354,71	99	43	0,56	32,31	18,25	275,31	18,03	282,60	0,56	32,21	2,02
20	14,61	1069,45	100	48	0,52	28,09	14,61	211,55	14,49	172,29	0,63	23,00	0,42
21	11,49	579,75	94	39	0,58	19,63	11,48	105,25	11,31	80,95	0,75	15,09	0,42
22	7,48	362,98	99	65	0,34	21,78	7,47	113,85	7,37	101,35	0,41	17,99	0,47
23	9,5	423,86	100	34	0,66	14,39	9,50	55,96	9,31	50,68	0,94	9,91	0,26
24	4,21	221,82	100	72	0,28	15,03	4,20	49,75	3,94	47,93	0,28	14,10	1,01
25	1,75	65,00	99	80	0,19	9,11	1,74	14,50	1,73	18,26	0,19	9,11	0,46
26	6,8	286,41	95	57	0,40	17,00	6,8	76,16	2,54	76,25	0,41	16,21	1,10

Pode-se observar, que a estimativa do parâmetro  $\hat{\omega}$  no modelo ZINB, é sempre maior ou igual à estimativa obtida no modelo ZIP.

Com relação às estimativas do parâmetro  $\hat{\lambda}$ , observa-se um comportamento oposto, comparado ao parâmetro  $\hat{\omega}$ , ou seja, para o modelo ZIP, as estimativas observadas de  $\hat{\lambda}$  são sempre maiores ou iguais às estimativas do modelo ZINB. Verificando também, valores altos dessas estimativas nos dois modelos.

As médias estimadas pelo modelo ZIP para todas as quinzenas, estão mais próximas da média amostral do número de ovos do que, as médias estimadas pelo modelo ZINB.

Já a variância amostral do número de ovos apresenta valores muito maiores do que a média amostral, indicando assim uma superdispersão nos dados. As estimativas para a variância através dos dois modelos em estudo (ZIP e ZINB), não conseguem captar toda essa variabilidade, porém, na maioria das quinzenas, a variância estimada pelo modelo ZIP é maior do que a estimada pelo modelo ZINB.

## 4.6.2 Intervalos de Confiança

A metodologia descrita na Seção 2.5 foi aplicada aos dados coletados considerando as 26 quinzenas e o número de observações variando entre  $94 < n \leq 100$ .

Para o modelo ZIP, dado por (3.1.1), a matriz de informação é dada por

$$J = \begin{bmatrix} \frac{n\theta (1 - e^{-\lambda})^2}{(1 - \omega + \omega e^{-\lambda})^2} + \frac{n - n\theta}{\omega^2} & \frac{n\theta e^{-\lambda}}{(1 - \omega + \omega e^{-\lambda})^2} \\ \frac{n\theta e^{-\lambda}}{(1 - \omega + \omega e^{-\lambda})^2} & -\frac{n\theta \omega e^{-\lambda} (1 - \omega)}{(1 - \omega + \omega e^{-\lambda})^2} + \frac{nY}{\lambda^2} \end{bmatrix}. \quad (4.6.1)$$

Substituindo as estimativas dos parâmetros do modelo ZIP, que se encontram na Tabela 4.1, foi possível obter a matriz inversa e assim calcular para cada quinzena os intervalos de confiança assintóticos para os parâmetros  $\omega$  e  $\lambda$  através das expressões dada por (2.5.3) e (2.5.4).

Para o modelo ZINB, dado por (2.4.9), os elementos da matriz de informação observada são

$$J_{11} = \frac{n_0(1 - (\frac{k}{k+\lambda})^k)^2}{(1 - \omega + \omega(\frac{k}{k+\lambda})^k)^2} + \frac{(n - n_0)}{\omega^2} \quad (4.6.2)$$

$$J_{12} = J_{21} = \frac{n_0 k (\frac{k}{k+\lambda})^k}{(k + \lambda)(1 - \omega + \omega(\frac{k}{k+\lambda})^k)} \left( 1 - \frac{\omega(-1 + (\frac{k}{k+\lambda})^k)}{1 - \omega + \omega(\frac{k}{k+\lambda})^k} \right) \quad (4.6.3)$$

$$J_{13} = J_{31} = \frac{n_0 (\frac{k}{k+\lambda})^k \left( \ln(\frac{k}{k+\lambda}) + (\frac{1}{k+\lambda} - \frac{k}{(k+\lambda)^2})(k + \lambda) \right)}{1 - \omega + \omega(\frac{k}{k+\lambda})^k} \left( -1 - \frac{\omega(1 - (\frac{k}{k+\lambda})^k)}{1 - \omega + \omega(\frac{k}{k+\lambda})^k} \right) \quad (4.6.4)$$

$$J_{22} = \frac{n_0 \omega k (\frac{k}{k+\lambda})^k}{(k + \lambda)^2 (1 - \omega + \omega(\frac{k}{k+\lambda})^k)} \left( -k - 1 + \frac{\omega k (\frac{k}{k+\lambda})^k}{(1 - \omega + \omega(\frac{k}{k+\lambda})^k)} \right) - \frac{(n - n_0)k}{(k + \lambda)^2} + \frac{n\bar{y}k}{\lambda(k + \lambda)} \left( \frac{1}{k + \lambda} + \frac{1}{\lambda} \right) \quad (4.6.5)$$

$$J_{33} = \frac{n_0 \omega (\frac{k}{k+\lambda})^k \left( \ln(\frac{k}{k+\lambda}) + (\frac{1}{k+\lambda} - \frac{k}{(k+\lambda)^2})(k + \lambda) \right)^2}{1 - \omega + \omega(\frac{k}{k+\lambda})^k} \left( -1 + \frac{\omega (\frac{k}{k+\lambda})^k}{1 - \omega + \omega(\frac{k}{k+\lambda})^k} \right) - \sum_{y=1}^{\infty} \Psi(1, y + k) \quad (4.6.6)$$

$$- \frac{n\bar{y}}{(k + \lambda)^2} - (n - n_0)\Psi(1, k) - \frac{(n - n_0) \left( (\frac{1}{k+\lambda} - \frac{k}{(k+\lambda)^2})(k + \lambda)k^{-1} + (-\frac{2}{(k+\lambda)^2} + \frac{2k}{(k+\lambda)^3})(k + \lambda) + \frac{1}{k+\lambda} - \frac{k}{(k+\lambda)^2} \right)}{1 - \omega + \omega(\frac{k}{k+\lambda})^k}, \quad (4.6.7)$$



em que  $\bar{y}$  é a média amostral das observações do número de ovos e  $\Psi(1, \cdot)$  é a primeira derivada da função  $\Psi$  calculada no ponto.

Substituindo as estimativas dos parâmetros  $k$ ,  $\lambda$  e  $\omega$  do modelo ZINB, que se encontram na Tabela 4.1, foi possível calcular a matriz inversa e assim calcular para cada quinzena os intervalos de confiança assintóticos para os parâmetros.

Os intervalos de confiança para o modelo ZINB podem ser obtidos através das expressões (2.5.3) e (2.5.4), para os parâmetros  $\lambda$  e  $\omega$  e, para o parâmetro  $k$ ,

$$IC(k; 100(1 - \alpha)\%) = \hat{k} \pm z_{\frac{\alpha}{2}} \sqrt{Var(\hat{k})}. \quad (4.6.8)$$

Na Tabela 4.2 são apresentados os resultados dos intervalos considerando 95% de confiança para os parâmetros dos modelos ZIP e ZINB.

**Tabela 4.2:** Intervalos de Confiança para os modelos ZIP e ZINB

Quinzena	ZIP				ZINB					
	$\hat{\omega}$		$\hat{\lambda}$		$\hat{\omega}$	$\hat{\lambda}$		$\hat{k}$		
1	[0,21	0,40]	[22,37	25,89]	[0,22	0,43]	[13,83	31,83]	[0,29	1,48]
2	[0,48	0,67]	[28,13	30,93]	[0,51	0,75]	[18,08	35,95]	[0,30	1,00]
3	[0,32	0,52]	[9,34	11,30]	[0,30	0,54]	[5,57	15,07]	[0,26	0,98]
4	[0,20	0,38]	[16,67	19,78]	[0,21	0,44]	[8,62	23,90]	[0,12	1,27]
5	[0,25	0,44]	[33,65	37,66]	[0,26	0,45]	[23,84	46,16]	[0,53	1,78]
6	[0,50	0,69]	[34,10	38,11]	[0,51	0,72]	[25,56	44,39]	[0,54	1,36]
7	[0,76	0,90]	[29,06	31,45]	[0,79	0,97]	[21,32	5,74]	[0,46	1,12]
8	[0,69	0,85]	[42,22	45,17]	[0,71	0,89]	[31,49	52,54]	[0,50	1,14]
9	[0,58	0,77]	[43,83	47,11]	[0,58	0,77]	[36,02	54,38]	[0,93	2,03]
10	[0,66	0,83]	[29,85	32,41]	[0,70	0,92]	[20,63	37,07]	[0,37	1,01]
11	[0,76	0,91]	[39,04	41,77]	[0,79	0,98]	[28,12	47,99]	[0,42	1,00]
12	[0,78	0,92]	[37,51	40,19]	[0,81	0,97]	[28,42	46,14]	[0,53	1,17]
13	[0,63	0,81]	[28,90	31,48]	[0,67	1,12]	[14,16	34,30]	[0,14	0,64]
14	[0,76	0,91]	[38,03	40,73]	[0,78	0,95]	[29,11	46,87]	[0,55	1,02]
15	[0,81	0,94]	[46,88	49,79]	[0,84	1,00]	[34,25	57,44]	[0,44	0,97]
16	[0,59	0,77]	[32,62	35,39]	[0,55	0,81]	[21,27	46,74]	[0,29	0,67]
17	[0,55	0,74]	[24,68	27,16]	[0,53	0,79]	[16,01	34,00]	[0,30	0,75]
18	[0,30	0,49]	[11,57	13,77]	[0,28	0,52]	[6,30	18,02]	[0,23	0,84]
19	[0,46	0,66]	[30,82	33,80]	[0,46	0,66]	[26,09	38,33]	[0,55	3,50]
20	[0,42	0,61]	[26,65	29,53]	[0,44	0,82]	[12,42	33,59]	[0,11	0,73]
21	[0,48	0,68]	[18,46	20,80]	[0,46	1,05]	[7,18	23,01]	[0,01	0,84]
22	[0,24	0,43]	[20,21	23,34]	[0,25	0,57]	[8,26	27,72]	[0,05	0,90]
23	[0,56	0,75]	[13,47	15,30]	[0,70	1,12]	[5,49	14,33]	[0,15	0,37]
24	[0,19	0,36]	[13,59	16,47]	[0,18	0,38]	[8,41	19,80]	[0,06	2,09]
25	[0,11	0,26]	[7,75	10,47]	[0,10	0,28]	[1,49	6,73]	[0,09	0,83]
26	[0,30	0,49]	[15,68	18,31]	[0,30	0,52]	[10,96	1,45]	[0,22	1,99]

Observou-se que a amplitude dos intervalos de confiança para o modelo ZIP são menores, comparativamente, do que o modelo ZINB e que o zero não está contido em nenhum dos intervalos.

### 4.6.3 Resultados do Teste de Wald

Após a estimação dos parâmetros deve-se proceder à investigação da significância estatística dos mesmos. O Teste de Wald é utilizado para avaliar se o parâmetro é estatisticamente significativo.

Os resultados do Teste de Wald apresentados nesta seção para os modelos ZIP e ZINB estão relacionados com a metodologia descrita na Seção 2.4.1. Esta estatística teste sob  $H_0$  tem distribuição  $\chi^2_{(1)}$ , com um grau de liberdade, pois estamos tratando dos modelos sem considerar as covariáveis. Sendo seu valor comparado com valores tabulados de acordo com o nível de significância definido.

A estatística do teste de Wald dada por (2.4.3) foi calculada para todas as quinzenas para os modelos ZIP e ZINB considerando para o parâmetro  $\omega$  as hipóteses nulas de que  $H_0 : \omega = 0$  e  $H_0 : \omega = 1$ , considerando a restrição de que  $0 < \omega < 1$  para os modelos ZIP e ZINB. Na Tabela 4.3 encontram-se as estatísticas de Wald para os testes do parâmetro  $\omega$ .

**Tabela 4.3:** Teste Wald para o parâmetro  $\omega$  dos modelos ZIP e ZINB

Quinzena	ZIP		ZINB	
	$W_{\omega=0}$	$W_{\omega=1}$	$W_{\omega=0}$	$W_{\omega=1}$
1	43,6	211,2	40,3	166,0
2	138,0	72,4	102,8	34,3
3	71,5	128,7	46,8	84,2
4	41,0	238,9	31,0	130,0
5	52,6	175,0	51,9	163,1
6	150,0	67,8	137,1	52,7
7	491,0	19,1	347,0	5,8
8	346,5	28,2	303,0	16,9
9	201,2	45,7	200,5	43,9
10	259,0	31,8	204,9	11,0
11	513,5	19,0	359,2	5,52
12	575,0	16,3	475,0	7,0
13	251,4	37,4	61,0	0,7
14	513,5	19,0	439,6	9,97
15	733,3	13,6	501,8	3,0
16	217,1	45,1	111,5	23,19
17	185,7	53,8	102,6	25,6
18	66,6	149,9	43,2	91,2
19	128,9	76,0	127,8	75,0
20	108,3	92,3	43,0	14,2
21	132,5	66,6	25,4	2,5
22	51,7	189,2	26,1	51,6
23	194,1	51,5	58,8	0,16
24	38,8	257,1	31,9	193,4
25	23,5	416,7	17,0	300,7
26	63,3	142,5	55,4	112,6

Considerando um nível de significância  $\alpha = 0,05$ , e sendo que o percentil 0,95 de uma distribuição  $\chi^2$  com um grau de liberdade é 3,84, pode-se observar pelos resultados da Tabela 4.3 que, em relação a hipótese nula  $H_0 : \omega = 0$  nos dois modelos ZIP e ZINB, o valor obtido da estatística Wald rejeita fortemente a hipótese nula  $\omega = 0$ , indicando que existe a inflação de zeros. Em relação à hipótese nula  $H_0 : \omega = 1$ , para o modelo ZIP, os valores obtidos para a estatística do Teste Wald também rejeita esta hipótese e, para o modelo ZINB não rejeitamos a hipótese nula para as quinzenas 13, 15, 21 e 23, o que significa, que para essas quinzenas o modelo mais adequado para o conjunto de dados é o modelo Binomial Negativo.

#### 4.6.4 Resultado do Teste Escore ZIP

Nesta Seção é feita uma aplicação da metodologia descrita na Seção 3.1.2 referente ao modelo Poisson Inflacionado de Zeros considerando os dados do número de ovos postos pelo *Aedes aegypti*.

A Tabela 4.4 mostra os resultados obtidos do Teste Escore ZIP calculado para as 26 quinzenas do número de ovos do mosquito *Aedes aegypti*, onde foi utilizada a estatística escore dada pela equação (3.2.24).

**Tabela 4.4:** Teste Escore para o modelo ZIP

Quinzena	Estatística Escore ( $S_1$ )	Quinzena	Estatística Escore ( $S_1$ )
1	$0,15 \times 10^{20}$	14	$0,38 \times 10^{42}$
2	$0,26 \times 10^{36}$	15	$0,56 \times 10^{78}$
3	$0,11 \times 10^{14}$	16	$0,71 \times 10^{46}$
4	$0,17 \times 10^{16}$	17	$0,39 \times 10^{35}$
5	$0,14 \times 10^{29}$	18	$0,27 \times 10^{15}$
6	$0,14 \times 10^{44}$	19	$0,31 \times 10^{38}$
7	$0,10 \times 10^{49}$	20	$0,15 \times 10^{32}$
8	$0,10 \times 10^{82}$	21	$0,95 \times 10^{31}$
9	$0,36 \times 10^{65}$	22	$0,10 \times 10^{20}$
10	$0,33 \times 10^{46}$	23	$0,27 \times 10^{22}$
11	$0,13 \times 10^{64}$	24	$0,14 \times 10^{14}$
12	$0,60 \times 10^{62}$	25	$0,15 \times 10^9$
13	$0,71 \times 10^{43}$	26	$0,36 \times 10^{18}$

O percentil da  $\chi^2_{(1)}$  considerando o nível de 5% de significância, tem valor crítico do teste igual a 3,84 e, observa-se que, para todas as quinzenas os valores obtidos através da estatística escore (3.2.24) são bem maiores que o valor crítico, rejeitando a hipótese nula  $H_0 : \theta = 0$ . Pode-se concluir que o modelo Poisson Inflacionado de Zeros é mais apropriado do que o modelo Poisson padrão, que é a distribuição sob a hipótese nula para estes dados.

#### 4.6.5 Resultado do Teste Escore ZINB

Esta seção traz uma aplicação referente à metodologia descrita sobre o Teste Escore ZINB apresentado na Seção 3.3.2 para verificar a suposição de que o modelo ZINB é mais apropriado para ajustar o número de ovos do que o modelo ZIP.

Para cada quinzena do número de ovos, foi calculada a estatística escore para o modelo ZINB, considerando as estimativas dos parâmetros  $\hat{\lambda}$  e  $\hat{\omega}$  obtidas na Seção 4.2.1. A Tabela 4.5 mostra os resultados obtidos da estatística do Teste Escore ZINB para as 26 quinzenas.

**Tabela 4.5:** Teste Escore para o modelo ZINB

Quinzena	Estatística Escore ( $S_2$ )	Quinzena	Estatística Escore ( $S_2$ )
1	43,44	14	26,89
2	45,35	15	66,19
3	29,54	16	69,99
4	34,98	17	58,62
5	29,55	18	38,00
6	47,92	19	75,15
7	14,82	20	77,62
8	42,71	21	67,89
9	27,50	22	46,41
10	39,23	23	57,39
11	30,32	24	52,60
12	49,91	25	38,34
13	95,31	26	46,16

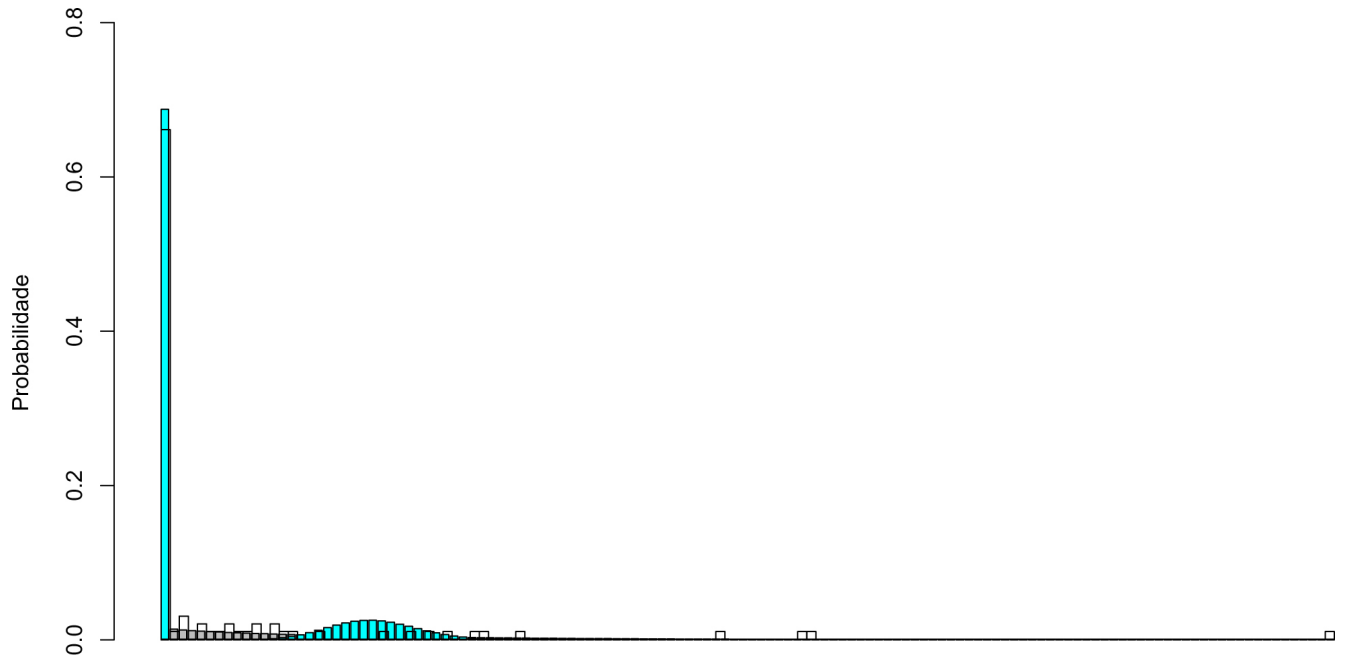
Para todas as quinzenas os valores obtidos através da estatística escore (3.4.20) são bem maiores do que o percentil da Normal  $(0, 1)$  para um nível de 5% de significância, cujo valor crítico é 1,64 (pois o teste é unilateral), rejeitando fortemente a hipótese nula  $H_0 : \lim_{\alpha \rightarrow 0} \alpha$ . Conclui-se que o modelo Binomial Negativa Inflacionado de Zeros, neste caso, é mais apropriado do que o modelo Poisson Inflacionado de Zeros, que é a distribuição sob a hipótese nula.

#### 4.6.6 Análise Gráfica

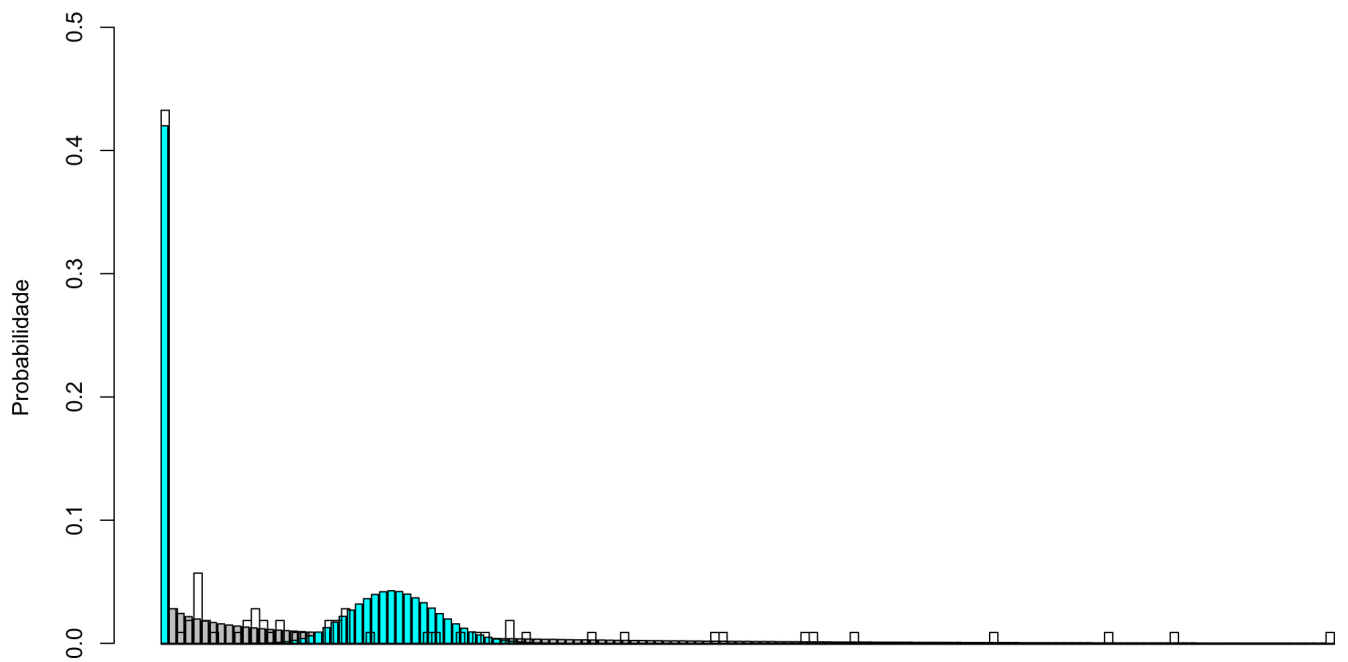
Com o objetivo de verificar qual modelo representa melhor, graficamente, o número de ovos posto pelo *aedes*, entre os modelos ZIP e ZINB para modelar o número de ovos considerando as covariáveis climáticas, foi utilizada uma implementação gráfica para analisar o comportamento destes modelos comparado com ao número de ovos.

A partir das estimativas dos parâmetros dos modelos ZIP e ZINB dados na Tabela 4.1, foram construídos os gráficos destas distribuições, na forma de histogramas, das 26 quinzenas, para se comparar estas distribuições com o histograma dos dados de cada quinzena.

As figuras de 4.1 à 4.26, apresentam estes gráficos.

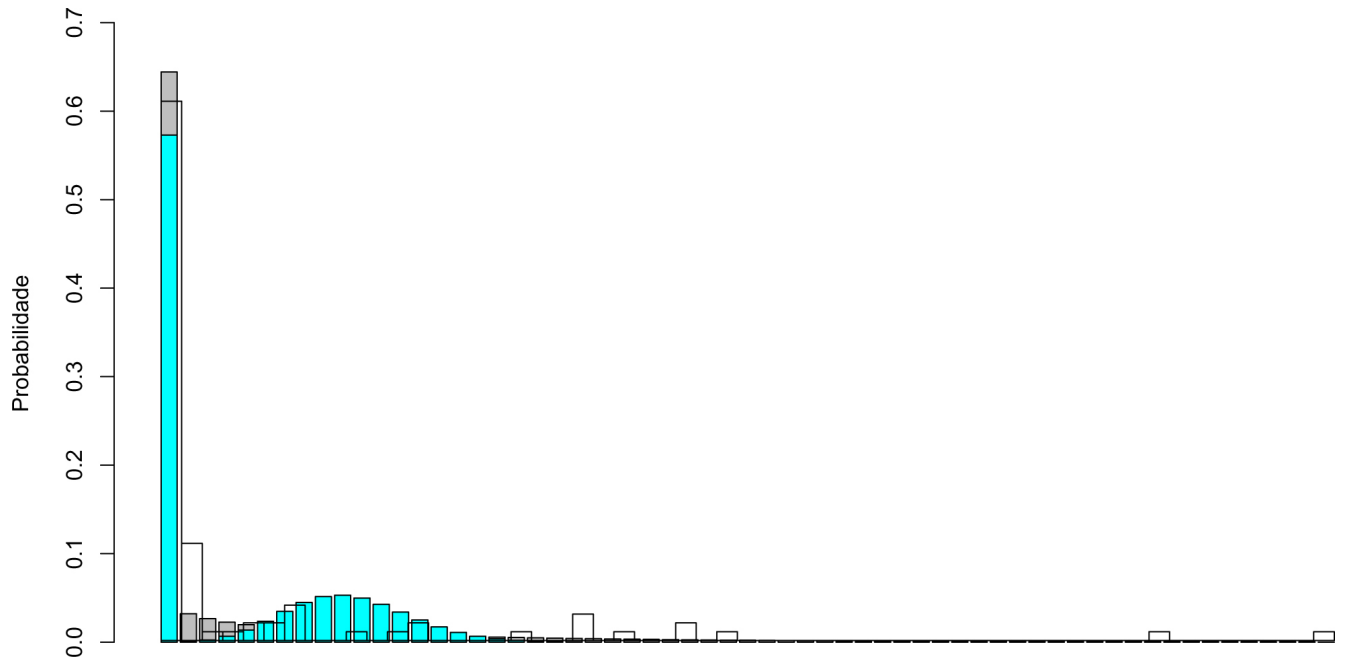


**Figura 4.1:** Histograma da observação do número de ovos e histograma dos modelos ZIP e ZINB a partir das estimativas da Tabela 4.1 da 1ª Quinzena

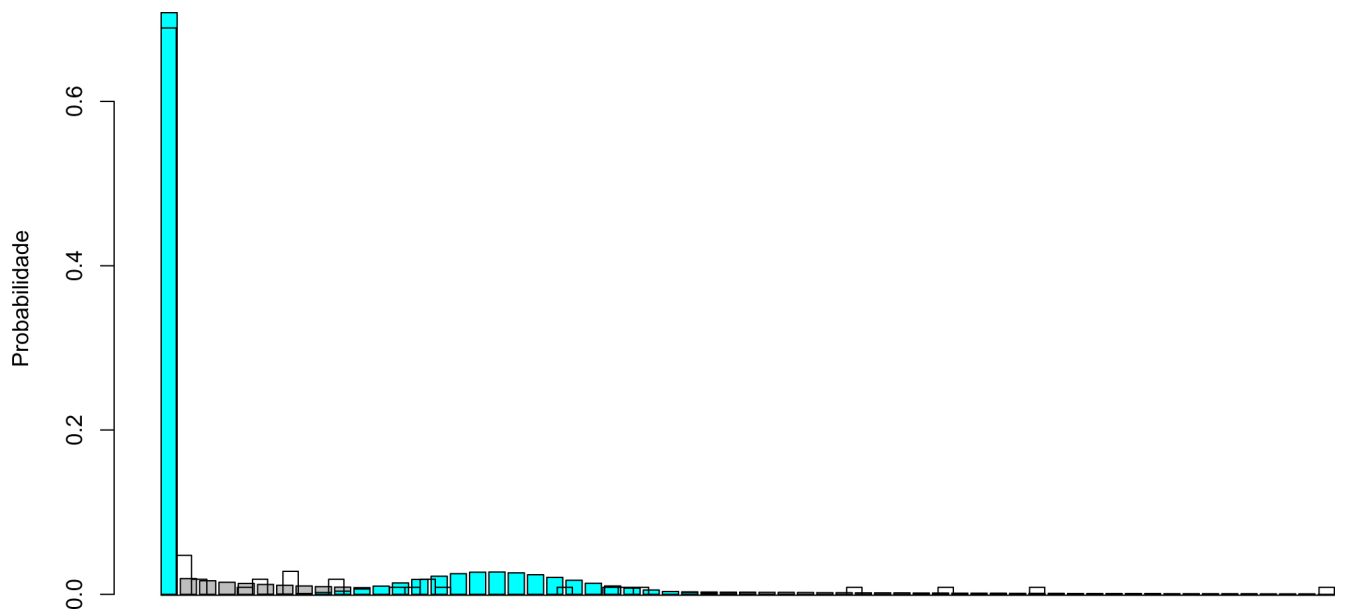


**Figura 4.2:** Histograma dos valores observados de número de ovos e histogramas dos modelos ZIP e ZINB a partir das estimativas da Tabela 4.1 da 2ª Quinzena

- - Histograma dos Dados
- (azul) - Gráfico do modelo ZIP
- (cinza) - Gráfico do modelo ZINB

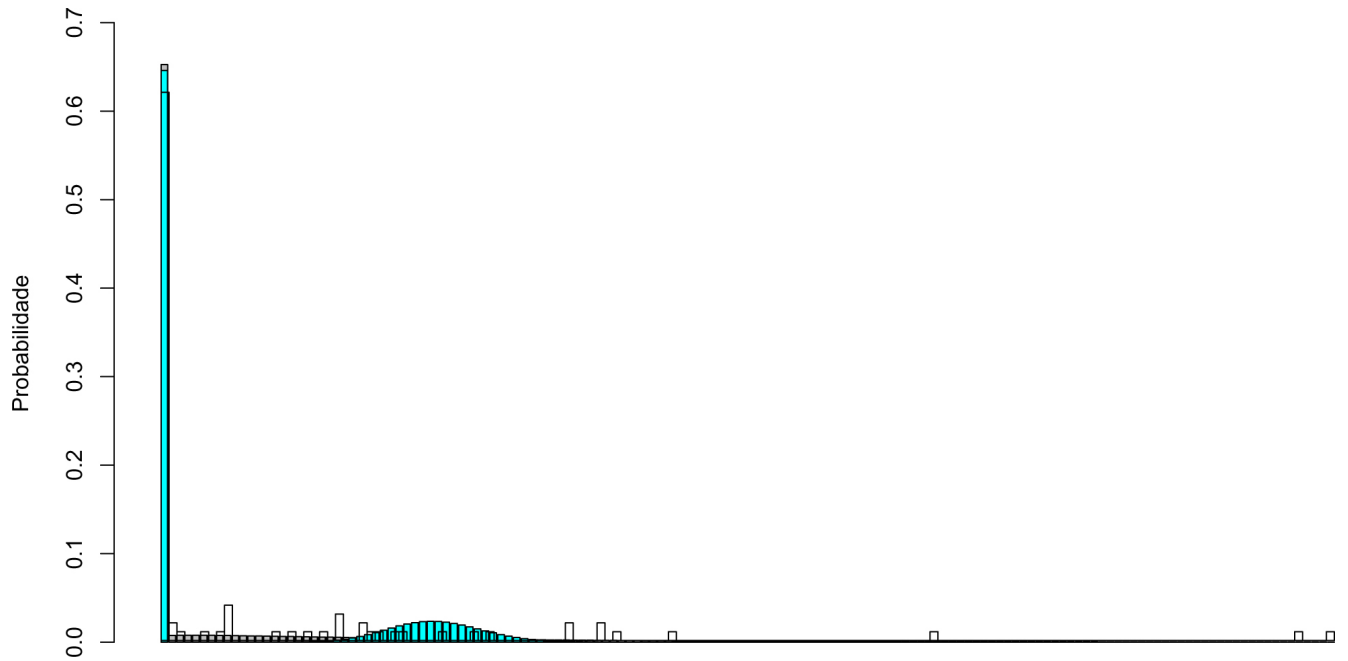


**Figura 4.3:** Histograma dos valores observados de número de ovos e histogramas dos modelos ZIP e ZINB a partir das estimativas da Tabela 4.1 da 3ª Quinzena

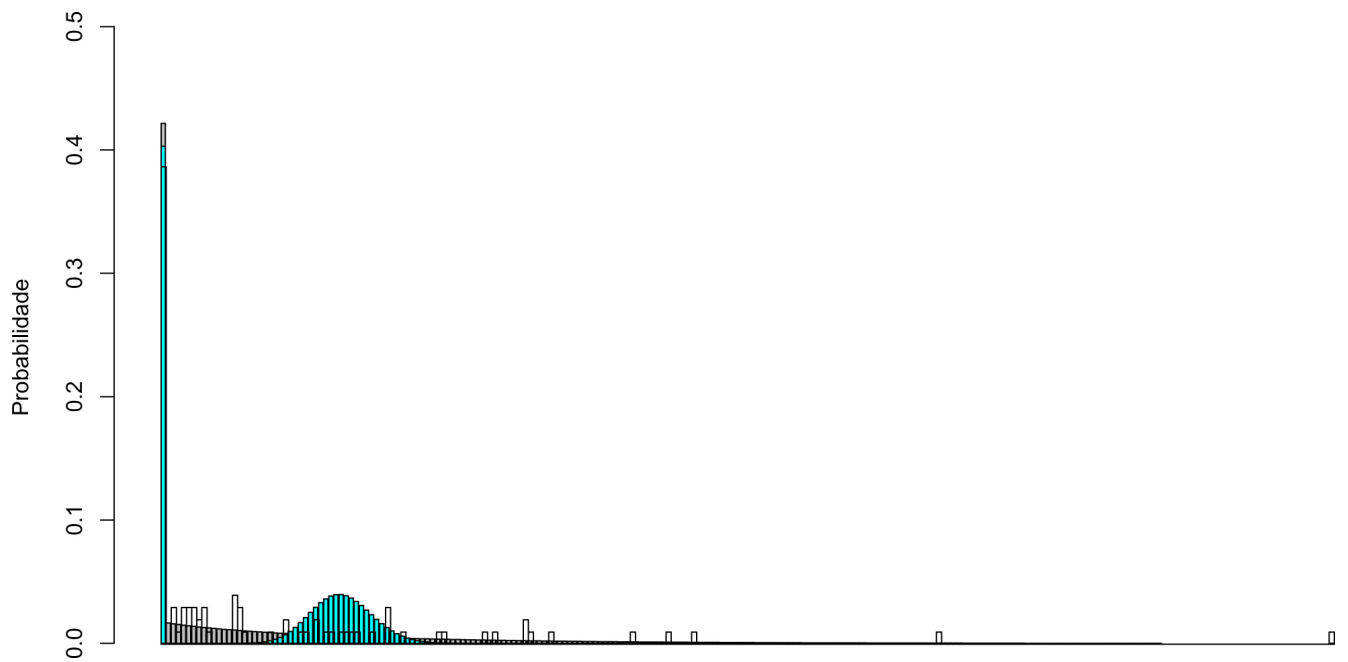


**Figura 4.4:** Histograma dos valores observados de número de ovos e histogramas dos modelos ZIP e ZINB a partir das estimativas da Tabela 4.1 da 4ª Quinzena

- - Histograma dos Dados
- - Gráfico do modelo ZIP
- - Gráfico do modelo ZINB



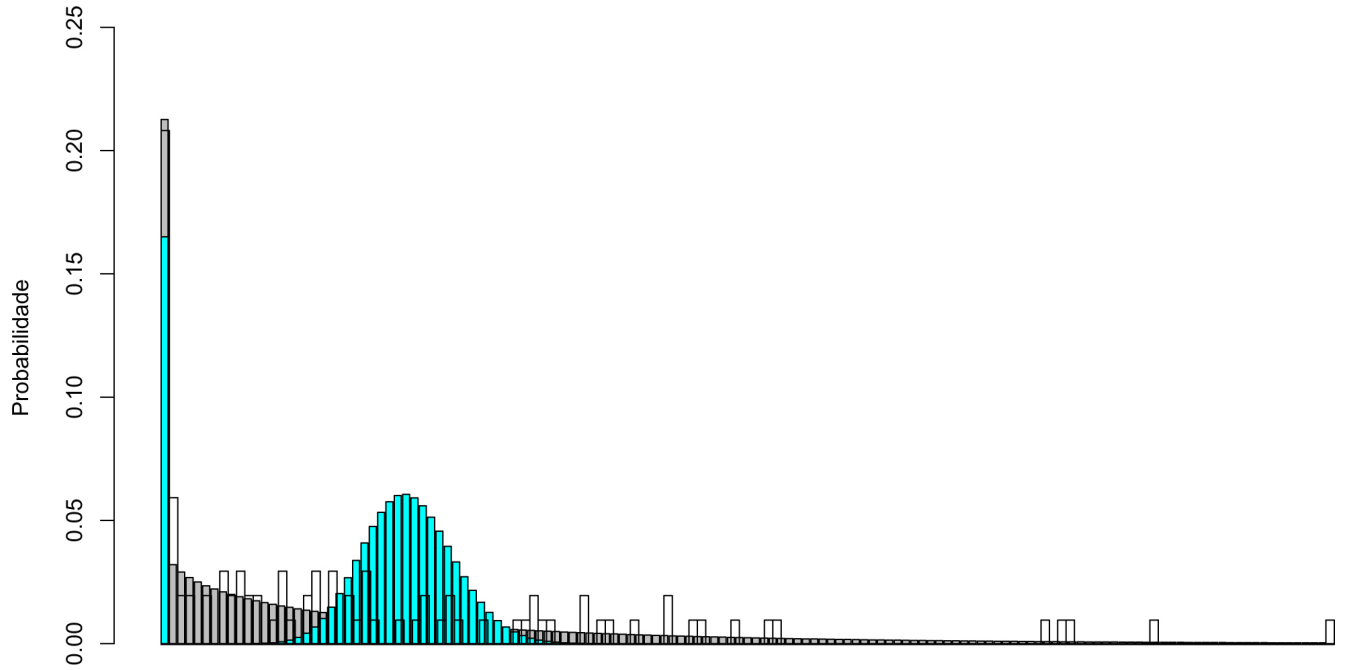
**Figura 4.5:** Histograma dos valores observados de número de ovos e histogramas dos modelos ZIP e ZINB a partir das estimativas da Tabela 4.1 da 5ª Quinzena



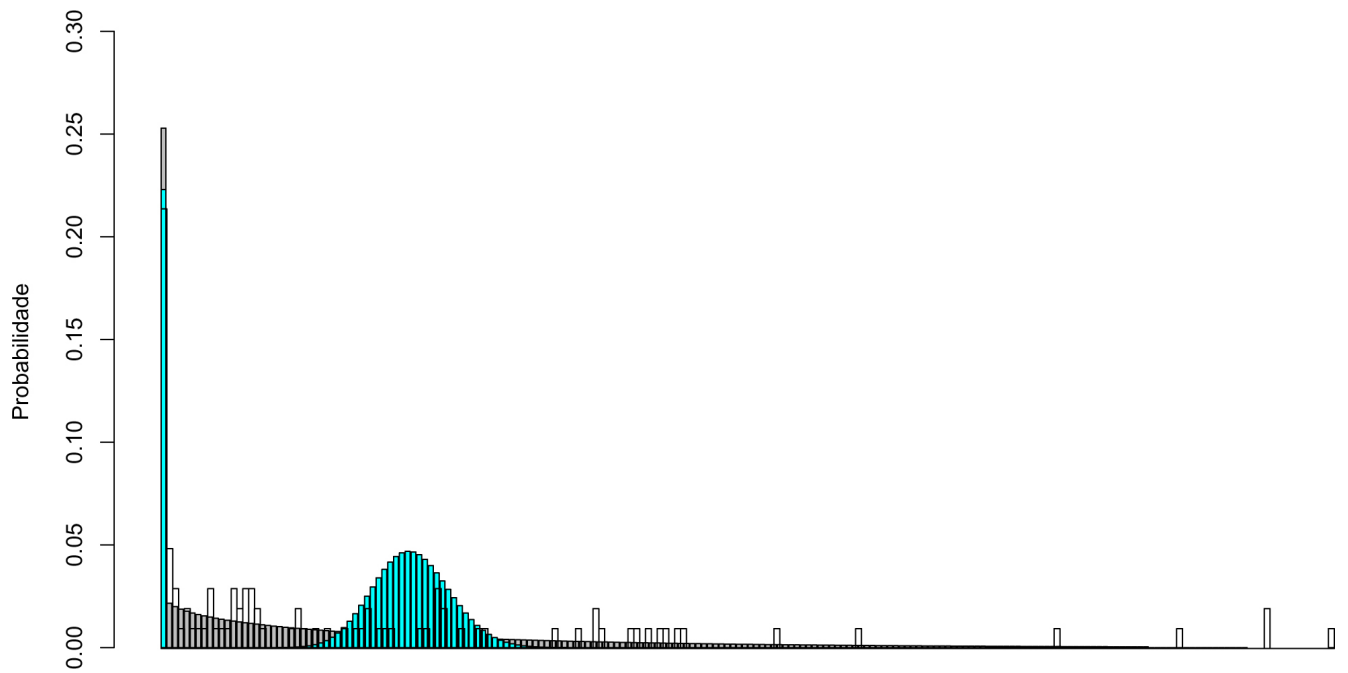
**Figura 4.6:** Histograma dos valores observados de número de ovos e histogramas dos modelos ZIP e ZINB a partir das estimativas da Tabela 4.1 da 6ª Quinzena

- - Histograma dos Dados
- - Gráfico do modelo ZIP
- - Gráfico do modelo ZINB



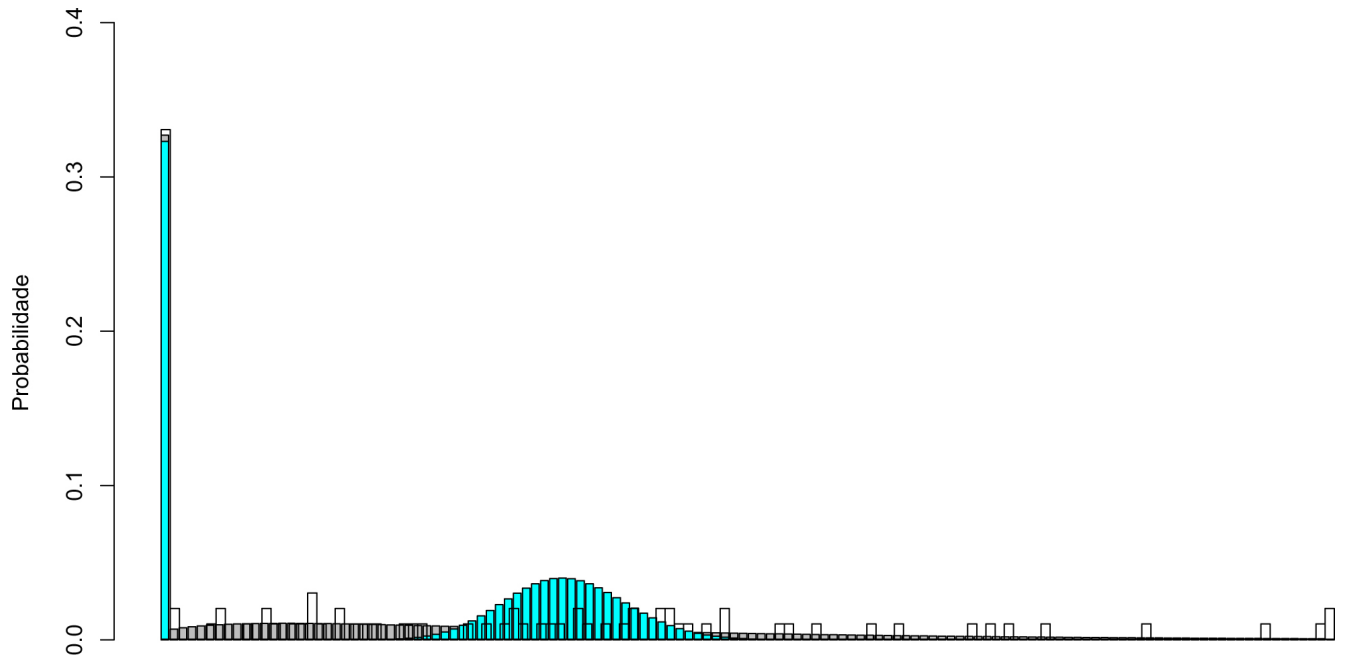


**Figura 4.7:** Histograma dos valores observados de número de ovos e histogramas dos modelos ZIP e ZINB a partir das estimativas da Tabela 4.1 da 7ª Quinzena

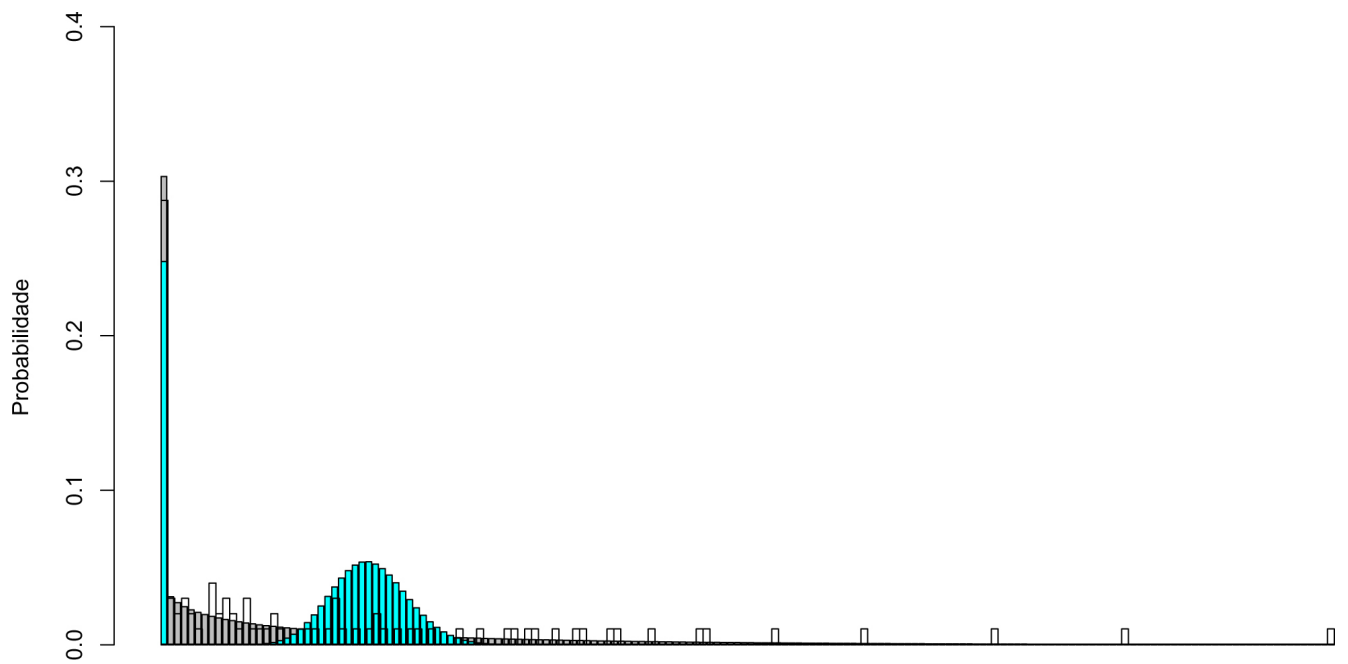


**Figura 4.8:** Histograma dos valores observados de número de ovos e histogramas dos modelos ZIP e ZINB a partir das estimativas da Tabela 4.1 da 8ª Quinzena

- - Histograma dos Dados
- - Gráfico do modelo ZIP
- - Gráfico do modelo ZINB

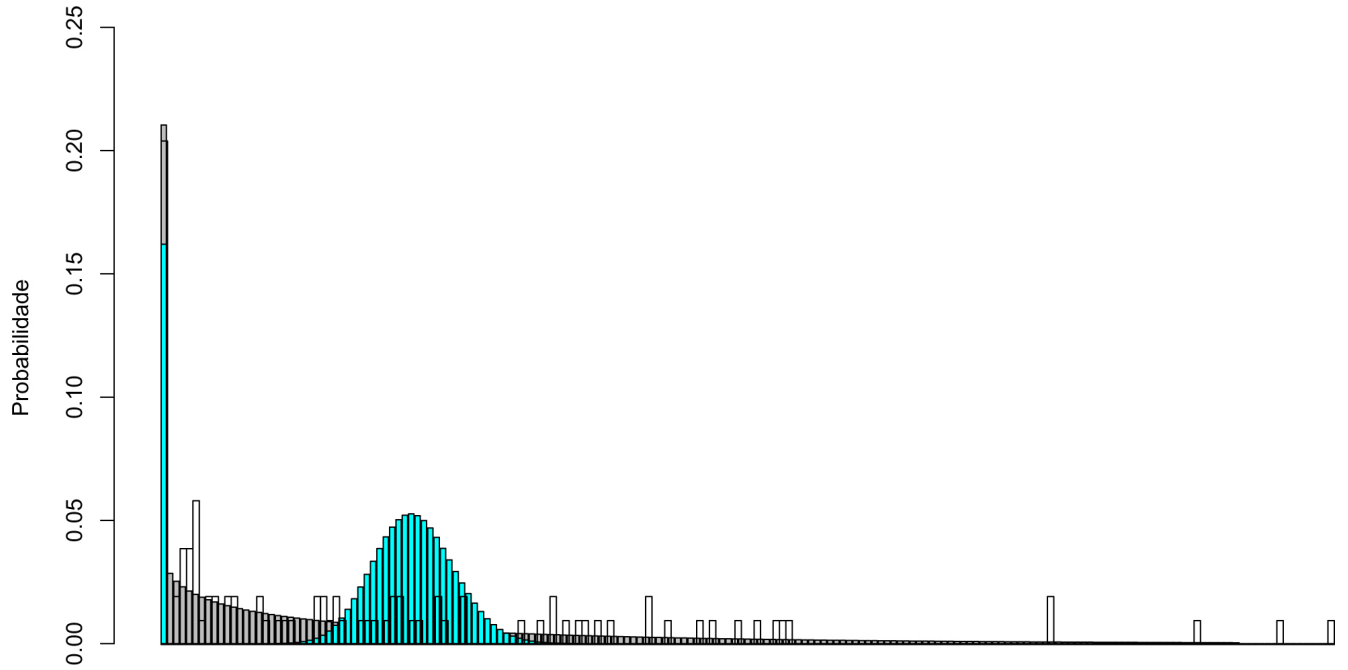


**Figura 4.9:** Histograma dos valores observados de número de ovos e histogramas dos modelos ZIP e ZINB a partir das estimativas da Tabela 4.1 da 9ª Quinzena

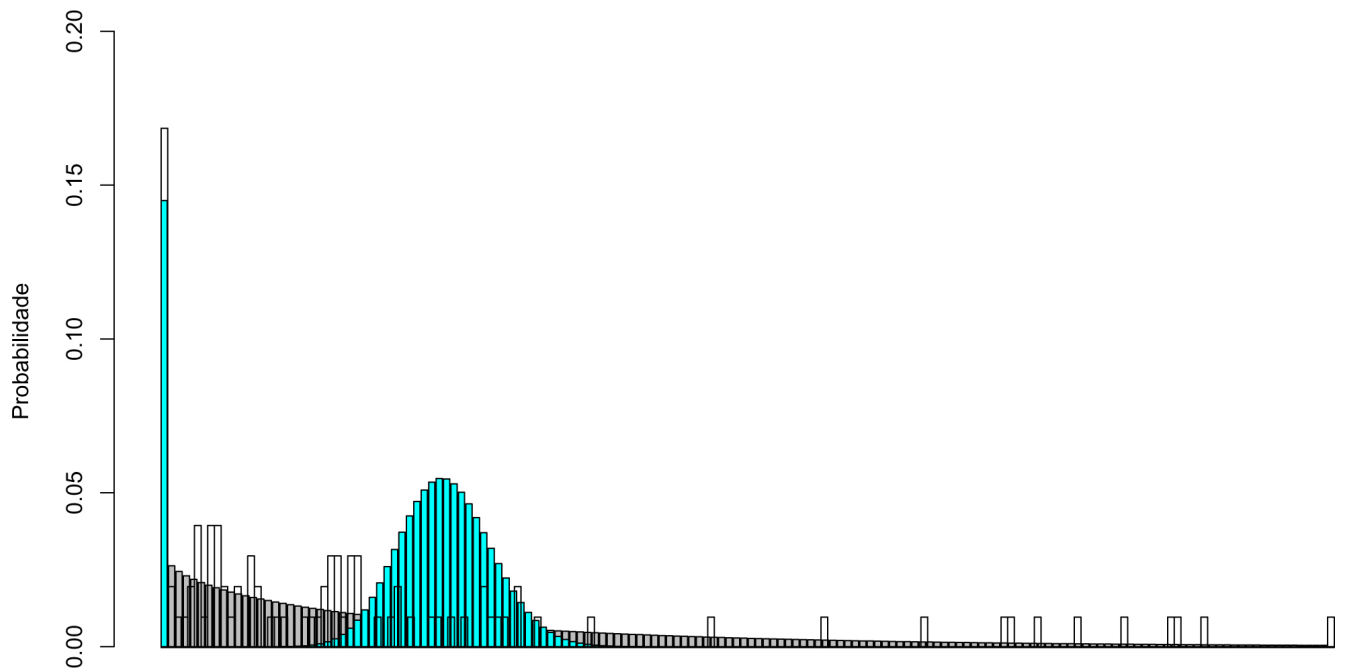


**Figura 4.10:** Histograma dos valores observados de número de ovos e histogramas dos modelos ZIP e ZINB a partir das estimativas da Tabela 4.1 da 10ª Quinzena

- - Histograma dos Dados
- (verde) - Gráfico do modelo ZIP
- (cinza) - Gráfico do modelo ZINB

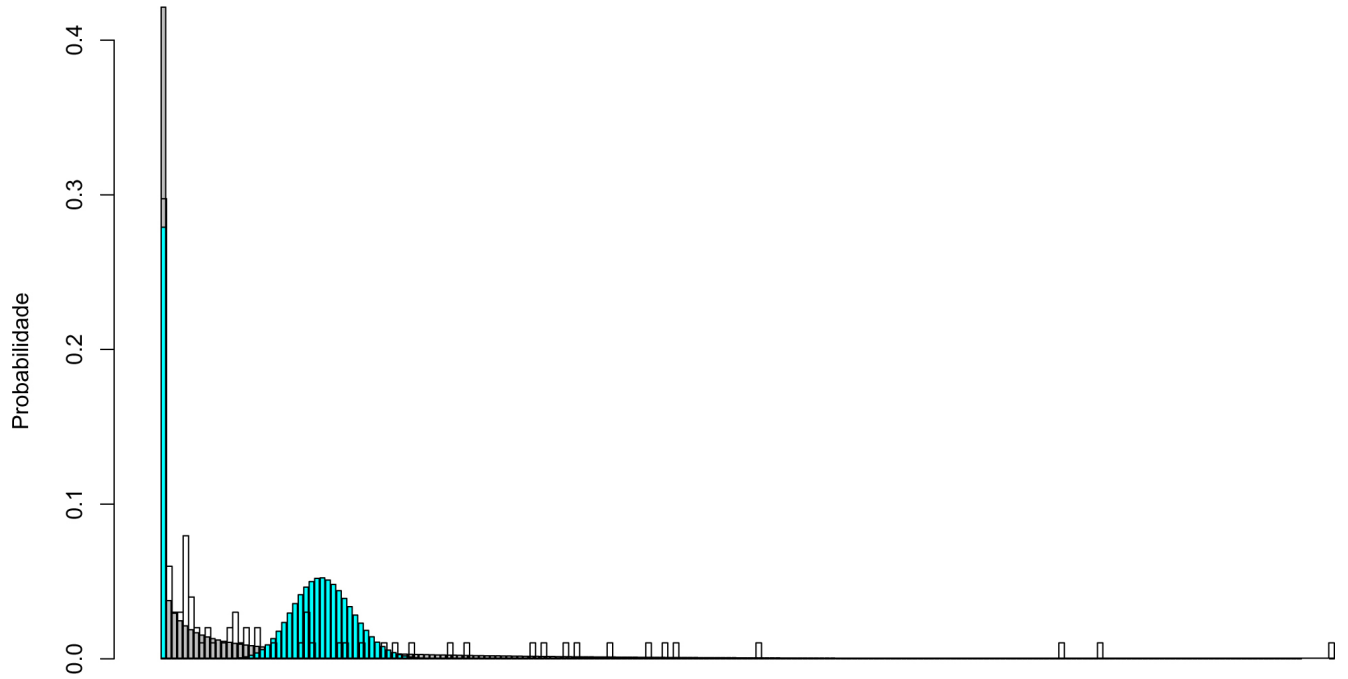


**Figura 4.11:** Histograma dos valores observados de número de ovos e histogramas dos modelos ZIP e ZINB a partir das estimativas da Tabela 4.1 da 11ª Quinzena

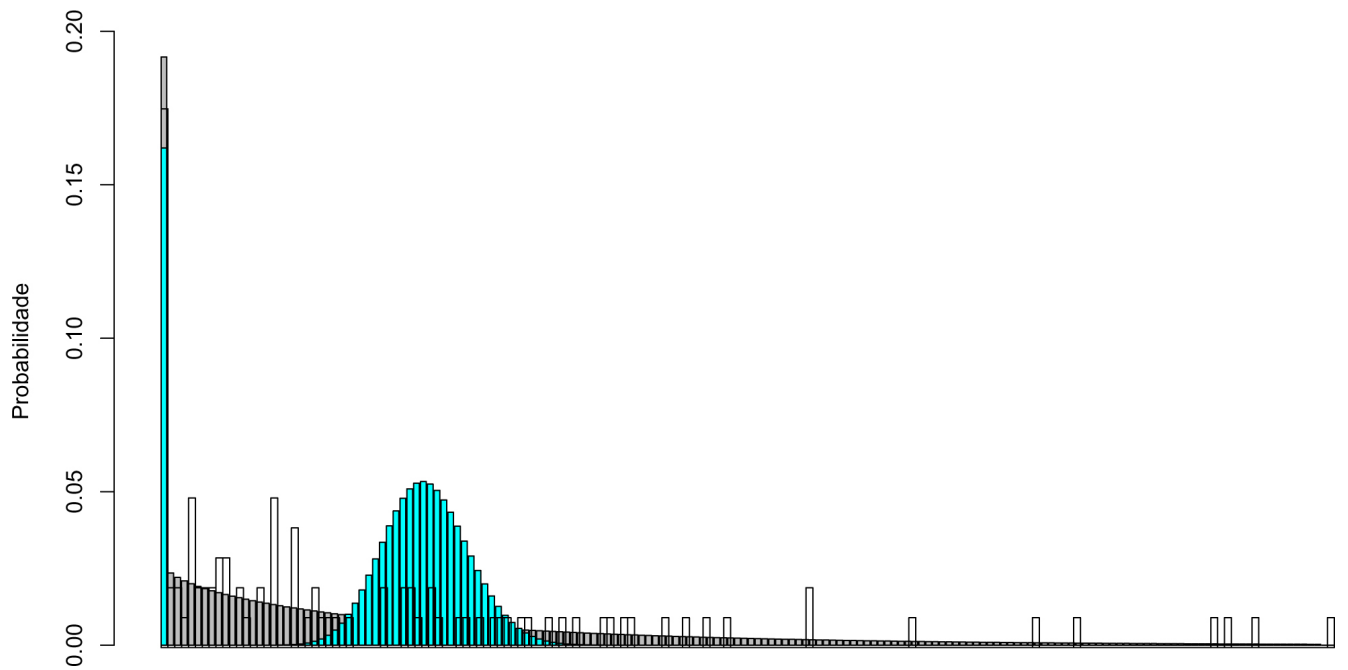


**Figura 4.12:** Histograma dos valores observados de número de ovos e histogramas dos modelos ZIP e ZINB a partir das estimativas da Tabela 4.1 da 12ª Quinzena

- - Histograma dos Dados
- (Ciano) - Gráfico do modelo ZIP
- (Cinza) - Gráfico do modelo ZINB

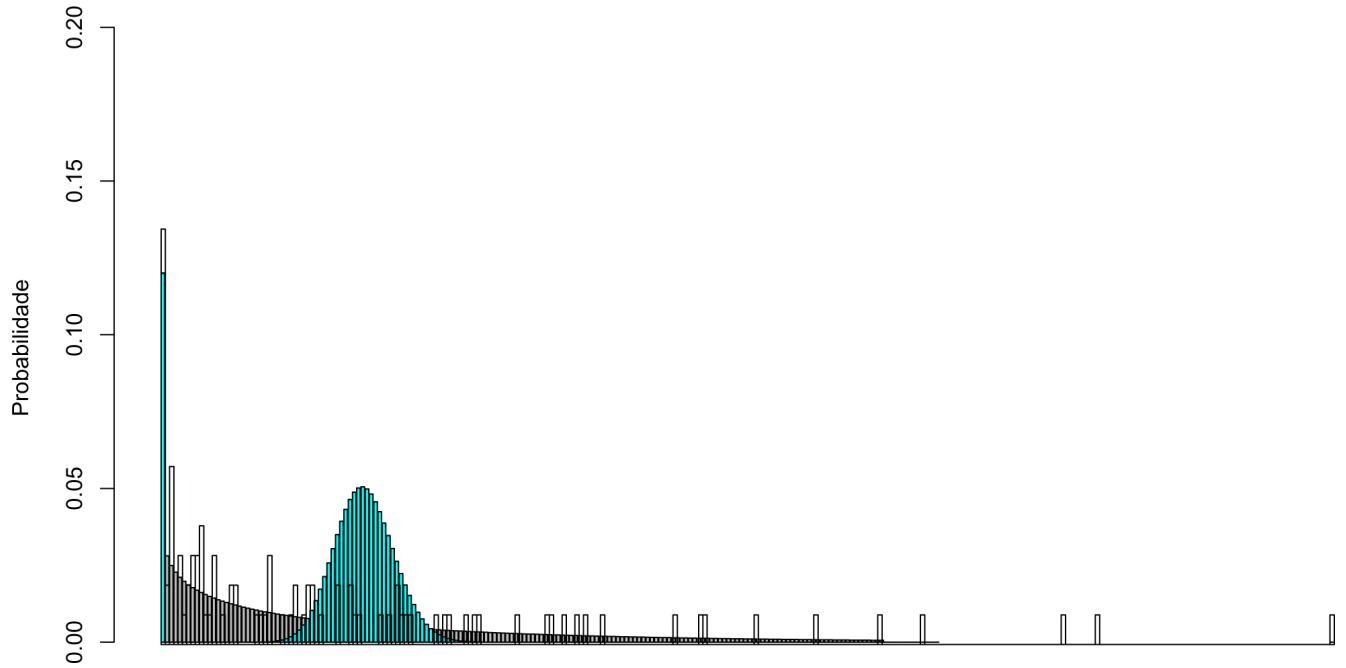


**Figura 4.13:** Histograma dos valores observados de número de ovos e histogramas dos modelos ZIP e ZINB a partir das estimativas da Tabela 4.1 da 13ª Quinzena

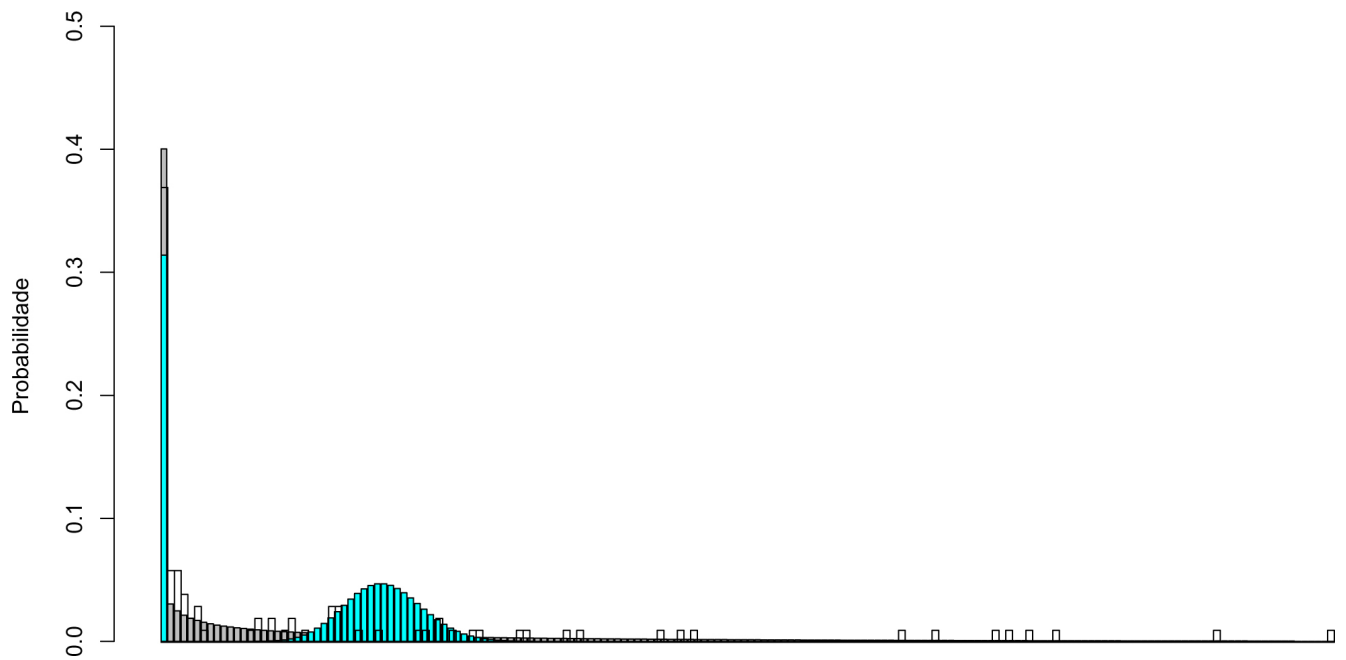


**Figura 4.14:** Histograma dos valores observados de número de ovos e histogramas dos modelos ZIP e ZINB a partir das estimativas da Tabela 4.1 da 14ª Quinzena

- - Histograma dos Dados
- - Gráfico do modelo ZIP
- - Gráfico do modelo ZINB

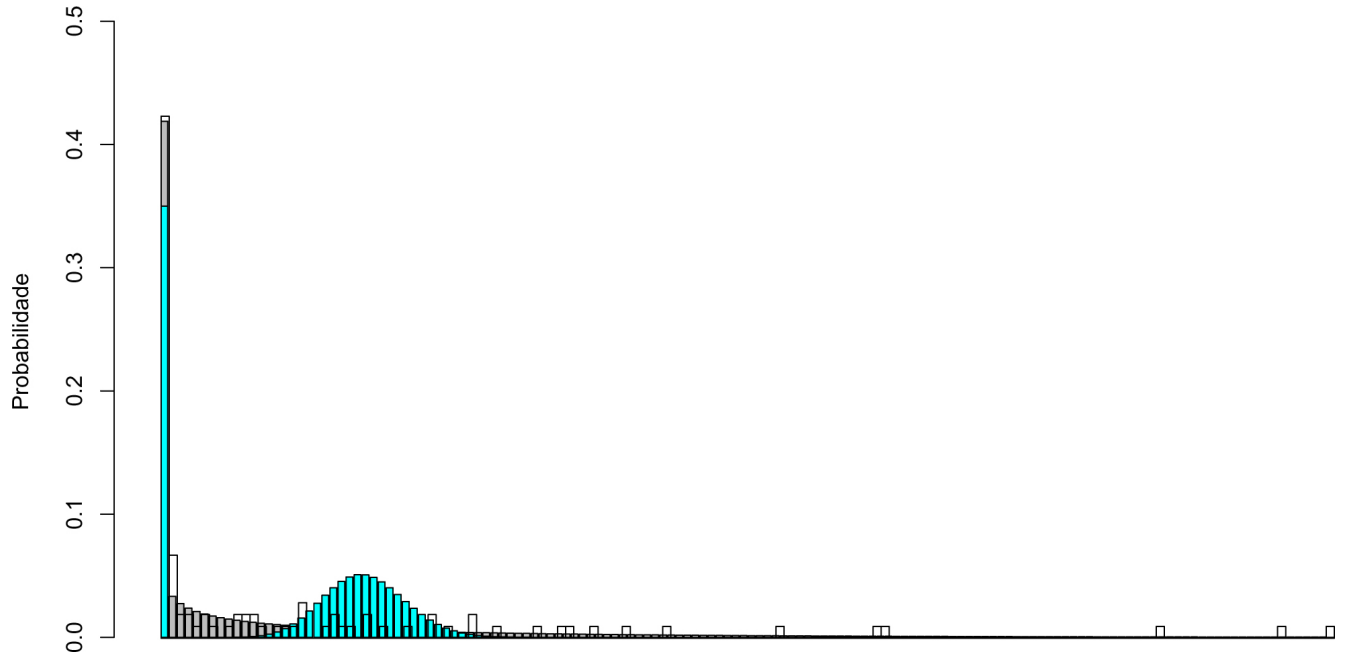


**Figura 4.15:** Histograma dos valores observados de número de ovos e histogramas dos modelos ZIP e ZINB a partir das estimativas da Tabela 4.1 da 15ª Quinzena

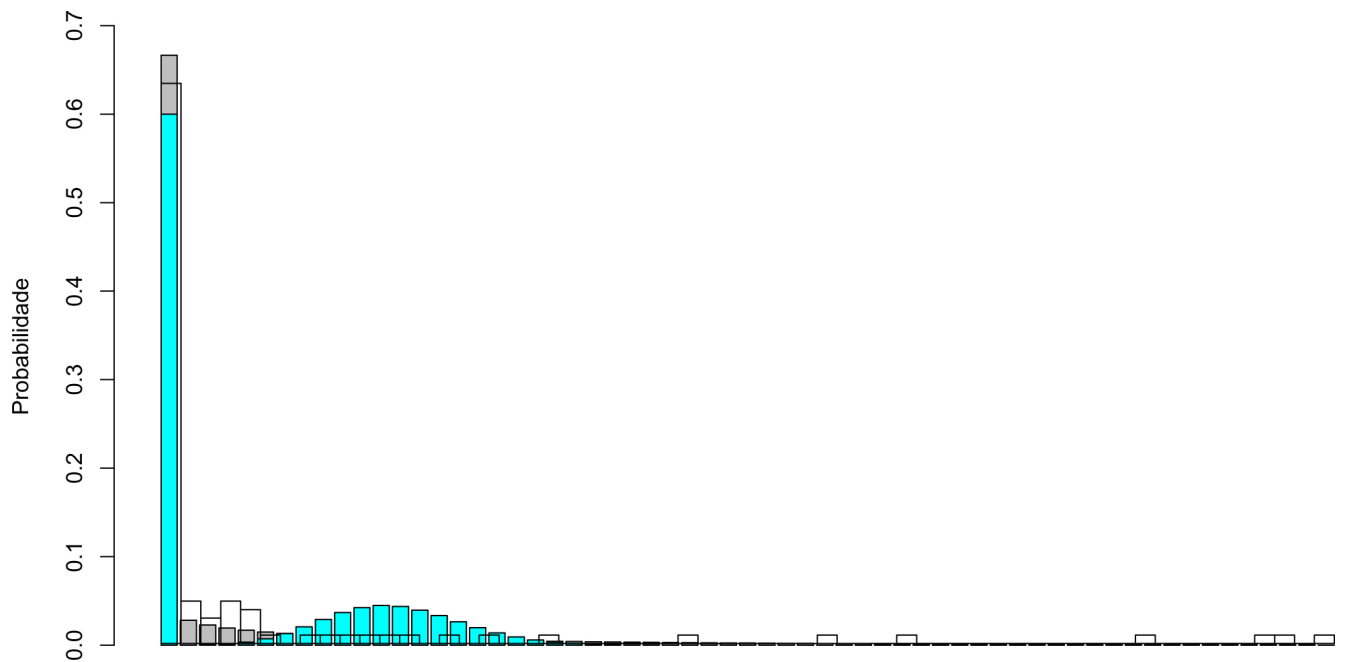


**Figura 4.16:** Histograma dos valores observados de número de ovos e histogramas dos modelos ZIP e ZINB a partir das estimativas da Tabela 4.1 da 16ª Quinzena

- - Histograma dos Dados
- (azul) - Gráfico do modelo ZIP
- (cinza) - Gráfico do modelo ZINB

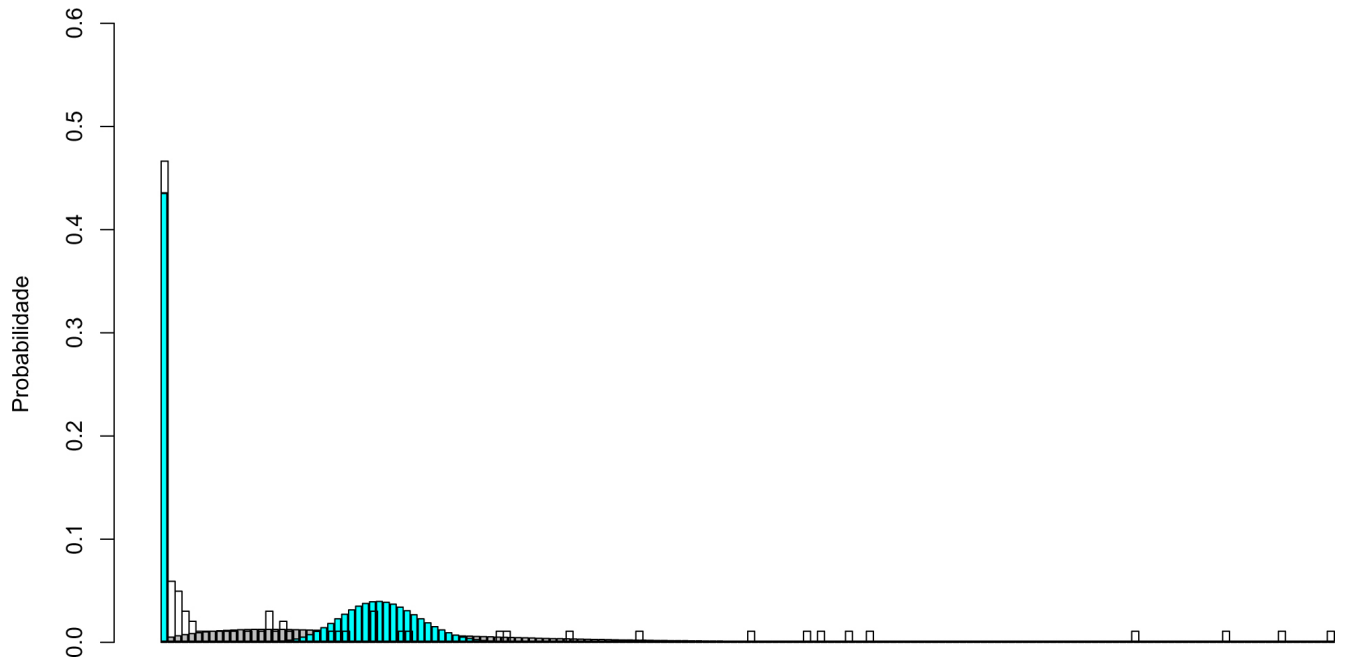


**Figura 4.17:** Histograma dos valores observados de número de ovos e histogramas dos modelos ZIP e ZINB a partir das estimativas da Tabela 4.1 da 17ª Quinzena

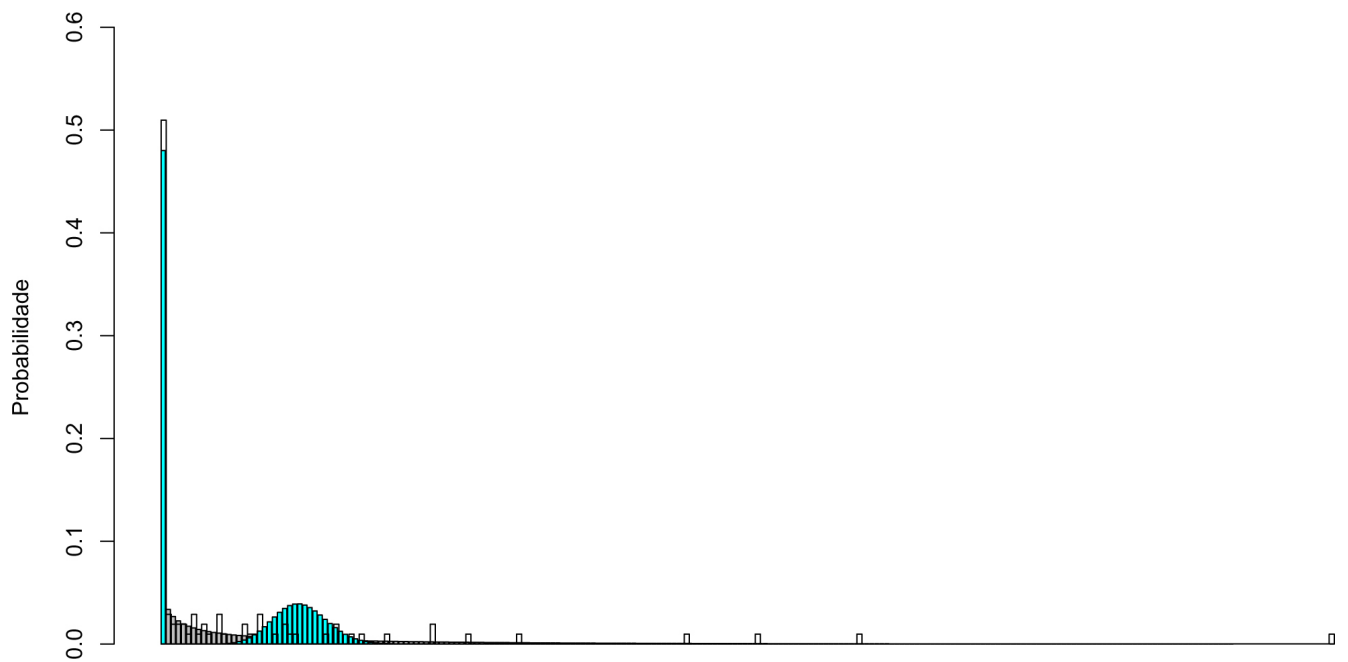


**Figura 4.18:** Histograma dos valores observados de número de ovos e histogramas dos modelos ZIP e ZINB a partir das estimativas da Tabela 4.1 da 18ª Quinzena

- - Histograma dos Dados
- (Ciano) - Gráfico do modelo ZIP
- (Cinza) - Gráfico do modelo ZINB

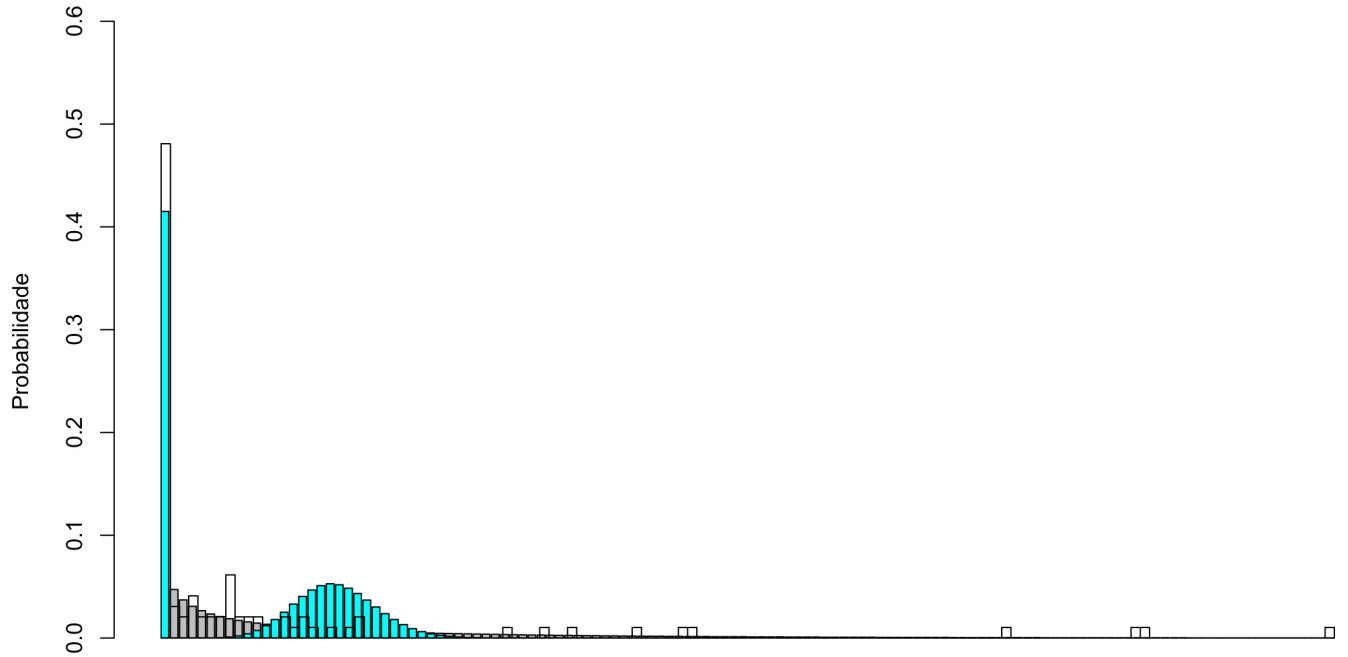


**Figura 4.19:** Histograma dos valores observados de número de ovos e histogramas dos modelos ZIP e ZINB a partir das estimativas da Tabela 4.1 da 19ª Quinzena

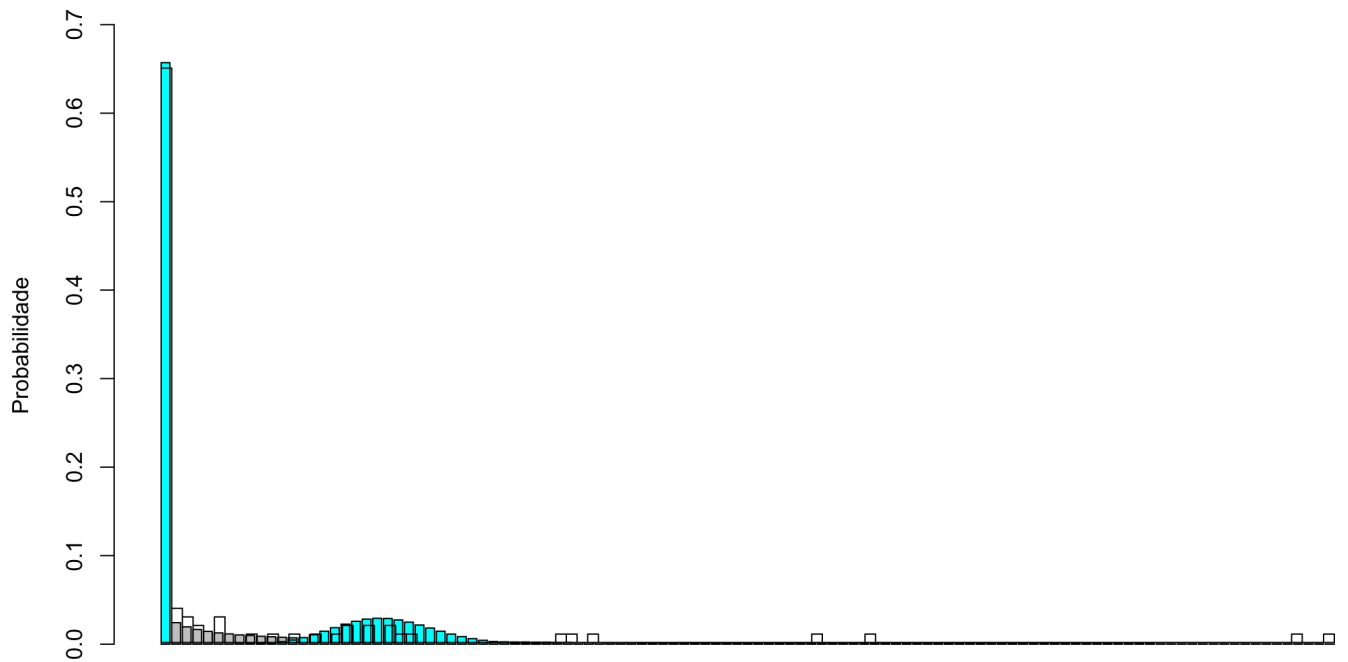


**Figura 4.20:** Histograma dos valores observados de número de ovos e histogramas dos modelos ZIP e ZINB a partir das estimativas da Tabela 4.1 da 20ª Quinzena

- - Histograma dos Dados
- (verde) - Gráfico do modelo ZIP
- (cinza) - Gráfico do modelo ZINB



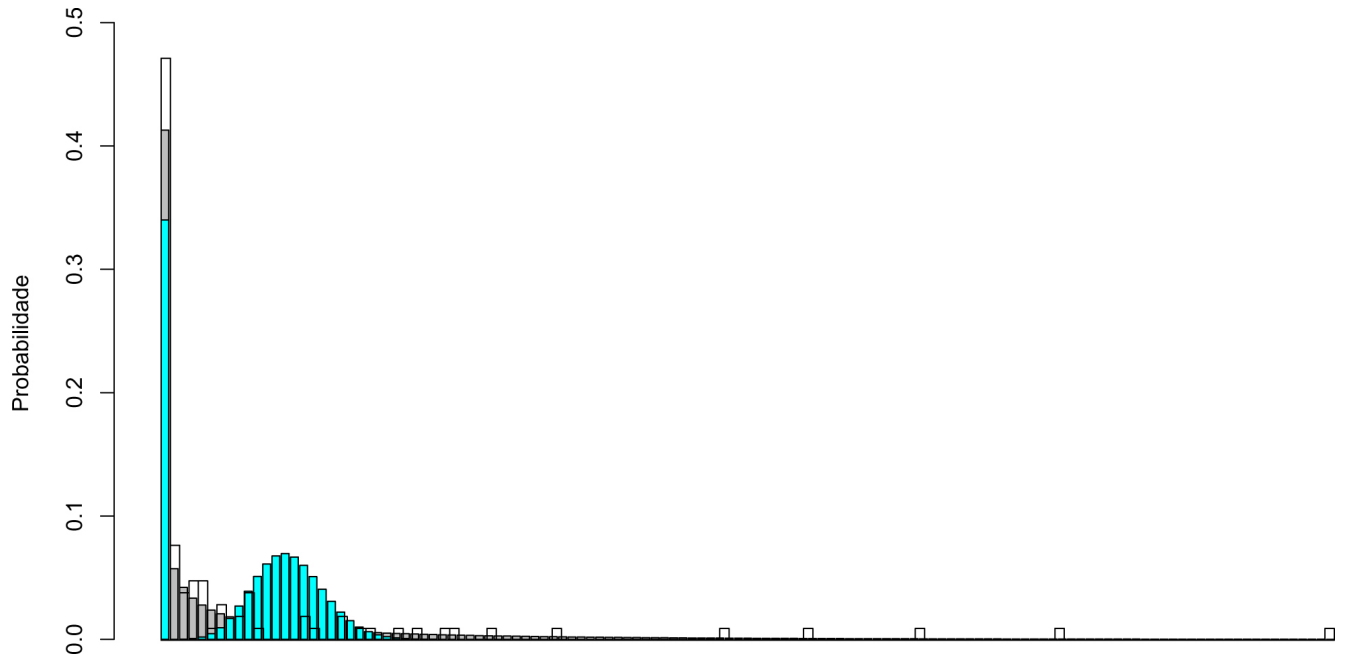
**Figura 4.21:** Histograma dos valores observados de número de ovos e histogramas dos modelos ZIP e ZINB a partir das estimativas da Tabela 4.1 da 21<sup>a</sup> Quinzena



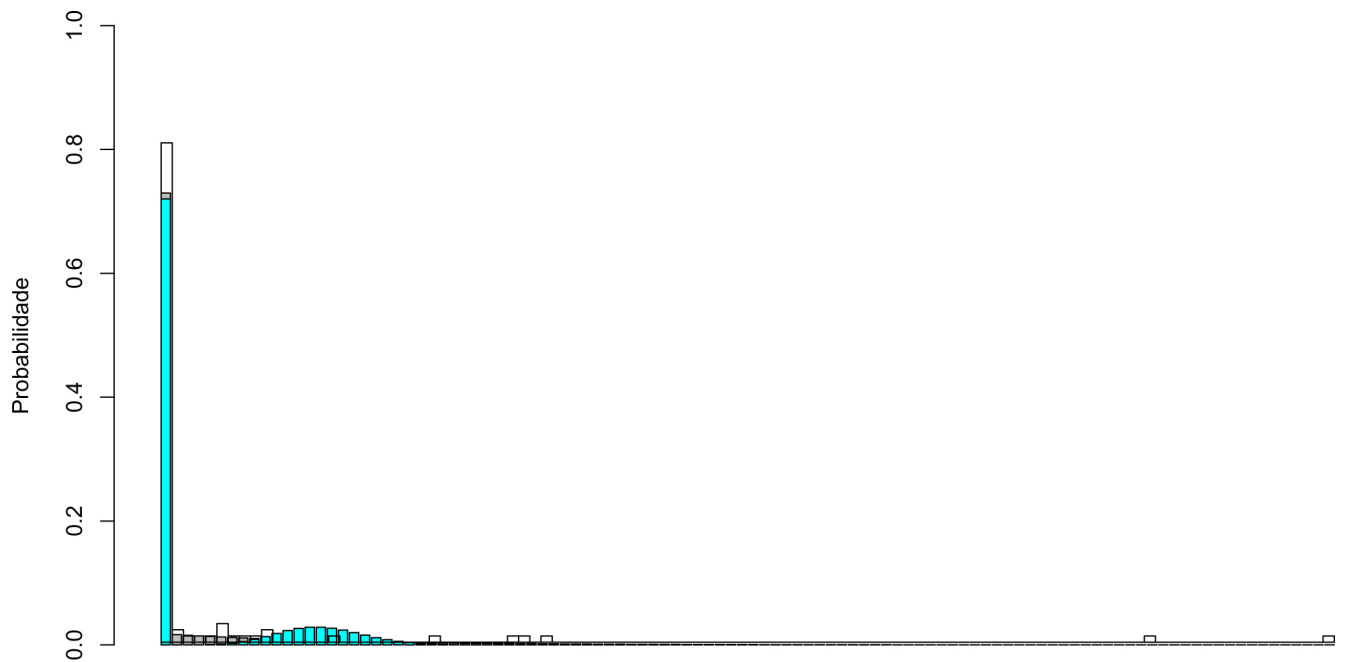
**Figura 4.22:** Histograma dos valores observados de número de ovos e histogramas dos modelos ZIP e ZINB a partir das estimativas da Tabela 4.1 da 22<sup>a</sup> Quinzena

- - Histograma dos Dados
- - Gráfico do modelo ZIP
- - Gráfico do modelo ZINB



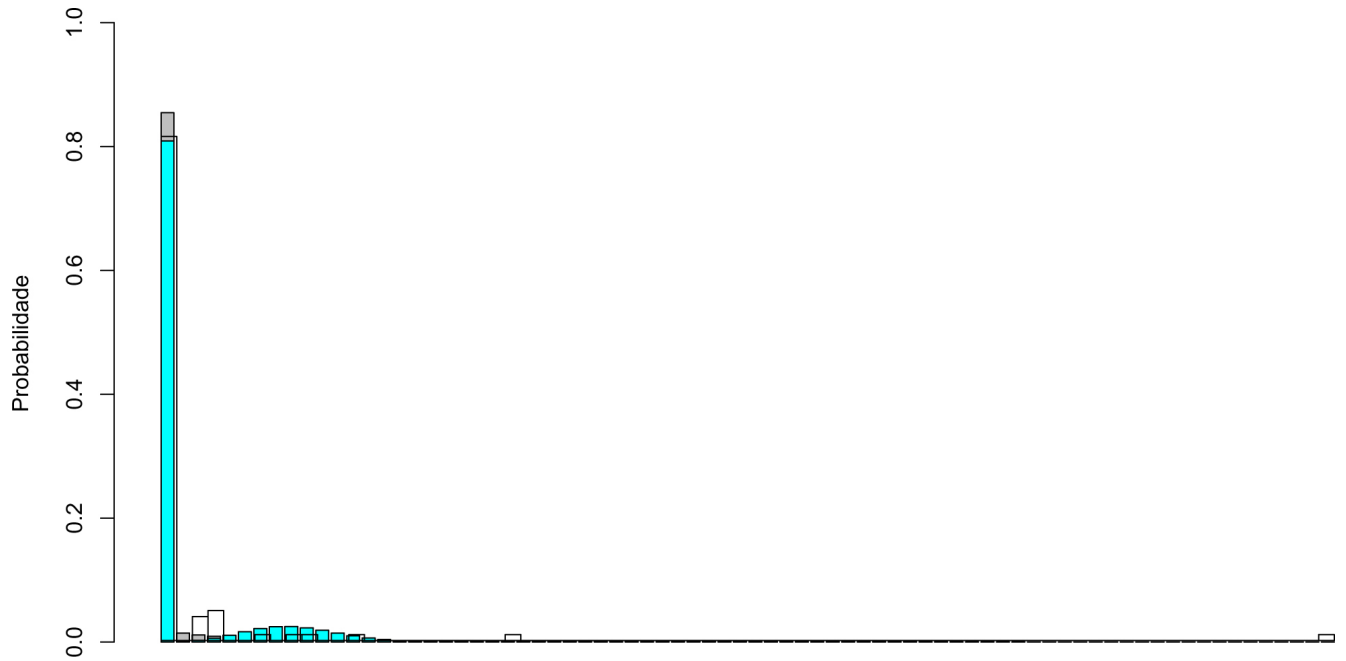


**Figura 4.23:** Histograma dos valores observados de número de ovos e histogramas dos modelos ZIP e ZINB a partir das estimativas da Tabela 4.1 da 23ª Quinzena

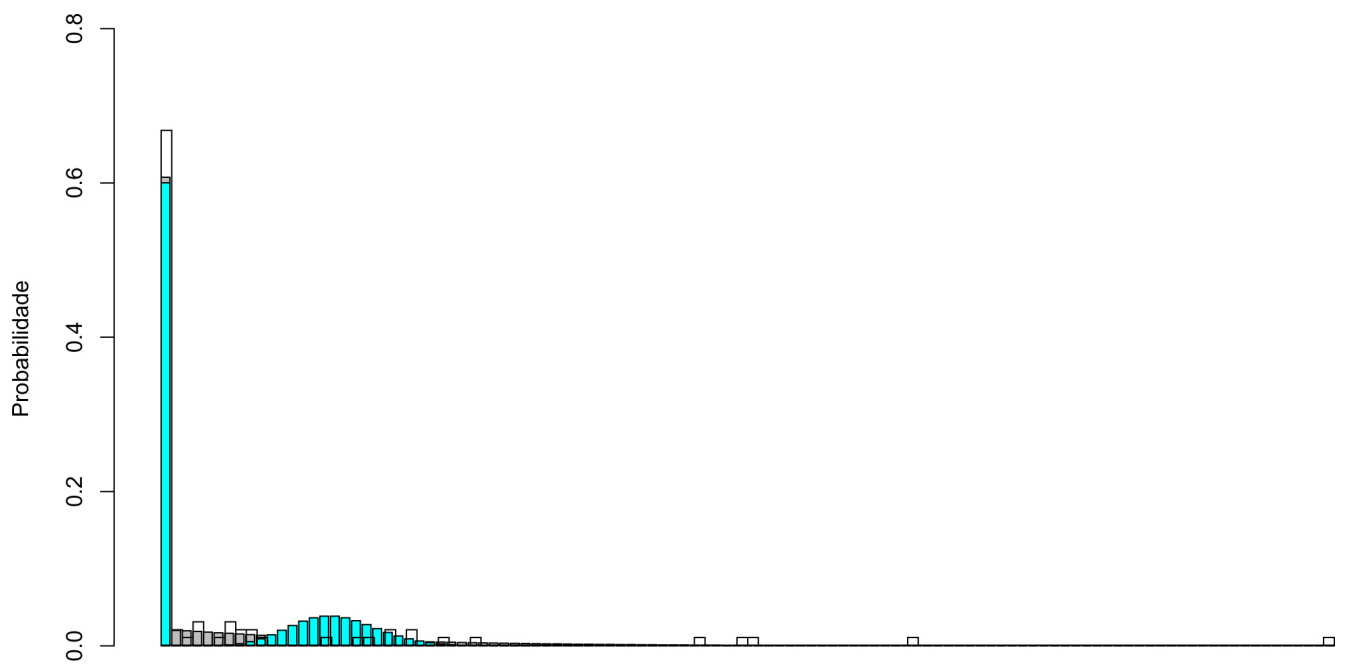


**Figura 4.24:** Histograma dos valores observados de número de ovos e histogramas dos modelos ZIP e ZINB a partir das estimativas da Tabela 4.1 da 24ª Quinzena

- - Histograma dos Dados
- - Gráfico do modelo ZIP
- - Gráfico do modelo ZINB



**Figura 4.25:** Histograma dos valores observados de número de ovos e histogramas dos modelos ZIP e ZINB a partir das estimativas da Tabela 4.1 da 25ª Quinzena



**Figura 4.26:** Histograma dos valores observados de número de ovos e histogramas dos modelos ZIP e ZINB a partir das estimativas da Tabela 4.1 da 26ª Quinzena

- - Histograma dos Dados
- (Azul) - Gráfico do modelo ZIP
- (Cinza) - Gráfico do modelo ZINB

Nota-se que nos histogramas dos valores observados do número de ovos, para todas as quinzenas, existe uma grande variabilidade que os modelos ZIP e ZINB não conseguem acompanhar, como pode ser visto nos resultados apresentados na Tabela 4.1. No entanto, o modelo Binomial Negativo Inflacionado de Zeros (ZINB) mostra uma semelhança maior com o comportamento dos histogramas dos dados do número de ovos, comparativamente com o modelo Poisson Inflacionado de Zeros (ZIP), indicando assim, que apenas uma análise das medidas descritivas não é suficiente para direcionar uma escolha do modelo a ser ajustado.

## 4.7 Modelo com Covariáveis

O banco de dados utilizado para estudar o modelo com covariáveis foi elaborado no Programa Microsoft Excel com dados diários das temperaturas máxima, médias e mínimas, das umidades relativas do ar máximas, médias e mínimas, da pluviosidade e do número de ovos. O banco de dados contém também, a indicação das quadras trabalhadas, as datas em que foram colocadas e em que foram retiradas as armadilhas (ovitrampa) e as medidas das variáveis climáticas até 21 dias antes de colocarem as armadilhas. Foram coletadas medidas quinzenais do número de ovos, entre outubro de 2003 a setembro de 2004, com colocação das armadilhas na primeira semana e retirada na segunda.

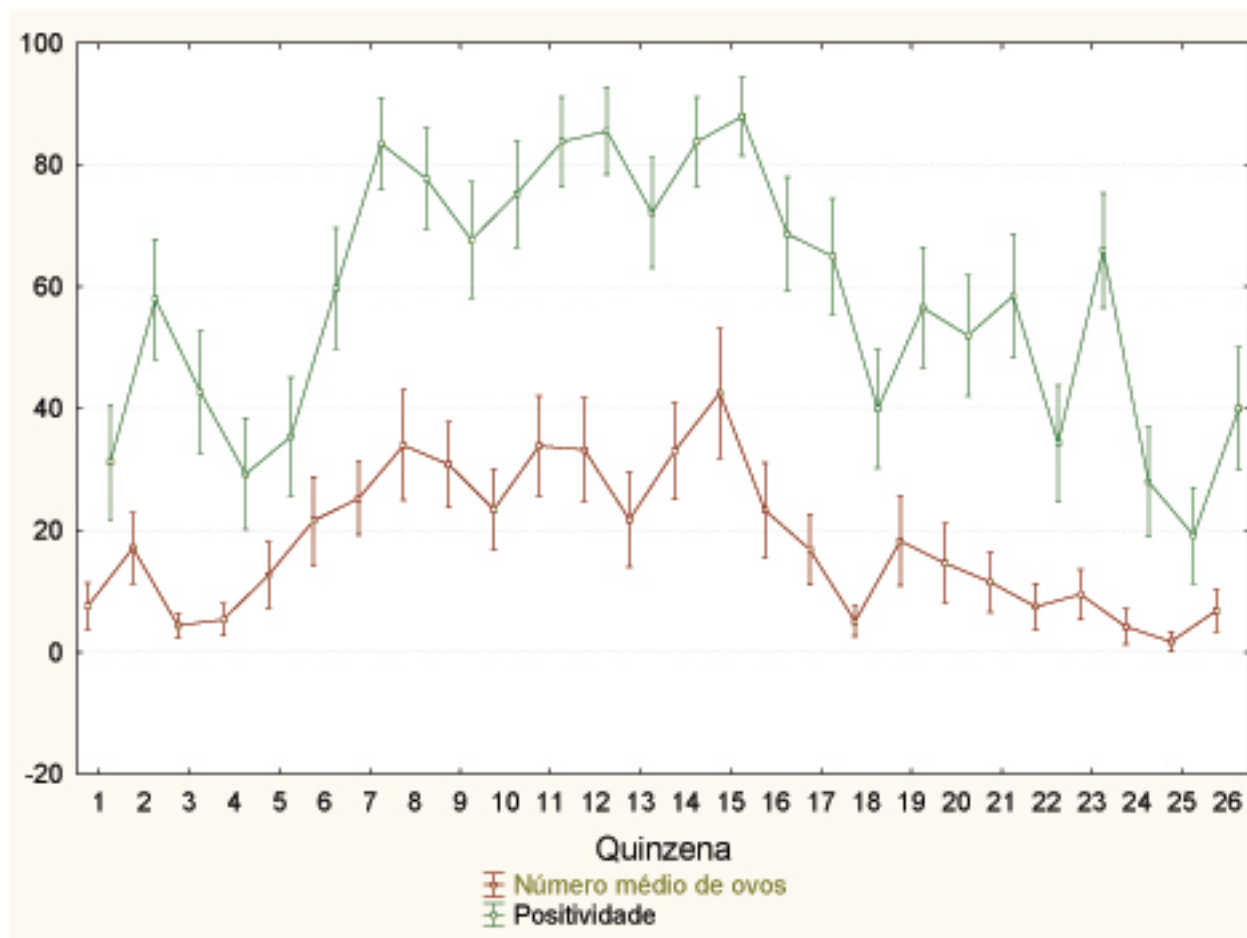
Para análise estatística do número de ovos postos pelo mosquito *Aedes aegypti* em função das covariáveis climáticas, alguns modelos são ajustados, discutidos e comparados, sob a abordagem do modelo Binomial Negativo Inflacionados de Zeros. As análises foram feitas usando o programa Stata, seguindo o procedimento citado na Seção 3.5. A seguir, apresentam-se a descrição dos modelos e os resultados obtidos.

### 4.7.1 Descrição dos Modelos

Barbosa (2006) fez um estudo sobre o comportamento do número médio de ovos por armadilha (número total de ovos coletados dividido pelo número de armadilhas trabalhadas) com a positividade das armadilhas (número de palhetas com a presença de ovos vezes 100 dividido pelo número de armadilhas trabalhadas) e, observou um comportamento diferenciado em três períodos.

O primeiro período, corresponde à quinzena de 1 a 6 (13 de outubro a 24 de dezembro), o segundo, da quinzena 7 a 18 (29 de dezembro a 7 de junho), e o terceiro, da quinzena 19 a 26 (14 de junho a 27 de setembro). Esses períodos serão respectivamente chamados de Período 1, Período 2 e Período 3. Barbosa (2006) também conclui que os períodos de outubro a dezembro e de junho a setembro foram identificados como os mais propícios para utilização da ovitrampa como medidas de vigilância entomológica, pois, existe correlação entre a positividade e a existência de ovos e, o período de dezembro a junho como ideal para a realização de medidas larvárias, ou seja, como existem muitos recipientes de ovos positivos, a contagem de larvas se torna mais eficaz que a contagem de ovos para indicar uma possível epidemia.

A Figura 4.27 descreve o comportamento da positividade das armadilhas e do número médio de ovos, onde nota-se a diferença entre os três períodos.



**Figura 4.27:** Positividade e número médio de ovos das armadilhas de oviposição segundo quinzenas e limites superiores e inferiores dos intervalos de 95% de confiança no período de outubro de 2003 a setembro de 2004

Com base nestes resultados o banco de dados foi dividido nos períodos citados acima, pois, uma questão importante e de interesse neste trabalho é saber qual(is) período(s) do ano é mais indicado para o uso da ovitrampa, e o mais indicado para modelar o número de ovos, e, se os modelos nos três períodos são diferentes.

Para os três períodos foram considerados quatro modelos a serem ajustados, porém, em todos eles o objetivo foi de explicar a variável resposta  $y_i$ , número de ovos, em função das variáveis climáticas: temperaturas máximas, médias e mínimas, das umidades relativas do ar máximas, médias e mínimas e da pluviosidade.

A partir de discussões com biólogos, epidemiologistas e pesquisadores do mosquito *Aedes aegypti* e da dengue foram elaborados os potenciais modelos, considerando o ciclo de vida do mosquito. Tais modelos são descritos como segue.

### Modelo 1

O Modelo 1 tem por finalidade verificar se o período em que ficou exposta a armadilha tem influência na postura do número de ovos. Dessa forma, os valores das covariáveis utilizadas neste modelo, são referentes ao intervalo de tempo entre o dia que foram colocadas as armadilhas e o dia em que as mesmas foram retiradas. Foi considerado o valor máximo, o valor mínimo e o valor médio da temperatura e umidade obtido em cada intervalo que ficaram expostas as ovitrampas.

### Modelo 2

Neste modelo, o interesse foi predizer se os valores das covariáveis referentes à 3 dias antes de colocarem as armadilhas têm mais influência para o número de ovos do que as covariáveis referentes à 3 dias antes de retirarem as armadilhas. Dessa forma, o Modelo 2 é composto pelas covariáveis referentes à temperatura, umidade e pluviosidade, observadas 3 dias antes de serem colocadas as armadilhas e 3 dias antes de serem retiradas as armadilhas.

### Modelo 3

Um outro modelo considerado teve como covariáveis as medidas observadas de temperatura, umidade e pluviosidade coletadas um dia antes de serem colocadas as armadilhas.

### Modelo 4

O último modelo, teve como covariáveis as medidas observadas de temperatura, umidade e pluviosidade coletadas um dia antes de serem retiradas as armadilhas.

Na Tabela 4.6 encontra-se as covariáveis correspondente aos quatro modelos citados acima.

**Tabela 4.6:** Covariáveis existentes nos quatro modelos analisados

Modelos	Covariáveis
Modelo 1	Tmax, Tmin, Tmed, Umax, Umin, Umed
Modelo 2	Tmax3, Tmin3, Tmed3, Chuv3, Umax3, Umin3, Umed3 Tmax10, Tmin10, Tmed10, Chuv10, Umax10, Umin10, Umed10
Modelo 3	Tmax8, Tmin8, Tmed8, Chuv8, Umax8, Umin8, Umed8
Modelo 4	Tmax1, Tmin1, Tmed1, Chuv1, Umax, Umin1, Umed1

No Modelo 1,  $T_{max}$ , denota o valor máximo das temperaturas máximas obtida no intervalo em que ficaram expostas as armadilhas,  $T_{min}$ , o valor mínimo das temperaturas mínimas obtida no intervalo em que ficaram expostas as armadilhas,  $T_{med}$ , é a média das temperaturas médias obtida no intervalo em que ficaram expostas as armadilhas,  $U_{max}$ , o valor máximo das umidades máximas obtida no intervalo em que ficaram expostas as armadilhas,  $U_{min}$ , o valor mínimo das

umidades mínimas obtida no intervalo em que ficaram expostas as armadilhas e  $U_{med}$ , a média das umidades médias obtida no intervalo em que ficaram expostas as armadilhas.

No Modelo 2,  $T_{max3}$ , corresponde à temperatura máxima obtida três dias antes de retirar as armadilhas,  $T_{min3}$ , a temperatura mínima obtida três dias antes de retirar as armadilhas,  $T_{med3}$ , a temperatura média obtida três dias antes de retirar as armadilhas,  $U_{max3}$ , a umidade máxima três dias antes de retirar as armadilhas,  $U_{min3}$ , a umidade mínima três dias antes de retirar as armadilhas,  $U_{med3}$ , a umidade média três dias antes de retirar as armadilhas e  $Chuv3$ , a medida de chuva três dias antes de retirar as armadilhas,  $T_{max10}$ , corresponde à temperatura máxima obtida três dias antes de colocar as armadilhas,  $T_{min10}$ , a temperatura mínima obtida três dias antes de colocar as armadilhas,  $T_{med10}$ , a temperatura média obtida três dias antes de colocar as armadilhas,  $U_{max10}$ , a umidade máxima três dias antes de colocar as armadilhas,  $U_{min10}$ , a umidade mínima três dias antes de colocar as armadilhas,  $U_{med10}$ , a umidade média três dias antes de colocar as armadilhas e  $Chuv10$ , a medida de chuva três dias antes de colocar as armadilhas.

No Modelo 3,  $T_{max8}$ , indica a temperatura máxima obtida um dia antes de colocar as armadilhas,  $T_{min8}$ , a temperatura mínima obtida um dia antes de colocar as armadilhas,  $T_{med8}$ , a temperatura média obtida um dia antes de colocar as armadilhas,  $U_{max8}$ , a umidade máxima um dia antes de colocar as armadilhas,  $U_{min8}$ , a umidade mínima um dia antes de colocar as armadilhas,  $U_{med8}$ , a umidade média um dia antes de colocar as armadilhas e  $Chuv8$ , a medida de chuva um dia antes de colocar as armadilhas.

No Modelo 4,  $T_{max1}$  denota a temperatura máxima obtida um dia antes de retirar as armadilhas,  $T_{min1}$ , a temperatura mínima obtida um dia antes de retirar as armadilhas,  $T_{med1}$ , a temperatura média obtida um dia antes de retirar as armadilhas,  $U_{max1}$ , a umidade máxima um dia antes de retirar as armadilhas,  $U_{min1}$ , a umidade mínima um dia antes de retirar as armadilhas,  $U_{med1}$ , a umidade média um dia antes de retirar as armadilhas e  $Chuv1$ , a medida de chuva um dia antes de retirar as armadilhas.

Os quatro modelos citados acima foram ajustados igualmente, para os três períodos analisados em questão, considerando, que a variável resposta  $Y$ , número de ovos posto pelo *Aedes aegypti*, seguisse uma distribuição Binomial Negativa Inflacionada de Zeros (ZINB).

A escolha de um modelo adequado para explicar o número de ovos para cada período foi feita através do critério da parcimônia e do procedimento *Stepwise*. Primeiramente foram ajustados os quatro modelos separadamente para os três períodos e depois, identificou-se as covariáveis que foram significativas em cada modelo. Em seguida, foi construído para cada período, um outro modelo (Modelo 5), que contém todas as covariáveis que foram significativas nos quatro modelos ajustados. Por fim, o Modelo 5 é ajustado em cada período e através do procedimento *Stepwise*, constrói-se o Modelo 6, em que mantêm-se apenas as covariáveis que foram significativas em cada parcela do modelo (parte do modelo de regressão binomial negativo e parte inflacionada).

## 4.7.2 Resultados e Discussões

Nesta seção são apresentados os resultados das estimativas dos parâmetros, testes e medidas de ajuste do Modelo 6 para cada período. As tabelas trazem os resultados das estimativas do modelo ZINB, com covariáveis, sendo que, o primeiro bloco apresenta os coeficientes, testes e medidas de ajuste que correspondem ao modelo de regressão binomial negativo. O segundo bloco mostra os coeficientes, testes e medidas de ajuste que correspondem à parte inflacionada.

### Modelo do Período 1: Outubro a Dezembro

Neste período, o número total de observações é de 584 com 254 observações diferentes de zero e 334 observações iguais a zero. A covariável  $U_{med}$  corresponde a média das umidades médias obtida no intervalo em que ficaram expostas as armadilhas. Como no Modelo 6 a única covariável significativa para esse período foi a  $U_{med}$ , a Tabela 4.7, traz o ajuste do modelo ZINB apenas com essa covariável.

**Tabela 4.7:** Resultado das estimativas do Modelo 6 para o Período 1

y	Coef. $\beta's$	Erro	Z	$P >  Z $	$e^b$
$U_{med}$	0.1017	0.0423	2.40	0.016	1.1070
cons	-0.6115	1.5796	-0.39	0.699	
inflação	Coef. $\gamma's$				
cons	0.0834	0.1089	0.77	0.444	

### Modelo do Período 2: Dezembro a Junho

No Período 2, o número total de observações é de 1180 com 876 observações diferentes de zero e 304 observações iguais a zero. As covariáveis que compõem esse modelo são:  $T_{max10}$  temperatura máxima de 3 dias antes de colocar as armadilhas;  $T_{med10}$ , temperatura média de 3 dias antes de colocar as armadilhas;  $U_{max1}$ , umidade máxima um dia antes de retirar as armadilhas;  $U_{med1}$ , umidade média um dia antes de retirar as armadilhas e  $T_{med}$ , temperatura média obtida no intervalo em que ficaram expostas as armadilhas. A Tabela 4.8 apresenta o resultado do ajuste com as covariáveis acima citadas.

**Tabela 4.8:** Resultado das estimativas do Modelo 6 para o Período 2

y	Coef. $\beta's$	Erro	Z	$P >  Z $	$e^b$
<i>Tmax10</i>	-0.1004	0.0334	-3.00	0.003	0.9044
<i>Tmed10</i>	0.1776	0.0408	4.35	0.000	1.1943
cons	2.1547	0.4283	5.03	0.000	
inflação	Coef. $\gamma's$				
<i>Umax1</i>	0.0901	0.0289	3.11	0.002	1.0943
<i>Umed1</i>	-0.0669	0.0210	-3.18	0.001	0.9352
<i>Tmed</i>	-0.1733	0.0314	-5.51	0.000	0.8408
cons	-0.4194	1.8362	-0.23	0.819	

### Modelo do Período 3: Junho a Setembro

No Período 3, o número total de observações é de 786 com 348 observações diferentes de zero e 438 observações iguais a zero. As covariáveis são: *Umax8*, umidade máxima um dia antes de colocar as armadilhas e *Umin*, umidade mínima obtida no intervalo em que ficaram expostas as armadilhas. A Tabela 4.9 mostra o ajuste feito com as covariáveis citadas acima.

**Tabela 4.9:** Resultado das estimativas do Modelo 6 para o Período 3

y	Coef. $\beta's$	Erro	Z	$P >  Z $	$e^b$
<i>Umax8</i>	0.0212	0.0088	2.41	0.016	1.0214
<i>Umin</i>	0.0390	0.0140	2.79	0.005	1.0397
cons	-0.3327	0.5247	-0.63	0.526	
inflação	Coef. $\gamma's$				
cons	-11.3747	239.9291	-0.05	0.962	

Observa-se que nos modelos correspondente aos Períodos 1 e 3, a parte inflacionada é modelada apenas por uma constante, e, no modelo correspondente ao Período 2 a parte inflacionada é modelada pelas covariáveis que indicam umidade máxima e média observadas um dia antes de retirar as armadilhas (*Umax1* e *Umed1*) e, pela média das temperaturas médias obtida no intervalo em que ficaram expostas as armadilhas (*Tmed*). Já na parte que envolve o modelo de regressão binomial negativo, o Período 1 foi explicado pela média das umidades médias obtida no intervalo em que ficaram expostas as armadilhas (*Umed*), no Período 2, pelas temperaturas máximas e médias observadas três dias antes de colocar as armadilhas (*Tmax10* e *Tmed10*), e, no Período 3, pela umidade máxima observada um dia antes de colocar as armadilhas (*Umax8*) e, pelo mínimo das umidades mínimas obtida no intervalo em que ficaram expostas as armadilhas (*Umin*).



### Medidas de Ajuste

As Tabelas 4.10 , 4.11 e 4.12 a seguir, trazem algumas medidas de ajuste para os modelos que foram analisados nos três períodos estudados.

**Tabela 4.10:** Medidas de Ajuste para os Modelos do Período 1

Período 1	$n^0$ de parâmetros no modelo	Deviance	AIC	BIC
Modelo 1	5	2871.34	4.941	-804.092
Modelo 2	21	2841.25	4.944	-732.261
Modelo 3	11	2858.46	4.939	-778.752
Modelo 4	Nenhuma covariável foi significativa			
Modelo 6	2	2911.64	4.999	-782.895

**Tabela 4.11:** Medidas de Ajuste para os Modelos do Período 2

Período 2	$n^0$ de parâmetros no modelo	Deviance	AIC	BIC
Modelo 1	3	9223.56	7.825	912.473
Modelo 2	5	9222.82	7.828	925.881
Modelo 3	5	9225.34	7.830	928.400
Modelo 4	7	9226.90	7.835	944.101
Modelo 6	6	9205.44	7.815	915.512

**Tabela 4.12:** Medidas de Ajuste para os Modelos do Período 3

Período 3	$n^0$ de parâmetros no modelo	Deviance	AIC	BIC
Modelo 1	11	3649.49	4.976	-1504.062
Modelo 2	Nenhuma covariável foi significativa			
Modelo 3	9	3652.42	4.675	-1514.467
Modelo 4	13	3656.43	4.690	-1483.792
Modelo 6	3	3702.39	4.723	-1504.502

Para os três períodos analisados, o Modelo 6 é considerado o mais indicado para ajustar o número de ovos, pois é o modelo mais parcimonioso e comporta as covariáveis que foram significativas.

---

## Conclusões

---

Neste trabalho foi apresentado o desenvolvimento analítico para os modelos Poisson Inflacionado de Zeros e Binomial Negativo Inflacionado de Zeros, e, com base nesses dois modelos, foram calculados alguns resultados importantes para desenvolver alguns testes e fazer uma análise gráfica para a escolha de uma distribuição adequada com o objetivo de modelar o número de ovos postos pelo mosquito *Aedes aegypti*. Estes resultados, mostraram que a distribuição que apresentou uma semelhança melhor com a do número de ovos foi a distribuição ZINB.

A partir desse fato, foi considerado que a variável resposta número de ovos seguisse uma distribuição ZINB, que foi utilizada para a modelagem do número de ovos em função das covariáveis climáticas. O modelo ZINB foi ajustado para quatro modelos apresentados na Seção 4.7.1, em três períodos distintos, denotados por, Período 1 (outubro a dezembro), Período 2 (dezembro a junho) e Período 3 (junho a setembro), seguindo estudos anteriores apresentado por Barbosa (2006).

O modelo que se mostrou mais adequado ao ajuste ZINB, chamado de Modelo 6, foi comparado entre os três períodos, onde, observou-se que todas as constantes que envolvem a parte inflacionada foram não significativas e, apenas no Período 2, a constante que não envolve a parte inflacionada foi a única significativa.

Os Períodos 1 e 3 mostraram um comportamento semelhante no que diz respeito à modelagem, visto que, no Modelo 6 de ambos os Períodos, as covariáveis estão relacionadas à parte que envolve o modelo de regressão Binomial Negativo e a parte inflacionada é modelada apenas por uma constante. Nos dois períodos, as constantes dos modelos foram não significativas, e as covariáveis são relacionadas à umidade. No Período 1, à umidade média obtida no intervalo em que ficaram expostas as armadilhas, e, no Período 3, a umidade máxima obtida um dia antes de colocar as armadilhas e a umidade mínima obtida no intervalo em que ficaram expostas as armadilhas.

Já no Período 2, as covariáveis envolvidas na parte do modelo de regressão Binomial Negativo, são relacionadas à temperatura máxima e média de três dias antes de serem colocadas as armadilhas. A parte inflacionada no Período 2 tem como covariáveis a temperatura média obtida no intervalo em que ficaram expostas as armadilhas, a umidade média de três dias antes de serem retiradas e a umidade máxima um dia antes de serem colocadas as armadilhas. Apenas a constante da parte inflacionada foi não significativa.

Através dos resultados obtidos, concluímos que o modelo Binomial Negativo Inflacionado de Zeros (ZINB) vem colaborar na explicação do número de ovos para cada período estudado, e, conseqüentemente mostra resultados importantes para a decisão de políticas públicas para uma utópica modificação do controle da dengue.

---

## Perspectivas Futuras

---

---

Para melhorar as medidas de ajustes obtidas nos três períodos, uma alternativa é analisar outras covariáveis que deveriam estar no modelo e não foram observadas. Sendo assim, podemos avaliar outras covariáveis com ou sem interação entre elas e construir outros modelos.

Outra possibilidade que pretendemos fazer a curto prazo, é implementar o ajuste desses modelos no programa R que é um software livre, desenvolvendo programas que consigam estimar valores preditos em toda a amplitude do conjunto de dados, pois, no software utilizado (*Stata*) essa foi uma das limitações encontradas.

Como as medidas das covariáveis ambientais foram colhidas ao longo do tempo, outra idéia é ajustar os modelos com a inclusão de efeitos aleatórios para modelar a dependência no tempo, que poderia favorecer um ajuste mais adequado.

E por fim, podemos trabalhar com outras distribuições inflacionadas ou não, que resolva o problema da superdispersão existente e possam se ajustar melhor a esse conjunto de dados.

# Referências Bibliográficas

---

---

- [1] Akaike, H. (1973). *Information theory and an extension of the maximum likelihood principle*, in B.N. Pretov and F. Csaki, eds., Second International Symposium on Information Theory. Budapest: Akademiai Kaido, p.267-281.
- [2] Barbosa, A. A. C. (2006). *Avaliação do uso de armadilhas de oviposição como método de vigilância entomológica para Aedes aegypti*, 2006. 65p. Dissertação (MSC) - Faculdade de Medicina de São José do Rio Preto.
- [3] Böhning, D.; Dietz, E.; Schlattmann, P.; Mendonça, L.; Kirchner, U. (1999). *The zero-inflated Poisson model and the decayed, missing and filled teeth index in dental epidemiology*. Journal Royal Statistical Society A, v.162, p.195-209.
- [4] Cameron, A. C.; Trivedi, P. K. (1998). *Regression analysis of count data*. New York: Cambridge University Press.
- [5] Cheung, Y. B. (2002). *Zero-inflated models for regression analysis of count data: a study of growth and development*. Statistics in Medicine, v.21, p.1461-1469.
- [6] Cox, D. R.; Hinkley, D. V. (1974). *Theoretical statistics*. London: Chapman and Hall.
- [7] Dantas, C. A. B. (2000). *Probabilidade: um curso introdutório*. Edusp: Editora da Universidade de São Paulo.
- [8] Famoye, F.; Singh, K. P. (2006). *Zero-inflated generalized Poisson regression model with an application to domestic violence data*. Journal of Data Science, v.4, p.117-130.
- [9] F. Chiaravalloti, N. (2004). *Estudo da relação entre as medidas entomológicas para os vetores da dengue obtidas de armadilhas adulticidas, de oviposição e de coleta de adultos, mirassol-são paulo*. Projeto de Pesquisa do departamento de Epidemiologia e Saúde Coletiva da Faculdade de Medicina de São José do Rio Preto.

- [10] Gupta, P. L.; Tripathi, R.C. (1996). *Analysis of zero-adjusted count data*. Computational Statistics & Data Analysis, v.23, p.207-218.
- [11] Hamilton, L. C. (1990). *Statistics with Stata*. California: Pacific Grove.
- [12] James, B.R. (1981). *Probabilidade: um curso em nível intermediário*. IMPA: Projeto Euclides.
- [13] Johnson, N. L.; Kotz, S. (1969). *Discrete Distributions: distributions in statistics*. New York: John Wiley & Sons.
- [14] Jull, S. (2003). *Introduction to Stata 8*. University of Aarhus.
- [15] Jung, B. C.; Jhun, M.; Lee, J. W. (2005). *Bootstrap tests for overdispersion in a zero inflated Poisson regression model*. Biometrics, v.61, p.626-629.
- [16] Khoshgoftaar, T. M.; Szabo, R. M. (2001). *An application of zero-inflated Poisson regression for software fault prediction*. Software e reliability Engineering, 2001, ISSRE 2001, *Proceedings, 12th International Symposium on*, v.66-73.
- [17] Lambert, D. (1992). *Zero-inflated Poisson regression, with an application to defects in manufacturing*. Technometrics, v.34, p.1-14.
- [18] Lewsey, J. D.; Thomson, W. M. (2004). *The utility of the zero-inflated Poisson and zero-inflated negative binomial models: a case study of cross-sectional and longitudinal DMF data examining the effect of socio-economic status*. Community Dentistry and Oral Epidemiology, v.32, p.183-189.
- [19] Long, J. S.; Freese, J. (2001). *Regression models for categorical dependent variables using stata*. Stata Press.
- [20] Meyer, P. L. (1984). *Probabilidade: aplicações à estatística*. Livros Técnicos e Científicos.
- [21] McCulloch, C. E.; Searle, S. R. (2001). *Generalized, linear, and mixed models*. John Wiley e Sons: INC.
- [22] Mood, A. M.; Graybill, F. A.; Boes, D. C. (1974). *Introduction to the theory of statistics*. McGraw-Hill.
- [23] Mwalili, S. M.; Lesaffre, E.; Declerck, D. *The zero-inflated negative binomial regression model with correction for misclassification: an example in caries research*. Technical Report, <http://www.stat.ucl.ac.be/IAP>, v.0462.
- [24] Nelder, J. A.; Wedderburn, R. W. M. (1972). *Generalized linear models*. Journal of the Royal Statistical Society A, v.135, n.3, p.370-384.

- [25] Osanai, C. H.; Travassos da Rosa, A. P. A.; Tang, A. T.; Amaral, R. S.; Passos, A. D. C.; Tauil, P. L. (1983). *Surto de Dengue em Boa Vista-Roraima*. Revista Inst. Medicina Tropical, v.25, p.53-54.
- [26] Paula, G. A. (2004). *Modelos de regressão com apoio computacional*. São Paulo: IME/USP.
- [27] Ridout, M. S.; Hinde, J.; Demétrio, C. G. B. (2001). *A score test for testing a zero-inflated Poisson regression model against zero-inflated negative binomial alternatives*. Biometrics, v.57, p.219-223.
- [28] Ridout, M.; Demétrio, C. G. B.; Hinde, J. (1998). *Models for count data with many zeros*. International Biometric Conference, In: Cape Town, Dezembro.
- [29] Schwarz, G. (1978). *Estimating the dimension of a model*. Annals of Statistics, v.6, p.461-464.
- [30] Statistical Software for Professionals. <http://www.stata.com/>
- [31] Smyth, G. K. (2003). *Pearson's goodness of fit statistic as a score test statistic*. Science and Statistics, v.40.
- [32] Tauil, P. L. (2002). *Aspectos críticos do controle do dengue no brasil*. Cad. Saúde Pública, v.18, n.3, p.867-871.
- [33] van den Broek, J. (1995). *A score test for zero inflation in a Poisson distribution*. Biometrics, v.51, p.738-743.
- [34] Xie, M.; He, B.; Goh, T. N. (2001). *Zero-inflated Poisson model in statistical process control*. Computational Statistics & Data Analysis, v.38, p.191-201.
- [35] Yau, K. W.; Wang, K.; Lee, A. H. (2003). *Zero-inflated negative binomial mixed regression modeling of over-dispersed count data with extra zeros*. Biometrical Journal, v.45, p.437-452.
- [36] Welsh, A. H. (1996). *Aspects of Statistical Inference*. John Wiley e Sons: INC.

## Apêndice

### A.1 EMV para os modelos ZIP e ZINB

O estimador de máxima verossimilhança do modelo ZIP para o parâmetro  $\omega$  é obtido igualando a expressão (3.1.6) a zero, resultando em

$$\begin{aligned}
 \frac{n_0 (1 - e^{-\lambda})}{(1 - \hat{\omega} + \hat{\omega} e^{-\lambda})} &= \sum_{y=1}^{\infty} \frac{n_y}{\hat{\omega}} \\
 \hat{\omega} \left[ n_0 (1 - e^{-\lambda}) + \sum_{y=1}^{\infty} n_y (1 - e^{-\lambda}) \right] &= \sum_{y=1}^{\infty} n_y \\
 \hat{\omega} [(1 - e^{-\lambda})(n_0 + (n - n_0))] &= (n - n_0) \\
 \hat{\omega} [n(1 - e^{-\lambda})] &= n - n_0 \\
 \hat{\omega} &= \frac{n - n_0}{n(1 - e^{-\lambda})} \\
 \hat{\omega} &= \frac{1 - \frac{n_0}{n}}{1 - e^{-\lambda}}. \tag{1.1.1}
 \end{aligned}$$



O estimador de máxima verossimilhança do modelo ZIP para o parâmetro  $\lambda$  é obtido igualando a expressão (3.1.7) a zero, resultando em

$$\begin{aligned}
 \frac{n_0 \hat{\omega} e^{-\hat{\lambda}}}{(1 - \hat{\omega} + \hat{\omega} e^{-\hat{\lambda}})} &= \sum_{y=1}^{\infty} n_y \left( \frac{y - \hat{\lambda}}{\hat{\lambda}} \right) \\
 \frac{n_0 \hat{\omega} e^{-\hat{\lambda}}}{(1 - \hat{\omega} + \hat{\omega} e^{-\hat{\lambda}})} &= \frac{n \bar{y}}{\hat{\lambda}} - (n - n_0) \\
 \hat{\lambda} &= \frac{n \bar{y} (1 - \hat{\omega} + \hat{\omega} e^{-\hat{\lambda}})}{n (1 - \hat{\omega} + \hat{\omega} e^{-\hat{\lambda}}) - n_0 (1 - \hat{\omega})} \\
 \hat{\lambda} &= \frac{n \bar{y}}{n - \frac{n_0 (1 - \hat{\omega})}{\hat{\omega} (e^{-\hat{\lambda}} - 1) + 1}} \\
 \hat{\lambda} &= \frac{n \bar{y}}{\frac{n - n_0}{1 - e^{-\hat{\lambda}}}} = \frac{\bar{y}}{\hat{\omega}}.
 \end{aligned} \tag{1.1.2}$$

O estimador de máxima verossimilhança do modelo ZINB para o parâmetro  $\omega$  é obtido igualando a expressão (3.3.6) a zero, resultando em

$$\begin{aligned}
 \frac{n_0 (1 - \hat{\alpha}^k)}{(1 - \hat{\omega} + \hat{\omega} \hat{\alpha}^k)} &= \sum_{y=1}^{\infty} \frac{n_y}{\hat{\omega}} \\
 \hat{\omega} n_0 (1 - \hat{\alpha}^k) &= \sum_{y=1}^{\infty} n_y - \hat{\omega} \sum_{y=1}^{\infty} n_y + \hat{\omega} \alpha^k \sum_{y=1}^{\infty} n_y \\
 \hat{\omega} [n_0 - n_0 \alpha^k + n - n_0 - n \alpha^k - n_0 \alpha^k] &= n - n_0 \\
 \hat{\omega} &= \frac{1 - \frac{n_0}{n}}{1 - \hat{\alpha}^k}.
 \end{aligned} \tag{1.1.3}$$

## A.2 Logaritmo da função de verossimilhança do modelo de regressão ZIP

$$\begin{aligned}
l(\theta_i, \lambda_i, y_i) &= \log \mathcal{L}(y_i, \theta_i, \lambda_i) \\
&= \log \left[ \prod_{i=1}^n f(y_i, \theta_i, \lambda_i) \right] \\
&= \sum_{i=1}^n \left\{ \log \left[ \frac{1}{1 + \theta_i} (\theta_i + e^{-\lambda_i}) \right] I_{y_i=0} + \log \left[ \frac{1}{1 + \theta_i} \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} \right] I_{y_i>0} \right\} \\
&= \sum_{i=1}^n \left\{ \log \left( \frac{1}{1 + \theta_i} \right) I_{y_i=0} + \log(\theta_i + e^{-\lambda_i}) I_{y_i=0} + \log \left( \frac{1}{1 + \theta_i} \right) I_{y_i>0} - \lambda_i I_{y_i>0} + \right. \\
&\quad \left. + y_i \log(\lambda_i) I_{y_i>0} - \log(y_i!) I_{y_i>0} \right\} \\
&= - \sum_{i=1}^n \log(1 + \theta_i) + \sum_{i:y_i=0} \log(\theta_i + e^{-\lambda_i}) - \sum_{i:y_i>0} \log(y_i!) + \sum_{i:y_i>0} [y_i \log(\lambda_i) - \lambda_i]
\end{aligned}$$

## A.3 Logaritmo da função de verossimilhança do modelo de regressão ZINB

$$\begin{aligned}
l(\theta_i, \alpha_i, k; y_i) &= \log \mathcal{L}(y_i; \theta_i, \alpha_i, k) \\
&= \log \left[ \prod_{i=1}^n f(y_i, \theta_i, \alpha_i, k) \right] \\
&= \sum_{i=1}^n \left\{ \log \left[ \frac{1}{1 + \theta_i} \left( \theta_i + \left( \frac{k}{k + \lambda_i} \right)^k \right) \right] I_{y_i=0} + \log \left[ \frac{1}{1 + \theta_i} \frac{\Gamma(y_i + k)}{\Gamma(k) y_i!} \left( \frac{k}{k + \lambda_i} \right)^k \frac{\lambda_i^{y_i}}{(k + \lambda_i)^{y_i}} \right] I_{y_i>0} \right\} \\
&= \sum_{i=1}^n \left\{ \log \left( \frac{1}{1 + \theta_i} \right) I_{y_i=0} + \log \left( \theta_i + \left( \frac{k}{k + \lambda_i} \right)^k \right) I_{y_i=0} + \log \left( \frac{1}{1 + \theta_i} \right) I_{y_i>0} \right. \\
&\quad \left. + \log \left( \frac{\Gamma(y_i + k)}{\Gamma(k) y_i!} \right) I_{y_i>0} + k \log \left( \frac{k}{k + \lambda_i} \right) I_{y_i>0} + y_i \log \left( \frac{\lambda_i}{k + \lambda_i} \right) I_{y_i>0} \right\} \\
&= - \sum_{i=1}^n \log(1 + \theta_i) + \sum_{i:y_i=0} \log \left[ \theta_i + \left( \frac{k}{k + \lambda_i} \right)^k \right] + \sum_{i:y_i>0} \left[ k \log \left( \frac{k}{k + \lambda_i} \right) + y_i \log \left( \frac{\lambda_i}{k + \lambda_i} \right) \right] \\
&\quad - \sum_{i:y_i>0} [\log \Gamma(y_i + 1) + \log \Gamma(k) - \log \Gamma(y_i + k)].
\end{aligned}$$

# Apêndice

## B.1 Programa - EMV ZIP

Nesta seção encontra-se o programa do método iterativo para calcular os estimadores de máxima verossimilhança do modelo ZIP através do programa Maple<sup>1</sup>.

O programa começa com um chute inicial, o do método dos momentos para  $\omega$  e com os valores para número de zeros ( $n_0$ ), tamanho da amostra ( $n$ ) e a média  $Y$  para cada quinzena. Abaixo está o programa para a 1ª *Quinzena*, sendo que para as outras quinzenas o programa difere apenas dos valores iniciais de  $\omega$ ,  $n_0$ ,  $n$  e  $Y$ .

### 1 Quinzena:

```
> omega:=0.14:          n0:=66:          n:=96:          Y:=7.542:
> lambda:=Y/omega;
> omega:=(1-n0/n)/(1-exp(-lambda));
> parada1:=abs(aux1-lambda):
> parada2:=abs(aux2-omega):
> aux1:=lambda:
> aux2:=omega:
> end do:
> print('lambda'=aux1,'omega'=aux2);
> restart:
```

$\lambda = 24.13440000, \omega = 0.3125000000$

<sup>1</sup>Waterloo Maple Inc., Ontario, Canada

## B.2 Programa - EMV ZINB

Nesta seção encontra-se o programa do método iterativo para calcular os estimadores de máxima verossimilhança do modelo ZINB através do programa Maple.

O programa começa com um chute inicial, o do método dos momentos para  $\omega$ ,  $k$  e  $\lambda$  e com os valores para número de zeros ( $n_0$ ), tamanho da amostra ( $n$ ) e a média  $Y$  para cada quinzena. Para as outras quinzenas o programa é semelhante.

### Quinzena 1:

```
> k:=.9624912809: lambda:=26.82753063: omega:=.2811167045:
> n0:=66:n:=96:Y:=7.542:nM:=25:
> omega:=(1-n0/n)/(1-(k/(k+lambda))^k):
> lambda:='lambda':
> expr1:=-n0*omega*(k/(k+lambda))^k*\
> k/(k+lambda)/(1-omega+omega*(k/(k+lambda))^k)\
> +n*Y*k/lambda/(k+lambda)-k*(n-n0)/(k+lambda):
> lambda:=fsolve(expr1,lambda):
> k:='k':
> soma:=sum(M[2,i]*evalf(Psi(M[1,i]+k)),i=2..nM):
> expr2:=(n0*omega*(k/(k+lambda))^k*(ln(k/(k+lambda))+1/(k+lambda)-k/
> +(((soma)*(k+lambda)-n*Y-evalf(Psi(k))*(n-n0)*(k+lambda)+\
> ln(k/(k+lambda))*(n-n0)*(k+lambda)+lambda*(n-n0))/(k+lambda));
> k:=fsolve(expr2,k):
> aux1:=k:
> aux2:=omega:
> aux3:=lambda:
> end do:
> print('k'=aux1,'omega'=aux2,'lambda'=aux3);
> restart:
> k = .8897422652, omega = .3302905142, lambda = 22.83444324
```

## B.3 Algoritmo IC Assintótico

Seja o vetor de parâmetros  $\beta = (\lambda, \omega, k)$ , podemos construir o intervalo de confiança assintótico para o parâmetro  $\lambda$  do modelo Binomial Negativa Inflacionado de Zeros seguindo os seguintes passos:

i) Calcular os EMV ( $\hat{\beta}$ ) para a amostra original;

ii) Calcular a matriz de informação de Fisher observada, que é dada por

$$-J(\hat{\beta}) = \begin{pmatrix} \frac{\partial^2 l(\beta)}{\partial \hat{\lambda}^2} & \frac{\partial^2 l(\beta)}{\partial \hat{\lambda} \partial \hat{\omega}} & \frac{\partial^2 l(\beta)}{\partial \hat{\lambda} \partial \hat{k}} \\ \frac{\partial^2 l(\beta)}{\partial \hat{\omega} \partial \hat{\lambda}} & \frac{\partial^2 l(\beta)}{\partial \hat{\omega}^2} & \frac{\partial^2 l(\beta)}{\partial \hat{\omega} \partial \hat{k}} \\ \frac{\partial^2 l(\beta)}{\partial \hat{k} \partial \hat{\lambda}} & \frac{\partial^2 l(\beta)}{\partial \hat{k} \partial \hat{\omega}} & \frac{\partial^2 l(\beta)}{\partial \hat{k}^2} \end{pmatrix} \quad (2.3.1)$$

iii) Inverter a matriz  $J(\hat{\beta})$ ;

iv) Utilizar a posição  $J^{-1}[1, 1]$  da matriz inversa  $J^{-1}(\hat{\beta})$  calculada no passo (iii);

v) Calcular então os limites de confiança superior e inferior respectivamente, através de

$$\hat{\lambda} + z_{(1-\frac{\alpha}{2})} \sqrt{J^{-1}[1, 1]}$$

e

$$\hat{\lambda} - z_{(1-\frac{\alpha}{2})} \sqrt{J^{-1}[1, 1]},$$

onde  $\hat{\lambda}$  é o EMV para  $\lambda$ ,  $z_{(1-\frac{\alpha}{2})}$  é o quantil  $(1 - \frac{\alpha}{2})$  da distribuição normal padrão.

Intervalos de confiança assintóticos  $100(1 - \alpha)\%$  para os demais parâmetros são obtidos de forma análoga.