



MARCOS JARDEL HENRIQUES

**MODELOS HIERÁRQUICOS APLICADOS A
DADOS BIOLÓGICOS DE CONTAGEM**

Maringá - Paraná
Fevereiro de 2017

MARCOS JARDEL HENRIQUES

**MODELOS HIERÁRQUICOS APLICADOS A DADOS
BIOLÓGICOS DE CONTAGEM**

Dissertação apresentada ao Programa de Pós-Graduação em Bioestatística do Centro de Ciências Exatas da Universidade Estadual de Maringá, como requisito parcial para a obtenção do título de Mestre em Bioestatística.
Orientador: Prof. Dr. Vanderly Janeiro
Coorientador: Prof. Dr. Josmar Mazucheli

Maringá - Paraná

Fevereiro de 2017

Dados Internacionais de Catalogação na Publicação (CIP)
(Biblioteca Central - UEM, Maringá, PR, Brasil)

H519m Henriques, Marcos Jardel
 Modelos hierárquicos aplicados a dados biológicos
de contagem / Marcos Jardel Henriques. -- Maringá,
2017.
 80 f. : il. color., figs., tabs.

 Orientador: Prof. Dr. Vanderly Janeiro.
 Coorientador: Prof. Dr. Josmar Mazucheli.
 Dissertação (mestrado) - Universidade Estadual de
Maringá, Departamento de Estatística, Programa de
Pós-Graduação em Bioestatística, 2017.

 1. Dados de contagem (Estatística)- Fertilização
in vitro. 2. Fertilização *in vitro* - Bioestatística.
3. Embrião *in vitro* - Bovino - Bioestatística. 4.
Complexos *cumullus-oophorus*. I. Janeiro, Vanderly,
orient. II. Mazucheli, Josmar, coorient. III.
Universidade Estadual de Maringá. Departamento de
Estatística Programa de Pós-Graduação em
Bioestatística. IV. Título.

CDD 23.ed. 570.15195

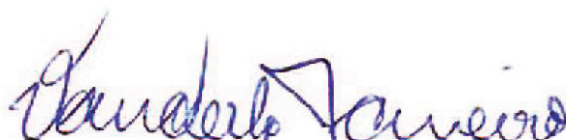
GV5-003719

MARCOS JARDEL HENRIQUES

MODELOS HIERÁRQUICOS APLICADOS A DADOS BIOLÓGICOS DE CONTAGEM

Dissertação apresentada ao Programa de Pós-Graduação em Bioestatística do Centro de Ciências Exatas da Universidade Estadual de Maringá, como requisito parcial para a obtenção do título de Mestre em Bioestatística.

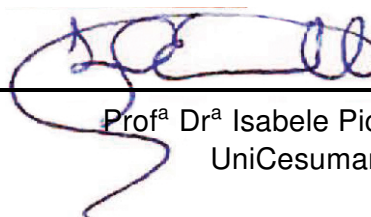
BANCA EXAMINADORA



Prof. Dr. Vanderly Janeiro
Universidade Estadual de Maringá – Maringá/PR



Prof. Dr. Carlos Aparecido dos Santos
Universidade Estadual de Maringá – Maringá/PR



Prof^a Dr^a Isabele Picada Emanuelli
UniCesumar – Maringá/PR

Maringá - Paraná
Fevereiro de 2017

Dedico este trabalho a todos que contribuíram para sua realização.

Agradecimentos

Agradeço a todas as pessoas que me orientaram e tornaram possível a realização deste trabalho. Em especial aos professores doutores Vanderly Janeiro, Josmar Mazucheli, Terezinha Aparecida Guedes, Carlos Aparecido dos Santos e Isabele Picada Emanuelli pelos ensinamentos, orientações, contribuições e correções. Agradeço imensamente aos amigos Oilson Alberto Gonzatto Junior, Edilênia Queiroz Pereira e Aline Edlaine de Medeiros. Agradeço a todas as outras pessoas, que agora não me recordo, pela contribuição na realização deste trabalho. Agradeço também à sociedade que através da CAPES me deram apoio financeiro.

“The problem with the world is that the intelligent people are full of doubts, while the stupid ones are full of confidence.”
(Charles Bukowski)

Resumo

Neste trabalho analisou-se um banco de dados, proveniente de um experimento com dois grupos, de 15 vacas da raça Nelore (*Bos Tauros indicus*), onde, um grupo recebeu uma suplementação controle (Tratamento 1), e o outro recebeu uma suplementação com inclusão de sais de ácidos graxos de cadeia longa (Tratamento 2). Este banco de dados é composto por dados de contagem, e possui, a quantidade de oócitos aspirados e a quantidade de oócitos viáveis, em um total de coletas que variou entre 2 a 8 aspirações por doadora em um experimento de aproximadamente 120 dias. Desta forma, objetivando-se avaliar a existência de possíveis diferenças entre os tratamentos, em relação à quantidade e a viabilidade dos oócitos, produzidos pelas vacas, bem como, identificar os modelos que melhor ajustaram os dados, por meio de abordagens Frequentista, Bayesiana e Bayesiana Longitudinal. Fez-se uso de alguns modelos hierárquicos e algumas distribuições de probabilidade. Foi possível concluir, a não existência de diferença significativa entre os dois tratamentos pelas três abordagens. Além disso, pela análise Frequentista, os melhores ajustes se deram, respectivamente, por meio dos modelos Binomial-Negativo, Beta-Binomial e Poisson-Generalizado. Já pela análise Bayesiana os melhores ajustes se deram pelos modelos Binomial, Beta-Binomial, Binomial-Negativo, Poisson-Generalizado e Binomial-Poisson-Generalizado. Por outro lado, pela abordagem Bayesiana Longitudinal, todos os modelos se mostraram adequados.

Palavras-chaves: Dados de contagem; Fertilização *in vitro*; Embriões; Complexos *cumulus-oophorus*.

Abstract

A database from an experiment with two groups of 15 Nelore cows (*Bos Tauros indicus*) was analyzed. One group received a control supplementation (treatment 1) and the other received a supplementation with inclusion of long chain fatty acids salts (treatment 2). The data has the amount of aspirated and viable oocytes in a total of collects ranging from 2 to 8 aspirations per cow in a 120 days experiment, approximately. The objective of this study was to evaluate possible differences between treatments in relation to oocytes quantity and viability produced, as well as to identify best fit models through the Frequent, Bayesian and Bayesian Longitudinal approaches. We used some hierarchical models and probability distributions. It was possible to conclude that there was no significant difference between treatments by the three approaches. In addition, by the Frequentist analysis, the best adjustments were made, respectively, by Binomial-Negative, Beta-Binomial and Poisson-Generalized models. In Bayesian analysis, the best adjustments were the Binomial, Beta-Binomial, Binomial-Negative, Poisson-Generalized and Binomial-Poisson-Generalized models. On the other hand, by the Bayesian Longitudinal approach, all models were adequate.

Keywords: Counting data; *in vitro* fertilization; Embryos; Complexes *cumullus-oophorus*.

Lista de ilustrações

- Figura 1 – Média de oócitos aspirados ($M_A = N/o$) e viáveis ($M_V = Y/o$) para as doadoras que receberam o Tratamento 1. Os valores numéricos acima das barras indicam a proporção de oócitos viáveis ($p_V = Y/N$). 33
- Figura 2 – Média de oócitos aspirados ($M_A = N/o$) e viáveis ($M_V = Y/o$) para as doadoras que receberam o Tratamento 2. Os valores numéricos acima das barras indicam a proporção de oócitos viáveis ($p_V = Y/N$). 33
- Figura 3 – Perfis individuais (em cinza) e linhas de tendência média (em preto) para o Total de Oócitos Aspirados (a), o Total de Oócitos Viáveis (b) e a Proporção de Oócitos Viáveis (c). 34
- Figura 4 – Gráficos Diagnósticos para o Ajuste do Modelo Binomial 36
- Figura 5 – Worm-plot do ajuste pelo Modelo Binomial 37
- Figura 6 – Gráficos Diagnósticos para o Ajuste do Modelo Poisson 39
- Figura 7 – Worm-plot do ajuste pelo Modelo Poisson 40
- Figura 8 – Gráficos Diagnósticos para o Ajuste do Modelo Binomial-Poisson 43
- Figura 9 – Gráficos Diagnósticos para o Ajuste do Modelo Binomial-Negativo 46
- Figura 10 – Gráficos Diagnósticos para o Ajuste do Modelo Poisson-Generalizado 49
- Figura 11 – Gráficos Diagnósticos para o Ajuste do Modelo Beta-Binomial 52
- Figura 11 – [ESQUERDA] Bandas de reamostragem simulada, determinadas com base nas médias a posteriori para os parâmetros associados aos modelos ajustados. Foram tomados 200 grupos de 30 observações amostradas dos modelos para a criação das bandas. Os pontos em preto correspondem à amostra original; [CENTRO] Quantis (2.5% e 97.5%) da distribuição a posteriori dos resíduos padronizados por doadora. Os pontos em preto e a linha tracejada representam as médias residuais por doadora e geral, respectivamente; [DIREITA] Quantis (2.5% e 97.5%) das distribuições preditivas a posteriori, por doadora, para a quantidade de oócitos viáveis por aspiração. Os pontos em preto e a linha tracejada representam a mediana das predições por doadora e geral, respectivamente. 59

Figura 10 – [ESQUERDA] Quantis (2.5% e 97.5%) da distribuição a posteriori dos resíduos padronizados por doadora. Os pontos em preto e a linha tracejada representam as médias residuais por doadora e geral, respectivamente; [DIREITA] Quantis (2.5% e 97.5%) das distribuições preditivas a posteriori, por doadora, para a quantidade de oócitos viáveis por aspiração. Os pontos em preto e a linha tracejada representam a mediana das predições por doadora e geral, respectivamente. 66

Sumário

1	Introdução	13
2	Materiais e Métodos	15
2.1	Descrição do Banco de Dados	15
2.2	Descrição de Alguns Modelos Probabilísticos	17
2.2.1	Modelo Binomial	17
2.2.2	Modelo Poisson	18
2.2.3	Modelo Poisson-Generalizado	19
2.3	Modelos Hierárquicos	20
2.3.1	Modelo Binomial-Poisson	21
2.3.2	Modelo Binomial-Negativo	23
2.3.3	Modelo Beta-Binomial	25
2.3.4	Modelo Binomial-Poisson-Generalizado	28
2.4	Inferência Bayesiana	29
2.4.1	Teorema de Bayes	29
2.4.2	Distribuição a Posteriori	29
2.4.3	Seleção de Modelos	30
2.4.4	Critérios de Convergência	30
2.4.5	Procedimento Computacional	30
3	Resultados	32
3.1	Análise Descritiva	32
3.2	Modelos Ajustados com Metodologia Frequentista	34
3.2.1	Ajuste com base no Modelo de Regressão Binomial	34
3.2.2	Ajuste com base no Modelo de Regressão Poisson	37
3.2.3	Ajuste com base no Modelo de Regressão Binomial-Poisson	40
3.2.4	Ajuste com base no modelo de Regressão Binomial-Negativo	44
3.2.5	Ajuste com base no Modelo Poisson-Generalizado	47
3.2.6	Ajuste com base no Modelo Beta-Binomial	50
3.2.7	Resumo dos Resultados Frequentistas	53
3.3	Modelos Ajustados com Metodologia Bayesiana	55
3.4	Modelos Ajustados com Metodologia Bayesiana considerando uma Estrutura Longitudinal	60
4	Conclusões	68

Referências	70
Apêndices	73
APÊNDICE A Códigos – Abordagem Frequentista	74
APÊNDICE B Códigos – Abordagem Bayesiana	75
APÊNDICE C Códigos – Abordagem Bayesiana Longitudinal	78

Capítulo 1

Introdução

As estruturas hierárquicas estão presentes nos mais variados contextos, modelados por diferentes distribuições de probabilidade. Por mais que uma variável aleatória possa ser descrita por uma única distribuição de probabilidade, muitas vezes é mais viável modelá-la considerando uma ou mais hierarquias (CASELLA; BERGER, 2011). Estimar efeitos de tratamento sobre observações de uma variável aleatória Binomial, por exemplo, leva em conta a suposição de que o número de tentativas em cada observação binomial seja conhecida. Um modelo hierárquico poderia enriquecer a modelagem com a suposição de que este número de tentativas seja também uma variável aleatória, Poisson, por exemplo. Um modelo como este foi utilizado por Shkedy et al. (2005).

Em reprodução de bovinos, variáveis relacionadas com a quantidade e viabilidade de oócitos possuem grande relevância, principalmente em animais de produção. Pesquisadores da área costumam modelar as proporções de oócitos viáveis ou não viáveis, e a quantidade de oócitos aspirados. Quando modelam a taxa de oócitos acabam fazendo uso apenas do modelo de Poisson, eventualmente incluindo efeitos aleatórios (CAVALIERI et al., 2005; OLIVEIRA, 2012; ZANIBONI et al., 2012; VIEIRA, 2013; VIEIRA et al., 2014; SOUZA et al., 2016).

Os sais de ácidos graxos de cadeia longa de óleo de soja são considerados gorduras protegidas. Estas são resistentes à degradação ruminal por serem fontes de gorduras praticamente inertes ou sobrepassantes no rúmen. Ou seja, quando ingeridas por ruminantes, não sofrem lipólise e biohidrogenação pelos micro-organismos no rúmen. Desta maneira, aumenta a quantidade de lipídeos digeridos e absorvidos no intestino delgado, melhorando o aproveitamento da densidade energética das dietas e da quantidade de ácidos graxos insaturados disponíveis no intestino delgado, que contribui para a melhoria da reprodução bovina. Porém, os mecanismos pelos quais a suplementação com gordura melhora o desempenho reprodutivo não têm sido elucidados. Constata-se que são poucos os

trabalhos que avaliaram o efeito da suplementação com gordura sobre a reprodução de vacas Nelore (*Bos taurus indicus*) doadoras de oócitos (SOUZA et al., 2016).

Para Ramos (2015), a coleta de oócitos possibilita a reprodução *in vitro* de embriões, aumentando a produção bovina tanto em quantidade, quanto em qualidade e velocidade. Atualmente, o Brasil é o maior produtor mundial de embriões *in vitro*, bem como, de oócitos bovinos, fazendo com que o país seja destaque no cenário internacional como referência na produção *in vitro* de embriões. A fertilização *in vitro*, só ocorre, graças a possibilidade da aspiração de oócitos de vacas para viabilizar a produção de um grande número de embriões, o que é fundamental para: transferência comercial, auxílio para superar problemas de infertilidade adquirida em doadoras de alto valor, produção de embriões de animais jovens, redução do intervalo de gerações e produção de animais clones e transgênicos, entre outros benefícios, que geram milhões de dólares, a esta cadeia produtiva, todos os anos Viana e Figueiredo (2016). Apesar dos estudos sobre fertilização *in vitro* nos últimos anos apontarem que a produção de embriões viáveis dificilmente supera a taxa de 40% Sirard et al. (2006), e parte destes resultados podem estar relacionados com a capacidade dos oócitos aspirados se desenvolverem *in vitro* Rizos et al. (2002), poucos trabalhos têm sido direcionados para os possíveis efeitos da nutrição sobre os parâmetros reprodutivos em bovinos (SOUZA et al., 2016).

Por meio de modelos hierárquicos e distribuições de probabilidade, serão propostos alguns ajustes/modelagens para algumas quantidades de interesse do pesquisador, tais como: a *proporção esperada* de oócitos viáveis, a *taxa esperada* de oócitos viáveis e/ou aspirados e o *total esperado* de oócitos viáveis e/ou aspirados, produzidos por vacas da raça Nelore (*Bos Taurus indicus*). Tais modelagens, conseqüentemente darão suporte para avaliar se há diferença entre os dois tratamentos quanto a quantidade e a viabilidade de oócitos produzidos por estas vacas. Além disso, será verificado, quais modelos melhor descrevem os dados. Para isso, irá se trabalhar com modelagens Frequentista, Bayesiana e Bayesiana Longitudinal, com os modelos hierárquicos e algumas distribuições de probabilidade, tais como, os modelos: Binomial, Poisson, Beta-Binomial, Binomial-Poisson, Binomial-Negativo, Poisson-Generalizado e Binomial-Poisson-Generalizado.

No segundo capítulo, Materiais e Métodos, estão descritos: o banco de dados; todas as distribuições utilizadas nesta dissertação; os critérios de convergência e seleção de modelos, bem como os aspectos computacionais adotados para a realização dos ajustes. No terceiro capítulo, Resultados, expõem-se as análises descritivas e os resultados obtidos com os ajustes: Frequentista, Bayesiano e Bayesiano Longitudinal. Por fim, no capítulo quatro, Conclusões, estão expostas as conclusões gerais acerca das análises.

Capítulo 2

Materiais e Métodos

2.1 Descrição do Banco de Dados

Este banco de dados é o resultado de um experimento realizado, ao longo de aproximadamente 120 dias, para avaliar os efeitos de uma dieta a base de sais de ácidos graxos de cadeia longa, derivados de óleo de soja, sobre a produção e viabilidade de oócitos aspirados de lactação Nelore (*Bos taurus indicus*). O ensaio foi composto por 2 tratamentos de 15 repetições cada. Trinta vacas Nelore lactantes foram distribuídas aleatoriamente para os dois tratamentos. Além dos tratamentos, durante todo o ensaio, as vacas pastaram em capim Tifton 85. Os dois tratamentos consistiram em: tratamento controle (tratamento 1) e um suplemento contendo sais de ácidos graxos de cadeia longa (tratamento 2). Após 14 dias de tratamento, iniciou-se a captação de oócitos, o que variou entre 2 e 8 captações por animal no decorrer do ensaio. O tempo de duração, entre uma aspiração e outra, na mesma bovina variou entre 10 e 20 dias, ou seja, trata-se de um banco de dados desbalanceado. A cada aspiração o oócitos eram quantificados e classificados morfolologicamente em microscópio óptico em viáveis e inviáveis.

Tabela 1 – Esboço do banco de dados.

x_i	D_i	$n_{i,1}$	$y_{i,1}$	$d_{i,1}$	$n_{i,2}$	$y_{i,2}$	$d_{i,2}$	$n_{i,3}$	$y_{i,3}$	$d_{i,3}$	$n_{i,4}$	$y_{i,4}$	$d_{i,4}$	$n_{i,5}$	$y_{i,5}$	$d_{i,5}$	$n_{i,6}$	$y_{i,6}$	$d_{i,6}$	$n_{i,7}$	$y_{i,7}$	$d_{i,7}$	$n_{i,8}$	$y_{i,8}$	$d_{i,8}$	o_i
1	1	19	16	14	26	23	26	22	21	54	20	19	81													2
	2	15	12	14	33	30	28	38	33	111																4
	3	47	44	26	36	33	54	65	62	84	41	39	111													3
	4	68	65	14	84	79	25	58	56	54	37	33	81	34	31	119										4
	5	59	52	14	71	69	26	24	20	42	7	6	53	13	11	84	8	4	98	9	8	119				5
	6	19	15	14	28	23	28	23	20	81	20	16	119													7
	7	17	16	42	25	22	53	23	20	14	32	30	82	34	29	98	32	27	110							4
1	8	45	41	28	31	45	21	19	14	54	17	7	67	25	18	81	19	14	98	10	8	119				6
	9	29	21	28	35	28	42	21	18	53	17	7	67	25	18	81	19	14	98	10	8	119				7
	10	27	23	14	26	24	25	24	22	53	22	19	67	16	14	98	13	9	119							6
	11	24	19	54	22	18	84	25	19	110																6
	12	34	35	14	18	15	26	31	29	54	35	32	84	22	19	119										3
	13	25	22	14	29	27	26	33	31	53	39	34	119													5
	14	19	16	14	27	22	26	24	21	41	11	10	54	13	11	81	12	9	111							4
	15	11	10	14	14	12	28	10	8	41	11	10	53													6
	16	13	12	41	17	12	41																			4
	17	6	2	14	14	12	26	13	12	54	9	7	81	3	3	111										2
	18	26	25	28	36	34	42	20	19	53	23	20	67	20	19	81	9	8	98	19	16	111				5
	19	22	20	14	23	21	28	36	34	41	35	33	67	27	26	119										7
	20	27	25	14	27	24	28	42	40	42	21	16	67	29	26	84	23	21	119							5
	21	29	28	14	35	32	25	34	33	53	36	33	84	25	22	110										6
	22	6	5	14	7	5	28	7	4	42	8	6	67	8	7	98	7	5	119							5
2	23	20	19	14	51	43	28	63	59	53	10	8	67	22	16	81	33	24	98	21	16	119				6
	24	21	20	14	21	19	25	30	22	42	24	22	53	20	18	67	16	12	82	13	10	98	14	10	119	8
	25	13	11	53	30	26	67	13	11	98	10	9	119													4
	26	20	16	25	35	32	53	17	13	110																4
	27	48	41	42	37	34	53	49	40	82	67	50	119													3
	28	20	19	53	28	25	67	32	23	81	35	26	110													4
	29	61	53	14	39	36	25	43	40	42	40	35	67	34	32	82	24	19	110							4
	30	17	14	14	25	18	28	18	14	42	19	16	67	14	10	84	11	10	98							6

x_i denota o tratamento recebido pela doadora i , com $i = 1, 2, \dots, 30$: (1) Controle; (2) Suplementação alternativa.

D_i é o identificador da doadora, $i = 1, \dots, 30$.

$n_{i,j}$ representa o total de óbitos aspirados da doadora i , com $i = 1, \dots, 30$, na aspiração j , com $j = 1, \dots, o_i$.

$y_{i,j}$ representa o total de óbitos viáveis, isto é, número de sucessos entre os óbitos aspirados, $n_{i,j}$, da doadora i , com $i = 1, 2, \dots, 30$, na aspiração j , com $j = 1, 2, \dots, o_i$.

$d_{i,j}$ indica o número de dias transcorridos entre o início da suplementação alimentar e a aspiração j , com $j = 1, 2, \dots, o_i$, na doadora i , com $i = 1, 2, \dots, 30$.

o_i denota o número de vezes que a doadora i , com $i = 1, \dots, 30$, foi aspirada.

2.2 Descrição de Alguns Modelos Probabilísticos

2.2.1 Modelo Binomial

Segundo Morettin (2010), para determinar a função de probabilidade Binomial, considerando-se Y o número de sucessos em n tentativas e um resultado particular (RP), pode-se determinar a função de probabilidade da variável aleatória Y , ou seja, $\Pr(Y = y)$. Assim, procedendo-se da seguinte maneira:

Fazendo-se,

$$\underbrace{SSS \dots S}_y \underbrace{FFF \dots F}_{n-y},$$

em que S e F denotam, respectivamente, sucesso e fracasso em um experimento de Bernoulli. Assim, tem-se que

$$\begin{aligned} \Pr(RP) &= \Pr(\underbrace{S \cap S \cap S \cap \dots \cap S}_y \cap \underbrace{F \cap F \cap F \cap \dots \cap F}_{n-y}) \\ &= \underbrace{\Pr(S) \Pr(S) \Pr(S) \dots \Pr(S)}_y \underbrace{\Pr(F) \Pr(F) \Pr(F) \dots \Pr(F)}_{n-y} \\ &= \underbrace{ppp \dots p}_y \underbrace{qqq \dots q}_{n-y}. \end{aligned}$$

Considerando-se agora, todas as n -uplas com Y sucessos, chega-se a função de probabilidade:

$$\Pr(Y = y) = \binom{n}{y} p^y (1-p)^{n-y}, \quad \begin{array}{l} y = 0, 1, \dots \\ 0 < p < 1. \end{array}$$

A média e a variância podem ser obtidas de uma forma simples, como sendo $\mathbb{E}(Y) = np$ e a $\text{Var}(Y) = np(1-p)$. A distribuição Binomial tem duas características importantes, que consiste em ser aproximada da distribuição de Poisson quando se tem uma quantidade suficientemente grande de observações, com a probabilidade de ocorrência do evento de interesse, suficientemente pequena, e também ser gerada a partir de provas de Bernoulli independentes.

O modelo de regressão Binomial, possui sua variável aleatória pertencente à família exponencial, que por sua vez é a componente aleatória do modelo. Para se construir uma regressão, muitas vezes é necessário fazer uso dos modelos lineares generalizados que é composto por uma componente aleatória, uma função de ligação e um componente sistemático, onde, a componente aleatória é ligada a componente sistemática através da função de ligação. Como já é de notório saber na Estatística, assim serão construídos os modelos de regressão no decorrer desta dissertação.

Em relação à regressão Binomial, considerando-se um conjunto de observações independentes $\mathbf{Y} = (Y_1, Y_2, Y_3, \dots, Y_n)$, tal que $Y_i \sim \text{BI}(n_i, p_i)$. O modelo de Regressão

Binomial é definido por

$$\begin{aligned} Y_i &\sim \text{BI}(n_i, p_i), \\ g(p_i) &= \mathbf{x}_i^\top \boldsymbol{\beta}, \end{aligned} \quad i = 1, 2, \dots, n,$$

em que $\mathbf{x}_i = (1, x_{1i}, x_{2i}, \dots, x_{ki})^\top$ denota o vetor de covariáveis para o indivíduo i ; $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_k)^\top$ denota o vetor de parâmetros associados as covariáveis; $g(\cdot)$ denota a função de ligação (comumente a função *logit*, na Regressão Binomial) que associa as componentes aleatória e sistemática do modelo.

Nesse contexto, a esperança e a variância de Y_i são expressas por

$$\mathbb{E}(Y_i) = n_i p_i, \quad \text{e} \quad \text{Var}(Y_i) = n_i p_i (1 - p_i),$$

em que $p_i = g^{-1}(\mathbf{x}_i^\top \boldsymbol{\beta})$.

Uma estimativa para $\boldsymbol{\beta}$ pode ser obtida pela determinação do ponto $\hat{\boldsymbol{\beta}}$ que maximiza a função de log-verossimilhança do modelo de Regressão Binomial, que é expressa por

$$l(\boldsymbol{\beta}; \mathbf{Y}) = \sum_{i=0}^n \left\{ \log(n_i!) - \log(y_i!) - \log[(n_i - y_i)!] + y_i \log(p_i) + (n_i - y_i) \log(1 - p_i) \right\},$$

com $p_i = g^{-1}(\mathbf{x}_i^\top \boldsymbol{\beta})$.

2.2.2 Modelo Poisson

Para Bussab e Morettin (2010), se a variável aleatória Y tiver distribuição de Poisson, com parâmetro $\lambda > 0$, então sua função de probabilidade poderá ser expressa da forma,

$$\Pr(Y = y) = \frac{e^{-\lambda} \lambda^y}{y!}, \quad \begin{aligned} y &= 0, 1, 2, \dots, \\ \lambda &> 0, \end{aligned}$$

em que $\mathbb{E}(Y) = \lambda = \text{Var}(Y)$.

A distribuição de Poisson é uma base padrão para o ajuste de dados de contagens. Porém, quando a variância observada é muito discrepante das médias observadas, o que acontece na maioria das vezes em "dados reais", ocorrem os fenômenos da super ou da subdispersão. Nesses casos o modelo Binomial-Negativo e Poisson-Generalizado, entre outros, podem ser mais apropriados, por possibilitarem que se modele a variância de modo mais adequado (NAGAMINE, 2007). Estes, e demais modelos, serão descritos posteriormente.

No contexto de um modelo de Regressão de Poisson, considerando-se um conjunto de observações independentes $\mathbf{Y} = (Y_1, Y_2, Y_3, \dots, Y_n)$, tal que $Y_i \sim \text{PO}(\lambda_i)$. O modelo de Regressão de Poisson é definido por

$$\begin{aligned} Y_i &\sim \text{PO}(\lambda_i), \\ g(\lambda_i) &= \mathbf{x}_i^\top \boldsymbol{\beta}, \end{aligned} \quad i = 1, 2, \dots, n,$$

em que $\mathbf{x}_i = (1, x_{1i}, x_{2i}, \dots, x_{ki})^\top$ denota o vetor de covariáveis para o indivíduo i ; $\beta = (\beta_0, \beta_1, \dots, \beta_k)^\top$ denota o vetor de parâmetros associados as covariáveis; $g(\cdot)$ denota a função de ligação (comumente a função \log , na Regressão de Poisson) que associa as componentes aleatória e sistemática do modelo.

A esperança e a variância de Y_i são expressas por

$$\mathbb{E}(Y_i) = \lambda_i, \quad \text{e} \quad \text{Var}(Y_i) = \lambda_i,$$

em que $\lambda_i = g^{-1}(\mathbf{x}_i^\top \beta)$.

Uma estimativa para β pode ser obtida pela determinação do ponto $\hat{\beta}$ que maximiza a função de log-verossimilhança do modelo de Regressão Poisson, que é expressa por

$$l(\beta; \mathbf{Y}) = \sum_{i=1}^n \left\{ y_i \log(\lambda_i) - \lambda_i - \log(y_i!) \right\},$$

com $\lambda_i = g^{-1}(\mathbf{x}_i^\top \beta)$.

2.2.3 Modelo Poisson-Generalizado

Uma grande vantagem da distribuição de Poisson-Generalizada, é que, assim como a distribuição Binomial-Negativa, o modelo Poisson-Generalizado, consegue se ajustar bem a variáveis que seguem função de probabilidade de Poisson com dados sobredispersos (CONSUL; FAMOYE, 1992).

Entre as distribuições Lagrangianas e/ou generalizadas propostas por Consul e Joint (1973), está a distribuição Generalizada de Poisson, que tem a forma,

$$\Pr(Y = y) = \frac{\lambda(\lambda + y\theta)^{y-1} e^{-(\lambda+y\theta)}}{y!}, \quad y = 0, 1, 2, \dots, \quad \theta > 0 \quad \text{e} \quad \max\left(-1, \frac{-\theta}{4}\right) \leq \lambda \leq 1. \quad (2.1)$$

A esperança e variância da distribuição Poisson-Generalizada são, respectivamente,

$$\mathbb{E}(Y) = \frac{\lambda}{1 - \theta} \quad \text{e} \quad \text{Var}(Y) = \frac{\lambda}{(1 - \theta)^3}.$$

No contexto de um modelo de Regressão Poisson-Generalizado é interessante realizar uma reparametrização, a fim de obter parâmetros que representem o valor médio e uma medida de precisão da distribuição, Consul e Famoye (2007) assim o fizeram, denominando-se, respectivamente,

$$\mu = \frac{\lambda}{1 - \theta} \quad \text{e} \quad \phi = \frac{1}{1 - \theta}.$$

A partir desta reparametrização, obteve-se uma nova expressão para a função de probabilidade 2.1, dada por

$$\Pr(Y = y) = \frac{\mu \left[\mu + (\phi - 1)y \right]^{y-1} e^{-\frac{\mu + (\phi - 1)y}{\phi}}}{y! \phi^y} \quad \mu = 0, 1, 2, \dots, \quad \mu > 0 \quad \text{e} \quad \phi \geq \max\left(\frac{1}{2}, 1 - \frac{\mu}{4}\right),$$

em que,

$$\mathbb{E}(Y) = \mu \quad \text{e} \quad \text{Var}(Y) = \phi^2 \mu.$$

Desta forma, o modelo de Regressão Poisson-Generalizado, pode ser descrito, considerando-se um conjunto de observações independentes $\mathbf{Y} = (Y_1, Y_2, Y_3, \dots, Y_n)$, tal que $Y_i \sim \text{PG}(\mu_i, \phi_i)$. O modelo de Regressão Poisson-Generalizado é definido por

$$\begin{aligned} Y_i &\sim \text{PG}(\mu_i, \phi_i), \\ g_\mu(\mu_i) &= \mathbf{x}_{\mu i}^\top \boldsymbol{\beta}_\mu, & i = 1, 2, \dots, n, \\ g_\phi(\phi_i) &= \mathbf{x}_{\phi i}^\top \boldsymbol{\beta}_\phi, \end{aligned}$$

em que $\mathbf{x}_{\mu i} = (1, x_{\mu 1i}, x_{\mu 2i}, \dots, x_{\mu k_\mu i})^\top$ e $\mathbf{x}_{\phi i} = (1, x_{\phi 1i}, x_{\phi 2i}, \dots, x_{\phi k_\phi i})^\top$ denotam os vetores de covariáveis associadas aos parâmetros μ e ϕ , respectivamente, para o indivíduo i ; $\boldsymbol{\beta}_\mu = (\beta_{\mu 0}, \beta_{\mu 1}, \dots, \beta_{\mu k_\mu})^\top$ e $\boldsymbol{\beta}_\phi = (\beta_{\phi 0}, \beta_{\phi 1}, \dots, \beta_{\phi k_\phi})^\top$ são os vetores de parâmetros associados às covariáveis referentes aos parâmetros μ e ϕ , respectivamente; $g_\mu(\cdot)$ e $g_\phi(\cdot)$ denotam as funções de ligação (comumente a função *log*, para ambas) que associam as componentes aleatória e sistemática do modelo.

A esperança e a variância de Y_i são expressas por

$$\mathbb{E}(Y_i) = \mu_i \phi_i, \quad \text{e} \quad \text{Var}(Y_i) = \mu_i \phi_i (1 + \phi_i),$$

com, $\mu_i = g_\mu^{-1}(\mathbf{x}_{\mu i}^\top \boldsymbol{\beta}_\mu)$ e $\phi_i = g_\phi^{-1}(\mathbf{x}_{\phi i}^\top \boldsymbol{\beta}_\phi)$.

Uma estimativa para $\boldsymbol{\beta} = (\boldsymbol{\beta}_\mu^\top, \boldsymbol{\beta}_\phi^\top)^\top$ pode ser obtida pela determinação do ponto $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_\mu^\top, \hat{\boldsymbol{\beta}}_\phi^\top)^\top$ que maximiza a função de log-verossimilhança do modelo de Regressão Poisson-Generalizado, que é expressa por

$$\begin{aligned} l(\boldsymbol{\beta}; \mathbf{Y}) &= \\ &= \sum_{i=1}^n \left\{ \log(\mu_i) + (y_i - 1) \log [\mu_i + (\phi_i - 1)y_i] - \frac{\mu_i + (\phi_i - 1)y_i}{\phi_i} - \log(y_i!) - y_i \log(\phi_i) \right\}, \end{aligned}$$

sendo, $\mu_i = g_\mu^{-1}(\mathbf{x}_{\mu i}^\top \boldsymbol{\beta}_\mu)$ e $\phi_i = g_\phi^{-1}(\mathbf{x}_{\phi i}^\top \boldsymbol{\beta}_\phi)$.

2.3 Modelos Hierárquicos

Os modelos hierárquicos possuem dois ou mais níveis de hierarquia. Mesmo após o agrupamento dos níveis em um único modelo, ele ainda permite a incorporação de efeitos aleatórios associados a cada nível hierárquico. Estes efeitos nada mais são que erros aleatórios das diferenças entre as unidades de cada nível, relacionados com a área de interesse a ser estudada (GOLDSTEIN, 1999).

Teoricamente, os níveis de um modelo hierárquico são formados por uma amostra aleatória das unidades a serem consideradas. Assim, estas estruturas de dados podem ser

estudadas pelos modelos lineares hierárquicos, tornando possível especificações separadas de cada nível para posteriormente serem reunidas em um único modelo (RAUDEN-BUSH; BRYK, 2002). Nas próximas secções desta dissertação, direcionado pelo banco de dados a ser modelado e embasado na teoria estatística de modelos hierárquicos, serão realizadas modelagens com base em algumas funções de probabilidade.

Para melhor entendermos o tipo de modelagem hierárquica que será realizada neste trabalho, iremos nos basear no modelo hierárquico descrito por Casella e Berger (2011). Trata-se de um exemplo que contempla resultados de probabilidades condicionais. Estes autores exemplificaram em sua obra, a hierarquia Binomial-Poisson, com algumas considerações, no seguinte contexto:

Um inseto põe um grande número de ovos, cada um deles com uma probabilidade de sobrevivência p . Na média quantos ovos sobreviverão?

O “grande número” de ovos colocados é uma variável aleatória, geralmente assumindo que seja de $PO(\lambda)$. Além disso, se assumirmos que a sobrevivência de cada ovo é independente, então, temos provas de Bernoulli. Portanto, se considerarmos X como número de sobreviventes e Y o número de ovos que foram colocados, temos,

$$\begin{aligned} X|Y &\sim \text{BI}(Y, p), \\ Y &\sim \text{PO}(\lambda) \end{aligned}$$

um modelo hierárquico. Lembrando que, por simplificação, utilizou-se a notação $X|Y \sim \text{BI}(Y, p)$, com o significado da distribuição condicional de X dado que $Y = y$ é $\text{BI}(y, p)$ (CASELLA; BERGER, 2011).

2.3.1 Modelo Binomial-Poisson

Para se obter a função de probabilidade para uma variável aleatória Y com distribuição Binomial-Poisson, parte-se da ideia de uma variável aleatória com distribuição $\text{BI}(n, p)$, condicionada à suposição de que a quantidade n é uma realização de uma variável aleatória N com distribuição $\text{PO}(\lambda)$, e então, determina-se a probabilidade marginal de Y (CASELLA; BERGER, 2011). A hierarquia desse modelo é expressa por

$$\begin{aligned} Y | N = n &\sim \text{BI}(n, p) \\ N &\sim \text{PO}(\lambda). \end{aligned}$$

Considerando a função de probabilidade conjunta do par (Y, N) cuja forma é dada por,

$$\begin{aligned} \Pr(Y = y, N = n) &= \Pr(Y = y | N = n) \Pr(N = n) \\ &= \underbrace{\Pr(y; p, n)}_{\text{BI}(p, n)} \underbrace{\Pr(n; \lambda)}_{\text{PO}(\lambda)} \\ &= \binom{n}{y} p^y (1-p)^{n-y} \frac{e^{-\lambda} \lambda^n}{n!}, \quad \text{para } n \geq y, \end{aligned}$$

a função de probabilidade de Y é obtida determinando a função de probabilidade marginal,

$$\begin{aligned}
 \Pr(Y = y) &= \sum_{n=y}^{\infty} \binom{n}{y} p^y (1-p)^{n-y} \frac{e^{-\lambda} \lambda^n}{n!} \\
 &= \sum_{n=y}^{\infty} \binom{n}{y} p^y (1-p)^{n-y} \frac{e^{-\lambda} \lambda^n}{n!} \frac{\lambda^y}{\lambda^y} \\
 &= \sum_{n=y}^{\infty} \frac{n!}{(n-y)! y!} p^y (1-p)^{n-y} \frac{e^{-\lambda} \lambda^n}{n!} \frac{\lambda^y}{\lambda^y} \\
 &= \frac{(\lambda p)^y e^{-\lambda}}{y!} \sum_{n=y}^{\infty} \frac{[(1-p)\lambda]^{n-y}}{(n-y)!}.
 \end{aligned}$$

Assim, chamando-se $n - y$ de x , tem-se que,

$$\Pr(Y = y) = \frac{(\lambda p)^y e^{-\lambda}}{y!} \sum_{x=0}^{\infty} \frac{[(1-p)\lambda]^x}{x!},$$

note que o termo,

$$\sum_{x=0}^{\infty} \frac{[(1-p)\lambda]^x}{x!},$$

é a representação em série de $e^{(1-p)\lambda}$, portanto

$$\Pr(Y = y) = \frac{(\lambda p)^y e^{-\lambda}}{y!} e^{(1-p)\lambda} = \frac{(\lambda p)^y}{y!} e^{-\lambda p}.$$

Assim, uma variável aleatória Y com distribuição Binomial-Poisson, tem como função de probabilidade

$$\Pr(Y = y) = \frac{(\lambda p)^y}{y!} e^{-\lambda p} \quad \begin{array}{l} y = 0, 1, 2, \dots, \\ \lambda > 0 \quad \text{e} \quad 0 < p < 1, \end{array}$$

o que, na prática, é a própria distribuição de Poisson, com parâmetro expresso por λp .

Uma vez que $Y \sim \text{BP}(\lambda p)$, a esperança e variância dessa distribuição são dadas por

$$\mathbb{E}(Y) = \lambda p \quad \text{e} \quad \text{Var}(Y) = \lambda p.$$

Percebe-se até aqui, que a introdução da quantidade aleatória N na hierarquia foi principalmente para ajudar a entender o modelo. Nota-se que existe uma vantagem a mais ao se considerar que o parâmetro da distribuição de Y é o produto de dois parâmetros (λ e p), sendo cada um deles relativamente simples de se entender (o “tamanho esperado” da amostra e a proporção binomial). Ou seja, pode existir uma vantagem interpretativa dos parâmetros, porém possivelmente indiferente para a estimação.

Em relação ao modelo de Regressão Binomial-Poisson, considerando-se um conjunto de observações independentes $\mathbf{Y} = (Y_1, Y_2, Y_3, \dots, Y_n)$, tal que $Y_i \sim \text{BP}(\lambda_i p_i)$. O modelo de Regressão Binomial-Poisson é definido por

$$\begin{aligned} Y_i &\sim \text{PO}(\lambda_i p_i), \\ g_\lambda(\lambda_i) &= \mathbf{x}_{\lambda_i}^\top \boldsymbol{\beta}_\lambda, \\ g_p(p_i) &= \mathbf{x}_{p_i}^\top \boldsymbol{\beta}_p, \end{aligned} \quad i = 1, 2, \dots, n,$$

em que $\mathbf{x}_{\lambda_i} = (1, x_{\lambda_{1i}}, x_{\lambda_{2i}}, \dots, x_{\lambda_{k_\lambda i}})^\top$ e $\mathbf{x}_{p_i} = (1, x_{p_{1i}}, x_{p_{2i}}, \dots, x_{p_{k_p i}})^\top$ denotam os vetores de covariáveis associadas aos parâmetros λ e p , respectivamente, para o indivíduo i ; $\boldsymbol{\beta}_\lambda = (\beta_{\lambda_0}, \beta_{\lambda_1}, \dots, \beta_{\lambda_{k_\lambda}})^\top$ e $\boldsymbol{\beta}_p = (\beta_{p_0}, \beta_{p_1}, \dots, \beta_{p_{k_p}})^\top$ são os vetores de parâmetros associados às covariáveis referentes aos parâmetros λ e p , respectivamente; $g_\lambda(\cdot)$ e $g_p(\cdot)$ denotam as funções de ligação (comumente as funções *log* e *logit*, respectivamente) que associam as componentes aleatória e sistemática do modelo.

A esperança e a variância de Y_i são expressas por

$$\mathbb{E}(N_i) = \lambda_i p_i, \quad \text{e} \quad \text{Var}(N_i) = \lambda_i p_i,$$

em que $\lambda_i = g_\lambda^{-1}(\mathbf{x}_{\lambda_i}^\top \boldsymbol{\beta}_\lambda)$ e $p_i = g_p^{-1}(\mathbf{x}_{p_i}^\top \boldsymbol{\beta}_p)$.

Uma estimativa para $\boldsymbol{\beta} = (\boldsymbol{\beta}_\lambda^\top, \boldsymbol{\beta}_p^\top)^\top$ pode ser obtida pela determinação do ponto $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_\lambda^\top, \hat{\boldsymbol{\beta}}_p^\top)^\top$ que maximiza a função de log-verossimilhança do modelo de Regressão Binomial-Poisson, que é expressa por

$$l(\boldsymbol{\beta}; \mathbf{Y}) = \sum_{i=1}^n \left\{ y_i \log(\lambda_i p_i) - \lambda_i p_i - \log(y_i!) \right\},$$

com $\lambda_i = g_\lambda^{-1}(\mathbf{x}_{\lambda_i}^\top \boldsymbol{\beta}_\lambda)$ e $p_i = g_p^{-1}(\mathbf{x}_{p_i}^\top \boldsymbol{\beta}_p)$.

2.3.2 Modelo Binomial-Negativo

Para Magalhães (2011), a variável aleatória da distribuição Binomial-Negativa é o número de tentativas até o r -ésimo sucesso, ou seja, conta-se o número de provas de Bernoulli necessárias para se obter um número fixo de sucessos. Alguns autores definem a Binomial-Negativa como número de ensaios anteriores ao r -ésimo sucesso, na sua variável. Dessa forma, eles definem a Binomial-Negativa como o número de ensaios necessários para a obtenção do r -ésimo sucesso.

Considerando-se uma sequência de ensaios de Bernoulli independentes, ao se definir Y como o número de fracassos anteriores ao r -ésimo sucesso, pode-se afirmar que a variável aleatória Y segue a função de probabilidade Binomial-Negativa com r e p parâmetros, sendo, $0 < p < 1$ e $r > 0$, com função de probabilidade dada por:

$$\Pr(Y = y) = \binom{r+y-1}{r-1} \binom{p}{p+1}^y \frac{1}{(1+p)^r}, \quad y = 0, 1, 2, \dots$$

Tem-se que a esperança e variância são $E[Y] = rp$ e a $Var[Y] = rp(1 + p)$. Note, também, que se $y = 1$, será obtido o modelo Geométrico.

Também é possível entender a distribuição Binomial-Negativa como uma composição hierárquica resultante das distribuições Poisson e Gama. Nesse contexto,

$$\begin{aligned} Y | \Lambda = \lambda &\sim \text{PO}(\mu\lambda), \\ \Lambda &\sim \text{GA}(1, \sigma^{\frac{1}{2}}), \end{aligned}$$

em que, λ tem distribuição gama com média 1 e parâmetro de escala $\sigma^{1/2}$.

Sob essas considerações, deseja-se determinar a distribuição marginal de Y . Para isso, tem-se que

$$\begin{aligned} \Pr(Y = y, \Lambda = \lambda) &= \Pr(Y = y | \Lambda = \lambda) \Pr(\Lambda = \lambda) \\ &= \underbrace{\Pr(y; \mu\lambda)}_{\text{PO}(\mu\lambda)} \underbrace{f(\lambda; 1, \sigma^{1/2})}_{\text{GA}(1, \sigma^{1/2})} \\ &= \frac{(\mu\lambda)^y e^{-\mu\lambda}}{y!} \frac{1}{\sigma^{\frac{1}{\sigma}}} \frac{\lambda^{\frac{1}{\sigma}-1} e^{-\frac{\lambda}{\sigma}}}{\Gamma\left(\frac{1}{\sigma}\right)} \\ &= \frac{1}{\Gamma\left(\frac{1}{\sigma}\right) \Gamma(y+1)} \frac{\mu^y}{\sigma^{\frac{1}{\sigma}}} \lambda^{y+\frac{1}{\sigma}-1} e^{-(\mu+\frac{1}{\sigma})\lambda} \end{aligned}$$

segue, portanto, que uma expressão para a função de probabilidade marginal de Y é dada por

$$\begin{aligned} \Pr(Y = y) &= \int_0^{\infty} \frac{1}{\Gamma\left(\frac{1}{\sigma}\right) \Gamma(y+1)} \frac{\mu^y}{\sigma^{\frac{1}{\sigma}}} \lambda^{y+\frac{1}{\sigma}-1} e^{-(\mu+\frac{1}{\sigma})\lambda} d\lambda \\ &= \frac{1}{\Gamma\left(\frac{1}{\sigma}\right) \Gamma(y+1)} \frac{\mu^y}{\sigma^{\frac{1}{\sigma}}} \int_0^{\infty} \underbrace{\lambda^{y+\frac{1}{\sigma}-1} e^{-(\mu+\frac{1}{\sigma})\lambda}}_{\text{Núcleo de GA}(y+\frac{1}{\sigma}, \mu+\frac{1}{\sigma})} d\lambda \\ &= \frac{1}{\Gamma\left(\frac{1}{\sigma}\right) \Gamma(y+1)} \frac{\mu^y}{\sigma^{\frac{1}{\sigma}}} \frac{\Gamma\left(y+\frac{1}{\sigma}\right)}{\left(\mu+\frac{1}{\sigma}\right)^{y+\frac{1}{\sigma}}} \\ &= \frac{\Gamma\left(y+\frac{1}{\sigma}\right)}{\Gamma\left(\frac{1}{\sigma}\right) \Gamma(y+1)} \left(\frac{\sigma\mu}{\sigma\mu+1}\right)^y \left(\frac{1}{\sigma\mu+1}\right)^{\frac{1}{\sigma}}. \end{aligned}$$

Assim, tem-se que a função de probabilidade marginal de Y é dada por

$$\Pr(Y = y) = \frac{\Gamma\left(y+\frac{1}{\sigma}\right)}{\Gamma\left(\frac{1}{\sigma}\right) \Gamma(y+1)} \left(\frac{\sigma\mu}{1+\sigma\mu}\right)^y \left(\frac{1}{1+\sigma\mu}\right)^{\frac{1}{\sigma}}, \quad y = 0, 1, 2, \dots, \\ \mu > 0 \text{ e } \sigma > 0.$$

Nesse caso, a média e variância de Y são calculadas por

$$\mathbb{E}(Y) = \mathbb{E}[\mathbb{E}(Y|\Lambda)] = \mathbb{E}[\mu\Lambda] = \mu,$$

e

$$\text{Var}(Y) = \mathbb{E}[\text{Var}(Y|\Lambda)] + \text{Var}[\mathbb{E}(Y|\Lambda)] = \mathbb{E}(\mu\Lambda) + \text{Var}(\mu\Lambda) = \mu + \mu^2\sigma.$$

No contexto de um modelo de regressão, esta última parametrização apresenta a vantagem de expressar seu valor esperado como o parâmetro μ . Considerando-se um conjunto de observações independentes $Y_1, Y_2, Y_3, \dots, Y_n$, tal que $Y_i \sim \text{BN}(\mu_i, \sigma_i)$. O modelo de Regressão Binomial-Negativo é definido por

$$\begin{aligned} Y_i &\sim \text{BN}(\mu_i, \sigma_i), \\ g_\mu(\mu_i) &= \mathbf{x}_{\mu i}^\top \boldsymbol{\beta}_\mu, & i = 1, 2, \dots, n, \\ g_\sigma(\sigma_i) &= \mathbf{x}_{\sigma i}^\top \boldsymbol{\beta}_\sigma, \end{aligned}$$

em que $\mathbf{x}_{\mu i} = (1, x_{\mu 1i}, x_{\mu 2i}, \dots, x_{\mu k_\mu i})^\top$ e $\mathbf{x}_{\sigma i} = (1, x_{\sigma 1i}, x_{\sigma 2i}, \dots, x_{\sigma k_\sigma i})^\top$ denotam os vetores de covariáveis associadas aos parâmetros μ e σ , respectivamente, para o indivíduo i ; $\boldsymbol{\beta}_\mu = (\beta_{\mu 0}, \beta_{\mu 1}, \dots, \beta_{\mu k_\mu})^\top$ e $\boldsymbol{\beta}_\sigma = (\beta_{\sigma 0}, \beta_{\sigma 1}, \dots, \beta_{\sigma k_\sigma})^\top$ são os vetores de parâmetros associados às covariáveis referentes aos parâmetros μ e σ , respectivamente; $g_\mu(\cdot)$ e $g_\sigma(\cdot)$ denotam as funções de ligação (comumente a função \log , para ambas) que associam as componentes aleatória e sistemática do modelo.

A esperança e a variância de Y_i são expressas por

$$\mathbb{E}(Y_i) = \mu_i, \quad \text{e} \quad \text{Var}(Y_i) = \mu_i + \sigma_i \mu_i^2,$$

onde, $\mu_i = g_\mu^{-1}(\mathbf{x}_{\mu i}^\top \boldsymbol{\beta}_\mu)$ e $\sigma_i = g_\sigma^{-1}(\mathbf{x}_{\sigma i}^\top \boldsymbol{\beta}_\sigma)$.

Uma estimativa para $\boldsymbol{\beta} = (\boldsymbol{\beta}_\mu^\top, \boldsymbol{\beta}_\sigma^\top)^\top$ pode ser obtida pela determinação do ponto $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_\mu^\top, \hat{\boldsymbol{\beta}}_\sigma^\top)^\top$ que maximiza a função de log-verossimilhança do modelo de Regressão Binomial-Negativo, que é expressa por

$$\begin{aligned} l(\boldsymbol{\beta}; \mathbf{Y}) &= \sum_{i=1}^n \left\{ \log \left[\Gamma \left(y_i + \frac{1}{\sigma_i} \right) \right] - \log \left[\Gamma \left(\frac{1}{\sigma_i} \right) \right] - \log \left[\Gamma (y_i + 1) \right] + \right. \\ &\quad \left. + y_i \left[\log (\sigma_i) + \log (\mu_i) - \log (1 + \sigma_i \mu_i) \right] - \frac{1}{\sigma_i} \log (1 + \sigma_i \mu_i) \right\}, \end{aligned}$$

com $\mu_i = g_\mu^{-1}(\mathbf{x}_{\mu i}^\top \boldsymbol{\beta}_\mu)$ e $\sigma_i = g_\sigma^{-1}(\mathbf{x}_{\sigma i}^\top \boldsymbol{\beta}_\sigma)$.

2.3.3 Modelo Beta-Binomial

A hierarquia que define o modelo Beta-Binomial parte da ideia de uma variável aleatória com distribuição $\text{BI}(n, p)$ condicionada à suposição de que a quantidade p é uma realização de uma variável aleatória P com distribuição $\text{BE}(\alpha, \beta)$, e então, determina-se a probabilidade marginal de Y (SCHERVISH, 1995; NAVARRO; PERFORS, 2012). A hierarquia desse modelo é expressa por

$$\begin{aligned} Y | P = p &\sim \text{BI}(n, p) \\ P &\sim \text{BE}(\mu, \sigma), \end{aligned}$$

deseja-se determinar a distribuição marginal de Y . Para isso, tem-se que

$$\begin{aligned}
 \Pr(Y = y, P = p) &= \Pr(Y = y | P = p) \Pr(P = p) \\
 &= \underbrace{\Pr(y; n, p)}_{\text{BI}(n,p)} \underbrace{f(p; \mu, \sigma)}_{\text{BE}(\mu,\sigma)} \\
 &= \binom{n}{y} p^y (1-p)^{n-y} \frac{\Gamma\left(\frac{1}{\sigma}\right)}{\Gamma\left(\frac{\mu}{\sigma}\right) \Gamma\left(\frac{1-\mu}{\sigma}\right)} p^{\frac{\mu}{\sigma}-1} (1-p)^{\frac{1-\mu}{\sigma}-1} \\
 &= \frac{\Gamma(n+1)}{\Gamma(y+1)\Gamma(n-y+1)} \frac{\Gamma\left(\frac{1}{\sigma}\right)}{\Gamma\left(\frac{\mu}{\sigma}\right) \Gamma\left(\frac{1-\mu}{\sigma}\right)} p^{y+\frac{\mu}{\sigma}-1} (1-p)^{n-y+\frac{1-\mu}{\sigma}-1},
 \end{aligned}$$

e portanto, uma expressão para a função de probabilidade marginal de Y é

$$\begin{aligned}
 \Pr(Y = y) &= \int_0^1 \frac{\Gamma(n+1)}{\Gamma(y+1)\Gamma(n-y+1)} \frac{\Gamma\left(\frac{1}{\sigma}\right)}{\Gamma\left(\frac{\mu}{\sigma}\right) \Gamma\left(\frac{1-\mu}{\sigma}\right)} p^{y+\frac{\mu}{\sigma}-1} (1-p)^{n-y+\frac{1-\mu}{\sigma}-1} dp \\
 &= \frac{\Gamma(n+1)}{\Gamma(y+1)\Gamma(n-y+1)} \frac{\Gamma\left(\frac{1}{\sigma}\right)}{\Gamma\left(\frac{\mu}{\sigma}\right) \Gamma\left(\frac{1-\mu}{\sigma}\right)} \int_0^1 \underbrace{p^{y+\frac{\mu}{\sigma}-1} (1-p)^{n-y+\frac{1-\mu}{\sigma}-1}}_{\text{Núcleo de BE}(\mu,\sigma)} dp \\
 &= \frac{\Gamma(n+1)}{\Gamma(y+1)\Gamma(n-y+1)} \frac{\Gamma\left(\frac{1}{\sigma}\right)}{\Gamma\left(\frac{\mu}{\sigma}\right) \Gamma\left(\frac{1-\mu}{\sigma}\right)} \frac{\Gamma\left(y+\frac{\mu}{\sigma}\right) \Gamma\left(n-y+\frac{1-\mu}{\sigma}\right)}{\Gamma\left(n+\frac{1}{\sigma}\right)}.
 \end{aligned}$$

Desta forma, sob essas considerações, a distribuição marginal de Y tem função de probabilidade dada por

$$\Pr(Y = y) = \frac{\Gamma(n+1)}{\Gamma(y+1)\Gamma(n-y+1)} \frac{\Gamma\left(\frac{1}{\sigma}\right) \Gamma\left(y+\frac{\mu}{\sigma}\right) \Gamma\left(n+\frac{1-\mu}{\sigma}-y\right)}{\Gamma\left(n+\frac{1}{\sigma}\right) \Gamma\left(\frac{\mu}{\sigma}\right) \Gamma\left(\frac{1-\mu}{\sigma}\right)},$$

em que $y = 0, 1, 2, \dots, 0 < \mu < 1$ e $\sigma > 0$.

A esperança de Y é expressa por

$$\mathbb{E}(Y) = \mathbb{E}[\mathbb{E}(Y|P)] = \mathbb{E}(nP) = n\mu,$$

e a variância, por

$$\begin{aligned}
\text{Var}(Y) &= \mathbb{E}[\text{Var}(Y|P)] + \text{Var}[\mathbb{E}(Y|P)] \\
&= \mathbb{E}[nP(1-P)] + \text{Var}[nP] \\
&= n\mathbb{E}(P) - n\mathbb{E}(P^2) + n^2 \text{Var}(P) \\
&= n\mu - n \left\{ \text{Var}(P) + [\mathbb{E}(P)]^2 \right\} + n^2 \text{Var}(P) \\
&= n\mu - n \left\{ \frac{\sigma}{\sigma+1} \mu(1-\mu) + \mu^2 \right\} + \frac{\sigma}{\sigma+1} \mu(1-\mu)n^2 \\
&= n\mu(1-\mu) \left[1 + \frac{\sigma}{1+\sigma}(n-1) \right].
\end{aligned}$$

Em relação ao modelo de Regressão Beta-Binomial, considerando-se um conjunto de observações independentes $Y_1, Y_2, Y_3, \dots, Y_n$, tal que $Y_i \sim \text{BB}(\mu_i, \sigma_i)$, o modelo de Regressão Beta-Binomial ficará definido por,

$$\begin{aligned}
Y_i &\sim \text{BB}(\mu_i, \sigma_i), \\
g_\mu(\mu_i) &= \mathbf{x}_{\mu i}^\top \boldsymbol{\beta}_\mu, & i = 1, 2, \dots, n, \\
g_\sigma(\sigma_i) &= \mathbf{x}_{\sigma i}^\top \boldsymbol{\beta}_\sigma,
\end{aligned}$$

em que $\mathbf{x}_{\mu i} = (1, x_{\mu 1i}, x_{\mu 2i}, \dots, x_{\mu k_\mu i})^\top$ e $\mathbf{x}_{\sigma i} = (1, x_{\sigma 1i}, x_{\sigma 2i}, \dots, x_{\sigma k_\sigma i})^\top$ denotam os vetores de covariáveis associadas aos parâmetros μ e σ , respectivamente, para o indivíduo i ; $\boldsymbol{\beta}_\mu = (\beta_{\mu 0}, \beta_{\mu 1}, \dots, \beta_{\mu k_\mu})^\top$ e $\boldsymbol{\beta}_\sigma = (\beta_{\sigma 0}, \beta_{\sigma 1}, \dots, \beta_{\sigma k_\sigma})^\top$ são os vetores de parâmetros associados às covariáveis referentes aos parâmetros μ e σ , respectivamente; $g_\mu(\cdot)$ e $g_\sigma(\cdot)$ denotam as funções de ligação (comumente as funções *logit* e *log*) que associam as componentes aleatória e sistemática do modelo.

A esperança e a variância de Y_i ficam expressas por

$$\mathbb{E}(Y_i) = n\mu_i \quad \text{e} \quad \text{Var}[Y_i] = n_i \mu(1-\mu_i) \left[1 + \frac{\sigma_i}{1+\sigma_i}(n_i-1) \right].$$

onde, $\mu_i = g_\mu^{-1}(\mathbf{x}_{\mu i}^\top \boldsymbol{\beta}_\mu)$ e $\sigma_i = g_\sigma^{-1}(\mathbf{x}_{\sigma i}^\top \boldsymbol{\beta}_\sigma)$.

Uma estimativa para $\boldsymbol{\beta} = (\boldsymbol{\beta}_\mu^\top, \boldsymbol{\beta}_\sigma^\top)^\top$ pode ser obtida pela determinação do ponto $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_\mu^\top, \hat{\boldsymbol{\beta}}_\sigma^\top)^\top$ que maximiza a função de log-verossimilhança do modelo de Regressão Binomial-Negativo, que é expressa por

$$\begin{aligned}
l(\boldsymbol{\beta}; \mathbf{Y}) &= \sum_{i=1}^n \left\{ \log [\Gamma(n_i + 1)] - \log [\Gamma(y_i + 1)] - \log [\Gamma(n_i - y_i + 1)] + \right. \\
&\quad + \log \left[\Gamma\left(\frac{1}{\sigma_i}\right) \right] + \log \left[\Gamma\left(y_i + \frac{\mu_i}{\sigma_i}\right) \right] + \log \left[\Gamma\left(n_i + \frac{1-\mu_i}{\sigma_i} - y_i\right) \right] + \\
&\quad \left. - \log \left[\Gamma\left(n_i + \frac{1}{\sigma_i}\right) \right] - \log \left[\Gamma\left(\frac{\mu_i}{\sigma_i}\right) \right] - \log \left[\Gamma\left(\frac{1-\mu_i}{\sigma_i}\right) \right] \right\},
\end{aligned}$$

com $\mu_i = g_\mu^{-1}(\mathbf{x}_{\mu i}^\top \boldsymbol{\beta}_\mu)$ e $\sigma_i = g_\sigma^{-1}(\mathbf{x}_{\sigma i}^\top \boldsymbol{\beta}_\sigma)$.

2.3.4 Modelo Binomial-Poisson-Generalizado

Levando-se em consideração o exposto para os modelos Binomial-Poisson, Binomial-Negativo e Beta-Binomial, e lembrando-se que a hierarquia que define o modelo Binomial-Poisson-Generalizado é dada por

$$\begin{aligned} Y | N = n &\sim \text{BI}(n, p) \\ N &\sim \text{PG}(\mu, \phi), \end{aligned}$$

deseja-se agora, encontrar a distribuição marginal de Y . Para isso, veja que

$$\begin{aligned} \Pr(Y = y, N = n) &= \Pr(Y = y | N = n) \Pr(N = n) \\ &= \underbrace{\Pr(y; n, p)}_{\text{BI}(n,p)} \underbrace{\Pr(n; \mu, \phi)}_{\text{PG}(\mu,\phi)} \\ &= \binom{n}{y} p^y (1-p)^{n-y} \frac{\mu [\mu + (\phi - 1)n]^{n-1} e^{-\frac{\mu + (\phi-1)n}{\phi}}}{n! \phi^n}, \end{aligned}$$

e portanto, a distribuição marginal de Y é determinada por

$$\begin{aligned} \Pr(Y = y) &= \sum_{n=y}^{\infty} \Pr(Y = y, N = n) \\ &= \sum_{n=y}^{\infty} \binom{n}{y} p^y (1-p)^{n-y} \frac{\mu [\mu + (\phi - 1)n]^{n-1} e^{-\frac{\mu + (\phi-1)n}{\phi}}}{n! \phi^n}. \end{aligned}$$

Diferente do modelo Binomial-Poisson, em que foi possível determinar o resultado da soma infinita necessária para encontrar uma expressão fechada para $\Pr(Y = y)$, neste caso, os termos envolvidos no somatório são muito difíceis de serem trabalhados, ou seja, não se consegue uma expressão fechada para a função de probabilidade marginal de Y . Assim, neste modelo em especial, será dado um enfoque bayesiano.

Ainda que a função densidade de probabilidade de Y seja desconhecida, é possível determinar seu valor esperado e variância. Dessa forma

$$\mathbb{E}(Y) = \mathbb{E}[\mathbb{E}(Y|N)] = \mathbb{E}[Np] = p\mu,$$

e também

$$\begin{aligned} \text{Var}[Y] &= \mathbb{E}[\text{Var}(Y|N)] + \text{Var}[\mathbb{E}(Y|N)] \\ &= \mathbb{E}[Np(1-p)] + \text{Var}(Np) \\ &= p(1-p)\mu + p^2\mu\phi^2. \end{aligned}$$

2.4 Inferência Bayesiana

A metodologia Bayesiana se baseia no fato de que toda incerteza precisa ser modelada por meio de probabilidades e, desta maneira, a inferência estatística tira conclusões com base nas leis de probabilidade (CHRISTENSEN et al., 2010). Na inferência Bayesiana o parâmetro de interesse é tomado como uma quantidade aleatória, tendo uma distribuição de probabilidades associada, isso distingue a inferência Bayesiana da Inferência Frequentista. A combinação de informação a priori com a informação proveniente dos dados amostrais, através do Teorema de Bayes, origina a distribuição a posteriori. Por meio da distribuição a posteriori pode-se observar a comportamento dos parâmetros após a inclusão das novas informações, se esse comportamento for adequado concluí-se sobre os resultados.

2.4.1 Teorema de Bayes

Considerando um espaço de probabilidades (Ω, E, \Pr) tal que uma partição do espaço amostral Ω é formada pelos eventos E_1, \dots, E_n , em que para todo $i = 1, \dots, n$ os eventos E_i são dois a dois disjuntos e $\Pr(E_i) > 0$, além disso, $\cup_{i=1}^n E_i = \Omega$. Dado um outro evento B qualquer, $B \subset \Omega$, e para todo $j = 1, 2, \dots, n$, tem-se que o Teorema de Bayes é dado por (deGroot & Schervish, 2011)

$$\Pr(E_i|B) = \frac{\Pr(B|E_i) \Pr(E_i)}{\Pr(B)} = \frac{\Pr(B|E_i) \Pr(E_i)}{\sum_{j=1}^n \Pr(B|E_j) \Pr(E_j)}, \quad i = 1, \dots, n.$$

O Teorema de Bayes pode ser interpretado como uma atualização de uma informação inicial e subjetiva acerca do evento E_i , representada pelas probabilidades $\Pr(E_i)$, e essa probabilidade será modificada assim que a ocorrência do evento B for conhecida. Com isso é possível perceber que a ocorrência do evento B modificará as probabilidades *a priori* sobre E_i , e esse resultado é conhecido por probabilidades *a posteriori*, denotada por $\Pr(E_i|B)$.

2.4.2 Distribuição a Posteriori

Considere-se Y_1, \dots, Y_n uma amostra aleatória e independente da variável aleatória Y , que tem uma dada distribuição de probabilidade associada a um parâmetro θ desconhecido. Qualquer informação subjetiva atribuída a esse parâmetro é chamada de informação *a priori*, e essa informação é quantificada formalmente por uma distribuição de probabilidades $\pi(\theta)$ *a priori* para o parâmetro θ (BOX; TIAO, 1992).

Probabilisticamente, a distribuição *a priori* representa qualquer conhecimento prévio acerca de θ , anterior a análise sobre os dados. Por meio, do Teorema de Bayes a distribuição $\pi(\theta)$ é combinada com a distribuição dos dados amostrados e, desta forma, origina a distribuição *a posteriori*, denotada por $\pi(\theta|Y)$, o que permite realizar toda a inferência.

2.4.3 Seleção de Modelos

A seleção de modelos constitui uma importante ferramenta da modelagem estatística, por meio dela o pesquisador tem um embasamento adicional para justificar a escolha de um modelo adequado. Para comparar os modelos ajustados nesta pesquisa utilizou-se o Critério de Informação da Deviance (DIC), Critério de Informação de Akaike Estendido (EAIC) e o Critério de Informação Bayesiano Estendido (EBIC).

Com base nesses critérios, escolhe-se o modelo que apresentar a menor medida. Definindo $D(\bar{\theta})$ como a *média a posteriori da deviance* e $D(\hat{\theta})$ como a *medida a posteriori da qualidade do ajuste do modelo para os dados*, as definições de cada critério são as seguintes:

- DIC (Spiegelhalter, Best, Carlin, & Linde, 2002): Um dos mais utilizados dentro da estatística Bayesiana, definido por $DIC = 2D(\bar{\theta}) - D(\hat{\theta})$.
- EAIC (Brooks S. , 2002): Definido como $EAIC = D(\bar{\theta}) + 2k$, sendo k o número de parâmetros.
- EBIC (Carlin & Louis, 2008): Definido por $EBIC = D(\bar{\theta}) + k \ln(n)$, sendo k o número efetivo de parâmetros e n o tamanho da amostra.

2.4.4 Critérios de Convergência

Para verificar a convergência das cadeias utilizou-se dois critérios: o critério de Geweke (GEWEKE, 1992), que baseia-se no teste de igualdade das médias da primeira e última cadeia de Markov. Sob a hipótese nula de convergência, assume-se que as amostras são tomadas de uma distribuição estacionária da cadeia; o critério de Heidelberger e Welch (HEIDELBERGER; WELCH, 1983), testa a hipótese nula de estacionariedade. Se a hipótese nula for rejeitada para um dado valor, repete-se o teste desconsiderando 10% das primeiras iterações. Caso a rejeição ocorra novamente, outras 10% iniciais são descartadas e o teste é repetido. Isso é feito até eliminar 50% da cadeia. Ao final, se a hipótese nula continuar sendo rejeitada, tem-se um indício da falta de estacionariedade, sendo então necessário o aumento do número de interação. Mais detalhes sobre esse critério podem ser vistos em (BROOKS; ROBERTS, 1998).

2.4.5 Procedimento Computacional

A obtenção das distribuições *a posteriori* para os parâmetros se deu com auxílio do pacote `rjags` (Plummer, R package version 4.6, 2016) do *software* R (R Core, 2016). Foram gerados 50000 valores em um processo MCMC, descartando os 10000 valores iniciais, a amostra foi construída considerando saltos de tamanho 10, com o intuito de eliminar autocorrelação. Ao final do processo, para cada parâmetro, obteve-se uma cadeia

resultante com 4000 observações, cuja convergência foi atestada com base nos critérios de Gelman-Rubin (incluindo a medida multivariada proposta por Gelman-Brooks), Geweke e também Heidelberger e Welch implementados no pacote *coda* (Plummer, Best, Cowles, & Vines, 2006) do *software* R. Os parâmetros foram considerados significativos se o valor zero não pertencesse aos seus respectivos intervalos com 95% de credibilidade. Para qualquer parâmetro θ , foi atribuída uma distribuição *a priori* não informativa, isto é $\theta \sim N(0, 100^2)$.

Capítulo 3

Resultados

Nesta seção será apresentada uma análise descritiva dos dados para que haja um correto direcionamento/posicionamento quanto ao desfecho de possíveis modelagens a serem aqui utilizadas.

3.1 Análise Descritiva

Uma vez que os dados são provenientes de uma contagem, é comum que a primeira distribuição a ser ajustada seja a distribuição de Poisson. Entretanto, como pode ser observado na Tabela 2, tanto as observações de oócitos aspirados quanto as de oócitos viáveis são sobredispersas, o que compromete o ajuste através desta distribuição (e quaisquer outras que não contemplem esta situação), pois, esta exige igualdade entre a média e a variância. Isso faz com que tenha-se que recorrer a outros modelos de probabilidade.

Pode-se notar também (Tabela 2) que, de modo geral, a proporção média de oócitos viáveis é próxima de 87%. Além disso, a variabilidade entre o total de oócitos aspirados e viáveis, indicada pelo coeficiente de variação, são muito próximas, tanto para o total coletado no decorrer do experimento quanto a quantidade coletada por aspiração.

Tabela 2 – Medidas Descritivas da Quantidades de Oócitos Aspirados e Viáveis

Quantidade de Oócitos		Mínimo	Máximo	Média	Variância	Desvio Padrão	Coeficiente de Variação
em todo o experimento	Aspirados	30,00	259,00	128,43	4233,90	65,07	0,5066
	Viáveis	24,00	245,00	111,83	3617,59	60,15	0,5378
por aspiração	Aspirados	3,00	84,00	26,03	212,47	14,58	0,5599
	Viáveis	2,00	79,00	22,67	190,21	13,79	0,6084

Observe na Figura 1 que as proporções de oócitos viáveis entre as doadoras que receberam o Tratamento 1 não esboçam muita variabilidade, independente da doadora e

da quantidade de aspirações realizadas. O mesmo ocorre com as doadoras que receberam o Tratamento 2 (Figura 2). Além disso, as taxas médias de aspiração também são semelhantes, em torno de 20 e 25, para oócitos aspirados e viáveis, respectivamente, independente do tratamento recebido.

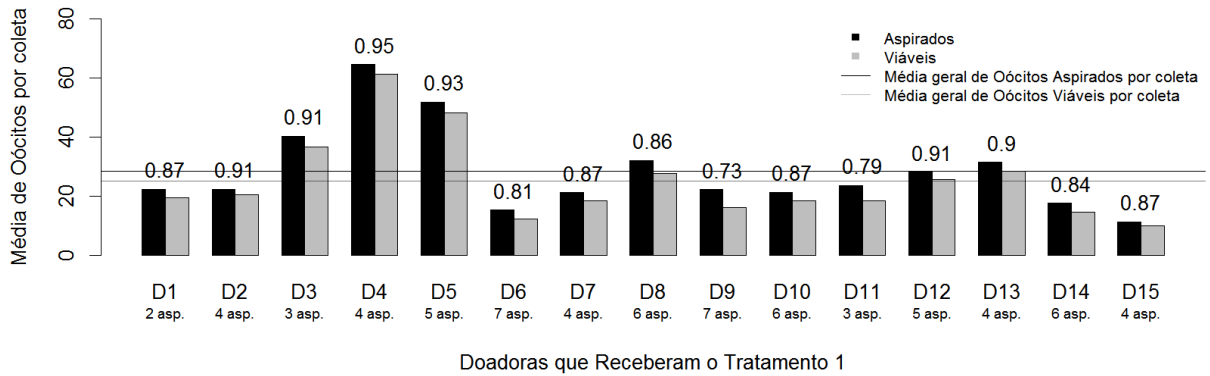


Figura 1 – Média de oócitos aspirados ($M_A = N/o$) e viáveis ($M_V = Y/o$) para as doadoras que receberam o Tratamento 1. Os valores numéricos acima das barras indicam a proporção de oócitos viáveis ($p_V = Y/N$).

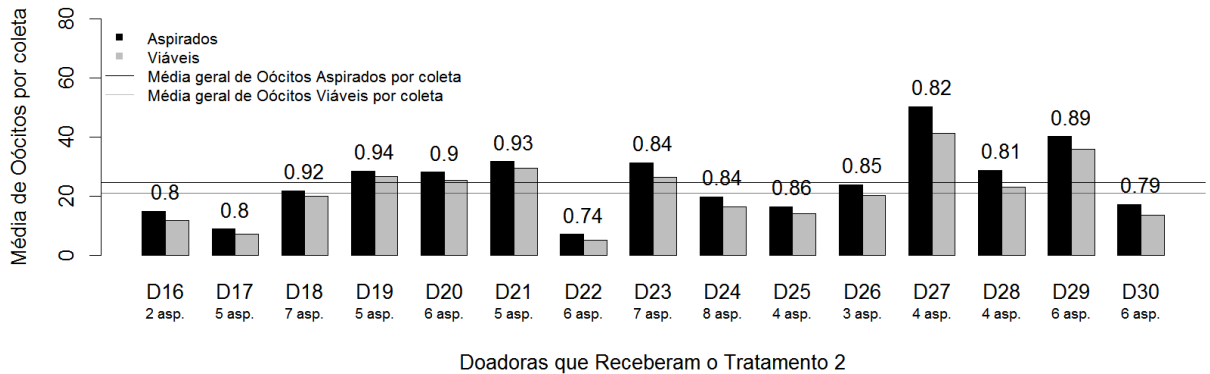


Figura 2 – Média de oócitos aspirados ($M_A = N/o$) e viáveis ($M_V = Y/o$) para as doadoras que receberam o Tratamento 2. Os valores numéricos acima das barras indicam a proporção de oócitos viáveis ($p_V = Y/N$).

O comportamento da quantidade de Oócitos Aspirados, Oócitos Viáveis e da Proporção de Oócitos Viáveis ao longo dos dias em que as coletas foram realizadas estão expostos na Figura 3 (a), (b) e (c), respectivamente. Pode-se notar que as tendências médias observadas são muito próximas, independente do tratamento. Além disso, as linhas de tendência média indicam, para as três medidas observadas, um leve declínio com o passar dos dias.

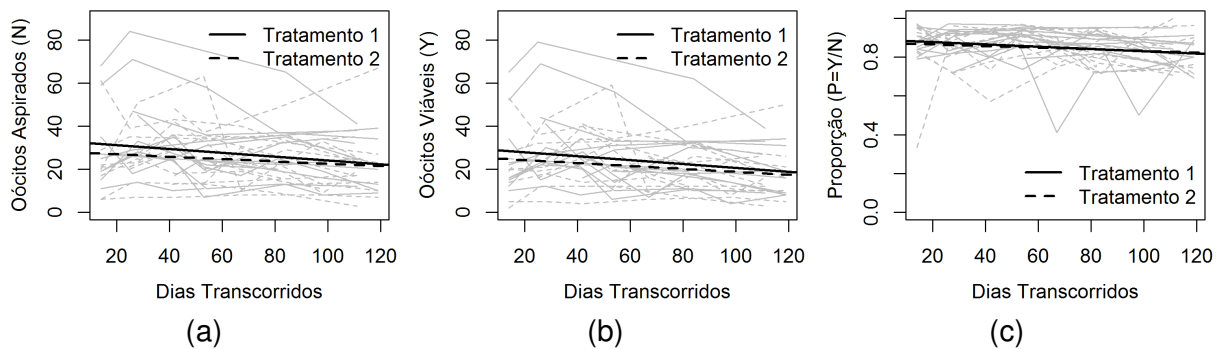


Figura 3 – Perfis individuais (em cinza) e linhas de tendência média (em preto) para o Total de Oócitos Aspirados (a), o Total de Oócitos Viáveis (b) e a Proporção de Oócitos Viáveis (c).

Do exposto anteriormente, para o ajuste é preciso considerar que:

- a adequação do modelo Poisson somente será possível devido à inclusão do efeito aleatório, inerente da modelagem longitudinal. Caso contrário, alternativas a essa modelagem se embasam em modelos que acomodam adequadamente a superdispersão inerente aos dados resumidos pelas somas (sem considerar a estrutura longitudinal), como Binomial-Negativo e Poisson-Generalizado;
- para se aproximar ainda mais da realidade, deve-se entender que tanto o total de oócitos aspirados quanto o de viáveis são quantidades aleatórias. Essa premissa só é considerada nos modelos Binomial-Poisson e Binomial-Poisson-Generalizada, uma vez que somente tais modelos subentendem essa hierarquia;
- se for entendido que, além do total de oócitos viáveis, sua proporção é, também, uma quantidade aleatória, um modelo adequado pode ser o Beta-Binomial.

3.2 Modelos Ajustados com Metodologia Frequentista

3.2.1 Ajuste com base no Modelo de Regressão Binomial

Sob a suposição de que a proporção de oócitos viáveis, dentre os aspirados, se mantém constante (exceto pelo efeito da covariável) independente da quantidade de coletas realizadas, da quantidade de oócitos aspirados em cada coleta, e até mesmo da doadora, é possível entender que o número total de oócitos viáveis por doadora tem comportamento regido por uma distribuição Binomial com parâmetro p . Veja que todas estas suposições não parecem demasiadas, uma vez que tais características foram observadas na análise descritiva. Entretanto, a questão da variabilidade fica ainda em aberto.

O Modelo de Regressão Binomial para este caso pode ser definido por

$$Y_i \sim \text{BI}(n_i, p_i),$$

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 T_i,$$

em que Y_i , com $i = 1, \dots, 30$, representa o total de oócitos viáveis, em uma amostra de tamanho n_i , associado à i -ésima doadora, cuja proporção de oócitos viáveis é igual a p_i . A relação funcional entre essa proporção e a covariável 'tratamento' é estabelecida pela função de ligação *logit*, que associa o tratamento recebido pela i -ésima doadora, denotado por T_i , ao parâmetro p_i , da distribuição $\text{BI}(n_i, p_i)$.

A proporção estimada de oócitos viáveis, \hat{p}_i , pode ser obtida pela relação inversa

$$\hat{p}_i = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 T_i}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 T_i}},$$

de onde segue que o valor esperado e a medida de variabilidade para a quantidade de oócitos estimada são dados, respectivamente, por

$$\hat{\mathbb{E}}(Y_i) = n_i \hat{p}_i \quad \text{e} \quad \widehat{\text{Var}}(Y_i) = n_i \hat{p}_i (1 - \hat{p}_i).$$

As estimativas obtidas para o ajuste estão expostas na Tabela 3. Com base no teste- t associado ao parâmetro β_1 (cujo valor- p é igual a $0.195 > 0.05$), entende-se que a influência exercida pelo tratamento 2 não difere estatisticamente da exercida pelo tratamento 1.

Tabela 3 – Estimativas obtidas para o ajuste do Modelo Binomial.

	Estimativa	Erro Padrão	Valor- t	Valor- p
β_0	1,97279	0,06947	28,396	<2e-16
β_1	-0,12776	0,09616	-1,329	0,195

Uma vez que o parâmetro β_1 se mostrou não significativo para o modelo, um novo ajuste, desconsiderando a influência do tratamento, obteve a estimativa $\hat{\beta}_0 = 1,90761$ com erro padrão $se(\hat{\beta}_0) = 0,04802$. Nesse caso, tem-se que uma estimativa para a proporção populacional de oócitos viáveis é $\hat{p} = 0,87075$.

A qualidade do ajuste pode ser percebida com base nos gráficos dos resíduos aleatorizados resultantes do processo de ajuste. Os gráficos diagnósticos estão expostos na Figura 4, onde podem ser vistos: os gráficos dos Resíduos vs. Valores Ajustados (a) e Resíduos vs. Índice da Observação (b) que indicam uma dispersão levemente superior associada aos resíduos para as observações provenientes das doadoras tratadas com o segundo tratamento (com destaque para as doadoras 25 e 30, possivelmente *outliers*). Os

gráficos para a Densidade Estimada (c) e o Q-Q Plot Normal (d) dos resíduos indicam que o modelo não está bem ajustado.

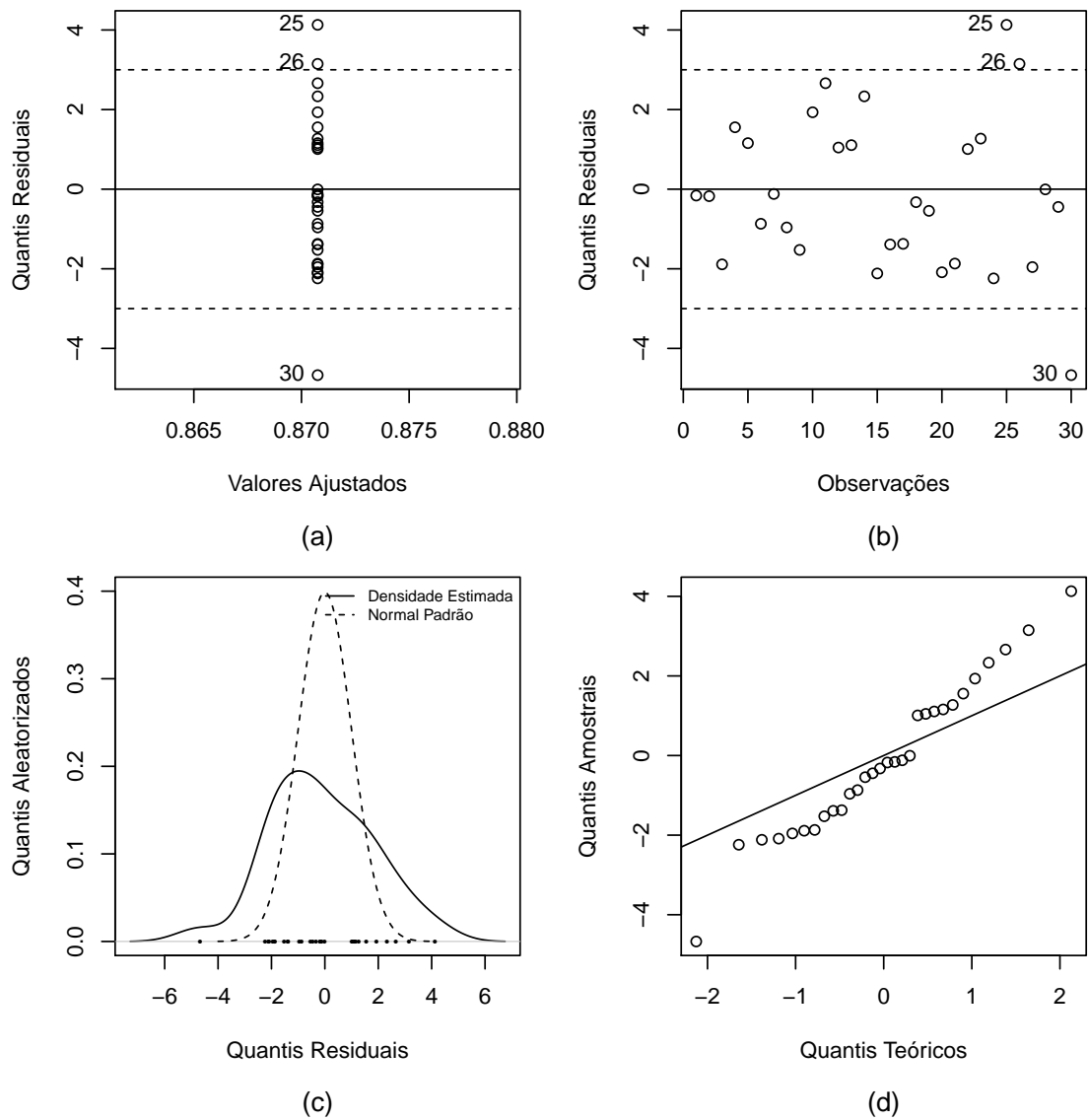


Figura 4 – Gráficos Diagnósticos para o Ajuste do Modelo Binomial

O mau comportamento dos resíduos aleatorizados pode ser constatado com base no gráfico Worm-plot (Figura 5) cuja disposição dos pontos, próximos a uma “reta” crescente, evidencia alta variabilidade residual, apontando que o modelo pode precisar de uma componente de variância para acomodar a superdispersão inerente aos dados de oócitos.

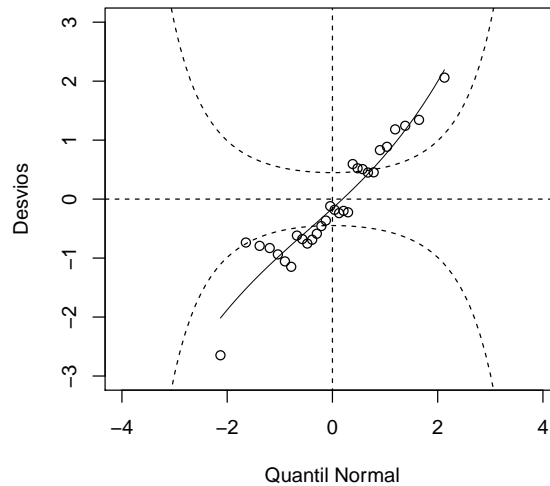


Figura 5 – Worm-plot do ajuste pelo Modelo Binomial

Na Tabela 4 são apresentadas algumas medidas resumo para os resíduos aleatorizados. Uma vez que, com um bom ajuste, é esperado que tais resíduos tenham distribuição normal padrão, percebe-se que o ajuste do Modelo Binomial não é suficiente para entender o comportamento do número de óocitos viáveis.

Tabela 4 – Medidas Resumo para os Quantis Residuais - Modelo Binomial

	Medida	Estimativa
	Média	-0,10584
	Variância	3,70691
	Coefficiente de Assimetria	-0,06061
	Coefficiente de Curtose	2,76563

3.2.2 Ajuste com base no Modelo de Regressão Poisson

O modelo de Regressão Poisson apresenta uma suposição ainda mais problemática do que as adotadas para o modelo de Regressão Binomial. Nesse caso, por hipótese, a variância do número total de óocitos viáveis deve ser igual à sua média. Como foi observado no decorrer da análise descritiva, este não é o caso dos dados aqui estudados. No entanto, a título de exploração, considere o modelo de Regressão Poisson, para a taxa de óocitos viáveis, definido por

$$Y_i \sim PO(\lambda_i),$$

$$\log\left(\frac{\lambda_i}{o_i}\right) = \beta_0 + \beta_1 T_i,$$

em que Y_i , com $i = 1, \dots, 30$, representa o total de oócitos viáveis, em uma quantidade o_i de coletas, associado à i -ésima doadora, cuja quantidade de oócitos viáveis esperada é igual a λ_i . A relação funcional entre λ_i e a covariável 'tratamento' é estabelecida pela função de ligação *log*, que associa o tratamento recebido pela i -ésima doadora, denotado por T_i , ao parâmetro λ_i , da distribuição $PO(\lambda_i)$.

O número estimado de oócitos viáveis, $\hat{\lambda}_i$, é obtido pela relação inversa

$$\hat{\lambda}_i = o_i e^{\hat{\beta}_0} e^{\hat{\beta}_1 T_i},$$

segue, portanto, que o valor esperado, a medida de variabilidade para a quantidade de oócitos estimada são dados, respectivamente, por

$$\hat{\mathbb{E}}(Y_i) = \hat{\lambda}_i = \widehat{\text{Var}}(Y_i).$$

e ainda, a taxa de oócitos viáveis pode ser estimada por

$$\hat{\mathbb{E}}\left(\frac{Y_i}{o_i}\right) = \frac{1}{o_i} \hat{\mathbb{E}}(Y_i) = \frac{\hat{\lambda}_i}{o_i}.$$

As estimativas obtidas para esse ajuste estão expostas na Tabela 5. O teste-*t* associado ao parâmetro β_1 (valor-*p* é igual a $0.000703 < 0.05$) indica que a influência exercida pelo tratamento 2 é estatisticamente distinta da exercida pelo tratamento 1. Entretanto, qualquer conclusão só pode ser tomada após a validação do modelo.

Tabela 5 – Estimativas do Ajuste pelo Modelo de Regressão Poisson

	Estimativa	Erro Padrão	Valor- <i>t</i>	Valor- <i>p</i>
β_0	3,18812	0,02427	131,334	<2e-16
β_1	-0,13146	0,03453	-3,807	0,0007

Se o modelo for adequado, entende-se que o valor esperado de oócitos viáveis para as vacas tratadas com os tratamentos 1 e 2, respectivamente, é dado por

$$\hat{\lambda}_{\text{Trat1}} = o_i e^{\hat{\beta}_0} \approx 24,24286 o_i \quad \text{e} \quad \hat{\lambda}_{\text{Trat2}} = o_i e^{\hat{\beta}_0 + \hat{\beta}_1} \approx 21,25641 o_i,$$

e portanto, a taxa de oócitos viáveis por aspiração, para as doadoras sujeitas aos tratamentos 1 e 2, respectivamente, é igual a 24,24286 e 21,25641.

A Figura 6, que apresenta os gráficos de diagnóstico para o Modelo de Regressão Poisson, através dos gráficos Densidade Estimada (c) e Q-Q Plot Normal (d), deixa claro que os resíduos não apresentam normalidade. Além disso, os gráficos Quantis Residuais vs. Valores Ajustados (a) e Quantis Residuais vs. Observações (b) apresentam resíduos aleatorizados, mas nada próximos do que se espera para uma distribuição normal padrão.

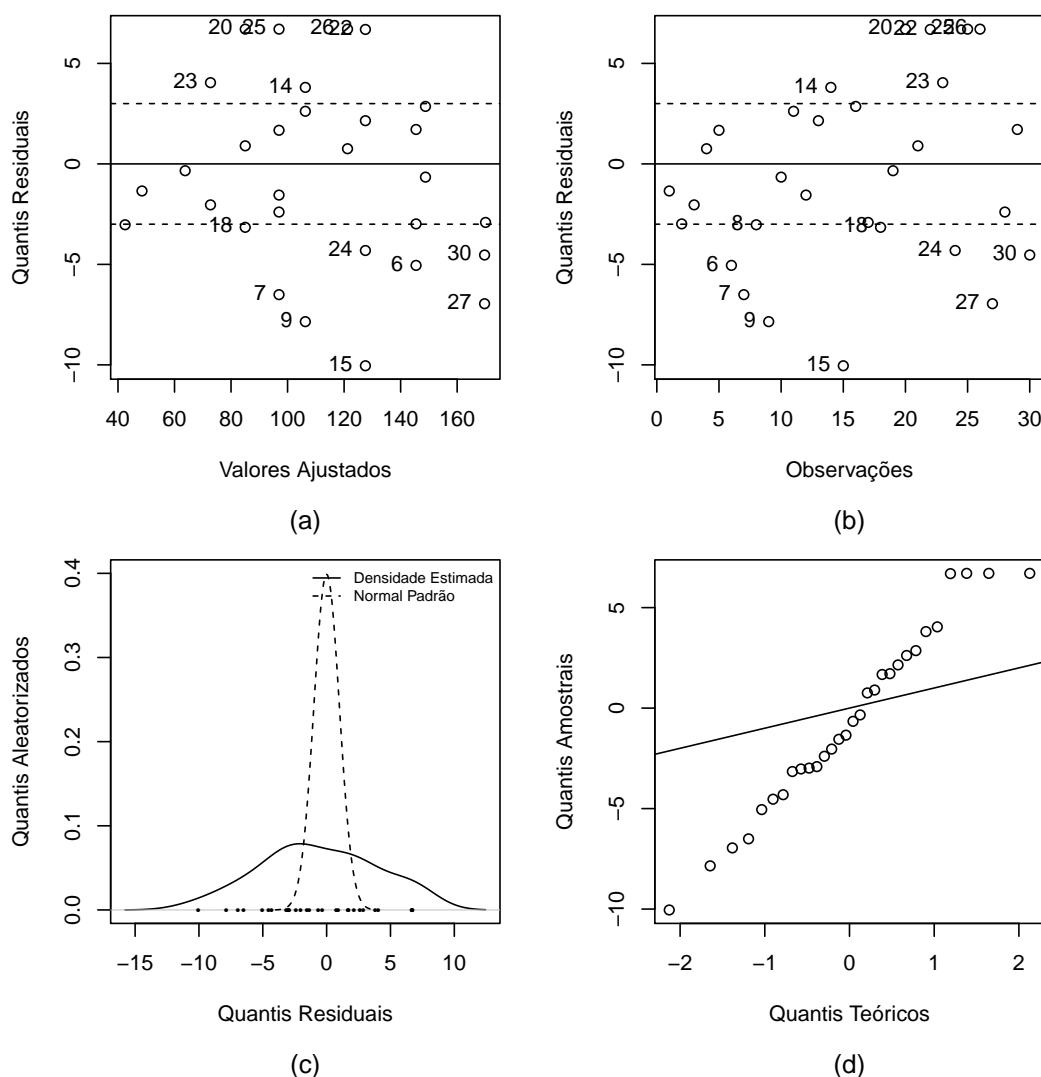


Figura 6 – Gráficos Diagnósticos para o Ajuste do Modelo Poisson

Com base nas medidas resumo da Tabela 6, nota-se que apenas o coeficiente de assimetria se aproxima do esperado. Além disso, novamente a superdispersão característica do fenômeno se evidencia no resíduo.

Tabela 6 – Medidas Resumo para os Quantis Residuais - Modelo Poisson

Medida	Estimativa
Média	- 0,6226
Variância	20,6025
Coefficiente de Assimetria	- 0,0204
Coefficiente de Curtose	2,1427

A superdispersão também é confirmada pelo gráfico Worm-Plot da Figura 7. Além disso, o comportamento em forma de S crescente (levemente percebido nesse caso) também indica que a cauda da distribuição ajustada está demasiadamente leve.

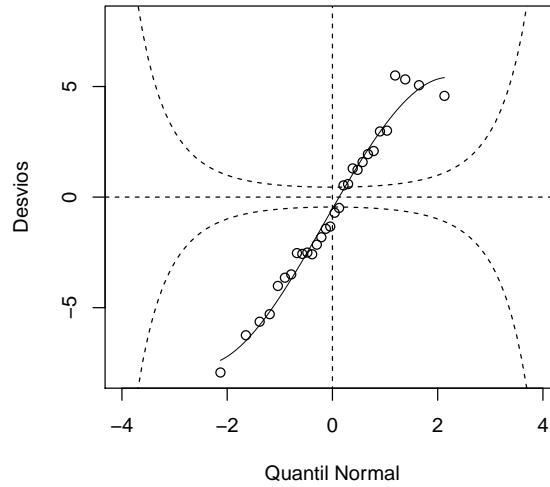


Figura 7 – Worm-plot do ajuste pelo Modelo Poisson

3.2.3 Ajuste com base no Modelo de Regressão Binomial-Poisson

O modelo de Regressão Binomial-Poisson inicialmente leva em conta todas as suposições aplicáveis ao modelo de Regressão Binomial. Entretanto, o modelo utiliza a informação proveniente das medidas de Y , o número de oócitos viáveis, acrescentando a ideia da aleatoriedade inerente de N , o número de oócitos aspirados, em um modelo hierárquico como o descrito na Seção 2.3.1.

Embora a distribuição resultante da composição (Poisson) esteja limitada às mesmas dificuldades discutidas anteriormente, este modelo é interessante no sentido em que possibilita a interpretação dos agentes influentes, tanto sobre a proporção de oócitos viáveis quanto total de oócitos aspirados. Um modelo como esse pode ser definido por

$$\begin{aligned} Y_i | N_i = n_i &\sim \text{BI}(n_i, p_i), \\ N_i &\sim \text{PO}(\lambda_i), \end{aligned} \iff Y_i \sim \text{PO}(\lambda_i p_i),$$

em que Y_i , com $i = 1, \dots, 30$, representa o total de oócitos viáveis associado à i -ésima doadora, cuja proporção de oócitos viáveis é igual a p_i . Além disso, N_i denota o número de oócitos aspirados, cujo valor esperado é λ_i .

As relações funcionais entre os parâmetros p_i e λ_i com a covariável 'tratamento' foram estabelecidas pelas funções de ligação *logit* e *log*, respectivamente, que associam o tratamento recebido pela i -ésima doadora, denotado por T_i , aos parâmetros p_i e λ_i , da

distribuição $PO(\lambda_i p_i)$.

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_{p0} + \beta_{p1} T_i,$$

$$\log\left(\frac{\lambda_i}{o_i}\right) = \beta_{\lambda0} + \beta_{\lambda1} T_i.$$

De maneira similar aos modelos de Regressão Binomial e Poisson, a proporção de oócitos viáveis e o número médio de oócitos aspirados podem ser estimados por \hat{p}_i e $\hat{\lambda}_i$, respectivamente, e são obtidos pelas relações inversas

$$\hat{p}_i = \frac{e^{\hat{\beta}_{p0} + \hat{\beta}_{p1} T_i}}{1 + e^{\hat{\beta}_{p0} + \hat{\beta}_{p1} T_i}} \quad \text{e} \quad \hat{\lambda}_i = o_i e^{\hat{\beta}_{\lambda0} + \hat{\beta}_{\lambda1} T_i},$$

de onde segue que uma estimativa para a esperança e a variância da quantidade de oócitos viáveis é dada por

$$\hat{\mathbb{E}}(Y_i) = \hat{\lambda}_i \hat{p}_i = \widehat{\text{Var}}(Y_i).$$

Na Tabela 7 podem ser vistas as estimativas obtidas para os parâmetros associados a p_i e λ_i , para ambos, o teste t associado indica que o tratamento 2 não exerce qualquer influência estatisticamente distinta do tratamento 1.

Tabela 7 – Estimativas do ajuste pelo Modelo Binomial-Poisson

	Estimativa	Erro Padrão	Valor- t	Valor- p
β_{p0}	-17,02417	0,0188	-905,400	<2e-16
β_{p1}	-0,01602	5811,7981	0,0000	1,0000
$\beta_{\lambda0}$	20,21229	0,01535	1317,0	<2e-16
$\beta_{\lambda1}$	-0,11545	5811,7981	0,0000	1,0000

Visto que os parâmetros β_{p1} e $\beta_{\lambda1}$ não foram considerados significativos, um novo ajuste foi realizado, desconsiderando-os. Os resultados do ajuste estão expostos na Tabela 8. Nesse caso, é possível estimar a proporção de oócitos viáveis, o valor médio de oócitos aspirados e viáveis,

$$\hat{p}_i = \frac{e^{\hat{\beta}_{p0}}}{1 + e^{\hat{\beta}_{p0}}} \approx 3,83154e - 08 \quad \text{e} \quad \hat{\lambda}_i = o_i e^{\hat{\beta}_{\lambda0}} \approx o_i 2.918.755.632 .$$

Tabela 8 – Estimativas para o segundo ajuste pelo Modelo Binomial-Poisson

	Estimativa	Erro Padrão	Valor- t	Valor- p
β_{p0}	-17,03064	0,01726	-986,500	<2e-16
$\beta_{\lambda0}$	20,15164	0,00198	10177,0	<2e-16

Veja que, mesmo antes de analisar os gráficos diagnósticos é possível perceber que o modelo não é adequado. As estimativas obtidas para p e λ são incapazes de, individualmente, representar a realidade. Isso pode ter ocorrido devido à provável não identificabilidade do modelo.

A Figura 8, assim como as Figuras 4 e 6, provenientes do modelo binomial e do modelo Poisson, mostra quatro gráficos onde nota-se que, o gráfico dos Quantis Residuais vs. Valores Ajustados (a) e (b); Quantis Residuais vs. Observações (c), indicam que os resíduos são aleatórios. Porém, Os gráficos da Densidade Estimada (d) e do Q-Q Plot Normal (e) apontam que os resíduos aleatorizados não seguem distribuição normal padrão. O diagnóstico com base no Worm-plot (f) é, também, análogo aos casos anteriores.

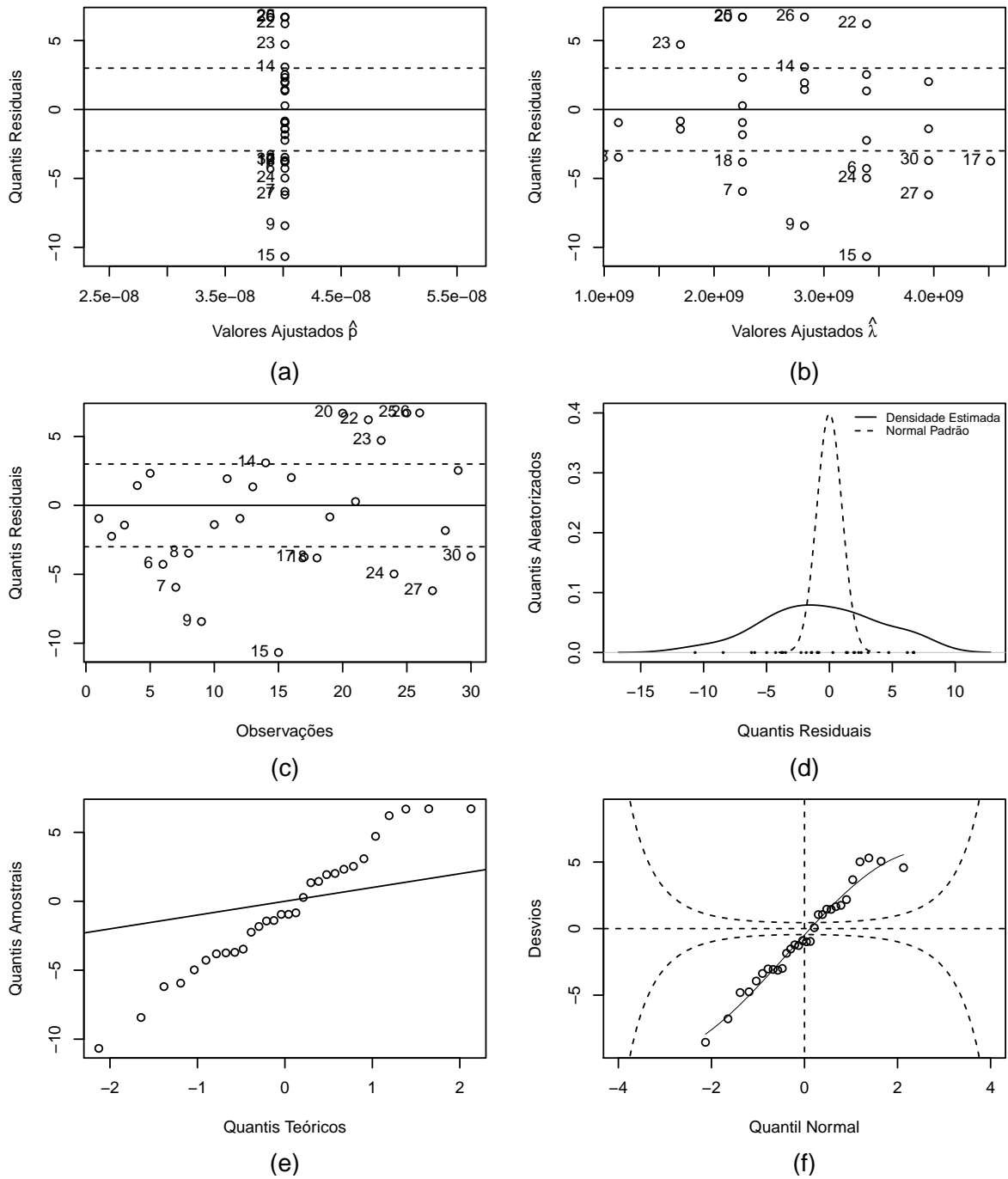


Figura 8 – Gráficos Diagnósticos para o Ajuste do Modelo Binomial-Poisson

As medidas resumo apresentadas na Figura 9, apresentam o único coeficiente, o de assimetria, próximo ao que se espera em uma distribuição normal padrão. As demais medidas além de serem discrepantes do que se almeja, indicam superdispersão e baixa curtose nos resíduos.

Tabela 9 – Medidas Resumo para os Quantis Residuais - Modelo Binomial-Poisson

	Medida	Estimativa
	Média	-0.13623
	Variância	20,59328
	Coefficiente de Assimetria	-0,02478564
	Coefficiente de Curtose	2,161135

3.2.4 Ajuste com base no modelo de Regressão Binomial-Negativo

O modelo de Regressão Binomial-Negativo caracteriza uma das possíveis maneiras de contornar a ocorrência de superdispersão em alguns conjuntos de dados, como o aqui estudado. Este modelo pode ser definido por

$$\begin{aligned}
 Y_i | \Lambda = \lambda &\sim \text{PO}(\mu_i \lambda), \\
 \Lambda &\sim \text{GA}(1, \sigma_i^{1/2}),
 \end{aligned}
 \iff
 Y_i \sim \text{BN}(\mu_i, \sigma_i),$$

em que Y_i , com $i = 1, \dots, 30$, representa o total de oócitos viáveis, em uma quantidade o_i de coletas, associado à i -ésima doadora, cuja quantidade esperada é dada por μ_i .

As relações funcionais entre os parâmetros μ_i e σ_i com a covariável 'tratamento' foram estabelecidas pela função de ligação *log*, para ambos os parâmetros. Uma vez que deseja-se verificar se o tratamento recebido tem influência sobre o número esperado de oócitos viáveis, somente o parâmetro μ_i foi associado ao tratamento recebido pela i -ésima doadora, denotado por T_i . De modo que as relações estabelecidas têm a forma

$$\begin{aligned}
 \log\left(\frac{\mu_i}{o_i}\right) &= \beta_{\mu 0} + \beta_{\mu 1} T_i, \\
 \log(\sigma_i) &= \beta_{\sigma 0},
 \end{aligned}$$

O número médio de oócitos viáveis, μ_i , e o parâmetro de escala, σ_i , podem ser estimados por $\hat{\mu}_i$ e $\hat{\sigma}_i$, respectivamente, e são obtidos pelas relações inversas

$$\hat{\mu}_i = o_i e^{\hat{\beta}_{\mu 0} + \hat{\beta}_{\mu 1} T_i} \quad \text{e} \quad \hat{\sigma}_i = e^{\hat{\beta}_{\sigma 0}},$$

segue, portanto, que as estimativas para a esperança e a variância da quantidade de oócitos viáveis são dadas, respectivamente, por

$$\widehat{\mathbb{E}}(Y_i) = \hat{\mu}_i \quad \text{e} \quad \widehat{\text{Var}}(Y_i) = \hat{\mu}_i + \hat{\sigma}_i \hat{\mu}_i^2.$$

Na Tabela 10 estão as estimativas para os parâmetros associados a μ_i e σ_i . O teste-*t* associado ao parâmetro $\beta_{\mu 1}$ atesta que o tratamento 2 não exerce qualquer influência estatisticamente distinta do tratamento 1.

Tabela 10 – Estimativas do Ajuste pelo Modelo Binomial-Negativo

	Estimativa	Erro Padrão	Valor- <i>t</i>	Valor- <i>p</i>
$\beta_{\mu 0}$	3,2244	0,1266	25,465	<2e-16
$\beta_{\mu 1}$	-0,1692	0,1791	-0,944	0,353
$\beta_{\sigma 0}$	-1,4661	0,2604	-5,63	5,64e-06

Uma vez que o parâmetro $\beta_{\mu 1}$ foi considerado não significativo, um novo modelo desconsiderando-o foi proposto, as estimativas resultantes estão na Tabela 11.

Tabela 11 – Estimativas do ajuste para o segundo modelo Binomial-Negativo

	Estimativa	Erro Padrão	Valor- <i>t</i>	Valor- <i>p</i>
$\beta_{\mu 0}$	3,14312	0,09078	34,62	<2e-16
$\beta_{\sigma 0}$	-1,4381	0,2599	-5,533	6,47e-06

Pode-se estimar a quantidade média de oócitos viáveis, via parâmetro μ_i , e o parâmetro de escala da distribuição $\text{BN}(\mu_i, \sigma_i)$. Suas estimativas são dadas por

$$\hat{\mu}_i = e^{\hat{\beta}_{\mu 0}} \approx 23,17599 o_i \quad \text{e} \quad \hat{\sigma}_i = e^{\hat{\beta}_{\sigma 0}} \approx 0,2373708 ,$$

e, portanto, uma vez que o modelo seja validado, a quantidade média de oócitos aspirados em uma coleta ($o = 1$), independente da doadora ter recebido o tratamento 1 ou 2, é igual a $23,17599 \approx 23$ unidades.

Ao contrário dos outros modelos (Binomial, Poisson e Binomial-Poisson), como pode ser visto na Figura 9, o modelo parece se ajustar bem às observações. Além de apresentar comportamentos aleatórios nos gráficos Quantis Residuais vs. Valores Ajustados para μ e σ (a) e (b) e Quantis Residuais vs. Observações (c), os gráficos para a densidade estimada (d) e o Q-Q Plot Normal (e) indicam que os resíduos são normalmente distribuídos.

A qualidade do ajuste pode também ser avaliada com base no Worm-plot, Figura 9 (f), esse gráfico não indica qualquer mau comportamento dos resíduos resultantes do processo de ajuste, uma vez que os pontos se concentram em torno do eixo horizontal (Desvio igual a zero) e não extrapolam as bandas de confiança.

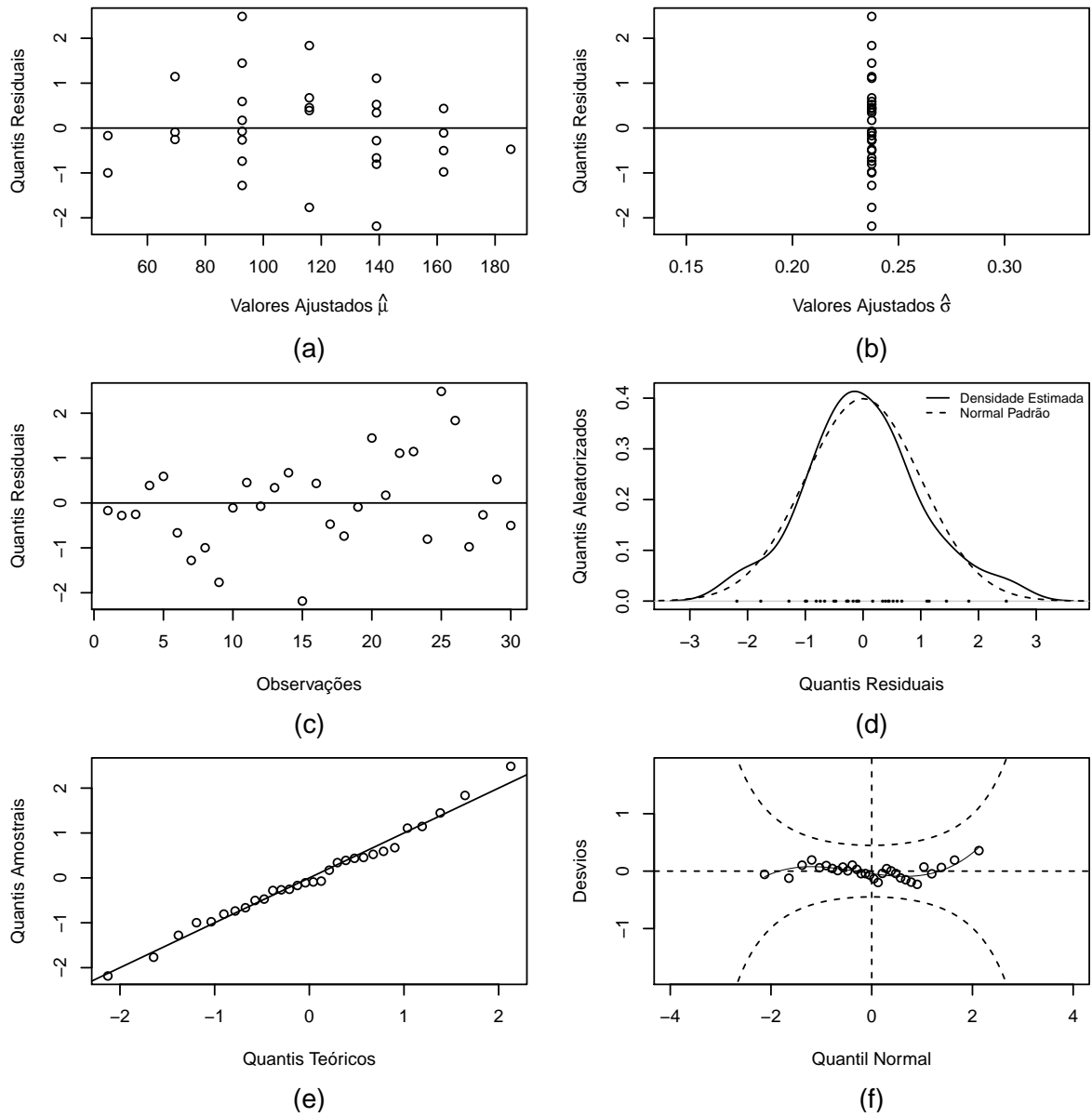


Figura 9 – Gráficos Diagnósticos para o Ajuste do Modelo Binomial-Negativo

As medidas resumo expostas na Tabela 12, apresentam média residual próxima de zero, variância próxima de um, coeficiente de curtose próximo de três. Talvez o coeficiente de assimetria ainda não esteja tão próximo de zero, como o desejado. Contudo, os testes de normalidade de Shapiro-Wilk e Lilliefors retornam um valor- p igual a 0,9604 e 0,6500, respectivamente, não rejeitando, portanto, a normalidade dos resíduos.

Tabela 12 – Medidas Resumo para os Quantis Residuais - Modelo Binomial-Negativo

	Medida	Estimativa
	Média	-0,00134
	Variância	1,02948
	Coefficiente de Assimetria	0,22222
	Coefficiente de Curtose	3,02631

3.2.5 Ajuste com base no Modelo Poisson-Generalizado

O modelo de Regressão Poisson-Generalizado também é proposto como uma alternativa à limitação da distribuição Poisson. O parâmetro de dispersão ϕ é o responsável por acomodar a sub ou superdispersão inerente ao fenômeno estudado. Sob a suposição de que o parâmetro de escala, ϕ , é comum à todas as doadoras, independente do tratamento, o modelo aqui proposto é definido como

$$Y_i \sim \text{PG}(\mu_i, \phi),$$

em que Y_i , com $i = 1, \dots, 30$, representa o total de oócitos viáveis, em uma quantidade o_i de coletas, associado à i -ésima doadora, cuja quantidade esperada é dada por μ_i .

As relações funcionais entre os parâmetros μ_i e ϕ e a estrutura de regressão foram estabelecidas pela função de ligação \log , para ambos os parâmetros. Visto que deseja-se avaliar se o tratamento recebido tem influência sobre o número esperado de oócitos viáveis, somente o parâmetro μ_i foi associado ao tratamento recebido pela i -ésima doadora. Deste modo, as relações estabelecidas têm a forma

$$\log\left(\frac{\mu_i}{o_i}\right) = \beta_{\mu 0} + \beta_{\mu 1} T_i,$$

$$\log(\phi) = \beta_{\phi 0}.$$

O número médio de oócitos aspirados, μ_i , e o parâmetro de escala, ϕ , podem ser estimados por $\hat{\mu}_i$ e $\hat{\phi}$, respectivamente, cujas estimativas são obtidas pelas relações inversas

$$\hat{\mu}_i = o_i e^{\hat{\beta}_{\mu 0} + \hat{\beta}_{\mu 1} T_i} \quad \text{e} \quad \hat{\phi} = e^{\hat{\beta}_{\phi 0}},$$

de onde segue que uma estimativa para a esperança e a variância da quantidade de oócitos viáveis é dada por

$$\widehat{\mathbb{E}}(Y_i) = \hat{\mu}_i \quad \text{e} \quad \widehat{\text{Var}}(Y_i) = \hat{\mu}_i \hat{\phi}^2.$$

Na Tabela 13 estão expostas as estimativas obtidas para os parâmetros associados a μ_i e ϕ . O teste- t associado ao parâmetro $\beta_{\mu 1}$ atesta que o tratamento 2 não exerce qualquer influência estatisticamente distinta do tratamento 1.

Tabela 13 – Estimativas do ajuste pelo modelo de Regressão Poisson-Generalizada.

	Estimativa	Erro Padrão	Valor- <i>t</i>	Valor- <i>p</i>
$\beta_{\mu 0}$	3,1992	0,1231	25,979	<2e-16
$\beta_{\mu 1}$	-0,1539	0,1632	-0,943	0,354
$\beta_{\phi 0}$	1.6930	0,1548	10,94	2,01e-11

Visto que o parâmetro $\beta_{\mu 1}$ foi considerado não significativo, um novo modelo, desconsiderando tal parâmetro, foi proposto. As estimativas resultantes estão na Tabela 14, e agora, uma vez que o efeito do tratamento foi desconsiderado, o modelo indica que existe um comportamento geral que ocorre independente do tratamento recebido pela doadora.

Tabela 14 – Estimativas do ajuste para o segundo modelo de Regressão Poisson-Generalizada.

	Estimativa	Erro Padrão	Valor- <i>t</i>	Valor- <i>p</i>
$\beta_{\mu 0}$	3,1210	0,0948	32,93	<2e-16
$\beta_{\phi 0}$	1,7029	0,1557	10,94	1,28e-11

Pode-se estimar a quantidade média de oócitos viáveis, via parâmetro μ_i , e o parâmetro de escala da distribuição $PG(\mu_i, \phi)$. Suas estimativas são dadas por

$$\hat{\mu}_i = o_i e^{\hat{\beta}_{\mu 0}} \approx 22,6695 o_i \quad \text{e} \quad \hat{\phi} = e^{\hat{\beta}_{\phi 0}} \approx 5,4898 ,$$

e, portanto, uma vez que o modelo seja validado, a quantidade média de oócitos aspirados em uma coleta ($o = 1$), independente da doadora ter recebido o tratamento 1 ou 2, é igual a $22,6695 \approx 23$ unidades.

Como pode ser visto na Figura 10, o modelo parece se ajustar bem às observações. Os gráficos Quantis Residuais vs. Valores Ajustados para μ e ϕ (a) e (b) e Quantis Residuais vs. Observações (c) apresentam comportamento aleatório, como o desejado. Os gráficos para a densidade estimada (d) e o Q-Q Plot Normal (e) indicam que os resíduos são normalmente distribuídos.

Com base no Worm-plot, Figura 9 (f), o comportamento dos resíduos é adequado, uma vez que os pontos se concentram em torno do eixo horizontal (Desvio igual a zero) e não extrapolam as bandas de confiança.

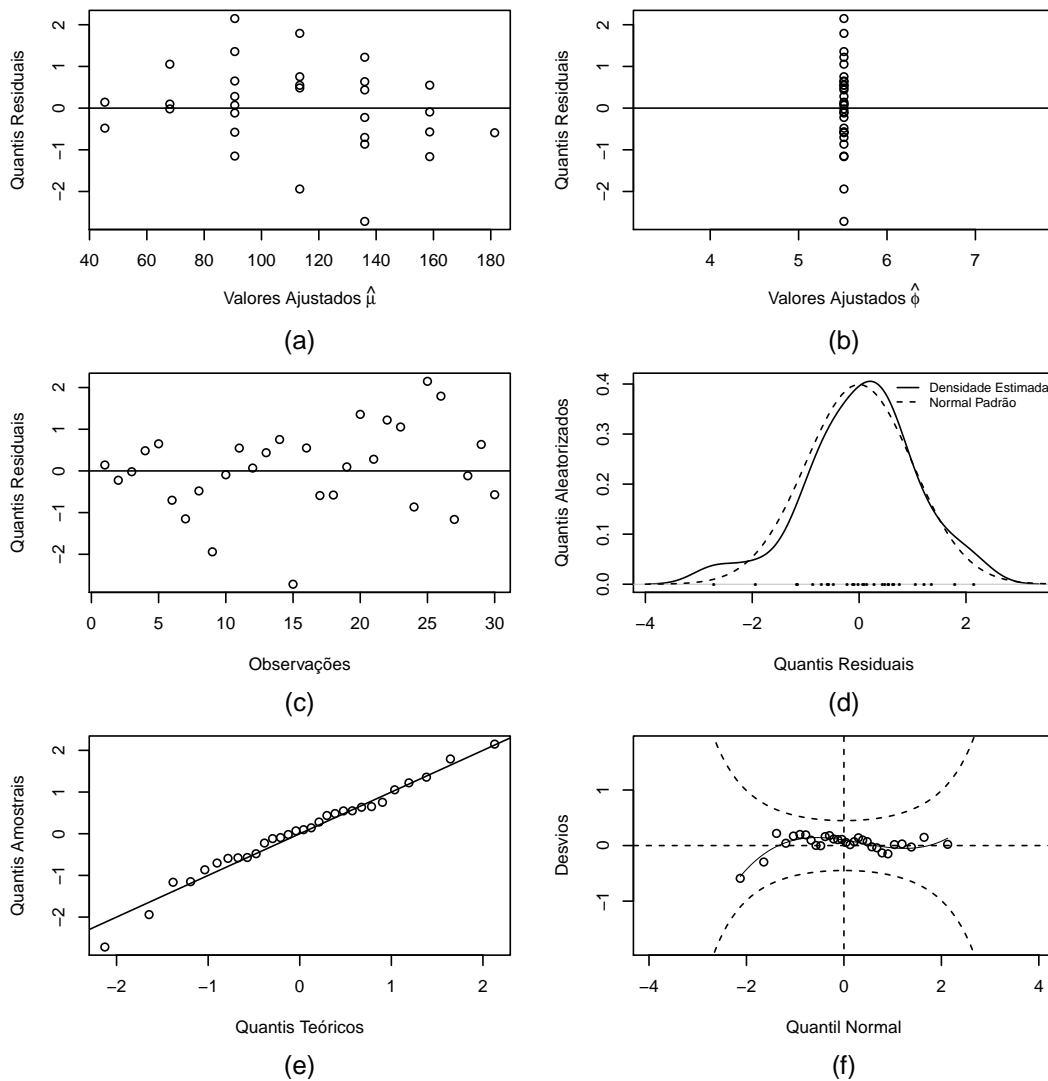


Figura 10 – Gráficos Diagnósticos para o Ajuste do Modelo Poisson-Generalizado

Similar ao modelo de Regressão Binomial-Negativo, as medidas resumo para os resíduos do modelo de Regressão Poisson-Generalizado, expostas na Tabela 15, apresentam média residual próxima de zero e variância próxima de um. Os coeficientes de assimetria e curtose talvez ainda não estejam tão próximos de zero e três, como o desejado. Contudo, os testes de normalidade de Shapiro-Wilk e Lilliefors retornam um valor- p igual a 0,8290 e 0,9269, não rejeitando, portanto, a normalidade dos resíduos.

Tabela 15 – Medidas Resumo para os Quantis Residuais - Modelo Poisson-Generalizado

Medida	Estimativa
Média	0,03355
Variância	1,07478
Coefficiente de Assimetria	-0,3777
Coefficiente de Curtose	3,3046

3.2.6 Ajuste com base no Modelo Beta-Binomial

O modelo de Regressão Beta-Binomial também contorna o problema da superdispersão. Diferente do modelo de Regressão Binomial, que supunha uma proporção de óocitos viáveis constante (exceto sobre o efeito do tratamento) independente da quantidade de coletas realizadas, da quantidade de óocitos aspirados em cada coleta, e até mesmo da doadora, o modelo de Regressão Beta-Binomial permite entender que a proporção de óocitos viáveis a cada coleta é não constante. Nesse caso, pode-se dizer que o número total de óocitos viáveis por doadora tem comportamento regido por uma distribuição Beta-Binomial com parâmetros μ e σ , cujo modelo de regressão é definido por

$$\begin{aligned} Y_i | P_i = p_i &\sim \text{BI}(n_i, p_i), \\ P_i &\sim \text{BE}(\mu_i, \sigma_i), \end{aligned} \quad \Leftrightarrow \quad \begin{aligned} Y_i &\sim \text{BB}(\mu_i, \sigma), \end{aligned}$$

em que Y_i , com $i = 1, \dots, 30$, representa o total de óocitos viáveis, em uma quantidade n_i de óocitos coletados, associado à i -ésima doadora. O valor esperado de Y_i é dado por μ_i .

As relações funcionais entre os parâmetros μ_i e σ com a covariável 'tratamento' foram estabelecidas pelas funções de ligação *logit* e *log*, respectivamente. Uma vez que deseja-se verificar se o tratamento recebido tem influência sobre o número esperado de óocitos viáveis, somente o parâmetro μ_i foi associado ao tratamento recebido pela i -ésima doadora. De modo que as relações estabelecidas têm a forma

$$\begin{aligned} \log\left(\frac{\mu_i}{1 - \mu_i}\right) &= \beta_{\mu 0} + \beta_{\mu 1} T_i, \\ \log(\sigma) &= \beta_{\sigma 0}, \end{aligned}$$

A proporção média de óocitos viáveis, μ_i , e o parâmetro de precisão, σ_i , podem ser estimados por $\hat{\mu}_i$ e $\hat{\sigma}_i$, respectivamente, e são obtidos pelas relações inversas

$$\hat{\mu}_i = \frac{e^{\hat{\beta}_{\mu 0} + \hat{\beta}_{\mu 1} T_i}}{1 + e^{\hat{\beta}_{\mu 0} + \hat{\beta}_{\mu 1} T_i}} \quad \text{e} \quad \hat{\sigma} = e^{\hat{\beta}_{\sigma 0}},$$

segue, portanto, que as estimativas para a esperança e a variância da quantidade de óocitos viáveis são dadas, respectivamente, por

$$\widehat{\mathbb{E}}(Y_i) = n_i \hat{\mu}_i \quad \text{e} \quad \widehat{\text{Var}}(Y_i) = n_i \hat{\mu}_i (1 - \hat{\mu}_i) \left[1 + \frac{\hat{\sigma}}{1 + \hat{\sigma}} (n_i - 1) \right].$$

Na Tabela 16 estão as estimativas para os parâmetros associados a μ_i e σ . O teste- t associado ao parâmetro $\beta_{\mu 1}$ atesta que o tratamento 2 não exerce qualquer influência estatisticamente distinta do tratamento 1.

Tabela 16 – Estimativas do Ajuste pelo Modelo Binomial-Negativo

	Estimativa	Erro Padrão	Valor- <i>t</i>	Valor- <i>p</i>
$\beta_{\mu 0}$	1,9085	0,1255	15,21	4,64e-15
$\beta_{\mu 1}$	-0,1341	0,1742	-0,77	0,448
$\beta_{\sigma 0}$	-3,940	0,421	-9,36	2,9e-10

Visto que o parâmetro $\beta_{\mu 1}$ foi considerado não significativo, um novo modelo, desconsiderando tal parâmetro, foi proposto. As estimativas resultantes estão na Tabela 17, e agora, uma vez que o efeito do tratamento foi desconsiderado, o modelo indica que existe um comportamento geral que ocorre independente do tratamento recebido pela doadora.

Tabela 17 – Estimativas do ajuste para o segundo modelo de Regressão Poisson-Generalizada.

	Estimativa	Erro Padrão	Valor- <i>t</i>	Valor- <i>p</i>
$\beta_{\mu 0}$	1,84003	0,08765	20,99	<2e-16
$\beta_{\sigma 0}$	-3,9182	0,4368	-8,971	7,31e-10

Pode-se estimar a proporção média de oócitos viáveis, via parâmetro μ_i , e o parâmetro de precisão da distribuição $BB(\mu_i, \sigma)$. Suas estimativas são dadas por

$$\hat{\mu}_i = \frac{e^{\hat{\beta}_{\mu 0}}}{1 + e^{\hat{\beta}_{\mu 0}}} \approx 0,86295 \quad \text{e} \quad \hat{\sigma} = e^{\hat{\beta}_{\sigma 0}} \approx 0,01988,$$

e, portanto, uma vez que o modelo seja validado, a proporção média de oócitos viáveis, independente da doadora ter recebido o tratamento 1 ou 2, é igual a 86,3.

Como pode ser visto na Figura 11, o modelo parece se ajustar bem às observações. Os gráficos Quantis Residuais vs. Valores Ajustados para μ e σ (a) e (b) e Quantis Residuais vs. Observações (c) apresentam comportamento aleatório, como o desejado. Os gráficos para a densidade estimada (d) e o Q-Q Plot Normal (e) estão próximos da normalidade, embora um pouco mais discrepantes que os modelos de Regressão Binomial-Negativo e Poisson-Generalizado.

Com base no Worm-plot, Figura 11 (f), o comportamento dos resíduos é adequado, visto que os pontos se concentram em torno do eixo horizontal (Desvio igual a zero) e não extrapolam as bandas de confiança.

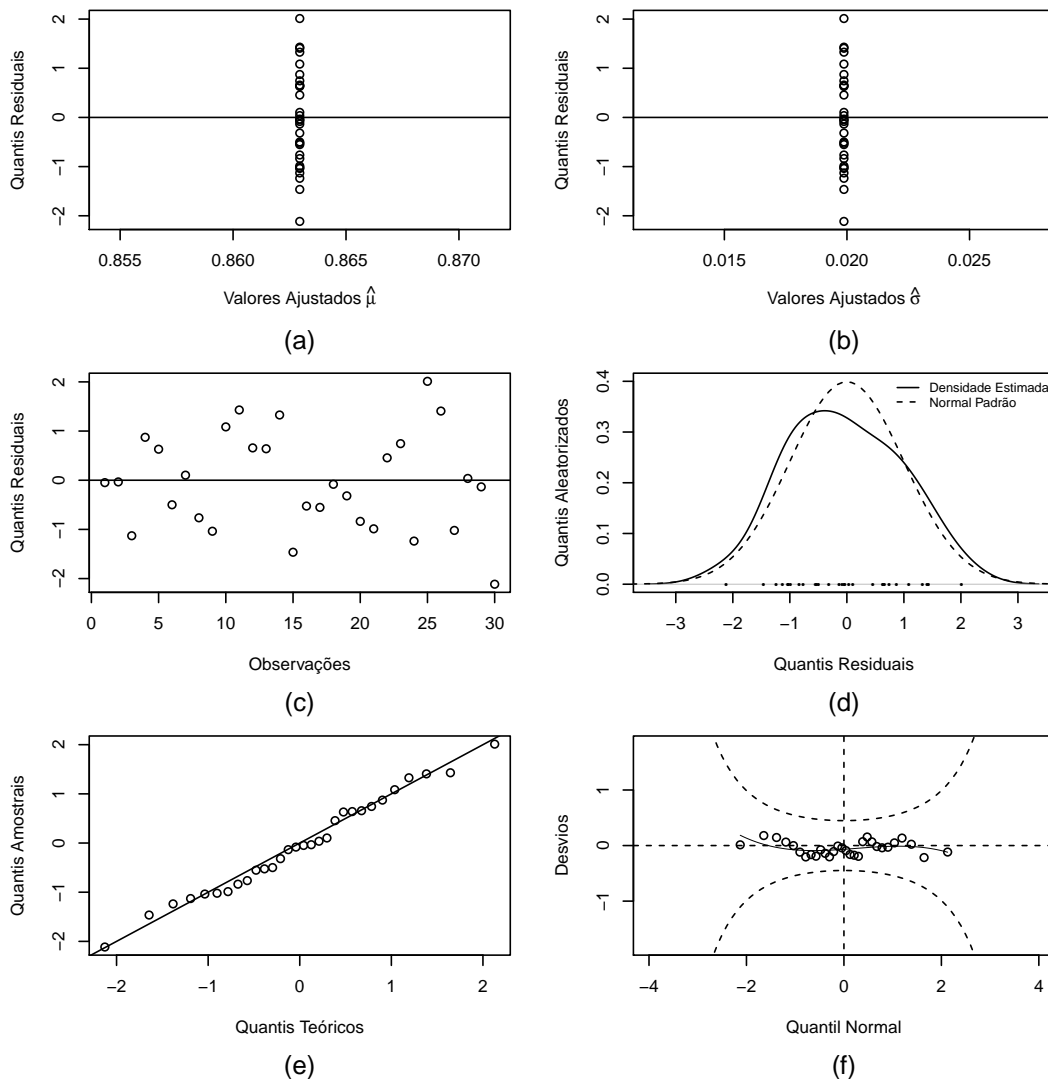


Figura 11 – Gráficos Diagnósticos para o Ajuste do Modelo Beta-Binomial

Similar aos modelo de Regressão Binomial-Negativo e Poisson-Generalizado, as medidas resumo expostas na Tabela 18 indicam que os resíduos tem o comportamento desejado. A média residual próxima de zero, variância próxima de um, coeficiente de assimetria próximo de zero. O coeficiente de curtose talvez ainda não esteja tão próximo de três, como o desejado. Contudo, os testes de normalidade de Shapiro-Wilk e Lilliefors retornam um valor- p igual a 0,9130 e 0,6517, respectivamente, não rejeitando, portanto, a normalidade dos resíduos.

Tabela 18 – Medidas Resumo para os Quantis Residuais - Modelo Poisson-Generalizado

Medida	Estimativa
Média	-0,04637
Variância	0,98200
Coefficiente de Assimetria	0,09608
Coefficiente de Curtose	2,19105

3.2.7 Resumo dos Resultados Frequentistas

Analisando-se a Tabela 19, que contém as estimativas dos ajustes dos seis modelos utilizados, nota-se que apenas o ajuste com o modelo Poisson apontou diferença entre os tratamentos, porém, este modelo não ficou bem ajustado, já que esta distribuição de probabilidade não consegue modelar a superdispersão existente nos dados. Todos os ajustes consideraram a influência do tratamento sobre as respostas. Desta forma, chega-se a conclusão, até aqui, que não houve diferença entre os tratamentos, já que os outros cinco ajustes não foram significativos. Desta forma, desconsiderando o efeito do tratamento, realizou-se novamente os ajustes.

Tabela 19 – Estimativas dos parâmetros dos modelos ajustados, considerando a covariável tratamento.

Modelo	Estimativa	Erro Padrão	Valor- <i>t</i>	Valor- <i>p</i>	
Binomial	β_0	1,9730	0,0695	28,3960	< 2e-16
	β_1	-0,1278	0,0962	-1,3290	0,1947
Poisson	β_0	3,1881	0,0243	131,3340	< 2e-16
	β_1	-0,1315	0,0345	-3,8070	0,0007
Binomial-Poisson	$\beta_{\lambda 0}$	-17,0242	0,0188	-905,4000	< 2e-16
	$\beta_{\lambda 1}$	-0,0160	5811,7981	0,0000	1,0000
	β_{p0}	20,2123	0,0154	1317,0000	< 2e-16
	β_{p1}	-0,1155	5811,7981	0,0000	1,0000
Binomial-Negativo	$\beta_{\mu 0}$	3,2244	0,1266	25,4650	< 2e-16
	$\beta_{\mu 1}$	-0,1692	0,1791	-0,9444	0,3534
	β_{σ}	-1,4661	0,2604	-5,6300	5,64e-06
Poisson-Generalizada	$\beta_{\mu 0}$	3,1992	0,1231	25,9790	< 2e-16
	$\beta_{\mu 1}$	-0,1539	0,1632	-0,9430	0,3545
	β_{ϕ}	1,6930	0,1548	10,9400	2,01e-11
Beta-Binomial	$\beta_{\mu 0}$	1,9085	0,1255	15,2100	4,64e-15
	$\beta_{\mu 1}$	-0,1341	0,1742	-0,7700	0,4482
	β_{σ}	-3,9400	0,4210	-9,3600	2,9e-10

Na Tabela 20, encontra-se as estimativas para os novos ajustes já desconsiderando o efeito do tratamento realizado. As estimativas obtidas para os demais parâmetros se mantiveram muito próximas das anteriores. E as conclusões se mantiveram, ou seja, não houve diferença entre os tratamentos para nenhum dos ajustes, exceto pelo modelo Poisson, que foi descartado.

Tabela 20 – Estimativas dos parâmetros dos modelos ajustados, desconsiderando a covariável tratamento.

Modelo	Estimativa	Erro Padrão	Valor- <i>t</i>	Valor- <i>p</i>
Binomial	β_0 1,9076	0,0480	39,7200	< 2e-16
Poisson	β_0 3,1210	0,0173	180,8000	< 2e-16
Binomial-Poisson	β_{λ_0} -17,0306	0,1726	986,500	< 2e-16
	β_{p_0} 20,1516	0,0020	10177,0	< 2e-16
Binomial-Negativo	β_{μ_0} 3,1431	0,0908	34,6200	< 2e-16
	β_{σ} -1,4381	0,2599	-5,5330	6,47e-06
Poisson-Generalizado	β_{μ_0} 3,1210	0,0948	32,9300	< 2e-16
	β_{ϕ} 1,7029	0,1557	10,9400	1,28e-11
Beta-Binomial	β_{μ_0} 1,8400	0,0877	20,9900	< 2e-16
	β_{σ} -3,9182	0,4368	-8,9710	7,31e-10

A Tabela 21 apresenta o ranking dos modelos que melhor ajustaram os dados analisados pelos critérios de AIC, AICc, BIC e GD. Porém, ao desconsiderarmos os modelos que não passaram nos diagnósticos, ou seja, Binomial, Poisson e Binomial Poisson, restam os modelos que realmente podemos considerar: Beta-Binomial, Binomial-Negativo e Poisson-Generalizado, apresentaram os melhores ajustes, respectivamente nesta ordem.

Tabela 21 – Estimativas dos parâmetros dos modelos ajustados, considerando a covariável Tratamento.

Modelo	g.l.	Critérios de Informação				Ranking Médio
		AIC	AICc	BIC	GD	
Beta-Binomial	3.00	200.58 ¹	201.50 ¹	204.78 ¹	194.58 ¹	1.00 ¹
Binomial	2.00	241.63 ²	242.07 ²	244.43 ²	237.63 ²	2.00 ²
Binomial-Negativo	3.00	324.25 ³	325.17 ³	328.45 ³	318.25 ³	3.00 ³
Poisson-Generalizado	3.00	325.78 ⁴	326.70 ⁴	329.98 ⁴	319.78 ⁴	4.00 ⁴
Poisson	2.00	986.25 ⁵	986.69 ⁵	989.05 ⁵	982.25 ⁵	5.00 ⁵
Binomial-Poisson	4.00	990.25 ⁶	991.85 ⁶	995.85 ⁶	982.25 ⁶	6.00 ⁶

Três modelos se mostraram adequados para abordar questões de interesses práticos, cada um com suas particularidades respondendo questões distintas acerca dos dados em estudo, em resumo (sintetizado na Tabela 22):

- Os modelos Binomial-Negativo e Poisson-Generalizado foram utilizados para ajustar o número esperado e a taxa média de oócitos viáveis. De acordo com tais modelos, independente do tratamento recebido pelas doadoras, espera-se aspirar em média uma quantidade aproximada de 23 oócitos por aspiração;
- O modelo Beta-Binomial foi utilizado para modelar a proporção média de oócitos viáveis, sob a suposição de que a quantidade de oócitos aspirados é uma constante

conhecida. De acordo com este modelo, independente do tratamento recebido pelas doadoras, espera-se que 86,3% dos oócitos aspirados sejam viáveis.

Tabela 22 – Resumo das respostas práticas que cada modelo válido é capaz de fornecer.

Modelo	Respostas Associadas (por aspiração)		
	Total de Oócitos Aspirados (N)	Total de Oócitos Viáveis (Y)	Proporção de Oócitos Viáveis (P)
Binomial-Negativo	–	23.1760 \approx 23	–
Poisson-Generalizado	–	22.6695 \approx 23	–
Beta-Binomial	conhecido	$N \times 0.8630$	$0.8630 \approx 86.30\%$

3.3 Modelos Ajustados com Metodologia Bayesiana

Os dados foram ajustados aos sete modelos propostos, na Tabela 23 estão expostas as estimativas para os parâmetros associados aos sete modelos propostos. Todos os ajustes consideraram a influência do tratamento sobre as respostas. As medidas de diagnóstico indicam convergência das cadeias para todos os parâmetros, de todos os modelos. Com exceção do modelo Poisson, todos os outros indicaram que o tratamento não exerce qualquer influência estatisticamente significativa. Sendo assim, os ajustes foram refeitos, desconsiderando a influência do tratamento.

Tabela 23 – Estimativas Bayesianas dos parâmetros, levando em consideração a influência da covariável Tratamento

Modelo	Estimativa	Erro Padrão	Quantis da Distribuição a Posteriori			Diagnósticos de Convergência*		
			2,5%	50,0%	97,5%	Valor-p HW	Valor-p Geweke	
Binomial	β_0	2.10446	0.15235	1.8076	2.1038	2.4048	0.735	0.6228
	β_1	-0.12902	0.09520	-0.3144	-0.1286	0.0593	0.710	0.6645
Poisson	β_0	3.32005	0.05466	3.2112	3.3205	3.4259	0.134	0.5476
	β_1	-0.13178	0.03454	-0.2005	-0.1315	-0.0630	0.109	0.6015
Binomial-Poisson	β_{p0}	2.10296	0.15360	1.8071	2.1019	2.4154	0.865	0.5538
	β_{p1}	-0.12830	0.09601	-0.3198	-0.1282	0.0588	0.802	0.5627
	$\beta_{\lambda 0}$	3.43241	0.05199	3.3292	3.4323	3.5367	0.509	0.4977
	$\beta_{\lambda 1}$	-0.11469	0.03295	-0.1810	-0.1148	-0.0500	0.221	0.2983
Binomial-Negativo	$\beta_{\mu 0}$	3.40395	0.28743	2.8483	3.4029	3.9774	0.885	0.7339
	$\beta_{\mu 1}$	-0.17015	0.18194	-0.5319	-0.1722	0.1907	0.775	0.6503
	β_{σ}	-1.44093	0.26486	-1.9317	-1.44850	-0.89790	0.750	0.5045
Poisson-Generalizado	$\beta_{\mu 0}$	3.36932	0.28048	2.8115	3.3693	3.9212	0.922	0.6082
	$\beta_{\mu 1}$	-0.15671	0.17524	-0.5007	-0.1553	0.1925	0.925	0.5368
	β_{ϕ}	1.75876	0.16983	1.4647	1.7450	2.1312	0.650	0.7900
Binomial-Poisson-Generalizado	β_{p0}	2.10296	0.15360	1.8071	2.1019	2.4154	0.865	0.5538
	β_{p1}	-0.12830	0.09601	-0.3198	-0.1282	0.0588	0.802	0.5627
	$\beta_{\mu 0}$	3.43241	0.05199	3.3292	3.4323	3.5367	0.509	0.4977

	$\beta_{\mu 1}$	-0.11469	0.03295	-0.1810	-0.1148	-0.0500	0.221	0.2983
	β_{ϕ}	1.73265	0.16347	1.4452	1.7207	2.0868	0.4713	0.4354
Beta-Binomial	$\beta_{\mu 0}$	2.04073	0.29852	1.4699	2.0372	2.6337	0.2376	0.1277
	$\beta_{\mu 1}$	-0.13319	0.18671	-0.4988	-0.1292	0.2314	0.2049	0.0892
	β_{σ}	-3.87980	0.36494	-4.5975	-3.8779	-3.1627	0.7350	0.9024

* Em todos os casos, os resultados obtidos com o critério de Gelman-Rubin indicam um limite superior (percentil 97,5%) para o Fator de Redução de Escala Potencial (e sua versão multivariada) igual, ou inferior, a 1,00.

Os resultados para os novos ajustes podem ser consultados na Tabela 24. As estimativas obtidas para os demais parâmetros se mantiveram muito próximas das anteriores, como esperado. As medidas de diagnóstico continuam atestando a estacionariedade das cadeias.

Tabela 24 – Estimativas Bayesianas dos parâmetros, desconsiderando a influência da co-variável Tratamento.

Modelo	Estimativa	Erro Padrão	Quantis da Distribuição a Posteriori			Diagnósticos de Convergência*		
			2,5%	50,0%	97,5%	Valor-p HW	Valor-p Geweke	
Binomial	β_0	1.90867	0.04789	1.8153	1.9083	2.0040	0.271	0.8320
Poisson	β_0	3.12062	0.01734	3.0863	3.1207	3.1546	0.719	0.3660
Binomial-Poisson	β_{p0}	1.90853	0.04811	1.8151	1.9084	2.0053	0.490	0.7690
	$\beta_{\lambda 0}$	3.25902	0.01622	3.2268	3.2589	3.2911	0.479	0.4340
Binomial-Negativo	$\beta_{\mu 0}$	3.14738	0.09187	2.9671	3.1465	3.3302	0.247	0.6895
	β_{σ}	-1.45028	0.25979	-1.9297	-1.4590	-0.9142	0.1980	0.4897
Poisson-Generalizado	$\beta_{\mu 0}$	3.13355	0.10236	2.9460	3.1298	3.3461	0.984	0.3177
	β_{ϕ}	1.74976	0.16564	1.4596	1.7380	2.1069	0.819	0.5991
Binomial-Poisson-Generalizado	β_{p0}	1.90883	0.04801	1.8167	1.9083	2.0044	0.965	0.3653
	$\beta_{\mu 0}$	3.27058	0.09379	3.0966	3.2663	3.4675	0.580	0.2467
	β_{ϕ}	1.72955	0.16178	1.4464	1.7177	2.0746	0.603	0.2758
Beta-Binomial	$\beta_{\mu 0}$	1.83648	0.09453	1.6434	1.8379	2.0169	0.597	0.7061
	β_{σ}	-3.89727	0.36562	-4.6320	-3.8950	-3.1840	0.6610	0.1871

* Em todos os casos, os resultados obtidos com o critério de Gelman-Rubin indicam um Fator de Redução de Escala Potencial (e sua versão multivariada) igual, ou inferior, a 1,01.

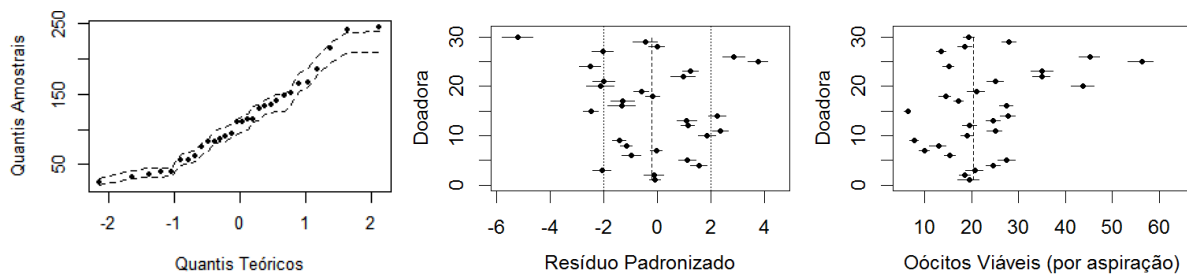
Na Tabela 25 estão expostas as medidas obtidas para os critérios de seleção entre os modelos, a fim de explicitar algum modelo que se destaque dos demais, no sentido de possuir maior parcimônia e ser capaz de resumir as informações de modo tão satisfatório quanto os outros. Note que, o modelo que apresenta o menor valor para os critérios DIC, EAIC e EBIC é o modelo Beta-Binomial, seguido pelo modelo Binomial e Binomial-Negativo.

Tabela 25 – Seleção de modelos com base nos critérios de informação Bayesianos DIC, EAIC e EBIC.

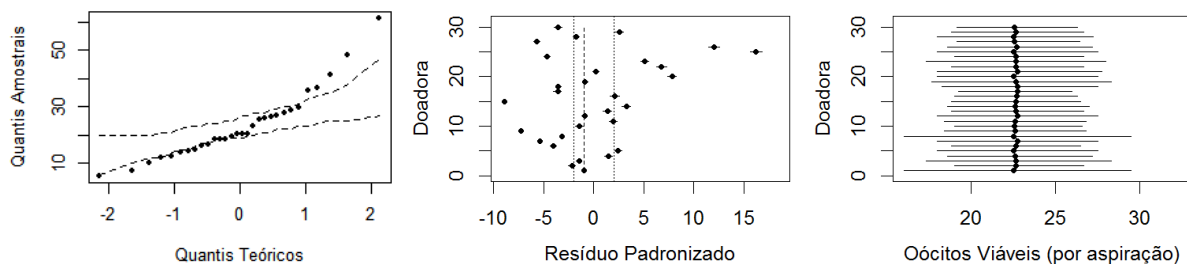
Modelo	Critérios de Informação			Ranking Médio entre os Critérios
	DIC	EAIC	EBIC	
Beta-Binomial	181.3 ¹	164.3 ¹	167.1 ¹	1.00 ¹
Binomial	241.4 ²	242.4 ²	243.8 ²	2.00 ²
Binomial-Negativo	323.1 ⁴	325.1 ³	327.9 ³	3.33 ³
Poisson-Generalizado	323.0 ³	326.7 ²⁴	329.5 ⁴	3.67 ⁴
Binomial-Poisson-Generalizado	568.5 ⁵	573.2 ⁵	577.4 ⁵	5.00 ⁵
Poisson	998.7 ⁶	999.7 ⁶	1001.1 ⁶	6.00 ⁶
Binomial-Poisson	1226.0 ⁷	1228.0 ⁷	1230.8 ⁷	7.00 ⁷

Um modo de verificar a validade de um ajuste pode ser observado na Figura 3, em que estão expostos os gráficos para: As bandas de reamostragem para cada modelo, que fornece indícios sobre a origem da amostra observada na pesquisa. É importante ressaltar que tal critério não é suficiente para justificar a exclusão de um modelo, especialmente devido ao fato de que sua construção não leva em conta a incerteza inerente aos parâmetros, isto é, somente a estimativa pontual da média a posteriori foi utilizada; Os gráficos *caterpillar* para os resíduos padronizados, por doadora e modelo, e; Os gráficos *caterpillar* para as distribuições preditivas *a posteriori*, para o número de oócitos viáveis, também de acordo com a doadora e o modelo. Destes gráficos, nota-se que:

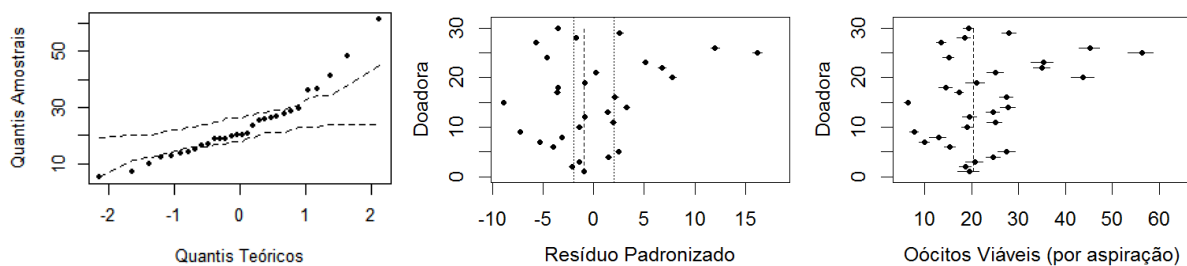
- Considerando as bandas de reamostragem, os modelos Poisson e Binomial-Poisson apresentaram as maiores inconcordâncias entre o ajuste e os dados observados, além disso, os modelos apresentaram uma quantidade considerável de discrepâncias entre as distribuições residuais obtidas pelo ajuste e o que se espera delas (concentração de observações entre -2 e 2).
- Os modelos Beta-Binomial e Binomial se destacam com as maiores concordâncias entre o ajuste e o que foi observado, considerando as bandas de reamostragem. De acordo com distribuições residuais, o modelo Beta-Binomial apresentou maior adequação, com uma única média residual fora do intervalo (-2,2), enquanto o modelo Binomial apresentou aproximadamente nove, praticamente um terço do banco de dados, sugerindo uma possível inadequação do modelo.
- Os demais modelos, Binomial-Negativo, Poisson-Generalizado e Binomial-Poisson-Generalizado não apresentaram resultados bons, considerando as bandas de reamostragem, mas apresentaram distribuições residuais adequadas, somente duas médias residuais fora do intervalo (-2,2), em cada modelo.



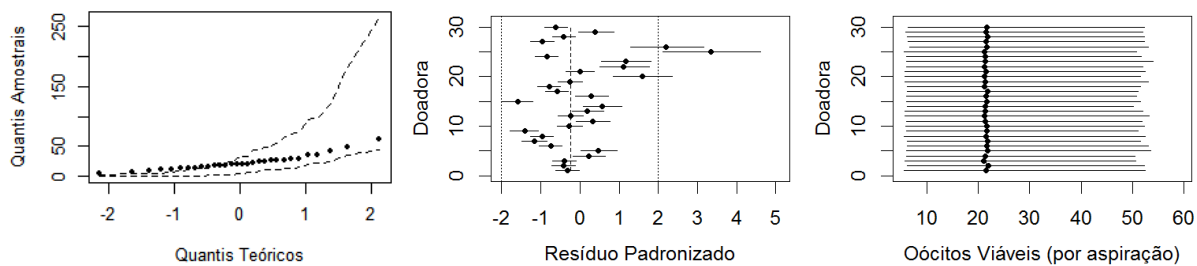
(a) Modelo Binomial.



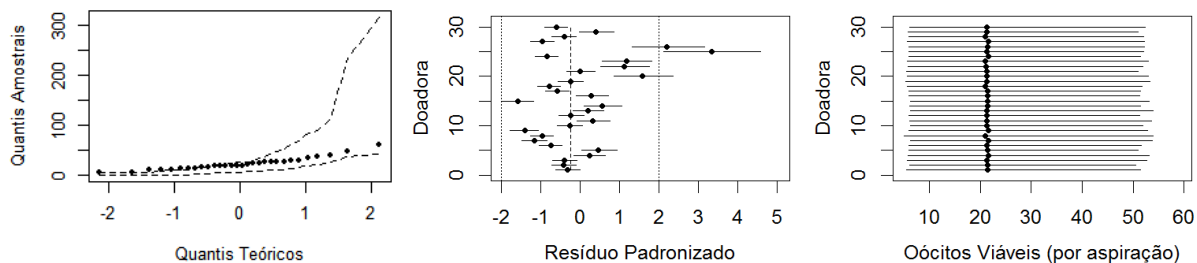
(b) Modelo Poisson.



(c) Modelo Binomial-Poisson.



(d) Modelo Binomial-Negativo.



(e) Modelo Poisson-Generalizado.

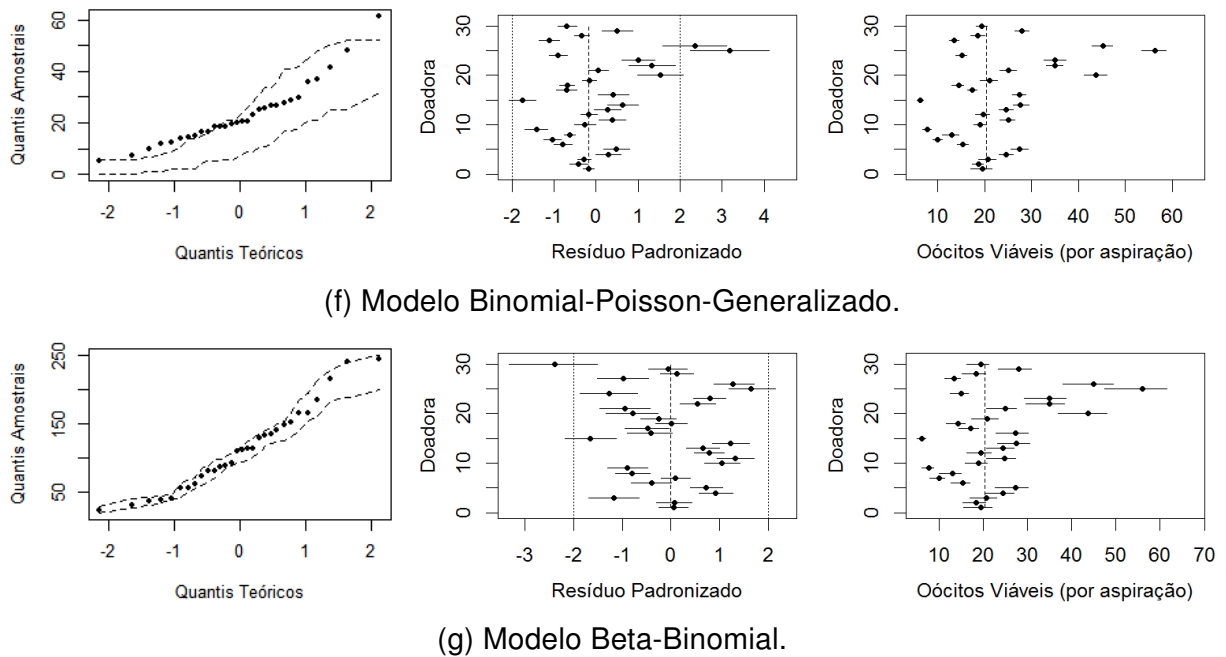


Figura 11 – [ESQUERDA] Bandas de reamostragem simulada, determinadas com base nas médias a posteriori para os parâmetros associados aos modelos ajustados. Foram tomados 200 grupos de 30 observações amostradas dos modelos para a criação das bandas. Os pontos em preto correspondem à amostra original; [CENTRO] Quantis (2.5% e 97.5%) da distribuição a posteriori dos resíduos padronizados por doadora. Os pontos em preto e a linha tracejada representam as médias residuais por doadora e geral, respectivamente; [DIREITA] Quantis (2.5% e 97.5%) das distribuições preditivas a posteriori, por doadora, para a quantidade de óocitos viáveis por aspiração. Os pontos em preto e a linha tracejada representam a mediana das predições por doadora e geral, respectivamente.

Nesse sentido, somente os modelos Poisson e Binomial-Poisson se mostraram inadequados tanto para as bandas de reamostragem, quanto para distribuições residuais. Cinco dos modelos (Binomial, Binomial-Negativo, Poisson-Generalizado, Binomial-Poisson-Generalizado e Beta-Binomial), se mostraram razoáveis para abordar questões de interesse prático, respondendo questões distintas acerca dos dados em estudo:

- Os modelos Binomial-Negativo e Poisson-Generalizado ajustaram o número esperado e a taxa média de óocitos viáveis. Nesse contexto, independente do tratamento, espera-se aspirar, em média, uma quantidade aproximada de 23 (mediana $21.33 \approx 21.43 \approx 21$, para cada modelo) óocitos viáveis por aspiração;
- Os modelos Binomial e Beta-Binomial modelaram a proporção esperada de óocitos viáveis, sob a suposição de que a quantidade de óocitos aspirados é conhecida. Independente do tratamento recebido pelas doadoras, espera-se, de acordo com cada modelo, que aproximadamente 87.07% e 86.21%, respectivamente, dos óocitos aspi-

rados sejam viáveis e uma quantidade mediana, por aspiração, de aproximadamente $20.33 \approx 20$ oócitos viáveis, para ambos os modelos;

- O modelo Binomial-Poisson-Generalizado ajustou tanto a proporção esperada de oócitos viáveis, quanto o número esperado de oócitos aspirados (e sua respectiva taxa de aspiração) e, conseqüentemente, o número de oócitos viáveis. De acordo com tal modelo, espera-se aspirar, em média, 26 oócitos por coleta, independente do tratamento. Espera-se também que dos 26 oócitos aspirados 87.08%, isto é, 23 (mediana $20.29 \approx 20$), sejam viáveis.

Um resumo sobre as medidas de interesse associadas a cada modelo está exposto na Tabela 26. Note que os modelos Binomial-Negativo e Poisson-Generalizado respondem somente sobre o total de oócitos viáveis, sendo incapazes de responder qualquer coisa sobre a quantidade esperada de oócitos aspirados ou a proporção esperada de oócitos viáveis. Analogamente, o modelo Beta-Binomial responde exclusivamente sobre a proporção esperada de oócitos viáveis. O modelo Binomial-Poisson-Generalizado, por outro lado, é capaz de fazer inferência sobre qualquer uma destas quantidades, mostrando ser o mais completo, embora esteja entre as últimas colocações, de acordo com os critérios de informação.

Tabela 26 – Resumo das respostas práticas que cada modelo válido é capaz de fornecer.

Modelo	Respostas Associadas (por aspiração)		
	Total de Oócitos Aspirados (N)	Total de Oócitos Viáveis (Y)	Proporção de Oócitos Viáveis (P)
Binomial	conhecido	$N \times 0.8708$	$0.8708 \approx 87.08\%$
Binomial-Negativo	–	$23.37159 \approx 23$	–
Poisson-Generalizado	–	$23.12403 \approx 23$	–
Binomial-Poisson-Generalizado	$26.45277 \approx 26$	$26 \times 0.87 \approx 23$	$0.8707 \approx 87.07\%$
Beta-Binomial	conhecido	$N \times 0.8621$	$0.8621 \approx 86.21\%$

3.4 Modelos Ajustados com Metodologia Bayesiana considerando uma Estrutura Longitudinal

Considerando os modelos já descritos anteriormente, suas respectivas versões longitudinais estão sumarizadas na Tabela 27.

Tabela 27 – Estruturas de regressão adotadas para cada modelo e suas respectivas esperanças e variâncias.

Modelos	Estruturas de Regressão	$\mathbb{E}(Y b)$	$\text{Var}(Y b)$
<i>Binomial</i> $Y b \sim \text{BI}(n, p)$ $b \sim \text{N}(0, \theta^2)$	$\text{logit}(p) = \beta_0 + \beta_1 D + \beta_2 T + b$	np	$np(1 - p)$
<i>Poisson</i> $Y b \sim \text{PO}(\lambda)$ $b \sim \text{N}(0, \theta^2)$	$\log(\lambda) = \beta_0 + \beta_1 D + \beta_2 T + b$	λ	λ
<i>Binomial-Poisson</i> $Y b_\lambda, b_p \sim \text{BP}(\lambda, p)$ $b_\lambda \sim \text{N}(0, \theta_\lambda^2)$ $b_p \sim \text{N}(0, \theta_p^2)$	$\text{logit}(p) = \beta_{p0} + \beta_{p1} D + \beta_{p2} T + b_p$ $\log(\lambda) = \beta_{\lambda0} + \beta_{\lambda1} D + \beta_{\lambda2} T + b_\lambda$	λp	λp
<i>Binomial-Negativo</i> $Y b \sim \text{BN}(\mu, \sigma)$ $b \sim \text{N}(0, \theta^2)$	$\log(\mu) = \beta_{\mu0} + \beta_{\mu1} D + \beta_{\mu2} T + b$ $\log(\sigma) = \beta_\sigma$	μ	$\mu + \mu^2 \sigma$
<i>Poisson-Generalizado</i> $Y b \sim \text{PG}(\mu, \phi)$ $b \sim \text{N}(0, \theta^2)$	$\log(\mu) = \beta_{\mu0} + \beta_{\mu1} D + \beta_{\mu2} T + b$ $\log(\phi) = \beta_\phi$	μ	$\mu \phi^2$
<i>Binomial-Poisson-Generalizado</i> $Y b_\mu, b_p \sim \text{BPG}(\mu, p, \phi)$ $b_\mu \sim \text{N}(0, \theta_\mu^2)$ $b_p \sim \text{N}(0, \theta_p^2)$	$\text{logit}(p) = \beta_{p0} + \beta_{p1} D + \beta_{p2} T + b_p$ $\log(\mu) = \beta_{\mu0} + \beta_{\mu1} D + \beta_{\mu2} T + b_\mu$ $\log(\phi) = \beta_\phi$	μp	$p(1 - p)\mu + p^2 \mu \phi^2$
<i>Beta-Binomial</i> $Y b \sim \text{BB}(\mu, \sigma)$ $b \sim \text{N}(0, \theta^2)$	$\text{logit}(\mu) = \beta_{\mu0} + \beta_{\mu1} D + \beta_{\mu2} T + b$ $\log(\sigma) = \beta_\sigma$	$n\mu$	$n\mu(1-\mu) \left[1 + \frac{\sigma(n-1)}{\sigma+1}\right]$

* A covariável T , que denota o tratamento, assume os valores $T = 0$ (Tratamento Controle) e $T = 1$ (Novo Tratamento).

** A covariável $D > 0$ denota a quantidade de dias transcorridos desde o início da suplementação até o dia de cada coleta.

*** A distribuição Normal $\text{N}(0, 100^2)$ foi utilizada como distribuição *a priori* para todos os parâmetros β 's estudados. A distribuição Gama $\text{GA}(0.001, 0.001)$ foi utilizada como distribuição *a priori* para todos os parâmetros θ 's estudados.

Os dados foram ajustados aos sete modelos propostos, na Tabela 28 estão expostas as estimativas obtidas. Todos os ajustes consideraram a influência do tratamento sobre as respostas. As medidas de diagnóstico indicam convergência das cadeias para todos os parâmetros, de todos os modelos. Neste caso, todos os modelos indicaram que o tratamento não exerce qualquer influência estatisticamente significativa. Sendo assim, os ajustes foram refeitos, desconsiderando a influência do tratamento.

Tabela 28 – Estimativas Bayesianas dos parâmetros, levando em consideração a influência da covariável Tratamento.

Modelo	Estimativa	Erro Padrão	Quantis da Distribuição a Posteriori			Diagnósticos de Convergência*		
			2,5%	50,0%	97,5%	Valor-p HW	Valor-p Geweke	
Binomial	β_0	2.41644	0.31593	1.8004	2.4151	3.0459	0.1413	0.4917
	β_1	-0.00546	0.00148	-0.0084	-0.0055	-0.0026	0.5762	0.8230
	β_2	-0.11953	0.19351	-0.5035	-0.1180	0.2610	0.1126	0.5547
	θ^2	0.19096	0.07662	0.0810	0.1778	0.3785	0.4573	0.9006
Poisson	β_0	3.48298	0.30210	2.9162	3.4821	4.1045	0.4648	0.3358
	β_1	-0.00396	0.00054	-0.0050	-0.0040	-0.0029	0.1056	0.1741

	β_2	-0.15320	0.18886	-0.5429	-0.1536	0.1920	0.5523	0.3202
	θ^2	0.29319	0.08891	0.1644	0.2784	0.5127	0.3965	0.3341
Binomial-Poisson	β_{p0}	2.40647	0.30619	1.8028	2.4088	3.0109	0.5782	0.4320
	β_{p1}	-0.00545	0.00150	-0.0084	-0.0054	-0.0025	0.2276	0.1251
	β_{p2}	-0.11388	0.18746	-0.4831	-0.1152	0.2608	0.5058	0.4411
	$\beta_{\lambda 0}$	3.58170	0.27123	3.0546	3.5777	4.0994	0.1991	0.9409
	$\beta_{\lambda 1}$	-0.00335	0.00049	-0.0043	-0.0034	-0.0024	0.0765	0.0594
	$\beta_{\lambda 2}$	-0.14018	0.16912	-0.4650	-0.1376	0.2028	0.2195	0.9693
	θ_p^2	0.19102	0.07673	0.0807	0.1783	0.3740	0.3274	0.9713
	θ_λ^2	0.25052	0.07440	0.1428	0.2377	0.4278	0.2703	0.7627
Binomial-Negativo	$\beta_{\mu 0}$	3.50395	0.31526	2.9008	3.4993	4.1166	0.6230	0.8805
	$\beta_{\mu 1}$	-0.00410	0.00082	-0.0057	-0.0041	-0.0025	0.5440	0.3780
	$\beta_{\mu 2}$	-0.15565	0.19961	-0.5442	-0.1530	0.2290	0.4306	0.6288
	β_σ	-2.89065	0.24867	-3.4052	-2.8812	-2.4313	0.1713	0.1823
	θ^2	0.28264	0.08863	0.1549	0.2682	0.4960	0.1889	0.6174
	Poisson-Generalizado	$\beta_{\mu 0}$	3.47858	0.31728	2.8382	3.4787	4.1033	0.3384
$\beta_{\mu 1}$		-0.00376	0.00080	-0.0053	-0.0038	-0.0022	0.2714	0.3761
$\beta_{\mu 2}$		-0.15087	0.19777	-0.5444	-0.1493	0.2441	0.2825	0.4378
β_ϕ		0.43516	0.06891	0.3041	0.4338	0.5759	0.9030	0.8728
θ^2		0.26345	0.08418	0.1399	0.2497	0.4664	0.6370	0.3473
Binomial-Poisson-Generalizado		β_{p0}	2.40494	0.31489	1.7803	2.4045	3.0189	0.7776
	β_{p1}	-0.00546	0.00149	-0.0084	-0.0055	-0.0025	0.4023	0.2487
	β_{p2}	-0.11280	0.19359	-0.4916	-0.1149	0.2750	0.6887	0.3582
	$\beta_{\mu 0}$	3.57330	0.29510	2.9933	3.5720	4.1636	0.2317	0.0662
	$\beta_{\mu 1}$	-0.00321	0.00076	-0.0047	-0.0032	-0.0017	0.1721	0.1400
	$\beta_{\mu 2}$	-0.13196	0.18458	-0.4998	-0.1306	0.2309	0.2540	0.0855
	β_ϕ	0.44177	0.06812	0.3135	0.4400	0.5813	0.6346	0.8007
	θ_p^2	0.19071	0.07685	0.0821	0.1773	0.3754	0.1493	0.7771
	θ_μ^2	0.22326	0.07188	0.1194	0.2111	0.3998	0.0643	0.1731
	Beta-Binomial	$\beta_{\mu 0}$	2.35670	0.40495	1.2111	2.4002	2.9906	0.3400
$\beta_{\mu 1}$		-0.00548	0.00153	-0.0085	-0.0055	-0.0026	0.0701	0.4091
$\beta_{\mu 2}$		-0.08687	0.25849	-0.4529	-0.1113	0.6845	0.1149	0.2974
β_σ		-6.23829	0.83648	-7.9730	-6.2491	-4.6881	0.7764	0.1154
θ^2		0.1853	0.08346	0.0671	0.1660	0.3851	0.4567	0.2773

* Em todos os casos, os resultados obtidos com o critério de Gelman-Rubin indicam um limite superior (percentil 97,5%) para o Fator de Redução de Escala Potencial (e sua versão multivariada) igual, ou inferior, a 1,07.

Os resultados para os novos ajustes podem ser consultados na Tabela 29. As estimativas obtidas para os demais parâmetros se mantiveram muito próximas das anteriores, como esperado. As medidas de diagnóstico continuam atestando a estacionariedade das cadeias.

Tabela 29 – Estimativas Bayesianas dos parâmetros, desconsiderando a influência da co-variável Tratamento.

Modelo	Estimativa	Erro Padrão	Quantis da Distribuição a Posteriori			Diagnósticos de Convergência*	
			2,5%	50,0%	97,5%	Valor-p HW	Valor-p Geweke

Binomial	β_0	2.23779	0.13329	1.9788	2.2360	2.4992	0.1553	0.2234
	β_1	-0.00550	0.00148	-0.0084	-0.0055	-0.0026	0.1712	0.3576
	θ^2	0.18475	0.07472	0.0781	0.1720	0.3683	0.2434	0.7027
Poisson	β_0	3.25510	0.10341	3.0458	3.2558	3.4571	0.4986	0.5585
	β_1	-0.00396	0.00053	-0.0050	-0.0040	-0.0029	0.4829	0.4440
	θ^2	0.28735	0.08630	0.1633	0.2733	0.4940	0.8254	0.6860
Binomial-Poisson	β_{p0}	2.23644	0.13176	1.9814	2.2371	2.4961	0.6080	0.7244
	β_{p1}	-0.00550	0.00148	-0.0084	-0.0055	-0.0026	0.4498	0.7574
	$\beta_{\lambda 0}$	3.37278	0.09592	3.1831	3.3731	3.5630	0.0629	0.1283
	$\beta_{\lambda 1}$	-0.00335	0.00049	-0.0043	-0.0033	-0.0024	0.3150	0.8402
	θ_p^2	0.18391	0.07407	0.0786	0.1711	0.3644	0.6567	0.8710
	θ_λ^2	0.24609	0.07331	0.1393	0.2342	0.4225	0.7387	0.3846
Binomial-Negativo	$\beta_{\mu 0}$	3.27289	0.11292	3.0522	3.2731	3.4950	0.9363	0.9592
	$\beta_{\mu 1}$	-0.00414	0.00082	-0.0058	-0.0041	-0.0025	0.2434	0.8579
	β_σ	-2.89314	0.24848	-3.3973	-2.8849	-2.4242	0.1684	0.1556
	θ^2	0.27763	0.08663	0.1525	0.2624	0.4875	0.1021	0.2108
Poisson-Generalizado	$\beta_{\mu 0}$	3.25801	0.10763	3.0464	3.2581	3.4719	0.1073	0.8757
	$\beta_{\mu 1}$	-0.00380	0.00080	-0.0054	-0.0038	-0.0022	0.4166	0.5838
	β_ϕ	0.43641	0.06922	0.3063	0.4344	0.5773	0.2485	0.4271
	θ^2	0.25658	0.08074	0.1386	0.2431	0.4515	0.5287	0.8627
Binomial-Poisson-Generalizada	β_{p0}	2.23884	0.13342	1.9785	2.2395	2.5022	0.4918	0.4271
	β_{p1}	-0.00549	0.00150	-0.0084	-0.0055	-0.0026	0.8827	0.8627
	$\beta_{\mu 0}$	3.37698	0.10045	3.1796	3.3784	3.5716	0.5406	0.9105
	$\beta_{\mu 1}$	-0.00323	0.00075	-0.0047	-0.0032	-0.0018	0.3309	0.6992
	β_ϕ	0.44245	0.06884	0.3122	0.4412	0.5814	0.1719	0.3759
	θ_μ^2	0.18367	0.07349	0.0782	0.1711	0.3633	0.2418	0.4664
Beta-Binomial	$\beta_{\mu 0}$	2.22037	0.13244	1.9785	2.2209	2.4691	0.1165	0.3412
	$\beta_{\mu 1}$	-0.00530	0.00158	-0.0082	-0.0055	-0.0021	0.1848	0.1375
	β_σ	-6.25961	0.80104	-7.7090	-6.2585	-4.7696	0.1032	0.1950
	θ^2	0.17656	0.07274	0.0789	0.1613	0.3432	0.8246	0.2659

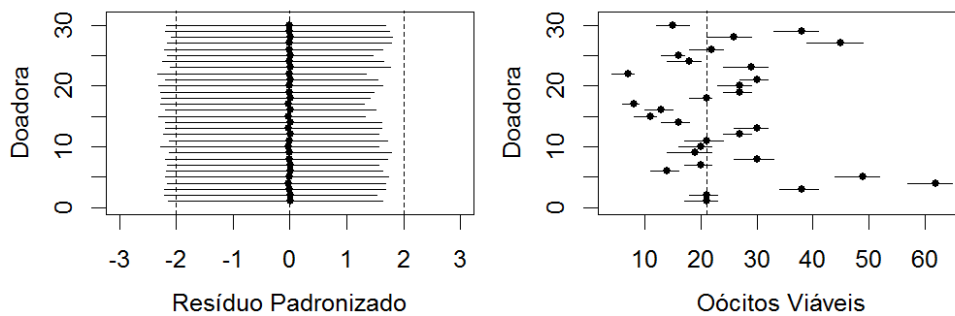
* Em todos os casos, os resultados obtidos com o critério de Gelman-Rubin indicam um limite superior (percentil 97,5%) para o Fator de Redução de Escala Potencial (e sua versão multivariada) igual, ou inferior, a 1,06.

Na Tabela 30 estão expostas as medidas obtidas para os critérios de seleção entre os modelos, a fim de explicitar algum modelo que se destaque dos demais, no sentido de possuir maior parcimônia e ser capaz de resumir as informações de modo tão satisfatório quanto os outros. Note que, o modelo que apresenta o menor valor para os critérios DIC, EAIC e EBIC é o modelo Beta-Binomial, seguido pelo modelo Binomial e Binomial-Negativo.

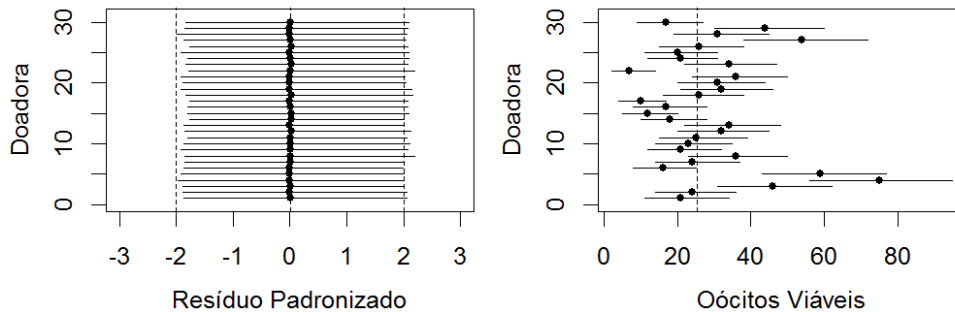
Tabela 30 – Seleção de modelos com base nos critérios de informação Bayesianos DIC, EAIC e EBIC.

Modelo	Critérios de Informação			Ranking Médio entre os Critérios
	DIC	EAIC	EBIC	
<i>Binomial</i>	564.7 ¹	570.7 ¹	579.7 ¹	1.00 ¹
<i>Beta-Binomial</i>	568.4 ²	574.4 ²	583.4 ²	2.00 ²
<i>Poisson-Generalizado</i>	989.4 ³	995.4 ³	1004.4 ³	3.00 ³
<i>Binomial-Negativo</i>	1009.0 ⁴	1015.0 ⁴	1024.0 ⁴	4.00 ⁴
<i>Poisson</i>	1040.0 ⁵	1046.0 ⁵	1055.0 ⁵	5.00 ⁵
<i>Binomial-Poisson-Generalizado</i>	1579.0 ⁶	1593.0 ⁶	1614.0 ⁶	6.00 ⁶
<i>Binomial-Poisson</i>	1634.0 ⁷	1646.0 ⁷	1664.0 ⁷	7.00 ⁷

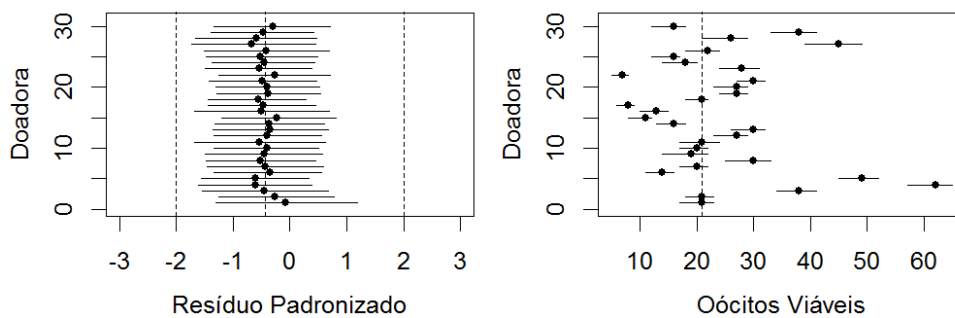
Um modo de verificar a validade de um ajuste pode ser observado na Figura 10, em que estão expostos os gráficos para: Os gráficos *caterpillar* para os resíduos padronizados, por doadora e modelo e; Os gráficos *caterpillar* para as distribuições preditivas a posteriori, para o número de oócitos viáveis, também de acordo com a doadora e o modelo. O comportamento dos resíduos sugere boa adequação de todos os modelos, além disso, as distribuições preditivas parecem concordar entre si.



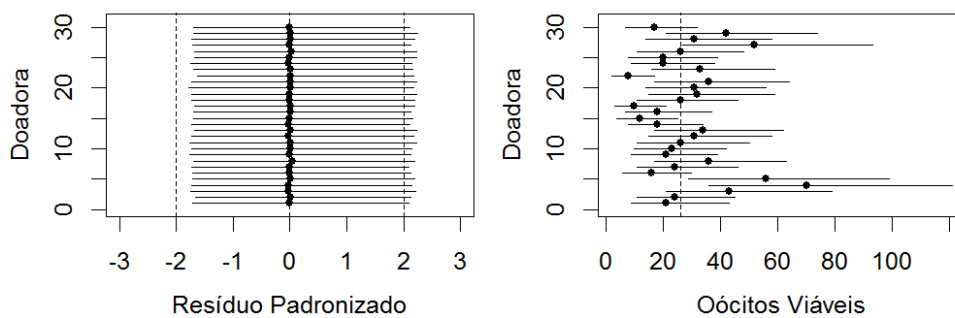
(a) Modelo Binomial.



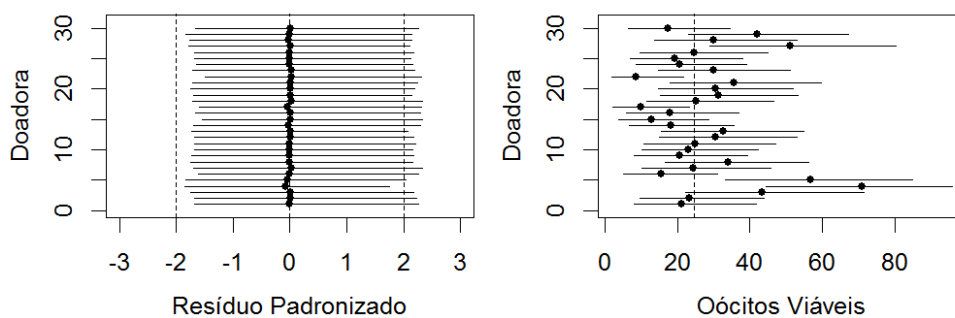
(b) Modelo Poisson.



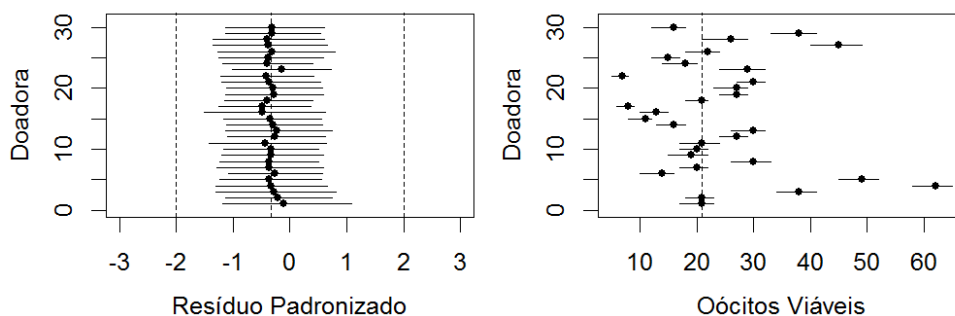
(c) Modelo Binomial-Poisson.



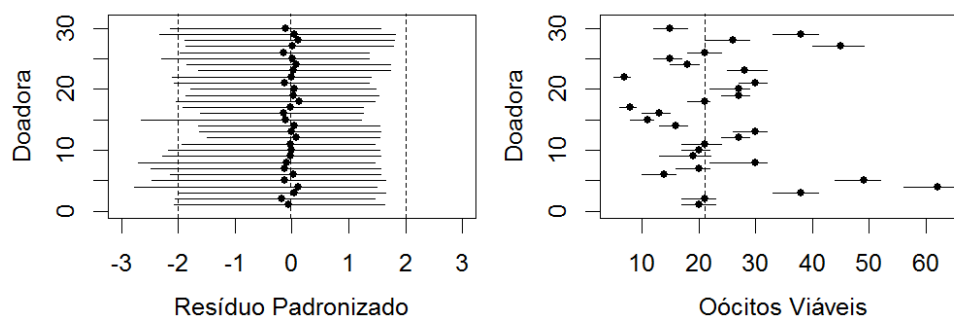
(d) Modelo Binomial-Negativo.



(e) Modelo Poisson-Generalizado.



(f) Modelo Binomial-Poisson-Generalizado.



(g) Modelo Binomial.

Figura 10 – [ESQUERDA] Quantis (2.5% e 97.5%) da distribuição a posteriori dos resíduos padronizados por doadora. Os pontos em preto e a linha tracejada representam as médias residuais por doadora e geral, respectivamente; [DIREITA] Quantis (2.5% e 97.5%) das distribuições preditivas a posteriori, por doadora, para a quantidade de óocitos viáveis por aspiração. Os pontos em preto e a linha tracejada representam a mediana das predições por doadora e geral, respectivamente.

Nesse sentido, todos os modelos se mostraram razoáveis para abordar questões de interesse prático, respondendo questões distintas acerca dos dados em estudo:

- Os modelos Poisson, Binomial-Negativo e Poisson-Generalizado ajustaram o número esperado de óocitos viáveis. Nesse contexto, independente do tratamento, espera-se aspirar em média, respectivamente, uma quantidade aproximada de 25.922, 26.387 e 25.997, isto é, aproximadamente 26 óocitos viáveis por aspiração, para qualquer modelo. Essa quantidade decresce com de acordo com um fator multiplicativo igual a 0.9960, 0.9959 e 0.9962, respectivamente;
- Os modelos Binomial e Beta-Binomial modelaram a proporção esperada de óocitos viáveis, sob a suposição de que a quantidade de óocitos aspirados é conhecida. Independente do tratamento recebido pelas doadoras, espera-se, de acordo com cada modelo, que aproximadamente 90.36% e 90.21%, respectivamente, dos óocitos aspirados sejam viáveis. Além disso, de acordo com cada modelo, a chance de observar um óocito viável é incrementada por um fator multiplicativo igual a 0.9945 e 0.9947, respectivamente, para cada dia após o início da suplementação;
- Os modelos Binomial-Poisson e Binomial-Poisson-Generalizado ajustaram tanto a proporção esperada de óocitos viáveis, quanto o número esperado de óocitos aspirados e, conseqüentemente, o número de óocitos viáveis. Respectivamente, de acordo com tais modelos, espera-se aspirar, em média, 29.1595 e 29.2822 óocitos por coleta, independente do tratamento. Espera-se também que dos óocitos aspirados, aproximadamente 90.35% e 90.37%, isto é, aproximadamente 26.3448 e

26.4619, sejam viáveis. A chance de coletar um oócito viável decresce com base em um fator multiplicativo igual a 0.9945 e 0.9945, enquanto o número de oócitos aspirados decresce com base em um fator multiplicativo igual a 0.9978 e 0.9968, para cada dia transcorrido após o início da suplementação.

Um resumo sobre as medidas de interesse associadas a cada modelo está exposto na Tabela 31. Note que os modelos Poisson, Binomial-Negativo e Poisson-Generalizado respondem somente sobre o total de oócitos viáveis, sendo incapazes de responder qualquer coisa sobre a quantidade esperada de oócitos aspirados ou a proporção esperada de oócitos viáveis. Analogamente, o modelo Binomial e Beta-Binomial respondem exclusivamente sobre a proporção esperada de oócitos viáveis. Os modelos Binomial-Poisson e Binomial-Poisson-Generalizado, por outro lado, são capazes de fazer inferência sobre qualquer uma destas quantidades, mostrando que são os mais completos, embora estejam entre as últimas colocações, de acordo com os critérios de informação.

Tabela 31 – Resumo das respostas práticas que cada modelo válido é capaz de fornecer.

Modelo	Respostas Associadas (por aspiração)		
	Total de Oócitos Aspirados (N)	Total de Oócitos Viáveis (Y)	Proporção de Oócitos Viáveis (P)
Binomial	conhecido	$N \times 0.8967$	$0.8967 \approx 89.67\%$
Poisson	–	$24.524 \approx 25$	–
Binomial-Poisson	$28.275 \approx 28$	$28 \times 0.8965 \approx 25$	$0.8965 \approx 89.65\%$
Binomial-Negativo	–	$25.213 \approx 25$	–
Poisson-Generalizado	–	$24.651 \approx 25$	–
Binomial-Poisson-Generalizado	$27.988 \approx 28$	$28 \times 0.8968 \approx 25$	$0.8968 \approx 89.68\%$
Beta-Binomial	conhecido	$N \times 0.8953$	$0.8953 \approx 89.53\%$

Capítulo 4

Conclusões

Pôde-se perceber que todos os modelos concordam entre si em relação a não existência de diferença entre os tratamentos pelas três abordagens: Frequentista, Bayesiana e Bayesiana Longitudinal, excetuando-se o modelo Poisson, nas duas primeiras. Além disso, pela análise Frequentista, os melhores ajustes se deram, em ordem decrescente de classificação, respectivamente por meio dos modelos Beta-Binomial, Binomial-Negativo e Poisson-Generalizado, pelos critérios de AIC, AICc, BIC e GD. Já pela análise Bayesiana os melhores ajustes se deram pelos modelos Beta-Binomial, Binomial, Binomial-Negativo, Poisson-Generalizado e Binomial-Poisson-Generalizado, de acordo com os critérios DIC, EAIC e EBIC. Pela abordagem Bayesiana Longitudinal, todos os modelos foram bem ajustados e, também de acordo com os critérios DIC, EAIC e EBIC melhor se classificam na seguinte ordem: Binomial, Beta-Binomial, Poisson-Generalizado, Binomial-Negativa, Poisson, Binomial-Poisson-Generalizado e Binomial-Poisson.

Nas abordagens Frequentista e Bayesiana, de acordo com os critérios de seleção, o modelo Beta-Binomial é o que resume a realidade de modo mais apropriado. Por outro lado na abordagem Bayesiana Longitudinal, o modelo vencedor, é o modelo Binomial. Entretanto o modelo Binomial-Poisson-Generalizado, em qualquer das três abordagens, é capaz de ajustar conjuntamente a proporção média e o total esperado (por aspiração) de oócitos aspirados e viáveis, inclusive, levando em conta também a aleatoriedade inerente a ambos os totais: o total de oócitos aspirados e viáveis, o que os modelos Binomial e Beta-Binomial não fazem.

Como os modelos que se ajustaram razoavelmente foram unânimes em descartar o efeito da nova dieta, seja sobre a quantidade de oócitos aspirados ou viáveis e também sobre a proporção de oócitos viáveis, levando-se em conta somente estas quantidades, recomenda-se utilizar a dieta que for mais conveniente ao pesquisador da área. Note que esta análise não permite concluir nada acerca das etapas posteriores às quantificações

dos oócitos, ou seja, nada pode-se afirmar sobre as próximas etapas de fertilização e reprodução em bovinos.

Referências

- BOX, G. E.; TIAO, G. C. *Bayesian Inference in Statistical Analysis*. New York: John Wiley & Sons INC, 1992. 608 p.
- BROOKS, S. P.; ROBERTS, G. O. Assessing convergence of markov chain monte carlo algorithms. *Statistics and Computing*, v. 8, p. 319–335, 1998.
- BUSSAB, W. D.; MORETTIN, P. A. *Estatística Básica*. [S.l.]: Saraiva, 2010. 161 p.
- CASELLA, G.; BERGER, R. L. *Inferência Estatística*. 2. ed. [S.l.]: Cengage Learning, 2011. Único.
- CAVALIERI, F. L. B.; SANTOS, G. T.; PETIT, H.; MORGAN, J.; RIGOLON, L. P. R.; RAMOS, F. S. R.; CAPOVILLA, L. C. T.; ALBUQUERQUE, K. P. Efeitos de duas fontes de gordura (megalac® ou linhaça em grão) na dieta na produção de embriões em vacas leiteiras da raça holandesa. *Acta Scientiae Veterinariae*, v. 33, n. 1, p. 217, 2005.
- CHRISTENSEN, R.; JOHNSON, W. O.; BRANSCUM, A. J.; HANSON, T. E. *Bayesian Ideas and Data Analysis*. New York: Taylor & Francis Inc, 2010. 473 p.
- CONSUL, P.; FAMOYE, F. Generalized poisson regression model. *Communications in Statistics - Theory and Methods*, Michigan - USA, v. 21, n. 1, p. 89–109, Dec. 1992.
- CONSUL, P.; FAMOYE, F. Generalized poisson regression model. *Communications in Statistics - Theory and Methods*, London - England, v. 21, n. 1, p. 89–109, July 2007.
- CONSUL, P. C.; JAINT, G. C. A generalization of the poisson distribution. *Technometrics*, Alberta - Canada, v. 15, n. 4, p. 791–799, Nov. 1973.
- GEWEKE, J. F. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: PRESS, U. (Ed.). *Bayesian Statistics*. [S.l.]: F. R. Minneapolis, 1992. p. 169–193.
- GOLDSTEIN, H. *Multilevel Statistical Models*. [S.l.]: Institute of Education, Multilevel Models Project - Edward Arnold, 1999.
- HEIDELBERGER, P.; WELCH, P. D. Simulation run length control in the presence of an initial transient. *Operations Research*, v. 31, p. 1109–1144, 1983.
- MAGALHÃES, M. N. *Probabilidade e Variáveis Aleatórias*. 3ª edição. ed. São Paulo - SP: EDUSP, 2011. Único. 411 p.

- MORETTIN, L. G. *Estatística Básica: probabilidade e inferência*. São Paulo - SP: Pearson Prentice Hall, 2010. Volume Único. 378 p.
- NAGAMINE, C. M. L. *Modelos para Dados de Contagem: um estudo sobre o número de ovos do mosquito Aedes aegypti*. 85 f. Dissertação (Dissertação) — Universidade Federal de São Carlos, São Carlos - SP, 2007.
- NAVARRO, D.; PERFORS, A. *An introduction to the Beta-Binomial model*. Adelaide - Austrália, 2012. 1-9 p. Notas de aula.
- OLIVEIRA, F. A. M. de. *Efeito da Adição da Mistura de Própolis na Dieta de Novilhas Nelore na Qualidade de Oócitos e Produção In Vitro de Embriões*. Dissertação (Mestrado) — Universidade Estadual de Maringá, 2012.
- RAMOS, R. dos S. *Brasil se destaca no cenário mundial da produção in vitro de embriões bovinos*. Cravinho, São Paulo, 2015. Disponível em: <<http://www.ourofinosaudeanimal.com/blog/brasil-se-destaca-no-cenario-mundial-da-producao-v/>>.
- RAUDENBUSH, S. W.; BRYK, A. S. *Hierarchical Linear Models: applications and data analysis methods*. London: Sage Publications, 2002.
- RIZOS, D.; WARD, F.; DUFFY, P.; BOLAND, M. P.; LONERGAN, P. Consequences of bovine oocytes maturation, fertilization and early embryo development in vitro versus in vivo: implications for blastocyst yield and blastocys quality. *Mol. Reprod. Dev.*, v. 61, p. 234–248, 2002.
- SCHERVISH, M. J. *Theory of Statistics*. Pittsburgh - USA: [s.n.], 1995. 702 p.
- SHKEDY, Z.; MOLENBERGHS, G.; CRAENENDONCK, H. V.; STECKLER, T.; BIJNENS, L. A hierarchical binomial-poisson model for the analysis of a crossover design for correlated binary data when the number of trials is dose-dependent. *Journal of Biopharmaceutical Statistics*, v. 15, p. 225–239, 2005.
- SIRARD, M.-A.; RICHARD, F.; BLONDIN, P.; ROBERT, C. Contribution of the oocyte to embryo quality. *Theriogenology*, v. 65, p. 126–136, 2006.
- SOUZA, M. R. de; SANTOS, G. T. dos; LIMA, L. S. de; MAZUCHELI, J.; ZEOULA, L. M.; RIGOLON, L. P.; CAVALIERI, F. L. B. Oocyte production in nellore cows supplemented with long-chain fatty acid soaps. *Revista Brasileira de Zootecnia*, v. 45, n. 3, p. 138–143, 2016.
- VIANA, J. H. M.; FIGUEIREDO, A. C. S. Produção de embriões bovinos em 2014 e 2015. *Jornal O Embrião*, v. 58, p. 6–8, 2016.
- VIEIRA, L. M. *Fatores associados à produção e à sobrevivência embrionária em programas de superovulação e transferência de embriões em fêmeas da raça Holandesa em clima tropical*. Dissertação (Mestrado) — Faculdade de Medicina Veterinária e Zootecnia da Universidade de São Paulo, 2013.
- VIEIRA, L. M.; RODRIGUES, C. A.; MENDANHA, M. F.; FILHO, M. F. S.; SALES, J. N. S.; SOUZA, A. H.; SANTOS, J. E. P.; BARUSELLI, P. S. Donor category and seasonal climate associated with embryo production and survival in multiple ovulation and embryo transfer programs in holstein cattle. *Theriogenology*, v. 30, n. 1, p. 1–9, 2014.

ZANIBONI, L.; COLOMBO, A. H. B.; SOUZA, A. B. de; AKIMOTO¹, M.; CAVALIERI, F. L. B.; RIGOLON, L. P.; SEKO, M. B.; SANTOS, J. M. G. dos. Desempenho reprodutivo na produção in vitro de embriões de novilhas cruzadas (nelore x red angus), submetidas à inclusão de ionóforos na dieta. *Anais Eletrônico - VI Mostra Interna de Trabalhos de Iniciação Científica do Cesumar*, v. 6, n. 1, p. 1–14, 2012.

Apêndices

APÊNDICE A

Códigos – Abordagem Frequentista

Considere os dados estruturados do seguinte modo:

```
      y   n t o
1  39  45 1 2
2 111 128 1 6
3  56  71 1 3
4 129 141 1 5
5 114 126 1 4
6  89 106 1 6
.
.
.
```

em que n e y denotam, respectivamente, a soma total de oócitos e oócitos viáveis, coletados no decorrer do experimento para cada uma das 30 doadoras, t denota o tratamento recebido e o denota a quantidade de vezes que a respectiva doadora foi aspirada.

As estimativas e gráficos diagnósticos para a abordagem frequentista foram obtidos com auxílio do pacote GAMLSS. Em particular, as estimativas do modelo Poisson, por exemplo, foram determinadas com base nos seguintes comandos:

```
mod <- gamlss(formula=y~offset(log(o))+factor(t),
              family=PO(mu.link='log'),
              data=dados)

summary(mod)
plot(mod)
wp(mod)
```

APÊNDICE B

Códigos – Abordagem Bayesiana

Considere os dados estruturados do seguinte modo:

```

y   n t o
1  39 45 1 2
2 111 128 1 6
3  56  71 1 3
4 129 141 1 5
5 114 126 1 4
6  89 106 1 6
.
.
.

```

em que n e y denotam, respectivamente, a soma total de oócitos e oócitos viáveis, coletados no decorrer do experimento para cada uma das 30 doadoras, t denota o tratamento recebido e o denota a quantidade de vezes que a respectiva doadora foi aspirada.

As estimativas e gráficos diagnósticos para a abordagem Bayesiana foram obtidos com auxílio do pacote RJAGS e CODA. Em particular, as estimativas do modelo Binomial, por exemplo, foram determinadas com base nos seguintes comandos:

```

# ===== Modelo BINOMIAL =====
# ----- Definição do modelo -----
cat("
  model {
    for (i in 1:N) {
      y[i] ~ dbin(p[i],n[i])

```

```

        logit(p[i]) <- beta[1] + beta[2]*t[i]
        res[i] <- (y[i] - n[i]*p[i]) / sqrt(n[i]*p[i]*(1-p[i]))
    }
    for (j in 1:2){
        beta[j] ~ dnorm( 0.0, 0.0001 )
    }
    p1 <- 1/(1+exp(-beta[1]))
    p2 <- 1/(1+exp(-beta[1]-beta[2]))
} ", file="modelo.jag"
)
# ----- Dados -----
jagsdata <-list(
  N = nrow(dados),
  y = dados$y,
  n = dados$n,
  t = dados$t)
# ----- Chutes iniciais -----
inits <- list(beta=rep(0,2))
# ----- Definido o modelo jags dentro do R -----
model <- jags.model(
  file="modelo.jag",
  data=jagsdata ,
  inits=inits,
  n.chains=5,
  n.adapt=100)
update(model, n.iter=10000)

params <- c("beta","p1","p2","res","y")
mod.post <- coda.samples(model, params, n.iter=40000, thin=10)

dic.samples(model, n.iter=1000, type="pD")
EAIC = X + 2*k      ; EAIC
EBIC = X + k*log(n) ; EBIC
cbind(EAIC=EAIC, EBIC=EBIC, DIC=X)
# ----- Resumos da distribuição a posteriori -----
mod1.p = summary(mod.post)
cbind(
  round(mod1.p$statistics,5)[,1:2],
  round(mod1.p$quantiles,4)[,c(1,3,5)])

```

```
# Cadeia 2, por exemplo
plot(mod.post[[2]][,1:2])
# ----- Diagnósticos de convergência das cadeias -----
autocorr.plot(mod.post[[2]][,1:2])
gelman.plot(mod.post)
gelman.diag(mod.post)
# ----- Critério de Convergência HW -----
heidel.diag(mod.post)
# ----- Critério de Convergência de Geweke -----
round(pnorm(abs(geweke.diag(mod.post[[2]]))$z),lower.tail=FALSE)*2,4)
```

APÊNDICE C

Códigos – Abordagem Bayesiana Longitudinal

Considere os dados estruturados do seguinte modo:

n	y	d	o	t	D
1	19	16	14	2	1 1
2	26	23	26	2	1 1
3	15	12	14	4	1 2
4	33	30	28	4	1 2
5	22	21	54	4	1 2
6	20	19	81	4	1 2
.					
.					
.					

em que n e y denotam, respectivamente, a soma total de oócitos e oócitos viáveis, coletados no decorrer do experimento para cada uma das 30 doadoras, d denota a quantidade de dias transcorridos entre o início da suplementação e a coleta dos oócitos, t denota o tratamento recebido e o denota a quantidade de vezes que a respectiva doadora foi aspirada e D denota a doadora.

As estimativas e gráficos diagnósticos para a abordagem Bayesiana foram obtidos com auxílio do pacote RJAGS e CODA. Em particular, as estimativas do modelo Binomial, por exemplo, foram determinadas com base nos seguintes comandos:

```
# ----- Definição do modelo -----
cat("

```

```

model {
  for (i in 1:N) {
    y[i] ~ dbin(p[i],n[i])
    logit(p[i]) <- beta[1] + beta[2]*t[i] + beta[3]*d[i] + g[D[i]]
    res[i] <- (y[i] - n[i]*p[i])/sqrt(n[i]*p[i]*(1-p[i]))
  }
  for (i in 1:3){
    beta[i] ~ dnorm( 0.0, 0.0001 )
  }
  for (i in 1:30){
    g[i] ~ dnorm(0, tau2)
  }
  sig2 ~ dgamma( 0.001, 0.001 )
  tau2 <- 1/sig2
  p1 <- 1/(1+exp(-beta[1]))
  p2 <- 1/(1+exp(-beta[1]-beta[2]))
} ", file="modelo.jag"
)
# ----- Dados -----
jagsdata <-list(
  N = nrow(dados),
  y = dados$y,
  n = dados$n,
  t = dados$t,
  o = dados$o,
  d = dados$d,
  D = dados$D)
# ----- Chutes iniciais -----
inits <- list(beta=rep(0,3))
# ----- Definido o modelo jags dentro do R -----
model <- jags.model(
  file="modelo.jag",
  data=jagsdata ,
  inits=inits,
  n.chains=5,
  n.adapt=100)
update(model, n.iter=10000)
params <- c("beta","sig2","p1","p2","res","y")
mod.post <- coda.samples(model, params, n.iter=50000, thin=10)

```



```
dic.samples(model, n.iter=1000, type="pD")
EAIC = X + 2*k      ; EAIC
EBIC = X + k*log(n) ; EBIC
cbind(EAIC=EAIC, EBIC=EBIC, DIC=X)
# ----- Resumos da distribuição a posteriori -----
mod.BI = summary(mod.post)
s.mod.BI <- cbind(
  round(mod1.p$statistics,5)[-c(4,5),1:2],
  round(mod1.p$quantiles,4)[-c(4,5),c(1,3,5)])
plot(mod.post[[2]][,c(1:3,6)])
# ----- Diagnósticos de convergência das cadeias -----
autocorr.plot(mod.post[[2]][,c(1:3,6)])
gelman.plot(mod.post)
gelman.diag(mod.post)
# ----- Critério de Convergência HW -----
s.mod.BI <- cbind(s.mod.BI, HW=heidel.diag(mod.post)[[2]][-c(4,5),3])
# ----- Critériode convergência de Geweke -----
s.mod.BI <- cbind(
  s.mod.BI,
  GE=round(pnorm(abs(geweke.diag(mod.post[[2]]))$z),
            lower.tail=FALSE)*2,4)[-c(4,5)])
```