



UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Câmpus de Presidente Prudente

DENISE CAROLINE BOSISIO DA SILVA

**APLICAÇÃO DE MODELOS LINEARES GENERALIZADOS PARA O
ESTUDO DO PESO DE *NASCIDOS VIVOS* DO DATASUS NO
PERÍODO DE 2016 ATÉ 2021 NO ESTADO DE SÃO PAULO**

PRESIDENTE PRUDENTE

2023

DENISE CAROLINE BOSISIO DA SILVA

**APLICAÇÃO DE MODELOS LINEARES GENERALIZADOS PARA O
ESTUDO DO PESO DE *NASCIDOS VIVOS* DO DATASUS NO
PERÍODO DE 2016 ATÉ 2021 NO ESTADO DE SÃO PAULO**

Trabalho de Conclusão de Curso
apresentado ao Curso de Graduação
em Estatística da FCT/Unesp como
parte dos requisitos para obtenção do
título de bacharelado em estatística.

Orientador(a): Prof. Dr. Mário
Hissamitsu Tarumoto.

PRESIDENTE PRUDENTE

2023

S586a	Silva, Denise Caroline Bosisio da Aplicação de modleos lineares generalizados para o estudo do peso de nascidos vivos do DATASUS no período de 2016 até 2021 no estado de São Paulo / Denise Caroline Bosisio da Silva. -- Presidente Prudente, 2023 51 p. Trabalho de conclusão de curso (Bacharelado - Estatística) - Universidade Estadual Paulista (Unesp), Faculdade de Ciências e Tecnologia, Presidente Prudente Orientador: Mário Hissamitsu Tarumoto 1. Modelos lineares generalizados. 2. Normal Inversa. 3. Gama. 4. Regressão logística. 5. Nascidos Vivos. I. Título.
-------	--

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Faculdade de Ciências e Tecnologia, Presidente Prudente. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

TERMO DE APROVAÇÃO

Denise Caroline Bosisio da Silva

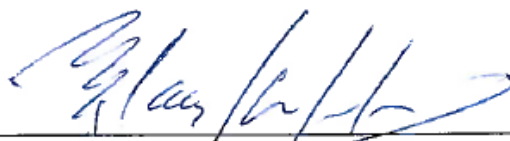
**APLICAÇÃO DE MODELOS LINEARES GENERALIZADOS PARA O
ESTUDO DO PESO DE NASCIDOS VIVOS DO DATASUS NO
PERÍODO DE 2016 ATÉ 2021 NO ESTADO DE SÃO PAULO**

Relatório de Final de Trabalho de Conclusão de Curso aprovado como requisito para obtenção de créditos na disciplina Trabalho de Conclusão do curso de graduação em Estatística da Faculdade de Ciências e Tecnologia da Unesp, pela seguinte banca examinadora:

Orientador:



Prof. Dr. Mário Hissamitsu Tarumoto
Departamento de Estatística



Prof. Dr. Klaus Schlünzen Junior
Departamento de Estatística



Prof. Dr. Miriam Rodrigues Silvestre
Departamento de Estatística

Presidente Prudente, 05 de Dezembro de 2023.

Aos meus pais, minhas irmãs que me apoiaram durante toda a graduação.
Aos meus amigos e ao meu namorado pelo incentivo e apoio de sempre.

“A vitalidade é demonstrada não apenas pela persistência,
mas pela capacidade de começar de novo.”
(F. Scott Fitzgerald)

RESUMO

O presente trabalho analisou dados de Nascidos Vivos disponíveis no Departamento de Informática do Sistema Único de Saúde, denominado DATASUS. Os dados analisados contemplaram os anos de 2016 a 2021 no estado de São Paulo. Por meio da aplicação de modelos lineares generalizados estudou-se o peso dos bebês ao nascer e observou as principais características que apresentaram significância estatística para a estimação do peso do bebê ao nascer. O resultado obtido foi: bebês do sexo masculino tendem a ter um maior peso do que bebês do sexo feminino; quanto maior o tempo de gestação, maior a probabilidade do bebê nascer com um maior peso; e gravidez duplas ou triplas influenciam no peso do bebê. Já para o modelo de regressão logística que estudou a probabilidade do bebê nascer abaixo do peso, ou seja, peso igual ou inferior a 2500 gramas (OPAS, 2019), o resultado obtido foi que bebês do sexo feminino tem 3,685 mais de chances de nascer com baixo peso do que bebês do sexo masculino; e com as variáveis sexo do bebê, número de consultas pré-natal, tipo de gravidez e anomalia congênita apresentando significância estatística.

Palavras-chave: nascidos vivos, datasus, modelos lineares generalizados, regressão logística, normal inversa, distribuição gama, baixo peso ao nascer.

ABSTRACT

The present work analyzed live birth data available in the Information Technology Department of the Unified Health System, called DATASUS. The data analyzed covered the years 2016 to 2021 in the state of São Paulo. Through the application of generalized linear models, the birth weight of babies was studied and the main characteristics that presented statistical significance for estimating the baby's birth weight were observed. The result obtained was: male babies tend to have a greater weight than female babies; the longer the gestation period, the greater the probability of the baby being born with a higher weight; and double or triple pregnancies influence the baby's weight. For the logistic regression model that studied the probability of a baby being born underweight, that is, weighing less than or equal to 2500 grams (OPAS, 2019), the result obtained was that female babies have 3.685 more chances of being born with low birth weight than male babies; and with the variables baby's sex, number of prenatal consultations, type of pregnancy and congenital anomaly presenting statistical significance.

Keywords: live births, datasus, generalized linear models, logistic regression, inverse normal, gamma distribution, low birth weight.

RESUMO	6
ABSTRACT	7
1 INTRODUÇÃO	9
1.2 JUSTIFICATIVA	10
2 REVISÃO BIBLIOGRÁFICA	11
2.1 SINASC	11
2.2 MODELOS LINEARES GENERALIZADOS	11
3 METODOLOGIA	13
3.1 MODELO DE REGRESSÃO LINEAR MULTIVARIADA	14
3.3 MODELO COM DISTRIBUIÇÃO GAMA	15
3.3 MODELO DE REGRESSÃO LOGÍSTICA	15
3.4 AJUSTE DOS MODELOS	16
3.4.1 Modelo normal inversa	16
3.4.1.1 Qualidade do ajuste	16
3.4.1.2 Diagnóstico do modelo	17
3.4.2 Modelo gama	17
3.4.2.1 Qualidade do ajuste	17
3.4.2.2 Diagnóstico do modelo	17
3.4.3 Modelo de regressão logística	18
3.4.3.1 Qualidade do ajuste	18
3.4.3.2 Diagnóstico do modelo	19
3.5 TRATAMENTO DOS DADOS	19
4 RESULTADOS	21
4.1 ANÁLISE EXPLORATÓRIA	21
4.1.1 Amostra	21
4.2 MODELO COM DISTRIBUIÇÃO NORMAL INVERSA	26
4.2.1 Normal Inversa ligação canônica	26
4.2.2 Normal Inversa ligação log	31
4.3 MODELO DE REGRESSÃO MÚLTIPLA GAMA	34
4.3.1 Gama com ligação canônica	35
4.3.2 Gama com ligação logarítmica	38
4.4 COMPARAÇÃO DOS MODELOS	41
4.4 MODELO DE REGRESSÃO LOGÍSTICA	42
5 CONSIDERAÇÕES FINAIS	47
REFERÊNCIA	49

1 INTRODUÇÃO

Atualmente, muitas empresas públicas e privadas armazenam grande quantidade de dados, necessitando de tratamento e análises estatísticas desses dados. Dentre os vários conjuntos de dados disponíveis para consulta pública no Brasil, tem-se o DATASUS. Nele há diversas informações armazenadas sobre o sistema de saúde do país, podendo ser usado para construção de modelos estatísticos.

O DATASUS - denominado Departamento de Informática do Sistema Único de Saúde -, surgiu no início dos anos 90 com a criação da Fundação Nacional de Saúde. Esse repositório tinha como principal objetivo a coleta, processamento e disseminação de informações sobre os índices de saúde pública do Brasil. Hoje, esse departamento pertence à Secretaria Executiva do Ministério da Saúde (MS), tendo como objetivo estruturar sistemas de informação, integrar dados em saúde e auxiliar na gestão de diversos níveis de atenção à saúde.

Entre os vários sistemas administrativos existentes no DATASUS, o presente projeto utilizará os dados dos arquivos disponíveis para download do Sistema de Informações de Nascidos Vivos (SINASC), referente aos anos de 2016 a 2021 no estado de São Paulo. Por tratar-se de bases grandes, há a necessidade de um tratamento de dados iniciais para, assim, poder aplicar técnicas de análise estatística.

O objetivo do presente trabalho é utilizar modelos lineares generalizados para estudar o peso dos bebês ao nascer e as principais características que podem ser utilizadas para estimar essa variável, assim como estimar a possibilidade do bebê nascer com baixo peso - identificado por bebês com um peso inferior a 2500 gramas (ou 2,5 kg) ao nascer (Organização Pan-Americana de Saúde - OPAS, 2019). Para isso, utilizou-se os modelos lineares generalizados com distribuição normal inversa e distribuição gama e o modelo de regressão logística.

Ademais, estudar e conhecer o cenário da saúde pública - especificamente sobre os nascidos vivos -, nos aproxima do Objetivo de Desenvolvimento Sustentável número 3, relativo à Saúde e o Bem-estar - que tem como propósito garantir uma vida saudável, promovendo o bem-estar para todos, em todas as idades. Além do mais, recém nascidos com baixo peso ou prematuros representam

um maior gasto para o sistema de saúde, pois precisam permanecer por mais tempo nas unidades de saúde.

1.2 JUSTIFICATIVA

Na construção de modelos de regressão linear, simples ou multivariada, uma das suposições básicas é a necessidade de que a variável resposta tenha uma distribuição normal. Existem várias técnicas que podem ser utilizadas para transformar estas variáveis não normais em normais. Em outras situações, como por exemplo em situações em que temos dados de contagem, a justificativa para a transformação pode não fazer sentido.

Em função desta problemática, Nelder e Wedderburn propuseram os modelos lineares generalizados (MLG) em 1972 (Paula, 2013). Estes modelos têm a vantagem de poder acomodar várias situações de dados, com diferentes distribuições de probabilidades. A exigência é que a função densidade de probabilidade possa ser escrita como a forma de uma família exponencial. Este modelo permite a construção de modelos de regressão para diferentes distribuições de probabilidade, mesmo sendo dados assimétricos, dados dicotômicos e até dados de contagem. O modelo de regressão logístico, muito conhecido e aplicado na prática, é um caso particular de modelo linear generalizado.

O histograma de pesos de bebês nascidos vivos do DATASUS não apresenta simetria, indicando que não poderá ser considerado como proveniente de uma distribuição normal. Desta forma, justifica-se o estudo desta área de pesquisa para construção de MLG para o estudo de dados de nascidos vivos para modelar o peso do bebê ao nascer, bem como para estimar a probabilidade de nascer o bebê com baixo peso - ou seja, peso inferior a 2500 gramas (OPAS, 2019).

2 REVISÃO BIBLIOGRÁFICA

2.1 SINASC

As bases de dados disponíveis no DATASUS são amplamente estudadas, pois fornecem informações variadas sobre diferentes aspectos da saúde no Brasil. Portanto, há inúmeros trabalhos disponíveis que abordam essas variáveis e com variados focos.

O presente trabalho utilizou-se da base de dados do Sistema de Informações de Nascidos Vivos (SINASC), sendo assim, os trabalhos pesquisados foram relacionados a esse banco de dados específico.

É vasto a quantidade de trabalhos que utilizam as bases de dados do SINASC. Um dos trabalhos estudou os fatores determinantes para o baixo peso dos bebês no município de Campinas-SP no ano de 2001, o resultado obtido foi recém nascidos nascidos por cesárea, prematuros, gestão gemelar ou múltiplas, os femininos e filhos de mulheres com pré-natal inadequado (Emília de Faria Carniel, Maria de Lourdes Zanolli, et al., 2002).

Por outro lado, o estudo de um modelo multivariado log-normal dos fatores de risco para baixo peso na cidade de São Paulo - SP, indicam que dentre os fatores estipulados a priori, os que foram considerados fatores de atenção foram ausência de assistência pré-natal, peso pré-gestacional menor que 50 quilos, tabagismo na gestação e idade materna menor que 20 anos (Benicio, Maria Helena D'Aquino; Monteiro, Carlos Augusto; et al, 1985).

Por sua vez, um estudo realizado no estado de São Paulo, no ano de 1992, com dados coletados durante 6 meses; obteve como resultado estatísticos significativos as seguintes variáveis: sexo feminino, prematuridade, mãe adolescente, mãe idosa e paridade materna (Costa, Cristina Elizabeth; Gotlieb, Sabina Léa Davidson, 1998).

2.2 MODELOS LINEARES GENERALIZADOS

As referências de textos utilizadas para o estudo dos modelos lineares generalizados são os livros de McCullagh e Nelder (1989) e Cordeiro (1986). Posteriormente, podemos citar outras referências como Turkman e Silva (2000), Cordeiro e Demétrio (2008) e Demétrio (2001).

No livro *Generalized Linear Models*, McCullagh e Nelder abordam métodos para análise de diversos tipos de dados, dando uma maior ênfase em casos que a dependência entre as variáveis ocorre por meio de combinações lineares desconhecidas das variáveis explicativas. A adição de tópicos relacionados a vários modelos que são amplamente utilizados hoje nas técnicas de Modelos Lineares Generalizados fizeram com que o livro continuasse atual.

Os demais autores utilizam-se dos tópicos usuais, com o incremento de tópicos voltados para o diagnóstico e ajuste dos modelos, importante para sabermos a adequação dos modelos aos dados e a determinadas situações, assim como a proporção de acerto.

3 METODOLOGIA

Com o propósito de alcançar os objetivos apresentados anteriormente, foi realizada uma revisão dos conceitos básicos de análise de dados, visualização de dados e modelagem estatística - em especial os modelos lineares generalizados.

A base de dados escolhida como objeto de estudo foi o Sistema de Informação de Nascidos Vivos (SINASC) do datasus, referente aos anos de 2019 e 2020 no estado de São Paulo e, portanto, foi realizado também um estudo sobre como os dados estão sendo apresentados para que, assim, seja possível analisá-los e manipulá-los durante a modelagem.

O trabalho foi desenvolvido utilizando o software RStudio para manipulação dos dados e o desenvolvimento dos gráficos, assim como a posterior aplicação do modelo.

Os modelos lineares generalizados podem ser utilizados em situações em que a variável resposta não segue uma distribuição normal, podendo ser um conjunto de dados assimétricos ou discretos. A exigência é que estes dados possam ser escritos como uma família exponencial:

$$f(y_i; \theta_i, \phi) = \exp[\phi\{\theta_i - b(\theta_i)\} + c(y_i, \phi)].$$

Um dos principais pontos a ser considerado nestes modelos, é a função de ligação. O objetivo desta função é fazer a ligação entre a média da distribuição e as variáveis independentes, também conhecidas como covariáveis. Ou seja, podemos escrever $\mu_i = E(Y_i)$, no preditor linear:

$$g(\mu_i) = \eta_i = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} e_i,$$

como a função é inversível, também pode-se escrever:

$$\mu_i = g^{-1}(\eta_i) = g^{-1}(\alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} e_i).$$

3.1 MODELO DE REGRESSÃO LINEAR MULTIVARIADA

Em geral, os dados sobre peso corporal comportam-se como um modelo normal. Temos que se $Y \sim N(\mu, \sigma^2)$, com distribuição densidade de probabilidade parecida com

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(y - \mu)^2\right\},$$

com $-\infty < \mu, y < \infty$ e $\sigma^2 > 0$.

Ao aplicarmos o exponencial na função $f(y)$ chegamos a conclusão que a distribuição pode ser escrita na forma da família exponencial, ou seja,

$$f(y_i; \theta_i, \phi) = \exp[\phi\{\theta_i - b(\theta_i)\} + c(y_i, \phi)]$$

com $\theta = \mu$, $b(\theta) = \frac{\theta^2}{2}$, $\phi = \sigma^{-2}$ e $c(y, \phi) = \frac{1}{2} \log \frac{\phi}{2\pi} - \frac{\phi y^2}{2}$.

Sendo assim, o modelo escolhido para estimar o peso do bebê ao nascer é o modelo de regressão linear multivariada,

$$Y_i = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + e_i.$$

Que é um caso particular dos modelos lineares generalizados.

Entretanto, como não há a possibilidade de termos um peso negativo, o ideal é utilizar o modelo log-normal ou o modelo da normal inversa - mais adequado para dados não negativos e com assimetria positiva.

A distribuição de probabilidade da normal inversa é denotado por uma variável aleatória Y com média μ e com parâmetro de dispersão dado por ϕ^{-1} , portanto, $Y \sim NI(\mu, \phi^{-1})$. Portanto,

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

com $y > 0$, $\mu > 0$. E, quanto mais o parâmetro de dispersão ϕ aumenta mais a distribuição fica mais simétrica em torno da média.

3.3 MODELO COM DISTRIBUIÇÃO GAMA

O modelo Gama será utilizado para fins de comparação com o modelo de regressão linear múltipla. Para compararmos o melhor modelo para estimar o peso do bebê ao nascer.

Esse modelo é indicado para dados contínuos positivos e, também, para dados assimétricos. Apesar de não termos dados assimétricos para o peso do bebê ao nascer, temos dados contínuos positivos.

Temos que $Y \sim G(\mu, \phi)$, com a função densidade de probabilidade escrita como

$$f(y_i; \mu, \phi) = \frac{1}{\Gamma(\phi)} \left(\frac{\phi y}{\mu} \right)^\phi \exp\left(-\frac{\phi y}{\mu}\right) d(\log y).$$

E, ao aplicarmos o exponencial, temos:

$$= \exp[(\phi\{-y/\mu\} - \log \mu) - \log \Gamma(\phi) + \phi \log(\phi y) - \log y].$$

em que $y > 0$, $\phi > 0$, $\mu > 0$ e $\Gamma(\phi) = \int_0^\infty t^{\phi-1} e^{-t} dt$ é a função gama.

Ela é utilizada para comparação com o modelo normal, pois a medida que ϕ aumenta a distribuição gama fica mais simétrica, ou seja, quanto mais ϕ aumenta, mais a variável resposta Y se aproxima da distribuição normal com os parâmetros média μ e variância $\mu^2 \phi^{-1}$.

3.3 MODELO DE REGRESSÃO LOGÍSTICA

Temos o interesse de estudar a probabilidade do bebê nascer com baixo peso, isto é, com peso igual ou inferior a 2500 gramas. Nesse caso, o método escolhido para desenvolver o estudo foi criar uma variável igual a 1 ou 0, ou seja,

- 1 = peso $\leq$ 2500 gramas;
- 0 = peso $>$ 2500 gramas.

Em vista disso, teremos variáveis dicotômicas e o modelo escolhido para trabalhá-las foi o modelo de regressão logística, que tem como objetivo encontrar a relação entre os dois fatores nos dados. O modelo é dado por:

$$\log\left\{\frac{\pi(X)}{1-\pi(X)}\right\} = \alpha + \beta_1 x_1 + \dots + \beta_k x_k,$$

com $X = (1, x_2, \dots, x_k)^T$ contendo os valores observados das variáveis explicativas.

3.4 AJUSTE DOS MODELOS

Estudar o ajuste de modelos é parte importante da verificação da existência de possíveis afastamentos dos pressupostos feitos para o modelo. Além disso, nos ajuda a saber se o modelo escolhido é o ideal para determinados dados.

3.4.1 Modelo normal inversa

3.4.1.1 Qualidade do ajuste

A qualidade do ajuste é medida através do desvio do modelo, no caso do modelo com distribuição normal inversa é dados por $D^*(y; \hat{\mu}) = \phi D^*(y; \hat{\mu})$, com

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

com $\hat{\mu}_i = g^{-1}(\hat{\eta}_i)$, $\hat{\eta}_i = X_i^T \hat{\beta}$ e $y_i > 0$.

Levando-se em consideração que ϕ é desconhecido, pode-se estimar por

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

3.4.1.2 Diagnóstico do modelo

A principal variável para o diagnóstico de modelos é o resíduo, nesse caso, o resíduo - como componente do desvio padronizado - é dado por:

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

em que $y_i > 0$, e $\hat{h}_{ii}$ é o i -ésimo elemento da diagonal principal da matriz

$$H = W^{1/2} X (X^T W X)^{-1} X^T W^{1/2} \text{ com } \omega_i = \frac{(d\mu_i/d\eta_i)^2}{\mu_i^3}.$$

3.4.2 Modelo gama

3.4.2.1 Qualidade do ajuste

Para modelos com distribuição gama, o desvio é dado por $D^*(y; \hat{\mu}) = \phi D^*(y; \hat{\mu})$, com

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

com $\hat{\mu}_i = g^{-1}(\hat{\eta}_i)$, $\hat{\eta}_i = X_i^T \hat{\beta}$.

Como ϕ é desconhecido, pode-se estimar através da máxima verossimilhança que pode ser dada por

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

com $\psi(\phi) = \Gamma'(\phi)/\Gamma(\phi)$. Ou utilizar a estimativa consistente

$$\hat{\phi}^{-1} = (n - p)^{-1} \sum_{i=1}^n (y_i - \hat{\mu}_i)^2 / \hat{\mu}_i.$$

3.4.2.2 Diagnóstico do modelo

Para o modelo gama, o resíduo componente do desvio padronizado é dado por:

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

em que $y_i > 0$, e h_{ii} é o i -ésimo elemento da diagonal principal da matriz

$$H = W^{1/2} X (X^T W X)^{-1} X^T W^{1/2} \text{ com } \omega_i = \frac{(d\mu_i/d\eta_i)^2}{\mu_i^3}.$$

3.4.3 Modelo de regressão logística

3.4.3.1 Qualidade do ajuste

A qualidade do ajuste é definida pela estatística que compara dois grupos. O primeiro é o grupo contendo o número observado, portanto, deve ter n_1' elementos correspondente a n_1' menores probabilidades ajustadas que são denotadas por $\hat{\pi}_{(1)} \leq \hat{\pi}_{(2)} \leq \dots \leq \hat{\pi}_{(n_1')}$. Já o segundo grupo irá conter os n_2' elementos correspondentes às probabilidades ajustadas $\hat{\pi}_{(n_2'+1)} \leq \hat{\pi}_{(n_2'+2)} \leq \dots \leq \hat{\pi}_{(n_1'+n_2')}$.

Até o último grupo que deverá conter as n_g' maiores probabilidades ajustadas por $\hat{\pi}_{(n_1'+\dots+n_{g-1}'+1)} \leq \hat{\pi}_{(n_1'+\dots+n_{g-1}'+2)} \leq \dots \leq \hat{\pi}_{(n)}$.

Portanto, o número de sucessos observado no primeiro grupo será dados por

$$O_1 = \sum_{j=1}^{n_1'} y_{(j)},$$

em que $y_{(j)} = 0$ se o elemento correspondente for fracasso e $y_{(j)} = 1$ se for sucesso. Generalizando, a expressão toma a seguinte forma:

$$O_i = \sum_{j=n_1'+\dots+n_{i-1}'+1}^{n_1'+\dots+n_i'} y_{(j)},$$

com $2 \leq i \leq g$.

Sendo assim, a estatística é definida por:

$$\hat{C} = \sum_{i=1}^g \frac{(O_i - n_i' \bar{\pi}_i)^2}{n_i' \bar{\pi}_i (1 - \bar{\pi}_i)},$$

com

$$\bar{\pi}_1 = \frac{1}{n_1} \sum_{j=1}^{n_1} \hat{\pi}_{(j)}$$

e

$$\bar{\pi}_i = \frac{1}{n_i} \sum_{j=n_1+\dots+n_{i-1}+1}^{n_1+\dots+n_i} \hat{\pi}_{(j)}.$$

3.4.3.2 Diagnóstico do modelo

Para o modelo de regressão logística, o desvio do resíduo é dado por:

$$t_{D_i} = \pm \sqrt{\frac{2}{1-h_{ii}}} \left\{ y_i \log \left(\frac{y_i}{n_i \hat{\pi}_i} \right) + (n_i - y_i) \log \left(\frac{n_i - y_i}{n_i - n_i \hat{\pi}_i} \right) \right\}^{1/2},$$

com sinal igual a $y_i - \hat{y}_i$.

Entretanto, quando $y_i = 0$ ou $y_i = n_i$, o componente do desvio padronizado pode ser calculado por meio de

$$t_{D_i} = - \frac{\{2n_i |\log(1-\hat{\pi}_i)|\}^{1/2}}{\sqrt{1-\hat{h}_{ii}}} \text{ e } t_{D_i} = - \frac{\{2n_i |\log \hat{\pi}_i|\}^{1/2}}{\sqrt{1-\hat{h}_{ii}}}.$$

Já o resíduo Studentizado, utilizado para observar pontos destoantes, é dado por:

$$t_{S_i} = \frac{1}{\sqrt{1-\hat{h}_{ii}}} \frac{(y_i - n_i \hat{\pi}_i)}{\{n_i \hat{\pi}_i (1-\hat{\pi}_i)\}^{1/2}}.$$

3.5 TRATAMENTO DOS DADOS

Para a limpeza dos dados do DATASUS, Sistema de Informações de Nascidos Vivos (SINASC), foi utilizado o pacote *dplyr* do software RStudio.

O conjunto total dos dados conta com aproximadamente 4 milhões de linhas e, para que fosse possível rodar todos os modelos, estudar e implementar os ajustes necessários, optou-se por trabalhar com uma amostra.

O método de amostragem escolhido foi a amostragem aleatória simples sem repetição de tamanho 5000. Entretanto, como algumas variáveis estavam com

ausência de informação transformou-se em 0 os espaços vazios. Isso foi feito pois não interferem no modelo, já que a ausência de informações não estava presente nas variáveis explicativas. Ademais, foram retiradas 66 observações da amostra que continham erro, ficando com 4934.

É importante ressaltar que a variável peso está na forma de gramas e não de quilos, ou seja, um bebê que nasceu com 3,5 quilos estará representado na base com 3500 gramas.

Para as variáveis no formato factor do modelo utilizou-se o default automático do R que transforma a variável em dummy e tem como referência o menor número.

4 RESULTADOS

4.1 ANÁLISE EXPLORATÓRIA

4.1.1 Amostra

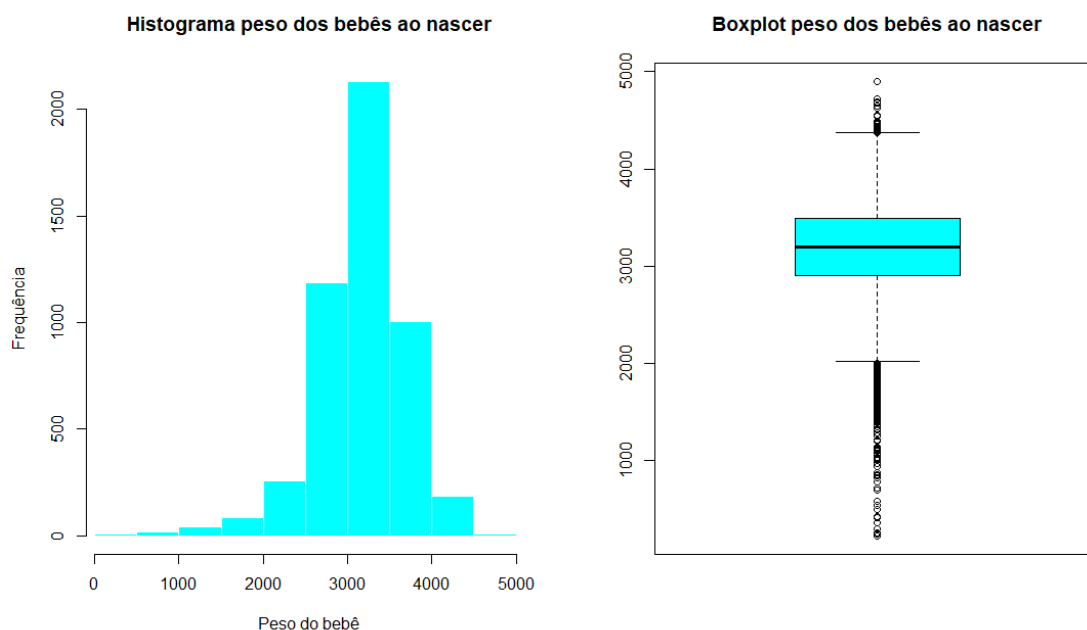
Podemos observar na tabela 1 que o número de nascimentos, entre os anos de 2016 e 2021, teve uma variação no número de nascimentos no período de 2016 e 2021, sendo que o menor número de nascimentos foi observado no ano de 2021, seguido de nascimentos no ano de 2020.

Tabela 1: Número de observações por ano na amostra e sua proporção.

Ano	Número de observações	Proporção (em %)
2016	837	17
2017	905	18
2018	901	18
2019	820	17
2020	768	16
2021	703	14

Fonte: elaboração própria com base nos dados do SINASC de 2016 a 2021 no estado de São Paulo.

Como variável resposta do modelo - ou variável dependente, será utilizado o peso dos bebês ao nascerem. O histograma e o boxplot desses dados podem ser visualizadas na figura 1:

Figura 1: histograma e boxplot do peso dos bebês ao nascer.

Observa-se que há assimetria nos dados, portanto, não podemos dizer que segue aproximadamente uma distribuição normal. Há uma grande presença de dados discrepantes, porém, a presença deles na modelagem é essencial, pois o intuito é estimar o peso do bebê ao nascer e, se retirarmos essas observações, teremos perda de informações e os dados não mais serão representativos da realidade.

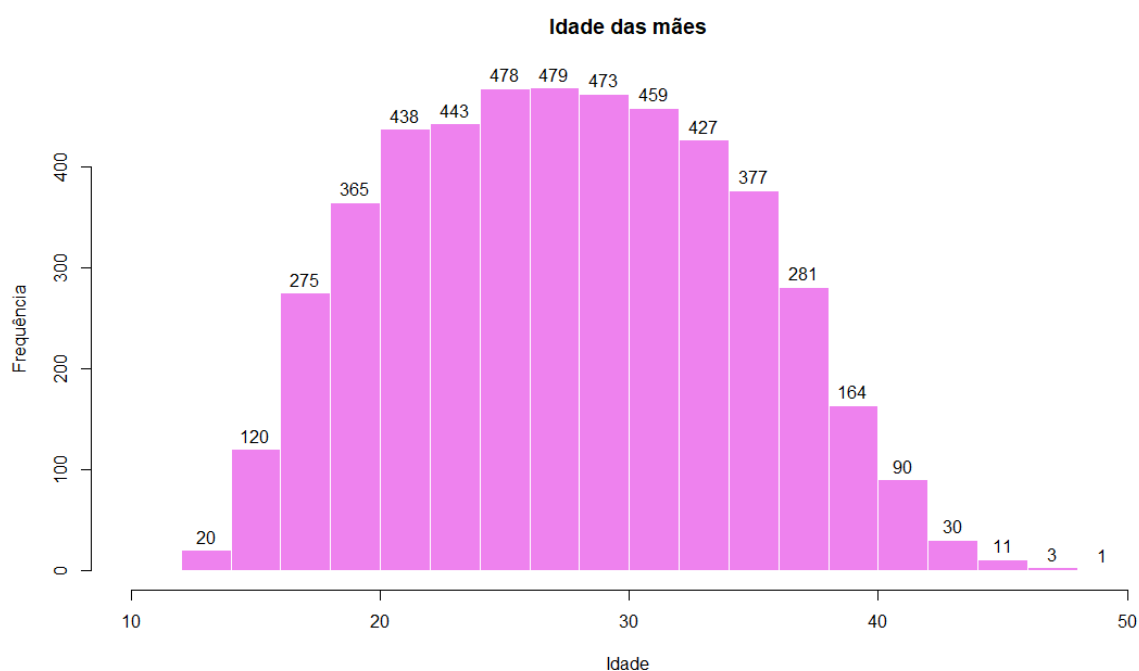
A média amostral calculada foi de 3.163 gramas e a mediana de 3.200 gramas, como pode ser observado na tabela 2. Observou-se ainda que nasceu um bebê com 230 gramas e o máximo observado na amostra foi de 4900 gramas.

Tabela 2: Medidas centrais e de variabilidade do peso do bebê ao nascer (em gramas).

Mínimo	Média	Mediana	Variância	Desvio-padrão	Máximo
230	3163	3200	296163,1	544,21	4900

Fonte: elaboração própria com base nos dados do SINASC de 2016 a 2021 no estado de São Paulo.

Na figura 2, é apresentado o histograma das idades das mães na amostra, pode-se observar um comportamento próximo da simetria.

Figura 2: gráfico da idade das mães ao darem a luz.

Na tabela 3, vemos que a idade média das mães que tiveram bebê entre os anos de 2016 e 2021 é de 28 anos, que é igual a mediana, indicando a simetria. Temos também que a idade mínima presente na amostra é de 13 anos, evidenciando mães que tiveram seus filhos no final da infância e começo da adolescência. Na amostra ainda, foi observado mãe com 49 anos.

Tabela 3: Medidas de tendência central e variabilidade da idade da mãe.

Mínimo	Média	Mediana	Variância	Desvio-padrão	Máximo
13	28,01	28	45,63	6,75	49

Fonte: elaboração própria com base nos dados do SINASC de 2016 a 2021 no estado de São Paulo.

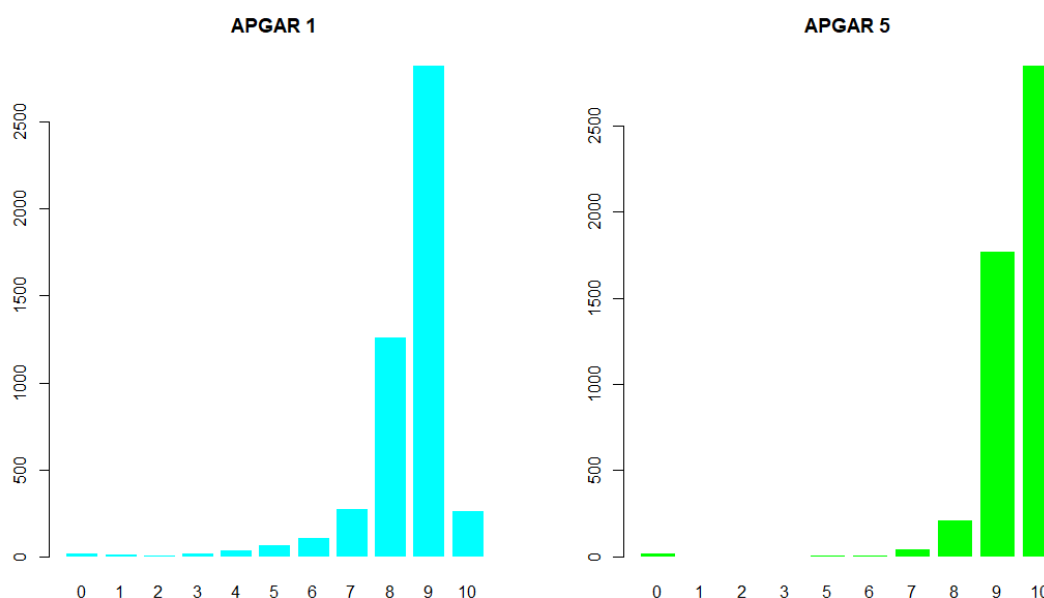
O índice de apgar é uma métrica que expressa a condição fisiológica do bebê nos primeiros minutos de vida e varia de 0 a 10, ou seja, o apgar 1 é o primeiro minuto de vida e o apgar 5 é o quinto minuto de vida. Essa métrica é composta pelo somatório de notas - que varia de 0 à 2 - dos seguintes sinais: força muscular, frequência de batimentos do coração, reflexo, respiração e cor.

Ao analisarmos os valores do apgar no primeiro minuto e no quinto minuto de vida do bebê, é perceptível que os valores estão mais concentrados em torno dos

maiores valores 7, 8 e 9 - evidenciando que mais da metade dos bebês apresentam uma boa vitalidade.

Observa-se também que, enquanto no apgar 1 há muitos bebês com índices baixos, no apgar 5 esse número diminuiu e concentra-se mais nas notas 8, 9 e 10.

Figura 3: Valores do índice de apgar no primeiro e no quinto minuto de vida do bebê.



Para os dados de tempo de gravidez, temos o seguinte cenário:

Tabela 4: Distribuição de frequência do tempo de gravidez (em semanas) e sua proporção (em %).

Semanas de gestação	≤ 22	22 à 27	28 à 31	32 à 36	37 à 41	≥ 42
Número de observações	3	29	43	452	4340	67
Proporção (em %)	0,06	0,59	0,87	9,16	87,96	1,36

Fonte: elaboração própria com base nos dados do SINASC de 2016 a 2021 no estado de São Paulo.

Mais da metade dos partos do período, 87,96%, ocorreram quando a mãe estava entre 37 e 41 semanas, ou seja, partos considerados não prematuros. Entretanto, do total de partos realizados nesse período, quase 11% foram

prematturos, isto é, com tempo de gestação igual ou menor que 36 semanas. Já os bebês considerados prematturos extremos - com tempo de gestação igual ou menor que 30 semanas - representaram quase 2% se comparado ao total de nascimentos e mais de 13% ao compararmos com o total de bebês prematturos.

Entre 2016 e 2021, o tipo de gravidez mais incidente foi gestão única - representando mais de 97%, como mostrado na tabela 5.

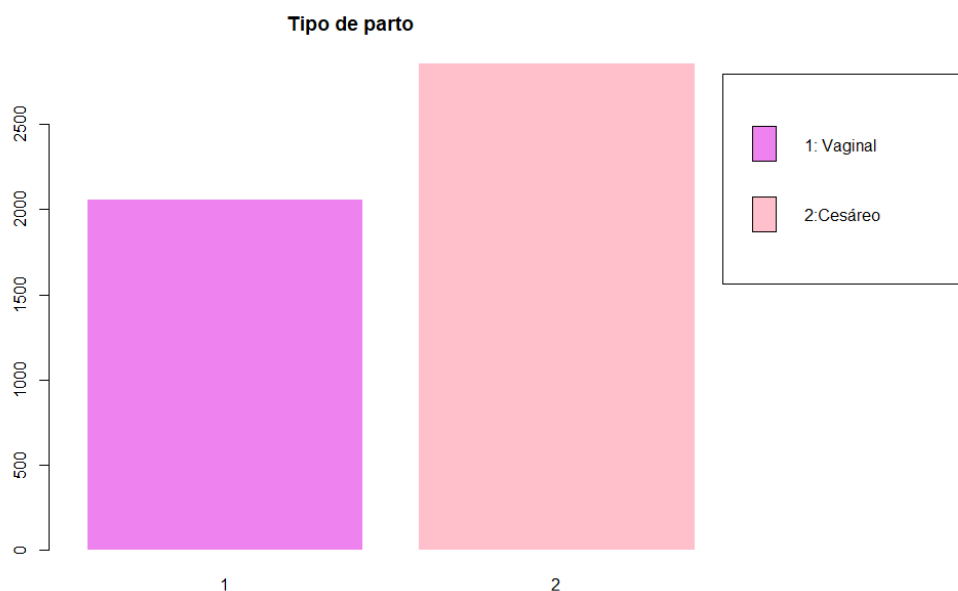
Tabela 5: Tipos de gravidez e sua proporção (em %).

Tipo de gravidez	Única	Dupla	Tripla ou mais
Número de observações	4811	117	6
Proporção (em %)	97,51	2,37	0,12

Fonte: elaboração própria com base nos dados do SINASC de 2016 a 2021 no estado de São Paulo.

Nesse mesmo período, o tipo de parto cesáreo representou mais de 58% do total de parto realizados.

Figura 4: Tipos de parto realizados entre os anos 2016 e 2021 no estado de São Paulo.



Quando olhamos o número de consultas pré-natais realizadas, temos o seguinte cenário:

Tabela 6: Quantidade de consultas de pré-natal realizadas e sua proporção (em %).

Consultas pré-natal	Nenhuma	de 1 a 3	de 4 a 6	7 e mais
Número de observações	39	176	749	3970
Proporção (em %)	0,79	3,57	15,18	80,46

Fonte: elaboração própria com base nos dados do SINASC de 2016 a 2021 no estado de São Paulo.

Portanto, entre 2016 e 2021, 80,46% das mulheres que tiveram filhos realizaram um acompanhamento igual ou superior a sete consultas. Entretanto, 4,3% das mulheres tiveram entre 0 e 3 consultas de pré-natal no período.

Para estudar o baixo peso ao nascer, foi criada uma variável dicotômica com base no peso do bebê, e obteve-se a distribuição de frequência mostrada na tabela 7, portanto, mais de 90% dos bebês nascem com o peso adequado.

Tabela 7: Distribuição de frequência da variável baixo peso ao nascer.

	≤ 2500 gramas	> 2500 gramas
Número de observações	416	4518
Proporção (em %)	8,43	91,57

Fonte: elaboração própria com base nos dados do SINASC de 2016 a 2021 no estado de São Paulo.

4.2 MODELO COM DISTRIBUIÇÃO NORMAL INVERSA

Para rodarmos o modelo de regressão generalizada da normal inversa utilizou-se o pacote estatístico *MASS* e a função *GLM* contida nele.

Para esse modelo foi utilizado as variáveis independentes: ano de nascimento, idade da mãe, quantidade de filhos vivo, quantidade de filhos morto, semanas de gestação, tipo de parto, tipo de gravidez, quantidade de consultas, sexo do bebê, apgar 1 e 5, anomalia congênita, quantidade de gestação.

4.2.1 Normal Inversa ligação canônica

Para esse primeiro modelo, a ligação escolhida foi a canônica - também chamada de recíproca quadrática, portanto,

$$\mu_i = \eta_i^{-2},$$

com $\eta_i = X_i^T \beta$, $X_i = (x_{i1}, \dots, x_{ip})^T$ e $\beta = (\beta_1, \dots, \beta_p)^T$.

Tabela 8: Estimativas dos parâmetros do modelo normal inversa com a ligação canônica.

Variável	Estimativa	Erro Padrão	t valor	P-Valor
Intercept	1,141e-5	1,303e-06	8,757	2e-16
Ano	-1,434e-10	2,871e-10	-0,499	0,6175
Idade da mãe	-4,042e-11	8,171e-11	-0,495	0,6208
Quantidade de filhos mortos	1,717e-09	3,648e-09	0,471	0,6379
Quantidade de filhos vivos	-8,887e-10	3,583e-09	-0,248	0,8041
Semanas de gestação - 22s a 27s	-1,011e-05	1,168e-06	-8,657	2e-16
Semanas de gestação - 28s a 31s	-1,065e-05	1,167e-06	-9,128	2e-16
Semanas de gestação - 32s a 36s	-1,096e-05	1,167e-06	-9,389	2e-16
Semanas de gestação - 37s a 41s	-1,101e-05	1,167e-06	-9,434	2e-16
Semanas de gestação - +42s	-1,101e-05	1,167e-06	-9,438	2e-16
Tipo de gravidez - dupla	5,109e-08	5,033e-09	10,151	2e-16
Tipo de gravidez - tripla ou mais	1,652e-07	4,591e-08	3,599	0,0003
Parto cesáreo	-3,204e-09	1,023e-09	-3,133	0,0017
Nº de consultas pré-natal - 1 a 3	3,773e-09	6,709e-09	0,562	0,5739
Nº de consultas pré-natal - 4 a 6	-2,428e-09	6,229e-09	-0,390	0,6967
Nº de consultas pré-natal - mais de 7	-5,957e-09	6,135e-09	-0,971	0,3314
Sexo do bebê: feminino	7,586e-09	9,618e-10	7,888	3,76e-15

Apgar 1	-2,896e-10	2,660e-10	-1,088	0,2764
Apgar 5	1,761e-10	3,053e-10	0,577	0,5639
Ausência de anomalia congênita	-1,380e-08	5,148e-09	-2,682	0,0074
Quantidade de gestação	-7,965e-10	3,609e-09	-0,221	0,8254

Fonte: elaboração própria com base nos dados do SINASC de 2016 a 2021 no estado de São Paulo.

Com base nas estimativas do modelo obtido na tabela 8 percebe-se que bebês do sexo feminino apresentam um menor peso em relação aos do sexo masculino; as semanas de gestação que o bebê nasce tem relação com o peso. Bebês nascidos de parto cesáreo apresentam um maior peso do que os bebês nascidos de parto normal.

O tipo de gestação também apresenta significância, ou seja, bebês de gestações duplas ou triplas tendem a nascer com um peso menor do que bebês de gravidez única. A variável anomalia congênita apresenta significância estatística, ou seja, bebês que não apresentam anomalias tendem a ter um peso maior.

Segundo o modelo, os índices de apgar 1 e 5 não apresentam significância estatística, o que é verossímil, já que esses valores são observados após o nascimento do bebê.

Sendo assim, o modelo que iremos utilizar levará as variáveis que apresentaram significância, e também com a inclusão da idade da mãe. Podemos ver os coeficientes na tabela 9.

Tabela 9: Estimativas dos parâmetros das variáveis com significância estatística do modelo normal inversa com ligação canônica.

Variável	Estimativa	Erro Padrão	t valor	P-Valor
Intercept	1,112e-05	1,68e-06	9,522	2e-16
Idade da mãe (x_1)	-1,646e-10	7,333e-11	-2,245	0,0248
Semanas de gestação - 22s a 27s (x_2)	-1,011e-05	1,169e-06	-8,648	2e-16
Semanas de gestação - 28s a 31s (x_3)	-1,065e-05	1,168e-06	-9,120	2e-16

Semanas de gestação - 32s a 36s (x_4)	-1,096e-05	1,168e-06	-9,383	2e-16
Semanas de gestação - 37s a 41s (x_5)	-1,101e-05	1,168e-06	-9,429	2e-16
Semanas de gestação - +42s (x_6)	-1,102e05	1,168e-06	-9,433	2e-16
Tipo de gravidez - dupla (x_7)	5,079e-08	5,035e-09	10,088	2e-16
Tipo de gravidez - tripla ou mais (x_8)	1,620e-07	4,594e08	3,526	0,0004
Parto cesáreo (x_9)	-2,812e-09	1,004e-09	-2,801	0,0051
Sexo do bebê: feminino (x_{10})	7,639e-09	9,616-10	7,945	2,4e-15
Anomalia congênita (x_{11})	-1,330e-08	5,148e09	-2,583	0,0098

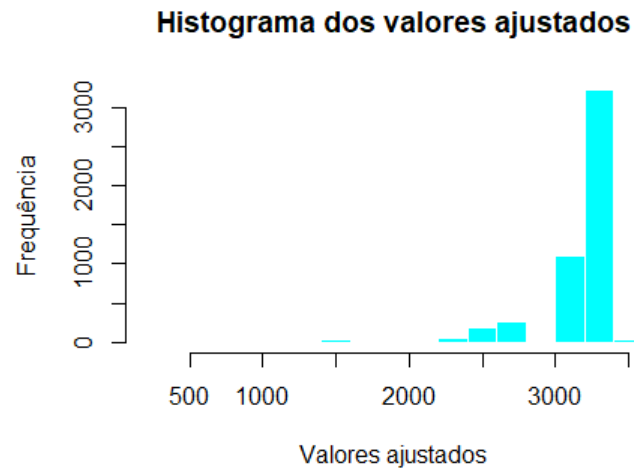
Fonte: elaboração própria com base nos dados do SINASC de 2016 a 2021 no estado de São Paulo.

Portanto, o modelo estimado é dado por:

$$\begin{aligned}
 Y_i = & 1,112204-05 - 1,646e-10x_1 - 1,011e-05x_2 - \\
 & 1,065e-05x_3 - 1,096e-05x_4 - 1,101e-05x_5 - 1,102e05x_6 + \\
 & 5,079e-08x_7 + 1,620e-07x_8 - 2,812e-09x_9 + \\
 & 7,639e-09x_{10} - 1,330e-08x_{11} + \varepsilon_{ij}.
 \end{aligned}$$

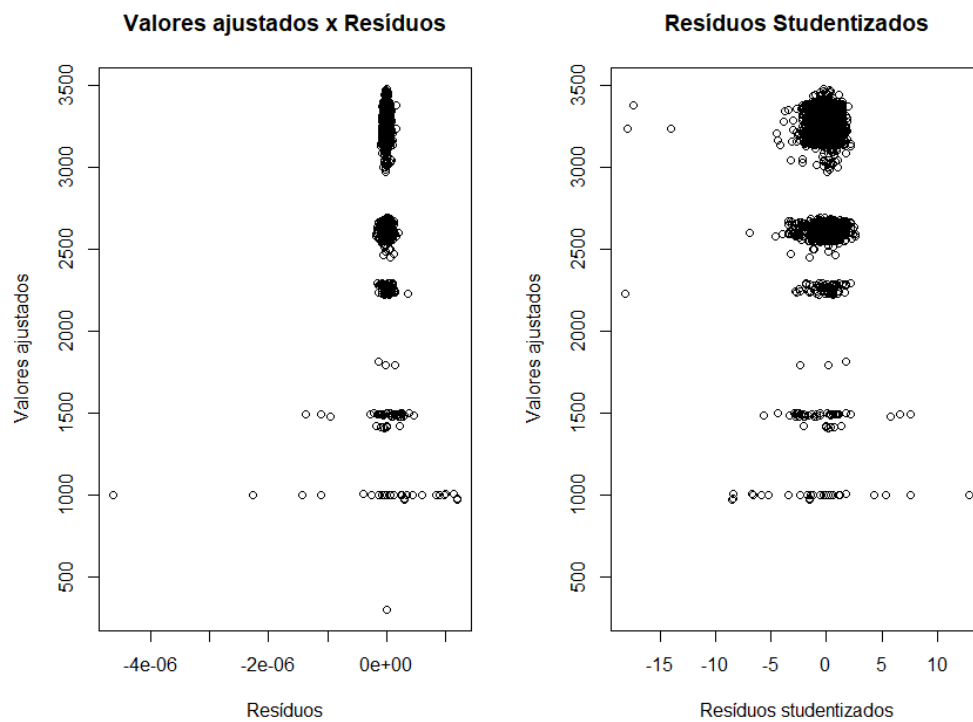
Ao observarmos a figura 5 temos que a distribuição dos dados difere da distribuição do histograma da amostra, podendo indicar uma inadequação do modelo.

Figura 5: Histograma dos valores ajustados para Normal Inversa com ligação canônica.



Na figura 6, temos o gráfico dos valores ajustados pelo resíduo e os resíduos studentizados. É possível observar que os gráficos mostram uma maior concentração de pontos entre os valores 2500g e 3500g.

Figura 6: Gráfico dos Valores Ajustados x Resíduos e Valores Ajustados x Resíduos Studentizados para Normal Inversa com ligação canônica.



Os pontos influentes pela medida h são: #480, #2635 e #4864; os três pontos são bebês do sexo feminino que nasceram com peso inferior a 2500g - em especial o ponto #4864 que nasceu com 300g.

Na tabela 10 temos a variância dos dados ajustados, indicando que é um valor maior do que o esperado para um ajuste adequado.

Tabela 10: Medidas para os valores ajustados do modelo.

Mínimo	Máximo	Mediano	Média	Variância	Desvio Padrão
300	3476	3234	3163	104.744,1	323,642

Fonte: elaboração própria com base nos dados do SINASC de 2016 a 2021 no estado de São Paulo.

4.2.2 Normal Inversa ligação log

Para esse primeiro modelo, a ligação escolhida foi a logarítmica, portanto,

$$\log \mu_i = \eta_i,$$

com $\eta_i = X_i^T \beta$, $X_i = (x_{i1}, \dots, x_{ip})^T$ e $\beta = (\beta_1, \dots, \beta_p)^T$.

Tabela 11: Estimativas dos parâmetros das variáveis com significância estatística do modelo normal inversa com ligação log.

Variável	Estimativa	Erro Padrão	t valor	P-Valor
Intercept	5,6685	0,0579	97,837	2e-16
Idade da mãe	0,0003	0,0004	0,733	0,4638
Semanas de gestação - 22s a 27s	1,1729	0,0551	21,271	2e-16
Semanas de gestação - 28s a 31s	1,5876	0,0552	28,765	2e-16
Semanas de gestação - 32s a 36s	2,1359	0,0527	40,510	2e-16
Semanas de gestação - 37s a 41s	2,3546	0,0522	45,072	2e-16
Semanas de gestação - +42s	2,3825	0,0564	42,254	2e-16
Tipo de gravidez - dupla	-0,1894	0,0138	-13,704	2e-16

Tipo de gravidez - tripla ou mais	-0,3819	0,0737	-5,180	2,31e-07
Parto cesáreo	0,0173	0,0049	3,460	0,0005
Sexo do bebê: feminino	-0,0448	0,0047	-9,400	2e-16
Anomalia congênita	0,0716	0,0219	3,271	0,0011

Fonte: elaboração própria com base nos dados do SINASC de 2016 a 2021 no estado de São Paulo.

Portanto, o modelo é dado por:

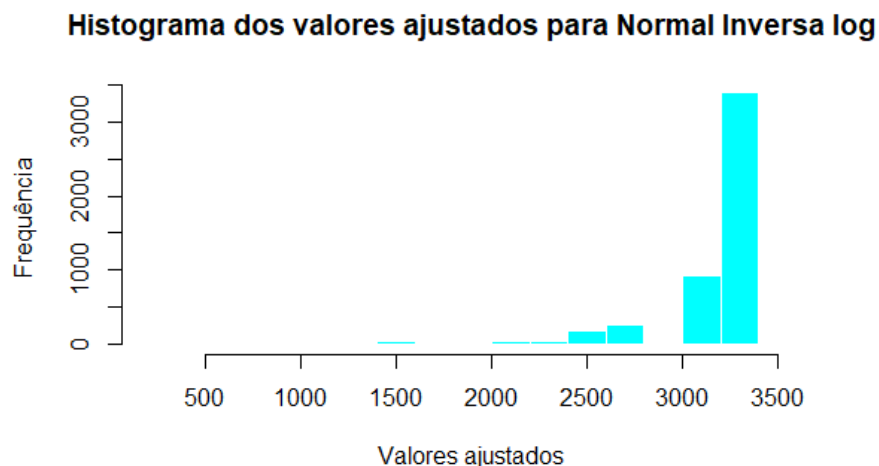
$$Y_i = \exp(5,6685 + 0,0003x_1 + 1,1729x_2 + 1,5876x_3 + 2,1359x_4 + 2,3546x_5 + 2,3825x_6 - 0,1894x_7 - 0,3819x_8 + 0,0173x_9 - 0,0448x_{10} + 0,0716x_{11} + \varepsilon_{ij}).$$

Com base nas estimativas obtidas na tabela 11, tem-se que o tempo de gestação apresenta significância estatística. O mesmo acontece com a variável tipo de gravidez, ou seja, bebês nascidos de gravidezes duplas e triplas apresentam um menor peso corporal.

Segundo o modelo, é mais comum bebês nascerem com um peso maior quando o tipo de parto foi cesáreo. Crianças nascidas com o sexo feminino apresentam um menor peso quando comparadas com crianças do sexo masculino. E, por fim, a ausência de anomalias congênitas em nascidos vivos indica um peso superior a crianças que nascem com anomalias congênitas.

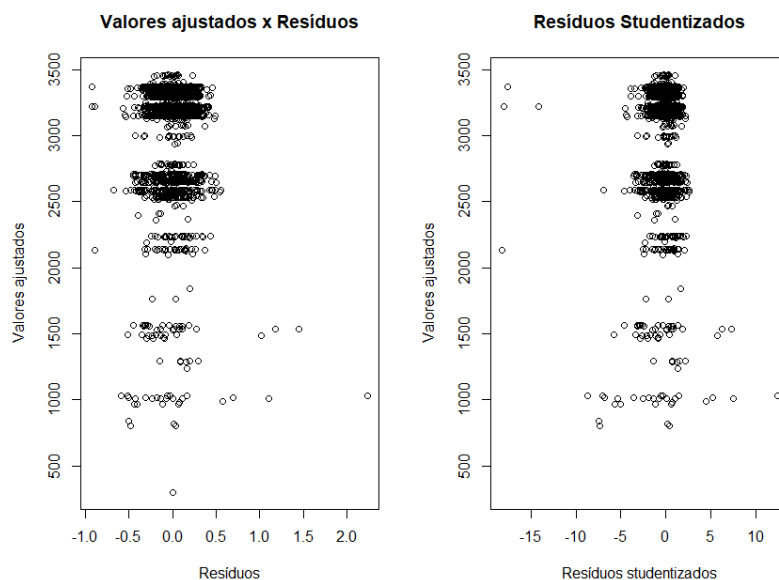
Na figura 7 pode-se observar que, os dados ajustados estão concentrados no intervalo de 3000g a 3500g.

Figura 7: Histograma dos valores ajustados para Normal Inversa com ligação canônica.

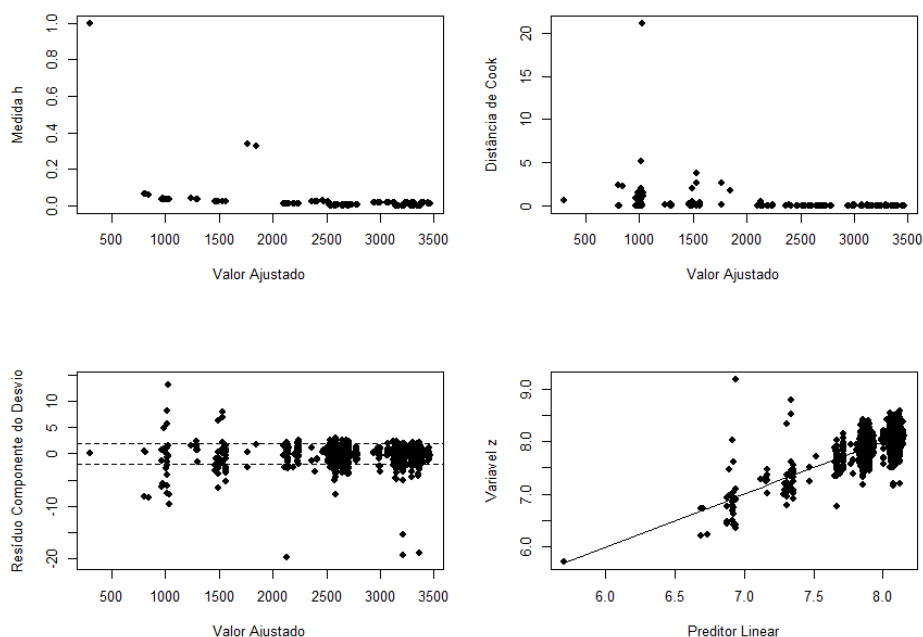


Assim como mostrado na figura 7, a figura 8 mostra uma concentração dos dados em valores acima de 2500g, indicando que a maior parte dos bebês nascidos nos anos de 2016 a 2021 tinham o peso adequado.

Figura 8: Gráfico dos Valores Ajustados x Resíduos e Valores Ajustados x Resíduos Studentizados para Normal Inversa com ligação log.



Os pontos influentes pela distância de Cook são os pontos: #832 e #2194. O mais influente foi o ponto #832, pertencente a um bebê do sexo masculino que nasceu com peso inferior a 2500g.

Figura 9: Gráficos dos pontos de influência para Normal Inversa com ligação log.

Na tabela 12 temos a variância dos dados ajustados, o que indica que a variação dos dados é maior do que o esperado para um ajuste adequado.

Tabela 12: Medidas para os valores ajustados do modelo.

Mínimo	Máximo	Mediana	Média	Variância	Desvio Padrão
300	3476	3218	3163	105.781,6	325,241

Fonte: elaboração própria com base nos dados do SINASC de 2016 a 2021 no estado de São Paulo.

4.3 MODELO DE REGRESSÃO MÚLTIPLA GAMA

Para rodarmos o modelo de regressão generalizada com distribuição Gama inversa utilizou-se o pacote estatístico MASS e a função GLM contida nele.

Para esse modelo foi utilizado as variáveis independentes: ano de nascimento, idade da mãe, quantidade de filhos vivo, quantidade de filhos morto, semanas de gestação, tipo de parto, tipo de gravidez, quantidade de consultas, sexo do bebê, apgar 1 e 5, anomalia congênita, quantidade de gestação.

4.3.1 Gama com ligação canônica

Para esse primeiro modelo, a ligação escolhida foi a ligação canônica, portanto,

$$\mu_i = \eta_i^{-1},$$

com $\eta_i = X_i^T \beta$, $X_i = (x_{i1}, \dots, x_{ip})^T$ e $\beta = (\beta_1, \dots, \beta_p)^T$.

Tabela 13: Estimativas dos parâmetros do modelo gama com ligação canônica.

Variável	Estimativa	Erro Padrão	t valor	P-Valor
Intercept	0,00382	0,00096	3,995	6.57e-05
Ano	-2,307e-07	4,030e-07	-0,572	0,5671
Idade da mãe	-6,811e-08	1,144e-07	-0,595	0,5516
Quantidade de filhos mortos	2,818e-06	5,161e-06	0,546	0,5851
Quantidade de filhos vivos	-1,370e-06	5,071e-06	-0,270	0,7871
Semanas de gestação - 22s a 27s	-0,0023	0,0005	-4,659	3,26e-06
Semanas de gestação - 28s a 31s	-0,0026	0,000502	-5,310	1,15e-07
Semanas de gestação - 32s a 36s	-0,0029	0,00049	-5,879	4,40e-09
Semanas de gestação - 37s a 41s	-0,00301	0,00049	-6,028	1,78e-09
Semanas de gestação - +42s	-0,00302	0,00049	-6,044	1,61e-09
Tipo de gravidez - dupla	6,828e-05	6,142e-06	11,118	2e-16
Tipo de gravidez - tripla ou mais	0,00017	4,821e-05	3,700	0,0002
Parto cesáreo	-5,150e-06	1,433e-06	-3,593	0,0003
Nº de consultas pré-natal - 1 a 3	6,039e-06	9,129e-06	0,661	0,5083

Nº de consultas pré-natal - 4 a 6	-3,464e-06	8,479e-06	-0,409	0,6829
Nº de consultas pré-natal - mais de 7	-9,157e-06	8,346e-06	-1,097	0,2726
Sexo do bebê: feminino	1,218e-05	1,348e-06	9,041	2e-16
Apgar 1	-4,866e-07	3,946e-07	-1,233	0,2176
Apgar 5	3,037e-07	4,494e-07	0,676	0,4992
Ausência de anomalia congênita	-2,133e-05	6,938e-06	-3,075	0,002119
Quantidade de gestação	-1,324e-06	5,108e-06	-0,259	0,7955

Fonte: elaboração própria com base nos dados do SINASC de 2016 a 2021 no estado de São Paulo.

Assim como o modelo da normal inversa dado nas tabelas 9 e 11, tem-se que bebês do sexo feminino apresentam um menor peso em relação aos do sexo masculino e as semanas de gestação que o bebê nasce tem relação com o peso, assim como bebês nascidos de parto cesáreo.

O tipo de gestação também apresenta significância, ou seja, bebês de gestações duplas ou triplas tendem a nascer com um peso menor do que bebês de gravidez única. A variável anomalia congênita apresenta significância estatística, ou seja, bebês que não apresentam anomalias tendem a ter um peso maior.

Sendo assim, o modelo que iremos utilizar levará as variáveis que apresentaram significância. Podemos ver os coeficientes na tabela 14.

Tabela 14: Estimativas dos parâmetros das variáveis com significância estatística do modelo gama com ligação canônica.

Variável	Estimativa	Erro Padrão	t valor	P-Valor
Intercept	0,0033	0,0005	6,657	2,90e-11
Semanas de gestação - 22s a 27s (x_1)	-0,0023	0,0005	-4,639	3,59e-06
Semanas de gestação - 28s a 31s (x_2)	-0,0026	0,0005	-5,294	1,25e-07

Semanas de gestação - 32s a 36s (x_3)	-0,0029	0,0005	-5,869	4,66e-09
Semanas de gestação - 37s a 41s (x_4)	-0,0030	0,0005	-6,021	1,86e-09
Semanas de gestação - +42s (x_5)	-0,0030	0,0005	-6,037	1,69e-09
Tipo de gravidez - dupla (x_6)	6,766e-05	6,152e-06	10,999	2e-16
Tipo de gravidez - tripla ou mais (x_7)	0,0001	4,831e-05	3,583	0,0003
Parto cesáreo (x_8)	-5,387e-06	1,371e-06	-3,929	8,63e-05
Sexo do bebê: feminino (x_9)	1,226e-05	1,349e-06	9,088	2e-16
Ausência de anomalia congênita (x_{10})	-2,009e-05	6,944e-06	-2,893	0,0038

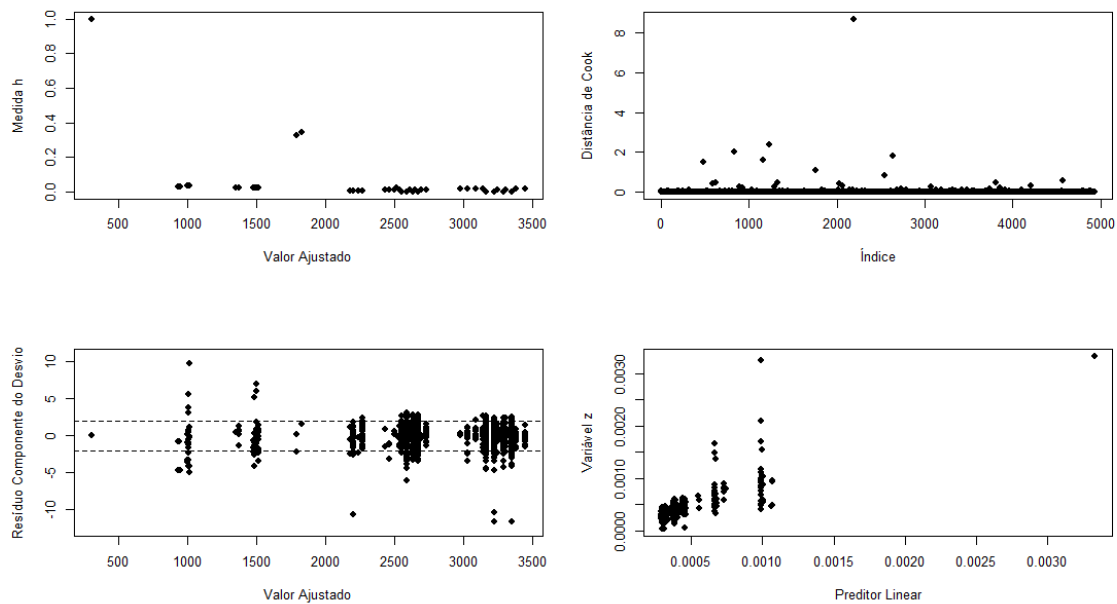
Fonte: elaboração própria com base nos dados do SINASC de 2016 a 2021 no estado de São Paulo.

Portanto, o modelo estimado é dado por:

$$Y_i = 0,0033 - 0,0023x_1 - 0,0026x_2 - 0,0029x_3 - 0,0030x_4 - 0,0030x_5 + 6,766e-05x_6 + 0,0001x_7 - 5,387e-06x_8 + 1,226e-05x_9 - 2,009e-05x_{10} + \varepsilon_{ij}.$$

O parâmetro de dispersão do modelo é dado por $\hat{\phi} = 41,6115$ e o $\hat{\mu}_i = g^{-1}(\eta_i) = 0,000323043$.

Pelo gráfico da distância de Cook mostrado na figura 12, há pontos de influência, o maior deles é o ponto #2194, entretanto, ele não se destaca pelo peso ou pelas características das variáveis independentes.

Figura 10: Gráficos dos pontos de influência para Gama com ligação canônica.

Ao testarmos a aderência do modelo pelo teste qui-quadrado, com grau de liberdade 4933 e $X^2=16196368$, obtivemos o p-valor $< 2,2e-16$. Portanto, o modelo é adequado.

A tabela 15 mostra a média e a variância dos dados ajustados do modelo gama com link de ligação canônica, e assim como obtido para os modelos da normal inversa, também obtivemos uma grande variação.

Tabela 15: Medidas para os valores ajustados do modelo.

Mínimo	Máximo	Mediana	Média	Variância	Desvio Padrão
300	3163	3217	3163	104.317,9	322,9828

Fonte: elaboração própria com base nos dados do SINASC de 2016 a 2021 no estado de São Paulo.

4.3.2 Gama com ligação logarítmica

Para o segundo modelo com distribuição gama, a ligação escolhida foi a logarítmica, portanto,

$$\log \mu_i = \eta_i$$

com $\eta_i = X_i^T \beta$, $X_i = (x_{i1}, \dots, x_{ip})^T$ e $\beta = (\beta_1, \dots, \beta_p)^T$.

Tabela 16: Estimativas dos parâmetros das variáveis com significância estatística do modelo gama com ligação log.

Variável	Estimativa	Erro Padrão	t valor	P-Valor
Intercept	5,6789	0,1513	37,526	2e-16
Semanas de gestação - 22s a 27s (x_1)	1,1819	0,1526	7,746	1,15e-14
Semanas de gestação - 28s a 31s (x_2)	1,5843	0,1517	10,443	2e-16
Semanas de gestação - 32s a 36s (x_3)	2,1357	0,1501	14,227	2e-16
Semanas de gestação - 37s a 41s (x_4)	2,3556	0,1499	15,709	2e-16
Semanas de gestação - +42s (x_5)	2,3824	0,1511	15,771	2e-16
Tipo de gravidez - dupla (x_6)	-0,1797	0,0147	-12,239	2e-16
Tipo de gravidez - tripla ou mais (x_7)	-0,3773	0,0868	-4,344	1,43e-05
Parto cesáreo (x_8)	0,0167	0,0044	3,842	0,0001
Sexo do bebê: feminino (x_9)	-0,0411	0,0043	-9,608	2e-16
Ausência de anomalia congenita (x_{10})	0,0659	0,0204	3,238	0,0012

Fonte: elaboração própria com base nos dados do SINASC de 2016 a 2021 no estado de São Paulo.

Portanto, o modelo estimado é dado por:

$$Y_i = \exp(5,6789 + 1,1819x_1 + 1,5843x_2 + 2,1357x_3 + 2,3556x_4 + 2,3824x_5 - 0,1797x_6 - 0,3773x_7 + 0,0167x_8 - 0,0411x_9 + 0,0659x_{10} + \varepsilon_{ij}).$$

O parâmetro de dispersão do modelo é dado por $\hat{\phi} = 41,7335$ e o $\hat{\mu}_i = g^{-1}(\eta_i) = 8,0513$.

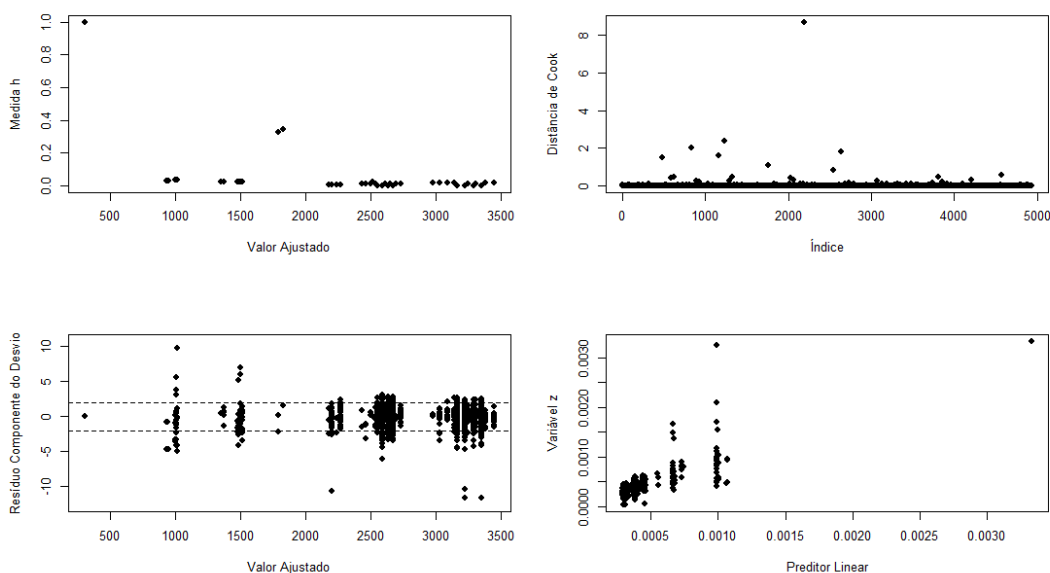
Segundo o modelo e suas estimativas, quanto mais perto do tempo normal de gravidez um bebê nascer, maior será o seu peso. Entretanto, bebês que nascem por parto cesáreo tem uma maior chance de ter um peso mais alto.

Outro ponto comum em todos os modelos gama e normal inversa é que fetos do sexo feminino apresentam um maior peso ao nascer se comparado com fetos do sexo masculino. A ausência de anomalias congênitas indicam um bebê com mais chances de nascer com peso saudável.

Entretanto, bebês provenientes de gravidez duplas ou triplas apresentam um menor peso ao nascer. Na tabela 16 temos os coeficientes do valor t menores para gravidez dupla do que para tripla, não necessariamente um bebê de gravidez tripla terá um peso maior do que o da gravidez dupla, há poucos dados desse tipo de gravidez para uma comparação significativa entre eles.

Na figura 13, podemos observar - pelo gráfico da distância de Cook - que há pontos influentes no modelo.

Figura 13: Gráficos dos pontos de influência para Gama com ligação canônica.



O ponto com maior influência é o ponto #4864 é um bebê nascido do sexo feminino e com apenas 300 gramas.

Ao testarmos a adequação do modelo pelo teste de qui-quadrado com grau de liberdade 4933 e $\chi^2=162277$, obtivemos o p-valor $< 2,2e-16$. Portanto, o modelo é adequado.

4.4 COMPARAÇÃO DOS MODELOS

Em modelagem estatística, o ideal é que o modelo escolhido consiga explicar o comportamento da variável de interesse, nesse caso o peso do bebê ao nascer, com o mínimo possível de parâmetros a serem estimados. Na literatura, diversos critérios são utilizados para isso, entretanto, o critério escolhido para este trabalho é o critério de informação de Akaike (AIC).

O AIC entende que existe um modelo real que descreve os dados desconhecidos e, a partir disso, tenta escolher um modelo que minimize a divergência de Kullback-Leibler (K-L). Essa divergência está relacionada à informação que perde-se por não utilizar um modelo real e sim um modelo estimado. Portanto, essa estimativa é dada por

$$AIC = -2L - 2K,$$

em que L é o máximo do logaritmo natural da função de verossimilhança (MLFV) com os parâmetros θ e K sendo o número de parâmetros. Portanto, o modelo que apresenta o menor AIC é considerado o modelo com o melhor ajuste.

Para os modelos ajustados com a distribuição normal inversa e a distribuição gama, temos o seguinte valor para AIC descritos na tabela 17.

Tabela 17: Comparação do modelo Normal Inversa e modelo Gama com ligações log e inversa.

Modelo	AIC	Deviance
Normal Inversa Canônica	76,786	0,0553
Normal Inversa Log	76,739	0,0548
Gama Canônica	75.001	119,05
Gama Log	74.987	118,70

Fonte: elaboração própria com base nos dados do SINASC de 2016 a 2021 no estado de São Paulo.

Com base nos valores de AIC, o modelo com melhor adequação é o modelo com distribuição gama e link de ligação log.

4.4 MODELO DE REGRESSÃO LOGÍSTICA

O primeiro passo para implementação do modelo foi a criação de uma nova variável binária, onde:

- 1 para peso ≤ 2500 gramas;
- 0 para peso > 2500 gramas.

O modelo é dado por:

$$\log\left\{\frac{\pi(X)}{1-\pi(X)}\right\} = \alpha + \beta_1 x_1 + \dots + \beta_k x_k,$$

com $X = (1, x_2, \dots, x_k)^T$ contendo os valores observados das variáveis explicativas. E com o apoio do pacote *glm* do software RStudio, obtivemos o seguinte modelo:

Tabela 18: Estimativas dos parâmetros do modelo de regressão logística.

Variável	Estimativa	Erro Padrão	t valor	P-Valor
Intercept	99,1187	886,5378	0,112	0,9109
Ano	-0,0414	0,0406	-1,021	0,3073
Idade da mãe	0,0101	0,0107	0,945	0,3449
Quantidade de filhos mortos	-0,3455	0,4715	-0,733	0,4637
Quantidade de filhos vivos	-0,6427	0,4613	-1,393	0,1635
Semanas de gestação - 22s a 27s	-11,6337	882,7440	-0,013	0,9895
Semanas de gestação - 28s a 31s	-12,2018	882,7436	-0,014	0,9889
Semanas de gestação - 32s a 36s	-14,7848	882,7434	-0,017	0,9866
Semanas de gestação - 37s a 41s	-17,7090	882,7434	-0,020	0,9839

Semanas de gestação - +42s	-17,7397	882,7437	-0,020	0,9839
Tipo de gravidez - dupla	2,2038	0,2736	8,052	8,13e-16
Tipo de gravidez - tripla ou mais	15,5865	508,1633	0,031	0,9755
Parto cesáreo	0,0769	0,1423	0,541	0,5886
Nº de consultas pré-natal - 1 a 3	0,3917	0,6124	0,640	0,5224
Nº de consultas pré-natal - 4 a 6	-0,1263	0,5710	-0,221	0,8249
Nº de consultas pré-natal - mais de 7	-0,9557	0,5649	-1,692	0,0907
Sexo do bebê: feminino	0,6620	0,1369	4,834	1,34e-06
Apgar 1	-0,2254	0,0674	-3,346	0,0008
Apgar 5	0,2417	0,0705	3,426	0,0006
Ausência de anomalia congênita	-1,5162	0,4011	-3,780	0,0002
Quantidade de gestação	0,3887	0,4606	0,844	0,3986

Fonte: elaboração própria com base nos dados do SINASC de 2016 a 2021 no estado de São Paulo.

Entretanto, nem todas as variáveis têm significância estatística, sendo assim, um novo modelo foi gerado. E, as variáveis apgar 1 e apgar 5, apesar de apresentarem significância estatística, não é possível utilizá-las para predizer o peso do bebê pois são métricas tomadas unicamente após o nascimento.

Tabela 19: Estimativas dos parâmetros das variáveis com significância estatística do modelo de regressão logística.

Variável	Estimativa	Erro Padrão	t valor	P-Valor
Intercept	0,133	0,524	0,254	0,7997
Tipo de gravidez - dupla (x_1)	3,228	0,208	15,500	2e-16

Tipo de gravidez - tripla ou mais (x_2)	16,594	308,744	0,054	0,9571
Nº de consultas pré-natal - 1 a 3 (x_3)	0,166	0,451	0,369	0,712
Nº de consultas pré-natal - 4 a 6 (x_4)	-0,218	0,423	-0,517	0,605
Nº de consultas pré-natal - mais de 7 (x_5)	-1,599	0,417	-3,836	0,000125
Sexo do bebê: feminino (x_6)	0,265	0,1109	2,396	0,0165
Ausência de anomalia congênita (x_7)	-1,733	0,333	-5,197	2,03e-07

Fonte: elaboração própria com base nos dados do SINASC de 2016 a 2021 no estado de São Paulo.

Portanto, obtemos o seguinte modelo:

$$\log\left\{\frac{\pi(X)}{1-\pi(X)}\right\} = 0,133 + 3,228x_1 + 16,594x_2 + 0,166x_3 - \\ 0,218x_4 - 1,599x_5 + 0,265x_6 - 1,733x_7.$$

Como a variável dicotômica utiliza a presença do baixo peso como sucesso, ou seja, igual a 1; a interpretação do modelo dá-se por: quanto mais perto de um a estimativa de uma variável estiver, maior a chance do bebê nascer com baixo peso.

Bebês nascidos de gravidez dupla apresentam uma maior chance de nascer abaixo do peso, entretanto, não há dados suficientes de gravidez triplas ou mais para ter significância estatística.

Assim como os modelos normal inversa e gama, o modelo de regressão logística aponta que crianças nascidas do sexo feminino apresentam uma maior chance de nascer com baixo peso, ou seja, com peso igual ou inferior a 2500 gramas.

As variáveis apgar 1, apgar 5 e anomalia congênita apresentam significância estatística. No caso da ausência de anomalia congênita tem-se que bebês que nascem com anomalia congênita apresentam menor peso corporal.

Utilizando a função `stargazer`, podemos obter os valores organizados na tabela 20.

Tabela 20: Estimativas dos parâmetros do segundo modelo de regressão logística.

Variável	Variável dependente: baixo peso
Tipo de gravidez - dupla (x_1)	3,228 (0,208)
Tipo de gravidez - tripla ou mais (x_2)	16,594 (308,744)
Nº de consultas pré-natal - 1 a 3 (x_3)	0,166 (0,451)
Nº de consultas pré-natal - 4 a 6 (x_4)	-0,219 (0,423)
Nº de consultas pré-natal - mais de 7 (x_5)	-1,599 (0,417)
Sexo do bebê: feminino (x_6)	0,266 (0,111)
Ausência de anomalia congênita (x_7)	-1,733 (0,333)
Constante	0,133 (0,524)
Observações	4927
Log da probabilidade	-1218,501
Critério de informação de Akaike (AIC)	2453,002

Fonte: elaboração própria com base nos dados do SINASC de 2016 a 2021 no estado de São Paulo.

Na tabela 21, temos os valores da razão de chances (odds ratio - OR) transformada, ou seja, tiramos da forma logarítmica por meio da função `logitor` que estima um modelo de regressão logística binária e calcula as razões de probabilidade correspondentes.

Tabela 21: Estimativas da razão de chances para o modelo de regressão logística.

Variável	Razão de Chances	Erro Padrão	z	P-Valor
----------	------------------	-------------	---	---------

Tipo de gravidez - dupla (x_1)	2,5233e+01	5,2551e+00	15,500	2,2e-16
Tipo de gravidez - tripla ou mais (x_2)	1,6093e+07	4,9686e+09	0,054	0,957
Nº de consultas pré-natal - 1 a 3 (x_3)	1,1810e+00	5,32114e-01	0,369	0,712
Nº de consultas pré-natal - 4 a 6 (x_4)	8,0352e-01	3,3975e-01	-0,517	0,605
Nº de consultas pré-natal - mais de 7 (x_5)	2,0201e-01	8,4239e-02	-3,836	0,00012
Sexo do bebê: feminino (x_6)	1,3042e+00	1,4458e-01	2,396	0,0165
Ausência de anomalia congênita (x_6)	1,7682e-01	5,8953e-02	-5,197	2,028e-07

Fonte: elaboração própria com base nos dados do SINASC de 2016 a 2021 no estado de São Paulo.

De maneira prática, a chance de $y = 1$ é 2,5233e+01 maior quando x_1 aumenta em uma unidade - levando-se em conta que as outras variáveis se mantêm constantes. Portanto, para cada variação unitária em x_1 o log das chances varia em 3,228, como visto na tabela 20.

Portanto, temos que a razão de chances de um bebê do sexo feminino nascer abaixo do peso é de $\exp(1,3042e+00) = 3,685$ maior do que de um bebê do sexo masculino.

5 CONSIDERAÇÕES FINAIS

Com base nas análises feita através dos modelos para estimar o peso do bebê ao nascer, tem-se que as principais características que podem ser utilizadas para estudar o comportamento do peso ao nascer são: semanas de gestação, sexo do bebê, tipo de gravidez, tipo de parto e a presença ou ausência de anomalia congênita.

Em contrapartida, as variáveis com significância estatística para estimar a probabilidade do bebê nascer com baixo peso são: tipo de gravidez, quantidade de consultas pré-natal, sexo do bebê e anomalia congênita. E, apesar de apgar 1 e apgar 5 darem significância estatística, não é factível utilizá-las para predizer o peso do bebê porque são medidas tomadas unicamente após o nascimento.

Em vista disso, pode-se afirmar que quanto maior o tempo de gestação maior a probabilidade do bebê nascer com um peso considerado adequado, já que bebês que nascem prematuros, em geral, apresentam um peso menor do que bebês que nascem com mais de 36 semanas.

Da mesma forma, bebês do sexo feminino tendem a ter uma maior probabilidade de apresentar um menor peso ao compararmos com bebês do sexo masculino. Segundo a razão de chances do modelo de regressão logística, a chance de que um bebê do sexo feminino nasça com baixo peso é 3,685 maior do que um bebê do sexo masculino.

Segundo o modelo com distribuição normal inversa e link de ligação logarítmica, bebês nascidos de gravidez duplas ou triplas têm maior probabilidade de apresentar baixo peso do que bebês de gravidez única. Entretanto, é válido ressaltar que esse tipo de gravidez requer maiores cuidados da gestante e um acompanhamento médico cuidadoso, já que a incidência de prematuridade é fator relevante para gestações múltiplas.

A ausência de anomalia congênita - são definidas como um grupo de alterações estruturais ou funcionais que ocorrem durante a vida intrauterina e que podem ser detectadas antes, durante ou após o nascimento (OMS, 2023) - apresenta significância estatística para o peso do bebê, a razão de chances de um bebê com ausência de anomalia congênita não ter um baixo peso é 1,2623 maior do que um bebê que tenha algum tipo de anomalia congênita. Ademais, o parto cesáreo apresenta relevância para a estimação do peso do bebê.

Entre as características utilizadas para estimar o peso e o baixo peso ao nascer, apenas o tempo de gestação, a quantidade de consultas pré-natal e o tipo de parto é um fator controlável. Entretanto, nem sempre o tipo de parto cesáreo pode ser evitado, principalmente quando há a necessidade de partos de emergência, que geralmente culminam em partos cesáreos ou quando o bebê tem um peso maior do que o esperado e o parto natural não é recomendado.

O número de consultas de pré-natal nem sempre é controlável, já que no Brasil não existe obrigatoriedade e há mães que descobrem a gravidez com um tempo já avançado, diminuindo assim a quantidade de consultas de acompanhamento da gravidez.

Entretanto, as consultas de pré-natais são altamente recomendadas por especialistas pois, por meio desse acompanhamento, é possível detectar anomalias congênitas, possíveis patologias maternas e do feto, prevenir partos prematuros e o baixo peso do bebê. Há também o alto custo envolvido no cuidado neonatal com bebês que nascem com baixo peso e/ou prematuros, já que precisam ficar na instituição médica por mais tempo, até que alcancem um peso ideal e tenham alta.

Segundo o critério de informação de Akaike (AIC), o modelo com melhor adequação para estimar o peso do bebê ao nascer é o modelo com distribuição gama e link de ligação log. Entretanto, o modelo ainda precisa de ajustes e não está completo, pois para uma estimação mais precisa e com menor variação é altamente indicado a inserção de dados sobre os pais da criança, já que fatores genéticos podem influenciar na variável de interesse peso do bebê ao nascer.

REFERÊNCIA

Bussab, Wilton de O.; MORETTIN, Pedro A. Estatística Básica. 8a. ed. São Paulo: Saraiva, 2013.

McCullagh, P., Nelder, J.A. Generalized Linear Models. 2a. ed. EUA. 1989.

Turkman, M.A.A e Silva, G.L. Modelos Lineares Generalizados - da teoria à prática (2000). Disponível em: https://www.researchgate.net/profile/Antonia-Turkman/publication/267948318_Modelos_Lineares_Generalizados_-da_teor%C3%ADa_a_pr%C3%A1tica/links/5f2d09cb92851cd302e50cf2/Modelos-Lineares-Generalizados-da-teoria-a-pratica.pdf. Acesso em 15 mar. 2023.

Cordeiro e Demétrio (2008). Modelos Lineares Generalizados e Extensões.

Disponível em: [https://www.ufjf.br/clecio_ferreira/files/2013/05/](https://www.ufjf.br/clecio_ferreira/files/2013/05/Livro-Gauss-e-Clarice.pdf)

Livro-Gauss-e-Clarice.pdf. Acesso em 15 mar. 2023.

Demétrio (2001). Modelos Lineares Generalizados em Experimentação Agrônômica.

Disponível em: [https://www.researchgate.net/profile/Clarice-Demetrio/](https://www.researchgate.net/profile/Clarice-Demetrio/publication/266053942_Modelos_Lineares_Generalizados_em_Experimentacao_Agronomica/links/56cda8a408ae85c8233e6014/Modelos-Lineares-Generalizados-em-Experimentacao-Agronomica.pdf)

publication/266053942_Modelos_Lineares_Generalizados_em_Experimentacao_Agronomica/links/56cda8a408ae85c8233e6014/Modelos-Lineares-Generalizados-em-Experimentacao-Agronomica.pdf. Acesso em 15/03.2023.

Paula, G. (2013). Modelo de Regressão com Apoio Computacional. Disponível em:

https://www.ime.usp.br/~giapaula/texto_2013.pdf. Acesso em 15 mar. 2023.

Schardosim, Juliana Machado; Rodrigues, Nayara Lauane de Araújo; Rattner, Daphne. Parâmetros utilizados na avaliação do bem estar do bebê no nascimento. 2017. Disponível em:

<http://www.scielo.org.co/pdf/aven/v36n2/0121-4500-aven-36-02-197.pdf>. Acesso em 15 jul. 2023.

De Almeida, André Henrique do Vale; Da Gama, Silvana Granado Nogueira; et al. Prematuridade e gravidez na adolescência no Brasil, 2011-2012. 2020. Disponível em: <https://doi.org/10.1590/0102-311X00145919>. Acesso em 20 jul. 2023.

Silveira, Leonardo Ito; Andrade, Fernanda; Dea, Bruna Di; et al. Fatores associados ao número de consultas no pré-natal: Análise segundo a autopercepção de usuárias da atenção primária no Brasil. 2020. Disponível em: <https://revista.acm.org.br/index.php/arquivos/article/view/565>. Acesso em: 05 ago. 2023.

Carniel, Emília de Faria; Zanolli, Maria de Lurdes; Antônio, Maria Ângela Reis de Góes Monteiro; Morcillo, André Moreno. Determinantes do baixo peso ao nascer a partir das Declarações de Nascidos Vivos. 2002. Disponível em: https://www.scielo.org/article/ssm/content/raw/?resource_ssm_path=/media/assets/rbepid/v11n1/16.pdf. Acesso em: 31 jul. 2023.

Benicio, Maria Helena D'Aquino; Monteiro, Carlos Augusto; et al. Análise multivariada de fatores de risco para o baixo peso ao nascer em nascidos vivos do município de São Paulo, SP (Brasil). 1985. Disponível em: <https://doi.org/10.1590/S0034-89101985000400004>. Acesso em: 05 ago. 2023.

Costa, Cristina Elizabeth; Gotlieb, Sabina Léa Davidson. Estudo epidemiológico do peso ao nascer a partir da Declaração de Nascido Vivo. 1998. Disponível em: <https://www.scielo.org/pdf/rsp/v32n4/a2459.pdf>. Acesso em: 10 ago. 2023.

De Melo, Thamires Francelino Mendonça; Carregaro, Rodrigo Luiz; et al. Custos diretos da prematuridade e fatores associados ao nascimento e condições maternas. 2022. Disponível em: <https://doi.org/10.11606/s1518-8787.2022056003657>. Acesso em: 12 ago. 2023.

Organização Pan-Americana de Saúde. Um em cada sete bebês em todo o mundo nascem com baixo peso. 2019. Disponível em: <https://www.paho.org/pt/noticias/16-5-2019-um-em-cada-sete-bebes-em-todo-mundo-nascem-com-baixo-peso>. Acesso em: 12 set. 2023.

Bozdongan. H. Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. *Psychometrika*. v.52, n.3, 345-370, setembro, 1987.

Ahumada-Barros, Margarita E.; Alvarado, German F. Fatores de Risco para parto prematuro em um hospital. 2016. Disponível em: <https://www.scielo.br/j/rlae/a/tZHBBYdTZjKNFyhvt75rmBm/?format=pdf&lang=pt>. Acesso em: 10 nov. 2023.

Marchetti, J. R.; Corrêa Oliveira, V.; Fernandes, B. C.; Reginatti Nicolao, G. A importância do pré-natal. *Anuário Pesquisa e Extensão Unoesc Xanxerê*, [S. l.], v. 5, p. e24175, 2020. Disponível em: <https://periodicos.unoesc.edu.br/apeux/article/view/24175>. Acesso em: 20 nov. 2023.

World Health Organization. Congenital disorders, Disponível em: <https://www.who.int/news-room/fact-sheets/detail/birth-defects>. Acesso em: 25 nov. 2023.