

Thiago Nascimento da Silva

Modelagem de Dados de Contagem com
Excesso de Zeros: Uma aplicação com
Sintomas Respiratórios

Niterói - RJ, Brasil
07 de julho de 2015

Thiago Nascimento da Silva

Modelagem de Dados de Contagem
com Excesso de Zeros: Uma
Aplicação com Sintomas
Respiratórios

Trabalho de Conclusão de Curso

Monografia apresentada para a obtenção do grau de Bacharel em
Estatística pela Universidade Federal Fluminense.

Orientadora: Profa. Ludmilla da Silva Viana Jacobson

Niterói - RJ, Brasil
07 de julho de 2015

Thiago Nascimento da Silva

Modelagem de Dados de Contagem com
Excesso de Zeros: Uma Aplicação com
Sintomas Respiratórios

Monografia de Projeto Final de Graduação sob o título *“Modelagem de Dados de Contagem com Excesso de Zeros: Uma Aplicação com Sintomas Respiratórios”*, defendida por Thiago Nascimento da Silva e aprovada em 07 de Julho de 2015, na cidade de Niterói, no Estado do Rio de Janeiro, pela banca examinadora constituída pelos professores:

Profa. Dra. Ludmilla da Silva Viana Jacobson
Orientadora
Departamento de Estatística – UFF

Prof. Dr. Jony Arrais Pinto Junior
Co-orientador
Departamento de Estatística – UFF

Profa. Dra. Jessica Quintanilha Kubrusly
Departamento de Estatística – UFF

Niterói, 07 de Julho de 2015

Silva, Thiago Nascimento da
Modelagem de dados de contagem com excesso de zeros: uma
aplicação com sintomas respiratórios / Thiago Nascimento da
Silva; Ludmilla da Silva Viana Jacobson, orientadora.
Niterói, 2015.
59 f. : il.

Trabalho de Conclusão de Curso (Graduação em
Estatística) - Universidade Federal Fluminense,
Instituto de Matemática e Estatística, Niterói, 2015.

1. Sintomas respiratórios. 2. Dados de contagem. 3. Modelo
de regressão. 4. Dados inflacionados de zeros. I. Jacobson,
Ludmilla da Silva Viana, orientadora. II. Universidade
Federal Fluminense. Instituto de Matemática e Estatística.
III. Título.

CDD -

Resumo

Nos dias atuais, a poluição atmosférica se apresenta como um fator determinante para a saúde das pessoas. O aumento da emissão desses poluentes afeta diretamente a saúde respiratória da população. Por isso faz-se cada vez mais necessário o estudo dos efeitos adversos desses poluentes no sistema respiratório. O objetivo desse trabalho é avaliar o excesso de zeros na variável de Sintomas Respiratórios, mais especificamente a Tosse e a Coriza, além de verificar o efeito defasado do poluente atmosférico PM_{10} na ocorrência dos sintomas. Para essa avaliação foi utilizado um conjunto de dados de um estudo de painel realizado na cidade de Tangará da Serra, Mato Grosso com escolares, no período de agosto a dezembro de 2008, considerado período de seca na região. As variáveis explicativas utilizadas para a estimação dos modelos se referiam às características individuais (sexo, idade, asma), meteorológicas (temperatura e umidade, ambas com defasagem de 1 dia), e de poluição (PM_{10} com defasagem de 1, 2 e 3 dias). Foi proposto a utilização de quatro modelos distintos para análise da variável de contagem Sintoma Respiratório, os modelos de regressão de Poisson, de regressão Binomial Negativa, de regressão de Poisson Inflacionado de Zeros e de regressão Binomial Negativa Inflacionado de Zeros. Para a escolha do modelo mais adequado, foi utilizado o Critério de Informação de Akaike (AIC), além dos resíduos da Deviance. Para o sintoma Tosse, que apresentou uma proporção de zeros no período de 11%, o modelo mais adequado foi o modelo de regressão Binomial Negativa. Além disso, o poluente atmosférico PM_{10} não apresentou efeito significativo na ocorrência da Tosse, para um nível de significância de 5%. Para o sintoma Coriza, que apresentou uma proporção de zeros de 22%, o modelo mais adequado foi o modelo de regressão Binomial Negativa Inflacionado de Zeros. Neste caso, para um nível de significância de 5%, o poluente atmosférico PM_{10} apresentou efeito significativo, porém com uma contribuição negativa, o que faz dela um efeito protetor para a ocorrência do sintoma Coriza.

Palavras-chaves: Sintomas Respiratórios, Dados de Contagem, Modelo de Regressão, Dados Inflacionados de Zeros.

*“Cada sonho que você deixa
para trás, é um pedaço do seu
futuro que deixa de existir.”*

(Steve Jobs)

Agradecimentos

Como diz meu avô: “a gratidão é a demonstração mais sincera de humildade, pois nos mostra que sozinho não somos ninguém”. Por isso gostaria de externar meu agradecimento a todas as pessoas que, de alguma forma, contribuíram para que eu chegasse onde estou.

Agradeço antes de tudo à minha família, principalmente a minha mãe Adriana, meu pai Paulo e meus irmãos Ana Paula e Rafael, que estão do comigo em todos os momentos da minha vida, sejam eles felizes ou tristes, me dando força e me incentivando a seguir em frente.

À meus avós, Joel e Amália, que sempre foram a minha base e meu porto seguro. Agradeço pela criação que me deram, pelo exemplo que vocês são em minha vida. Que meu filho possa aprender com vocês o tanto que aprendi.

À minha esposa Camile, pela paciência, ajuda, companheirismo, carinho. Agradeço a Deus por ter posto você ao meu lado. Sei que a vida é cheia de surpresas, desafios, perdas e decepções. Muitas decepções. Mas eu sei que juntos passaremos por tudo isso. Obrigado por ter me dado essa benção que é o nosso filho. Eu amo você!

Agradeço meu filho, Matheus, por me mostrar que a vida é muito mais do que eu imaginava. Por fazer de mim uma pessoa mais feliz, mais realizada. Você me mostrou o que é realmente ter uma responsabilidade. Mas uma responsabilidade que eu tenho o prazer de carregar. Que eu possa ser pra você não somente um pai, mas um amigo, um espelho, um exemplo. Você é tudo na minha vida.

Agradeço também aos amigos que sempre estiveram ao meu lado. Todas as demonstrações de carinho e incentivos que me deram não foram esquecidas. Obrigado por tudo.

Gostaria de expressar meu carinho e admiração a todos que foram importantes e participaram da minha vida durante esse período na faculdade.

Agradeço a minha orientadora, Ludmilla, por não ter desistido de mim e da monografia, quando até eu mesmo já estava para desistir. Obrigado pelo apoio, por todos os conselhos e ajudas na conclusão deste trabalho. Agradeço também por compreender esse atual momento na minha vida. Ninguém melhor do que a gente para

saber que ser pai/mãe de primeira viagem é desgastante, mesmo que maravilhoso. Obrigado por tudo. Com certeza acertei na minha escolha.

Agradeço aos professores do Departamento de Estatística que foram importantes para a minha formação. Jony, Jessica, Núbia, José Rodrigo, Ana Maria, Valentin, Adrian, Guillermo, Mariana e muitos outros que sempre estiveram dispostos a me ajudar e me aconselhar.

À todos os amigos e colegas que fiz durante o curso. Obrigado por todos os momentos que tivemos. Os ENESTEs, SINAPEs, estudos, chopadas. Tenham certeza de que serão lembrados para sempre. Obrigado por ter feito da faculdade um dos melhores períodos da minha vida.

Em especial, agradeço aos amigos/irmãos que a faculdade me proporcionou: Bruno Cidade, Bruno Lucian, Carolina Valani, Evandro Dalbem, Fábio Mascarenhas, João Vitor Bastos, Juliana Freitas, Kiese Quiavauca, Leonor Lobato, Marcela Martins, Pablo Casais e Roger Barros. Obrigado pelo companheirismo, pela força, pela ajuda e principalmente por estarem ao meu lado. Essa conquista tem um pouco de cada um de vocês. Vamos comemorar!

Também agradeço aos amigos de trabalho da Prefeitura Municipal de Rio Bonito. Obrigado pelos ensinamentos e por me cobrir sempre que ficava preso na faculdade. E não foram poucas às vezes. Minha chefe/amiga Sônia Maria Pinheiro, obrigado por tudo. Essa conquista também é sua.

E por último, mas não menos importante, quero agradecer a Deus, pelas bênçãos derramadas em minha vida. Agradeço por me dar discernimento, sabedoria, paciência e força para continuar. Por todas as pessoas que colocou no meu caminho e que de alguma forma contribuíram para que chegasse até aqui.

Sumário

Lista de Figuras

Lista de Tabelas

1. Introdução	14
1.1. Justificativa	15
1.2. Revisão de Literatura sobre Dados de Contagem com Excesso de Zeros	17
1.3. Revisão de Literatura sobre Poluição Atmosférica e Sintomas Respiratórios	19
2. Objetivos	23
2.1. Objetivos Específicos	23
3. Material e Métodos	24
3.1. Sintoma Respiratório na Amazônia	24
3.2. Modelos Lineares Generalizados	26
3.2.1. Definição	26
3.2.2. Família Exponencial	28
3.2.2.1. Distribuição de Poisson	29
3.2.2.2. Distribuição Binomial Negativa	30
3.2.3. Modelos de Regressão para Dados de Contagem	31
3.2.3.1. Modelo de Regressão de Poisson	31
3.2.3.2. Modelo de Regressão Binomial Negativa	33
3.2.3.3. Modelo de Regressão ZIP	34
3.2.3.4. Modelo de Regressão ZINB	35
4. Análise de Resultados	36
4.1. Análise Exploratória	36
4.2. Modelagem dos Dados	37

5. Conclusão	42
Referências	44
Anexo A – Modelagem dos dados	47
Anexo B – Gráfico dos Resíduos	55

Lista de Figuras

Figura 1:	Frequência da ocorrência do sintoma Tosse.	16
Figura 2:	Frequência da ocorrência do sintoma Coriza.	16
Figura 3:	Gráfico dos resíduos da Deviance do Modelo de Regressão Binomial Negativa com variável de poluição PM10 lag3 para o sintoma Tosse.	38
Figura 4:	Gráfico dos resíduos da Deviance do Modelo de Regressão ZINB com variável de poluição PM10 lag3 para o sintoma Coriza	40
Figura 5:	Gráfico dos resíduos da Deviance do Modelo de Regressão Binomial Negativa com variável de poluição PM10 lag1 para o sintoma Tosse.	55
Figura 6:	Gráfico dos resíduos da Deviance do Modelo de Regressão Binomial Negativa com variável de poluição PM10 lag2 para o sintoma Tosse.	55
Figura 7:	Gráfico dos resíduos da Deviance do Modelo de Regressão de Poisson com variável de poluição PM10 lag1 para o sintoma Tosse.	55
Figura 8:	Gráfico dos resíduos da Deviance do Modelo de Regressão de Poisson com variável de poluição PM10 lag2 para o sintoma Tosse	55
Figura 9:	Gráfico dos resíduos da Deviance do Modelo de Regressão de Poisson com variável de poluição PM10 lag3 para o sintoma Tosse.	56
Figura 10:	Gráfico dos resíduos da Deviance do Modelo de Regressão ZINB com variável de poluição PM10 lag1 para o sintoma Tosse.	56
Figura 11:	Gráfico dos resíduos da Deviance do Modelo de Regressão ZINB com variável de poluição PM10 lag2 para o sintoma Tosse.	56
Figura 12:	Gráfico dos resíduos da Deviance do Modelo de Regressão ZINB com variável de poluição PM10 lag3 para o sintoma Tosse.	56
Figura 13:	Gráfico dos resíduos da Deviance do Modelo de Regressão ZIP com variável de poluição PM10 lag1 para o sintoma Tosse.	56
Figura 14:	Gráfico dos resíduos da Deviance do Modelo de Regressão ZIP com variável de poluição PM10 lag2 para o sintoma Tosse.	56
Figura 15:	Gráfico dos resíduos da Deviance do Modelo de Regressão ZIP com variável de poluição PM10 lag3 para o sintoma Tosse.	57
Figura 16:	Gráfico dos resíduos da Deviance do Modelo de Regressão Binomial Negativa com variável de poluição PM10 lag1 para o sintoma Coriza.	57
Figura 17:	Gráfico dos resíduos da Deviance do Modelo de Regressão Binomial Negativa com variável de poluição PM10 lag2 para o sintoma Coriza.	57
Figura 18:	Gráfico dos resíduos da Deviance do Modelo de Regressão Binomial	57

Negativa com variável de poluição PM10 lag3 para o sintoma Coriza.

Figura 19:	Gráfico dos resíduos da Deviance do Modelo de Regressão de Poisson com variável de poluição PM10 lag1 para o sintoma Coriza.	57
Figura 20:	Gráfico dos resíduos da Deviance do Modelo de Regressão de Poisson com variável de poluição PM10 lag2 para o sintoma Coriza	58
Figura 21:	Gráfico dos resíduos da Deviance do Modelo de Regressão de Poisson com variável de poluição PM10 lag3 para o sintoma Coriza.	58
Figura 22:	Gráfico dos resíduos da Deviance do Modelo de Regressão ZINB com variável de poluição PM10 lag1 para o sintoma Coriza.	58
Figura 23:	Gráfico dos resíduos da Deviance do Modelo de Regressão ZINB com variável de poluição PM10 lag2 para o sintoma Coriza.	58
Figura 24:	Gráfico dos resíduos da Deviance do Modelo de Regressão ZIP com variável de poluição PM10 lag1 para o sintoma Coriza.	58
Figura 25:	Gráfico dos resíduos da Deviance do Modelo de Regressão ZIP com variável de poluição PM10 lag2 para o sintoma Coriza.	58
Figura 26:	Gráfico dos resíduos da Deviance do Modelo de Regressão ZIP com variável de poluição PM10 lag3 para o sintoma Coriza.	59

Lista de Tabelas

Tabela 1:	Distribuição de frequência da amostra de criança e adolescente	23
Tabela 2:	Resumo da construção das variáveis	25
Tabela 3:	Análise descritiva das variáveis quantitativas agregadas para o sintoma Tosse	35
Tabela 4:	Análise descritiva das variáveis quantitativas agregadas para o sintoma Coriza	35
Tabela 5:	Análise descritiva das variáveis qualitativas	36
Tabela 6:	Análise descritiva da variável resposta Sintoma Respiratório	36
Tabela 7:	Valores do AIC referente ao sintoma Tosse	
Tabela 8:	Modelo de Regressão Binomial Negativa utilizando o PM10lag3 para a variável Tosse	38
Tabela 9:	Valores do AIC referente ao sintoma Coriza	39
Tabela 10:	Modelo de Regressão ZINB utilizando o PM10lag3 para a variável Coriza	40
Tabela 11:	Modelo de Regressão ZINB com interação entre o PM10lag3 variável Umidade do Ar	40
Tabela 12:	Modelo de Regressão Binomial Negativa utilizando o PM10lag1 para a variável Tosse	47
Tabela 13:	Modelo de Regressão Binomial Negativa utilizando o PM10lag2 para a variável Tosse	47
Tabela 14:	Modelo de Regressão de Poisson utilizando o PM10lag1 para a variável Tosse	47
Tabela 15:	Modelo de Regressão de Poisson utilizando o PM10lag2 para a variável Tosse	48
Tabela 16:	Modelo de Regressão de Poisson utilizando o PM10lag3 para a variável Tosse	48
Tabela 17:	Modelo de Regressão ZINB utilizando o PM10lag1 para a variável Tosse	48
Tabela 18:	Modelo de Regressão ZINB utilizando o PM10lag2 para a variável Tosse	49
Tabela 19:	Modelo de Regressão ZINB utilizando o PM10lag3 para a variável Tosse	49
Tabela 20:	Modelo de Regressão ZIP utilizando o PM10lag1 para a variável Tosse	49
Tabela 21:	Modelo de Regressão ZIP utilizando o PM10lag2 para a variável Tosse	50
Tabela 22:	Modelo de Regressão ZIP utilizando o PM10lag3 para a variável Tosse	50
Tabela 23:	Modelo de Regressão Binomial Negativa utilizando o PM10lag1 para a variável Coriza	50
Tabela 24:	Modelo de Regressão Binomial Negativa utilizando o PM10lag2 para a variável Coriza	51

Tabela 25:	Modelo de Regressão Binomial Negativa utilizando o PM10lag3 para a variável Coriza	51
Tabela 26:	Modelo de Regressão de Poisson utilizando o PM10lag1 para a variável Coriza	51
Tabela 27:	Modelo de Regressão de Poisson utilizando o PM10lag2 para a variável Coriza	52
Tabela 28:	Modelo de Regressão de Poisson utilizando o PM10lag3 para a variável Coriza	52
Tabela 29:	Modelo de Regressão ZINB utilizando o PM10lag1 para a variável Coriza	52
Tabela 30:	Modelo de Regressão ZINB utilizando o PM10lag2 para a variável Coriza	53
Tabela 31:	Modelo de Regressão ZIP utilizando o PM10lag1 para a variável Coriza	53
Tabela 32:	Modelo de Regressão ZIP utilizando o PM10lag2 para a variável Coriza	53
Tabela 33:	Modelo de Regressão ZIP utilizando o PM10lag3 para a variável Coriza	54

1 Introdução.

A poluição atmosférica é um fator determinante na saúde da população nos dias atuais, principalmente nas áreas urbanas. O crescimento exacerbado das cidades, ligada à industrialização e ao crescimento populacional, acarretou em um aumento considerável na emissão de poluentes, elevando a chance de se adquirir problemas respiratórios. Estima-se que cerca de 50% da população mundial viva em centros urbanos, onde a exposição a poluentes apresentam níveis exorbitantes, segundo Arbex (2012).

O CONAMA (2008) (Conselho Nacional de Meio Ambiente) define a poluição atmosférica como sendo “Qualquer forma de matéria ou energia com intensidade e em quantidade, concentração, tempo ou características em desacordo com os níveis estabelecidos e que torne ou possa tornar o ar: impróprio, nocivo, ou ofensivo à saúde; inconveniente ao bem estar público; danoso aos materiais, à fauna e flora; prejudicial à segurança, ao uso e gozo da propriedade e às atividades normais da comunidade”.

Estudos recentes da OMS (2008) mostram que a poluição atmosférica, tanto interna como externa, causa anualmente cerca de 3,15 milhões de óbitos no mundo todo, o que representa 2% do total de óbitos. Especificamente para o Brasil, estima-se que 30,7 mil pessoas morrem por ano em decorrência da poluição do ar.

Com o aumento do nível de emissão dos poluentes atmosféricos, cresce também o número de casos de sintomas respiratórios. Segundo Brook (2010), uma grande parte dos poluentes inalados afeta diretamente o funcionamento dos órgãos do sistema respiratório e, por esse motivo, muitos estudos vêm sendo realizados para medir a importância da qualidade do ar para a saúde respiratória da população.

Segundo Rodriguez-Villamizar (2008), o efeito Negativa da poluição atmosférica afeta todas as faixas etárias, porém crianças, idosos e indivíduos com doenças crônicas, tendem a ser mais vulneráveis aos efeitos dos poluentes. Estudos mostram que o baixo e/ou imaturo sistema imunológico respondem o fato deste grupo ser mais suscetível aos efeitos Negativas da poluição do ar, como explicado por Brook (2010) e Rodriguez-Villamizar (2008).

Segundo Arbex (2012), os estudos dos sintomas respiratórios causados pela poluição atmosférica são de suma importância, pois ajudam a comprovar os efeitos adversos dos poluentes na população, além de ajudar na criação de políticas públicas para auxiliar na prevenção dos sintomas e na diminuição dos poluentes.

No Brasil diversos estudos avaliam o efeito da poluição atmosférica na saúde. Seus efeitos são verificados tanto em áreas urbanas quanto em áreas rurais. No entanto, os estudos sobre a ocorrência de sintomas respiratórios foram pouco explorados na região da Amazônia, onde a poluição atmosférica é decorrente de queimadas no período de seca, durante os meses de Junho a Outubro, segundo Pereira (2011).

Os Capítulos seguintes estão organizados da seguinte ordem: O Capítulo 2 irá apresentar os objetivos do trabalho, tanto geral quanto específicos. O Capítulo 3 irá se aprofundar no banco de dados utilizado no trabalho, explicitando as variáveis que serão estudadas, bem como foram coletadas suas observações. Além disso, os modelos utilizados no trabalho também serão apresentados e explicados nesta seção. No Capítulo 4 serão apresentados os resultados obtidos através da aplicação dos modelos e da interpretação dos dados. A conclusão e a escolha do melhor modelo serão apresentadas no Capítulo 5.

1.1 Justificativa.

Os dados relacionados aos sintomas respiratórios são basicamente contagens originárias da frequência com que cada indivíduo da amostra relatou ter o sintoma, durante o período de estudo. Logo, o pensamento inicial supõe que a distribuição desses dados siga uma distribuição de Poisson. Porém, dados de contagem podem apresentar uma variabilidade maior do que o valor esperado, fenômeno denominado de superdispersão. Além disso, há a possibilidade dos dados de contagem apresentar uma massa de probabilidade considerável no ponto zero.

As Figuras (1) e (2) se referem às variáveis de sintomas respiratórios, mais especificamente o número de ocorrência de Tosse e o número de ocorrência de Coriza, respectivamente, de um estudo realizado no município de Tangará da Serra, MT. Mais detalhes deste estudo serão abordados no capítulo 3.

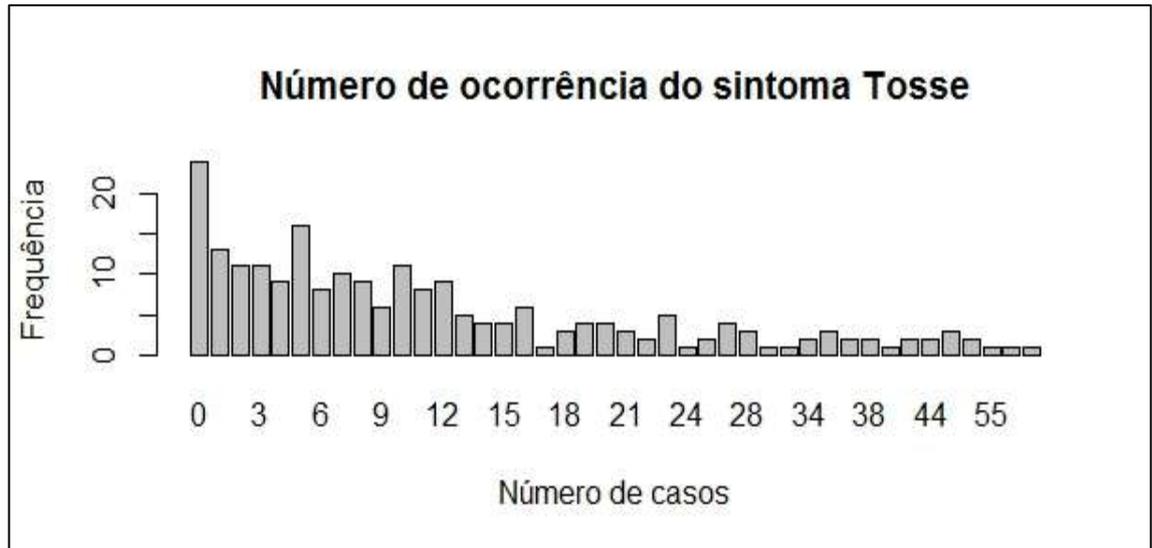


Figura1 – Frequência da ocorrência do sintoma Tosse, durante os 82 dias de estudo.

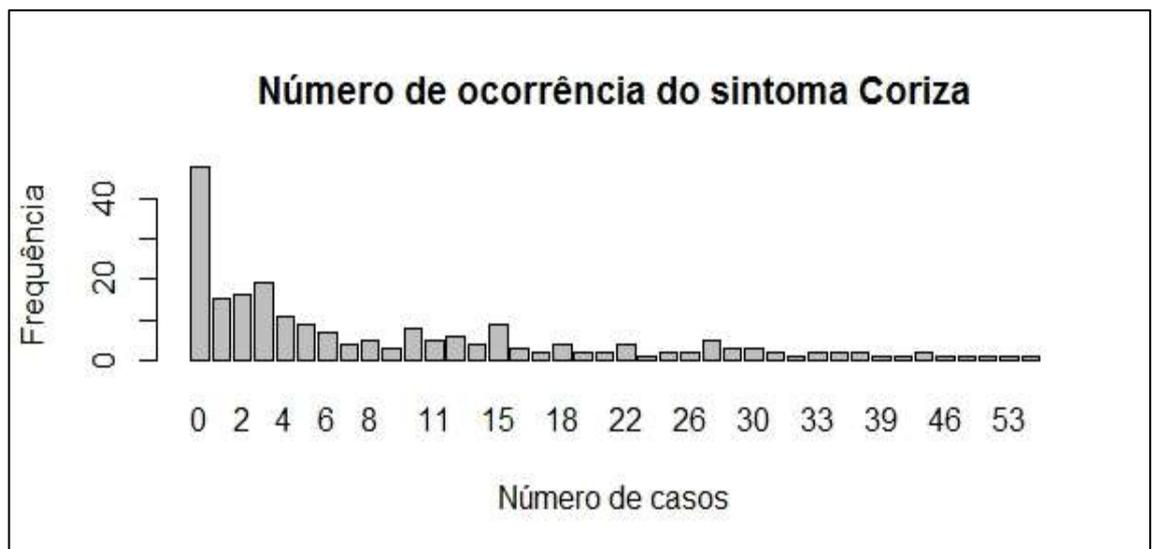


Figura 2 – Frequência da ocorrência do sintoma Coriza, durante os 82 dias de estudo.

Visto isso, Lambert et al., (1992) e Slymen et al., (2006) mostraram que nesses casos, modelos de regressão inflacionados de zeros são muito úteis para a análise dos dados.

Portanto, este trabalho tem como um dos objetivos comparar os Modelos de Regressão de Poisson, de regressão Binomial Negativa, de regressão de Poisson Inflacionado de Zeros (ZIP) e de regressão Binomial Negativa Inflacionado de Zeros

(ZINB) a fim de decidir qual que melhor se ajusta aos dados, para cada sintoma respiratório em estudo (tosse e coriza).

1.2 Revisão de Literatura sobre Dados de Contagem com Excesso de Zeros.

Famoye (2006) realizaram um estudo com o objetivo de observar qual modelo é o mais adequado para se aplicar em um banco de dados de violência doméstica. Foram testados os Modelos de Regressão de Poisson, Binomial Negativa, Poisson Generalizado, ZIP, ZINB e Poisson Generalizado Inflacionados de Zeros (ZIGP). Os dados foram coletados a partir de registros oficiais de agressores e de inquiridos sobre as vítimas entre os anos de 1996 e 1997. Para estimar os parâmetros da distribuição utilizada no estudo, foi realizado o método de máxima verossimilhança. A estatística de teste utilizada como critério de decisão de escolha do modelo foi o teste de Escore. Após o estudo, concluiu-se que para este referido banco de dados, o Modelo de Regressão ZIGP foi o que melhor se ajustou aos dados.

Fumes (2010) realizaram um estudo utilizando os modelos inflacionados de zeros para analisar um questionário de frequência alimentar de idosos do município de Avaré, no interior de São Paulo, e verificar qual modelo melhor se ajustava aos dados. Foram utilizados os Modelos de Regressão de Poisson, Binomial Negativa, ZIP e ZINB. Foi utilizado o Teste de Vuong, além do critério de informação de Akaike (AIC) e do critério de informação de Bayes (BIC) para a comparabilidade dos modelos. Como conclusão, observou-se que para variáveis que apresentavam até 10% na proporção de zeros, o Modelo de Regressão Binomial Negativa se mostrou adequado. Para porcentagens entre 10% e 50% na proporção de zeros das variáveis, os modelos inflacionados de zeros se mostraram mais adequados. Para proporções de zeros acima de 50%, os modelos propostos não são adequados.

Nagamine (2007) realizou um estudo para avaliar o número de ovos do mosquito *aedes aegypti* na cidade de São José do Rio Preto, em São Paulo. Foram utilizados os Modelos de Regressão de Poisson, Binomial Negativa, ZIP e ZINB. Os dados relativos ao estudo estão explicados de forma detalhada em Barbosa (2006). Foram utilizados nas análises os testes de razão de verossimilhança, teste de Wald e teste de Escore. Para melhorar a qualidade de ajuste do modelo foram utilizados o AIC, o BIC e a

Deviance. Após a realização dos testes, concluiu-se que o Modelo de Regressão ZINB foi o que melhor colaborou na explicação da variável dependente.

Potts (2006) utilizaram cinco modelos de regressão para analisar relação entre a abundância de uma determinada planta com o ambiente. Foram eles: Poisson, Binomial Negativa, Quasi-Poisson, *hurdle model* e o ZIP. Eles foram avaliados teoricamente, e seu desempenho preditivo foi analisado com estatísticas de correlação, de calibragem e de erro calculadas dentro de um processo de avaliação de *bootstrap* que simulava desempenho para dados independentes. Um dos métodos utilizados neste trabalho foi o Coeficiente de Correlação de Pearson e Correlação de Spearman. Por fim, concluiu-se que, para este trabalho, o *hurdle model* foi a melhor alternativa para o método de modelagem.

Ridout et al. (2001) realizou um estudo para testar qual é o melhor modelo a ser utilizado considerando como critério de decisão o teste de Escore: o Modelo ZIP ou o Modelo ZINB. Para tal, ele realizou dois exemplos distintos: um sem covariáveis e outro com covariáveis. Ele concluiu que a vantagem da utilização do teste de Escore, comparado ao teste de razão de verossimilhança é que não se faz necessário que o modelo mais complexo seja utilizado. Já a desvantagem do teste de Escore, também comparado ao teste de razão de verossimilhança, é que a distribuição assintótica da estatística de Escore é abordada de forma mais lenta. Isso pode influenciar em uma significância errada para os testes, principalmente para amostras pequenas. Observando esse fato, percebeu-se que, comparando o Modelo ZIP com o Modelo ZINB, o teste de Escore tende a apresentar poucos resultados significativos. Após os testes, concluiu-se que se o teste de Escore indicar que o Modelo ZIP é inadequado, ao nível de significância de 10%, o Modelo ZINB deve ser utilizado para modelar os dados. Por outro lado, se o teste de Escore apresentar qualquer indicação de falta de ajuste, inferências com base no Modelo ZIP podem ser feitas com uma maior confiança.

Sergio (2012) realizou um estudo para identificar o comportamento da variável Contagem de Células Somáticas e encontrar uma distribuição de probabilidade que se ajustasse melhor aos dados. Foi utilizada inicialmente uma família de distribuições para dados inflacionados de zeros, e observou-se seu comportamento em relação às covariáveis idade e dias de lactação. Os dados foram adquiridos através da EMBRAPA. Para a tomada de decisão do melhor modelo foi utilizado o AIC. Após as análises concluiu-se que o Modelo de Regressão ZINB foi o que melhor se ajustou aos dados, apesar da estimativa para a proporção de zeros não se apresentar significativa.

1.3 Revisão de Literatura sobre Poluição Atmosférica e Sintomas Respiratórios

Por ser um tema bastante amplo, há muitos artigos referentes à associação entre a poluição atmosférica e sintomas respiratórios. O foco desta revisão de literatura foi nos poluentes CO, NO₂, SO₂, O₃, PM₁₀, PM_{2.5} e Idade.

Em âmbito internacional, destacam-se três estudos que avaliaram os efeitos da poluição atmosférica em escolares (Sarnat (2012); Epton (2008); Lee (2005))

Sarnat (2012) realizou um estudo de painel a fim de medir o impacto da poluição do ar em crianças asmáticas de Ciudad Juarez e El Paso, cidades situadas na fronteira do México e EUA, respectivamente. Participaram do estudo 56 escolares divididos em quatro escolas, sendo duas em cada cidade. Selecionou-se as escolas de pontos distintos, sendo uma situada em zona de tráfego leve e a outra situada em zona de tráfego intenso. Os dados relacionados à saúde dos escolares foram coletados diariamente nas escolas utilizando um monitor de inflamação de vias aéreas (NIOX MINO®). Os dados referentes à qualidade do ar foram coletados de cinco estações de monitoramento contínuo do ar, distribuído em pontos estratégicos de El Paso. As análises estatísticas foram realizadas usando técnicas de Modelos Lineares Generalizados, mais precisamente: o modelo logístico e o modelo misto. Os resultados obtidos no estudo apresentaram evidências de um efeito adverso da poluição atmosférica (NO) sobre a saúde dos asmáticos. Mais especificamente, observou-se diferenças estatisticamente significante na ocorrência de sintomas respiratórios, segundo a zona de tráfego.

Epton (2008) realizou um estudo de painel com o objetivo de analisar os efeitos da poluição do ar proveniente da queimada de madeiras em chaminés, na saúde de alunos de idade escolar de Christchurch, Nova Zelândia. Participaram do estudo 93 estudantes de sexo masculino, sendo 26 deles asmáticos, com idade variando entre 12 a 18 anos. Os estudantes selecionados residiam no alojamento de uma escola. Os dados referentes aos sintomas respiratórios dos participantes foram registrados diariamente e medidos utilizando aparelhos específicos. Os dados referentes à qualidade do ar também foram medidos aparelhos e técnicas específicas. As análises estatísticas foram realizadas utilizando testes não paramétricos (Friedman, Wilcoxon, Spearman, Mann-Whitney). Os resultados obtidos no estudo observaram efeito significativo da poluição do ar por material particulado (PM) nos escolares, mais especificamente a tosse.

Lewis (2013) realizou um estudo de painel a fim de avaliar o efeito de material particulado (PM) e ozônio (O_3) em sintomas respiratórios de crianças asmáticas de baixa renda, que vivem em duas regiões distintas de Detroit, Michigan. A população de estudo foi preenchida por 298 crianças asmáticas por um período de quatro anos. Os dados das crianças foram obtidos através de questionário, e os de sintomas foram obtidos através de um diário de sintomas. Os níveis dos poluentes foram medidos com uma rede de monitoramento distribuída em cada região. As análises estatísticas foram realizadas utilizando modelo de regressão logística, utilizando equações de estimação generalizada. Os resultados obtidos apresentaram uma relação entre o aumento no nível de ozônio (O_3) com os níveis de chiado.

No Brasil, três estudos em áreas urbanas avaliaram o efeito da poluição em amostras de crianças e adolescentes (Castro (2009); Rosa, (2009); Moura, (2009)).

Castro (2009) avaliou 118 crianças, através de um estudo de painel, entre seis e quinze anos de idade, de uma escola da rede pública de ensino, do Complexo de Mangueiras, no Rio de Janeiro. O estudo visava mensurar o efeito da poluição do ar na saúde respiratória dos escolares. Os dados sobre as características das crianças foram obtidos através de um questionário, que incluía o International Study of Asthma and Allergies in Childhood (ISAAC). Os níveis dos poluentes foram medidos a partir de uma unidade móvel de monitoramento dos poluentes da Secretaria Municipal de Meio Ambiente do Rio de Janeiro, no local de estudo, e foram realizados exames diários de pico de fluxo para medir a função respiratória dos escolares. Além disso, as condições meteorológicas foram obtidas através de medidores localizados no aeroporto do Galeão. A análise estatística explorou a hierarquia natural dos dados. Logo, para as medidas repetidas foi utilizado um modelo multinível gaussiano, onde as unidades de primeiro nível foram os dias de avaliação pulmonar, e as unidades de segundo nível foram as crianças. Os resultados obtidos no trabalho mostraram que apesar dos níveis de poluição estar estabilizados na maior parte do período de estudo, o PM_{10} e o NO_2 se associaram à diminuição da função respiratória dos escolares.

Rosa (2009) realizou um estudo com 3362 estudantes residentes no município de Tangará da Serra, Mato Grosso. O estudo visava analisar a prevalência de asma e seus sintomas em estudantes de duas faixas etárias específicas: Escolares (6-7 anos) e adolescentes (13-14 anos). Os dados sobre as características das crianças foram coletadas através de um questionário padronizado do International Study of Asthma and Allergies in Childhood (ISAAC). Os bairros do município foram agrupados em três

categorias: urbano, periurbano e rural. Foi realizado um estudo transversal de base populacional. Para a análise dos dados foi utilizado o teste de Qui-Quadrado para comparar diferenças de proporções. Os resultados obtidos observaram uma maior prevalência de asma e de sintomas relacionados ao grupo de escolares, o que mostra uma redução na prevalência de asma na adolescência.

Moura (2009) realizou um estudo com 45.595 visitas pediátricas emergenciais devido a problemas respiratórios que deram entrada em um hospital público de Jacarepaguá, no Rio de Janeiro. Seu objetivo era estimar a associação entre os níveis dos poluentes atmosféricos e sintomas respiratórios nas crianças. Os dados foram obtidos do prontuário médico de crianças entre um mês e 12 anos de idade, no período de 01 de abril de 2002 a 31 de março de 2003. Os níveis dos poluentes foram registrados a partir de uma estação de monitoramento automático, localizado no centro de Jacarepaguá. As análises estatísticas foram realizadas utilizando técnicas de séries temporais. A regressão de Poisson semiparamétrica foi utilizada para modelar as séries. Os resultados obtidos mostraram que a exposição à poluentes atmosféricos foram associadas a um aumento significativo nas consultas pediátricas de emergência devido a sintomas de obstrução brônquica em crianças menores de 2 anos de idade. Além disso, o aumento no nível de PM10 estava altamente relacionado com as visitas das crianças a emergência.

Na Amazônia Brasileira, Jacobson e colaboradores (2012; 2014) avaliaram o efeito da poluição em dois estudos de painel com escolares residentes em municípios do Mato Grosso.

No estudo realizado em Alta Floresta, Jacobson (2012) visou investigar se a exposição ao material particulado fino (PM_{2.5}) tinha efeito adverso no pico de fluxo expiratório diário de crianças em idade escolar. Foi realizado um estudo de painel com 309 escolares com idade entre 6 a 15 anos, no período de agosto a dezembro de 2006, período conhecido como de seca na região amazônica. Os picos de fluxo expiratório foram medidos diariamente, exceto em feriados e finais de semana. Análises estatísticas foram realizadas através de modelos de efeitos mistos, considerando os níveis de poluição do ar do dia atual, com defasagem de um e dois dias. Foram encontrados resultados estatisticamente significantes, onde se associou a exposição ao PM_(2.5) com a redução da função pulmonar dos escolares.

No estudo realizado em Tangará da Serra, Jacobson (2014) teve por objetivo analisar os efeitos adversos da poluição atmosférica devido às queimadas na saúde de

crianças. O estudo visou estimar os efeitos da poluição no período de seca da Amazônia através do pico de fluxo expiratório. Foi realizado um estudo de painel com 234 escolares com idade entre 6 e 15 anos, no ano de 2008. Os picos de fluxo expiratórios foram coletados através de um aparelho específico. Os dados referentes aos poluentes foram coletados utilizando materiais específicos. Os dados meteorológicos foram obtidos junto ao INMET (Instituto Nacional de Meteorologia). Foi implementado neste estudo cinco tipos de análises estatísticas em ordem de complexidade: análise do banco de dados separados por crianças; modelos de efeitos mistos relativo ao clima e covariáveis individuais; estimação do efeito da poluição do ar para defasagens simples; estimação do efeito da poluição do ar para PDLM; análise de sensibilidade. Os resultados encontrados mostraram que as reduções do pico de fluxo expiratório se associaram com os níveis da poluição do ar, principalmente quando observadas com defasagens de 3-5 dias e para com crianças mais novas.

Não existem estudos publicados no Brasil que relacionam a ocorrência diária de sintomas respiratórios em escolares com a poluição atmosférica.

Neste trabalho será usado o mesmo banco de dados utilizado no estudo de Jacobson et al. (2014).

2 Objetivos.

1.1 Geral

Avaliar o excesso de zeros na variável de sintomas respiratórios (Tosse, Coriza), em estudantes de 6 a 15 anos residentes em Tangará da Serra, no Estado do Mato Grosso.

1.2 Específicos

- Criar um banco de dados com as variáveis de pesquisa agregadas;
- Estimar os quatro Modelos de Regressão (Poisson, Binomial Negativa, ZIP e ZINB) para cada sintoma respiratório, tendo como variáveis resposta número de ocorrências de sintomas Tosse ou Coriza durante o período de seca;
- Comparar os modelos e escolher o que melhor se ajusta aos dados de cada sintoma respiratório;
- Verificar quais variáveis estão associadas à ocorrência de cada sintoma respiratório;
- Verificar o efeito defasado da exposição ao poluente atmosférico PM10.

3 Material e Métodos

3.1 Sintomas Respiratórios na Amazônia

Foi realizado um estudo de painel em estudantes de uma Escola no município de Tangará da Serra, localizado no estado do Mato Grosso, na região amazônica, entre os meses de agosto e dezembro do ano de 2008 (considerado período de seca na região). Foram selecionados alunos de 6 a 15 anos de idade da Escola Pedro Alberto Tayano, situada no bairro Vila Esmeralda. A ideia inicial consistia em selecionar 25 alunos de cada idade, porém somente 234 alunos iniciaram o estudo, conforme apresentado na Tabela abaixo. Os responsáveis dos referidos alunos assinaram um Termo de Consentimento Livre e Esclarecido.

Tabela1: Distribuição de frequência da amostra de crianças e adolescentes

	Idade										Total
	06	07	08	09	10	11	12	13	14	15	
Meninas	07	19	18	09	09	16	13	18	13	03	125
Meninos	05	11	12	16	17	07	17	10	12	02	109
Total	12	30	30	25	26	13	30	28	25	05	234

Foram retiradas do banco de dados inicial, as crianças que apresentaram dado faltante referente a apresentar histórico de asma e as que saíram da escola durante o período de estudo. Com isso, o estudo foi realizado com um total de 220 estudantes.

Este projeto considerou relevante para o estudo as seguintes variáveis:

- Variáveis de sintomas respiratórios (Tosse e Coriza);
- Variáveis de poluição (representada pelo nível do poluente PM10);
- Variáveis meteorológicas (representada pelos níveis de Temperatura e Umidade do Ar);
- Características individuais (obtidas através de um questionário).

Para esse estudo, foi necessário realizar algumas modificações no banco de dados. A primeira foi agregar as variáveis de sintomas, visto que este projeto visa avaliar o número de sintomas respiratórios de cada estudante durante o período de estudo.

A variável Sintoma Respiratório é definida por y_{it} e será 1 se o i -ésimo indivíduo teve o sintoma no dia t . Com isso, a variável resposta para o banco agregado será o número de sintomas do indivíduo i no período de estudo:

$$y_i = \sum_{t=1}^{82} y_{it}$$

As p variáveis relacionadas aos níveis de poluição, temperatura e umidade do ar serão denotadas da seguinte maneira:

$$X_i = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1j} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2j} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{t1} & x_{t2} & \dots & x_{tj} & \dots & x_{tp} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{82\ 1} & x_{82\ 2} & \dots & x_{82\ j} & \dots & x_{82\ p} \end{pmatrix}_{82 \times p}$$

sendo que x_{tj} está representando uma covariável j no dia t , onde $t = 1, 2, \dots, 82$ e $j = 1, 2, \dots, p$. Por exemplo, $x_{3,4}$ representa o valor da 3ª covariável (Temperatura por exemplo), referente ao 4º dia de estudo.

A metodologia utilizada para agregar as variáveis acima foi a mesma utilizada no estudo de Hwang et al. (2000). Essa metodologia consiste em calcular o nível de exposição da variável para cada indivíduo, ponderado pelos dias que o mesmo apresentou a característica de interesse. Nesse projeto, foi calculado o nível de exposição da variável x_j , $j = 1, 2, \dots, p$, para cada indivíduo i , ponderado pelos dias que o estudante apresentou o sintoma. Pode também ser visto como a média de x_j , ponderado pelos dias que o indivíduo apresentou o sintoma.

$$w_{ij} = \frac{1}{\sum_{t=1}^{82} y_{it}} \sum_{t=1}^{82} (y_{it} x_{tj})$$

Podem ocorrer casos em que a ocorrência do sintoma está influenciada pelos valores dos níveis da exposição de dias anteriores. Com isso, calculou-se o nível de exposição considerando uma defasagem de 1 a 3 dias para as variáveis de poluição e uma defasagem de 1 a 2 dias para as variáveis meteorológicas.

Para tal, calculamos o nível de exposição da variável j com defasagem de n dias para o indivíduo i da seguinte forma:

$$z_{ij} = \frac{1}{\sum_{t=1}^{82} y_{it}} \sum_{t=1}^{82} (y_{it} x_{(t-n)j})$$

A Tabela 2 nos mostra o resultado final da construção das variáveis:

Tabela 2: Resumo da construção das variáveis.

Variável Original	Variável Transformada	Explicação
Se apresentou o sintoma	Y_i	Número de sintomas do indivíduo no período de estudo.
Valores diários de Temperatura Média, Umidade do ar e PM_{10} .	W_i	Média das variáveis meteorológicas e de poluição em que o i -ésimo indivíduo foi exposto nos dias em que apresentou o sintoma.
Defasagem das variáveis Temperatura Média, Umidade do ar e PM_{10} .	Z_i	Média do nível de exposição do i -ésimo indivíduo com n dias de defasagem.

As variáveis relacionadas às características de cada criança foram coletadas através de um questionário. Para este trabalho foram selecionadas as variáveis que, segundo a literatura, podem se relacionar com a nossa variável de interesse. As variáveis selecionadas para este estudo foram Idade, Sexo e Histórico de Asma.

3.2 Modelos Lineares Generalizados

3.2.1 Definição

Os modelos lineares generalizados são uma extensão do modelo linear clássico, desenvolvido por Nelder & Wedderburn (1972), para uma família mais geral de distribuições. Eles são definidos a partir de distribuição de probabilidade para a variável resposta Y , que se associa a um conjunto de variáveis explicativas $(X_1, X_2, X_3, \dots, X_p)$ e também por uma função de ligação.

Esses modelos são comumente utilizados quando os valores possíveis para a variável resposta tomam valores binários, quando são provenientes de contagens ou também quando a variância da variável resposta depende da média.

Os modelos lineares generalizados são definidos pelos seguintes componentes:

- Componente Aleatório;
- Componente Sistemático;
- Função de Ligação.

1º) Componente Aleatório: É composta por uma variável aleatória Y com n valores independentes, de um vetor de média e uma distribuição de probabilidade pertencente à família exponencial.

2º) Componente sistemática: É composta por variáveis explicativas $x_{i1}, x_{i2}, \dots, x_{ip}$ que produzem um preditor linear η_i , dado por:

$$\eta_i = x_i^T \beta = [1 \ x_{i1} \ x_{i2} \ \dots \ x_{ip}] \cdot \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} = \sum_{j=0}^p \beta_j x_{ij}$$

onde β é o vetor de parâmetros desconhecido.

3º) Função de ligação: É uma função monótona e diferenciável que relaciona o componente aleatório ao componente sistemático através da função $g(\mu_i) = x_i^T \beta$, isto é:

$$g(\mu_i) = \eta_i = x_i^T \beta, \quad \forall i=1,2,\dots,n$$

3.2.2 Família Exponencial

Algumas características importantes do Modelo Linear Generalizado é que elas partem da hipótese de independência (ou não correlação) das observações, além do fato da variável resposta pertencer à família exponencial.

Definição: Seja Y uma variável aleatória discreta (contínua) com função de probabilidade (densidade) $P(Y = y, \theta)$ ($f(y, \theta)$). Dizemos que a distribuição de probabilidade de Y pertence a família exponencial se sua função de probabilidade puder ser expressa da forma:

$$P(Y = y, \theta) = e^{[a(y).b(\theta)+c(\theta)+d(y)]} \quad , \text{no caso discreto}$$

$$f(y, \theta) = e^{[a(y).b(\theta)+c(\theta)+d(y)]} \quad , \text{no caso contínuo}$$

onde $a(y)$, $b(\theta)$, $c(\theta)$, $d(y)$ são funções conhecidas.

Algumas observações relevantes sobre a família exponencial de distribuições nos mostram que se $a(y) = y$, temos a forma canônica da função de probabilidade (ou função densidade de probabilidade) de Y . Com isso, temos que $b(\theta)$ é o parâmetro natural da distribuição. Além disso, assumindo a forma canônica de Y , podemos calcular sua esperança e variância da seguinte forma:

$$E(a(Y)) = E(Y) = -\frac{c'(\theta)}{b'(\theta)}$$

$$VAR(a(Y)) = VAR(Y) = \frac{b''(\theta).c'(\theta) - b'(\theta).c''(\theta)}{[b'(\theta)]^3}$$

Onde: $b'(\theta)$, $b''(\theta)$, $c'(\theta)$ e $c''(\theta)$ são a primeira e segunda derivada de $b(\theta)$ e $c(\theta)$, respectivamente.

Abaixo segue demonstração de que as distribuições de Poisson e Binomial Negativa, distribuições que serão utilizadas neste trabalho, pertencem à família exponencial de distribuições.

3.2.2.1 Distribuição de Poisson

Uma variável aleatória Y segue distribuição de Poisson se sua função de probabilidade é dada por:

$$P(Y = y; \lambda) = \frac{e^{-\lambda} \lambda^y}{y!}, y = 0, 1, 2 \dots \text{ e } \lambda > 0$$

Escrevendo na forma da família exponencial, temos que:

$$P(Y = y; \lambda) = e^{[y \cdot \ln(\lambda) - \lambda - \ln(y!)]}$$

Neste caso, $a(y) = y$, $b(\lambda) = \ln(\lambda)$, $c(\lambda) = -\lambda$ e $d(y) = -\ln(y!)$

Observando que Y apresenta forma canônica, podemos afirmar que $\ln(\lambda)$ é o parâmetro natural de Y , e sua esperança e variância são, respectivamente:

$$E(Y) = -\frac{c'(\lambda)}{b'(\lambda)} = -\frac{-1}{\frac{1}{\lambda}} = \lambda$$

$$VAR(Y) = \frac{b''(\lambda) \cdot c'(\lambda) - b'(\lambda) \cdot c''(\lambda)}{[b'(\lambda)]^3} = \frac{\frac{-1}{\lambda^2} \cdot (-1) - \frac{1}{\lambda} \cdot 0}{\left[\frac{1}{\lambda}\right]^3} = \frac{\frac{1}{\lambda^2}}{\frac{1}{\lambda^3}} = \frac{\lambda^3}{\lambda^2} = \lambda$$

$$\text{Logo; } E(Y) = \lambda = Var(Y)$$

3.2.2.2 Distribuição Binomial Negativa

Uma variável aleatória Y segue distribuição Binomial Negativa se sua função de probabilidade é dada por:

$$P(Y = y|k, p) = \binom{y+k-1}{k-1} p^k (1-p)^y, \quad y = 1, 2, 3 \dots$$

Onde k = número de sucessos

Escrevendo na forma da família exponencial, temos que:

$$P(Y = y|k, p) = e^{[y \cdot \ln(1-p) + k \cdot \ln(p) + \ln\left(\frac{y+k-1}{k-1}\right)]}$$

Neste caso, $a(y) = y$, $b(p) = \ln(1-p)$, $c(p) = k \cdot \ln(p)$ e $d(y) = \ln\left(\frac{y+k-1}{k-1}\right)$

Observando que Y apresenta forma canônica, podemos afirmar que $\ln(1-p)$ é o parâmetro natural de Y , e sua esperança e variância são, respectivamente:

$$E(Y) = -\frac{c'(p)}{b'(p)} = -\frac{\frac{k}{p}}{\frac{-1}{(1-p)}} = \frac{k(1-p)}{p}$$

$$VAR(Y) = \frac{b''(p) \cdot c'(p) - b'(p) \cdot c''(p)}{[b'(p)]^3} = \frac{\frac{-1}{(1-p)^2} \cdot \frac{k}{p} - \frac{-1}{(1-p)} \cdot \frac{-k}{p^2}}{\left[\frac{-1}{(1-p)}\right]^3} = \frac{\frac{-kp - k(1-p)}{p^2(1-p)^2}}{\frac{-1}{(1-p)^3}} = \frac{k \cdot (1-p)}{p^2}$$

Logo; $E(Y) = \frac{k(1-p)}{p}$ e $V(Y) = \frac{k(1-p)}{p^2}$.

3.2.3 Modelos de Regressão para Dados de Contagem

Nesta seção, foram usados como referências os textos de Santos (2013), Fumes (2010) e Sergio (2012).

Define-se contagem de dados como sendo o número de vezes que ocorre determinado evento, em uma mesma unidade de observação durante um intervalo de tempo.

Usualmente, o modelo de regressão de Poisson é utilizado para modelar o tipo de dado acima explicitado. Porém, podem ocorrer situações onde a utilização do modelo de Poisson não é o mais adequado, visto que em alguns casos a variância da variável dependente (resposta) é superior a sua média, fenômeno conhecido como superdispersão. Outro problema que pode ocorrer é o excesso de zeros nos dados, o que afeta diretamente no ajustamento do modelo de regressão de Poisson.

Um método comumente utilizado para resolver o problema da superdispersão é a utilização do modelo de regressão Binomial Negativa. Ele é uma generalização do modelo de regressão de Poisson. Porém, na ocorrência de excesso de zeros na variável resposta, esse modelo também não é o mais adequado.

Por esse motivo, os modelos de regressão de zeros inflacionados foram desenvolvidos. Eles têm o intuito de tratar separadamente a massa de probabilidade no ponto zero. Para tal, dois modelos são frequentemente utilizados, o modelo de regressão de Poisson de zeros inflacionados (modelo ZIP) e o modelo de regressão Binomial Negativa de zeros inflacionados (modelo ZINB). Ambos modelam utilizando uma combinação de duas distribuições distintas, uma para tratar o excesso de zeros, assumindo que a única observação possível é zero, com *probabilidade* = p e outra que trata das contagens, com *probabilidade* = $1 - p$.

3.2.3.1 Modelo de Regressão de Poisson

Suponha Y_1, \dots, Y_n são variáveis aleatórias independentes tais que $Y_i \sim \text{Poisson}(\lambda_i)$. A sua função de Probabilidade é dada por:

$$P(Y_i = y_i | \lambda_i) = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!}, y_i = 0, 1, 2, \dots \quad \lambda_i > 0$$

onde λ_i representa o número médio de ocorrência de determinado evento. Como mostrado em 1.1.1, a média e variância de Y_i é dada por:

$$E(Y_i) = VAR(Y_i) = \lambda_i$$

Considere a variável aleatória Y que representa o número de ocorrências de um acontecimento em um determinado espaço de tempo. Dado $X = (X_1, X_2, \dots, X_p)^T$ um vetor de covariáveis e $x_i^T = (x_{i1}, x_{i2}, \dots, x_{ip})$ as observações do indivíduo i , assume-se

$$Y | X = x_i \sim \text{Poisson}(\lambda(x_i))$$

onde $\lambda_i = \lambda(x_i)$ é o número médio de ocorrência de um dado acontecimento, dada a observação x_i .

Com isso, para modelar $E(Y | X = x_i)$ poderia utilizar um modelo linear da forma:

$$\lambda(x_i) = Z_i^T \beta = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}$$

onde β é o vetor de coeficiente de regressão e $Z_i = (1, X_i^T)^T$ é a componente sistemática do modelo.

Porém este modelo não é apropriado, pois o preditor linear pode assumir qualquer valor real, enquanto λ_i só pode assumir valores não Negativas.

Como mostrado em 3.2.1.1., o parâmetro natural de Y é dado por $\ln(\lambda)$. Logo a função de ligação para a componente aleatória com distribuição de Poisson é a função logarítmica. Com isso solucionamos o problema e assim, temos que o Modelo de Regressão de Poisson é definido como:

$$\ln(\lambda(x_i)) = \ln(\eta_i) + Z_i^T \beta = \ln(\eta_i) + \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}$$

O termo $\ln(\eta_i)$ é conhecido como offset. Ele é uma constante conhecida que se incorpora ao modelo, utilizada como uma forma de padronizar a variável resposta. Neste trabalho, o offset representa o tempo de exposição de cada estudante. Ou seja, é o tempo que cada estudante ficou no estudo durante os 82 dias.

3.2.3.2 Modelo de Regressão Binomial Negativa

Vamos supor que Y_1, \dots, Y_n são variáveis aleatórias independentes tais que $Y_i \sim \text{BinNeg}(k, p)$, temos que a função de probabilidade de Y é dada por:

$$P(Y_i = y_i | p_i, k) = \binom{y_i + k - 1}{k - 1} p_i^k (1 - p_i)^{y_i}, y_i = 1, 2, 3 \dots$$

Onde k é o número de sucessos.

Realizando as substituições $k = \frac{1}{\lambda}$ e $p = \frac{1}{1 + \lambda\beta}$, chegamos a seguinte função:

$$P(Y_i = y_i | \lambda_i, \beta) = \binom{y_i + \frac{1}{\beta} - 1}{\frac{1}{\beta} - 1} \left(\frac{1}{1 + \beta\lambda_i} \right)^{\frac{1}{\beta}} \left(\frac{\beta\lambda_i}{1 + \beta\lambda_i} \right)^{y_i}, y_i = 1, 2, 3 \dots$$

onde β é o parâmetro de heterogeneidade.

Com as substituições, temos que a esperança e a variância de Y_i são dadas por:

$$E[Y_i] = \lambda_i$$

$$\text{Var}[Y_i] = \lambda_i + \beta\lambda_i^2$$

podemos observar que a variância da distribuição Binomial Negativa apresenta um termo adicional $\beta\lambda^2$ que, comparado à variância da distribuição Poisson, ajuda em um melhor ajuste de uma variável com superdispersão.

Assumindo Y como sendo uma variável aleatória que representa o número de ocorrências de um evento, $X = (X_1, X_2, \dots, X_p)^T$ um vetor de covariáveis e $x_i^T = (x_{i1}, x_{i2}, \dots, x_{ip})$ uma observação relacionada ao indivíduo i , temos que

$$Y | X = x_i \sim \text{BN}(\lambda(x_i), \beta)$$

Onde $\lambda_i = \lambda(x_i)$ representa o número médio de ocorrência de um evento, dada a observação (x_i) .

O Modelo de Regressão Binomial Negativa, assim como o Modelo de Regressão de Poisson, apresenta a necessidade de uma função de ligação para a componente

sistemática do modelo. Ela também é representada pela função logarítmica. Assim, o Modelo de Regressão Binomial Negativa é dado por:

$$Y|X = x_i \sim BN(\lambda(x_i), \beta)$$

e

$$\ln(\lambda(x_i)) = \ln(\eta_i) + Z_i^T \beta = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}$$

Onde $\ln(\eta_i)$ representa o offset, β é o vetor de coeficiente de regressão e $Z_i = (1, X_i^T)^T$ é a componente sistemática do modelo.

3.2.3.3 Modelo de Regressão de Poisson Inflacionado de Zeros (ZIP)

Supondo y_1, \dots, y_n observações da variável resposta Y_i , $i = 1, 2, \dots, n$, temos que o Modelo de Regressão de Poisson de Zeros Inflacionados é dado por:

$$P(Y_i = y_i | x_i) = \begin{cases} p_i + (1 - p_i)e^{-\lambda_i}, & y_i = 0 \\ (1 - p_i) \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!}, & y_i > 0 \end{cases}$$

Onde $0 < p_i < 1$ e $\lambda_i > 0$. O parâmetro p pode ser interpretado como a proporção de zeros e λ_i como o parâmetro de intensidade.

A média e a variância da variável Y_i são dadas por:

$$E[Y_i] = (1 - p_i)\lambda_i$$

$$Var[Y_i] = (1 - p_i)(1 + p_i\lambda_i)\lambda_i$$

O Modelo de Regressão de Poisson de Zeros Inflacionados modela os parâmetros de forma distinta. A média λ_i é modelada através de uma regressão de Poisson e a probabilidade p_i através de uma regressão Logística, usando como função de ligação $\eta_i = \text{logit}(p_i)$. Então temos que:

$$\log(\lambda_i) = X_i \beta$$

$$\text{logit}(p_i) = \frac{p_i}{1 - p_i} = G_i \gamma$$

Onde X_i e G_i são as matrizes de covariáveis. Portanto podemos concluir que:

$$Y|X = x_i \sim ZIP(\lambda_i, p_i)$$

3.2.3.4 Modelo de Regressão Binomial Negativa Inflacionado de Zeros (ZINB)

Supondo y_1, \dots, y_n observações da variável resposta Y_i , $i = 1, 2, \dots, n$, temos que o Modelo de Regressão Binomial Negativa de Zeros Inflacionados é dado por:

$$P(Y_i = y_i | x_i) = \begin{cases} p_i + (1 - p_i) \left(\frac{1}{1 + \beta \lambda_i} \right)^{\frac{1}{\beta}}, & y_i = 0 \\ (1 - p_i) \binom{y + \frac{1}{\beta} - 1}{\frac{1}{\beta} - 1} \left(\frac{1}{1 + \beta \lambda_i} \right)^{\frac{1}{\beta}} \left(\frac{\beta \lambda_i}{1 + \beta \lambda_i} \right)^{y_i}, & y_i > 0 \end{cases}$$

Onde $0 \leq p_i < 1$, $\lambda_i > 0$ e $\beta > 0$. O parâmetro p pode ser interpretado como a proporção de zeros, λ como o parâmetro da média e β o parâmetro de dispersão.

A média e variância de Y_i são dadas por

$$E[Y_i] = (1 - p_i)\lambda_i$$

$$Var[Y_i] = (1 - p_i)(1 + p_i\lambda_i + \beta\lambda_i)\lambda_i$$

O Modelo de Regressão Binomial Negativa de Zeros Inflacionados modela os parâmetros de forma distinta. O parâmetro λ_i é modelado utilizando a função de ligação logarítmica, assim como o parâmetro β . Já o parâmetro p_i é através da função de ligação logito. Portanto podemos concluir que:

$$Y|X = x_i \sim ZINB(\lambda_i, p_i, \beta)$$

4 Resultados

4.1 Análise Exploratória

Fez-se necessário um estudo inicial das variáveis do banco de dados. Realizou-se então uma análise exploratória, a fim de avaliar suas principais características e seu comportamento.

As Tabelas 3 e 4 apresentam uma análise descritiva das variáveis quantitativas, com as variáveis agregadas a cada tipo de sintoma. Para tal, foram utilizadas as seguintes medidas: Número de observações, Mínimo, Máximo, Média e Desvio Padrão.

Tabela 3 – Análise descritiva das variáveis quantitativas agregadas para o sintoma Tosse.

Variáveis	Nº de Observações	Mínimo	Máximo	Média	Desvio Padrão
Idade	220	6	15	10,26	2,524
Temperatura lag1	220	0	29,6	23,43	8,299
Umidade Relativa lag 1	220	0	85,9	46,94	18,680
PM ₁₀ lag 1	220	0	139,25	51,02	27,981
PM ₁₀ lag 2	220	0	154,28	51,59	28,997
PM ₁₀ lag 3	220	0	164,48	53,65	31,834

Tabela 4 – Análise descritiva das variáveis quantitativas agregadas para o sintoma Coriza

Variáveis	Nº de Observações	Mínimo	Máximo	Média	Desvio Padrão
Idade	220	6	15	10,26	2,528
Temperatura lag1	220	0	29,6	20,64	10,98
Umidade Relativa lag 1	220	0	83	41,41	23,82
PM ₁₀ lag 1	220	0	164,5	43,77	32,11
PM ₁₀ lag 2	220	0	144,1	44,57	32,23
PM ₁₀ lag 3	220	0	156,87	47,22	34,29

Podemos observar que as médias das variáveis de poluição tendem a aumentar, conforme o aumento de sua defasagem. Isso nos leva a dizer, inicialmente, que o efeito

do poluente tende a aumentar a incidência de tosse, conforme o passar dos dias, para ambos os sintomas. Observa-se para ambos os sintomas que os níveis médios de PM₁₀ estão próximos ao limite estabelecido pela Organização Mundial de Saúde para efeitos adversos na saúde, que estabelece média diária de 50µg/m³.(OMS 2005)

A Tabela 5 apresenta uma análise descritiva das variáveis qualitativas. Foram utilizadas as seguintes medidas: a frequência absoluta e as porcentagens. Verificou-se que a distribuição por sexo é semelhante, enquanto que a distribuição de asmáticos foi distinta, tendo 18% de estudantes asmáticos.

Tabela 5 – Análise descritiva das variáveis qualitativas

Variáveis		Observações	Porcentagem
Sexo	Masculino	101	45,9%
	Feminino	119	54,1%
Asma	Sim	40	18,2%
	Não	180	81,8%

Já a Tabela 6 apresenta uma análise descritiva das variáveis de Sintomas Respiratórios. Podemos observar que a variância de ambos os sintomas são consideravelmente maiores que sua média, mostrando, assim, que esse é um caso de superdispersão. Além disso, o percentual de zeros foi de 11% para Tosse e 22% para coriza.

Tabela 6 – Análise descritiva da variável resposta Sintoma Respiratório

Sintomas	Nº de Observações	Mínimo	Porcentagem de zero	Máximo	Média	Variância
Tosse	220	0	11%	63	12,43	167,6
Coriza	220	0	22%	55	10,05	150,2

4.2 Modelagem dos Dados

O objetivo aqui é estimar os modelos para as quatro distribuições definidas no capítulo 3. Todas as modelagens foram feitas utilizando o Software R versão 3.1.0.

Inicialmente, verificou-se quais variáveis eram significativas para explicar a ocorrência de sintomas respiratórios. Como o objetivo é comparar os modelos, todas as variáveis serão utilizadas para comparar todos os modelos. As variáveis mais importantes foram sexo, idade, asma, PM_{10} com defasagem de 1, 2 e 3 dias, temperatura com defasagem de 1 e 2 dias e umidade do ar com defasagem de 1 e 2 dias.

Após feita todas as modelagens (anexo A), temos os valores do AIC para cada modelo com sintoma Tosse, conforme apresentado na Tabela 7. Observamos que o modelo que apresenta o menor AIC é o modelo de regressão Binomial Negativa, com defasagem do sintoma de poluição igual a 3. A Figura 3 ajuda a confirmar a escolha deste modelo, apresentando os resíduos da Deviance distribuídos aleatoriamente em torno do zero.

Tabela 7 – Valores do AIC referente ao sintoma Tosse

Modelos	AIC
BN (PM_{10} lag1)	1399,8
BN (PM_{10} lag2)	1400,5
BN (PM_{10} lag3)	1399,4
Poisson (PM_{10} lag1)	2423,8
Poisson (PM_{10} lag2)	2424,4
Poisson (PM_{10} lag3)	2416,2
ZIP (PM_{10} lag1)	2425,8
ZIP (PM_{10} lag2)	2426,4
ZIP (PM_{10} lag3)	2418,2
ZINB (PM_{10} lag1)	1401,8
ZINB (PM_{10} lag2)	1402,5
ZINB (PM_{10} lag3)	1401,3

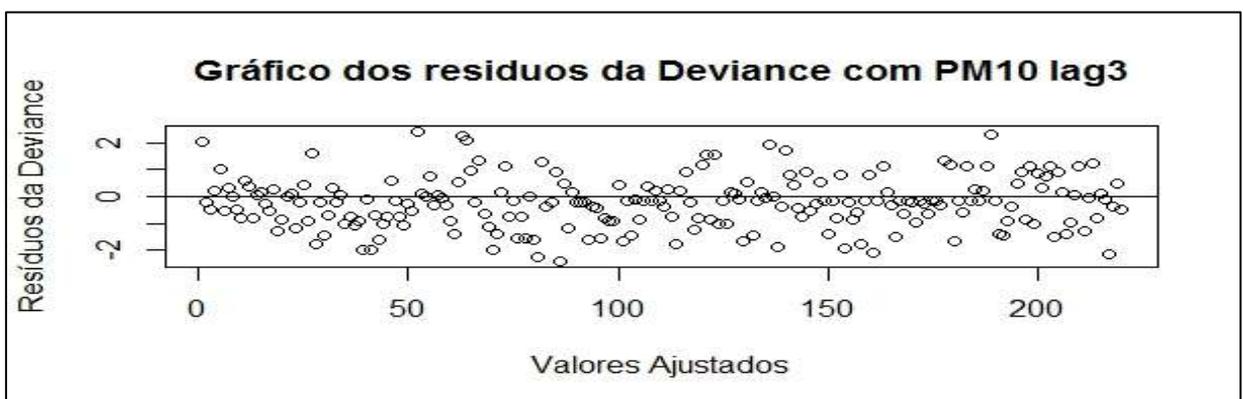


Figura 3 – Gráfico dos resíduos da Deviance para o Modelo de Regressão Binomial Negativa com a variável de poluição PM_{10} lag3.

Encontrando o modelo mais adequado para o sintoma Tosse, apresentado na Tabela 8 avaliamos o efeito do PM_{10} nos estudantes. Ao nível de significância de 5%, não foi observado efeito significativo do PM_{10} na ocorrência de Tosse.

Com relação às outras variáveis, verificou-se que com exceção da variável sexo, todas foram significativas ao nível de 5%. Para idade o Risco Relativo estimado (RR) foi de 0,90 (IC95%: 0,86; 0,94), ou seja, a cada aumento da idade em 1 ano, o risco de ocorrência de tosse diminui. Quanto à asma, nos estudantes asmáticos o risco de Tosse é 42% (RR=1,416) maior do que nos não asmáticos (IC95%: 1,071; 1,878). A temperatura apresenta um RR de 1,194 (IC95%: 1,102; 1,294), o que mostra que a cada aumento de 1°C da temperatura, o risco de ocorrência de tosse aumenta em 19%. Já para a umidade, o RR foi de 1,039 (IC95%: 1,024;1,055), ou seja, a cada aumento de 1% no nível de umidade, o risco de tosse aumenta em 4%.

Tabela 8 – Modelo de Regressão Binomial Negativa utilizando o PM_{10} lag3 para Tosse

	Coeficientes	P-Valor	IC 95%		Parâmetro de Dispersão	AIC
Intercepto	-7,538	<0,001	-10,123	-4,954		
Sexo	0,182	0,1133	-0,043	0,408		
Idade	-0,108	<0,001	-0,155	-0,063		
Asma	0,348	0,0147	0,069	0,630		
PM_{10} Lag3	-0,004	0,0750	-0,009	0,0004	1,808	1399,4
Temperatura Média Lag1	0,177	<0,001	0,097	0,258		
Umidade Lag1	0,038	<0,001	0,0234	0,054		

Para o sintoma Coriza, observamos através da Tabela 9 que o modelo que apresenta o menor AIC é o Modelo de Regressão ZINB, com defasagem do sintoma de poluição igual a 3. A Figura 4 ajuda a confirmar a escolha deste modelo, apresentando os resíduos da Deviance distribuídos aleatoriamente em torno do zero.

Tabela 9 – Valores do AIC referente ao sintoma Coriza

Modelos	AIC
BN (PM ₁₀ lag1)	1226,9
BN (PM ₁₀ lag2)	1227,2
BN (PM ₁₀ lag3)	1221,7
Poisson (PM ₁₀ lag1)	2263,9
Poisson (PM ₁₀ lag2)	2268,3
Poisson (PM ₁₀ lag3)	2240
ZIP (PM ₁₀ lag1)	2265,9
ZIP (PM ₁₀ lag2)	2270,3
ZIP (PM ₁₀ lag3)	2242,1
ZINB (PM ₁₀ lag1)	1228,9
ZINB (PM ₁₀ lag2)	1229,2
ZINB (PM ₁₀ lag3)	1188,7

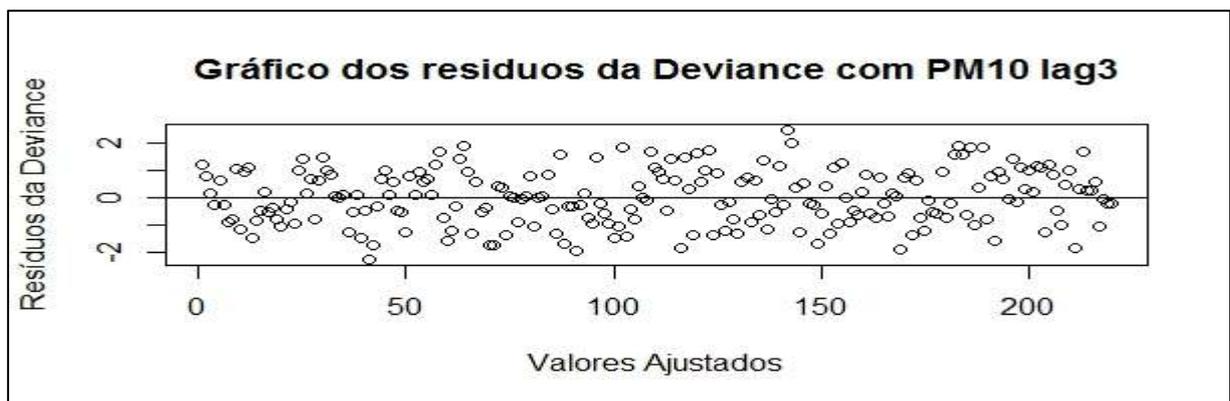


Figura 4 – Gráfico dos resíduos da Deviance para o Modelo de Regressão ZINB com a variável de poluição PM₁₀ lag3.

Encontrando nosso modelo mais adequado para o sintoma Coriza, conforme apresentado na Tabela 10, avaliamos o efeito do PM₁₀ nos estudantes. Ao nível de significância de 5%, a variável de poluição se apresentou significativa, porém sua contribuição é negativa. Logo ela apresenta um efeito protetor para o sintoma coriza.

Com relação às outras variáveis, se apresentaram significantes a idade e a Umidade do ar. Para idade o Risco Relativo estimado (RR) foi de 0,89 (IC95%: 0,84; 0,95), ou seja, a cada aumento da idade em 1 ano, o risco de ocorrência de coriza diminui. Já para a umidade, o RR foi de 1,05 (IC95%: 1,021;1,070), ou seja, a cada aumento de 1% no nível de umidade, o risco de coriza aumenta em 5%.

Tabela 10 – Modelo de Regressão ZINB utilizando o PM₁₀ lag3 para Coriza

	Coeficientes	P-Valor	IC 95%		Parâmetro de Dispersão	AIC
Intercepto	-7,136	0,016	-12,875	-1,399		
Sexo	0,046	0,753	-0,239	0,330		
Idade	-0,112	<0,001	-0,175	-0,049		
Asma	0,377	0,056	-0,008	0,762		
PM ₁₀ Lag3	-0,013	0,005	-0,022	-0,004	1	1188,677
Temperatura Média Lag1	0,173	0,071	-0,014	0,361		
Umidade Lag1	0,045	<0,001	0,021	0,068		

Para investigar o efeito protetor da poluição atmosférica, foi estimado o modelo apresentado na Tabela 11, onde se acrescenta uma interação entre as variáveis PM₁₀ e Umidade. O resultado deste modelo confirma que a variável Umidade é uma variável de confusão na relação entre Poluição atmosférica e ocorrência de Coriza

Tabela 11 – Modelo de Regressão ZINB com a interação entre PM₁₀ e Umidade do ar.

	Coeficientes	P-Valor	IC 95%	
Intercepto	-7,661	0,007	-13,242	-2,080
Sexo	0,036	0,791	-0,229	0,300
Idade	-0,078	0,011	-0,137	-0,019
Asma	0,333	0,056	-0,007	0,672
Temperatura Média Lag1	0,025	0,779	-0,152	0,203
PM ₁₀ lag3	0,060	<0,001	0,031	0,089
Umidade Lag1	0,121	<0,001	0,084	0,157
Interação PM ₁₀ lag3 x Umidade lag1	-0,001	<0,001	-0,002	-0,001

5 Conclusão

Conforme os resultados explicitados no capítulo anterior, podemos concluir que, para o conjunto de dados definidos para este trabalho, o Modelo de Regressão Binomial Negativa e o Modelo de Regressão ZINB são os que melhores explicam as variáveis de sintoma Tosse e Coriza, respectivamente.

Os modelos provenientes da distribuição de Poisson foram os que apresentaram os maiores valores para o AIC e, conseqüentemente os que pior se ajustaram aos dados, conforme observado nas Tabelas 7 e 9. Isso era de se esperar, visto que as variáveis dependentes (contagem) apresentaram além de uma proporção de zeros consideráveis, uma variabilidade muito alta (superdispersão, apresentada na Tabela 4).

Visto isso, os modelos provenientes da distribuição Binomial Negativa foram os que melhor se ajustaram aos dados. A variável de sintoma Coriza, cuja proporção de zeros era de 22% apresentou um melhor ajuste através do Modelo de Regressão ZINB. Esse resultado corrobora com os estudos de Fumes (2010) que também mostraram que o Modelo de Regressão ZINB é o modelo que melhor se ajusta aos dados, quando sua variável resposta apresenta uma proporção de zeros entre 10% e 50%.

Já a variável de sintoma Tosse, cuja proporção de zeros era de 11% apresentou um melhor ajuste através do Modelo de Regressão Binomial Negativa. Isso se aproxima bastante dos resultados encontrados por Fumes (2010) que chegaram a conclusão de que as variáveis dependentes (contagem) com proporção de zeros acima de 10% são melhores ajustadas pelo Modelo de Regressão Binomial Negativa.

Nagamine (2007) e Sergio (2012) utilizaram, assim como neste trabalho, o AIC e a Deviance como critério de seleção do melhor modelo. Nossos resultados corroboram, visto que em seus referidos trabalhos, o Modelo de Regressão ZINB também foi o que melhor se ajustou aos dados, uma vez que suas variáveis dependentes também apresentaram uma proporção de zeros considerável.

Quanto aos dados de saúde, os níveis de PM_{10} não apresentaram efeito significativo na ocorrência do sintoma Tosse. Isso vai contra os resultados encontrados por Epton (2008), que mostraram que aumento da ocorrência de tosse está associado aos níveis do PM_{10} .

Entretanto, a variável asma se mostrou significativa. Concluiu-se que o risco de tosse para os estudantes asmáticos era 42% maiores, comparado aos não asmáticos. Esse resultado corrobora com os dos trabalhos de Sarnat (2012) e Rosa (2009), que mostraram que crianças asmáticas são mais suscetíveis ao efeito dos poluentes.

Já para o sintoma Coriza, encontrou-se efeito significativo na variável de poluição PM_{10} . Porém, ela apresenta uma contribuição negativa, o que faz dela um efeito protetor para o sintoma. Isso é explicado pela presença da variável meteorológica Umidade do Ar. Ela age, neste caso como uma variável de confusão na modelagem dos dados. Isso é comprovado pela Tabela 11, onde apresenta a modelagem dos dados assumindo a interação entre o PM_{10} e a Umidade do Ar. Com isso, vemos que a variável de poluição além de significativa, apresenta uma contribuição positiva, o que faz dela um efeito causador para o sintoma.

Sumariamente, este trabalho buscou ajustar os modelos de regressão para verificar os efeitos da exposição ao PM_{10} sobre o número de ocorrência dos sintomas respiratório Tosse e Coriza, cuja frequência de zeros nestas variáveis era superior a 10%. A literatura sugeriu quatro tipos de modelos. Os resultados apontaram o Modelo de Regressão Binomial Negativa para os dados referente ao sintoma Tosse e o Modelo de Regressão ZINB para os dados referentes ao sintoma Coriza, corroborando com a literatura sobre o excesso de zeros e a superdispersão na variável dependente. Quanto à poluição, verificou-se efeito significativo apenas na ocorrência do sintoma Coriza.

Referências

- ARBEX, Marcos A. et al.. *“Air pollution and the respiratory system”*. Jornal Brasileiro de Pneumologia. 2012.
- BROOK, R. D.; RAJAGOPALAN, S. et al.. *“Particulate Matter Air Pollution and Cardiovascular Disease: An Update to the Scientific- Statement From the American Heart Association”*, <http://dx.doi.org/10.1161/CIR.0b013e3181d8e3e1>, 2010.
- OMS. *“Global Health Risk: Mortality and burden of disease attributable to selected major risk.”*. World Health Organization. 2009.
- RODRIGUEZ-VILLAMIZAR, L. A.; REY-SERRANO, J. J. et al.. *“The effects of air pollution on respiratory health in susceptible populations: a multilevel study in Bucaramanga”*. Cad. Saúde Pública. 2008.
- MOURA, M.; JUNGER, W. L. et al.. *“Qualidade do ar e transtornos respiratórios agudos em crianças”*. Revista Saúde Pública. 2008.
- CASTRO, H. A. et al.. *“Efeitos da poluição do ar na função respiratória de escolares, Rio de Janeiro, RJ”*. Academic Press. 4ed. 2006.
- CARMO, C. N. et al.. *“Associação entre material particulado de queimadas e doenças respiratórias na região sul da Amazônia brasileira”*. Rev. Panam Salud Publica. 2010.
- MOURA, M.; JUNGER, W. L. et al.. *“Qualidade do ar e atendimento pediátricos de emergência por sintomas de obstrução brônquica categorizadas por faixas etárias no Rio de Janeiro, Brasil”*. Caderno Saúde Pública. Rio de Janeiro. 2009.
- MASCARENHAS, M. D. M. et al.. *“Poluição Atmosférica devido a queima de biomassa florestal e atendimentos de emergência por doença respiratória em Rio Branco, Brasil”*. Jornal Brasileiro de Pneumologia. 2008.
- ROSA, A. M. et al.. *“Prevalencia de asma em escolares e adolescentes em um municipio na regio da Amazonia brasileira”*. Jornal Brasileiro de Pneumologia. 2009.

- SOLE, D. et al.. *“Prevalencia de sintomas de asma, rinite, e eczema atopico entre crianças e adolescentes brasileiros identificados pelo International Study of Asthma and Allergies (ISAAC)”*. Jornal de Pediatria. Rio de Janeiro. 2006.
- PEREIRA, V. S. et al.. *“Análise dos atendimentos ambulatoriais por doenças respiratórias no Município de Alta Floresta - Mato Grosso - Amazônia Brasileira”*. Epidemiol. Serv. Saúde. 2011.
- SARNAT, S. E. et al.. *“Air Pollution and Acute Respiratory Response in a Panel of Asthmatic Children Along the U.S.-Mexico Border”*. Environmental Health Perspectives. 2012.
- EPTON, M. J. et al.. *“The effect of ambient air pollution on respiratory health of school children: a panel study”*. Environmental Health. 2008.
- LEWIS, T. C. et al.. *“The effect of ambient air pollution on respiratory health of school children: a panel study”*. Science of the Total Environment. 2013.
- JACOBSON, L. S. V. et al.. *“Association between fine particulate matter and the peak expiratory flow of school children in the Brazilian subequatorial Amazon: A panel study”*. Environmental Research. 2012.
- JACOBSON, L. S. V. et al.. *“Acute Effects of Particulate Matter and Black Carbon from Seasonal Fires on Peak Expiratory Flow of Schoolchildren in the Brazilian Amazon”*. www.plosone.org. 2014.
- FAMOYE, F.; SINGH, K. P.. *“Zero-Inflated Generalized Poisson Regression Model with an Application to Domestic Violence Data”*. Journal of Data Science. p117-130. 2006.
- FUMES, G.; CORRENTE, J. E.. *“Modelos Inflacionados de Zeros: Aplicações na Análise de um Questionário de Frequência Alimentar”*. Rev. Bras. Biom. São Paulo. p24-38. 2010.
- NAGAMINE, C. M. L.. *“Modelo para Dados de Contagem: Um estudo sobre o número de ovos do mosquito aedes aegypti”*. 2007
- POTTS, J. M.; ELITH, J.. *“Comparing species abundance models”*. journal homepage: www.elsevier.com/locate/ecolmodel. p153-163. 2006.

RIDOUT, M. et al.. “*A Score Test for Testing a Zero-Inflated Poisson Regression Model Against Zero-Inflated Negative Binomial Alternatives*”. *Biometrics* 57. p219-223. 2001.

SERGIO, V. F. N.. “*Utilização das Distribuições Inflacionadas de Zeros no Monitoramento da Qualidade do Leite*”. UFMG: Minas Gerais. 2012.

JACOBSON, L. S. V.. “*Efeitos adversos da poluição atmosférica em crianças e adolescentes devido a queimadas na Amazônia: uma abordagem de modelos mistos em estudos de painel*”. UERJ: Rio de Janeiro, 2013.

KUBRUSLY, J. “*Notas de aula - Modelos Lineares I*”. Niterói, RJ: Departamento de Estatística - UFF, 2014.

CONAMA, Resoluções do CONAMA 1984-2008, 2^a. Edição, Brasília, 2008

SANTOS, J. H. P.. “*Modelos para dados de contagem com excesso de zeros*”. Universidade de Minho, <http://hdl.handle.net/1822/29402>, Portugal, 2013.

Anexo A – Modelagem dos Dados

Tabela 12 - Modelo Binomial Negativa Com PM₁₀ lag1 para o sintoma Tosse

	Coeficientes	P-Valor	IC 95%		Parâmetro de Dispersão	AIC
Intercepto	-7,447	<0,001	-9,941	-4,954	1,801	1399,8
Sexo	0,194	0,0931	-0,032	0,420		
Idade	-0,109	<0,001	-0,155	-0,063		
Asma	0,348	0,0152	0,067	0,628		
PM ₁₀ Lag1	-0,005	0,1606	-0,011	0,002		
Temperatura Média Lag1	0,172	<0,001	0,094	0,250		
Umidade Lag1	0,039	<0,001	0,023	0,056		

Tabela 13 - Modelo Binomial Negativa Com PM₁₀ lag2 para o sintoma Tosse

	Coeficientes	P-Valor	IC 95%		Parâmetro de Dispersão	AIC
Intercepto	-7,428	<0,001	-9,926	-4,930	1,7962	1400,5
Sexo	0,185	0,109	-0,041	0,411		
Idade	-0,110	<0,001	-0,156	-0,063		
Asma	0,349	0,015	0,068	0,630		
PM ₁₀ Lag2	-0,003	0,378	-0,008	0,003		
Temperatura Média Lag1	0,174	<0,001	0,096	0,252		
Umidade Lag1	0,036	<0,001	0,020	0,051		

Tabela 14 - Modelo de Poisson Com PM₁₀ lag1 para o sintoma Tosse

	Coeficientes	P-Valor	IC 95%		Parâmetro de Dispersão	AIC
Intercepto	-5,026	<0,001	-5,883	-4,170	1	2423,800
Sexo	0,147	<0,001	0,070	0,223		
Idade	-0,118	<0,001	-0,135	-0,102		
Asma	0,332	<0,001	0,246	0,417		
PM ₁₀ Lag1	-0,003	0,005	-0,006	-0,001		
Temperatura Média Lag1	0,109	<0,001	0,082	0,136		
Umidade Lag1	0,026	<0,001	0,021	0,032		

Tabela 15 - Modelo de Poisson Com PM₁₀ lag2 para o sintoma Tosse

	Coefficientes	P-Valor	IC 95%		Parâmetro de Dispersão	AIC
Intercepto	-5,030	<0,001	-5,892	-4,169	1	2424,400
Sexo	0,142	<0,001	0,066	0,219		
Idade	-0,119	<0,001	-0,135	-0,103		
Asma	0,332	<0,001	0,246	0,418		
PM ₁₀ Lag2	-0,003	0,006	-0,005	-0,001		
Temperatura Média Lag1	0,111	<0,001	0,083	0,138		
Umidade Lag1	0,025	<0,001	0,020	0,031		

Tabela 16 - Modelo de Poisson Com PM₁₀ lag3 para o sintoma Tosse

	Coefficientes	P-Valor	IC 95%		Parâmetro de Dispersão	AIC
Intercepto	-5,119	<0,001	-6,006	-4,233	1	2416,200
Sexo	0,139	<0,001	0,063	0,216		
Idade	-0,119	<0,001	-0,135	-0,102		
Asma	0,330	<0,001	0,244	0,416		
PM ₁₀ Lag3	-0,004	<0,001	-0,005	-0,002		
Temperatura Média Lag1	0,113	<0,001	0,085	0,141		
Umidade Lag1	0,027	<0,001	0,022	0,032		

Tabela 17 - Modelo ZIP Com PM₁₀ lag1 para o sintoma Tosse

	Coefficientes	P-Valor	IC 95%		Parâmetro de Dispersão	AIC
Intercepto	-5,027	<0,001	-5,601	-4,453	1	2425,812
Sexo	0,147	<0,001	0,121	0,173		
Idade	-0,118	<0,001	-0,124	-0,113		
Asma	0,332	<0,001	0,304	0,360		
PM ₁₀ Lag1	-0,003	<0,001	-0,004	-0,002		
Temperatura Média Lag1	0,109	<0,001	0,091	0,127		
Umidade Lag1	0,027	<0,001	0,024	0,029		

Tabela 18 - Modelo ZIP Com PM₁₀ lag2 para o sintoma Tosse

	Coefficientes	P-Valor	IC 95%		Parâmetro de Dispersão	AIC
Intercepto	-5,031	<0,001	-5,604	-4,459	1	2426,398
Sexo	0,143	<0,001	0,117	0,168		
Idade	-0,119	<0,001	-0,125	-0,114		
Asma	0,332	<0,001	0,304	0,360		
PM ₁₀ Lag2	-0,003	<0,001	-0,004	-0,002		
Temperatura Média Lag1	0,110	<0,001	0,092	0,128		
Umidade Lag1	0,026	<0,001	0,023	0,028		

Tabela 19 - Modelo ZIP Com PM₁₀ lag3 para o sintoma Tosse

	Coefficientes	P-Valor	IC 95%		Parâmetro de Dispersão	AIC
Intercepto	-5,120	<0,001	-5,686	-4,555	1	2418,181
Sexo	0,140	<0,001	0,114	0,165		
Idade	-0,119	<0,001	-0,124	-0,113		
Asma	0,330	<0,001	0,302	0,358		
PM ₁₀ Lag3	-0,004	<0,001	-0,004	-0,003		
Temperatura Média Lag1	0,113	<0,001	0,095	0,131		
Umidade Lag1	0,027	<0,001	0,025	0,029		

Tabela 20 - Modelo ZINB Com PM₁₀ lag1 para o sintoma Tosse

	Coefficientes	P-Valor	IC 95%		Parâmetro de Dispersão	AIC
Intercepto	-7,504	0,001	-11,950	-3,057	1	1401,831
Sexo	0,193	0,099	-0,035	0,422		
Idade	-0,109	<0,001	-0,155	-0,062		
Asma	0,348	0,019	0,059	0,638		
PM ₁₀ Lag1	-0,005	0,240	-0,013	0,003		
Temperatura Média Lag1	0,173	0,015	0,035	0,312		
Umidade Lag1	0,040	<0,001	0,017	0,062		

Tabela 21 - Modelo ZINB Com PM₁₀ lag2 para o sintoma Tosse

	Coeficientes	P-Valor	IC 95%		Parâmetro de Dispersão	AIC
Intercepto	-7,464	0,001	-11,870	-3,058	1	1402,531
Sexo	0,185	0,112	-0,042	0,412		
Idade	-0,110	<0,001	-0,156	-0,063		
Asma	0,349	0,019	0,060	0,638		
PM ₁₀ Lag2	-0,003	0,463	-0,010	0,004		
Temperatura Média Lag1	0,175	0,013	0,038	0,312		
Umidade Lag1	0,036	0,001	0,014	0,058		

Tabela 22 - Modelo ZINB Com PM₁₀ lag3 para o sintoma Tosse

	Coeficientes	P-Valor	IC 95%		Parâmetro de Dispersão	AIC
Intercepto	-7,484	0,001	-11,998	-2,969	1	1401,351
Sexo	0,183	0,115	-0,044	0,409		
Idade	-0,109	<0,001	-0,155	-0,062		
Asma	0,348	0,019	0,059	0,638		
PM ₁₀ Lag3	-0,004	0,130	-0,010	0,001		
Temperatura Média Lag1	0,176	0,014	0,036	0,315		
Umidade Lag1	0,038	<0,001	0,016	0,060		

Tabela 23 - Modelo Binomial Negativa Com PM₁₀ lag1 para o sintoma Coriza

	Coeficientes	P-Valor	IC 95%		Parâmetro de Dispersão	AIC
Intercepto	-6,991	<0,001	-8,988	-4,994	1,371	1226,900
Sexo	0,018	0,899	-0,255	0,290		
Idade	-0,119	<0,001	-0,175	-0,064		
Asma	0,334	0,049	0,002	0,666		
PM ₁₀ Lag1	-0,008	0,075	-0,017	0,001		
Temperatura Média Lag1	0,175	<0,001	0,110	0,241		
Umidade Lag1	0,037	<0,001	0,017	0,057		

Tabela 24 - Modelo Binomial Negativa Com PM₁₀ lag2 para o sintoma Coriza

	Coeficientes	P-Valor	IC 95%		Parâmetro de Dispersão	AIC
Intercepto	-7,036	<0,001	-9,026	-5,045	1,369	1227,200
Sexo	0,037	0,793	-0,237	0,310		
Idade	-0,117	<0,001	-0,173	-0,061		
Asma	0,345	0,042	0,013	0,677		
PM ₁₀ Lag2	-0,006	0,159	-0,015	0,003		
Temperatura Média Lag1	0,176	<0,001	0,111	0,241		
Umidade Lag1	0,035	<0,001	0,015	0,055		

Tabela 25 - Modelo Binomial Negativa Com PM₁₀ lag3 para o sintoma Coriza

	Coeficientes	P-Valor	IC 95%		Parâmetro de Dispersão	AIC
Intercepto	-7,117	<0,001	-9,104	-5,129	1,408	1221,700
Sexo	0,046	0,739	-0,224	0,316		
Idade	-0,112	<0,001	-0,167	-0,056		
Asma	0,377	0,025	0,047	0,708		
PM ₁₀ Lag3	-0,013	<0,001	-0,020	-0,006		
Temperatura Média Lag1	0,173	<0,001	0,108	0,237		
Umidade Lag1	0,044	<0,001	0,028	0,061		

Tabela 26 - Modelo de Poisson Com PM₁₀ lag1 para o sintoma Coriza

	Coeficientes	P-Valor	IC 95%		Parâmetro de Dispersão	AIC
Intercepto	-4,578	<0,001	-5,209	-3,947	1	2263,900
Sexo	0,030	0,480	-0,054	0,115		
Idade	-0,131	<0,001	-0,150	-0,113		
Asma	0,389	<0,001	0,296	0,482		
PM ₁₀ Lag1	-0,005	<0,001	-0,008	-0,002		
Temperatura Média Lag1	0,103	<0,001	0,083	0,124		
Umidade Lag1	0,025	<0,001	0,019	0,031		

Tabela 27 - Modelo de Poisson Com PM₁₀ lag2 para o sintoma Coriza

	Coeficientes	P-Valor	IC 95%		Parâmetro de Dispersão	AIC
Intercepto	-4,578	<0,001	-5,206	-3,951	1	2268,300
Sexo	0,035	0,419	-0,050	0,119		
Idade	-0,130	<0,001	-0,149	-0,112		
Asma	0,386	<0,001	0,293	0,479		
PM ₁₀ Lag2	-0,004	0,002	-0,007	-0,002		
Temperatura Média Lag1	0,104	<0,001	0,083	0,125		
Umidade Lag1	0,024	<0,001	0,018	0,030		

Tabela 28 - Modelo de Poisson Com PM₁₀ lag3 para o sintoma Coriza

	Coeficientes	P-Valor	IC 95%		Parâmetro de Dispersão	AIC
Intercepto	-4,520	<0,001	-5,138	-3,902	1	2240,000
Sexo	0,027	0,531	-0,057	0,111		
Idade	-0,131	<0,001	-0,150	-0,112		
Asma	0,408	<0,001	0,314	0,501		
PM ₁₀ Lag3	-0,007	<0,001	-0,009	-0,005		
Temperatura Média Lag1	0,103	<0,001	0,083	0,123		
Umidade Lag1	0,027	<0,001	0,022	0,032		

Tabela 29 - Modelo ZIP Com PM₁₀ lag1 para o sintoma Coriza

	Coeficientes	P-Valor	IC 95%		Parâmetro de Dispersão	AIC
Intercepto	-4,577	<0,001	-5,126	-4,028	1	2265,943
Sexo	0,029	0,039	0,002	0,057		
Idade	-0,131	<0,001	-0,138	-0,125		
Asma	0,389	<0,001	0,359	0,419		
PM ₁₀ Lag1	-0,005	<0,001	-0,006	-0,004		
Temperatura Média Lag1	0,103	<0,001	0,085	0,122		
Umidade Lag1	0,025	<0,001	0,023	0,028		

Tabela 30 - Modelo ZIP Com PM₁₀ lag2 para o sintoma Coriza

	Coeficientes	P-Valor	IC 95%		Parâmetro de Dispersão	AIC
Intercepto	-4,578	<0,001	-5,144	-4,012	1	2270,330
Sexo	0,034	0,016	0,007	0,062		
Idade	-0,130	<0,001	-0,137	-0,124		
Asma	0,386	<0,001	0,357	0,416		
PM ₁₀ Lag2	-0,004	<0,001	-0,005	-0,003		
Temperatura Média Lag1	0,104	<0,001	0,085	0,123		
Umidade Lag1	0,024	<0,001	0,022	0,026		

Tabela 31 - Modelo ZIP Com PM₁₀ lag3 para o sintoma Coriza

	Coeficientes	P-Valor	IC 95%		Parâmetro de Dispersão	AIC
Intercepto	-4,521	<0,001	-5,088	-3,954	1	2242,050
Sexo	0,027	0,059	-0,001	0,055		
Idade	-0,131	<0,001	-0,137	-0,124		
Asma	0,407	<0,001	0,377	0,438		
PM ₁₀ Lag3	-0,007	<0,001	-0,008	-0,006		
Temperatura Média Lag1	0,103	<0,001	0,084	0,122		
Umidade Lag1	0,027	<0,001	0,025	0,029		

Tabela 32 - Modelo ZINB Com PM₁₀ lag1 para o sintoma Coriza

	Coeficientes	P-Valor	IC 95%		Parâmetro de Dispersão	AIC
Intercepto	-6,954	0,017	-12,595	-1,312	1	1228,850
Sexo	0,018	0,900	-0,262	0,298		
Idade	-0,119	<0,001	-0,182	-0,057		
Asma	0,334	0,081	-0,040	0,708		
PM ₁₀ Lag1	-0,008	0,117	-0,018	0,002		
Temperatura Média Lag1	0,174	0,065	-0,010	0,358		
Umidade Lag1	0,037	0,004	0,012	0,062		

Tabela 33 - Modelo ZINB Com PM₁₀ lag2 para o sintoma Coriza

	Coeficientes	P-Valor	IC 95%		Parâmetro de Dispersão	AIC
Intercepto	-6,997	0,017	-12,676	-1,317		
Sexo	0,037	0,797	-0,247	0,321		
Idade	-0,117	<0,001	-0,180	-0,054		
Asma	0,345	0,072	-0,029	0,720		
PM ₁₀ Lag2	-0,006	0,242	-0,017	0,004	1	1229,176
Temperatura						
Média Lag1	0,175	0,068	-0,012	0,361		
Umidade Lag1	0,035	0,008	0,009	0,061		

Anexo B – Gráfico dos resíduos

Sintoma Tosse

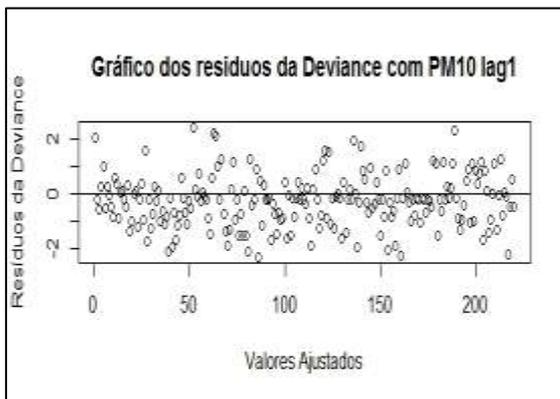


Figura 5: Gráfico dos resíduos da Deviance do Modelo de Regressão Binomial Negativa com PM10 lag1 para o sintoma Tosse.

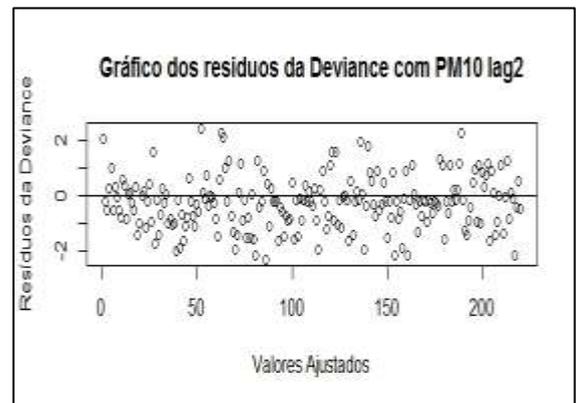


Figura 6: Gráfico dos resíduos da Deviance do Modelo de Regressão Binomial Negativa com PM10 lag2 para o sintoma Tosse.

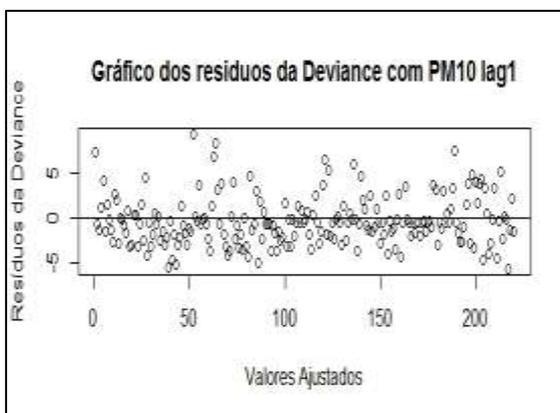


Figura 7: Gráfico dos resíduos da Deviance do Modelo de Regressão de Poisson com PM10 lag1 para o sintoma Tosse.

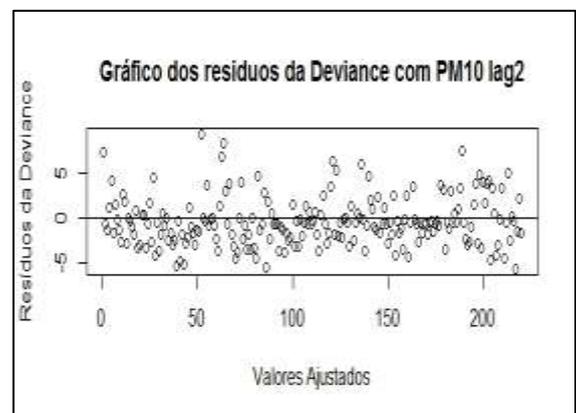


Figura 8: Gráfico dos resíduos da Deviance do Modelo de Regressão de Poisson com PM10 lag2 para o sintoma Tosse.

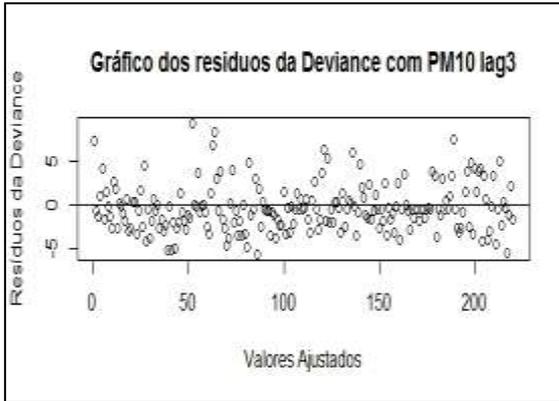


Figura 9: Gráfico dos resíduos da Deviance do Modelo de Regressão de Poisson com PM10 lag3 para o sintoma Tosse.

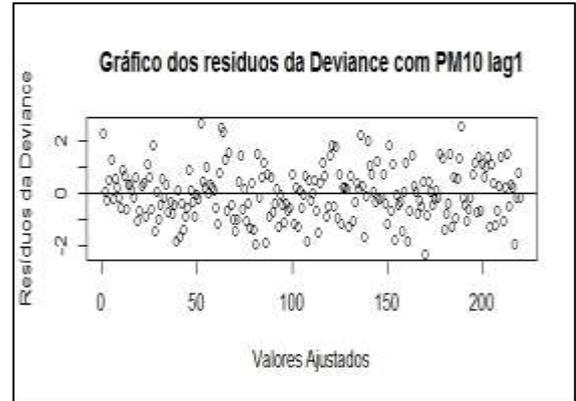


Figura 10: Gráfico dos resíduos da Deviance do Modelo de Regressão ZINB com PM10 lag1 para o sintoma Tosse.

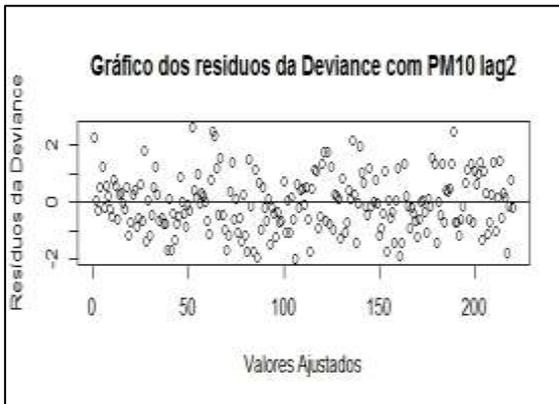


Figura 11: Gráfico dos resíduos da Deviance do Modelo de Regressão ZINB com PM10 lag2 para o sintoma Tosse.

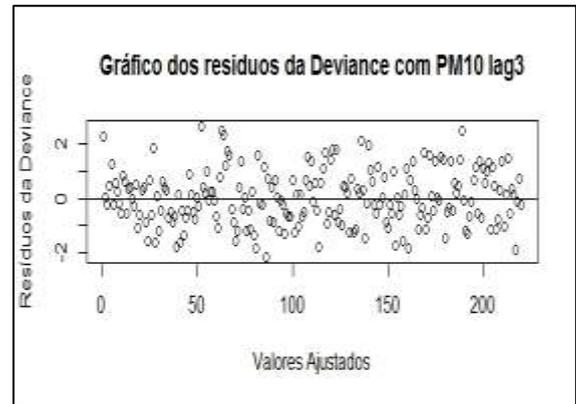


Figura 12: Gráfico dos resíduos da Deviance do Modelo de Regressão ZINB com PM10 lag3 para o sintoma Tosse.

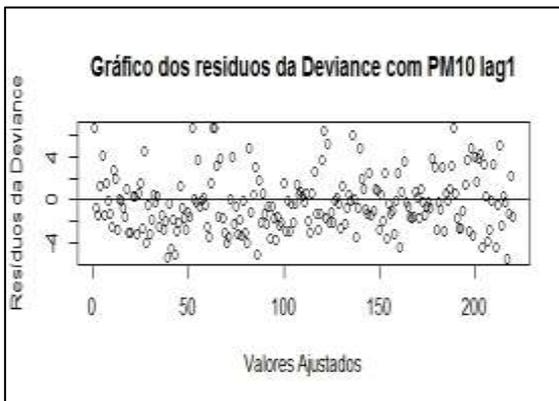


Figura 13: Gráfico dos resíduos da Deviance do Modelo de Regressão ZIP com PM10 lag1 para o sintoma Tosse.

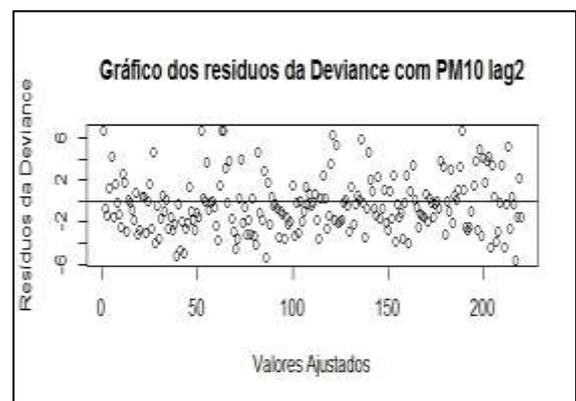


Figura 14: Gráfico dos resíduos da Deviance do Modelo de Regressão ZIP com PM10 lag2 para o sintoma Tosse.

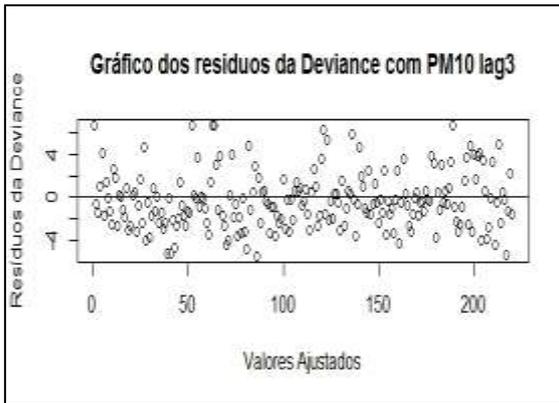


Figura 15: Gráfico dos resíduos da Deviance do Modelo de Regressão ZIP com PM10 lag3 para o sintoma Tosse.

Sintoma Coriza

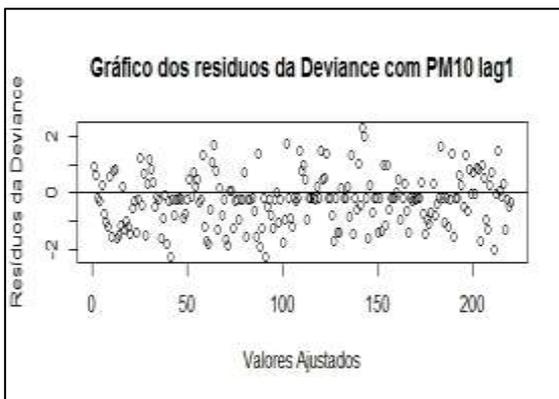


Figura 16: Gráfico dos resíduos da Deviance do Modelo de Regressão Binomial Negativa com PM10 lag1 para o sintoma Coriza.

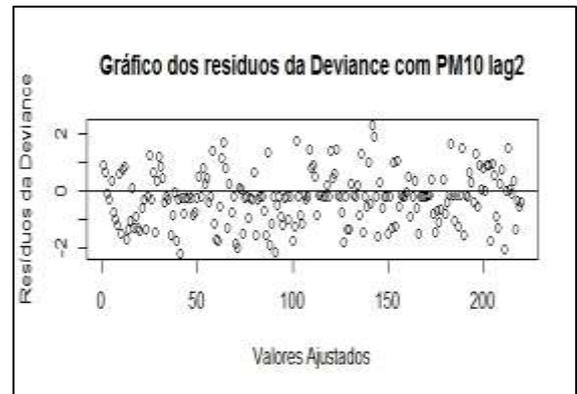


Figura 17: Gráfico dos resíduos da Deviance do Modelo de Regressão Binomial Negativo com PM10 lag2 para o sintoma Coriza.

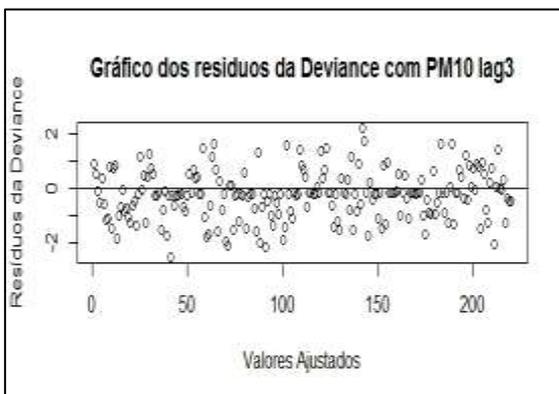


Figura 18: Gráfico dos resíduos da Deviance do Modelo de Regressão Binomial Negativa com PM10 lag3 para o sintoma Coriza.

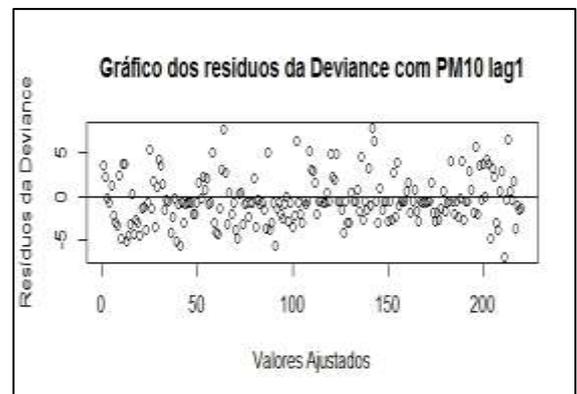


Figura 19: Gráfico dos resíduos da Deviance do Modelo de Regressão de Poisson com PM10 lag1 para o sintoma Coriza.

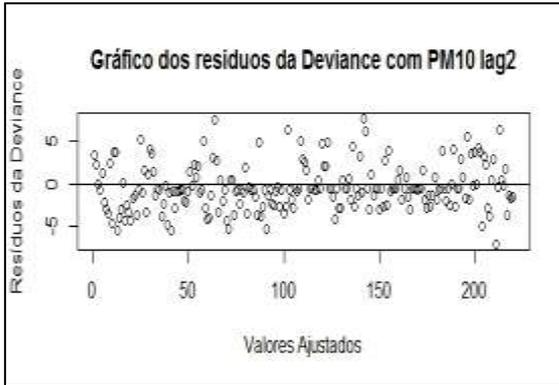


Figura 20: Gráfico dos resíduos da Deviance do Modelo de Regressão de Poisson com PM10 lag2 para o sintoma Coriza.

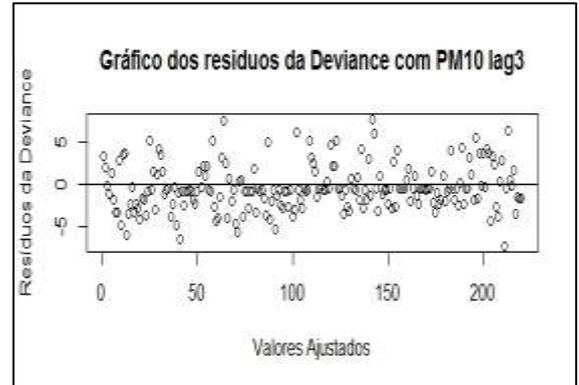


Figura 21: Gráfico dos resíduos da Deviance do Modelo de Regressão de Poisson com PM10 lag3 para o sintoma Coriza.

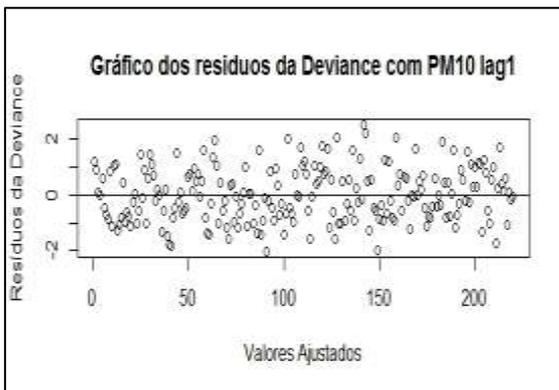


Figura 22: Gráfico dos resíduos da Deviance do Modelo de Regressão ZINB com PM10 lag1 para o sintoma Coriza.

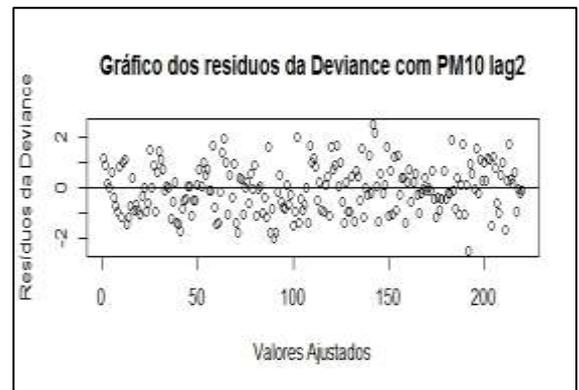


Figura 23: Gráfico dos resíduos da Deviance do Modelo de Regressão ZINB com PM10 lag2 para o sintoma Coriza.

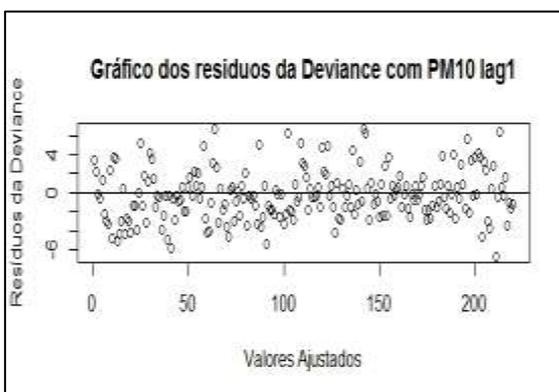


Figura 24: Gráfico dos resíduos da Deviance do Modelo de Regressão ZIP com PM10 lag1 para o sintoma Coriza.

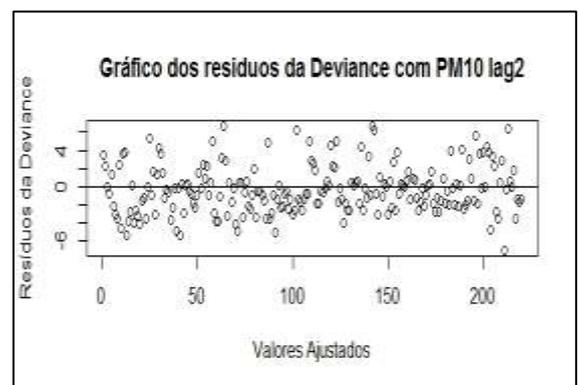


Figura 25: Gráfico dos resíduos da Deviance do Modelo de Regressão ZIP com PM10 lag2 para o sintoma Coriza.

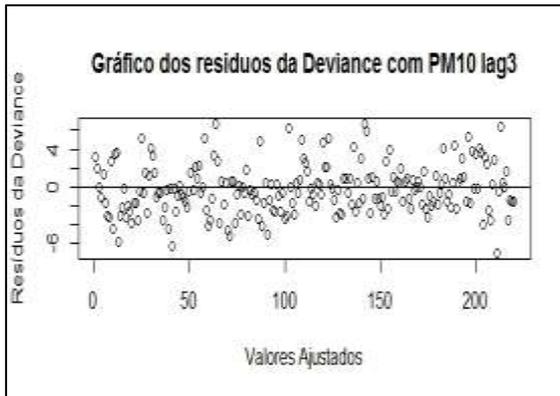


Figura 26: Gráfico dos resíduos da Deviance do Modelo de Regressão ZIP com PM10 lag3 para o sintoma Coriza.