

RAE – CEA – 17P12

RELATÓRIO DE ANÁLISE ESTATÍSTICA SOBRE O PROJETO:

“Clima e correlações com a morbidade respiratória em crianças na cidade de São Paulo”

Camila Yukie Takiyama

Fernando Silveira Marques

Prof. Dr. Alexandre Galvão Patriota

Prof^a. Dr^a. Elisete da Conceição Quintaneiro Aubin

São Paulo

2017

CENTRO DE ESTATÍSTICA APLICADA – CEA – USP
RELATÓRIO DE ANÁLISE ESTATÍSTICA

TÍTULO DO PROJETO: Relatório de Análise Estatística sobre o Projeto: “Clima e correlações com a morbidade respiratória em crianças na cidade de São Paulo”

PESQUISADORES: Sara Lopes de Moraes
Prof. Dr. Emerson Galvani

INSTITUIÇÃO: Faculdade de Filosofia, Letras e Ciências Humanas (FFLCH) - USP

RESPONSÁVEIS PELA ANÁLISE: Camila Y. Takiyama
Fernando S. Marques
Prof. Dr. Alexandre G. Patriota
Prof^a. Dr^a. Elisete C. Q. Aubin

REFERÊNCIA DESTE TRABALHO: PATRIOTA, A.G., AUBIN, E.C.Q., TAKIYAMA, C.Y., MARQUES, F.S. (2017). **Relatório de Análise Estatística sobre o Projeto: “Clima e correlações com a morbidade respiratória em crianças na cidade de São Paulo”**. São Paulo, IME-USP, RAE-CEA-17P12.

FICHA TÉCNICA

REFERÊNCIAS BIBLIOGRÁFICAS

AGRESTI, A. (2002). **Categorical Data Analysis**. 2ª Edição, John Wiley & Sons, Inc.

DATASUS (2014). **Indicadores de morbidade e fatores de risco**. Disponível em: <http://tabnet.datasus.gov.br/cgi/idb2000/fqd14.htm>. Acesso em 9 de junho de 2017.

FOX, J. (2002). **An R and S-Plus Companion to Applied Regression**. 1ª Edição, Sage.

JOHNSON, R.A. e WICHERN, D.W. (2008). **Applied Multivariate Statistical Analysis**. 6ª Edição, Pearson.

KUTNER, M.H., NACHTSHEIM, C.J., NETER, J. e LI, W. (2004). **Applied linear statistical models**. 5ª Edição, Boston: McGraw-Hill Irwin.

MONTEIRO, C.A.F. (1976). **Teoria e clima urbano**. São Paulo. 181p. Tese (Livre Docência em Geografia). Faculdade de Filosofia, Letras e Ciências Humanas - USP.

MONTGOMERY, D.C., PECK, E.A., VINING, G.G. (2012). **Introduction to Linear Regression Analysis**. 5ª Edição, Wiley Series in Probability and Statistics.

PAULA, G.A. (2013). **Modelos de Regressão com Apoio Computacional**. São Paulo. Instituto de Matemática e Estatística - USP.

PROGRAMAS COMPUTACIONAIS UTILIZADOS

Microsoft Office Excel for Windows, versão 2013

Microsoft Office Word for Windows, versão 2013

Software R for Windows, versão 3.3.3

Software R for Linux, versão 3.3.3

Google Docs

TÉCNICAS ESTATÍSTICAS UTILIZADAS (classificação ISI entre parênteses)

Análise Descritiva Unidimensional (03:010)

Análise Descritiva Multidimensional (03:020)

Análise de Associação e Dependência de Dados Quantitativos (06:010)

Análise de Componentes Principais (06:070)

Análise de Regressão Clássica (07:020)

ÁREA DE APLICAÇÃO

Outros (14:990)

Sumário

Resumo	7
1. Introdução.....	10
2. Objetivos.....	10
3. Descrição do estudo	11
4. Descrição das variáveis.....	12
4.1. Variável resposta.....	12
4.2. Variáveis explicativas	13
4.2.1. Variáveis quantitativas.....	13
4.2.2. Variáveis qualitativas	13
5. Análise descritiva.....	14
5.1. Variáveis qualitativas	15
5.2. Variáveis quantitativas.....	16
5.3. Da exposição à internação	18
5.3.1. Lags.....	19
5.3.2. Exposição crônica	19
6. Análise inferencial	21
6.1 Multicolinearidade.....	21
6.2. Modelo de regressão de componentes principais	23
6.2.1 Discussão dos resultados.....	24
6.3. Modelos de regressão com base em variáveis selecionadas.....	25
6.3.1. Verão	26
6.3.1.1 Interpretação dos coeficientes.....	27
6.3.2. Outono.....	29
6.3.2.1 Interpretação dos coeficientes.....	29
6.3.3. Inverno.....	32
6.3.3.1 Interpretação dos coeficientes.....	32
6.3.4. Primavera	35
6.3.4.1 Interpretação dos coeficientes.....	36
7. Conclusões.....	38

APÊNDICE A - Tabelas 40

APÊNDICE B - Gráficos 49

APÊNDICE C - Modelo..... 66

Resumo

Com o crescimento das cidades, a dinâmica atmosférica e os tipos de tempo no espaço urbano começaram a se modificar. Essas alterações climáticas acabam gerando ambientes confortáveis e desconfortáveis. Diversos atributos climáticos, como a temperatura do ar, velocidade do vento, umidade relativa, radiação solar, além da emissão excessiva de poluentes, influenciam no conforto térmico, no bem-estar e na saúde do homem. Sabe-se, por exemplo, que determinadas condições climáticas favorecem o agravamento de doenças do aparelho respiratório, as quais têm grande impacto no Brasil, em especial nos grupos mais suscetíveis (crianças e idosos), sendo a segunda maior causa de internações nos hospitais do Sistema Único de Saúde (SUS).

Para melhorar a qualidade de vida e o bem-estar da população, este estudo visa verificar a influência do clima e da poluição na morbidade respiratória em crianças na Cidade de São Paulo.

O total de internações diário de crianças em hospitais públicos, no período de 2003 a 2013, foi utilizado para representar uma medida de morbidade respiratória. Tais internações foram divididas por tipos de doenças do aparelho respiratório que eventualmente estejam relacionadas com o clima. Os dados de clima coletados na região próxima de residência dessas crianças foram a precipitação, a temperatura, a umidade relativa e a velocidade do vento. A concentração de material particulado inalável no ar foi considerada como medida do parâmetro de poluição.

As análises descritivas mostram que o outono é a estação com o maior número de registros de internações. Verificou-se também que o total de internações diárias parece estar associado ao aumento da concentração de material particulado, e o aumento da precipitação ou da velocidade do vento, aparentemente, ocasiona uma diminuição no total dessas internações. As análises sugerem também que o número de internações observado em um dia pode estar relacionado tanto à poluição e ao clima do referido dia, como dos dias anteriores.

Para a análise inferencial, foram ajustados modelos de regressão com resposta binomial negativa. A fim de contornar o problema de multicolinearidade entre as variáveis explicativas, foram propostas duas estratégias: regressão de componentes principais e seleção de variáveis.

Na regressão de componentes principais, observou-se que segunda-feira e a estação outono estão associados a um aumento na média do total de internações, e que as componentes principais têm importância menor na média do total de internações, quando comparadas com as variáveis qualitativas. Além disso, a variável quantitativa que mais contribuiu com o aumento no total de internações foi a concentração de MP10 média. Uma vez que esse modelo não permite a interpretação dos coeficientes, foram ajustados modelos por estação, a partir da seleção de variáveis.

Para todos os modelos de variáveis selecionadas, assim como no modelo de componentes principais, observou-se um aumento na média do total de internações para todos os dias da semana em relação a domingo, além de um aumento na média do total de internações para dias normais, em relação a feriados durante as estações verão e inverno. Observou-se também uma diminuição na média do total de internações para dias de onda de calor durante o outono em relação a dias sem onda de calor, e para dias de onda de frio durante o inverno em relação a dias sem onda de frio.

Para as estações verão, outono e inverno, o aumento da concentração de MP10 média, e/ou da média de concentração de MP10 média dos últimos 7 dias ocasionou um aumento na média do total de internações. Já para a primavera, observou-se um efeito contrário, ou seja, o aumento da concentração de MP10 média ocasionou uma diminuição na média do total de internações.

Observou-se também que um aumento na precipitação durante o verão e durante o outono ocasionou uma diminuição na média do total de internações. Já um aumento na média de precipitação dos últimos 9 dias ocasionou um aumento na média do total de internações durante o inverno.

Durante a primavera, com o aumento na temperatura efetiva, observou-se um aumento na média do total de internações. A temperatura efetiva observada 8 dias antes, no entanto, tem efeito contrário, diminuindo a média do total de internações.

Algumas variáveis quantitativas apresentaram coeficientes não esperados pela expertise dos pesquisadores. Os ajustes dos modelos se mostraram adequados e contribuem na identificação e mensuração da influência de condições climáticas e de qualidade de ar no total de internações por doenças respiratórias.

1. Introdução

O clima e os diversos tipos de tempo exercem grande influência no ambiente físico e nas atividades humanas. O tipo de moradia, alimentação, vestimenta, meios de transporte e a saúde são exemplos da atuação direta ou indireta da variabilidade climática.

A partir da Revolução Industrial na Europa, no século XVIII, foi possível verificar que a dinâmica atmosférica e os tipos de tempo no espaço urbano começaram a se modificar devido ao crescimento das cidades. Com isso, o clima urbano das cidades potencializou as alterações térmicas, pluviométricas, bem como a qualidade do ar, sendo capaz de interferir na organização da vida na cidade e na qualidade de vida dos seus habitantes (MONTEIRO, 1976).

Essas alterações climáticas acabam gerando ambientes confortáveis e desconfortáveis. Diversos atributos climáticos, como a temperatura do ar, velocidade do vento, umidade relativa, radiação solar, além da emissão excessiva de poluentes, influenciam no conforto térmico, no bem-estar e na saúde do homem. Sabe-se, por exemplo, que determinadas condições climáticas favorecem o agravamento de doenças do aparelho respiratório.

Segundo o DATASUS (2014), as doenças respiratórias no Brasil têm grande impacto, em especial nos grupos mais suscetíveis (crianças e idosos), sendo a segunda maior causa de internações nos hospitais do Sistema Único de Saúde (SUS).

Para melhorar a qualidade de vida e o bem-estar da população, este relatório visa estudar, por meio de análises descritivas e inferenciais, a influência do clima na morbidade respiratória em crianças, possibilitando assim a implementação de políticas públicas que previnam e/ou minimizem os efeitos da interferência do homem na dinâmica climática, bem como os impactos do clima na saúde humana.

2. Objetivos

Este trabalho tem como finalidade:

- Compreender como a dinâmica climática e os tipos de tempo estão associados às internações hospitalares por doenças respiratórias em crianças menores de 10 anos de idade em 14 distritos da Cidade de São Paulo;
- Verificar a relação entre a poluição atmosférica pela emissão de material particulado de até 10µm de diâmetro (MP10) e a qualidade do ar com a morbidade respiratória em crianças;
- Encontrar limites específicos de temperatura, precipitação, umidade relativa e faixa de conforto térmico que ocasionam o aumento ou diminuição das internações por doenças respiratórias.

3. Descrição do estudo

A amostra é composta por observações diárias do clima, da poluição e do número de internações, coletadas de 01 de janeiro de 2003 a 31 de dezembro de 2013, totalizando 4018 dias.

Como discutido anteriormente, foi considerado o número de internações diárias nos hospitais da rede pública por doenças respiratórias que pudessem estar relacionadas ao clima. As doenças respiratórias escolhidas para este estudo e seus respectivos códigos da Classificação Internacional de Doenças (CID 10) foram:

- Infecções agudas das vias aéreas superiores (J00-J06);
- Influenza (gripe) e pneumonia (J09-J18);
- Outras infecções agudas das vias aéreas inferiores (J20-J22);
- Outras doenças das vias aéreas superiores (J30-J39, exceto J33 a J39);
- Doenças crônicas das vias aéreas inferiores (J40-J47).

As doenças classificadas J33 a J39, J60-J70 (doenças pulmonares devidas a agentes externos), J80-J84 (outras doenças respiratórias que afetam principalmente o interstício), J85-J86 (afecções necróticas e supurativas das vias aéreas inferiores), J90-J94 (outras doenças da pleura) e J95-J99 (outras doenças do aparelho respiratório) não foram consideradas, pois não são grupos relevantes ou influenciados diretamente pela condição atmosférica.

Os dados foram obtidos por meio do banco de dados do Sistema de Informações Hospitalares do Sistema Único de Saúde (SIH/SUS), disponíveis no website do DATASUS por código de endereçamento postal (CEP). Esses dados são registros das Autorizações de Internação Hospitalar (AIH) pagas pelo SUS, contendo: Cadastro do Contribuinte (CGC) do hospital, cidade, dados do paciente, como idade, sexo, CEP de residência, data de entrada e saída de internação, diagnóstico principal, entre outras informações. Apenas crianças menores de 10 anos foram consideradas no estudo, por não apresentarem hábitos que pudessem contribuir com o surgimento ou o agravamento de doenças do aparelho respiratório. Na data de coleta dos dados, essas crianças residiam nos distritos Brás, Belém, Bom Retiro, Casa Verde, Limão, Pari, República, Santa Cecília, Santana, Sé, Tucuruvi, Vila Guilherme, Vila Maria e Vila Medeiros, na Cidade de São Paulo.

Os dados climáticos da região considerados no estudo foram fornecidos pela estação meteorológica do Instituto Nacional de Meteorologia – INMET.

Os índices de conforto térmico foram obtidos por meio de cálculos utilizando as variáveis climáticas.

Como medida de poluição, foi considerada a concentração média diária de material particulado com até 10µm de diâmetro (MP10), por ser inalável e, portanto, apresentar riscos à saúde. Os dados de qualidade do ar e do poluente MP10 foram disponibilizados pela Companhia Ambiental de São Paulo - CETESB.

4. Descrição das variáveis

Variáveis de interesse em um experimento são chamadas de variáveis-resposta. As variáveis que podem afetar a resposta são chamadas de variáveis explicativas.

4.1. Variável resposta

Neste estudo, a variável resposta é *Total de Internações* diário, que é o número de internações de crianças menores de 10 anos de idade, registradas em hospitais públicos da Cidade de São Paulo por doenças respiratórias em um determinado dia.

4.2. Variáveis explicativas

As variáveis explicativas podem ser classificadas em dois tipos: quantitativas ou qualitativas.

4.2.1. Variáveis quantitativas

As variáveis quantitativas apresentam como possíveis realizações números resultantes de uma contagem ou mensuração. As variáveis quantitativas deste estudo são descritas abaixo:

- Precipitação diária (mm);
- Temperatura máxima diária (°C);
- Temperatura mínima diária (°C);
- Temperatura média diária (°C);
- Amplitude térmica diária (°C);
- Umidade relativa média diária (%);
- Velocidade do vento média diária (m/s);
- TE diária (°C): temperatura efetiva – índice diário de conforto térmico, considerando-se a temperatura média e a umidade relativa média no dia;
- Tev diária (°C): temperatura efetiva com o vento – índice diário de conforto térmico, considerando-se a temperatura média, a umidade relativa média, e a velocidade média do vento no dia;
- MP10 média diária ($\mu\text{g}/\text{m}^3$): média diária da concentração de material particulado de até $10\mu\text{m}$ de diâmetro.

4.2.2. Variáveis qualitativas

As variáveis qualitativas apresentam como possíveis realizações uma qualidade ou categoria. Neste estudo, algumas variáveis qualitativas são decorrentes da categorização em intervalos de variáveis quantitativas. As variáveis qualitativas deste estudo são descritas abaixo:

- Data: dia, mês e ano da observação, abrangendo o período de 01/01/2003 a 31/12/2013;
- Dia da semana na data: domingo, segunda-feira, terça-feira, quarta-feira, quinta-feira, sexta-feira ou sábado;
- Feriado: dia é feriado (sim ou não);
- Estação do ano na data: verão, outono, inverno ou primavera;
- Onda de frio: dia faz parte de uma onda de frio (sim ou não);
- Onda de calor: dia faz parte de uma onda de calor (sim ou não);
- Faixa de precipitação diária: dias sem chuva < 1,0 mm; chuva entre 1,1 mm e 10,0 mm; chuva entre 10,1 mm e 30,0 mm; chuva entre 30,1 mm e 50,0 mm, chuva entre 50,1 mm e 80,0 mm; chuva > 80 mm;
- Faixa de conforto TE diária: muito frio ($TE \leq 7,9^{\circ}\text{C}$), frio ($7,9^{\circ}\text{C} < TE \leq 17,0^{\circ}\text{C}$), um pouco de frio ($17,0^{\circ}\text{C} < TE \leq 21,0^{\circ}\text{C}$), nem frio nem calor ($21,0^{\circ}\text{C} < TE \leq 25,4^{\circ}\text{C}$), um pouco de calor ($25,4^{\circ}\text{C} < TE \leq 26,9^{\circ}\text{C}$), calor ($26,9^{\circ}\text{C} < TE \leq 33,9^{\circ}\text{C}$) ou muito calor ($TE > 33,9^{\circ}\text{C}$);
- Faixa de conforto T_{ev} diária: muito frio ($T_{ev} \leq 5^{\circ}\text{C}$), frio ($5^{\circ}\text{C} < T_{ev} \leq 10^{\circ}\text{C}$), moderadamente frio ($10^{\circ}\text{C} < T_{ev} \leq 13^{\circ}\text{C}$), ligeiramente frio ($13^{\circ}\text{C} < T_{ev} \leq 16^{\circ}\text{C}$), pouco frio ($16^{\circ}\text{C} < T_{ev} \leq 19^{\circ}\text{C}$), ligeiramente fresco ($19^{\circ}\text{C} < T_{ev} \leq 22^{\circ}\text{C}$), confortável ($22^{\circ}\text{C} < T_{ev} \leq 25^{\circ}\text{C}$), ligeiramente quente ($25^{\circ}\text{C} < T_{ev} \leq 28^{\circ}\text{C}$), quente moderado ($28^{\circ}\text{C} < T_{ev} \leq 31^{\circ}\text{C}$), quente ($31^{\circ}\text{C} < T_{ev} \leq 34^{\circ}\text{C}$) ou muito quente ($T_{ev} > 34^{\circ}\text{C}$);
- Qualidade do ar para MP10 diária: boa ($0 \mu\text{g}/\text{m}^3 \leq \text{MP10 média} \leq 50 \mu\text{g}/\text{m}^3$), moderada ($50 \mu\text{g}/\text{m}^3 < \text{MP10 média} \leq 100 \mu\text{g}/\text{m}^3$), ruim ($100 \mu\text{g}/\text{m}^3 < \text{MP10 média} \leq 150 \mu\text{g}/\text{m}^3$), muito ruim ($150 \mu\text{g}/\text{m}^3 < \text{MP10 média} \leq 250 \mu\text{g}/\text{m}^3$) ou péssima ($\text{MP10 média} > 250 \mu\text{g}/\text{m}^3$).

5. Análise descritiva

A base de dados, composta por 4018 registros, possui 14 dados faltantes com relação à concentração de material particulado (MP10). Portanto, a medida de qualidade do ar, baseada nessa variável, também estará ausente. Para as análises que

envolverem essas variáveis, será considerada a base de dados sem essas observações, ou seja, a base composta por 4004 registros.

5.1. Variáveis qualitativas

Algumas categorias das variáveis *Faixa de Conforto TE*, *Faixa de Conforto Tev* e *Qualidade do Ar para MP10* não foram observadas. Para a variável *Faixa de Conforto TE* não foram observados dias de pouco frio, calor ou muito calor, e foi registrado apenas um dia de muito frio. Para a variável *Faixa de Conforto Tev* foi registrado apenas um dia ligeiramente quente, e não foram registrados dias quentes moderados, quentes ou muito quentes. Já para a variável *Qualidade do Ar para MP10*, foram registrados apenas quatro dias de qualidade do ar muito ruim, e nenhum dia de qualidade do ar péssima.

Para realizar a análise, algumas categorias pouco observadas das variáveis *Faixa de Conforto TE*, *Faixa de Conforto Tev* e *Qualidade do Ar para MP10* foram reagrupadas. As categorias não observadas foram desconsideradas do estudo. O Gráfico B.1 mostra as frequências absolutas das variáveis qualitativas reagrupadas.

Pelo Gráfico B.1, na maioria dos dias (aproximadamente 78%), a qualidade do ar foi boa. Observou-se também que em 2860 dias (aproximadamente 71%) não choveu, e em apenas 11 dias (dos 4018 dias) a taxa de precipitação superou 80 mm.

No Gráfico B.2, temos os boxplots do total de internações segundo as variáveis qualitativas. O gráfico sugere um maior número de internações na segunda-feira e terça-feira, e um menor número de internações durante o fim de semana. Com relação às estações, nota-se um maior número de internações durante o outono, e o gráfico sugere ainda que o número de internações aumenta com a piora da qualidade do ar.

Para faixas de conforto Tev de muito frio a pouco frio, parece que as medianas do total de internações diárias são aproximadamente iguais, enquanto que para as faixas ligeiramente fresco e confortável, ocorre uma queda na mediana. Observa-se alguns dias sem chuva com um número grande de internações (“outliers”). O mesmo ocorre em dias sem onda de calor ou frio, e os totais de internações são semelhantes em dias de ondas de calor ou de frio.

Nos Gráficos B.3 e B.4, apresentamos os boxplots do total de internações segundo algumas variáveis qualitativas, por ano. Observa-se que os anos de 2005 e 2006 não apresentaram ondas de frio, nem de calor. Em geral, a variabilidade do número de internações é maior para dias sem ondas de frio nem de calor. No Gráfico B.4, nota-se que em 2003 e em 2007, a piora na qualidade do ar aparentemente ocasionou um aumento no total de internações. Além disso, os dias com maiores números de internações aconteceram em 2003.

5.2. Variáveis quantitativas

A Tabela A.1 e os Histogramas apresentados no Gráfico B.5 apresentam medidas descritivas e distribuição das variáveis quantitativas presentes no estudo, respectivamente. Temos que o total de internações diárias variou de 0 a 17, com mediana de 4 internações por dia. Além disso, as variáveis *Total de Internações*, *Precipitação* e *MP10 Média* apresentam uma distribuição assimétrica à direita, o que se nota pelas caudas pesadas à direita presentes nos histogramas. Analogamente, as variáveis *Temperatura Máxima*, *Temperatura Mínima*, *Temperatura Média*, *Umidade Relativa Média*, *TE* e *Tev* apresentam distribuição assimétrica à esquerda.

As frequências absolutas das internações por estação são dadas na Tabela 5.2.1.

Tabela 5.2.1. Frequências absolutas das internações por estação

	Total de Internações
Verão	2781
Outono	6009
Inverno	4707
Primavera	4020

Para extrair mais algumas informações a respeito do que torna o outono a estação com maior número de casos, construímos boxplots das variáveis quantitativas segundo as estações. Os resultados encontram-se no Gráfico B.6.

Os boxplots por si só não permitem extrair informações a respeito das características do outono que poderiam elevar o número de internações. No entanto,

observa-se que os maiores valores de concentrações de MP10 média (“outliers”) aconteceram durante o outono. Além disso, é possível que combinações entre as variáveis criem uma condição favorável para o surgimento ou o agravamento de doenças do aparelho respiratório.

Para avaliar a influência das variáveis quantitativas no total de internações diárias, apresentamos os gráficos de dispersão (Gráfico B.7), que sugerem que algumas variáveis podem influenciar fracamente no número de internações. O aumento do número de internações parece estar associado a um aumento da concentração de MP10 no ar, o que estaria de acordo com a hipótese de que o material particulado é prejudicial à saúde. Os gráficos sugerem também que o aumento da velocidade do vento ou da precipitação ocasiona uma diminuição no número de internações diárias. Supõe-se que o vento e a chuva ajudam a diminuir a concentração de poluentes no ar, melhorando a qualidade do ar.

Com o intuito de avaliar a associação entre as variáveis quantitativas, é apresentada a matriz de correlação de Spearman no Gráfico B.8. O coeficiente de Spearman ρ é uma medida de correlação não-paramétrica que mede a força e direção de associação monotônica entre duas variáveis ordinais, sem fazer suposições sobre suas distribuições. Diferentemente do coeficiente de Pearson, o coeficiente de Spearman não requer a suposição de que a relação entre as variáveis seja linear, sendo menos restritivo. Suas interpretações, no entanto, são semelhantes. Quanto mais próximo de ± 1 , mais forte é a associação monotônica entre as variáveis (KUTNER et al., 2004). Na Tabela A.2, apresentamos um guia para interpretar os valores dos coeficientes do Gráfico B.8.

Observamos algumas correlações positivas muito fortes, entre os índices TE e Tev (0,9), entre o índice TE e as variáveis *Temperatura Média* (1,0), *Temperatura Mínima* (0,9) e *Temperatura Máxima* (0,9) e entre o índice Tev e as variáveis *Temperatura Mínima* (0,8), *Temperatura Média* (0,9) e *Temperatura Máxima* (0,9), além das correlações positivas muito fortes entre a temperatura média e as temperaturas máxima (0,9) e mínima (0,9). Como o cálculo dos índices envolve temperatura, essa correlação era esperada. Da mesma forma, o cálculo da temperatura média envolve as temperaturas máxima e mínima.

As variáveis *Amplitude Térmica* e *Umidade Relativa Média* apresentaram forte correlação negativa (-0,7), ou seja, altos valores de amplitude térmica estão relacionados a baixos valores de umidade relativa média. Além disso, a variável *Amplitude Térmica* se correlaciona forte e positivamente com a variável *MP10 Média* (0,6), indicando que o aumento da amplitude térmica está associado a um aumento na média de material particulado. As variáveis *Precipitação* e *Umidade Relativa Média* (0,5) e *Amplitude Térmica* e *Tev* (0,4) se mostraram positiva e moderadamente correlacionadas, indicando que maiores valores de precipitação estão associados a maiores valores de umidade relativa média, e maiores valores de amplitude térmica estão associados a maiores valores de *Tev*.

A variável *Precipitação* se correlaciona negativa e moderadamente com a variável *Amplitude Térmica* (-0,4), ou seja, maiores valores de precipitação estão associados a menores valores de amplitude térmica. A variável *MP10 Média* apresentou também correlação moderada negativa com as variáveis *Precipitação* (-0,5), *Umidade Relativa Média* (-0,5) e *Velocidade do Vento Média* (-0,4), ou seja, maiores valores de precipitação, umidade relativa média e velocidade do vento média estão associados a menores médias de material particulado. Da mesma forma, as variáveis *Umidade Relativa Média* e *Temperatura Máxima* mostraram-se negativa e moderadamente correlacionadas (-0,5), indicando que maiores valores de temperatura máxima estão associados a baixas umidades relativas médias.

5.3. Da exposição à internação

Nas Seções 5.1 e 5.2, analisamos descritivamente de que forma as variáveis climáticas e de poluição estariam relacionadas com o total de internações do mesmo dia. No entanto, a internação pode não ser imediata, podendo ocorrer alguns dias após a exposição aos fatores de risco (condição climática e qualidade do ar). Em outras palavras, um indivíduo pode ter sido internado em resposta à sua exposição a algum fator de risco, por exemplo, concentração de material particulado (MP10), que ocorreu dias antes.

Para ponderar essa hipótese do ponto de vista descritivo, consideramos dois cenários:

- A manifestação clínica pode surgir após alguns dias da exposição. Esse intervalo de tempo é chamado de lag (atraso);
- A internação é resultado de uma exposição crônica.

Neste estudo, segundo a pesquisadora, supõe-se que as internações possam ocorrer até 28 dias após a exposição aos fatores de risco, e que o número de dias até a internação pode ser diferente para cada fator de risco.

5.3.1. Lags

Supondo-se que exista uma defasagem entre o momento de exposição aos fatores e a manifestação clínica, espera-se que as correlações entre a variável resposta e as variáveis explicativas aumentem com a escolha dos lags mais adequados para cada variável explicativa. Dessa forma, o Gráfico B.9 apresenta os coeficientes de correlação de Spearman entre o total de internações e as variáveis quantitativas, para lags de 0 a 28 dias.

A Tabela A.3 apresenta os lags que resultaram em maiores valores absolutos dos coeficientes de correlação. Para a análise descritiva, consideraremos esses lags como os mais adequados, ou seja, supõe-se que o total de internações de um dia seja influenciado pela precipitação do 4º dia anterior a ele (lag 4), e pela concentração de material particulado do próprio dia (lag 0), por exemplo.

Nota-se pelos gráficos de dispersão (Gráfico B.10) que a inclusão dos lags acentuou a inclinação das retas de tendência, mas a interpretação se manteve. No Gráfico B.11, com a incorporação do lag 4 para a variável *Faixa de Precipitação*, o total de internações parece decrescer com o aumento da precipitação, exceto para a faixa *Chuva > 80*, cuja mediana aumentou para 4 internações por dia. Neste caso, como a categoria apresentou apenas 11 observações com alta variabilidade, talvez não seja comparável com as demais categorias.

5.3.2. Exposição crônica

Supondo-se agora que a internação seja decorrente da exposição em dias consecutivos aos fatores de risco analisados, seria necessária uma medida que

considerasse não só o dia da internação, mas também os j dias que o antecedem. Para isso, calculamos médias móveis no dia t , $m(t)$, das variáveis quantitativas usando a expressão:

$$m(t) = \frac{1}{j+1} \sum_{i=0}^j y(t-i),$$

em que $y(t-i)$ representa o valor observado da variável quantitativa y no dia $(t-i)$, t é a data da internação e $j = 0, \dots, 28$ é o número de dias anteriores ao dia da internação considerados na média móvel.

O Gráfico B.12 mostra que a força de correlação de Spearman varia com o aumento do número de dias anteriores ao dia da internação considerados na média móvel. Em geral, existe um acumulado de número de dias (dentro desse intervalo de 28 dias) no qual as correlações são maiores. A Tabela A.4 apresenta o número de dias anteriores ao dia da internação considerados nas médias móveis que resultaram em maiores valores absolutos dos coeficientes de correlação. Para a análise descritiva, consideraremos esses números como a quantidade de dias mais adequados para serem considerados como efeito de exposição crônica, ou seja, supõe-se que o total de internações de um dia seja influenciado pela precipitação do mesmo dia e dos 9 dias que o antecedem (MA 9), e pela concentração de material particulado do mesmo dia e dos 7 dias que o antecedem (MA 7), por exemplo.

Os gráficos de dispersão (Gráfico B.13) usando a média móvel sugerem que os resultados encontrados sem as médias móveis se mantêm.

Como a variável *Velocidade do Vento Média* apresentou valores muito altos e possivelmente incorretos, para a análise inferencial, optou-se por retirá-los da base de dados, conforme sugestão da pesquisadora de que velocidades acima de 12 m/s podem ser consideradas anormais. Calculando-se novamente os lags e dias anteriores ao dia da internação considerados nas médias móveis que resultam em maiores valores absolutos dos coeficientes de correlação de Spearman, obtivemos uma alteração para o número de dias considerados nas médias móveis de 28 dias para 4 dias.

6. Análise inferencial

Nesta seção, são propostos modelos de regressão para estimar o número de internações diárias por problemas respiratórios (variável resposta) com base nas variáveis climáticas e de poluição (variáveis explicativas), detalhados no Apêndice C.

Antes de tratar dos modelos, faremos algumas considerações sobre multicolinearidade entre as variáveis explicativas. Pelos resultados encontrados na análise descritiva, notamos a existência de correlações muito fortes entre elas, o que pode ser um indício de multicolinearidade, podendo afetar o modelo de regressão a ser estimado.

6.1 Multicolinearidade

A multicolinearidade acontece no caso de uma variável explicativa ser escrita como uma combinação linear de outras variáveis explicativas.

Ela pode ser classificada em dois tipos:

- Estrutural: quando uma variável explicativa é criada a partir de variáveis explicativas já existentes:
 - Amplitude Térmica é função da Temperatura Máxima e da Temperatura Mínima;
 - Temperatura Média é função da Temperatura Máxima e da Temperatura Mínima;
 - TE é função da Temperatura Média e da Umidade Relativa Média;
 - Tev é função da Temperatura Média, da Umidade Relativa Média e da Velocidade do Vento Média.
- Baseada nos dados: comum em estudos observacionais, que não permitem manipular o sistema do qual se coletam os dados e algumas variáveis são correlacionadas naturalmente, por exemplo, umidade relativa do ar e precipitação.

A presença de multicolinearidade produz os seguintes efeitos num modelo de regressão:

- Os coeficientes estimados de preditores se alteram com a presença ou ausência de variáveis correlacionadas. Quanto maior a correlação entre as variáveis preditoras, maior a influência que uma exerce sobre a outra em termos de coeficientes;
- A mesma conclusão para os coeficientes serve para a soma dos quadrados dos resíduos atribuídos a um preditor;
- Quanto mais preditores correlacionados no modelo, menor a precisão dos estimadores de seus coeficientes, pois os erros padrão aumentam, acarretando um aumento dos intervalos de confiança;
- Como consequência direta dos efeitos anteriores, quando as variáveis explicativas são correlacionadas, o teste de hipótese para os coeficientes $\beta_k = 0$ pode produzir resultados diferentes dependendo de quais variáveis correlacionadas estão no modelo, ou seja, algumas variáveis podem ser importantes numa composição do modelo, mas não em outra.

Assim, um dos efeitos indesejados da multicolinearidade é a incapacidade de interpretar as relações estimadas entre as variáveis explicativas e a variável resposta.

Um método formal para detectar a presença de multicolinearidade é o uso do fator de inflação de variância generalizada (Generalized Variance Inflation Factor - GVIF), que mede o quanto a variância do estimador de um coeficiente é inflacionada pela existência de multicolinearidade em modelos lineares generalizados (FOX, 2002). Sua interpretação é semelhante à do fator de inflação da variância (Variance Inflation Factor - VIF). Valores acima de 5 indicam problemas de multicolinearidade (MONTGOMERY et. al., 2012).

Uma alternativa para contornar os efeitos da multicolinearidade é construir “novas” variáveis explicativas, combinações lineares das variáveis explicativas existentes, através de uma transformação ortogonal, utilizando a técnica de análise de componentes principais (ACP), de maneira que essas novas variáveis (chamadas de componentes principais (CPs) sejam linearmente independentes. A ACP procura explicar o máximo da variabilidade dos dados com um número de componentes principais menor ou igual ao número de variáveis explicativas originais, possibilitando a

redução de dimensionalidade. Essa técnica associada aos modelos de regressão é conhecida como Regressão de Componentes Principais.

Outra alternativa para lidar com a multicolinearidade é excluir uma ou mais variáveis explicativas altamente correlacionadas. Para simplificar o método de escolha, o Gráfico B.20 apresenta um esquema para seleção de variáveis, visando eliminar a multicolinearidade estrutural.

6.2. Modelo de regressão de componentes principais

Neste tópico, utilizaremos a ACP padronizada com todas as variáveis quantitativas, incluindo as de lag e MA com base na Tabela A.5 (que apresentaram maior correlação de Spearman).

O resultado da ACP está apresentado nas Tabelas A.6, A.7, A.8. e no Gráfico B.14.

Decidimos usar no ajuste do modelo as 10 primeiras componentes principais, que explicam, aproximadamente 90,21% da variabilidade dos dados. Embora a escolha da quantidade de CPs seja arbitrária, a decisão foi baseada em Johnson e Wichern (2008), ou seja, de 29 componentes utilizaremos 10 para a construção do modelo. Os coeficientes das variáveis explicativas que compõem cada componente principal (CP1, ..., CP10) estão apresentados na Tabela A.9. Para ter uma ideia do peso de cada variável explicativa na componente principal, no Gráfico B.15 está representada a intensidade de cada coeficiente na composição da componente, sendo que quanto mais forte a cor, maior o coeficiente (positivo ou negativo).

Construímos o modelo com as variáveis qualitativas pertinentes e com as 10 primeiras componentes principais.

O resultado do modelo ajustado está na Tabela A.10. e indica que as variáveis *CP7* (sétimo componente principal), *Feriado* e *Onda de Calor* não possuem coeficientes estatisticamente significativos a 5% de significância (valor-p > 0,05).

Realizamos a seleção de variáveis com base no critério de Akaike (*Akaike Information Criterion* - *AIC*), processo de minimização que não envolve testes estatísticos, e seleciona o modelo que apresenta o menor AIC. A ideia básica desse critério é selecionar um modelo que esteja bem ajustado e tenha um número reduzido

de parâmetros (PAULA, 2013). Foram eliminadas as variáveis que não foram estatisticamente significativas no modelo completo e o resultado é apresentado na Tabela A.11 e nos Gráficos B.16 e B.17. Verificamos que o ajuste foi razoável com a análise gráfica dos resíduos (Gráficos B.18 e B.19) e com o teste qui-quadrado de Pearson. Esse teste compara a distribuição de frequências observada com a distribuição determinada pelo modelo ajustado (AGRESTI, 2002). O teste não rejeitou a hipótese nula (valor-p = 0,11) de que o ajuste do modelo é adequado.

6.2.1 Discussão dos resultados

Embora a estratégia adotada nesta seção não permita a interpretação clássica para as variáveis mantidas no modelo, construímos os Gráficos B.16 e B.17, que permitem, através dos padrões de cor, identificar as variáveis quantitativas e qualitativas que mais contribuem com o aumento (vermelho) ou diminuição (azul) no total de internações.

Corroborando com alguns resultados obtidos na análise descritiva, nota-se que:

- Segunda-feira e a estação outono estão associados com um aumento do número de internações (cores vermelhas mais acentuadas);
- Os componentes principais têm importância menor no total de internações, quando comparadas com as variáveis qualitativas (*Dia da Semana*, *Estações*, *Onda de Frio* e *Onda de Calor*). Como as componentes principais são uma composição das variáveis quantitativas, isso implica que as variáveis quantitativas também têm menor influência no total de internações. Este resultado está de acordo com a associação fraca encontrada nas análises de correlação;
- Vale lembrar que a componente principal 10, que possui o maior coeficiente estimado entre as componentes principais, possui como variável mais importante a MP10 média (Gráfico B.15), sendo que na composição das componentes, essa variável quantitativa é a que mais contribui para o aumento do número total de internações (Gráfico B.17), também apontado pela análise de correlação (Gráfico B.8).

Algumas variáveis quantitativas podem apresentar coeficientes que não são os esperados pela expertise dos pesquisadores. A nossa suposição é de que os dados coletados podem não representar perfeitamente as condições climáticas da residência da criança internada e, como consequência, leva os coeficientes das variáveis climáticas a terem menos importância ou até mesmo valores contrários ao esperado pelos pesquisadores.

De qualquer maneira, vale lembrar que o modelo está com ajuste adequado e contribui na identificação e mensuração de condições climáticas e de qualidade de ar que influenciam no total de internações por doenças respiratórias.

6.3. Modelos de regressão com base em variáveis selecionadas

A partir do esquema do Gráfico B.20, foram escolhidos, pela pesquisadora, os seguintes conjuntos de variáveis:

- Conjunto 1: Precipitação, Temperatura Máxima, Temperatura Mínima, Umidade Relativa Média, Velocidade do Vento Média e MP10 Média;
- Conjunto 2: Precipitação, Temperatura Média, Amplitude Térmica, Umidade Relativa Média, Velocidade do Vento Média e MP10 Média;
- Conjunto 3: Precipitação, Temperatura Mínima, TE, Velocidade do Vento Média e MP10;
- Conjunto 4: Precipitação, Temperatura Máxima, TE, Velocidade do Vento Média e MP10.

Ao considerarmos os lags e as médias móveis das variáveis explicativas selecionadas, possivelmente introduzimos uma multicolinearidade não prevista pelo esquema. Dessa forma, para cada conjunto de variáveis selecionadas, incluímos os respectivos lags e médias móveis, e verificamos a existência de multicolinearidade, calculando o GVIF para cada variável explicativa. No caso de conjuntos com valores do GVIF maiores que 5, removemos a variável com o maior valor, e recalculamos os

GVIFs para o conjunto reduzido, procedendo dessa forma, até que os GVIFs de todas as variáveis de cada conjunto estivessem dentro do limite. As variáveis retiradas por esse procedimento para cada conjunto foram:

- Conjunto 1: Temperatura Mínima MA10, Umidade Relativa Média MA6;
- Conjunto 2: Temperatura Média MA14, Umidade Relativa Média MA6;
- Conjunto 3: Temperatura Mínima MA10, TE MA13, Temperatura Mínima;
- Conjunto 4: TE MA13, TE Lag8, TE.

Como a interpretação dos coeficientes de um modelo de regressão é feita observando-se a mudança na resposta a cada unidade de mudança nas observações de uma variável explicativa quando as demais variáveis são mantidas fixas, para cada conjunto de variáveis selecionadas, foram ajustados modelos por estação do ano, já que as condições climáticas diferem para cada estação, como foi observado no Gráfico B.6. A seleção de variáveis foi realizada pelo AIC.

Assumindo que, para cada estação, o total de internações possa ser melhor explicado por um determinado conjunto de variáveis selecionadas, como forma de avaliar a qualidade do ajuste dos 16 modelos, realizamos o teste qui-quadrado de Pearson, cujos valores-p são apresentados na Tabela A.12. Observa-se um melhor ajuste dos modelos das estações outono e inverno para o conjunto 1, e um melhor ajuste dos modelos das estações verão e primavera para o conjunto 3.

Os ajustes desses modelos são apresentados a seguir.

6.3.1. Verão

O modelo ajustado é apresentado na Tabela A.13.

Os Gráficos de diagnóstico B.21 não sugerem afastamento das suposições do modelo. Como ponto potencialmente influente, destaca-se a observação 1512. Trata-se de um feriado com alto número de internações (9), comportamento não esperado de acordo com o modelo. O ajuste do modelo sem essa observação, apresentado na Tabela A.14, mostra uma alteração no coeficiente associado à variável Feriado (de 0,1989 para 0,2777), e a torna estatisticamente significativa ao nível de 5%. Portanto,

para a interpretação dos coeficientes, consideraremos o modelo ajustado sem essa observação.

O Gráfico B.22 mostra o envelope simulado, e indica um bom ajuste do modelo.

6.3.1.1 Interpretação dos coeficientes

Precipitação Lag4:

- O aumento de 1 mm na precipitação de 4 dias antes multiplica por $e^{-0,0037} = 0,9963$ a média do total de internações, mantendo-se todas as outras variáveis constantes, ou seja, há uma diminuição de aproximadamente 0,37% na média do total de internações;
- Como regra geral, aumento de x mm na precipitação de 4 dias antes multiplica por $e^{-0,0037x}$ a média do total de internações, mantendo-se todas as outras variáveis constantes.

MP10 Média MA7:

- O aumento de $1 \mu\text{g}/\text{m}^3$ na média de MP10 Média dos últimos 7 dias multiplica por $e^{0,0112} = 1,0113$ a média do total de internações, mantendo-se todas as outras variáveis constantes, ou seja, há um aumento de aproximadamente 1% na média do total de internações;
- Como regra geral, o aumento de $x \mu\text{g}/\text{m}^3$ na média de MP10 Média dos últimos 7 dias multiplica por $e^{0,0110x}$ a média do total de internações, mantendo-se todas as outras variáveis constantes.

Dia da Semana:

Consideramos domingo o dia da semana de referência, na estação verão:

- Segunda-feira: a média do total de internações em uma segunda-feira é $e^{0,3624} = 1,4368$ vezes a média do total de internações em um domingo, mantendo-se todas as outras variáveis constantes, ou seja, há um

aumento de aproximadamente 44% na média do total de interações em relação a domingo;

- Terça-feira: a média do total de interações em uma terça-feira é $e^{0,3084} = 1,3611$ vezes a média do total de interações em um domingo, mantendo-se todas as outras variáveis constantes, ou seja, há um aumento de aproximadamente 36% na média do total de interações em relação a domingo;
- Quarta-feira: a média do total de interações em uma quarta-feira é $e^{0,3260} = 1,3853$ vezes a média do total de interações em um domingo, mantendo-se todas as outras variáveis constantes, ou seja, há um aumento de aproximadamente 39% na média do total de interações em relação a domingo;
- Quinta-feira: a média do total de interações em uma quinta-feira é $e^{0,2866} = 1,3318$ vezes a média do total de interações em um domingo, mantendo-se todas as outras variáveis constantes, ou seja, há um aumento de aproximadamente 33% na média do total de interações em relação a domingo;
- Sexta-feira: a média do total de interações em uma sexta-feira é $e^{0,0972} = 1,1020$ vezes a média do total de interações em um domingo, mantendo-se todas as outras variáveis constantes, ou seja, há um aumento de aproximadamente 10% na média do total de interações em relação a domingo;
- Sábado: a média do total de interações em um sábado é $e^{0,1046} = 1,1102$ vezes a média do total de interações em um domingo, mantendo-se todas as outras variáveis constantes, ou seja, há um aumento de aproximadamente 11% na média do total de interações em relação a domingo.

Feriado:

- A média do total de interações em um dia normal é $e^{0,2777} = 1,3201$ vezes a média do total de interações em um feriado, mantendo-se todas

as outras variáveis constantes, ou seja, há um aumento de aproximadamente 32% na média do total de internações em relação ao feriado.

6.3.2. Outono

O modelo ajustado é apresentado na Tabela A.15.

Os Gráficos de diagnóstico B.23 não sugerem afastamento das suposições do modelo. Como ponto de alavanca, destaca-se a observação 119 e como ponto potencialmente influente, a observação 114, que é um dia com alta concentração de MP10 média ($97,8 \mu\text{g}/\text{m}^3$), e alto número de internações (16). O ajuste do modelo sem essa observação, no entanto, não altera a inferência.

O Gráfico B.24 mostra o envelope simulado, e indica um bom ajuste do modelo.

6.3.2.1 Interpretação dos coeficientes

Precipitação MA9:

- O aumento de 1 mm na média de precipitação dos últimos 9 dias multiplica por $e^{-0,0094} = 0,9907$ a média do total de internações, mantendo-se todas as outras variáveis constantes, ou seja, há uma diminuição de aproximadamente 0,93% na média do total de internações;
- Como regra geral, o aumento de x mm na média de precipitação dos últimos 9 dias multiplica por $e^{-0,0094x}$ a média do total de internações, mantendo-se todas as outras variáveis constantes.

•

Temperatura Máxima Lag8:

- O aumento de 1°C na temperatura máxima de 8 dias antes multiplica por $e^{-0,0110} = 0,9890$ a média do total de internações, mantendo-se todas as outras variáveis constantes, ou seja, há uma diminuição de aproximadamente 1% na média do total de internações;

- Como regra geral, no outono, o aumento de $x^{\circ}\text{C}$ na temperatura máxima de 8 dias antes multiplica por $e^{-0,0110x}$ a média do total de interações, mantendo-se todas as outras variáveis constantes.

Temperatura Mínima:

- O aumento de 1°C na temperatura mínima multiplica por $e^{0,0174} = 1,0175$ a média do total de interações, mantendo-se todas as outras variáveis constantes, ou seja, há um aumento de aproximadamente 2% na média do total de interações;
- Como regra geral, o aumento de $x^{\circ}\text{C}$ na temperatura mínima multiplica por $e^{0,0174x}$ a média do total de interações, mantendo-se todas as outras variáveis constantes.

Velocidade do Vento Média MA4:

- O aumento de 1 m/s na média da velocidade do vento dos últimos 4 dias multiplica por $e^{0,0460} = 1,0471$ a média do total de interações, mantendo-se todas as outras variáveis constantes, ou seja, há um aumento de aproximadamente 5% na média do total de interações;
- Como regra geral, o aumento de x m/s na média da velocidade do vento dos últimos 4 dias multiplica por $e^{0,0460x}$ a média do total de interações, mantendo-se todas as outras variáveis constantes.

MP10 Média:

- O aumento de $1 \mu\text{g}/\text{m}^3$ na concentração de MP10 média multiplica por $e^{0,0023} = 1,0023$ a média do total de interações, mantendo-se todas as outras variáveis constantes, ou seja, há um aumento de aproximadamente 0,23% na média do total de interações;
- Como regra geral, o aumento de $x \mu\text{g}/\text{m}^3$ na concentração de MP10 média multiplica por $e^{0,0023x}$ a média do total de interações, mantendo-se todas as outras variáveis constantes.

Dia da Semana:

Consideramos domingo o dia da semana de referência, na estação outono:

- Segunda-feira: a média do total de interações em uma segunda-feira é $e^{0,4179} = 1,5188$ vezes a média do total de interações em um domingo, mantendo-se todas as outras variáveis constantes, ou seja, há um aumento de aproximadamente 52% na média do total de interações em relação a domingo;
- Terça-feira: a média do total de interações em uma terça-feira é $e^{0,2766} = 1,3186$ vezes a média do total de interações em um domingo, mantendo-se todas as outras variáveis constantes, ou seja, há um aumento de aproximadamente 32% na média do total de interações em relação a domingo;
- Quarta-feira: a média do total de interações em uma quarta-feira é $e^{0,2506} = 1,2848$ vezes a média do total de interações em um domingo, mantendo-se todas as outras variáveis constantes, ou seja, há um aumento de aproximadamente 28% na média do total de interações em relação a domingo;
- Quinta-feira: a média do total de interações em uma quinta-feira é $e^{0,2619} = 1,2994$ vezes a média do total de interações em um domingo, mantendo-se todas as outras variáveis constantes, ou seja, há um aumento de aproximadamente 30% na média do total de interações em relação a domingo;
- Sexta-feira: a média do total de interações em uma sexta-feira é $e^{0,1733} = 1,1892$ vezes a média do total de interações em um domingo, mantendo-se todas as outras variáveis constantes, ou seja, há um aumento de aproximadamente 19% na média do total de interações em relação a domingo;
- Sábado: a média do total de interações em um sábado é $e^{0,0646} = 1,0667$ vezes a média do total de interações em um domingo, mantendo-se

todas as outras variáveis constantes, ou seja, há um aumento de aproximadamente 7% na média do total de interações em relação a domingo.

Onda de Calor:

- A média do total de interações em um dia de onda de calor é $e^{-0,1389} = 0,8704$ vezes a média do total de interações em um dia sem onda de calor, mantendo-se todas as outras variáveis constantes, ou seja, há uma diminuição de aproximadamente 13% na média do total de interações em relação a um dia sem onda de calor.

6.3.3. Inverno

O modelo ajustado é apresentado na Tabela A.16.

Os Gráficos de diagnóstico B.25 não sugerem afastamento das suposições do modelo. O Gráfico B.26 mostra o envelope simulado, e indica um ajuste razoável do modelo.

6.3.3.1 Interpretação dos coeficientes

Precipitação MA9:

- O aumento de 1 mm na média de precipitação dos últimos 9 dias multiplica por $e^{0,0125} = 1,0126$ a média do total de interações, mantendo-se todas as outras variáveis constantes, ou seja, há um aumento de aproximadamente 1% na média do total de interações;
- Como regra geral, O aumento de x mm na média de precipitação dos últimos 9 dias multiplica por $e^{0,0125x}$ a média do total de interações, mantendo-se todas as outras variáveis constantes.

Temperatura Máxima

- O aumento de 1°C na temperatura máxima multiplica por $e^{0,0118} = 1,0119$ a média do total de interações, mantendo-se todas as outras variáveis

constantes, ou seja, há um aumento de aproximadamente 1% na média do total de internações;

- Como regra geral, o aumento de $x^{\circ}\text{C}$ na temperatura máxima multiplica por $e^{0,0118x}$ a média do total de internações, mantendo-se todas as outras variáveis constantes.

Temperatura Máxima Lag8:

- o aumento de 1°C na temperatura máxima de 8 dias antes multiplica por $e^{-0,0109} = 0,9892$ a média do total de internações, mantendo-se todas as outras variáveis constantes, ou seja, há uma diminuição de aproximadamente 1% na média do total de internações;
- Como regra geral, o aumento de $x^{\circ}\text{C}$ na temperatura máxima de 8 dias antes multiplica por $e^{-0,0109x}$ a média do total de internações, mantendo-se todas as outras variáveis constantes.

Temperatura Mínima Lag6:

- o aumento de 1°C na temperatura mínima de 6 dias antes multiplica por $e^{0,0169} = 1,0170$ a média do total de internações, mantendo-se todas as outras variáveis constantes, ou seja, há uma diminuição de aproximadamente 2% na média do total de internações;
- Como regra geral, o aumento de $x^{\circ}\text{C}$ na temperatura mínima de 6 dias antes multiplica por $e^{0,0169x}$ a média do total de internações, mantendo-se todas as outras variáveis constantes.

MP10 Média MA7:

- O aumento de $1 \mu\text{g}/\text{m}^3$ na média de MP10 Média dos últimos 7 dias multiplica por $e^{0,0040} = 1,0040$ a média do total de internações, mantendo-se todas as outras variáveis constantes, ou seja, há um aumento de aproximadamente 0,40% na média do total de internações;

- Como regra geral, o aumento de $x \mu\text{g}/\text{m}^3$ na média de MP10 Média dos últimos 7 dias multiplica por $e^{0,0040x}$ a média do total de internações, mantendo-se todas as outras variáveis constantes.

Dia da Semana:

Consideramos domingo o dia da semana de referência, na estação inverno:

- Segunda-feira: a média do total de internações em uma segunda-feira é $e^{0,4329} = 1,5418$ vezes a média do total de internações em um domingo, mantendo-se todas as outras variáveis constantes, ou seja, há um aumento de aproximadamente 54% na média do total de internações em relação a domingo;
- Terça-feira: a média do total de internações em uma terça-feira é $e^{0,3311} = 1,3925$ vezes a média do total de internações em um domingo, mantendo-se todas as outras variáveis constantes, ou seja, há um aumento de aproximadamente 39% na média do total de internações em relação a domingo;
- Quarta-feira: a média do total de internações em uma quarta-feira é $e^{0,3120} = 1,3661$ vezes a média do total de internações em um domingo, mantendo-se todas as outras variáveis constantes, ou seja, há um aumento de aproximadamente 37% na média do total de internações em relação a domingo;
- Quinta-feira: a média do total de internações em uma quinta-feira é $e^{0,2823} = 1,3262$ vezes a média do total de internações em um domingo, mantendo-se todas as outras variáveis constantes, ou seja, há um aumento de aproximadamente 33% na média do total de internações em relação a domingo;
- Sexta-feira: a média do total de internações em uma sexta-feira é $e^{0,2487} = 1,2824$ vezes a média do total de internações em um domingo, mantendo-se todas as outras variáveis constantes, ou seja, há um aumento de

aproximadamente 28% na média do total de interações em relação a domingo;

- Sábado: a média do total de interações em um sábado é $e^{0,1803} = 1,1976$ vezes a média do total de interações em um domingo, mantendo-se todas as outras variáveis constantes, ou seja, há um aumento de aproximadamente 20% na média do total de interações em relação a domingo.

Feriado:

- A média do total de interações em um dia normal é $e^{0,1880} = 1,2068$ vezes a média do total de interações em um feriado, mantendo-se todas as outras variáveis constantes, ou seja, há um aumento de aproximadamente 21% na média do total de interações em relação ao feriado.

Onda de Frio:

- A média do total de interações em um dia de onda de frio é $e^{-0,1632} = 0,8494$ vezes a média do total de interações em um dia sem onda de frio, mantendo-se todas as outras variáveis constantes, ou seja, há uma diminuição de aproximadamente 15% na média do total de interações em relação ao feriado.

6.3.4. Primavera

O modelo ajustado é apresentado na Tabela A.17.

Os Gráficos de diagnóstico B.27 não sugerem afastamento das suposições do modelo. Como ponto potencialmente influente, destaca-se a observação 267. Trata-se de um dia com alta concentração de MP10 média, e alto número de interações (11), comportamento não esperado de acordo com o modelo. O ajuste do modelo sem essa observação, no entanto, não altera a inferência.

O Gráfico B.28 mostra o envelope simulado, e indica um ajuste razoável do modelo.

6.3.4.1 Interpretação dos coeficientes

TE:

- O aumento de 1°C na temperatura efetiva multiplica por $e^{0,0113} = 1,0113$ a média do total de interações, mantendo-se todas as outras variáveis constantes, ou seja, há um aumento de aproximadamente 1% na média do total de interações;
- Como regra geral, o aumento de x°C na temperatura efetiva multiplica por $e^{0,0113x}$ a média do total de interações, mantendo-se todas as outras variáveis constantes.

TE Lag8:

- O aumento de 1°C na temperatura efetiva de 8 dias antes multiplica por $e^{-0,0252} = 0,9751$ a média do total de interações, mantendo-se todas as outras variáveis constantes, ou seja, há uma diminuição de aproximadamente 2% na média do total de interações;
- Como regra geral, o aumento de x°C na temperatura efetiva de 8 dias antes multiplica por $e^{-0,0252x}$ a média do total de interações, mantendo-se todas as outras variáveis constantes.

MP10 Média:

- O aumento de 1 $\mu\text{g}/\text{m}^3$ na concentração de MP10 média multiplica por $e^{0,0021} = 0,9979$ a média do total de interações, mantendo-se todas as outras variáveis constantes, ou seja, há uma diminuição de aproximadamente 0,21% na média do total de interações;
- Como regra geral, o aumento de x $\mu\text{g}/\text{m}^3$ na concentração de MP10 média multiplica por $e^{0,0021x}$ a média do total de interações, mantendo-se todas as outras variáveis constantes.

MP10 Média MA7:

- O aumento de $1 \mu\text{g}/\text{m}^3$ na média de MP10 Média dos últimos 7 dias multiplica por $e^{0,0060} = 1,0060$ a média do total de interações, mantendo-se todas as outras variáveis constantes, ou seja, há um aumento de aproximadamente 0,60% na média do total de interações;
- Como regra geral, o aumento de $x \mu\text{g}/\text{m}^3$ na média de MP10 Média dos últimos 7 dias multiplica por $e^{0,0060x}$ a média do total de interações, mantendo-se todas as outras variáveis constantes.

Dia da Semana:

Consideramos domingo o dia da semana de referência, na estação primavera:

- Segunda-feira: a média do total de interações em uma segunda-feira é $e^{0,4335} = 1,5426$ vezes a média do total de interações em um domingo, mantendo-se todas as outras variáveis constantes, ou seja, há um aumento de aproximadamente 54% na média do total de interações em relação a domingo;
- Terça-feira: a média do total de interações em uma terça-feira é $e^{0,4458} = 1,5617$ vezes a média do total de interações em um domingo, mantendo-se todas as outras variáveis constantes, ou seja, há um aumento de aproximadamente 56% na média do total de interações em relação a domingo;
- Quarta-feira: a média do total de interações em uma quarta-feira é $e^{0,3259} = 1,3853$ vezes a média do total de interações em um domingo, mantendo-se todas as outras variáveis constantes, ou seja, há um aumento de aproximadamente 39% na média do total de interações em relação a domingo;
- Quinta-feira: a média do total de interações em uma quinta-feira é $e^{0,2576} = 1,2938$ vezes a média do total de interações em um domingo,

mantendo-se todas as outras variáveis constantes, ou seja, há um aumento de aproximadamente 29% na média do total de interações em relação a domingo;

- Sexta-feira: a média do total de interações em uma sexta-feira é $e^{0,2928} = 1,3402$ vezes a média do total de interações em um domingo, mantendo-se todas as outras variáveis constantes, ou seja, há um aumento de aproximadamente 34% na média do total de interações em relação a domingo;
- Sábado: a média do total de interações em um sábado é $e^{0,0088} = 1,0089$ vezes a média do total de interações em um domingo, mantendo-se todas as outras variáveis constantes, ou seja, há um aumento de aproximadamente 0,89% na média do total de interações em relação a domingo.

7. Conclusões

Para todos os modelos de variáveis selecionadas ajustados por estação, assim como no modelo de componentes principais, observa-se um aumento na média do total de interações para todos os dias da semana em relação a domingo.

Nota-se também um aumento na média do total de interações para dias normais, em relação a feriados durante as estações verão e inverno, o que pode ser decorrente tanto dos efeitos do feriado, quanto do desbalanceamento da base de dados com relação a essa variável, que apresenta apenas 160 dias feriados contra 3858 dias normais. Em estudos observacionais, esse desbalanceamento é inevitável e pode prejudicar o desenvolvimento do modelo, já que o número de dias feriados pode ser insuficiente para se estabelecer uma relação com o total de interações. Da mesma forma, observa-se uma diminuição na média do total de interações para dias de onda de calor durante o outono, e para dias de onda de frio durante o inverno. A base de dados conta com apenas 89 dias de onda de calor durante as estações verão e outono, e 101 dias de onda de frio, durante as estações inverno e primavera, contra 3828 dias sem onda de frio nem de calor.

Para as estações verão, outono e inverno, o aumento da concentração de MP10 média, e/ou da média de concentração de MP10 média dos últimos 7 dias ocasiona um aumento na média do total de internações. Já para a primavera, o aumento da concentração de MP10 média ocasiona uma diminuição na média do total de internações, efeito contrário ao esperado, o que pode ser devido à interação entre as fontes de poluição e a atmosfera, ou ao efeito de variáveis não consideradas neste estudo que possam ter influência nas internações por doenças respiratórias.

Observa-se também que um aumento na precipitação durante o verão e durante o outono ocasiona uma diminuição na média do total de internações. Já um aumento na média de precipitação dos últimos 9 dias ocasiona um aumento na média do total de internações durante o inverno.

Durante a primavera, com o aumento na temperatura efetiva, observa-se um aumento na média do total de internações. A temperatura efetiva observada 8 dias antes, no entanto, tem efeito contrário, diminuindo a média do total de internações, o que pode ser devido a uma relação com outras variáveis de clima e de poluição neste período de defasagem, que pode criar uma condição favorável para a diminuição na média do total de internações.

O modelo de regressão de componentes principais, apesar de não possibilitar a interpretação clássica dos coeficientes, é útil para a previsão da média do total de internações, a partir das variáveis climáticas e de poluição observadas no referido dia, e nos dias anteriores a ele.

Os modelos de variáveis selecionadas, apesar de possibilitarem a interpretabilidade dos coeficientes, dependem da escolha das variáveis, podendo apresentar diferentes interpretações de acordo com o conjunto considerado. Para cada estação do ano, a média do total de internações pode ser melhor explicada por um determinado conjunto de variáveis. Nesse contexto, outros modelos podem ser ajustados.

Como apontado anteriormente, os modelos estão com ajustes adequados e contribuem na identificação e mensuração de condições climáticas e de qualidade de ar que influenciam no total de internações por doenças respiratórias.

APÊNDICE A

Tabelas

Tabela A.1. Medidas descritivas das variáveis quantitativas.

Variáveis	Mín.	1º Q	Mediana	Média	3º Q	Máx.	Desvio Padrão
Total de Internações	0,00	2,00	4,00	4,36	6,00	17,00	2,61
Temperatura Máxima	8,50	23,10	26,35	26,05	29,30	36,60	4,27
Temperatura Mínima	5,20	14,30	16,70	16,43	18,90	24,30	3,17
Temperatura Média	7,10	18,70	21,15	20,94	23,50	29,00	3,38
Amplitude Térmica	0,20	7,40	9,60	9,62	11,90	21,30	3,17
Precipitação	0,00	0,00	0,00	4,65	2,30	140,40	11,72
Umidade Relativa Média	28,00	67,00	75,30	73,58	81,80	98,30	11,59
Velocidade do Vento Média	0,00	1,93	2,55	2,62	3,23	50,33	1,49
TE	7,30	17,80	19,90	19,73	22,00	26,60	2,91
Tev	-5,80	11,70	14,90	14,49	17,90	25,10	4,46
MP10 Média	5,80	23,70	33,60	38,18	47,52	185,40	20,16

Tabela A.2. Força de associação do coeficiente de correlação de Spearman.

Coeficiente (módulo)	Força de associação
$0,0 \leq \rho < 0,2$	Muito fraca
$0,2 \leq \rho < 0,4$	Fraca
$0,4 \leq \rho < 0,6$	Moderada
$0,6 \leq \rho < 0,8$	Forte
$0,8 \leq \rho \leq 1,0$	Muito forte

Tabela A.3. Lags que resultam em maiores coeficientes de correlação.

Variáveis	Coeficiente de correlação máximo (módulo)	Lag
Precipitação	-0,1380	4
Temperatura Máxima	-0,1484	8
Temperatura Média	-0,1643	8
Temperatura Mínima	-0,1551	6
Amplitude Térmica	-0,0619	10
Umidade Relativa Média	-0,0755	3
Velocidade do Vento Média	-0,0873	2
TE	-0,1667	8
Tev	-0,1486	9
MP10 Média	0,1882	0

Tabela A.4. Número de dias anteriores ao dia da internação que resultam em maiores coeficientes de correlação.

Variáveis	Coeficiente de correlação máximo (módulo)	MA
Precipitação	-0,2024	9
Temperatura Máxima	-0,1688	18
Temperatura Média	-0,1611	14
Temperatura Mínima	-0,1566	10
Amplitude Térmica	-0,0551	28
Umidade Relativa Média	-0,0893	6
Velocidade do Vento Média	-0,1052	28
TE	-0,1611	13
Tev	-0,1459	15
MP10 Média	0,2031	7

Tabela A.5. Variáveis lag e MA.

Variável	Lag	MA
Precipitação	4	9
Temperatura Máxima	8	18
Temperatura Média	8	14
Temperatura Mínima	6	10
Amplitude Térmica	10	28
Umidade Relativa Média	3	6
Velocidade do Vento Média	2	4
TE	8	13
Tev	9	15
MP10 Média	0	7

Tabela A.6. Componente Principais (1 a 10) e Variância explicada.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
Desvio Padrão	3,33	2,25	1,54	1,39	1,23	1,05	0,97	0,91	0,85	0,76
Proporção da Variância	0,38	0,18	0,08	0,07	0,05	0,04	0,03	0,03	0,02	0,02
Proporção Acumulada	0,38	0,56	0,64	0,71	0,76	0,80	0,83	0,86	0,88	0,90

Tabela A.7. Componente Principais (11 a 20) e Variância explicada.

	PC11	PC12	PC13	PC14	PC15	PC16	PC17	PC18	PC19	PC20
Desvio Padrão	0,76	0,70	0,62	0,59	0,54	0,44	0,40	0,36	0,34	0,24
Proporção da Variância	0,02	0,02	0,01	0,01	0,01	0,01	0,01	0,00	0,00	0,00
Proporção Acumulada	0,92	0,94	0,95	0,96	0,97	0,98	0,99	0,99	0,99	1,00

Tabela A.8. Componente Principais (21 a 29) e Variância explicada.

	PC21	PC22	PC23	PC24	PC25	PC26	PC27	PC28	PC29
Desvio Padrão	0,19	0,16	0,11	0,10	0,09	0,05	0,05	0,03	0,00
Proporção da Variância	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Proporção Acumulada	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00

Tabela A.9. Composição das Componentes Principais (1 a 10).

Variável	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
Temperatura Máxima	0.17	-0.33	0.11	-0.21	0.10	-0.04	0.06	-0.03	0.10	-0.05
Temperatura Mínima	0.25	-0.12	0.14	-0.05	-0.24	0.01	0.02	-0.22	-0.09	0.11
Temperatura Média	0.22	-0.26	0.13	-0.15	-0.05	-0.02	0.04	-0.11	0.02	0.03
Amplitude Térmica	-0.02	-0.31	0.00	-0.22	0.37	-0.07	0.07	0.17	0.22	-0.17
TE	0.24	-0.21	0.16	-0.12	-0.08	-0.01	0.07	-0.14	-0.03	0.04
Tev	0.20	-0.28	0.22	-0.02	0.01	0.03	0.14	-0.06	-0.16	-0.05
Precipitação	0.07	0.11	0.10	0.21	-0.15	-0.44	0.53	0.40	0.24	0.17
Umidade Relativa Média	0.04	0.31	0.08	0.21	-0.22	0.03	0.11	-0.16	-0.28	0.04
Velocidade do Vento Média	0.03	0.18	-0.26	-0.33	-0.19	-0.15	-0.31	-0.15	0.49	0.28
MP10 Média	-0.10	-0.31	-0.02	0.06	0.20	0.05	-0.08	0.17	-0.17	0.56
Precipitação MA9	0.15	0.15	0.22	0.04	0.12	-0.47	-0.01	0.11	0.06	0.08
Temperatura Máxima MA18	0.27	-0.06	-0.14	0.03	-0.10	-0.09	-0.01	-0.04	0.02	-0.01
Temperatura Média MA14	0.29	-0.02	-0.07	0.02	-0.11	0.00	-0.04	-0.01	0.01	0.04
Temperatura Mínima MA10	0.29	0.03	0.01	0.01	-0.13	0.06	-0.05	0.00	-0.01	0.07
Amplitude Térmica MA28	-0.03	-0.19	-0.37	0.10	0.05	-0.36	0.13	-0.17	0.05	-0.18
Umidade Relativa Média MA6	0.07	0.32	0.29	-0.01	0.25	0.07	0.07	-0.13	0.02	0.12
Velocidade do Vento Média MA4	0.04	0.20	-0.27	-0.51	-0.05	-0.06	0.09	0.02	-0.08	0.12
TE MA13	0.29	0.00	-0.02	0.02	-0.11	0.01	-0.04	-0.01	0.01	0.05
Tev MA15	0.28	-0.03	-0.02	0.11	-0.09	0.01	-0.08	0.01	0.00	0.05
MP10 Média MA7	-0.13	-0.25	-0.19	0.21	-0.11	-0.02	-0.13	0.13	-0.15	0.53
Precipitação Lag4	0.08	0.08	0.14	-0.05	0.22	-0.53	-0.56	0.05	-0.39	-0.09
Temperatura Máxima Lag8	0.21	0.05	-0.27	0.14	0.23	0.13	-0.03	0.24	0.02	-0.10
Temperatura Média Lag8	0.25	0.08	-0.20	0.10	0.20	0.15	-0.04	0.19	0.02	-0.06
Temperatura Mínima Lag6	0.25	0.08	-0.03	0.00	0.00	0.09	-0.09	0.12	0.00	0.07
Amplitude Térmica Lag10	0.02	-0.01	-0.31	0.21	0.31	-0.19	0.24	-0.61	-0.08	0.11
Umidade Relativa Média Lag3	0.05	0.19	0.28	-0.11	0.42	0.13	0.09	-0.17	0.21	0.36
Velocidade do Vento Média Lag2	0.03	0.13	-0.19	-0.47	0.05	0.00	0.36	0.17	-0.52	0.07
TE Lag8	0.26	0.08	-0.15	0.08	0.17	0.15	-0.04	0.17	0.02	-0.05
Tev Lag9	0.22	0.07	-0.16	0.16	0.21	0.08	0.01	-0.04	0.00	0.02

Tabela A.10. Estimativa para os parâmetros do modelo de componentes principais.

Nível de referência	Variável	Estimativa	Erro Padrão	Valor-p
	(Intercepto)	0,7779	0,0538	2,19554E-47
	CP1	-0,0087	0,0039	0,026556671
	CP2	-0,0207	0,0037	3,39237E-08
	CP3	0,0118	0,0057	0,038559648
	CP4	-0,0026	0,0062	0,673183575
	CP5	-0,0338	0,0069	1,01209E-06
	CP6	0,0254	0,0085	0,002834343
	CP7	0,0030	0,0091	0,739268528
	CP8	0,0122	0,0092	0,183397263
	CP9	-0,0023	0,0098	0,817760667
	CP10	0,0610	0,0105	5,78298E-09
Domingo	Segunda-feira	0,4108	0,0315	6,93647E-39
	Terça-feira	0,3257	0,0320	2,23694E-24
	Quarta-feira	0,2865	0,0323	6,72826E-19
	Quinta-feira	0,2612	0,0324	8,2204E-16
	Sexta-feira	0,2016	0,0329	8,51242E-10
	Sábado	0,0798	0,0336	0,017585355
Feriado: Sim	Feriado: Não	0,0501	0,0434	0,249088881
Verão	Outono	0,7021	0,0276	3,7056E-143
	Inverno	0,4496	0,0392	1,79777E-30
	Primavera	0,3558	0,0317	3,38163E-29
Onda de Frio: Não	Onda de Frio: Sim	-0,1537	0,0605	0,011035223
Onda de Calor: Não	Onda de Calor: Sim	-0,0443	0,0634	0,484962874

Tabela A.11. Estimativa para os parâmetros do submodelo de componentes principais.

Nível de referência	Variável	Estimativa	Erro Padrão	Valor-p
	(Intercepto)	0,8258	0,0339	2,723E-131
	CP1	-0,0094	0,0038	0,014292495
	CP2	-0,0203	0,0037	3,32227E-08
	CP3	0,0117	0,0057	0,03924262
	CP5	-0,0333	0,0069	1,44456E-06
	CP6	0,0253	0,0085	0,002910406
	CP10	0,0616	0,0105	4,28572E-09
Domingo	Segunda-feira	0,4121	0,0315	3,69898E-39
	Terça-feira	0,3266	0,0320	1,65329E-24
	Quarta-feira	0,2885	0,0323	3,70709E-19
	Quinta-feira	0,2621	0,0324	6,66955E-16
	Sexta-feira	0,2034	0,0328	5,83676E-10
	Sábado	0,0815	0,0336	0,015244779
Verão	Outono	0,7003	0,0273	1,8335E-145
	Inverno	0,4471	0,0383	2,07199E-31
	Primavera	0,3527	0,0314	2,35417E-29
Onda de Frio: Não	Onda de Frio: Sim	-0,1535	0,0603	0,010907569

Tabela A.12. Valores-p do teste qui-quadrado de Pearson para os ajustes dos modelos por estação do ano, para cada conjunto de variáveis selecionadas.

Conjunto	Estação			
	Verão	Outono	Inverno	Primavera
1	0,3486196	0,4514672	0,3709484	0,3227950
2	0,3567634	0,4209749	0,3209940	0,2980561
3	0,3697443	0,4434056	0,3678387	0,3358796
4	0,3503649	0,4352898	0,3101214	0,2833866

Tabela A.13. Modelo ajustado para a estação verão.

Nível de referência	Variável	Estimativa	Erro Padrão	Valor-p
	(Intercepto)	0,3123	0,1573	0,047034973
	Precipitação Lag4	-0,0038	0,0015	0,011842241
	MP10 Média MA7	0,0112	0,0027	3,34403E-05
Domingo	Segunda-feira	0,3638	0,0837	1,38542E-05
	Terça-feira	0,3221	0,0843	0,000132213
	Quarta-feira	0,3271	0,0840	9,85719E-05
	Quinta-feira	0,2866	0,0846	0,000702705
	Sexta-feira	0,0980	0,0876	0,262938191
	Sábado	0,1050	0,0876	0,231036326
Feriado: Sim	Feriado: Não	0,1989	0,1258	0,113716718

Tabela A.14. Modelo ajustado para a estação verão sem a observação 1512

Nível de referência	Variável	Estimativa	Erro Padrão	Valor-p
	(Intercepto)	0,2425	0,1614	0,132985947
	Precipitação Lag4	-0,0037	0,0015	0,013493536
	MP10 Média MA7	0,0110	0,0027	4,67421E-05
Domingo	Segunda-feira	0,3624	0,0834	1,38911E-05
	Terça-feira	0,3084	0,0843	0,000254549
	Quarta-feira	0,3260	0,0837	9,86648E-05
	Quinta-feira	0,2866	0,0843	0,000675377
	Sexta-feira	0,0972	0,0873	0,265523425
	Sábado	0,1046	0,0874	0,231262173
Feriado: Sim	Feriado: Não	0,2777	0,1315	0,034650507

Tabela A.15. Modelo ajustado para a estação outono.

Nível de referência	Variável	Estimativa	Erro Padrão	Valor-p
	(Intercepto)	1,3991	0,1484	4,28279E-21
	Precipitação MA9	-0,0094	0,0030	0,001480812
	Temperatura Máxima Lag8	-0,0110	0,0042	0,009261446
	Temperatura Mínima	0,0174	0,0057	0,00235282
	Velocidade do Vento Média MA4	0,0460	0,0258	0,074004732
	MP10 Média	0,0023	0,0008	0,004733624
Domingo	Segunda-feira	0,4179	0,0531	3,44546E-15
	Terça-feira	0,2766	0,0547	4,26866E-07
	Quarta-feira	0,2506	0,0550	5,19329E-06
	Quinta-feira	0,2619	0,0552	2,06638E-06
	Sexta-feira	0,1733	0,0562	0,002031122
	Sábado	0,0646	0,0571	0,257816151
Onda de Calor: Não	Onda de Calor: Sim	-0,1389	0,0918	0,130595845

Tabela A.16. Modelo ajustado para a estação inverno.

Nível de referência	Variável	Estimativa	Erro Padrão	Valor-p
	(Intercepto)	0,6259	0,2096	0,002829396
	Precipitação MA9	0,0125	0,0058	0,030340564
	Temperatura Máxima	0,0118	0,0044	0,008031551
	Temperatura Máxima Lag8	-0,0109	0,0047	0,020578779
	Temperatura Mínima Lag6	0,0169	0,0078	0,02940401
	MP10 Média MA7	0,0040	0,0010	0,000131377
Domingo	Segunda-feira	0,4329	0,0593	2,95741E-13
	Terça-feira	0,3311	0,0605	4,39745E-08
	Quarta-feira	0,3120	0,0607	2,76122E-07
	Quinta-feira	0,2823	0,0615	4,37466E-06
	Sexta-feira	0,2487	0,0613	5,014E-05
	Sábado	0,1803	0,0623	0,00380891
Feriado: Sim	Feriado: Não	0,1880	0,1218	0,122856481
Onda de Frio: Não	Onda de Frio: Sim	-0,1632	0,0621	0,008583224

Tabela A.17. Modelo ajustado para a estação primavera.

Nível de referência	Variável	Estimativa	Erro Padrão	Valor-p
	(Intercepto)	1,2492	0,1904	5,40647E-11
	TE	0,0113	0,0072	0,120408556
	TE Lag8	-0,0252	0,0066	0,000132219
	MP10 Média	-0,0021	0,0013	0,102028748
	MP10 Média MA7	0,0060	0,0017	0,000386736
Domingo	Segunda-feira	0,4335	0,0638	1,07597E-11
	Terça-feira	0,4458	0,0640	3,2455E-12
	Quarta-feira	0,3259	0,0658	7,22079E-07
	Quinta-feira	0,2576	0,0667	0,00011242
	Sexta-feira	0,2928	0,0666	1,10173E-05
	Sábado	0,0088	0,0702	0,900036642

APÊNDICE B

Gráficos

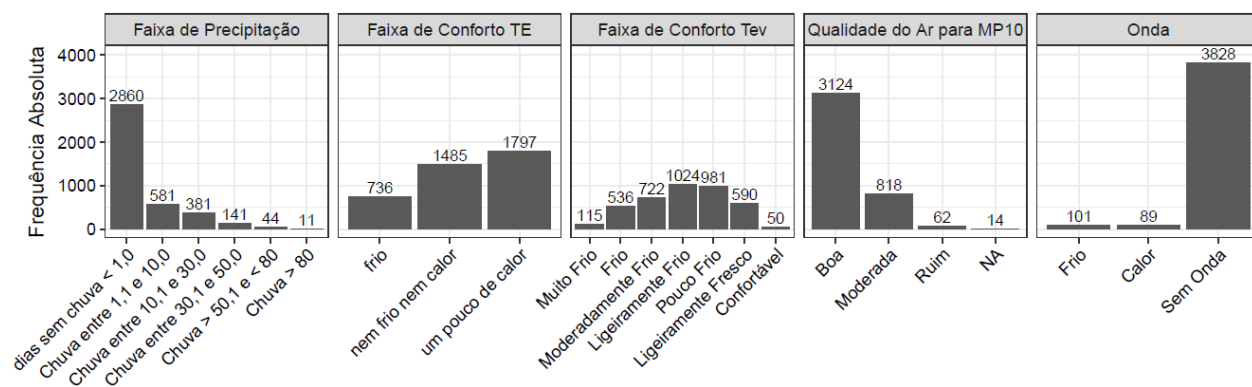
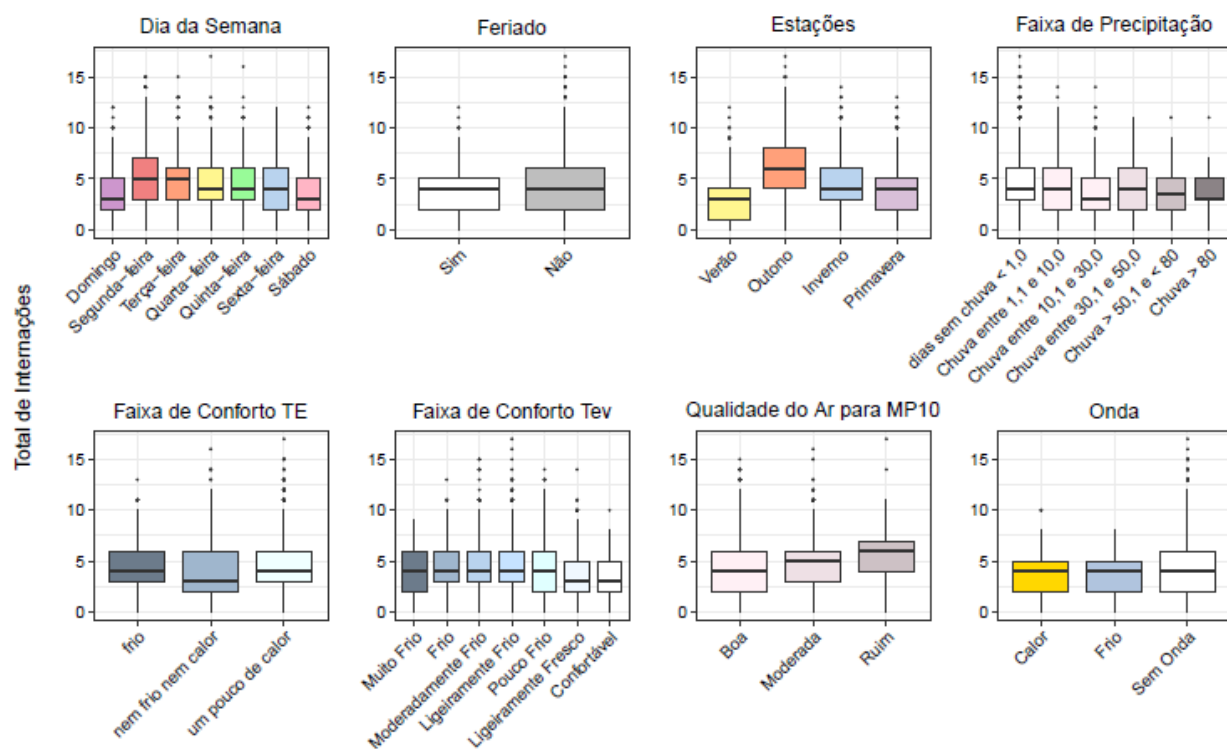
Gráfico B.1. Frequências absolutas das variáveis qualitativas.**Gráfico B.2.** Boxplots do total de interações segundo as variáveis qualitativas.

Gráfico B.3. Boxplots do total de interações segundo *Dia da Semana*, *Feriado*, e *Onda* ao longo dos anos.

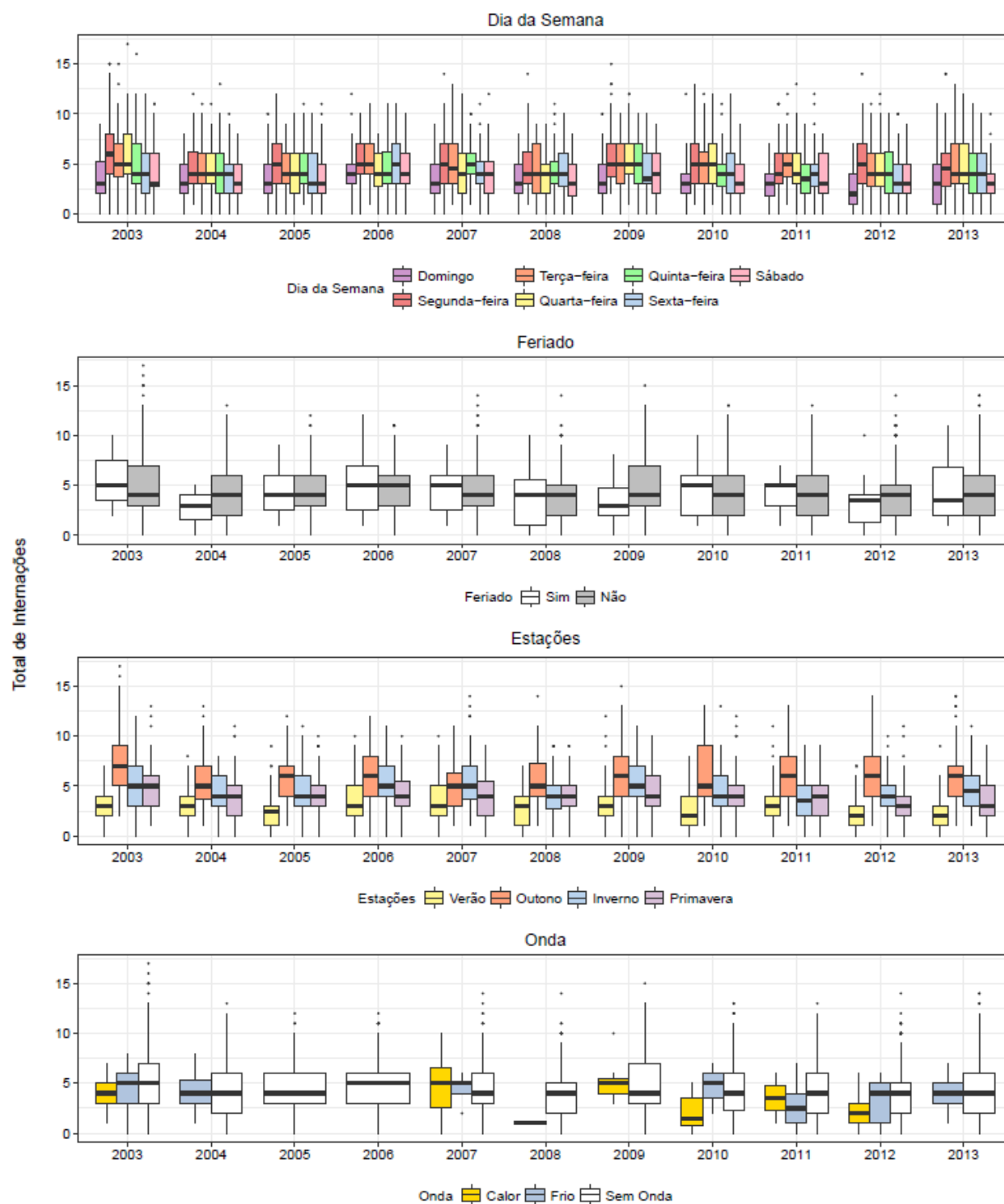


Gráfico B.4. Boxplots do total de interações segundo *Qualidade do Ar para MP10*, *Faixa de Conforto TE*, *Faixa de Conforto Tev* e *Faixa de Precipitação* ao longo dos anos.

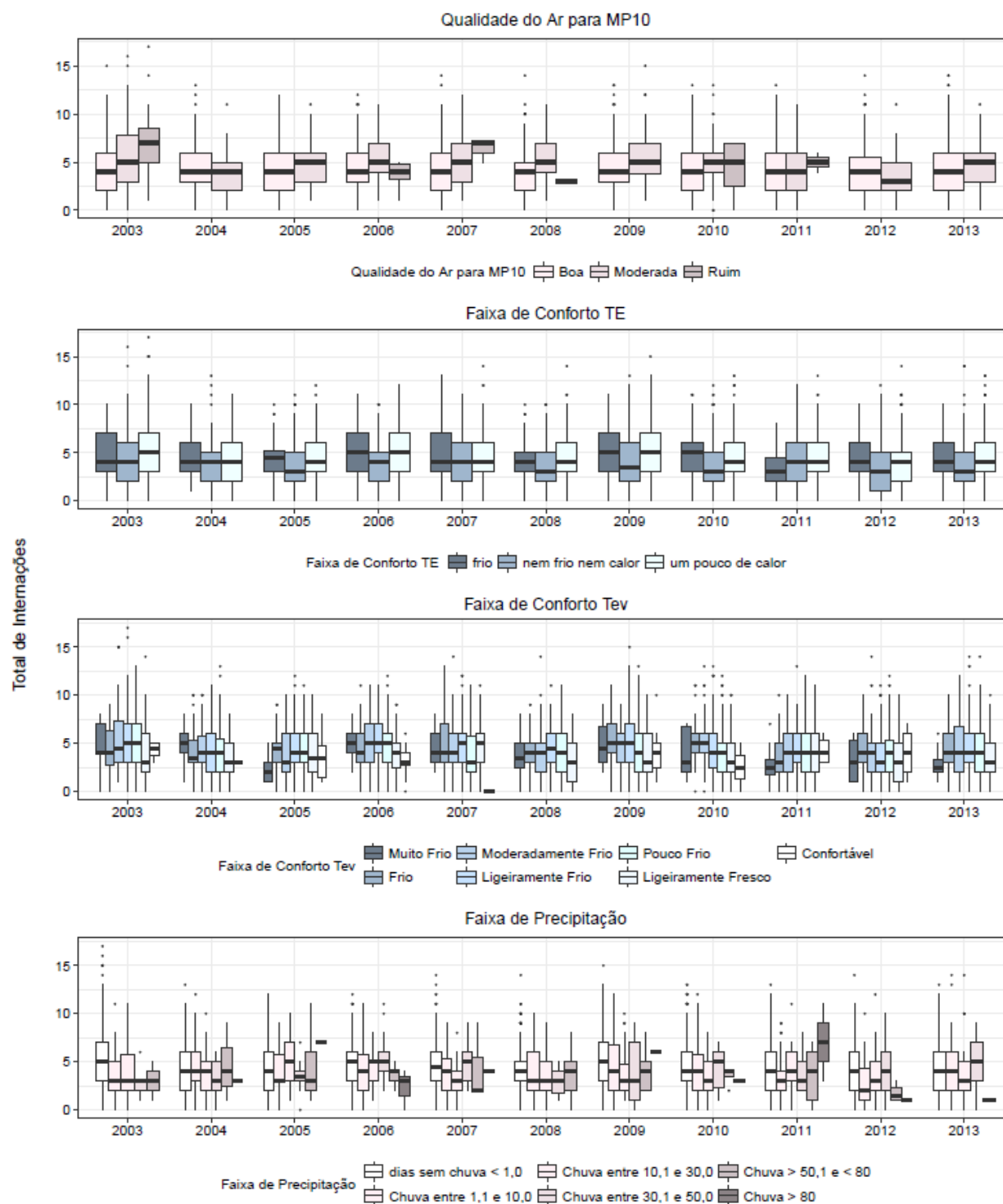


Gráfico B.5. Histogramas das variáveis quantitativas.

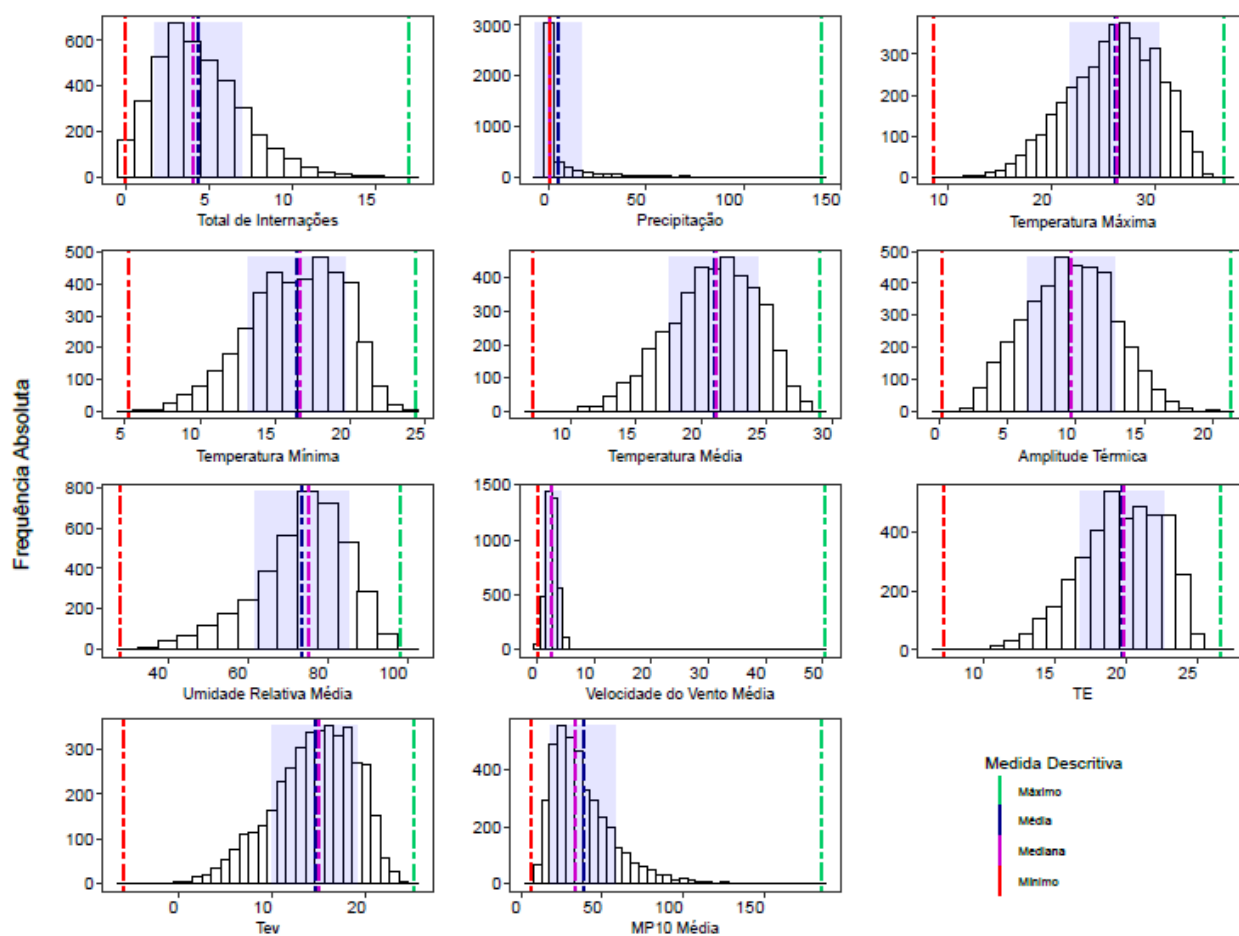


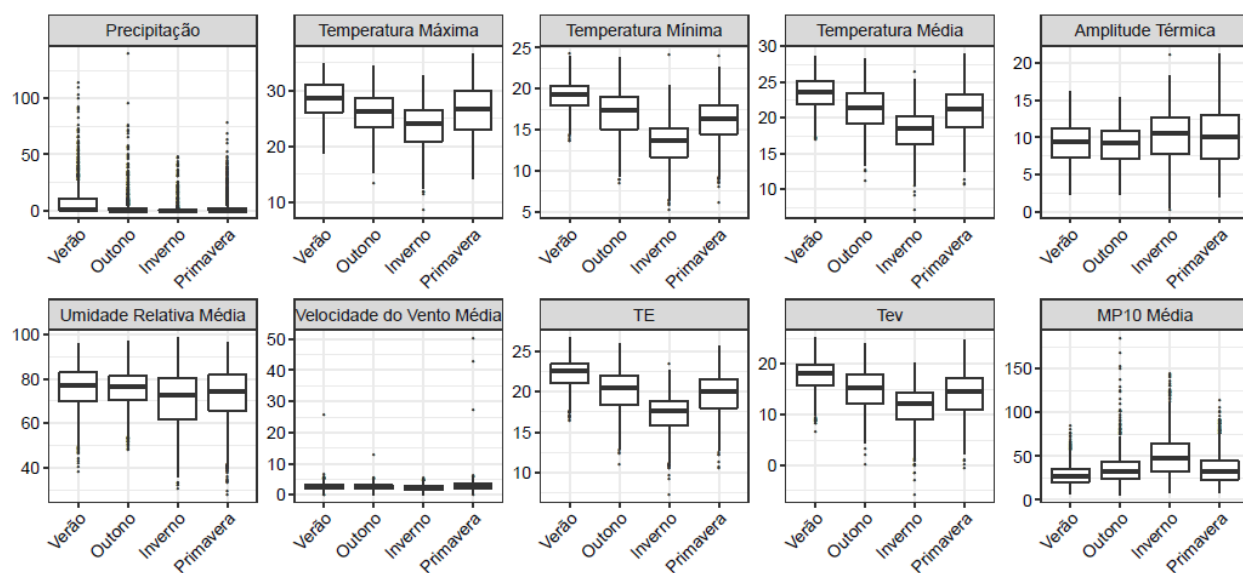
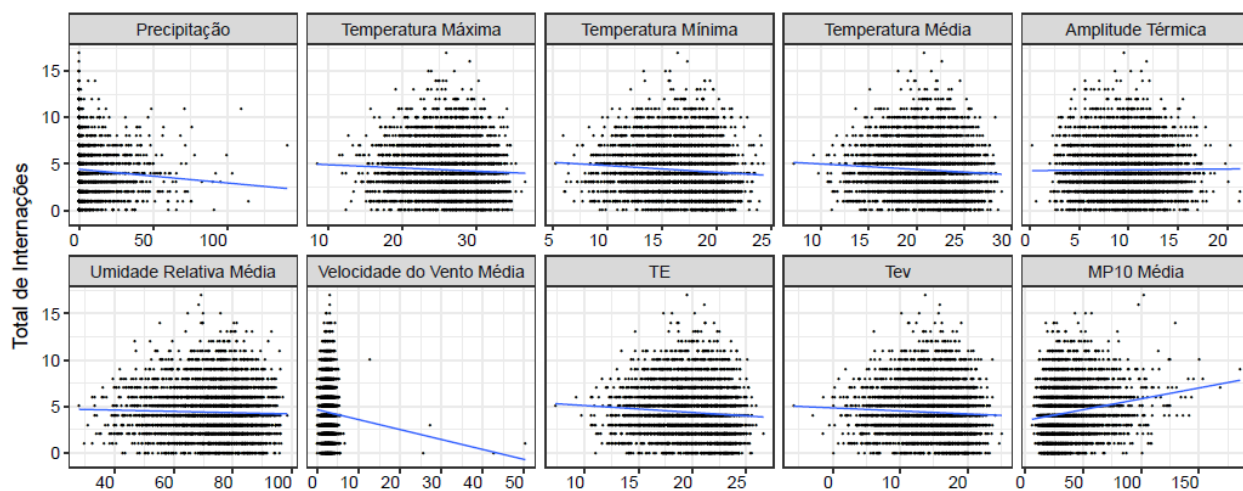
Gráfico B.6. Boxplots das variáveis quantitativas segundo as estações.**Gráfico B.7.** Gráficos de dispersão do total de interações pelas variáveis quantitativas.

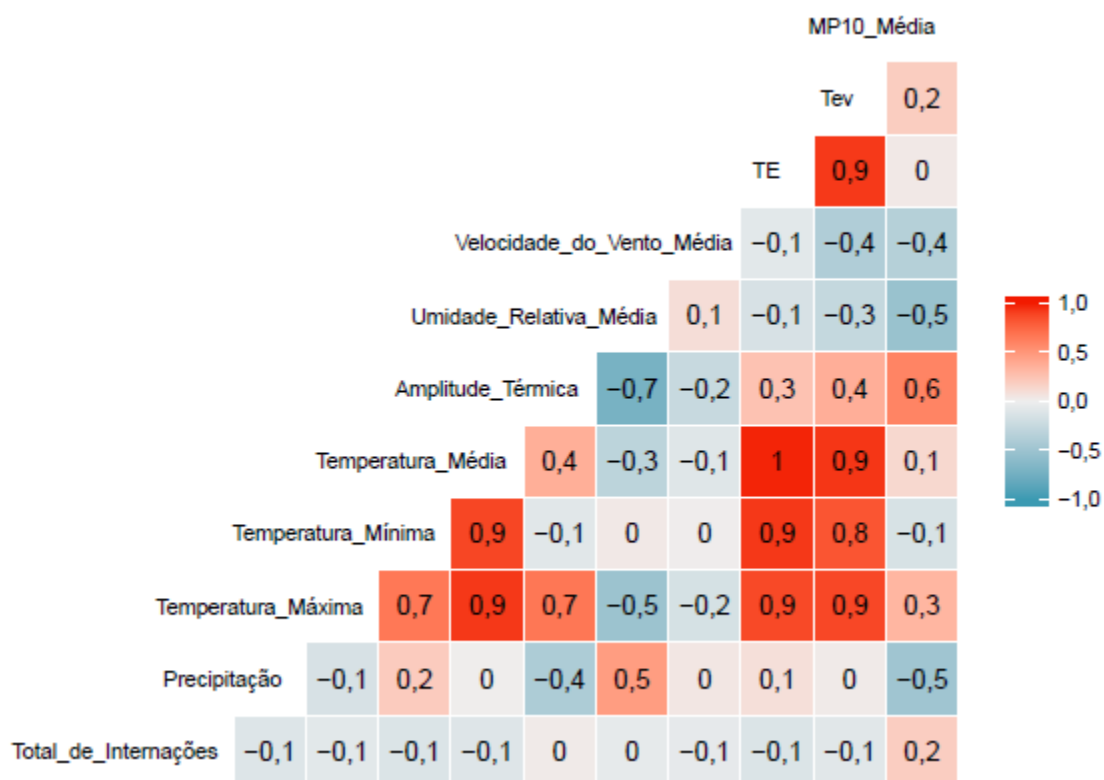
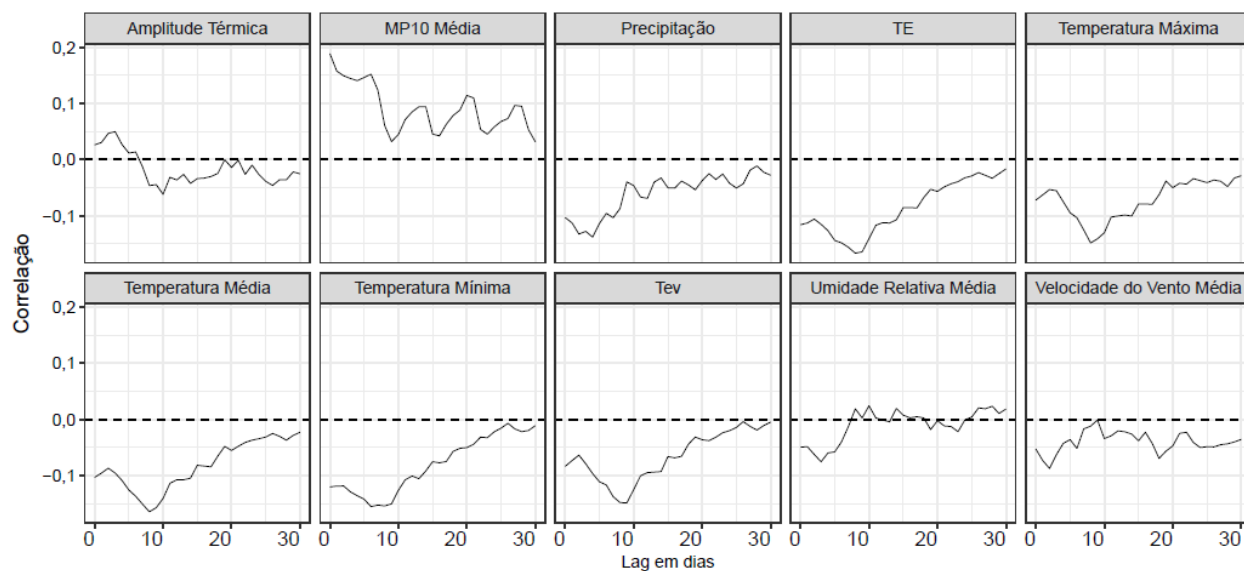
Gráfico B.8. Matriz de correlação de Spearman das variáveis quantitativas.**Gráfico B.9.** Correlações de Spearman entre o total de interações e as variáveis quantitativas para vários valores de lag.

Gráfico B.10. Dispersão do total de interações pelas variáveis quantitativas considerando-se os lags.

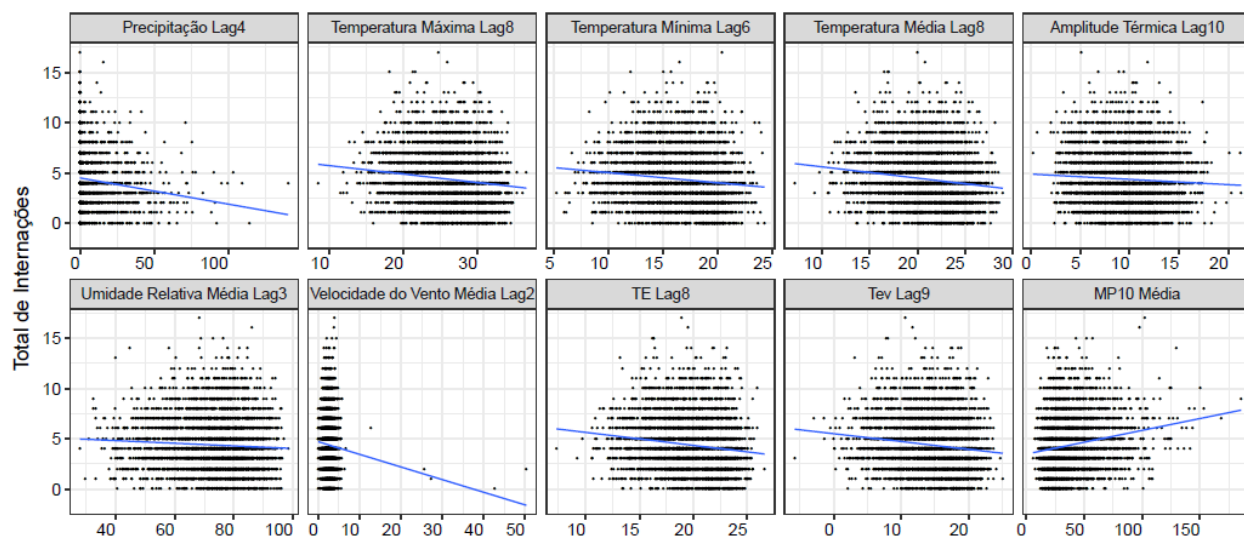


Gráfico B.11. Boxplots do total de interações pelas variáveis qualitativas considerando-se os lags.

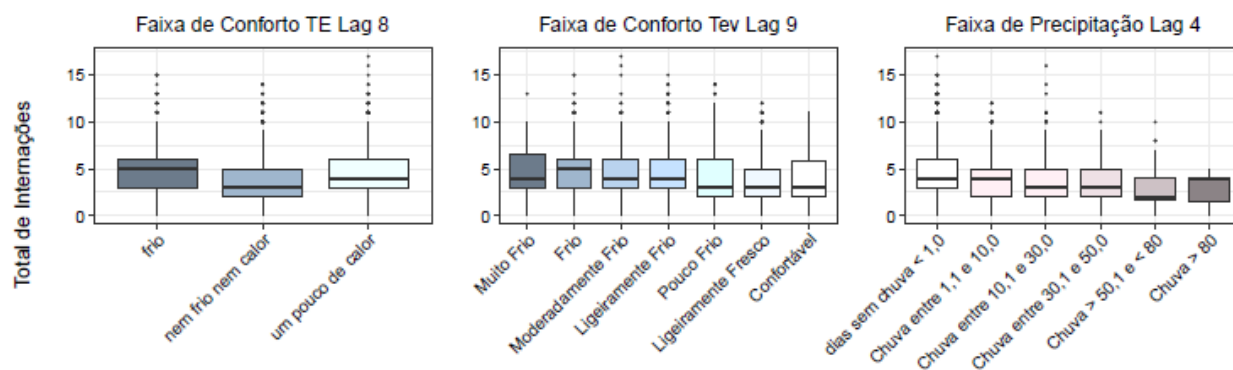


Gráfico B.12. Correlações de Spearman entre o total de interações e as variáveis quantitativas para vários valores de dias anteriores ao dia da interação considerados na média móvel.

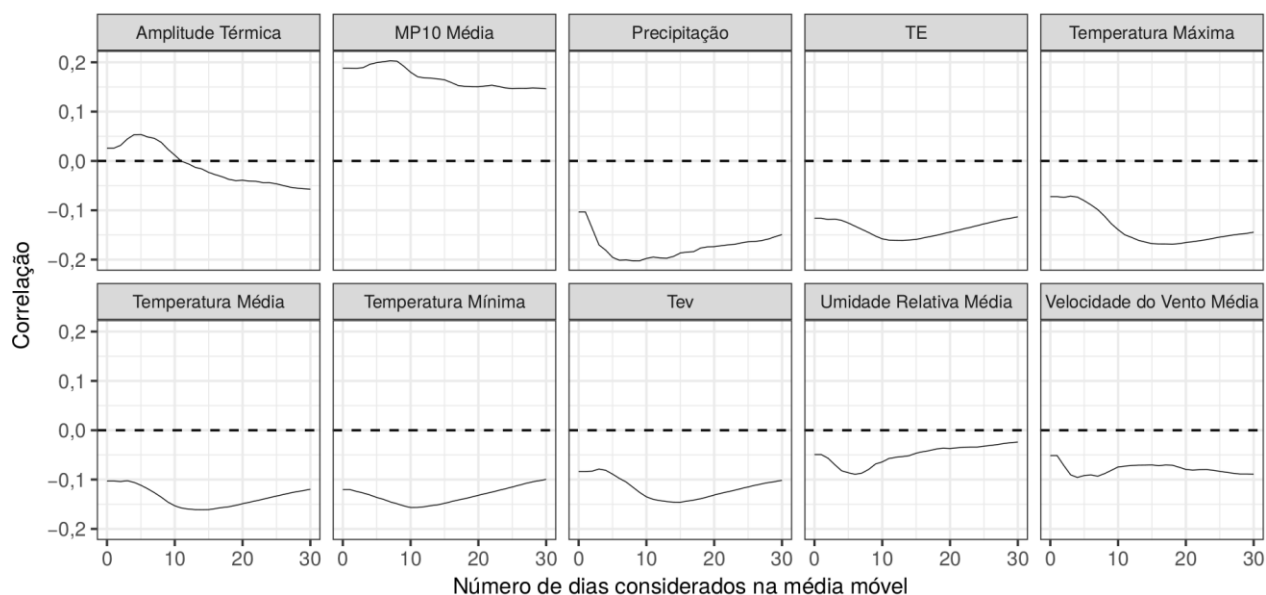


Gráfico B.13. Dispersão do total de interações pelas variáveis quantitativas considerando médias móveis.

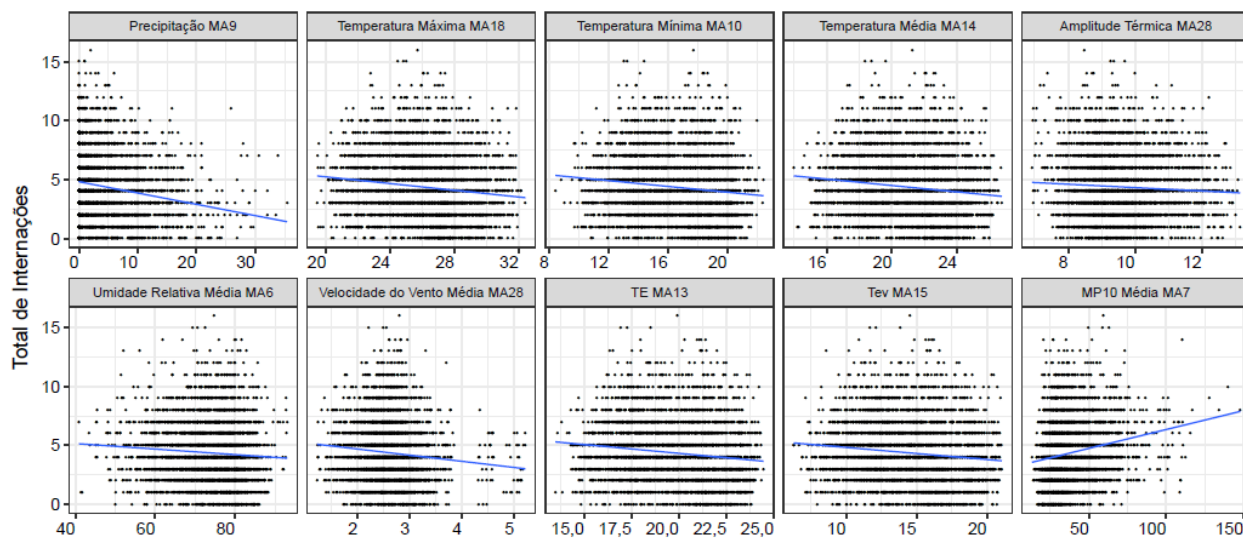


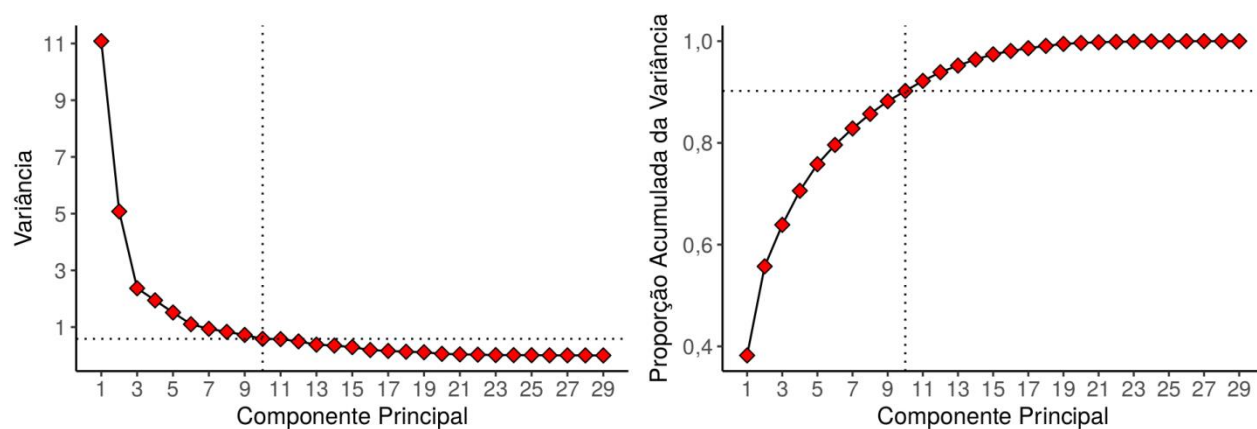
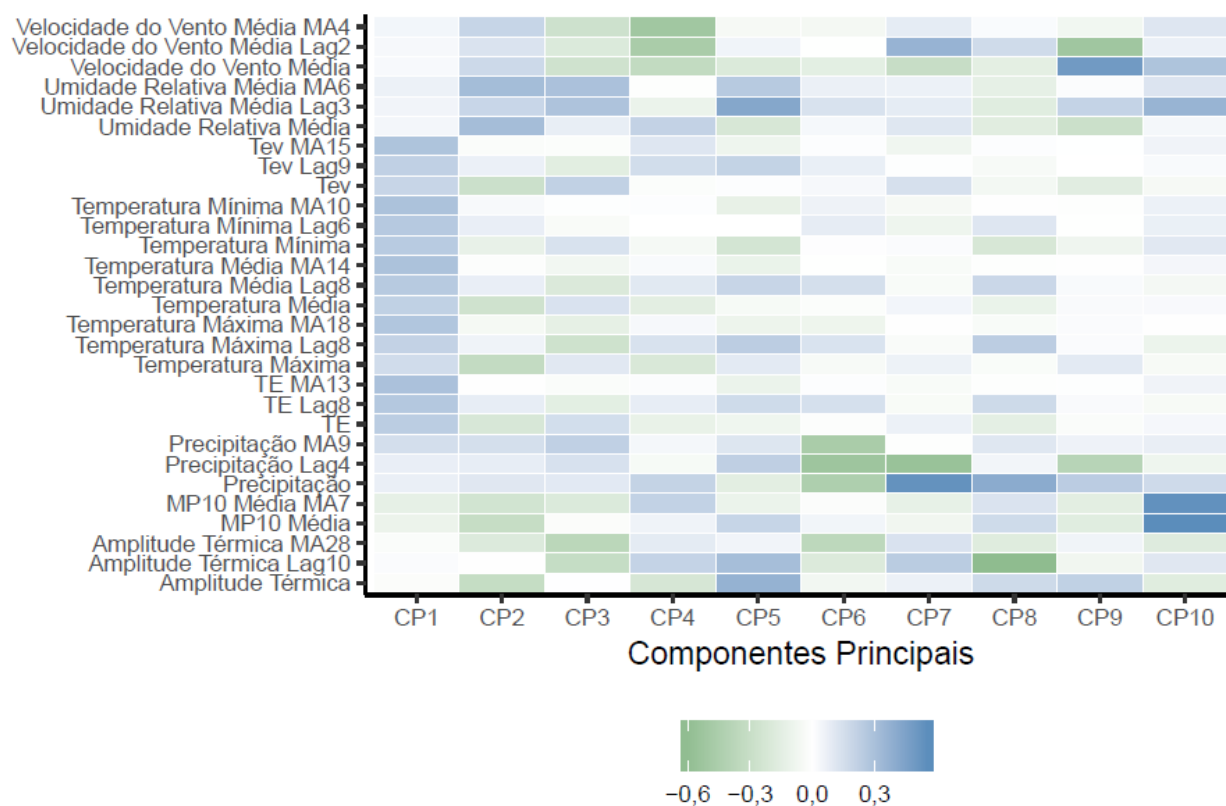
Gráfico B.14. Scree plot e proporção acumulada da variância.**Gráfico B.15.** Visualização da composição das componentes principais.

Gráfico B.18. Análise de Resíduos para o modelo de componentes principais: Medida h , Distância de Cook, Resíduo Componente do Desvio e Preditor Linear.

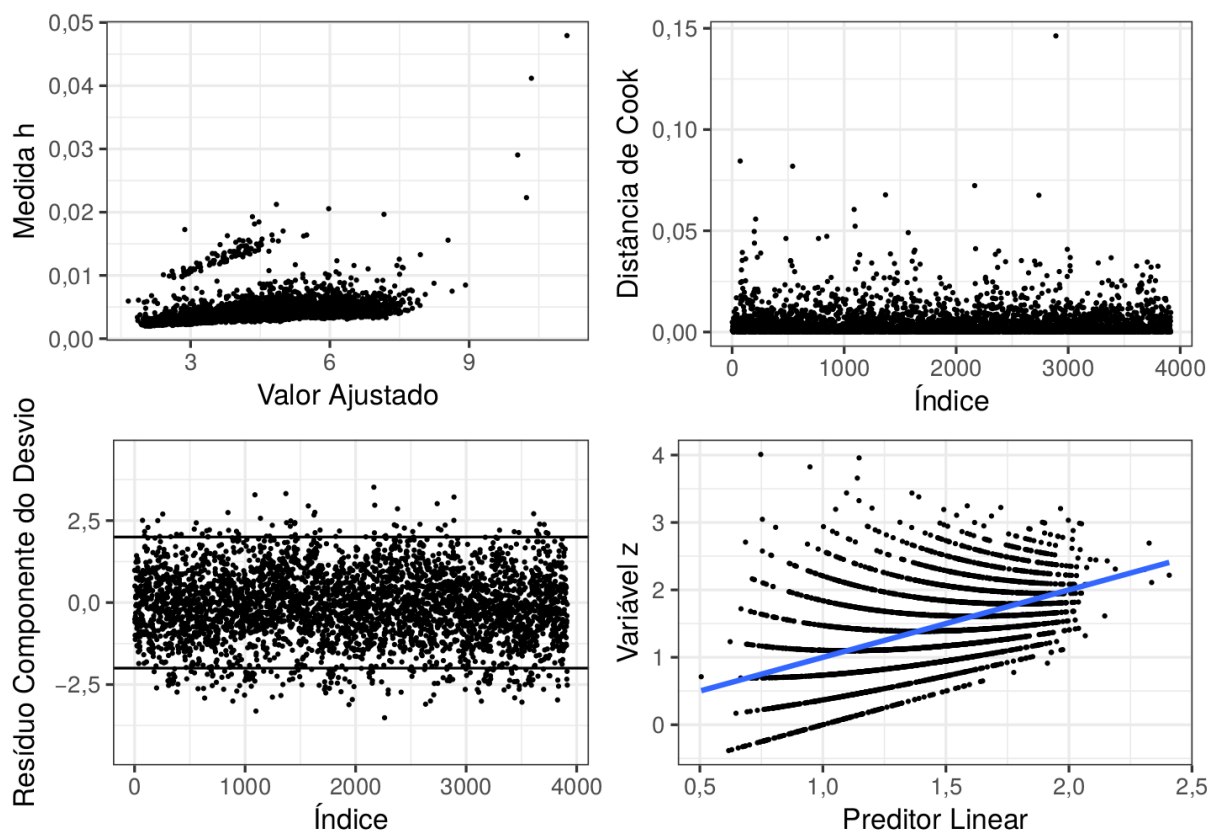


Gráfico B.19. Análise de Resíduos para o modelo de componentes principais: qq plot. Há 295 (7,54 %) pontos fora do envelope.

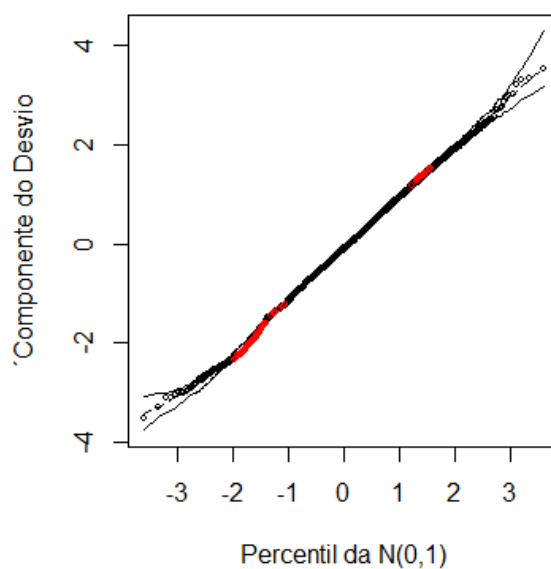


Gráfico B.20. Esquema para seleção de variáveis que elimina a multicolinearidade estrutural.

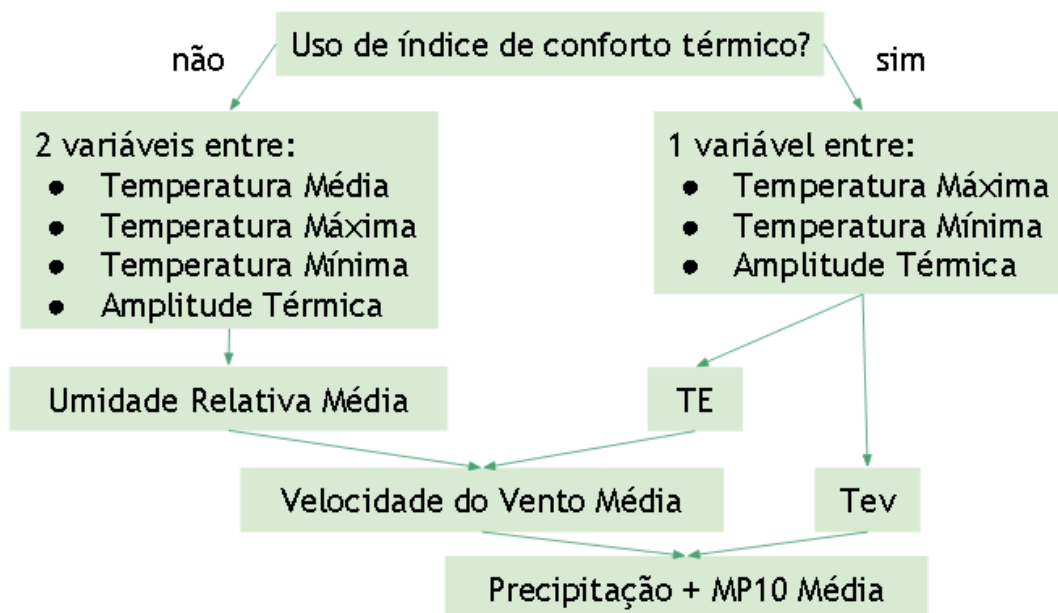


Gráfico B.21. Análise de Resíduos para o modelo ajustado para a estação verão: Medida h, Distância de Cook, Resíduo Componente do Desvio e Preditor Linear.

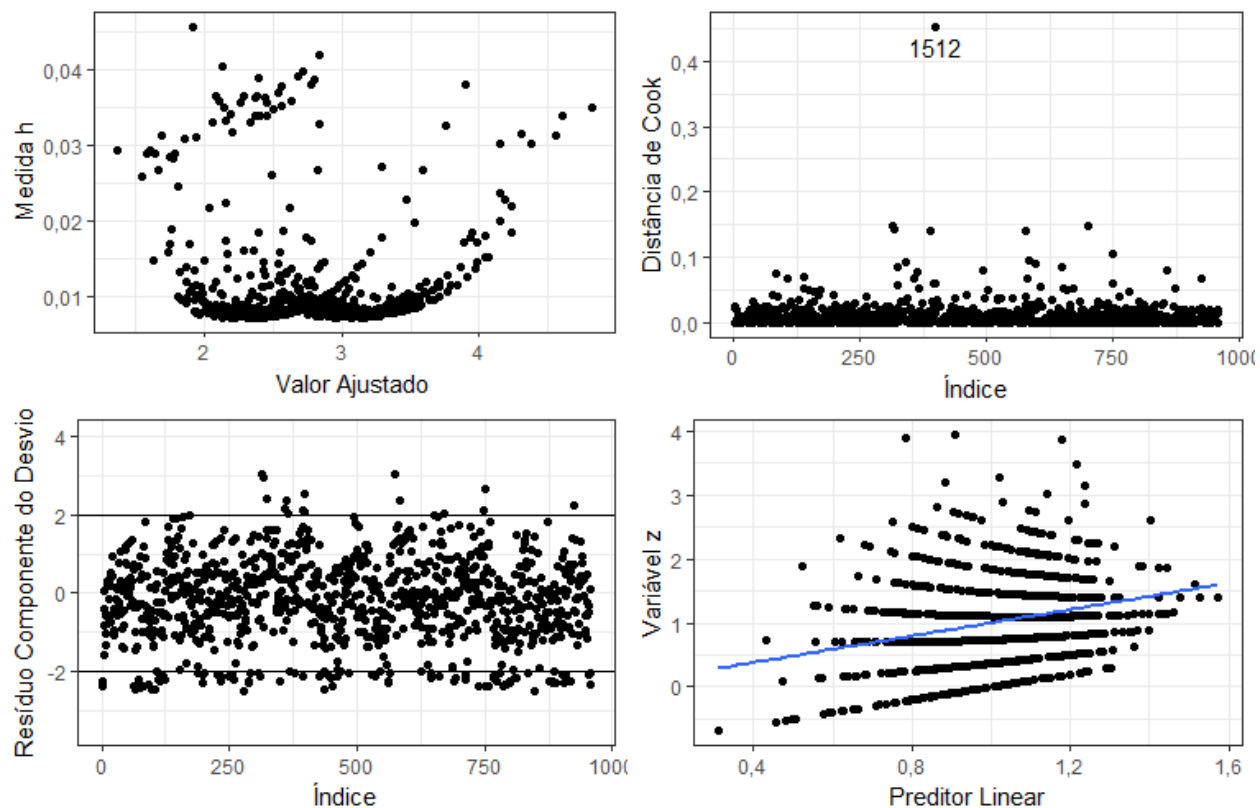


Gráfico B.22. Análise de Resíduos para o modelo da estação verão: qq plot. Há 18 (1,88 %) pontos fora do envelope.

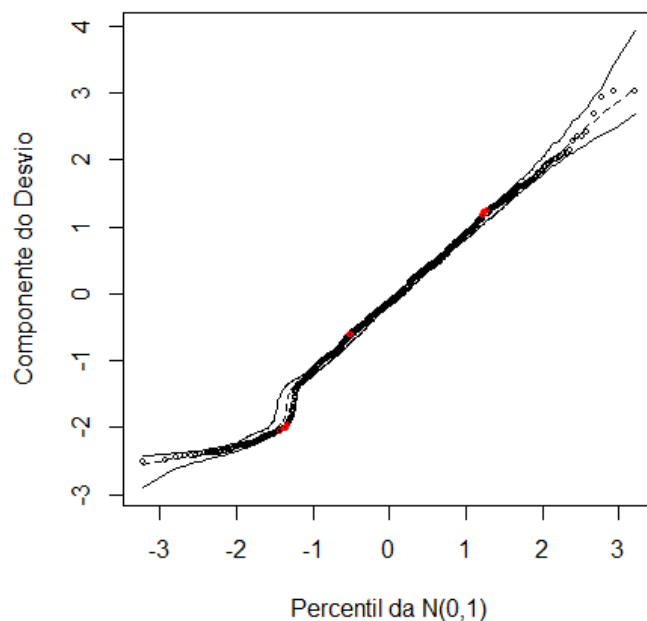


Gráfico B.23. Análise de Resíduos para o modelo ajustado para a estação outono: Medida h, Distância de Cook, Resíduo Componente do Desvio e Preditor Linear.

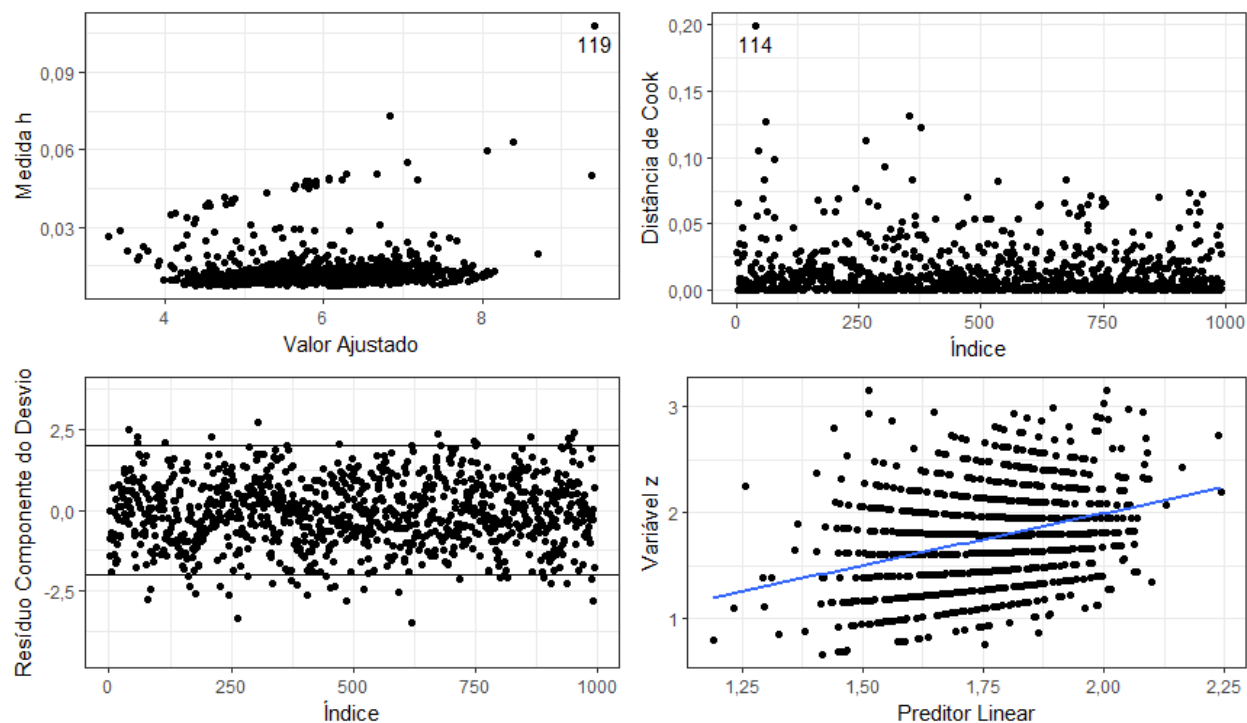


Gráfico B.24. Análise de Resíduos para o modelo da estação outono: qq plot. Há 4 (0,4%) pontos fora do envelope.

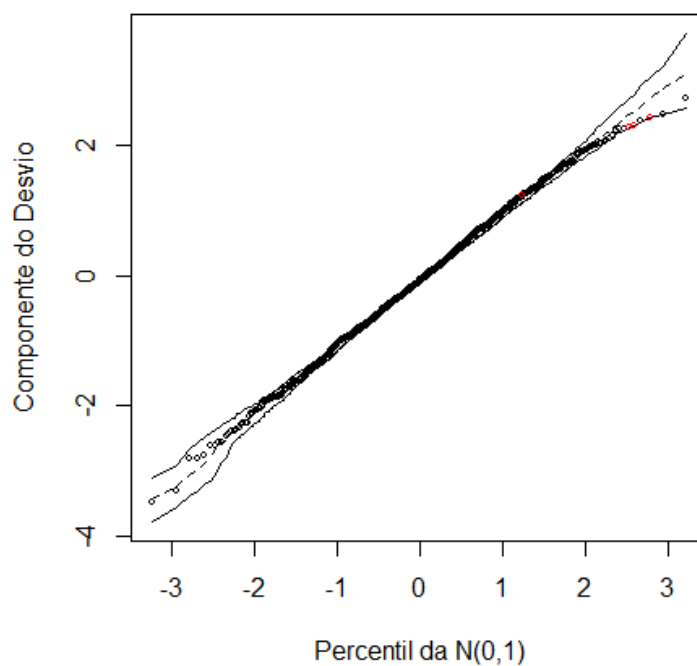


Gráfico B.25. Análise de Resíduos para o modelo ajustado para a estação inverno: Medida h, Distância de Cook, Resíduo Componente do Desvio e Preditor Linear.

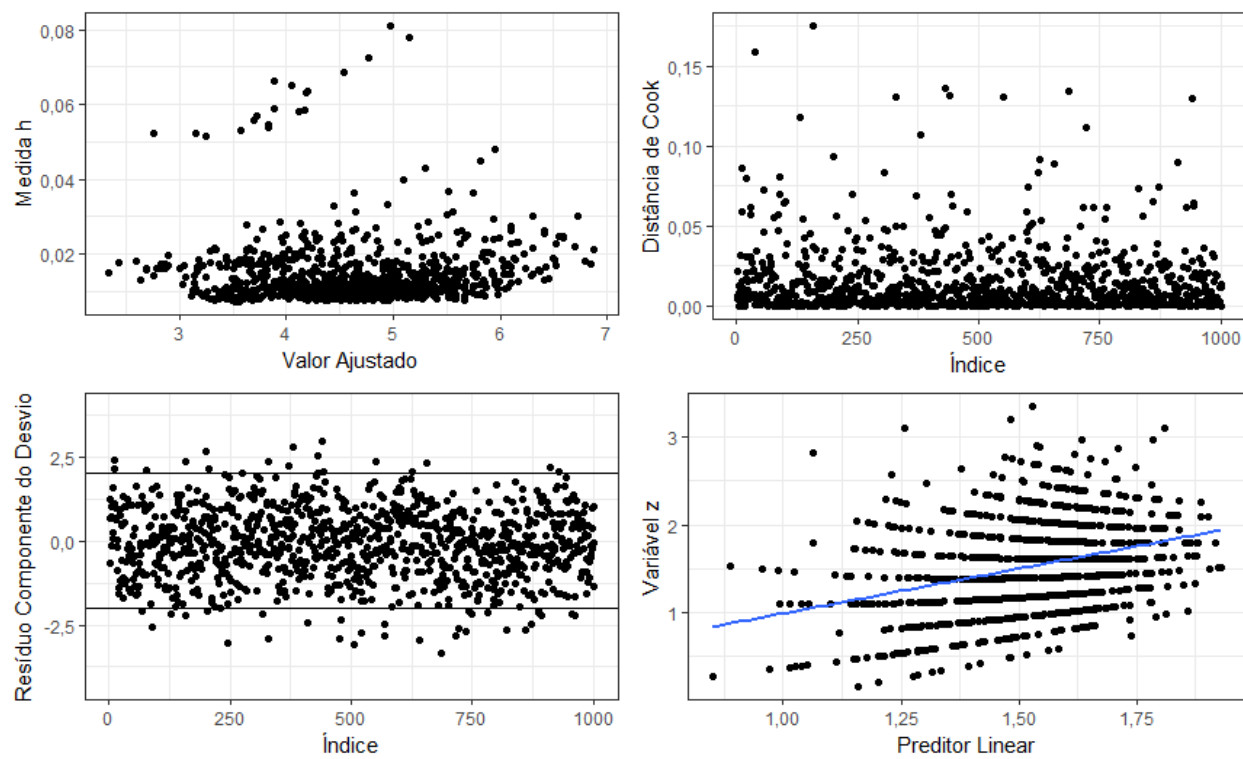


Gráfico B.26. Análise de Resíduos para o modelo da estação inverno: qq plot. Há 121 (12,08%) pontos fora do envelope.

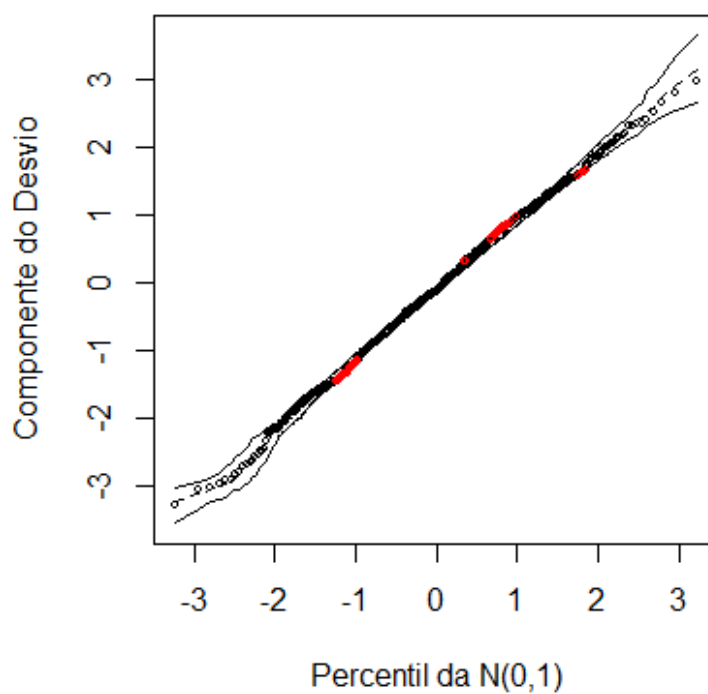
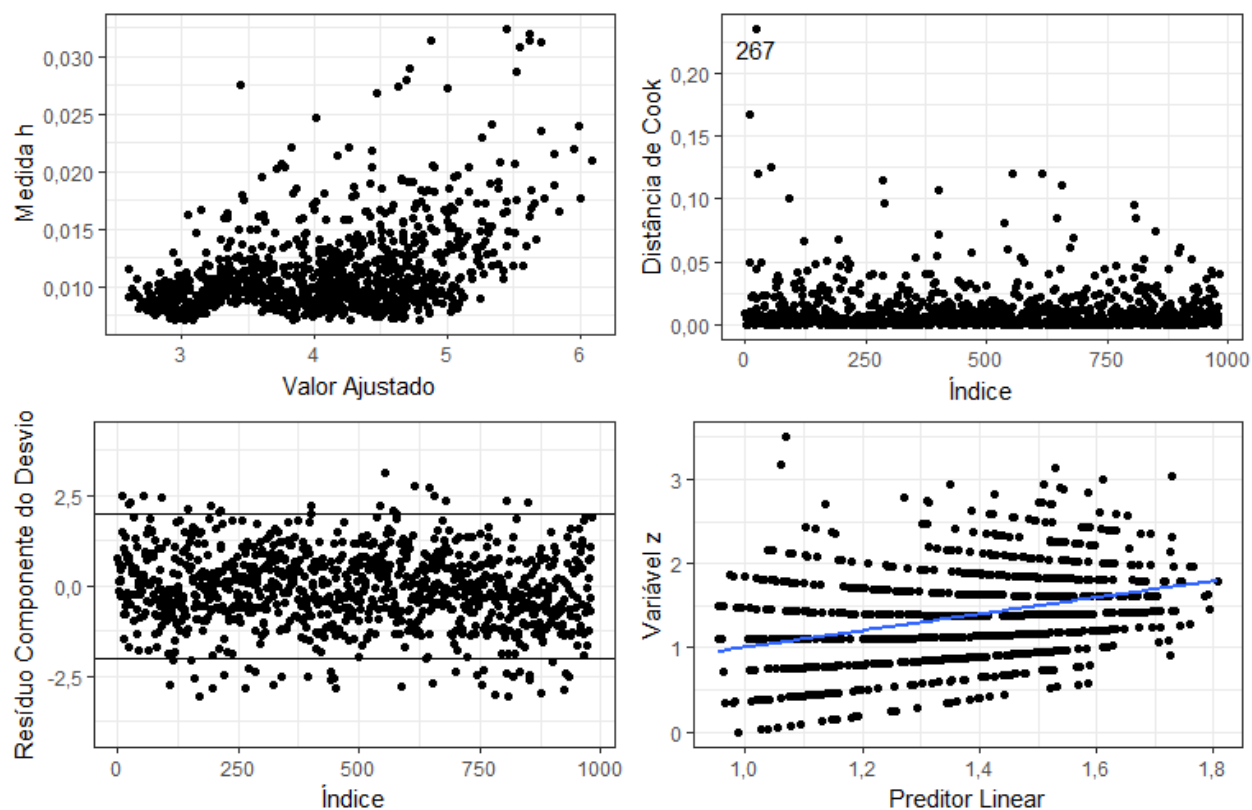
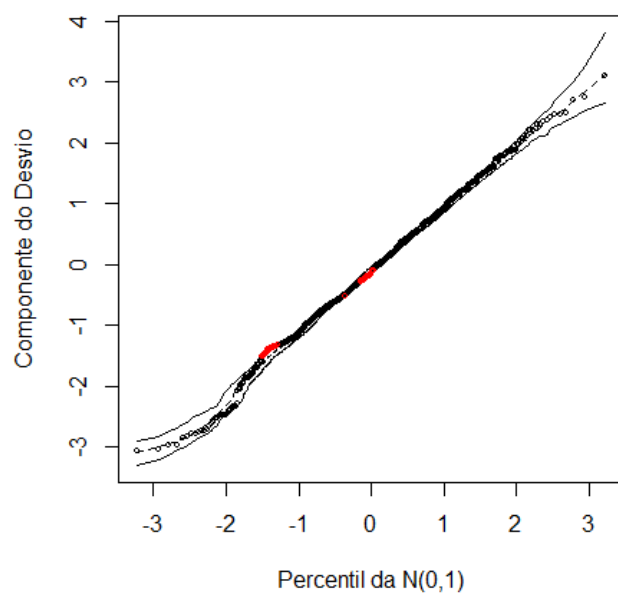


Gráfico B.27. Análise de Resíduos para o modelo ajustado para a estação primavera: Medida h, Distância de Cook, Resíduo Componente do Desvio e Preditor Linear.



67

Gráfico B.28. Análise de Resíduos para o modelo da estação primavera: qq plot. Há 102 (10,4 %) pontos fora do envelope.



APÊNDICE C

Modelo

Modelo de regressão com resposta binomial negativa

Seja y_{ijklmn} o total de interações observado no i -ésimo dia, $i = 1, \dots, 4018$, no j -ésimo dia da semana (domingo ($j=0$), segunda-feira ($j=1$), terça-feira ($j=2$), quarta-feira ($j=3$), quinta-feira ($j=4$), sexta-feira ($j=5$) ou sábado ($j=6$)), no tipo de dia k (feriado ($k = 0$) ou dia normal ($k = 1$)), na l -ésima estação do ano (verão ($l = 0$), outono ($l = 1$), inverno ($l = 2$) ou primavera ($l = 3$)), na onda de calor m (não ($m = 0$) ou sim ($m = 1$)), e na onda de frio n (não ($m = 0$) ou sim ($m = 1$)).

Vamos supor que a resposta tem distribuição binomial negativa (BN), isto é:

$$y_{ijklmn} \sim \text{BN}(\mu_{ijklmn}, \Phi),$$

$$g(\mu_{ijklmn}) = \log(\mu_{ijklmn}) = x_{ijklmn}^T \beta,$$

em que $x_{ijklmn} = (x_{ijklmn1}, \dots, x_{ijklmnp})^T$ contém os valores das p variáveis explicativas, $\beta = (\beta_1, \dots, \beta_p)^T$ é o vetor de coeficientes das variáveis explicativas, $g(\mu_{ijklmn})$ é a função de ligação logarítmica, μ_{ijklmn} é a média da variável resposta y_{ijklmn} ; Φ é seu parâmetro de dispersão; $i = 1, \dots, 4018$; $j = 0, \dots, 6$; $k = 0, 1$; $l = 0, \dots, 3$; $m = 0, 1$ e $n = 0, 1$.