

**Carolina Martins de Medeiros**

**Exposição materna aguda a altas  
temperaturas e o risco de parto prematuro**

Niterói - RJ, Brasil

02 de Agosto de 2022

**Carolina Martins de Medeiros**

**Exposição materna aguda a altas  
temperaturas e o risco de parto  
prematuro**

**Trabalho de Conclusão de Curso**

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

Orientador(a): Profa. Dra. Ludmilla da Silva Viana Jacobson

Niterói - RJ, Brasil

02 de Agosto de 2022

**Carolina Martins de Medeiros**

**Exposição materna aguda a altas  
temperaturas e o risco de parto prematuro**

Monografia de Projeto Final de Graduação sob o título “*Exposição materna aguda a altas temperaturas e o risco de parto prematuro*”, defendida por Carolina Martins de Medeiros e aprovada em 02 de Agosto de 2022, na cidade de Niterói, no Estado do Rio de Janeiro, pela banca examinadora constituída pelas professoras:

---

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

---

**Profa. Dra. Mariana Albi de Oliveira Souza**  
Departamento de Estatística – UFF

---

**Profa. Dra. Patrícia Lusié Velozo da Costa**  
Departamento de Estatística – UFF

Niterói, 02 de Agosto de 2022

Ficha catalográfica automática - SDC/BIME  
Gerada com informações fornecidas pelo autor

M488e Medeiros, Carolina Martins de  
Exposição materna aguda a altas temperaturas e o risco de parto prematuro / Carolina Martins de Medeiros ; Ludmilla da Silva Viana Jacobson, orientadora. Niterói, 2022.  
43 f.

Trabalho de Conclusão de Curso (Graduação em Estatística)-Universidade Federal Fluminense, Instituto de Matemática e Estatística, Niterói, 2022.

1. Prematuridade. 2. Altas Temperaturas. 3. Regressão Logística. 4. Produção intelectual. I. Jacobson, Ludmilla da Silva Viana, orientadora. II. Universidade Federal Fluminense. Instituto de Matemática e Estatística. III. Título.

CDD -

# Resumo

Estima-se que, no mundo, ocorram 15 milhões de nascimentos antes de 37 semanas de gestação, ditos prematuros. Suas consequências para a expectativa de vida do bebê são grandes, podendo levar a morte antes do primeiro mês de vida, ou então aparecer a longo prazo, fazendo com que a qualidade de vida como adulto seja bem difícil. Não há ao certo uma causa para a prematuridade, podendo ser algum fator genético ou fator externo. O objetivo desse trabalho é avaliar se existe associação entre a exposição materna ao calor (altas temperaturas) nos últimos dias de gestação e o nascimento de bebês prematuros na cidade do Rio de Janeiro. Para isso, foram utilizados Modelos de Regressão Logística. Foram utilizados dados do Sistema Nacional de Nascidos Vivos (SINASC) de nascimentos de uma única gestação (não gemelar), com peso acima de 1.500g e sem anomalia congênita do ano de 2019 e os dados do Instituto Nacional de Meteorologia (INMET) do mesmo ano para a verificação da temperatura média, mínima, máxima e aparente dos dias dos nascimentos. A variável resposta foi prematuridade, dividida em dois níveis: os nascimentos antes e depois de 37 semanas de gestação. As variáveis explicativas foram divididas em características da mãe e de nascimento e indicadores de temperatura. Com relação ao primeiro caso, as variáveis foram: idade, estado civil, escolaridade, raça/cor, município de residência, tipo de parto e quantidade de consultas pré-natal. Já para os indicadores de temperatura, as variáveis explicativas foram: amplitude da temperatura, temperatura média, mediana, percentis 75 e 90, temperatura aparente e também suas defasagens de 1 dia (lag). Ao estimar modelos de regressão logística simples para os indicadores de temperatura, foi observado que mediana lag foi o melhor indicador que explicou a prematuridade (OR = 1,020; intervalo de confiança de 95% (IC 95% = 1,010 - 1,029)). A partir disso, foram adicionados as características da mãe e de nascimento e o modelo com as variáveis explicativas mediana lag e quantidade de consultas pré-natal foi o melhor modelo. Para avaliar o efeito do calor na prematuridade, a variável mediana lag foi categorizada com um ponto de corte de 26°C (OR = 1,143; intervalo de confiança de 95% (IC 95% = 1,079 - 1,209)). A variável estação do ano foi uma modificadora de efeito do calor.

Palavras-chave: Prematuridade, Altas Temperaturas, Regressão Logística.

# Dedicatória

Dedico esse trabalho a mim, mesmo com adversidades ao longo dos anos, consegui fazer um bom trabalho.

# Agradecimentos

Primeiramente, quero agradecer a Deus e a Nossa Senhora por me acompanhar em todo esse processo.

Meus maiores agradecimentos são para todos da minha família pela ajuda e apoio a minha educação. Agradeço aos meus pais por sempre colocarem minha educação em primeiro lugar. E principalmente aos meus irmãos por todo companheirismo e por estarem comigo em todos os momentos.

Quero agradecer a todos os amigos que fiz na UFF, por dividirem comigo as alegrias, desesperos e reclamações desses anos de faculdade. Agradeço a Débora Zava, Karolina Mello, Vanessa Medeiros, Jessyka Goltara, Bruna Marques, Mônica Calado por terem feito esses anos os melhores! E a Giuliana Morel, sua confiança em mim me fez acreditar mais ainda que eu conseguiria terminar.

Esse projeto não seria possível sem a Whika Mello, que me ajudou a organizar os últimos semestres, compartilhou comigo os sofrimentos e alegrias desses últimos meses, e também histórias de vida e de pets.

Agradeço aos amigos que fiz no Iepi, Ythalo, Maradona, Clara, Rafaella, Ábia Mariane e Romeika. Vocês fizeram meus dias mais felizes em um dos momentos mais tensos que já vivi e me ajudaram bastante com dúvidas de estatística e referências. Cresci muito com vocês!

Agradeço também à minha orientadora Profa. Dra. Ludmilla da Silva Viana Jacobson pela paciência e por me ajudar a fazer um trabalho incrível.

# Sumário

## Lista de Figuras

## Lista de Tabelas

<b>1</b>	<b>Introdução</b>	p. 11
1.1	Objetivos . . . . .	p. 13
1.2	Organização do Texto . . . . .	p. 14
<b>2</b>	<b>Materiais e Métodos</b>	p. 15
2.1	Área de estudo . . . . .	p. 15
2.2	SINASC - Sistema de Informações sobre Nascidos Vivos . . . . .	p. 16
2.3	INMET - Instituto Nacional de Meteorologia . . . . .	p. 17
2.4	Banco de dados . . . . .	p. 17
2.5	Regressão Logística . . . . .	p. 18
2.5.1	Modelo de Regressão Logística . . . . .	p. 20
2.5.2	Interpretação dos parâmetros do modelo . . . . .	p. 22
2.5.3	Ajustando o Modelo de Regressão Logística Múltipla . . . . .	p. 23
2.5.4	Outras Inferências do Modelo . . . . .	p. 24
2.5.5	Bondade de Ajuste . . . . .	p. 26
2.5.6	Análise de Resíduos . . . . .	p. 26
<b>3</b>	<b>Análise dos Resultados</b>	p. 29
3.1	Análise Descritiva . . . . .	p. 29



3.2	Definições do modelo . . . . .	p. 30
<b>4</b>	<b>Conclusões</b>	p. 40
	<b>Referências</b>	p. 41

# Lista de Figuras

1	Gráfico da sensibilidade (sensitivity) versus 1-especificidade (1-specificity) (JR; LEMESHOW; STURDIVANT, 2013) . . . . .	p. 28
2	Gráficos do efeito do calor com as características da mãe e de nascimento	p. 38
3	Gráficos das curvas ROC dos modelos selecionados. . . . .	p. 39

# Lista de Tabelas

1	Categorização das variáveis das características da mãe e de nascimento	p. 19
2	Tabela dos indicadores de temperatura . . . . .	p. 20
3	Categorias observadas versus previstas . . . . .	p. 26
4	Distribuição de frequências das variáveis de características da mãe e de nascimentos do município do Rio de Janeiro em 2019 . . . . .	p. 31
5	Estatísticas descritivas dos indicadores de temperatura do município do Rio de Janeiro em 2019 . . . . .	p. 32
6	Razão de chances (OR), intervalo de confiança, p-valor do teste de significância Wald e AIC dos modelos de regressão logística simples para as características da mãe e de nascimento . . . . .	p. 33
7	Razão de chances (OR), intervalo de confiança, p-valor do teste de significância Wald e AIC dos modelos de regressão logística simples dos indicadores de temperatura . . . . .	p. 34
8	Razão de chances (OR), intervalo de confiança, p-valor do teste de significância Wald e AIC dos modelos de regressão logística do indicadores mediana lag e características da mãe e de nascimento . . . . .	p. 35
9	Tabela com novas categorizações das variáveis de características da mãe e de nascimento . . . . .	p. 36

# 1 Introdução

Um bebê é dito prematuro quando seu nascimento ocorre antes de completar as 37 semanas gestacionais. Quando nasce entre 32 e 35 semanas de gestação, é considerado como uma criança de risco, e antes de 32 semanas, é considerado de alto risco (VIDAL, 2011). A Organização Mundial de Saúde (OMS) estima que ocorram 15 milhões de partos prematuros todos os anos no mundo, sendo essa a maior causa de morte de crianças com menos de 5 anos de idade (OMS, 2018). O Brasil fica entre os 10 países com o maior número de casos de prematuridade (TABNET, 2020a), e em 2019 esses casos correspondem a cerca de 11% dos nascidos vivos. O Rio de Janeiro obteve taxa parecida no mesmo ano, com 10,86% de partos prematuros dentre os nascidos vivos (TABNET, 2020b).

As consequências de um nascimento pré-termo podem ser definidas a curto, médio e a longo prazo, podendo ser irreversíveis. Além de maior chance de morte no período neonatal, intervalo de tempo desde o nascimento até o momento em que a criança atinge 27 dias, há também chances de intercorrências graves, como enterocolite necrotizante (inflamação que acontece em partes da superfície interna do intestino, o que pode causar infecção por bactérias e levar a necrose neste órgão), hemorragia intraventricular (hemorragia dentro ou ao redor do cérebro), problemas respiratórios e de desenvolvimento e hipertensão pulmonar persistente, entre outras, podendo repercutir ao decorrer da vida, chegando até a vida adulta. Conhecer e compreender o processo do nascimento e os fatores que nele interferem, incluindo aqueles que aumentam o risco de parto prematuro, é fundamental para o cuidado eficiente da mãe e do filho (BALBI; CARVALHAES; PARADA, 2016).

Podem ser inúmeras as causas que acarretam um bebê nascer antes do tempo. Relacionadas à mãe, podem ocorrer problemas associados ao aparelho genital feminino, alterações placentárias (placenta prévia e descolamento prematuro) e excesso de líquido amniótico. Outros fatores incluem: a idade materna (maior incidência em mães mais jovens), infecções maternas e primiparidade (condição de parir ou ter o primeiro parto) (RAMOS; CUMAN, 2009). Através de um estudo de revisão, Carolan-Olah e Frankowska (2014) examinou evidências de que altas temperaturas podem ter relação aos grandes números de

nascimentos prematuros. Suas conclusões apoiaram essa associação, apontando também o estresse por calor, causado por aumento repentino na temperatura ou casos de calor extremo, como uma das causas já que o corpo pode não ser apto a se adaptar rapidamente. Sun et al. (2019) reforçou tais evidências em um estudo retrospectivo com 32 milhões de nascimentos únicos nos EUA, onde foi concluído que dias de calor extremo, mas não frio extremo, estão ligados a maiores riscos de partos prematuros. Também nos EUA, em uma análise case-crossover (comparação do status de exposição do caso imediatamente antes da ocorrência do desfecho com ele mesmo em um momento anterior), Basu, Malig e Ostro (2010) utilizou-se de casos de partos prematuros de um registro estadual de nascimentos da Califórnia, combinados com dados meteorológicos e de monitoramento de poluição do ar com base no código postal residencial, para concluir que alta temperatura ambiental foi significativamente associada com parto prematuro de todas as mães, independentemente do grupo racial/étnico, idade e educação materna ou sexo do bebê. Ainda na Califórnia, Ilango et al. (2020) estimou o risco de parto prematuro em relação a exposição de calor extremo durante a última semana de gestação, tendo descobertas de que exposição aguda ao calor extremo durante a última semana de gestação pode desencadear um parto prematuro.

Na Espanha, Vicedo-Cabrera et al. (2014) avaliou o risco de prematuridade em diferentes valores de temperatura nos meses mais quentes de Valência, (maio à setembro, 2006 - 2010) até 3 semanas antes do parto, tendo como referência o valor médio anual geral. A partir disso, seus resultados indicaram que a exposição a elevadas temperaturas foi associada com o aumento do risco de parto prematuro. Em Brisbane, Austrália, examinando os efeitos da temperatura ambiente nos três trimestres da gravidez, não só nascimentos prematuros como também os natimortos, Li et al. (2018) pôde concluir que os efeitos da baixa temperatura foram mais fortes do que os efeitos da alta temperatura. Em Seoul, na Coreia, Son et al. (2019) investigou associações entre calor e resultados adversos de nascimentos e como características individuais e comunitárias afetam essas associações. Já na China, Wang et al. (2020), por meio de um estudo de coorte com 1.281.859 gravidezes unifetais entre 2013 e 2014, descobriu que a exposição a temperaturas relativamente baixas ou altas durante toda a gravidez aumentou significativamente o risco de partos prematuros. Nessa mesma linha de pensamento, Guo et al. (2018) investigaram as exposições de temperaturas e resultados de nascimentos em 132 cidades, dividindo as áreas de estudo em três categorias: área fria, média e quente, de acordo com a média local de temperatura. Foi concluído que a exposição aguda e crônica à temperaturas extremas pode afetar o risco de prematuridade. Li et al. (2021) investigou como a variação da

temperatura também pode ser um fator de risco de prematuridade, chegando à conclusão de que o maior risco ocorre na estação mais quente.

Estudos têm sugerido o aumento da temperatura global (KRAAIJENBRINK et al., 2017). A Organização Meteorológica Mundial (OMM) considerou a década entre 2011 e 2020 a mais quente da história (MUNDO..., 2020), tendo 2020 empatado com 2016 como o ano com maiores temperaturas pela Copernicus Climate Change Service (COPERNICUS..., 2021). Segundo dados das estações meteorológicas do Instituto Nacional de Meteorologia (INMET), a temperatura do Brasil vem aumentando desde a década de 90 (SANTOS, 2021), sendo o ano mais quente já registrado o de 2019 (FIORAVANTI, 2020).

No estado do Rio de Janeiro, a temperatura varia com relação a sua região, podendo fazer frio na Região Serrana enquanto faz sol no Norte Fluminense. Na região Metropolitana, há um fenômeno chamado Ilhas de Calor, que são fenômenos climáticos onde o elevado grau de urbanização faz com que a temperatura média seja mais alta do que nas zonas rurais. A cidade do Rio de Janeiro é um exemplo de ilha de calor, devido a sua grande concentração de asfalto e concreto nas ruas e prédios. A concentração de calor é grande e, com isso, a temperatura fica acima da média na cidade do Rio de Janeiro e nos municípios no entorno, assim como a umidade do ar fica baixa nessas áreas <sup>1</sup>.

## 1.1 **Objetivos**

Para esse projeto, o objetivo geral foi investigar a existência de associação de nascimentos de bebês prematuros e exposição materna ao calor (altas temperaturas) nos últimos dias de gestação no município do Rio de Janeiro. Já os objetivos específicos temos listados a seguir:

- avaliar diferentes variáveis de exposição relacionadas a temperatura como a temperatura média, amplitude da temperatura (que é a diferença entre a temperatura máxima e mínima) e a temperatura aparente, que leva em consideração não só a temperatura do ar como também a umidade e o vento; e
- identificar as variáveis modificadoras do efeito da temperatura, como características da mãe e local de nascimento.

---

<sup>1</sup>RIO DE JANEIRO, Prefeitura da Cidade do. *O Rio Visto do Espaço: Ilhas de Calor Urbano e Mudanças Climáticas*. Rio de Janeiro, 2016. Disponível em: <https://www.data.rio/apps/o-rio-visto-do-espaco-ilhas-de-calor-urbano-e-mudancas-climaticas/explore>. Acesso em janeiro de 2022

## 1.2 Organização do Texto

O Capítulo 2 mostra os materiais e métodos utilizados neste trabalho como a área de estudo, no caso o Rio de Janeiro, as origens das bases de dados SINASC e INMET, a organização da base utilizada e o método modelo de regressão logística, que foi o método utilizado neste estudo. No Capítulo 3, contém os resultados das análises descritiva e também da implementação dos modelos de regressão logística. Já o Capítulo 4 mostra as conclusões deste estudo e indicações de trabalhos futuros.

## 2 Materiais e Métodos

Neste capítulo serão apresentados os materiais e métodos utilizados neste trabalho. Na Seção 2.1 contém informações sobre o município do Rio de Janeiro, a área de interesse desse estudo. As Seções 2.2 e 2.3 são apresentados o Sistema de Informação sobre Nascidos Vivos (SINASC) e o Instituto Nacional de Meteorologia (INMET), de onde foram selecionados os dados de nascimentos e temperatura. Na Seção 2.4 são mostradas as variáveis de interesse deste estudo. Já na Seção 2.5 é desenvolvido a regressão logística, o método estatístico utilizado neste trabalho.

### 2.1 Área de estudo

A área de estudo deste projeto foi o município do Rio de Janeiro. Localizado na região sudeste brasileira, é a capital do estado que recebe o mesmo nome e que é a segunda maior área urbanizada do Brasil (MAPAS..., 2018). Sua área territorial é de 1.200,329 km<sup>2</sup><sup>1</sup>, sendo desta, 600 km<sup>2</sup> de área urbanizada<sup>2</sup>. Sua população é estimada em 6.775.561 pessoas<sup>3</sup> e sua densidade demográfica é de 5.265,82 hab/km<sup>2</sup><sup>4</sup>.

O clima da cidade é dito como do tipo tropical, quente e úmido, com variações locais, devido às diferenças de altitude, vegetação e proximidade do oceano. A temperatura média anual é de 22° centígrados, com médias diárias elevadas no verão (de 30° a 32°); as chuvas variam de 1.200 a 1.800 mm anuais. Nos quatro meses do chamado alto verão (de dezembro a março) os dias muito quentes são sempre seguidos de tardes luminosas,

---

<sup>1</sup>IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. *Áreas Territoriais*. Rio de Janeiro, 2021. Disponível em: <https://www.ibge.gov.br/geociencias/organizacao-do-territorio/estrutura-territorial/15761-areas-dos-municipios.html?=&t=acesso-ao-produto>. Acesso em outubro de 2021

<sup>2</sup>INSTITUTO PEREIRA PASSOS. *Rio em Síntese*. Rio de Janeiro, 2019. Disponível em: <https://www.data.rio/pages/rio-em-sntese-2>. Acesso em janeiro de 2022

<sup>3</sup>IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. *Estimativas da População*. 2021. Disponível em: <https://www.ibge.gov.br/estatisticas/sociais/populacao/9103-estimativas-de-populacao.html?=&t=resultados>. Acesso em janeiro de 2022.

<sup>4</sup>IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. *Censo Brasileiro de 2010*. Rio de Janeiro: IBGE, 2012.



quando em geral caem chuvas fortes e rápidas, trazendo noites frescas e estreladas <sup>5</sup>.

Segundo Fiocruz (2014), que mapeou o estado do Rio de Janeiro de modo a indicar a exposição dos municípios às mudanças climáticas previstas para os próximos 30 anos, a população do município do Rio teve elevados índices de vulnerabilidade da saúde e do ambiente quando comparado ao dos demais municípios do Estado, fazendo com que ficassem entre os mais vulneráveis.

## 2.2 SINASC - Sistema de Informações sobre Nascidos Vivos

O Sistema de Informações sobre Nascidos Vivos (SINASC) foi implantado pela primeira vez em 1990 pelo Ministério Público, tendo como um dos principais objetivos o de permitir obter um perfil epidemiológico dos nascimentos, segundo variáveis como: peso ao nascer, índice de Apgar, duração de gestação, tipo de parto e paridade. Além disso, como muitas dessas informações também estão registradas na Declaração de Óbito (DO) torna-se possível a obtenção de coeficientes específicos de mortalidade infantil, necessários para análises mais minuciosas, na área de saúde materno-infantil (JORGE et al., 1993).

O SINASC baseia-se nos dados contidos na Declaração de Nascido Vivo (DNV), documento oficial e padronizado para todo o país. A DNV é de emissão obrigatória nos hospitais e outras instituições de saúde nos quais são realizados partos. Os Cartórios do Registro Civil também apresentam tal obrigatoriedade relacionada com os partos ocorridos no domicílio e cuja informação chega a esse tipo de instituição (PEDRAZA, 2012).

De uma maneira geral, as Secretarias Estaduais de Saúde enviam a sua Base de Dados para o Ministério da Saúde, quando consideram a coleta completa. A Organização Mundial de Saúde (OMS) só pode considerar a Base Nacional completa quando todas as UF enviarem seus dados. A partir daí, é feita a consolidação, inclusive com a redistribuição dos dados pelo local de residência, a qual é a forma tradicional de apresentar os dados de Nascidos Vivos. Eventualmente, são feitas algumas correções nas informações (DATASUS, 2017).

---

<sup>5</sup>RIO DE JANEIRO, Prefeitura da Cidade do. Rio de Janeiro, 2009. Disponível em: <https://www.rio.rj.gov.br/web/riotur/caracteristicas-geograficas>. Acesso em janeiro de 2022.

## 2.3 INMET - Instituto Nacional de Meteorologia

O Instituto Nacional de Meteorologia (INMET) é um órgão do Ministério da Agricultura, Pecuária e Abastecimento e tem como objetivo prover informações meteorológicas à sociedade brasileira e influir construtivamente no processo de tomada de decisão, contribuindo para o desenvolvimento sustentável do País. Esta missão é alcançada por meio de monitoramento, análise e previsão de tempo e de clima, que se fundamentam em pesquisa aplicada, trabalho em parceria e compartilhamento do conhecimento, com ênfase em resultados práticos e confiáveis <sup>6</sup>.

No Brasil, o INMET administra mais de 400 estações meteorológicas. Possui 10 Distritos Regionais que recebem, processam e enviam estes dados para a Sede, localizada em Brasília-DF. A sede, por sua vez, processa estes dados e os envia por satélite para todo o mundo <sup>7</sup>.

Existem dois tipos de estações meteorológicas: as estações convencionais, que utilizam equipamentos mecânicos e necessitam de pessoal para observar e registrar manualmente os dados, três vezes ao dia (12:00, 18:00 e 24:00 UTC), e as estações automáticas, com equipamentos eletrônicos programáveis para a observação e o registro nos intervalos de tempo previamente escolhidos para arquivamento de valores médios, máximos, mínimos ou totais, geralmente a cada 1 hora (KNIPPELBERG, 2019).

## 2.4 Banco de dados

Os dados de nascimentos utilizados neste trabalho foram originários do SINASC, onde foram selecionados os bebês que nasceram com mais de 28 semanas (nascimentos de menos de 28 semanas são considerados extremamente prematuros (OMS, 2018)), de uma única gestação (não gemelar), com peso acima de 1.500g (segundo Wardlaw (2004), o peso normal é acima de 2.500g e extremamente baixo é menos de 1.500g ao nascer) e sem anomalia congênita do ano de 2019. As covariáveis associadas à gestante e ao nascimento são idade, estado civil, escolaridade, raça/cor, município de residência, tipo de parto, quantidade de consultas pré-natal. A variável dependente foi prematuridade, de modo que:

---

<sup>6</sup>INMET - INTITUTO NACIONAL DE METEOROLOGIA. Disponível em: <https://portal.inmet.gov.br/sobre>. Acesso em janeiro de 2022

<sup>7</sup>INMET - INTITUTO NACIONAL DE METEOROLOGIA. Disponível em: <https://portal.inmet.gov.br/sobre-meteorologia>. Acesso em janeiro de 2022

$$\text{Prematuridade} = \begin{cases} 1, & \text{se a gestação durou menos de 37 semanas} \\ 0, & \text{se a gestação durou mais ou igual à 37 semanas} \end{cases}$$

A temperatura foi obtida a partir da base do INMET do mesmo ano. Foram utilizados os dados de estações automáticas, sendo elas um total de quatro estações localizadas em Marambaia, Jacarepaguá, Vila Militar e Forte de Copacabana. Os dados foram coletados de hora em hora, e foram utilizados para os cálculos de indicadores de temperatura como amplitude da temperatura, temperatura média, mediana, percentis 75 e 90, temperatura aparente e também suas defasagens.

A temperatura aparente (ou sensação térmica) é dita como a temperatura percebida derivada da combinação da temperatura e do vento (*Wind Chill*) ou da temperatura com a umidade relativa do ar (*Heat Index*) para hora indicada. Quando a temperatura cai para 10°C ou menos em um momento em particular, *wind chill* será usado para determinar a temperatura aparente desse momento. Quando a temperatura no momento sobe para acima de 26,6°C, o *heat index* que será utilizado. Entre 11°C e 26,6°C a temperatura aparente será a temperatura ambiente do ar<sup>8</sup>. Como a temperatura média anual do Rio é de 22°C<sup>9</sup>, iremos fazer uso da temperatura ambiente e do *heat index* para determinar a temperatura aparente. O cálculo do *heat index* é dado como:

$$\begin{aligned} HI = & -42,38 + 2,049T + 10,14R - 0,2248TR - 0,006838\sqrt{T} \\ & -0,05482\sqrt{R} + 0,001228T^2R + 0,0008528TR^2 - 0,00000199T^2R^2 \end{aligned} \quad (2.1)$$

onde  $HI$  = Heat Index,  $T$  = temperatura (°C) e  $R$  = umidade relativa (%)<sup>8</sup>.

As variáveis utilizadas foram apresentadas nas Tabelas 1 e 2.

## 2.5 Regressão Logística

Para se fazer uma análise de dados onde é procurada uma relação entre uma variável resposta (variável de interesse do experimento) e duas ou mais variáveis explicativas, geralmente se recorre à regressão linear, seja ela simples ou múltipla. O termo “regressão” foi usado pela primeira vez quando o Sir Francis Galton, em 1885, fez um estudo ma-

<sup>8</sup> *What is Apparent Temperature?* Disponível em: <https://meteor.geol.iastate.edu/~ckarsten/bufkit/apparent.temperature.html>. Acesso em janeiro de 2022.

<sup>9</sup> RIO DE JANEIRO, Prefeitura da Cidade do. Rio de Janeiro, 2009. Disponível em: <https://www.rio.rj.gov.br/web/riotur/caracteristicas-geograficas>. Acesso em janeiro de 2022.

Tabela 1: Categorização das variáveis das características da mãe e de nascimento

Variável	Categorias
Idade da mãe	Até 19 anos Entre 20 e 29 anos Entre 30 e 39 anos A partir de 40 anos
Raça/Cor	Branca Preta Amarela Parda Indígena
Estado civil	Solteira Casada Viúva Separada judicialmente/Divorciada
Escolaridade	Até Ensino Fundamental, 0 a 7 anos de estudos Ensino Médio, 8 a 11 anos de estudos Superior, maior que 12 anos de estudos
Município de Residência	Rio De Janeiro Outro
Tipo de parto	Normal Cesáreo
Quantidade de Consultas Pré-Natal	Nenhuma consulta Realizou 1 a 3 consultas Realizou 4 a 6 consultas Realizou 7 ou mais consultas
Estações do Ano	Verão Outono Inverno Primavera

temático querendo demonstrar que a altura dos filhos não está relacionada com a altura de seus pais, mas que tem a tendência de “regredir” ou aproximar a altura média da população (SARMENTO; COSTA, 2017). No caso do Sir Galton, a variável resposta (altura dos filhos) é do tipo quantitativa contínua, onde para cada filho pesquisado, houve apenas uma resposta. No caso da variável resposta de uma regressão linear ser categórica, usar a regressão logística é o natural a se fazer (EL-HABIL, 2012).

Tabela 2: Tabela dos indicadores de temperatura

<b>Indicador</b>	<b>Identificação</b>
Amplitude	Média das amplitudes diárias das estações meteorológicas em °C
Média	Média das médias diárias das estações meteorológicas em °C
Mediana	Mediana das médias diárias das estações meteorológicas em °C
Máximo	Máximo das médias diárias das estações meteorológicas em °C
Percentil 75	Percentil 75 das médias diárias das estações meteorológicas em °C
Percentil 90	Percentil 90 das médias diárias, temperatura aparente das estações meteorológicas em °C
Temperatura Aparente	Temperatura aparente das médias diárias das estações meteorológicas em °C
Amplitude Lag	Amplitude do dia anterior em °C
Média Lag	Média do dia anterior em °C
Mediana Lag	Mediana do dia anterior em °C
Máximo Lag	Máximo do dia anterior em °C
Percentil 75 Lag	Percentil 75 do dia anterior em °C
Percentil 90 Lag	Percentil 90 do dia anterior em °C
Temperatura Aparente Lag	Temperatura Aparente do dia anterior em °C

### 2.5.1 Modelo de Regressão Logística

O modelo de regressão logística é considerado um Modelo Linear Generalizado em que a variável resposta é mensurada de modo binário (DOBSON; BARNETT, 2018). Para o caso deste trabalho, a variável resposta é prematuridade, como foi mostrado na seção anterior, sendo assim, uma variável binária. Dentre os modelos existentes, o escolhido foi

o modelo de regressão logística. Neste modelo, a prematuridade é mostrada tal que

$$Y = \begin{cases} 1, & \text{se a gestação durou menos de 37 semanas} \\ 0, & \text{se a gestação durou mais ou igual à 37 semanas} \end{cases} \quad (2.2)$$

com probabilidades  $P(Y = 1) = \pi$  a probabilidade de sucesso e  $P(Y = 0) = 1 - \pi$  a probabilidade de fracasso. Nesse caso,  $Y$  tem distribuição Bernoulli com média  $\pi$  ( $Y \sim \text{Bern}(\pi)$ ). A função de probabilidade de  $Y$  é dada por

$$P(Y = y) = \pi^y(1 - \pi)^{1-y}, \quad y = 0, 1. \quad (2.3)$$

A distribuição Bernoulli pertence a família exponencial, um conjunto de distribuições com características em comum. A definição de família exponencial é dada a seguir (DOBSON; BARNETT, 2018):

**Definição 2.1** *Considere a variável aleatória  $Y$  em que a distribuição de probabilidade depende de um único parâmetro  $\theta$ . A distribuição pertence a família exponencial se pode ser escrita da forma*

$$f(y; \theta) = \exp[a(y)b(\theta) + c(\theta) + d(y)]. \quad (2.4)$$

onde  $a$ ,  $b$ ,  $c$  e  $d$  são funções conhecidas.

Se  $a(y) = y$ , a distribuição é dita como forma canônica da função de probabilidade e  $b(\theta)$  ocasionalmente é chamada de parâmetro natural da distribuição.

Assim, a função de probabilidade de  $Y$  pode ser escrita como

$$\begin{aligned} f(y; \pi) &= \pi^y(1 - \pi)^{1-y} \\ &= \exp[y \log(\pi) + (1 - y) \log(1 - \pi)] \\ &= \exp[y \log(\pi) + \log(1 - \pi) - y \log(1 - \pi)] \\ &= \exp[y(\log(\pi) - \log(1 - \pi)) + \log(1 - \pi)] \\ &= \exp \left[ y \log \left( \frac{\pi}{1 - \pi} \right) + \log(1 - \pi) \right] \end{aligned} \quad (2.5)$$

em que  $a(y) = y$ ,  $b(\pi) = \log \left( \frac{\pi}{1 - \pi} \right)$ ,  $c(\pi) = \log(1 - \pi)$  e  $d(y) = 0$ .

Um Modelo Linear Generalizado é composto por 3 componentes: componente aleatório, componente sistemático e função de ligação. No caso do modelo de regressão logística, os componentes são:

- **Componente aleatório:** composto pela variável aleatória  $Y_i$ ,  $i = 1, \dots, n$ , que

neste trabalho representa se o parto ocorreu antes ou depois de 37 semanas de gestação, com distribuição Bernoulli com média  $E(Y_i) = \mu_i = \pi_i$  pertencente a família exponencial.

- **Componente sistemático:** composto por um preditor linear  $\eta_i = \mathbf{X}_i^T \boldsymbol{\beta}$ , onde  $\boldsymbol{\beta} = (\beta_0, \dots, \beta_p)$  é um vetor de parâmetros desconhecidos e  $\mathbf{X}_i = (1, X_1, \dots, X_p)$  vetor de covariáveis.
- **Função de ligação:** é uma função monôtonica diferenciável,  $g(\mu_i)$ , que liga a média da variável resposta  $E(Y_i) = \mu_i$  ao preditor linear  $\eta_i = \mathbf{X}_i^T \boldsymbol{\beta}$ , ou seja,  $g(\mu_i) = \eta_i = \mathbf{X}_i^T \boldsymbol{\beta}$ ,  $\forall i = 1, \dots, n$ . No caso da regressão logística, a função de ligação escolhida foi a função logit, dada como

$$\text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1 - \pi_i}\right) \quad (2.6)$$

Como

$$\mu_i = E(Y_i) = \beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip} = \mathbf{X}_i^T \boldsymbol{\beta} \quad (2.7)$$

temos que

$$\text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1 - \pi_i}\right) = \mathbf{X}_i^T \boldsymbol{\beta} \quad (2.8)$$

Existe uma fórmula correspondente para  $\pi_i$  aplicando exponencial nos dois lados da equação e fazendo as operações matemáticas necessárias. Logo,

$$\pi_i = \frac{e^{\mathbf{X}_i^T \boldsymbol{\beta}}}{1 + e^{\mathbf{X}_i^T \boldsymbol{\beta}}} \quad (2.9)$$

## 2.5.2 Interpretação dos parâmetros do modelo

Uma forma de interpretar os resultados dos parâmetros de um modelo de regressão logística é a partir da razão de chances (*odds ratio* em inglês). A chance é a probabilidade de um evento ocorrer ( $\pi$ ) dividido pela probabilidade desse evento não ocorrer ( $1 - \pi$ ).

$$\text{chance} = \frac{\pi}{1 - \pi}$$

A razão de chances (*OR*) de dois grupos A e B é definida como:

$$OR = \frac{\text{chance}_A}{\text{chance}_B}$$

Quando a *OR* for maior que 1, significa que a chance do evento ocorrer no grupo A é

maior do que no grupo B. Já quando a  $OR$  for menor que 1, a chance do evento ocorrer no grupo A é menor do que no grupo B. Ao fim, para  $OR = 1$ , a chance do evento ocorrer é igual para os grupos A e B.

A partir das estimativas do modelo, podemos fazer uma estimativa da  $OR$ . Para uma variável  $X_j$  categórica, estimamos a  $OR$  entre uma categoria de interesse e uma segunda categoria da variável. O valor estimado é  $\widehat{OR} = \exp(\hat{\beta}_j)$ . Já para uma variável  $X_j$  contínua, a estimativa continua sendo  $\widehat{OR} = \exp(\hat{\beta}_j)$ , a interpretação muda, a  $\widehat{OR}$  aumenta (ou diminui) ao aumentar em uma unidade a variável. O intervalo de confiança de  $\widehat{OR}$  é dado ao exponenciar o intervalo de confiança de  $\hat{\beta}$  para os dois casos.

### 2.5.3 Ajustando o Modelo de Regressão Logística Múltipla

Nesse momento, é preciso estimar os valores de  $\beta$  e o método escolhido para tal estimação foi o de máxima verossimilhança. Primeiro temos que construir a *função de verossimilhança*, que expressa a probabilidade dos dados observados como função dos parâmetros desconhecidos. Os *estimadores de máxima verossimilhança* dos parâmetros são os valores que maximizam a função de máxima verossimilhança (AGRESTI, 2018).

A função de verossimilhança,  $L$ , dos parâmetros foi definida como um produto das probabilidades de sucesso e fracasso de todas as observações da amostra.

$$\begin{aligned} L(\beta|\mathbf{y}) &= \prod_{i=1}^n P(y_i|x_{i1}, \dots, x_{ip}) \\ &= \prod_{i=1}^n [\pi_i^{y_i} \times (1 - \pi_i)^{1-y_i}] \end{aligned} \quad (2.10)$$

Aplicando o logaritmo natural de ambos os lados da equação, temos a função de log-verossimilhança:

$$\begin{aligned} \log(L(\beta|\mathbf{y})) &= \log \left[ \prod_{i=1}^n [\pi_i^{y_i} \times (1 - \pi_i)^{1-y_i}] \right] \\ &= \sum_{i=1}^n [y_i \log(\pi_i) + (1 - y_i) \log(1 - \pi_i)] \end{aligned} \quad (2.11)$$

Geralmente, o estimador  $\hat{\beta}$  é obtido ao derivar a função de log-verossimilhança em relação a cada elemento  $\beta_j$  de  $\beta$  e resolvendo simultaneamente as equações: (DOBSON;



BARNETT, 2018)

$$\frac{\partial \log(L(\boldsymbol{\beta}|\mathbf{y}))}{\partial \beta_j} = 0 \quad j = 0, \dots, p \quad (2.12)$$

É necessário checar se as soluções correspondem aos máximos de  $\log(L(\boldsymbol{\beta}|\mathbf{y}))$  verificando se a matrix de segundas derivadas

$$\frac{\partial^2 \log(L(\boldsymbol{\beta}|\mathbf{y}))}{\partial \beta_j \partial \beta_k}$$

avaliado em  $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$  é definida como negativa. Por exemplo, se  $\boldsymbol{\beta}$  tem apenas um elemento  $\beta$ , significa que é necessário checar se

$$\left[ \frac{\partial^2 \log(L(\boldsymbol{\beta}|\mathbf{y}))}{\partial^2 \beta_j} \right]_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} < 0$$

Para o caso do modelo de regressão logística, temos que:

$$\frac{\partial \log(L(\boldsymbol{\beta}|\mathbf{y}))}{\partial \beta_j} = \sum_{i=1}^n \left[ x_{ij} y_i - x_{ij} \frac{e^{\mathbf{x}_i^T \boldsymbol{\beta}}}{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}} \right], \quad j = 0, \dots, p. \quad (2.13)$$

Igualando a equação 2.13 a 0, são encontradas  $p + 1$  equações não-lineares que são resolvidas a partir de método numérico iterativo (método de escore de Fisher). Estas soluções resultam nos estimadores de máxima verossimilhança  $\hat{\beta}_j$ ,  $j = 0, \dots, p$ .

Logo, a função logit é estimada como:

$$\text{logit}(\hat{\pi}_i) = \log \left( \frac{\hat{\pi}_i}{1 - \hat{\pi}_i} \right) = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_p x_{ip} = \mathbf{x}_i^T \hat{\boldsymbol{\beta}} \quad (2.14)$$

e também,

$$\hat{\pi}_i = \frac{e^{\mathbf{x}_i^T \hat{\boldsymbol{\beta}}}}{1 + e^{\mathbf{x}_i^T \hat{\boldsymbol{\beta}}}} \quad (2.15)$$

## 2.5.4 Outras Inferências do Modelo

Após a estimação dos coeficientes, há o interesse em assegurar a significância das variáveis do modelo através de formulação e testes de hipóteses estatísticos para determinar se as variáveis independentes são significativamente relacionadas com a variável dependente. O teste Wald a seguir foi baseado no texto de Taconeli (2019)

### Teste Wald

O teste Wald é um teste simples que explora a normalidade de  $\boldsymbol{\beta}$  estimado pelo método

de máxima verossimilhança. Seja SE o erro padrão irrestrito de  $\beta$  avaliado ao substituir a estimativa de máxima verossimilhança para o parâmetro desconhecido na expressão para o verdadeiro erro padrão. Para  $n \rightarrow \infty$ , temos que:

$$\hat{\beta}_j \sim Normal(\beta_j, SE(\hat{\beta}_j)) \quad (2.16)$$

onde o erro padrão de  $\hat{\beta}_j$  foi denotado por  $SE(\hat{\beta}_j) = \sqrt{Var(\hat{\beta}_j)}$ .

Considere as hipóteses a seguir:

$$\begin{cases} H_0 : \beta_j = 0 \\ H_1 : \beta_j \neq 0 \end{cases}, \quad j = 0, \dots, p$$

O teste Wald se baseia na seguinte estatística de teste, sob a hipótese nula:

$$Z_t = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \sim Normal(0, 1) \quad (2.17)$$

Para a tomada de decisão, teremos um p-valor, uma probabilidade de se observar um valor para a estatística de teste tão ou maior que o encontrado. O nível  $\alpha$  atua como uma espécie de medida de corte, significando que rejeitaremos a hipótese nula se p-valor for menor que  $\alpha$ , isto é, no nosso caso, queremos dizer que há evidências de que a variável explicativa está estatisticamente relacionada com a variável resposta.

Uma vez que, assintoticamente,

$$\frac{\hat{\beta}_j - \beta_j}{SE(\hat{\beta}_j)} \sim Normal(0, 1)$$

pode-se determinar quantis  $z_{\alpha/2}$  e  $z_{1-\alpha/2}$  tais que,  $0 < \alpha < 1$

$$P\left(z_{\alpha/2} < \frac{\hat{\beta}_j - \beta_j}{SE(\hat{\beta}_j)} < z_{1-\alpha/2}\right) = 1 - \alpha \quad (2.18)$$

Isolando  $\beta_j$  no centro da desigualdade, temos:

$$P\left(\hat{\beta}_j - z_{1-\alpha/2}SE(\hat{\beta}_j) < \beta_j < \hat{\beta}_j + z_{1-\alpha/2}SE(\hat{\beta}_j)\right) = 1 - \alpha \quad (2.19)$$

Assim, um intervalo de confiança  $1 - \alpha$  para  $\beta_j$  fica dado por:

$$IC(\beta_j; 1 - \alpha) = (\hat{\beta}_j \pm z_{1-\alpha/2}SE(\hat{\beta}_j)) \quad (2.20)$$

### 2.5.5 Bondade de Ajuste

Um modo de avaliar se o modelo esta bem ajustado é a partir do Critério de Informação Akaike (AIC) ou Critério de Informação Bayesiano (BIC). Eles são estatísticas de bondade de ajuste baseadas na função de log-verossimilhança com ajuste para o número de parâmetros estimados e para a quantidade de dados (DOBSON; BARNETT, 2018). Estas estatísticas são geralmente definidas como:

$$AIC = -2l(\hat{\boldsymbol{\pi}}; \mathbf{y}) + 2p \quad (2.21)$$

$$BIC = -2l(\hat{\boldsymbol{\pi}}; \mathbf{y}) + p \times (\log)(n) \quad (2.22)$$

onde  $\hat{\boldsymbol{\pi}}$  é o vetor de estimativas de  $\pi_i$ ,  $i = 1, \dots, n$ ,  $p$  é o número de parâmetros estimados e  $n$  o número de observações.

### 2.5.6 Análise de Resíduos

Em regressão linear, medidas resumidas de ajuste, assim como diagnósticos para efeitos caso a caso do ajuste, são funções de um resíduo definido como diferença entre o valor observado e o ajustado ( $y - \hat{y}$ ). Em regressão logística, existem várias maneiras possíveis para mensurar a diferença entre os valores observados e ajustados, sendo uma delas a curva ROC (Receiver Operating Characteristic),

#### Tabela de Classificação e curva ROC

Uma maneira de examinar a performance do modelo é a partir de uma tabela de classificação (ou matriz de confusão), que corresponde a uma tabela 2x2 onde é possível comparar ocorrências reais com as predições (JONG; HELLER et al., 2008). A Tabela 3 a seguir ilustra esse conceito para os casos de nascimentos prematuros.

Tabela 3: Categorias observadas versus previstas

Categorias Previstas	Categorias Observados	
	Nasceu prematuro ( $Y_i = 1$ )	Não nasceu prematuro ( $Y_i = 0$ )
Nasceu prematuro ( $\hat{Y}_i = 1$ )	A	B
Não nasceu prematuro ( $\hat{Y}_i = 0$ )	C	D

A interpretação da tabela é dada de modo que: A é o número que foi previsto de crianças que nasceram prematuras e que realmente nasceram prematuras (verdadeiro positivo), B é o número que foi previsto de crianças que nasceram prematuras mas, na verdade, não nasceram prematuras (falso positivo), C é o número que foi previsto de crianças

que não nasceram prematuras mas, na verdade, nasceram prematuras (falso negativo) e D é o número que foi previsto de crianças que não nasceram prematuras e realmente não nasceram prematuras (verdadeiro negativo).

Dada a tabela de classificação, a utilidade preditiva de um modelo é frequentemente resumida usando as duas medidas a seguir (JONG; HELLER et al., 2008):

- Sensibilidade: é a probabilidade de prever um evento quando ele ocorre (proporção de verdadeiros positivos dentre o total de observações que realmente ocorreram)

$$S = P(\hat{Y}_i = 1 | Y_i = 1) = \frac{P(\hat{Y}_i = 1 \cap Y_i = 1)}{P(Y_i = 1)} = \frac{A}{A + C}.$$

- Especificidade: é a probabilidade de prever um não-evento quando não ocorre o evento (proporção de verdadeiros negativos dentre o total de observações que não ocorreram)

$$E = P(\hat{Y}_i = 0 | Y_i = 0) = \frac{P(\hat{Y}_i = 0 \cap Y_i = 0)}{P(Y_i = 0)} = \frac{D}{B + D}.$$

Uma melhor e mais completa descrição da classificação da Tabela 3 é a área embaixo da curva ROC (Receiver Operating Characteristic). Essa curva, originária da teoria de detecção de sinal, mostra como o receptor detecta a existência de sinal na presença de ruído. Ela traça a probabilidade de detectar sinal verdadeiro (sensibilidade) e sinal falso (1 - especificidade) para toda uma gama de possíveis pontos de corte. Esta medida tornou-se agora o padrão para avaliar uma capacidade do modelo ajustado de atribuir, em geral, maiores probabilidades do resultado para o subgrupo que desenvolve o resultado ( $y = 1$ ) do que para o subgrupo que não desenvolva o resultado ( $y = 0$ ). (JR; LEMESHOW; STURDIVANT, 2013)

Todas as curvas ROC começam em (0,0) e terminam em (1,1), já que esses pontos correspondem às probabilidades limites 0 e 1, respectivamente. Uma curva ROC “boa” cresce rapidamente para 1: quanto mais a curva se inclina para o canto superior esquerdo, melhor é a predição do modelo. Isso é quantificado pelo cálculo da área sob a curva ROC (AUC), como uma medida da capacidade preditiva do modelo (JONG; HELLER et al., 2008). A AUC tem valor máximo de 1, indicando 100% de sensibilidade e 100% de especificidade. Um modelo em que a curva ROC é uma linha com 45° tem uma AUC = 0,5, o que indica nenhum valor discriminatório. (FAN; UPADHYE; WORSTER, 2006).

A capacidade de um modelo ajustado para discriminar entre os dois resultados é mais uma função da diferença entre os grupos e magnitudes dos coeficientes de inclinação

do que o próprio modelo logístico. Assim, como observado, podemos ter modelos bem ajustados que discriminam mal, assim como poderíamos ter modelos com ajuste ruim que discriminam bem (JR; LEMESHOW; STURDIVANT, 2013).

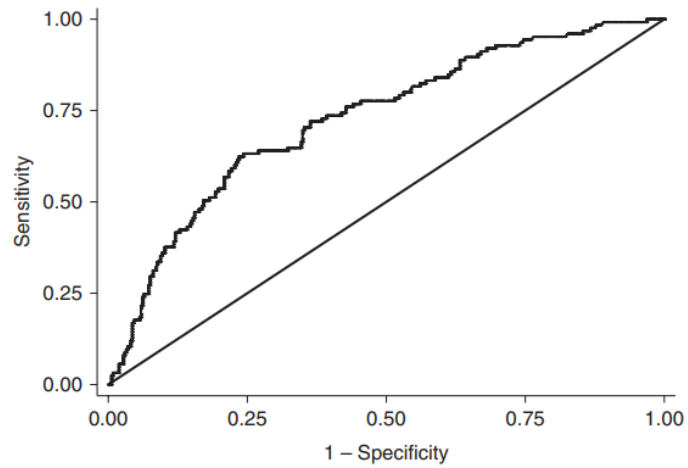


Figura 1: Gráfico da sensibilidade (sensitivity) versus 1-especificidade (1-specificity) (JR; LEMESHOW; STURDIVANT, 2013)

## 3 Análise dos Resultados

Neste capítulo serão mostrados o tratamento da base de dados, e das variáveis citadas no Capítulo 2, assim como os resultados das análises descritivas e inferência do modelo de regressão. Na Seção 3.1 são mostradas as análises descritivas dos dados e na Seção 3.2 as descrições dos modelos feitos. Para tais análises, foi utilizado a linguagem de software RStudio na versão 4.2.1, para os ajustes dos modelos de regressão logística, foi utilizado o comando *glm* do pacote *stats*.

### 3.1 Análise Descritiva

O tratamento das bases de dados começou com a base dos nascidos vivos. A base de dados não precisou de muitas modificações com relação aos dados iniciais. Foram selecionadas as variáveis idade, raça/cor, estado civil, escolaridade e município de residência da mãe, além das variáveis tipo de parto e quantidade de consultas pré-natal. A variável zona de residência não foi possível de ser utilizada pois a base de dados do SINASC não tinha a variável bairro de residência. Contudo, ficou decidido que também seria utilizada a variável estação do ano, para saber em qual estação do ano a criança nasceu e assim analisar o efeito que isso poderia trazer no momento do parto. Foram dispensados os dados faltantes e também a categoria “ignorado” da base de dados.

Com relação à base de dados de temperatura do INMET, inicialmente os dados estavam por hora. Foi feito o cálculo de média e amplitude diária para as quatro estações meteorológicas para então serem feitos indicadores de temperatura. O motivo se dá pois as estações meteorológicas são distantes umas das outras e não foi encontrada uma relação com as gestantes já que não existe a variável do seu bairro de residência. Os indicadores criados foram amplitude, criado a partir da média das amplitudes diárias; média, criado a partir da média das médias diárias; mediana, criado a partir da mediana das médias diárias; máximo, criado a partir do máximo das médias diárias; percentil 75, criado a partir do percentil 75 das médias diárias; percentil 90, criado a partir do percentil 90

das médias diárias; temperatura aparente, criado a partir do modo dito na Seção 2.4 e com as médias diárias da umidade e temperatura. Também foram criados os indicadores com defasagem (lag), que selecionaram os valores do dia anterior ao parto, para analisar o quanto a temperatura no dia anterior ao nascimento impacta na chance de prematuridade.

Ao juntar a base de dados de nascimentos com a base de temperatura, ficaram 72.862 observações e 23 variáveis. Inicialmente foi feita a análise descritiva dos dados, cujos resultados são mostrados na Tabela 4. O número de nascimentos prematuros foi de 6.358, totalizando 8,73% de nascimentos no ano de 2019. A maior parte das mulheres que tiveram filhos em 2019 tinha entre 20 e 29 anos, 45,62%. Com relação ao estado civil, 61,88% das mulheres eram solteiras, sendo seguidas pelas casadas, 32,23%. 56,04% das gestantes tinham o ensino médio completo e 50,23% eram pardas. O município de residência foi predominantemente o Rio de Janeiro, com 88,43% contra 11,57% de mulheres tendo filhos na capital do estado e residindo em outro município. Quanto às características gestacionais, uma grande parte teve 7 ou mais consultas de pré-natal e um pouco mais da metade dos nascimentos, 51,17%, foi via parto cesáreo. O número de partos ficou bem distribuído com relação à estação do ano, sendo o outono a época em primeiro lugar com 26,59% dos partos.

A Tabela 5 mostra o comportamento das médias e desvios padrões dos indicadores de temperatura em relação à prematuridade. Não há muita diferença das médias dos indicadores quando se observa separadamente as categorias de prematuridade, acontecendo o mesmo com os desvios padrões. A amplitude média ficou em 7,98°C nos dias de nascimentos prematuros. Sua média ficou com uma média 24,26°C, bem perto da mediana média, que ficou em 24,21°C. A média do máximo ficou em 25,26°C e da temperatura aparente ficou em 22,66°C. Os percentis 75 e 90 ficaram, respectivamente, com uma média de 24,66°C e 25,02°C nos dias de nascimentos prematuros. Ao analisar as estatísticas das defasagens (lag), as médias e desvios padrões não variaram muito, apenas  $\pm 0,01^\circ\text{C}$  nos indicadores para a temperatura do dia anterior.

## 3.2 Definições do modelo

Após as análises descritivas, foram feitas análises para especificação de modelos de regressão logística. Primeiramente, foram construídos modelos com o intuito de avaliar a prematuridade com relação às características da mãe e do nascimento. Em seguida, considerou-se os indicadores de temperatura e verificou-se o que melhor avaliasse a tem-

Tabela 4: Distribuição de frequências das variáveis de características da mãe e de nascimentos do município do Rio de Janeiro em 2019

Variável	Nascimento - n(%)	Prematuro - n(%)	Não prematuro - n(%)
<b>Prematuridade</b>	72.865 (100,00)	6.358 (8,73)	66.504 (91,27)
<b>Idade da Mãe</b>			
Até 19 anos	8.979 (12,32)	869 (9,68)	8.110 (90,32)
Entre 20 e 29 Anos	33.237 (45,62)	2.509 (7,55)	30.728 (92,45)
Entre 30 e 39 Anos	27.349 (37,54)	2.549 (9,32)	24.800 (90,68)
A partir de 40 anos	3.297 (4,52)	431 (13,07)	2.866 (86,93)
<b>Estado Civil da Mãe</b>			
Solteira	45.088 (61,88)	3.891 (8,63)	41.197 (91,37)
Casada	23.485 (32,23)	2.067 (8,80)	21.418 (91,20)
Separada Judicialmente/Divorciada	1.112 (1,53)	123 (11,06)	989 (88,94)
União Estável	3.099 (4,25)	260 (8,39)	2.839 (91,61)
Viúva	78 (0,11)	17 (21,79)	61 (78,21)
<b>Escolaridade da Mãe</b>			
Até o Ensino Fundamental	11.213 (15,39)	1.080 (9,63)	10.133 (90,37)
Ensino Médio	40.833 (56,04)	3.474 (8,51)	37.359 (91,49)
Ensino Superior	20.816 (28,57)	1.804 (8,67)	19.012 (91,33)
<b>Raça/Cor</b>			
Branca	26.731 (36,69)	2.348 (8,78)	24.383 (91,22)
Preta	9.128 (12,53)	796 (8,72)	8.332 (91,28)
Amarela	350 (0,48)	35 (10,00)	315 (90,00)
Parda	36.595 (50,23)	3.173 (8,67)	33.422 (91,33)
Indígena	58 (0,08)	6 (10,34)	52 (89,66)
<b>Município de Residência</b>			
Rio De Janeiro	64.434 (88,43)	5.515 (8,56)	58.919 (91,44)
Outro	8.428 (11,57)	843 (10,00)	7.585 (90,00)
<b>Tipo de Parto</b>			
Normal	35.577 (48,83)	2.571 (7,23)	33.006 (92,77)
Cesáreo	37.285 (51,17)	3.787 (10,16)	33.498 (89,84)
<b>Quantidade de Consultas Pré-Natal</b>			
Nenhuma	637 (0,87)	90 (14,13)	547 (85,87)
De 1 a 3 Consultas	2.365 (3,25)	441 (18,65)	1.924 (81,35)
De 4 a 6 Consultas	9.833 (13,50)	1.628 (16,56)	8.205 (83,44)
7 ou mais Consultas	60.027 (82,38)	4.199 (7,00)	55.828 (93,00)
<b>Estações do Ano</b>			
Verão	18.582 (25,5)	1.801 (9,69)	16.781 (90,31)
Outono	19.372 (26,59)	1.701 (8,78)	17.671 (91,22)
Inverno	18.593 (25,52)	1.590 (8,55)	17.003 (91,45)
Primavera	16.315 (22,39)	1.266 (7,76)	15.049 (92,24)

peratura e o parto prematuro. A partir dele, foram estimados modelos para identificar as variáveis modificadoras do efeito da temperatura, como características da mãe e de nascimento.

Na Tabela 6, temos os modelos logísticos simples de cada uma das variáveis de características da mãe e do nascimento. Para cada um deles, avaliou-se o teste de Wald para a significância das estimativas do modelo. A um nível de significância de 5%, a variável raça/cor não obteve nenhuma categoria com p-valor menor que o nível de significância, logo há evidências de que essa variável não está estatisticamente relacionada com a prematuridade. Além disso, as variáveis escolaridade, município de residência da mãe e estações do ano são as únicas em que todos os p-valores são menores que o nível de significância, logo há evidências de que essas variáveis estão estatisticamente relacionadas com a prematuridade.



Tabela 5: Estatísticas descritivas dos indicadores de temperatura do município do Rio de Janeiro em 2019

Variável	Prematuro		Não Prematuro	
	Média	Desvio padrão	Média	Desvio padrão
Amplitude	7,98	2,85	7,88	2,94
Média	24,26	2,92	24,11	2,85
Mediana	24,21	2,89	24,06	2,82
Máximo	25,26	3,07	25,10	3,00
Percentil 75	24,66	2,95	24,51	2,87
Percentil 90	25,02	3,01	24,86	2,94
Temperatura Aparente	22,66	2,52	22,70	2,46
Amplitude Lag	7,97	2,87	7,88	2,94
Média Lag	24,26	2,92	24,11	2,85
Mediana Lag	24,22	2,89	24,06	2,82
Máximo Lag	25,25	3,07	25,10	3,00
Percentil 75 Lag	24,66	2,94	24,51	2,87
Percentil 90 Lag	25,01	3,01	24,86	2,94
Temperatura Aparente Lag	22,67	2,54	22,70	2,46

As mulheres com mais de 40 anos têm 40,4% mais chances de ter parto prematuro com relação às mulheres com menos de 19 anos. Já as mulheres viúvas têm quase 3 vezes mais chances de seus filhos nascerem antes de 37 semanas do que as solteiras. As chances de parto prematuro diminuem em 11% em mulheres com ensino superior comparadas às mulheres que têm até o ensino fundamental. Com relação à raça/cor, mulheres indígenas têm quase 20% mais chances de prematuridade do que mulheres brancas. Quem tem residência no município do Rio de Janeiro tem uma chance diminuída em 15,8% de ter parto pré-termo daquelas que moram fora do município do Rio. As chances de ter parto prematuro aumentam em 45,1% em mulheres que tiveram parto cesáreo. Quanto à quantidade de consultas pré-natal, quem não teve nenhuma consulta tem 54,5% mais chance de ter um filho prematuro comparado com quem teve 7 ou mais consultas pré-natal. No que diz respeito às estações do ano, as chances de prematuridade são diminuídas em quase 13% no inverno, tendo como referência o verão.

Para escolher o melhor modelo, usamos o critério de Akaike (AIC), sendo o melhor modelo aquele com menor valor de AIC. A partir disso, o melhor modelo que explica a prematuridade com relação às características da mãe e do nascimento é aquele com a variável quantidade de consultas pré-natal, sendo seguido pelo tipo de parto e idade da mãe. O pior modelo foi da variável raça/cor, corroborando o que já foi afirmado com o teste Wald.

Para encontrar o melhor indicador de temperatura, também foram feitos modelos

Tabela 6: Razão de chances (OR), intervalo de confiança, p-valor do teste de significância Wald e AIC dos modelos de regressão logística simples para as características da mãe e de nascimento

Variável	OR (IC 95%)	p-valor	AIC
<b>Idade da Mãe</b>			43013,88
Até 19 anos	1,00		
Entre 20 e 29 Anos	0,762 (0,703 - 0,827)	< 0,001	
Entre 30 e 39 Anos	0,959 (0,885 - 1,041)	0,314	
A partir de 40 anos	1,404 (1,240 - 1,586)	< 0,001	
<b>Estado Civil da Mãe</b>			43146,30
Solteira	1,00		
Casada	1,022 (0,966 - 1,080)	0,449	
Separada Judicialmente/Divorciada	1,317 (1,670 - 4,938)	< 0,001	
União Estável	0,970 (1,084 - 1,586)	0,005	
Viúva	2,951 (0,849 - 1,103)	0,645	
<b>Escolaridade da Mãe</b>			43149,02
Até o Ensino Fundamental	1,00		
Ensino Médio	0,873 (0,812 - 0,938)	< 0,001	
Ensino Superior	0,890 (0,823 - 0,964)	0,004	
<b>Raça/Cor</b>			43165,66
Branca	1,00		
Preta	0,992 (0,912 - 1,079)	0,853	
Amarela	1,154 (0,798 - 1,616)	0,425	
Parda	0,986 (0,932 - 1,043)	0,618	
Indígena	1,198 (0,460 - 2,576)	0,675	
<b>Município de Residência</b>			43141,98
Outro	1,00		
Rio De Janeiro	0,842 (0,781 - 0,910)	< 0,001	
<b>Tipo de Parto</b>			42963,17
Normal	1,00		
Cesáreo	1,451 (1,377 - 1,529)	< 0,001	
<b>Quantidade de Consultas Pré-Natal</b>			42063,52
Nenhuma	1,00		
De 1 a 3 Consultas	1,393 (1,094 - 1,790)	0,008	
De 4 a 6 Consultas	1,206 (0,964 - 1,526)	0,109	
7 ou mais Consultas	0,457 (0,367 - 0,576)	< 0,001	
<b>Estações do Ano</b>			43123,09
Verão	1,00		
Outono	0,897 (0,837 - 0,962)	0,002	
Inverno	0,871 (0,812 - 0,935)	< 0,001	
Primavera	0,784 (0,727 - 0,845)	< 0,001	

para cada indicador. A Tabela 7 mostra os resultados desses modelos. A partir do teste Wald e com o mesmo nível de significância que foi usado anteriormente, as variáveis temperatura aparente e temperatura aparente lag tiveram p-valores maiores que 5%, logo há evidências de que estes indicadores não estão estatisticamente relacionados com a

prematuridade. As chances de ocorrer parto prematuro diminuem 0,7% ao aumentar 1°C a temperatura aparente, e da temperatura aparente do dia anterior, aumenta em 4,7% as chances ao aumentar 1°C. Os outros indicadores tiveram p-valores menores que 5%, logo há evidências de que estes indicadores estão estatisticamente relacionados com a prematuridade. As chances de haver nascimento pré-termo aumentam em torno de 1% ao aumentar 1°C na amplitude e na amplitude do dia anterior. Os outros indicadores tiveram razões de chances bem parecidas, aumentando as chances em cerca de 2% ao aumentar em 1°C. Para avaliar o melhor modelo, foi utilizado novamente o critério AIC. Os indicadores média lag e mediana lag tiveram valores do AIC bem parecidos, e por serem os menores valores, foram considerados os melhores modelos. Para decidir o melhor, foi usado o p-valor. O p-valor da média lag é  $2.86 \times 10^{-5}$  e o da mediana lag é  $3.18 \times 10^{-5}$ . Apesar do p-valor da média lag ser menor que o da mediana lag, o indicador escolhido foi a mediana lag, já que mostra como 50% das médias de temperatura do dia anterior se comportam.

Tabela 7: Razão de chances (OR), intervalo de confiança, p-valor do teste de significância Wald e AIC dos modelos de regressão logística simples dos indicadores de temperatura

Variável	OR (IC 95%)	p-valor	AIC
Amplitude	1,011 (1,003 - 1,020)	0,012	43154,46
Média	1,019 (1,010 - 1,028)	< 0,001	43144,58
Mediana	1,018 (1,009 - 1,028)	< 0,001	43145,38
Máximo	1,018 (1,010 - 1,027)	< 0,001	43143,59
Percentil 75	1,018 (1,009 - 1,028)	< 0,001	43144,75
Percentil 90	1,019 (1,010 - 1,027)	< 0,001	43143,95
Temperatura Aparente	0,993 (0,983 - 1,003)	0,177	43158,96
Amplitude Lag	1,010 (1,001 - 1,019)	0,027	43155,71
Média Lag	1,020 (1,010 - 1,029)	< 0,001	43143,05
Mediana Lag	1,020 (1,010 - 1,029)	< 0,001	43143,25
Máximo Lag	1,017 (1,008 - 1,026)	< 0,001	43145,58
Percentil 75 Lag	1,019 (1,010 - 1,028)	< 0,001	43144,38
Percentil 90 Lag	1,018 (1,009 - 1,027)	< 0,001	43145,03
Temperatura Aparente Lag	1,047 (0,951 - 1,151)	0,375	43159,81

Com o melhor indicador de temperatura escolhido, vamos agora analisar qual variável é modificadora do efeito da temperatura. Para isso, foram feitos modelos em que as variáveis de características da mãe e de nascimento foram adicionadas ao indicador mediana lag. A Tabela 8 mostra os resultados desses modelos. Analisando a razão de chances (OR) dos modelos, percebeu-se que em quase todos, as chances de prematuridade continuaram de 20% ao aumentar 1°C a temperatura mediana do dia anterior; apenas no modelo com a variável estação do ano que as chances de partos antes de 37 semanas foram de 0,8% ao aumentar 1°C a temperatura da mediana do dia anterior. A variável raça/cor

continua com valores de p-valor maior que o nível de significância de 5%, logo, pelo teste Wald, há evidências de que esta variável não está estatisticamente relacionada com a prematuridade. Foi utilizado o critério de Akaike (AIC) e o menor valor foi do modelo mediana lag + quantidade de consultas pré-natal, sendo seguido pelo modelo mediana lag + tipo de parto e o modelo mediana lag + idade da mãe.

Tabela 8: Razão de chances (OR), intervalo de confiança, p-valor do teste de significância Wald e AIC dos modelos de regressão logística do indicadores mediana lag e características da mãe e de nascimento

Variável	OR (IC 95%)	p-valor	AIC
<b>Mediana Lag + Idade Da Mãe</b>			42998,03
Mediana Lag	1,020 (1,011 - 1,029)	< 0,001	
Até 19 anos	1,00		
Entre 20 e 29 anos	0,763 (0,704 - 0,828)	< 0,001	
Entre 30 e 39 anos	0,961 (0,887 - 1,043)	0,338	
A partir de 40 anos	1,407 (1,243 - 1,590)	< 0,001	
<b>Mediana Lag + Estado Civil da Mãe</b>			43130,79
Mediana Lag	1,020 (1,010 - 1,029)	< 0,001	
Solteira	1,00		
Casada	1,022 (0,967 - 1,081)	0,442	
Viúva	2,955 (1,672 - 4,946)	< 0,001	
Separada judicialmente/Divorciada	1,316 (1,083 - 1,585)	0,005	
União estável	0,971 (0,850 - 1,105)	0,659	
<b>Mediana Lag + Escolaridade da Mãe</b>			43133,52
Mediana Lag	1,020 (1,010 - 1,029)	< 0,001	
Até o Ensino Fundamental	1,00		
Ensino Médio	0,872 (0,812 - 0,938)	< 0,001	
Ensino Superior	0,892 (0,825 - 0,966)	0,005	
<b>Mediana Lag + Raça/Cor</b>			43150,05
Mediana Lag	1,020 (1,010 - 1,029)	< 0,001	
Branca	1,00		
Preta	0,991 (0,910 - 1,077)	0,824	
Amarela	1,156 (0,799 - 1,618)	0,421	
Parda	0,984 (0,931 - 1,041)	0,577	
Indígena	1,192 (0,458 - 2,564)	0,684	
<b>Mediana Lag + Município de Residência</b>			43126,10
Mediana Lag	1,020 (1,011 - 1,029)	< 0,001	
Rio de Janeiro	0,841 (0,780 - 0,908)	< 0,001	
Outro	1,00		
<b>Mediana Lag + Tipo de Parto</b>			42945,52
Mediana Lag	1,021 (1,012 - 1,030)	< 0,001	
Normal	1,00		
Cesáreo	1,454 (1,380 - 1,533)	< 0,001	
<b>Mediana Lag + Quantidade de Consultas Pré-Natal</b>			42048,03
Mediana Lag	1,020 (1,010 - 1,029)	< 0,001	
Nenhuma	1,00		
De 1 a 3 consultas	1,391 (1,092 - 1,787)	0,009	
De 4 a 6 consultas	1,204 (0,962 - 1,523)	0,113	
7 ou mais consultas	0,456 (0,366 - 0,575)	< 0,001	
<b>Mediana Lag + Estações do Ano</b>			43123,30
Mediana Lag	1,008 (0,995 - 1,022)	0,207	
Verão	1,00		
Outono	0,918 (0,849 - 0,993)	0,033	
Inverno	0,913 (0,825 - 1,011)	0,079	
Primavera	0,807 (0,739 - 0,882)	< 0,001	

Nos modelos mostrados na Tabela 8, o indicador mediana lag está da forma contínua. Para avaliar os efeitos do calor na prematuridade, foi feita uma categorização deste indicador. O valor de corte escolhido foi 26°C e o indicador ficou de modo que, para valores maiores ou iguais a 26°C, a nova variável recebia o valor 1, caso contrário, recebia 0. As variáveis características da mãe também foram categorizadas para melhor mostrar os efeitos, suas novas categorias estão na Tabela 9.

Tabela 9: Tabela com novas categorizações das variáveis de características da mãe e de nascimento

Variável	Novas Categorias
Idade da Mãe	1 – Até 30 anos 2 – 30 anos ou mais
Estado Civil da Mãe	1 – Casada + União Estável 2 – Solteira + Divorciada + Viúva
Escolaridade da Mãe	1 – Até o Ensino Fundamental 2 – Ensino Médio + Ensino Superior
Raça/Cor	1 – Branca 2 – Preta + Amarela + Parda + Indígena
Município de Residência	1 – Rio De Janeiro 2 – Outro
Tipo de Parto	1 – Normal 2 – Cesáreo
Quantidade de Consultas Pré-Natal	1 – Até 3 consultas 2 – 4 ou mais consultas
Estações do Ano	1 – Verão 2 – Outono + Inverno + Primavera

Foram estimados, então, modelos com a prematuridade como variável resposta e as variáveis explicativas sendo a mediana lag categórica e as características da mãe e de nascimento com as novas categorias. Para melhor visualizar o efeito do calor, foram feitos gráficos dos intervalos de confiança das razões de chance. A Figura 2 mostra que em quase todas as variáveis de características da mãe e de nascimento o intervalo de confiança da razão de chances de suas categorias são bem próximas, a razão de chances de uma categoria está incluída no intervalo de confiança da outra. Apenas a variável estação do ano que as categorias não estão contidas nos seus intervalos de confiança. Logo, pode-se dizer que essa variável é modificadora de efeito do calor.

Com relação às razões de chances, as chances de ter filho com mais de 30 anos aumenta em quase 24% as chances de nascer prematuro com relação às mulheres com menos de 30 anos. Com relação ao estado civil destas mulheres, as chances de ter filho prematuro são quase 10 vezes menores para quem é casada ou está em uma união estável do que em outro tipo de estado civil. Já mulheres com nível de escolaridade até o ensino fundamental

têm 8,8 vezes menos chances de ter parto prematuro do que mulheres com ensino médio e ensino superior. No que se refere à raça/cor, mulheres que se declaram pretas, amarelas, pardas ou indígenas tiveram quase 10 vezes mais chances de ter filhos prematuros do que mulheres que se declaram brancas. Quanto a residência, moradoras do município do Rio de Janeiro teve 8,4 vezes mais chances de ter parto prematuro do que as moradoras de outros municípios. No que diz respeito às características de nascimento, ter parto cesáreo aumenta em 45,50% as chances do bebê nascer antes do tempo do que ter parto normal. E quem fez 4 ou mais consultas pré-natal tem 4,23 vezes mais chances de ter parto prematuro do que quem teve menos de 4 consultas. No verão, as mulheres têm quase 9 vezes menos chances de ter filhos antes de 37 semanas do que em outras estações do ano.

A partir dos modelos estimados anteriormente, foram feitas análises a partir da curva ROC para mostrar a capacidade preditiva dos modelos que melhor mostram a associação entre a prematuridade e altas temperaturas. O critério utilizado foi a área abaixo da curva (AUC). A Figura 3(a) mostra a curva ROC do modelo Mediana Lag + Quantidade de Consultas Pré-Natal e tem  $AUC = 0.601$  ou 60,1%. Já a Figura 3(b) apresenta a curva ROC do modelo Modelo Mediana Lag Categórica + Estações do Ano, sua AUC é igual a 0,524 ou 52,4%. As duas AUC estão próximas do valor 0,5, logo os valores não são muito discriminatórios, sendo o modelo Modelo Mediana Lag Categórica + Estações do Ano com valores menos discriminatórios que o modelo Mediana Lag + Quantidade de Consultas Pré-Natal.

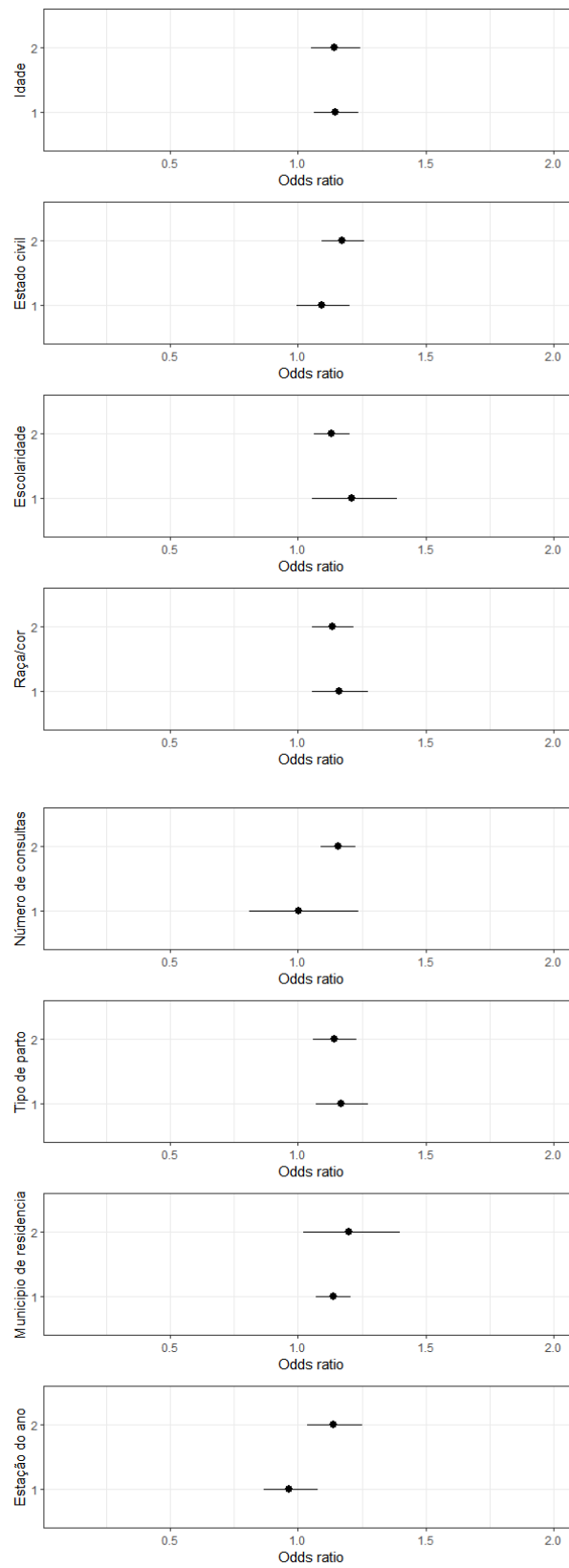
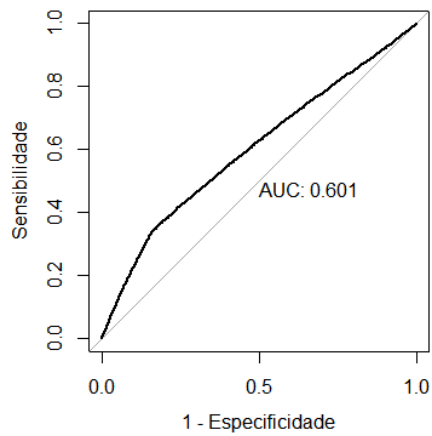
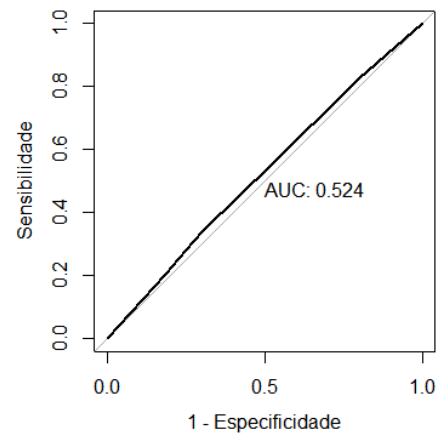


Figura 2: Gráficos do efeito do calor com as características da mãe e de nascimento



(a) Modelo Mediana Lag + Quantidade de Consultas Pré-Natal



(b) Modelo Mediana Lag Categórica + Estações do Ano

Figura 3: Gráficos das curvas ROC dos modelos selecionados.



## 4 Conclusões

O presente trabalho mostrou que em 2019, os casos de prematuridade no município do Rio de Janeiro foram maiores em mulheres entre 30 e 39 anos, solteiras, pardas, com ensino médio completo, que têm residência no próprio município, que tiveram parto cesáreo, que fizeram 7 ou mais consultas e tiveram partos no verão.

Através de indicadores de temperaturas de quatro estações meteorológicas, foram estimados modelos de regressão logística simples e concluiu-se que o melhor indicador de temperatura para explicar a associação com prematuridade foi a mediana do dia anterior, com 20% de chance de ocorrer prematuridade ao aumentar 1°C a temperatura do dia anterior.

Para avaliar possíveis modificadores de efeito desse indicador de temperatura, associada à prematuridade, foram adicionadas as variáveis das características da mãe e de nascimento à mediana lag. Não houve modificação no indicador ao incluir as variáveis, continuando com o aumento de 20% na chance de ocorrer prematuridade ao aumentar 1°C a temperatura do dia anterior. Já ao categorizar a mediana lag de modo que pudesse ver o calor, sendo 26°C o ponto de corte para essa categorização, a variável estação do ano foi uma modificadora do efeito.

Sendo assim, os resultados sugerem associação entre o aumento da temperatura do dia anterior e o parto pré-termo, sendo a estação do ano e a quantidade de consultas pré-natal fatores que intensificam esse desfecho.

Para continuação deste estudo, sugere-se a inclusão de todas as variáveis de características da mãe em um modelo logístico múltiplo, a fim de investigar se mais de uma característica pode aumentar as chances de prematuridade e modificar o efeito da mediana da temperatura do dia anterior. Além disso, aumentar a área de estudo para todo o estado, a fim de identificar lugares com maiores diferenças de temperatura.

# Referências

- AGRESTI, A. *An introduction to categorical data analysis*. [S.l.]: John Wiley & Sons, 2018.
- BALBI, B.; CARVALHAES, M. A. d. B. L.; PARADA, C. M. G. d. L. Tendência temporal do nascimento pré-termo e de seus determinantes em uma década. *Ciência & Saúde Coletiva*, SciELO Public Health, v. 21, p. 233–241, 2016.
- BASU, R.; MALIG, B.; OSTRO, B. High ambient temperature and the risk of preterm delivery. *American Journal of Epidemiology*, Oxford University Press, v. 172, n. 10, p. 1108–1117, 2010.
- CAROLAN-OLAH, M.; FRANKOWSKA, D. High environmental temperature and preterm birth: a review of the evidence. *Midwifery*, Elsevier, v. 30, n. 1, p. 50–59, 2014.
- COPERNICUS: 2020 warmest year on record for Europe; globally, 2020 ties with 2016 for warmest year recorded. *Copernicus*, 2021.
- DATASUS. *Nascidos Vivos: Notas Técnicas*. 2017.
- DOBSON, A. J.; BARNETT, A. G. *An introduction to generalized linear models*. [S.l.]: Chapman and Hall/CRC, 2018.
- EL-HABIL, A. M. An application on multinomial logistic regression model. *Pakistan journal of statistics and operation research*, College of Statistical and Actuarial Sciences, p. 271–291, 2012.
- FAN, J.; UPADHYE, S.; WORSTER, A. Understanding receiver operating characteristic (roc) curves. *Canadian Journal of Emergency Medicine*, Cambridge University Press, v. 8, n. 1, p. 19–20, 2006.
- FIOCRUZ. *Mapa da Vulnerabilidade da População dos Municípios do Estado do Rio de Janeiro Frente às Mudanças Climáticas*. [S.l.]: Fiocruz Rio de Janeiro, 2014.
- FIORAVANTI, C. 2019 foi o ano mais quente já registrado no brasil. *Pesquisa FAPESP*, 2020.
- GUO, T. et al. The association between ambient temperature and the risk of preterm birth in china. *Science of The Total Environment*, Elsevier, v. 613, p. 439–446, 2018.
- ILANGO, S. D. et al. Extreme heat episodes and risk of preterm birth in california, 2005–2013. *Environment international*, Elsevier, v. 137, p. 105541, 2020.
- JONG, P. D.; HELLER, G. Z. et al. *Generalized linear models for insurance data*. Cambridge Books, Cambridge University Press, 2008.

- JORGE, M. H. P. d. M. et al. Avaliação do sistema de informação sobre nascidos vivos e o uso de seus dados em epidemiologia e estatísticas de saúde. *Revista de Saúde Pública, SciELO Public Health*, v. 27, p. 1–46, 1993.
- JR, D. W. H.; LEMESHOW, S.; STURDIVANT, R. X. *Applied logistic regression*. [S.l.]: John Wiley & Sons, 2013. v. 398.
- KNIPPELBERG, F. de M. *ESTAÇÕES METEOROLÓGICAS: FUNCIONAMENTO E A SUA IMPORTANCIA NA AGRICULTURA*. 2019.
- KRAAIJENBRINK, P. D. et al. Impact of a global temperature rise of 1.5 degrees celsius on asia's glaciers. *Nature*, Nature Publishing Group, v. 549, n. 7671, p. 257–260, 2017.
- LI, C. et al. Temperature variation and preterm birth among live singleton deliveries in shenzhen, china: A time-to-event analysis. *Environmental Research*, Elsevier, v. 195, p. 110834, 2021.
- LI, S. et al. Temporal change in the impacts of ambient temperature on preterm birth and stillbirth: Brisbane, 1994–2013. *Science of the Total Environment*, Elsevier, v. 634, p. 579–585, 2018.
- MAPAS revelam maiores áreas urbanizadas do país: São Paulo, Rio e Brasília. *CAU/RJ - Conselho de Arquitetura e Urbanismo do Rio de Janeiro*, 2018.
- MUNDO teve a década mais quente da história entre 2011 e 2020, diz OMM. *ONU News - Perspectiva Global Reportagens Humanas*, 2020.
- OMS. *Nascimento prematuro*. 2018. Disponível em: <http://www.who.int/mediacentre/factsheets/fs363/en/index.html>. Acesso em fevereiro de 2021.
- PEDRAZA, D. F. Qualidade do sistema de informações sobre nascidos vivos (sinasc): análise crítica da literatura. *Ciência & Saúde Coletiva*, SciELO Brasil, v. 17, p. 2729–2737, 2012.
- RAMOS, H. d. C.; CUMAN, R. K. N. Fatores de risco para prematuridade: pesquisa documental. *Esc Anna Nery Rev Enferm*, SciELO Brasil, v. 13, n. 2, p. 297–304, 2009.
- SANTOS, A. C. C. dos. Inmet chama atenção para elevação de temperatura no país. *Instituto Nacional de Meteorologia - INMET*, 2021.
- SARMENTO, R.; COSTA, V. Introduction to linear regression. In: \_\_\_\_\_. [S.l.: s.n.], 2017. ISBN 9781522519898.
- SON, J.-Y. et al. Impacts of high temperature on adverse birth outcomes in seoul, korea: Disparities by individual-and community-level characteristics. *Environmental research*, Elsevier, v. 168, p. 460–466, 2019.
- SUN, S. et al. Ambient temperature and preterm birth: a retrospective study of 32 million us singleton births. *Environment international*, Elsevier, v. 126, p. 7–13, 2019.
- TABNET. 2020. Disponível em: <http://tabnet.datasus.gov.br/cgi/tabcgi.exe?sinasc/cnv/nvuf.def>. Acesso em fevereiro de 2021.

TABNET. 2020. Disponível em: (<http://tabnet.datasus.gov.br/cgi/tabcgi.exe?sinasc/cnv/nvrj.def>). Acesso em fevereiro de 2021.

TACONELI, C. A. *TESTES DE HIPÓTESES E INTERVALOS DE CONFIANÇA EM MODELOS LINEARES GENERALIZADOS*. 2019. 27 slides.

VICEDO-CABRERA, A. M. et al. Exposure to elevated temperatures and risk of preterm birth in valencia, spain. *Environmental research*, Elsevier, v. 134, p. 210–217, 2014.

VIDAL, M. Alta hospitalar e reinternação de bebê prematuro: uma reflexão sobre o acesso aos serviços de saúde. *Mental*, Universidade Presidente Antônio Carlos, v. 9, n. 17, p. 537–558, 2011.

WANG, Y.-Y. et al. Ambient temperature and the risk of preterm birth: A national birth cohort study in the mainland china. *Environment International*, Elsevier, v. 142, p. 105851, 2020.

WARDLAW, T. M. *Low birthweight: country, regional and global estimates*. [S.l.]: Unicef, 2004.