

**Universidade de São Paulo  
Instituto de Matemática e Estatística**

**Centro de Estatística Aplicada**

**Relatório de Análise Estatística**

RAE-CEA-24P14

**RELATÓRIO DE ANÁLISE ESTATÍSTICA SOBRE O PROJETO:**

**“Estudo de correlação clínica, patológica e radiológica de neoplasias diagnosticadas em autopsias realizadas pelo Departamento de Patologia da Universidade de São Paulo, no período de 2007 a 2023”**

**Ana Caroline Oliveira de Freitas**

**Luís Gustavo Esteves**

**Rodrigo Chan Seng Chon**

**São Paulo, abril de 2024**

**CENTRO DE ESTATÍSTICA APLICADA - CEA – USP**

**TÍTULO:** Relatório de Análise Estatística sobre o Projeto: “Estudo de correlação clínica, patológica e radiológica de neoplasias diagnosticadas em autopsias realizadas pelo Departamento de Patologia da Universidade de São Paulo, no período de 2007 a 2023”.

**PESQUISADOR:** Amaro Nunes Duarte Neto

**COLABORADOR:** João Carlos Geber Júnior

**INSTITUIÇÃO:** FMUSP

**FINALIDADE DO PROJETO:** Doutorado

**RESPONSÁVEIS PELA ANÁLISE:** Ana Caroline Oliveira de Freitas

Luís Gustavo Esteves

Rodrigo Chan Seng Chon

**REFERÊNCIA DESTE TRABALHO:** CHON, R.C.S.; ESTEVES, L.G.; FREITAS, A.C.O. Relatório de análise estatística sobre o projeto: “Estudo de correlação clínica, patológica e radiológica de neoplasias diagnosticadas em autopsias realizadas pelo Departamento de Patologia da Universidade de São Paulo, no período de 2007 a 2023”. São Paulo, IME-USP, 2024. (RAE–CEA-24P14)

## FICHA TÉCNICA

### REFERÊNCIAS BIBLIOGRÁFICAS:

AGRESTI, A. (2019). **An Introduction to Categorical Data Analysis**. 3ª ed. Hoboken, NJ: John Wiley & Sons.

MORETTIN, P. A.; BUSSAB, W. O. (2024). **Estatística Básica**. 10.ed., São Paulo: Editora Saraiva.

PEREIRA, C. A. De B.; STERN, J. (1999). **Evidence and Credibility**: a full Bayesian test of precise hypothesis. *Entropy*, **1**, 104-115.

PREFEITURA DO MUNICÍPIO DE SÃO PAULO. Secretaria Municipal da Saúde. **Atlas da Saúde da Cidade de São Paulo**, 2011. Disponível em: [https://www.prefeitura.sp.gov.br/cidade/secretarias/upload/saude/arquivos/publicacoes/atlas\\_da\\_saude\\_da\\_cidade\\_de\\_sao\\_paulo\\_2011.pdf](https://www.prefeitura.sp.gov.br/cidade/secretarias/upload/saude/arquivos/publicacoes/atlas_da_saude_da_cidade_de_sao_paulo_2011.pdf). Acesso em: 7 jul. 2024.

PYTHON SOFTWARE FOUNDATION. **Python Documentation. Versão 3.9**. [S.l.]: Python Software Foundation, 2021. Disponível em: <https://docs.python.org/3/>. Acesso em: 4 mar. 2024.

R CORE TEAM. **R: A language and environment for statistical computing**. Viena, Áustria: R Foundation for Statistical Computing, 2021. Disponível em: <https://www.R-project.org/>. Acesso em: 4 mar. 2024.

RIBEIRO, M. P.; DUARTE-NETO, A. N.; DOLHNIKOFF, M.; LINDOSO, L.; LOURENÇO, B.; MARQUES, H. H.; PEREIRA, M. F. B.; CRISTOFANI, L. M.; ODONE-FILHO, V.; CAMPOS, L. M. A.; SALLUM, A. M. E.; CARNEIRO-SAMPAIO, M.; DELGADO, A. F.; CARVALHO, W. B.; MAUAD, T.; SILVA, C. A. (2023). Major discrepancy between clinical diagnosis of death and anatomopathological findings in adolescents with chronic diseases during 18-years. **Clinics**, **78**, n. 100184. Disponível em: <https://www.journals.elsevier.com/clinics>. Acesso em: 4 mar. 2024.

**STATISTICS HOW TO.** Moran's I. [S.l.]: Statistics How To, s.d. Disponível em: <https://www.statisticshowto.com/morans-i/>. Acesso em: 4 mai. 2024.

#### **PROGRAMAS COMPUTACIONAIS UTILIZADOS:**

Microsoft Word for Windows (versão 2016)

Microsoft Excel for Windows (versão 2016)

Python 3.12.2 for Windows

R-4.3.3 for Windows

#### **TÉCNICAS ESTATÍSTICAS UTILIZADAS**

03:010 - Análise Descritiva Unidimensional

03:020 - Análise Descritiva Multidimensional

03:990 – Outros

05:060 – Testes Bayesianos

05:070 – Testes de Hipóteses Não Paramétricas

06:020 - Associação e Dependência dos Dados Qualitativos

06:030 - Análise de Dados Categorizados

07:090 – Regressão Logística

### **ÁREA DE APLICAÇÃO**

14:040 - Medicina

## Resumo

Este relatório apresenta uma análise estatística detalhada do projeto intitulado “Estudo de correlação clínica patológica e radiológica de neoplasias diagnosticadas em autopsias realizadas pelo Departamento de Patologia da Universidade de São Paulo no período de 2007 a 2023”. O estudo tem como objetivo principal caracterizar casos de neoplasias diagnosticadas postumamente, que aqui denominaremos como diagnóstico pós-morte, analisando dados demográficos, clínicos, patológicos e radiológicos, e avaliando a associação entre diagnósticos de neoplasia pré-morte e pós-morte com as características do paciente.

Foram examinados 326 prontuários de pacientes autopsiados nos anos de 2007 e 2008, residentes predominantemente no estado de São Paulo. As análises incluíram variáveis como idade, sexo, ano da autópsia, distância da residência do paciente até as unidades de saúde mais próximas, e classificação do diagnóstico de neoplasia. As técnicas estatísticas aplicadas variaram de análises descritivas a testes bayesianos e de testes exatos de Fisher.

Os resultados indicam a falta de associação entre diagnóstico pós-morte e a idade, o sexo, e as distâncias até as unidades de saúde mais próximas dos pacientes. Inicialmente também não encontrou-se associação espacial entre Diagnóstico pós-morte e localidade geográfica. No entanto, quando separamos em regiões, ou seja, separamos os pacientes em residentes de fora do estado de São Paulo, e os pacientes dentro da cidade e do estado de São Paulo, existe associação entre diagnóstico pós-morte e a separação por região. Possível associação também foi detectada ao separar os pacientes da cidade de São Paulo nas áreas: Centro-Oeste, Leste, Norte, Sudeste e Sul.

## **Sumário**

<b>1. Introdução</b>	9
<b>2. Objetivos</b>	9
<b>3. Descrição do estudo</b>	10
<b>4. Descrição das variáveis</b>	10
<b>5. Análise estatística</b>	12
<b>6. Conclusões</b>	19
<b>APÊNDICE A</b>	21
<b>APÊNDICE B</b>	27



## 1. Introdução

A compreensão detalhada das neoplasias, especialmente daquelas diagnosticadas postumamente, é crucial para aprimorar tanto as estratégias de diagnóstico quanto de tratamento no âmbito da medicina moderna. O estudo conduzido pelo Departamento de Patologia da Universidade de São Paulo entre 2007 e 2017 se debruça sobre esta necessidade, buscando estabelecer associações clínica, patológica e radiológica das condições diagnosticadas através de autópsias. Esta investigação visa não apenas caracterizar a casuística de casos com diagnóstico de neoplasias em termos demográficos, clínicos, patológicos e radiológicos, mas também se propõe a determinar o grau de concordância entre os diagnósticos clínicos pré-morte e os achados pós-morte, desvendando a real contribuição dessas neoplasias para os óbitos registrados.

O valor deste estudo é ampliado pela sua capacidade de elucidar aspectos fundamentais das neoplasias que, de outra forma, permaneceriam obscuros, especialmente aqueles diagnósticos realizados exclusivamente no exame *post-mortem* e sua relação direta ou indireta com a causa *mortis*. A análise detalhada dos achados anatomopatológicos, juntamente com os registros clínicos e radiológicos dos pacientes, permite uma compreensão mais aprofundada sobre a incidência, a progressão e o impacto real dessas condições sobre o sistema de saúde pública. Além disso, ao examinar a potencial associação entre a localização residencial dos pacientes e o acesso aos serviços de saúde, este estudo também pretende descobrir informações novas valiosas sobre as desigualdades no diagnóstico e tratamento de neoplasias, ressaltando a importância de estratégias de saúde pública mais inclusivas, acessíveis e eficientes.

## 2. Objetivos

- Caracterizar a casuística de casos com diagnóstico de neoplasias autopsiados quanto às características demográficas, clínicas, patológicas e radiológicas.
- Descrever as características demográficas gerais da população do estudo.

- Analisar a associação entre a distância da residência do paciente e a instituição de saúde do SUS mais próxima e seu impacto no diagnóstico e tratamento de casos de neoplasia.
- Avaliar a existência de outros possíveis fatores que tenham associação com o diagnóstico de neoplasias.
- Georreferenciar a amostra obtida no estudo.

### **3. Descrição do estudo**

Este estudo retrospectivo e descritivo abordou uma amostra significativa de autópsias de pessoas que faleceram no Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, enfocando aquelas com diagnósticos de neoplasias. O período de análise estende-se de janeiro de 2007 a janeiro de 2023. Entretanto, é importante ressaltar que nesse relatório foi analisada apenas uma parte inicial da amostra, que consiste nos diagnósticos realizados nos anos de 2007 e 2008. Foram analisados então 326 prontuários de pacientes com diagnóstico de neoplasia realizado antes ou após a autópsia do paciente, sendo a maioria dos pacientes residentes no estado de São Paulo e a totalidade deles residentes no Brasil.

### **4. Descrição das variáveis**

Variáveis que caracterizam a amostra:

- Idade: idade em anos com que o paciente veio a óbito
- Sexo: masculino e feminino
- Ano da autópsia: 2007 ou 2008

- Diagnóstico pós-morte: sim, para pacientes com neoplasia diagnosticada somente após a morte; e não caso a neoplasia tenha sido diagnosticada tanto antes quanto depois do óbito.
- Distância da 1ª porta mais próxima a pé (km): variável quantitativa que representa a menor distância geodésica por deslocamento a pé da residência do paciente até uma porta de entrada de uma Unidade Básica de Saúde do Sistema Único de Saúde.
- Distância da 1ª porta mais próxima de carro (km)
- Distância da 2ª porta mais próxima a pé (km)
- Distância da 2ª porta mais próxima de carro(km)
- Distância da 3ª porta mais próxima a pé (km)
- Distância da 3ª porta mais próxima de carro(km)
- Nome da 1ª porta mais próxima
- Nome da 2ª porta mais próxima
- Nome da 3ª porta mais próxima
- CEP: do endereço do paciente
- Latitude: coordenada latitudinal do endereço do paciente
- Longitude: coordenada longitudinal do endereço do paciente
- Classificação: variável que representa a classificação do paciente quanto à situação do diagnóstico de neoplasia na autópsia, variando de 1 a 4, sendo:
  - 1: Neoplasias diagnosticadas somente na autópsia, com contribuição direta no óbito, porém sem suspeita pre mortem do diagnóstico de neoplasia.
  - 2: Neoplasias diagnosticadas somente na autópsia, com contribuição direta no óbito, porém houve suspeita pre mortem do diagnóstico de neoplasia, mas a investigação diagnóstica da neoplasia (sítio primário e histogênese) foi incompleta (por gravidade do caso, vindo a óbito antes do diagnóstico patológico final; impossibilidade de medidas invasivas diagnósticas, por recusa do paciente e/ou familiares ou por terminalidade).
  - 3: Neoplasia diagnosticada antes da morte corretamente e na autópsia, com o óbito decorrente de neoplasia em estágio avançado, eventos adversos de medidas terapêuticas na internação (tais como os decorrentes de procedimentos

cirúrgicos ou minimamente invasivos, infecções nosocomiais e secundários a medicamentos prescritos).

4: Neoplasia diagnosticada antes da morte corretamente e na autópsia, com o óbito decorrente de outras causas não relacionadas à neoplasia.

- São Paulo (capital): variável categórica indicadora: Sim, se o paciente reside em São Paulo; ou Não, se não reside em São Paulo.
- Região: variável que representa a região do Brasil de origem do paciente: São Paulo (capital), se o paciente reside no município de São Paulo; São Paulo (estado), se o paciente não reside no município de São Paulo mas reside no estado de São Paulo; ou Externo, se o paciente não reside no estado de São Paulo.
- Criança: variável categórica indicadora: Criança, se o paciente tem 14 anos de idade ou menos; ou Adulto, se o paciente tem mais de 14 anos de idade.
- Área de saúde: variável que representa a região da área de saúde do Atlas da Saúde 2011, do paciente residente no município de São Paulo, sendo estas regiões Centro-Oeste, Leste, Norte, Sudeste, ou Sul.

## 5. Análise estatística

Nesta seção, apresentamos, primeiramente, a análise descritiva dos dados, que nos permite ter uma visão inicial dos resultados do estudo (Bussab e Morettin, 2024). Na sequência, apresentamos medidas de evidência e valores-p de testes estatísticos que visam detectar a associação entre a variável de interesse, diagnóstico pós-morte, e cada variável do estudo.

Na Tabela A.1 e na Figura B.3, verifica-se que a faixa etária dos pacientes da amostra se concentra entre 50 e 70 anos, com mediana de 59 anos. Observa-se, também, no *box plot*, um grupo considerável de crianças e adolescentes (cujas idades podem ser consideradas valores discrepantes ou *outliers*). Com relação ao sexo, a amostra se caracteriza como quase igualmente distribuída entre os dois sexos, como

podemos ver na Tabela A.5, e o mesmo acontece com o ano da autópsia na Tabela A.4, dividido entre os anos de 2007 e 2008.

Quanto à classificação, podemos observar, na Tabela A.2, que as quantidades de indivíduos classificados nas categorias 1 e 4 são pequenas (grupos que totalizam juntos menos de 30 pacientes) o que pode inviabilizar o uso de certos métodos de análise estatística. Sendo assim, optamos por usar na análise a variável diagnóstico pós-morte. Tal variável agrupa os pacientes classificados como 1 ou 2 na categoria “sim” de diagnóstico pós-morte, ou seja, os pacientes com neoplasia diagnosticada somente depois do óbito do paciente, e os pacientes classificados como 3 ou 4 na categoria “não” de diagnóstico pós-morte, ou seja, os pacientes com neoplasia diagnosticada tanto antes e depois do óbito do paciente. Dessa forma, temos duas categorias razoavelmente grandes como mostrado na Tabela A.3.

Na Tabela A.6 podemos ver as medidas-resumo para as 6 variáveis quantitativas de distância (em km), que correspondem às distâncias até as 3 portas de atendimento de saúde pública mais próximas, a pé ou de carro. As variáveis foram coletadas em 2024 através da API (sigla) do Busca Saúde, que abrange em seu mapa a cidade de São Paulo. Considerando que as autópsias correspondentes à amostra foram feitas nos anos de 2007 e 2008 e sabendo que o estudo busca entender a associação entre o diagnóstico tardio e tais distâncias, deve ser levado em conta que pode haver inconsistências quanto à existência de tais unidades de saúde nos anos de realização das autópsias (2007 e 2008).

Na Figura B.5 é apresentado o *box plot* para a distância da 1ª porta mais próxima a pé, onde percebe-se a existência de muitos valores outliers. Esses valores demasiadamente grandes são, provavelmente, de pacientes residentes fora de São Paulo.

Na Figura B.6 é apresentado o *box plot* para a distância da 1ª porta mais próxima a pé por grupo de diagnóstico pós-morte. Observa-se que mesmo separando por grupo de diagnóstico pós-morte, as distâncias se concentram numa faixa muito pequena de valores e que ainda há valores muito discrepantes da maioria dos dados. Isso justifica o

uso da variável São Paulo (capital) na análise da variável distância: os dados dos pacientes que não residem em São Paulo não serão considerados na análise das variáveis de distância. Esse fato é constatado também na Tabela A.6. Há uma série de *outliers* em relação à distância máxima e isso ocorre pelo fato de que na execução da API já comentada, não foi levado em consideração que o algoritmo abrange apenas o município de São Paulo. Desse modo, para pacientes provenientes de outros estados foram registradas distâncias entre seus locais de origem (fora de São Paulo) e a unidade de saúde mais próxima no município de São Paulo, o que gerou distorções para a análise dessas variáveis. A distribuição de frequências da variável São Paulo (capital) é apresentada na Tabela A.7. Por essa tabela, observa-se a amostra de 216 (66%) residentes de São Paulo, de um total de 326. A maioria de pacientes da amostra morava em São Paulo, mas nota-se que é razoavelmente grande a porção de pacientes que não reside em São Paulo. Novas medidas-resumo de distância foram obtidas para esse subgrupo de acordo com a Tabela A.8. Agora os valores das distâncias para os *outliers* são bem menores.

Na Figura B.7 é apresentado o *box plot* para a distância da 1ª porta mais próxima a pé, somente para os residentes em São Paulo, enquanto na Figura B.8 temos o histograma por grupo de diagnóstico para Distância da 1ª porta mais próxima a pé. Nesse caso, a distância máxima observada foi 6 km e a maioria dos valores estão situados entre zero e 2 quilômetros. Também observamos pouca diferença entre os grupos de diagnóstico com relação à Distância: as medianas e o primeiro quartil das distâncias são quase iguais, enquanto que o terceiro quartil dos pacientes com diagnóstico da neoplasia antes do óbito é pouco maior do que o terceiro quartil dos pacientes com diagnóstico da neoplasia somente depois do óbito.

Na Figura B.1, observamos a matriz de correlação para as variáveis quantitativas, onde temos os valores da correlação linear entre as variáveis. A figura também indica além do coeficiente de correlação, a natureza de associação linear entre as variáveis, sendo mais vermelho ao ser mais próximo de 1 (forte associação positiva), e mais azul ao ser próximo de -1 (forte associação negativa). Uma correlação fraca seria próxima de zero, com a cor próxima de cinza. Deve-se notar que com exceção da variável Idade, as

variáveis de distâncias, latitude e longitude possuem correlação próxima de 1. Optou-se então por usar apenas uma dessas variáveis altamente correlacionadas para avaliar a variável de interesse classificação, ou diagnóstico pós-morte. Na Figura B.2, temos os gráficos de dispersão entre as variáveis de distância, que novamente indicam a alta correlação linear entre tais variáveis. Sendo assim, optou-se por usar somente a variável distância da 1ª porta mais próxima a pé para explicar a variável de interesse, no caso diagnóstico pós-morte. Considerando a variável distância da 1ª porta mais próxima a pé, mais de 75% dos pacientes residentes em São Paulo moram no máximo a 2 km de distância da Unidade Básica de Saúde mais próxima. A maior distância observada foi de 6,59 km (Tabela A.8).

Na Figura B.4 temos o histograma da variável Idade em relação ao diagnóstico pós-morte. O objetivo deste gráfico é averiguar visualmente se existe alguma diferença entre as idades dos pacientes segundo a neoplasia ter sido diagnosticada antes da morte ou diagnosticada somente depois da morte. A distribuição dos 2 grupos são parecidas, concentrando-se mais na faixa etária de 50 a 75 anos, embora o grupo “Não” tenha um maior número de pacientes concentrada nesta faixa etária que o grupo “Sim”. Nota-se que há mais pacientes com idade zero com diagnóstico de neoplasia somente depois do óbito, provavelmente devido aos pacientes que nascem natimortos em decorrência da neoplasia. Na Tabela A.11 podemos comparar as medidas-resumo da idade por diagnóstico pós-morte, e parece não ter grandes diferenças entre os grupos de diagnóstico pós-morte: em ambos os grupos, as idades dos pacientes se concentram na faixa de de 50 a 70 anos, mas nota-se que o 1º quartil, 43 anos, e a média, 54,45 anos, dos pacientes em que não se diagnosticou neoplasia somente depois da morte possuem valores menores.

Na Figura B.9, podemos ver a distribuição geográfica com a latitude e a longitude dos pacientes em relação ao diagnóstico pós-morte “sim” (em vermelho) e “não” (em verde). Podemos notar uma maioria de pontos verdes entre os pontos fora do estado de São Paulo. Esse fato pode indicar a situação de pessoas que viajam para encontrar tratamento mais avançado e moderno para as suas condições de saúde na cidade de São Paulo, após serem diagnosticadas em outros estados do Brasil.

Nas Figuras B.10 e B.11, é notável que a maioria dos indivíduos da amostra reside no estado de São Paulo, sendo a maior parte no município de São Paulo (66%), como descrito na Tabela A.7. Com uma visão mais ampla da distribuição dos pontos, não foi possível identificar nenhum tipo de padrão que sugerisse relação entre diagnóstico pós-morte e a localização geográfica. Adicionalmente, para avaliar associação entre tais variáveis, foi calculado o índice de Moran, uma medida estatística para indicar a autocorrelação espacial (Statistics How To), e obtido um valor próximo de zero (0,05), indicando a inexistência de associação entre localização geográfica e diagnóstico pós-morte.

Na Tabela A.9, nota-se que a distribuição de frequências de diagnóstico pós-morte não parece muito homogênea entre os anos de autópsia do estudo (2007 e 2008). Em 2007, a proporção dos pacientes com diagnóstico pós-morte era de 44%, enquanto que em 2008, esta proporção era de 34%. Testando a hipótese de independência (existência de não associação) entre as variáveis diagnóstico pós-morte e ano da autópsia com a estatística qui-quadrado de Pearson, obtém-se o valor-p  $< 0,001$ , indicando que há evidência de associação entre os anos de autópsia e o diagnóstico pós-morte.

Na Tabela A.10, observa-se que a distribuição de frequência de diagnóstico pós-morte parece homogênea entre os sexos: dentre os homens, 38% tiveram diagnóstico de neoplasia pós-morte enquanto essa proporção foi de 41% entre as mulheres. Testando a hipótese de independência entre as variáveis diagnóstico pós-morte e sexo com a estatística qui-quadrado de Pearson, obtivemos o valor-p de 0,577, o que fortemente nos indica a inexistência de associação entre o sexo do paciente e o diagnóstico pós-morte.

Na Tabela A.12 temos a separação dos pacientes por região. Entre os grupos do estado de São Paulo e da capital São Paulo não se nota grande diferença entre as proporções de pacientes em relação ao diagnóstico pós-morte (nos dois grupos, a proporção de pacientes com diagnóstico pós-morte é de aproximadamente 41%). Entretanto, para o grupo fora do estado, a maioria massiva (92%) não apresenta diagnóstico pós-morte, o que corrobora com a ideia de que pacientes fora do estado foram transferidos para os hospitais públicos de São Paulo já com o diagnóstico de



neoplasia. Porém é importante ressaltar que esse grupo conta com apenas 13 pacientes, de modo que as conclusões obtidas das análises devem ser vistas com certa cautela.

Na Tabela A.13, temos a avaliação do diagnóstico pós-morte separadamente para crianças (menores de 15 anos) e adultos. É fácil observar que tanto para crianças quanto para adultos, as distribuições do diagnóstico pós-morte são similares. Sendo assim não aparenta haver relação entre diagnóstico pós-morte e a distinção entre adultos e crianças. Mais uma vez é importante observar que a quantidade de unidades amostrais no grupo de crianças é pequeno (15).

Na Figura B.12 temos o mapa dos Distritos administrativos e Coordenadorias Regionais de Saúde do município de São Paulo. Com este mapa, foi criada a variável área de saúde. A Tabela A.14 exhibe a distribuição de frequências da variável diagnóstico pós-morte segundo área de saúde. Nesse caso, as proporções de pacientes com diagnóstico pós-morte diferem entre as áreas de saúde: nas regiões Centro-Oeste e Sudeste, a proporção de pacientes com diagnóstico pós-morte é de 45% e 47%, respectivamente, sendo que na região Leste é 59% (maior proporção). Já na região Norte tal proporção é de aproximadamente 38%, no Sul é de 23%. Principalmente quando se olha para a região Sul, essa região é a única que intuitivamente não vai de acordo com o índice de Moran apresentado anteriormente.

Em algumas tabelas, como as Tabelas A.12 e A.13, há caselas com poucas observações (por exemplo, há um paciente com diagnóstico pós-morte residente fora do estado de São Paulo e apenas 5 crianças com diagnóstico pós-morte). Esses valores reduzidos de observações nessas categorias podem comprometer o uso adequado do teste de independência baseado na estatística qui-quadrado (Agresti, 2019) para averiguar associação de diagnóstico pós-morte com outra variável.

Assim, adicionalmente ao teste usual baseado na estatística de qui-quadrado, foram conduzidos o teste exato de Fisher (Agresti, 2019) e o teste FBST (Pereira e Stern, 1999) baseado na medida de evidência de Pereira-Stern por não envolverem resultados aproximados (aproximações) para grandes amostras.

Na Tabela A.15, apresentamos medidas de evidência para a hipótese de inexistência de associação entre diagnóstico pós-morte e uma outra variável sob consideração. Primeiro, é apresentada a medida de evidência de Pereira-Stern (Pereira e Stern, 1999) sob a perspectiva bayesiana. Em seguida, são apresentados os valores-p do teste de qui-quadrado e do teste exato de Fisher. Em todos os casos, quanto maior o valor do índice, maior a indicação de inexistência de associação entre as variáveis. Ou, quanto menor o valor do índice, maior o indício de associação entre diagnóstico pós-morte e a outra variável sob consideração.

Para o ano, a evidência do teste FBST (0,355) indica alguma evidência de associação entre diagnóstico pós-morte e ano. O teste qui-quadrado (valor-p < 0,001) e o teste exato de Fisher (valor p < 0,001) indicam a existência de associação entre tais variáveis.

Para sexo, a evidência FBST (0,957) indica alta evidência de independência (inexistência de associação) entre diagnóstico pós-morte e sexo. O teste qui-quadrado (valor-p = 0,577) e o teste exato de Fisher (valor p = 0,650) indicam o mesmo.

Considerando a região, a evidência FBST (0,231) sugere a existência de associação entre diagnóstico pós-morte e região. O teste qui-quadrado (valor-p = 0,056) e o teste exato de Fisher (valor p = 0,045) apontam o mesmo, ou seja, indicam que diagnóstico pós-morte depende da região de procedência do paciente. Deve-se salientar que os resultados desses testes devem ser tomados com cautela pois há poucos pacientes provenientes de outros estados na amostra, dos quais apenas um teve diagnóstico pós-morte.

Para a variável criança, a evidência FBST (0,968) indica alta evidência de inexistência de associação entre diagnóstico pós-morte e ser criança ou não. O teste qui-quadrado (valor-p = 0,613) e o teste exato de Fisher (valor p = 0,788) indicam o mesmo.

Para a área de saúde, a evidência FBST (0,298) sugere a existência de associação entre diagnóstico pós-morte e área de saúde. O teste qui-quadrado (valor-p = 0,033) e o teste exato de Fisher (valor p = 0,031) apontam o mesmo, indicando que o diagnóstico pós-morte depende da região de área de saúde do paciente, ou seja, o

diagnóstico pós-morte depende do paciente ser do Centro-Oeste, Leste, Norte, Sudeste, ou Sul da região de saúde da cidade de São Paulo.

Foram ajustados ainda modelos de regressão logística (Agresti, 2019) para a variável diagnóstico pós-morte segundo as variáveis idade, sexo, ano, distância da 1ª porta mais próxima a pé (km), São Paulo (capital), e interação entre distância da 1ª porta mais próxima a pé (km) com São Paulo (capital).

Na Tabela A.16, são apresentados os modelos ajustados (os diferentes modelos incorporam conjuntos diferentes de variáveis explicativas), as estimativas para os correspondentes parâmetros e os valores-p. Observa-se que para os modelos de 1 a 5, somente a variável idade apresenta valor-p menor que 0,05, o que indica que somente a idade é uma variável significativa ao nível de 5% no modelo de regressão logística. Nota-se que a variável ano, é significativa apenas ao nível de 10%. para avaliar a qualidade do ajuste desses modelos, calculamos a área sob a curva ROC (Agresti, 2019). Essa curva para o modelo que inclui apenas a variável explicativa idade é apresentada na Figura B.13. Dos valores da última linha da Tabela A.16, percebe-se que nenhum dos modelos ajustados seria adequado para estimar a probabilidade de o paciente ter diagnóstico pós-morte a partir das variáveis explicativas consideradas.

## 6. Conclusões

A análise dos dados aponta que os grupos de classificação 1 e 4 são pequenos, com predominância dos grupos 2 e 3 na amostra. Sendo assim, os pacientes foram reclassificados em apenas dois grupos maiores, segundo a ocorrência ou não de diagnóstico pós-morte. Seria necessária uma amostra maior para ter os 4 grupos de classificação com um tamanho razoável para inferência.

Inicialmente, não foram encontradas associações da variável de interesse, diagnóstico pós-morte, e as demais variáveis como idade, sexo e distância da 1ª porta mais próxima a pé. Deve-se salientar a possibilidade de inconsistência na análise dessa última variável pois os dados relativos aos pontos do SUS mais próximos do paciente

foram obtidos em 2024 sendo que os dados de diagnóstico pós-morte da amostra advêm dos anos de 2007 e 2008. Esse fato pode inviabilizar que os dados forneçam informação exata sobre as portas do SUS mais próximas do paciente na época.

Não se obteve associação espacial entre os grupos de diagnóstico pós-morte e localização geográfica. Talvez com mais informações relacionadas aos distritos geográficos do município de São Paulo, como o IDH por distrito, uma associação espacial entre classificação e outros fatores possa ser detectada.

Verificamos que parece existir associação entre diagnóstico pós-morte e região, ou seja, o diagnóstico pós-morte parece depender se o paciente era residente do município de São Paulo, era residente do estado mas não do município de São Paulo, ou se era residente fora do estado de São Paulo. Essa associação se deve possivelmente pelo fato de que pacientes fora do estado procuram hospitais do sistema de saúde da cidade de São Paulo para tratamento de saúde já tendo o diagnóstico de neoplasia..

Existe associação entre diagnóstico pós-morte e área de saúde, ou seja, o diagnóstico pós-morte parece depender da área de saúde do paciente na cidade de São Paulo: Centro-Oeste, Leste, Norte, Sudeste, ou Sul, segundo o Atlas da Saúde 2011. Nota-se que a área Norte e, especialmente, a área Sul têm uma porcentagem menor de diagnóstico pós-morte (Norte: 38%, Sul: 23%), as áreas Centro-Oeste (45%) e Sudeste (47%) têm porcentagens intermediárias, ao passo que a área Leste (59%) tem uma porcentagem maior de diagnóstico pós-morte.

Apesar de existir alguma associação estatística entre ano e diagnóstico pós-morte, deve-se notar que a princípio o ano não deveria afetar o diagnóstico pós-morte, de acordo com os pesquisadores.

Uma sugestão de análise adicional seria avaliar os dados segundo o tipo de neoplasia diagnosticada. Há muitos tipos diferentes de neoplasias, a velocidade de crescimento das diferentes neoplasias depende de diversos fatores, o que pode contribuir para o sucesso ou a falha no diagnóstico da neoplasia antes do óbito.

# APÊNDICE A

## Tabelas

**Tabela A.1** Medidas-resumo para a variável Idade.

<b>Idade (anos)</b>	
<b>mínimo</b>	0,00
<b>1º quartil</b>	48,00
<b>média</b>	56,13
<b>mediana</b>	59,00
<b>3º quartil</b>	69,75
<b>máximo</b>	90,00

**Tabela A.2** Distribuição de frequências da variável Classificação .

<b>Classificação</b>	<b>Frequência</b>
<b>1</b>	10 (3%)
<b>2</b>	119 (36%)
<b>3</b>	182 (56%)
<b>4</b>	15 (5%)
<b>Total</b>	326(100%)

**Tabela A.3** Distribuição de frequências da variável Diagnóstico pós-morte .

<b>Diagnóstico pós-morte</b>	<b>Frequência</b>
<b>Sim</b>	129 (40%)
<b>Não</b>	197 (60%)
<b>Total</b>	326 (100%)

**Tabela A.4** Distribuição de frequências da variável Ano da autópsia .

<b>Ano da autópsia</b>	<b>Frequência</b>
<b>2007</b>	172 (53%)
<b>2008</b>	154 (47%)
<b>Total</b>	326 (100%)

**Tabela A.5** Distribuição de frequências da variável Sexo .

<b>Sexo</b>	<b>Frequência</b>
<b>Masculino</b>	173 (53%)
<b>Feminino</b>	153 (47%)
<b>Total</b>	326 (100%)

**Tabela A.6** Medidas-resumo para as variáveis quantitativas de distância em km.

	<b>Distância da 1ª porta mais próxima a pé</b>	<b>Distância da 2ª porta mais próxima a pé</b>	<b>Distância da 3ª porta mais próxima a pé</b>
<b>Mínimo</b>	0,11	0,23	0,67
<b>1º quartil</b>	0,64	1,12	1,43
<b>Média</b>	44,05	44,60	44,98
<b>Mediana</b>	1,03	1,59	1,92
<b>3º quartil</b>	2,60	3,11	3,84
<b>Máximo</b>	3097,03	3097,69	3097,74
	<b>Distância da 1ª porta mais próxima de carro</b>	<b>Distância da 2ª porta mais próxima de carro</b>	<b>Distância da 3ª porta mais próxima de carro</b>
<b>Mínimo</b>	0,11	0,36	0,65
<b>1º quartil</b>	0,72	1,27	1,65
<b>Média</b>	39,07	39,71	40,17
<b>Mediana</b>	1,13	1,87	2,24
<b>3º quartil</b>	2,89	3,49	4,53
<b>Máximo</b>	2730,68	2731,16	2732,55

**Tabela A.7** Distribuição de frequências da variável São Paulo (capital).

<b>São Paulo (capital)</b>	<b>Frequência</b>
<b>Sim</b>	216 (66%)
<b>Não</b>	110 (34%)
<b>Total</b>	326(100%)

**Tabela A.8** Medidas-resumo para as variáveis quantitativas de distância em km para a amostra reduzida aos residentes em São Paulo (capital).

	<b>Distância da 1ª porta mais próxima a pé</b>	<b>Distância da 2ª porta mais próxima a pé</b>	<b>Distância da 3ª porta mais próxima a pé</b>
<b>Mínimo</b>	0,11	0,23	0,67
<b>1º quartil</b>	0,56	0,99	1,26
<b>Média</b>	0,89	1,38	1,71
<b>Mediana</b>	0,79	1,25	1,58
<b>3º quartil</b>	1,07	1,66	1,99
<b>Máximo</b>	6,59	6,87	7,18

**Tabela A.8(continuação)** Medidas-resumo para as variáveis quantitativas de distância em km para a amostra reduzida aos residentes em São Paulo (capital).

	Distância da 1ª porta mais próxima de carro	Distância da 2ª porta mais próxima de carro	Distância da 3ª porta mais próxima de carro
Mínimo	0,11	0,36	0,65
1º quartil	0,60	1,10	1,51
Média	0,99	1,60	1,99
Mediana	0,86	1,45	1,84
3º quartil	1,21	1,92	2,32
Máximo	6,93	7,07	7,57

**Tabela A.9** Distribuição de frequências de Diagnóstico pós-morte por Ano da autópsia.

Diagnóstico pós-morte	Ano da autópsia		
	2007	2008	Total
Não	96 (55%)	101 (66%)	197 (60%)
Sim	76 (44%)	53 (34%)	129 (40%)
Total	172 (100%)	154 (100%)	326 (100%)

**Tabela A.10** Distribuição de frequências de Diagnóstico pós-morte por Sexo.

Diagnóstico pós-morte	Sexo		
	Masculino	Feminino	Total
Não	107 (62%)	90 (59%)	197(60%)
Sim	66 (38%)	63 (41%)	129(40%)
Total	173 (100%)	153 (100%)	326 (100%)

**Tabela A.11** Medidas-resumo para Idade (anos) por Diagnóstico pós-morte.

	Diagnóstico pós-morte	
	Não	Sim
Mínimo	0,00	0,00
1º quartil	43,00	52,00
Média	54,45	58,99
Mediana	58,00	61,00
3º quartil	68,00	71,00
Máximo	89,00	90,00



**Tabela A.12** Distribuição de frequências de Diagnóstico pós-morte por Região.

	São Paulo (capital)	São Paulo (estado)	Externo	Total
<b>Não</b>	127 (59%)	58 (59%)	12 (92%)	197 (60%)
<b>Sim</b>	88 (41%)	40 (41%)	1 (8%)	129 (40%)
<b>Total</b>	215 (100%)	98 (100%)	13 (100%)	326 (100%)

**Tabela A.13** Distribuição de frequências de Diagnóstico pós-morte por Criança.

	Criança	Adulto	Total
<b>Não</b>	10 (67%)	186 (60%)	197 (60%)
<b>Sim</b>	5 (33%)	124 (40%)	129 (40%)
<b>Total</b>	15 (100%)	311 (100%)	326 (100%)

**Tabela A.14** Distribuição de frequências de Diagnóstico pós-morte por Área de saúde.

	Centro-Oeste	Leste	Norte	Sudeste	Sul	Total
<b>Não</b>	33 (55%)	12 (41%)	30 (63%)	18 (53%)	33 (77%)	126 (59%)
<b>Sim</b>	27 (45%)	17 (59%)	18 (38%)	16(47%)	10 (23%)	88 (41%)
<b>Total</b>	60 (100%)	29 (100%)	48 (100%)	34 (100%)	43 (100%)	214 (100%)

**Tabela A.15** Evidência de independência FBST, qui-quadrado de Pearson e teste exato de Fisher, para as tabelas das variáveis por Diagnóstico pós-morte.

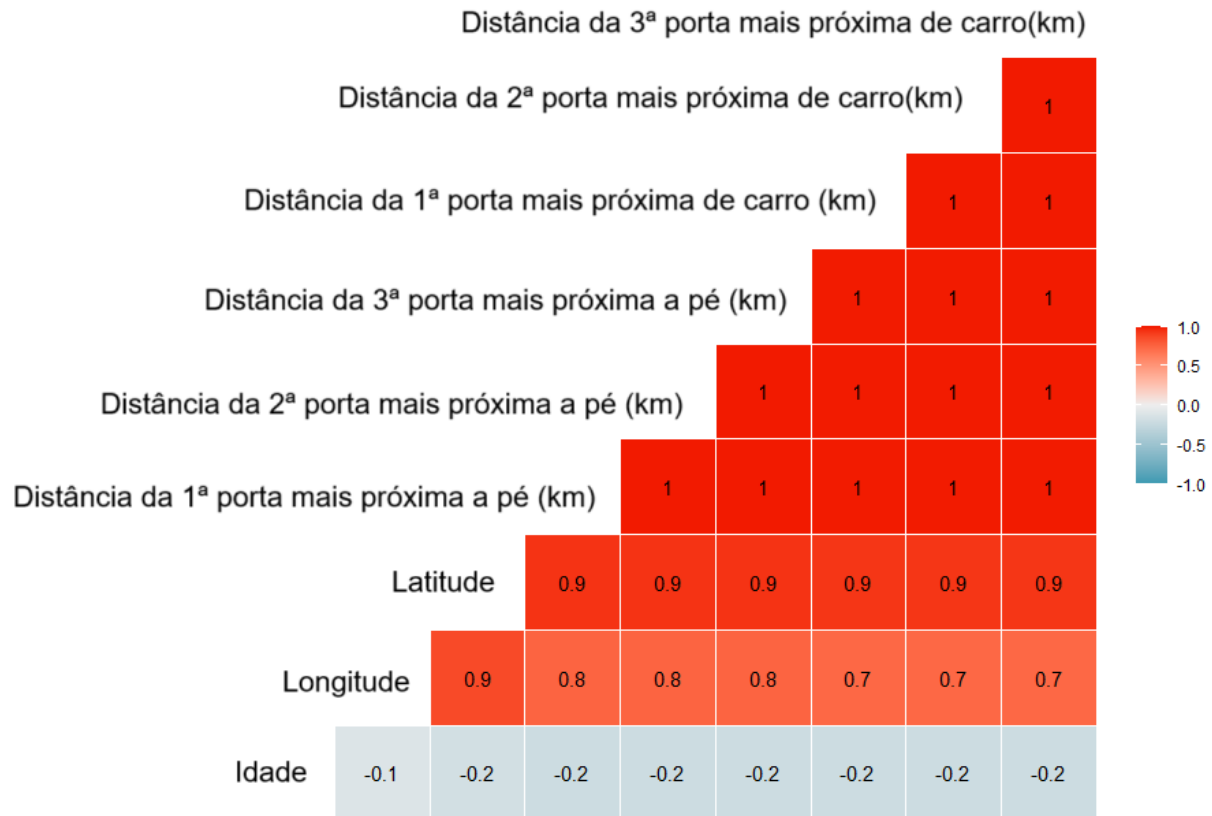
Tabela	Variável	Evidência de independência FBST	Qui-quadrado valor-p	Fisher valor-p
A.9	Ano	0,355	0,000	0,000
A.10	Sexo	0,957	0,577	0,650
A.12	Região	<b>0,231</b>	<b>0,056</b>	<b>0,045</b>
A.13	Criança	0,968	0,613	0,788
A.14	Área de saúde	<b>0,298</b>	<b>0,033</b>	<b>0,031</b>

Tabela A.16 Modelos de regressão logística

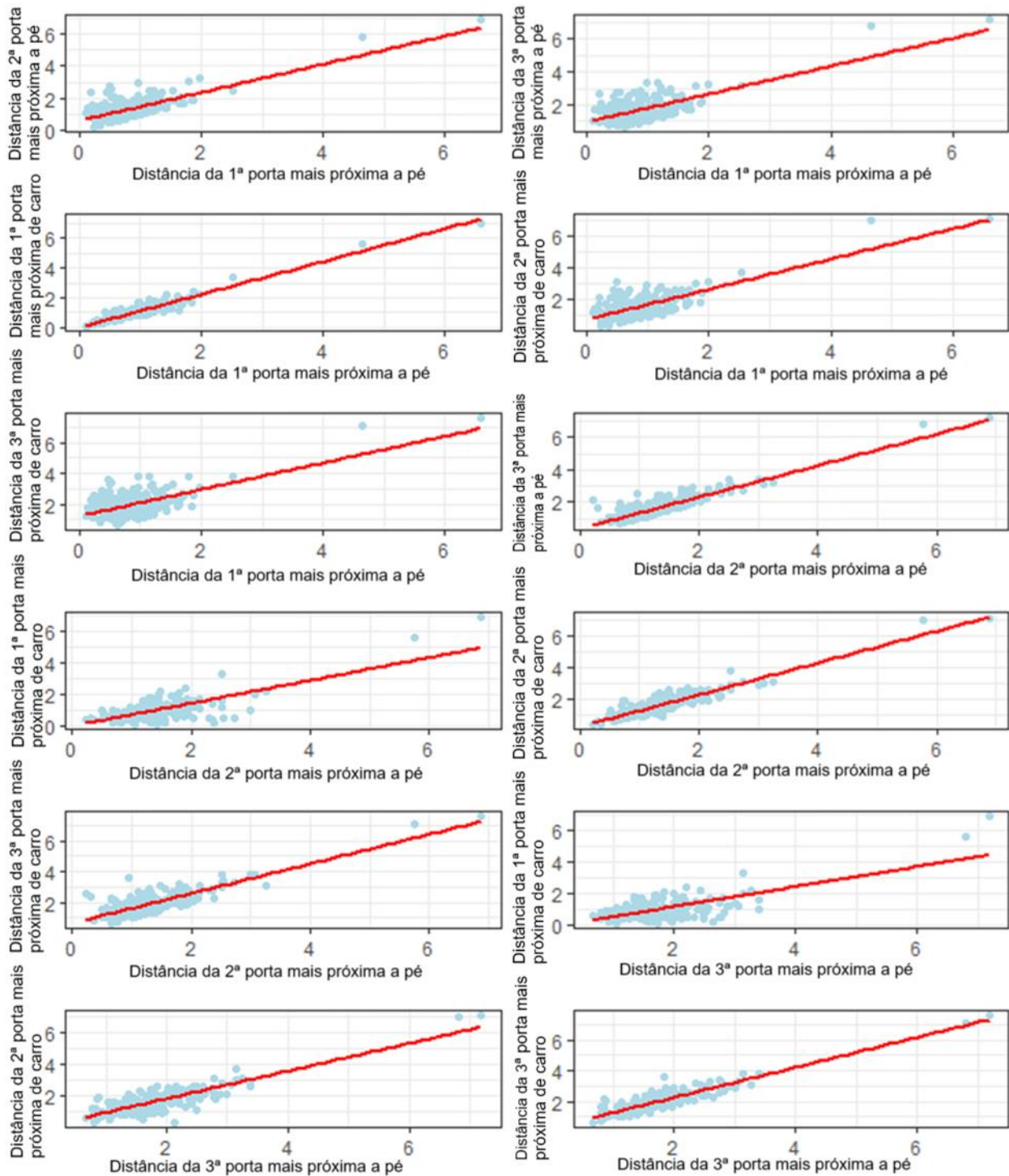
	<b>Modelo 1</b>	<b>Modelo 2</b>	<b>Modelo 3</b>	<b>Modelo 4</b>	<b>Modelo 5</b>	<b>Modelo 6</b>
<b>Intercepto</b>	-0,862	-1,009	-1,248	-0,997	-1,196	-0,423
<b>Desvio padrão</b>	0,430	0,403	0,391	0,395	0,377	0,113
<b>valor-p</b>	0,045	0,012	0,001	0,012	0,002	0,000
<b>Idade</b>	0,012	0,012	0,014	0,013	0,014	
<b>Desvio padrão</b>	0,006	0,006	0,006	0,006	0,006	
<b>valor-p</b>	0,074	0,069	0,031	0,034	0,030	
<b>SexoMasculino</b>	0,107		0,119			
<b>Desvio padrão</b>	0,232		0,229			
<b>valor-p</b>	0,644		0,604			
<b>Ano2008</b>	-0,386			-0,397		
<b>Desvio padrão</b>	0,233			0,230		
<b>valor-p</b>	0,097			0,085		
<b>Distância</b>	-0,002	-0,002				
<b>Desvio padrão</b>	0,002	0,002				
<b>valor-p</b>	0,233	0,234				
<b>SP (capital)</b>	0,130	0,140				
<b>Desvio padrão</b>	0,342	0,338				
<b>valor-p</b>	0,705	0,679				
<b>Distância*SP</b>	-0,228	-0,217				
<b>Desvio padrão</b>	0,264	0,257				
<b>valor-p</b>	0,387	0,398				
<b>AIC</b>	438,41	437,45	437,44	435,71	436,7	439,64
<b>Área ROC</b>	0,604	0,585	0,574	0,595	0,573	0,500

# APÊNDICE B

## Figuras



**Figura B.1** Heatmap das correlações entre as variáveis quantitativas, 2 a 2.



**Figura B.2** Gráficos de dispersão entre as variáveis quantitativas de distância, 2 a 2.

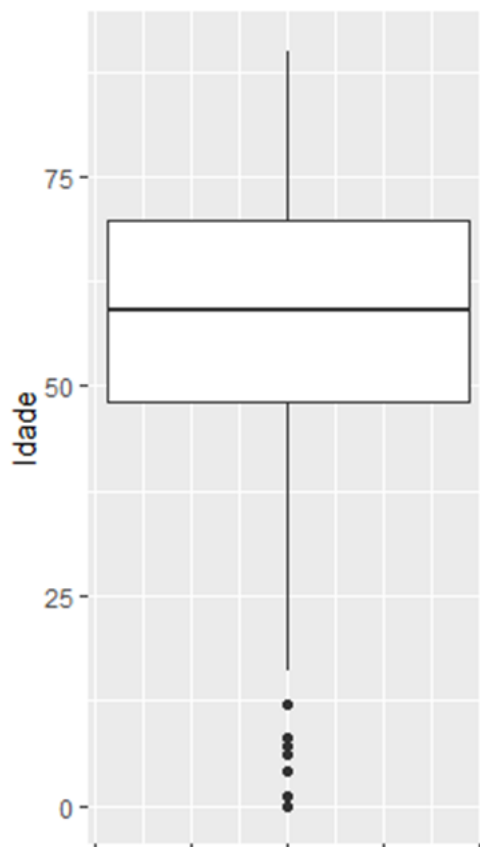


Figura B.3 *Box plot* da variável Idade.

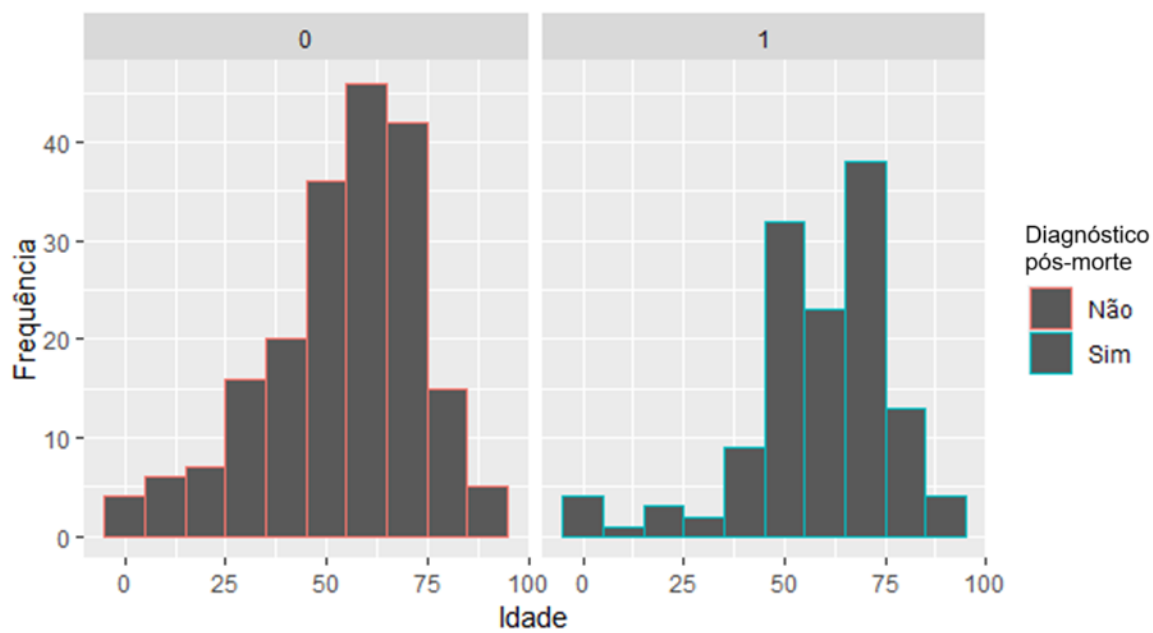
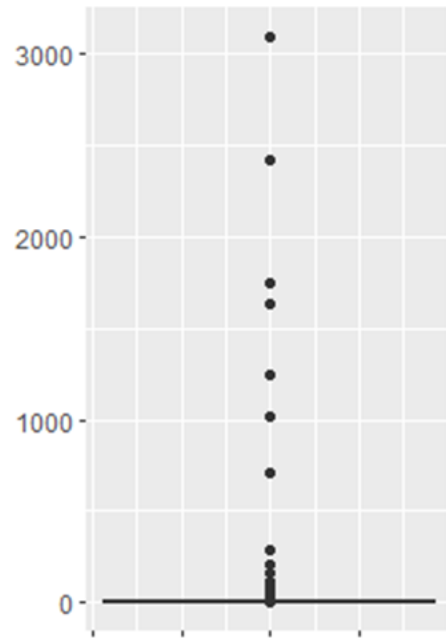


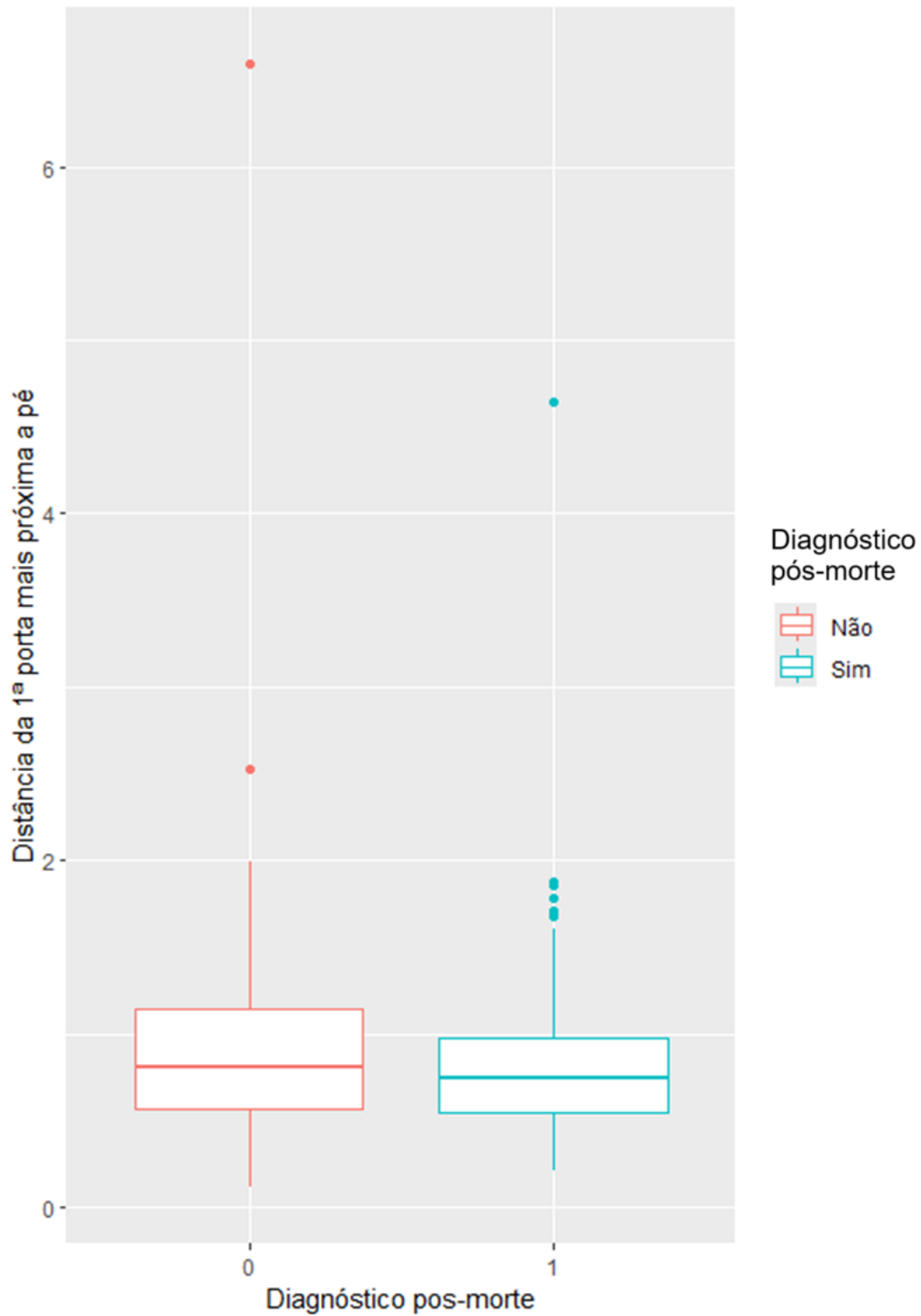
Figura B.4 Histograma da variável Idade em relação ao Diagnóstico pós-morte.



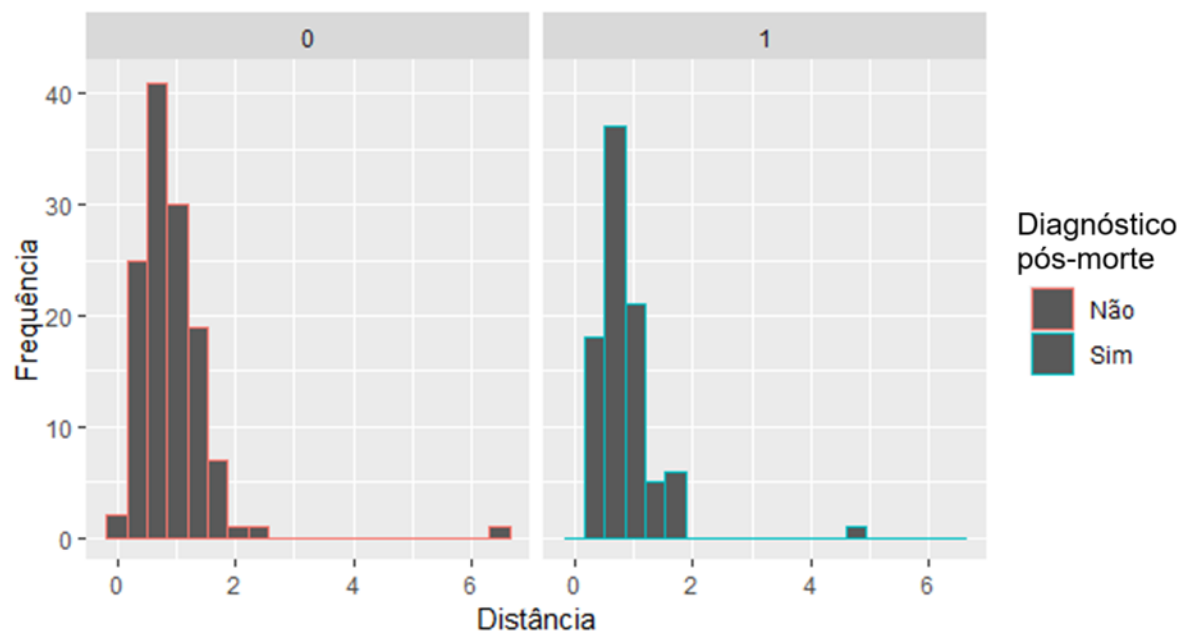
**Figura B.5** *Box plot* da variável Distância da 1ª porta mais próxima a pé.



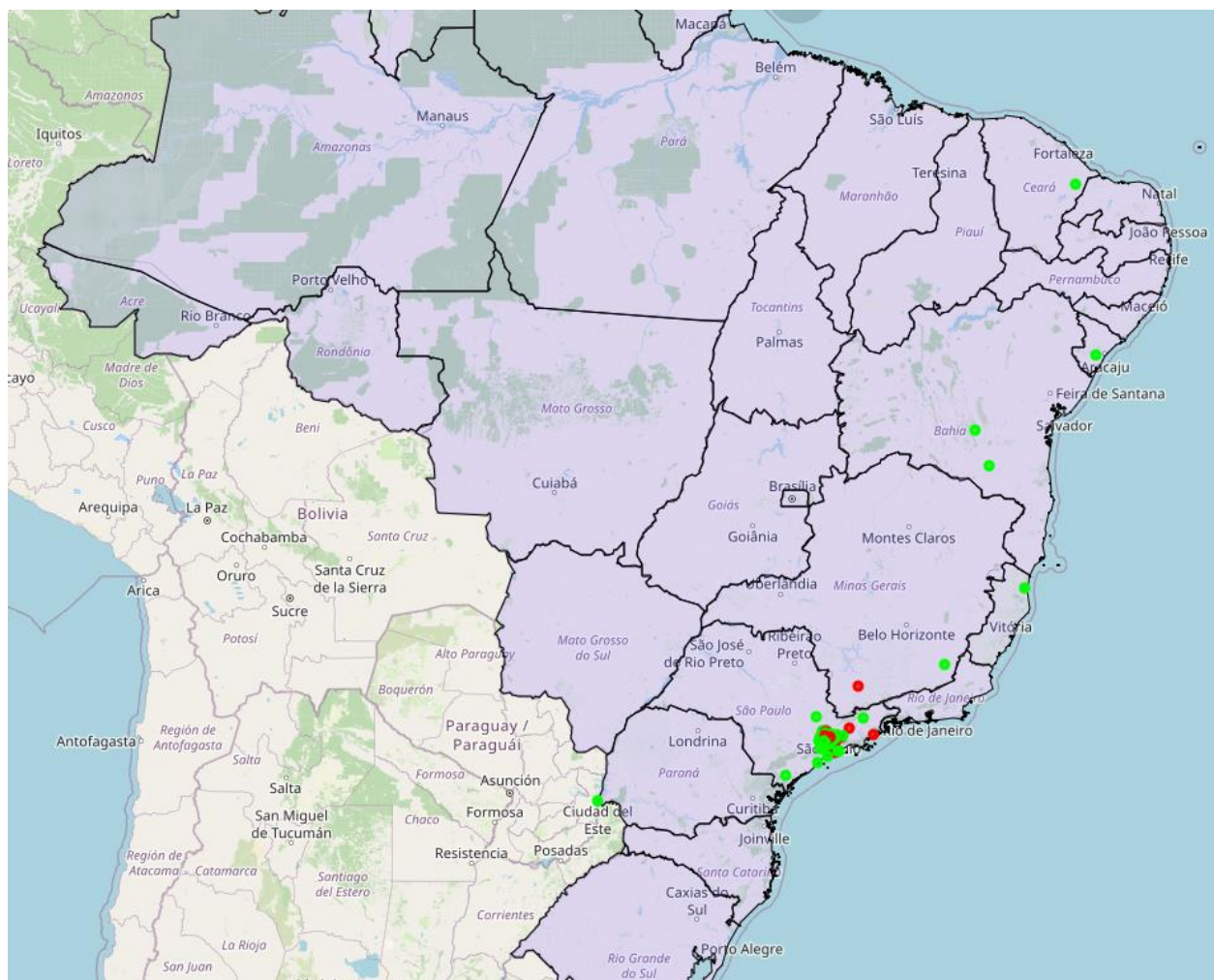




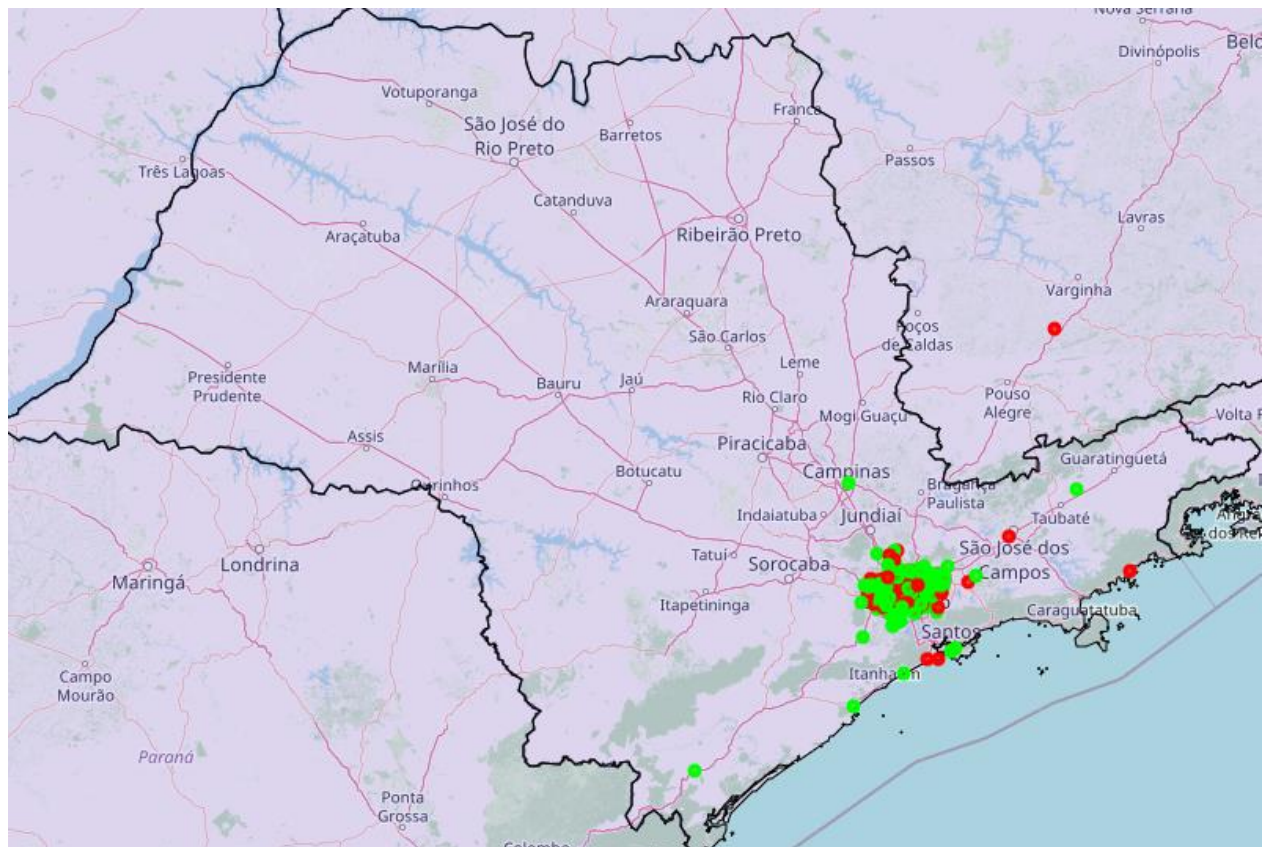
**Figura B.7** *Box plot* da variável Distância da 1ª porta mais próxima a pé em relação ao Diagnóstico pós-morte para a amostra reduzida aos residentes em São Paulo (capital).



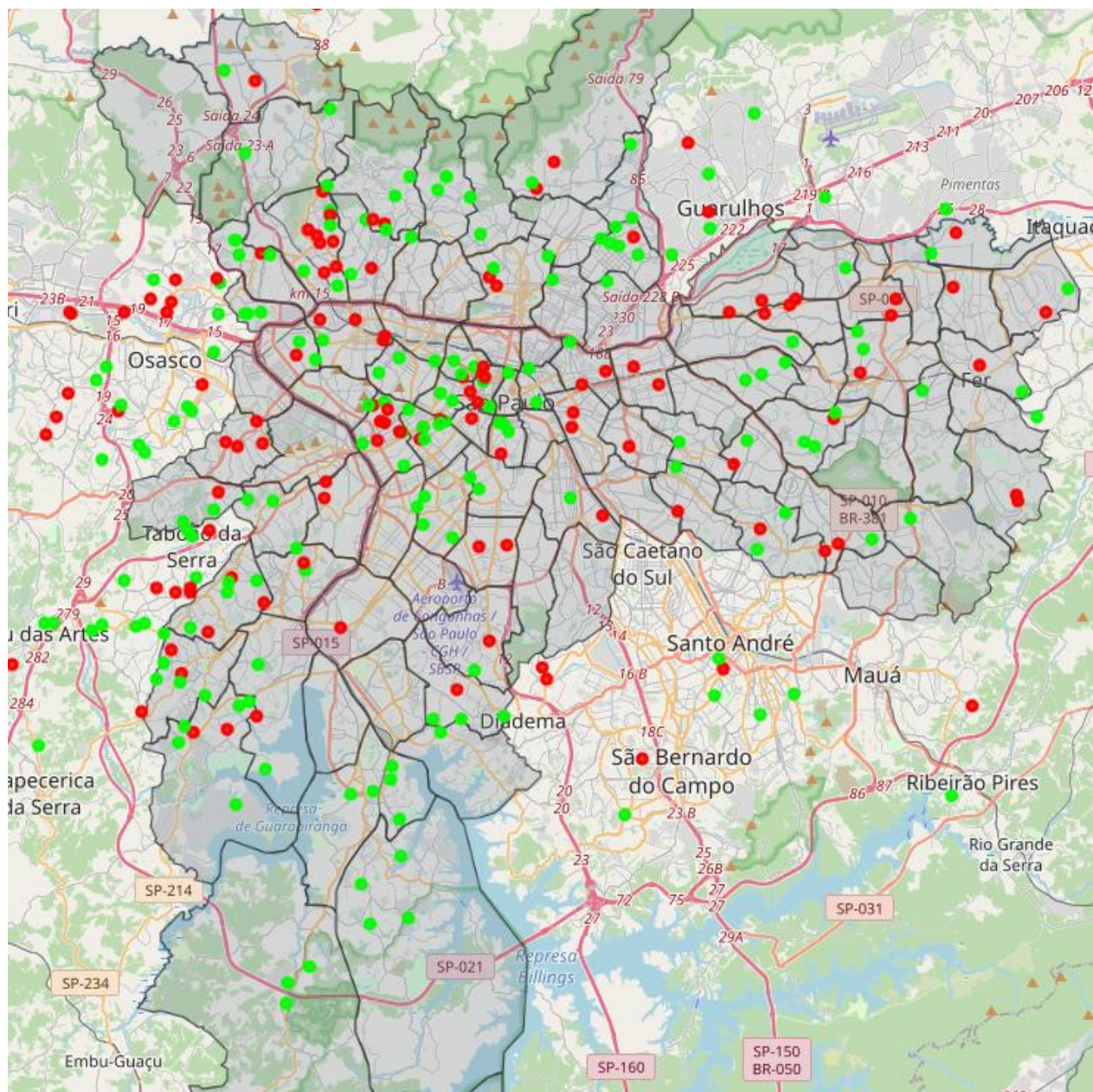
**Figura B.8** Histograma da variável Distância da 1ª porta mais próxima a pé em relação ao Diagnóstico pós-morte para a amostra reduzida aos residentes em São Paulo (capital).



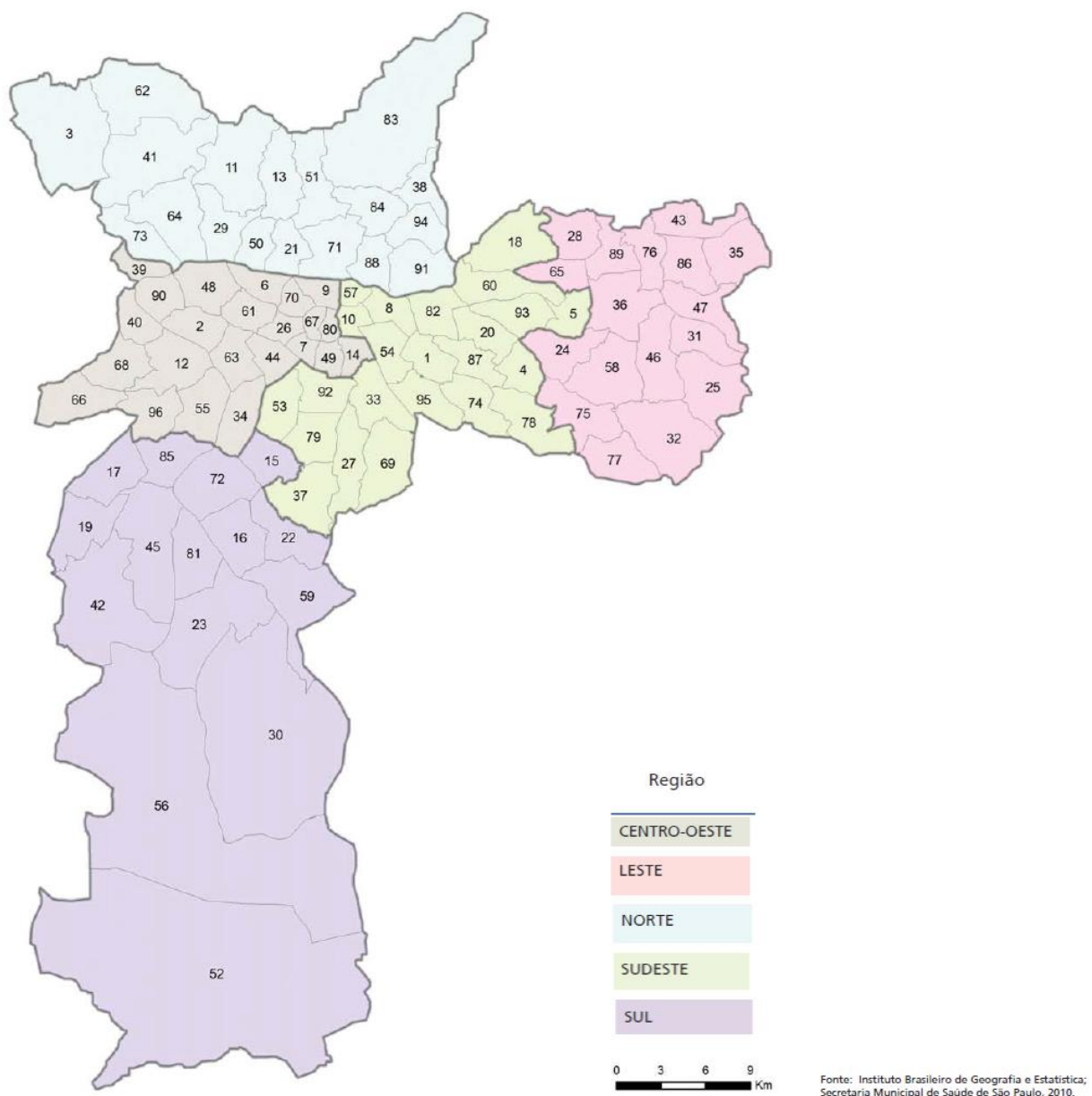
**Figura B.9** Distribuição geográfica dos pacientes com foco no Brasil em relação ao Diagnóstico pós-morte: “sim” (em vermelho) e “não” (em verde).



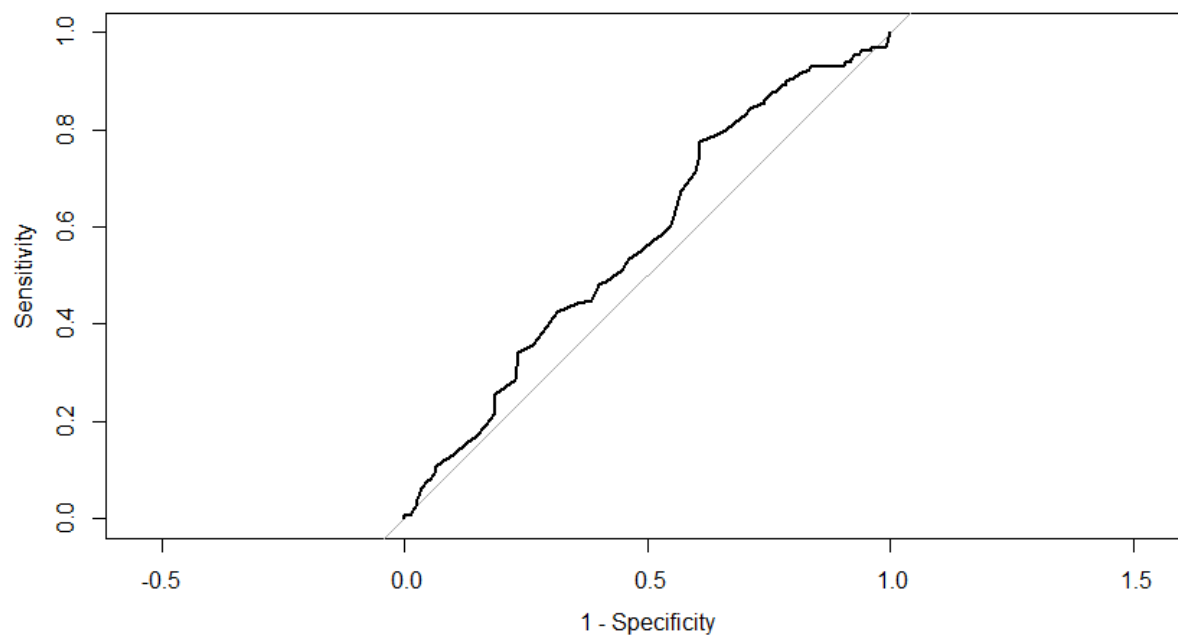
**Figura B.10** Distribuição geográfica dos pacientes com foco no estado de São Paulo em relação ao Diagnóstico pós-morte: “sim” (em vermelho) e “não” (em verde).



**Figura B.11** Distribuição geográfica dos pacientes com foco no município de São Paulo em relação ao Diagnóstico pós-morte: “sim” (em vermelho) e “não” (em verde).



**Figura B.12** Mapa dos Distritos administrativos e Coordenadorias Regionais de Saúde do município de São Paulo.



**Figura B.13** Curva ROC do modelo de regressão com variável resposta Diagnóstico pós-morte contra variável explicativa Idade.