

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

Centro de Estatística Aplicada

Relatório de Análise Estatística

RAE-CEA–20P14

RELATÓRIO DE ANÁLISE ESTATÍSTICA SOBRE O PROJETO:

“Mortalidade no longo prazo de pacientes que receberam atendimento de hospital acadêmico da rede SUS voltado para o atendimento médico cardiológico”

Guilherme Jordan Castro Nascimento

Lúcia Pereira Barroso

Nicole Zukowski Luduvise

São Paulo, outubro de 2020

CENTRO DE ESTATÍSTICA APLICADA - CEA – USP

CENTRO DE ESTATÍSTICA APLICADA - IME/ USP

TÍTULO: Relatório de Análise Estatística sobre o Projeto: “Mortalidade no longo prazo de pacientes que receberam atendimento de hospital acadêmico da rede SUS voltado para o atendimento médico cardiológico”.

PESQUISADOR(A): Carlos Lederman

ORIENTADOR(A): Prof. Dr. Alfredo J. Mansur

INSTITUIÇÃO: InCor / FMUSP

FINALIDADE DO PROJETO: Doutorado

RESPONSÁVEIS PELA ANÁLISE: Guilherme Jordan Castro Nascimento

Lúcia Pereira Barroso

Nicole Zukowski Luduvise

REFERÊNCIA DESTE TRABALHO: NASCIMENTO, G.J.C.; BARROSO, L.P.; LUDUVISE, N.Z. (2020). **Relatório de análise estatística sobre o projeto: “Mortalidade no longo prazo de pacientes que receberam atendimento de hospital acadêmico da rede SUS voltado para o atendimento médico cardiológico”**. São Paulo, IME-USP, 2020. (RAE–CEA-20P14)

FICHA TÉCNICA

REFERÊNCIAS BIBLIOGRÁFICAS:

FUNDAÇÃO SEADE (2020). **Retratos de São Paulo**. Disponível em: <<http://produtos.seade.gov.br/produtos/retratosdesp/>>. Acesso em: 16 de outubro de 2020.

GRAY, R.J. (1988). A class of K-sample tests for comparing the cumulative incidence of a competing risk. **The Annals of Statistics**. v. 16, n. 3, p. 1141-1154.

JOHNSON, R.A.; WICHERN, D.W. (2007). **Applied multivariate statistical analysis**. Upper Saddle River, N.J. Prentice Hall. p. 716-726.

KAPLAN, E.L.; MEIER, P. (1958). Nonparametric estimation from incomplete observations. **J. Amer. Statist. Assoc.** v. 53, n. 282, p. 457-481.

KLEIN, J. P.; MOESCHBERGER, M. L.; (2010). **Survival Analysis: Techniques for Censored and Truncated Data**. Springer, 1 ed.

LOEKITO, E.; BAILEY, J.; BELLOMO, R.; HART, G.K.; HEGARTY, C.; DAVEY, P.; BAIN, C.; PILCHER, D.; SCHNEIDER, H. (2013). Common laboratory tests predict imminent death in ward patients. **Resuscitation**, v. 84, n. 3, p. 280-285.

MANTEL, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. **Cancer Chemotherapy Reports**. v. 50, n.3, p. 163-70.

IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (2020). **Expectativa de vida dos brasileiros aumenta para 76,3 anos em 2018**. Disponível em: <<https://censo2020.ibge.gov.br/2012-agencia-de-noticias/noticias/26103-expectativa-de-vida-dos-brasileiros-aumenta-para-76-3-anos-em-2018.html>>. Acesso em: 16 de outubro de 2020.

PRENTICE, R.L.; KALBFLEISCH, J.D.; PETERSON, A.V.; FLOURNOY Jr, N.; FAREWELL, V.T.; BRESLOW, N.E. (1978). The analysis of failure times in the presence of competing risks. **Biometrics**. v.34, n.4 p. 541-554.

TUKEY, J.W. (1977). **Exploratory data analysis**. Reading, Mass.: Addison-Wesley.

WHO - WORLD HEALTH ORGANIZATION (2019). **International Classification of Diseases, 10th revision**. Disponível em: <<https://icd.who.int/browse10/2019/en>>. Acesso em: 16 de outubro de 2020.

PROGRAMAS COMPUTACIONAIS UTILIZADOS:

R for Windows, versão 3.6.3.

RStudio for Windows, versão 1.2.5042.

Microsoft Office Excel for Windows, versão 2019

TÉCNICAS ESTATÍSTICAS UTILIZADAS

Análise Descritiva Unidimensional (03:010)

Análise Descritiva Multidimensional (03:020)

Análise de Sobrevida (13:070)

Associação e Dependência de Dados Qualitativos (06:020)

ÁREA DE APLICAÇÃO

Bioestatística (14:030)

Resumo

A fim de entender quais os pacientes que procuram os serviços do InCor, e estudar a mortalidade no longo prazo destes pacientes, este estudo foi realizado com dados de prontuários eletrônicos de 78.059 pacientes atendidos no InCor e que faleceram por qualquer causa no período entre 2010 e 2016. Para alcançar este objetivo, análises descritivas foram utilizadas para entender o comportamento das variáveis. Além disso, Análise de Correspondência, gráficos de Kaplan-Meier e gráficos assumindo riscos competitivos também foram utilizados.

Em linhas gerais, concluiu-se que, considerando o óbito por doenças cardíacas e por outras causas como riscos competitivos, há diferenças nas funções de incidência acumulada para estado civil, raça ou cor e local de óbito. Aparentemente os solteiros, pardos e pretos e óbitos em domicílio, para causa cardíaca, apresentam função de incidência acumulada maior que as outras categorias respectivas. Para o local de óbito, quando se considera outras causas de morte, essa relação se inverte, sendo que a função de incidência acumulada é menor para os que faleceram em casa. Quanto ao sexo, a diferença ocorre somente para outras causas de óbito que não as cardíacas, sendo que os homens apresentam maior incidência.

A Análise de Correspondência mostrou associação entre os capítulos 16 (Afecções originadas no período perinatal) e 17 (Malformações genéticas) do CID10, ambos como diagnóstico e causa de óbito, o mesmo ocorrendo para o capítulo 2 (Neoplasmas).

Sumário

1. Introdução.....	9
2. Objetivo	9
3. Descrição do estudo	9
4. Descrição das variáveis.....	10
5. Análise estatística.....	12
5.1 Análise descritiva unidimensional	12
5.2 Análise de sobrevivência.....	14
5.3 Análise de correspondência.....	17
6 Conclusões.....	18
APÊNDICE A - Tabelas	20
APÊNDICE B - Figuras.....	27

1. Introdução

Com os avanços tecnológicos, os prontuários médicos passaram a ser registrados em meios eletrônicos, trazendo novas possibilidades para a atuação nos atendimentos hospitalares. As informações registradas nos prontuários eletrônicos hospitalares dizem respeito tanto ao período de internação dos pacientes quanto ao período de atendimento ambulatorial e podem incluir dados clínicos, dados de exames laboratoriais, intervenções realizadas e o desfecho da hospitalização. Desta forma, é possível acompanhar a evolução dos pacientes e estudar a mortalidade no longo prazo de pacientes que receberam atendimento hospitalar, bem como a mortalidade na evolução tardia depois da alta hospitalar.

Estudos com dados de registros eletrônicos têm sido realizados ao redor do mundo apresentando diversas descobertas. Um exemplo disso é o estudo realizado na Austrália, em que se concluiu que os resultados de testes de laboratório rotineiros podem sinalizar como desfecho o óbito em pacientes hospitalizados (Loekito et al., 2013).

Neste contexto, o projeto visa avaliar variáveis de exames clínicos e de laboratório rotineiros que poderiam servir como preditores úteis de morte iminente em pacientes hospitalizados, dando enfoque à saúde cardiovascular, dado que as doenças cardiovasculares são importantes causas de morbidade e mortalidade na população brasileira.

2. Objetivo

Este projeto visa estudar a relação de variáveis clínicas de interesse como perfil do paciente e diagnósticos de doenças com a mortalidade no longo prazo, de pacientes que receberam atendimento médico hospitalar no Instituto do Coração do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo (InCor).

3. Descrição do estudo

O estudo foi realizado com dados de prontuários eletrônicos de 78.059 pacientes atendidos no InCor que faleceram por qualquer causa no período entre 2010 e 2016. A base de dados analisada foi construída por meio da vinculação de prontuários eletrônicos do InCor com dados de declaração de óbito da Fundação Sistema Estadual de Análise de Dados (SEADE). Foram excluídos, pelo pesquisador,

os registros tecnicamente inacessíveis ou com inconsistências que limitam a análise, tais como repetições, incorreções e outras falhas que impeçam a correta identificação e interpretação dos dados.

4. Descrição das variáveis

As variáveis consideradas neste estudo podem ser agrupadas como: demográficas, hospitalares e referentes ao óbito.

Variáveis demográficas:

Idade (na data do óbito, em anos)

Sexo

masculino

feminino

Raça ou cor

branca

parda

preta

amarela

indígena

Município de residência:

São Paulo

Osasco

Guarulhos

Outros municípios de São Paulo

Outros estados

Variáveis hospitalares:

Número de registros na base de dados: 1; 2; 3 ou mais

Tempo entre a primeira e a última admissão (anos)

Tempo entre a primeira admissão e o óbito (anos)

Tempo entre a última admissão e o óbito (anos)

Número de diagnósticos distintos em vida: 1, 2, 3, 4, 5 ou mais

Indicador ambulatorial

1 se recebeu atendimento ambulatorial e 0 em caso contrário

Indicador pronto socorro

1 se recebeu atendimento no pronto socorro e 0 em caso contrário

Indicador internação

1 se recebeu atendimento de internação e 0 em caso contrário

Indicador SA

1 se recebeu atendimento do serviço de apoio ao diagnóstico e terapêutica e 0 em caso contrário

Indicador cirurgia cardíaca

1 se fez alguma cirurgia cardíaca e 0 em caso contrário

Indicador outras cirurgias

1 se fez alguma cirurgia (não cardíaca) e 0 em caso contrário

Diagnósticos InCor - códigos das doenças agrupados por capítulos do CID-10 (WHO, 2019) dos diagnósticos feitos para cada paciente no InCor

Variáveis referentes ao óbito:

Local do óbito:

hospital

domicílio

outros estabelecimentos de saúde

via pública

outros

Indicador de óbito no InCor: sim ou não

Causa básica do óbito

Doenças infecciosas e parasitárias

Neoplasias

Doenças do sangue

Doenças endócrinas

Transtornos mentais e comportamentais

Doenças do sistema nervoso

Doenças do olho e anexos

Doenças do ouvido e da apófise mastoide

Doenças do aparelho circulatório

Doenças do aparelho respiratório

Doenças do aparelho digestivo
Doenças da pele e do tecido subcutâneo
Doenças osteomusculares e do tecido conjuntivo
Doenças do aparelho geniturinário
Gravidez
Afecções originadas no período perinatal
Malformações genéticas
Sintomas de exames não classificados
Lesões
Causas externas de morbidade e de mortalidade
Fatores que influenciam o estado de saúde

5. Análise estatística

5.1 Análise descritiva unidimensional

Com base na Tabela A.1 e na Figura B.1, verifica-se que a população em estudo tem idade que varia entre 0 e 111 anos e é composta majoritariamente por idosos, com idade média de 69 anos no óbito, cerca de sete anos a menos que a expectativa de vida dos brasileiros em 2018 (IBGE, 2020). Pacientes com zero anos de idade são, em geral, recém nascidos atendidos no InCor, que não sobreviveram. Há 54% de homens na população de estudo (Tabela A.2), valor maior que os 49% da população do estado de São Paulo (SEADE, 2020). A Tabela A.3 mostra que a proporção de brancos (76%) também é maior que a do estado (64%). A Tabela A.4 mostra que a maior parte (60%) dos pacientes residia no município de São Paulo, seguido por Osasco e Guarulhos (4% e 3%, respectivamente). Observa-se também que apenas 1% dos indivíduos residiam em outros estados.

A fim de estudar a relação entre as variáveis clínicas e a mortalidade no longo prazo, é importante entender como são as interações dos pacientes com o hospital. Em relação ao número de registros na base de dados, cerca de 94% dos indivíduos têm apenas um registro, 5% tem dois registros e menos de 1% tem três registros ou mais (Tabela A.5). Além disso, observa-se que aproximadamente 86% dos pacientes

não fizeram cirurgia e apenas 11% fizeram cirurgia cardíaca (Tabela A.6). Quanto ao tipo de atendimento recebido pelos pacientes (Tabela A.7), as proporções de atendimento ambulatorial, pronto socorro e serviço de apoio ao diagnóstico e terapêutica são bastante equilibradas (34%, 32% e 30%, respectivamente), o que sugere que esses serviços têm um alcance parecido com a população em estudo. Em contrapartida, apenas 6% dos pacientes receberam algum atendimento em internação. É importante mencionar que, apesar de alguns pacientes terem sido atendidos em mais de um serviço, não há registro do tipo de atendimento para 12.312 pacientes.

Outra informação relevante para o estudo da mortalidade no longo prazo consiste nos diagnósticos recebidos pelos pacientes ao longo de suas vidas. Observa-se, pela Tabela A.8, que 44% dos pacientes não têm nenhum diagnóstico codificado registrado no prontuário e 21% têm um diagnóstico. Além disso, 17% dos pacientes têm 5 ou mais diagnósticos registrados. A Tabela A.9 mostra que 43% dos pacientes receberam pelo menos um diagnóstico de doenças do aparelho circulatório no InCor. Esse percentual é de 12% para doenças endócrinas, nutricionais e metabólicas e 11% para doenças do aparelho respiratório.

Quanto às condições dos óbitos, 87% deles ocorreram em hospitais ou outros estabelecimentos de saúde (Tabela A.10). Entretanto, do total de óbitos, apenas 8% ocorreu no InCor (Tabela A.11). Para 40% dos pacientes, a causa básica do óbito foi alguma doença do aparelho circulatório. Para 19% foi alguma neoplasia e para 13% alguma doença do sistema respiratório (Tabela A.12).

Buscou-se também estudar o tempo entre alguns eventos de interesse, a saber, primeira admissão no InCor, última admissão no InCor e óbito. Apenas para pacientes com pelo menos duas admissões em datas distintas foi calculado o tempo entre a primeira e a última admissão. Para todos os pacientes com dados disponíveis, foi calculado o tempo entre a primeira admissão e o óbito. Também foi calculado o tempo entre a última admissão e o óbito. Para pacientes com um único atendimento, esse valor é igual ao tempo entre a primeira admissão e o óbito. A Tabela A.13 apresenta estatísticas descritivas para essas variáveis e as Figuras B.2 a B.4 apresentam histogramas e *box plots* (Tukey, 1977). As três variáveis apresentam comportamento bastante assimétrico, com grande concentração de valores próximos

a zero. Os seus valores variam de zero a aproximadamente 17 anos. Em termos de média e mediana, os tempos entre a última admissão e o óbito são menores que os tempos entre a primeira e a última admissão (médias de 2.7 e 4.2 anos; medianas de 0.9 e 3.0 anos, respectivamente). Com base no coeficiente de variação, o tempo entre a última admissão e o óbito possui a maior variabilidade (134%) se comparado às outras duas variáveis, que ainda assim têm alta variabilidade. Cabe ressaltar que estas análises foram baseadas em parte da amostra, pois faltam dados de admissões para grande parte dos pacientes.

5.2 Análise de sobrevivência

A fim de estudar a mortalidade no longo prazo, foram feitas análises utilizando recursos da Análise de Sobrevivência. O tempo entre o primeiro atendimento e o óbito foi estudado por meio do estimador de Kaplan-Meier (Kaplan e Meier, 1958) e de análises descritivas considerando as possíveis causas de morte como riscos competitivos (Prentice, 1978). É importante destacar que as interpretações destas análises devem ser feitas com cautela, pois, dada a natureza dos dados, o que está sendo estudado é o tempo de acompanhamento, desde a primeira admissão no InCor, de uma parte dos pacientes que vieram a óbito entre 2010 e 2016, o que por sua vez é diferente do tempo de sobrevida após o diagnóstico de uma doença específica. Além disso, não estão sendo considerados os pacientes vivos ou que vieram a óbito antes de 2010 ou após 2016, o que limita a generalização dos resultados.

Os estimadores de Kaplan Meier são estimadores não paramétricos usados, neste contexto, para estimar a função de distribuição do tempo de acompanhamento dos pacientes até seu óbito. As Figuras B.5 a B.9 apresentam gráficos de Kaplan-Meier para o tempo de acompanhamento, agrupando-se por diferentes variáveis de interesse. As figuras também exibem o valor-p do teste log-rank (Mantel, 1966) indicando se pelo menos uma curva é estatisticamente diferente das demais. Observa-se que, para pacientes do sexo masculino, o valor da função de sobrevivência é menor que o de pacientes do sexo feminino, para qualquer valor de

tempo, ou seja, o tempo de acompanhamento de homens é, em geral, menor do que o tempo de acompanhamento de mulheres (Figura B.5). Pela Figura B.6, verifica-se que as curvas para indivíduos brancos e amarelos aparentam estar bem próximas e, em geral, sugerem tempo de acompanhamento mais longos para indivíduos dessas cores, em relação a pretos e pardos. Considerando o estado civil, pacientes viúvos têm maior tempo de acompanhamento, seguido de pacientes com separação judicial (apenas 187 indivíduos), união consensual e casados. Os indivíduos solteiros têm o menor tempo de acompanhamento (Figura B.7). A Figura B.8 apresenta as curvas de Kaplan Meier agrupadas por local de ocorrência do óbito. Observa-se que o tempo de acompanhamento é, em geral, maior para pacientes que faleceram em casa ou em outros locais, seguido de pacientes que morreram em via pública. Os pacientes com menor tempo de acompanhamento são os que morreram em hospitais ou outros estabelecimentos de saúde (Figura B.8). Pacientes que vieram a óbito com diagnóstico da causa básica de doenças do aparelho circulatório têm um tempo de acompanhamento maior que o de pacientes com outros tipos de doenças até aproximadamente os 10 anos de acompanhamento, quando ambas as curvas se aproximam (Figura B.9).

Foi feita uma análise, considerando o paradigma de riscos competitivos dadas as diversas, porém únicas, possíveis causas básicas de óbito dos pacientes. Como o foco do atendimento no InCor são pacientes de doenças cardíacas, decidiu-se separar os indivíduos entre pacientes com causa básica do óbito por doenças do aparelho circulatório e pacientes com causa básica do óbito por outras doenças. As Figuras B.10 a B.13 mostram os gráficos de incidência acumulada (IA) em função do tempo de acompanhamento até o óbito para cada risco competitivo. No estudo, a função de incidência acumulada corresponde à probabilidade do tempo de acompanhamento para uma determinada causa ser menor ou igual a um valor de tempo t . Foram realizados testes de Gray (Gray, 1988) para verificar se há diferença estatisticamente significativa entre as funções de incidência acumulada em diferentes grupos (Tabelas A.14 a A.17).

Observa-se que as funções de incidência acumulada de homens e mulheres são bem próximas, tanto para causas do aparelho circulatório como para outras causas. Com base no resultado do teste de Gray (Tabela A.14) as curvas de incidência acumulada de óbitos causados por doenças do aparelho circulatório são

diferentes entre homens e mulheres ($p < 0.01$). O mesmo não é válido para outras causas de óbito ($p = 0.64$).

Os testes de Gray para as curvas de incidência acumulada dos diferentes estados civis (Tabela A.15) indicam diferença significativa para os dois riscos analisados. A Figura B.11 ilustra essas diferenças, de forma que o comportamento das curvas para óbito por causas diferentes daquelas por doenças do aparelho circulatório é parecido com o da Figura B.7, sugerindo um tempo de acompanhamento menor para solteiros.

Em geral, o valor da função de incidência acumulada é menor para pacientes brancos e amarelos (Figura B.12). Há apenas 11 pacientes que se declaram indígenas, de forma que as curvas estimadas para este grupo podem não ser suficientemente representativas. Indivíduos declarados amarelos têm a menor incidência de óbito por causas que não doenças do aparelho circulatório entre 0 a aproximadamente 12 anos de acompanhamento, quando essa incidência aumenta consideravelmente em relação às demais cores. Ademais, as diferenças nas curvas de incidência acumulada entre cores são significantes para ambos os riscos (Tabela A.16).

Nota-se que as curvas de função de incidência acumulada agrupadas por local de ocorrência do óbito (Figura B.13), se comportam de forma diferente dependendo da causa do óbito. Para pacientes com doenças do aparelho circulatório como causa básica de óbito, o valor da função de incidência acumulada é maior para o local “Domicílio” em todos os instantes. Para as outras causas, a incidência é menor neste local. Isso pode ser um reflexo do tipo de atendimento oferecido pelo InCor, que muitas vezes prolonga o tempo de vida de pacientes de doenças circulatórias. Além disso, as diferenças nas curvas de incidência acumulada entre locais são significantes para ambos os riscos (Tabela A.17).

Vale ressaltar que há curvas bem próximas, por exemplo nas Figuras B.10 e B.11, cujos testes levam a valores-p pequenos, indicando diferença estatística. Como a amostra é grande, é possível detectar pequenas diferenças, resta saber se são significativas do ponto de vista médico.

5.3 Análise de correspondência

Também foram realizadas análises de correspondência (Johnson e Wichern, 2007) para estudar a relação entre os diagnósticos dos pacientes em vida e as suas causas de óbito. A Análise de Correspondência é uma abordagem geométrica para visualizar as linhas e colunas de uma tabela de contingência como pontos num plano, de forma que as posições das categorias de linha e coluna sejam consistentes com suas associações na tabela. Desta forma, é possível identificar associações entre categorias que não seriam facilmente notadas sem o auxílio de um recurso visual.

Foram construídas duas tabelas de contingência: causa básica do óbito por todos os diagnósticos em vida (Tabela A.18); demais causas de óbito por todos os diagnósticos em vida (Tabela A.19). Os diagnósticos foram agrupados de acordo com os capítulos do CID-10 e não foram consideradas repetições. Como exemplo, tomando um paciente com diagnósticos C15.1, C15.9, I35.1 e I30.1 em toda sua vida, indica-se um diagnóstico do capítulo 2 e um diagnóstico do capítulo 9. Desta forma, se a sua causa básica do óbito for uma doença do capítulo 9, acrescenta-se uma unidade na casela representada pela linha do capítulo 2 e coluna do capítulo 9 e uma unidade na casela representada pela linha do capítulo 9 e coluna do capítulo 9. Desta forma, cabe frisar que estas tabelas não apresentam contagens de indivíduos, mas sim de diagnósticos. Assim, tem-se que a soma dos valores em cada tabela é maior que o número de pacientes na base de dados.

Uma análise inicial da Tabela A.18 por meio do teste qui-quadrado de independência indica que a causa básica do óbito e os diagnósticos em vida não são independentes ($N = 283.163$, $\chi^2 = 38935$, g.l. = 400, $p < 0.0001$). Este resultado não é suficiente para entender como essas duas variáveis se relacionam. Neste aspecto, a Análise de Correspondência permite compreender melhor essa relação por meio do gráfico *biplot*. O gráfico para a Tabela A.18 pode ser observado na Figura B.14. O primeiro aspecto a ser analisado é o percentual da variabilidade explicada pelas duas dimensões apresentadas no gráfico, que pode ser observado nos eixos coordenados e totaliza 85%. Nota-se também os pontos correspondentes aos capítulos 16 (Afecções originadas no período perinatal) e 17 (Malformações genéticas), com valores altos para a primeira coordenada do gráfico, indicando que o número de

registros que apresentam as doenças nesses capítulos em vida e como causa de óbito é maior do que o esperado sob a condição de independência das variáveis analisadas. O mesmo pode ser afirmado, em menor grau, sobre os pontos referentes ao capítulo 2 (Neoplasmas), que apresentam valores mais baixos no eixo vertical. O aglomerado de pontos próximos à origem do gráfico sugere que as relações entre os demais capítulos não são tão determinantes para a associação das variáveis quanto às discutidas anteriormente. Comportamento bastante semelhante é notado no gráfico *biplot* (Figura B.15) para a Análise de Correspondência para a Tabela A.19, com exceção do que parece ser uma inversão da direção do eixo vertical, mas ainda assim preservando a interpretação dos resultados, mostrando que mesmo considerando apenas as causas de óbito da parte II e das linhas A, B, C e D da declaração de óbito, a relação entre diagnósticos em vida e causa do óbito se mantém semelhante.

6 Conclusões

Com base nos resultados apresentados, percebe-se que a base de dados coletada pode ser analisada sob várias óticas, desde uma perspectiva descritiva até utilizando metodologias mais sofisticadas como a Análise de Sobrevida. Esta última pôde revelar aspectos relevantes sobre o comportamento de diferentes grupos dentro da população de estudo, em termos da mortalidade e do tempo de acompanhamento, bem como fornecer insumos para a reflexão sobre a evolução dos pacientes atendidos pelo hospital. As análises descritivas unidimensionais elencam fatos importantes para compreender o público estudado, permitindo reconhecer previamente fatores que podem influenciar análises futuras, como a distribuição assimétrica da idade. Quanto à Análise de Correspondência, esta se mostrou como mais um recurso para a análise exploratória dos dados, permitindo a identificação de relações nos dados que poderiam passar despercebidas com menor escrutínio.

Considerando o óbito por doenças cardíacas e por outras causas como riscos competitivos, há diferenças nas funções de incidência acumulada para estado civil, raça ou cor e local de óbito. Aparentemente os solteiros, pardos e pretos e óbitos em domicílio, para causa cardíaca, apresentam função de incidência acumulada maior que as outras categorias respectivas. Para o local de óbito, quando se considera outras causas de morte, essa relação se inverte, sendo que a função de incidência

acumulada é menor para os que faleceram em casa. Quanto ao sexo, a diferença ocorre somente para outras causas de óbito que não as cardíacas, sendo que os homens apresentam maior incidência.

A Análise de Correspondência mostrou associação entre os capítulos 16 (Afecções originadas no período perinatal) e 17 (Malformações genéticas) do CID10, ambos como diagnóstico e causa de óbito, o mesmo ocorrendo para o capítulo 2 (Neoplasmas).

Haverá ainda mais possibilidades para a análise do conjunto de dados completo, quando disponível. Algumas sugestões podem ser adotadas, como a expansão das análises com estimadores de Kaplan Meier com a presença de censuras no conjunto de dados. Outra possível abordagem é o estudo de metodologias para dados truncados (Klein e Moeschberger, 2010), que pode se adequar à situação estudada. Havendo quantidade suficiente de pacientes vivos, é possível avaliar técnicas de modelagem preditiva, buscando padrões que possam ajudar a identificar o risco iminente de morte e fazer comparações entre o grupo de pacientes que faleceram e o grupo dos que continuavam vivos ao término do período de coleta dos dados.

APÊNDICE A

Tabelas

Tabela A.1 Estatísticas descritivas para a Idade (anos)

Mínimo	Q1	Média	Mediana	Q3	Máximo	Desvio padrão	Coefficiente de Variação	n	Dados faltantes
0	60.4	69.4	72.2	82.2	111.3	17.6	25%	78,059	0

Tabela A.2 Tabela de frequências de Sexo

Sexo	n	%
Masculino	42,007	53.8
Feminino	36,052	46.2
Total	78,059	100

Tabela A.3 Tabela de frequências de Raça ou cor

Raça ou cor	n	%
Branca	59,095	75.7
Parda	11,556	14.8
Preta	4,828	6.2
Amarela	1,917	2.5
Indígena	11	< 0.1
Sem resposta	652	0.8
Total	78,059	100

Tabela A.4 Tabela de frequências de Município de residência

Município de residência	n	%
São Paulo	46,950	60.1
Osasco	2,748	3.5
Guarulhos	2,541	3.3
Outros municípios de SP	24,910	31.9
Outros estados	910	1.2
Total	78,059	100

Tabela A.5 Tabela de frequências do Número de registros na base

Número de	n	%
-----------	---	---

registros na base		
1	73,806	94.5
2	3,951	5.1
3 ou mais	302	0.4
Total	78,059	100

Tabela A.6 Tabela de frequências das Cirurgias realizadas pelo paciente

Cirurgias	n	%
Cardíaca	8,767	11.2
Cardíaca e outras	4	< 0.1
Outras	2,466	3.3
Nenhuma	66,822	85.6
Total	78,059	100

Tabela A.7 Tabela de frequências do tipo de atendimento

Tipo de atendimento	Número de atendimentos	% de atendimentos
Ambulatorial	22,305	33.9
Pronto Socorro	21,355	32.5
S. Apoio ao Diagnóstico e Terapêutica	19,947	30.3
Internação	4,293	6.5
Total de pacientes	65,747	

OBS: não há registro do tipo de atendimento para 12.312 pacientes

O total de pacientes não é igual à soma do número de atendimentos, pois eles podem ter sido atendidos em mais de um serviço.

Tabela A.8 Tabela de frequências do número de diagnósticos recebidos no InCor

Número de diagnósticos	n	%
0	34,110	43.7
1	16,052	20.6
2	6,805	8.7
3	4,554	5.8
4	3,488	4.5
5 ou mais	13,050	16.7
Total	78,059	100.0

Tabela A.9 Número e percentual de pacientes diagnosticados por capítulo do CID-10 (número de pacientes = 78.059)

Capítulo	Número de pacientes diagnosticados	% de pacientes diagnosticados
9. Doenças do aparelho circulatório	33,282	42.6
18. Sintomas de exames não classificados	16,133	20.7
4. Doenças endócrinas, nutricionais e metabólicas	9,037	11.6
10. Doenças do aparelho respiratório	8,903	11.4
21. Fatores que influenciam o estado de saúde	7,325	9.4
14. Doenças do aparelho geniturinário	5,194	6.7
1. Doenças infecciosas e parasitárias	4,325	5.5
13. Doenças osteomusculares e do tecido conjuntivo	2,723	3.5
2. Neoplasmas	2,547	3.3
19. Lesões, envenenamentos, causas externas	2,545	3.3
11. Doenças do aparelho digestivo	2,487	3.2
5. Transtornos mentais e comportamentais	1,709	2.2
6. Doenças do sistema nervoso	1,373	1.8
17. Malformações genéticas	1,300	1.7
3. Doenças do sangue, hematopoiéticas, imunitárias	908	1.2
12. Doenças da pele e do tecido subcutâneo	770	1.0
8. Doenças do ouvido e da apófise mastoide	444	0.6
20. Causas externas de morbidade e de mortalidade	249	0.3
7. Doenças do olho e anexos	141	0.2
15. Gravidez, parto e puerpério	46	0.1
16. Afecções originadas no período perinatal	41	0.1

OBS: a soma das porcentagens não é 100%, pois os pacientes podem ter mais de um diagnóstico

Tabela A.10 Tabela de frequências do Local de óbito

Local de óbito	n	%
Hospital	64,213	82.3
Domicílio	8,705	11.2
Outros Estab. de Saúde	3,909	5.0
Via Pública	386	0.5
Outros	800	1.0
Sem resposta	46	0.1
Total	78,059	100.0

Tabela A.11 Tabela de frequências da variável Óbito no InCor

Óbito no InCor	n	%
Não	71,995	92.2
Sim	6,064	7.8
Total	78,059	100.0

Tabela A.12 Tabela de frequências da Causa básica do óbito por capítulo do CID-10

Capítulo	n	%
9. Doenças do aparelho circulatório	30,863	39.6
2. Neoplasmas	15,160	19.4
10. Doenças do aparelho respiratório	9,873	12.6
11. Doenças do aparelho digestivo	3,643	4.7
4. Doenças endócrinas, nutricionais e metabólicas	3,529	4.5
1. Doenças infecciosas e parasitárias	3,347	4.3
14. Doenças do aparelho geniturinário	2,822	3.6
20. Causas externas de morbidade e de mortalidade	2,506	3.2
6. Doenças do sistema nervoso	2,039	2.6
18. Sintomas de exames não classificados	1,201	1.5
17. Malformações genéticas	978	1.3
5. Transtornos mentais e comportamentais	763	1.0
13. Doenças osteomusculares e do tecido conjuntivo	530	0.7
3. Doenças do sangue, hematopoiéticas, imunitárias	411	0.5
12. Doenças da pele e do tecido subcutâneo	330	0.4
15. Gravidez, parto e puerpério	41	0.1
16. Afecções originadas no período perinatal	13	< 0.1
8. Doenças do ouvido e da apófise mastoide	9	< 0.1
Zika vírus	1	< 0.1
Total	78,059	100.0

Tabela A.13 Estatísticas descritivas para as variáveis de tempo (anos)

Variável	Mínimo	Q1	Média	Mediana	Q3	Máximo	Desvio Padrão	Coef. Var.	n	Dados faltantes
Tempo entre a primeira e a última admissão	0.0	0.6	4.2	3.0	7.0	17.4	4.0	95%	43,050	35,009
Tempo entre a primeira admissão e o óbito	0.0	1.4	5.5	4.8	8.8	17.5	4.3	78%	64,986	13,073
Tempo entre a última admissão e o óbito	0.0	0.1	2.7	0.9	4.1	16.9	3.5	134%	64,986	13,073

Tabela A.14 Teste de Gray para variável Sexo

	Estimativa	Valor-p	Graus de liberdade
Doenças do aparelho circulatório	9.92	< 0.01	1
Outras doenças	0.22	0.64	1

Tabela A.15 Teste de Gray para variável Estado civil

	Estimativa	Valor-p	Graus de liberdade
Doenças do aparelho circulatório	38.49	< 0.01	4
Outras doenças	336.03	< 0.01	4

Tabela A.16 Teste de Gray para variável Raça ou cor

	Estimativa	Valor-p	Graus de liberdade
Doenças do aparelho circulatório	60.59	< 0.01	4
Outras doenças	17.54	<0.01	4

Tabela A.17 Teste de Gray para variável Local de ocorrência do óbito

	Estimativa	Valor-p	Graus de liberdade
Doenças do aparelho circulatório	1 138.85	< 0.01	4
Outras doenças	1 182.38	< 0.01	4

Tabela A.18 Tabela de contingência de diagnóstico no InCor por causa básica do óbito

Causa básica do óbito (capítulo do CID-10)

Diagnóstico no InCor (capítulo do CID-10)		1	2	3	4	5	6	8	9	10	11	12	13	14	15	16	17	18	20
	1	739	335	18	113	25	60	1	1,843	534	153	22	29	145	3	3	184	42	76
	2	33	1,952	6	33	1	18	0	267	119	46	3	4	24	0	0	4	20	17
	3	43	134	34	21	7	17	0	401	115	56	3	2	25	0	0	19	13	18
	4	361	1,092	36	707	78	141	2	4,369	1,114	375	40	29	345	0	0	13	115	224
	5	76	190	7	54	42	77	1	791	227	91	9	7	51	1	0	4	24	57
	6	60	157	10	57	22	99	0	638	187	47	6	3	36	2	0	14	8	27
	7	10	15	0	8	2	1	0	67	15	6	2	1	1	0	0	8	1	4
	8	23	63	0	17	7	15	0	209	60	15	1	1	16	0	0	2	2	13
	9	1,539	3,946	130	1,519	258	626	5	16,867	3,977	1,360	135	194	1,151	15	1	271	499	789
	10	429	1,289	43	269	68	136	1	3,498	2,066	276	32	74	252	7	1	217	74	171
	11	174	312	4	68	28	55	1	1,044	311	243	15	11	83	0	2	55	24	57
	12	43	65	7	36	12	12	0	383	83	24	24	15	31	0	1	13	8	13
	13	134	349	13	102	34	60	0	1,220	372	136	11	87	85	0	0	4	34	82
	14	317	358	23	313	39	84	0	2,531	660	174	20	28	346	2	2	162	44	91
	15	4	3	0	0	0	0	0	26	4	1	0	2	0	2	0	3	1	0
	16	0	2	1	0	0	0	0	7	5	0	0	0	0	0	3	22	0	1
	17	31	52	8	9	4	18	0	283	54	27	1	2	12	2	8	744	20	25
	18	833	2,169	66	636	163	346	2	7,584	2,031	636	67	77	535	12	3	382	182	409
	19	200	227	9	88	31	50	0	1,261	326	94	18	8	106	0	0	19	33	77
	20	21	17	2	5	4	9	0	152	35	6	2	0	6	0	0	4	5	12
	21	457	1,018	32	272	71	128	0	3,567	896	273	24	38	222	4	1	43	97	182

Tabela A.19 Tabela de contingência de diagnóstico no InCor por demais causas do óbito

Demais causas de óbito (capítulo do CID-10)

Diagnóstico no InCor (capítulo do CID-10)		1	2	3	4	5	7	6	8	9	10	11	12	13	14	15	16	17	18	19	20	21
	1	2,015	392	74	586	114	0	141	1	3,174	1,755	369	75	48	1,144	3	13	207	1,497	653	757	1
	2	714	2,080	108	211	87	2	62	1	757	1,290	184	10	15	298	0	1	7	591	129	187	0
	3	278	160	78	115	27	0	39	1	640	387	119	12	4	208	0	2	22	288	100	126	0
	4	2,687	1,274	141	2,574	244	3	372	5	6,722	3,864	775	146	76	2,185	0	0	21	2,823	663	862	1
	5	501	217	27	303	127	1	140	2	1,151	772	161	27	12	320	1	0	4	521	107	138	1
	6	422	179	23	226	65	1	178	1	942	639	104	21	8	268	2	0	20	413	97	128	1
	7	51	20	2	35	7	2	4	0	95	54	10	6	2	36	0	0	8	44	13	20	0
	8	146	74	7	69	23	0	29	0	323	189	35	5	6	92	0	0	2	131	33	40	0
	9	9,621	4,538	603	6,019	1,039	11	1,486	8	24,790	14,128	2,871	487	372	7,116	16	15	332	10,633	2,784	3,513	3
	10	3,071	1,429	160	1,318	379	2	342	3	5,837	5,083	617	94	137	1,855	7	10	245	2,563	845	1,029	2
	11	843	374	40	381	70	3	112	2	1,754	1,017	474	39	29	573	0	7	70	839	246	318	0
	12	252	83	14	152	25	0	31	0	603	306	57	57	26	216	0	3	15	231	74	94	0
	13	848	395	65	489	91	1	133	0	1,921	1,262	274	43	145	548	0	1	6	796	219	301	0
	14	1,830	442	95	1,023	127	3	206	1	3,985	2,107	383	76	56	2,229	2	15	180	1,750	663	817	1
	15	14	4	0	11	2	0	0	0	39	15	3	1	2	8	2	1	4	20	5	7	0
	16	9	2	1	3	0	0	2	0	13	16	2	0	0	5	0	9	25	12	12	13	0
	17	248	58	51	62	13	0	51	0	571	362	84	8	2	220	2	72	823	355	477	546	0
	18	4,918	2,442	291	2,695	554	1	772	5	11,349	6,884	1,387	250	177	3,434	13	39	420	5,137	1,769	2,140	1
	19	864	270	46	381	83	2	116	1	1,943	1,090	201	46	21	572	0	0	23	837	312	381	1
	20	87	22	5	35	12	0	13	0	214	122	13	2	2	68	0	0	4	102	41	44	0
	21	2,275	1,155	171	1,191	243	2	305	1	5,312	3,128	602	110	78	1,533	4	1	52	2,247	777	1,000	2

APÊNDICE B

Figuras

Observação: nos histogramas a seguir, a linha tracejada indica a média da variável.

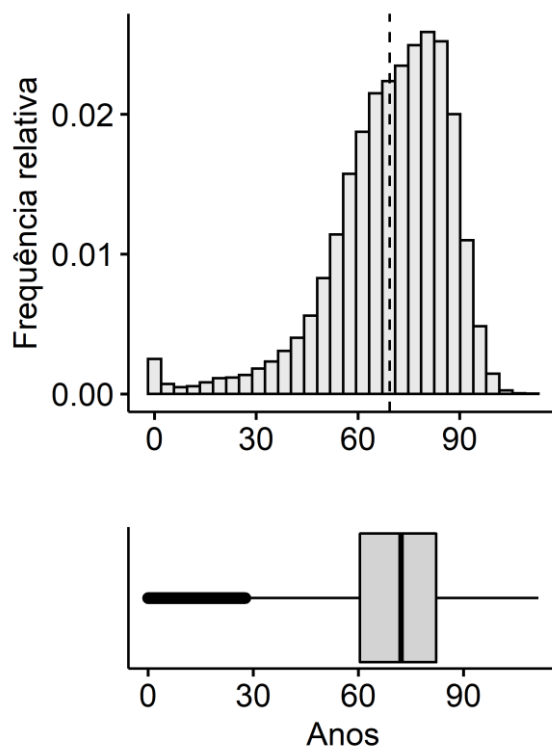


Figura B.1 Histograma e *box plot* da Idade (anos) dos pacientes na data do óbito

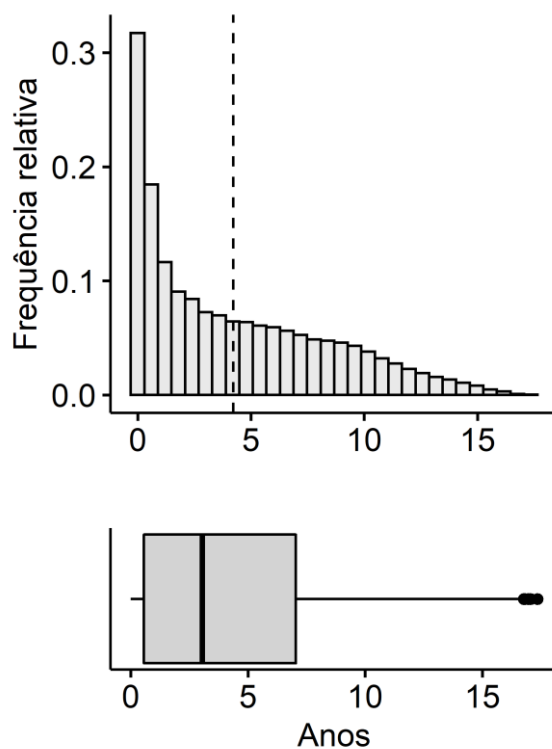


Figura B.2 Histograma e *box plot* do Tempo entre a primeira e a última admissão (anos) - apenas pacientes com pelo menos duas admissões em dias distintos

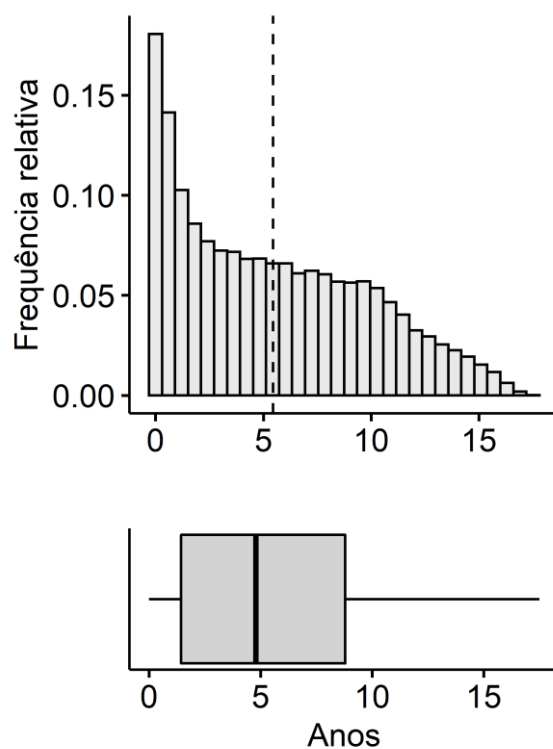


Figura B.3 Histograma e *box plot* do Tempo entre a primeira admissão e o óbito (anos)

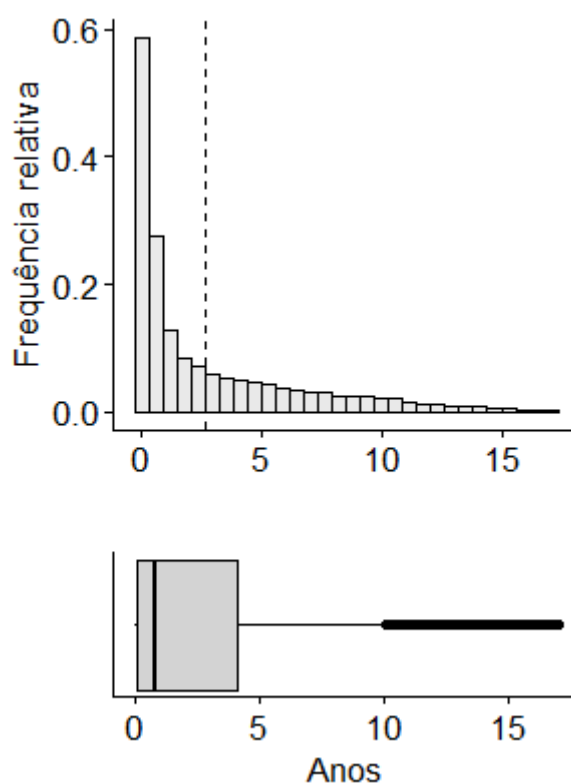


Figura B.4 Histograma e *box plot* do Tempo entre a última admissão e o óbito (anos)

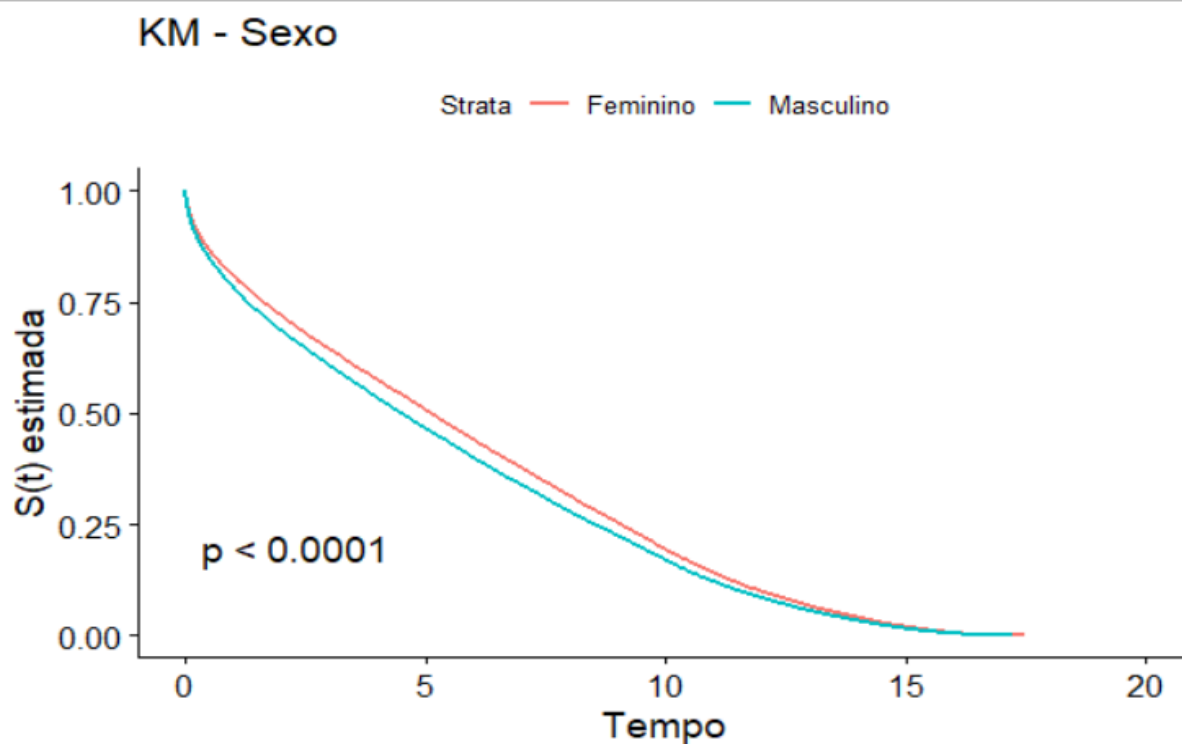


Figura B.5 Gráfico de Kaplan-Meier para variável Sexo

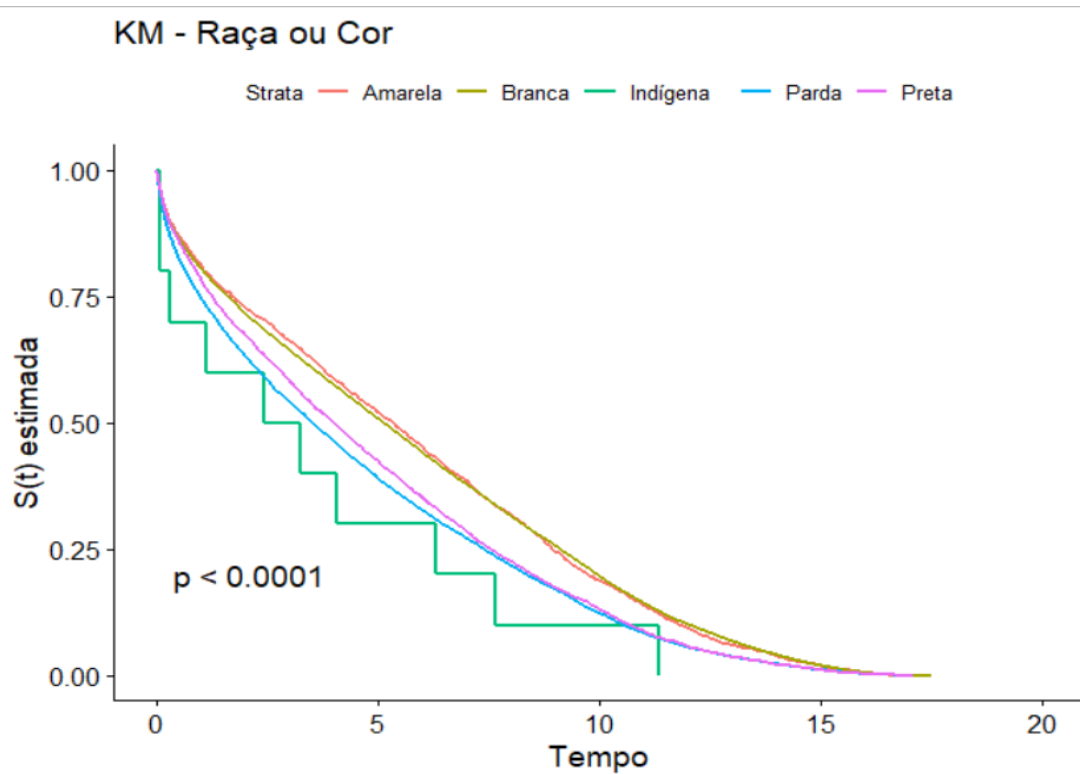


Figura B.6 Gráfico de Kaplan-Meier para variável Raça ou cor

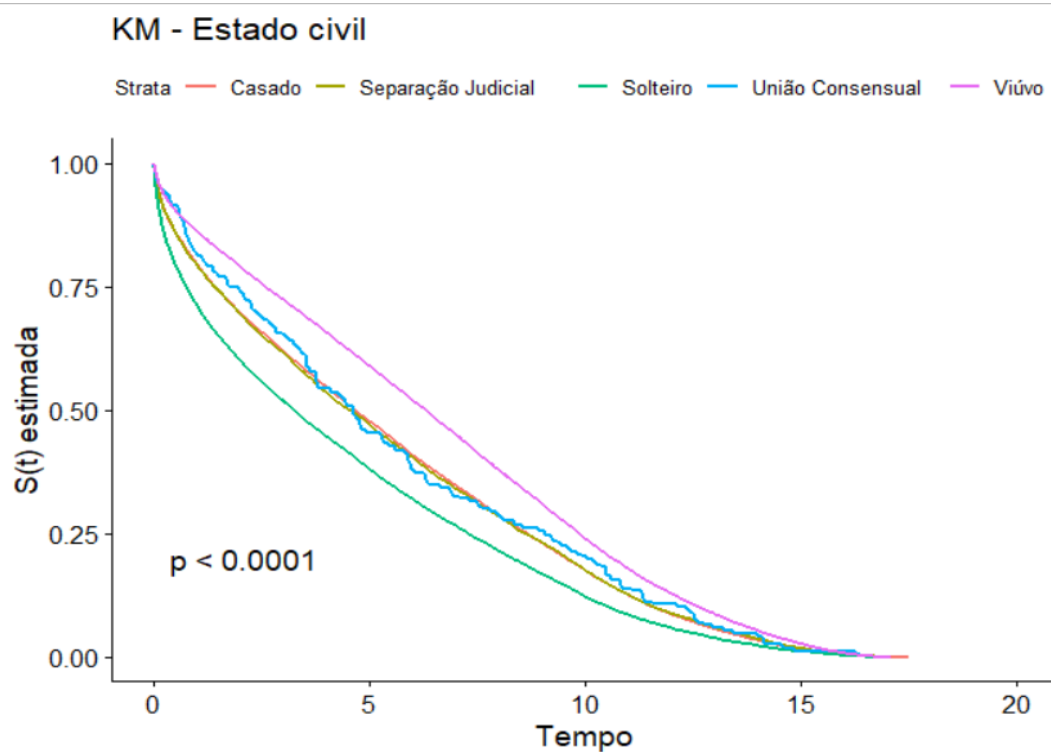


Figura B.7 Gráfico de Kaplan-Meier para variável Estado civil

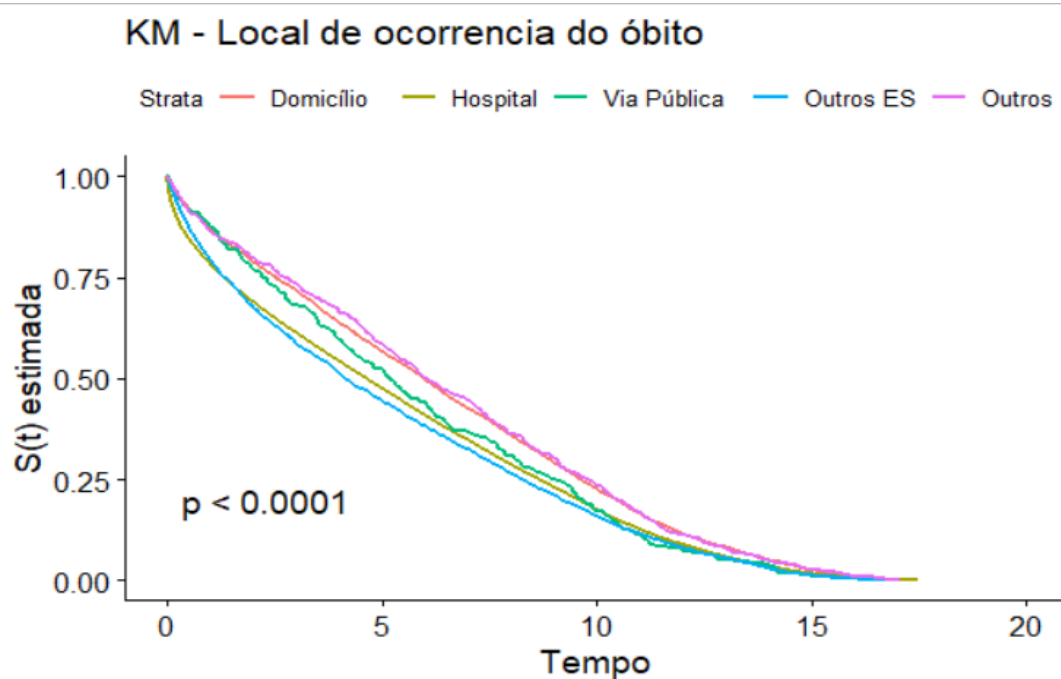


Figura B.8 Gráfico de Kaplan-Meier para o local de ocorrência do óbito

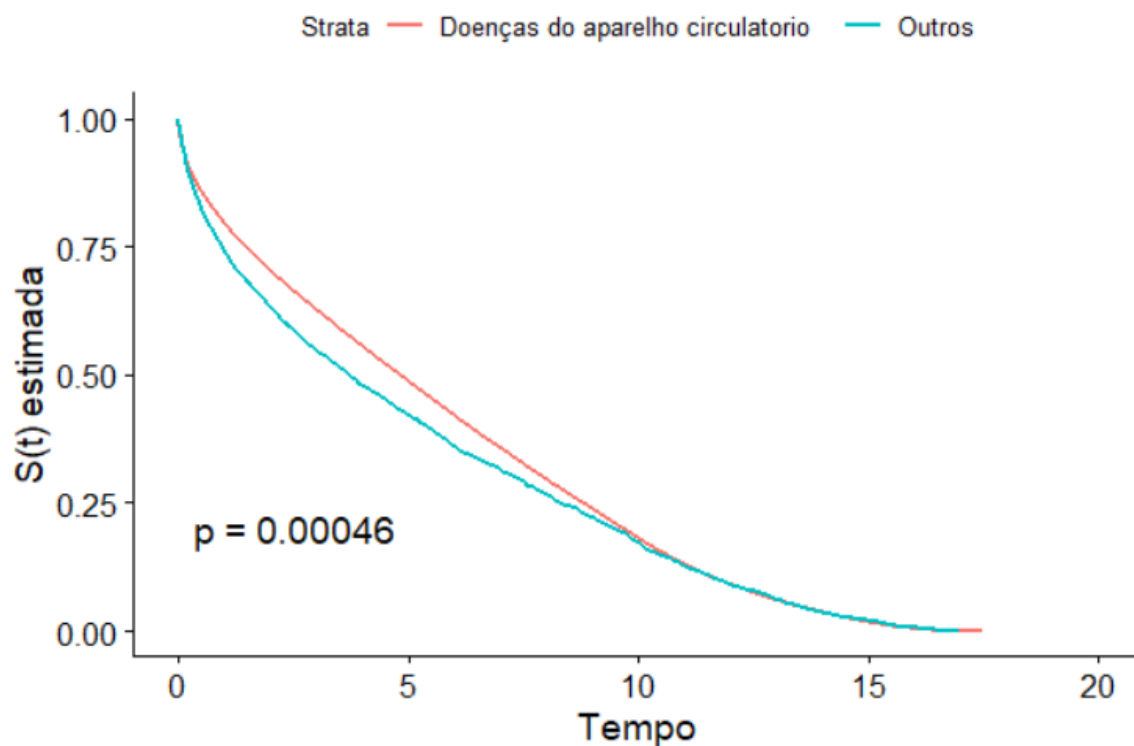


Figura B.9 Gráfico de Kaplan-Meier para diagnóstico da causa básica do óbito

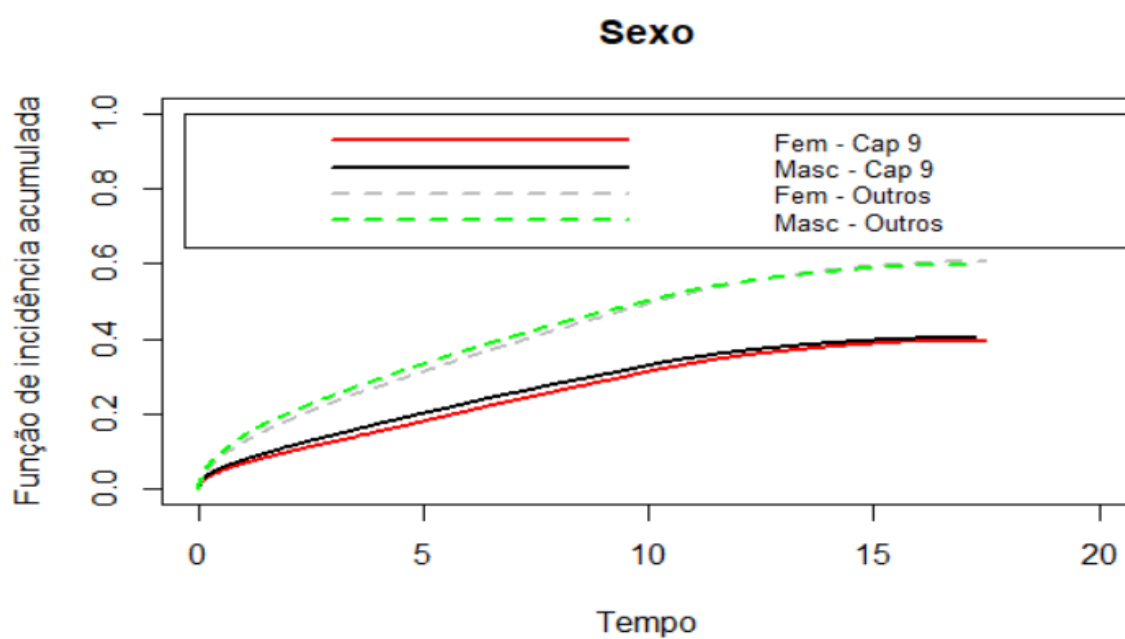


Figura B.10: Gráfico de incidência acumulada para a variável Sexo

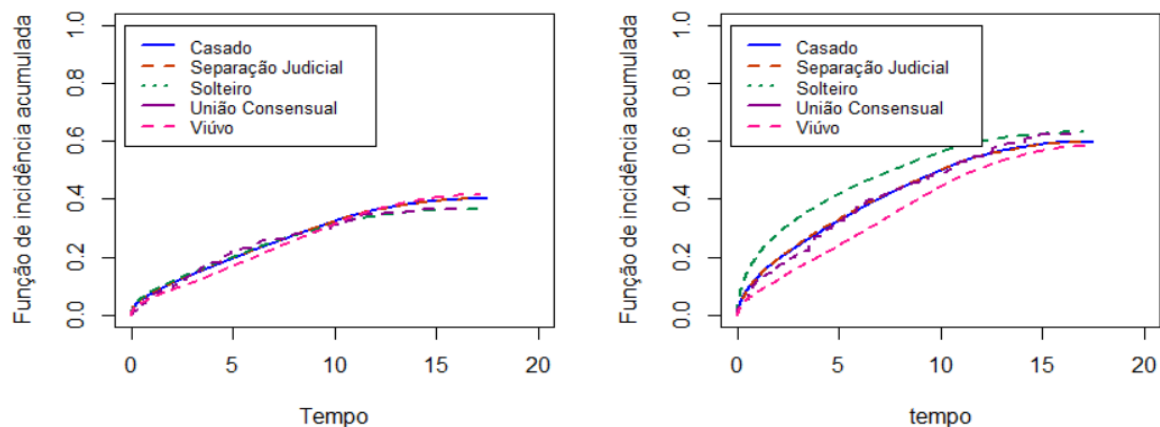


Figura B.11: Gráficos de incidência acumulada para a variável Estado civil. À esquerda para pacientes com doenças do aparelho circulatório como causa básica do óbito e à direita para pacientes com outras doenças.

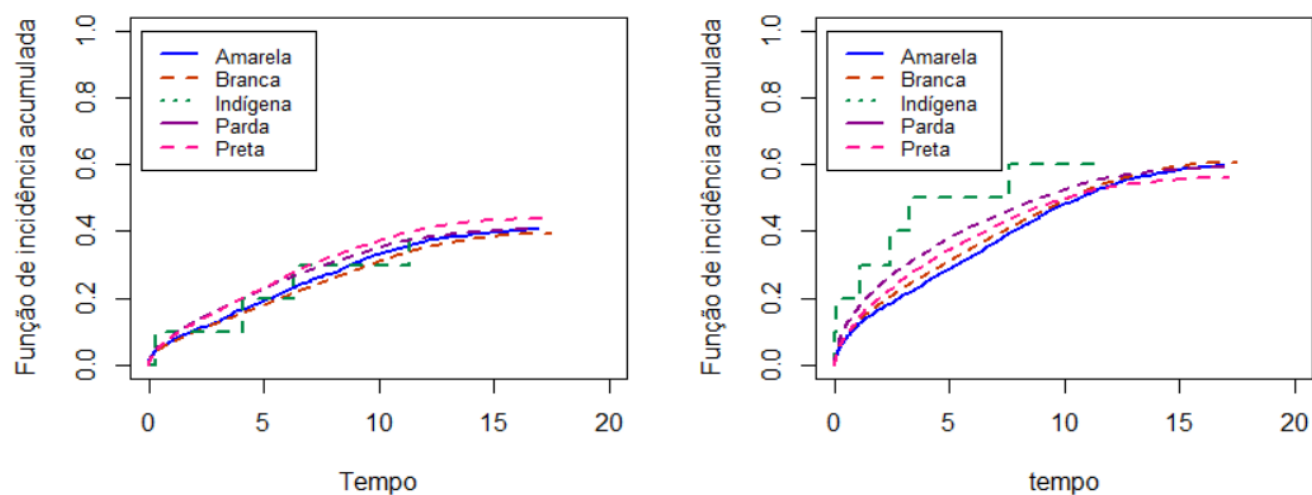


Figura B.12: Gráficos de incidência acumulada para a variável Raça ou Cor. À esquerda para pacientes com doenças do aparelho circulatório como causa básica do óbito e à direita para pacientes com outras doenças.

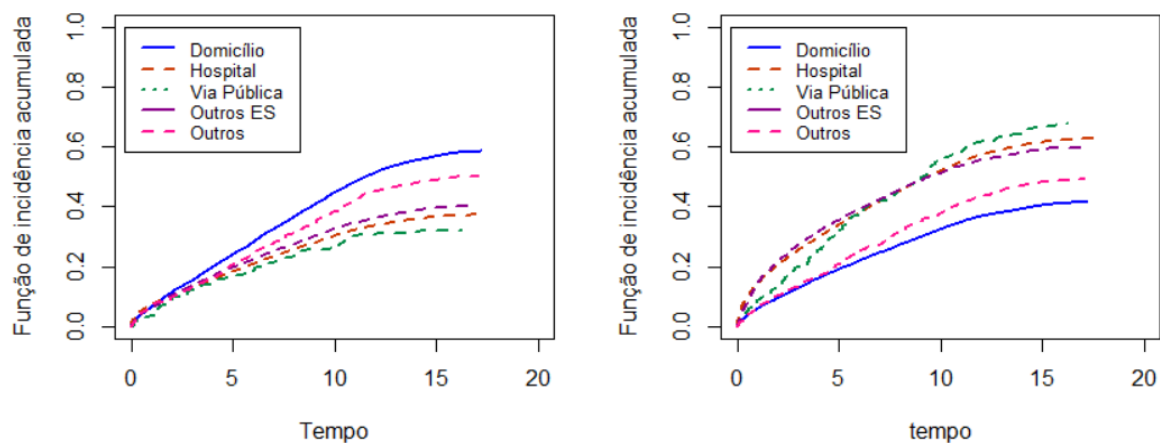


Figura B.13: Gráficos de incidência acumulada para a variável Local de ocorrência do óbito. À esquerda para pacientes com doenças do aparelho circulatório como causa básica do óbito e à direita para pacientes com outras doenças.

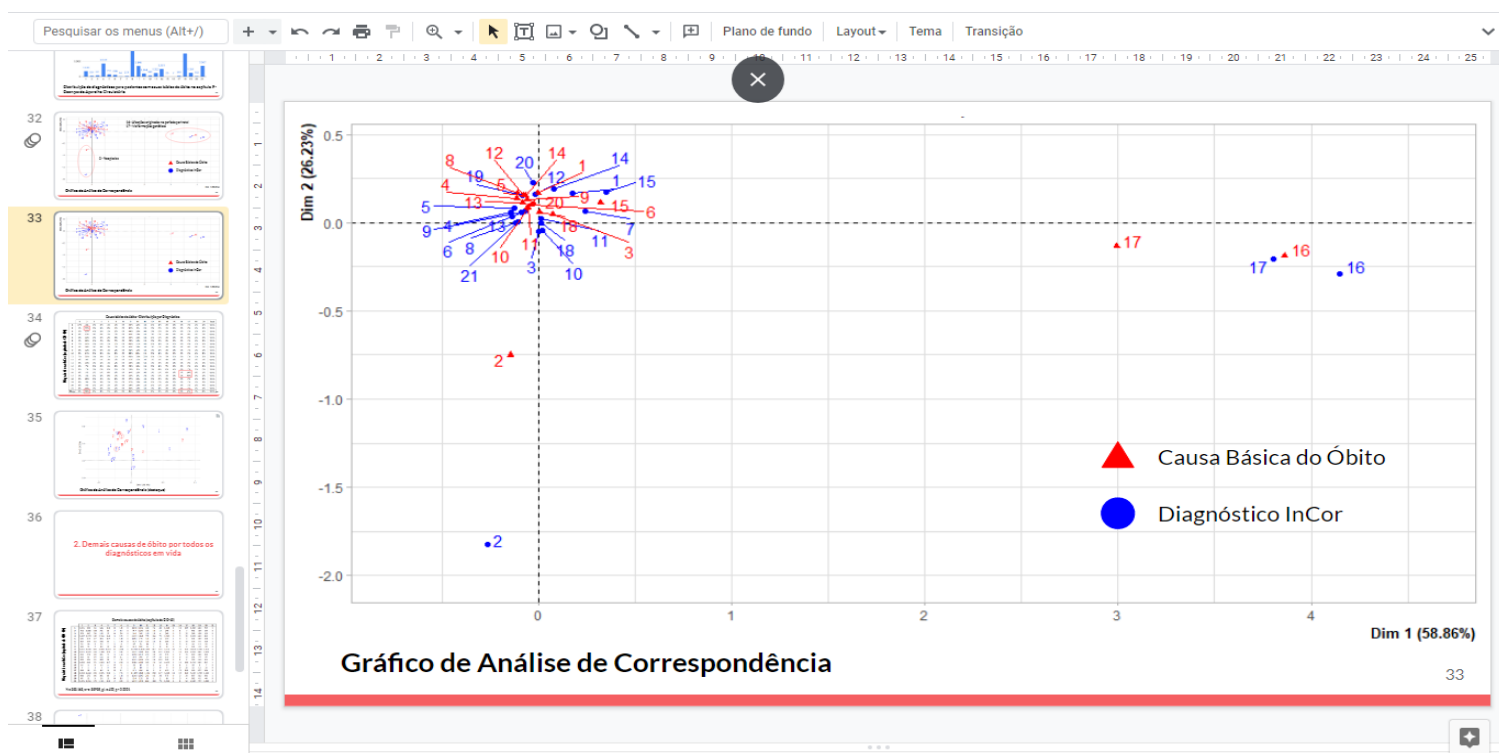


Figura B.14: Gráfico *biplot* para a Análise de Correspondência da Tabela A.17

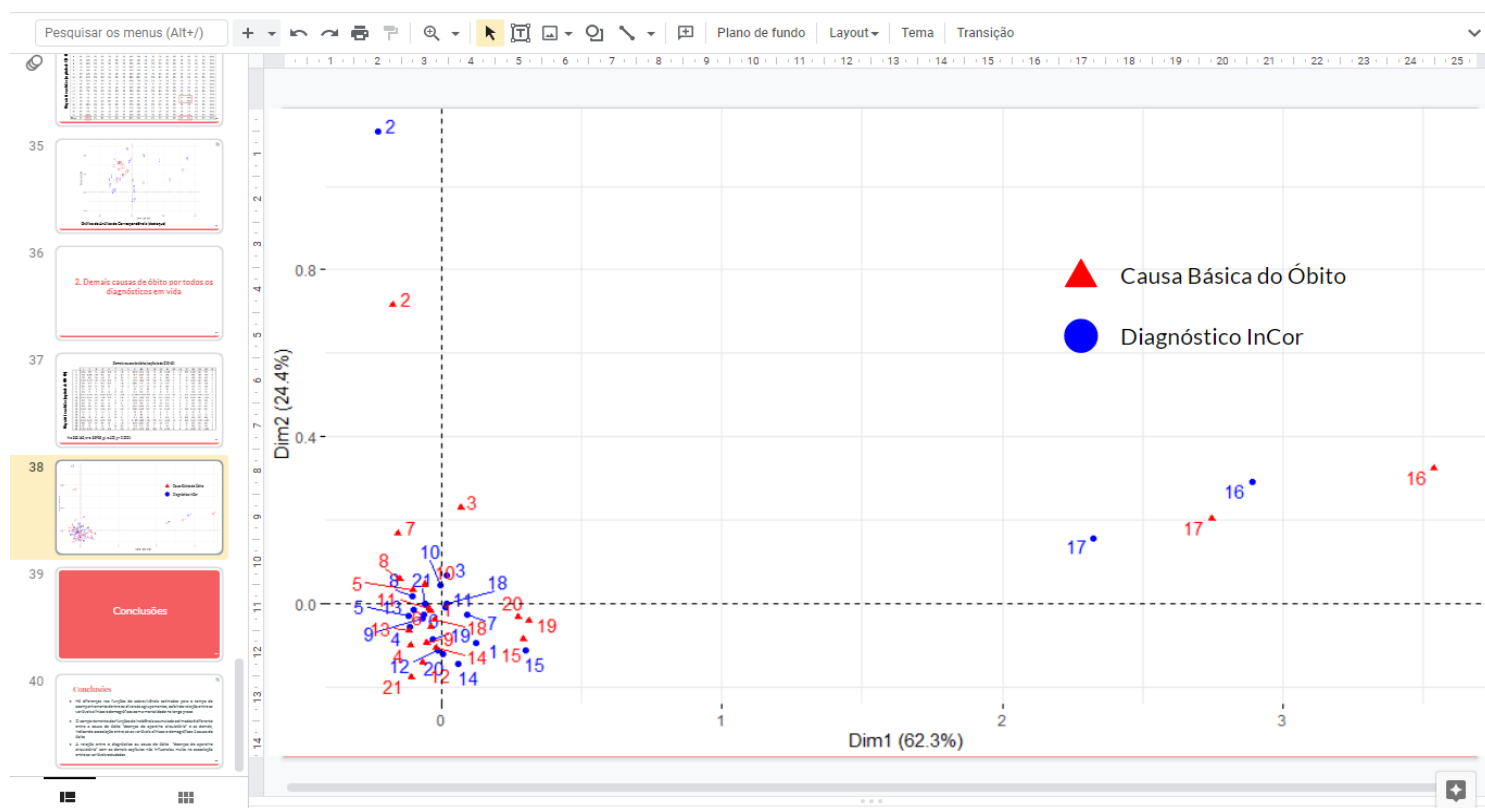


Figura B.15: Gráfico *biplot* para a Análise de Correspondência da Tabela A.18