Universidade Federal de São Carlos

Centro de Ciências Exatas e de tecnologia

Departamento de Estatística

Seminário de Análise de Sobrevivência: Análise de Sinusite em Pacientes com HIV

Henrique Salviano Nº 614440

Renato Trevisol Nº 758562

Samuel Treméa Nº 632090

Thais Cristina Cardozo de Souza Nº 770656

Victor Alves Dogo Martins N° 744878

São Carlos

2021

Sumário

1 Resumo

O conjunto de dados obtido para este trabalho foi coletado na década de 90 com o objetivo de elucidar o desenvolvimento de sinusite em pacientes com HIV. Mais especificamente, procurou-se fatores que pudessem influenciar na chance de um indivíduo desenvolver a doença, como uso de drogas, idade e estágio da infecção por HIV.

De início, notamos algumas peculiaridades dos dados: devido à época em que foram coletados, algumas dinâmicas tiveram um aparente envelhecimento frente às condições atuais de quem convive com o vírus do HIV.

Começamos por estimar as taxas de sobrevivência ao evento de interesse (infecção de sinusite) através do Estimador de Kaplan-Meier. Percebemos um aumento na chance de diagnóstico de sinusite quanto o indivíduo atingia a marca de 600 dias após o diagnóstico com HIV, por exemplo.

Após isto, foram feitas curvas segundo o mesmo estimador para compararmos as diversas covariáveis presentes no estudo. Algumas aparentaram maior influência no tempo de sobrevivência (como a contagem de leucócitos) e outras, nem tanto. Interessante ressaltarmos que pessoas que usaram drogas injetáveis e pessoas que usaram cocaína possuíram comportamentos praticamente idênticos.

Após a análise não-paramétrica, partimos para as análises Paramétrica e dos Modelos de Cox. A intenção foi buscar um modelo confiável o bastante que pudesse predizer em quanto tempo determinado indivíduo desenvolveria sinusite dentro do contexto da década de 90 com base no conjunto de dados.

A Análise Paramétrica nos trouxe a informação de que o Modelo Log-Logístico levando em conta contagem dos leucócitos CD4 e CD8 mais a idade do indivíduo seria melhor ajustado do que o Modelo Gompertz apresentado anteriormente.

Após isto, foi feito o ajuste dos dados seguindo três modelos: um Modelo de Cox Semi-Paramétrico, o Modelo de Cox-Gompertz e o Modelo de Cox Log-Logístico. As comparações foram feitas considerando todas as variáveis do modelo e, enquanto que algumas delas não possuíram significância satisfatória, os modelos no geral pareceram se sair melhor do que os ajustes paramétricos obtidos.

Por fim, temos que o trabalho por si só foi um exercício interessante dado o banco de dados com evento de interesse atípico. A temática tem aplicação contemporânea se levarmos em conta uma possível nova coleta dos dados e atualização das variáveis a serem coletadas.

2 Introdução

2.1 Descrição do Banco de Dados

O banco de dados fornecido para este trabalho foi obtido através do livro **Análise de Sobrevivência Aplicada** de Enrico Antônio Colosimo e Suely Ruiz Giolo.

É dito que o banco de dados foi construído com o intuito de identificar a ocorrência de manifestações otorrinolaringológicas em pacientes HIV positivos. Mais especificamente, a intenção das pesquisas seria compreender se a infecção pelo HIV aumenta a incidência de sinusite e se a incidência e gravidade da doença cresce de acordo com a progressão da imunodeficiência.

Foram coletados dados entre março de 1993 até fevereiro de 1995 através de consultas trimestrais, com frequência mediana de 4 consultas. Ao contrário da maior parte das vezes em que estudamos situações parecidas na disciplina de Análise de Sobrevivência, aqui temos que o evento de interesse é a ocorrência de sinusite e não a morte do paciente. Com isso, o tempo presente no banco de dados indica se o paciente contraiu sinusite ou se foi censurado através do estudo.

Para este trabalho, fizemos uma limpeza do banco de dados devido à presença de diversas variáveis faltantes que poderiam prejudicar nossas conclusões. Com isso, tivemos em mãos para este trabalho **52 observações** de pacientes HIV positivos ou negativos com dados coletados de acordo com diversas variáveis. O banco de dados limpo utilizado encontra-se nas referências deste trabalho.

A descrição completa das variáveis consta na seguinte tabela:

Variável	Descrição
ID	Número de Identificação do Paciente
Idade	Medida em anos no início da pesquisa
Gênero	1 - Masculino e 0 - Feminino
Grupos de Risco	1 - Paciente HIV Soronegativo; 2 - Paciente HIV Soropositivo Assintomatico; 3 - Paciente com ARC; 4 - Paciente com AIDS.
Tempo	Tempo entre o início da pesquisa e o desenvolvimento de sinusite; medida em dias
Censura	1 para falha e 0 para censura
CD4	Contagem do Linfócito CD4
CD8	Contagem do Linfócito CD8
Orientação Sexual	1 - Homossexual; 2 - Bissexual; 3 - Heterossexual;
Uso de Drogas Injetáveis	1 - Não, 2 - Sim
Uso de Cocaína por Aspiração	1 - Não, 2 - Sim

Tabela 1: Descrição do Banco de Dados de Sinusite em Pacientes com HIV após limpeza

2.2 Objetivo

Com as variáveis apresentadas, o objetivo deste trabalho é compreender se há relação entre o tempo do desenvolvimento de sinusite entre pacientes HIV positivos e negativos e as variáveis presentes no banco de dados em estudo.

Para isso, aplicaremos alguns conhecimentos apresentados na disciplina **Análise de Sobrevivência** ministrada pela Profa. Dra. Vera Lucia Damasceno Tomazella, entre eles: análise paramétrica do tempo de sobrevivência, análise não-paramétrica e ajuste à Modelos de Cox. Nesta última parte, inclusive, é de interesse testarmos se é possível predizer o tempo

de desenvolvimento de sinusite a partir dos dados que possuímos em mão, tudo a partir do momento do diagnóstico de HIV.

3 Aplicação

3.1 Análise Não-Paramétrica

3.1.1 Tabela das Estimativas de Kaplan-Meier

Antes de realizarmos as curvas de sobrevivência utilizando o estimador de Kaplan-Meier, dispomos a seguir as estimativas numa tabela de fácil compreensão.

t_j	n_j	d_j	n_{cens}	$\hat{S}(t)$	$\sqrt{\hat{V}(\hat{S}(t))}$	$\hat{H}(t)$	$\sqrt{\hat{V}(\hat{H}(t))}$	IC(95%)
0	52	0	0	1	0	0	0	[1, 1]
75	48	0	4	1	0	0	0	[1, 1]
150	46	2	0	0.958	0.028	0.042	0.029	[0.903, 1]
225	45	0	1	0.958	0.028	0.042	0.029	[0.903, 1]
300	31	1	13	0.93	0.038	0.07	0.041	[0.857, 1]
375	24	1	6	0.898	0.049	0.105	0.053	[0.807, 1]
450	21	1	2	0.857	0.061	0.15	0.07	[0.745, 0.987]
525	17	2	3	0.767	0.081	0.258	0.103	[0.623, 0.945]
600	9	2	5	0.665	0.097	0.396	0.142	[0.498, 0.887]
675	2	2	5	0.394	0.177	0.841	0.379	[0.163, 0.951]

Tabela 2: Tabela de Estimativas segundo estimador de Kaplan-Meier

É interessante evidenciarmos que, apesar de não termos uma fácil visualização destes conceitos apenas através da tabela, existe uma queda na chance de desenvolvimento de sinusite estável até a última linha.

É possível interpretarmos que, após pouco menos de dois anos, a chance de alguém com HIV soropositivo ou negativo desenvolver sinusite dado as condições do estudo chega a mais de 60%. A estimativa da função de risco acumulado, inclusive, nos diz que essa chance pode chegar a ser maior. Se compararmos com outras faixas de tempo, três meses antes disso a chance do desenvolvimento de sinusite é algo próximo de 40%; na linha anterior, 25%.

No entanto, vale ressaltar que o Intervalo de Confiança desta última linha apresenta uma amplitude atípica dada a progressão das linhas anteriores. É de interesse nosso que outras

análises sejam feitas para verificarmos estas questões.

3.1.2 Curvas de Kaplan-Meier

Para compreendermos o comportamento das variáveis, convém plotarmos curvas de sobrevivência. Faremos isso para todo o conjunto de dados e para cada covariável na intenção de entender a diferença entre elas.

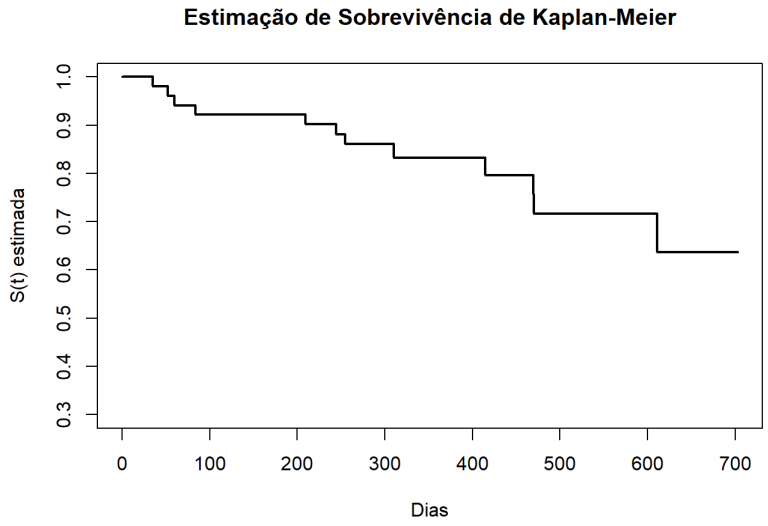

Figura 1: Curva de Kaplan-Meier para dados completos

No geral, segundo o gráfico acima, a chance de desenvolvimento de sinusite não chega a ser maior do que 30% ao longo do tempo. É possível que, dos estratos presentes no nosso banco de dados, algum deles diferencie-se de forma que justifique os resultados obtidos na tabela anteriormente.

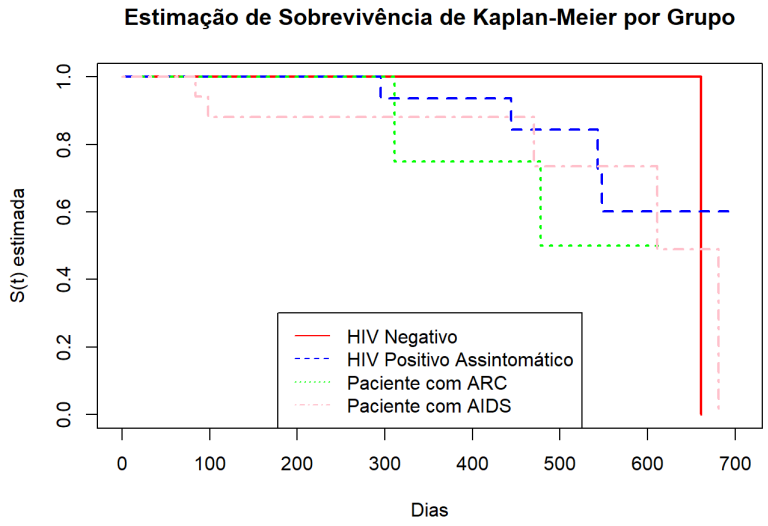

Figura 2: Curvas de Kaplan-Meier divididas por grupo de risco

Na medida em que a gravidade do grupo de risco avança, podemos perceber que a chance de desenvolver sinusite aumenta, ainda que seja relativamente pouco. Se compararmos os pontos quando estamos em 600 dias, pacientes sem HIV possuem chance nula de desenvolver sinusite, enquanto que os outros grupos de risco possuem chances similares, próximas de 40%.

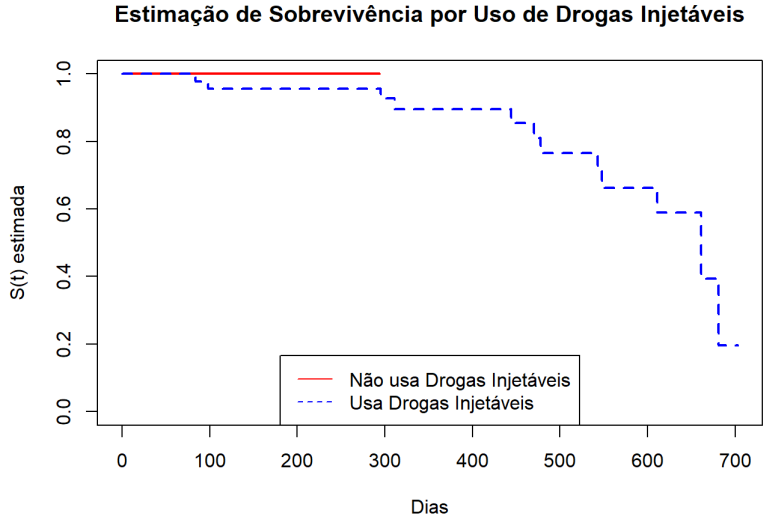

Figura 3: Curvas de Kaplan-Meier divididas por uso de drogas injetáveis

É possível que pela falta de dados, a curva do estrato que não utiliza drogas injetáveis tenha ficado estranha como é possível observar. No entanto, fica nítida a diferença se

compararmos com a curva dos que utilizam drogas injetáveis. O uso de drogas injetáveis, conhecido como contribuinte para o aumento da transmissão de HIV, pode ter uma influência semelhante ao estágio da AIDS ou até mesmo ao uso de cocaína, como veremos a seguir.

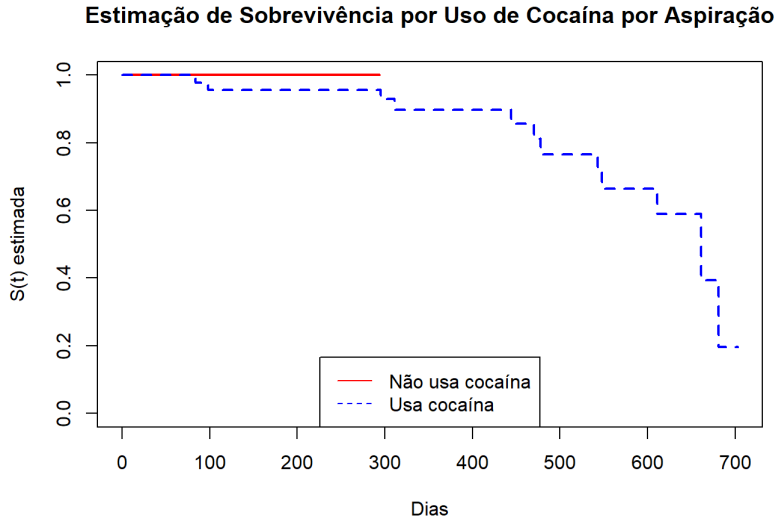

Figura 4: Curvas de Kaplan-Meier divididas por uso de cocaína

É curioso notarmos como as curvas relativas ao uso de drogas injetáveis e ao uso de cocaína são semelhantes, praticamente iguais. É de se pensar que sejam duas variáveis com alta correlação (na prática, a pessoa que usa drogas injetáveis também usa cocaína e vice-versa).

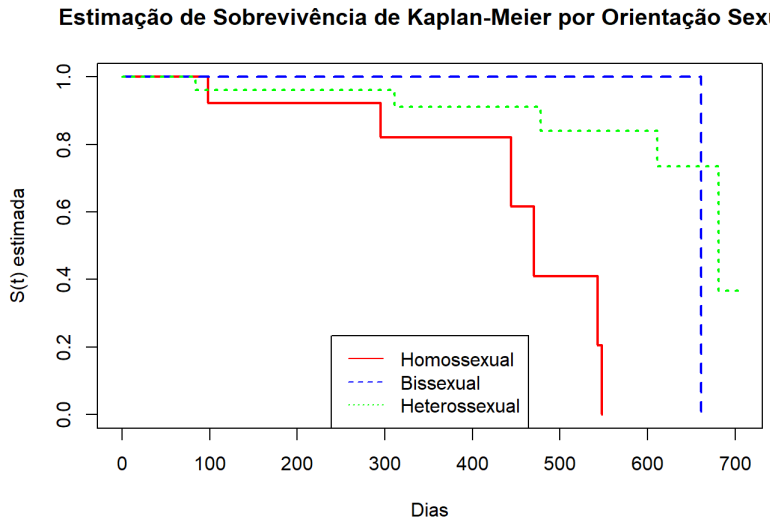

Figura 5: Curvas de Kaplan-Meier divididas por Orientação Sexual

Como dito anteriormente, o banco de dados foi obtido no meio da década de 90. Diferente dos dias de hoje, onde ainda existe um grande estigma mas o controle e tratamento de pessoas com HIV/AIDS é maior e mais acessível, na época de coleta dos dados a população homoafetiva era tida como grupo marginalizado no que diz respeito à estruturas básicas de saúde.

Podemos enxergar que as curvas de pessoas hetero e bissexuais são próximas, enquanto que as de pessoas homossexuais descreve uma queda mais acentuada, chegando a praticamente 100% de chance de desenvolver sinusite em decorrência da infecção por HIV.

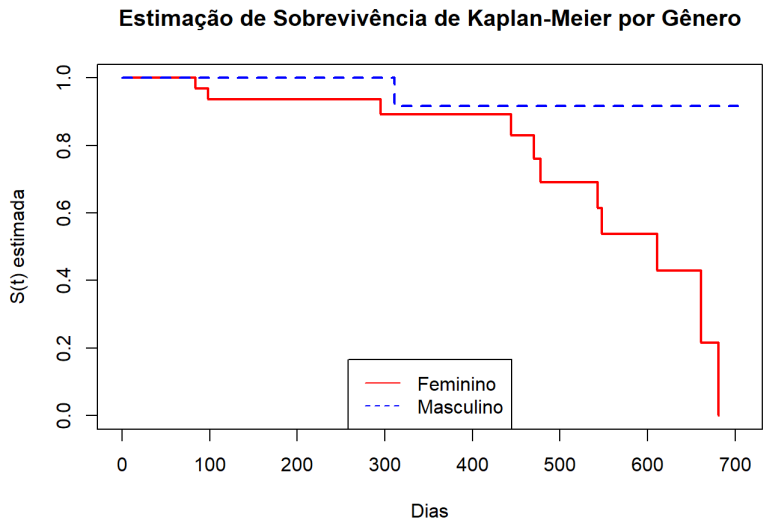

Figura 6: Curvas de Kaplan-Meier divididas por Gênero

Curioso levantarmos que a diferença da taxa de desenvolvimento de sinusite por gêneros é tão distoante. No entanto, essa realidade independe da infecção por HIV ou não: é tido como consenso na comunidade médica que mulheres são mais acometidas pela doença do que homens, no geral e por mais que exista a possibilidade de homens serem mais acometidos por outras condições devido a falta da procura por tratamento.

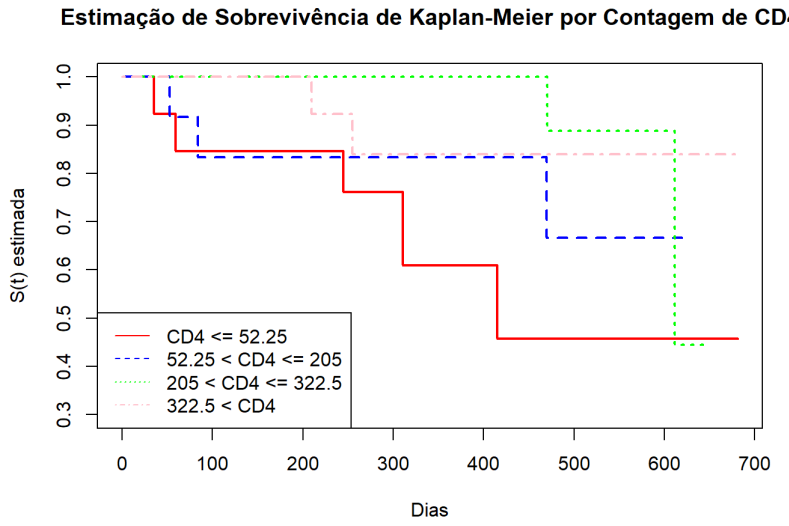

Figura 7: Curvas de Kaplan-Meier divididas por Contagem de CD4

Leucócitos CD4 são tidos como indicativo do estágio de infecção de HIV: quanto menor a contagem, mais avançado é o estágio. Conceitualmente, portanto, esta covariável estaria ligada ao estágio de infecção de HIV mostrado anteriormente.

Não parece haver uma relação clara do tempo de desenvolvimento de sinusite: as curvas se misturam sem ordem ou critério aparente, possivelmente indicando que não haja razão para levarmos a contagem de CD4 em consideração em detrimento de outras variáveis.

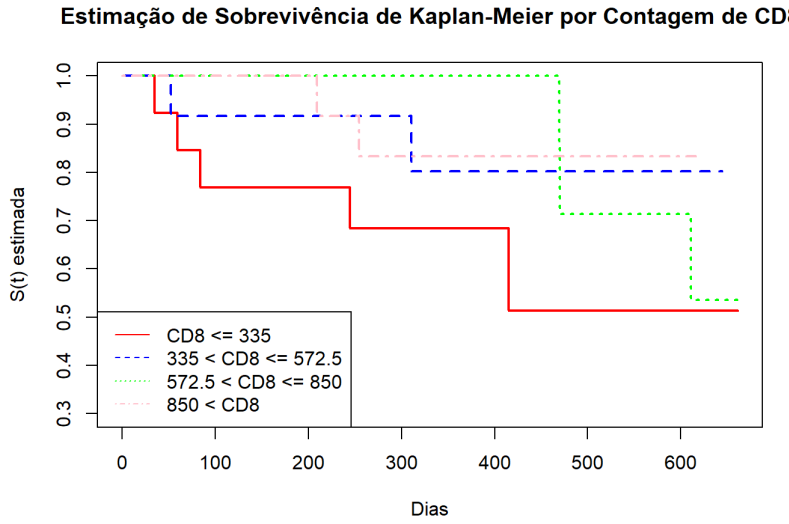

Figura 8: Curvas de Kaplan-Meier divididas por Contagem de CD8

A contagem de CD8, semelhante aos leucócitos CD4, é um indicativo do progresso da

infecção por HIV. Como vimos no gráfico anterior à este, não parece haver uma relação muito clara da diminuição da contagem com o tempo de desenvolvimento de sinusite.

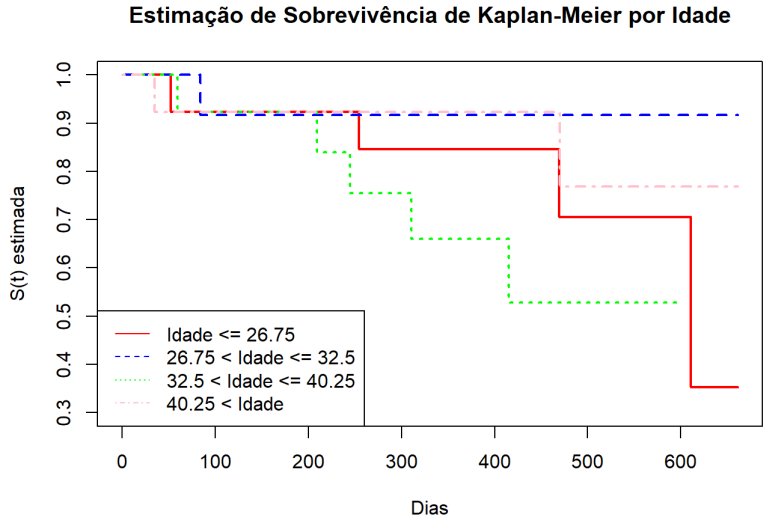

Figura 9: Curvas de Kaplan-Meier divididas por Idade

Por fim, temos a última covariável presente em nosso banco de dados. Não parece haver relação clara entre o tempo de desenvolvimento de sinusite e idade do paciente: a curva de pacientes com menos de 26 anos encontra-se acima da curva de pacientes entre 32 e 40 anos, mas abaixo de pacientes com mais de 40 anos, indicando que essa covariável não influencie no nosso evento de interesse.

A seguir, para termos conclusões mais sólidas, realizaremos Testes Log-Rank para cada covariável.

3.1.3 Teste Log-Rank para Comparação de Grupos

As conclusões acerca dos estimadores de Kaplan-Meier segundo os testes log-rank serão dispostos de acordo com cada estrato utilizado para comparação. Nossa intenção foi encontrarmos p-valores significativos mostrando a diferença real entre os estratos segundo o estimador de Kaplan-Meier.

Os resultados obtidos foram o seguinte:

Estrato	**p-valor**
Grupo de Risco	0,8
Uso de Drogas Injetáveis	0,7
Uso de Cocaína	0,8
Orientação Sexual	0,001
Gênero	0,03
CD4	>0,0001
CD8	>0,0001
Idade	0,009

Tabela 3: Resultado dos testes log-rank para cada estrato das covariáveis

Temos que, das covariáveis presentes no estudo, as que apresentam diferença significativa nos tempos de sobrevivência se analisadas lado a lado com um nível de significância de 0.05 são Orientação Sexual, Gênero, Contagem de CD4, Contagem de CD8 e Idade. No entanto, no caso do Gênero, vale ressaltarmos que seu p-valor obtido pode ser dado como insatisfatório devido à sua proximidade com o nível de significância.

Com os resultados acima, podemos encontrar modelagens mais críveis para nossos dados de forma que seja possível predizer certos comportamentos da forma mais fidedigna possível.

3.2 Análise Paramétrica

3.2.1 TTT-Plot

Para nos auxiliar na escolha de qual modelo é mais adequado para realizar uma análise dos dados em questão, utilizamos o método gráfico conhecido como *TTT-Plot*. Assim, por meio de tal técnica podemos ter indícios da forma da função de risco do problema em questão e assim utilizar tal fato como indicador de qual modelo é adequado ou não para ser utilizado.

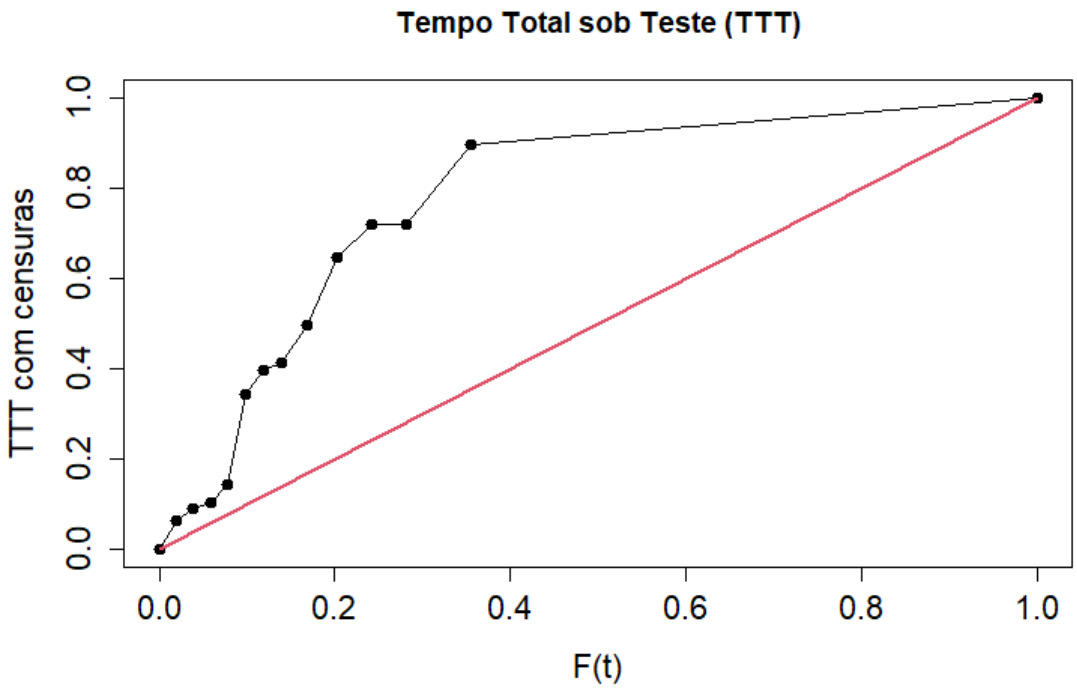

Figura 10: Gráfico do tempo total sob teste até a ocorrência de sinusite.

Por meio da Figura 10, podemos observar que a forma da função de risco se caracteriza por sua característica crescente, assim temos evidências de que a o possível modelo adequado para o problema em questão terá que possuir uma função de risco crescente.

3.2.2 Ajuste para a distribuição Gompertz

Uma distribuição que possui função risco com comportamento crescente, de tal forma que seja plausível realizarmos um ajuste levando ela em conta, é a distribuição Gompertz.

Ela é caracterizada, primeiramente, pela sua função densidade de probabilidade, definida como:

$$f(t|\eta, b) = b \cdot \eta \cdot e^{\eta + bt - \eta e^{bt}}$$

Onde η é o parâmetro de forma e b o parâmetro escala. A função de distribuição acumulada de probabilidade, por sua vez, é dada por:

$$F(t|\eta, b) = 1 - e^{-\eta \cdot (e^{bt} - 1)}$$

Como aprendemos em sala, é intuitivo obtermos a função de sobrevivência através da relação $S(t) = 1 - F(t)$:

$$S(t|\eta, b) = e^{-\eta \cdot (e^{bt} - 1)}$$

Por consequência, obtemos $h(t) = \frac{f(t)}{S(t)}$, neste caso dado por:

$$\begin{aligned} h(t|\eta, b) &= \frac{b \cdot \eta \cdot e^{\eta + bt - \eta e^{bt}}}{e^{-\eta \cdot (e^{bt} - 1)}} \\ &= b \cdot \eta \cdot exp\left[\eta + bt - \eta e^{bt} + \eta e^{bt} - \eta\right] \\ &= b \cdot \eta \cdot e^{bt} \end{aligned}$$

Logo, realizamos o Teste de Wald para um modelo com todas as covariáveis e analisamos quais delas foram significativas. No caso da distribuição Gompertz, obtivemos o resultado atípico de que apenas a idade seria relevante para nosso estudo, então ajustamos um modelo considerando a distribuição Gompertz utilizando a função `flexsurvreg` do pacote `flexsurv` do software estatístico R. As estimativas para os parâmetros de forma e escala foram $\eta = 0,00112$ e $b = -6,81504$. Os demais resultados obtidos foram:

Covariáveis	Coeficientes
Idade	β_1 = -0.02691
Log-Verossimilhança	-100.561
AIC	207.122

Para verificarmos se o modelo é adequado, convém plotarmos a curva de sobrevivência da distribuição Gompertz sobreposta à curva de sobrevivência estimada dos dados, como

podemos ver a seguir:

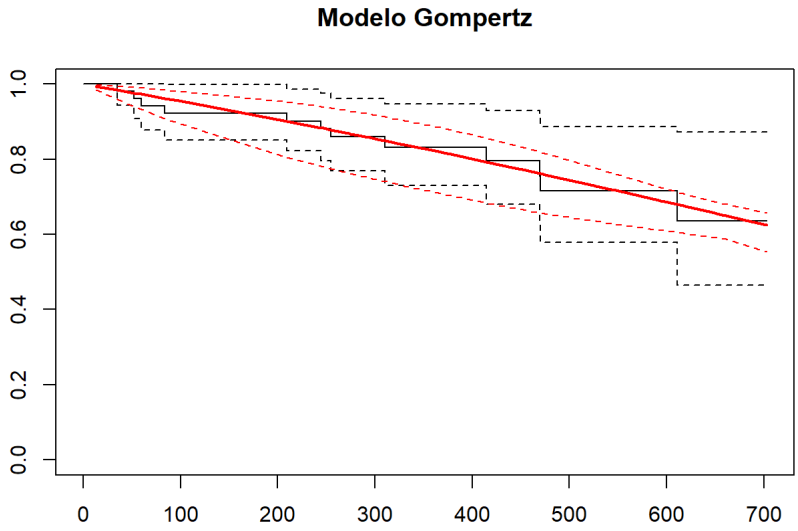

Figura 11: Ajuste via Distribuição Gompertz

É possível enxergarmos na Figura 11 que a curva da distribuição Gompertz se ajusta relativamente bem às estimativas de nossos dados, indicando que o Modelo apresentado seja adequado. Os limites do Intervalo de Confiança (representados numa linha tracejada) não se comportam tão bem, mas acompanham o traçado das estimativas de sobrevivência de alguma forma.

Para fins de comparação, no entanto, é interessante que nós realizemos um ajuste levando em consideração outra distribuição. O fato de termos levado em conta apenas uma covariável devido aos resultados do teste de wald nos faz suspeitar do modelo.

3.2.3 Ajuste para a distribuição Log-Logística

Uma outra distribuição que possui uma função risco com comportamento crescente é a Distribuição Log-Logística. Tal distribuição apresenta função densidade de probabilidade dada por

$$f(t \mid \alpha,\ \beta) = \frac{\frac{\alpha}{\beta}(\frac{t}{\beta}^{\alpha-1})}{(1+(\frac{t}{\beta})^{\alpha})^2}$$

onde α é o parâmetro de forma e β é o parâmetro de escala. Também temos que a função distribuição acumulada de probabilidade

$$F(t \mid \alpha,\ \beta) = 1 - \frac{1}{(1 + (\frac{t}{\beta})^{\alpha})}.$$

Assim, podemos obter a função de sobrevivência $S(t)$ por meio da relação $S(t) = 1 - F(t)$, logo temos que

$$S(t \mid \alpha,\ \beta) = \frac{1}{(1 + (\frac{t}{\beta})^{\alpha})}$$

e, consequentemente, obtemos $h(t) = \frac{f(t)}{S(t)}$ que é dado por

$$h(t \mid \alpha,\ \beta) = \frac{\frac{\alpha}{\beta}(\frac{t}{\beta})^{\alpha-1}}{(1 + (\frac{t}{\beta})^{\alpha})}.$$

Inicialmente, antes de ajustar o modelo considerando todas as cováriaveis e observamos o resultado do Teste Wald para verificar a significância das mesmas. Assim, ao nível de significância de 5% verificou-se que apenas as variáveis Idade, CD4 e CD8 são significantes para a adequação do modelo log-logístico. Portanto, considerando a distribuição Log-Logística podemos obter o modelo ajustado por meio da função `flexsurvreg` do pacote `flexsurv` do `R`. Assim obtemos as estimativas dos parâmetros da distribuição em questão, $\alpha = 0.3004$ e $\beta = 4.2798$ e os coeficientes do modelo estão expressos a seguir.

Covariáveis	Coeficientes
CD4	$\beta_3 = 0.0024$
CD8	$\beta_4 = 0.0015$
Idade	$\beta_5 = 0.0407$
Log-Verossimilhança	-97.32646
AIC	204.6529

Logo, ajustando a curva obtida (Figura 12), pode-se verificar graficamente o comportamento do modelo em relação ao tempo de sobrevivência estimado por meio da distribuição log-logística.

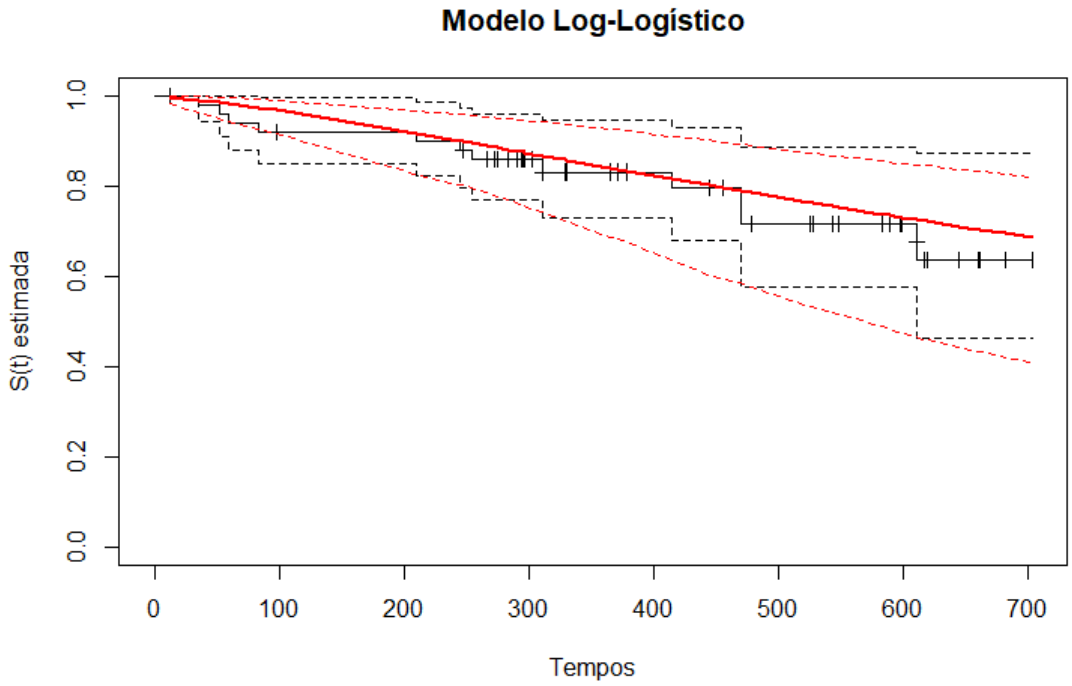

Figura 12: Ajuste via Distribuição Log-Logística.

Oberseva-se na Figura 12 que o modelo ajustado aparenta ser uma boa solução para o problema em questão, pois o tempo de sobrevivência estimado, dado pela linha vermelha, acompanham bem a linha preta que representa o estimado, via modelo não paramétrico, do tempo de sobrevivência e, além disso, pode-se observar os limites inferiores e superiores do intervalo de confiança para as estimativas apresentam um comportamento semelhante.

3.2.4 Comparação

A estatística AIC apresentada em ambos os modelos acima é conhecida como Critério de Informação de Akaike. Ela mede o quão bem ajustado aos dados um modelo é: quanto menor, melhor é o modelo à ela associado.

Temos que, nos modelos apresentados acima, um deles (da distribuição Gompertz) possuiu comportamento atípico pois, segundo a formulação conceitual de nosso problema, era

de se esperar que mais covariáveis fossem significativas.

O segundo (da distribuição Log-Logística) leva em conta a contagem dos Leucócitos CD4 e CD8, o que faz sentido: são diretamente ligados à deficiência imunológica do paciente que, por consequência, tem mais chances de desenvolver sinusite, por exemplo.

Os valores AIC obtidos, se comparados, mostram que o modelo ajustado à distribuição Log-Logística é mais adequado para nossas finalidades. Portanto, a conclusão de nossa Análise Paramétrica é que, dentre os modelos escolhidos, é melhor utilizarmos o Modelo Log-Logístico com as covariáveis "CD4", "CD8"e Idade.

3.3 Ajuste aos Modelos de Cox

Os modelos de regressão de Cox, também chamados de modelos de riscos proporcionais de Cox (Cox, 1972), é essencialmente um modelo de regressão estatística comumente usado em pesquisas médicas para investigar a associação entre o tempo de sobrevida dos pacientes e uma ou mais variáveis preditoras. No caso do estudo o usaremos para determinar a associação entre o tempo do desenvolvimento de sinusite entre pacientes HIV positivos e negativos com as demais covariáveis já citadas.

Os métodos utilizados anteriormete no relatóri, são exemplos de análise univariada. Eles descrevem a sobrevivência de acordo com um fator sob investigação, mas ignoram o impacto das demais.

Além disso, esses testes são úteis apenas quando a variável preditora é categórica, o que não é o caso do estudo. Sendo assim, podemos ter o modelo de Cox como um método alternativo que modela tanto variáveis preditoras quantitativas quanto variáveis categóricas. Tendo também como extensão de sua característica os métodos de análise de sobrevivência para avaliar simultaneamente o efeito de vários fatores de risco no tempo de sobrevivência.

3.3.1 Análise de diagnóstico

O modelo de Cox não se ajusta a qualquer situação, como qualquer outro modelo estatístico, requer o uso de técnicas para avaliar a sua adequação, faremos algumas análises para averiguar se a suposição de proporcionalidade do risco, esta satisfeita.

Resíduos de Schoenfeld

A seguir, apresentamos os Graficos de Resíduos de Schoenfeld, para cada uma das covariaveis, afim de analisar a suposição de proporcionalidade do risco.

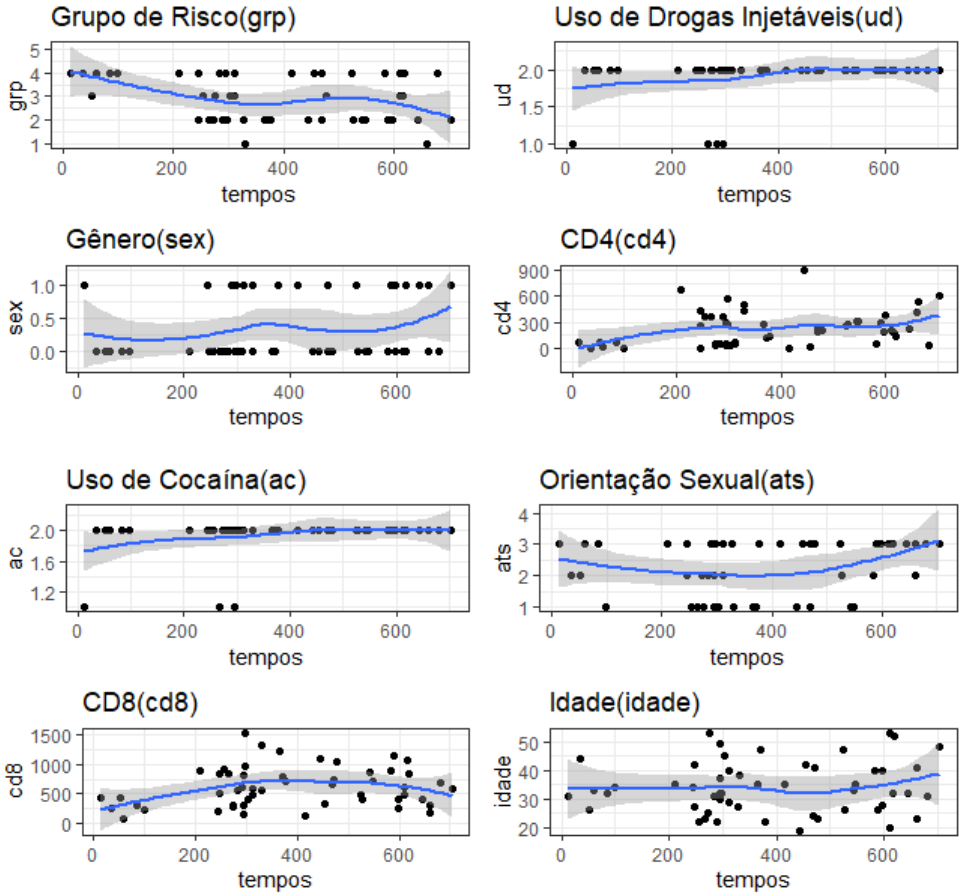

Figura 13: Graficos de Resíduos de Schoenfeld

Note que, apenas os gráficos referentes as variáveis Grupo de Risco e a Orientação sexual, apresentam um leve indício de que talvez ocorra uma quebra na suposição de risco proporcional, os demais gráficos não possuem tendências marcantes, ou concentração de pontos que puxem o gráfico.

3.3.2 Modelo Semi-Paramétrico

A partir das informações ditas anteriormente sobre o Modelo de Cox, temos então que é possível estimar os efeitos das covariáveis sem qualquer suposição a respeito da distribuição do tempo de sobrevida, e por isso podemos dizer que o modelo é dito semi-paramétrico, pois,

não assume qualquer distribuição para a função de risco basal $\lambda_0(t)$, apenas que as covariáveis agem multiplicativamente sobre o risco sendo esta a parte paramétrica do modelo.

Para todas as variáveis preditoras citadas na Subseção 2.1, com as ressalvas feitas na Subsubseção 3.3.1 e com auxílio do software RStudio, obteve-se os seguintes resultados:

Tabela 4: Tabela com os coeficientes estimados.

Covariáveis	**Coeficientes**	**Razão de riscos**	**p-valor**
Idade	-1.007	1.106	0.0078
Gênero Masculino	5.054	1.658	0.6215
Grupos de Risco 2	2.107	1.415	0.9984
Grupos de Risco 3	2.602	2.007	0.9980
Grupos de Risco 4	2.610	2.165	0.9980
Contagem do Linfócito CD4	7.567	1.008	0.0469
Contagem do Linfócito CD8	-6.546	1.007	0.0171
Orientação Sexual Bissexual	1.326	3.767	0.2181
Orientação Sexual Heterosexual	-1.287	3.622	0.2382
Usa de Drogas Injetáveis	2.122	1.640	0.9996
Usa de Cocaína por Aspiração	7.152	2.045	1.000

Como podemos ver na Tabela 4, à um nível de significância de 5 %, temos que apenas as covariáveis idade, contagem do Linfócito CD4 e CD8 se mostram significativas para ocorrência do evento de interesse (ter sinusite) e para o objetivo do estudo em verificar a associação entre o tempo do desenvolvimento de sinusite entre pacientes HIV positivos e negativos. Já as demais covariáveis não se mostraram significativas, ou seja, não são importantes fatores de risco. Este modelo adquiriu um valor da estatística do TRV de 30.77.

Dadas as conclusões acima, encontramos que a covariável idade e Contagem do Linfócito CD8 apresentam-se como fatores de redução de risco de -1.007 e -6.546 vezes, respectivamente. Isto é, a cada ano de idade a mais implica em uma redução de risco de cerca de 10% de não se ter sinusite e para a covariável com efeito protetor Contagem do Linfócito

CD8 temos que conforme essa contagem aumenta o indivíduo tem cerca de 0.7% de chances de não ter sinusite. Já a variável Contagem do Linfócito CD4 indica que um paciente com contagem baixa desse linfócito tem 1.008 vezes mais chance de ter sinusite, por unidade de tempo.

3.3.3 Modelo Paramétrico 1

O primeiro modelo Cox paramétrico a ser analisado, será o modelo de Cox-Gompertz, levando em conta o grupo de Risco, o uso de drogas injetáveis, o uso de Cocaína, a orientação sexual,o gênero , contagem do Linfócito CD4,CD8 e a idade. Os resultados do Modelo Cox-Gompertz, estão na tabela abaixo :

Tabela 5: Tabela com os coeficientes estimados.

Covariáveis	**Coeficientes**	**Erro**	**p-valor**
Idade	-0.107	0.037	0.004
Gênero Masculino	0.315	1.008	0.755
Grupos de Risco 2	16.136	755.956	0.983
Grupos de Risco 3	21.441	755.960	0.977
Grupos de Risco 4	21.445	755.959	0.977
Contagem do Linfócito CD4	0.008	0.004	0.037
Contagem do Linfócito CD8	-0.007	0.003	0.010
Orientação Sexual Bissexual	1.189	1.081	0.271
Orientação Sexual Heterosexual	-1.520	1.138	0.182
Usa de Drogas Injetáveis	11.175	267.769	0.967
Usa de Cocaína por Aspiração	5.614	1110.256	0.996
Estatística do TRV =	33.60		
Log-Verossimilhança =	-84.125		
P-valor geral =	0.00042006		

Como podemos ver na Tabela 5, à um nível de significância de 5 %, temos que apenas as covariáveis idade, contagem do Linfócito CD4 e CD8 se mostram significativas para ocorrên-

cia do evento de interesse (ter sinusite) e objetivo do estudo em verificar a associação entre o tempo do desenvolvimento de sinusite entre pacientes HIV positivos e negativos. Já as demais covariáveis não se mostraram significativas, ou seja, não são importantes fatores de risco.

Dadas as conclusões acima encontramos que a covariável idade e Contagem do Linfócito CD8 apresentam-se como fatores de redução de risco de -0.107 e -0.007, respectivamente. Isto é, a cada ano de idade a mais implica em uma redução de risco de cerca de 11% de não se ter sinusite e para a covariável com efeito protetor Contagem do Linfócito CD8 temos que conforme essa contagem aumenta o indivíduo tem cerca de 0.7% de chances de não ter sinusite. Já a variável Contagem do Linfócito CD4 indica que um paciente com contagem baixa desse linfócito tem 0.004 vezes mais chance de ter sinusite, por unidade de tempo.

3.3.4 Modelo Paramétrico 2

Por fim, o segundo modelo Cox paramétrico a ser analisado, será o modelo de Cox log-logístico, que levará em consideração as mesmas variáveis do modelo anterior. Seus resultados estão na tabela abaixo:

Tabela 6: Tabela com os coeficientes estimados.

Covariáveis	**Coeficientes**	**Razão de Riscos**	**p-valor**
Idade	-0.111	0.039	0.004
Gênero Masculino	0.078	1.002	0.938
Grupos de Risco 2	31.450	1878961.365	1.000
Grupos de Risco 3	36.616	1878961.365	1.000
Grupos de Risco 4	36.574	1878961.365	1.000
Contagem do Linfócito CD4	0.007	0.004	0.054
Contagem do Linfócito CD8	-0.007	0.003	0.020
Orientação Sexual Bissexual	1.258	1.129	0.265
Orientação Sexual Heterossexual	-1.226	1.122	0.274
Usa de Drogas Injetáveis	28.889	1716893.528	1.000
Usa de Cocaína por Aspiração	1.745	1994426.249	1.000

Como podemos ver na Tabela 6, à um nível de significância de 5 %, temos que, semelhantemente ao que acontece no primeiro modelo, apenas as covariáveis idade, contagem do Linfócito CD4 e CD8 se mostram significativas para ocorrência do evento de interesse (ter sinusite); enquanto as outra variáveis não demonstraram indícios de serem fatores de risco.

Dadas as conclusões acima, observou-se que, para este modelo, a covariável Idade e Contagem do Linfócito CD8 apresentam-se como fatores de redução de risco de -0.111 e -0.007, respectivamente. Isto é, a cada ano de idade a mais implica em uma redução de risco de cerca de 11% de não se ter sinusite e para a covariável com efeito protetor Contagem do Linfócito CD8 temos que conforme essa contagem aumenta o indivíduo tem cerca de 0.7% de chances de não ter sinusite. Já a variável Contagem do Linfócito CD4 indica que um paciente com contagem baixa desse linfócito tem 0.004 vezes mais chance de ter sinusite, por unidade de tempo.

3.3.5 Comparação

No que diz respeito aos Modelos de Cox, resolvemos visualizar um cenário em que utilizamos todas as variáveis presentes no nosso banco de dados. Tomando as estatísticas do TRV e p-valores gerais de cada modelo, o modelo que mais nos chamou a atenção foi o Modelo Cox-Gompertz ajustado para todas as variáveis.

Temos que algumas de suas variáveis possuem p-valor grande, denotando algo que já suspeitávamos após a análise paramétrica anteriormente neste trabalho: que algumas variáveis no banco de dados não nos são úteis, seja pelo fato do banco de dados ter passado por uma limpeza ou pelo fato de ser necessária uma nova coleta com o objetivo de trazer mais informações.

4 Conclusão

Ao longo de todas as análises feitas neste trabalho, o principal intuito foi aplicarmos os conceitos aprendidos até aqui, algo realizado sem maiores problemas. O modelo paramétrico Log-Logístico se mostrou uma ferramenta interessante para o contexto fora dos ajustes aos Modelos de Cox, enquanto que o Modelo Cox-Gompertz se destacou.

Com conhecimentos mais avançados (e, possivelmente, uma nova coleta para o banco de dados), temos ciência de que seja possível atingirmos conclusões mais sólidas. Devido ao tempo e à ementa da disciplina, no entanto, isso fica como um "gostinho"de novos métodos que possivelmente serão vistos mais à frente, ainda mais no que diz respeito à manipulação dos dados faltantes e que variam com o passar do tempo.

Por fim, o aprendizado sumarizado neste trabalho não foi apenas o dos conceitos de análise de sobrevivência, mas também os de contextualização: o banco de dados utiliza diversos termos datados e que, num contexto atual, seriam utilizados de outras formas. A possibilidade de aplicação deste tema nos dias de hoje, no entanto, continua interessante.

5 Referências Bibliográficas

- COLOSIMO, Enrico Antônio; GIOLO, Suely Ruiz. Análise de Sobrevivência Aplicada. Brasil: Blucher, 2006. 392 p. ISBN 9788521203841.

- TOMAZELLA, Vera Lucia Damasceno. Slides de Aula da Disciplina de Análise de Sobrevivência. Slides. Disponível em: Google Classroom. Acesso em: Junho de 2021.

6 Apêndice: Código Utilizado

6.1 Análise Não Paramétrica

6.1.1 Tabela de Estimativas de Kaplan-Meier

```
library(tidyverse)
library(survival)

#LIMPEZA DOS DADOS

dados <- read.table("aids.txt", header=T) %>%
  rename("ID"="pac",
         "idade"="id",
         "tempos"="tf") %>%
  drop_na()  %>%
  filter(tempos != 0) %>%
  select(-ti) %>%
  slice(-c(6, 8, 14, 16, 24,
           26, 36, 38, 39, 52,
           58, 60, 63, 64)) %>%
  mutate(ID=row_number())

attach(dados)

#tabela kaplan-meier para dados completos

tabela <- survfit(Surv(tempos, cens)~1, type="kaplan-meier") %>%
  summary(times=seq(0, 800, 75))
tabela
```

```
#montando tabela latex

tabela.latex <- data.frame(tabela$time, tabela$n.risk,
                           tabela$n.event, tabela$n.censor,
                           tabela$surv, tabela$std.err,
                           tabela$cumhaz,tabela$std.chaz,
                           tabela$lower, tabela$upper)

knitr::kable(tabela.latex, "latex")
```

6.1.2 Curvas de Kaplan-Meier

```
#grafico de sobrevivencia para dados completos

survfit(Surv(tempos, cens)~1, type="kaplan-meier") %>%
  plot(lwd=2, ymin=0.3, conf.int=FALSE,
       main="Estimação de Sobrevivência de Kaplan-Meier",
       xlab="Dias",
       ylab="S(x) estimada")

#grafico de sobrevivencia por grupos de risco

fit1 <- survfit(Surv(tempos, cens)~grp,
                type="kaplan-meier")

#plot

plot(fit1, col=c("red", "blue", "green", "pink"),
     lwd=2, conf.int=FALSE, lty=1:4,
     main="Estimação de Sobrevivência de Kaplan-Meier por Grupo",
```

```
     xlab="Dias",
     ylab="S(x) estimada")

legend("bottom",
       legend=c("HIV Negativo",
                "HIV Positivo Assintomático",
                "Paciente com ARC",
                "Paciente com AIDS"),
       col=c("red", "blue", "green", "pink"), lty=1:4)


#por uso ou nao de droga

fit2 <- survfit(Surv(tempos, cens)~ud,
                type="kaplan-meier")


#plot

plot(fit2, col=c("red", "blue"),
     lwd=2, conf.int=FALSE, lty=1:2,
     main="Estimação de Sobrevivência por Uso de Drogas Injetáveis",
     xlab="Dias",
     ylab="S(x) estimada")

legend("bottom",
       legend=c("Não usa Drogas Injetáveis", "Usa Drogas Injetáveis"),
       col=c("red", "blue"), lty=1:2)
```

```
#por uso ou nao de cocaina por aspiracao

fit22 <- survfit(Surv(tempos, cens)~ac,
                 type="kaplan-meier")



#plot

plot(fit22, col=c("red", "blue"),
     lwd=2, conf.int=FALSE, lty=1:2,
     main="Estimação de Sobrevivência por Uso de Cocaína por Aspiração",
     xlab="Dias",
     ylab="S(x) estimada")

legend("bottom",
       legend=c("Não usa cocaína", "Usa cocaína"),
       col=c("red", "blue"), lty=1:2)

#por atividade sexual

fit3 <- survfit(Surv(tempos, cens)~ats,
                 type="kaplan-meier")



#plot

plot(fit3, col=c("red", "blue", "green"),
     lwd=2, ymin=0, conf.int=FALSE, lty=1:3,
     main="Estimação de Sobrevivência de Kaplan-Meier por Orientação Sexual",
```

```
     xlab="Dias",
     ylab="S(x) estimada")

legend("bottom",
       legend=c("Homossexual", "Bissexual", "Heterossexual"),
       col=c("red", "blue", "green"), lty=1:3)

#por genero

fit4 <- survfit(Surv(tempos, cens)~sex,
                type="kaplan-meier")



#plot

plot(fit4, col=c("red", "blue"),
     lwd=2, conf.int=FALSE, lty=1:2,
     main="Estimação de Sobrevivência de Kaplan-Meier por Gênero",
     xlab="Dias",
     ylab="S(x) estimada")

legend("bottom",
       legend=c("Feminino", "Masculino"),
       col=c("red", "blue"), lty=1:2)



#sobrevivencia por contagem de CD4

cd.dados1 <- dados %>% filter(cd4<=52.25)
```

```
cdfit1 <- survfit(Surv(cd.dados1$tempos, cd.dados1$cens)~1,
                  type="kaplan-meier")

cd.dados2 <- dados %>% filter(cd4<=205, 52.25<cd4)
cdfit2 <- survfit(Surv(cd.dados2$tempos, cd.dados2$cens)~1,
                  type="kaplan-meier")

cd.dados3 <- dados %>% filter(cd4<=322.5, 205<cd4)
cdfit3 <- survfit(Surv(cd.dados3$tempos, cd.dados3$cens)~1,
                  type="kaplan-meier")

cd.dados4 <- dados %>% filter(322.5<cd4)
cdfit4 <- survfit(Surv(cd.dados4$tempos, cd.dados4$cens)~1,
                  type="kaplan-meier")

#plot

colors <- c("red", "blue", "green", "pink")
groups <- c("CD4 <= 52.25", "52.25 < CD4 <= 205",
              "205 < CD4 <= 322.5", "322.5 < CD4")

plot(cdfit1, col=colors[1],
     lwd=2, ymin=0.3, conf.int=FALSE, lty=1,
     main="Estimação de Sobrevivência de Kaplan-Meier por Contagem de CD4",
     xlab="Dias",
     ylab="S(x) estimada")
lines(cdfit2, col=colors[2], lty=2, lwd=2, conf.int=FALSE)
lines(cdfit3, col=colors[3], lty=3, lwd=2, conf.int=FALSE)
lines(cdfit4, col=colors[4], lty=4, lwd=2, conf.int=FALSE)
```

```
legend("bottomleft",
       legend=groups,
       col=colors, lty=1:4)

#sobrevivencia por contagem de CD8

cd.dados1 <- dados %>% filter(cd8<=335)
cdfit1 <- survfit(Surv(cd.dados1$tempos, cd.dados1$cens)~1,
                  type="kaplan-meier")

cd.dados2 <- dados %>% filter(cd8<=572.5, 335<cd8)
cdfit2 <- survfit(Surv(cd.dados2$tempos, cd.dados2$cens)~1,
                  type="kaplan-meier")

cd.dados3 <- dados %>% filter(cd8<=850, 572.5<cd8)
cdfit3 <- survfit(Surv(cd.dados3$tempos, cd.dados3$cens)~1,
                  type="kaplan-meier")

cd.dados4 <- dados %>% filter(850<cd8)
cdfit4 <- survfit(Surv(cd.dados4$tempos, cd.dados4$cens)~1,
                  type="kaplan-meier")

#plot

colors <- c("red", "blue", "green", "pink")
groups <- c("CD8 <= 335", "335 < CD8 <= 572.5",
            "572.5 < CD8 <= 850", "850 < CD8")
```

```
plot(cdfit1, col=colors[1],
     lwd=2, ymin=0.3, conf.int=FALSE, lty=1,
     main="Estimação de Sobrevivência de Kaplan-Meier por Contagem de CD8",
     xlab="Dias",
     ylab="S(x) estimada")
lines(cdfit2, col=colors[2], lty=2, lwd=2, conf.int=FALSE)
lines(cdfit3, col=colors[3], lty=3, lwd=2, conf.int=FALSE)
lines(cdfit4, col=colors[4], lty=4, lwd=2, conf.int=FALSE)


legend("bottomleft",
       legend=groups,
       col=colors, lty=1:4)


#por idade

id.dados1 <- dados %>% filter(idade<=26.75)
idfit1 <- survfit(Surv(id.dados1$tempos, id.dados1$cens)~1,
                  type="kaplan-meier")

id.dados2 <- dados %>% filter(idade<=32.5, 26.75<idade)
idfit2 <- survfit(Surv(id.dados2$tempos, id.dados2$cens)~1,
                  type="kaplan-meier")

id.dados3 <- dados %>% filter(idade<=40.25, 32.5<idade)
idfit3 <- survfit(Surv(id.dados3$tempos, id.dados3$cens)~1,
                  type="kaplan-meier")
```

```
id.dados4 <- dados %>% filter(40.25<idade)
idfit4 <- survfit(Surv(id.dados4$tempos, id.dados4$cens)~1,
                  type="kaplan-meier")

#plot

colors <- c("red", "blue", "green", "pink")
groups <- c("Idade <= 26.75", "26.75 < Idade <= 32.5",
            "32.5 < Idade <= 40.25", "40.25 < Idade")

plot(idfit1, col=colors[1],
     lwd=2, ymin=0.3, conf.int=FALSE, lty=1,
     main="Estimação de Sobrevivência de Kaplan-Meier por Idade",
     xlab="Dias",
     ylab="S(x) estimada")
lines(idfit2, col=colors[2], lty=2, lwd=2, conf.int=FALSE)
lines(idfit3, col=colors[3], lty=3, lwd=2, conf.int=FALSE)
lines(idfit4, col=colors[4], lty=4, lwd=2, conf.int=FALSE)



legend("bottomleft",
       legend=groups,
       col=colors, lty=1:4)
```

6.1.3 Teste Log-Rank

```
#grupo de risco
survdiff(Surv(tempos, cens)~grp)
```

```
#uso de drogas
survdiff(Surv(tempos, cens)~ud)

#cocaina
survdiff(Surv(tempos, cens)~ac)

#orientacao sexual
survdiff(Surv(tempos, cens)~ats)

#genero
survdiff(Surv(tempos,cens)~sex)

#cd4
survdiff(Surv(tempos, cens)~cd4)

#cd8
survdiff(Surv(tempos, cens)~cd8)

#idade
survdiff(Surv(tempos, cens)~idade)
```

6.2 Análise Paramétrica

6.2.1 TTT-Plot

```
##### TTT-PLOT ########################################################################

# TTT-PLOT COM CENSURA

o = order(dados$tempos)
t = dados$tempos[o]
```

```
cens = dados$cens[o]

n = length(t)
r = sum(cens)

j = 1

TF = numeric()
MON = numeric()
I = numeric()
TTT = numeric()
Fi = numeric()
F = numeric()
S = numeric()

TF[j] = 0
MON[j] = 0
F[j] = 0
S[j] = 1
TTT[j] = 0
i = 1

while(i < (n+1)){
        if(cens[i] == 1){
                j = j + 1
                TF[j] = t[i]
                NI = n - i + 1
                I = ((n + 1) -MON[j - 1])/(1 + NI)
                MON[j] = MON[j - 1] + I
```

```
                F[j] = MON[j]/n
                S[j] = 1 - F[j]
        }
        i = i + 1
}

TF[r + 2] = t[n]
F[r + 2] = 1
TTT[1] = 0

for(j in 2:(r+2)){
        TTT[j] = TTT[j-1]+n*S[j-1]*(TF[j]-TF[j-1])
}

for(j in 1:(r+2)){
        Fi[j] = TTT[j]/TTT[r+2]
}

plot(F, Fi, xlim=c(0,1), ylim=c(0,1), ylab='TTT com censuras', xlab='F(t)',
     lwd=2, pch=19, cex.axis=1.4, cex.lab=1.4,
     main = "Tempo Total sob Teste (TTT)")
lines(F, Fi, type='l', lwd=1)
x = c(0,1)
y = c(0,1)
lines(x,y,lwd=2, col=2)
```

6.2.2 Ajuste para a distribuição Gompertz

```
#Encontrando p-valores do teste de wald
```

```
fit.gompertz <- flexsurvreg(Surv(tempos, cens)~idade+as.factor(sex)+as.factor(grp)+cd
                            dist="gompertz")

fit.gompertz
fit.gompertz$coefficients
fit.gompertz %>% plot(type="survival", main="Modelo Gompertz")

invgompertz.res <- fit.gompertz$res
invgompertz.wald <- invgompertz.res[,1]/invgompertz.res[,4]
invgompertz.p <- 2*pnorm(-abs(invgompertz.wald))

invgompertz.p

#Ajuste para variavel significativa

fit.gompertz.new <- flexsurvreg(Surv(tempos, cens)~idade,
                            dist="gompertz")

fit.gompertz.new
fit.gompertz.new$coefficients
fit.gompertz.new %>% plot(type="survival", main="Modelo Gompertz")
```

6.2.3 Ajuste para a distribuição Log-Logística

```
library(survival)
##### LOG-LOGIS ##########################################################
library(flexsurv)

loglogis = flexsurvreg(formula = Surv(dados$tempos, dados$cens) ~ dados$idade +
                    dados$sex + dados$grp + dados$cd4 + dados$cd8 + dados$ats,
```

```
                    data = dados, dist = "gompertz")
loglogis



plot(loglogis,
     main="Modelo Log-Logístico",
     xlab = "Tempos", ylab = "S(t) estimada", mark.time = TRUE, lwd = 2)

##############################################################################

loglogis = flexsurvreg(formula = Surv(dados$tempos, dados$cens) ~ dados$idade +
                                  dados$sex +
                                  dados$grp +
                                  dados$ats +
                                  dados$ac +
                                  dados$ud +
                                  dados$cd4 +
                                  dados$cd8,
                       data = dados, dist = "llogis")
loglogis$coefficients
plot(loglogis,
     main="Modelo Log-Logístico",
     xlab = "Tempos", ylab = "S(t) estimada", mark.time = TRUE, lwd = 2)



loglogis.res = loglogis$res
loglogis.wald = loglogis.res[,1]/loglogis.res[,4]
loglogis.wald_pvalue = 2*pnorm(-abs(loglogis.wald))
round(loglogis.wald_pvalue, digits = 4)
```

```
loglogis = flexsurvreg(formula = Surv(dados$tempos, dados$cens) ~ dados$idade +
                                 dados$cd4 +
                                 dados$cd8,
                       data = dados, dist = "llogis")
loglogis$coefficients
plot(loglogis,
     main="Modelo Log-Logístico",
     xlab = "Tempos", ylab = "S(t) estimada", mark.time = TRUE, lwd = 2)
```

6.3 Modelos de Cox

```
require(eha)
library(ggplot2)
library(gridExtra)

## Analise de residuos



## Resíduos de Schoenfeld

fit1 <- coxph(Surv(tempos, cens)~grp+ud+ac+ats+sex+cd4+cd8+idade,
              data    = dados,
              ties    = c("efron","breslow","exact")[1])
summary(fit1)

dados$resid_mart <- residuals(fit1, type = "martingale")



# grafico covariavel grp
```

```
grafico_1 <- ggplot(data = dados, mapping = aes(x = tempos, y = grp)) +
  geom_point() +
  geom_smooth() +
  labs(title = "Grupo de Risco(grp)") +
  theme_bw() + theme(legend.key = element_blank())

# grafico covariavel ud

grafico_2 <- ggplot(data = dados, mapping = aes(x = tempos, y = ud)) +
  geom_point() +
  geom_smooth() +
  labs(title = "Uso de Drogas Injetáveis(ud)") +
  theme_bw() + theme(legend.key = element_blank())

# grafico covariavel ac

grafico_3 <- ggplot(data = dados, mapping = aes(x = tempos, y = ac)) +
  geom_point() +
  geom_smooth() +
  labs(title = "Uso de Cocaína(ac)") +
  theme_bw() + theme(legend.key = element_blank())
# grafico covariavel ats

grafico_4 <- ggplot(data = dados, mapping = aes(x = tempos, y = ats)) +
  geom_point() +
  geom_smooth() +
  labs(title = "Orientação Sexual(ats)") +
  theme_bw() + theme(legend.key = element_blank())
```

```
# grafico covariavel sex

grafico_5 <- ggplot(data = dados, mapping = aes(x = tempos, y = sex)) +
  geom_point() +
  geom_smooth() +
  labs(title = "Gênero(sex)") +
  theme_bw() + theme(legend.key = element_blank())

# grafico covariavel cd4

grafico_6 <- ggplot(data = dados, mapping = aes(x = tempos, y = cd4)) +
  geom_point() +
  geom_smooth() +
  labs(title = "CD4(cd4)") +
  theme_bw() + theme(legend.key = element_blank())

# grafico covariavel cd8

grafico_7 <- ggplot(data = dados, mapping = aes(x = tempos, y = cd8)) +
  geom_point() +
  geom_smooth() +
  labs(title = "CD8(cd8)") +
  theme_bw() + theme(legend.key = element_blank())

# grafico covariavel idade

grafico_8 <- ggplot(data = dados, mapping = aes(x = tempos, y = idade)) +
  geom_point() +
```

```
  geom_smooth() +
  labs(title = "Idade(idade)") +
  theme_bw() + theme(legend.key = element_blank())


#Plotando os graficos Resíduos de Schoenfeld


grid.arrange(grafico_1 , grafico_2 ,grafico_3 , grafico_4 , grafico_5, grafico_6
             , grafico_7, grafico_8,
             ncol=4, nrow=4)
```

6.3.1 Modelo Semi-Paramétrico

require(survival)

fit2<-coxph(Surv(tempos,cens) idade + factor(sex) + factor(grp) + cd4 + cd8 + factor(ats) + factor(ud) + factor(ac), data=base,x = T, method="breslow")

summary(fit2)

6.3.2 Modelo Paramétrico 1

```
### Modelo Paramétrico 1


##Cox-Gompertz


mod <- phreg(Surv(tempos, cens)~factor(grp)+factor(ats)+factor(ud)+factor(ac)
+factor(sex)+cd4+cd8+idade, data=dados, dist = "gompertz", shape=0)
mod
```

6.3.3 Modelo Paramétrico 2

```
mod <- phreg(Surv(tempos, cens)~factor(grp)+factor(ats)+factor(ud)+factor(ac)+
```

```
factor(sex)+cd4+cd8+idade, data=dados, dist = "loglogistic", shape=0)

mod
```