

Apresentação do Trabalho de Análise de Sobrevivência

Guilherme Mendes, Rafael Ribeiro e Tonny Barbosa

Universidade de Brasília

27/11/2018

Análise de Sobrevivência

- Estuda informações relacionadas ao tempo até a ocorrência de um evento de interesse (falha).
- O tempo de vida dos indivíduos é afectado por variáveis que são observadas no estudo.
- O primeiro passo para iniciarmos esta análise consiste em ter uma base de dados.

Dados para Aplicação

A base que nos foi passasda pela professora Juliana provêm de uma coorte de 6.805 pacientes que foram submetidos a hemodiálise em 67 unidades de atendimento no Rio de Janeiro, no período de janeiro de 1998 a outubro de 2001. Os dados foram originados pelo sistema Apac (Autorização de Procedimentos de Alta Complexidade – DATASUS). Uma discussão detalhada do tema e da modelagem pode ser encontrada em Carvalho e cols. (2003). Cada paciente possui um registo que apresenta dados para cada variável. A tabela a seguir contém a descrição das variáveis.

Variáveis

- **unidade** = número do centro de diálise
- **idade** = idade ao iniciar a diálise (0 a 97 anos)
- **inicio** = data do início da primeira diálise
- **fim** = data da interrupção do acompanhamento
- **status** = (0 = censura, 1 = óbito)
- **tempo** = tempo de sobrevivência (meses) (fim - inicio)
- **causa** = (hip = hipertensão, dia = diabetes, ren = renal, con = congênita, out = outras)
- **grande** = número de salas de diálise na unidade de tratamento: 0 = uma ou duas salas; e 1 = três salas ou mais
- **cdiab** = (1 = diabetes como causa da insuficiência renal e 0 = não)
- **crim** = (1 = causas renais e 0 = não)
- **congenita** = (1 = causas congênitas e 0 = não)

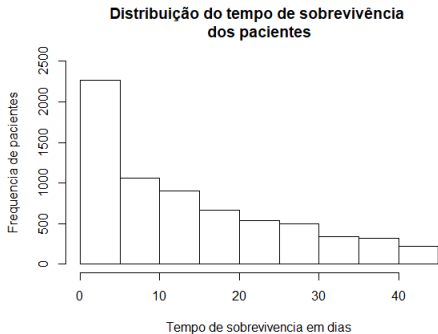
Variáveis

Logo as variáveis ficaram da seguinte forma:

- **unidade** = número do centro de diálise
- **idade** = idade ao iniciar a diálise (0 a 97 anos)
- **inicio** = data do início da primeira diálise
- **fim** = data da interrupção do acompanhamento
- **status** = (0 = censura, 1 = óbito)
- **tempo** = tempo de sobrevivência (meses) (fim - inicio)
- **grande** = número de salas de diálise na unidade de tratamento: 0 = uma ou duas salas; e 1 = três salas ou mais
- **cdiab** = (1 = diabetes como causa da insuficiência renal e 0 = não)
- **crim** = (1 = causas renais e 0 = não)
- **congenita** = (1 = causas congênitas e 0 = não)
- **hip** = (1 = causas hipertensão e 0 = não)
- **out** = (1 = outras causas e 0 = não)

Análise de Dados

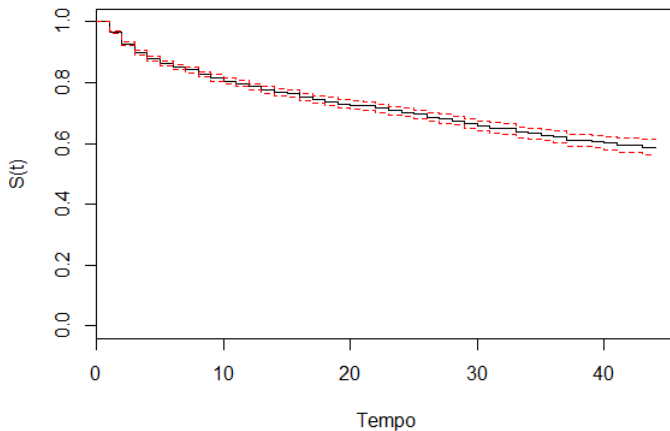
Primeiramente, fizemos uma análise exploratória dos dados. Criamos um histograma para ver a disposição dos dados da variável tempo:



Dado que os dados apresentaram censura, utilizaremos o estimador de Kaplan-Meier para estimar a função de sobrevivência.

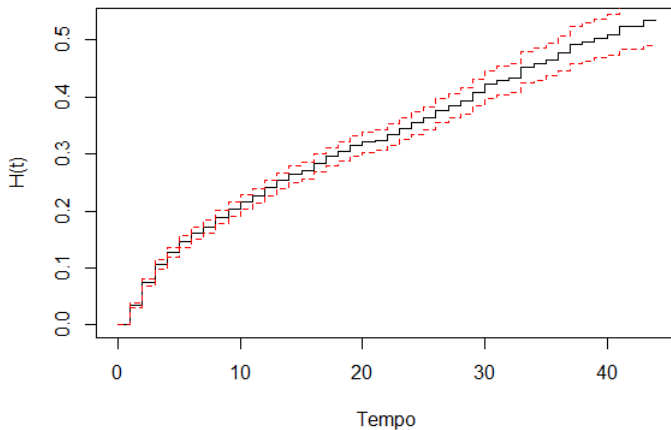
Análise Exploratória

Função de Sobrevivência



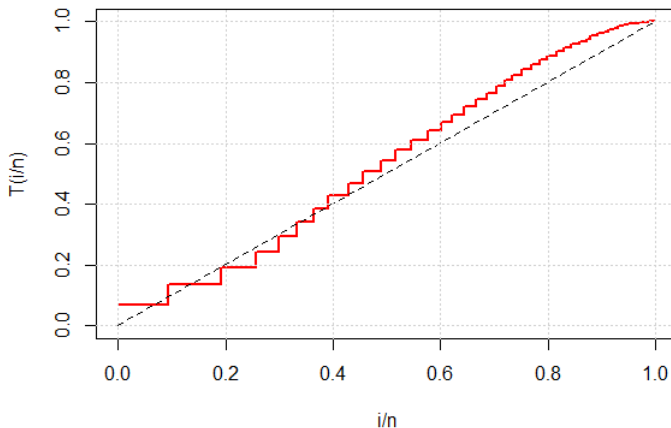
Análise Exploratória

Função de risco acumulada

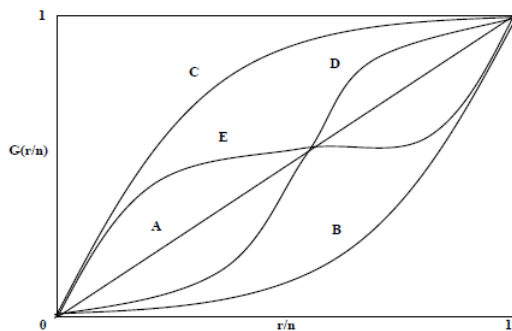


Análise Exploratória

Tempo total em teste (TTT)



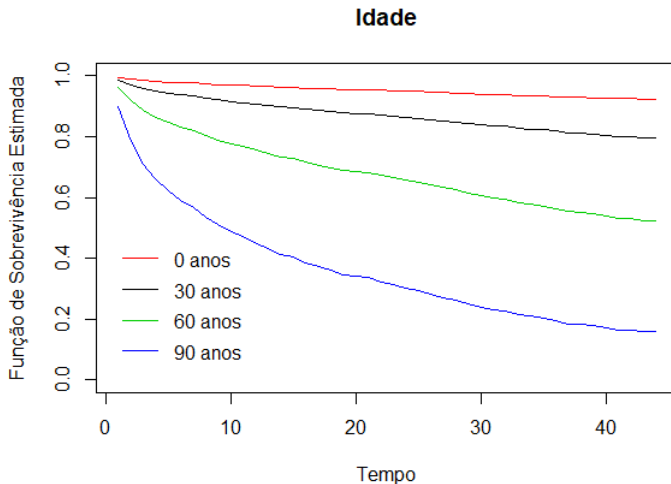
Análise Exploratória



Análise Exploratória de Cada Variável

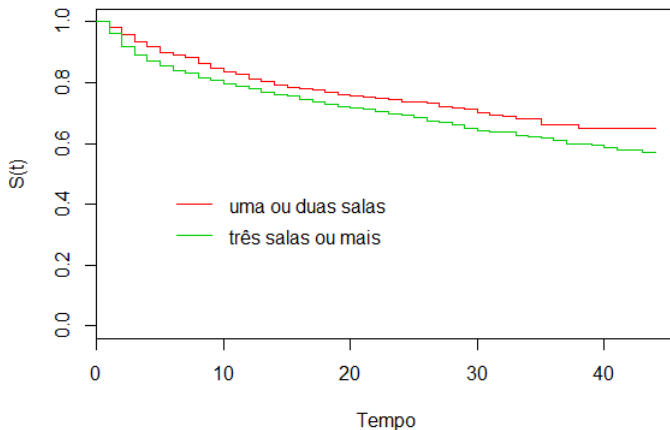
Fizemos uma análise exploratória de cada covariável com a variável resposta, ou seja, a estimativa do Kaplan-Meier e o gráfico TTT para cada fator de cada covariável. Porém a quantidade de gráficos ficou grande, então trouxemos apenas os KM.

Análise Exploratória de Cada Variável (idade)

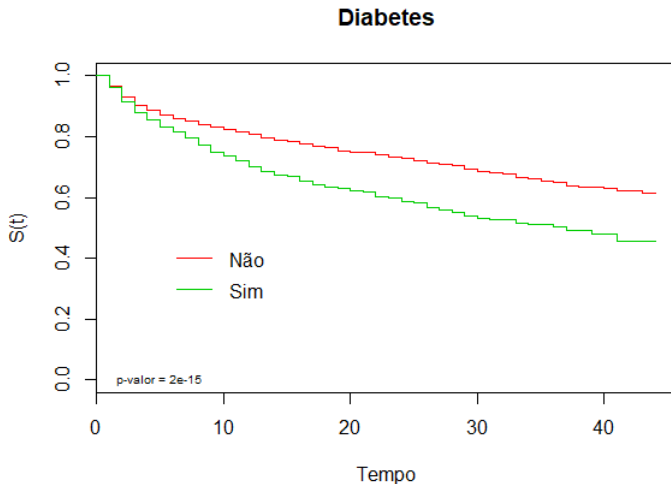


Análise Exploratória de Cada Variável (grande)

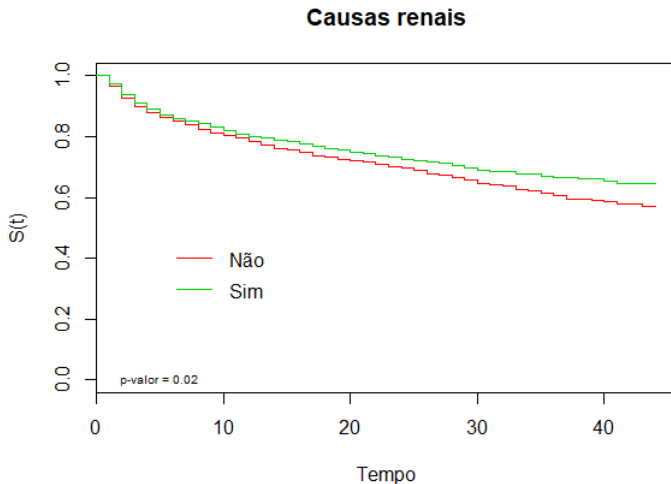
Número de salas de diálise na unidade de tratamento



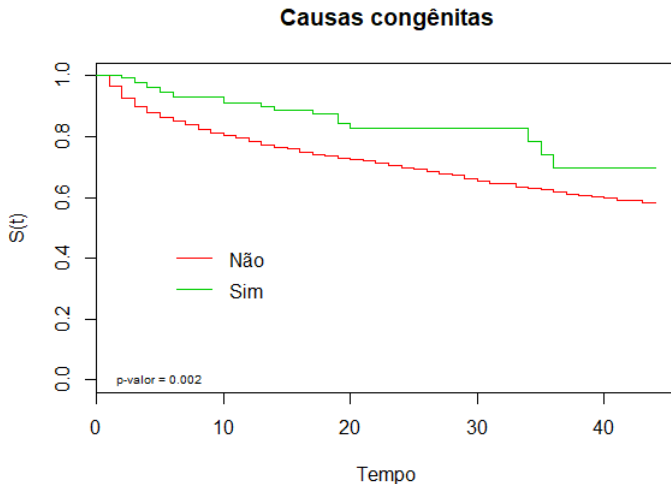
Análise Exploratória de Cada Variável (diabetes)



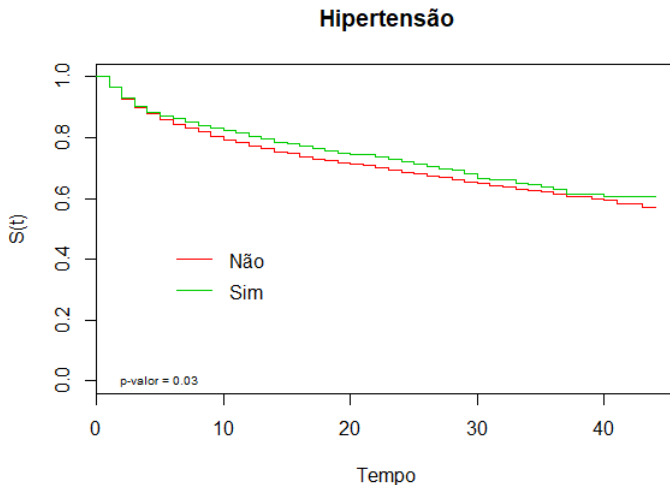
Análise Exploratória de Cada Variável(causas renais)



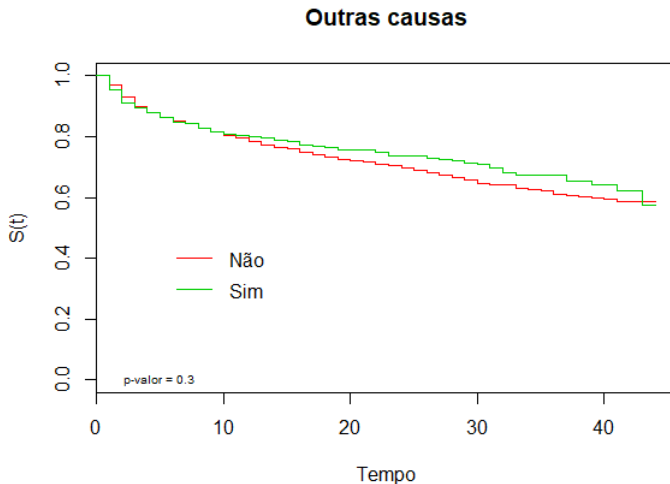
Análise Exploratória de Cada Variável (congenita)



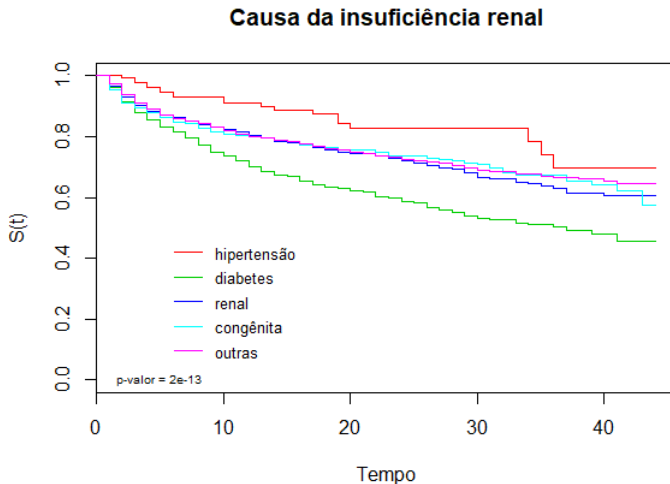
Análise Exploratória de Cada Variável (hipertensão)



Análise Exploratória de Cada Variável (outros)



Análise Exploratória de Cada Variável (Causa)



Análise da Distribuição

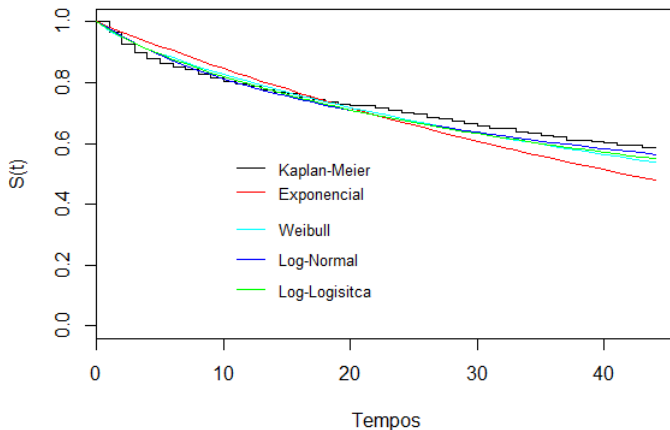
Após verificarmos cada KM e cada TTT de suas respectivas covariáveis vimos que houveram casos onde suspeitamos que a distribuição que poderia modelar o tempo de sobrevivência seria a Log-normal. Porém não temos certeza da distribuição. Então fizemos um estudo entre as distribuições estatísticas para modelar o tempo de sobrevivência: Exponencial, Weibull, Log-normal e Log-logística. Calculamos as regressões AIC, AICc e BIC para cada distribuição.

Distribuição

Distribuição	AIC	AICc	BIC
Exponencial	16339.96	16339.97	16344.67
Weibull	16212.48	16212.48	16221.89
Log-normal	16017.70	16017.72	16027.12
Log-logístico	16156.33	16156.34	16165.74

Ajuste do Modelo

Ajuste dos modelos



Modelagem

- Dado que concluímos a análise exploratória dos dados e selecionamos uma distribuição, passamos para etapa de selecionar as variáveis que serão incluídas na modelagem estatística.
- Para verificarmos o ajuste utilizamos a função *survreg*. Primeiramente testamos as variáveis uma de cada vez. Logo após testamos os conjuntos e depois testamos as interações entre elas. Verificamos que o melhor modelo encontrado foi com as variáveis **grande+cdiab+congenita+grande*congenita**. Sendo que, nesse modelo a interação de grande e congenita deu $p - \text{valor} = 0.024$, assim foi considerada no modelo.

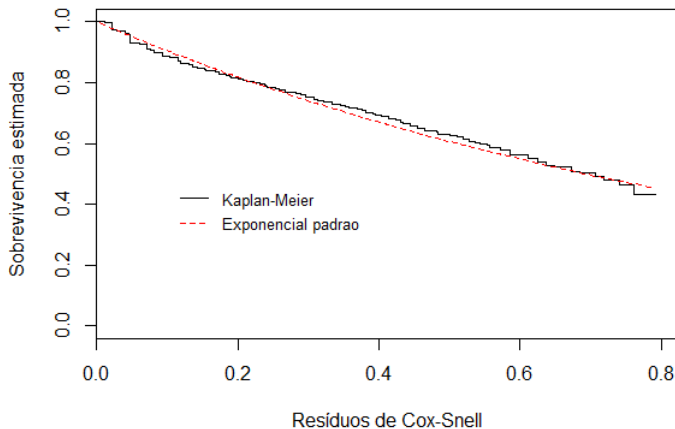
Modelagem

Modelo final: A variável resposta **tempo** trouxe como explicação para o modelo mais ajustado as variáveis: grande, diabetes, congenita e a interação entre grande e congenita.

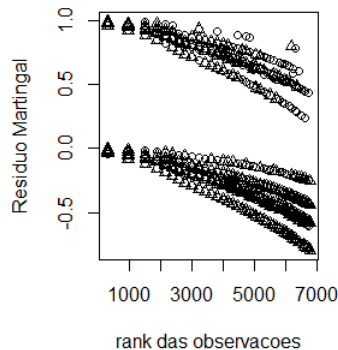
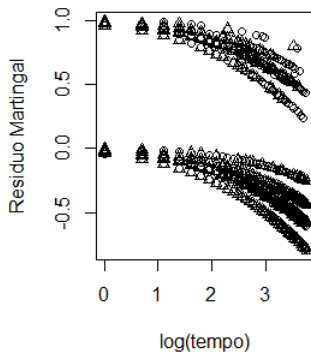
Modelagem

- Para validação dos resultados obtidos na realização de inferências para um determinado modelo é necessário verificar se ele é adequado. Para isso as técnicas de análise de resíduos e sensibilidade são algumas das formas para avaliar a adequabilidade dos modelos considerados.
- Em análise de sobrevivência os resíduos mais utilizados são os de Cox-Snell, martingal e deviance.

Cox-Snell



Martingal



Deviance

