Análise de Sobrevivência

Modelagem paramétrica

Valeska Andreozzi[1]

valeska.andreozzi@fc.ul.pt

&

Marilia Sá Carvalho[2]

cavalho@fiocruz.br

[1]Centro de Estatística e Aplicações da Universidade de Lisboa, Portugal

[2]Escola Nacional de Saúde Pública e Programa de Computação Científica da Fundação Oswaldo Cruz, Brasil

Julho, 2008

Programa

Paramétrica e Não-paramétrica

Os estimadores de Kaplan-Meier e Nelson-Aalen para as funções $S(t)$ e $\lambda(t)$ são obtidos a partir dos dados, supondo que a cada momento do tempo existe um processo diferente gerando as observações.

Como cada intervalo de tempo é estimado de forma independente, a estimação não-paramétrica possui tantos parâmetros quantos intervalos de tempo.

Distribuições

- distribuições estatísticas para modelar as funções de sobrevida:
    - Exponencial
    - Weibull
    - Lognormal

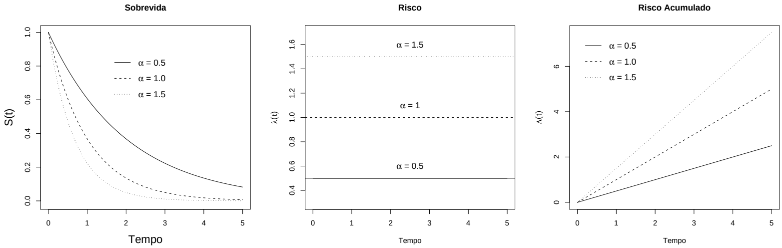
Distribuição Exponencial

Se a variável T possui uma distribuição exponencial,

- Densidade de probabilidade: $f(t) = \alpha \exp(-\alpha t), \qquad \alpha > 0$
- Função de sobrevida: $S(t) = \exp(-\alpha t)$
- A função risco é constante para todo o tempo de observação t, ou seja: $\lambda(t) = \frac{f(t)}{S(t)} = \alpha =$ constante
- A função de risco acumulado é uma função linear no tempo e é dada por: $\Lambda(t) = -\ln S(t) = \alpha t$

Algumas exponenciais

Função de sobrevida, de risco e de risco acumulado para a distribuição exponencial considerando diferentes valores de α



A distribuição exponencial é conhecida como distribuição exponencial padrão quando $\alpha = 1$.

Interpretando risco exponencial

- média e a variância: $\bar{T} = \frac{1}{\alpha}$ e $var(T) = \frac{1}{\alpha^2}$
- quanto maior o risco, menor o tempo médio de sobrevida e menor a variabilidade deste em torno da média
- como a distribuição do tempo de sobrevida T é assimétrica, usa-se mais o tempo mediano
- o modelo exponencial é matematicamente simples, mas a suposição de risco constante no tempo é pouco plausível
- é aplicável quando o tempo de acompanhamento é curto o suficiente para que o risco naquele período possa ser considerado constante (por ex., o risco de óbito de crianças entre dois e cinco anos, independente da causa, pode ser considerado constante neste intervalo)

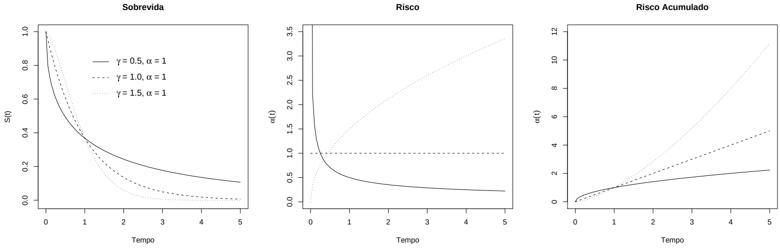
Distribuição Weibull

- permite variação do risco no tempo
- é uma generalização da distribuição exponencial, sendo a densidade $f(t) = \gamma\alpha^{\gamma}t^{\gamma-1}\exp(-(\alpha t)^{\gamma})$ e a sobrevida $S(t) = \exp(-(\alpha t)^{\gamma})$ $(\alpha > 0$ e $\gamma > 0)$
- o parâmetro γ determina a forma da função de risco sendo chamado de parâmetro de forma:

  $\gamma < 1$ função de risco decrescente
  $\gamma > 1$ função de risco crescente
  $\gamma = 1$ função de risco constante (equivalente ao modelo exponencial)
- o parâmetro α determina a escala da distribuição
- a função de risco é: $\Lambda(t) = -\ln S(t) = (\alpha t)^{\gamma-1}$

Algumas Weibull

Função de sobrevida, de risco e de risco acumulado com parâmetro escala $\alpha = 1$ e diferentes valores do parâmetro de forma γ



Regressão Paramétrica

- Nos modelos paramétricos, a inclusão de covariáveis segue a forma utilizada em modelos lineares generalizados, podendo ser tanto contínuas – pressão sanguínea, idade, dosagens bioquímicas – como categóricas – gênero, tratamento, comportamentos.
- O objetivo de um modelo de regressão é o de estimar o efeito de covariáveis (ou variáveis independentes ou preditores), $x_1, x_2, \cdots, x_p$, sobre uma variável resposta (ou variável dependente), Y.
- Supondo uma distribuição da família exponencial para a variável resposta teremos um modelo linear generalizado.
- Ainda que a distribuição exponencial e a Weibull sejam parte desta família, os modelos de regressão paramétricos para tempo de sobrevida não são parte dos GLM por causa de dados censurados.

Exemplo

Assumindo que o risco de morrer é constante ao longo do tempo, pode-se estimar o efeito da $idade$ na sobrevida e no risco de 6.805 pacientes em diálise acompanhados durante um ano (1.603 morreram) através do modelo exponencial:

$$\lambda(t|idade) = \exp(\beta_0 + idade\beta_1)$$

Os parâmetros estimados são: $\beta_0 = 6,14$ e $\beta_1 = -0,04$, ou seja, para cada ano a mais de vida o risco aumenta de $exp(0,04) = 0,96$.
Pode-se comparar o risco constante de morte no tempo, entre dois indivíduos submetidos à diálise, um com 30 anos e outro com 70, substituindo as estimativas dos parâmetros $\boldsymbol{\beta}$:

$$\frac{\lambda(t|x_1 = 30)}{\lambda(t|x_1 = 70)} = \frac{\exp(\beta_0 + 30\beta_1)}{\exp(\beta_0 + 70\beta_1)} = \frac{0,006}{0,029} = 4,39$$