ANÁLISE DE SOBREVIVÊNCIA APLICADA

Empates, Censura Intervalar e Dados Grupados- Cap. 8

Enrico A. Colosimo/UFMG

Depto. Estatística - ICEx - UFMG

Censura Intervalar

- O tempo exato em que o evento ocorreu é desconhecido.
- Sabe-se que ele ocorreu no intervalo $(L, U]$ em que $L < T \leq U$.
- Tempo exato de ocorrência: $T = L = U$.
- Censura à direita: $T \in (L, \infty)$, L é o tempo da última observação.
- Existem técnicas especializadas para censura intervalar
  - não-paramétrica: estimador de Turnbull (1976);
  - escrever a função de verossimilhança para censura intervalar.

Presença de Empates

- Empates: ocorrência de mais de um evento exatamente no mesmo tempo.
- Aproximação em uma escala contínua.
  - Utilização de unidades grosseiras.
  - Censura Intervalar (fazer imputação, meio do intervalo).
- Empates podem ser legítimos.

Como tratar Empates ou/e Censura Intervalar?

- A própria censura intervalar, sem aproximação.
- Modelos Paramétricos (empates ou censura intervalar).
- Aproximação na Verossimilhança Parcial (modelo de Cox) para acomodar empates.
- Modelos Discretos para dados grupados.

Censura Intervalar: Estimador de Turnbull (1976)

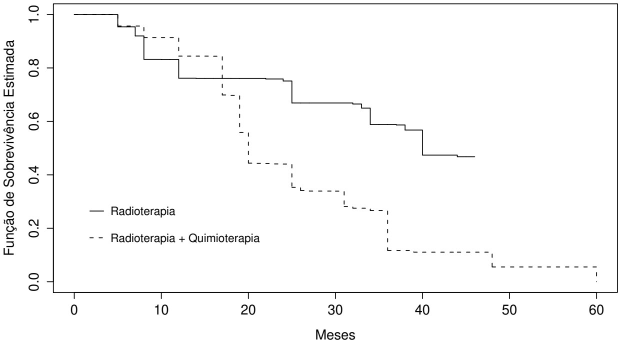
- Estimar a função S(t) de forma não paramétrica em dados de sobrevivência intervalar.
- Turnbull (1976) propôs um estimador não paramétrico (modificação do estimador proposto por Kaplan-Meier).
- O estimador de Turnbull é baseado em um procedimento iterativo, não apresenta uma forma analítica fechada.

Censura Intervalar: Exemplo de Câncer de Mama

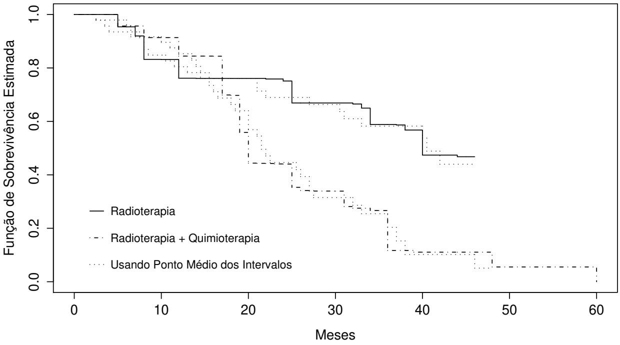
- Estudo retrospectivo: 94 mulheres com câncer de mama (46 receberam radioterapia e 48 radioterapia e quimioterapia). Inicialmente foram observadas a cada 4 ou 6 meses, ocorrendo aumento do tempo entre as visitas quando observado quadro de melhora da paciente.

- Evento de interesse: primeira ocorrência (moderada ou severa) de retração da mama. Devido ao tempo entre as visitas o tempo exato do evento é desconhecido, mas quando ocorreu foi entre duas visitas consecutivas.

Estimativas de Turnbull



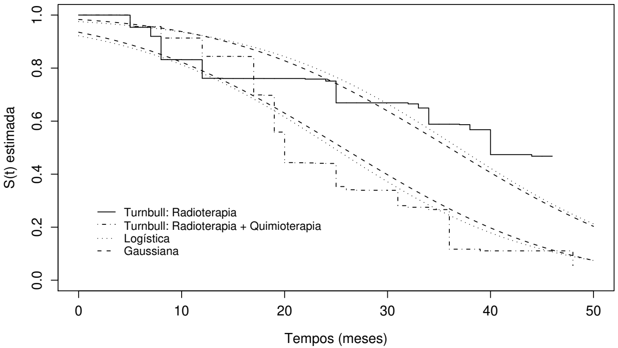
Estimativas de Turnbull



Contirbuição para a função de verossimilhança.

Indivíduo	T	contribuição
Falha	$T = t$	$f(t)$
Censura à direita	$T \in (L, \infty)$	$S(l)$
Censura à esquerda	$T \in (0, U)$	$1 - S(u)$
Censura intervalar	$T \in (LU)$	$S(l) - S(u)$

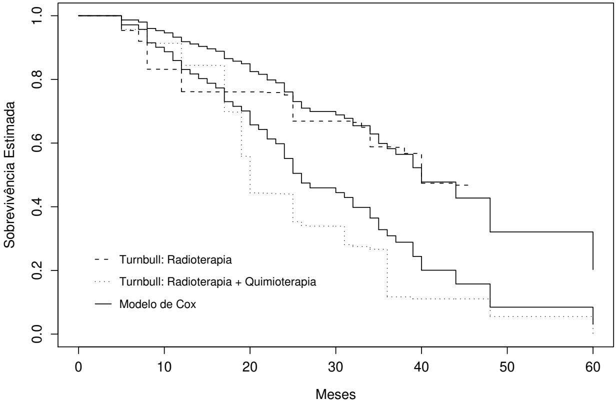
Estimativas de Turnbull e paramétricas



Verossimilhança Parcial na Presença de Empates

- Aproximações apresentadas no Capítulo 5.
  - Breslow.
  - Efron (default do pacote survival no R).
- Verossimilhança parcial exata
  - computacionamente intensiva, principalmente na presença de muitos empates em banco de dados grande.
  - diferenças pequenas das aproximações em banco de dados com muitos indivíduos e poucos empates.
- A verossimihança parcial deve ser evitada na presença de muitos empates.

Modelos de Cox



Modelos Discretos para Dados Grupados

- Dados grupados: caso particular de dados de sobrevivência intervalar.
- Ocorrem quando todas as unidades a mostrais são avaliadas nos mesmos instantes de tempo.
- Por exemplo, nos dias 7, 14, 21, 28 e 35.
- Estão associados, em geral, às situações com excesso de empates.
- Modelos de Regressão Discretos: (1) taxas de falhas proporcionais; (2) chances proporcionais.

$0 \qquad a_1 \qquad\qquad\qquad a_{i-1} \qquad a_i$

A Função de Verossimilhança é escrita em termos de:

$$p_i(x_l) = P[T_l \leq a_i \mid T_l \geq a_{i-1}, x_l].$$

- a contribuição de uma observação censurada no intervalo I_i para a função de verossimilhança é

$$\{(1 - p_1(x_l)) \ldots (1 - p_{i-1}(x_l))\} p_i(x_l)$$

- a contribuição de uma observação não censurada no intervalo I_i para a função de verossimilhança é

$$[(1 - p_1(x_l)) \ldots (1 - p_i(x_l))].$$

A Função de Verossimilhança é dada por

$$\prod_{i=1}^{k} \prod_{l \in R_i} (p_i(x_l))^{\Delta_{li}} (1 - p_i(x_l))^{(1-\Delta_{li})}.$$

Esta é a função de verossimilhança é de uma distribuição de Bernoulli com probabilidade de resposta igual a $p_i(x_l)$.

A estrutura de regressão pode ser modelada segundo o modelo de Cox, ou seja

$$p_i(x_l) = 1 - \gamma_i^{\exp(\beta' x_l)},$$

em que $\gamma_i = S_0(a_i)/S_0(a_{i-1})$ é o efeito do i-ésimo intervalo.

Ou segundo o modelo logístico,

$$p_i(x_l) = 1 - \{1 + \gamma_i \exp(\beta' x_l)\}^{-1},$$

em que $\gamma_i = p_i(0)/\{1 - p_i(0)\}$.

Modelos de Regressão Discretos

- Inserindo estes termos na função de verossimilhança temos os modelos de sobrevivência para tempos discretos ou grupados.

- O Modelo de Regressão para Tempos Discretos somente deve ser utilizado na presença de MUITOS empates.

Qual modelo utilizar na presença de Empates?

Proporção de Empates:

$$PE = \frac{NF - NFD}{N} \times 100\%$$

$PE > 25\%$ Use Modelos Discretos
$PE < 25\%$ Use as Aproximações.

Chalita, Colosimo e Demétrio (2002)

MOTIVAÇÃO

Tempo de Vida de Mangueiras

- Experimento Completamente Aleatorizado em blocos (fazendas);
- Realizado na região de Piracicaba, SP entre 1971 e 1992;
- Resposta: tempo de vida da mangueira;
- Condução do Experimento: visitas em doze ocasiões;
- Fatores: copa (6) e porta-enxerto (7)
- Objetivo: identificar a combinação mais resistente.
- Os efeitos de porta-enxerto e a da interação foram não significativos.

MOTIVAÇÃO

Tempo de Vida de Mangueiras

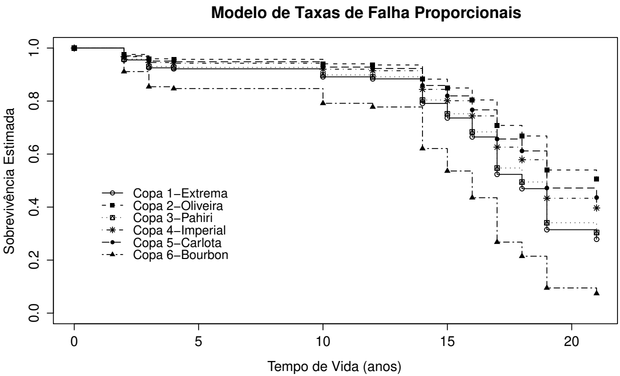
Visita	sob risco	Morte	Sobrevida	Visita	sob risco	Morte	Sobrevida
73	210	12	198	86	156	13	143
74	198	08	190	87	143	16	127
75	190	01	189	88	127	28	99
81	189	08	181	89	99	10	89
83	181	02	179	90	89	27	62
85	179	23	156	92	62	06	56

Ajuste dos Modelos de Regressão para Tempos Discretos

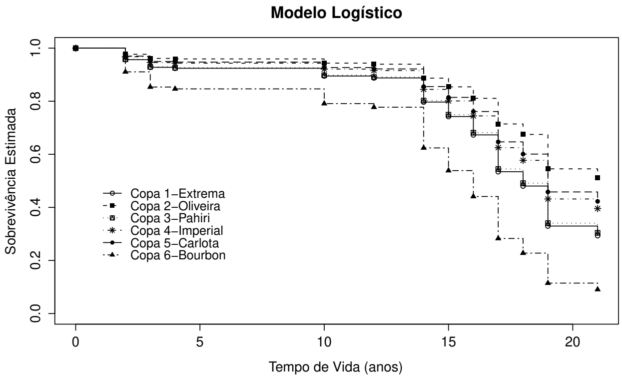
Parameter	logistic model estimate	logistic model p-value	Cox model estimate	Cox model p-value
γ_1^*	-3.089	$<$0.001	-3.086	$<$0.001
γ_2^*	-3.471	$<$0.001	-3.449	$<$0.001
γ_3^*	-5.563	$<$0.001	-5.508	$<$0.001
γ_4^*	-3.427	$<$0.001	-3.401	$<$0.001
γ_5^*	-4.806	$<$0.001	-4.751	$<$0.001
γ_6^*	-2.171	$<$0.001	-2.201	$<$0.001
γ_7^*	-2.609	$<$0.001	-2.630	$<$0.001
γ_8^*	-2.276	$<$0.001	-2.280	$<$0.001
γ_9^*	-1.349	$<$0.001	-1.433	$<$0.001
γ_{10}^*	-2.186	$<$0.001	-2.217	$<$0.001
γ_{11}^*	-0.782	0.029	-0.916	0.004
γ_{12}^*	-2.093	$<$0.001	-2.104	$<$0.001
β_1: Block 2	-0.001	0.999	-0.025	0.927
β_2: Block 3	0.013	0.965	0.012	0.965
β_3: Block 4	0.615	0.028	0.576	0.023
β_4: Block 5	0.630	0.026	0.577	0.025
β_5: Scion 2	-0.653	0.040	-0.629	0.030
β_6: Scion 3	-0.033	0.914	-0.072	0.796
β_7: Scion 4	-0.304	0.324	-0.324	0.251
β_8: Scion 5	-0.384	0.227	-0.432	0.141
β_9: Scion 6	0.768	0.009	0.711	0.006

$\gamma_i^* = \log(\gamma_i)$ is the ith interval effect.

Modelos de Regressão Discretos



Modelos de Regressão Discretos



Ajuste dos Modelos de Regressão para Tempos Discretos

- Os modelos são consistentes;
- As copas foram ordenadas de forma similar em todos os modelos;
- Inferência para censura intervalar é a forma correta de analisar os dados;
- Modelos discretos são recomendados para dados grupados.