

ANÁLISE DE SOBREVIVÊNCIA APLICADA

Extensões do Modelo de Cox - Cap. 6

Enrico A. Colosimo/UFMG

Depto. Estatística - ICEx - UFMG

EXTENSÕES DO MODELO DE COX

- ▶ Algumas covariáveis podem ser monitoradas durante o estudo. Por exemplo, a dose de quimioterapia aplicada em pacientes com câncer pode sofrer alterações durante o curso do tratamento. Tais covariáveis são chamadas de dependentes do tempo.
- ▶ Em outras situações a suposição de RTP pode ser violada. O modelo de Cox estratificado é uma alternativa para enfrentar esta situação.

Modelo de Cox com Covariáveis Dependentes do Tempo

- ▶ Covariáveis que alteram seu valor ao longo do período podem ser incorporadas ao modelo de regressão de Cox:

$$\lambda(t) = \lambda_0(t) \exp \{ \mathbf{x}'(t)\beta \}.$$

- ▶ Definido desta forma, este modelo não é mais de taxas de falha proporcionais pois a razão das funções de taxa de falha no tempo t para dois indivíduos quaisquer i e j fica sendo

$$\frac{\lambda_i(t)}{\lambda_j(t)} = \exp \{ \mathbf{x}'_i(t)\beta - \mathbf{x}'_j(t)\beta \},$$

que é dependente do tempo.

Modelo de Cox com Covariáveis Dependentes do Tempo

- ▶ A interpretação dos coeficientes β do modelo deve considerar o tempo t . Cada coeficiente $\beta_l, l = 1, \dots, p$, pode ter a interpretação usual, mantendo as covariáveis as demais fixas no mesmo tempo.
- ▶ O ajuste do modelo é obtido estendendo o logaritmo da função de verossimilhança parcial. Isto é feito usando

$$U(\beta) = \sum_{i=1}^n \delta_i \left[x_i(t_i) - \frac{\sum_{j \in R(t_i)} x_j(t_i) \exp \{ \mathbf{x}'_j(t_i) \hat{\beta} \}}{\sum_{j \in R(t_i)} \exp \{ \mathbf{x}'_j(t_i) \hat{\beta} \}} \right] = 0$$

- ▶ Os estimadores obtidos são consistentes e assintoticamente normais, sob certas condições de regularidade. As estatísticas de Wald e da razão de verossimilhança são usadas de forma usual.

Modelo de Cox Estratificado

- ▶ O modelo de Cox não pode ser usado se a suposição de RTP for violada. Nestes casos, uma solução é estratificar os dados de modo que a suposição seja válida em cada estrato.
- ▶ A análise estratificada consiste em dividir os dados de sobrevivência em m estratos, de acordo com uma indicação de violação da suposição. O modelo é então expresso como

$$\lambda_{ij}(t) = \lambda_{0_j}(t) \exp \{ \mathbf{x}'_{ij} \boldsymbol{\beta} \},$$

para $j = 1, \dots, m$ e $i = 1, \dots, n_j$, sendo n_j o n^o de observações no j -ésimo estrato. As funções de base $\lambda_{0_1}, \dots, \lambda_{0_m}$, são arbitrárias e completamente não relacionadas.

Inferência no modelo de Cox Estratificado

- ▶ A estratificação não cria complicações na estimação do vetor de parâmetros β .
- ▶ Uma função de verossimilhança parcial é construída para cada estrato e a estimação dos β 's é baseada na soma dos logaritmos das funções de verossimilhanças parciais, isto é, em:

$$\ell(\beta) = [\ell_1(\beta) + \cdots + \ell_m(\beta)],$$

com $\ell_j(\beta) = \log(L_j(\beta))$ obtida usando somente os dados dos indivíduos no j -ésimo estrato. $\ell(\beta)$ é maximizada com respeito a β .

- ▶ As propriedades assintóticas dos estimadores são obtidas a partir dos estimadores do modelo não estratificado (Colosimo, 1997).

Modelo de Cox Estratificado

- ▶ O modelo de Cox estratificado assume que as covariáveis atuam de modo similar na função de taxa de falha base de cada estrato. Ou seja, β é assumido ser comum para todos os estratos.
- ▶ Esta suposição pode ser testada usando, por exemplo, o teste da razão de verossimilhança cuja estatística de teste é dada por:

$$TRV = -2 \left[\ell(\hat{\beta}) - \sum_{j=1}^m \ell_j(\hat{\beta}_j) \right],$$

sendo $\ell(\hat{\beta})$ o logaritmo da função de verossimilhança parcial sob o modelo que assume β 's comuns e $\sum_{j=1}^m \ell_j(\hat{\beta}_j)$ o logaritmo da função de verossimilhança parcial sob o modelo que assume β 's distintos em cada estrato.

Modelo de Cox Estratificado

- ▶ Sob a hipótese nula e para grandes amostras, a estatística TRV segue distribuição qui-quadrado com $(m - 1)p$ graus de liberdade em que m é o número de estratos e p a dimensão do vetor β .
- ▶ O modelo estratificado deve somente ser usado caso realmente necessário, ou seja, na presença de violação da suposição de taxas de falha proporcionais.
- ▶ O uso desnecessário da estratificação acarreta em uma perda de eficiência das estimativas obtidas.

APLICAÇÃO: Sinusite em Pacientes HIV

- ▶ Neste estudo foram utilizadas informações provenientes de 91 pacientes HIV positivos e 21 HIV negativos.
- ▶ Pacientes forma acompanhados no período entre março de 1993 a fevereiro de 1995.
- ▶ Todos os pacientes incluídos no estudo foram encaminhados ao Centro de Treinamento e Referência em Doenças Infecto-parasitárias (CTR-DIP) da cidade de Belo Horizonte-MG, por pertencerem a grupos de comportamento de risco para adquirir o HIV ou por terem um exame elisa HIV positivo.
- ▶ Após a primeira consulta clínica, os pacientes foram encaminhados ao Serviço de Otorrinolaringologia da UFMG.

APLICAÇÃO: Sinusite em Pacientes HIV

- ▶ As doenças otorrinolaringológicas (ORL) avaliadas foram definidas com base nos estudos de prevalência destas manifestações na literatura em pacientes infectados pelo HIV. Os resultados sobre a infecção *sinusite* são apresentados a seguir.

- ▶ Quanto à infecção pelo HIV, os pacientes foram classificados de acordo ao seu Grupo de Risco, isto é,
 1. HIV soronegativo: não possuem o HIV;
 2. HIV soropositivo assintomático: possuem o vírus mas não desenvolveram o quadro clínico de AIDS e apresentam um perfil imunológico estável;
 3. Com ARC: apresentam baixa imunidade e outros indicadores clínicos que antecedem o quadro clínico de AIDS e,
 4. Com AIDS: já desenvolveram infecções oportunistas que definem esta doença, segundo os critérios do CDC de 1987.

APLICAÇÃO: Sinusite em Pacientes HIV

- ▶ A cada consulta, a classificação do paciente foi reavaliada e, sendo assim, esta covariável depende do tempo pois os pacientes podem mudar de classificação ao longo do estudo.
- ▶ A resposta de interesse foi o tempo, em dias, contado a partir da primeira consulta, até a ocorrência da sinusite. O objetivo foi identificar fatores de risco para esta manifestação.
- ▶ As contagens de CD4 e CD8 também foram medidas no início do estudo mas não foram incluídas nas análises devido a falta de registro de ambas para em torno de 37% dos pacientes.

APLICAÇÃO: Sinusite em Pacientes HIV

- Os fatores considerados na análise foram:

Idade do Paciente	medida em anos
Sexo do Paciente	0 - Masculino 1 - Feminino
Grupos de Risco	1 - Paciente HIV Soronegativo 2 - Paciente HIV Soropositivo Assintomático 3 - Paciente com ARC 4 - Paciente com AIDS
Atividade Sexual	1 - Homossexual 2 - Bissexual 3 - Heterossexual
Uso de Droga Injetável	1 - Sim 2 - Não
Uso de Cocaína por Aspiração	1 - Sim 2 - Não

APLICAÇÃO: Sinusite em Pacientes HIV

- ▶ Os resultados do ajuste do modelo de Cox incluindo a covariável Grupos de Risco, que depende do tempo foram:

Covariável	Coefficiente Estimado	Valor-p
Idade	-0,101	0,0210
Sexo	1,036	0,1700
HIV soropositivo assintomático	-0,308	0,8300
com ARC	3,074	0,0094
com AIDS	3,842	0,0015
Atividade Bissexual	0,344	0,6500
Heterossexual	-0,853	0,2900
Uso de droga	-0,152	0,9000
Aspira cocaína	1,454	0,3200

- ▶ As covariáveis idade e grupos de risco parecem ser não significativas. Removendo-se estas covariáveis gradativamente chegou-se no modelo final para a ocorrência de sinusite.

APLICAÇÃO: Sinusite em Pacientes HIV

- ▶ Para o modelo final ajustado tem-se as estimativas:

Covariável	Coefficiente de Regressão	Erro Padrão	Valor-p	Razão de Taxas de falha Estimado (I.C. 95%)
Idade	-0,077	0,0313	0,014	0,926 (0,871; 0,984)
HIV soropos. assint.	-0,730	1,0006	0,470	0,482 (0,067; 3,424)
com ARC	2,273	0,8371	0,006	9,705 (1,881; 50,064)
com AIDS	2,649	0,7897	<0,001	14,141 (3,008; 66,473)

- ▶ Idade e grupos de risco foram identificados como fatores de risco para a ocorrência de sinusite.
- ▶ O aumento de 10 anos na idade do paciente, o risco de desenvolver sinusite diminui em 54% ($1 - \exp\{-0,077 * 10\} \approx 0,54$). Ou seja, pacientes mais jovens estão mais sujeitos a esta infecção.

APLICAÇÃO: Sinusite em Pacientes HIV

- ▶ A taxa de ocorrência de sinusite para os pacientes HIV soropositivos assintomáticos não difere significativamente do grupo HIV soronegativo.
- ▶ A taxa de ocorrência do grupo com ARC é $\exp\{2,273\} = 9,7$ vezes àquela do grupo HIV soronegativo.
- ▶ Para o grupo com AIDS, o risco de desenvolver sinusite é 14,1 vezes o risco do grupo HIV soronegativo. A precisão das estimativas associadas a estas duas últimas razões de taxas de falha é bastante reduzida.
- ▶ Mais informações sobre esse estudo bem como os resultados encontrados para outras manifestações ORL em pacientes HIV positivos podem ser encontradas em Gonçalves (1995).

APLICAÇÃO: Hormônio do Crescimento

- ▶ O Hormônio do Crescimento (GH) é um importante agente do desenvolvimento humano e, quando sua deficiência é diagnosticada, são ministradas doses periódicas.
- ▶ A deficiência de GH pode se manifestar em graus variados e ter muitas causas diferentes. Em crianças com defasagem no crescimento, é feita uma avaliação para verificar se a baixa estatura é devida à má secreção/ação do hormônio do crescimento.
- ▶ Em caso positivo, são ministradas doses de hormônio sintetizado e o desenvolvimento de cada indivíduo é acompanhado em intervalos regulares de tempo.

APLICAÇÃO: Hormônio de Crescimento

- ▶ Um estudo foi realizado com 80 crianças participantes do Programa Hormonal do Crescimento da Secretaria de Saúde de Minas Gerais, diagnosticadas com deficiência do hormônio de crescimento. As mesmas tiveram um período de acompanhamento de no mínimo 19 meses em 31 de dezembro de 2002.
- ▶ O objetivo do estudo consistiu em identificar fatores determinantes do crescimento de crianças com deficiência do GH. Foram coletadas 16 covariáveis que foram consideradas potencialmente importantes para descrever o crescimento de crianças.

Covariáveis Medidas: Hormônio de Crescimento

Código	Variável	Descrição
V1	Sexo	1 se Masculino e 2 se Feminino
V2	Raça	1 se Branca e 2 se Negra
V3	Naturalidade	1 se Grande BH e 2 se Interior
V4	Tipo de parto	1 se Normal, 2 se Cesário e 3 se Fórceps
V5	Parto traumático	1 se Sim e 2 se Não
V6	Recém nascido	1 se AIG e 2 se PIG
V7	Apresentação	1 se Cefálica e 2 se Pélvica
V8	Renda	1: \leq 2SM, 2: 2 a 5SM, 3: 5 a 10SM e 4: $>$ 10SM
V9	Diagnóstico/origem	1 se Idiopático e 2 se Orgânico
V10	Grau de deficiência	1 se Isolado e 2 se DMHH
V11	Peso ao nascimento	Entre 1250 e 4240 g
V12	Velocidade	Entre 0,5 e 5 cm/ano
V13	Idade óssea	Entre 0,3 e 13 anos
V14	Idade cronológica/inicial	Entre 2 e 21 anos
V15	Altura inicial	Entre 71 e 154,8 cm
V16	Dose inicial	Entre 0,20 e 0,71 mml/kg

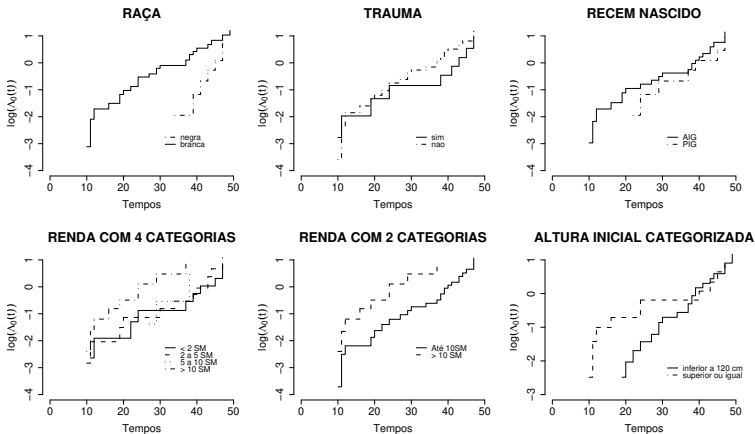
AIG = adequado para a idade gestacional e PIG = pequeno para a idade gestacional,
SM = salário mínimo, BH = Belo Horizonte, DMHH = Deficiência múltipla de hormônios hipofisários.

PLICAÇÃO: Hormônio de Crescimento

- ▶ Uma variável que norteia a decisão de alta clínica é a altura alvo. Ela é definida como a média da altura dos pais subtraída de 7 cm para meninas e somada de 7 cm para meninos. A variável resposta considerada é o tempo até a altura alvo ser atingida.
- ▶ Para seleção das covariáveis foi utilizado o modelo de Cox e a estratégia de construção de modelos proposta por Collett (1994).
- ▶ Após utilização dessa estratégia e discussões com os pesquisadores, as covariáveis selecionadas foram: raça, ocorrência de parto traumático, recém nascido, renda e altura inicial, com a presença das possíveis interações entre essas covariáveis.

APLICAÇÃO: Hormônio de Crescimento

- ▶ Ajustado o modelo de Cox com as covariáveis selecionadas, foi inicialmente avaliada a suposição de taxas de falha proporcionais por meio dos gráficos dos tempos *versus* $\log(\hat{\Lambda}_0(t))$.

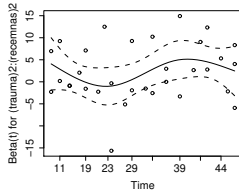
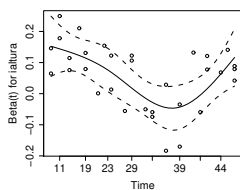
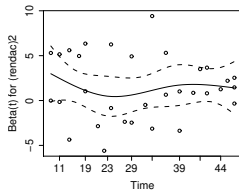
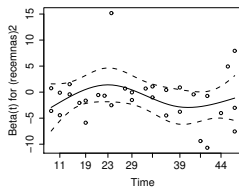
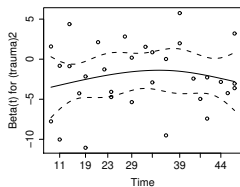
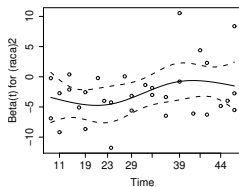


APLICAÇÃO: Hormônio de Crescimento

- ▶ As curvas para as covariáveis raça, trauma e recém-nascido não mostram situações de cruzamentos extremos.
- ▶ As curvas para as três categorias que envolvem renda igual ou inferior a 10SM possivelmente não diferem entre si. O gráfico desta covariável considerando as categorias $\leq 10SM$ e $>10SM$ também não sugere violação da suposição de taxas de falha proporcionais.
- ▶ O mesmo não pode ser concluído para a altura inicial uma vez que o gráfico correspondente a esta covariável, que foi categorizada em dois níveis de acordo com seu valor mediano, mostra que as curvas apresentam desvios quanto ao paralelismo que sugerem a violação desta suposição.

APLICAÇÃO: Hormônio de Crescimento

- ▶ Os gráficos dos resíduos padronizados de Schoenfeld a seguir,



APLICAÇÃO: Hormônio de Crescimento

Testes associados aos resíduos de Schoenfeld. :

Covariável	rho (ρ)	chisq	valor p
Raça	0,2686	2,6404	0,1042
Trauma	0,0606	0,1319	0,7164
Recém nascido	-0,0971	0,3042	0,5813
Renda	-0,0177	0,0126	0,9105
Altura Inicial	-0,4108	5,6068	0,0179
Trauma*Recém nascido	0,1450	0,7504	0,3863
GLOBAL	–	10,1775	0,1174

indica violação da suposição de taxas de falha proporcionais associado a covariável altura inicial estaria causando esta violação.

Resultados do Modelo de Cox Estratificado

- ▶ Considerando o modelo de Cox estratificado em que as crianças são consideradas de acordo com os estratos formados pelo valor mediano da altura inicial (< 120 cm e ≥ 120 cm) e assumindo que o vetor β é comum para esses dois estratos, tem-se o modelo:

$$\lambda_{ij}(t) = \lambda_{0_j}(t) \exp \{ \mathbf{x}'_{ij} \beta \}$$

para $j = 1, 2$ e $i = 1, \dots, n_j$, em que n_j é o número de crianças no j -ésimo estrato.

APLICAÇÃO: Hormônio de Crescimento

- ▶ As estimativas obtidas para este modelo foram:

Covariável	Coefficiente Estimado	Valor-p
Raça	-1,98	0,0062
Trauma	-1,35	0,0210
Recém nascido	-0,47	0,5800
Renda	1,19	0,0240
Trauma*recém nascido	1,26	0,2300

- ▶ Removendo os termos não significativos, chegou-se ao modelo final cujas estimativas obtidas foram:

Covariável	Coefficiente de Regressão	Erro Padrão	Valor-p	Razão de Taxas de falha Estimado(I.C. 95%)
Raça (2: negra)	-1,96	0,687	0,0043	0,141 (0,0367; 0,541)
Trauma (2: não)	-1,01	0,527	0,0560	0,365 (0,1299; 1,026)
Renda (2:>10SM)	1,03	0,505	0,0410	2,812 (1,0451; 7,565)

APLICAÇÃO: Hormônio de Crescimento

- ▶ Os resíduos padronizados de Schoenfeld e respectivos testes associados não mostraram evidência contra a hipótese de proporcionalidade das taxas de para cada um dos estratos formado pelas duas categorias consideradas para a covariável altura inicial.
- ▶ As seguintes interpretações podem ser obtidas a partir das estimativas fornecidas pelo ajuste do modelo de Cox estratificado.
 - i) a taxa de falha de uma criança branca atingir a altura alvo é $\exp\{1,96\} = 7,1$ (IC 95% 1,85 e 27,27);

APLICAÇÃO: Hormônio de Crescimento

- ii) Para crianças que tiveram parto traumático, a taxa de atingir a altura alvo é cerca de $\exp\{1,01\} = 2,74$ (IC 95% 0,97 e 7,7);
- ii) e, finalmente, a taxa de uma criança de família com renda superior a 10 SM atingir a altura alvo é 2,81 (IC 95% 1,04 e 7,5).

OUTRAS EXTENSÕES

▶ E se nada der certo?

- A suposição de taxas de falhas proporcionais não é satisfeita.
- Não podemos utilizar o modelo estratificado.
- Os modelos paramétricos não se ajustam aos dados.

Uma solução: O modelo aditivo (não-paramétrico) de Aalen
Capítulo 7 do Livro ASA.

▶ Como lidar com respostas altamente discretas?

- Poucos empates: aproximações para a verossimilhança parcial (Breslow e Efron).
- Muitos empates: aproximações ficam ruins. Temos que reconhecer a natureza discreta das respostas.

Uma solução: Modelos Discretos e Censura Intervalar
Capítulo 8 do Livro ASA.

OUTRAS EXTENSÕES

▶ **Como lidar com respostas dependentes: mesmo indivíduo ou conglomerado?**

- Modelos de efeitos aleatórios (Modelo de Fragilidade).
- Modelos do tipo GEE. Corrigindo a variância dos estimadores.

Capítulo 9 do Livro ASA.

▶ **Como lidar com respostas dependentes: eventos recorrentes?**

- Processo de Contagem (Processo de Poisson).
- O objetivo do estudo deve nortear a modelagem estatística.
 - Identificação de marcadores de risco: usar "gap times": modelos de fragilidade e GEE;
 - Caracterização de sistemas reparáveis: tempo global: modelagem de processos de contagem (lei de potência, mais utilizado).

Diferentes Desenhos e Objetivos

▶ Desenho do Estudo

- Assume-se que a amostra foi coletada em um esquema de "amostragem aleatória simples";
- O estudo longitudinal pode ser observado uma única vez: "Current status data".

▶ Objetivo do Estudo

- Usualmente o objetivo é identificar marcadores ou comparar grupos;
- Predições.
 - Predição Clínica;
 - Caracterização de produtos industriais: Testes de Degradação.

▶ Conceitos Importantes em Análise de Sobrevida

- "sob risco": assume-se que ao ser censurado a observação estava sob risco do evento.
- Censura não-informativa: T e C são independentes.