

Métodos Estatísticos Avançados em Epidemiologia

Análise de Sobrevivência - Modelo de Cox

Enrico A. Colosimo

Departamento de Estatística
Universidade Federal de Minas Gerais
<http://www.est.ufmg.br/~enricoc>

- ▶ Modelo de Tempos de Vida Acelerados ou Modelos Paramétricos.
- ▶ Modelo de Taxas de Falha Proporcionais ou Modelo Semiparamétrico de Cox.

Modelo de Regressão de Cox

O modelo de taxas de falha proporcionais, também chamado de modelo de Cox (Cox, 1972):

- abriu uma nova fase na modelagem de dados clínicos;
- é o mais utilizado na análise de dados de sobrevivência;
- permite incorporar facilmente covariáveis dependentes do tempo, que ocorrem com frequência em estudos clínicos.

MODELO DE REGRESSÃO DE COX

- ▶ Assume-se, nesse modelo, que os tempos t_i , $i = 1, \dots, n$, são independentes e que a taxa de falha (risco) tem a seguinte forma:

$$\lambda(t) = \lambda_0(t) \exp\{\beta_1 x_1 + \dots + \beta_p x_p\}.$$

- ▶ O componente não-paramétrico, $\lambda_0(t)$, não é especificado e é uma função não-negativa do tempo. Ele é usualmente chamado de função de base.
- ▶ O componente paramétrico $\exp\{x'\beta\}$ é o nosso interesse, em especial no vetor de parâmetros β .

MODELO DE REGRESSÃO DE COX

- ▶ O modelo é conhecido por ter taxas de falha proporcionais. Este fato é conveniente na sua interpretação.
- ▶ Ou seja, a razão das taxas de falha de dois indivíduos diferentes i e j é,

$$\frac{\lambda_i(t)}{\lambda_j(t)} = \frac{\lambda_0(t) \exp\{\mathbf{x}'_i \boldsymbol{\beta}\}}{\lambda_0(t) \exp\{\mathbf{x}'_j \boldsymbol{\beta}\}} = \exp\{\mathbf{x}'_i \boldsymbol{\beta} - \mathbf{x}'_j \boldsymbol{\beta}\},$$

que não depende do tempo.

- ▶ Assim, se um indivíduo no início do estudo tem um risco de morte igual a duas vezes o risco de um segundo indivíduo, então esta razão de riscos será a mesma para todo o período de acompanhamento.

ESTIMAÇÃO NO MODELO DE COX

- ▶ O modelo de regressão de Cox é caracterizado pelos coeficientes β 's, que medem os efeitos das covariáveis sobre a função de taxa de falha.
- ▶ Um método de estimação é necessário para se fazer inferências no modelo.
- ▶ O método de máxima verossimilhança não é adequado devido a presença do componente não-paramétrico $\lambda_0(t)$.

FUNÇÃO DE VEROSSIMILHANÇA PARCIAL

- ▶ Cox (1975) propôs então uma solução alternativa: verossimilhança parcial.
- ▶ Este método consiste em condicionar a construção da função de verossimilhança ao conhecimento da história passada de falhas e censuras
- ▶ Desta forma, elimina-se o componente não-paramétrico da função de verossimilhança.

Função de Verossimilhança Parcial

- ▶ A função de verossimilhança parcial é utilizada para fazer inferência no modelo de Cox.
- ▶ A função de verossimilhança parcial é então, formada pelo produto de todos os indivíduos da amostra:

$$L(\beta) = \prod_{i=1}^k \frac{\exp\{\mathbf{x}'_i\beta\}}{\sum_{j \in R(t_i)} \exp\{\mathbf{x}'_j\beta\}} = \prod_{i=1}^n \left(\frac{\exp\{\mathbf{x}'_i\beta\}}{\sum_{j \in R(t_i)} \exp\{\mathbf{x}'_j\beta\}} \right)^{\delta_i}$$

em que δ_i é o indicador de falha.

- ▶ Os valores de β que maximizam $L(\beta)$ são os estimadores de máxima verossimilhança parcial.

Inferência no Modelo de Cox

- i) **Teste Usual de Wald:** é geralmente o mais usado para testar hipóteses relativas a um único parâmetro, isto é,

$$H_0: \beta = 0$$

$$z = \frac{\hat{\beta}}{\widehat{EP}(\hat{\beta})} \sim N(0, 1)$$

Valores de $z > 1,96$ ou $z < -1,96$ indicam a rejeição de H_0 .

- **Teste da Razão de Verossimilhanças:** envolve a comparação de funções de verossimilhanças parciais: sem restrição e sob H_0 .

Interpretação dos Coeficientes Estimados

- ▶ O efeito das covariáveis é de acelerar ou desacelerar a função de risco.
- ▶ A propriedade de taxas proporcionais é extremamente útil na interpretação dos coeficientes estimados.
- ▶ A razão das taxas de falha de dois indivíduos i e j que têm os mesmos valores para as covariáveis com exceção da l -ésima, tem-se

$$\frac{\lambda_i(t)}{\lambda_j(t)} = \exp \left\{ \beta_l (x_{il} - x_{jl}) \right\},$$

que é interpretado como a razão de taxas de falha.

INTERPRETAÇÃO DOS PARÂMETROS

- ▶ Por exemplo, suponha que x_l seja uma covariável dicotômica indicando pacientes hipertensos. A taxa de morte entre os hipertensos é $\exp(\beta_l)$ vezes a taxa daqueles com pressão normal, mantida fixas as outras covariáveis.
- ▶ Uma interpretação similar é obtida para covariáveis contínuas. Se, por ex., o efeito de idade é significativo e $e^{\hat{\beta}} = 1,05$ para este termo, tem-se com o aumento de 1 ano na idade, que a taxa de morte aumenta em 5%.

ADEQUAÇÃO DO MODELO DE COX

- ▶ O modelo de Cox não se ajusta a qualquer situação clínica e, como qualquer outro modelo estatístico, requer o uso de técnicas para avaliar a sua adequação.
- ▶ A violação da suposição básica, que é a de taxas de falha proporcionais, pode acarretar em sérios vícios na estimação dos coeficientes do modelo (Struthers e Kalbfleisch, 1986).
- ▶ Diversos métodos para avaliar a adequação desse modelo encontram-se disponíveis na literatura e baseiam-se, essencialmente, nos resíduos de Schoenfeld.

Avaliação da Proporcionalidade dos Riscos

Para esse propósito encontram-se disponíveis na literatura, técnicas gráficas e testes estatísticos. Dentre eles:

■ Método gráfico descritivo:

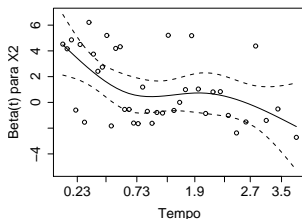
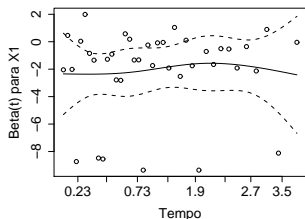
- dividir os dados em m estratos, usualmente de acordo com alguma covariável;
- em seguida, estima-se $\hat{\Lambda}_{0_j}(t)$ para cada estrato usando o estimador de Breslow;
- analisa-se as curvas do logaritmo de $\hat{\Lambda}_{0_j}(t) \times t$, ou $\log(t)$.
Curvas não paralelas significam desvios da suposição de riscos proporcionais. Situações extremas de violação da suposição ocorrem quando as curvas se cruzam.

Avaliação da Proporcionalidade das Taxas de Falha

- **Método com coeficiente dependente do tempo:** uma proposta adicional de análise da suposição de RP é fazer uso dos resíduos de Schoenfeld (1982).
- Existe um conjunto de resíduos para cada covariável;
- Usar o gráfico dos resíduos padronizados contra o tempo para cada covariável.

Avaliação da Proporcionalidade das Taxas de Falha

- ▶ Para auxiliar na detecção de uma possível falha da suposição de riscos proporcionais, uma curva suavizada, com bandas de confiança, é adicionada a este gráfico. A figura a seguir ilustra tais gráficos.



Avaliação da Proporcionalidade das Taxas de Falha

- **Medidas estatísticas e testes de hipóteses:** as técnicas gráficas envolve uma interpretação com carácter subjetivo. Testes de hipóteses podem auxiliar neste processo de decisão.
- ▶ O coeficiente de correlação de Pearson (ρ) entre os resíduos padronizados de Schoenfeld e $g(t)$ para cada covariável é uma dessas medidas. Valores de ρ próximos de zero mostram evidências em favor da suposição de RP.
- ▶ Um teste hipótese global de proporcionalidade de riscos.

Estudo de Caso: Aleitamento Materno

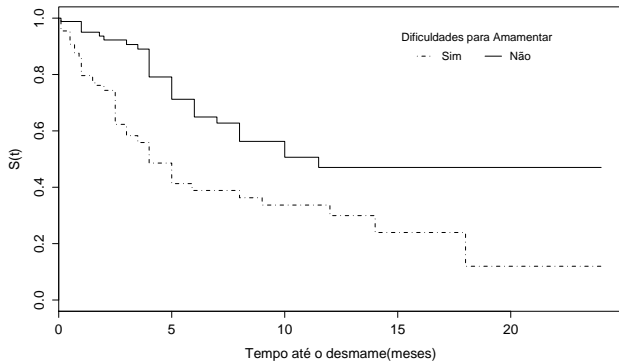
- Estudo realizado pelos Profs. Eugênio Goulart e Cláudia Lindgren do Departamento de Pediatria da UFMG.
- O estudo foi realizado no Centro de Saúde São Marcos, ambulatório municipal de BH, que atende, essencialmente a população de baixa renda.
- O objetivo principal do estudo é conhecer a prática do aleitamento materno de mães que utilizam este centro, assim como os possíveis fatores de risco ou de proteção para o desmame precoce.
- Um inquérito epidemiológico composto por questões demográficas e comportamentais foi aplicado a 150 mães de crianças menores de 2 anos de idade.
- A variável resposta de interesse foi o tempo máximo de aleitamento materno, ou seja, o tempo contado a partir do nascimento até o desmame completo da criança.

Estudo de Caso: Avaliação da Prática de Aleitamento Materno

Código	Covariável
V1	Experiência anterior de amamentação
V2	Número de filhos vivos (2)
V3	Conceito materno sobre o tempo ideal de amamentação (6 meses)
V4	Dificuldade para amamentar nos primeiros dias pós-parto
V5	Tipo de serviço em que realizou o (público) pré-natal
V6	Recebeu exclusivamente leite materno na maternidade
V7	A criança tem contato com o pai
V8	Renda per capita (1 SM)
V9	Peso ao nascimento (2,5 kgs)
V10	Tempo se separação mãe-filho pós-parto (6 horas)
V11	Permanência no berçário

ANÁLISE EXPLORATÓRIA: Prática de Aleitamento Materno

- Curvas de Kaplan-Meier
- Testes de Wilcoxon e log-rank.



TESTES DE IGUALDADE DE CURVAS DE SOBREVIVÊNCIA

Covariável	Testes (valor-p)	
	logrank	Wilcoxon
Experiência de amamentação	3,95 (0,047)	6,73 (0,010)
Número de filhos vivos	2,60 (0,107)	2,02 (0,155)
Tempo de amamentação	6,15 (0,013)	8,54 (0,004)
Dificuldade para amamentar	12,26 (0,001)	15,45 (< 0,001)
Tipo de serviço do pré-natal	1,38 (0,241)	1,09 (0,296)
Recebeu leite materno	7,47 (0,006)	6,31 (0,012)
Contato com o pai	1,84 (0,175)	0,90 (0,344)
Renda per capita	2,11 (0,146)	2,60 (0,107)
Peso ao nascimento	1,87 (0,171)	2,59 (0,108)
Separação mãe-filho	2,60 (0,107)	0,97 (0,325)
Permanência no berçário	2,93 (0,087)	0,90 (0,343)

Estudo de Caso: Avaliação da Prática de Aleitamento Materno

CRITÉRIO: Manter todas as covariáveis com pelo menos um valor-p $< 0,25$.

ESTRATÉGIA PARA A SELEÇÃO DE VARIÁVEIS (Collett, 1994)

- 1 Ajustar todos os modelos contendo uma única covariável. Incluir todas as covariáveis que forem significativas ao nível de 0,10.
- 2 As covariáveis significativas no passo 1 são então ajustadas conjuntamente. Ajustamos modelos reduzidos, excluindo uma única covariável.
- 3 Neste passo as covariáveis excluídas no passo 2 retornam ao modelo para confirmar que elas não são estatisticamente significantes.
- 4 Neste passo retornamos com as covariáveis excluídas no passo 1 para confirmar que elas não são estatisticamente significantes.
- 5 Ajustamos um modelo incluindo as covariáveis significativas no passo 4. Neste passo testamos se alguma delas podem ser retiradas do modelo.
- 6 Ajustamos o modelo final para os efeitos principais. Devemos verificar a possibilidade de inclusão de termos de interação.

PONTOS IMPORTANTES:

- Incluir as informações clínicas no processo de decisão;
- Evitar ser muito rigoroso ao testar cada nível individual de significância. É recomendado um valor próximo de 0,10.

MODELO DE REGRESSÃO DE COX

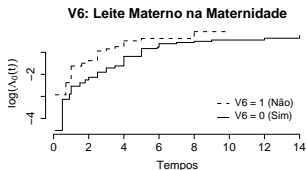
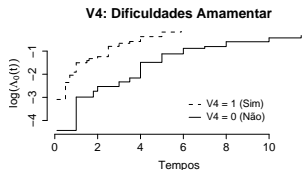
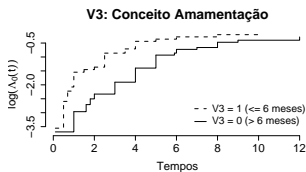
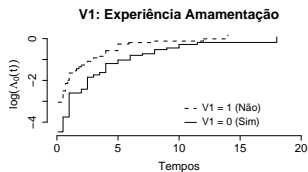
Nome	Modelo	-2 log L	Estatística	Valor-p
Passo 1	Nulo	560,628	—	—
	V1	556,958	3,670	0,0554
	V2	557,922	2,706	0,1000
	V3	554,920	5,708	0,0169
	V4	549,455	11,173	0,0008
	V5	559,402	1,226	0,2682
	V6	554,008	6,620	0,0101
	V7	558,420	2,208	0,1373
	V8	558,617	2,011	0,1562
	V9	558,597	2,031	0,1541
	V10	558,137	2,491	0,1145
Passo 2	V11	557,872	2,756	0,0969
	V1+V2+V3+V4+V6+V11	536,196	—	—
	V2+V3+V4+V6+V11	538,771	2,575	0,1085
	V1+V3+V4+V6+V11	536,196	0,000	1,0000
	V1+V2+V4+V6+V11	541,104	4,908	0,0267
	V1+V2+V3+V6+V11	543,629	7,433	0,0064
	V1+V2+V3+V4+V11	540,242	4,046	0,0443
	V1+V2+V3+V4+V6	536,346	0,150	0,6985
Passo 3	V3+V4+V6	539,433	—	—
	V3+V4+V6+V1	536,347	3,086	0,0790
	V3+V4+V6+V2	538,823	0,610	0,4348
	V3+V4+V6+V11	539,359	0,074	0,7856

MODELO DE REGRESSÃO DE COX

Passo 4	V3+V4+V6+V1	536,347	—	—
	V3+V4+V6+V1+V5	536,076	0,271	0,6027
	V3+V4+V6+V1+V7	534,108	2,239	0,1346
	V3+V4+V6+V1+V8	533,257	3,090	0,0788
	V3+V4+V6+V1+V9	535,012	1,335	0,2479
	V3+V4+V6+V1+V10	536,268	0,079	0,7787
Passo 5	V1+V3+V4+V6+V8	533,257	—	—
	V3+V4+V6+V8	534,492	1,235	0,2497
	V1+V4+V6+V8	538,540	5,283	0,0215
	V1+V3+V6+V8	542,136	8,879	0,0029
	V1+V3+V4+V8	538,172	4,915	0,0266
	V1+V3+V4+V6	536,347	3,090	0,0788
Passo 6	V1+V3+V4+V6	536,347	—	—
	V1+V3+V4+V6+V1*V3	535,922	0,425	0,5145
	V1+V3+V4+V6+V1*V4	536,123	0,224	0,6360
	V1+V3+V4+V6+V1*V6	536,005	0,342	0,5587
	V1+V3+V4+V6+V3*V4	535,136	1,211	0,2711
	V1+V3+V4+V6+V3*V6	534,673	1,674	0,1957
	V1+V3+V4+V6+V4*V6	535,873	0,474	0,4912
Modelo Final	V1+V3+V4+V6	536,347		

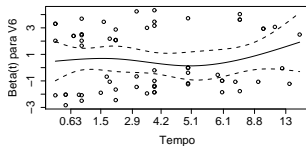
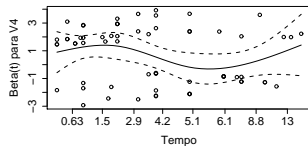
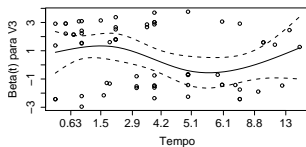
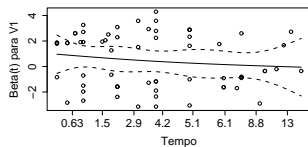
VERIFICAÇÃO DA ADEQUAÇÃO DO MODELO

Curvas de $\log \hat{H}(t)$ versus t para a Covariável Conceito Materno sobre o Tempo Ideal de Amamentação em dois Níveis (... ≤ 6 meses, — > 6 meses).



VERIFICAÇÃO DA ADEQUAÇÃO DO MODELO

Situações similares foram observadas para as outras curvas das covariáveis.



VERIFICAÇÃO DA ADEQUAÇÃO DO MODELO

Teste de proporcionalidade via resíduos de Schoenfeld

Covariável	rho (ρ)	χ^2	Valor-p
Exper. Amam. (V1)	-0,1098	0,754	0,385
Conc. Amam. (V3)	-0,1289	1,083	0,298
Dific. Amam. (V4)	-0,1047	0,653	0,419
Leite Excl. (V6)	0,0918	0,608	0,435
Global	-	3,232	0,520

RESULTADOS FINAIS E INTERPRETAÇÃO

Covariável	Estimativa	Valor-p	RTF	IC(RR, 95%)
Exper. Amam. (V1)	0,471	0,079	1,60	(0,94; 2,71)
Conc. Amam. (V3)	0,579	0,027	1,79	(1,07; 2,99)
Dific. Amam. (V4)	0,716	0,007	2,05	(1,22; 3,43)
Leite Excl. (V6)	0,578	0,029	1,78	(1,06; 2,99)

RESULTADOS FINAIS E INTERPRETAÇÃO

- A taxa de desmame precoce em mães que não tiveram experiência anterior de amamentação é 1,6 vezes a taxa das mães que tiveram essa experiência.
- A taxa de desmame precoce em mães que acreditam que o tempo ideal de amamentação é menor ou igual a 6 meses é aproximadamente 1,8 vezes a taxa das mães que acreditam que o tempo ideal de amamentação é superior a 6 meses.
- A taxa de desmame precoce em mães que apresentaram dificuldades de amamentar nos primeiros dias pós-parto é aproximadamente 2 vezes a taxa das mães que não apresentaram essas dificuldades.
- A taxa de desmame precoce em crianças que não receberam exclusivamente leite materno na maternidade é 1,8 vezes a taxa de desmame precoce em crianças que receberam exclusivamente o leite materno.

Observações

- Não existe estimador para o modelo de Cox quando em um nível de uma covariável categórica não observamos eventos.
- Exemplo: Na covariável Mitose nos dados de Melanoma não observamos o evento Metástase quando não ocorreu Mitose.
- Solução: ignorar esta covariável ou utilizar o estimador penalizado no pacote coxphf.

EXTENSÕES DO MODELO DE COX

- ▶ Algumas situações práticas envolvem covariáveis que são monitoradas durante o estudo, e seus valores podem mudar ao longo desse período. 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** e o modelo de Cox pode ser estendido para incorporá-las.
- ▶ Em outras situações a suposição de proporcionalidade das taxas de falha é violada e o modelo de Cox não é adequado. Modelos alternativos existem para enfrentar esta situação. Um deles é uma extensão do próprio modelo de Cox chamado de **modelo de taxas de falha proporcionais estratificado**.

Modelo com Covariáveis Dependentes do Tempo

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

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

- ▶ Definido desta forma, este modelo não é mais de taxas de falha proporcionais pois a razão das funções de risco 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.

- ▶ A interpretação dos coeficientes β do modelo deve considerar o tempo t .
- ▶ O ajuste do modelo é obtido, de forma natural, estendendo a função de verossimilhança parcial.

Modelo de Cox Estratificado

- ▶ O modelo de Cox não pode ser usado se a suposição de proporcionalidade das taxas de falha 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

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

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.

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. As derivadas são encontradas por meio da soma das derivadas obtidas para cada estrato e, então, $\ell(\beta)$ é maximizada com respeito a β .

Modelo de Cox Estratificado

- ▶ O modelo estratificado deve somente ser utilizado caso realmente necessário, ou seja, na presença de violação da suposição de riscos proporcionais. O uso desnecessário da estratificação acarreta em uma perda de eficiência das estimativas obtidas.
- ▶ Outro inconveniente do modelo de Cox estratificado é não permitir avaliar o efeito da covariável que gerou a estratificação.

UM ROTEIRO PARA A CONSTRUÇÃO DE MODELOS

- 1 Descrever o Problema: Importância e Objetivos.
- 2 Desenho do Estudo:
- 3 Exploração e Verificação da Consistência do Banco de Dados.
- 4 Análise Descritiva
- 5 Análise Univariada
 - Kaplan-Meier;
 - log-rank, Wilcoxon.

UM ROTEIRO PARA A CONSTRUÇÃO DE MODELOS

- 8 Regra Empírica: excluir covariáveis com valor- $p > 0,25$
- 9 Seleção de Modelos de Regressão
 - Utilizar algum tipo de stepwise.
 - Utilizar as covariáveis não excluídas no passo anterior;
 - Utilizar de preferência o Teste da RV;
 - Manter no modelo as covariáveis, já sabidamente, relevante em termos clínicos;
 - Investigar possíveis associações entre as covariáveis (colinearidade);
 - Investigar a possibilidade de categorizar covariáveis contínuas;
 - Obter um "Modelo Final" utilizando algum método de construção de modelos.
- 10 Incluir possíveis termos de interação.
- 11 Verificar a adequação do modelo ajustado.
- 12 Interpretar o modelo final apresentando intervalos de confiança para as quantidades de interesse.