



Tempo, tempo, tempo: regressão de Cox na ciência política

Rodrigo Lins
(linsprodrigo@gmail.com)

**Dossiê de Metodologia
2020**

Análise de sobrevivência

- Não se preocupa apenas com "se" um evento ocorre;
- Mas "quando" tal evento ocorre.

O que é o modelo de Cox?

- Modelo semiparamétrico;
- Permite a inclusão de covariáveis;
- Efeito das variáveis é multiplicativo;
- Não carrega pressuposto no que diz respeito à distribuição estatística para a função de risco basal.

O que é o modelo de Cox?

- Modelo de Riscos Proporcionais;
- Para quaisquer duas observações em qualquer ponto no tempo, a razão de seus riscos deve ser constante;
- A razão dos riscos de uma observação A para uma observação B deve ser o mesmo tanto em t_0 quanto em t_{100} , por exemplo;

Modelo de Cox

$$\log h(t) = a(t) + b_1x_1 + b_nx_n$$

- Lado esquerdo da equação é o logaritmo natural da função de risco;
- A função de risco basal $a(t)$ pode ser qualquer função de tempo;
- Em b_1x_1 , b representa o coeficiente e x o valor da variável independente 1.

Modelo de Cox

E se houver uma covariável que varie no tempo?

$$\log h(t) = a(t) + b_1x_1 + b_nx_n(t)$$

- $a(t)$ segue podendo ser qualquer função de tempo;
Variável X_1 segue constante no tempo;
- Variável X_n varia em função do tempo.

Resíduos

- Existem diferentes formas de testar o ajuste de uma regressão de Cox;
- No entanto, considerando que a proporcionalidade dos riscos é uma das características definidoras do modelo, é importante avaliar se tal pressuposto está sendo cumprido

Resíduos de Schoenfeld

- Apresenta a proporcionalidade de cada covariável definido para cada indivíduo

$$r_{ik} = \delta_i(x_{ik} - a_{ik})$$

- Ocorrência da falha para a observação i é simbolizada por δ_i ;
- Valor da variável específica se dá por x_{ik} ;
- Média ponderada das covariáveis em risco em um determinado tempo t é a_{ik} .

Resíduos de Schoenfeld

- A maneira mais intuitiva de compreender a análise é a de que os resíduos devem ser independentes do tempo;
- Se uma correlação entre os resíduos e o tempo não for estatisticamente significativa, isso quer dizer que os coeficientes não são dependentes do tempo e, assim, o pressuposto foi observado;
- Caso seja estatisticamente significativa, é indicação de que a proporcionalidade dos riscos não foi atendida.

Resíduos de Schoenfeld

- Em caso de não observância da proporcionalidade de risco, o primeiro passo é verificar se isso acontece de forma natural ou artificial;
- Caso a quebra de proporcionalidade seja real, duas abordagens são possíveis:
 - (1) buscar um novo modelo, como o modelo de Cox estratificado;
 - (2) adicionar um efeito de interação entre a variável que fere o pressuposto de proporcionalidade e alguma função de tempo (em geral, o logaritmo natural).

Pequeno glossário de sobrevivência

- Como os modelos de sobrevivência lidam com vocábulos que são pouco usuais a outras técnicas estatísticas, apresento um pequeno glossário;
- Os conceitos apresentados nessa seção são válidos não só para os modelos de Cox, mas para análise de sobrevivência em geral;
- Esse pequeno glossário, no entanto, não tem intenção de ser exaustivo.

Censurados/censored

- Censura à direita: acontece quando a observação chega ao fim do estudo sem experimentar o evento ($T > t^-$);
- Censura à esquerda: não sabemos quando o evento aconteceu, mas sabemos que ocorreu antes de um dado momento ($T < t+$);
- Censura intervalar: indica que o evento aconteceu dentro de um período de tempo, mas o pesquisador não sabe exatamente quando ($t^- < T < t+$).

Censurados/censored

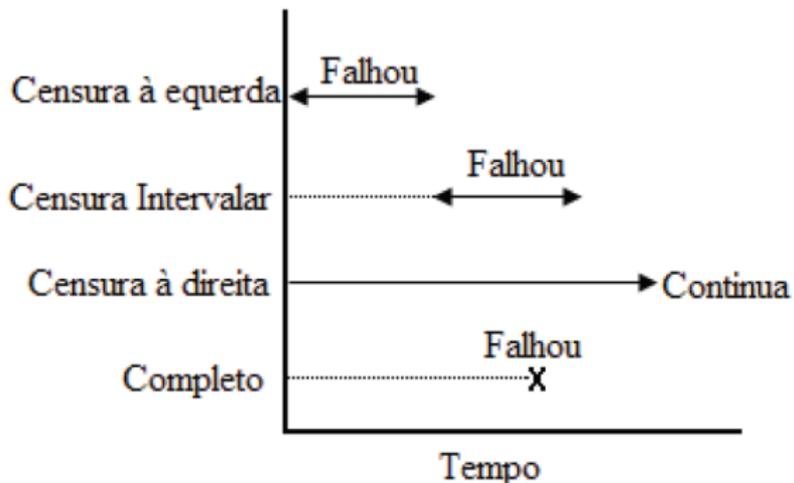


Figura: Lins et al. (2017)

Censurados/censored

- Tipo I: acontece quando o estudo é encerrado após um tempo pré-estabelecido e, neste momento, algumas observações ainda não tinham sofrido o evento de interesse;
- Tipo II: ocorre quando o estudo é encerrado depois que um número pré-estabelecido de eventos acontece;
- Aleatória: É aquela em que uma observação é retirada do estudo sem que a falha tenha acontecido.

Truncados/truncated

- À direita: Acontece quando são selecionados casos que sofreram o evento. Isso implica dizer que todas as observações sofreram o evento de interesse antes do limite final do estudo;
- À esquerda: Conta com os casos que sofreram o evento após o limite inferior da janela de estudo;
- Entrada atrasada: Casos que entraram no banco de dados uma vez que a pesquisa já havia iniciado;

Função de sobrevivência/survival function - $S(t)$

- É a probabilidade de que uma observação sobreviva mais tempo do que um algum período específico t .

$$S(t) = P(T > t)$$

Função de risco/hazard function – $h(t)$

- Apresenta o potencial de falha em cada unidade de tempo, considerando que a observação sobreviveu até o tempo sendo observado t .

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$$

Verossimilhança parcial/partial likelihood

- Os coeficientes dos modelos de Cox são estimados a partir de uma verossimilhança parcial;
- São consideradas apenas as informações das observações que estão sob risco a cada tempo t

$$PL = \prod_{j=1}^K L_j$$

- K é o número total de eventos na amostra;
- L_j denota a verossimilhança de falhar no tempo j , dado que a observação sobreviveu até este ponto.

Modelo de Cox tradicional

Tabela 1 – Banco de dados

ID	Início	Fim	Tempo	Status	X1	X2
A	0	30	30	0	1	1
B	0	27	27	1	1	0
C	1	15	14	1	0	0
D	2	30	28	0	1	1
E	3	15	12	1	0	1

Fonte: autor.

Modelo de Cox estendido

**Tabela 2 – Banco de dados
(Variáveis dependente do tempo)**

ID	Início	Fim	Status	X1	X2
A	0	11	0	1	100
A	11	22	1	1	120
B	0	15	0	0	40
B	12	23	0	0	42
B	23	30	0	0	47

Fonte: autor.

Modelo de Cox com eventos recorrentes

**Tabela 3 – Banco de dados
(Eventos recorrentes)**

ID	Início	Fim	Status
ARG	1946	1955	1
ARG	1958	1962	1
ARG	1963	1966	1
ARG	1973	1976	1
ARG	1983	2008	0

Fonte: autor.

Exemplo com quebras de regimes políticos Usando bando do V-Dem

Quadro 2 – Desenho do modelo

Modelo	Evento recorrente	Tempo	Evento	Unidade
Cox estendido	Sim	Ano	Quebra de regime	<u>País-ano</u>

Fonte: autor.

Exemplo com quebras de regimes políticos

Variáveis utilizadas

Quadro 3 – Variáveis

Variável	No banco
Regime político	democracy
Início	start
Fim	stop
Institucionalização partidária	v2xps_party
Presidencialismo	v2xnp_pres
PIB per capita (ln)	e_migdppln

Fonte: autor.

Aplicação no R

Pacotes e opções

```
## Pacotes e opções ####  
options(scipen = 999)  
install.packages("survival")  
library("survival")
```

Aplicação no R

Selecionando variáveis

Quadro 3 – Variáveis

Variável	No banco
Regime político	democracy
Início	start
Fim	stop
Institucionalização partidária	v2xps_party
Presidencialismo	v2xnp_pres
PIB per capita (ln)	e_migdppln

Fonte: autor.

Aplicação no R

Criando variáveis 'start' e 'stop'

```
## Criando a variável 'start' ##  
vdem$start <- as.numeric(as.factor(vdem$year))  
## Criando a variável 'stop' ##  
vdem$stop <- vdem$start + 1
```

Aplicação no R

Checando as primeiras linhas do banco

```
> head(vdem)
  country country_text_id country_id year democracy start stop v2x_polyarchy v2x_regime
207 Mexico             MEX           3 1995           0    73   74     0.514           2
208 Mexico             MEX           3 1996           0    74   75     0.528           2
209 Mexico             MEX           3 1997           0    75   76     0.546           2
210 Mexico             MEX           3 1998           0    76   77     0.596           2
211 Mexico             MEX           3 1999           0    77   78     0.606           2
212 Mexico             MEX           3 2000           0    78   79     0.619           2
v2xps_party v2xnp_pres e_migdppln e_regionpol
207      0.869      0.530      9.116           2
208      0.879      0.404      9.121           2
209      0.894      0.381      9.177           2
210      0.894      0.381      9.232           2
211      0.894      0.381      9.249           2
212      0.918      0.211      9.336           2
\ |
```

Aplicação no R

Regressão de Cox estendido

```
## Regressao de Cox ####  
modelo <- coxph(Surv(start, stop, democracy) ~ v2xps_party + v2xnp_pres + e_migdppcln,  
               data=vdem)  
modelo
```

Aplicação no R

Resultado

```
Call:
coxph(formula = Surv(start, stop, democracy) ~ v2xps_party +
      v2xnp_pres + e_migdppcln, data = vdem)

      coef exp(coef)    se(coef)      z      p
v2xps_party  1.5569    4.7442    3.5509  0.438 0.66106
v2xnp_pres  10.1272 25014.0175    3.1730  3.192 0.00141
e_migdppcln -0.6959    0.4986    1.0544 -0.660 0.50925

Likelihood ratio test=22.99 on 3 df, p=0.00004055
n= 609, number of events= 14
(1 observation deleted due to missingness)
```

Aplicação no R

Resultado

- O resultado vem com quatro informações: o coeficiente, o exponencial do coeficiente, erro padrão, a estatística z e o p-valor;
- O exponencial do coeficiente pode ser interpretado como a razão de risco:
 - Valores entre 0 e 1 indicam proteção; acima de 1, maior risco;
 - Pode ser lido como percentual ao se aplicar a simples fórmula $(1-\exp(\text{coef})) * 100$

Aplicação no R

Resíduos de Schoenfeld

```
## Resíduos de Schoenfeld ####  
res.modelo <- cox.zph(modelo, transform = "identity")  
res.modelo
```

Aplicação no R

Resíduos de Schoenfeld (output)

```
> res.modelo
      chisq df    p
v2xps_party 0.0137 1 0.91
v2xnp_pres  0.9335 1 0.33
e_migdppc1n 0.2364 1 0.63
GLOBAL      2.0072 3 0.57
```

Aplicação no R

Resíduos de Schoenfeld (output)

- É possível notar que nenhuma variável apresentou uma correlação significativa entre seus coeficientes e o tempo;
- Assim, não se exclui a hipótese nula que afirma não haver relação entre tempo e variáveis.

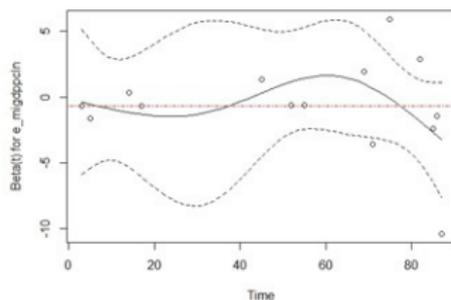
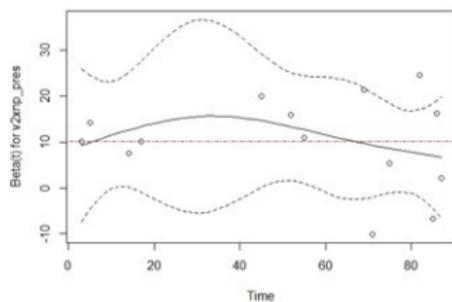
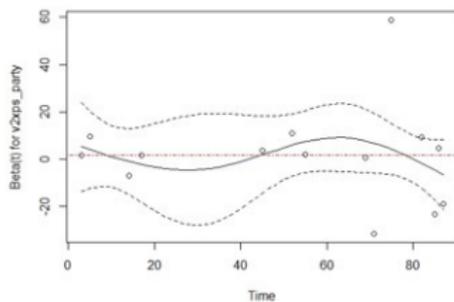
Aplicação no R

Resíduos de Schoenfeld (plot)

```
plot(res.modelo[1])
abline(h=modelo$coefficients[1], lty=4, col=2) ## Institucioalização (partidos) ##
plot(res.modelo[2])
abline(h=modelo$coefficients[2], lty=4, col=2) ## Presidencialismo ##
plot(res.modelo[3])
abline(h=modelo$coefficients[3], lty=4, col=2) ## Ln PIB per capita ##
```

Aplicação no R

Resíduos de Schoenfeld (plot)



Aplicação no R

Resíduos de Schoenfeld (plot)

- Se a linha, assim como a linha suavizada, estiver dentro do intervalo de confiança, é possível atestar que o pressuposto foi observado