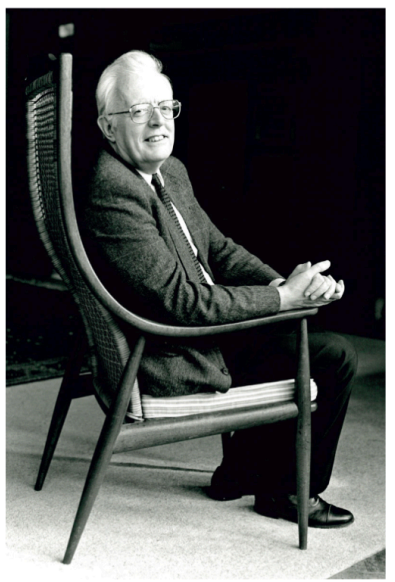
Regresión de Riesgos Proporcionales de Cox

Jorge Mario Estrada Alvarez PhD. MSc. FETP

Sir David R. Cox (15 Julio 1924 — 18 Enero 2022)



“David relató que trabajó en el artículo durante algunos años, y que la idea crucial que condujo al análisis de verosimilitud parcial le llegó mientras estaba en cama con fiebre, y que al recuperarse le costó cierto esfuerzo recordar el argumento preciso.

http://doi.org/10.1098/rsbm.2023.0052

Introducción

- El modelo de **Cox (1972)** permite evaluar la relación entre múltiples predictores y la función hazard o tasa de fallo instantanea, considerando datos **censurados**.
- Es el modelo más utilizado en **Investigacion epidemiologica, clinica y salud publica**, por su flexibilidad y bajo número de supuestos paramétricos.
- Se considera **semi-paramétrico**, ya que no especifica la forma del riesgo basal $\mathrm{h}_0(\mathrm{t})$.

Fundamento del modelo

El modelo relaciona el **riesgo instantáneo** $h(t|x)$ con los predictores $x_1, x_2, \ldots, x_p$:

D. Cox penso en como comparar las tasas de evento entre dos tratamientos (segun entrevista dada por él)

$$HR(t) = \frac{h_1(t)}{h_0(t)},$$

estable en el tiempo entonces $HR(t) \equiv HR$

De aqui surge el supuesto de riesgo proporcionales en el tiempo.

$$\log(HR) = \log(\frac{h_1(t)}{h_0(t)})$$

El modelo de D. Cox (1972)

$$\log \frac{h(t \mid \mathbf{x})}{h_0(t)} = \beta_1 x_1 + \ldots + \beta_p x_p$$

$$\log h(t \mid \mathbf{x}) - \log h_0(t) = \beta_1 x_1 + \ldots + \beta_p x_p$$

$$\log h(t \mid \mathbf{x}) = \log h_0(t) + \beta_1 x_1 + \ldots + \beta_p x_p$$

$$h(t|x) = h_0(t) \cdot e^{\beta_1 x_1 + \cdots + \beta_p x_p}$$

- $h_0(t)$: riesgo instantáneo del grupo de referencia (base).
- e^{β_i}: cambio relativo en el riesgo instantáneo asociado a una unidad de cambio en (x_i).

Comparación con Otros Modelos

Tipo de modelo	Relación	Variable dependiente
Lineal	$E[Y\|x] = \beta_0 + \beta_1 x_1 + \dots$	Valor esperado (media)
Logístico	$\log \frac{p(x)}{1-p(x)} = \beta_0 + \beta_1 x_1 + \dots$	Probabilidad (odds)
Cox (PH)	$\log h(t\|x) = \log h_0(t) + \beta_1 x_1 + \dots + \beta_p x_p$	**Riesgo instantáneo**

Supuesto de Riesgos Proporcionales

- El **hazard ratio (HR)** entre dos individuos es **constante en el tiempo**:

$$\mathrm{HR}(t) = \frac{h(t|x_1)}{h(t|x_2)} = \exp[\beta(x_1 - x_2)] = \mathrm{HR}$$

- Esto implica que las **curvas de riesgo** de ambos grupos son proporcionales, y sus **curvas de supervivencia no se cruzan**.

Tipos de Modelos Relacionados

Modelo	**Riesgo basal** $h_0(t)$	**Tipo**	**Observaciones**
Exponencial	Constante	Paramétrico	Riesgo constante
Weibull	Monótono (↑ o ↓)	Paramétrico	Requiere forma del riesgo
Cox	No especificado	**Semi-paramétrico**	Flexible, robusto ante mala especificación

Estimación en el Modelo de Cox

- Los coeficientes β se estiman **sin necesidad de conocer** $h_0(t)$.
- La estimación usa la **verosimilitud parcial** (partial likelihood), considerando solo los momentos donde ocurre un evento.
- Luego se puede estimar $H_0(t)$ (riesgo acumulado) y $S_0(t)$ (supervivencia basal) para obtener curvas ajustadas.

Ventajas del Modelo de Cox

- Maneja **censura a la derecha** eficientemente.
- No requiere suponer una forma funcional para $h_0(t)$.
- Permite ajustar por múltiples covariables.

Trabajaremos con los datos del ensayo PBC

Consideramos una cohorte de 312 participantes en un ensayo clínico controlado con placebo de D-penicilamina (DPCA) para la cirrosis biliar primaria (CBP) (Dickson et al., 1989).

Los investigadores recopilaron datos sobre la edad, niveles de bilirrubina sérica, presencia de hepatomegalia, edema de miembros inferiores y venas varicosas visibles en el pecho y los hombros (arañas vasculares).

Algunas descriptivas de sobrevida de los participantes

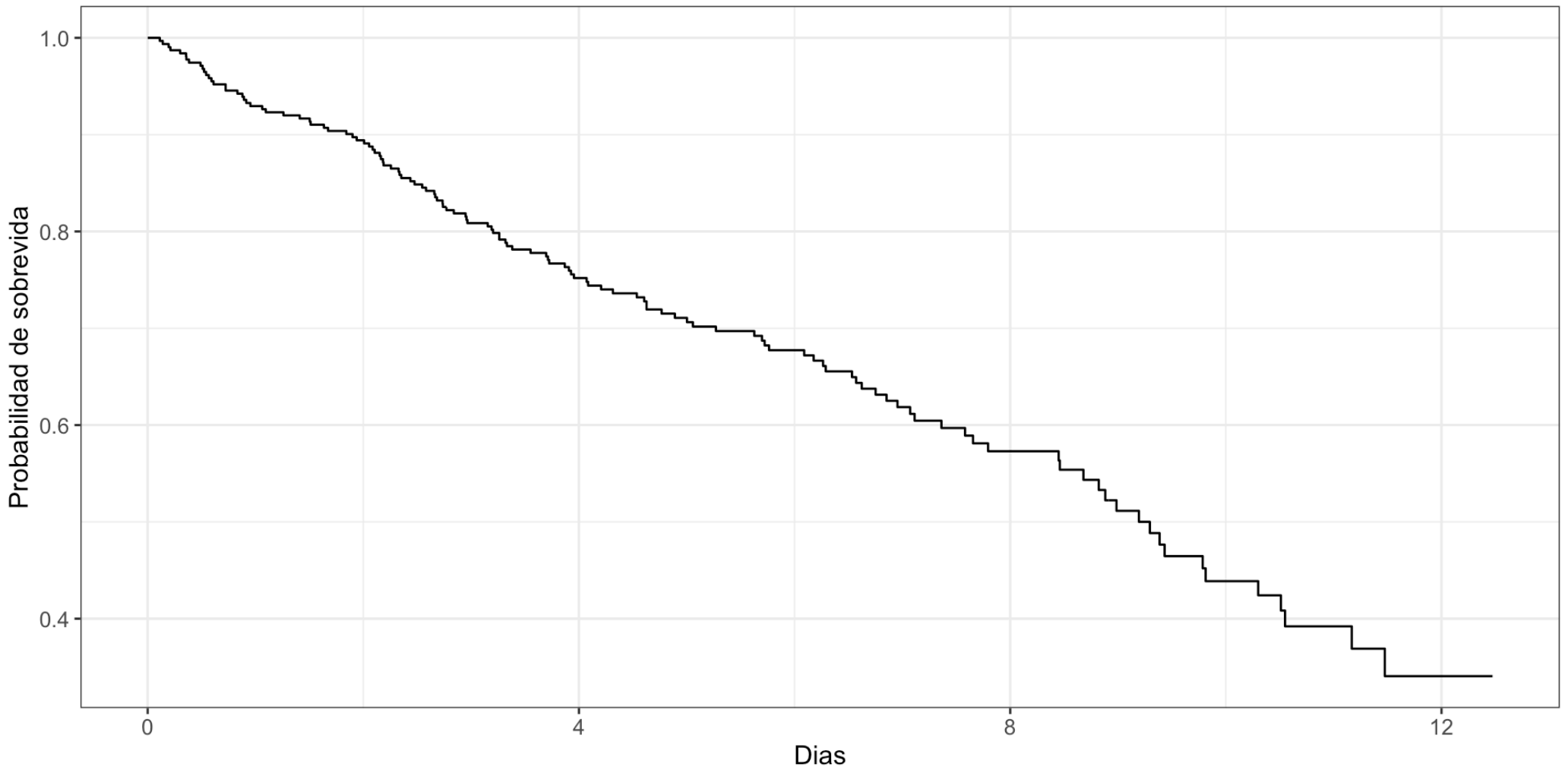
```r
model <- survfit(Surv(years, status) ~ 1, data = pbc)
model
```

```
Call: survfit(formula = Surv(years, status) ~ 1, data = pbc)

      n events median 0.95LCL 0.95UCL
[1,] 312    125    9.3    8.45    10.5
```

```r
summary(model, times = c(1,3,6,9,12,15))
```

```
Call: survfit(formula = Surv(years, status) ~ 1, data = pbc)

 time n.risk n.event survival std.err lower 95% CI upper 95% CI
    1    290      22    0.929  0.0145        0.902        0.958
    3    240      37    0.809  0.0224        0.766        0.854
    6    130      33    0.677  0.0283        0.624        0.735
    9     47      22    0.511  0.0387        0.441        0.593
   12      8      11    0.341  0.0528        0.251        0.461
```

Algunas descriptivas de sobrevida de los participantes



Interpretacion del modelo: Variable dicotomica (1/0)

$$\log h(t \mid \mathbf{x}) = \log h_0(t) + \beta_1 x_1$$

$$\log\left(\frac{h(t|x=1)}{h(t|x=0)}\right) = h_0(t) + \beta_1(1) - h_0(t) - \beta_1(0)$$

$$= \beta_1$$

$$\log(HR) = \beta_1$$

$$e^{\log(HR)} = e^{\beta_1}$$

$$HR = e^{\beta_1}$$

```r
model_cox <- coxph(Surv(years, status) ~ rx, data = pbc)
model_cox
```

```
Call:
coxph(formula = Surv(years, status) ~ rx, data = pbc)

       coef exp(coef) se(coef)      z     p
rx -0.05722   0.94438  0.17916 -0.319 0.749

Likelihood ratio test=0.1  on 1 df, p=0.7494
n= 312, number of events= 125
```

Interpretacion del modelo: Variable politomica

Como la variables asume distintos valores se debe crear dummy

$$X = C_1, C_2, C_3$$

$$X_1 : C_1 = 1, \text{resto } 0$$

$$X_2 : C_2 = 1, \text{resto } 0$$

$$X_3 : C_3 = 1, \text{resto } 0$$

$$\log h(t \mid \mathbf{x}) = \log h_0(t) + \beta_2 x_2 + \beta_3 x_3$$

$$\log\left(\frac{h(t|x=2)}{h(t|x=1)}\right) = h_0(t) + \beta_2(1) + \beta_3(0) - h_0(t) -$$

$$= \beta_2$$

$$\log(HR) = \beta_2$$

$$e^{\log(HR)} = e^{\beta_2}$$

$$HR = e^{\beta_2}$$

```
Call:
coxph(formula = Surv(years, status) ~ factor(histol), data = pbc)

                   coef exp(coef) se(coef)     z      p
factor(histol)2  1.607     4.988    1.031 1.559 0.1191
factor(histol)3  2.150     8.581    1.012 2.124 0.0337
factor(histol)4  3.063    21.387    1.009 3.036 0.0024

Likelihood ratio test=52.74  on 3 df, p=2.085e-11
n= 312, number of events= 125
```

Interpretacion del modelo: Variable continua

Miremos para un cambio en un año en la edad:

$$\log\left(\frac{h(t|x = 36)}{h(t|x = 35)}\right) = h_0(t) + \beta_1(36) - h_0(t) - \beta_1(35)$$

$$= \beta_1(36) - \beta_1(35)$$

$$\log(HR) = \beta_1(36) - \beta_1(35)$$

$$e^{\log(HR)} = e^{\beta_1(36) - \beta_1(35)}$$

$$HR = e^{\beta_1(36-35)}$$

$$HR = e^{\beta_1}$$

```
Call:
coxph(formula = Surv(years, status) ~ age, data = pbc)

        coef exp(coef) se(coef)     z        p
age 0.039995  1.040806 0.008811 4.539 5.65e-06

Likelihood ratio test=20.51  on 1 df, p=5.947e-06
n= 312, number of events= 125
```

Podría calcularse para un cambio mayor en la edad, ejemplo: 5 años

$$\log\left(\frac{h(t|x=35)}{h(t|x=30)}\right) = h_0(t) + \beta_1(35) - h_0(t) - \beta_1(30)$$

$$= \beta_1(35) - \beta_1(30)$$

$$\log(HR) = \beta_1(35) - \beta_1(30)$$

$$e^{\log(HR)} = e^{\beta_1(35)-\beta_1(30)}$$

$$HR = e^{\beta_1(35-30)}$$

$$HR = e^{\beta_1 k} = (e^{\beta_1})^k$$

$$HR = e^{0.039995\times 5} = 1.22$$

Modelo Multivariado

```
Call:
coxph(formula = Surv(years, status) ~ rx + sex + bilirubin, data =

              coef exp(coef) se(coef)      z       p
rx         -0.1914    0.8258   0.1835 -1.043  0.2971
sex        -0.5795    0.5602   0.2397 -2.418  0.0156
bilirubin   0.1548    1.1675   0.0135 11.470 <2e-16

Likelihood ratio test=91.05  on 3 df, p=< 2.2e-16
n= 312, number of events= 125
```

```
Call:
coxph(formula = Surv(years, status) ~ rx + sex + bilirubin +
    age, data = pbc)

                coef exp(coef)  se(coef)      z        p
rx         -0.069523  0.932839  0.186771 -0.372   0.7097
sex        -0.418928  0.657751  0.242596 -1.727   0.0842
bilirubin   0.150663  1.162605  0.013324 11.308  < 2e-16
age         0.037108  1.037805  0.009148  4.056 4.99e-05

Likelihood ratio test=107.5  on 4 df, p=< 2.2e-16
n= 312, number of events= 125
```

Test de razon de verosimilitud entre modelos

- Se requiere comparar dos modelos anidado uno en el otro

```
Analysis of Deviance Table
 Cox model: response is  Surv(years, status)
 Model 1: ~ rx + sex + bilirubin + age
 Model 2: ~ rx + sex + bilirubin
   loglik  Chisq Df Pr(>|Chi|)
1 -586.23
2 -594.44 16.428  1  5.054e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Confusión - Interacción - Predicción

Confusión

En un modelo para efecto que tiene bilirrubina sobre el hazard de muerte seria:

```
Call:
coxph(formula = Surv(years, status) ~ bilirubin, data = pbc)

             coef exp(coef) se(coef)     z      p
bilirubin 0.14892   1.16058  0.01301 11.44 <2e-16

Likelihood ratio test=84.65  on 1 df, p=< 2.2e-16
n= 312, number of events= 125
```

Algunos sintomas estan asociados con el hazard de muerte, y se requiere evaluar el efecto de la bilirrubina controlando otras variables:

```
Call:
coxph(formula = Surv(years, status) ~ bilirubin + edema + hepatom +
    spiders, data = pbc)

              coef exp(coef) se(coef)     z        p
bilirubin  0.11186   1.11836  0.01487 7.524 5.31e-14
edema      0.75581   2.12934  0.22215 3.402 0.000668
hepatomYes 0.71818   2.05070  0.21185 3.390 0.000699
spidersYes 0.38905   1.47558  0.19475 1.998 0.045754

Likelihood ratio test=119  on 4 df, p=< 2.2e-16
n= 312, number of events= 125
```

Interacción

- Ejemplo: proponemos estimar si hay una heterogeneidad del efecto de tratamiento según la presencia de hepatomegalia

$$\log h(t|x) = \log h_0(t) - 0.178\text{rx} + 1.14\text{hepatom} + 0.095\text{hepatom} \times \text{rx}$$

```
Call:
coxph(formula = Surv(years, status) ~ rx + hepatom + rx:hepatom,
    data = pbc)

                   coef exp(coef) se(coef)      z        p
rx              -0.17856   0.83648  0.33216 -0.538    0.591
hepatomYes       1.14808   3.15214  0.26591  4.318 1.58e-05
rx:hepatomYes    0.09501   1.09968  0.39490  0.241    0.810

Likelihood ratio test=40.56  on 3 df, p=8.126e-09
n= 312, number of events= 125
```

Calculemos efectos heterogeneos

$$\log h(t|x) = \log h_0(t) - 0.178\text{rx} + 1.14\text{hepatom} + 0.095\text{hepatom} \times \text{rx}$$

$$HR_0 = e^{\beta_1} = e^{-0.178} = 0.8369424$$

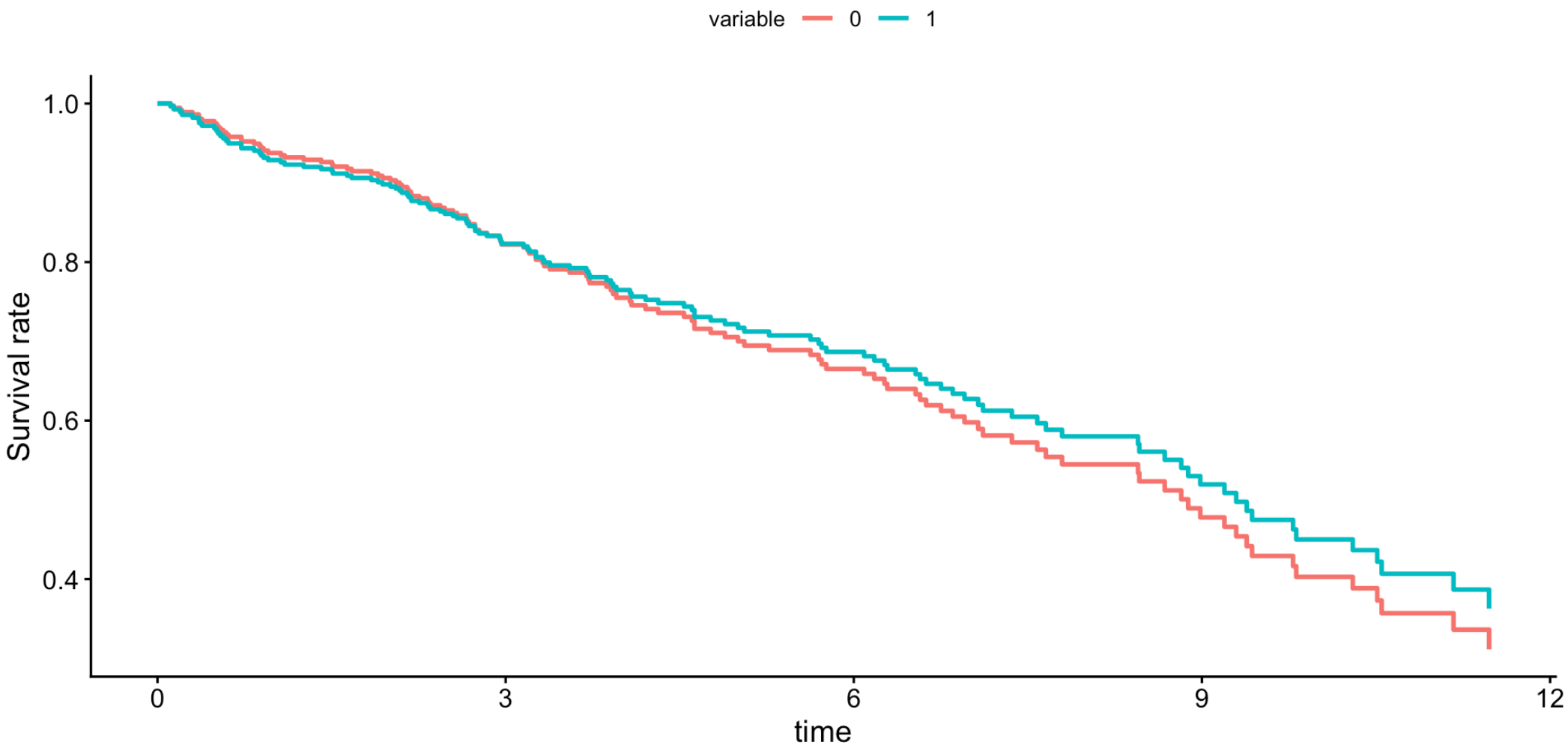
$$HR_1 = e^{\beta_1+\beta_3} = e^{-0.178+0.095} = 0.9203511$$

Predicción

- Segun el modelo de Cox, el mismo no estima $h_0(t)$ que es el hazard de base, solo se estima los β_i
- Sin embargo hay una relación entre $S(t)$ y $h_0(t)$ que permite estimar, pero requiere integración.
- No estima valores puntuales de sobrevida para un sujeto, sino la curva completa ajustada por una combinación de covariables.

Curvas de sobrevida ajustadas

```
1 ggadjustedcurves(model_cox, method = "average", fun = "pct", var
```



Gracias