

Análisis de Supervivencia

Modelo de Riesgos Proporcionales de Cox

Sergio M. Nava Muñoz

2025-06-01

Table of contents

1	Introducción	2
2	Fundamento Matemático del Modelo	2
3	Suposición de Riesgos Proporcionales (PH)	2
3.1	Ejemplo concreto	2
3.2	Ejemplo cuando no se cumple la proporcionalidad	3
4	Características del Modelo	4
5	Interpretación de los coeficientes	4
5.1	Razón de Riesgo (HR)	4
5.2	Ejemplos de interpretación de HR	4
6	Ejemplo computacional: Modelo de Cox PH	5
6.1	Datos de Leucemia	6
6.2	Curvas de supervivencia ajustadas	10
7	Evaluación de la Suposición PH	10
7.1	Evaluación de proporcionalidad: Curvas log(-log)	11
7.2	Residuos de Schoenfeld	12
7.3	Test global de PH con <code>cox.zph()</code>	12
7.4	¿Qué evalúa?	12
7.5	Ejemplo en R	13
8	Soluciones a Violaciones de PH	13
9	Conclusiones	13
10	Referencias	13

1 Introducción

- El modelo de riesgos proporcionales de Cox es el más utilizado para analizar datos de supervivencia con múltiples covariables.
 - Modelo de regresión **semiparamétrico**.
 - Basado en Klein & Moeschberger (2003).
-

2 Fundamento Matemático del Modelo

$$h(t|X) = h_0(t) \cdot \exp(\beta^T X)$$

- $h(t|X)$: función de riesgo condicional.
 - $h_0(t)$: riesgo base (no especificado).
 - X : vector de covariables.
 - β : coeficientes a estimar.
-

3 Suposición de Riesgos Proporcionales (PH)

- La razón de riesgos entre dos individuos:

$$\frac{h(t|X_1)}{h(t|X_2)} = \exp(\beta^T (X_1 - X_2))$$

compara el riesgo de dos individuos con distintos valores de covariables X_1 y X_2 , en el mismo tiempo t .

- No depende del tiempo \rightarrow **proporcionalidad**.
- Si las funciones de riesgo se cruzan, la suposición PH se viola.

3.1 Ejemplo concreto

Supongamos un modelo con dos covariables:

- **tratamiento**: 0 = control, 1 = experimental
- **edad**: en años

Y los coeficientes estimados son:

- $\beta = (-0.5, 0.04)$

```
# Vectores de covariables para dos individuos
X1 <- c(tratamiento = 1, edad = 60)
X2 <- c(tratamiento = 0, edad = 60)

# Coeficientes estimados del modelo
```

```

beta <- c(-0.5, 0.04)

# Cálculo de la razón de riesgos
HR <- exp(sum(beta * (X1 - X2)))
HR

[1] 0.6065307

```

3.2 Ejemplo cuando no se cumple la proporcionalidad

```

library(survival)
library(survminer)

# Cargar datos
data(lung)

# Re-codificar status: 2 = evento, 1 = censura → convertir a 1/0
lung$status2 <- ifelse(lung$status == 2, 1, 0)

# Verificar niveles de sexo
table(lung$sex) # 1 = hombre, 2 = mujer

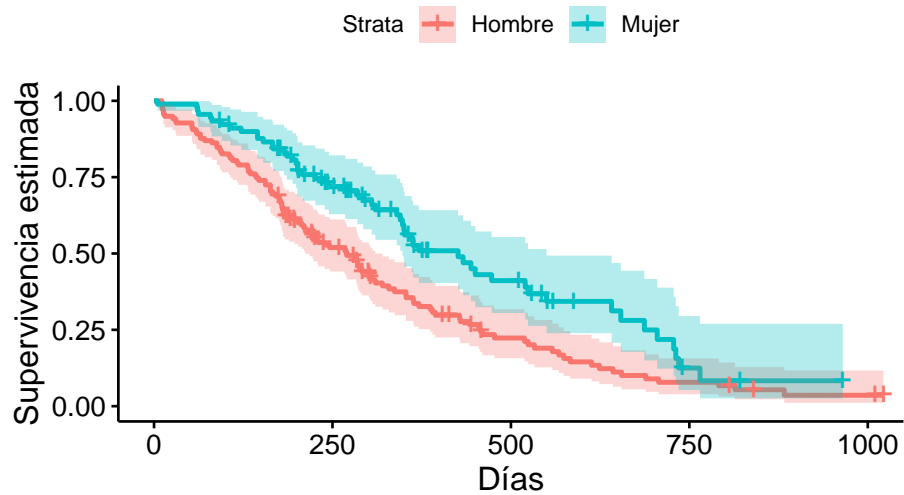
  1    2
138  90

fit_km <- survfit(Surv(time, status2) ~ factor(sex), data = lung)

ggsurvplot(fit_km,
            data = lung,
            conf.int = TRUE,
            legend.labs = c("Hombre", "Mujer"),
            xlab = "Días", ylab = "Supervivencia estimada",
            title = "Curvas Kaplan-Meier por sexo")

```

Curvas Kaplan–Meier por sexo



Si las curvas se cruzan, puede indicar que la suposición de riesgos proporcionales se viola.

4 Características del Modelo

- **Semiparamétrico:** no se asume forma para $h_0(t)$.
 - Estimación de coeficientes mediante **verosimilitud parcial**.
 - Robusto y flexible ante diferentes tipos de datos censurados.
-

5 Interpretación de los coeficientes

5.1 Razón de Riesgo (HR)

- $HR = \exp(\beta)$.
- $HR > 1$: mayor riesgo.
- $HR < 1$: efecto protector.
- $HR = 1$: no afecta el riesgo.

5.2 Ejemplos de interpretación de HR

Table 1: Ejemplos de interpretación de razones de riesgo (HR)

Variable	Coficiente	HR	Interpretación
tratamiento (experimental vs control)	-0.510	0.600	40% menos riesgo en grupo experimental
edad (años)	0.050	1.051	Cada año adicional → +5.1% riesgo
karno (índice Karnofsky)	-0.032	0.969	Cada punto adicional → -3.2% riesgo

6 Ejemplo computacional: Modelo de Cox PH

Analizaremos el modelo de Cox PH usando una base de datos de remisión en pacientes con leucemia (Freireich et al. (1963)).

- Dos grupos con 21 pacientes cada uno: **Grupo 1** (Tratamiento), **Grupo 2** (Placebo)
- Covariable adicional: **log WBC**, un importante predictor pronóstico en leucemia.

Pregunta de investigación: Comparar la supervivencia entre grupos, **ajustando por posibles efectos de confusión o interacción con log WBC.**

Table 2: Leukemia Remission Data: Group 1 (Treatment) vs Group 2 (Placebo)

t.Grupo.1.	log.WBC..Grupo.1.	t.Grupo.2.	log.WBC..Grupo.2.
6	2.31	1	2.80
6	4.06	1	5.00
6	3.28	2	4.91
7	4.43	2	4.48
10	2.96	3	4.01
13	2.88	4	4.36
16	3.60	4	2.42
22	2.32	5	3.49
23	2.57	5	3.97
6+	3.20	8	3.52
9+	2.80	8	3.05
10+	2.70	8	2.32
11+	2.60	8	3.26

17+	2.16	11	3.49
19+	2.05	11	2.12
20+	2.01	12	1.50
25+	1.78	12	3.06
32+	2.20	15	2.30
32+	2.53	17	2.95
34+	1.47	22	2.73
35+	1.45	23	1.97

6.1 Datos de Leucemia

```
leukemia <- data.frame(
  time = c(6, 6, 6, 7, 10, 13, 16, 22, 23, 6, 9, 10, 11, 17, 19, 20, 25, 32, 32, 34, 35,
           1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23),
  status = c(rep(1, 9), rep(0, 12), rep(1, 21)),
  group = c(rep(1, 21), rep(2, 21)),
  logWBC = c(2.31, 4.06, 3.28, 4.43, 2.96, 2.88, 3.60, 2.32, 2.57,
             3.20, 2.80, 2.70, 2.60, 2.16, 2.05, 2.01, 1.78, 2.20, 2.53, 1.47, 1.45,
             2.80, 5.00, 4.91, 4.48, 4.01, 4.36, 2.42, 3.49, 3.97,
             3.52, 3.05, 2.32, 3.26, 3.49, 2.12, 1.50, 3.06, 2.30, 2.95, 2.73, 1.97)
)
```

Explicación: Este conjunto representa a pacientes con leucemia, donde **time** es el tiempo hasta recaída (o censura), **status** indica si ocurrió el evento (1) o no (0), **group** es el tratamiento (1 = tratado, 2 = placebo) y **logWBC** es el logaritmo del conteo de glóbulos blancos.

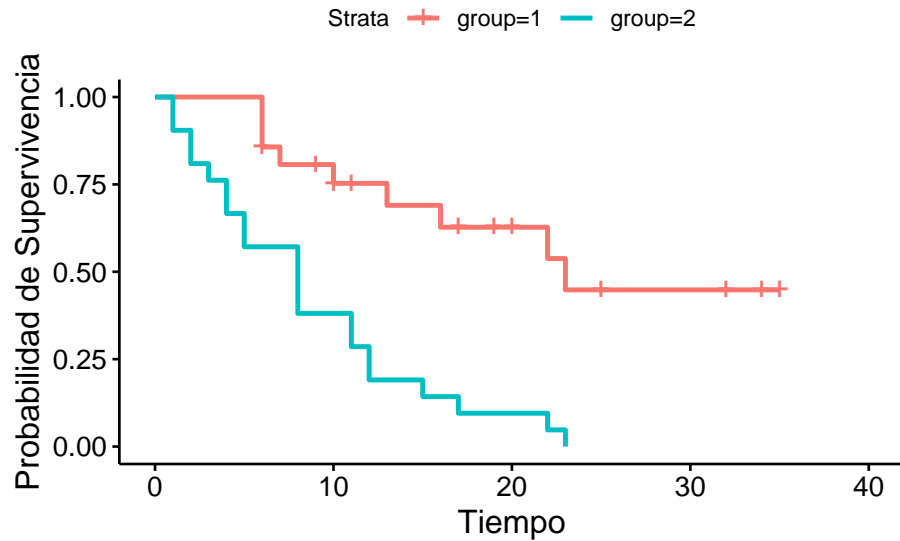
```
summary(leukemia)
```

time	status	group	logWBC
Min. : 1.00	Min. :0.0000	Min. :1.0	Min. :1.450
1st Qu.: 6.00	1st Qu.:0.0000	1st Qu.:1.0	1st Qu.:2.303
Median :10.50	Median :1.0000	Median :1.5	Median :2.800
Mean :12.88	Mean :0.7143	Mean :1.5	Mean :2.930
3rd Qu.:18.50	3rd Qu.:1.0000	3rd Qu.:2.0	3rd Qu.:3.490
Max. :35.00	Max. :1.0000	Max. :2.0	Max. :5.000

```
table(leukemia$status,leukemia$group)
```

1	2	
0	12	0
1	9	21

```
fit <- survfit(Surv(time, status) ~ group, data = leukemia)
ggsurvplot(fit, xlab = "Tiempo", ylab = "Probabilidad de Supervivencia")
```



```
cox_model <- coxph(Surv(time, status) ~ factor(group) + logWBC, data = leukemia)
summary(cox_model)
```

Call:

```
coxph(formula = Surv(time, status) ~ factor(group) + logWBC,
      data = leukemia)
```

n= 42, number of events= 30

	coef	exp(coef)	se(coef)	z	Pr(> z)
factor(group)2	1.3861	3.9991	0.4248	3.263	0.0011 **
logWBC	1.6909	5.4243	0.3359	5.034	4.8e-07 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
factor(group)2	3.999	0.2501	1.739	9.195
logWBC	5.424	0.1844	2.808	10.478

Concordance= 0.852 (se = 0.04)

Likelihood ratio test= 46.71 on 2 df, p=7e-11

Wald test = 33.6 on 2 df, p=5e-08

Score (logrank) test = 46.07 on 2 df, p=1e-10

Explicación: El modelo de Cox ajusta el riesgo de recaída según el tratamiento y logWBC. `factor(group)` permite comparar placebo contra tratamiento. La salida incluye coeficientes `beta`, errores estándar, valor z y p-valor.

Interpretación del modelo (HR e IC)

```
exp(coef(cox_model))      # Razón de riesgo (HR)
```

```
factor(group)2      logWBC
      3.999125      5.424308
```

```
exp(confint(cox_model))   # Intervalo de confianza del 95% para HR
```

```
                2.5 %    97.5 %
factor(group)2 1.739306  9.195048
logWBC         2.808199 10.477578
```

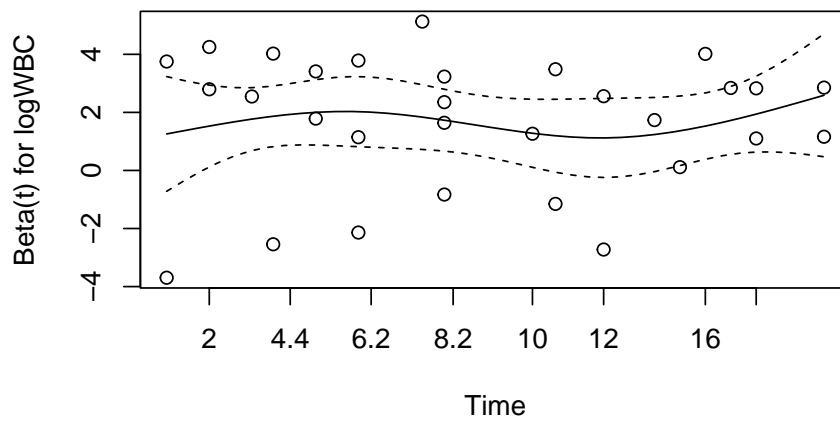
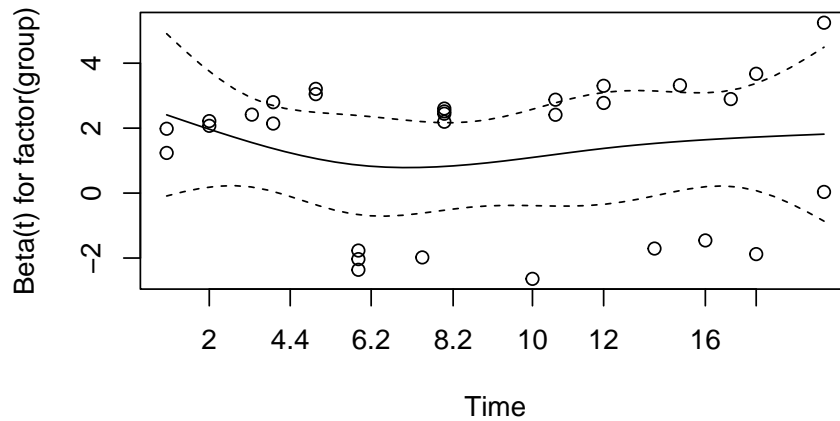
Explicación: Aquí se presentan los coeficientes exponenciados, que se interpretan como razones de riesgo (HR). Un $HR > 1$ indica mayor riesgo relativo; $HR < 1$ sugiere efecto protector. El intervalo de confianza permite evaluar si el efecto es estadísticamente significativo (no debe incluir 1).

Evaluación de la suposición de riesgos proporcionales

```
test_ph <- cox.zph(cox_model)
test_ph
```

```
              chisq df    p
factor(group) 8.27e-05  1 0.99
logWBC        4.00e-02  1 0.84
GLOBAL        4.02e-02  2 0.98
```

```
plot(test_ph)
```

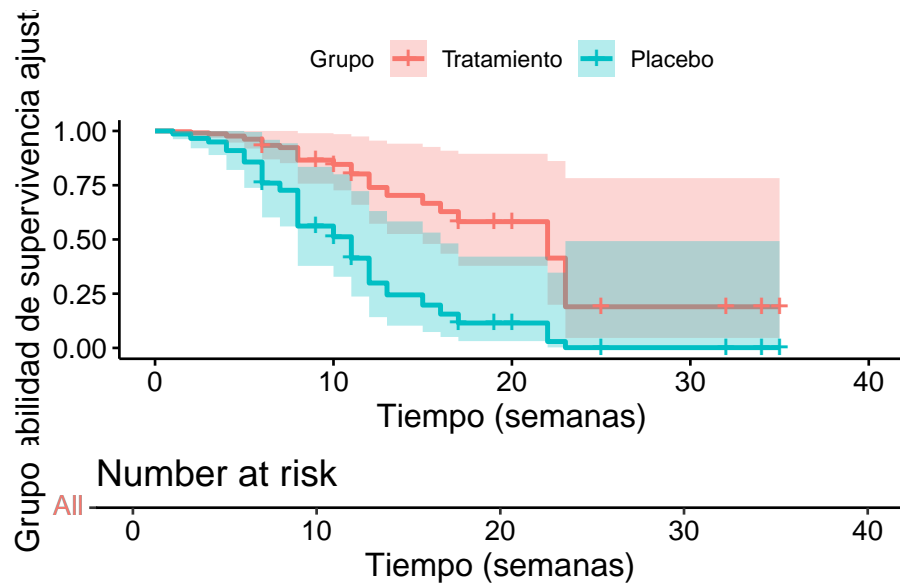



i Hipótesis evaluadas con `cox.zph()`

- **Hipótesis nula (H_0):** la covariable cumple la suposición de riesgos proporcionales (el efecto de la covariable es constante en el tiempo).
- **Hipótesis alternativa (H_1):** la covariable no cumple la suposición de riesgos proporcionales (el efecto cambia con el tiempo).

Un p-valor menor a 0.05 indica que se rechaza la hipótesis nula, sugiriendo que la suposición de riesgos proporcionales **no se cumple** para esa covariable. El gráfico asociado muestra residuos de Schoenfeld.

6.2 Curvas de supervivencia ajustadas



Explicación: Se grafican las curvas de supervivencia estimadas para un paciente con nivel medio de logWBC, comparando tratamiento vs placebo. El `risk.table` muestra cuántos pacientes permanecen en riesgo a lo largo del tiempo.

7 Evaluación de la Suposición PH

1. Gráficas:

- Curvas log(-log) paralelas.
- Gráficas de Schoenfeld residuals.

2. Pruebas formales:

- Test global de PH (e.g., `cox.zph` en R).

3. Extensión con covariables dependientes del tiempo:

- Incluir interacción con función del tiempo.

7.1 Evaluación de proporcionalidad: Curvas log(-log)

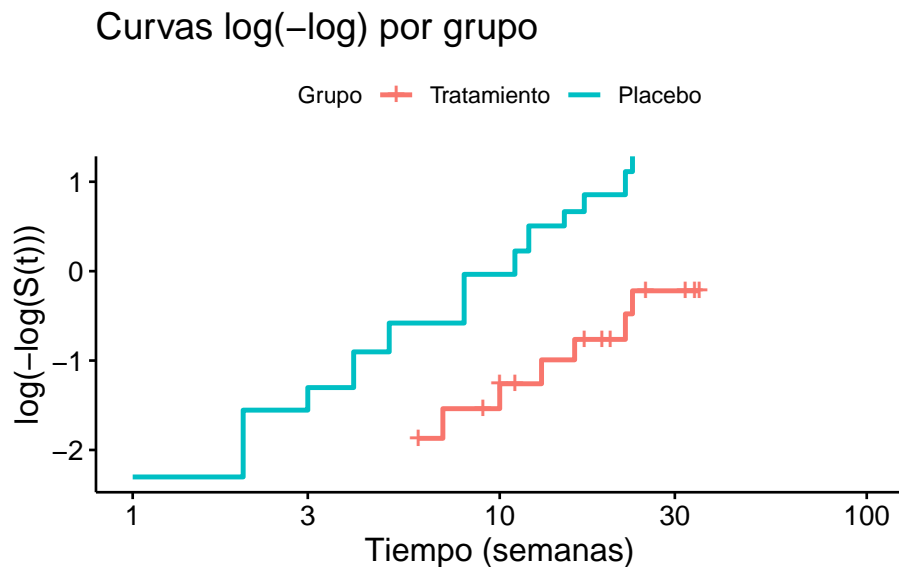
- Otra forma gráfica de verificar la suposición de riesgos proporcionales.
- Se grafican curvas:

$$\log\{-\log[\hat{S}(t)]\}$$

- Se esperan **curvas paralelas** si la suposición PH se cumple.
- Se usa típicamente para comparar grupos categóricos (ej. tratamiento vs placebo).

```
fit_loglog <- survfit(Surv(time, status) ~ factor(group), data = leukemia)

ggsurvplot(
  fit_loglog,
  fun = "cloglog", # complementary log-log
  data = leukemia,
  legend.labs = c("Tratamiento", "Placebo"),
  legend.title = "Grupo",
  xlab = "Tiempo (semanas)",
  ylab = "log(-log(S(t)))",
  title = "Curvas log(-log) por grupo"
)
```



- Si las curvas **son aproximadamente paralelas**, la suposición PH se considera razonable.
 - Si **se cruzan** o divergen significativamente, puede haber violación.
-

7.2 Residuos de Schoenfeld

- Son residuos calculados **solo en los tiempos de evento**.
- Se usan para evaluar si el **efecto de una covariable es constante en el tiempo** (suposición de riesgos proporcionales).

Definición:

$$\text{Residuo de Schoenfeld} = X_{\text{observado}} - \mathbb{E}[X \mid \text{riesgo}]$$

- Donde X es una covariable.
- Se calcula en cada tiempo de evento.

Interpretación gráfica

- Los residuos se grafican contra el tiempo.
- Se ajusta una curva de suavizado (por ejemplo, LOESS):
 - Si la curva es **horizontal**, el efecto de la covariable es constante.
 - Si tiene **pendiente creciente o decreciente**, sugiere que el efecto **cambia con el tiempo** → **violación de la suposición PH**.

Ejemplo de interpretación:

- Línea plana: suposición PH razonable
 - Tendencia ascendente: el efecto crece con el tiempo
 - Tendencia descendente: el efecto decrece con el tiempo
-

7.3 Test global de PH con `cox.zph()`

7.4 ¿Qué evalúa?

- Contrasta la **hipótesis nula** de que el efecto de cada covariable es constante en el tiempo.
- Evalúa la **proporcionalidad de riesgos** para cada covariable y de forma global.

Hipótesis:

- H_0 : la covariable cumple la suposición PH (efecto constante en el tiempo)
- H_1 : el efecto varía con el tiempo

Un **p-valor bajo** (< 0.05) indica que **se viola** la suposición PH para esa covariable o globalmente.

7.5 Ejemplo en R

```
test_ph <- cox.zph(cox_model)
test_ph
```

	chisq	df	p
factor(group)	8.27e-05	1	0.99
logWBC	4.00e-02	1	0.84
GLOBAL	4.02e-02	2	0.98

Esto muestra una tabla con:

- Una fila por covariable y una para el test global
- Estadístico chi-cuadrado y p-valor asociado

Interpretación:

- Si el test global es significativo, el modelo **no cumple** con PH en general.
 - Si solo una covariable tiene $p < 0.05$, considerar transformaciones o modelos extendidos.
-

8 Soluciones a Violaciones de PH

- **Modelo estratificado:**
 - $h_0(t)$ específico por estrato.
 - **Modelo extendido:**
 - Términos dependientes del tiempo.
-

9 Conclusiones

- Modelo robusto y versátil.
- Permite ajustar múltiples covariables.
- Ideal para datos censurados.
- Evaluar la suposición PH es crucial.

10 Referencias

Freireich, E. J., Karon, M., Frei, E., Holland, J. F., Taylor, R., Hananian, J., Selawry, O., Hoogstraten, B., Wolman, I. J., Abir, E., Sawitsky, A., Lee, S., Mills, S. D., Burgert, E. O. J., Spurr, C. L., Patterson, R. B., Ebaugh, F. G., James, G. W., & Moon, J. H. (1963). The effect of 6-mercaptopurine on the duration of steroid-induced remissions in acute leukemia: A model for evaluation of antileukemic agents. *Blood*, 21(6), 699–716.

Klein, J. P., & Moeschberger, M. L. (2003). *Survival analysis: Techniques for censored and truncated data* (2nd ed.). Springer.