

Análisis de Supervivencia

Modelo de Riesgos Proporcionales de Cox

Sergio M. Nava Muñoz

2025-06-01

Table of contents

0.1	Introducción	1
0.2	Fundamento Matemático del Modelo	1
0.3	Suposición de Riesgos Proporcionales (PH)	2
0.4	Características del Modelo	4
0.5	Interpretación de los coeficientes	4
0.6	Ejemplo computacional: Modelo de Cox PH	5
0.7	Evaluación de la Suposición PH	12
0.8	Soluciones a Violaciones de PH	14
0.9	Introducción a la estratificación	15
0.10	Dataset ejemplo: <code>lung</code> (paquete <code>survival</code>)	15
0.11	Evaluación inicial: ¿sex viola PH?	17
0.12	Estratificación por <code>sex</code>	21
0.13	Comparación de modelos: sin estratificación vs con estratificación	22
0.14	Comparación visual de curvas ajustadas	24
0.15	Conclusiones	24
0.16	Referencias	24

0.1 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).
- El artículo original se puede ver en Cox (1972).

0.2 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.

0.3 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.

0.3.1 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
```

0.3.2 Ejemplo cuando no se cumple la proporcionalidad

```
library(survival)
library(survminer)
```

```
# Cargar datos
data(cancer, package = "survival")

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

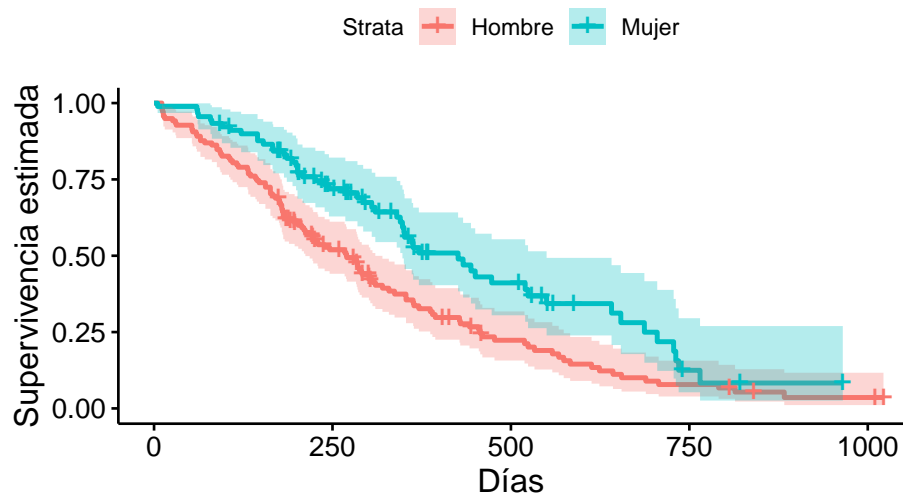
# Verificar niveles de sexo
table(lung$sex) # 1 = hombre, 2 = mujer

      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.

0.4 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.

0.5 Interpretación de los coeficientes

0.5.1 Razón de Riesgo (HR)

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

0.5.2 Ejemplos de interpretación de HR

```
# Crear tabla de ejemplos
ejemplos_hr <- data.frame(
  Variable = c("tratamiento (experimental vs control)",
               "edad (años)",
               "karno (índice Karnofsky)"),
  Coeficiente = c(-0.510, 0.050, -0.032),
  HR = round(exp(c(-0.510, 0.050, -0.032)), 3),
  Interpretación = c("40% menos riesgo en grupo experimental",
                    "Cada año adicional → +5.1% riesgo",
                    "Cada punto adicional → -3.2% riesgo")
)

# Mostrar tabla
knitr::kable(ejemplos_hr,
              caption = "Ejemplos de interpretación de razones de riesgo (HR)")
```

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

Variable	Coeficiente	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

0.6 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** (*log white blood cell count* o logaritmo del recuento de leucocitos), un importante predictor pronóstico en leucemia.

```

library(knitr)

# Crear los datos
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)
)

# Crear columnas de tiempo con o sin "+"
leukemia$t_weeks <- ifelse(leukemia$status == 0, paste0(leukemia$time, "+"), as.character)

# Separar por grupo
grupo1 <- subset(leukemia, group == 1, select = c(t_weeks, logWBC))
grupo2 <- subset(leukemia, group == 2, select = c(t_weeks, logWBC))

# Mostrar resultados
#print(grupo1)
#print(grupo2)

# Combinar las tablas lado a lado
tabla <- data.frame(
  "t(Grupo 1)" = grupo1$t_weeks,
  "log WBC (Grupo 1)" = grupo1$logWBC,
  "t(Grupo 2)" = grupo2$t_weeks,
  "log WBC (Grupo 2)" = grupo2$logWBC
)

# Mostrar con kable
kable(tabla, align = "c", caption = "Leukemia Remission Data: Group 1 (Treatment) vs Group 2 (Placebo)")

```

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

0.6.1 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

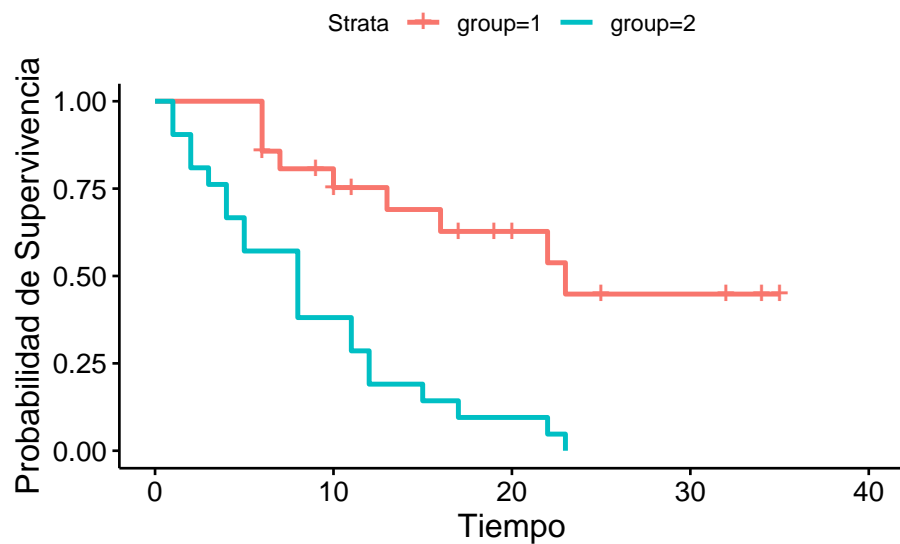
```
table(leukemia$status,leukemia$group)
```

```

      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

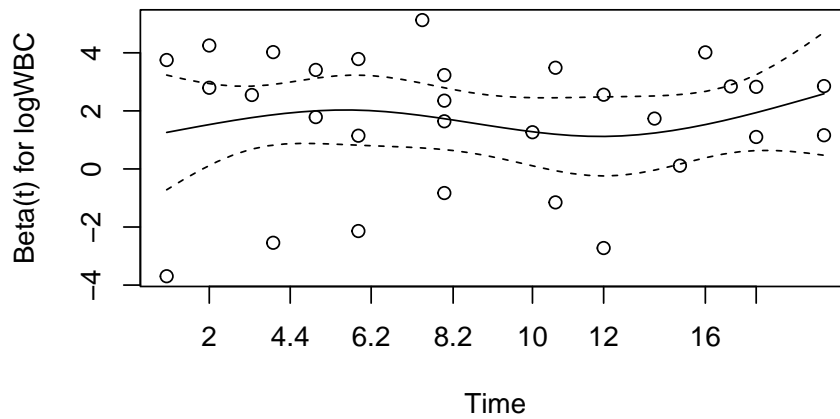
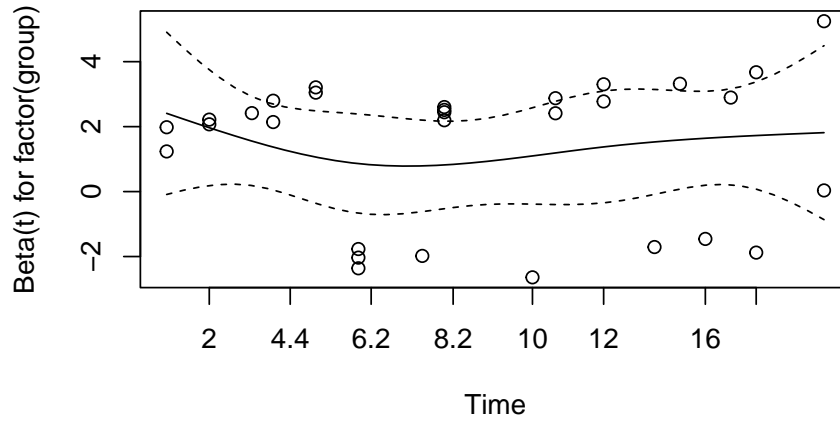
```
test_ph <- cox.zph(cox_model)
test_ph
```

```

              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.

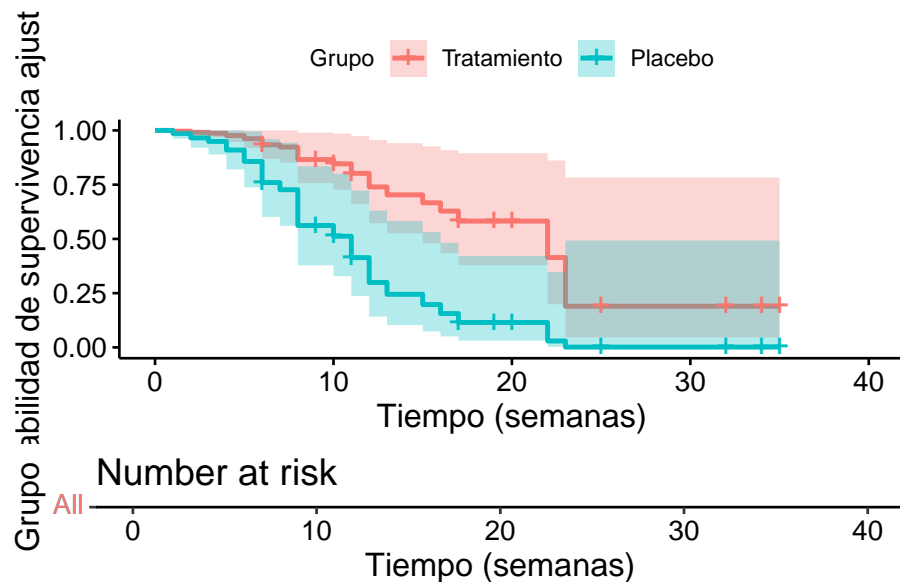
El gráfico asociado muestra residuos de Schoenfeld.

0.6.2 Curvas de supervivencia ajustadas

```
fit <- survfit(cox_model, newdata = data.frame(logWBC = median(leukemia$logWBC), group = c("Tratamiento", "Placebo")))

# Graficar curvas

ggsurvplot(
  fit,
  data = leukemia,
  legend.title = "Grupo",
  legend.labs = c("Tratamiento", "Placebo"),
  xlab = "Tiempo (semanas)",
  ylab = "Probabilidad de supervivencia ajustada",
  risk.table = TRUE
)
```



Explicación: Se grafican las curvas de supervivencia estimadas para un pa-

ciente con nivel medio de logWBC, comparando tratamiento vs placebo. El `risk.table` muestra cuántos pacientes permanecen en riesgo a lo largo del tiempo.

0.7 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.
-

0.7.1 Evaluación de proporcionalidad: Curvas log(-log)

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

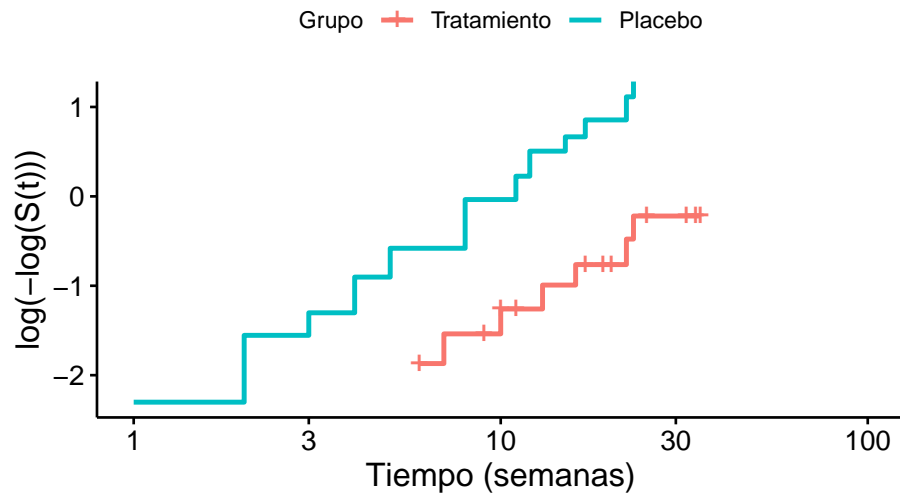
$$\log\{-\log[\hat{S}(t)]\}$$

- 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"
)
```

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.

0.7.2 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
-

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

0.7.4 ¿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.

0.7.5 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.
-

0.8 Soluciones a Violaciones de PH

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

0.9 Introducción a la estratificación

- El modelo de Cox supone que el efecto de cada covariable sobre el riesgo es **proporcional en el tiempo**.
- ¿Qué hacer si esta **suposición se viola para una covariable categórica**?
- **Solución práctica**: usar **estratificación**.

0.9.1 ¿Qué es la estratificación?

- Permite que la **función de riesgo base $h(t)$** sea **diferente para cada nivel** de una variable estratificadora.
- Se supone que **el efecto de otras covariables es el mismo** dentro de cada estrato.
- Se implementa con el argumento `strata()` en `coxph()`.

0.9.2 ¿Cuándo usarla?

- Cuando una covariable **no cumple la suposición de riesgos proporcionales**, pero sí se desea controlar su efecto.
- Ejemplos:
 - Hospital de origen
 - Sexo o edad agrupada
 - Centros clínicos

0.10 Dataset ejemplo: lung (paquete survival)

```
library(survival)
data(cancer, package="survival")
lung$status2 <- ifelse(lung$status == 2, 1, 0)
lung$sex <- factor(lung$sex, levels = c(1, 2), labels = c("Hombre", "Mujer"))
lung$ph.ecog2 <- factor(lung$ph.ecog,
                        levels = c(0,1,2,3,4),
                        labels = c("asymptomatic",
                                   "symptomatic but completely ambulatory",
                                   "in bed <50% of the day",
                                   "in bed > 50% of the day but not bedbound",
                                   "bedbound"))
head(lung)
```

	inst	time	status	age	sex	ph.ecog	ph.karno	pat.karno	meal.cal	wt.loss
1	3	306	2	74	Hombre	1	90	100	1175	NA
2	3	455	2	68	Hombre	0	90	90	1225	15

3	3	1010	1	56	Hombre	0	90	90	NA	15
4	5	210	2	57	Hombre	1	90	60	1150	11
5	1	883	2	60	Hombre	0	100	90	NA	0
6	12	1022	1	74	Hombre	1	50	80	513	0

	status2	ph.ecog2
1	1	symptomatic but completely ambulatory
2	1	asymptomatic
3	0	asymptomatic
4	1	symptomatic but completely ambulatory
5	1	asymptomatic
6	0	symptomatic but completely ambulatory

```
summary(lung)
```

inst	time	status	age	sex
Min. : 1.00	Min. : 5.0	Min. : 1.000	Min. : 39.00	Hombre:138
1st Qu.: 3.00	1st Qu.: 166.8	1st Qu.: 1.000	1st Qu.: 56.00	Mujer : 90
Median : 11.00	Median : 255.5	Median : 2.000	Median : 63.00	
Mean : 11.09	Mean : 305.2	Mean : 1.724	Mean : 62.45	
3rd Qu.: 16.00	3rd Qu.: 396.5	3rd Qu.: 2.000	3rd Qu.: 69.00	
Max. : 33.00	Max. : 1022.0	Max. : 2.000	Max. : 82.00	
NA's : 1				

ph.ecog	ph.karno	pat.karno	meal.cal
Min. : 0.0000	Min. : 50.00	Min. : 30.00	Min. : 96.0
1st Qu.: 0.0000	1st Qu.: 75.00	1st Qu.: 70.00	1st Qu.: 635.0
Median : 1.0000	Median : 80.00	Median : 80.00	Median : 975.0
Mean : 0.9515	Mean : 81.94	Mean : 79.96	Mean : 928.8
3rd Qu.: 1.0000	3rd Qu.: 90.00	3rd Qu.: 90.00	3rd Qu.: 1150.0
Max. : 3.0000	Max. : 100.00	Max. : 100.00	Max. : 2600.0
NA's : 1	NA's : 1	NA's : 3	NA's : 47

wt.loss	status2
Min. : -24.000	Min. : 0.0000
1st Qu.: 0.000	1st Qu.: 0.0000
Median : 7.000	Median : 1.0000
Mean : 9.832	Mean : 0.7237
3rd Qu.: 15.750	3rd Qu.: 1.0000
Max. : 68.000	Max. : 1.0000
NA's : 14	

	ph.ecog2
asymptomatic	: 63
symptomatic but completely ambulatory	: 113
in bed <50% of the day	: 50
in bed > 50% of the day but not bedbound:	1
bedbound	: 0
NA's	: 1

```
lung <- lung %>%
  mutate(surv_obj = Surv(time, status2))

print(lung$urv_obj)
```

```
[1] 306 455 1010+ 210 883 1022+ 310 361 218 166 170 654
[13] 728 71 567 144 613 707 61 88 301 81 624 371
[25] 394 520 574 118 390 12 473 26 533 107 53 122
[37] 814 965+ 93 731 460 153 433 145 583 95 303 519
[49] 643 765 735 189 53 246 689 65 5 132 687 345
[61] 444 223 175 60 163 65 208 821+ 428 230 840+ 305
[73] 11 132 226 426 705 363 11 176 791 95 196+ 167
[85] 806+ 284 641 147 740+ 163 655 239 88 245 588+ 30
[97] 179 310 477 166 559+ 450 364 107 177 156 529+ 11
[109] 429 351 15 181 283 201 524 13 212 524 288 363
[121] 442 199 550 54 558 207 92 60 551+ 543+ 293 202
[133] 353 511+ 267 511+ 371 387 457 337 201 404+ 222 62
[145] 458+ 356+ 353 163 31 340 229 444+ 315+ 182 156 329
[157] 364+ 291 179 376+ 384+ 268 292+ 142 413+ 266+ 194 320
[169] 181 285 301+ 348 197 382+ 303+ 296+ 180 186 145 269+
[181] 300+ 284+ 350 272+ 292+ 332+ 285 259+ 110 286 270 81
[193] 131 225+ 269 225+ 243+ 279+ 276+ 135 79 59 240+ 202+
[205] 235+ 105 224+ 239 237+ 173+ 252+ 221+ 185+ 92+ 13 222+
[217] 192+ 183 211+ 175+ 197+ 203+ 116 188+ 191+ 105+ 174+ 177+
```

0.11 Evaluación inicial: ¿sex viola PH?

```
mod1 <- coxph(surv_obj ~ sex + age, data = lung)
summary(mod1)
```

Call:

```
coxph(formula = surv_obj ~ sex + age, data = lung)
```

n= 228, number of events= 165

	coef	exp(coef)	se(coef)	z	Pr(> z)
sexMujer	-0.513219	0.598566	0.167458	-3.065	0.00218 **
age	0.017045	1.017191	0.009223	1.848	0.06459 .

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

	exp(coef)	exp(-coef)	lower .95	upper .95
sexMujer	0.5986	1.6707	0.4311	0.8311

```
age          1.0172      0.9831      0.9990      1.0357
```

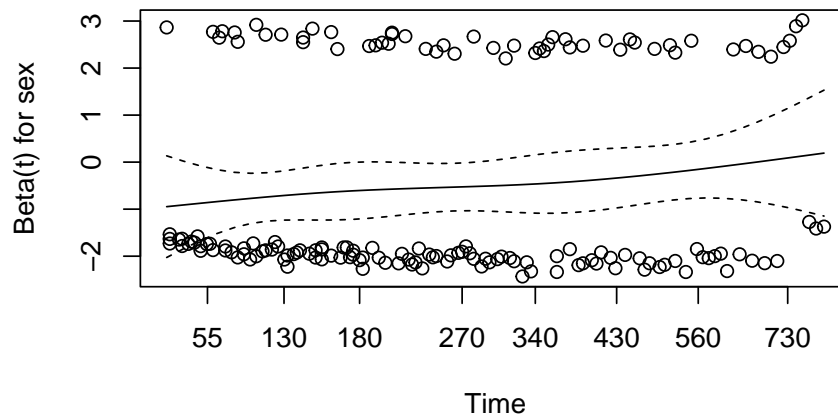
```
Concordance= 0.603 (se = 0.025 )
Likelihood ratio test= 14.12 on 2 df,  p=9e-04
Wald test               = 13.47 on 2 df,  p=0.001
Score (logrank) test = 13.72 on 2 df,  p=0.001
```

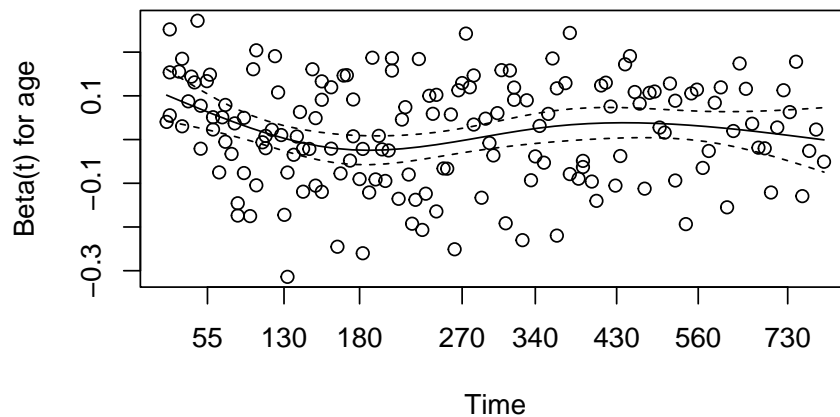
0.11.1 Evaluemos el supuesto de Proporcionalidad

```
zph1 <- cox.zph(mod1)
zph1
```

```
      chisq df    p
sex    2.608  1 0.11
age     0.209  1 0.65
GLOBAL 2.771  2 0.25
```

```
plot(zph1)
```





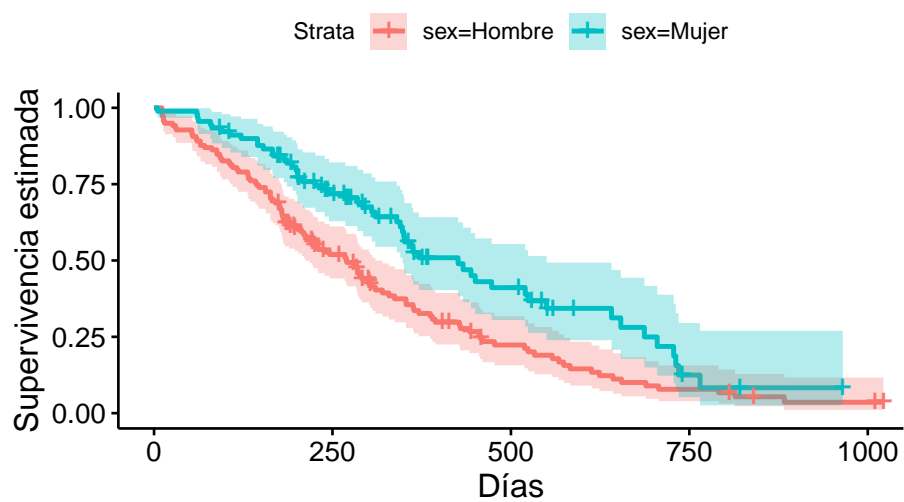
- Si el p-valor en la prueba `cox.zph()` para `sex` es significativo o el gráfico tiene tendencia → **violación de PH**.

0.11.2 Estimador de Kaplan-Meier por sexo

```
fit_km <- survfit(surv_obj ~ sex, data = lung)

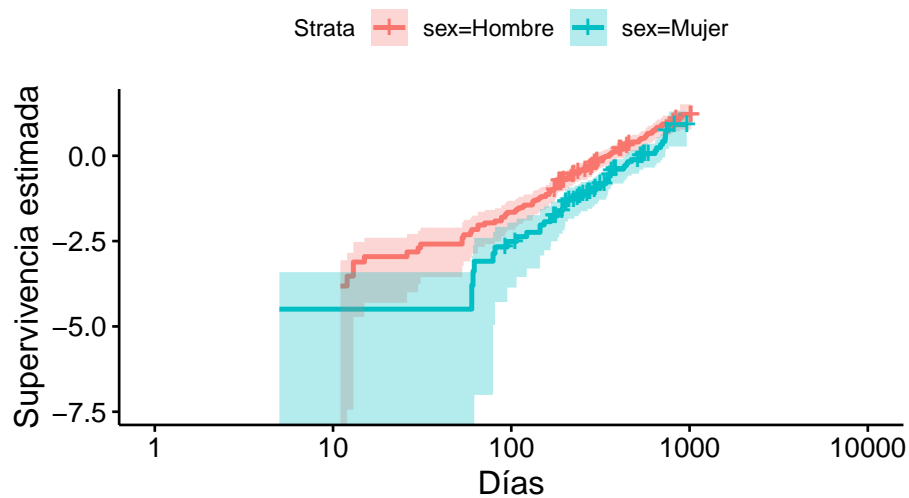
ggsurvplot(fit_km,
  data = lung,
  conf.int = TRUE,
  xlab = "Días", ylab = "Supervivencia estimada",
  title = "Curvas Kaplan-Meier por sexo")
```

Curvas Kaplan–Meier por sexo



```
ggsurvplot(fit_km,  
  data = lung,  
  fun = "cloglog",  
  conf.int = TRUE,  
  xlab = "Días", ylab = "Supervivencia estimada",  
  title = "Curvas Kaplan-Meier por sexo")
```

Curvas Kaplan–Meier por sexo



0.12 Estratificación por sex

```
mod_strat <- coxph(surv_obj ~ age + strata(sex), data = lung)
summary(mod_strat)
```

Call:

```
coxph(formula = surv_obj ~ age + strata(sex), data = lung)
```

n= 228, number of events= 165

	coef	exp(coef)	se(coef)	z	Pr(> z)
age	0.016215	1.016347	0.009187	1.765	0.0776 .

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

	exp(coef)	exp(-coef)	lower .95	upper .95
age	1.016	0.9839	0.9982	1.035

Concordance= 0.546 (se = 0.026)

Likelihood ratio test= 3.18 on 1 df, p=0.07

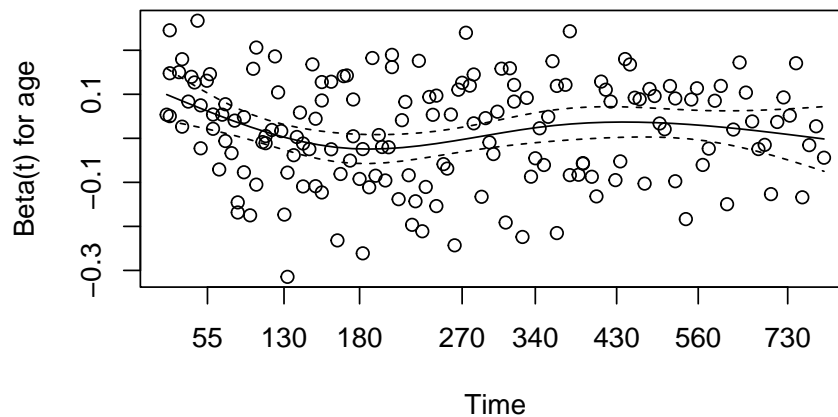
Wald test = 3.12 on 1 df, p=0.08

Score (logrank) test = 3.12 on 1 df, p=0.08

```
zph_strat <- cox.zph(mod_strat)
zph_strat
```

	chisq	df	p
age	0.146	1	0.7
GLOBAL	0.146	1	0.7

```
plot(zph_strat)
```



- La función de riesgo base es diferente para **hombres** y **mujeres**.
- La covariable **age** tiene el mismo efecto en ambos estratos.

0.13 Comparación de modelos: sin estratificación vs con estratificación

```
# Modelo sin estratificación
AIC(mod1)
```

```
[1] 1489.696
```

```
# Modelo con estratificación
AIC(mod_strat)
```

```
[1] 1285.692
```

```
# Ver ambos resúmenes
summary(mod1)
```

Call:

```
coxph(formula = surv_obj ~ sex + age, data = lung)
```

```
n= 228, number of events= 165
```

	coef	exp(coef)	se(coef)	z	Pr(> z)	
sexMujer	-0.513219	0.598566	0.167458	-3.065	0.00218	**
age	0.017045	1.017191	0.009223	1.848	0.06459	.

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

	exp(coef)	exp(-coef)	lower .95	upper .95
sexMujer	0.5986	1.6707	0.4311	0.8311
age	1.0172	0.9831	0.9990	1.0357

```
Concordance= 0.603 (se = 0.025 )
Likelihood ratio test= 14.12 on 2 df,  p=9e-04
Wald test               = 13.47 on 2 df,  p=0.001
Score (logrank) test = 13.72 on 2 df,  p=0.001
```

```
summary(mod_strat)
```

Call:

```
coxph(formula = surv_obj ~ age + strata(sex), data = lung)
```

```
n= 228, number of events= 165
```

	coef	exp(coef)	se(coef)	z	Pr(> z)
age	0.016215	1.016347	0.009187	1.765	0.0776 .

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

	exp(coef)	exp(-coef)	lower .95	upper .95
age	1.016	0.9839	0.9982	1.035

```
Concordance= 0.546 (se = 0.026 )
Likelihood ratio test= 3.18 on 1 df,  p=0.07
Wald test               = 3.12 on 1 df,  p=0.08
Score (logrank) test = 3.12 on 1 df,  p=0.08
```

```
anova(mod1, mod_strat, test = "LRT")
```

Analysis of Deviance Table

Cox model: response is surv_obj

Model 1: ~ sex + age

Model 2: ~ age + strata(sex)

	loglik	Chisq	Df	Pr(> Chi)
1	-742.85			
2	-641.85	202	1	< 2.2e-16 ***

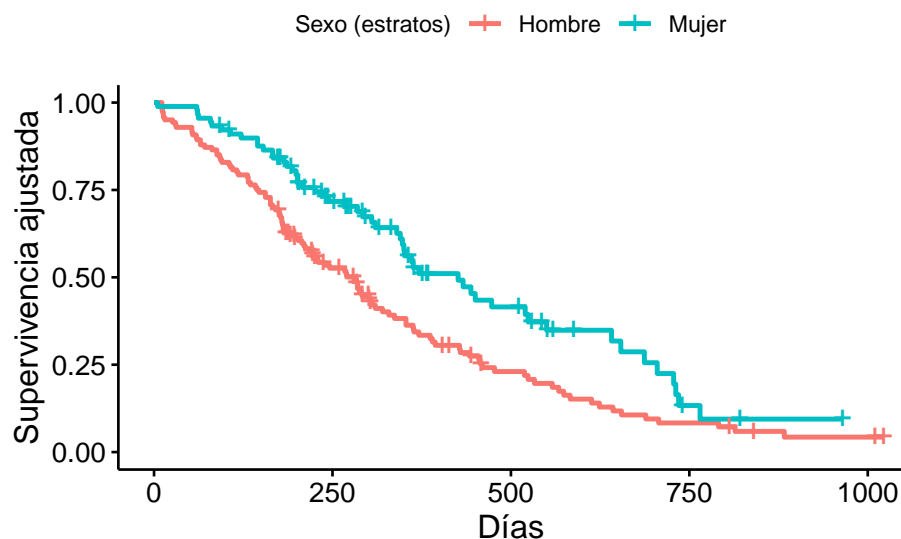
```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Se puede comparar el AIC para ver si mejora el ajuste.
- Nota: el coeficiente de **sex** desaparece en el modelo estratificado.

0.14 Comparación visual de curvas ajustadas

```
library(survminer)
fit_strat <- survfit(mod_strat)

ggsurvplot(fit_strat, data = lung,
            legend.title = "Sexo (estratos)",
            xlab = "Días", ylab = "Supervivencia ajustada")
```



0.15 Conclusiones

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

0.16 Referencias

- Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2), 187–220.
- 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.