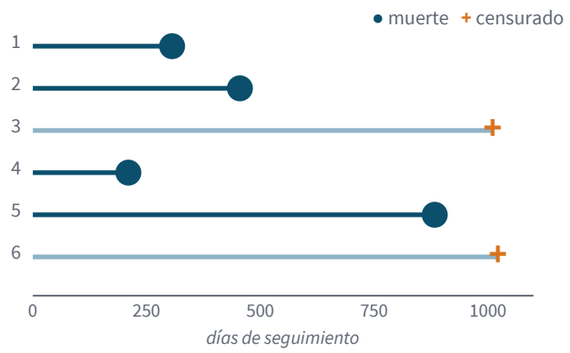
Análisis de supervivencia con survival :: **CHEAT SHEET**

survival modela el tiempo hasta un evento (muerte, recaída, alta) cuando algunos sujetos solo se siguen una parte del tiempo.

1 El objeto Surv

Cada sujeto aporta un **tiempo** y un **indicador de evento**. Si el evento no se observó, el tiempo está **censurado** (+): solo sabemos que sobrevivió al menos ese tiempo.




```
library(survival)
Surv(lung$time, lung$status)
#> 306   455   1010+   210   883   1022+ ...
```

CÓMO CODIFICAR EL EVENTO (status =)

`0 / 1`	1 = evento (lo habitual)
`1 / 2`	2 = evento (como en lung)
`FALSE / TRUE`	TRUE = evento

OTROS FORMATOS

`Surv(start, stop, event)` proceso de conteo: datos que cambian en el tiempo, entrada tardía

`Surv(time, time2, type = "interval2")` censura por intervalo (NA = extremo abierto)

CONSEJO La censura debe ser **no informativa**: si un paciente se pierde porque empeoró, todos los métodos de esta hoja fallan. Revisa por qué se pierde el seguimiento.

TIEMPO A PARTIR DE FECHAS

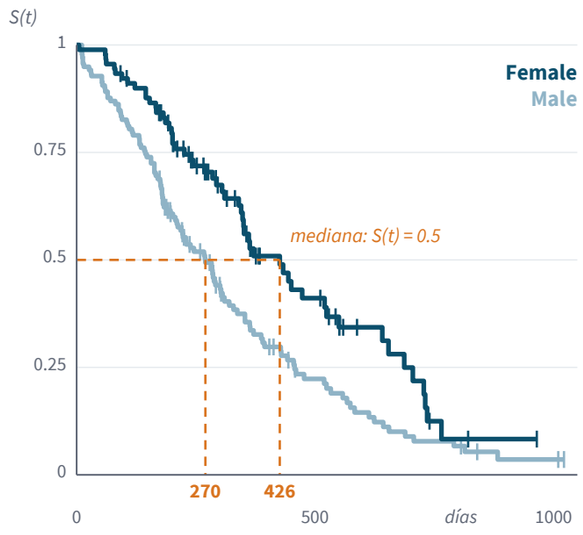
```
df$time <- as.numeric(difftime(
  df$end, df$start, units = "days"))
```

end = fecha del evento, o último contacto si está censurado.

2 Kaplan–Meier

```
lung$sex <- factor(lung$sex, 1:2,
  labels = c("Male", "Female"))
km <- survfit(Surv(time, status) ~ sex,
              data = lung)
km
```

	n	eventos	mediana	IC 95%
Male	138	112	**270**	212–310
Female	90	53	**426**	348–550



Los escalones marcan tiempos con eventos; las marcas, censuras. **Curva global: ~ 1** (mediana 310 días).

```
plot(km, col = 1:2, conf.int = TRUE,
     mark.time = TRUE, xlab = "Days")
```

Gráfico rápido. Para publicar: survminer o ggsurvfit.

INCIDENCIA ACUMULADA 1 − S(t)

```
plot(km, fun = "event")   # las curvas suben
```

Solo con un tipo de evento; con eventos competitivos, ver página 2.

REVISA RIESGOS PROPORCIONALES A OJO

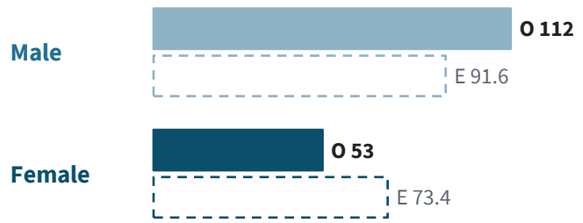
```
plot(km, fun = "cloglog")
```

log(−log S) vs log(tiempo): curvas aprox. paralelas apoyan un modelo de Cox.

3 Compara grupos

```
survdiff(Surv(time, status) ~ sex,
         data = lung)
```

La prueba compara las muertes observadas con las esperadas si ambos grupos compartieran una sola curva:



Hombres: más muertes de las esperadas · mujeres: menos
→ **$\chi^2(1) = 10.3$, $p = 0.001$**

ELIGE LA PONDERACIÓN: rho =

`rho = 0` **χ^2 10.3**

Log-rank (por defecto). Pesa igual todos los tiempos; más potencia si los riesgos son proporcionales.

`rho = 1` **χ^2 12.7**

Peto–Peto (tipo Gehan–Wilcoxon). Más peso a los eventos tempranos.

AJUSTA POR UN CONFUSOR

```
survdiff(Surv(time, status) ~ sex +
  strata(ph.ecog), data = lung)
```

Log-rank estratificado: compara por sexo dentro de cada nivel ECOG (χ^2 10.8).

CONSEJO Si las curvas se **cruzan**, no hay riesgos proporcionales: el log-rank puede no detectar una diferencia real. Reporta la supervivencia a tiempos fijos o el RMST (ver 4).

MÁS DE DOS GRUPOS

```
survdiff(Surv(time, status) ~ ph.ecog,
         data = lung)
```

Una prueba global ($\chi^2(3) = 22$, $p < 0.001$). Por pares: survminer::pairwise_survdiff().

4 Extrae los números

SUPERVIVENCIA A TIEMPOS ELEGIDOS

```
summary(km, times = c(180, 365))
```

	día 180	día 365
Male	64% (57–73)	34% (26–43)
Female	84% (77–92)	53% (42–66)
en riesgo	89 / 71	35 / 30

MEDIANA Y OTROS CUANTILES

```
quantile(km, probs = c(0.25, 0.5, 0.75))
```

Mediana Male 270 (212–310) · Female 426 (348–550). *Sin mediana si S(t) nunca llega a 0.5*: repórtala como "no alcanzada".

SUPERVIVENCIA MEDIA RESTRINGIDA (RMST)

```
print(km, rmean = 365)
```

Días vividos en promedio en el primer año (área bajo S(t) hasta 365): Male 241, Female 297. Sirve aunque las curvas se crucen.

COMO DATA FRAME

```
summary(km)$table   # n, eventos, mediana
broom::tidy(km)     # una fila por tiempo
```

ARGUMENTOS ÚTILES DE survfit()

`conf.type = "log-log"` IC que queda entre 0 y 1

`conf.int = 0.90` nivel del IC

`start.time = 30` condiciona a sobrevivir 30 días

CONSEJO Muestra siempre el **número en riesgo** bajo la curva de KM: la cola derecha se basa en pocos sujetos y es inestable.

SEGUIMIENTO MEDIANO (KM INVERSO)

```
survfit(Surv(time, status == 1) ~ 1,
        data = lung)   # censura = evento
```

→ **588 días** (511 – 1010). Repórtalo en todo análisis de supervivencia.

CC BY SA Anthony Romero-Cerdán • anthonyromero.netlify.app • Más información en `vignette("survival")` • versión de survival 3.8-12 • Actualizado: 2026-10

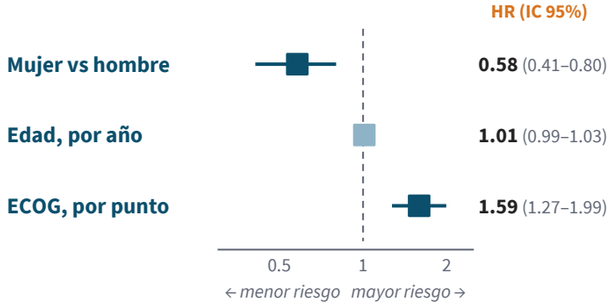
Análisis de supervivencia con survival :: **CHEAT SHEET**

De las curvas a los modelos: hazard ratios ajustados por covariables, sus supuestos y las extensiones que encontrarás en la práctica.

5 Regresión de Cox

```
cx <- coxph(Surv(time, status) ~
  sex + age + ph.ecog, data = lung)
summary(cx)
```

El hazard ratio (HR) compara el riesgo instantáneo del evento, en cualquier momento, entre grupos:



Léelo así: *en cualquier momento, las mujeres tuvieron un riesgo (hazard) de muerte 42% menor que los hombres (HR 0.58, IC 95% 0.41–0.80), ajustando por edad y ECOG.*

TAMBIÉN EN summary(cx)

Prueba de razón de verosimilitud 30.5 con 3 gl, p < 0.001

Concordancia (C) 0.64: discriminación

n / eventos 227 / 164 (1 fila con NA excluida)

TABLA ORDENADA DE HR

```
broom::tidy(cx, exponentiate = TRUE,
            conf.int = TRUE)
```

CURVAS DE SUPERVIVENCIA AJUSTADAS

```
nd <- data.frame(sex = c("Male", "Female"),
                 age = 62, ph.ecog = 1)
survfit(cx, newdata = nd)
```

→ **Supervivencia a 1 año 33% vs 53%** para alguien de 62 años con ECOG 1.

NIVEL DE REFERENCIA Y PRUEBAS GLOBALES

`relevel(sex, ref = "Female")` invierte el HR (1/0.58 = 1.7)

`anova(cx)` prueba RV secuencial por término, p. ej., para un factor con > 2 niveles

6 Revisa los supuestos

Cox asume **riesgos proporcionales**: el HR es constante durante el seguimiento. Pruébalo con los residuos de Schoenfeld escalados.

```
zph <- cox.zph(cx)
zph
```

	chisq	df	p
sex	2.31	1	0.13
age	0.19	1	0.66
ph.ecog	2.05	1	0.15
GLOBAL	**4.46**	**3**	**0.22**

Todos los p > 0.05 → sin evidencia contra RP. Un valor p pequeño señala esa variable.

MIRA, NO SOLO PRUEBES

```
plot(zph[1])   # residuos de sex vs tiempo
```



Esquema: una línea suavizada plana indica efecto constante; una tendencia indica que el HR cambia con el tiempo.

SI UNA VARIABLE NO CUMPLE RP

- `strata(ph.ecog)` riesgo basal propio por nivel; no da HR para ella, el resto sigue ajustado.
- `tt(age)` deja que el efecto cambie con el tiempo, p. ej. `tt = function(x, t, ...) x * log(t)`
- `survSplit(..., cut = 180)` divide el seguimiento y estima un HR por periodo.

```
coxph(Surv(time, status) ~ sex + age +
  strata(ph.ecog), data = lung)
```

7 Más allá de lo básico

COVARIABLES QUE CAMBIAN EN EL TIEMPO

Valores que cambian durante el seguimiento (laboratorio, tratamiento) requieren una fila por intervalo: (tstart, tstop].

```
temp <- subset(pbc, id <= 312,
  select = c(id:sex, stage))
pbc2 <- tmerge(temp, temp, id = id,
  death = event(time, status == 2))
pbc2 <- tmerge(pbc2, pbcseq, id = id,
  bili = tdc(day, bili))
coxph(Surv(tstart, tstop, death) ~
  log(bili), data = pbc2)
```

→ **HR 3.94** por unidad logarítmica de bilirrubina actual.

EFECTOS NO LINEALES

```
coxph(Surv(time, status) ~ sex +
  pspline(age) + ph.ecog, data = lung)
```

Prueba las partes lineal y no lineal (edad: no lineal p = 0.41).

RIESGOS COMPETITIVOS

Si otro evento impide el que estudias (muerte antes de la progresión), usa un **factor** como status: primer nivel = censurado.

```
etime <- with(mgus2, ifelse(pstat == 0,
  futime, ptime))
event <- with(mgus2, ifelse(pstat == 0,
  2 * death, 1))
event <- factor(event, 0:2, labels =
  c("censor", "pcm", "death"))
cr <- survfit(Surv(etime, event) ~ 1,
              data = mgus2)
summary(cr, times = 120)$pstate
```

→ **Incidencia acumulada a 10 años: PCM 6.4%, muerte 53%** (Aalen–Johansen).

Regresión sobre la incidencia acumulada (Fine–Gray): `finegray()` y luego `coxph(..., weight = fgwt)`.

DATOS AGRUPADOS O RECURRENTES

```
coxph(Surv(time, status) ~ sex + age,
      cluster = inst, data = lung)
```

Agrega EE robustos (sándwich) para pacientes anidados en centros o eventos repetidos por sujeto.

8 Modelos AFT y errores

TIEMPO DE FALLA ACELERADO (survreg)

```
wb <- survreg(Surv(time, status) ~
  sex + age + ph.ecog, data = lung,
  dist = "weibull")
exp(coef(wb))   # razones de tiempo
```

→ **Razón de tiempo 1.49:** el tiempo de supervivencia de las mujeres se alarga 49%. Aquí > 1 es protector (al revés que el HR).

`dist =` "weibull" · "exponential" · "lognormal" · "loglogistic"

ERRORES FRECUENTES

- **Demasiadas covariables.** Apunta a ≥ 10 eventos por parámetro: 164 eventos admiten unos 16 parámetros (cuenta eventos, no pacientes).
- **HR ≠ RR.** Un HR de 0.58 es una razón de riesgos instantáneos (tasas), no un riesgo acumulado 42% menor.
- **Sesgo de tiempo inmortal.** No clasifiques la exposición con información posterior al tiempo cero (p. ej., "alguna vez tratado"). Usa una covariable dependiente del tiempo.
- **1 − KM con riesgos competitivos** sobreestima la incidencia acumulada. Usa el status como factor (ver 7).
- **Filas excluidas en silencio.** Las filas con covariables faltantes se eliminan: revisa n y eventos en summary().
- **Empates.** coxph() usa Efron por defecto y Stata, Breslow, así que los resultados difieren un poco.

CHECKLIST Reporta: n, eventos, seguimiento mediano, curvas de KM con número en riesgo, HR con IC 95% y cómo se revisó RP.

PAQUETES COMPLEMENTARIOS

survminer · ggsurvfit curvas de KM con tabla de riesgo

gtsummary tbl_regression(cx, exponentiate = TRUE)

cmprsk · tidycmprsk prueba de Gray, regresión de Fine–Gray

flexsurv modelos paramétricos flexibles

CC BY SA Anthony Romero-Cerdán • anthonyromero.netlify.app • Más información en `vignette("survival")` • versión de survival 3.8-12 • Actualizado: 2026-10