Principios Básicos del Análisis de Supervivencia con R

Sergio M. Nava Muñoz
nava@cimat.mx
CIMAT

2025-06-01

Objetivo

- Comprender los fundamentos del análisis de supervivencia
- Distinguir entre las funciones de riesgo, supervivencia, densidad y distribución
- Introducir modelos paramétricos y el enfoque de máxima verosimilitud
- Aplicar estas ideas en R usando funciones del paquete `survival`

1. Función de Supervivencia $S(t)$

- Define la probabilidad de sobrevivir más allá del tiempo t:

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

- Es monótona decreciente
- Se relaciona con la función de distribución:

$$F(t) = 1 - S(t)$$

Estas funciones se discuten con detalle en Moore (2016).

2. Función de Riesgo $h(t)$

- Tasa instantánea de falla en t dado que se ha sobrevivido hasta t:

$$h(t) = \lim_{\delta \to 0} \frac{P(t < T < t + \delta \mid T > t)}{\delta}$$

- También llamada tasa de falla o función de intensidad

3. Relación entre funciones

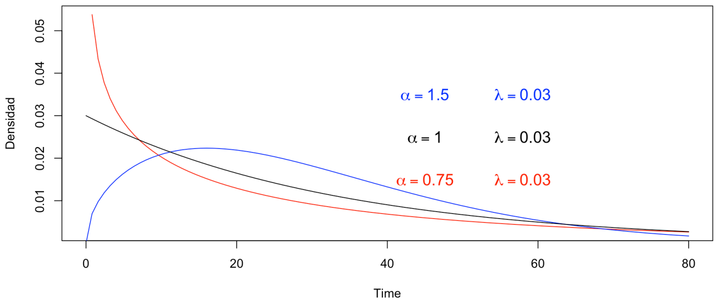
- $$f(t) = -\frac{d}{dt} S(t)$$
- $$h(t) = \frac{f(t)}{S(t)}$$
- $S(t) = \exp\left(-\int_0^t h(u)du\right)$

4. Distribuciones Paramétricas

- **Exponencial**:
  - $h(t) = \lambda$ constante
  - $S(t) = e^{-\lambda t}$
- **Weibull**:
  - $h(t) = \alpha\lambda^{\alpha} t^{\alpha-1}$
  - Controla crecimiento o decrecimiento del riesgo
- **Gamma** y otras

CIMAT

5 Ejemplo en R: Weibull



CIMAT

Hazard and survival functions for US males and females in 2004

Nombre del dataset:

`survexp.us` del paquete `survival` en R

Descripción:

Contiene tasas de riesgo (hazard rates) diarias por edad, sexo y año calendario en EE. UU., derivadas de las tablas de vida oficiales.

Dimensiones:

dim(survexp.us) = 111 x 2 x 65

- 110 edades: de 0 a 109 años
- 2 sexos: `"male"`, `"female"`
- 65 años calendario: de 1940 a 2004

```r
library(survival)
tabla=as.table(survexp.us[,,"2004"])

knitr::kable(tabla, booktabs = TRUE, digits = 5,format.args = list(scientific = FALSE),
             format = "simple",
             caption="Tasas de riesgo (hazard rates) de hombres y mujeres por edad para el año 2004")
```

Tasas de riesgo (hazard rates) de hombres y mujeres por edad para el año 2004

	male	female
0	0.0000205423075	0.0000167272438
1	0.0000013911816	0.0000012514838
2	0.0000008926849	0.0000007311038
3	0.0000006845483	0.0000005394097
4	0.0000005695322	0.0000004599976
5	0.0000005229794	0.0000004134467

CIMAT

	male	female
6	0.0000004983342	0.0000003778495
7	0.0000004682125	0.0000003532055
8	0.0000004161850	0.0000003285618
9	0.0000003422527	0.0000003066565
10	0.0000002874894	0.0000002929657
11	0.0000003039183	0.0000003093946
12	0.0000004435678	0.0000003696348
13	0.0000007502739	0.0000004873808
14	0.0000011802681	0.0000006489475
15	0.0000016651195	0.0000008379105
16	0.0000021281369	0.0000010159311
17	0.0000025610880	0.0000011528779
18	0.0000029146241	0.0000012213538
19	0.0000031941966	0.0000012405274
20	0.0000034683150	0.0000012487447
21	0.0000037260113	0.0000012706576
22	0.0000038877693	0.0000012898315
23	0.0000039316375	0.0000013172230
24	0.0000038877693	0.0000013473541
25	0.0000038082599	0.0000013857031
26	0.0000037397193	0.0000014295313
27	0.0000036821464	0.0000014815781
28	0.0000036602142	0.0000015473229
29	0.0000036739218	0.0000016240271
30	0.0000037040788	0.0000017171708
31	0.0000037561689	0.0000018267558
32	0.0000039014780	0.0000019664829

CIMAT

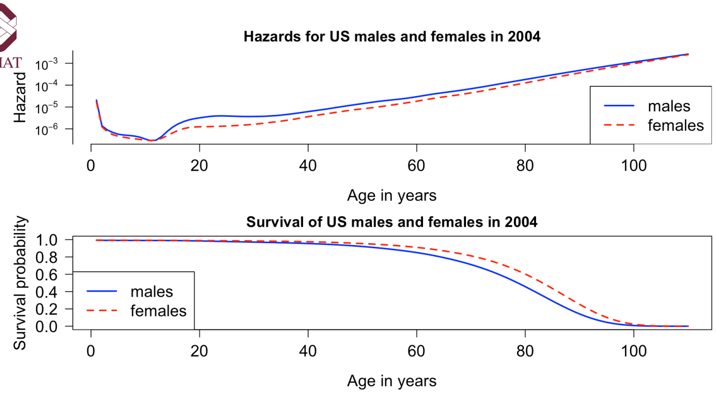
	male	female
33	0.0000040248598	0.0000020925174
34	0.0000042551873	0.0000022596591
35	0.0000045294119	0.0000024432528
36	0.0000048502897	0.0000026597446
37	0.0000052370384	0.0000029338095
38	0.0000056869426	0.0000032599825
39	0.0000061780810	0.0000036190918
40	0.0000066830301	0.0000039809901
41	0.0000072155231	0.0000043456784
42	0.0000078222521	0.0000047323559
43	0.0000085362230	0.0000051602329
44	0.0000093575199	0.0000056320723
45	0.0000102450172	0.0000061478968
46	0.0000111685396	0.0000066857747
47	0.0000121501236	0.0000072182681
48	0.0000131898315	0.0000077289004
49	0.0000142904817	0.0000082396279
50	0.0000155072150	0.0000087889035
51	0.0000168098840	0.0000094124649
52	0.0000181131731	0.0000101350948
53	0.0000193564221	0.0000109761134
54	0.0000205781679	0.0000119246306
55	0.0000218418554	0.0000129670037
56	0.0000232579887	0.0000141005879
57	0.0000249290293	0.0000153585353
58	0.0000269909242	0.0000167795823
59	0.0000294694636	0.0000183970583

CIMAT

	male	female
60	0.0000323962531	0.0000203050893
61	0.0000356423614	0.0000224518348
62	0.0000390422861	0.0000246886681
63	0.0000423852650	0.0000268969167
64	0.0000457128431	0.0000291235505
65	0.0000493037051	0.0000315735562
66	0.0000531532997	0.0000340950595
67	0.0000576401609	0.0000370573273
68	0.0000627954313	0.0000404867969
69	0.0000686902565	0.0000444047004
70	0.0000752027649	0.0000487991958
71	0.0000824643114	0.0000538177629
72	0.0000908487335	0.0000596233153
73	0.0001004259604	0.0000662993874
74	0.0001111229212	0.0000739057740
75	0.0001229182933	0.0000825658898
76	0.0001356362658	0.0000919753943
77	0.0001496335961	0.0001024375192
78	0.0001650314702	0.0001140641492
79	0.0001819644193	0.0001269818499
80	0.0002005695435	0.0001413240533
81	0.0002210021109	0.0001572405157
82	0.0002434250467	0.0001748926443
83	0.0002680161867	0.0001944546937
84	0.0002949577784	0.0002161240702
85	0.0003244532439	0.0002401024790
86	0.0003567077805	0.0002666125143

CIMAT

	male	female
87	0.0003919488210	0.0002958971187
88	0.0004304033923	0.0003282100755
89	0.0004723192404	0.0003638217296
90	0.0005179417112	0.0004030222174
91	0.0005675321498	0.0004461123850
92	0.0006213573981	0.0004934071813
93	0.0006796850660	0.0005452395431
94	0.0007427815967	0.0006019410035
95	0.0008109134282	0.0006638581520
96	0.0008843517255	0.0007313354039
97	0.0009633424670	0.0008047205492
98	0.0010481387132	0.0008843555073
99	0.0011389678015	0.0009705646602
100	0.0012392426794	0.0010666267047
101	0.0013475429313	0.0011714549915
102	0.0014653077900	0.0012865858234
103	0.0015933643890	0.0014130317366
104	0.0017326121470	0.0015519047795
105	0.0018840290851	0.0017044262930
106	0.0020486786957	0.0018719376514
107	0.0022277174124	0.0020559120598
108	0.0024224027319	0.0022579675101
109	0.0026341020468	0.0024798810105



6. Cálculo de la función de supervivencia $S(t)$ a partir de la función de riesgo $h(t)$

```r
library(survival)
tm <- c(0, # birth
        1:110) # subsequent years
hazMale <- survexp.us[,"male","2004"]
tm.diff <- diff(tm)
survMale <- exp(-cumsum(hazMale * tm.diff) * 365.24)
```


7. Tiempo Medio y Mediano de Supervivencia

- **Tiempo medio de supervivencia**:

$$E(T) = \int_0^\infty S(t)dt$$

Si $S(t)$ no tiende a 0, esta integral puede ser infinita.

```r
# Estimación numérica en R
sum(survMale * tm.diff)  # área bajo la curva de supervivencia
```

```
[1] 74.38014
```

- **Tiempo mediano de supervivencia**:

$$\text{Mediana} = \min\{t : S(t) \leq 0.5\}$$

```r
library(survival)
tt <- c(7,6,6,5,2,4)
status <- c(0,1,0,0,1,1)
fit <- survfit(Surv(tt, status) ~ 1)
summary(fit)$table["median"]
```

```
median
     6
```

- Si la curva $S(t)$ no cruza 0.5, la mediana no está definida.

Media y mediana en modelos paramétricos

Modelo Exponencial

- Si $T \sim \text{Exp}(\lambda)$:
  - Media:

$$E(T) = \frac{1}{\lambda}$$

  - Mediana:

$$\text{Mediana} = \frac{\ln(2)}{\lambda}$$

```r
lambda <- 0.1
media_exp <- 1 / lambda
mediana_exp <- log(2) / lambda
media_exp
```

```
[1] 10
```

```r
mediana_exp
```

```
[1] 6.931472
```

Podemos generar variables aleatorias de la distribución Weibull

```r
set.seed(137)
tt.exp <- rexp(1000, rate=lambda)
mean(tt.exp)
```

```
[1] 9.778199
```

```r
median(tt.exp)
```

```
[1] 6.907413
```


Modelo Weibull

- Si $T \sim \text{Weibull}(\alpha, \lambda)$:
  - Media:

$$E(T) = \lambda^{-1} \cdot \Gamma\left(1 + \frac{1}{\alpha}\right)$$

  - Mediana:

$$\text{Mediana} = \left(\frac{\ln(2)^{1/\alpha}}{\lambda}\right)$$

```r
alpha <- 1.5; lambda <- 0.03
media_weibull <- (1/lambda) * gamma(1 + 1/alpha)
mediana_weibull <- (log(2)^(1/alpha))/lambda
media_weibull
```

```
[1] 30.09151
```

```r
mediana_weibull
```

```
[1] 26.10733
```

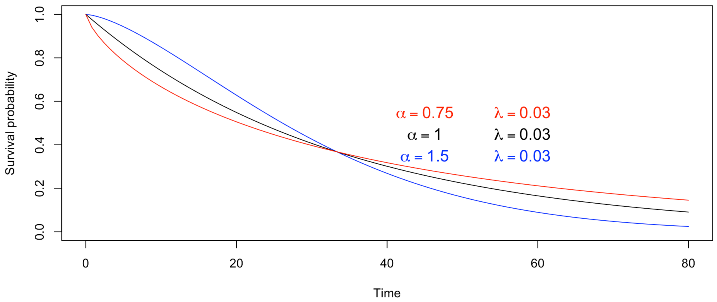
Podemos generar variables aleatorias de la distribución Weibull

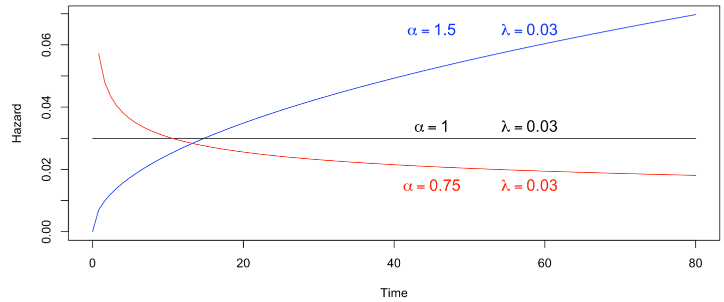
```r
set.seed(137)
tt.weib <- rweibull(1000,
                    shape=alpha,
                    scale=1/lambda)
mean(tt.weib)
```

```
[1] 29.90632
```

```r
median(tt.weib)
```

```
[1] 26.26265
```





Weibull hazard functions

8. Máxima Verosimilitud: Exponencial

- Si $f(t) = \lambda e^{-\lambda t}$:
  - $L(\lambda) = \prod_{i=1}^{n} f(t_i)^{\delta_i} S(t_i)^{1-\delta_i}$
  - $\hat{\lambda} = \frac{d}{\sum t_i}$

9. Código R: Estimación con censura

```r
tt <- c(7,6,6,5,2,4)
status <- c(0,1,0,0,1,1)
library(survival)
survfit(Surv(tt, status) ~ 1)
```

```
Call: survfit(formula = Surv(tt, status) ~ 1)

     n events median 0.95LCL 0.95UCL
[1,] 6      3      6       4      NA
```


10. Actividad práctica

- Simular datos censurados
- Estimar $S(t)$ para Weibull y Exponencial
- Comparar con estimación empírica (Kaplan-Meier)

11. Lecturas recomendadas

- Moore, D. F. (2016). *Applied Survival Analysis Using R*
- Klein & Moeschberger (2003). *Survival Analysis*

Moore, D. F. (2016). *Applied survival analysis using r* (2nd ed.). Springer. https://doi.org/10.1007/978-3-319-31245-3

Preguntas y discusión

- ¿Qué modelo parece más adecuado para tiempos de vida humanos?
- ¿Cuáles son los riesgos de usar modelos paramétricos sin validarlos?