

# Tutorial 4: Supervivencia con destino conocido\*

Curso: Métodos de captura-recaptura, UNAM. Abril, 2010.

Roberto E. Munguía-Steyer

Dpto. Ecología, IB-USP, Brasil.

rmunguia.steyer@gmail.com

## 1. Clases de análisis de supervivencia

La supervivencia es un componente de la adecuación de los individuos que conforman una población y por lo tanto objeto del interés de ecólogos poblacionales y evolutivos. Si encontramos diferencias sistemáticas entre individuos con determinados atributos (*e.g.* sexo, coloración, tamaño) que tienen un componente genético y heredable, podemos realizar inferencias sobre el mantenimiento y evolución sobre dicho rasgo morfológico, fisiológico o comportamental.

Los análisis de supervivencia pueden ser categorizados en dos principales grupos:

1. Destino conocido: Somos capaces de detectar la muerte (o falla) de los individuos estudiados. Ejemplos: Apertura de las flores, germinación de semillas en condiciones experimentales, eclosión de huevos de lagartos. Los análisis de supervivencia con destino conocido pueden ser de dos tipos:
  - Modelos lineares generalizados con distribución binomial (GLMs): Fotografía en el tiempo. Ejemplo: ¿Existe diferencia en la supervivencia de los individuos en la transición de ninfas a adultos en dos especies de hemipteros? Trabajo en condiciones de laboratorio.
  - Curva de supervivencia: Cómo los individuos mueren en función del tiempo. Ejemplo: Análisis del riesgo a morir en función de la edad del individuo. ¿Los individuos presentan diferentes curvas de supervivencia según el sexo al cual pertenezcan? Análisis Kaplan-Meier, análisis de supervivencia paramétrico, modelo con distribución exponencial, gamma o Weibull.
2. Destino desconocido: Existe un componente de detectabilidad asociada a la supervivencia que precisamos disociar de la misma. Estudios de captura-recaptura:



\* Bajo licencia: Creative Commons attribution-share alike.

Vamos a campo y generamos una historia de encuentros (una cadena formada por ceros y unos). A partir de la historia de encuentros disociamos la probabilidad de recaptura de la de supervivencia mediante métodos de máxima verosimilitud. Este tema será motivo del próximo tutorial.

## 2. Supervivencia: GLMs con distribución binomial

A veces, nuestro interés es determinar si la supervivencia de los organismos que estudiamos difiere sistemáticamente entre grupos en un momento dado (fotográfico).

Pensemos que en las chinches de agua gigantes (Hemiptera: Belostomatidae) las cuales se caracterizan por presentar cuidado paternal exclusivo. Las hembras ovipositan sobre los machos en una serie de cópulas y oviposiciones intercaladas y hay machos que cargan más de 100 huevos sobre su dorso. Los machos cargados de huevos pasan la mayor del tiempo bajo el agua y realizan una serie de movimientos de aireación que facilitan el transporte de oxígeno y modulan la viabilidad de los huevos que cargan [3].

Dado que el cuidado paternal en la forma de los movimientos de aireación representa un costo en términos de energía, procedemos a evaluar experimentalmente si los machos malnutridos realizan un cuidado paternal menos intenso y si eso redundaría en la viabilidad de los huevos que cargan. Manipulamos el estado nutricional experimentalmente regulando la cantidad de alimento (grillos) para cada tratamiento. A los machos paternales malnutridos los alimentamos una vez cada diez días mientras a los bien nutridos los alimentamos cada cuatro días. MN = Machos malnutridos, BN = Machos nutridos. 30 machos por tratamiento. Supongamos que los machos no difieren en el número de huevos que cargan:  $\mu = 30$ ,  $\sigma = 0.5$ .

```
> set.seed(2222)
> nind <- 60
> m.huevos <- 30
> sd.huevos = 2
> huevos <- round(rnorm(nind, m.huevos, sd = sd.huevos), digits = 0)
> trat <- sample(rep(c("BN", "MN"), c(30, 30)))
> bd <- data.frame(huevos, trat)
```

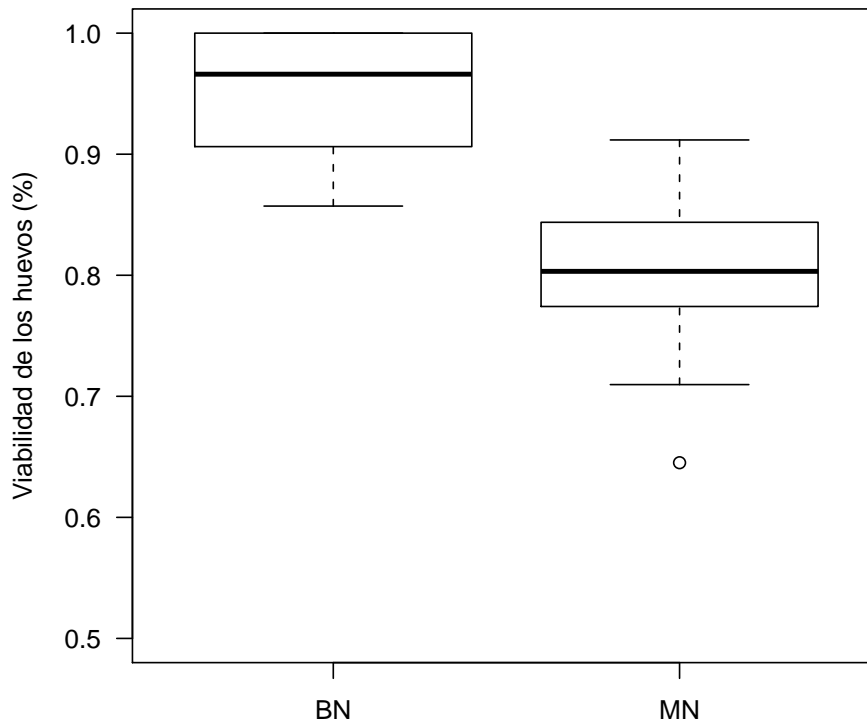
Asumamos que los machos malnutridos realizan menor esfuerzo en el cuidado parental (menor tasa en los movimientos de aireación) y por lo tanto la probabilidad de supervivencia (viabilidad) de los huevos de la oviposición a la eclosión es de 0.80, mientras que en los machos bien nutridos es de 0.95. Realicemos la simulación con los valores de los parámetros descritos y veamos las primeras filas de nuestra base de datos.

```
> n.filas <- dim(bd)[1]
> for (i in 1:n.filas) {
+   bd$h.viab[i] <- ifelse(bd$trat[i] == "BN", rbinom(1, bd$huevos[i],
+     0.95), rbinom(1, bd$huevos[i], 0.8))
+ }
> bd[1:10, ]
```

	huevos	trat	h.viab
1	29	BN	27
2	32	BN	29
3	33	BN	32
4	31	BN	28
5	31	MN	22
6	29	MN	23
7	30	MN	24
8	32	BN	31
9	34	MN	31
10	32	MN	28

Despues de contar el número de huevos que eclosionaron y no lo hicieron en la respectiva puesta de huevos que cada macho cargó, nos preguntamos: ¿Difieren la viabilidad de los huevos dependiendo si los cuida un macho malnutrido o uno bien nutrido?

Veamos en los diagramas de cajas los porcentajes de viabilidad de los huevos dependiendo del status nutricional del macho que los carga:



Determinemos si el tratamiento incide en la viabilidad de los huevos:

```
> m1 <- glm(cbind(h.viab, h.nviab) ~ factor(trat), binomial, data = bd)
> m.bin <- anova(m1, test = "Chi")
```

Existe un serio costo de la malnutrición de los machos parentales en la viabilidad de los huevos que cuidan.

	Df	Deviance	Resid. Df	Resid. Dev	P(> Chi )
NULL			59	155.84	
factor(trat)	1	96.49	58	59.35	0.0000

### 3. Curvas de supervivencia

#### 3.1. Mortalidad *per capita* constante en el tiempo: Distribución exponencial

Cuando analizamos las curvas de supervivencia es posible extraer más información a nuestros datos, por ejemplo determinar si el riesgo de morir varía en función del tiempo. También es posible estimar el tiempo medio de vida para los individuos de una población. ¿Cuál es el tiempo medio de vida de los individuos de una población de insectos que sufren una mortalidad mensual constante = 0.15? La población al inicio cuenta con 300 individuos, no presenta inmigración y no se encuentra en etapa reproductiva, por lo tanto no hay reclutamiento de individuos.

```
> t.vida <- rexp(300, 0.15)
> mean(t.vida)
```

```
[1] 6.162
```

El tiempo de vida esperado de una distribución exponencial es:

$$\mu = \frac{1}{\lambda} \tag{1}$$

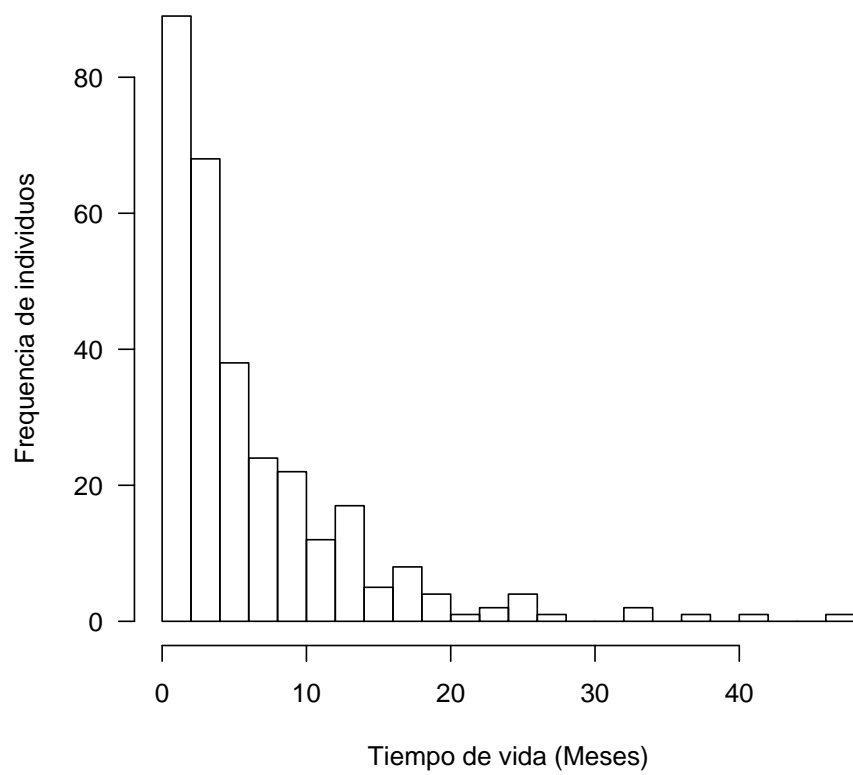
Donde  $\lambda$  es la tasa de mortalidad por unidad de tiempo.

```
> m.esp <- 1/0.15
> m.esp
```

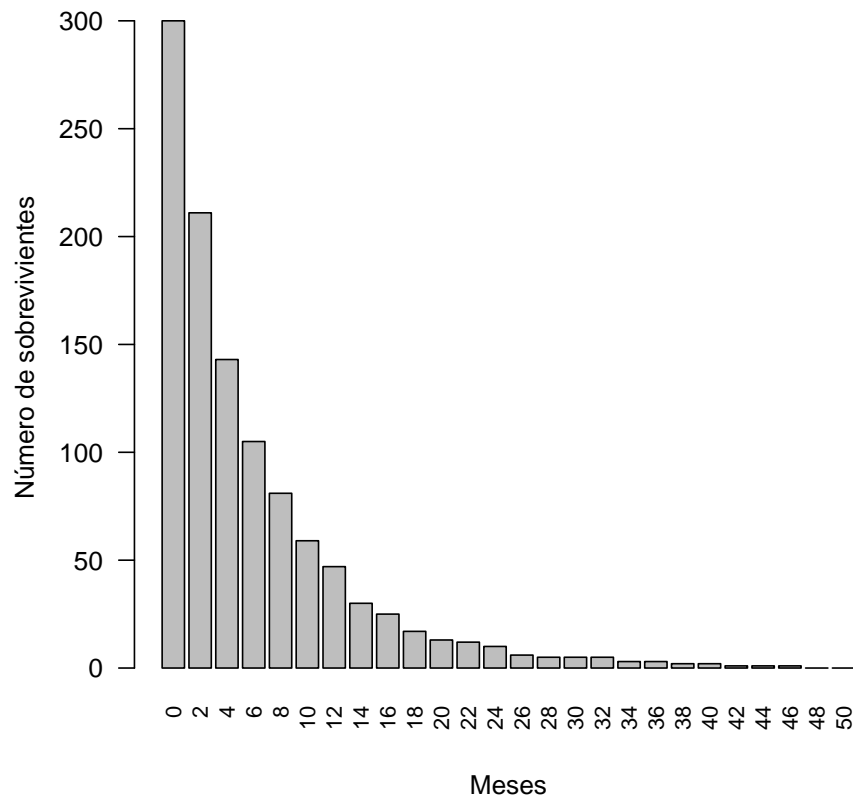
```
[1] 6.667
```

El valor esperado es similar al obtenido al valor empírico obtenido de la muestra de datos que simulamos a partir de nuestro parámetro poblacional.

Observemos la distribución de frecuencias del tiempo de vida de los individuos que componen la población.



Y el número de individuos que van quedando vivos a medida que pasa el tiempo.

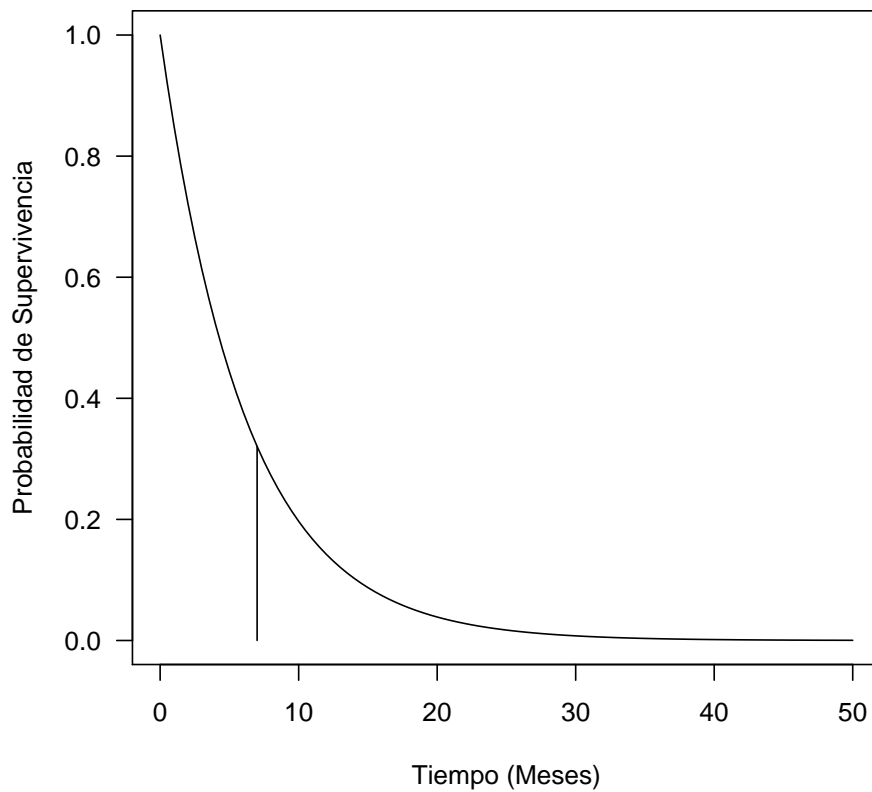


Podríamos estar interesados en conocer la probabilidad de supervivencia en un tiempo determinado. ¿Cuál es la probabilidad de que un individuo sobreviva 7 meses?

```
> 0.85^7
```

```
[1] 0.3206
```

Grafiquemos la probabilidad de supervivencia en función del tiempo.



Analicemos nuestros datos con un modelo de distribución exponencial, el cual asume una tasa constante en la tasa de mortalidad *per capita* [1].

```
> library(bbmle)
> exp.fun <- function(rate) {
+   -sum(dexp(t.vida, rate = rate, log = TRUE))
+ }
> m.exp <- mle2(exp.fun, start = list(rate = 1/mean(t.vida)))
> coef(m.exp)

rate
0.1623

> confint(m.exp)

2.5 % 97.5 %
0.1446 0.1814
```

La tasa de mortalidad estimada es ligeramente mayor que el valor usado en la simulación. Sin embargo, los intervalos de confianza al 95 % abarcan el valor de mortalidad usado en la simulación.

### 3.2. Mortalidad *per capita* variable en el tiempo: Distribución Weibull

Existen casos en donde la tasa de mortalidad *per capita* varía a lo largo del tiempo. Podemos pensar en especies animales (*e.g.* venados) en donde la vejez puede aumentar la probabilidad de los individuos a ser depredados.

```
> t.vidaw <- rweibull(300, shape = 1.3, scale = 7)
> mean(t.vidaw)
```

```
[1] 6.602
```

Para la distribución de tipo Weibull el valor esperado es igual a [1]:

$$\mu = \text{escala} \cdot \Gamma(1 + 1/\text{forma}) \quad (2)$$

donde  $\Gamma$  es la función gamma. Calculemos el tiempo de vida esperado para los datos con distribución Weibull con el parámetro de escala = 7 y de forma 1.3.

```
> tvid.weib = function(shape, scale) {
+   scale * gamma(1 + 1/shape)
+ }
> tvid.weib(1.3, 7)
```

```
[1] 6.465
```

Comparemos ahora el ajuste que los modelos con distribución exponencial y Weibull brindan a nuestros datos.

```
> library(bbmle)
> exp.fun <- function(rate) {
+   -sum(dexp(t.vidaw, rate = rate, log = TRUE))
+ }
> weibull.fun <- function(shape, scale) {
+   -sum(dweibull(t.vidaw, shape = shape, scale = scale, log = TRUE))
+ }
> m.exp <- mle2(exp.fun, start = list(rate = 1/mean(t.vidaw)))
> m.weib <- mle2(weibull.fun, start = list(shape = 1, scale = mean(t.vidaw)))
> AICtab(m.exp, m.weib, base = T)
```

	AIC	df	dAIC
m.weib	1713.4	2	0.0
m.exp	1734.5	1	21.1

```
> coef(m.weib)
```

```
shape scale
1.265 7.087
```



El modelo con distribución Weibull presenta un soporte mucho mayor que el modelo exponencial ( $\Delta AIC = 21.1$ ), por lo cual seleccionamos el modelo basado en la hipótesis de que existe variación en el riesgo de morir a lo largo del tiempo.

Veamos la incertidumbre asociada a nuestros parámetros estimados calculando el intervalo de confianza al 95 %.

```
> confint(m.weib)

      2.5 % 97.5 %
shape 1.153  1.383
scale 6.444  7.780
```

Los intervalos abarcan los valores usados en la simulación de los datos, forma = 1.3, escala = 7.

## 4. Análisis de Kaplan-Meier y datos censados

Existen ocasiones en las cuales a pesar de realizar un diseño en el cual teóricamente sabemos el destino todos los individuos, perdemos la información de algunos de ellos en un momento dado. Un ejemplo de esta situación ocurre con los análisis de supervivencia con métodos de telemetría. En estos estudios el radiotransmisor colocado en los individuos puede fallar y entonces solo tendremos certeza de que el individuo se encontraba vivo hasta el momento del último registro del radiotransmisor. En este caso decimos que el individuo fue censado. Simulemos una población de 50 individuos con una tasa de mortalidad mensual de 0.1 [4]. Los individuos tendrán la posibilidad de ser censados para lo cual crearemos una variable llamada estatus. Si conocemos el tiempo en que ocurrió la muerte de la individuo este tendrá un estatus = 1, caso contrario tendrá un estatus = 0.

```
> t.vidak <- rexp(50, rate = 0.1)
> t.censok <- 7 + 4 * runif(50)
> estatus <- as.numeric(t.censok > t.vidak)
> preced <- pmin(t.censok, t.vidak)
```

Mostramos las primeras filas de nuestra base de datos: la variable `preced` representa el tiempo de vida o el tiempo de vida registrado por el radiotransmisor en los individuos censados (el radiotransmisor falló antes de su muerte). Para distinguir entre esas dos posibilidades usamos la variable `estatus`:

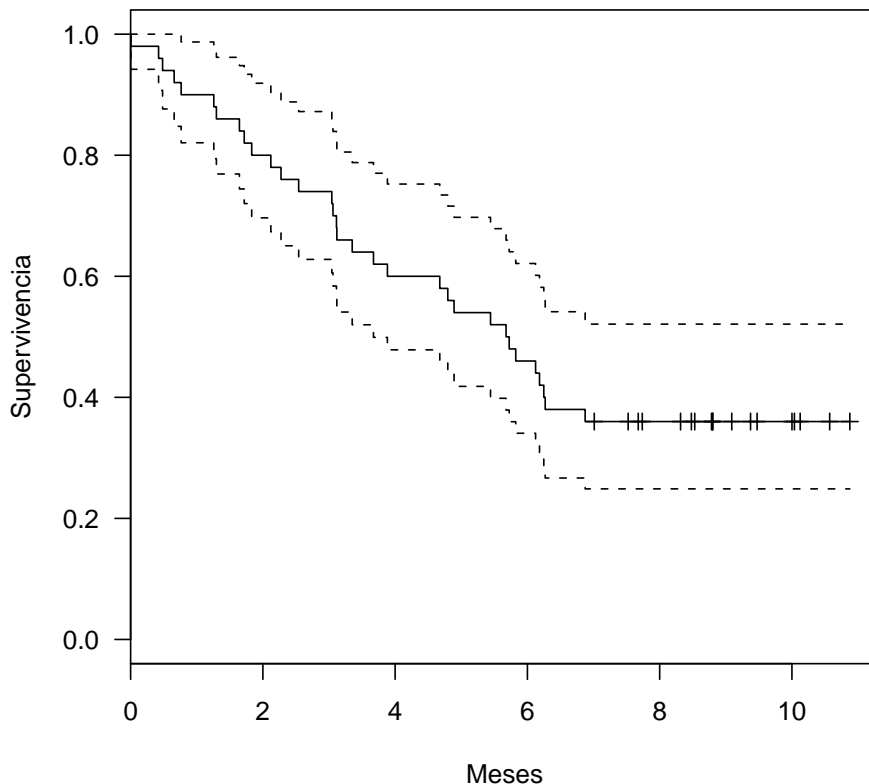
```
> telem <- data.frame(t.vidak, t.censok, preced, estatus)
> telem[1:10, ]
```

	t.vidak	t.censok	preced	estatus
1	5.7247	9.678	5.7247	1
2	6.2480	7.431	6.2480	1
3	17.6490	8.802	8.8024	0

4	0.4831	9.046	0.4831	1
5	8.9868	8.802	8.8017	0
6	0.4226	10.389	0.4226	1
7	33.4482	7.011	7.0109	0
8	3.0425	10.823	3.0425	1
9	2.1205	10.042	2.1205	1
10	1.8308	7.745	1.8308	1

Veamos la distribución de los tiempos de vida usando el estimador Kaplan-Meier. El análisis Kaplan-Meier es de tipo no paramétrico y considera en el análisis los datos censados. Usemos el paquete `survival` para tal fin:

```
> library(survival)
> par(las = 1)
> m.km <- survfit(Surv(telem$preced, telem$estatus) ~ 1)
> plot(m.km, xlab = "Meses", ylab = "Supervivencia")
```



En la gráfica observamos las probabilidades de supervivencia y los intervalos de confianza al 95% como líneas puntuadas. Cada línea vertical que corta la curva de supervivencia representa el momento en que algún individuo fue censado. Es posible incorporar predictores categóricos en los análisis Kaplan-Meier usando pruebas de rango

logarítmicas, aunque para tal fin es preferible utilizar los modelos paramétricos vistos anteriormente o semiparamétricos como usar la función Cox. Esto se debe a la versatilidad de los métodos paramétricos y semiparamétricos en incorporar múltiples predictores y sus interacciones. Para mayor información sobre los análisis de Cox ver el capítulo relativo a supervivencia del libro de John Fox [2] disponible en la red<sup>1</sup>.

## Referencias

- [1] B.M. Bolker. *Ecological models and data in R*. Princeton University Press, 2008.
- [2] J. Fox. *An R and S-Plus companion to applied regression*. Sage Publications, 2002.
- [3] R. Munguía-Steyer, M.E. Favila, and R. Macías-Ordóñez. Brood Pumping Modulation and the Benefits of Paternal Care in *Abedus breviceps* (Hemiptera: Belostomatidae). *Ethology*, 114(7):693–700, 2008.
- [4] M. Zhou. Use Software R to do Survival Analysis and Simulation. A tutorial, 1989.

---

<sup>1</sup><http://socserv.mcmaster.ca/jfox/Books/Companion/index.html>