

Tiempo hasta que ocurre un suceso Análisis de supervivencia

Preparado por Luis M. Molinero (Alce Ingeniería)

CorreoE: bioestadistica@alceingenieria.net

[Artículo en formato PDF](#)

Junio 2001

Descripción del problema

Es muy común que la variable que se desea estudiar sea el tiempo que tarda en ocurrir un suceso, ya sea éste beneficioso (curación, alta hospitalaria) o perjudicial (muerte, aparición de efecto adverso, rechazo del trasplante) o incluso indiferente (cambio de tratamiento). El conjunto de técnicas estadísticas que se emplean para analizar este tipo de datos se conoce en medicina como "análisis de supervivencia", debido a que se diseñaron inicialmente para estudiar el tiempo hasta el fallecimiento del paciente, fundamentalmente en el campo de la oncología.

Para centrar la exposición supongamos que el dato de interés es el tiempo hasta la aparición de problemas microvasculares en un conjunto de pacientes diabéticos tipo II hipertensos, en un estudio prospectivo durante 5 años. Escogemos intencionadamente un ejemplo en el que el suceso no es el fallecimiento del paciente para ilustrar precisamente el uso de estas técnicas fuera del ámbito del concepto estricto de supervivencia.

La característica más importante de este tipo de datos (tiempo hasta que ocurre un suceso) es que, muy probablemente, al final del periodo de observación no todos los pacientes habrán presentado el suceso objeto de estudio. Además puede ocurrir que algunos pacientes se hayan perdido por causas diversas, no habiendo sido posible determinar su estado. O cuando el suceso es la muerte pueden haber fallecido por causas diferentes a las que se analizan (por ejemplo en un accidente de coche).

También es habitual que los pacientes vayan incorporándose durante todo el periodo de observación, por lo que los últimos en hacerlo serán observados durante un periodo de tiempo menor que los que entraron al principio y por lo tanto la probabilidad de que les ocurra el suceso es menor.

Y por último, al final del estudio habrá pacientes que no presentan el suceso.

Es intuitivo que con este tipo de datos no podemos usar los métodos estadísticos habituales para variables cuantitativas, como pueden ser el cálculo de medias y su comparación mediante la t de Student. Así en el ejemplo planteado ¿qué sentido tendría calcular el tiempo medio hasta la aparición de problemas microvasculares, cuando no todos los pacientes han sido observados durante el mismo periodo y además hay pacientes que no llegan a tenerlos? ¿Y qué hacemos con los pacientes que no se observaron hasta el final?

Precisamente esas observaciones incompletas, todos los pacientes que hasta el último momento en que fueron observados (bien sea al final del estudio o antes, si se perdieron por alguna causa) y no habían desarrollado el suceso, tienen también importancia, y el tiempo durante el que fueron observados debe intervenir en el análisis. En la terminología inglesa se les denomina observaciones "**censuradas**" (**censored**).

Si todos los pacientes estudiados experimentan el suceso durante el periodo de observación (en nuestro ejemplo todos en algún momento llegan a presentar problemas microvasculares), es fácil calcular la proporción de pacientes que transcurrido un tiempo determinado desde que comenzó su observación no han llegado a tener problemas microvasculares. Podríamos representar gráficamente la proporción de pacientes sin

problemas en función del tiempo desde que entraron en el estudio.

Función de Supervivencia calculada por el método de Kaplan–Meier

Pero las cosas no son tan sencillas si no todos los pacientes presentan el suceso. Un paciente puede haber entrado a mitad del estudio, siendo observado durante 3 años, y al finalizar éste no presentar problemas microvasculares. No sabemos entonces que habría pasado con el paciente si lo siguiéramos 2 años más, hasta completar los 5 años. Ahora ya no es tan simple determinar la proporción de pacientes con problemas microvasculares a los 5 años desde la inclusión en el estudio. Por ello se utiliza el denominado método de **Kaplan–Meier**, que se basa en algo que es obvio: para sobrevivir un año hay que sobrevivir cada uno de los días de éste. Calculamos entonces para cada día la proporción de sucesos que se observan en ese día.

Utilizando el concepto de probabilidad condicional decimos que para vivir 31 días hay que vivir 30 días y luego un día más. En estadística esto se calcula multiplicando las probabilidades. La probabilidad de vivir una semana vendrá dada por

$$p_1 \times p_2 \times p_3 \times p_4 \times p_5 \times p_6 \times p_7$$

Siendo p_1 la tasa de supervivencia el primer día, p_2 la del segundo, etc. La tasa de supervivencia para un día dado, por ejemplo para el séptimo, se calcula como el cociente entre el número de pacientes vivos el día 7 (o que no experimentaron el suceso) de entre los que estaban vivos el día anterior, el 6. En nuestro ejemplo donde dice número de pacientes vivos leemos número de pacientes sin problemas microvasculares.

Así que la supervivencia se calculará mediante la siguiente fórmula recursiva:

$$S(t_i) = \frac{r_i - m_i}{r_i} \cdot S(t_{i-1})$$

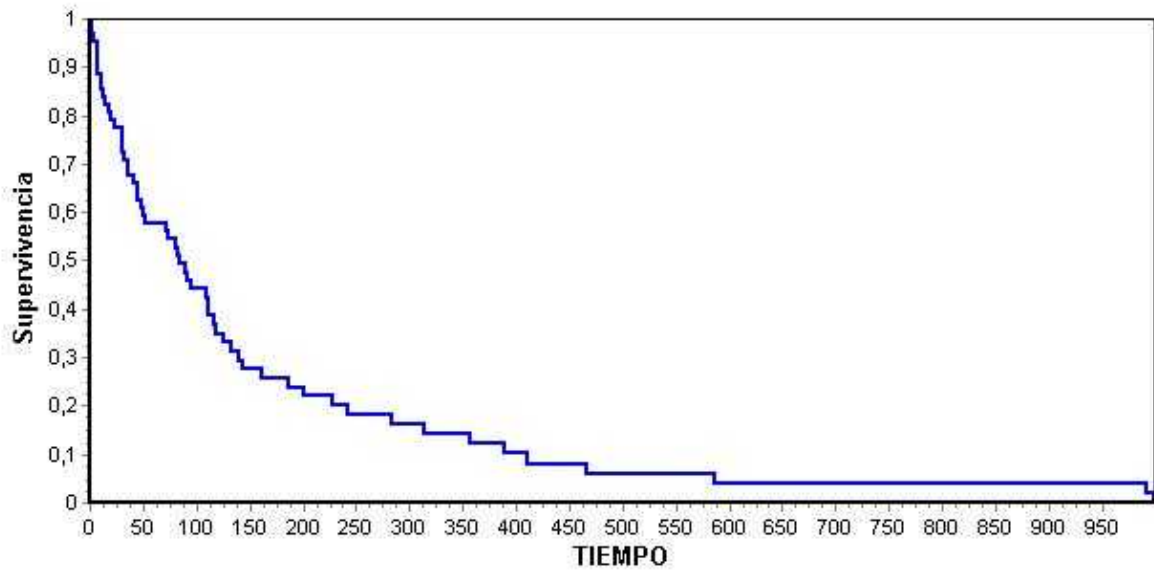
para cada instante de tiempo la supervivencia se calcula como la supervivencia en el instante anterior multiplicada por la tasa de supervivencia en ese instante. En el denominador tenemos el número de pacientes r que continuaban en el estudio en el instante anterior (expuestos al riesgo) y en el numerador a ese valor se resta el número de pacientes m que presentan el suceso en ese instante.

A lo largo del tiempo van saliendo del estudio los pacientes que presentan el suceso y aquellos para los que finaliza la observación por otras causas (incompletos).

Los valores de la curva de supervivencia sólo hay que calcularlos para aquellos momentos en los que se produce algún suceso, ya que en el resto de casos el numerador y el denominador coinciden y por tanto el cociente vale 1 y la supervivencia es igual que en el instante anterior, no cambia.

Si hay observaciones incompletas (censuradas o pérdidas) hacen que vaya cambiando el número de pacientes expuestos al riesgo, pero si en un instante determinado sólo hay observaciones incompletas y no hay sucesos la curva de supervivencia no cambia al ser $m = 0$, aunque r haya cambiado, lo que se tendrá en cuenta en el siguiente instante en el que se observen sucesos ya que habrá menos pacientes expuestos.

Aunque la tasa de supervivencia p calculada en nuestra muestra para cada instante individual es una estimación muy burda del valor verdadero, sin embargo el producto de muchas de esas tasas constituye una estimación bastante precisa de la curva de supervivencia.

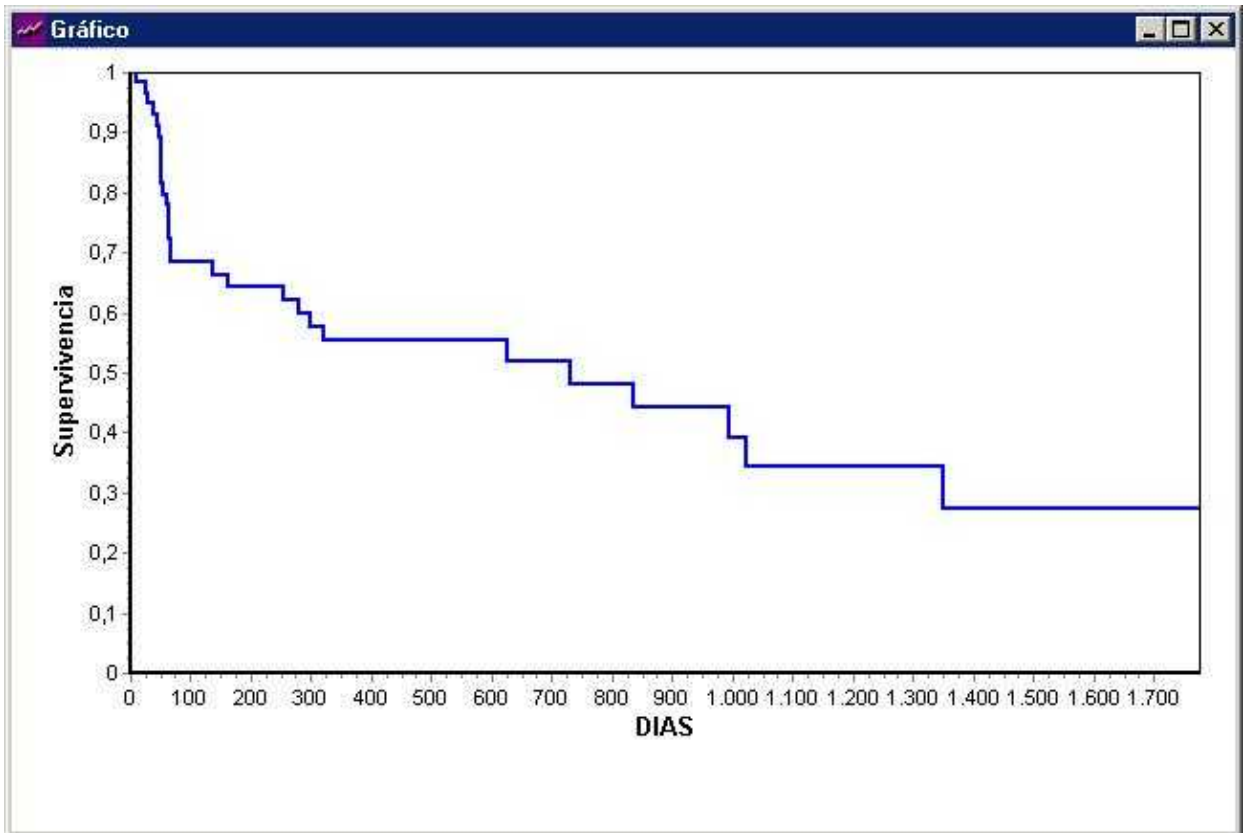


Ejemplo de curva de supervivencia

A pesar de que la información que proporciona la curva de supervivencia es mucho más fiable que las tasas individuales que se han utilizado para calcularla, se pueden producir en ésta grandes saltos o grandes zonas planas, cuyo intento de explicación puede ser complicado, sobre todo si el número de pacientes que permanecen es pequeño. Esta advertencia cobra especial importancia en las zonas de la derecha de la curva donde el efectivo de muestra puede ser muy pequeño

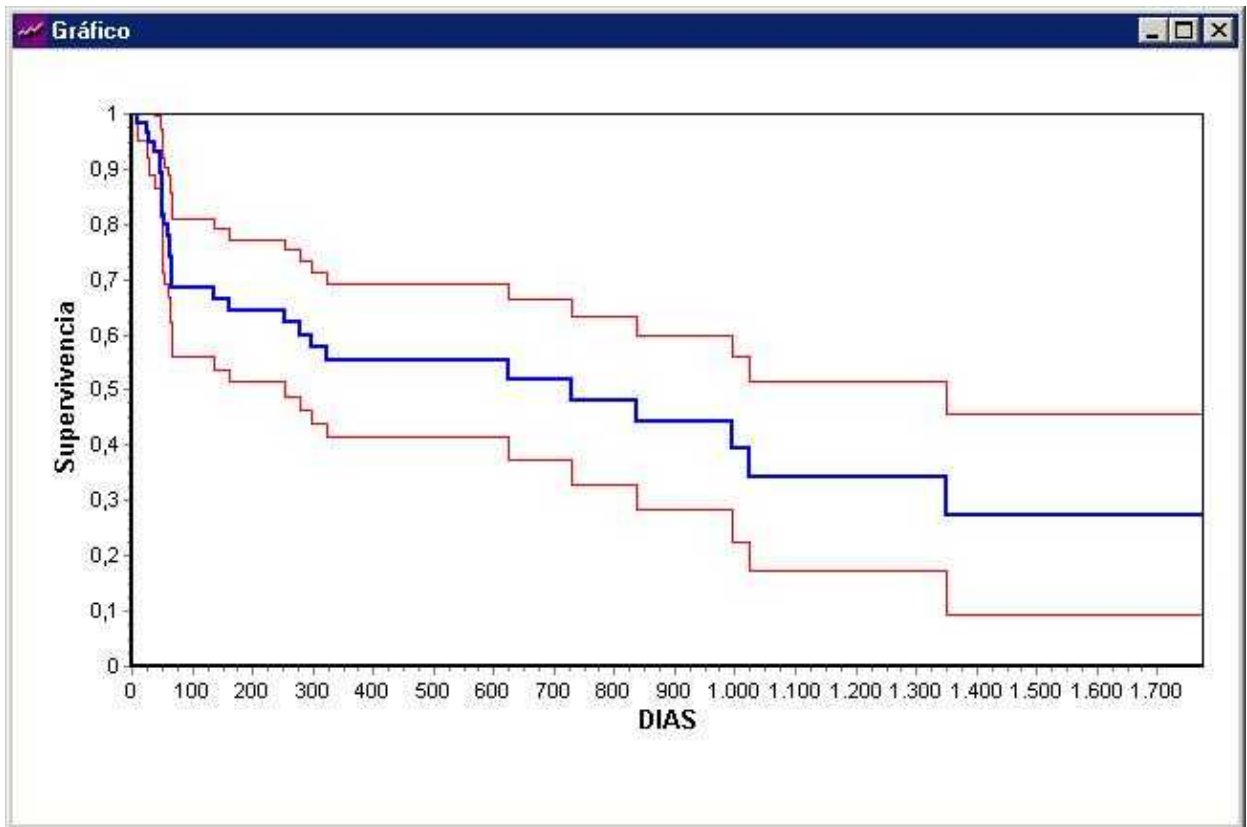
Seguidamente se presenta un ejemplo de una tabla con los datos numéricos de una curva de supervivencia

Tiempo	Supervivencia	Err.est.S.	Nº Sucesos	Expuestos
10	0,9836	0,0163	1	61
25	0,9660	0,0236	1	56
29	0,9482	0,0292	1	54
39	0,9299	0,0338	1	52
46	0,9113	0,0379	1	50
47	0,8927	0,0415	1	49
50	0,8547	0,0476	2	47
51	0,8167	0,0525	2	45
54	0,7978	0,0546	1	43
60	0,7788	0,0565	1	42
63	0,7598	0,0583	1	41
64	0,7408	0,0598	1	40
65	0,7218	0,0612	1	39
66	0,7028	0,0625	1	38
68	0,6838	0,0636	1	37
136	0,6637	0,0649	1	34
161	0,6436	0,0659	1	33
253	0,6214	0,0673	1	29
280	0,5992	0,0685	1	28
297	0,5770	0,0694	1	27
322	0,5539	0,0704	1	25
624	0,5193	0,0740	1	16
730	0,4822	0,0775	1	14
836	0,4420	0,0808	1	12
994	0,3929	0,0854	1	9
1024	0,3438	0,0877	1	8
1350	0,2750	0,0933	1	5
1775	0,2750	0,0933	0	0



Se trata de una curva obtenida con datos reales que presenta intervalos de tiempo largos en los que no aparece ningún suceso. Es esta una característica común a las curvas de supervivencia obtenidas a partir de muestras de pequeño tamaño (en nuestro caso comienzan el estudio 61 pacientes). Hay que destacar que **las conclusiones que se extraigan de los detalles particulares en diferentes puntos de este tipo de curvas de supervivencia no son fiables**, sobre todo en los extremos de ésta, donde la estimación es menos precisa. **La curva hay que interpretarla en su aspecto general.**

Con el fin de determinar la precisión de la estimación es posible calcular un intervalo de confianza aproximado a partir del error estándar, que se suele indicar en las salidas de ordenador. En la figura siguiente vemos representadas las bandas de confianza para el 95 % de la curva anterior



Vemos que a los 1.350 días la supervivencia cambia de 0.3438 a valer 0.275 ya que a los 1350 días quedan en el estudio (expuestos al riesgo) 5 pacientes y se observa 1 suceso por lo que el nuevo valor de supervivencia vendrá dado por

$$0.3438 \times \frac{5-1}{5}$$

que es igual a 0.275. Sin embargo se trata de una estimación muy imprecisa ya que el límite inferior del intervalo de confianza aproximado es de 0.092.

En las curvas a veces se representa también mediante un punto o una pequeña raya vertical los instantes para los que existen observaciones incompletas (censuradas).

Los programas de ordenador calculan el error estándar de la supervivencia mediante el método de Greenwood, pero si sólo se dispone de la tabla de supervivencia es posible calcular el error estándar de una forma más simple. Si en un momento determinado el valor de la curva de supervivencia es S y hay N pacientes que continúan en el estudio, el error estándar de S es aproximadamente

$$S \cdot \sqrt{\frac{(1-S)}{N}}$$

El razonamiento para esta fórmula es muy sencillo y se basa en que para que en ese momento la supervivencia sea S y continúen N pacientes, es necesario que inicialmente hubiera al menos N/S pacientes. Si fuera exactamente así (no hay observaciones incompletas), la fórmula anterior no es más que la habitual del error estándar de una proporción, basándonos en la distribución binomial.

Comparación de curvas de supervivencia

Para comparar dos curvas de supervivencia de forma global se suele emplear la prueba denominada **log-rank**.

El razonamiento empleado en esta prueba es bastante fácil de comprender. Si se está comparando dos grupos de pacientes, y por ejemplo en un momento del estudio del total de pacientes que permanecen 1/4 son del grupo A y el resto 3/4 son del grupo B, y no hubiera diferencias en cuanto a la supervivencia entre los grupos, es de esperar que del total de sucesos que ocurran en ese momento 1/4 correspondan al primer grupo y 3/4 al segundo. Aunque verdaderamente no hubiera diferencias en cuanto a la supervivencia, el número real de sucesos observados en cada momento de tiempo puede no coincidir con lo que es más probable, y así para el grupo A unas veces el número de sucesos será superior al esperado y otras será inferior, pero a la larga esas diferencias se equilibrarán, como las series de números en una ruleta.

Para efectuar los cálculos se ordenan cronológicamente las observaciones de los dos grupos de forma combinada, como si de un solo grupo se tratase, al igual que se hace para el cálculo de la curva de supervivencia. Para cada instante en el que se observa algún suceso se determina el número total de pacientes que continúan en el estudio **r** y cuántos son del primer grupo **a**. Si el número de sucesos que se observan en ese instante es **d**, el número esperado para el primer grupo es

$$E_A = \frac{a_i \cdot d_i}{r_i}$$

con el subíndice **i** se indica que este cálculo se repite para todos los instantes de tiempo en los que se produce algún suceso, pero no para aquellos en los que sólo hay observaciones incompletas, que sólo contribuirán a reducir el número de pacientes expuestos a considerar en el siguiente momento en el que se producen sucesos. Se calcula finalmente la suma de todos los sucesos esperados para el grupo A

$$E_A = \sum E_A$$

Si **D** es el número total de sucesos observados en el estudio, considerando ambos grupos, el número de sucesos esperado para el grupo B podemos calcularlo por diferencia

$$E_B = D - E_A$$

Para contrastar la hipótesis nula (hipótesis de que el riesgo es el mismo en ambos grupos) se calcula

$$\chi^2 = \frac{(O_A - E_A)^2}{E_A} + \frac{(O_B - E_B)^2}{E_B}$$

que se distribuye como una χ^2 con 1 grado de libertad.

Este método es también válido para comparar más de dos grupos.

En uno de los [enlaces](#) que se citan al final de esta página se puede ver un ejemplo de cálculo de esta prueba.

Enlaces de interés

- [Time to event \(survival\) data](#)
Altman DG, Bland JM. BMJ 1998;317:468–469
- [Survival probabilities \(the Kaplan–Meier method\)](#)
Bland JM, Altman DG. BMJ 1998;317:1572
- [Statistics at Square One: Survival analysis](#)
- [Kaplan–Meier Survival Analysis](#)
- [Survival in treated hypertension: follow up study after two decades](#)
Andersson OV et al. BMJ 1998;317:167–71
- [Calculadora on–line para comparar dos curvas de supervivencia](#)

Bibliografía seleccionada

- Peto R, Pike MC, Armitage P et al. Design and analysis of randomized clinical trials requiring prolonged observation of each patient: II. Analysis and examples. Br. J Cancer 1977; 35:1–39.
- Kleinbaum DG. Statistics in the health sciences: Survival analysis. New York: Springer_Verlag; 1996
- Parmar MKB, Machin D. Survival analysis: a practical approach. Chichester: Wiley.



[Indice de artículos](#)

[Principio de la página](#) ▲