



16

Análisis de supervivencia

Ágata Carreño Serra

16.1. Introducción

En variados estudios nuestro objetivo consiste en estimar el tiempo transcurrido hasta un suceso en concreto, dependiendo de otras variables explicativas. En un principio, cuando se estudiaban los éxitos de una determinada patología, se estudiaba el tiempo transcurrido hasta la muerte, de manera que se pretendía conocer el tiempo de supervivencia. Con esta particularidad, se ha generalizado el término *tiempo de supervivencia* para englobar cualquier análisis en que la variable de interés sea el tiempo hasta un evento determinado, sea muerte o, por ejemplo, tiempo hasta la curación de una determinada enfermedad con dos tratamientos alternativos o bien tiempo hasta la aparición de una enfermedad o recaída.

El tiempo no deja de ser una medida cuantitativa (número de años, meses o días) y podría ser estudiada mediante las técnicas que hemos visto en capítulos anteriores como el análisis de la varianza o los modelos de regresión. Sin embargo, existen algunas particularidades en este tipo de análisis que deben tenerse en cuenta. En primer lugar, la variable tiempo no acostumbra a seguir una distribución normal y, en segundo lugar, durante el estudio existen pérdidas que impo-

225



sibilitan la observación del evento de interés. Es decir, en un estudio de mortalidad, probablemente al finalizar el tiempo de seguimiento no conoceremos el estado de todos los pacientes, puesto que algunos de ellos no completarán el seguimiento, otros todavía estarán vivos y en otros su causa de muerte no será la que se quiere analizar en el estudio. Por lo tanto, conocemos el tiempo transcurrido hasta la muerte en aquellos pacientes con *éxitus* documentado y sabemos que en los pacientes en los que no se ha producido el evento, por lo menos sobrevivieron a la última vez que los observamos con vida.

16.2. Datos censurados

Como hemos comentado a modo de introducción, el tiempo de seguimiento puede variar considerablemente de un paciente a otro. Cuando el tiempo de seguimiento termina antes de producirse el evento o antes de completar el período de observación de nuestro estudio, podemos decir que ésta es una observación censurada. Las causas de la censura suelen ser varias pero siempre responde a alguno de estos criterios:

- o existe una pérdida de seguimiento (por ejemplo, el paciente decide no participar más en el estudio).
- o el estudio termina antes de que aparezca el evento.
- o se produce el evento sin ser éste motivo de estudio (por ejemplo, se está estudiando el tiempo hasta la muerte por una determinada enfermedad, y se produce un *éxitus* causado por accidente de tráfico).

Dado que estamos estudiando el tiempo de seguimiento, deben quedar muy claras las fechas de inicio del seguimiento y del último valor conocido para cada individuo, puesto que nos darán el tiempo en el cual por lo menos el paciente ha permanecido vivo.

Supongamos, por ejemplo que queremos estudiar el tiempo que transcurre hasta que hay una recaída o reinfección tras la administración de un antibiótico para el tratamiento de infecciones urinarias. Se ha previsto una duración del tiempo de seguimiento de 7 meses. El esquema del seguimiento se muestra en la figura 68. El paciente 1 se incluye en el estudio a los 3 meses y medio de haber comenzado el seguimiento y cuando finaliza el estudio todavía estaba libre de enfermedad, por lo que se ha seguido durante 2 meses y medio. Por el contrario, el paciente 2 ha iniciado el estudio correctamente y ha tenido una recaída a los 5 meses. El paciente 3 decidió abandonar el estudio al cabo de dos meses de iniciarlo, el paciente 4 tuvo una recaída al cabo de 3 meses y el paciente 5 completa todo el periodo de seguimiento sin haber tenido ninguna recaída.



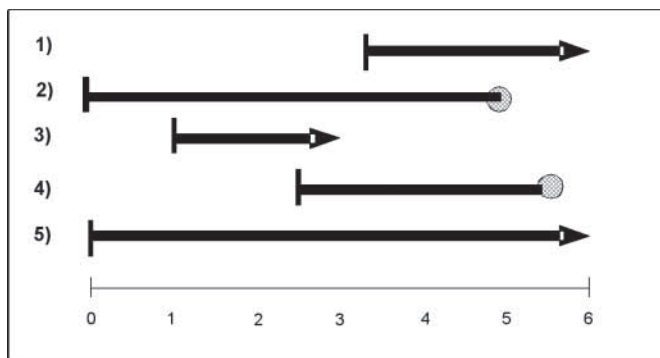


Figura 68. Esquema del periodo de seguimiento de los pacientes durante un estudio de supervivencia

De esta forma, sabemos que los pacientes 2 y 4 han tenido recaídas a los 5 y 3 meses de finalizar el tratamiento, respectivamente. Y los pacientes 1, 3 y 5 no sabemos si han tenido recaídas desde que completaron su tratamiento pero lo que sí sabemos es que en la última observación de la que disponemos todavía no habían recaído: 2.5+, 2+ y 6+ (el símbolo + indica que el tiempo de recaída es superior al indicado). Estos datos son datos censurados puesto que no se ha observado el evento.

Sin embargo, los datos censurados no son las únicas limitaciones de nuestros datos, puesto que también podemos encontrarnos con que lo que no está bien definido es el origen, es decir, en nuestro caso se finaliza el antibiótico anteriormente al inicio del estudio, por lo que el tiempo hasta la recaída, está “censurado por la izquierda”.

El acontecimiento o suceso estudiado también debe estar perfectamente definido para poder determinar exactamente la fecha del mismo. Este evento está casi siempre asociado a la muerte del paciente pero no tiene por que ser así, ya que puede hacer referencia también a la fecha de alta, la fecha de remisión de la enfermedad, la fecha de recidiva, la fecha de recaída o fallo, etc.

Para planificar correctamente un estudio de supervivencia, deberemos definir apropiadamente el origen o inicio del seguimiento, la escala de tiempo que vamos a utilizar y el evento que queremos analizar para que no se introduzcan eventos debidos a otras causas que no son de nuestro interés.

16.3. Metodología estadística

En la mayoría de los casos, los investigadores analizan los datos antes de que todos los pacientes hayan muerto o hayan presentado el evento de interés, ya





que si no habría que esperar muchos años para poder obtener todos los tiempos de supervivencia. Además, los pacientes no inician el tratamiento o entran al estudio al mismo tiempo, de manera que los datos aportados por los pacientes vivos, como se señaló previamente, son observaciones “censuradas” y deben considerarse como tales a la hora de analizarlas.

Por estos motivos, se necesitan métodos especiales de análisis, al igual que en la metodología estadística básica se señalaba la existencia de pruebas paramétricas y no paramétricas. En el análisis de supervivencia el análisis de los datos también puede ser realizado utilizando técnicas paramétricas, basadas en distribuciones más complejas que las estudiadas, como son la exponencial, Weibull y lognormal, así como las no paramétricas, entre los que se encuentra la metodología de Kaplan-Meier y la regresión de Cox.

Estos últimos son los métodos estadísticos más utilizados. Así, las curvas de supervivencia por lo general se producen usando uno de dos métodos: el análisis actuarial o el método de Kaplan-Meier, que son los que nos ocuparán en este capítulo.

El método actuarial implica dos premisas en los datos: la primera es que todos los abandonos durante un intervalo dado ocurren aleatoriamente durante dicho intervalo. Esta premisa es de escasa importancia cuando se analizan intervalos de tiempo cortos, sin embargo, puede haber un sesgo importante cuando los intervalos son grandes, si hay numerosos abandonos o si los abandonos no ocurren a mitad del intervalo. La segunda premisa es que aunque la supervivencia en un tiempo dado depende de la supervivencia en todos los períodos previos, la probabilidad de la misma en un período de tiempo es independiente de la probabilidad de supervivencia en los demás períodos. De esta forma, en el método actuarial se divide el tiempo en intervalos y se calcula la supervivencia en cada intervalo.

El método de Kaplan-Meier se utiliza cuando la muestra es menor de 30 o bien para muestras mayores de 30 y se conocen los tiempos individuales de los censurados y no censurados, de manera que se calcula la supervivencia cada vez que un paciente muere o alcanza el evento.

Toda la metodología del análisis de supervivencia se basa en obtener una función dependiente del tiempo, cuyos valores proporcionen la probabilidad de que el suceso final (muerte, recaída, etc.) no ocurra hasta pasado un periodo de tiempo superior o igual a t o, del mismo modo, si denominamos “fallecer” a la ocurrencia del suceso final, y “sobrevivir” a la no ocurrencia, se trata de obtener una función cuyos valores proporcionen la probabilidad de sobrevivir por lo menos hasta el instante t .





A simple vista, parece razonable estimar la proporción de pacientes que sobreviven a un determinado tiempo t como el número de individuos cuyo tiempo de observación ha sido superior o igual a t , es decir, los que “sobreviven” sobre todos los pacientes de los que disponemos información hasta por lo menos ese instante.

16.3.1. Método actuarial

Supongamos que, para los pacientes transplantados con insuficiencia renal terminal (IRT), se desea estimar la probabilidad de que el injerto sea rechazado en función del tiempo transcurrido desde su trasplante. Para ello, se obtienen las historias clínicas del registro español de pacientes con IRT, habiendo sido tratados mayoritariamente con dos tipos de terapias. Con esta información obtenemos si el paciente había sobrevivido en la última visita o bien había fallecido. Antes de ser tratados post-trasplante, algunos pacientes habían sido transplantados previamente, por lo que se sospecha que, para un mismo tiempo, la probabilidad de que el trasplante sea exitoso depende también de si ha habido un trasplante anterior.

El método actuarial, también conocido como el método de la tabla de vida consiste en estimar, para un determinado periodo de tiempo la probabilidad de que el suceso a analizar no suceda hasta transcurrido un periodo de tiempo como mínimo igual al que se está analizando. Es decir, si volvemos a nuestro ejemplo, se pretende conocer la probabilidad de que un individuo transplantado, sobreviva a los 12, 24, 36 meses, etc.

Para cada individuo de la muestra se dispone del tiempo de observación y una variable dicotómica que indica la ocurrencia o no del suceso final (1=fallecer y 0=sobrevivir). Con esta información, la probabilidad de que el suceso final ocurra hasta pasado un determinado período de tiempo superior o igual a t podría estimarse como el total de individuos cuyo tiempo de observación ha sido superior o igual a t (la proporción de pacientes que sobreviven al instante t), sobre todos los pacientes de los que disponemos información hasta por lo menos ese instante. Con el método actuarial, los tiempos de supervivencia se agrupan en intervalos, por lo que debe indicarse su longitud.

La tabla de cálculo para estimar la función de supervivencia por el método actuarial para nuestro ejemplo se solicita en el paquete SPSS mediante el cuadro de diálogo que aparece en la Figura 69, señalando como variable de estado nuestra variable EXITUS (donde el valor 1 indica que ha ocurrido el suceso), el factor por el que se quiere estratificar, en nuestro caso TRATAM y el tiempo (meses) transcurridos desde el trasplante a la última observación conocida, indicando





la longitud de los intervalos que queremos analizar. En nuestro caso, puesto que estamos analizando meses transcurridos hasta el *éxitus*, indicaremos una periodicidad de 6 meses.

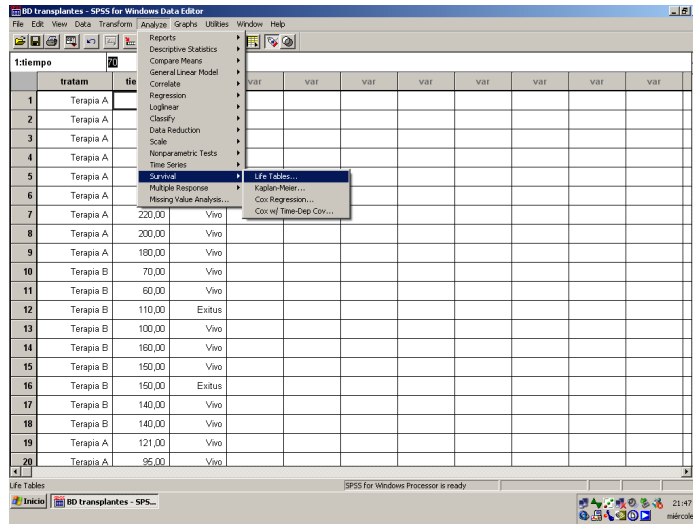


Figura 69. Obtención de las tablas de vida por el método actuarial en SPSS

16.3.1.1. Descripción de la información obtenida por el método actuarial

Para ilustrar el proceso de estimación de la función de supervivencia sólo mostraremos los resultados obtenidos para el primer grupo, es decir los pacientes con la terapia A. La información reflejada en forma de tabla muestra en intervalos de tiempo de 6 meses, iniciando en 0 y hasta el máximo definido, podemos explicarla tal como sigue a continuación (Figura 70). Por ejemplo, en la 5ª fila correspondiente a tiempos de supervivencia iguales o superiores a 24 meses e inferiores a 30, los elementos considerados son:

- o **Número entradas («Number Entrng this Intrvl»)**. Número de pacientes para los que, hasta ese instante, con certeza aún estaban vivos.
- o **Número de abandonos («Number Wdrawn During Intrvl»)**. Número de pacientes tales que se sabe que no han fallecido en un tiempo inferior a 18 meses, pero se desconoce si fallecerán en un tiempo inferior a 24 meses.
- o **Número de fallecimientos («Number of Termnl Events»)**. Número de pacientes que han fallecido en el intervalo comprendido entre los 18 y los 24 meses.





- o **Número de individuos expuestos a riesgo («Number Exposed to Risk»)**. Número estimado de pacientes para los que hasta los 18 meses no habían fallecido pero podrían hacerlo durante este intervalo (de 18 a 24 meses).

Teniendo en cuenta que, de entre los expuestos a riesgo (13), la proporción de pacientes que no han fallecido en el intervalo señalado («Propn Surviving») es igual a 0,9231; la probabilidad de que los pacientes sobrevivan hasta pasado un plazo de tiempo superior o igual a 24 meses («Cumul Propn Surv at End» o **Función de supervivencia**) desde el trasplante renal, se estima como la proporción acumulada de pacientes para los que, al final de este período todavía continuaban con vida. Es decir, la probabilidad de los pacientes con trasplante renal tratados con la terapia A sobreviven a los 24 meses, obteniendo una probabilidad de 0,8112.





Life Table
 Survival Variable TIEMPO Tiempo de supervivencia post-transplante
 for TRATAM Tratamiento
 = I Terapia A

Intrvl Start Time	Number Entrg This Intrvl	Number Wdrawn During Intrvl	Number Exposd to Risk	Number of Termnl Events	Propn Termi- nating	Propn Sur- viving	Cumul Propn Surv at End	Proba- bility Densty	Hazard Rate
,0	19,0	,0	19,0	,0	,0000	1,0000	1,0000	,0000	,0000
6,0	19,0	1,0	18,5	,0	,0000	1,0000	1,0000	,0000	,0000
12,0	18,0	1,0	17,5	,0	,0000	1,0000	1,0000	,0000	,0000
18,0	17,0	1,0	16,5	2,0	,1212	,8788	,8788	,0202	,0215
24,0	14,0	2,0	13,0	1,0	,0769	,9231	,8112	,0113	,0133
30,0	11,0	3,0	9,5	,0	,0000	1,0000	,8112	,0000	,0000
36,0	8,0	1,0	7,5	,0	,0000	1,0000	,8112	,0000	,0000
42,0	7,0	2,0	6,0	2,0	,3333	,6667	,5408	,0451	,0667
48,0	3,0	2,0	2,0	,0	,0000	1,0000	,5408	,0000	,0000
54,0+	1,0	,0	1,0	1,0	1,0000	,0000	,0000	**	**

** These calculations for the last interval are meaningless.
 The median survival time for these data is 54,00+

Intrvl Start Time	Se of Cumul Sur- viving	Se of Proba- bility Densty	Se of Hazard Rate
,0	,0000	,0000	,0000
6,0	,0000	,0000	,0000
12,0	,0000	,0000	,0000
18,0	,0803	,0134	,0152
24,0	,0986	,0109	,0133
30,0	0,986	,0000	,0000
36,0	0,986	,0000	,0000
42,0	,1694	,0266	,0462
48,0	,1694	,0000	,0000
54,0+	,0000	**	**

Figura 70. Obtención de las tablas de vida por el método actuarial en SPSS





En las tablas de vida, también se presenta la mediana del tiempo de supervivencia, («The median survival time for these data is»), es decir, el tiempo en el que se espera que hayan fallecido al menos la mitad de los pacientes. En nuestro caso es igual a más de 54 meses con la terapia A y, aunque no se muestran estos resultados, de 39,5 con la terapia B. De esta forma, parece que la terapia seguida tras el trasplante, influye en la supervivencia de los pacientes con IRT.

Junto con la función de supervivencia, la tabla de cálculo proporciona dos funciones más: la función de densidad de probabilidad y la función de riesgo («Probability Density» y «Hazard Rate», respectivamente), funciones equivalentes. Considerando que el error estándar es una medida del error cometido en la estimación, cuanto menor sea su valor más precisa será la estimación del valor de la función correspondiente. Dado que el error estándar de los valores estimados de la función de supervivencia («SE of Cumul Surviving») no son excesivamente grandes, podemos concluir que las estimaciones son bastante precisas.

16.3.1.2. Comparación de la supervivencia en distintos grupos de individuos por el método actuarial

Si el análisis se realiza considerando varios grupos de individuos, para comparar las distribuciones de la supervivencia en los diferentes grupos, a cada individuo de la muestra se le asigna una puntuación que refleje la posición relativa de su tiempo de supervivencia. Es decir, la puntuación de supervivencia para el i -ésimo paciente será: $p_i = \frac{a_i}{b_i}$, siendo a_i y b_i el número de individuos tales que su tiempo de observación es inferior o superior a t_i , respectivamente. Las puntuaciones de supervivencia se utilizarán en el cálculo del estadístico asociado a la prueba para contrastar la hipótesis nula de que la distribución de la supervivencia es la misma en los distintos grupos. Si el p-valor asociado al estadístico de contraste es menor que α se rechazará la hipótesis nula al nivel de significación α .

Comparando simultáneamente las dos funciones de supervivencia de las dos terapias («Overall comparison», en la Figura 71), el p-valor asociado al estadístico de contraste («Prob = 0,0431») es ligeramente menor que 0,05. Luego, al nivel de significación 0,05, se rechaza la hipótesis nula de igualdad de funciones. La tabla resumen muestra además que en los pacientes que recibieron la terapia A, sobre un total de 19 pacientes («Total N»), han fallecido 6 («Uncen»=No censurado) y han sobrevivido 13 casos («Cen»=Censurado). La media de las puntuaciones de supervivencia asignadas a dichos pacientes («Mean Score = 4,158») es positiva y superior a la obtenida en la terapia B.





Comparison of survival experience using the Wilcoxon (Gehan) statistic						
Survival Variable TIEMPO Tiempo de supervivencia post-transplante						
grouped by TRATAM Tratamiento						
Overall comparison	statistic	4,092	D.F.	1	Prob.	,0431
Group	label	Total N	Uncen	Cen	Pct Cen	Mean Score
1	Terapia A	19	6	13	68,42	4,1579
2	Terapia B	17	7	10	58,82	-4,6471

Figura 71. Obtención de las tablas de vida por el método actuarial en SPSS

La siguiente figura contiene las representaciones gráficas de las dos funciones de supervivencia correspondientes a las dos terapias. Como vemos, en el eje de las X's, se presenta el tiempo de supervivencia y en el eje de las Y's, se presenta la función de supervivencia acumulada, de manera que partiendo del tiempo 0, donde todos los pacientes están vivos, es decir tienen una probabilidad 1 de sobrevivir, a medida que pasa el tiempo y hay éxitus en el estudio, la probabilidad de sobrevivir disminuye, tendiendo a 0. Por ejemplo, la función que corresponde a los pacientes que recibieron la terapia A, permanecen más tiempo en el valor 1, y se mantiene hasta casi las 25 meses, donde empieza a bajar la supervivencia de forma escalonada. Por el contrario, la función de los pacientes tratados con la terapia B, poco después de los 10 meses del implante la supervivencia baja considerablemente, observando que a los 30 meses, sólo existe un 50% de probabilidades de sobrevivir, mientras que en la terapia anterior, la probabilidad de sobrevivir se sitúa entorno al 80%.

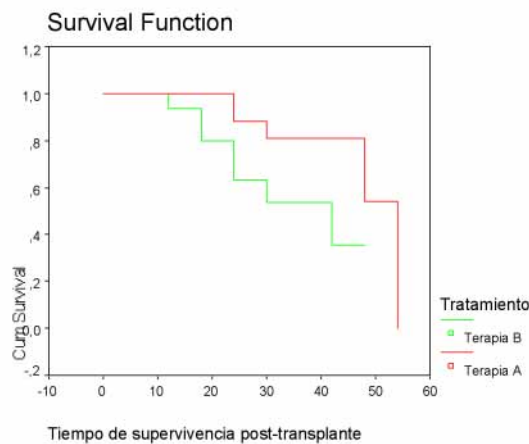


Figura 72. Función de supervivencia por el método actuarial





En términos generales, comparando funciones dentro de un mismo gráfico, puede afirmarse que con la terapia A disminuye más rápidamente la probabilidad de sobrevivir que con la obtenida con la terapia B.

Cabe decir que en ausencia de datos censurados, la prueba no paramétrica de Wilcoxon sería adecuada para comparar los tiempos de supervivencia entre grupos de estudio.

16.3.2. Kaplan-Meier

El método de Kaplan-Meier se conoce por ser el método en el que la proporción acumulada que sobrevive o función de supervivencia se calcula para el tiempo de supervivencia individual de cada paciente y no para la agrupación de los tiempos en intervalos. Por esta razón es especialmente útil para estudios que utilizan un número pequeño de pacientes.

La validez de este método descansa en dos suposiciones: Las personas que se retiran del estudio tienen un destino parecido a las que quedan y el período de tiempo durante el cual una persona entra en el estudio no tiene efecto independiente en la respuesta.

Para ilustrar el proceso de estimación de la función de supervivencia por el método de Kaplan-Meier consideraremos el mismo ejemplo de pacientes con IRT transplantados con distintas terapias paliativas. La tabla de cálculo para estimar la función de supervivencia por el método de Kaplan-Meier se solicita tal como aparece en la siguiente figura:

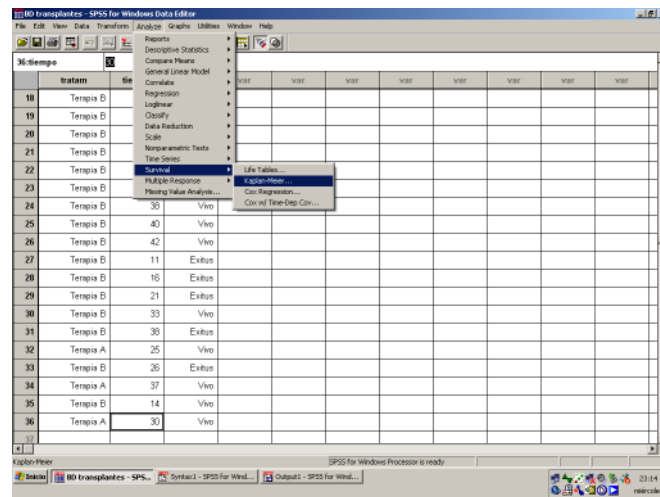


Figura 73. Análisis de supervivencia por el método Kaplan-Meier en SPSS





16.3.2.1. Descripción de la información obtenida por el método Kaplan-Meier

La tabla de cálculo contiene un listado con una línea por cada paciente observado (Figura 74). Las líneas se disponen ordenadas, de menor a mayor valor, en función del tiempo de observación del paciente correspondiente. Los distintos tiempos observados en el grupo objeto de análisis (en la columna «Time») son: 11, 15, 18, ... y se proporciona la siguiente información:

- o Estado del paciente. Si ha habido un “*éxitus*”, en la columna «Status» aparecerá el valor (1); en caso contrario, aparecerá 0.
- o Valor estimado de la función de supervivencia («Cumulative Survival»).
- o Error estándar de la estimación del valor correspondiente de la función de supervivencia («Standard Error»).
- o Número de pacientes que han fallecido en un período de tiempo inferior o igual a t_k («Cumulative events»).
- o Número de pacientes para los que el tiempo de observación ha sido superior a t_k («Number Remaining»).

El valor de la función de supervivencia en el instante t_k , probabilidad de sobrevivir hasta, al menos, el instante t_k («Cumulative Survival»), se estima como la proporción o porcentaje de supervivencia acumulada hasta t_k .



Survival Analysis for TIEMPO Tiempo de supervivencia post-transplante
Factor TRATAM = 1,00

Time	Status	Cumulative Survival	Standard Error	Cumulative Events	Number Remaining
11	,00			0	18
15	,00			0	17
18	1,00	,9412	,0571	1	16
20	1,00	,8824	,0781	2	15
20	,00			2	14
24	,00			2	13
25	,00			2	12
26	1,00	,8088	,1004	3	11
30	,00			3	10
30	,00			3	9
33	,00			3	8
37	,00			3	7
43	1,00	,6933	,1373	4	6
45	1,00	,5777	,1556	5	5
45	,00			5	4
47	,00			5	3
48	,00			5	2
53	,00			5	1
55	1,00	,0000	,0000	6	0

Number of Cases: 19 Censored: 13 (68,42%) Events: 6

	Survival Time	Standard Error	95% Confidence Interval
Mean:	46	4	(39; 53)
Median:	55	0	(;)





Survival Analysis for TIEMPO Tiempo de supervivencia post-transplante

Factor TRATAM = 2,00

Time	Status	Cumulative Survival	Standard Error	Cumulative Events	Number Remaining
9	,00			0	16
11	1,00	,9375	,0605	1	15
14	,00			1	14
14	,00			1	13
15	1,00	,8654	,0890	2	12
15	,00			2	11
16	1,00	,7867	,1103	3	10
18	,00			3	9
19	1,00	,6993	,1281	4	8
21	1,00	,6119	,1387	5	7
25	,00			5	6
26	1,00	,5099	,1484	6	5
33	,00			6	4
38	1,00	,3824	,1568	7	3
38	,00			7	2
40	,00			7	1
42	,00			7	0

Number of Cases: 17 Censored: 10 (58,82%) Events: 7

	Survival Time	Standard Error	95% Confidence Interval
Mean:	30	3	(24; 37)
(Limited to	42)		
Median:	38	12	(15; 61)

Figura 74. Análisis de supervivencia por el método Kaplan-Meier en SPSS

Por ejemplo, la probabilidad estimada de que los pacientes con terapia B(2) sobrevivan hasta pasado un plazo de tiempo superior o igual a $t_1 = 11$, a $t_3 = 16$ o a $t_6 = 26$ meses desde el transplante renal es igual a: 0,94, 0,79 y 0,51, respectivamente. La estimación de la mediana del tiempo de reaparición de la sintomatología («Median Survival Time»), tiempo en el que se espera que hayan fallecido





por lo menos la mitad de los pacientes es igual a 38 meses, frente a los 55 meses de la terapia A. Los límites inferior y superior del intervalo de confianza al 95 por 100 para la mediana vienen dados por los tiempos 15 y 61 meses, respectivamente. Por otro lado, la estimación del tiempo medio de supervivencia («Mean Survival Time») es igual a 30 meses, y su intervalo de confianza al 95 por 100 es de 24 y 37 meses.

16.3.2.2. Comparación de la supervivencia en distintos grupos de individuos por el método de Kaplan-Meier

Para evaluar por este método si la terapia recibida influye en el tiempo de supervivencia de los pacientes transplantados, observamos el resultado obtenido en el análisis Kaplan-Meier, tal como se observa en la siguiente figura. En primer lugar, para la terapia A (Tratam =1) el 68,42% de los casos estaban vivos al finalizarse el estudio. Mientras que este porcentaje desciende al 58,82% en el caso de la terapia B.

Survival Analysis for TIEMPO Tiempo de supervivencia post-transplante					
		Total	Number Events	Number Censored	Percent Censored
TRATAM	1,00	19	6	13	68,42
TRATAM	2,00	17	7	10	58,82
Overall		36	13	23	63,89
Test Statistics for Equality of Survival Distributions for TRATAM					
	Statistic	df	Significance		
Log Rank	4,66	1	,0308		
Breslow	4,33	1	,0375		
Tarone-Ware	4,51	1	,0338		

Figura 75. Análisis de supervivencia por el método Kaplan-Meier en SPSS

Si el análisis se realiza considerando más de un grupo de individuos, para contrastar la hipótesis nula de que la distribución de la supervivencia en los diferentes grupos es la misma, podemos utilizar tres pruebas: Mantel-Cox, Breslow y Tarone-Ware. Las diferencias entre ellas radican en las ponderaciones asignadas en la construcción del estadístico correspondiente. Si el p-valor aso-





ciado al estadístico de contraste es menor que α se rechazará la hipótesis nula al nivel de significación $\alpha = 0,05$.

Los resultados de las pruebas de Mantel-Cox, de Breslow y de Tarone-Ware para comparar las dos funciones de supervivencia correspondientes a las distintas terapias post-transplante, confirman lo que se sospechaba anteriormente: en los tres casos el p-valor asociado al estadístico de contraste es menor que 0,05, tanto en Log Rank («Significance = 0,0308»), Breslow («Significance = 0,0375») como Tarone-Ware («Significance = 0,0338»), luego, al nivel de significación 0,05, se rechaza la hipótesis nula de igualdad de las dos funciones.

Por otro lado, al igual que en el método Actuarial, podría haberse realizado una representación gráfica de las funciones de supervivencia, a partir de la que podría interpretarse en qué consisten las diferencias entre ellas (Figura 72).

16.4. Regresión de Cox. Modelo de Riesgos Proporcionales.

La Regresión de Cox es la generalización del análisis de Kaplan-Meier. En efecto, mientras que éste último permite realizar comparaciones entre grupos de población diferentes en un factor simple, la regresión de Cox ofrece la posibilidad de introducir en el modelo un conjunto de covariables, categóricas o continuas que, para cada combinación de sus distintos valores o niveles, definen múltiples subgrupos poblacionales. De esta forma, la regresión de Cox consiste en obtener una función lineal de las variables independientes que permita estimar, en función del tiempo, la probabilidad de que ocurra dicho suceso.

Continuemos con el ejemplo anterior: además de disponer de los registros de las historias clínicas de los pacientes con IRT transplantados y la terapia recibida, se dispone de información sobre el hábito tabáquico que parece estar relacionado con el éxito del transplante, por lo que se categorizará a pacientes que hayan fumado en los 3 años previos al transplante y los que no lo han hecho.

Para estimar la probabilidad de supervivencia en función del tipo de terapia recibida así como también del hábito de fumar, se aplicará el modelo de regresión de Cox.

16.4.1. Formulación del problema

En la regresión de Cox, a diferencia de los dos métodos anteriores, se supone que existe un conjunto de variables independientes, X_1, \dots, X_p , cuyos valores influyen en el tiempo que transcurre hasta que ocurre el suceso final.

El análisis consistirá entonces en estimar los parámetros desconocidos, como si de una regresión lineal o logística se tratara. Por ello, una vez estimados los





parámetros del modelo, además de la estimación de la función de riesgo se obtendrá la estimación del valor de la función de supervivencia para cada instante t . En nuestro ejemplo, tenemos las siguientes variables:

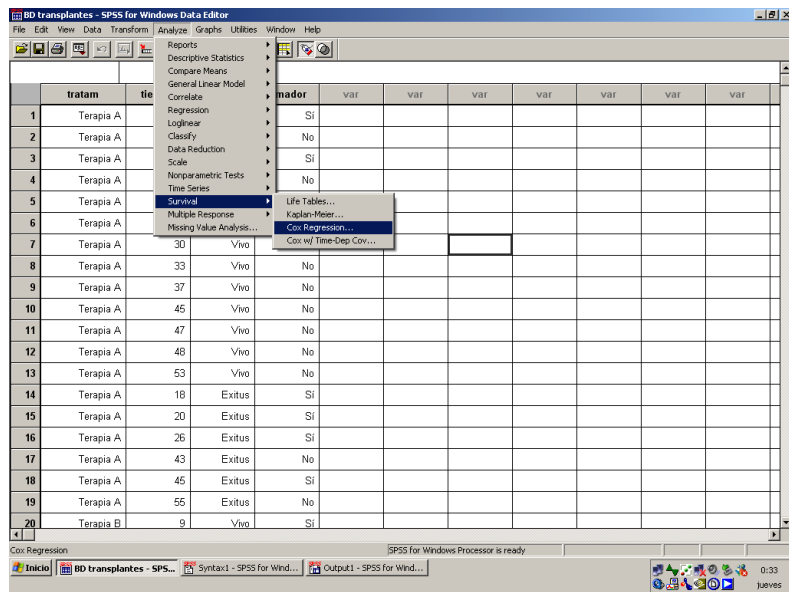
TRATAM 1=Terapia A, 2=Terapia B

TABACO 0=El paciente no ha fumado durante los 3 años previos al trasplante y 1=El paciente ha fumado durante los 3 años previos al trasplante.

EXITUS 0=Paciente Vivo, 1=Éxito.

TIEMPO Tiempo transcurrido desde el trasplante renal y el último dato conocido antes del cierre del estudio (meses).

Se trata de obtener una combinación lineal de las variables independientes **TRATAM** y **TABACO** a partir de la que se pueda estimar, para cada t , la probabilidad de que sobrevivir hasta pasado un período de tiempo superior o igual a t , siendo t los posibles valores que puede tomar la variable **TIEMPO**. La siguiente figura, muestra como realizar los pasos para obtener el análisis del tiempo de supervivencia con regresión de Cox.



16.4.2. Variables cualitativas en la regresión de Cox

En este tipo de análisis, al igual que en la regresión lineal y logística, las variables cualitativas deben ser tratadas de forma ligeramente distintas a las cuantitativas, debido a que sus valores sólo representan un código y no un valor por sí mismo.

En nuestro ejemplo, al introducir las covariables en el modelo, indicamos que tanto TRATAM como FUMADOR son categóricas. El programa creará las variables dummies necesarias para poder realizar el análisis, tomando por defecto la última categoría de referencia, es decir, la función de supervivencia será modelizada con lo que aporten las categorías no referenciadas. Sin embargo, la categoría de referencia puede cambiarse, de manera que podemos elegir la primera categoría como categoría de referencia: la variable TRATAM será modelizada con 2=Terapia B y FUMADOR con 1=Si.

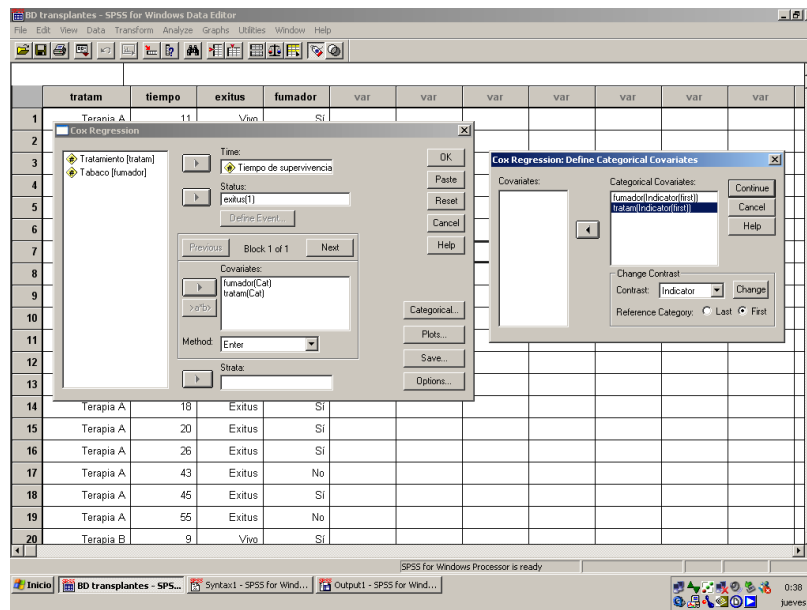


Figura 77. Cuadro de diálogo para obtener una regresión de Cox en SPSS

En el desarrollo del capítulo dedicado a la regresión logística, la situación planteada en el ejemplo para ilustrar la técnica era prácticamente la misma que la actual. Tanto el método Forward como los criterios basados en la Puntuación eficiente de Rao y en el estadístico de Wald para la selección y eliminación de variables, utilizados en la construcción de la ecuación de regresión logística, están disponibles en el caso de la regresión de Cox. Por ello, en la construcción



de la función Z para el modelo de regresión de Cox, tanto el método como los criterios para la selección y eliminación de variables serán los anteriormente citados en el capítulo 14 de regresión logística. Aunque en este ejemplo utilizaremos el método enter puesto que sólo existen dos covariables y no requiere de más complejidad, se describen los estadísticos más utilizados en la determinación de parámetros.

16.4.2.1. Estadístico de Wald

El estadístico de Wald, para las variables incluidas en la ecuación de regresión de Cox, juega exactamente el mismo papel que en la regresión logística. Es decir, para cualquier variable independiente X_i seleccionada, si β es el parámetro asociado en la ecuación de regresión, el estadístico de Wald permite contrastar la hipótesis nula de que $\beta=0$:

La interpretación de dicha hipótesis es que la información que se perdería al eliminar la variable X_i no es significativa. Si el p-valor asociado al estadístico de Wald es menor que α se rechazará la hipótesis nula al nivel de significación α . Bajo este punto de vista, en cada etapa del proceso de selección de variables, la candidata a ser eliminada será la que presente el máximo p-valor asociado al estadístico de Wald. Será eliminada si dicho máximo es mayor que un determinado valor crítico prefijado (si no se indica lo contrario, 0.1).

16.4.2.2. Puntuación eficiente de Rao

Si el estadístico de Wald en la regresión de Cox juega el mismo papel que en la regresión logística para las variables incluidas en la ecuación, la Puntuación eficiente de Rao juega exactamente el mismo papel que en la regresión logística para las variables no incluidas en la ecuación de regresión de Cox. Supongamos que β , es el parámetro asociado a la variable X_i supuesto que entrará en la ecuación de regresión en el siguiente paso. El estadístico Puntuación eficiente de Rao permite contrastar la hipótesis nula de que $\beta=0$.

La interpretación de dicha hipótesis es que, si la variable X_i fuera seleccionada en el siguiente paso, la información que aportaría no sería significativa. Si el p-valor asociado al estadístico Puntuación eficiente de Rao es menor que α se rechazará la hipótesis nula al nivel de significación α . Bajo este punto de vista, en cada etapa del proceso de selección de variables, la candidata a ser seleccionada será la que presente el mínimo p-valor asociado al estadístico Puntuación eficiente de Rao. Será seleccionada si dicho mínimo es menor que un determinado valor crítico prefijado (si no se indica lo contrario, 0,05).



16.4.3. Estimación de los parámetros

De acuerdo a la Figura 78, observamos los coeficientes de las variables FUMADOR y TRATAM y vemos que ambas son significativas con un p-valor < 0,05. Para la variable FUMADOR, $\beta=2,46$, de manera que $e^{\beta}=11,71$ es el Riesgo Relativo asociado al hecho ser fumador en los 3 años previos al trasplante, por lo que es un factor que aumenta el riesgo de fallecer. Por otro lado, la terapia B también presenta un mayor riesgo de fallecer con $\beta=1,57$, de manera que $e^{\beta}=4,83$, aunque su intervalo de confianza está muy cercano a 1.

Dado el resultado del análisis del tiempo de supervivencia mediante la regresión de Cox, observamos que tanto el ser fumador, como el tipo de terapia, influyen en la supervivencia post-trasplante. Por lo tanto la función estimada será:

$$Z(X) = 2,46 \text{ FUMADOR} + 1,58 \text{ TRATAM}$$

$$G(X) = e^{2,46 \text{ FUMADOR} + 1,58 \text{ TRATAM}} = 11,70 \text{ FUMADOR} \times 4,85 \text{ TRATAM}$$

Es decir, si por ejemplo, un paciente que ha sido tratado con la terapia B (terapia=2, codificado como 1) y además es fumador (fumador=1, codificado como 1), se obtendrá una $g(1) = 56,82$; pero si otro paciente también recibe la misma terapia pero no es fumador (fumador=0, codificado como 0), se obtendrá una $g(2)=4,85$. En otras palabras, la probabilidad estimada de fallecer en un pequeño intervalo de tiempo, se reducirá considerablemente si el paciente deja de fumar en los años previos al trasplante.

	Indicator Value	Parameter Freq	Coding (I)
TRATAM	Tratamiento		
	Terapia A	19	,000
	Terapia B	17	1,000
	Indicator Value	Parameter Freq	Coding (I)
FUMADOR	Tabaco		
	No	22	,000
	Sí	14	1,000
36 Total cases read			
0 Cases with missing values			
0 Valid cases with non-positive times			
1 Censored cases before the earliest event in a stratum			
1 Total cases dropped			
35 Cases available for the analysis			
Dependent Variable: TIEMPO Tiempo de supervivencia post-trasplante			





Events	Censored					
13	22 (62,9%)					
Beginning Block Number 0. Initial Log Likelihood Function						
-2 Log Likelihood	70,043					
Beginning Block Number 1. Method: Enter						
Variable(s)	Entered at Step Number 1..					
FUMADOR	Tabaco					
TRATAM	Tratamiento					
Coefficients converged after 4 iterations.						
-2 Log Likelihood	50,841					
	Chi-Square	df	Sig			
Overall (score)	21,477	2	,0000			
Change (-2LL) from						
Previous Block	19,203	2	,0001			
Previous Step	19,203	2	,0001			
----- Variables in the Equation -----						
Variable	B	S.E.	Wald	df	Sig	R
FUMADOR	2,4608	,7135	11,8961	1	,0006	,3759
TRATAM	1,5761	,7058	4,9875	1	,0255	,2065
95% CI for Exp(B)						
Variable	Exp(B)	Lower	Upper			
FUMADOR	11,7144	2,8934	47,4277			
TRATAM	4,8363	1,2128	19,2862			
Covariate Means and Pattern Values						
Variable	Mean	Pattern 1	Pattern 2			
FUMADOR	,3714	mean	mean			
TRATAM	,4571	,0000	1,0000			

Figura 78. Resultados obtenidos en la regresión de Cox

16.4.4. Bondad del ajuste

Comprobar la bondad del ajuste es analizar cuan probables son los resultados muestrales a partir del modelo ajustado. La probabilidad de los resultados





obtenidos se denomina verosimilitud. Cuanto más próximo a cero sea el valor del estadístico $-2LL$, más próxima a 1 será la verosimilitud y mejor será el modelo. Para comprobar si la verosimilitud difiere de 1 (que el modelo se ajusta perfectamente a los datos) se utiliza el estadístico:

$$-2LL = -2 \times \text{Logaritmo de la verosimilitud}$$

16.5. Consideraciones importantes

Día a día, un mayor número de investigadores realiza análisis de supervivencia con multitud de datos: desde valores clínicos tales como el fallecimiento, la recaída o aparición de un determinado acontecimiento adverso, a valores tales como tiempo entre una consulta y la siguiente, etc.

Las dos metodologías presentadas son útiles en el análisis de supervivencia, presentando de forma clara los resultados en cuanto al tiempo de supervivencia y en cuanto a la comparación de los valores respecto a distintos factores o estratos. Mientras que el método actuarial nos resuelve la supervivencia para intervalos de tiempo determinados, el método de Kaplan-Meier estima la supervivencia en función de una función obtenida a partir de todos los casos del estudio.

Paralelamente, la regresión de Cox nos ofrece una visión mucho más extensa del análisis de supervivencia, comparándolo con una regresión logística en cuanto a su metodología e interpretación de resultados. Por lo que será un análisis muy útil si queremos cuantificar riesgos de posibles factores asociados a la supervivencia o, en cualquier caso, al tiempo en que da lugar un determinado evento.

