

Resumen del análisis de supervivencia

A continuación, se listan los análisis de supervivencia realizados sobre el conjunto de datos de “Cáncer de Próstata”

En este estudio el evento de interés es la *muerte de un paciente que fue diagnosticado con cáncer*. Por lo tanto, tenemos las probabilidades:

- 1- Probabilidad de supervivencia (S), representa la probabilidad que tiene un paciente a permanecer en la población.
- 2- Probabilidad de riesgo (Hazard function), representa la probabilidad que tiene un paciente a tener el evento estudiado en un tiempo determinado.

El objetivo es analizar la relación que puede existir entre el evento de interés (muerte) y los dos grupos en los cuales está dividida la población. La población está dividida en dos grupos, los cuales fueron formados previamente por los expertos.

Características del conjunto de datos

- 1- Características de la variable Meses. Esta es la variable tiempo en el análisis.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.00	24.25	36.00	36.26	48.00	60.00

El menor tiempo registrado fue de 1 mes a partir del tiempo inicial, y el mayor tiempo fue de 60 meses. La media de los tiempos de registros es de aproximadamente 36 meses.

- 2- Características de la variable Grupo

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0000	0.0000	1.0000	0.5714	1.0000	1.0000

Aproximadamente, el 57% de los pacientes con cáncer pertenecen al grupo 1.

- 3- Características de la variable CCS (MUERTE CANCER ESPECIFICA).

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0000	0.0000	0.0000	0.1429	0.0000	1.0000

Aproximadamente solo el 14% de los pacientes murieron de cáncer de próstata.

- 4- Características de la variable OS (MUERTE). Variable que refleja la muerte de un paciente, no necesariamente murió de cáncer de próstata.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0000	0.0000	0.0000	0.3095	1.0000	1.0000

Aproximadamente el 30% de los pacientes murieron.

1ra parte del análisis

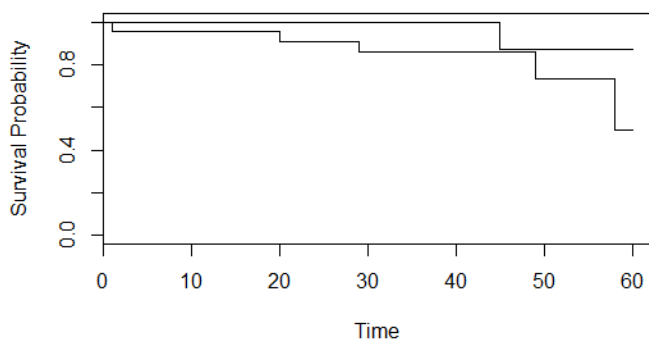
En esta primera consideramos como variables dependientes a las variables Meses y CCS, y CCS es nuestra variable evento. La variable independiente es la variable Grupo.

Análisis no-paramétrico

1- Kaplan-Meier non-parametric analysis

group=0						
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
45.000	8.000	1.000	0.875	0.117	0.673	1.000

group=1						
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
1	24	1	0.958	0.0408	0.882	1
20	20	1	0.910	0.0607	0.799	1
29	18	1	0.860	0.0755	0.724	1
49	7	1	0.737	0.1308	0.520	1
58	3	1	0.491	0.2187	0.205	1



Como podemos apreciar, solo un paciente del grupo 0 murió de cáncer de próstata, mientras que 5 pacientes del grupo 1 presentaron el evento. La gráfica muestra las curvas de supervivencia de los dos grupos de pacientes. A simple vista no existen grandes diferencias entre las dos curvas hasta pasado los 50 meses. En la curva correspondiente al grupo 1 se registra un descenso considerable en la probabilidad de supervivencia a partir del mes 50 (de un 73% a un 49% de probabilidad).

Análisis Semi-paramétrico

En esta parte del estudio se emplearon los modelos proporcionales de Cox para tiempos discretos, puesto que realmente la variable tiempo en nuestro estudio es discreta (datos agrupados por cantidad de meses).

1- Cox proportional hazard model - Breslow method for ties handling

	coef	exp(coef)	se(coef)	z	Pr(> z)
group	1.253	3.501	1.101 1.138		0.255

	exp(coef)	exp(-coef)	lower .95	upper .95
group	3.501	0.2857	0.4046	30.29

Concordance= 0.659 (se = 0.122)
 Rsquare= 0.038 (max possible= 0.563)
 Likelihood ratio test= 1.65 on 1 df, p=0.1991
 Wald test = 1.3 on 1 df, p=0.2551
 Score (logrank) test = 1.47 on 1 df, p=0.2257

En este caso, el modelo de Cox nos indica que la variable independiente Grupo tiene un coeficiente positivo, en otras palabras, esta variable está correlacionada positivamente con la variable evento CCS. Este coeficiente indica que aquellos pacientes que pertenecen al grupo 1 tienen más probabilidad de sufrir el evento estudiado, es decir tiene más probabilidad de morir por cáncer de próstata. El valor de 3.501 asociado a la variable Grupo nos indica que por cada tres individuos pertenecientes al grupo 1 se tiene un 50% mayor de ratio Hazard, en otras palabras, un 50% mayor de probabilidad de morir de cáncer de próstata.

Por otra parte, ninguno de los tres tests de comparaciones para curvas de supervivencia (Likelihood ratio test, Wald test, Score (logrank) test) detectan diferencias significativas entre las curvas a un nivel de significación de 0.10 (90% de confianza). Solo el Likelihood ratio test puede rechazar la hipótesis nula si se consideraría un nivel de significación de 0.20 (80% de confianza). Estos test tienden a encontrar diferencias cuando las curvas son consistentes en el tiempo. En mi opinión, aunque estos tests no llegan a rechazar la hipótesis nula, lo cual pudiera ser debido al tamaño de la muestra, creo que es interesante la tendencia que detecta el modelo de Cox entre la variable independiente Grupo y la variable dependiente CCS.

2- Cox proportional hazard model – Efron method for ties handling

No existen diferencias en los resultados respecto a cuando se ejecuta el modelo de Cox con el método Breslow.

3- Cox proportional hazard model -Exact method for ties handling

No existen diferencias en los resultados respecto a cuando se ejecuta el modelo de Cox con el método Breslow.

Análisis paramétrico

1- Exponential parametric model coefficients

	value	Std. Error	z	p
(Intercept)	6.50	1.0	6.50	8.21e-11
group	-1.35	1.1	-1.23	2.18e-01

Scale fixed at 1

Exponential distribution
 Loglik(model)= -38.2 Loglik(intercept only)= -39.2
 Chisq= 1.97 on 1 degrees of freedom, p= 0.16
 Number of Newton-Raphson Iterations: 6

n= 42

Con un análisis paramétrico era de esperar que se detectaran mayores diferencias significativas, ya que los análisis paramétricos son más robustos que los no paramétricos. Se vuelve a confirmar que la variable independiente Grupo está correlacionada positivamente con la variable dependiente evento. Note que, en este caso el coeficiente da con signo negativo, pero esto es 100% correcto, ya que deben dar con un signo contrario a los coeficientes de los modelos de Cox.

En este caso, si se puede rechazar la hipótesis nula en un nivel de significación de 0.2 (80% de nivel de confianza).

2- log-logistic parametric model coefficients

	value	Std. Error	z	p
(Intercept)	5.675	1.027	5.528	3.25e-08
group	-1.008	0.889	-1.133	2.57e-01
Log(scale)	-0.349	0.370	-0.945	3.45e-01

Scale= 0.705

Log logistic distribution
Loglik(model)= -38.2 Loglik(intercept only)= -39.1
 Chisq= 1.92 on 1 degrees of freedom, p= 0.17
Number of Newton-Raphson Iterations: 5
n= 42

Este modelo vuelve a ratificar las conclusiones que anteriormente se han expuesto.

2da parte del análisis

En esta 2da parte consideramos como variables dependientes a las variables Meses y OS, y OS es nuestra variable evento. La variable independiente es la variable Grupo.

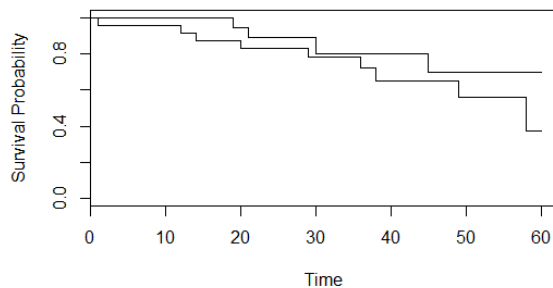
Análisis no-paramétrico

1- Kaplan-Meier non-parametric analysis

grupo=0							
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI	
19	18	1	0.944	0.0540	0.844		1
21	17	1	0.889	0.0741	0.755		1
30	10	1	0.800	0.1075	0.615		1
45	8	1	0.700	0.1327	0.483		1

grupo=1							
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI	
1	24	1	0.958	0.0408	0.882		1.000
12	23	1	0.917	0.0564	0.813		1.000
14	22	1	0.875	0.0675	0.752		1.000
20	20	1	0.831	0.0770	0.693		0.997

29	18	1	0.785	0.0855	0.634	0.972
36	13	1	0.725	0.0979	0.556	0.944
38	10	1	0.652	0.1118	0.466	0.913
49	7	1	0.559	0.1289	0.356	0.878
58	3	1	0.373	0.1747	0.149	0.934



La gráfica muestra las curvas de supervivencia de los dos grupos de pacientes. A simple vista no existen grandes diferencias entre las dos curvas. Sin embargo, vale la pena destacar que los pacientes del grupo 1 registran eventos a todo lo largo del tiempo de estudio.

Análisis Semi-paramétrico

1- Cox proportional hazard model - Breslow method for ties handling

n= 42, number of events= 13

	coef	exp(coef)	se(coef)	z	Pr(> z)
group	0.5171	1.6771	0.6040	0.856	0.392

	exp(coef)	exp(-coef)	lower .95	upper .95
group	1.677	0.5963	0.5134	5.479

Concordance= 0.557 (se = 0.078)
 Rsquare= 0.018 (max possible= 0.86)
 Likelihood ratio test= 0.77 on 1 df, p=0.3787
 Wald test = 0.73 on 1 df, p=0.392
 Score (logrank) test = 0.75 on 1 df, p=0.3868

En este caso, el modelo de Cox nos indica que la variable independiente Grupo tiene una correlación positiva con el evento estudiado, pero resulta relevante notar que la correlación es mucho menor que cuando se considera como evento la muerte por cáncer de próstata (ver la 1ra parte del estudio).

2- Cox proportional hazard model – Efron method for ties handling

No existen diferencias en los resultados respecto a cuando se ejecuta el modelo de Cox con el método Breslow.

3- Cox proportional hazard model -Exact method for ties handling

No existen diferencias en los resultados respecto a cuando se ejecuta el modelo de Cox con el método Breslow.

Análisis Paramétrico

1- Exponential parametric model coefficients

	Value	Std. Error	z	p
(Intercept)	5.110	0.500	10.221	1.60e-24
group	-0.551	0.601	-0.917	3.59e-01

Scale fixed at 1

Exponential distribution

Loglik(model)= -74.5 Loglik(intercept only)= -74.9
Chisq= 0.89 on 1 degrees of freedom, p= 0.34
Number of Newton-Raphson Iterations: 5
n= 42

Con un análisis paramétrico se confirman los resultados arrojados por el modelo de Cox. En este caso, no se puede rechazar la hipótesis nula, ni siquiera a un nivel de significación de 0.3 (70% de nivel de confianza).

2- log-logistic parametric model coefficients

	Value	Std. Error	z	p
(Intercept)	4.506	0.414	10.886	1.34e-27
group	-0.388	0.449	-0.865	3.87e-01
Log(scale)	-0.471	0.247	-1.910	5.61e-02

Scale= 0.624

Log logistic distribution

Loglik(model)= -73.8 Loglik(intercept only)= -74.2
Chisq= 0.81 on 1 degrees of freedom, p= 0.37
Number of Newton-Raphson Iterations: 4
n= 42

Este modelo vuelve a ratificar las conclusiones que anteriormente se han expuesto.

Conclusiones

- Los test de comparaciones de curvas de supervivencia no detectan diferencias significativas en niveles de significación bajos, por ejemplo 0.10 o 0.05. Sin embargo, para un nivel de confianza del 80% si se detectan en algunos casos.
- Tanto los modelos semi-paramétricos, como los paramétricos, reflejan que existe una correlación positiva considerable entre la variable grupo y la variable evento CCS, indicándonos que aquellos pacientes que pertenecen al grupo 1 tienen más probabilidades de morir de cáncer de próstata.

- Al considerar la variable OS como variable evento, la correlación entre la variable Grupo y la variable evento disminuye considerablemente, pero sigue siendo positiva. La disminución se debe al incremento de eventos registrados en el grupo 0; los eventos registrados son muertes por otras situaciones en esta parte del estudio. Esto es favorable para el análisis, puesto que la variable CCS es un subconjunto de la variable OS (todos los pacientes registrados como CCS son OS, pero no ocurre la situación inversa). Este resultado nos indica que realmente la correlación entre la variable Grupo y la variable CCS es muy fuerte, a pesar de que los test no llegan a detectar diferencias significativas para niveles de significación bajos.