

LICENCIATURA EN ESTADÍSTICA

“Análisis de Duración sobre el cáncer de mamas”

Trabajo Práctico: Segunda presentación



Autores: Tomás Anderson - Alejo Vaschetti

Docentes: Nora Arnesi, Gabriela Boggio, Guillermina Harvey y Victorio Costa

07/06/2024

Introducción

De acuerdo a datos de la Organización Mundial de la Salud (OMS), cada año se producen 1,38 millones de nuevos casos y 458.000 muertes por cáncer de mama. Por este motivo se dio el interés de hacer un análisis de supervivencia de las víctimas de esta patología.

Para ello se cuenta con un conjunto de datos de 444 mujeres que tuvieron cáncer de mamas y se observa su supervivencia en días desde la **operación primaria** del tumor hasta la **muerte** del paciente. Se toman aquellos pacientes que fueron operados en 1992 y 1993 en la ciudad de Rotterdam.

Si bien el 72% de los datos se encuentran censurados por derecha, el análisis de supervivencia nos permite continuar con el estudio.

Datos y objetivos

Dentro de la base de datos se cuentan con las siguientes variables observadas de cada paciente:

1. Tiempo: Cantidad de días desde la operación hasta la muerte del paciente o pérdida de seguimiento.
2. Muerte: Si el paciente murió o fue una censura.
3. Quimioterapia: Si el paciente hizo o no un tratamiento de quimioterapia.
4. Hormonas: Si el paciente usaba un tratamiento hormonal o no.
5. Menopausia: Si el paciente tenía menopausia o no.
6. Edad: Edad del paciente al hacerse la operación primaria.
7. Tamaño del tumor: Si el tamaño del tumor es menor o igual a 20mm, entre 20mm y 50mm o mayor a 50mm.
8. Grado: Grado del tumor (grado 2 o 3).
9. Receptores de progesterona: Cantidad de receptores de progesterona (en fmol/l).
10. Receptores de estrógeno: Cantidad de receptores de estrógeno (en fmol/l).

El objetivo principal de este informe es analizar la influencia de las variables provistas por el conjunto de datos en el tiempo de supervivencia de las mujeres con cáncer de mama. Observar bajo qué condiciones la mortalidad del cáncer de mama es menor. Realizar estimaciones puntuales y por intervalos de la supervivencia de los pacientes.

Selección de variables

Se tienen muchas variables que pueden llegar a considerarse para el análisis, pero se busca aquellas que puedan significar un cambio en la supervivencia de los pacientes al construir un modelo. Para ello se observa qué variable en forma marginal produce hay un cambio significativo de la supervivencia de los pacientes en un modelo de razón de hazards proporcionales de Cox.

Variables	gl	G^2	p-value
Edad	1	0.575	0.448
Menopausia	1	0.459	0.498
Tamaño del tumor	2	34.331	< 0.001
Grado del tumor	1	9.464	0.002
Receptores de progesterona	1	22.915	< 0.001
Receptores de estrógeno	1	7.792	0.005
Hormonas	1	7.246	0.007
Quimioterapia	1	3.092	0.079

Tabla 1: Modelos marginales

Al utilizar un $\alpha = 0.10$, se observa en la tabla 1 que las variables que podrían aportar a la construcción de un modelo de Cox son:

- El tamaño del tumor.
- El grado del tumor.
- Los receptores de progesterona.
- Los receptores de estrógeno.
- Si recibió tratamiento hormonal o no.
- Si recibió quimioterapia o no.

Por lo que se plantea un modelo con dichas variables.

$$h_i(t) = \exp(\beta_3 T_{1i} + \beta_4 T_{2i} + \beta_5 G_i + \beta_6 P_i + \beta_7 Est_i + \beta_8 H_i + \beta_9 Q_i) h_0(t)$$

Donde:

- T_{1i} : Dummy asociada al tamaño del tumor del i-ésimo paciente donde es 1 si está entre 20 y 50mm, y 0 en otro caso.
- T_{2i} : Dummy asociada al tamaño del tumor del i-ésimo paciente donde es 1 si es mayor a 50mm, y 0 en otro caso.

- G_i : Dummy asociada al grado de diferenciación del tumor del i-ésimo paciente donde es 1 si el grado es 3, y 0 si es 2.
- P_i : Cantidad de receptores de progesterona (en fmol/l) del i-ésimo paciente.
- Est_i : Cantidad de receptores de estrógeno (en fmol/l) del i-ésimo paciente.
- H_i : Dummy asociada a si el i-ésimo paciente recibió un tratamiento hormonal donde es 1 si lo recibió, y 0 si no.
- Q_i : Dummy asociada a si el i-ésimo paciente recibió quimioterapia donde es 1 si lo recibió, y 0 si no.

Para ver si las variables contribuyen significativamente al estar las demás presentes, se procede a sacar en orden las variables con menor significación marginal en los tests marginales y se compara con el modelo que sí la incluye. En el caso que el sacar la variable no represente un cambio significativo del modelo, se optará por dejarla afuera del modelo y seguir con los pasos. Una vez que se llegue a las variables que sí aportan al modelo, se probará las interacciones entre ellas.

Variables	Hipótesis	gl	G_{diff}^2	p-value
$\beta_3 T_{1i} + \beta_4 T_{2i} + \beta_5 G_i + \beta_6 P_i + \beta_7 Est_i + \beta_8 H_i + \beta_9 Q_i$	$H_0) \beta_9 = 0 \quad H_1) \beta_9 \neq 0$	1	2.768	0.096
$\beta_3 T_{1i} + \beta_4 T_{2i} + \beta_5 G_i + \beta_6 P_i + \beta_7 Est_i + \beta_8 H_i$	$H_0) \beta_8 = 0 \quad H_1) \beta_8 \neq 0$	1	2.655	0.103
$\beta_3 T_{1i} + \beta_4 T_{2i} + \beta_5 G_i + \beta_6 P_i + \beta_7 Est_i$	$H_0) \beta_7 = 0 \quad H_1) \beta_7 \neq 0$	1	1.401	0.236
$\beta_3 T_{1i} + \beta_4 T_{2i} + \beta_5 G_i + \beta_6 P_i$	$H_0) \beta_6 = 0 \quad H_1) \beta_6 \neq 0$	1	19.168	< 0.001
$\beta_3 T_{1i} + \beta_4 T_{2i} + \beta_5 G_i + \beta_6 P_i$	$H_0) \beta_3 = \beta_4 = 0$ $H_1) Al menos uno \neq de 0$	2	29.511	< 0.001
$\beta_3 T_{1i} + \beta_4 T_{2i} + \beta_5 G_i + \beta_6 P_i$	$H_0) \beta_5 = 0 \quad H_1) \beta_5 \neq 0$	1	3.564	0.059
$\beta_3 T_{1i} + \beta_4 T_{2i} + \beta_5 G_i + \beta_6 P_i + \beta_{36} T_{1i} P_i + \beta_{46} T_{2i} P_i$	$H_0) \beta_{36} = \beta_{46} = 0$ $H_1) Al menos uno \neq de 0$	2	1.894	0.388
$\beta_3 T_{1i} + \beta_4 T_{2i} + \beta_5 G_i + \beta_6 P_i + \beta_{56} G_i P_i$	$H_0) \beta_{56} = 0 \quad H_1) \beta_{56} \neq 0$	1	0.003	0.956
$\beta_3 T_{1i} + \beta_4 T_{2i} + \beta_5 G_i + \beta_6 P_i + \beta_{36} T_{1i} G_i + \beta_{46} T_{2i} G_i$	$H_0) \beta_{35} = \beta_{45} = 0$ $H_1) Al menos uno \neq de 0$	1	0.085	0.958

Tabla 2: Selección de variables hacia atrás

Por lo que se elige el siguiente modelo con las siguientes variables:

- Tamaño del tumor.
- Grado del tumor.
- Receptores de progesterona.

Para ver si se introduce al efecto de los receptores de progesterona como una variable cuantitativa, se realiza un test de linealidad.

Para crear grupos de la variable de receptores de progesterona que cumplan con una escala ordinal, se ordena de menor a mayor dichos números y se separa en 4 grupos aproximadamente con el mismo tamaño. Estos son:

Progesterona	Frecuencia absoluta
0	122
1-23	102
24-151	110
151	110

Tabla 3: Progesterona categorizada

Se crea el modelo con las demás variables que resultaron significativas y con esta nueva variable introducida con scores equiespaciados del 1 al 4, para luego comparar dicho modelo con uno donde se haya introducido la nueva variable como dummies.

Al hacer un test de razón de verosimilitud entre ambos modelos se espera probar lo siguiente:

H_0) El efecto de los receptores de progesterona es lineal

H_1) El efecto de los receptores de progesterona no es lineal

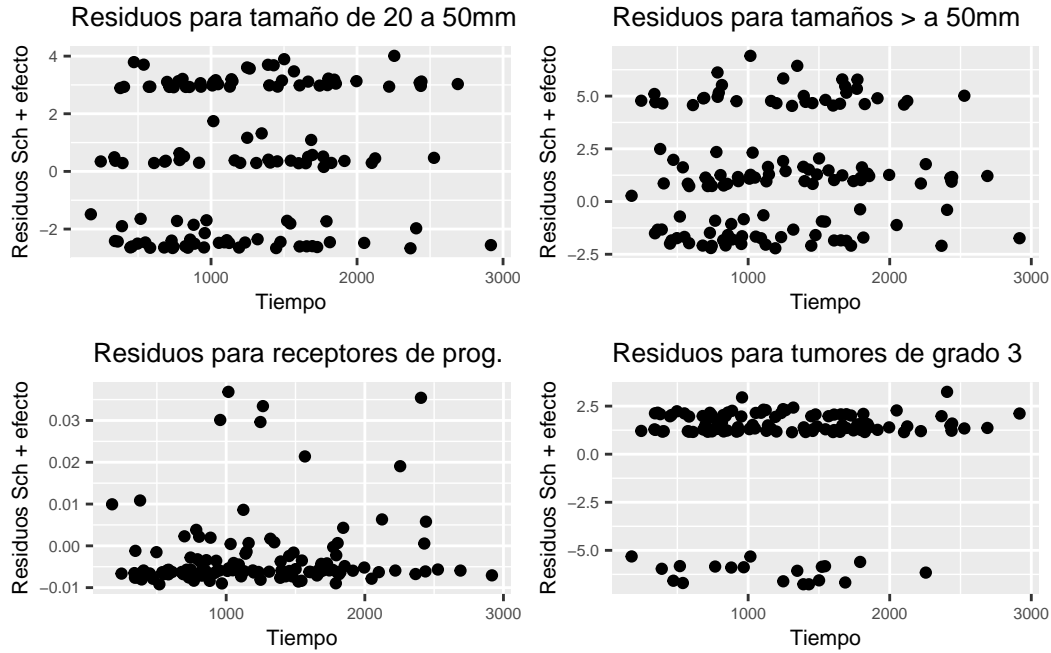
Se obtiene una estadística de 0.819 con un p-value asociado de 0.664, por lo que se decide conservar la variable como cuantitativa.

El modelo más parsimonioso y por lo tanto el que se decide usar para su interpretación resulta el siguiente:

$$h_i(t) = \exp(\beta_3 T_{1i} + \beta_4 T_{2i} + \beta_5 G_i + \beta_6 P_i) h_0(t)$$

Validación del modelo

Para validar el supuesto de que el efecto de las variables usadas en la construcción del modelo no dependen del tiempo se procede a hacer un análisis de residuos de Schoenfeld:



No se observa ninguna tendencia de los residuos a través del tiempo en cada variable por lo que pareciera estar cumpliéndose el supuesto planteado. Se procede a hacer test de hipótesis para probar el supuesto estadísticamente. Para ello se construyen nuevos modelos donde se añade a cada variable una por una como una variable que es función del tiempo y se trata de probar su significancia con un test de razón de verosimilitud.

H_0) El efecto de la j -ésima variable no cambia durante el tiempo

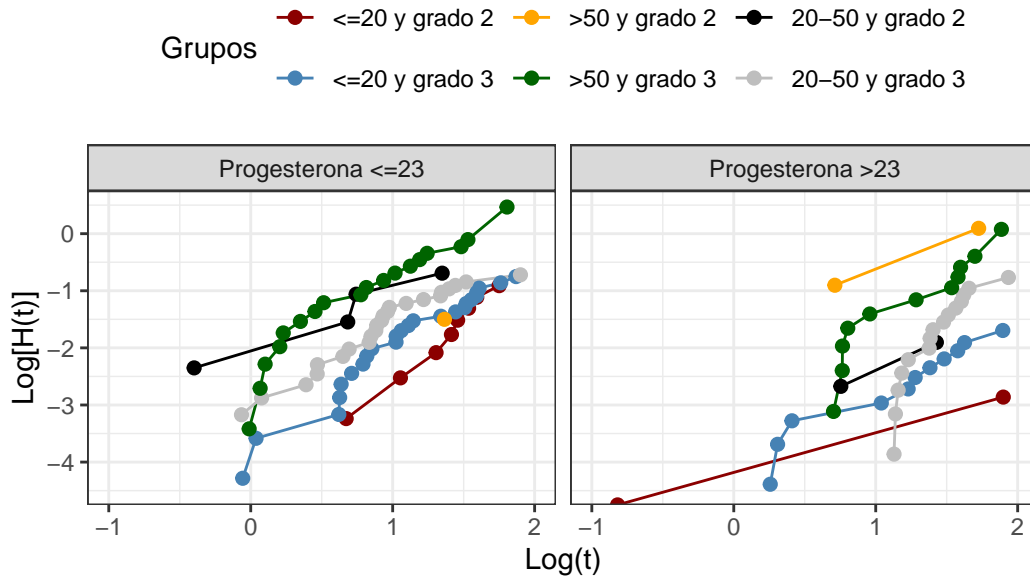
H_1) El efecto de la j -ésima variable cambia durante el tiempo

Variables	G^2	p-value
Tamaño	2.982	0.225
Receptores de progesterona	0.884	0.347
Grado 3 del tumor	1.055	0.304

Tabla 4: Efectos a través del tiempo

Al no rechazar ningún test, se considera que se cumple el supuesto de homogeneidad de efecto a medida que avanza el tiempo.

Logaritmo de la Función Hazard acumulada vs Logaritmo del tiempo



Parece que las distancias entre los grupos permanecen bastante constantes a través del logaritmo del tiempo, lo que indica que se está dando el cumplimiento de hazards proporcionales.

Resultados

Una vez comprobados los supuestos, se procede a interpretar las estimaciones del modelo final.

$$h_i(t) = \exp(0.497T_{1i} + 1.326T_{2i} + 0.453G_i + -0.003P_i)h_0(t)$$

donde $h_0(t)$ es la función hazard para un paciente con un tumor de tamaño menor a 20mm, de grado 2 y con 0 receptores de progesterona.

Parámetros	Estimación	2.75%	97.75%
$e^{\hat{\beta}_3}$	1.643	1.085	2.487
$e^{\hat{\beta}_4}$	3.768	2.388	5.944
$e^{\hat{\beta}_5}$	1.574	0.96	2.58
$e^{\hat{\beta}_6}$	0.997	0.996	0.999

Tabla 5: Estimaciones de las razones de hazard de la basal cambiando el valor de una variable

La tasa de muerte para los pacientes que tienen un tumor de grado 3 es 57.4% más grande que el de los que tienen uno de grado 2, manteniendo las demás variables constantes. El intervalo de confianza de la razón de hazard cubre el 1, pero puede interpretarse como que la tasa de muerte para los pacientes con un tumor de grado 3 es igual o mayor a las de los de grado 2.

Al aumentar en 10 unidades los receptores de progesterona, la tasa de muerte para los pacientes se ve disminuida en 2.6% al mantener las demás variables constantes.

La tasa de muerte para los pacientes que tienen un tumor de tamaño de 20 a 50 mm es un 64.3% mayor que el de los que tienen menor tamaño cuando las demás variables se mantienen constantes. En el caso de los pacientes con un tumor mayor a 50mm, su tasa es un 276.8% mayor que los que tienen 20mm o menos.

Conclusión

Al terminar el análisis, se puede afirmar lo siguiente:

La edad de los pacientes no influye en su supervivencia, es decir, que una persona joven tiene la misma supervivencia media que otra con una edad avanzada. Tampoco es afectada dicha supervivencia por si el paciente recibió quimioterapia, que es un resultado paupérrimo al ser un tratamiento con efectos secundarios problemáticos. Además, no se ve modificada por si son menopáusicas o si recibieron algún tratamiento hormonal. Los receptores de estrógeno no fueron significativos al ya estar analizando los de progesterona, ya que dichas variables están fuertemente relacionadas.

Las dos variables que están relacionadas con características del tumor resultaron ser importante al momento de medir la supervivencia del paciente, por lo que esto indica una relación con el tipo de tumor y la tasa de muerte. Entonces, es importante una detección temprana, ya que favorece a la supervivencia por tener un tumor con características menos malignas. Y por último, los receptores de progesterona afecta a la mortalidad, por lo que las hormonas forman parte de la explicación de la supervivencia.