

Licenciatura en Actuaría

Análisis de Supervivencia

Prueba de desempeño

Ejercicio1:

En la base de datos *pneumon* de la librería *KMsurv*, se tiene información acerca de la edad en meses en la que un bebe fue hospitalizado (*chldage*), así como si fue hospitalizado debido a neumonía o no (*hospital*). La variable *hospital* toma los valores 0 para no y 1 para si.

¿La variable **region** (1 noreste, 2 centro norte, 3 sur, 4 oeste) determina varias subpoblaciones de acuerdo al comportamiento de la edad al que fué hospitalizado el bebe? ¿Cómo están definidas?

Procedimiento estadístico con todos sus elementos incluidos (5 puntos)

Se realiza la prueba de hipótesis

$H_0: S_1(t)=S_2(t)=S_3(t)=S_4(t)$ para todo $t \geq 0$

H_1 : existe $i \neq j$ y $t_0 > 0$ tal que $S_i(t_0) \neq S_j(t_0)$.

survdiff(formula = pneumo.surv ~ region, data = datos, rho = 0)

N Observed Expected (O-E)^2/E (O-E)^2/V

region=1 517 13 10.9 0.406 0.479

region=2 868 25 18.1 2.633 3.513

region=3 1396 26 29.4 0.390 0.655

region=4 689 9 14.6 2.160 2.710

Chisq= 5.6 on 3 degrees of freedom, p= 0.1

En este caso el *p-valor* asociado a la prueba resultó ser de 0.1, el cual es mayor que 0.05, por lo que la muestra no presenta evidencia suficiente para rechazar la hipótesis nula.

Conclusión en términos del problema (2 puntos)

La muestra no presenta evidencia para decir que la región influye en el comportamiento del tiempo en meses al que fue hospitalizado el bebe, es decir, no se cuenta con información suficiente para decir que el comportamiento del tiempo de vida es diferente debido a la región de residencia.

Ejercicio 2

En la base de datos *pneumon* de la librería KMsurv, se tiene información acerca de la edad en meses en la que un bebe fue hospitalizado (*chldage*), así como si fue hospitalizado debido a neumonía o no (*hospital*). La variable *hospital* toma los valores 0 para no y 1 para si.

¿La variable *smoke* determina varias subpoblaciones de acuerdo al comportamiento de la edad al que fué hospitalizado el bebe? ¿Cómo están definidas?. La variable *smoke* se define como 0 si no fuma, 1 si fuma menos de una cajetilla por día, 2 si fuma 1 o más cajetillas por día)

Procedimiento estadístico con todos sus elementos incluidos (5 puntos)

Se realiza la prueba de hipótesis

$H_0: S_0(t)=S_1(t)=S_2(t)$ para todo $t \geq 0$

H_1 : existe $i \neq j$ y $t_0 > 0$ tal que $S_i(t_0) \neq S_j(t_0)$.

`survdifff(formula = neumo.surv ~ smoke, data = datos, rho = 0)`

N Observed Expected (O-E)²/E (O-E)²/V

smoke=0 2285 34 48.11 4.14 12.17

smoke=1 838 28 17.71 5.98 7.92

smoke=2 347 11 7.18 2.03 2.26

Chisq= 12.2 on 2 degrees of freedom, p= 0.002

Dado que el p-valor asociado a la prueba es de 0.002, que es menor que 0.05, se rechaza la hipótesis nula. Es decir, la variable fuma si afecta en el comportamiento del tiempo al que es hospitalizado el bebe por primera vez.

Por tanto es conveniente realizar pruebas de comparaciones múltiples. Utilizando el ajuste por Bonferroni se tiene el siguiente resultado:

Pairwise comparisons using Log-Rank test

data: datos and smoke

0 1

1 0.0035 -

2 0.0671 1.0000

Este resultado nos dice que existen diferencias entre el grupo de bebés con madres que no fuman y el grupo con las madres que fuman menos de una cajetilla (p-valor 0.0035). Para el grupo de madres que fuman más de una cajetilla no se puede afirmar que difiere de los otros dos grupos (p-valores 0.0671 y 1 respectivamente).

Una sugerencia de partición es como sigue: se puede hablar de dos grupos: no fuman y si fuman, con el grupo de si fuman definido con fumar desde un cigarro al día. Esta es una sugerencia basada en que se espera que el grupo conformado con bebés cuyas madres fuman más de una cajetilla al día tenga un efecto igual o peor que el efecto de fumar menos de una cajetilla al día, por lo que se esperaría que el grupo 0 también sea diferente al grupo 2, dado que fue diferente al grupo 1.

Conclusión en términos del problema (2 puntos)

Ejercicio 3

Si la probabilidad de sobrevivir 11 meses es 0.5, y 5 de los 10 sujetos restantes mueren en el doceavo mes de un estudio, ¿cuál es la probabilidad de sobrevivir un año completo? (1 punto)

- a) 0.4, b) 0.25, c) 0, d) 0.5

Respuestas: $S(11)=0.5$, y $S(12)=S(11)(5/10)=(0.5)(0.5)=0.25$

Ejercicio 4

En la siguiente gráfica explica porque la función de supervivencia KM del grupo 1 no se hace cero, mientras que la del grupo 2 sí. (1 punto)

En el grupo 1 el dato observado más grande corresponde a una censura, mientras que en el grupo 2 corresponde a una muerte.

Ejercicio 5.

Utilizando la información de la siguiente gráfica proporciona estimadores puntuales para la vida mediana de cada grupo. Justifica tu respuesta (2 puntos)

Grupo 2: cualquier dato alrededor del 230

Grupo 1: cualquier dato alrededor del 270

La mediana del grupo 1 es mayor que la mediana del grupo 2.

Ejercicio 6.

La siguiente salida de R corresponde al análisis de un tiempo de vida medido en semanas.

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
9	11	1	0.909	0.0867	0.7541	1.000
13	10	1	0.818	0.1163	0.6192	1.000
18	8	1	0.716	0.1397	0.4884	1.000
23	7	1	0.614	0.1526	0.3769	0.999
31	5	1	0.491	0.1642	0.2549	0.946
34	4	1	0.368	0.1627	0.1549	0.875
48	2	1	0.184	0.1535	0.0359	0.944

Realiza las siguientes actividades

- a) Estima la tasa instantánea de riesgo en el tiempo 23 semanas (1 punto)

Resp= $d_i/n_i = 1/7 = 0.1428$

- b) ¿Cuál es la probabilidad de que un sujeto sobreviva mas de 7 meses? (1 punto)

Resp: $S(23) = 0.614$, ya que 7 meses es aproximadamente 28 semanas. Y como 28 semanas esta entre 23 y 31 semanas, se toma el valor de la función de supervivencia en las 23 semanas.

Pregunta 7

Se realizó un estudio con un grupo de individuos adultos seleccionados al azar de una población de interés. El estudio tuvo una duración de 30 años y se estuvo interesado en la edad en la que por primera vez el individuo presentó síntomas de cierta enfermedad.

En cada una de las situaciones descritas a continuación determina si se presentó censura, truncamiento o muerte en el dato, así como el dato registrado en ese individuo. En caso de estar presente la censura, identificar el tipo de censura o truncamiento: por la derecha, por la izquierda o por intervalo.

- a) Un individuo entró al estudio a la edad de 45 años y cuando entró al estudio ya había presentado los síntomas (1 punto)
- b) Dos individuos entraron al estudio a la edad de 30 y 42 años. Al entrar al estudio estaban sanos (no tenían los síntomas), y nunca presentaron los síntomas durante el estudio (2 puntos)
- c) Dos individuos entraron al estudio a la edad de 35 y 40 respectivamente. Ellos mostraron los síntomas a los 6 y 15 años de que ingresaron al estudio, respectivamente. (2 puntos)

Respuestas:

- a) Presenta censura por la izquierda, pues ya había experimentado el evento de interés. Se registra el valor 45, la edad a la que entra al estudio
- b) Ambos presentan censura por la derecha, pues se terminó el estudio y no experimentaron el evento de interés. Los datos registrados son $60=30+30$ y $72=42+30$
- c) Se observaron muertes en ambos casos. Los datos registrados son $41=35+6=$ y $55=40+15$