UNIVERSIDAD DE COSTA RICA

ESCUELA DE ESTADÍSTICA

XS-3010 DEMOGRAFIA APLICADA

Profesor: Arodys Robles

Curvas de sobrevivencia. Kaplan-Meier

Tiempo en meses	Número en riesgo	Muertes	Salidas	Probabilidad de muerte en el intervalo		Probabilidad de sobrevivir el intervalo		Función de sobrevivencia S(t)
0	10	0	0					1
2	10	1	0	1/10	0,100	9/10	0,900	0,9000
4+	9	0	1	0	0,000	1	1,000	0,9000
5	8	1	0	1/8	0,125	7/8	0,875	0,7875
6	7	1	0	1/7	0,143	6/7	0,857	0,6750
9	6	2	0	2/6	0,333	4/6	0,667	0,4500
12	4	1	0	1/4	0,250	3/4	0,750	0,3375
12+	3	0	1	0	0,000	1	1,000	0,3375
15+	2	0	1	0	0,000	1	1,000	0,3375
17	1	1	0	1	1,000	0	0,000	0,0000

+ indica una salida de observación

Kaplan-Meier: la estimación de S(t) se hace solo para los tiempos en que ocurre un evento (o falla).

Tiempo en meses	Número en riesgo	Muertes	Salidas	Probabilidad de muerte en el intervalo		Probabilidad de sobrevivir el intervalo		Función de sobrevivencia S(t)
0	10	0	0					1
2	10	1	1	1/10	0,100	9/10	0,900	0,9000
5	8	1	0	1/8	0,125	7/8	0,875	0,7875
6	7	1	0	1/7	0,143	6/7	0,857	0,6750
9	6	2	0	2/6	0,333	4/6	0,667	0,4500
12	4	1	2	1/4	0,250	3/4	0,750	0,3375
17	1	1	0	1	1,000	0	0,000	0,0000

Curvas de sobrevivencia. Censura

Censura: cuando tenemos información del tiempo de sobrevivencia, pero no conocemos el tiempo exacto.

Sin censura: casos en que conocemos el inicio y el fin de la duración.

censura a la derecha: algunos individuos no experimentan el evento durante el tiempo de observación.

Ej. Transición al segundo hijo (información obtenida en forma retrospectiva). Mujeres que han tenido un primer hijo, pero al momento de la encuesta no han tenido un segundo.

Supuesto la censura es aleatoria e independiente de lo que determina la ocurrencia de un evento.

Censura tipo I.

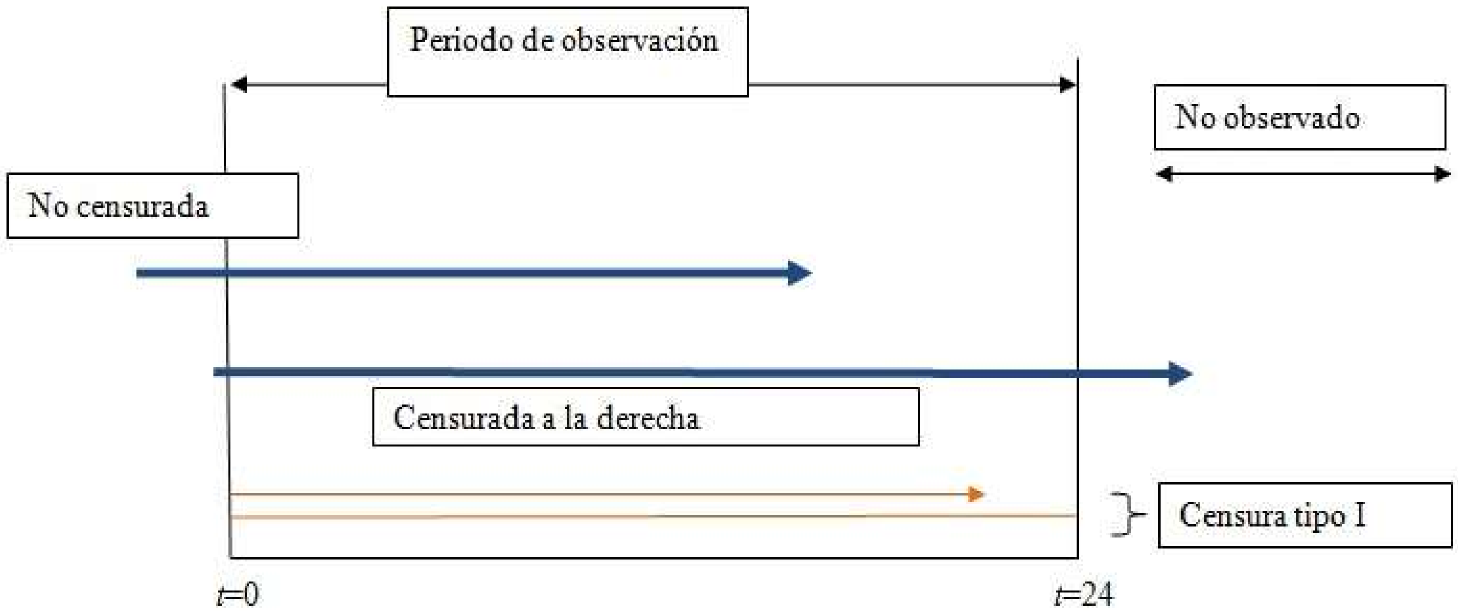
Censura no es aleatoria.

Ej. Pruebas de materiales. El seguimiento se detiene cuando algunos casos no han fallado.

Curvas de sobrevivencia.

Truncados: no hay información sobre la ocurrencia del evento: algunos sujetos no pueden ser observados.

Truncados a la derecha. Ej. Estudios basados en registros de muerte con información retrospectiva . Ej. Estudios SIDA y transfusión.


Periodo de observación
No observado
No censurada
Censurada a la derecha
Censura tipo I
t=0
t=24

Curvas de sobrevivencia. Kaplan-Meier

Estimación de la sobrevivencia método Kaplan Meier

p1= probabilidad de sobrevivir 1 día

p2= probabilidad condicional de sobrevivir el segundo día dado que sobrevivió el primero

p365= probabilidad condicional de sobrevivir el 365avo día dado que sobrevivió el día 364

S(365)=p1*p2*p3*…*p365

S(t)=p1*p2*…*pt

pt= pacientes vivos t-1 días que sobreviven el día t/ pacientes vivos al final del día t-1

$$pt = \frac{pacientes\ vivos\ t-1\ días\ que\ sobreviven\ el\ día\ t}{pacientes\ vivos\ al\ final\ del\ día\ t-1}$$

Tiempo t inicio de un intervalo que termina en t+1

nt número de sujetos vivos al inicio del intervalo y en riesgo de morir en el intervalo

Curvas de sobrevivencia. Kaplan Meier

Tiempo t inicio de un intervalo que termina en t+1

nt número de sujetos vivos al inicio del intervalo y en riesgo de morir en el intervalo

dt sujetos que mueren después del inicio del intervalo t

nt-dt sujetos que sobreviven el intervalo equivalente a los que inician t+1 (nt+1)

$$p_t = \frac{(n_t - d_t)}{n_t} = 1 - \frac{d_t}{n_t}$$

Si d_t=0 p_t=1

S(t) probabilidad de sobrevivir el momento t solo cambia cuando hay una muerte o falla.

$$S(t) = \left(1 - \frac{d_1}{n_1}\right)\left(1 - \frac{d_2}{n_2}\right)\dots\left(1 - \frac{d_t}{n_t}\right)$$

$$S(t) = \prod_t \left(1 - \frac{d_t}{n_t}\right)$$

$$S(t) = S(t-1)\left(1 - \frac{d_t}{n_t}\right)$$

Si t=0 entonces S(0)=1

Supuestos

Selección aleatoria o muestra representativa

Independencia: (cuando se juntan muestras de dos sitios o se toman muestras de una misma familia)

Consistencia del criterio de entrada al estudio (cambios en diagnóstico)

Fin de observación definido consistentemente

Momento de inicio del estudio claramente definido (intención de tratar)

Censura no relacionada con sobrevivencia (salida de los más enfermos o los que se sienten sanos: inclusión y exclusión crea un sesgo)

Sobrevivencia media no cambia durante el estudio (tiempo de reclutamiento)

Intervalos de confianza

Si los tiempos t tienen una distribución normal:

$$S(t) - 1.96ES(S(t))...S(t) + 1.96ES(S(t))$$

A medida que queden menos sujetos los intervalos se van haciendo más amplios $\left(\frac{Desv.Est.}{\sqrt{n}}\right)$

Alternativa es corregir el número en observación: Greenwood

$$ES(S(t)) = S(t)\left(\sum_{j=0}^{t-1} \frac{d_j}{n_j(n_j - d_j}\right)^{\frac{1}{2}}$$

d_j muertes en el día j

n_j sujetos vivos y en el estudio al inicio del día j

Ej. t=12 S(12)=0.6522

$$ES(S(12)) = .6522\left[\frac{4}{23(23-4)} + \frac{2}{19(19-2}\right]^{\frac{1}{2}}$$

ES(S(12)=.6522[.12387]=0.0808

Estimación de la función de sobrevivencia al momento t S(t)

S(t) probabilidad de sobrevivir el momento t solo cambia cuando hay una muerte .

$$S(t) = \left(1 - \frac{d_1}{n_1}\right)\left(1 - \frac{d_2}{n_2}\right)\ldots\left(1 - \frac{d_t}{n_t}\right)$$

$$S(t) = \prod_t \left(1 - \frac{d_t}{n_t}\right)$$

Kaplan-Meier

$$S(t) = S(t-1)\left(1 - \frac{d_t}{n_t}\right)$$

Si t=0 entonces S(0)=1

Medidas resumen:
mediana

Cuánto tarda en morirse la mitad de los sujetos.

No se puede definir si la mitad de los sujetos están vivos al final del estudio
(la mediana sería mayor que el último tiempo de sobrevivencia pero desconocida)

Cuando la curva de sobrevivencia es horizontal en 50% se pueden promediar los dos valores

Sobrevivencia después de x años (generalmente 5 en estudios bioestadísticos)

Proporción que sobrevive hasta los 5 años.

Comparación de 2 curvas de sobrevivencia

Supuestos:

1. Sujetos pertenecen a grupos distintos definidos antes de que empiece la recolección de información.
(no se pueden separar los grupos por algo que ocurre durante el periodo de observación (no aparición de un síntoma))

2. Consistencia en la definición de grupos a lo largo del estudio.
Cambio produce una paradoja la sobrevivencia de **todos los grupos** aumenta o disminuye
Ej. mejora en el diagnóstico

localizado	**Mejora en la identificación de metástasis**	**localizado**	**Localizado sin evidencia de metástasis**	**Mayor sobrevivenci a**
metástasis		metástasis	+ metástasis en fase inicial	Mayor sobrevivencia

3. Supuesto todos los sujetos reciben el tratamiento

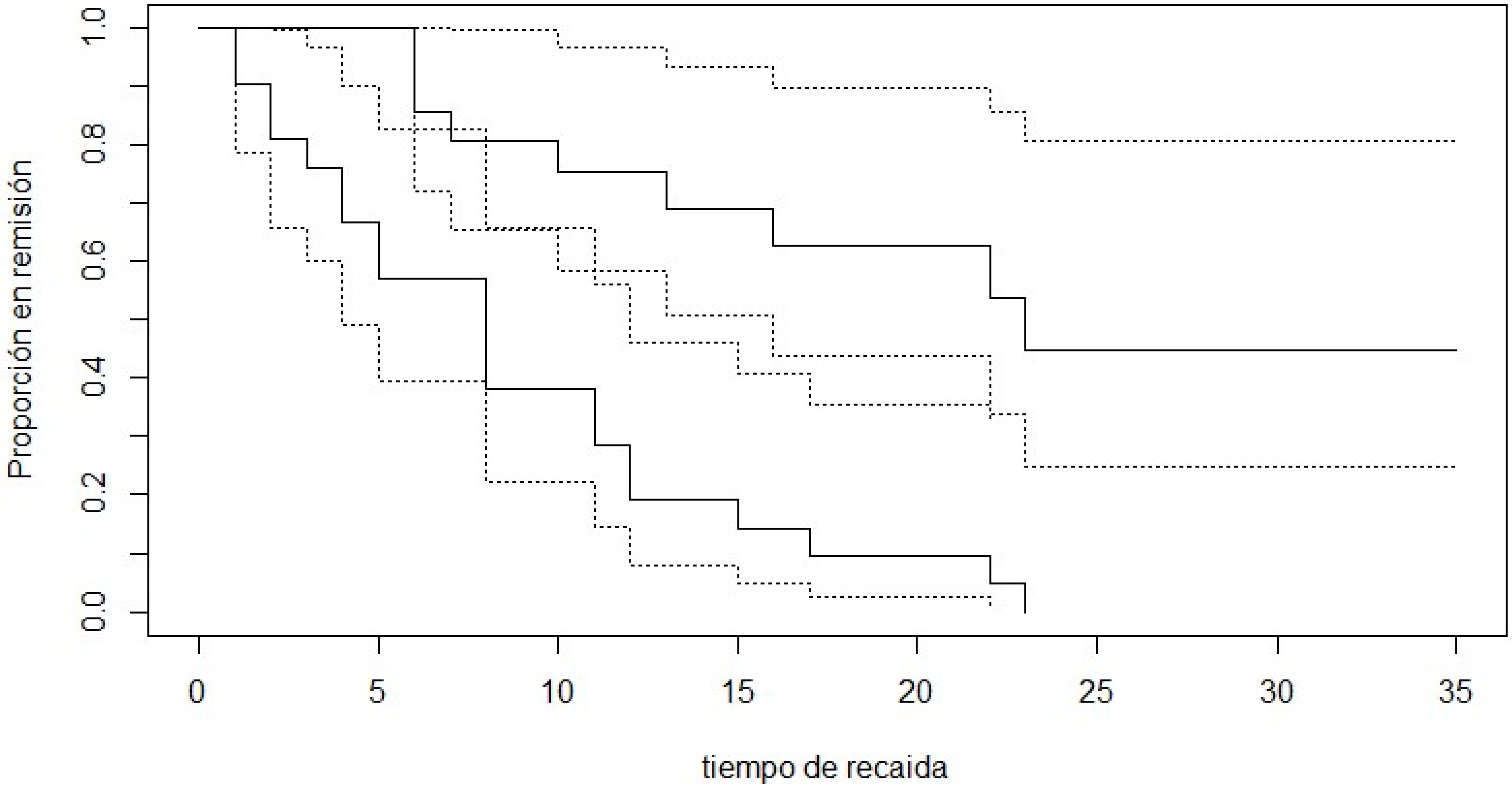
Tratamientos pueden no estar completos: ***intención de tratar***
Analizar: tratamiento y tratamiento incompleto y comparar.

4. riesgos proporcionales

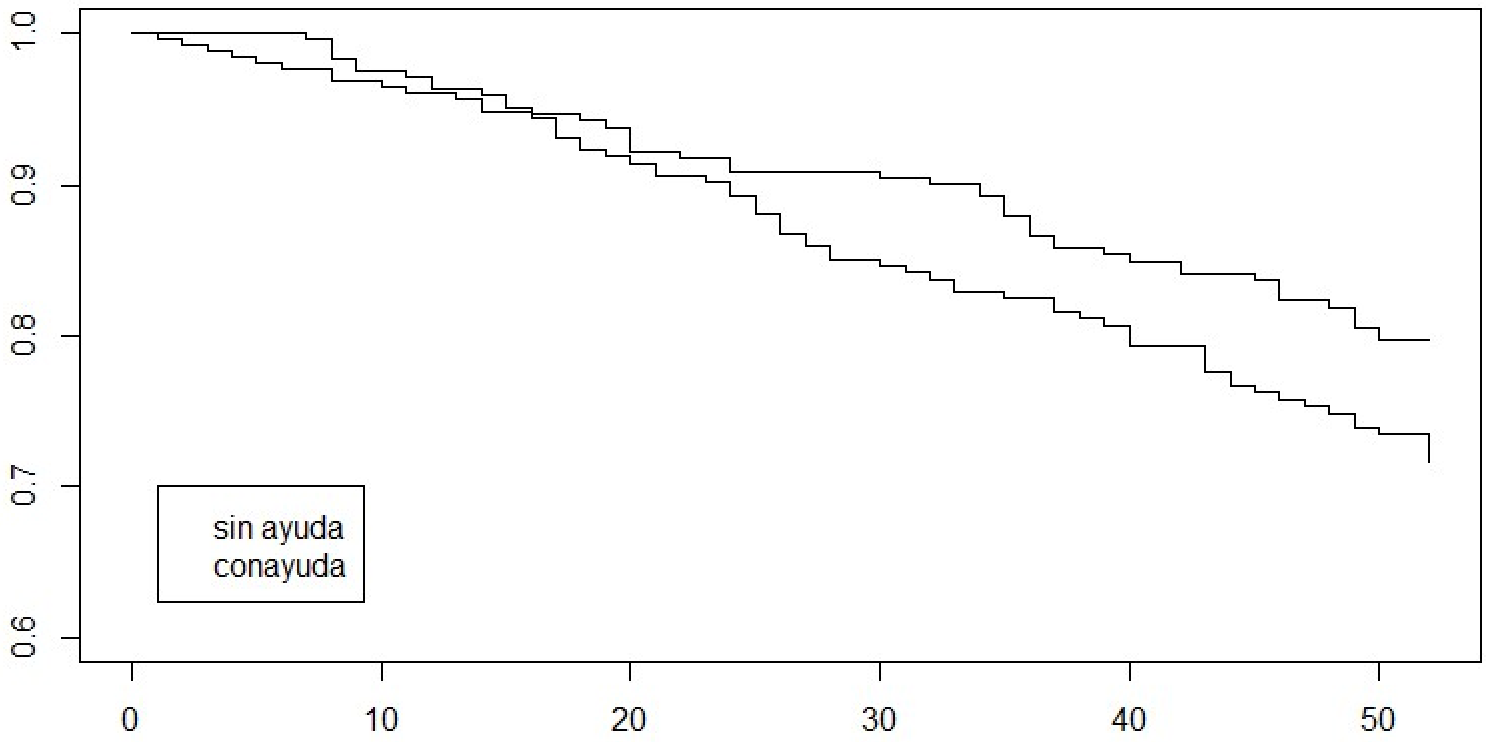
Razón de riesgos de los dos grupos se mantiene constante a lo largo de los valores de la curva

cirugía	Riesgo inicial alto	Riesgo baja
tratamiento	Riesgo inicial bajo	Riesgo aumenta

Estimador Kaplan Meier para dos tratamientos



Estimador Kaplan Meier reincidencia según ayuda económica o no



Comparación de dos curvas

Pruebas para comparar curvas

Log-rank

Comparación entre observados y esperados χ^2

Esperados si tenemos dos grupos:

$$e_{1f} = \left(\frac{n_{1f}}{n_{1f} + n_{2f}} \right) \left(m_{1f} + m_{2f} \right)$$

Proporción en el conjunto de riesgo * número de fallas en los dos grupos

$$e_{2f} = \left(\frac{n_{2f}}{n_{1f} + n_{2f}} \right) \left(m_{1f} + m_{2f} \right)$$

Comparación de dos curvas

$$O_i - E_i = \sum_{f=1}^{\# f} m_{if} - e_{if}$$

$$Log\_rank = \frac{\left(O_2 - E_2\right)^2}{Var\left(O_2 - E_2\right)}$$

$$Var\left(O_i - E_i\right) = \sum \frac{n_{1f} n_{2f} (m_{1f} + m_{2f})(n_{1f} + n_{2f} - m_{1f} - m_{2f})}{(n_{1f} + n_{2f})^2 (n_{1f} + n_{2f} - 1)}$$

n_{if} número en riesgo en cada grupo

m_{if} número de fallas en cada grupo en el momento f

Suma sobre todos los tiempos de falla

Comparación de dos curvas

H_0: no hay diferencia entre las curvas

Log-rank χ^2 con 1 grado de libertad

$$O - E = \sum (m_{if} - e_{if})$$

i grupo

f tiempos de falla

todos los tiempos de falla tienen el mismo peso.

Comparación de dos curvas

Problema de la reducción del tamaño de muestra en cada estrato, alternativas:

Wilcoxon: $w(t_{(f)}) = n_f$

peso es el número en riesgo n_f en todos los grupos al momento t_f

Fallas iniciales reciben más peso

Tarone-Ware mayor peso al inicio $\sqrt{n_f}$

Peto $\widehat{S}(t_{(f)})$ estimador de sobrevivencia calculado para todos los grupos

Flemington-Harrington

$\widehat{S}(t_{(f-1)})^p \left[1-\widehat{S}(t_{(f-1)})\right]^q$ p y q son asignados

$w(t) = \widehat{S}(t_{(f-1)})$ p=1,q=0 mayor peso a los primeros tiempos de sobrevivencia

$w(t) = 1-\widehat{S}(t_{(f-1)})$ p=0,q=1 tiempos mayores reciben mayor peso.