

Análisis de Supervivencia

Introducción y supuesto clave

Vamos a utilizar R como entorno de análisis porque es un software libre y ampliamente utilizado en bioestadística y salud pública. Trabajaremos con un dataset en formato CSV `datos_kaplan_meier.csv`, que podrás descargar y cargar en R. Te invito a explorarlo.

El estimador de Kaplan–Meier (K-M) es un método **no paramétrico** para describir la supervivencia a lo largo del tiempo. Por lo tanto, no requiere asumir una distribución específica para los tiempos de supervivencia.

El supuesto más importante es la **censura no informativa**: se asume que las personas censuradas (por ejemplo, abandonan el estudio por decisión propia o son retiradas por los investigadores) tendrían un comportamiento similar al de quienes continúan bajo observación, si pudiéramos seguirlas. Es decir, la censura no debe estar relacionada con el pronóstico. Si la censura estuviera asociada al riesgo (censura “informativa”), la validez del análisis se vería comprometida.

Construcción de la tabla Kaplan–Meier

En el script, la tabla `tabla_km_directa` se genera con `summary(km_fit)`. Es importante notar que:

- Solo se crea una fila cuando ocurre al menos una muerte (evento).
- Los censurados no generan un salto en la supervivencia, pero sí influyen en el número de personas en riesgo en tiempos posteriores.

Interpretación de las columnas:

- **tiempo**: momento en que ocurre una o más muertes.
- **ni (n.risk)**: número de individuos **en riesgo** justo antes de ese tiempo (vivos y no censurados previamente).
- **di (n.event)**: número de muertes en ese tiempo.
- **si/ni**: proporción que “sobrevive” ese instante. En términos de K-M equivale a

$$(n_i - d_i) / n_i$$

S_acum: supervivencia acumulada hasta ese tiempo, calculada como el **producto** de los cocientes $\prod (s_i / n_i)$ de ese tiempo y los anteriores

Interpretación de la curva de Kaplan–Meier

La curva siempre:

- **Empieza en 1** (100% de supervivencia) y se mantiene horizontal hasta la primera muerte.
- **Desciende en escalones** únicamente cuando ocurren muertes.
- Si el seguimiento termina con personas vivas (censura al final), la curva queda **horizontal** al final.
- Si el último sujeto observado **fallece** al final del seguimiento, la curva puede terminar con un descenso hasta 0.

En el ejemplo:

- Entre 0 y 4 años, $S(t) = 1$
- Al año 4, $S(t)$ desciende (por muertes) y se mantiene constante hasta el siguiente tiempo con evento, y así sucesivamente.
- Al año 14, ocurre el último evento observado y la supervivencia acumulada alcanza su valor final.

Tabla de riesgo y precisión de la estimación

Con el tiempo, el número “en riesgo” suele disminuir, por lo que la incertidumbre aumenta en la parte final de la curva. Por eso la tabla de riesgo bajo la gráfica es importante: ayuda a interpretar qué tan estable es la estimación en cada tramo del seguimiento.

Supervivencia mediana (y un error común)

La mediana de supervivencia es el primer tiempo en que $S(t) \leq 0.5$. Gráficamente, se obtiene trazando una línea horizontal desde 0.5 en el eje Y y viendo dónde corta la curva.

Un error frecuente es confundirla con la **mediana del tiempo de seguimiento**.

No son lo mismo: la mediana de supervivencia depende de la curva $S(t)$, no del orden estadístico de los tiempos.

Además, puede ocurrir que no exista mediana de supervivencia si $S(t)$ nunca baja a 0.5 durante el seguimiento.



¿Cómo influyen los censurados si no producen saltos?

Los censurados no generan un descenso inmediato de la curva, pero **reducen el denominador (n_i)** en tiempos posteriores. Esto puede hacer que los siguientes descensos (cuando haya muertes) sean más grandes, porque ocurren con menos personas en riesgo.

¿Qué sigue después de Kaplan–Meier?

Para comparar curvas entre grupos se ha usado tradicionalmente el log-rank test (contrasta diferencias globales de supervivencia a lo largo del tiempo). En la práctica, estas comparaciones suelen complementarse o reemplazarse por la regresión de Cox, que permite ajustar por varias covariables y estimar razones de riesgo (HR).

Referencias

- Martínez, Miguel, Almudena Sánchez-Villegas, Estefanía Toledo, y Javier Faulin. 2020. Bioestadística Amigable. 4a ed. Barcelona: Elsevier España.
- Dawson, Beth, y Robert G. Trapp. 2005. Bioestadística Médica. 4a ed. El Manual Moderno.