

Tema 5: Análisis de Supervivencia

Metodología de la Investigación en Fisioterapia

Miguel González Velasco
Departamento de Matemáticas. Universidad de Extremadura



- 1 Función de Supervivencia
- 2 Estimación de la Función de Supervivencia
 - Método de Kaplan-Meier
- 3 Comparación de Funciones de Supervivencia

Modelo Probabilístico

Modelizar el tiempo que transcurre hasta que se produce un determinado evento.

$T \equiv$ tiempo que tarda en producirse el evento. (si el evento es la muerte, T será el tiempo de vida o supervivencia del paciente)

- T es una variable aleatoria no negativa, habitualmente continua.
- Su distribución de probabilidad no se ajusta a una normal salvo en casos excepcionales.
- Sus distribuciones de probabilidad más frecuentes son la exponencial, la Weibull, la gamma o la log-normal.

Definición

La **Función de Supervivencia**, $S(t)$, se define como

$$S(t) = P(T > t), \quad t \geq 0$$

Intuitivamente $S(t)$ representa la **probabilidad de que un paciente sobreviva más de un tiempo t** . Si $F(t)$ es la función de distribución de T entonces $S(t) = 1 - F(t)$ y por tanto:

- $S(t)$ es decreciente. La velocidad a la que decrezca en un instante indica el riesgo de que el paciente muera en ese instante de tiempo.

Modelo Estadístico

- Los individuos son sometidos a un seguimiento a lo largo del tiempo hasta que se produzca determinado evento.
- Ha de existir un punto de comienzo del seguimiento bien definido (operación quirúrgica, iniciación del tratamiento, etc.)
- El evento al que se debe llegar debe estar también claramente definido (muerte, recaída de la enfermedad, recuperación de la enfermedad, etc.)
- La entrada de nuevos individuos en el estudio puede tener lugar en cualquier instante dentro del periodo de seguimiento, iniciado a partir de la primera entrada en el estudio.

Modelo Estadístico

El seguimiento de los individuos no siempre es posible llevarlo a cabo hasta que se produzca el evento. Puede ocurrir

- que se pierda algún individuo durante el seguimiento por alguna razón ajena al experimento (pacientes que hayan dejado de acudir a los controles, etc.).
- que el seguimiento concluya antes de que algunos individuos hayan alcanzado el evento.

Los individuos que se encuentren en alguna de estas dos situaciones diremos que han sido *censurados*. Podemos llevar a cabo el análisis de supervivencia aunque algunos individuos hayan sido censurados.

Estimación de la Curva de Supervivencia: Kaplan-Meier

Utiliza los **tiempos exactos** donde se han producido los eventos: $t_1 < t_2 < \dots < t_N$

- n_k = número de individuos en riesgo de sufrir el evento en el instante inmediatamente anterior a t_k .
- d_k = número de eventos ocurridos en el instante t_k .

$P(T \leq t_k | T > t_{k-1})$ = probabilidad de que un individuo que ha sobrevivido hasta el instante inmediatamente anterior al t_k sufra el evento en el instante t_k

$$\hat{P}(T \leq t_k | T > t_{k-1}) = \frac{d_k}{n_k}, \quad k = 1, \dots, N \quad (t_0 = 0)$$

Estimador de Kaplan-Meier:

$$\hat{S}(t) = 1 \quad \text{si} \quad 0 \leq t < t_1$$

$$\hat{S}(t) = \prod_{k=1}^i (1 - d_k/n_k) \quad \text{si} \quad t_i \leq t < t_{i+1} \quad i = 1, \dots, N$$

Comparación de Curvas de Supervivencia

Para **comparar las curvas de supervivencia** de dos grupos de pacientes, planteamos el contraste de hipótesis

$$H_0 : S_1(t) = S_2(t)$$

$$H_1 : S_1(t) \neq S_2(t)$$

Para resolverlo se utiliza el **test log-rank**, cuyo valor experimental se basa en las diferencias entre el número de muertes observadas en cada grupo (O) y el número de muertes que cabría esperar (E) si no hubiera diferencias entre dichos grupos, todo ello en los tiempos donde se ha producido el seguimiento.

Este valor experimental se compara con un valor teórico de una distribución chi cuadrado con un grado de libertad.