



Pontificia Universidad  
**JAVERIANA**  
Bogotá

MAESTRÍA EN   
**EPIDEMIOLOGÍA**  
CLÍNICA

**BIOESTADÍSTICA AVANZADA**

## **MÓDULO III**

**Semana 11**  
Análisis de Supervivencia

# Práctica 1

## Taller de aplicación semana 1

La base de datos **lung.xlsx** contiene información de pacientes diagnosticados con cáncer de pulmón, a quienes se les realizó seguimiento en el tiempo con el fin de evaluar la ocurrencia del evento de interés (muerte) o la finalización del estudio sin que este ocurriera<sup>1</sup>. Adicionalmente, se recolectaron algunas variables clínicas, demográficas y nutricionales consideradas como posibles factores asociados al riesgo de mortalidad.

Las variables recolectadas fueron las siguientes:

**id:** Identificador del paciente.

**time:** Tiempo de seguimiento (en días) desde el inicio del estudio hasta la ocurrencia del evento o censura.

**status:** Indicador del evento:

1 = Censurado (vivo al final del seguimiento)

2 = Evento (muerte)

**age:** Edad del paciente (en años).

**sex:** Sexo del paciente:

1 = Masculino

2 = Femenino

**ph.ecog:** Estado funcional según la escala ECOG<sup>2</sup>:

0 = Asintomático

1 = Sintomático leve

2 = Ambulatorio, incapaz de trabajar

3 = Limitado severamente

4 = Postrado

---

<sup>1</sup> Los datos utilizados corresponden a la base de datos lung disponible en el paquete survival de R, basado en información clínica del North Central Cancer Treatment Group (Therneau & Grambsch, 2000).

<sup>2</sup> ECOG: Eastern Cooperative Oncology Group

**ph.karno:** Índice de Karnofsky evaluado por el médico (0–100), donde valores más altos indican mejor estado funcional.

**pat.karno:** Índice de Karnofsky reportado por el paciente (0–100).

**meal.cal:** Consumo calórico diario estimado.

**wt.loss:** Pérdida de peso reciente (en kilogramos).

El propósito de este estudio es analizar el tiempo hasta la ocurrencia de la muerte e identificar los factores pronósticos asociados. A partir de esta información, realice las siguientes actividades:

1. Estime la función de supervivencia global mediante el método de Kaplan–Meier considerando todos los pacientes del estudio. Reporte las estimaciones de supervivencia junto con sus intervalos de confianza al 95%.
2. Interprete la probabilidad de permanecer vivo en los tiempos  $t = 12$  meses y  $t = 24$  meses. Explique el significado clínico de estas probabilidades.
3. Grafique la curva de supervivencia global de Kaplan–Meier para todos los pacientes del estudio.
4. A partir del modelo de Kaplan–Meier, obtenga la estimación del tiempo mediano de muerte junto con su intervalo de confianza al 95%. Interprete el resultado obtenido.
5. Calcule los percentiles 10, 25, 75 y 90 del tiempo hasta la muerte. Explique qué representan estos percentiles en términos de evolución clínica.
6. Estime la función de supervivencia estratificada por sexo y grafique las curvas de Kaplan–Meier para ambos grupos.
7. Con base en la gráfica de supervivencia por grupos, describa qué grupo presenta mayor probabilidad de permanecer vivo a lo largo del seguimiento y en qué periodo del seguimiento parecen separarse más las curvas.
8. Realice la prueba de log-rank para comparar las curvas de supervivencia entre los dos grupos. Reporte el estadístico de prueba, el valor p e interprete los resultados obtenidos.
9. Repita los puntos 6–8 para la variable estado funcional