MÉTODOS ECONOMÉTRICOS Y ORGANIZACIÓN INDUSTRIAL APLICADA

Trabajo Práctico 1
Métodos no paramétricos

Ma. Florencia Gabrielli
Diego Fernández Meijide

23 de junio de 2025

Modalidad: se deberá entregar un informe en el que comente los resultados de cada uno de los ejercicios y un archivo que contenga todos los códigos utilizados. Ambos archivos deberán ser enviados por mail a dfernandez-meijide@udesa.edu.ar con copia a florgabrielli@gmail.com al final del día (23/06/2025).

Objetivo: utilizar métodos no paramétricos para analizar la supervivencia de los pacientes y estimar la densidad de los tiempos de supervivencia. Se espera que incorpore la teoría discutida en clase a lo largo de su trabajo, identificando y enfocándose en los temas más relevantes.

Un grupo de investigadores está estudiando la eficacia de un nuevo tratamiento para una enfermedad crónica. Han recogido datos de un estudio clínico donde los pacientes fueron seguidos durante varios años. La base `resultados.csv` contiene el tiempo de muerte de cada paciente, edad, sexo, estado de la enfermedad, tipo de tratamiento y factores genéticos de cada paciente.

A continuación se presenta la lista de variables que van a encontrar en la base de datos:

Variable	Descripción	Unidades de medida
`time`	Tiempo de supervivencia del paciente luego del tratamiento	Días
`age`	Edad del paciente al comienzo del tratamiento	Años
`sex`	Sexo del paciente	Femenino, Masculino
`disease_state`	Estado de la enfermedad al comienzo del tratamiento	Avanzado, Moderado, Leve
`treatment`	Tipo de tratamiento	A, B
`genetic_factor`	Factores genéticos adversos presentes al comienzo del tratamiento	0 (no)/ 1 (si)

Cuadro 1: Descripción de las variables

Consignas

1. Estimen y grafiquen la función de distribución acumulada para toda la muestra (ecdf), sin distinguir entre diferentes grupos de personas. Interpreten el valor de la ECDF cuando el tiempo de supervivencia es 50 unidades de tiempo.
2. Definimos a la función de supervivencia, $S(t)$, como la probabilidad de que un individuo sobreviva más allá del tiempo t:

   $$S(t) = P(T > t)$$

   Donde T es una variable aleatoria que representa el tiempo de supervivencia. Estimen y grafiquen la función de supervivencia e interpreten los resultados obtenidos, ¿cómo describirían los resultados del tratamiento sobre la supervivencia? (mayores momentos de mortalidad, tiempo mediano de supervivencia, etc). [1] **Bonus**: pueden agregar intervalos de confianza.
3. Estimen la densidad de los tiempos de supervivencia utilizando alguno de los *kernels* vistos en clase (definan el *bandwith* con la Regla de Oro de Silverman) y grafiquen. ¿Cuáles son los tiempos de supervivencia más comunes? ¿Qué pueden decir sobre la variabilidad de los tiempos de supervivencia? Comparen la densidad estimada en el tiempo 50 y 10 e interpreten.

[1] Ayuda: escriban la expresión de la ECDF!

4. Generen un grupo de gráficos que muestre cómo cambia la estimacion de la función de distribución acumulada empírica a medida que aumenta la cantidad de observaciones. Para esto pueden sacar muestras aleatorias de la base con diferente cantidad de observaciones. Expliquen lo que observan.
5. Generen un grupo de gráficos que muestre cómo cambia la estimación de densidad de los tiempos de supervivencia para diferentes valores del *bandwith* (utilice un kernel Gaussiano). Expliquen lo que observan.
6. Tal como se explicó al comienzo del trabajo, en la base cuentan con variables que indican el tipo de tratamiento, edad del paciente, entre otras. Consideren al menos dos variables para crear subgrupos de pacientes y repita los puntos 1 y 2. En este inciso es importante que comparen detalladamente los diferentes subgrupos.