

03 – Análisis del dataset Titanic

Autor: Peitsch, Pablo mail: preynosopeitsch@estudiantes.unsam.edu.ar
Github: @PPeitsch Fecha: 2023-04-10

Mediante un breve análisis, anteriormente, se hizo un breve análisis del conjunto de datos de Titanic, que contiene datos de los pasajeros que viajaban en el RMS Titanic en abril del año 1912, cuando se hundió, y que se encuentra disponible de forma abierta en OpenML [1] y Encyclopedia Titanica [2]. Se observó cómo influyó el poder adquisitivo en la supervivencia del accidente: aquellos que contaban con tickets de primera clase tenían mayor tasa de supervivencia que los de segunda y tercera clase, Figura 1, a. En el presente análisis se pretende encontrar si existe una relación entre la supervivencia y el sexo de los pasajeros, así como la edad. Para ello, se importan los datos y se cambian algunos nombres de columna para mejorar la legibilidad. Luego, se utilizan funciones de agrupación y la librería ggplot2 para crear un gráfico de barra y un gráfico de densidad.

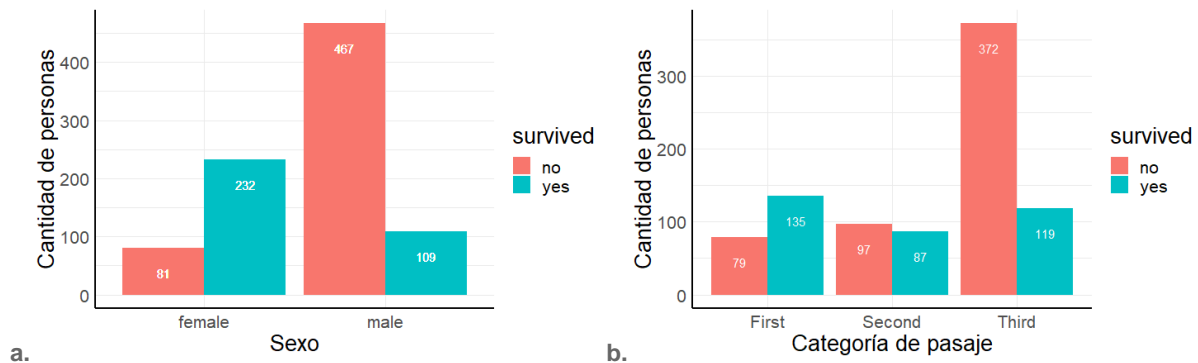


Figura 1. a. Número de supervivientes en el accidente del RMS Titanic en abril del año 1912, en función del a. Sexo y b. Categoría de pasaje.

El análisis puede resultar útil para identificar posibles patrones en los datos y entender cómo las categorías de clase de pasaje y el sexo influyeron en la supervivencia en el Titanic. En la Figura 1, b. los resultados muestran que las personas de sexo femenino tuvieron una mayor tasa de supervivencia. Luego, en la Figura 2, se puede apreciar que la densidad poblacional de pasajeros estaba concentrada entre los 20 y 40 años principalmente. Por otro lado, no hay mucha diferencia en la forma de la curva de densidad entre supervivientes y no, en población femenina; en cambio, la curva de supervivientes masculinos se corre ligeramente hacia la derecha, pronunciando los menores.

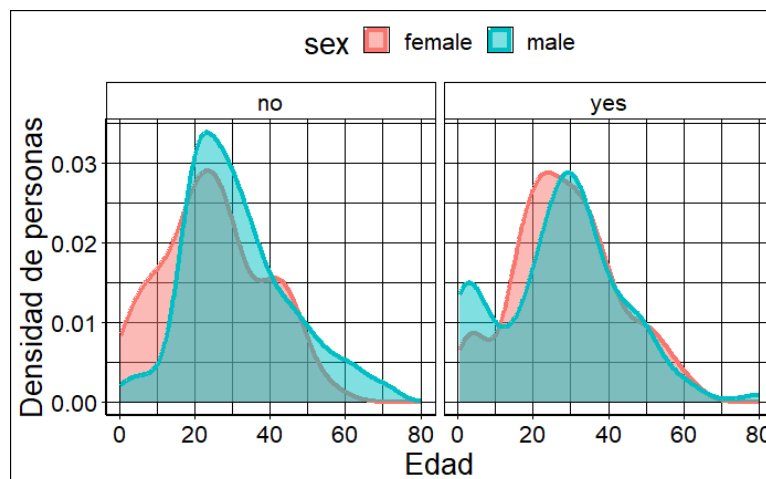


Figura 2. Densidad poblacional de pasajeros en el accidente del RMS Titanic en abril del año 1912, en función de la Edad, agrupados por Sexo y condición de supervivencia.

Los gráficos sugieren que principalmente los factores de supervivencia pueden haber sido el económico y el sexo. La edad parece no jugar un papel fundamental en esta ocasión, aunque en población masculina hay ligeras diferencias: mayor supervivencia en menores de 10 años y menor supervivencia entre 10 y 20 años.

REFERENCIAS.

1. OpenML, <https://www.openml.org/search?type=data&sort=runs&id=40945&status=active>
2. Encyclopedia Titanica, <https://www.encyclopedia-titanica.org/>