

Tarea 1: Probabilidades

DPDC 102 : Modelos Estadísticos

01.05.2022

Considere la base de datos 'titanic.csv', disponible en el apartado **Tarea 1** del aula virtual del curso.

Pregunta 1 (5 puntos)

Generar un histograma de frecuencias para la variable 'Age', que representa la edad de los pasajeros del *Titanic*. Establezca un número apropiado de marcas de clase (bins) para visualizar de buena forma la posible distribución de probabilidad de la variable.

Pregunta 2 (40 puntos, 10 puntos cada letra)

Suponga que esta variable tiene una distribución de probabilidad de tipo Normal, es decir $Age \sim N(\mu, \sigma^2)$, donde su media μ y σ^2 representan el promedio la varianza de las observaciones, respectivamente.

- Grafique la distribución teórica para la variable, sobre el histograma de frecuencia de los datos observados.
- Calcule la probabilidad de que las edades de los pasajeros del Titanic sean menores a 18 años $\mathbb{P}(Age \leq 18)$.
- Calcule la probabilidad de que hayan pasajeros mayores a 80 años $\mathbb{P}(Age \geq 80)$.
- Calcule la probabilidad de que los pasajeros tengan entre 20 y 40 años $\mathbb{P}(20 \leq Age \leq 40)$.

Pregunta 3 (15 puntos)

Use filtros para separar la variable 'Age' en las edades de las personas que sobrevivieron y las que no sobrevivieron. Puede utilizar el siguiente código como ayuda para establecer los filtros:

```
titanic[(titanic.Survived == 0)]['Age'] #Para las personas que no sobrevivieron
titanic[(titanic.Survived == 1)]['Age'] #Para las personas que si sobrevivieron
```

Se dice que las personas más jóvenes tuvieron una probabilidad más alta de sobrevivencia al accidente del Titanic. Para comprobar la veracidad de este comentario, suponga que las edades presentan una distribución de tipo Normal, encuentre los parámetros μ, σ^2 para la distribuciones de las edades de los pasajeros que sobrevivieron y los que no sobrevivieron. Elabore gráficos de las distribuciones teóricas encontradas, compare los parámetros y comente respecto a los resultados encontrados.

Condiciones:

- Fecha de entrega viernes 06 de mayo del 2022.
- Los grupos de trabajo son los mismos grupos ya establecidos en el módulo anterior.
- Formato de entrega: Un archivo de Python (.py) con los códigos ejecutados para cada respuesta, puede usar comentarios para separar los ejercicios y escribir sus respuestas dentro del código.