

## Desafío - Tabaquismo y gestación

En este desafío validaremos nuestros conocimientos sobre variables aleatorias. Para lograrlo, necesitarás aplicar lo visto en clases.

Lee todo el documento antes de comenzar el desarrollo **individual**, para asegurarte de tener el máximo de puntaje y enfocar bien los esfuerzos.

Tiempo asociado: 2 horas cronológicas

### Descripción

Un grupo de investigadores está interesado en analizar la relación entre el hábito de fumar con el crecimiento de los bebés en edad gestacional. Para esto han recopilado los datos guardados en el dataset baby.csv, que contiene los siguientes campos de datos:

**Birth.Weight** : Peso del bebe

**Gestational.Days**: Días de gestación

**Maternal.Age**: Edad de la madre

**Maternal.Height**: Altura de la madre

**Maternal.Pregnancy.Weight**: Peso de la madre

**Maternal.Smoker**: ¿Es la madre fumadora? (**True**: fumadora, **False**: no es fumadora)

A partir de estos datos se solicita:

1. Realiza una exploración de los datos, considerando sus tipos y los descriptivos correspondientes.
2. Considera los pesos de los bebés. ¿Se puede decir que siguen una distribución normal? Justifica tu respuesta creando los gráficos y modelos necesarios.
3. Considerando lo anterior, verifica el modelo de distribución normal de los pesos para 5 valores diferentes. Compara con los datos reales.
4. Compara los pesos de los bebés de madres fumadoras y no fumadoras. ¿Qué diferencias existen? Explica y construye los modelos correspondientes.
5. Considerando los datos dados:
  - Calcula la probabilidad de que una madre escogida al azar fume.
  - A partir del valor anterior, construye un modelo de distribución binomial que modele el experimento correspondiente a escoger 5 madres al azar y observar cuántas fuman.
  - Utilizando este modelo, simula 1000 muestras de tamaño 8, y calcula sus medias muestrales.

- Utilizando los datos directamente, extrae 1000 muestras de tamaño 8 y calcula sus respectivas medias muestrales.

Utilizando lo anterior:

- compara gráficamente los resultados obtenidos anteriormente.
- verifica el teorema del límite central, considerando las muestras de datos reales y las simuladas con el modelo binomial.
- construye un modelo de distribución normal derivado del modelo binomial, y gráficalo junto a los anteriores para comparar. Concluye.

## Requerimientos

1. Conoce y aplica la distribución normal de datos, con sus respectivos parámetros. **(3 puntos)**
2. Relaciona las distribuciones reales de datos con modelos de distribución binomial y normal. **(4 puntos)**
3. Comprende y verifica la ley de los grandes números y el teorema del límite central. **(3 puntos)**



¡Mucho éxito!

### Consideraciones y recomendaciones

Debes entregar tu trabajo en un archivo de Jupyter Notebook, con las respuestas, tu código y comentarios necesarios para comprender tu procedimiento.