

## Trabajo Práctico Final

### Introducción

La obesidad y el exceso de grasa corporal se han convertido en problemas de salud pública de primer orden.

Los métodos más precisos para medir la grasa corporal (como la DXA o la pletismografía) son costosos, requieren equipos especiales y casi siempre se limitan a los grandes centros urbanos.

En el otro extremo, índices simples como el índice de masa corporal (IMC) o la relación cintura-cadera son baratos pero ofrecen una estimación indirecta y a veces poco fiable de la grasa corporal real.

¿Podemos encontrar una forma de, a partir de un reducido número de mediciones antropométricas, predecir con suficiente precisión el porcentaje de grasa corporal de un sujeto?

### El objetivo

El conjunto de datos `bodyfat.csv` contiene observaciones de ~250 adultos junto con:

- ``BodyFat` (%)`: porcentaje de grasa corporal medido por densitometría hidrostática (variable dependiente)
- ``Density` (g/cc)`: densidad corporal estimada por desplazamiento de agua.
- ``Age``, ``Weight``, ``Height``
- Circunferencias (cm): ``Neck``, ``Chest``, ``Abdomen``, ``Hip``, ``Thigh``, ``Knee``, ``Ankle``, ``Biceps``

Su misión es construir y comparar modelos de regresión que prevean `BodyFat` usando solo mediciones que pueda tomar fácilmente un nutricionista o un preparador físico con una cinta métrica y una balanza.

Se sugiere abordar, entre otros, los siguientes interrogantes:

- ¿Qué subconjunto mínimo de circunferencias mantiene el error de predicción por debajo de un umbral clínicamente aceptable (por ejemplo,  $\pm 3\%$ )?
- ¿Existen transformaciones (logaritmos, relaciones, índices) que mejoren la capacidad predictiva?
- ¿Cuánto aporta la información adicional de la densidad corporal (``Density``) frente a las mediciones puramente antropométricas?
- ¿Cómo influyen variables demográficas como la edad?

Para ello, se recomienda que se tengan en consideración las siguientes pautas generales que hacen a un análisis bayesiano:

- Análisis exploratorio de datos
- Propuesta y ajuste de modelos
  - Descripción matemática
  - Elicitación de los *priors*
  - Pruebas predictivas *a priori*
  - Ajuste del modelo
  - Evaluación de la convergencia de las cadenas de Markov
  - Exploración de la distribución *a posteriori* de los parámetros

- Pruebas predictivas *a posteriori*
- Evaluación del ajuste del modelo
- Interpretación de parámetros
- Comparación de modelos
- Análisis final y conclusión

Se recomienda fuertemente que se haga uso de diferentes visualizaciones para comunicar los resultados de las diferentes etapas del análisis y que se propongan y evalúen un mínimo de tres modelos.

La presentación deberá incluir:

- Introducción clara al problema
- Definición de las preguntas de investigación
- Descripción del conjunto de datos a utilizar
- Análisis exploratorio de los datos
- Fundamentación y análisis de los modelos propuestos
- Hallazgos, resultados y conclusiones