



## **EJERCICIO PRÁCTICO 13: REGRESIÓN LOGÍSTICA**

### **CONTEXTO**

Conocemos varias herramientas que facilitan la búsqueda y construcción de modelos de regresión lineal, además del proceso iterativo para conseguir un modelo confiable.

El objetivo de este ejercicio es utilizar herramientas y procedimientos análogos para crear y evaluar modelos de regresión logística (RLog) para predecir una variable dicotómica.

### **OBJETIVOS DE APRENDIZAJE**

1. Preparar un conjunto de datos para la construcción de modelos RLog.
2. Iterar en el proceso de selección de variables, creación y evaluación de modelos RLog, hasta conseguir uno que sea confiable y satisfactorio.

### **ÉXITO DE LA ACTIVIDAD**

El equipo es capaz de encontrar modelos RLog confiables y de buen desempeño al predecir una variable dependiente.

### **ACTIVIDADES**

1. El equipo descarga desde el directorio compartido para este ejercicio práctico el enunciado (único) y el archivo con los datos a utilizar.
2. El equipo recupera la muestra usada en el ejercicio práctico anterior.
3. El equipo identifica a los integrantes con RUT, nombre y apellido, como comentario al inicio de un script.
4. El equipo lee el enunciado y selecciona las columnas para trabajar de acuerdo a las instrucciones.
5. El equipo construye los modelos solicitados usando la muestra correspondiente.
6. El equipo sube el script con las actividades anteriores comentando en detalle los pasos seguidos.

Antes de que venza el plazo para esta actividad, cada equipo debe subir el script realizado al correspondiente directorio compartido o espacio destinado para ello en UVirtual, con el nombre "EP13-respuesta-sala-i", donde i es el número de la sala asignada.

### **ENUNCIADO**

Para esta actividad usaremos la misma muestra de medidas anatómicas seleccionada para el ejercicio práctico anterior desde los datos recolectados por Heinz et al. (2003). Como este ejercicio requiere de una variable dicotómica, vamos a realizar lo siguiente:

1. Crear la variable IMC (índice de masa corporal) como el peso de una persona (en kilogramos) dividida por el cuadrado de su estatura (en metros).
2. Si bien esta variable se usa para clasificar a las personas en varias clases de estado nutricional (bajo peso, normal, sobrepeso, obesidad, obesidad mórbida), para efectos de este ejercicio, usaremos dos clases: sobrepeso ( $IMC \geq 25,0$ ) y no sobrepeso ( $IMC < 25,0$ )<sup>1</sup>.
3. Crear la variable dicotómica EN (estado nutricional) de acuerdo al valor de IMC de cada persona.

Ahora podemos construir un modelo de regresión logística para predecir la variable EN, de acuerdo con las siguientes instrucciones:

1. Recordar las ocho posibles variables predictoras seleccionadas de forma aleatoria en el ejercicio anterior.
2. Seleccionar, de las otras variables, una que el equipo considere que podría ser útil para predecir la clase EN, justificando bien esta selección.
3. Usando el entorno R, construir un modelo de regresión logística con el predictor seleccionado en el paso anterior.
4. Usando herramientas para la exploración de modelos del entorno R, buscar entre dos y cinco predictores de entre las variables seleccionadas al azar, recordadas en el punto 3, para agregar al modelo obtenido en el paso 3.
5. Evaluar los modelos y “arreglarlos” en caso de que tengan algún problema con las condiciones que deben cumplir.
6. Evaluar el poder predictivo de los modelos en datos no utilizados para construirlo (o utilizando validación cruzada) y revisar las respectivas curvas ROC.

## Referencias

Heinz, G., Peterson, L. J., Johnson, R. W., & Kerk, C. J. (2003). Exploring relationships in body dimensions. *Journal of Statistics Education*, 11(2).

---

<sup>1</sup> En el siguiente curso (Análisis de Datos) aprenderemos técnicas para enfrentar problemas de clasificación con más de dos clases.