

2025

# Trabajo práctico

Curso de nivelación  
Análisis de Regresión  
Maestría en Estadística Aplicada



**Docentes:**

**Lic. María Belén Allasia**

**Lic. Noelia Castellana**

## Introducción

La **hipertensión arterial** es el aumento de la presión arterial de forma crónica. Es una enfermedad que no da síntomas durante mucho tiempo y, si no se trata, aumenta la probabilidad de sufrir un accidente cerebrovascular, ataque cardíaco, insuficiencia cardíaca, enfermedad renal y muerte temprana.

La **tensión arterial** o presión arterial se mide en milímetros de mercurio (mmHg) y tiene dos componentes:

- La *tensión arterial sistólica (TAS)*: es el número más alto y representa la tensión que genera el corazón cuando bombea la sangre al resto del cuerpo.
- La *tensión arterial diastólica (TAD)*: es el número más bajo y se refiere a la presión en los vasos sanguíneos entre los latidos del corazón.

La **hipertensión arterial** o tensión arterial alta se diagnostica cuando uno de estos números o ambos son altos: TAS > 140 mmHg y/o TAD > 90 mmHg.

Muchos **factores** pueden influenciar en el aumento de la tensión arterial. Dentro de ellos se pueden mencionar: una edad avanzada, el consumo elevado de alcohol, el estrés, el exceso de peso, el sedentarismo, fumar y una dieta inadecuada. Muchos de estos factores se pueden controlar con un estilo de vida saludable.

Dentro de este contexto, un grupo de médicos, que trabajan en un centro de salud de nuestra ciudad, desea evaluar la posible relación entre ciertas variables y la tensión arterial. Para ello, se realizó un **estudio observacional** que incluyó pacientes que concurrieron durante el año 2019 a dicho centro de salud. A estos pacientes se les tomó la tensión arterial y se registraron las siguientes características: edad, peso, si es fumador, si realiza actividad física, altura y colesterol total.

Los datos relevados se encuentran en el archivo Excel que compartimos a cada grupo.

## Estructura del archivo

En la primera fila del archivo se encuentran los nombres de las variables. A partir de la segunda fila, cada una de ellas corresponde a un paciente.

En las columnas se encuentran los datos de las características relevadas que se describen a continuación:

Columna	Descripción
<b>Id</b>	Identificación única del número de paciente.
<b>Edad</b>	Edad en años cumplidos del paciente.
<b>Act_fisica</b>	Indica si el paciente realiza actividad física frecuentemente (2 veces o más por semana por al menos 1 hora). Sus categorías son: Si (1), No (0).
<b>TAS</b>	Tensión arterial sistólica del paciente medida en mmHg.
<b>Peso</b>	Peso del paciente medido en kg.
<b>Fuma</b>	Indica si el paciente es fumador. Sus categorías son: Si, No.
<b>Colesterol</b>	Colesterol Total del paciente medido en mg/dl (miligramos por decilitro).
<b>IMC</b>	Índice de masa corporal. Se obtiene como $imc = \text{peso} / \text{altura}^2$
<b>altura</b>	Altura del paciente en metros

## Preguntas

### Parte 1

1.1 ¿Qué características presentan los pacientes estudiados?

Realizar un análisis descriptivo univariado de las variables relevadas. ¿Hay valores atípicos? ¿Hay valores perdidos?

1.2 ¿Qué relación tienen las variables consideradas explicativas con TAS? ¿Entre ellas presentan alguna relación?

1.3 En el caso de identificar valores atípicos tomar una resolución acerca de cómo van a considerarse. Pueden consultar a las docentes como referentes de los médicos.

### Parte 2

En una primera instancia se desea estudiar la relación entre el **peso** de un paciente y su **tensión arterial sistólica**. Para ello se decide ajustar un modelo de regresión lineal simple (**Modelo 1**).

2.1 Escribir la ecuación y los supuestos para el Modelo 1.

2.2 ¿Existe regresión? Plantear la hipótesis correspondiente, escribir la estadística de prueba y su distribución y concluir en términos del problema.

2.3 Escribir la ecuación estimada e interpretar los parámetros estimados.

2.4 ¿Qué pueden comentar acerca del valor del coeficiente de determinación?

2.5 Evaluar si se cumplen los supuestos del modelo. En caso de no cumplirse comentar si se identifica la causa y cómo procederían para solucionarlo.

### Parte 3

En una segunda instancia se desea estudiar la relación entre **todas las variables** relevadas y la **tensión arterial sistólica**. Para ello se decide ajustar un modelo de **regresión lineal múltiple sin interacción (Modelo 2)**.

- 3.1 Escribir la ecuación y los supuestos para el Modelo 2.
- 3.2 ¿Existe regresión? Plantear las hipótesis correspondientes, construir el cuadro ANOVA e interpretar en términos del problema.
- 3.3 Realizar los test parciales. ¿Qué sugieren estos test?
- 3.4 Evaluar la multicolinealidad.
- 3.5 Evaluar si se cumplen los supuestos del modelo.

## Parte 4

Considerando las conclusiones de los puntos anteriores y utilizando métodos de selección de variables se les propone construir el mejor modelo que ustedes consideren que sea adecuado.

Puntos a tener en cuenta:

- Incluir o no interacciones
- Evaluar la presencia de multicolinealidad
- Evaluar el cumplimiento de supuestos
- Detectar observaciones atípicas.

Luego del proceso iterativo que conlleva la construcción de un modelo, al modelo final resultante se lo denominará **Modelo 3**.

- 4.1 Escribir el Modelo 3 (modelo final)
- 4.2 Interpretar los coeficientes estimados
- 4.3 ¿Es posible, con el modelo 3, hacer un ranking de las variables consideradas en el modelo según su impacto en la tensión arterial sistólica?
- 4.4. Predecir los valores de TAS media (con el modelo 3) para aquellos pacientes que tienen *missing* en este campo.

## Parte 5

El grupo de investigadores realizó un informe analizando los resultados del Modelo 3 en donde se incluyeron las siguiente afirmaciones. No todas ellas son correctas. Identificar cuáles son ciertas y cuáles son falsas. Justificar.

- 5.1 La actividad física reduce el aumento de la TAS media, cuando se consideran otros factores de riesgo relevantes (variables del modelo 3)

5.2 El IMC aporta en forma significativa a explicar la variabilidad de la TAS cuando se consideran las demás variables (modelo 3).

5.3 A medida que el peso del paciente aumenta, la TAS media disminuye, mientras las demás variables del modelo permanecen constantes.

5.4 Los fumadores presentan valores medios de TAS superiores a los no fumadores.

5.5 Los pacientes que realizan actividad física frecuente no presentan diferencias en los valores de TAS media con los pacientes que no realizan actividad física, cuando se controlan las demás variables del modelo.