

# Métricas de Evaluación para modelos de regresión

## Coeficiente de determinación - R2

Es una métrica que permite medir qué porcentaje de la variabilidad de la variable Y es explicada por el modelo. Para calcularlo son necesarios:

$$SumaCuadradosTotal = \sum_{i=1}^n (Y_i - \bar{Y})^2 \longrightarrow \text{Suma de las diferencias cuadráticas entre la observación y la predicción base (promedio)}$$

$$SumaCuadradosResiduos = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \longrightarrow \text{Suma de las diferencias cuadráticas entre la observación y la predicción}$$

$$SumaCuadradosModelo = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 \longrightarrow \text{Suma de las diferencias cuadráticas entre la predicción y predicción base}$$

## R2

$$SCT = SCM + SCR *$$

$$0 \leq R^2 \leq 1$$

$$R^2 = \frac{\text{SumaCuadradosModelo}}{\text{SumaCuadradosTotal}} = 1 - \frac{\text{SumaCuadradosResiduos}}{\text{SumaCuadradosTotal}}$$

Ratio variabilidad explicada

Ratio variabilidad no explicada

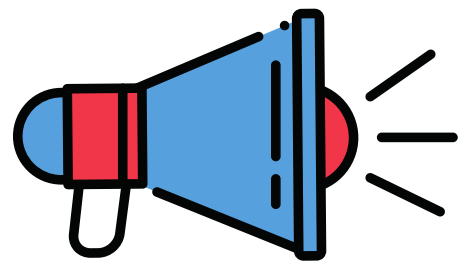
**Cuanto más cercano sea el valor del R2 a 1, mejor es el ajuste del modelo a los datos**

## R2 ajustado

$$R_A^2 = 1 - \underbrace{\frac{\text{SumaCuadradosResiduos}}{\text{SumaCuadradosTotal}}}_{\text{ratio variabilidad no explicada}} \cdot \underbrace{\frac{\text{NumObservaciones}-1}{\text{NumObservaciones}-\text{NumVariables}}}_{\text{"penalización"}}$$

Una limitación del R2 es que tiende a aumentar incluso cuando se agregan más variables predictoras al modelo, independientemente de si esas variables son relevantes o no. Para sobrellevar esta limitación podemos utilizar R2 ajustado, este penaliza el uso de variables predictoras adicionales que no mejoran significativamente la capacidad de predicción del modelo.

\* La nomenclatura puede variar



Cuanto más cercano sea a cero, mejor es el rendimiento del modelo, ya que significa que las predicciones se ajustan muy bien a los valores reales.

## Error Cuadrático Medio (Mean Squared Error)

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

Calcula la diferencia cuadrática media entre los valores estimados y los valores reales. Es decir, la media de los cuadrados de los residuos

## Raíz Error Cuadrático Medio (Root Mean Squared Error)

$$RMSE = \sqrt{MSE}$$

## Error Absoluto Medio (Mean Absolute Error)

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i|$$

Calcula la diferencia absoluta promedio entre las predicciones y los valores reales