

# Ciencia de datos

Para el sector público de salud

## Módulo 4

Sesión 6: Modelos de regresión



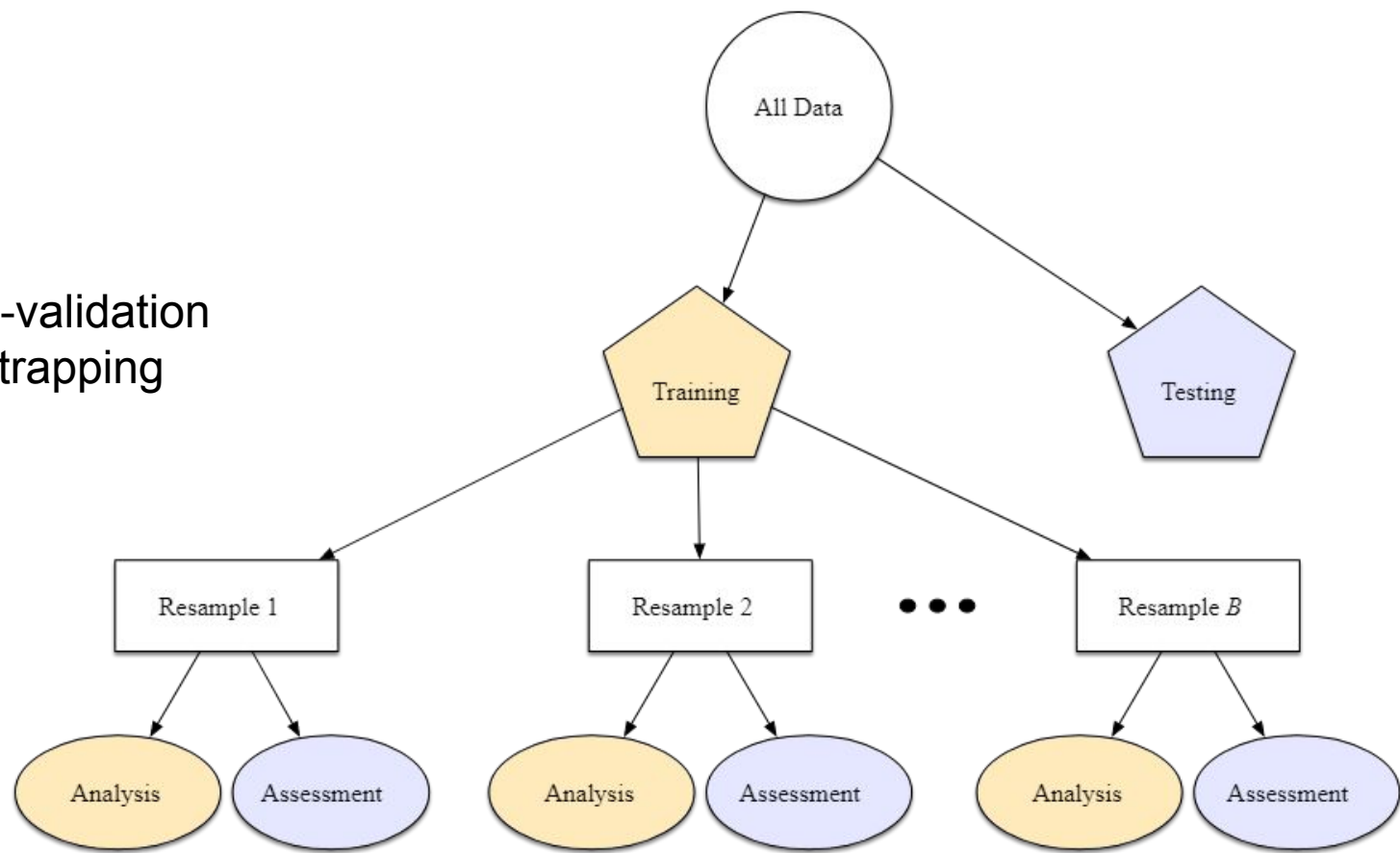
academia .opensaludlab.org

# opensaludlab.org

Twitter / Instagram / LinkedIn

# Resampleo

## Cross-validation Bootstrapping



# Métricas

## R<sup>2</sup> (coeficiente de determinación)

El coeficiente determina la calidad del modelo para replicar los resultados, y la proporción de variación de los resultados que puede explicarse por el modelo.

$$R^2 = 1 - \frac{SS_{RES}}{SS_{TOT}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}$$

# MAE (error absoluto medio)

Mide la magnitud promedio de los errores en un conjunto de predicciones, sin considerar su dirección.

El error absoluto medio es la distancia vertical promedio entre cada uno de los puntos y la recta identidad ( $y=x$ )

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$$

# RMSE (raíz del error cuadrático medio)

Compara un valor predicho y un valor observado o conocido.

RMSE cuantifica cuán diferente es un conjunto de valores. Cuanto más pequeño es un valor RMSE, más cercanos son los valores predichos y observados.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (P_i - O_i)^2}{n}}$$



# Machine Learning

## Aprendizaje supervisado

# CARET (Classification And REgression Training)



## **Modelos disponibles**

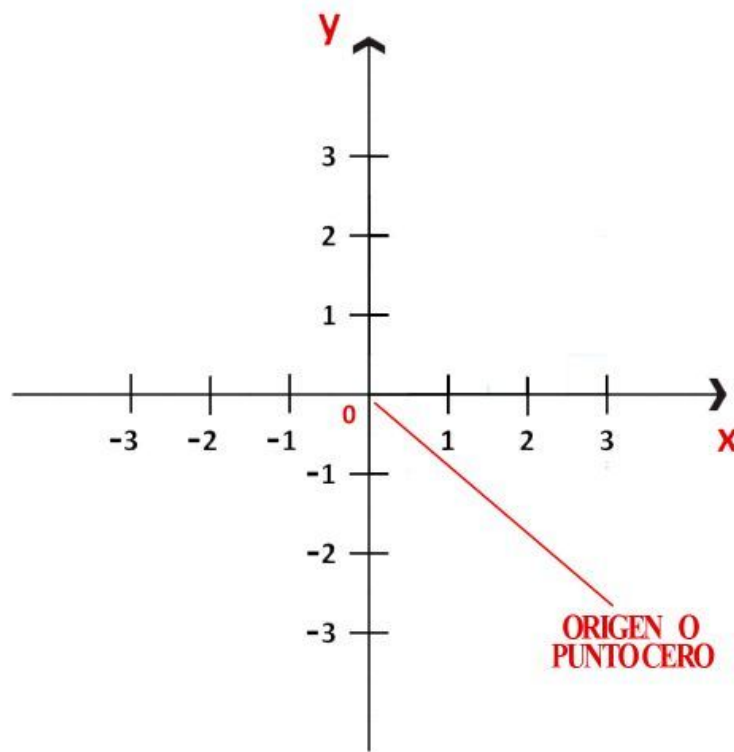
<https://topepo.github.io/caret/available-models.html>

## **Documentación oficial**

<https://topepo.github.io/caret/index.html>

# Modelos lineales

Se llama **plano cartesiano** al plano formado por 2 rectas numéricas perpendiculares, una horizontal y otra vertical, que se cortan en un punto llamado origen.



# Ecuación de una recta

Relación entre abscisas y ordenadas de todos los puntos pertenecientes a dicha recta.

$$y = mx + n$$

$m = pendiente$

$n = coeficiente\ de\ posición$

Veamos algo de código...