

**LAURA DANIELA DIAZ TORRES  
VIVIAN CABANZO FERNÁNDEZ  
ZENNETH OLIVERO TAPIAS  
CRISTIAN FELIPE MUÑOZ GUERRERO**

2025-02

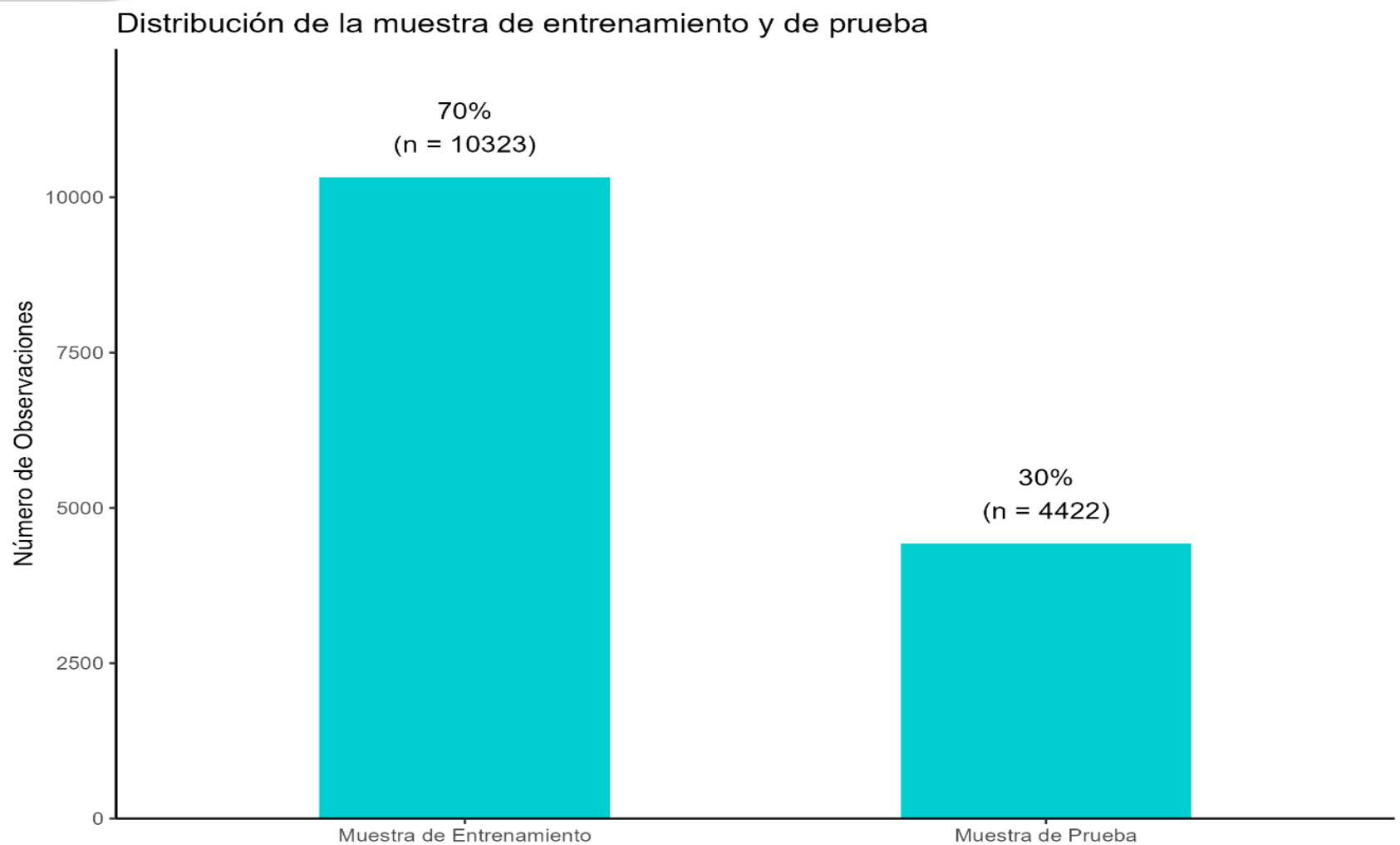
**BIG DATA Y MACHINE LEARNING PARA  
ECONOMIA APLICADA**

**Predicción**

## FORMULACIÓN DE MODELOS PARA VALIDACIÓN CRUZADA

Modelo	Ecuación
Modelo 1	$\ln(w) = \beta_1 + \beta_2 \cdot \text{age} + \beta_3 \cdot \text{age}^2 + u$
Modelo 2	$\ln(w) = \beta_1 + \beta_2 \cdot \text{sex} + u$
Modelo 3	$\ln(w) = \beta_1 + \beta_2 \cdot \text{sex} + \beta_3 \cdot \text{age} + \beta_4 \cdot \text{age}^2 + \beta_5 \cdot \text{educ} + \beta_6 \cdot \text{relab} + \beta_7 \cdot \text{formal} + \beta_8 \cdot \text{size\_firm} + u$
Modelo 4	$\ln(w) = \beta_1 + \beta_2 \cdot \text{age} + \beta_3 \cdot \text{sex} + \beta_4 \cdot (\text{age} \cdot \text{sex}) + \beta_5 \cdot \text{age}^2 + \beta_6 \cdot \text{educ} + u$
Modelo 5	$\ln(w) = \beta_1 + \beta_2 \cdot \text{sex} + \beta_3 \cdot \text{educ} + \beta_4 \cdot (\text{sex} \cdot \text{educ}) + \beta_5 \cdot \text{age} + u$
Modelo 6	$\ln(w) = \beta_1 + \beta_2 \cdot \text{age} + \beta_3 \cdot \text{relab} + u$
Modelo 7	$\ln(w) = \beta_1 + \beta_2 \cdot \text{age} + \beta_3 \cdot \text{formal} + \beta_4 \cdot \text{educ} + \beta_5 \cdot (\text{formal} \cdot \text{educ}) + u$
Modelo 8	$\ln(w) = \beta_1 + \beta_2 \cdot \text{educ} + \beta_3 \cdot \text{relab} + \beta_4 \cdot (\text{educ} \cdot \text{relab}) + \beta_5 \cdot \text{age} + u$

# DISTRIBUCIÓN DE MUESTRA PARA VSA



## RESULTADOS RMSE POR MÉTODOS DE VALIDACIÓN CRUZADA PARA SELECCIÓN

Modelo	RMSE (VSA)	RMSE (K-FOLD)
Modelo 1	0.8116	0.8104
Modelo 2	0.8180	0.8187
Modelo 3	0.6311	0.6373
Modelo 4	0.6857	0.6876
Modelo 5	0.6940	0.6956
Modelo 6	0.7780	0.7793
Modelo 7	0.6506	0.6563
Modelo 8	0.6704	0.6762

**Selección de modelo 3 y 7 para análisis por método de LOOCV, se eligen los modelos con menor RMSE, lo que implica mejor poder de predicción.**

Modelo	RMSE (VSA)	RMSE (K-FOLD)	RMSE (LOOCV)
Modelo 3	0.6311	0.6373	0.6376
Modelo 7	0.6506	0.6564	0.6566

# PREDICCIÓN DEL SALARIO POR HORA

Comparación del RMSE entre los Modelos 3 y 7

