

## Machine Learning para Inferencia Causal con Stata

### **MACHINE LEARNING PARA INFERENCIA CAUSAL CON STATA (microcredencial)**

**Eje estratégico:** Formación en competencias para la investigación

**Modalidad de formación:** Presencial

**Colectivo al que se dirige:** Personal docente e investigador

**Plazas ofertadas:** 40

La realización de esta acción formativa está condicionada a que haya un mínimo de 10 participantes.

**Requisitos previos:** Conocimientos básicos de aprendizaje automático (regresión regularizada, árboles de decisión) y de inferencia causal (modelo de outcomes potenciales, variables de control). Familiaridad con Stata.

**Es necesario que las personas participantes lleven sus equipos personales.**

**Duración de la formación:** 10 horas

**Calendario y horario:** 3 y 4 de julio de 2025, de 9:00 a 14:00 horas.

**Lugar:** Edificio 14 Aula de informática 1

### **Objetivos de la acción formativa:**

Este curso tiene como finalidad introducir a investigadores/as y profesionales en el uso de técnicas de aprendizaje automático aplicadas a la inferencia causal. Aunque muchos métodos de Machine Learning fueron diseñados para mejorar la predicción, este curso pone el foco en cómo adaptar dichas herramientas para identificar efectos causales, controlando por alta dimensionalidad y selección de variables. Se abordarán marcos como Post-Lasso y Double Machine Learning, y su implementación práctica con Stata.

### **Programa de contenidos:**

Estructura del Curso

1. Fundamentos del Machine Learning para la Inferencia

- Diferencias conceptuales: predicción vs inferencia causal

- El rol de la teoría y la interpretabilidad

- Sesgos por regularización y sobreajuste

- Evaluación del efecto causal vs ajuste predictivo

2. Selección de Variables con LASSO y Post-LASSO

- Principios de regularización y sparsity
- Estimación de modelos con alto número de controles
- Inference post-selección: validez y limitaciones
- Implementación en Stata

### 3. Introducción al Enfoque de Double Machine Learning (DML)

- Motivación: control flexible de covariables
- Framework teórico de DML
- División de muestra (sample splitting)
- Cross-fitting y estimación ortogonal

### 4. Estimación de Efectos Medios del Tratamiento con Ajuste Automático

- Ejercicios con Stata: estimación del ATT con ``ddml``, ``drdid`` y control de sesgo de selección
- Evaluación de robustez y validación cruzada en estimaciones causales
- Métodos híbridos: regresión ajustada con predictores seleccionados automáticamente
- Uso de Machine Learning para estimar funciones de resultado condicional y funciones de propensión.
- Definición e interpretación del efecto medio del tratamiento (ATE) y del efecto medio del tratamiento en los tratados (ATT)

### 5. Implementación Práctica en Stata

- Ejemplos con variables continuas y binarias
- Interpretación de resultados: efectos medios parciales (Partial Effects)
- Comparación con estimaciones OLS tradicionales.

### 6. Aplicaciones y Extensiones

- Estimación de efectos heterogéneos del tratamiento con ML
- Tratamientos múltiples y selección basada en covariables
- Integración con métodos clásicos (IPW, regresión ajustada)
- Casos prácticos con bases de datos reales

#### **Metodología de evaluación:**

Explicaciones teóricas guiadas por ejemplos empíricos. Aplicaciones prácticas con Stata y ejercicios reproducibles. Discusión de casos reales de investigación.

#### **Resultados de aprendizaje:**

1. Ser capaces de distinguir un problema de aprendizaje supervisado frente a uno no supervisado.
2. Ser capaz de elegir la mejor técnica del problema de aprendizaje dirigido a la predicción dependiendo del problema y de la naturaleza de los datos y aplicarla en un entorno de programación real.

3. Ser capaz de elegir la mejor técnica del problema de aprendizaje supervisado dirigido a la inferencia estadística dependiendo del problema y de la naturaleza de los datos y aplicarla en un entorno de programación real.

4. Ser capaz de evaluar y comparar los resultados obtenidos.

**Criterios para acreditar la formación realizada:** Superación de la evaluación.

**Plazo de presentación de solicitudes:** Hasta el día **27 de junio de 2025**.

**Persona formadora:** D.<sup>a</sup> Yolanda F. Rebollo Sanz. Profesora Titular, Universidad Pablo de Olavide. Experta en evaluación de políticas públicas, microeconometría avanzada e inferencia causal. Ha impartido formación en métodos cuantitativos en diversos programas de posgrado y ha publicado más de 20 artículos científicos en revistas indexadas.