

# Análisis Exploratorio del Gasto de los Hogares en Educación (EGHE 2019)

true true true

9 de Noviembre de 2025

## Abstract

El presente proyecto realiza un Análisis Exploratorio de Datos (AED) sobre la Encuesta de Gasto de los Hogares en Educación (EGHE) del año 2019, publicada por el Instituto Nacional de Estadística (INE). El objetivo es entender la estructura del gasto en educación en España, identificando sus principales componentes y las características de los hogares y estudiantes asociados a dicho gasto. Este informe detalla el proceso de importación, limpieza y transformación de los microdatos, seguido de un análisis univariante y bivalente para responder a las preguntas de investigación planteadas.

## Contents

<b>1. Introducción</b>	<b>1</b>
<b>2. Carga y Preparación de Datos</b>	<b>2</b>
<b>3. Descripción del Conjunto de Datos Limpio</b>	<b>2</b>

## 1. Introducción

La educación es ampliamente reconocida como uno de los pilares fundamentales para el desarrollo económico y social de un país. Sin embargo, el acceso y la permanencia en el sistema educativo conllevan una serie de costes directos e indirectos que son asumidos en gran medida por los hogares. Este gasto no se limita únicamente a las tasas y matrículas, sino que abarca un complejo ecosistema de bienes y servicios que incluye desde libros de texto y material escolar hasta servicios de comedor, transporte, actividades extraescolares y clases particulares.

Comprender la magnitud y la composición de este gasto es esencial para evaluar la equidad del sistema educativo y el esfuerzo económico que supone para las familias. El Instituto Nacional de Estadística (INE) proporciona una visión detallada de esta realidad a través de la Encuesta de Gasto de los Hogares en Educación (EGHE).

El presente proyecto tiene como objetivo realizar un Análisis Exploratorio de Datos (AED) sobre los microdatos de la EGHE 2019. Nuestro propósito es diseccionar la estructura del gasto en educación en España, identificando los patrones subyacentes, las diferencias entre distintos tipos de enseñanza y las características de los estudiantes que más gasto generan.

Para guiar nuestro análisis, planteamos las siguientes preguntas de investigación:

1. **¿Cómo se distribuye el gasto educativo entre los diferentes tipos de centros (públicos, concertados, privados) y qué factores explican estas diferencias?** Compararemos los patrones de gasto entre los tres regímenes de financiación, analizando no solo las diferencias en el gasto total sino también en su composición (matrículas, servicios complementarios, material).
2. **¿Qué características del hogar y del estudiante predicen un mayor gasto educativo? ¿Existen diferencias significativas por sexo, composición familiar o tamaño del municipio?** Identificaremos los principales predictores del gasto educativo mediante el análisis de variables demográficas y estructurales. Evaluaremos específicamente si existen brechas de gasto asociadas al sexo del estudiante y el efecto de la composición del hogar.
3. **¿Cómo evoluciona la estructura del gasto a lo largo del itinerario educativo, desde Infantil hasta la Universidad, y qué servicios adquieren mayor relevancia en cada etapa?** Compararemos el gasto medio en diferentes niveles, como Educación Infantil, Primaria, ESO, Bachillerato y Universidad.
4. **¿Qué peso real tienen los gastos complementarios (clases particulares, actividades extraescolares, material específico) en el presupuesto total y qué familias los asumen en mayor medida?** Cuantificaremos el impacto de partidas específicas como clases particulares, servicios de comedor, actividades extraescolares y uniformes escolares, que frecuentemente pasan desapercibidos en el debate público sobre el coste de la educación.

Este informe detallará el proceso de limpieza y transformación de los datos, seguido del análisis univariante y bivalente para dar respuesta a estas preguntas, concluyendo con los principales hallazgos de nuestro estudio.

## 2. Carga y Preparación de Datos

Este *chunk* es el “motor” del informe. Carga los datos crudos una vez y ejecuta los 3 scripts de limpieza en orden sobre una copia.

## 3. Descripción del Conjunto de Datos Limpio

Tras la ejecución de los scripts de limpieza, obtenemos el *dataset* `dataset_limpio`. Este conjunto de datos contiene *4903 observaciones* (cada una representando a un estudiante que reportó gastos) y *80 variables* limpias y listas para el análisis.