

## Exploración y visualización de datos para lo socioeconómico

Miguel Andrés Garzón Ramírez

16 de octubre de 2025

### Proyecto 1: EDA, visualización y estadística para respaldar hipótesis

El propósito de este proyecto es consolidar lo aprendido en la primera parte del curso, integrando el análisis exploratorio de datos (EDA) con la formulación de una prueba de hipótesis que respalde un hecho estilizado. El proyecto busca fortalecer tanto la coherencia conceptual del análisis como la argumentación basada en evidencia. Para esto, vamos a retomar lo desarrollado en la Actividad 2 y profundizaremos en la generación de evidencia sobre los hallazgos encontrados inicialmente, o su verificación y ajuste.

#### Condiciones

- Se debe trabajar en los equipos de la Actividad 2, con los mismos datos.
- Puede usar cualquier herramienta de su preferencia.
- Se debe entregar:
  - Código o proyecto (Stata, R, Python) con pasos reproducibles. Se debe observar un proceso de inicio a fin.
  - Documento escrito donde se dé respuesta a los requerimientos de cada una de las partes y pasos del enunciado. Si se utiliza un cuaderno (Jupyter Notebook o Markdown), asegurarse de que la lectura sea sencilla y se reporte solo lo necesario de acuerdo con lo solicitado.
- Fecha de entrega: sábado 1 de noviembre

#### Situación (de la Actividad 2)

Su grupo de trabajo hace parte de un equipo técnico que apoya a una alcaldía en un programa de mejora de la calidad de vida de hogares venezolanos. La alcaldía quiere mejorar la **focalización de su respuesta en vivienda y servicios básicos**, pero no sabe qué factores son los más importantes para identificar condiciones de vida deficientes.

#### Objetivo analítico

A partir del análisis exploratorio previamente realizado (EDA y visualización ajustado), el objetivo ahora es formular y poner a prueba una hipótesis estadística que respalde un hecho estilizado del contexto estudiado. Deberá identificar una relación o diferencia entre variables observadas en los datos que pueda expresarse mediante una hipótesis estadística contrastable. Esto es, una frase concreta, respaldada con evidencia.

#### Parte 1: EDA y visualización ajustada

Para realizar una prueba de hipótesis se debe realizar un EDA alrededor de una pregunta concreta bien definida. Para empezar este proyecto se debe revisar lo entregado en la Actividad 2 y corregir/ajustar los aspectos señalados en la retroalimentación recibida.

Estas correcciones deben reflejar mejoras (si son pertinentes) en:

- La formulación del problema de indagación y la claridad del contexto analítico.
- La definición de las variables utilizadas y la justificación de su selección.
- La consistencia entre las visualizaciones y la respuesta a la pregunta planteada.
- Precisión y pertinencia de la respuesta planteada de acuerdo con la situación de contexto.
- La coherencia narrativa y argumentativa en las conclusiones.

El nuevo documento deberá incluir una breve descripción (máximo media página) de los principales ajustes realizados a partir de la retroalimentación recibida.

### Parte 2: Evidencia empírica y estimación

1. Describa el diseño del contraste estadístico
  1. Seleccione la prueba de acuerdo con la situación (por ejemplo, prueba  $t$  de diferencia de medias, prueba de proporciones, etc.)
  2. Explique brevemente qué se está comparando y cuál sería la expectativa teórica o empírica según el hecho estilizado que se desea comprobar.
  3. Señale si el contraste es bilateral o unilateral y cuál sería la dirección esperada del efecto. Esto en línea con lo encontrado en el EDA
2. Revise los supuestos y ejecute la prueba
  1. Verifique las condiciones necesarias para aplicar la prueba elegida (normalidad, independencia, homogeneidad de varianzas, tamaño de muestra). En caso de que alguna condición no se cumpla, justifique la prueba alternativa empleada o las transformaciones realizadas sobre los datos.
  2. Ejecute la prueba y reporte los resultados relevantes: estadístico de prueba, el intervalo de confianza estimado.
  3. Reporte el  $p$ -valor de la prueba de hipótesis y explique qué significa en términos del riesgo de error tipo I.

### Parte 3: Interpretación y comunicación de resultados

1. Interpretación:
  1. Verifique si el resultado de la prueba y el intervalo conducen a la misma conclusión.
  2. Interprete el resultado en relación con la hipótesis nula y con el contexto del análisis
  3. Explique qué implican los resultados en términos sustantivos: ¿qué nos dicen sobre el fenómeno estudiado?
2. Comunicación: Acompañe el resultado del contraste con una o más visualizaciones que permitan interpretar la diferencia o relación observada (por ejemplo, barras con intervalos de confianza, densidades comparadas, gráficos de dispersión con tendencia, etc.) Si ya los ya creados en el EDA son oportunos, complémtelos e intégralos a la narrativa que permite realizar la prueba de hipótesis. La narrativa es un breve texto de apoyo que explique lo que muestra el gráfico y cómo respalda (o no) la hipótesis planteada.