

## Exploración y visualización de datos para lo socioeconómico

Miguel Andrés Garzón Ramírez

19 de octubre de 2025

### Actividad práctica 3: Pruebas de hipótesis y comparación de grupos

#### Situación

Usted forma parte de un equipo de análisis económico encargado de identificar patrones globales relacionados con el desarrollo sostenible. El equipo quiere analizar si el nivel de ingreso de los países se asocia con mayores emisiones de CO<sub>2</sub> per cápita, y para ello propone usar datos del Banco Mundial (World Development Indicators) y aplicar pruebas de hipótesis que permitan evaluar si las diferencias observadas son estadísticamente significativas o producto del azar muestral.

El propósito del ejercicio es vincular lo aprendido sobre intervalos de confianza con el razonamiento y la interpretación de una prueba de hipótesis, explorando cómo los errores tipo I y tipo II influyen en la toma de decisiones basadas en evidencia.

#### Condiciones

- Se debe trabajar en los equipos definidos en la clase 6. Los grupos que entregaron la actividad bono de la clase 6 pueden aprovechar lo avanzado en esta actividad.
- Descargue los datos más recientes del sitio web de World Development Indicators sobre el PIB per cápita y las emisiones de CO<sub>2</sub> per cápita.
- Puede usar cualquier herramienta de su preferencia: Excel, Stata, R, Python, Power BI u otro.
- Se debe entregar:
  - Código o proyecto (Stata, R, Python, Excel/Power Query o Power BI) 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.
- Fecha de entrega: domingo 17 de octubre

#### Objetivo analítico

Comparar las emisiones de CO<sub>2</sub> per cápita entre países con PIB per cápita alto y bajo, estimar la diferencia promedio, su error estándar y un intervalo de confianza del 95 %, para posteriormente evaluar si la diferencia observada es estadísticamente significativa mediante una prueba de hipótesis bilateral o unilateral, según la justificación teórica.

#### Parte 1: Planteamiento conceptual y formulación de hipótesis

1. **Contextualice el problema.** Explique brevemente por qué es plausible pensar que el nivel de ingreso de los países podría asociarse con sus emisiones per cápita.
  - ¿Qué teorías o intuiciones respaldan esta relación?
  - ¿Qué razones podrían llevar a que no exista una diferencia real?

2. **Formule el contraste de hipótesis:** Defina la hipótesis nula ( $H_0$ ) y la alternativa ( $H_1$ ). Justifique si el contraste debe ser bilateral (interesa cualquier diferencia) o unilateral (interesa una dirección específica: mayor contaminación en países ricos).
3. **Especifique cuál sería la consecuencia de cometer** un error tipo I (falso positivo) o un error tipo II (falso negativo), en este contexto.
4. **Defina los indicadores:**
  - Variable de comparación (medición - Y): emisiones de  $\text{CO}_2$  per cápita.
  - Variable de clasificación (X): PIB per cápita, dividiendo países en dos grupos según su criterio. Explique por qué usa ese criterio.
  - Explique qué representa esta forma de agrupar y qué ventajas o limitaciones tiene.

## Parte 2: Evidencia empírica y estimación

1. Describa brevemente los datos:
  1. Número de países, año de referencia, valores mínimos y máximos de PIB y  $\text{CO}_2$  per cápita. Esto es el análisis univariado que permite conocer las particularidades de los datos con los que va a trabajar.
  2. Estadísticas descriptivas de cada grupo.
2. Estime la diferencia promedio de emisiones entre los grupos:
  1. Interprete el sentido y magnitud de la diferencia (¿cuánto más emiten los países de PIB alto?).
  2. Estime un intervalo de confianza del 95 % para esa diferencia usando un procedimiento Bootstrap y una prueba t. ¿Qué diferencia hay entre los resultados de Bootstrap y la prueba t?
3. Evalúe la evidencia:
  1. Determine si el intervalo de confianza incluye o no el valor 0.
  2. Reporte el p-valor de la prueba de hipótesis y explique qué significa en términos del riesgo de error tipo I.
  3. Verifique si el resultado de la prueba y el intervalo conducen a la misma conclusión.

## Parte 3: Interpretación y reflexión sobre la decisión estadística

1. Analice conceptualmente la decisión:
  - Si el resultado fue significativo, ¿qué implicaría en términos de política ambiental o desarrollo sostenible?
  - Si no fue significativo, ¿significa que no existe relación o que la evidencia no es suficiente?
2. Reflexione sobre los errores posibles:
  - Describa qué sería un falso positivo y un falso negativo en este estudio, y cuáles consecuencias tendría cada uno.
  - ¿Qué factores metodológicos (tamaño de muestra, variabilidad, agrupación) influyen en la probabilidad de cometer estos errores?
3. Concluya sobre la robustez de la evidencia:

- ¿Cómo podría mejorarse el análisis (más datos, otras variables, nuevos años, ponderaciones)?
- ¿Qué limitaciones debemos reconocer al generalizar estos resultados?

El informe debe incluir:

1. Planteamiento conceptual: hipótesis, justificación teórica y discusión sobre los errores tipo I y II.
2. Resultados empíricos: tabla de medias, diferencia e intervalo de confianza
3. Reflexión final: interpretación sustantiva de los resultados y aprendizajes sobre la toma de decisiones estadísticas.