**Exploración y visualización de datos para lo socioeconómico**

Miguel Andrés Garzón

|  |
| --- |
| Taller 3: **Pruebas de hipótesis y comparación de grupos[[1]](#footnote-1)** |

**ELABORADO POR:**

Laura Sarif Rivera Sanabria

María Camila Caraballo Candela

Javier Antonio Amaya Nieto

# Enunciado del taller

**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 CO₂ 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.

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

**Contextualización del problema**

El impulso global hacia un mayor crecimiento económico ha estado históricamente acompañado por una intensificación en el uso de los recursos naturales, particularmente como insumos energéticos y productivos para la industria. Este patrón ha derivado en un incremento sostenido de las emisiones contaminantes, en especial de dióxido de carbono (CO₂), principal responsable del calentamiento global. La evidencia muestra que la mayor parte de estas emisiones proviene del consumo energético basado en combustibles fósiles, fuente predominante en la generación de energía a nivel mundial. En 2023, se estimó que cerca del 80 % de las emisiones históricas acumuladas de CO₂ derivadas de combustibles fósiles correspondían a los países del G20, siendo China, Estados Unidos y la Unión Europea los principales contribuyentes a dicha proporción (United Nations Environment Programme, 2023).

En el marco de las conferencias y debates de alto nivel promovidos por diversos organismos multilaterales que han situado el cambio climático en el centro de la agenda global, la investigación académica ha buscado comprender la relación entre las emisiones de carbono y el crecimiento económico. Este interés no se limita a establecer una mera correlación, sino a identificar y analizar los mecanismos a través de los cuales dicha relación se manifiesta (Monasterolo et al., 2019). En esta línea, una parte importante de la literatura se ha concentrado en el estudio de la Curva Ambiental de Kuznets (Environmental Kuznets Curve, EKC), la cual propone un vínculo entre el crecimiento económico y el deterioro ambiental. Esta formulación se deriva de la curva original de Kuznets (1955), que examinaba la relación entre el ingreso per cápita y la desigualdad en el proceso de desarrollo económico.

La adaptación ambiental de esta curva plantea que, a medida que el ingreso per cápita aumenta, las emisiones de carbono también se incrementan hasta alcanzar un punto de inflexión, tras el cual comienzan a disminuir, configurando una relación de tipo cóncavo. En otras palabras, las economías en etapas tempranas de desarrollo tienden a depender de actividades intensivas en energía y recursos, mientras que las economías más avanzadas logran desacoplar el crecimiento económico de la contaminación gracias a la innovación tecnológica, la eficiencia energética y una mayor regulación ambiental. No obstante, diversos estudios han señalado que las regulaciones específicas y las características institucionales de cada país introducen problemas de endogeneidad y variables omitidas, que limitan la capacidad para identificar una relación causal robusta (Mitic et al, 2023)

Desde otra perspectiva, algunos análisis vinculan el nivel de ingreso con el consumo energético, señalando que este se explica principalmente por el uso de combustibles fósiles. Aunque la justificación causal de este vínculo no siempre es concluyente, dichos estudios ofrecen indicios sobre los canales a través de los cuales el crecimiento económico afecta las emisiones. De forma complementaria, investigaciones más recientes han incorporado variables adicionales como la apertura comercial, la urbanización, el consumo de energías renovables e incluso la inversión extranjera directa, con el fin de capturar la complejidad del fenómeno (Kasman & Selman, 2015).

Sin embargo, como se ha señalado, aunque existe una literatura amplia que explora diversas variables capaces de explicar la relación entre ingreso y emisiones, persiste el interés en determinar si las diferencias observadas entre países son estadísticamente significativas. En esta línea, Ritchie (2023) presenta un análisis interactivo que relaciona los promedios de ingreso per cápita y emisiones de CO₂ por país, mostrando tendencias empíricas claras, pero sin una validación estadística formal.

Esto sugiere la necesidad de realizar un ejercicio que evalúe si las diferencias entre grupos de ingreso y niveles de emisiones son producto del azar o de un patrón sistemático. Tal aproximación permite además considerar factores estructurales que pueden distorsionar la relación, como el hecho de que no todos los países de altos ingresos presentan la misma composición productiva, o que algunos han adoptado en la última década fuentes de energía renovable que revierten parcialmente esta tendencia. Asimismo, las regulaciones ambientales más estrictas, la heterogeneidad institucional y los cambios estructurales hacia economías basadas en servicios también influyen en el comportamiento de las emisiones (Dong et al., 2018).

Por estas razones, este estudio propone estimar si existen diferencias estadísticamente significativas en las emisiones de CO₂ per cápita según el nivel de ingreso de los países, contribuyendo así a la comprensión empírica del vínculo entre desarrollo económico y sostenibilidad ambiental.

**Estrategia empírica**

Para responder a la pregunta de investigación se plantea una hipótesis unilateral, ya que la literatura teórica y empírica sugiere una dirección esperada en la relación entre el crecimiento económico, medido a través del ingreso per cápita, y las emisiones de CO₂ per cápita, de acuerdo con la teoría de Kuznets. La evidencia indica que las economías más desarrolladas, al alcanzar niveles más altos de ingreso, tienden también a registrar mayores emisiones por persona, especialmente cuando su matriz energética depende de los combustibles fósiles y su estructura productiva conserva sectores industriales con un alto consumo de energía.

En este contexto, una prueba bilateral no resulta adecuada, pues implicaría evaluar si las diferencias entre grupos podrían presentarse en cualquier dirección. Dado que la literatura anticipa un vínculo positivo entre el ingreso y las emisiones, el uso de un contraste unilateral se considera más consistente y metodológicamente coherente con los fundamentos teóricos y la evidencia empírica que sustentan este análisis.

Es así como para operacionalizar la pregunta, se deben clasificar los países en grupos discretos como ‘altos ingresos’ y ‘bajos ingresos’ y el parámetro de interés para esta comparación sería la mediana, considerando el sesgo que tiene la distribución de esta variable.

Con base en este objetivo, se formulan las siguientes hipótesis:

Donde:

* representa la media de emisiones de CO₂ per cápita en los países de altos ingresos
* representa la media correspondiente a los países de bajos ingresos.

Siendo la hipótesis nula la representación de un ‘no efecto’ o ‘no diferencia’ pues es bajo la suposición de que esto se mantiene a menos de que exista evidencia estadística que sea lo suficientemente significativa para rechazarla. Es así como la hipótesis nula en este caso establece que la mediana de las emisiones per cápita de los países con ingreso alto es menor o igual a la del grupo con ingresos bajos, determinando que la riqueza no conduce a mayores emisiones.

Por su parte, la hipótesis alternativa representa la afirmación que como investigadores buscamos respaldar con evidencia estadística, siendo en este caso la afirmación de que los países de ingreso alto tienen en efecto una mediana de emisiones de CO₂ per cápita significativamente mayor que la de los países de ingreso bajo.

Ahora bien, en el contexto de este análisis, es relevante considerar las implicaciones de cometer errores estadísticos al momento de interpretar los resultados. En primer lugar, un error tipo I, o falso positivo, ocurriría si se concluye que existen diferencias significativas en las emisiones de CO₂ per cápita entre países según su nivel de ingreso, cuando en realidad dichas diferencias no existen. En términos prácticos, esto supondría atribuir al nivel de ingreso una diferencia dado el comportamiento de las emisiones sin que haya evidencia que lo sustente. Una conclusión de este tipo podría conducir a interpretaciones erróneas sobre la relación entre crecimiento económico y contaminación, y en consecuencia, a la formulación de políticas públicas inadecuadas en materia de sostenibilidad ambiental.

Por otra parte, un error tipo II, o falso negativo, se presentaría al no reconocer una diferencia que en realidad sí existe. En este caso, se concluiría que los países presentan niveles similares de emisiones pese a sus diferencias de ingreso, cuando en realidad las economías más prósperas generan, en promedio, mayores emisiones per cápita. Este tipo de error también tendría consecuencias relevantes, pues llevaría a subestimar la influencia del desarrollo económico sobre la condición ambiental y a pasar por alto la necesidad de políticas más estrictas en las economías de mayores ingresos. En ambos casos, la incorrecta interpretación de los resultados afectaría la validez del análisis y la pertinencia de las recomendaciones derivadas del estudio.

**Respuesta:**

1. **Defina los indicadores:**

* Variable de comparación (medición - Y): emisiones de CO₂ per cápita.

**Respuesta:**

* Variable de clasificación (X): PIB per cápita, dividiendo países en dos grupos según su criterio. Explique por qué usa ese criterio.

**Respuesta**: Usamos la mediana del PIB per cápita para clasificar los países en dos grupos (alto/bajo ingreso). Este umbral que usa la mediana es más robusto y estable, porque produce grupos de tamaño comparable y reduce la influencia desproporcionada de economías extremadamente ricas (valores extremos), ubicando de una mejor forma la mitad de la distribución.

* Explique qué representa esta forma de agrupar y qué ventajas o limitaciones tiene.

**Respuesta:** Dividir los países en “alto” y “bajo” ingreso usando la mediana del PIB per cápita es una forma sencilla y justa de partir la muestra en dos mitades comparables. No intenta copiar ni reproducir la clasificación del Banco Mundial; más bien, busca un punto de corte claro y replicable para contrastar el comportamiento de las emisiones de CO₂ per cápita en países con un mayor nivel de desarrollo, medido por su PIB per cápita, frente a aquellos con menores ingresos. Una de las ventajas de este enfoque es que equilibra los grupos y funciona mejor como medida de tendencia central para una variable que no sigue una distribución normal, además de no se afecta tanto por casos extremos (economías muy ricas), por lo que las comparaciones suelen ser más estables y fáciles de explicar. Por otro lado, esta clasificación no está claramente alineada con la versión oficial desarrollada por el Banco Mundial, y al dividir solo en “altos” y “bajos” ingresos puede generar grupos en los que coexistan economías muy distintas.

**Parte 2: Evidencia empírica y estimación**

1. **Describa brevemente los datos:**

* Número de países, año de referencia, valores mínimos y máximos de PIB y CO₂ per cápita. Esto es el análisis univariado que permite conocer las particularidades de los datos con los que va a trabajar.
* Estadísticas descriptivas de cada grupo.

**Respuesta:** Con el propósito de explorar la relación entre el nivel de ingreso y las emisiones de CO₂, se presenta un análisis univariado entre el PIB per cápita y el CO₂ per cápita con base en los datos del Banco Mundial se dispone de información correspondiente al año 2023 sobre el PIB per cápita y las emisiones de CO₂ per cápita. En total, se registran 239 países, con un PIB per cápita mínimo de 192,07 y un máximo de 132.604. En cuanto a las emisiones, el CO₂ per cápita varía entre 0 y 81,21 toneladas, como se muestra en la Tabla 1.

Por otra parte, en la Tabla 2 presenta las estadísticas descriptivas diferenciadas por grupo de PIB per cápita (alto y bajo). El grupo de alto PIB per cápita incluye 120 países, con un promedio de 32.648, un mínimo de 7.195 y un máximo de 132.604. Las emisiones promedio de CO₂ en este grupo son de 7,43 toneladas, con un rango entre 0 y 81,21 toneladas. No obstante, el grupo de bajo PIB per cápita está compuesto por 119 países, con un PIB promedio de 2.917, un mínimo de 192 y un máximo de 7.133. Las emisiones de CO₂ en este grupo presentan un valor mínimo de 0 y un máximo de 8,60 toneladas.

En general, como se observa en la figura 1, se evidencia una correlación positiva: los países con mayor PIB per cápita tienden a registrar niveles más altos de emisiones, aunque con una dispersión considerable entre grupos. Este patrón sugiere que el crecimiento económico se asocia, en promedio, con mayores niveles de emisiones, especialmente en economías desarrolladas.

**Texto

El contenido generado por IA puede ser incorrecto.**

**Tabla 1.** Estadísticas descriptivas de análisis univariado. Fuente: Banco Mundial. Elaboración propia.

Tabla

El contenido generado por IA puede ser incorrecto.

**Tabla 2.** Estadísticas descriptivas por grupo. Fuente: Banco Mundial. Elaboración propia.

Gráfico, Gráfico de dispersión

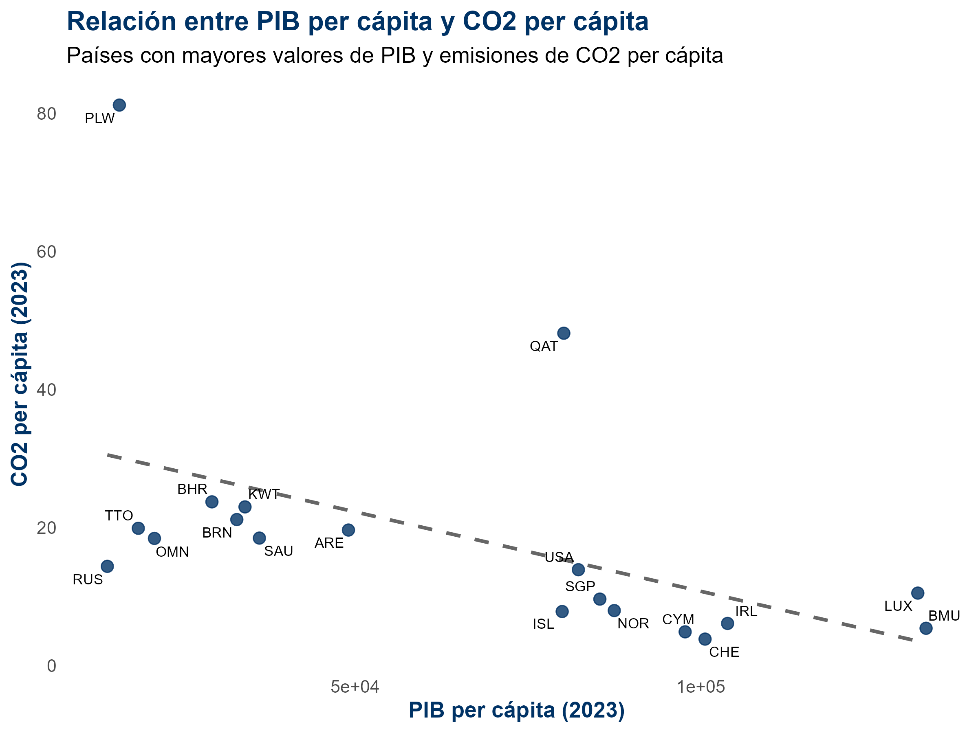
El contenido generado por IA puede ser incorrecto.

**Figura 1.** Relación entre PIB per cápita y emisiones de CO2 per cápita. Fuente: Banco Mundial. Elaboración propia.

Gráfico

El contenido generado por IA puede ser incorrecto.

**Figura 2.** Top 10 de países con mayores emisiones de CO2 per cápita. Fuente: Banco Mundial. Elaboración propia.



**Figura 3.** Top 10 de países- Relación entre PIB per cápita y de CO2 per cápita. Fuente: Banco Mundial. Elaboración propia.

1. **Estime la diferencia promedio de emisiones entre los grupos:**

* Interprete el sentido y magnitud de la diferencia (¿cuánto más emiten los países de PIB alto?).
* 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?

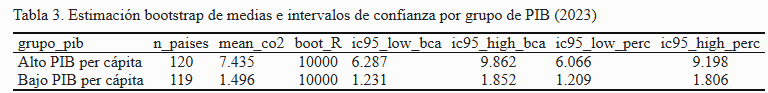
**Respuesta:** Los resultados del análisis muestran una diferencia promedio de 5,93 toneladas de CO₂ per cápita entre los países de alto PIB per cápita y los de bajo PIB per cápita, indicando que, en promedio, los países más ricos emiten casi seis toneladas más de CO₂ por persona que los países con menor ingreso.

El intervalo de confianza del 95%, estimado mediante el método bootstrap, oscila entre 4,75 y 8,46 toneladas (IC95% BCa), lo que sugiere que esta diferencia es estadísticamente significativa y robusta. De manera consistente, la prueba t clásica confirma el resultado, con un intervalo de confianza muy similar (4,27 a 7,61 toneladas) y un p-valor igual a 0.

La estimación bootstrap de las medias e intervalos de confianza para las emisiones de CO₂ per cápita en 2023 muestra diferencias marcadas entre los grupos de países según su nivel de ingreso. En promedio, los países con alto PIB per cápita registran 7,4 toneladas de CO₂ por persona, mientras que aquellos con bajo PIB per cápita apenas alcanzan 1,5 toneladas.

El método Bootstrap refuerza esta conclusión, al mostrar que la distribución de la diferencia de medias (Figura 4) se concentra por encima de cero, evidenciando que los países más ricos emiten sustancialmente más CO₂ per cápita que los de menor ingreso.

Estos resultados son coherentes con la evidencia internacional sobre la relación entre nivel de desarrollo económico y emisiones, donde un mayor ingreso suele estar asociado a mayores niveles de consumo energético, transporte e industria. No obstante, también sugiere la necesidad de transiciones energéticas diferenciadas, dado que los países de bajo ingreso contribuyen en menor medida a las emisiones globales, pero suelen ser los más vulnerables a los efectos del cambio climático.



**Tabla 3.** Diferencia de medias con Boostrap. Fuente: Banco Mundial. Elaboración propia.

Gráfico, Histograma

El contenido generado por IA puede ser incorrecto.

**Figura 4.** Distribución de la diferencia de emisiones CO2 per cápita. Fuente: Banco Mundial. Elaboración propia.

1. **Evalúe la evidencia:**

* Determine si el intervalo de confianza incluye o no el valor 0.
* Reporte el p-valor de la prueba de hipótesis y explique qué significa en términos del riesgo de error tipo I.
* Verifique si el resultado de la prueba y el intervalo conducen a la misma conclusión.

**Respuesta:** El intervalo de confianza al 95% tiene resultados entre **4,748 a 8,456 lo que indica que existe una diferencia estadísticamente significativa entre las emisiones promedio de CO2 per cápita de los países con alto y bajo PIB. En este intervalo se muestra que los piases con mayor PIB tiene valores más altos en CO2. Estos resultados se pueden comprobar también con el p-valor obtenido (1,0147e-10) y que es muy bajo del 5% . Por lo tanto, hay una probabilidad de cometer error tipo I, es decir, rechazar la hipótesis nula siendo verdadera. Por tanto, se rechaza se rechaza la hipótesis nula de la diferencia de medias y se concluye que las emisiones difieren de manera significativa entre los grupos.**

**Finalmente, el intervalo de confianza como la prueba de hipótesis es similar y se concluye que los países con mayor PIB per cápita emiten significativamente más CO₂ per cápita que los países con menor nivel de ingreso, y la magnitud de esta diferencia es robusta bajo ambos métodos de estimación (bootstrap y prueba t).**

**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?

1. **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?

1. **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?

# Referencias

Massey, D., Arango, J., Graeme, H., Kouaouci, A., Pellegrino, A., & Taylor, E. (1993). Theories of International Migration: A Review and Appraisal. *Population and Development Review*.

Migración Colombia . (2023). *Colombia, el país más solidario con la migración venezolana.*

Migración Colombia. (2024). *Informe de migrantes venezolanos en Colombia en febrero de 2024.*

Kasman, A., & Selman, Y. (2015). CO2 emissions, economic growth, energy consumption, trade and urbanization in new EU member and candidate countries: A panel data analysis. *Economic Modelling*, *44*.

Ritchie, H. (2023, August 31). *Global inequalities in CO2 emissions*. Our World in Data. https://ourworldindata.org/inequality-co2

United Nations Environment Programme. (2023). *Emissions Gap Report 2023*. UN Environment Programme. https://www.unep.org/interactives/emissions-gap-report/2023/

Dong K, Hochman G, Zhang Y, Sun R, Li H, Liao H. CO2 emissions, economic and population growth, and renewable energy: empirical evidence across regions. Energy Economics. 2018

Monasterolo I, Roventini A, Foxon TJ. Uncertainty of climate policies and implications for economics and finance: an evolutionary economics approach. Ecol Econ. 2019.

Mitić P, Fedajev A, Radulescu M, Rehman A. The relationship between CO2 emissions, economic growth, available energy, and employment in SEE countries. Environ Sci Pollut Res Int. 2023

Kuznets S. Economic growth and income inequality. Am. Econ. Rev. 1955

1. **https://github.com/Javier-Amaya-Nieto/StoryTelling3/tree/main** [↑](#footnote-ref-1)