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: Análisis Exploratorio de Datos

# Formulación del problema de indagación y la claridad del contexto analítico

## El fenómeno migratorio que atraviesa Colombia, aunque no es nuevo, ha marcado un precedente significativo para el Estado en términos de articulación institucional, al configurar un nuevo mapa social, económico y político que requiere atención frente a la llegada masiva de migrantes. En los últimos años, Colombia se ha consolidado como el principal país receptor de venezolanos en la región: para el año 2025, se estima que cerca de 2,8 millones de personas han ingresado al territorio nacional. Este movimiento humano sin precedentes, motivado por la búsqueda de mejores condiciones de vida, plantea desafíos de gran magnitud para la capacidad institucional y la respuesta de las autoridades tanto locales como nacionales. Se requiere no solo de voluntad política para diseñar e implementar respuestas integrales e incluyentes, sino también de recursos materiales y técnicos que permitan dar soluciones efectivas y sostenibles a este reto (Migración Colombia , 2023). Por esta razón, garantizar condiciones mínimas de bienestar resulta especialmente complejo ante un flujo tan voluminoso en un periodo tan corto y una alternativa para agilizar la asistencia es recurrir a herramientas estadísticas que permitan caracterizar con mayor precisión a los hogares más vulnerables y tomar decisiones basadas en evidencia. De este modo, es posible brindar soluciones oportunas y contribuir a una integración efectiva de la población migrante. Así, las autoridades locales estarían en mejores condiciones de atender las necesidades básicas de estas familias, sin que su calidad de vida se vea comprometida durante el proceso de adaptación. No obstante, el flujo migratorio continúa siendo dinámico y cambiante, lo que representa un desafío no solo para la regularización de las personas, sino también para su integración socioeconómica en Colombia. Según Migración Colombia, en 2023 la cifra de migrantes en el país se redujo en 1,10 %, independientemente de si ello obedeció al retorno a Venezuela o a que muchos utilizan a Colombia como territorio de tránsito (Migración Colombia, 2024). En el marco del trabajo técnico para la Alcaldía, la implementación de políticas de vivienda y acceso a servicios básicos se plantea como una prioridad estratégica. Un criterio clave para orientar estas políticas consiste en focalizar esfuerzos en ciudades donde los hogares migrantes presentan menores incentivos de movilidad posterior2, pues esto permite optimizar los recursos en poblaciones más estables y menos transitorias, aunque no necesariamente en mejores condiciones socioeconómicas. Este enfoque busca no solo atender necesidades inmediatas, sino también generar bases más sólidas para la integración de largo plazo. A partir de los datos de la Evaluación Conjunta de Necesidades del año 2025 (ECN2025), se observa que Cartagena de Indias es la ciudad con menor proporción de hogares que expresan intención de mudarse fuera de su territorio. Este hallazgo contrasta con municipios fronterizos como Pasto, Arauca, Riohacha y Maicao, caracterizados por la presencia de población migrante en tránsito, y con grandes urbes como Bogotá, Medellín o Pereira, que funcionan como destinos temporales donde los migrantes suelen instalarse inicialmente para generar ingresos y, posteriormente, desplazarse hacia otros lugares (Figura 1). A ello se suma la existencia de redes comunitarias que brindan apoyo en empleo y vivienda, consolidando dinámicas de arraigo que permiten aproximarse a una política pública más eficiente y de mayor impacto en términos de integración. Este patrón es consistente con la literatura sobre migraciones, que muestra cómo la combinación de oportunidades laborales inmediatas y redes de apoyo locales tiende a disminuir la movilidad secundaria y a favorecer la consolidación de proyectos de vida en el lugar de asentamiento (Massey, et al., 1993). Por ello, la decisión de focalizar el análisis en Cartagena no solo responde a la evidencia estadística sobre la intención de permanencia, sino también a un marco conceptual que entiende la integración como un proceso que requiere estabilidad territorial (Figura 2). Esta perspectiva contribuye a situar la política pública en un camino más eficiente respondiendo a necesidades inmediatas de alojamiento y servicios, mientras se sientan bases más sólidas para la inclusión social de la población migrante y refugiada.

Finalmente, dado el objetivo de orientar políticas públicas de alto impacto y la disponibilidad de datos de la ECN2025, el presente análisis se focalizará geográficamente en la ciudad de Cartagena de Indias, selección que responde a la baja intención de movilidad posterior3 de los hogares migrantes, sugiriendo una mayor estabilidad residencial propicia para intervenciones a largo plazo. En este contexto, este trabajo busca responder si las carencias en el acceso a servicios básicos de los hogares migrantes en Cartagena difieren según su nivel de vulnerabilidad. Para ello, se utilizará como unidad de análisis los hogares de migrantes provenientes de Venezuela residentes en la ciudad de Cartagena encuestadas en la ECN2025. De esta manera, la variable dependiente en este análisis es la carencia de acceso a servicios públicos y la variable independiente es el nivel de vulnerabilidad considerando situaciones de vulnerabilidad las siguientes: (i) Jefatura del hogar femenina, ii) hogar monoparental, iii) presencia de personas con discapacidad en el hogar y iv) antecedente de violencia. Los hallazgos de este análisis permitirán a la Alcaldía de Cartagena diseñar y focalizar políticas públicas más eficientes, asegurando que los recursos se dirijan a los hogares migrantes con mayores niveles de vulnerabilidad que al final representan mayores barreras para su integración a la sociedad.

* La definición de las variables utilizadas y la justificación de su selección

# Parte 2: Evidencia empírica y estimación

## Diseño del contraste estadístico

* Seleccionar la prueba de acuerdo con la situación
* Explicar brevemente qué se está comparando y cuál sería la expectativa teórica o empírica según el hecho estilizado que se desea comprobar.
* 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.

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 la medida que proponemos para determinar un punto de corte sería la mediana, considerando la distribución sesgada a la derecha de la variable de ingreso per cápita.

Dado que el interés del contraste es determinar si los países de altos ingresos presentan mayores emisiones de CO₂ per cápita que los de bajos ingresos, la prueba se formula como una prueba de una cola (superior). Por lo tanto, 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 media 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 media de emisiones de CO₂ per cápita significativamente mayor que la de los países de ingreso bajo.

## Revisar los supuestos y ejecutar la prueba

* 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
* Ejecutar la prueba y reporte de los resultados relevantes como el estadístico de prueba, el intervalo de confianza estimado
* Reporte del valor p de la prueba de hipótesis y explicar la significancia en términos de riesgo de error tipo I

# Parte 3: Interpretación y comunicación de los resultados

## Interpretación

* Verificar si el resultado de la prueba y el intervalo conducen a la misma conclusión
* Interpretar el resultado en relación con la hipótesis nula y con el contexto del análisis
* Explicar qué implican los resultados en términos sustantivos: ¿qué nos dicen sobre el fenómeno estudiado?

## 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, compleméntelos e integrarlos 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.

* Consistencia entre visualizaciones y la respuesta a la pregunta planteada
* Precisión y pertenencia de la respuesta planteada de acuerdo con la situación del contexto
* Coherencia narrativa y argumentativa en las conclusiones
* Añadir una breve descripción de los principales ajustes realizados a partir de la retroalimentación recibida.

# Referencias

Banco Mundial. (18 de Octubre de 2025). *World Development Indicators*.

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

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*.

Kuznets, S. (1955). Economic growth and income inequality. *The American Economic Review*.

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.*

Mitic, P., Fedajev, A., Radulescu, M., & Rehman, A. (2023). The relationship between CO₂ emissions, economic growth, available energy, and employment in SEE countries. *Environmental Science and Pollution Research*.

Monasterolo, I., Roventini , A., & Foxon, T. (2019). Uncertainty of climate policies and implications for economics and finance: An evolutionary economics approach. *Ecological Economics*.

OECD. (2018). Compendium of Productivity Indicators. *OECD comunications*, págs. https://www.oecd.org/content/dam/oecd/en/publications/reports/2018/06/oecd-compendium-of-productivity-indicators-2018\_g1g8e1ba/pdtvy-2018-en.pdf.

Ritchie, H. (2023). *Global inequalities in CO₂ emissions*. Obtenido de Our World in Data.

Ritchie, H., Rosado, P., & Roser, M. (26 de September de 2023). Per capita, national, historical: how do countries compare on CO2 metrics? *The World Bank*.

United Nations Environment Programme. (2023). *Emissions Gap Report 2023*. Obtenido de United Nations Environment Programme.

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