Exploración y visualización de datos para lo socioeconómico

Miguel Andrés Garzón

|  |
| --- |
| Proyecto 1: EDA, visualización y estadística para respaldar hipótesis**[[1]](#footnote-1)** |

**ELABORADO POR:**

Laura Sarif Rivera Sanabria

María Camila Caraballo Candela

Javier Antonio Amaya Nieto

**Enunciado del taller- situación**

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.

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.

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: 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 posterior[[2]](#footnote-2), 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.

Gráfico

El contenido generado por IA puede ser incorrecto.

**Figura 1.** Ranking de ciudades con mayor intención de movilidad. Fuente: Evaluación Conjunta de Necesidades. Elaboración propia

## 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.

Gráfico, Gráfico de barras

El contenido generado por IA puede ser incorrecto.

**Figura 2.** Ranking de ciudades con menor intención de movilidad. Fuente: Evaluación Conjunta de Necesidades. Elaboración propia

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 posterior[[3]](#footnote-3) 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

Considerando la pregunta de investigación planteada, en esta sección se definen y describen las variables centrales que orientan el análisis. La variable dependiente corresponde a la carencia de acceso a servicios públicos, entendida como un indicador sintético de las condiciones de vivienda. Por su parte, la variable independiente se construye como un índice de vulnerabilidad a nivel de hogar, elaborado a partir de cuatro dimensiones estructurales que reflejan riesgos sociales específicos. El objetivo de esta sección es caracterizar ambas variables, identificando su composición y lógica de construcción como paso previo al análisis comparativo de su relación.

**1.2.1 Variable dependiente-carencia de acceso a servicios públicos:** La variable dependiente del estudio es la carencia de acceso a servicios públicos, concebida como una medida integral de las condiciones básicas de la vivienda. Para su construcción se consideraron cinco dimensiones fundamentales: acceso a electricidad, gas, alcantarillado, acueducto y recolección de basuras. Cada servicio ausente en el hogar se contabiliza con un punto, de modo que la variable puede tomar valores entre 0, cuando el hogar dispone de todos los servicios, y 5, cuando carece de la totalidad. Así, este indicador refleja de forma clara el nivel de privación en infraestructura básica al que están expuestos los hogares migrantes en Cartagena. A continuación, se presentan las estadísticas descriptivas:

**Tabla 1.** Estadísticas descriptivas de la variable dependiente-carencia de servicios públicos.

La variable servicios es de tipo numérico discreto y cuenta con 163 observaciones sin valores faltantes. Representa el número de servicios disponibles en cada hogar, con un rango de 0 a 4, de manera que ningún hogar alcanza el máximo teórico de cinco. Los estadísticos descriptivos muestran un mínimo de 0, un primer cuartil de 0, una mediana de 1, una media de 1,34, un tercer cuartil de 2 y un máximo de 4, lo que evidencia que la distribución está sesgada hacia valores bajos y que la mayoría de los hogares cuentan con pocos servicios. Aunque se observa que algunos hogares tienen todos los servicios posibles dentro del rango definido.

**Tabla 2.** Estadísticas descriptivas básicas de la variable dependiente-carencia de servicios públicos. Representación categórica de la variable

La tabla muestra la distribución de los hogares según el número de servicios con los que cuentan, diferenciando entre proporciones sin ponderar y ponderadas con la variable de expansión. Se observa que la mayor concentración corresponde a los hogares que no disponen de ningún servicio, con un 41,1% sin ponderar y un 43,6% al considerar los factores de expansión. En contraste, los hogares con un solo servicio representan entre el 11,7% y el 13,2%, mientras que aquellos con dos servicios alcanzan cerca de una tercera parte de la población (29,4% sin ponderar y 27,1% ponderado). Los hogares con tres servicios son menos frecuentes, con proporciones de 8% y 5,7% respectivamente, y finalmente, quienes acceden a cuatro servicios corresponden a cerca del 10% de los hogares. Estos resultados evidencian una marcada concentración de hogares en condiciones de baja cobertura de servicios, lo cual refleja desigualdades en el acceso.

La Figura 3 muestra la distribución porcentual de hogares según el número de carencias en servicios públicos (0 a 4), comparando dos grupos (barras azul oscuro-usando el factor de expansión oscuro y gris-sin el factor de expansión). Se observa que en ambos grupos predomina la ausencia de carencias: más del 40 % de los hogares no presenta ninguna. En segundo lugar, cerca de un 25 %–30 % tiene carencia de dos servicios públicos, mientras que los hogares con una, tres o cuatro carencias representan proporciones menores. En conjunto, estos resultados sugieren que los resultados no cambian de forma considerable al usar el factor de expansión y que cerca del 60 % de los hogares tienen problemas de acceso a alguno de los servicios públicos mencionados.

**Figura 3.** Proporción de hogares con servicios públicos faltantes en la ciudad de Cartagena utilizando el factor de expansión (color azul) y sin utilizar (color gris). Fuente: Evaluación Conjunta de Necesidades-2025. Elaboración propia.

**1.2.3 Variable independiente- índice de vulnerabilidad:** La variable independiente corresponde a un índice de vulnerabilidad que sintetiza condiciones estructurales que incrementan el riesgo social de los hogares y que pueden afectar su acceso a servicios públicos. Este índice se construye a partir de cuatro dimensiones: (i) jefatura femenina del hogar, (ii) condición monoparental, (iii) presencia de personas con discapacidad y (iv) antecedentes de violencia o amenazas. Cada condición suma un punto, por lo que el índice varía entre 0 y 4. De esta manera, el indicador permite clasificar a los hogares de acuerdo con su nivel de vulnerabilidad acumulada y constituye la base para analizar en qué medida estas diferencias estructurales se relacionan con las carencias en servicios básicos.

**a) Jefatura femenina del hogar:** Identifica si la persona cabeza de hogar es mujer. Se realiza el análisis de frecuencias sin ponderación que se evidencia que el 61,2% de los hogares tienen jefatura femenina frente al 38,8% con jefatura masculina. Sin embargo, al realizar las frecuencias aplicando el factor de expansión, la distribución se invierte: el 57,2% de los hogares en el país presentan jefatura masculina, mientras que el 42,8% tienen jefatura femenina. Estos resultados reflejan el cambio que se tiene al aplicar la variable de expansión como pesos muestrales.

**Figura 4.** Proporción de hogares con jefatura del hogar masculina y femenina en la ciudad de Cartagena utilizando el factor de expansión (color azul) y sin utilizar este factor (color gris). Fuente: Evaluación Conjunta de Necesidades-2025. Elaboración propia.

**b) Hogar monoparental:** Determina si el hogar está conformado por un padre o madre sin cónyuge y con hijos, o si incluye un “Papá/Mamá” con estado civil soltero, separado, divorciado o viudo.

En relación con esta variable, se observa que no presenta valores faltantes y que sus únicos valores posibles corresponden a 0 y 1. En un primer análisis sin aplicar ponderación, se identifica que el 30,7% de los hogares está conformado por un solo adulto responsable de hijos, mientras que el 69,3% restante cuenta con dos o más adultos a cargo. Al incorporar el factor de expansión, la proporción de hogares monoparentales disminuye a 24,1%, mientras que el 75,9% restante cuenta con dos o más adultos a cargo, mayor a los hogares sin ponderar, lo que ofrece una visión más precisa de su peso en la población total. Este ajuste pone de relieve la representatividad real de estos hogares y permite dimensionar mejor su situación dentro del conjunto de la muestra. Estos resultados son relevantes para la formulación de políticas sociales, pues la presencia de un solo adulto responsable suele estar asociada a mayores retos en términos de ingresos, cuidado infantil y acceso a servicios.

**Figura 5**. Proporción de hogares con jefatura del hogar monoparental en la ciudad de Cartagena utilizando el factor de expansión (color azul) y sin utilizar este factor (color gris). Fuente: Evaluación Conjunta de Necesidades-2025. Elaboración propia.

**c) Presencia de personas con discapacidad:** Señala si al menos un miembro del hogar reporta alguna limitación o barrera física. La variable es de tipo entero, con valores 0 (sin discapacidad) y 1 (al menos un miembro con discapacidad). En el análisis sin ponderar, se observaron que el 99,4% de los hogares no tienen condición de discapacidad menos del 1% tienen una discapacidad. Una vez se aplicó el factor de expansión, se evidencia que la gran mayoría de los hogares no cuenta con miembros con discapacidad y que solo una proporción muy pequeña presenta al menos un miembro con discapacidad. No se encontraron valores faltantes en la variable.

**Figura 6.** Proporción de hogares con alguna persona con discapacidad física en la ciudad de Cartagena utilizando el factor de expansión (color azul) y sin utilizar este factor (color gris). Fuente: Evaluación Conjunta de Necesidades-2025. Elaboración propia.

**d) Antecedentes de violencia o amenazas:** Registra si algún miembro del hogar ha sido víctima de violencia o amenazas. Se trata de una variable de tipo entero con valores 0 y 1, donde el valor 1 indica que al menos un miembro del hogar ha sufrido algún tipo de violencia. En el análisis sin ponderar se identifican 151 hogares que no reportaron situaciones de violencia y 12 que sí lo hicieron, lo que corresponde al 92,6% y 7,4 % de la muestra, respectivamente. Una vez aplicado el factor de expansión el porcentaje de hogares sin violencia se reduce en 8 puntos mientras que el porcentaje de hogares sin violencia incrementa en 8 puntos respectivamente. La variable no presenta valores faltantes. Este hallazgo permite tener una visión más precisa de la prevalencia de este fenómeno dentro de la población estudiada.

**Figura 7.** Proporción de hogares con antecedentes de violencia o amenazas en la ciudad de Cartagena utilizando el factor de expansión (color azul) y sin utilizar este factor (color gris). Fuente: Evaluación Conjunta de Necesidades-2025. Elaboración propia.

**e) Índice de vulnerabilidad:** Síntesis de condiciones estructurales que incrementan el riesgo social de los hogares y que pueden afectar su acceso a servicios públicos. Variable independiente: Índice de vulnerabilidad Estadístico valor Min. 0 1st Qu. 1 Median 1 Mean 1,09202454 3rd Qu. 2 Max. 3

**Tabla 3.** Estadísticas descriptivas de la variable dependiente-carencia de servicios públicos.

El análisis descriptivo del índice de vulnerabilidad evidencia que los valores oscilan entre 0 y 3, donde 0 corresponde a hogares no vulnerables y 3 a los más vulnerables. La media se ubica en 1,09, muy cercana a la mediana (1), lo que indica una distribución concentrada en los niveles bajos de vulnerabilidad. El primer cuartil y la mediana coinciden en 1, mientras que el tercer cuartil se sitúa en 2, mostrando que al menos el 75% de los hogares presentan niveles de vulnerabilidad de 2 o inferiores. Finalmente, estas estadísticas sugieren que la mayoría de los hogares se ubican en condiciones de baja a media vulnerabilidad, con una proporción reducida alcanzando el nivel máximo. Variable dependiente: Índice de vulnerabilidad Servicios Sin ponderar Ponderado Prop Prop 0 24,5 40,3 1 44,8 36,4 2 27,6 20,9 3 3,1 2,4

**Tabla 4.** Estadísticas descriptivas de la variable dependiente-carencia de servicios públicos. Descripción categórica de la variable.

La distribución del índice de vulnerabilidad muestra que, al aplicar ponderaciones, se incrementa de manera significativa la proporción de hogares clasificados como no vulnerables (categoría 0), pasando de 24,5% sin ponderar a 40,3% ponderado. En contraste, la categoría 1, que concentra la mayor proporción de hogares en la estimación simple (44,8%), se reduce a 36,4% con ponderaciones de la variable de expansión. De igual forma, la proporción de hogares con vulnerabilidad media (categoría 2) disminuye de 27,6% a 20,9%, mientras que los altamente vulnerables (categoría 3) mantienen una presencia baja en la muestra, oscilando entre 3,1% y 2,4%. Estos resultados evidencian que, al considerar los factores de expansión, el peso poblacional de los hogares no vulnerables es mayor al que sugiere la muestra sin ponderar, lo que implica una menor incidencia relativa de la vulnerabilidad en la población total.

Al examinar en detalle el índice de vulnerabilidad (Figura 8), se observa que corresponde a una variable categórica con valores que van de 0 a 4, donde cada número refleja la cantidad de condiciones de vulnerabilidad presentes en el hogar. Se trata de una variable entera que no presenta valores faltantes, lo que asegura la integridad de la información para el análisis. En términos generales, la distribución ponderada muestra que el 40,3 % de los hogares no registra ninguna vulnerabilidad, mientras que el 36,4 % presenta al menos una. Por su parte, el 20,9 % acumula dos condiciones de vulnerabilidad y un 2,4 % alcanza tres. Aunque los hogares con niveles más altos de vulnerabilidad representan un porcentaje reducido, su identificación resulta clave, ya que son los más propensos a enfrentar mayores limitaciones en el acceso a servicios básicos y a requerir intervenciones más focalizadas.

**Figura 8.** Proporción de hogares con un índice de vulnerabilidad entre 0 y 4, en la ciudad de Cartagena utilizando el factor de expansión (color azul) y sin utilizar este factor (color gris). Fuente: Evaluación Conjunta de Necesidades-2025. Elaboración propia

**1.2.4 Análisis agregado y relación entre variables**

Con el propósito de representar de manera integral las distintas combinaciones de carencias en servicios públicos a través de los niveles de vulnerabilidad, se elaboró un gráfico de barras apiladas que muestra la distribución de privaciones en cada una de las categorías del índice de vulnerabilidad. Esta visualización resulta especialmente útil, pues permite observar con mayor claridad las diferencias en la incidencia de carencias según las condiciones estructurales de los hogares y, de este modo, avanzar en la respuesta a la pregunta de investigación planteada.

El análisis gráfico revela que la distribución de carencias en servicios públicos varía de manera sustancial entre los distintos grupos de hogares vulnerables, lo que permite identificar perfiles de mayor riesgo. Por ejemplo, entre los hogares sin vulnerabilidad, el 36,8 % no reporta carencias y apenas un 11,2 % acumula cuatro, mientras que en aquellos con jefatura femenina o en condición monoparental la proporción de hogares sin carencias disminuye de manera importante, observándose una mayor concentración en niveles intermedios de privación. Esto sugiere que las características estructurales de los hogares inciden en la probabilidad de enfrentar limitaciones en el acceso a servicios básicos, aunque la intensidad de las carencias no sea homogénea entre todos los hogares vulnerables.

Un patrón particular se observa en los hogares con presencia de personas con discapacidad: estos concentran principalmente dos carencias, un hallazgo que debe interpretarse considerando el bajo número de hogares en esta categoría, lo que explica la concentración en un único nivel. Por otro lado, los hogares víctimas de violencia muestran una distribución más dispersa: aunque el 54,1 % no presenta carencias, un 15,1 % acumula cuatro, lo que indica un mayor riesgo de privaciones extremas en este grupo. Estos resultados ponen de relieve la heterogeneidad entre los hogares vulnerables y la necesidad de que las políticas públicas no se enfoquen únicamente en la cobertura general, sino que prioricen a aquellos grupos con mayor exposición a múltiples carencias, tales como hogares monoparentales, con personas en condición de discapacidad o víctimas de violencia.

**Figura 9.** Distribución de carencias en servicios por condición de vulnerabilidad del hogar. Fuente: Evaluación Conjunta de Necesidades. Elaboración propia.

El segundo gráfico de barras apiladas muestra la distribución de las carencias en servicios básicos según el índice de vulnerabilidad de los hogares, que varía entre 0 para aquellos sin condiciones de vulnerabilidad y 3 para quienes enfrentan tres vulnerabilidades simultáneas. Un primer hallazgo es que, incluso en hogares sin vulnerabilidad, no existe garantía plena de acceso a los servicios: aunque el 36,8 % no presenta carencias, un 29,3 % enfrenta al menos una y un 11,2 % acumula cuatro. Esto evidencia que la ausencia de vulnerabilidades no elimina por completo la posibilidad de privaciones, lo que refleja limitaciones estructurales en la provisión de servicios públicos.

A medida que aumenta la vulnerabilidad, la distribución de carencias cambia de manera notable. En los hogares con una condición de vulnerabilidad, la proporción de aquellos sin carencias es mayor (42,8 %), pero al mismo tiempo crece el grupo que acumula cuatro carencias (13,2 %), lo que muestra una mayor polarización. En los hogares con dos condiciones, predomina la concentración en niveles intermedios: el 61 % presenta al menos una carencia y un 21,7 % alcanza dos, reforzando la idea de que la acumulación de vulnerabilidades incrementa la exposición a limitaciones en el acceso a servicios básicos.

Finalmente, los hogares con tres condiciones de vulnerabilidad exhiben el panorama más crítico. Si bien el 17,6 % no reporta carencias, el 57,5 % enfrenta al menos una y cerca de una cuarta parte (24,9 %) acumula dos o más. En términos de política pública, estos resultados señalan la necesidad de priorizar intervenciones focalizadas hacia los hogares con mayor acumulación de vulnerabilidades, pues son los más expuestos a situaciones de exclusión en el acceso a servicios esenciales. Atender a este grupo de manera diferenciada y eficiente constituye un criterio clave para orientar la acción de la alcaldía.

**Figura 10.** Distribución porcentual de hogares según la vulnerabilidad y el número de carencias en servicios básicos. Fuente: Evaluación Conjunta de Necesidades. Elaboración propia.

# 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.

En el taller 2, se dieron algunos comentarios y se hicieron los siguientes cambios:

1. En relación con la observación sobre la variable de discapacidad, se nos indicó que esta no contiene datos y que, además, no se realizó el análisis univariado correspondiente. Sin embargo, aclaramos que dicho análisis sí fue realizado y se evidencia en el gráfico respectivo. En este se observa que la mayoría de los hogares no cuentan con personas con discapacidad, razón por la cual se consideró que dicha variable no aporta información relevante para la construcción del índice de vulnerabilidad. Quedamos atentos a su recomendación sobre si conviene eliminarla definitivamente y conservar únicamente las otras tres variables.
2. Se ajustaron las tablas y los gráficos para mantener un formato homogéneo con el utilizado en la primera parte del trabajo (fase de construcción). No obstante, aún está pendiente aplicar el mismo ajuste a las tablas correspondientes al análisis univariado.

# 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.

1. https://github.com/Laurarivera34/Proyecto-1-EDA-estad-stica [↑](#footnote-ref-1)
2. Para efectos de este análisis y considerando la naturaleza de los datos proporcionados, se considera movilidad posterior como intención de movilidad fuera de la ciudad en los siguientes 6 meses. [↑](#footnote-ref-2)
3. Para efectos de este análisis y considerando la naturaleza de los datos proporcionados, se considera movilidad posterior como intención de movilidad fuera de la ciudad en los siguientes 6 meses. [↑](#footnote-ref-3)