

The background is a dark blue gradient. It features several vertical bars of varying heights and widths, some in a lighter blue and others in a darker blue. Scattered throughout are small squares, some solid red, some solid white, and some with red outlines. Small red triangles pointing upwards are also visible. The overall aesthetic is modern and data-oriented.

ITLA

CIENCIA DE DATOS

REPUBLICA DOMINICANA
INSTITUTO TECNOLÓGICO DE LAS AMÉRICAS



PROGRAMACION I

SECCION:

2025-C-1-2802-3089-TCD-101

PROYECTO FINAL PROGRAMACION I

Análisis del Conjunto de Datos ORGANDATA

Rosy Elania Arvelo Pérez

20240861

Jose Arcadio Garcia Garcia.

Indice

Introducción	4
Metodologia	¡Error! Marcador no definido.
Modelos de Regresión.....	6
Modelo Robusto (RLM).....	6
Modelo de Regresión Lineal Tradicional (LM)	6
Comparación de Modelos.....	7
Análisis de los Gráficos	8
Residuals vs Fitted (Residuos vs Ajustados).....	8
Q-Q Plot (Gráfico Cuantil-Cuantil de los Residuos) Compara la distribución de los residuos con una distribución normal teórica.	9
Scale-Location Plot (Gráfico de Escala-Localización)	9
Residuals vs Leverage (Residuos vs Palanca) Identifica observaciones influyentes que podrían tener un gran impacto en el ajuste del modelo.	9
Conclusión	13
Recomendaciones	14
Grafico 1: Gráfico de Diagnóstico	8
Grafico 2: Relación entre PIB (log) y Donaciones (log)	10
Grafico 3: Relación entre Población (log) y Donaciones (log).....	11
Grafico 4: Gráficos de Residuos	12

Introducción

En este informe se presenta un análisis detallado del conjunto de datos organdata, del paquete socviz, recopila información sobre las tasas de donaciones de órganos en varios países europeos. Entre las principales variables de interés están:

- **donors:** tasa de donantes por millón de habitantes.
- **gdp:** Producto Interno Bruto per cápita.
- **pop:** población del país.
- Además, variables de tipo categórico como el país y el sistema de consentimiento (presunto o explícito).

El conjunto de datos contiene mediciones para múltiples años por país, permitiendo analizar no solo diferencias entre países, sino también evolución temporal.

Antes del análisis, los datos fueron limpiados para eliminar registros con valores faltantes (NA) y se transformaron algunas variables aplicándoles logaritmos. Estas transformaciones permiten estabilizar varianzas y facilitar interpretaciones más lineales entre las variables.

Metodología

Para llevar a cabo este análisis, se utilizó el lenguaje de programación R, una herramienta ampliamente reconocida por su eficacia en el manejo y visualización de datos estadísticos.

La metodología aplicada sigue estos pasos:

- **Transformación de Variables:** Se aplicó logaritmo natural a las variables gdp, pop y donors (con ajuste de +1 para evitar logaritmo de cero).
- **Modelos de Regresión:** Se compararon dos modelos:
 - Modelo de regresión robusta (rlm), menos sensible a valores atípicos.
 - Modelo de regresión lineal tradicional (lm).
- **Evaluación de los residuos:** Se analizó la distribución de los residuos, homocedasticidad (varianza constante) y presencia de valores atípicos o puntos de alta influencia.
- **Visualización:** Se generaron gráficos de dispersión con líneas de tendencia para evaluar relaciones entre las variables.

Modelos de Regresión

Modelo Robusto (RLM)

El Modelo Robusto Lineal (RLM) es una técnica de regresión utilizada cuando se tienen datos con valores atípicos o outliers que podrían afectar los resultados de un modelo de regresión tradicional. A diferencia de la regresión lineal estándar, el RLM reduce el impacto de estos valores extremos y proporciona estimaciones más confiables.

En este análisis, el RLM se ha utilizado para ajustar un modelo de regresión lineal, asegurando que los resultados no se vean distorsionados por posibles outliers en los datos. Esto permite obtener una visión más precisa de la relación entre las variables, especialmente cuando los datos no siguen completamente los supuestos de la regresión lineal clásica.

Resultados Clave:

- El coeficiente de `log_gdp` es positivo (0.2076), sugiriendo que un mayor PIB per cápita tiende a asociarse con mayores tasas de donación.
- El coeficiente de `log_pop` es ligeramente negativo (-0.0196), pero no significativo.
- La prueba de Breusch-Pagan mostró **homocedasticidad** (p-valor = 0.8092).
- Aunque se aplicó el test de Jarque-Bera, recordamos que no es requerido en regresión robusta, pero igual indicó no evidencia de no normalidad (p-valor = 0.138).

Modelo de Regresión Lineal Tradicional (LM)

El Modelo de Regresión Lineal Tradicional (LM) es una técnica que busca encontrar la mejor línea recta que se ajuste a los datos, minimizando la suma de los cuadrados de los errores. Este modelo supone que los datos tienen una relación lineal entre las variables y que los errores siguen una distribución normal con varianza constante.

En este análisis, el LM se utiliza para estimar la relación entre las variables, asumiendo que los datos cumplen con los supuestos de normalidad y homocedasticidad. Este enfoque es útil cuando se tiene confianza en que los datos no contienen outliers significativos y siguen una distribución adecuada para la regresión lineal.

Resultados Clave:

- El PIB sigue siendo un predictor positivo y significativo ($p = 0.0193$).
- La población ahora también aparece como significativa pero negativa ($p = 0.0494$).
- El modelo explica un porcentaje muy pequeño de la variabilidad (R^2 ajustado = 0.0382), lo que indica que existen otros factores no considerados.

Comparación de Modelos

Aspecto	Modelo Robusto (RLM)	Modelo Tradicional (LM)
Sensibilidad a outliers	Baja	Alta
Significancia de PIB	Alta	Alta
Significancia de Población	Baja	Marginalmente significativa
Supuesto de normalidad residuos	No requerido	Requerido
Homocedasticidad	Cumple	Cumple
Interpretabilidad	Buena	Buena

El modelo robusto es más adecuado si sospechamos de la presencia de datos atípicos, lo cual es muy común en estudios internacionales donde algunos países pueden tener comportamientos extremos. El modelo tradicional es útil para comparación, pero puede ser afectado por esos valores.

Análisis de los Gráficos

Análisis del Gráfico de Diagnóstico

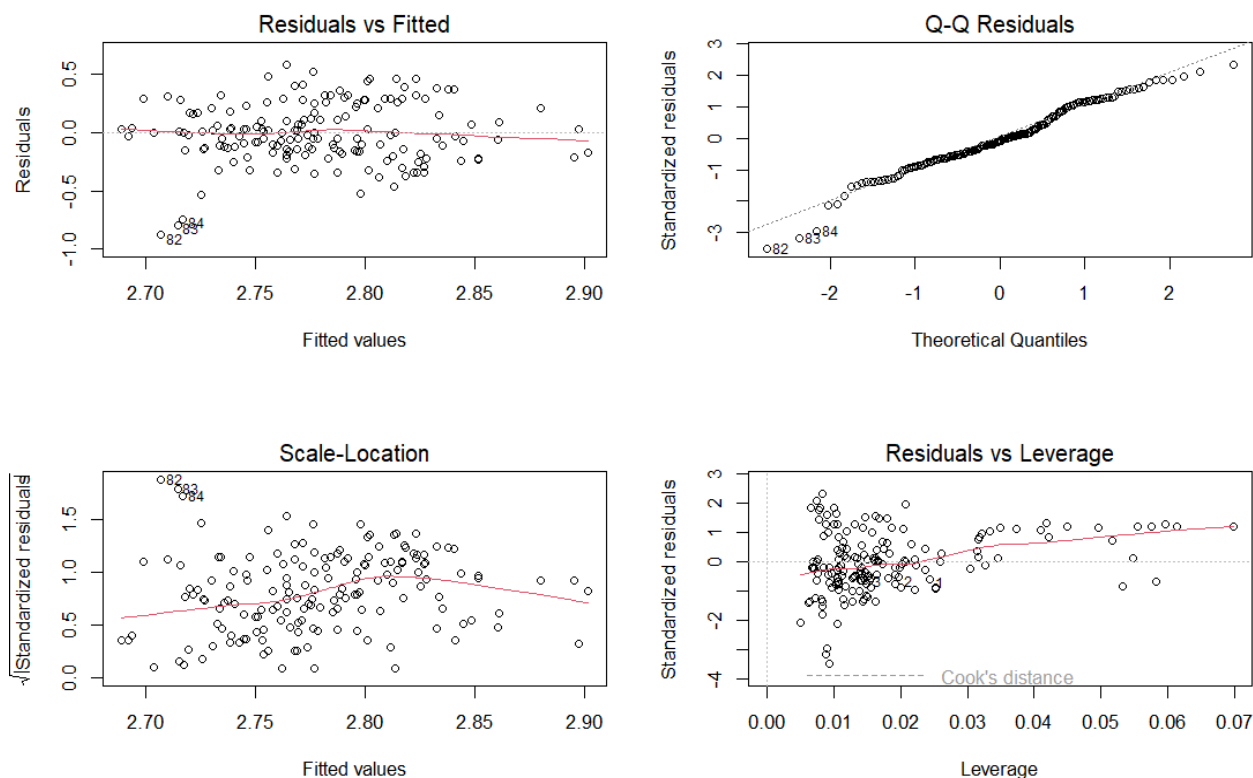


Gráfico 1: Gráfico de Diagnóstico

Residuals vs Fitted (Residuos vs Ajustados)

Detecta si hay patrones en los residuos, lo que podría indicar problemas como no linealidad o heterocedasticidad (varianza no constante).

Interpretación:

- Los residuos están dispersos alrededor de la línea horizontal (cero) sin un patrón claro.
- Esto sugiere que el supuesto de linealidad y homocedasticidad se cumple razonablemente.
- Sin embargo, vemos algunos puntos más alejados como los casos 82 y 84, que podrían ser valores atípicos.

Q-Q Plot (Gráfico Cuantil-Cuantil de los Residuos)

Compara la distribución de los residuos con una distribución normal teórica.

Interpretación:

- La mayoría de los puntos siguen bastante bien la línea diagonal, pero en las colas (extremos) se observa alguna desviación.
- Casos como 82 y 84 son outliers que no siguen perfectamente la normalidad.
- Conclusión: La normalidad de los residuos es aceptable pero no perfecta. Hay ligeras desviaciones en los extremos.

Scale-Location Plot (Gráfico de Escala-Localización)

Examina si los residuos tienen varianza constante (homocedasticidad).

Interpretación:

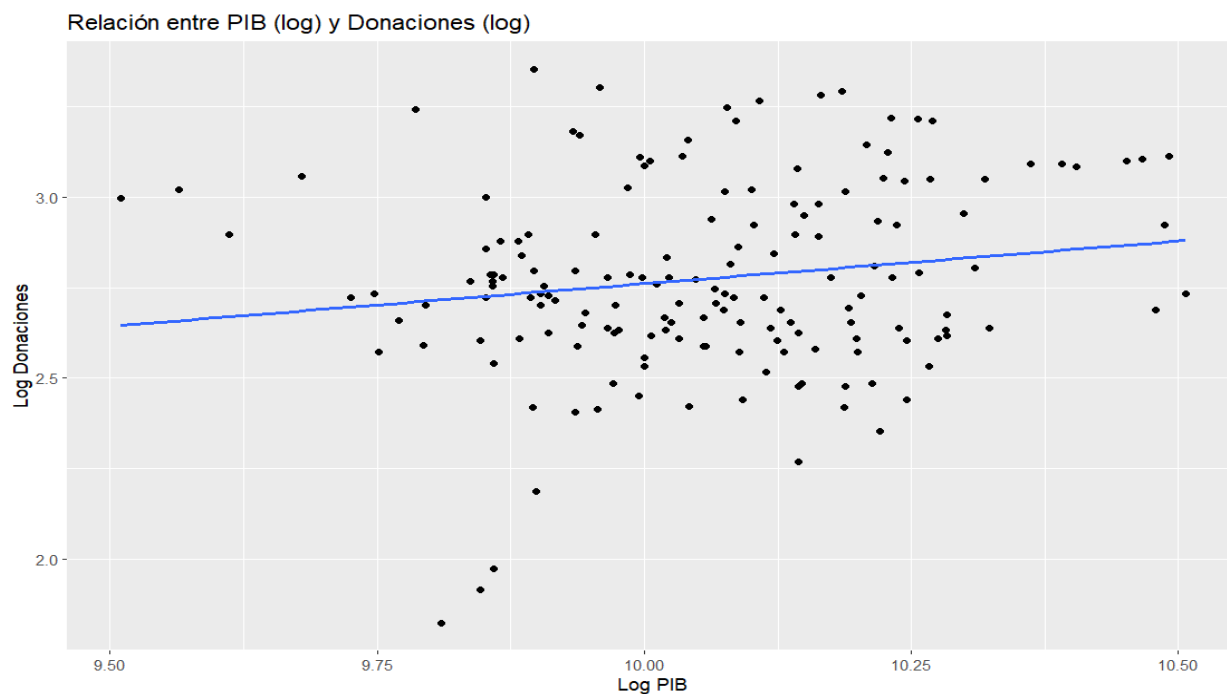
- La dispersión de los puntos parece bastante uniforme, sin un patrón de "embudo" claro (ni abriéndose ni cerrándose).
- Esto sugiere que la varianza de los residuos es aproximadamente constante.
- De nuevo, los casos 82 y 84 destacan ligeramente, pero no afectan seriamente la conclusión general.

Residuals vs Leverage (Residuos vs Palanca)

Identifica observaciones influyentes que podrían tener un gran impacto en el ajuste del modelo.

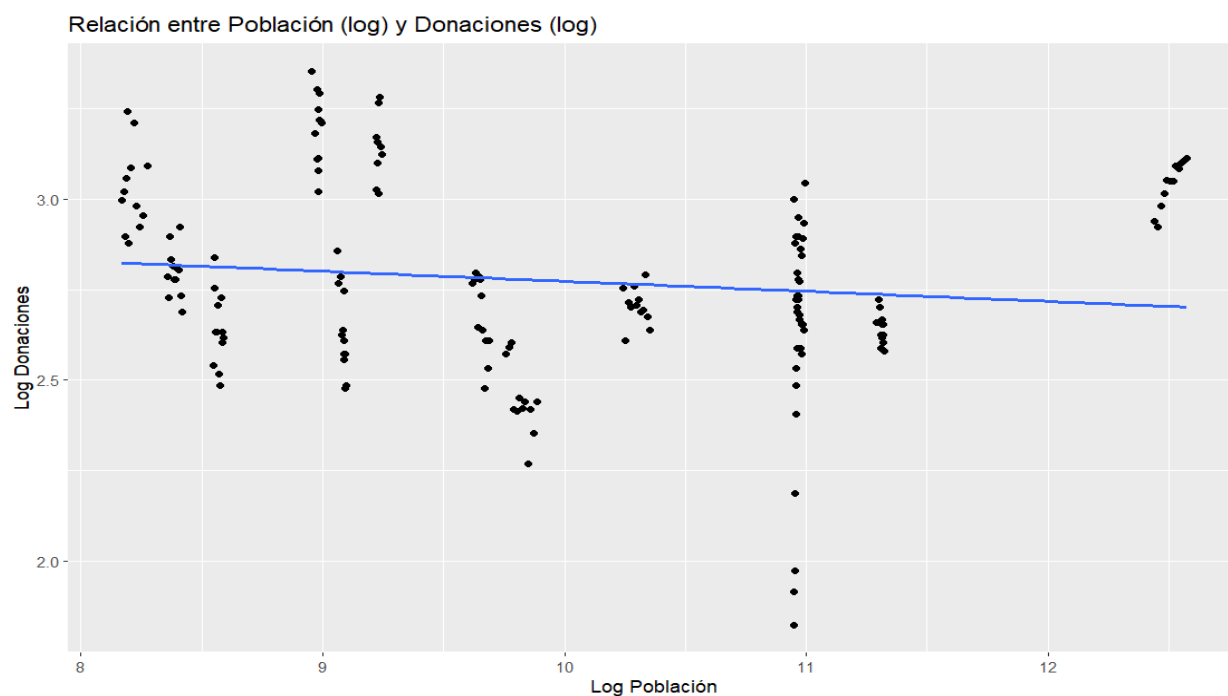
Interpretación:

- La mayoría de los puntos tienen baja influencia (baja leverage y residuos pequeños).
- Hay algunos puntos con un leverage mayor, pero ninguno parece superar claramente la distancia de Cook crítica (la línea punteada).
- Casos 82 y 84 vuelven a aparecer como potencialmente influyentes, pero no son lo suficientemente extremos como para preocuparnos demasiado.

PIB vs Donaciones:**Grafico 2: Relación entre PIB (log) y Donaciones (log)**

Se observa una tendencia positiva: a mayor PIB per cápita, mayor tasa de donación de órganos. Esto sugiere que los países más ricos podrían tener sistemas de salud más eficaces para promover y gestionar la donación.

Población vs Donaciones:



Gráficos de Residuos

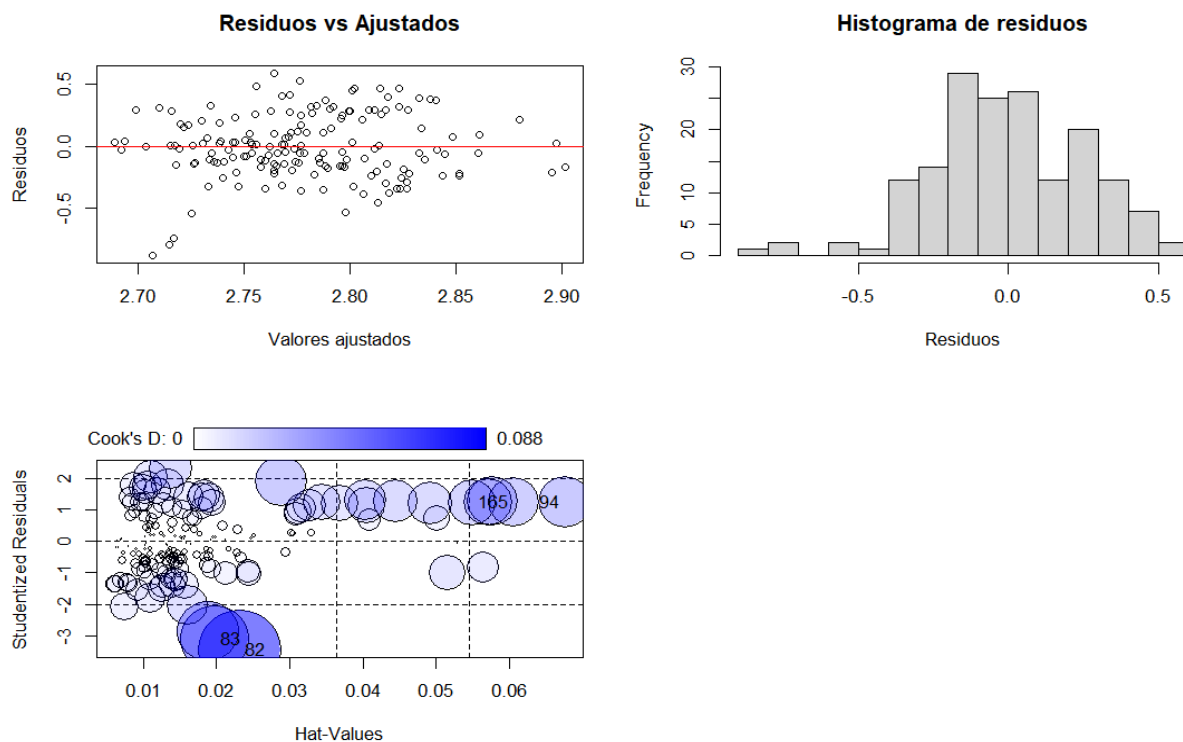


Grafico 4: Gráficos de Residuos

Residuos vs Ajustados:

No se observa un patrón claro, indicando que la suposición de homocedasticidad se cumple.

Histograma de Residuos:

La distribución de los residuos parece aproximadamente simétrica, sin colas extremas.

Influence Plot:

Se identificaron algunos puntos con alta influencia en el modelo lm, específicamente los casos 82, 83, 94 y 165. Esto respalda el uso del modelo robusto.

Conclusión

1. **Relación positiva significativa entre el PIB per cápita y las tasas de donación de órganos:**

Tanto el Modelo Robusto Lineal (RLM) como el Modelo de Regresión Lineal Tradicional (LM) muestran una relación positiva entre el PIB per cápita y las tasas de donación de órganos. En ambos modelos, el coeficiente de \log_gdp (PIB per cápita) es significativo y positivo. Esto sugiere que los países con un mayor PIB per cápita tienden a tener mayores tasas de donación de órganos. Esta relación puede estar vinculada a mejores sistemas de salud y mayores recursos disponibles para promover y gestionar las donaciones. Sin embargo, es importante destacar que el RLM proporciona un ajuste más confiable cuando existen puntos extremos.

2. **Efecto negativo de la población, aunque con menor influencia:**

En el RLM, el coeficiente para \log_pop (población) es negativo pero no significativo, lo que indica que la población no tiene un impacto claro sobre las tasas de donación en este modelo. Por otro lado, en el LM, la población es significativa pero con un efecto negativo leve ($p = 0.0494$). Esto sugiere que en algunos países más poblados, la tasa de donación de órganos podría ser menor, aunque la influencia de la población es mucho menos clara en comparación con el PIB. Esto podría deberse a factores adicionales, como la infraestructura de salud, que no están siendo capturados adecuadamente en el modelo.

3. **El modelo robusto proporciona un ajuste más confiable, menos afectado por valores extremos:**

El uso del RLM es especialmente útil cuando los datos contienen outliers o valores extremos, como se observa en los casos 82 y 84, que tienen una influencia significativa en el LM. El RLM reduce el impacto de estos valores extremos, lo que proporciona un ajuste más robusto y confiable. En cambio, el LM, al ser más sensible a los outliers, podría generar resultados distorsionados si hay puntos atípicos, como es el caso en este análisis.

4. **Baja variabilidad explicada por ambos modelos, lo que sugiere la influencia de otros factores no considerados:**

En ambos modelos, el R^2 ajustado es bajo (0.0382 en el LM), lo que indica que los factores incluidos en los modelos (PIB y población) explican solo una pequeña fracción de la variabilidad en las tasas de donación de órganos. Esto sugiere que hay otros factores, como elementos culturales, legales o institucionales, que influyen en las tasas de donación pero que no están siendo capturados por estos modelos. La inclusión de variables adicionales podría mejorar la capacidad explicativa de los modelos.

Recomendaciones

1. **Extender el modelo:** incluir variables como tipo de legislación sobre donación, gasto público en salud, educación, etc.: Para mejorar la capacidad explicativa del modelo y capturar factores adicionales que podrían influir en las tasas de donación de órganos, se recomienda extender el modelo actual incorporando variables adicionales. Por ejemplo, el tipo de legislación sobre donación de órganos podría tener un impacto significativo, ya que las políticas legales de consentimiento pueden influir directamente en las tasas de donación. Igualmente, el gasto público en salud y la educación podrían ser factores importantes, ya que un mayor financiamiento en estas áreas podría mejorar la infraestructura de salud y la concienciación sobre la importancia de la donación de órganos. Incorporar estas variables permitirá una visión más completa de los factores que influyen en las tasas de donación y puede mejorar la capacidad del modelo para explicar la variabilidad observada.
2. **Análisis multianual:** considerar efectos por año o tendencias temporales usando modelos de efectos mixtos: Otro paso importante es realizar un análisis multianual para capturar posibles efectos temporales. Las tasas de donación de órganos podrían variar a lo largo del tiempo debido a cambios en las políticas, la economía o los eventos sociales. Para esto, se recomienda utilizar modelos de efectos mixtos, que permiten modelar tanto los efectos fijos (por ejemplo, PIB, población) como los efectos aleatorios asociados a los años o periodos de tiempo. Este tipo de modelo puede ayudar a identificar tendencias temporales o efectos específicos de ciertos años que podrían estar influyendo en las tasas de donación, proporcionando un análisis más dinámico y realista de los datos.
3. **Estudio cualitativo:** investigar las políticas de donación en países que son outliers en el dataset: Los países que se identifican como outliers en el conjunto de datos, es decir, aquellos que presentan comportamientos atípicos en términos de sus tasas de donación, merecen un análisis más profundo. Se recomienda llevar a cabo un estudio cualitativo de las políticas de donación en estos países para comprender los factores detrás de estos comportamientos extremos. A menudo, los países con tasas de donación muy altas o bajas tienen políticas o contextos sociales únicos que influyen en los resultados. Al investigar estos factores cualitativos, se puede obtener una visión más completa y contextualizada que complementa los análisis cuantitativos y ofrezca recomendaciones más precisas para mejorar las políticas de donación a nivel global.

4. **Validar outliers:** revisar los países o años que presentan comportamientos atípicos para decidir si deben tratarse de manera especial: Es fundamental realizar una validación de outliers para determinar si los países o años que presentan comportamientos atípicos deben ser tratados de manera especial. No todos los outliers deben ser eliminados del análisis, ya que algunos pueden reflejar variaciones legítimas en las políticas o circunstancias que merecen una mayor atención. Se recomienda revisar estos puntos atípicos y analizar si son el resultado de factores excepcionales que deben ser considerados en el modelo, o si en cambio, son errores o datos inexactos que deberían ser excluidos. Esta validación asegurará que los outliers no distorsionen los resultados y ayudará a decidir cómo deben ser tratados en futuros análisis.

