



Máster en Estadística Aplicada para Data Science con R Software

Efectos de la descentralización política y la
actividad turística en la desigualdad
económica en Latinoamérica: estudio
longitudinal combinando técnicas de datos
de panel con aprendizaje no automatizado

AUTOR: Luis Eduardo Mella Gómez

DIRECTOR: Juan Luis López Garrancho

FECHA: 20/04/2024

ENTIDAD COLABORADORA



Resumen

En la literatura econométrica los datos de panel o muestras longitudinales de sección cruzada permiten el estudio de fenómenos complejos en el tiempo que afectan a múltiples unidades de análisis. Gran parte del armazón estadístico de estas técnicas intentan hacer un correcto modelamiento de las características individuales de las unidades de corte transversal en el tiempo (individuos, estados, condados, países, etc.) mediante dos estrategias conocidas: 1) dejar que el intercepto varíe para cada unidad de corte transversal mientras los demás parámetros se asumen como homogéneos (within group transformation) y 2) mientras que la otra estrategia implica tener un componente del error grupal, temporal o ambos además del error idiosincrático (random effects transformation¹).

Ambas estrategias de modelización para ajustar por la heterogeneidad no observable de los propios individuos tienen sus ventajas y desventajas: within effects o efectos fijos resulta en un estimador poco eficiente, aunque es consistente cuando $N \rightarrow \infty$, $T \rightarrow \infty$ y $NT \rightarrow \infty$, mientras que la transformación de efectos aleatorios es más eficiente y también consistente e insesgado en términos asintóticos. El principal obstáculo es que esta estrategia de modelización posee supuestos que son muy restrictivos (asume que los factores individuales no observados y las variables explicativas no están correlacionados). Intentar aproximar dicha heterogeneidad no observable es posible mediante el uso de técnicas de aprendizaje no automatizado al intentar caracterizar y explorar los datos con diferentes variables para generar clústeres.

El presente trabajo de fin de master pretende combinar herramientas de aprendizaje no automatizado con metodologías usadas en estudios econométricos de panel, en este caso, análisis de conglomerados por k-medias y k-mediods para intentar generar una categorización de esas características invariantes en el tiempo en un panel de 19 países desde 1995 a 2018, para estudiar el efecto que posee las capacidades institucionales de los gobiernos regionales (municipalidades, provincias, federaciones) de los países y la actividad turística en la desigualdad del ingreso, controlando por ciertos factores determinantes de la desigualdad del ingreso.

¹ Nota aclaratoria: la terminología econométrica difiere de la usada en modelos lineales generales o modelos mixtos, aunque usen similares conceptos. Efectos fijos en modelos mixtos refiere a parámetros que se asumen como constantes mientras que en econometría se refiere a controlar por factores no observables de las unidades de corte transversal (o within group effects). En econometría los efectos aleatorios refieren a modelar los componentes de la varianza mientras que en modelos mixtos los efectos aleatorios son parámetros que varían alrededor de 0 acorde a una distribución normal multivariante.

Todo el ejercicio se realizó con R Software y para ello se utilizó datos de los indicadores de desarrollo del Banco Mundial, el índice de Autoridad Regional (RAI), datos de desigualdad de la Base de datos mundial de desigualdad estandarizada de Solt (2020), así como indicadores de calidad institucional del Varieties of Democracy. Los datos sugieren, en consonancia con la literatura empírica de Latinoamérica, que existe una relación de Kuznets entre el desarrollo turístico y la desigualdad y que el desarrollo institucional de las regiones administrativas de los países parece tener poca relación o nula con la desigualdad del ingreso cuando se incluyen variables de control en el modelo, incluyendo las variables de clústeres generadas en los análisis de conglomerados efectuados, demostrando así como es posible hacer uso de modelos de efectos aleatorios y aprovechar su eficiencia en términos asintóticos y modelar los efectos grupales de tal manera para prevenir el incumplimiento del supuesto de exogeneidad estricta que posee esta técnica de estimación, mediante caracterizar parte de la heterogeneidad no observable de las unidades de corte transversal en el modelo, evitando la endogeneidad entre ese componente del error y las variables explicativas.

Agradecimientos

A mi tutor, Juan Luis López, por acompañarme en el proceso. Durante el master siempre recibí buena retroalimentación en los contenidos del master que hoy hacen posible este trabajo.

A la doctora Susana Ferrero por todo el material didáctico generado y por su consideración conmigo dada las situaciones personales que viví durante el master, pero que gracias al acompañamiento de Juan Luis pude superar.

A mi madre, Yodaska Lissette por apoyarme en todo el proceso.

A mi hermana de otra madre, María Victoria, por ser mi confidente y su amor incondicional.

A Patricia Rojas, por estar ahí para mí y escucharme y hacerme sentir validado y querido.

Índice de contenido	
Resumen	1
Agradecimientos	3
Introducción	9
Marco teórico y revisión de la literatura	10
Turismo y desigualdad	10
Posibles mecanismos.....	11
Instituciones deficientes y volátiles, propensas a la corrupción.....	11
Descentralización política y provisión de bienes públicos.....	12
Hipótesis	14
Material y métodos	14
Análisis de conglomerados	16
Correlación contemporánea y autocorrelación serial.....	18
Resultados	18
Descriptivos asociados al análisis de k-medias/k-medioides.....	18
Resultado de las regresiones	28
Caso de estudio: República Dominicana	33
Conclusiones	39
Referencias	41
Anexos	46
Código de R utilizado para el análisis del dato:	47

Índice de gráficos

<i>Ilustración 1 Matriz de correlación de las variables del Índice de Autoridad Regional (RAI). Fuente: Elaboración propia. Nota: x1= grado de desconcentración, x2= autonomía política, x3= autonomía fiscal, x4= capacidad de pedir préstamos, x5= capacidad representativa, x6= creación de piezas legislativas, x7= control en deliberaciones intergubernamentales, x8=Control fiscal sobre el gobierno nacional, x9=Control de endeudamiento sobre el gobierno nacional, x10=Control constitucional sobre el gobierno central.</i>	19
<i>Ilustración 2. Suma de cuadrados total intragrupo y número de clústeres óptimo. Fuente: Elaboración propia.</i>	20
<i>Ilustración 3. Gráfico de línea de estadístico Gap. Fuente. Elaboración propia.</i>	21
<i>Ilustración 4. Gráfico de dispersión de etiquetas, índice de Auto gobierno y índice de Gobierno compartido segregado por conglomerado. Fuente: Elaboración propia.</i>	22
<i>Ilustración 5. Gráfico de dispersión, Índice de Gini e índice de autoridad compartida (RAI). Fuente: Elaboración propia. Nota: El R^2 del es de un modelo que incluye la variable independiente más su término cuadrático</i>	22
<i>Ilustración 6. gráfico de dispersión, índice de Gini e índice de autogobierno. Fuente: Elaboración propia. Nota: El R^2 del es de un modelo que incluye la variable independiente más su término cuadrático y cúbico.</i>	23
<i>Ilustración 7. Gráfico de dispersión, Índice de Gini después de transferencias y Log(ingresos del turismo por artículos de viaje en USD). Fuente: Elaboración propia. Nota: El R^2 del es de un modelo que incluye la variable independiente más su término cuadrático.</i>	23
<i>Ilustración 8. Gráfico de dispersión, Log(ingresos del turismo por artículos de viaje en USD) por índice de Autoridad Regional. Fuente: Elaboración propia. Nota: El R^2 del es de un modelo que incluye la variable independiente más su término cuadrático. Los puntos del diagrama fueron ponderados por el índice de Gini post transferencias.</i>	24
<i>Ilustración 9. Gráfico de dispersión, Ingresos por artículos de viaje en USD e índice de autoridad compartida. Fuente: Elaboración propia. . Nota: El R^2 del es de un modelo que incluye la variable independiente más su término cuadrático. Los puntos del diagrama fueron ponderados por el índice de Gini post transferencias.</i>	24
<i>Ilustración 10. Gráfico de dispersión, Ingresos del turismo por artículos de viaje en USD e índice de Autogobierno ponderado por índice de Gini. Fuente: Elaboración propia.</i>	25

<i>Ilustración 11. Gráfico de dispersión, Log(PIB por persona empleada) e índice de corrupción pública (VDEM). Fuente: Elaboración propia. Nota: el diagrama fue ponderado por el índice de autogobierno del Regional Authority Index (RAI).</i>	<i>26</i>
<i>Ilustración 12. Gráfico de dispersión, Log(PIB por persona empleada) e índice de corrupción pública (VDEM). Fuente: Elaboración propia. Nota: el diagrama fue ponderado por el índice de gobierno compartido del Regional Authority Index (RAI)..</i>	<i>27</i>
<i>Ilustración 13. Gráfico de dispersión, Índice de componentes participativos (VDEM) e Índice de Autoridad Regional (RAI). Fuente: Elaboración propia. Nota: El diagrama fue ponderado por el índice de Gini post-transferencias.</i>	<i>27</i>
<i>Ilustración 14. Gráfico de dispersión, Índice de Gini después de transferencias e Índice de componentes participativos. Fuente: Elaboración propia. Nota: el diagrama fue ponderado por el Índice de Autoridad Regional (RAI).</i>	<i>28</i>
<i>Ilustración 15. Gráfico de dispersión, Predicción del índice de Gini por el Logaritmo de los ingresos del turismo por artículos de viaje en USD. Fuente: Elaboración propia. Nota: Ponderado por el índice de autoridad regional (RAI).</i>	<i>33</i>
<i>Ilustración 16. Gráfico de Dispersión, Índice de Palma y Log(Número de habitaciones para hospedaje), 2016-2022. Fuente: Elaboración propia con datos del Sistema de Indicadores Dominicano (SISDOM), datos administrativos del Ministerio de Turismo/ASONAHORES y del Sistema Municipal de la Administración Pública (SISMAP).....</i>	<i>34</i>
<i>Ilustración 17. Gráfico de Dispersión, Índice de Gini para hogares y Log(PIB per cápita (2010=100), 2016-2022. Fuente: Elaboración propia con datos del Sistema de Indicadores Dominicano (SISDOM) y Dirección General de Presupuesto (DIGEPRESS).</i>	<i>35</i>
<i>Ilustración 18. Gráfico de dispersión, Índice de Palma e Ingresos de los gobiernos locales en millones de RD\$ (2010=100), 2016-2022. Fuente: Elaboración propia con datos del Sistema de Indicadores Dominicano (SISDOM) y Dirección General de Presupuesto (DIGEPRESS).....</i>	<i>36</i>
<i>Ilustración 19. Gráfico de dispersión, Ratio entre Quintil 5 y 4 de mayores ingresos y egresos de los gobiernos locales en millones de RD\$ (2010=100), 2016-2022. Fuente: Elaboración propia con datos del Sistema de Indicadores Dominicano (SISDOM) y Dirección General de Presupuesto (DIGEPRESS).</i>	<i>37</i>
<i>Ilustración 20. Prueba de Breuch Pagan de Homocedasticidad para modelos de panel para Rep. Dom. Fuente: Elaboración propia.</i>	<i>46</i>

Ilustración 21. Tabla con resultados de prueba de Autocorrelación de Breuch-Godfrey/Wooldridge para datos en panel. Fuente: Elaboración propia. Nota: Se estimó para un componetne autorregresivo (AR) de orden 2. 46

Ilustración 22. Prueba de Hausman Robusta para modelos de panel de Rep. Dom. Fuente: Elaboración propia. 46

Índice de tablas

<i>Tabla 1. Recuento de valores extremos por variable. Fuente: Elaboración propia.</i>	<i>20</i>
<i>Tabla 2. Medidas de validación interna y externas de los análisis de conglomerados por k-medioídes. Fuente: Elaboración propia.</i>	<i>21</i>
<i>Tabla 3. Tabla de regresión. Fuente: Elaboración propia. Nota: Errores estándares robustos por Driscoll y Kraay (1998) entre paréntesis. Se incluyó dummies temporales. Log(Tur (1))= Gasto en turismo por motivos de transporte de artículos de pasajeros en USD.</i>	<i>29</i>
<i>Tabla 4. Tabla de regresión. Fuente: Elaboración propia. Nota: Errores estándares robustos por Driscoll y Kraay (1998) entre paréntesis. Se incluyó dummies temporales. Log(Tur(2))= Logaritmo de ingresos del turismo por artículos de viaje (USD).</i>	<i>30</i>
<i>Tabla 5. Tabla de regresión. Fuente: Elaboración propia. Nota: Errores estándares robustos por Driscoll y Kraay (1998) entre paréntesis. Se incluyó dummies temporales. Log(Tur(3))= Logaritmo de gastos de turismo internacional como % de las importaciones.</i>	<i>31</i>
<i>Tabla 6. Tabla de regresión. Fuente: Elaboración propia. Nota: Errores estándares robustos por Driscoll y Kraay (1998) entre paréntesis. Se incluyó dummies temporales. Log(Tur(4))= Logaritmo de gastos de turismo internacional (USD).</i>	<i>32</i>
<i>Tabla 7. Tabla de regresiones de panel para República Dominicana (2016-2022). Fuente: Elaboración propia. Nota: CR2= Corrección de Pustejovsky y Tipton, (2018).</i>	<i>38</i>
<i>Tabla 8. Prueba de Hausman Robusta. Fuente: Elaboración propia. Nota: CR2= Corrección de Pustejovsky & Tipton, (2018)</i>	<i>46</i>

Introducción

El turismo en la región de Latinoamérica es de los sectores económicos más populares debido a la capacidad de este sector de brindar crecimiento económico. Ahora bien, dado que Latinoamérica como región adolece de problemas de diseño institucional, se tiene cierto escepticismo de que los beneficios del turismo se repartan de manera eficiente, entiendo la eficiencia en términos de una distribución de ingresos Pareto-eficiente.

En la literatura empírica sobre turismo y desigualdad ha dado cuenta de que en casos de países en vías de desarrollo el turismo exacerba la desigualdad y apunta a diferentes causas: 1) corrupción; 2) Captura política del sector por parte de empresarios locales e internacionales; 3) Falta de capacidad estatal y 4) mercados oligopólicos, donde el turismo compite solo en precio y tiene una oferta homogénea (muy común en países en donde se fomenta modelos de turismo de sol y playa todo incluido). Dentro de la dimensión institucional se destacan posibles mecanismos. La provisión de bienes públicos es incapaz de transformar la estructura productiva del sector y de paso, mecanismos de revelación de preferencias constituidas por los gobiernos subnacionales o gobernaciones locales son incapaz de articular las demandas ciudadanas e incluir a la gente en el proceso generador de rentas del turismo, todo esto debido a la volatilidad institucional en la que la región está sumergida. Primeras aproximaciones fueron hechas por sobre que la hipótesis acerca de los problemas de diseño institucional en lo que refiere a la descentralización política era el mecanismo causal que por vía dos canales (provisión deficiente de bienes públicos e inadecuada inclusión de las localidades cercanas a la actividad turística) explica la relación entre ambas.

Una de las aproximaciones más recientes hechas sobre este problema es de Mella Gómez (2022). A través de datos de paneles robustos a sesgos de paneles dinámicos arrojó que en el corto plazo la falta de rendición de cuentas y el ingreso del turismo tienen un efecto negativo en la desigualdad mientras que en el largo plazo su efecto no es distinto de cero. Una de las limitaciones de esta investigación es que la especificación econométrica utilizada para capturar la capacidad del Gobierno de involucrar a la ciudadanía en asuntos relacionados a la política del turismo no aproxima bien el efecto de la descentralización política o de las capacidades de los gobiernos subnacionales. Esta investigación aporta a la literatura sobre desigualdad turismo e instituciones En Latinoamérica mediante una forma de aproximar el efecto de estas dos variables: las capacidades de los gobiernos locales y el desarrollo turístico en la desigualdad. Partiendo de la literatura empírica sobre institucionalismo, turismo y desigualdad, y combinando las técnicas de aprendizaje no automatizado con el estudio econométrico

de panel o de análisis de series de tiempo de cepción cruzada, se estima el efecto de las variables de interés en la desigualdad.

El artículo se organiza de la siguiente manera: una primera sección detalla el marco teórico y la literatura empírica consultada acerca de la relación entre instituciones, gobiernos locales y/o descentralización política, por un lado, el desarrollo turístico respecto de la desigualdad con énfasis en la región. La siguiente sección detalla la estrategia empírica utilizada, la muestra del estudio, fuentes de datos empleadas en el estudio, así como las técnicas y herramientas estadísticas a utilizar. Luego se detallarán los resultados tanto descriptivos como econométricos, comparando modelos tradicionales de datos en panel con la estrategia empírica detallada, así como un análisis de un caso de estudio de un país, República Dominicana, la cual es un caso que resulta adecuado debido a que es uno de los países que ha experimentado un crecimiento económico sostenido gracias a la expansión del turismo, divisas y zonas francas, un crecimiento por encima de la media de la región. Paradójicamente, al mirar sus indicadores institucionales y de desarrollo humano, así como sus capacidades institucionales en materia de sus gobiernos subnacionales, adolece de problemas y desigualdades estructurales. Por último, se detallan las conclusiones del estudio y recomendaciones.

Marco teórico y revisión de la literatura

Turismo y desigualdad

El sector turístico, a pesar de ser un motor clave para el crecimiento económico y social, presenta una relación compleja con la desigualdad. Esta relación varía según el contexto del país, donde en algunas situaciones la industria turística inicialmente puede generar desigualdad, para luego beneficiar a la sociedad en su conjunto, mientras que en otros casos, la brecha de desigualdad se amplía con el tiempo, siendo en etapas iniciales al fomento del turismo este contribuía reduciendo la desigualdad (Al Tal et al., 2021; Alam y Paramati, 2016; Ghosh y Mitra, 2021). Existen investigaciones que muestran cómo el turismo puede reducir la desigualdad en naciones con diferentes niveles de ingreso (Nguyen et al., 2020), aunque hay evidencias de que, en ciertos contextos, especialmente en Sudamérica, el turismo puede agravar las desigualdades estructurales (Büscher y Fletcher, 2016). Además, estudios adicionales destacan un vínculo negativo entre el turismo y la equidad en Latinoamérica (Fang et al., 2020), y se ha identificado que el turismo internacional podría incrementar la desigualdad de género (Nguyen, 2022).

En el caso de la República Dominicana, las perspectivas sobre el impacto del turismo difieren. Torres y Corporán (2021) sugieren que, en la República Dominicana, el modelo de turismo basado en resorts de “sol y playa” todo incluido contribuye a la persistencia de la pobreza, a pesar de la expansión económica generada por el sector a través del flujo de divisas que genera al país. Contrariamente, hay estudios que ofrecen una visión más optimista sobre el turismo, destacando sus ventajas competitivas y comparativas, como los trabajos de Acosta Guzmán (2017) y Caro-González et al. (2015), el primero enfocándose en el impacto positivo de los cruceros en la economía local sin abordar los aspectos redistributivos, y el segundo resaltando el potencial turístico debido a los recursos culturales y naturales del país, aunque reconociendo falencias en la planificación sectorial.

Sin embargo, como apuntan Alvarado y Oddone (2017), el crecimiento económico derivado del turismo no siempre se traduce en beneficios sociales o ambientales. El enfoque predominante en el turismo de “sol y playa”, caracterizado por grandes complejos hoteleros y una oferta homogénea, lleva a una competencia basada en precios más que en calidad. Esto ha situado a la República Dominicana lejos de alcanzar su pleno potencial en el sector a nivel mundial (Hernández R et al., 2019), y frecuentemente excluye a las comunidades locales de los beneficios (Ceara-Hatton (2005); Contreras (2011); Ceara-Hatton et al., 2008; Contreras, 2011). A pesar de la generación de divisas, esta dinámica ha resultado en desigualdades territoriales (Hernández R et al., 2019; Solano Cabrera, 2020) y, aunque contribuye al crecimiento económico (Ceara-Hatton, 2015), los beneficios están limitados por una distribución desigual de los mismos.

Posibles mecanismos

Instituciones deficientes y volátiles, propensas a la corrupción

Para comprender el proceso de formación institucional en América Latina hay que comprender primero qué son las instituciones y sus dimensiones. Para Douglass North, las instituciones son un conjunto de reglas o normas que se construyen socialmente y permiten a los individuos tomar decisiones estratégicas en contextos donde hay incertidumbre (North, 1990). Las instituciones pueden a grandes rasgos dimensionarse en dos según Dryzek (1996):

1. Instituciones formales: Aquellas que son reglas escritas y estructuradas por instituciones como el Estado, el cuerpo normativo, el régimen político.

2. Instituciones informales: Aquellas constituidas por normas no escritas, por formas de socialización y la cultura.

Las instituciones es uno de los factores más importantes en el desarrollo económico, en la reducción de desigualdades políticas, económicas y sociales (Acemoglu et al., 2014; Chong y Gradstein, 2007; Levitsky y Murillo, 2010; Mahoney et al., 2010; Przeworski, 1988; Thelen, 2009; Zhao, 2020). Ahora bien, la situación particular de Latinoamérica entrevé la existencia de instituciones débiles, con alta volatilidad y poco cumplimiento normativo (enforcement) Levitsky y Murillo (2012); Levitsky y Murillo (2005); Levitsky y Murillo (2010). Aún en las transiciones democráticas en la segunda mitad del siglo XX acontecidas en la región, como sugiere O'Donnell (2004), dio como resultados democracias delegativas que si bien poseen todos los elementos formales de una poliarquía en sentido de Robert Dahl, es decir, participación inclusiva, libertades civiles, oposición política competitiva, elecciones frecuentes, libres y justas (mecanismos de rendición de cuentas vertical), que los ciudadanos poseen control sobre las políticas públicas (Dahl, 2008, 2008), según O'Donnell, las democracias delegativas poseen vicios institucionales de las dictaduras pasadas, por lo que hay cierto nivel de discreción de parte del poder ejecutivo y que ciertos mecanismos que limitan su poder actúan de manera intermitente (lo que O'Donnell llama rendición de cuentas horizontal, es decir, que en las democracias delegativas hay una ausencia o una intermitencia en el funcionamiento de estos mecanismos). Estos factores explican por qué la corrupción (con todo y su heterogeneidad, ver Mella Gómez y Prestol Flaquer (2023)) es una práctica incrustada en la cultura política y en la administración pública y el por qué es una región muy pero muy desigual, con muy poca movilidad intergeneracional pero que se han notado indicios de reducciones significativas desde los 2000 (Birdsall et al., 2013; Gasparini y Lustig, 2011; Hoffman y Centeno, 2003; Torche, 2014).

Descentralización política y provisión de bienes públicos

La descentralización política en la región de Latinoamérica ha sufrido varios cambios importantes. Primeramente, tal como expone Ivan Finot (2001, 2002, 2005, 2007), el modelo centralista regional, el cual sirvió para la planificación y provisión de bienes públicos en los territorios se agotó especialmente tras las crisis económicas de 1980. Este periodo coincide con los procesos de transición democrática pos dictadura evidenciados por O'Donnell (2004). Tras la adopción de las reformas del consenso de Washington la región empezó un proceso de cambio institucional, la cual consistió en

términos generales en una transición del modelo centralista regional a un modelo más descentralizado (Finot, 2007). Ahora bien, varios problemas fueron presentándose. Primero las ventajas conocidas de la descentralización, las cuales son reducir costes de transacción en la provisión de bienes públicos (tanto sociales como territoriales), generar mayor gobernabilidad democrática al incentivar la participación política y ciudadana, aliviar malestares producto de los ajustes macro económicos que vivió la región durante 1980 y reducir las desigualdades territoriales. Estos resultados esperados fueron muy mixtos en la región (Finot, 2007).

Además, la descentralización efectuada en la región durante la época del Consenso de Washington tuvo varios problemas respecto de la forma en la que se ejecutó. Según la teoría de la descentralización política (Boettke y Marciano, 2016; Wallace E. Oates, 1968; Wallace E. Oates, 2008; Tiebout, 1956; Wallis y Oates, 1988) y teorema de descentralización de Wallace Oates hasta sus versiones modernas, la descentralización puede hacerse en tres aristas o dimensiones según Finot (2007):

1. Descentralización operativa o administrativa: las estructuras necesarias para ejecutar las decisiones políticas de provisión de bienes públicos.
2. Descentralización del gasto: En donde el gasto del gobierno central se mueve a las distintas unidades subnacionales.
3. Descentralización fiscal (o impositiva): Las unidades subnacionales ahora pueden cobrar ciertos tributos para invertirlos en el proceso de provisión (y producción) de bienes públicos.

En el caso específico de la región la descentralización fue solo en términos operativos y del gasto. Este tipo de arreglos no solo no genera peores resultados en cuanto a la reducción de disparidades territoriales, sino que tiene como resultado que fomenta el uso de los gobiernos locales para fines clientelares, así como mayor incidencia de corrupción (Finot, 2002). Una ventaja de la descentralización en la región ha sido la de aumentar la cobertura geográfica de las ayudas sociales, pero dejando por un lado unidades subnacionales con pocas capacidades institucionales (Finot, 2005).

En República Dominicana no fue la excepción. A pesar de las transiciones y los cambios en la legislación referente a las unidades subnacionales desde el 1990 a la actualidad, República Dominicana es uno de los países con el mayor grado de atomización municipal de la región, solo rivalizando con El Salvador (Martínez-Vázquez et al., 2017). Parte del problema ha sido que la descentralización solo ha sido en una dimensión del gasto y a nivel administrativo, pero hay pocas competencias a nivel fiscal y la propia normativa incentiva a los municipios a escalar a un nivel mayor respecto de la jerarquía de la división territorial. Esto es, pasar de municipio a distrito municipal o de distrito a provincia, para así obtener más recursos económicos. A pesar de que la ley de municipios establece un mínimo de población para elevar de categoría un municipio, no solo la ley se incumple en mucho de los casos, sino que esto trae como consecuencia la creación de más burocracia y de territorios sin mecanismos institucionales que garanticen procesos de desarrollo endógenos (Prestol Flaquer, 2022).

Hipótesis

H1: El turismo incide de manera positiva en la desigualdad.

H2: La actividad turística incide de manera negativa en la desigualdad.

H3: Las capacidades de los gobiernos locales no tienen un efecto significativo en la reducción de la desigualdad.

H4: La participación política tiene un efecto significativo en la reducción de la desigualdad.

Material y métodos

Esta investigación hará uso de dos fuentes de datos longitudinales (o de panel, o lo que llaman “series de tiempo de corte transversal”) para evaluar el efecto de las dos variables de interés (el turismo y las capacidades de las instituciones descentralizadas como gobiernos subnacionales, gobiernos locales, etc. La primera estrategia estudia el efecto a nivel de 19 países de Latinoamérica desde 1995 a 2018, cubriendo un periodo de 24 años. Esta investigación hace uso de datos de la Standardized World Inequality Database (Solt, 2020), para datos de PIB, población, otros datos demográficos y datos sobre el impacto del turismo en la nación se utilizó los Indicadores de Desarrollo del Banco Mundial. Para las variables de control institucional se utilizó un conjunto de indicadores del Varieties of Democracy (de ahora en adelante VDEM) (Coppedge et al., 2019). Por último, para aproximar las capacidades institucionales de los gobiernos

subnacionales o de aproximar su grado de descentralización política, se utilizó el Regional Authority Index (RAI) (HOOGHE et al., 2021).

Para el primer modelo partimos de un modelo de datos de panel:

$$Y_{it} = \alpha_i + \gamma Tur_{it} + \phi Desc_{it} + \omega Inst_{it} + \beta X_{it} + \epsilon_{it} \quad (1)$$

En donde Y_{it} es el índice de Gini post transferencias del gobierno, α_i es la heterogeneidad no observable del modelo, γTur_{it} es el vector que contiene los indicadores de turismo utilizados en el modelo. $\phi Desc_{it}$ refiere al vector de indicadores de descentralización política, $\omega Inst_{it}$ es el vector de variables institucionales elegidas del VDEM, βX_{it} es un vector que contiene todas las variables de control y ϵ_{it} es el error idiosincrático.

Por otro lado, para el caso de estudio se tomaron datos del Sistema de Indicadores Sociales Dominicano, así como datos poblacionales de la Oficina Nacional de Estadística (ONE), de ejecución presupuestaria de los gobiernos locales de la Dirección General de Presupuesto (DIGEPRES). Se compilaron datos del número de habitaciones como una variable proxy del desarrollo turístico. Algunas variables estaban a nivel de provincias y para fines de comparabilidad se agruparon en las 10 regiones administrativas según la ley 345-22 de regiones únicas de Planificación de la República Dominicana durante el periodo de 2016-2022. La ecuación es la siguiente:

$$Y_{it} = \alpha_i + \gamma Ln(Tur)_{it} + \phi Ln(PIB)_{it} + \theta Z_{it} + \beta X_{it} + \epsilon_{it} \quad (2)$$

En donde Y_{it} es la variable dependiente que representa diferentes indicadores de desigualdad (Índice de Gini, Índice de Palma y Ratio Quintil 5/Quintil 4, Quintil 4/Quintil 3, Quintil 3/Quintil 2, Quintil 2/Quintil 1), $\gamma Ln(Tur)_{it}$ es el logaritmo del número de habitaciones recopiladas por el Ministerio de Turismo y publicadas por la ONE, $\phi Ln(PIB)_{it}$ es el logaritmo del PIB per cápita regional deflactado (año de referencia 2010), θZ_{it} es un vector que contiene dos variables de carácter institucional: el desempeño administrativo de los ayuntamientos a nivel regional, recopilados por el Sistema de Monitoreo de la Administración Pública Municipal (SISMAP). Como los datos son extraídos a nivel de provincias (32) se calculó el promedio regional de las regiones

de planificación para representar el desempeño regional de estos desde 2016-2022. La otra variable del vector es la ratio de ingresos y egresos deflactado de los gobiernos locales para aproximar las capacidades fiscales de los gobiernos locales, datos recopilados por la Dirección General de Presupuesto (DIGEPRES) y compilados por la ONE. A su vez, dentro del vector $X\beta_{it}$ se encuentra una serie de variables de control normalmente usada en especificaciones econométricas sobre turismo y desigualdad, como lo son la tasa de desempleo abierto, el Logaritmo del total de la población urbana. Por último, ϵ_{it} es el error idiosincrático.

Análisis de conglomerados

Para mejorar la especificación del modelo de panel se necesita intentar controlar adecuadamente la heterogeneidad no observada (α_i) ya que este elemento hace inconsistente las estimaciones por mínimos cuadrado. Para corregir este problema, en la literatura convencional econométrica, se hace uso de técnicas convencionales de panel, las cuales son:

1. Estimación por efectos fijos.
2. Estimación por efectos aleatorios.

La estimación por efectos fijos implica controlar por las características individuales de las i observaciones. Analíticamente se ilustra con esta transformación:

$$(Y_{it} - \bar{Y}_i) = (\alpha_i - \bar{\alpha}_i) + \beta(X_{it} - \bar{X}_i) + (\epsilon_{it} - \bar{\epsilon}_i) \quad (3)$$

Esta transformación es el equivalente a colocar una variable dummy por cada observación i . Si bien asintóticamente esta transformación es consistente e insesgada conforme $N \rightarrow \infty$ y $T \rightarrow \infty$, es un estimador poco eficiente (y/o menos precisa) en comparación con efectos aleatorios.

Si cada una de las variables a transformar las incluimos dentro de un vector M en donde representamos cada variable con el sufijo j podemos decir que la transformación intra grupo (within transformation) requiere de restarle a las variables su respectiva media grupal, la cual podemos representar como:

$$\bar{M}_j = \sum_{i=1}^N M_j \quad (4)$$

En donde \bar{M}_j es el vector que contiene las medias agrupadas de las variables de la ecuación 1.

Efectos aleatorios (o lo que en modelos lineales generalizados es conocido como variance-component model o random intercept model de dos niveles). La diferencia en el caso de la literatura econométrica es que este tipo de modelos se estiman no por máxima verosimilitud restringida, sino por Mínimos Cuadrados Generalizados Factibles (FGLS)².

Una forma de definir la ecuación de un modelo de efectos aleatorios es:

$$(Y_{it} - \lambda \bar{Y}_i) = (\alpha_i - \lambda \bar{\alpha}_i) + \beta(X_{it} - \lambda \bar{X}_i) + (\epsilon_{it} - \lambda \bar{\epsilon}_i) \quad (5)$$

En donde λ^3 es:

$$\lambda = 1 - \left(\frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + \sigma_\alpha^2} \right)^{\frac{1}{2}} \quad (6)$$

Este es un parámetro que es desconocido pero que se estima usando los residuos de una primera ecuación estimada por MCO agrupado y se introduce en la ecuación de la transformación de efectos fijos.

Uno de los supuestos fundamentales de la estimación de efectos aleatorios es que no existe una relación entre α_i y las variables explicativas ($Cov(\alpha_i, X_{it}\beta) = 0$). En contextos en donde se incumple este hecho, efectos fijos es consistente mientras que la transformación de efectos aleatorios es inconsistente y sesgada ya que existe problemas de endogeneidad.

Por otro lado, el aprendizaje no automatizado puede utilizarse para caracterizar una serie de datos y es posible agruparlos adecuadamente en base a ciertas características propias de la data. En estudios longitudinales o de datos en panel, el uso del análisis de

² Tanto los comandos xtreg, mle (si se estima por máxima verosimilitud restringida), xtreg, re (por fgls), como xtmixed, reml son equivalentes en Stata, así como en R los comandos lmer como plm, model="random" de los paquetes lme4 y plm respectivamente.

³ Cuando $\lambda = 0$ efectos aleatorios es igual a mínimos cuadrados agrupados (MCOA), mientras que si $\lambda = 1$ efectos aleatorios es igual a efectos fijos.

conglomerados es una forma de como caracterizar o aproximar la heterogeneidad no observada, así como de generar una interpretación más sencilla de dicha heterogeneidad (ejemplo en LOMBARDO et al. (2011). Para capturar la heterogeneidad no observada de los diferentes regímenes subnacionales para el primer modelo que compara datos de países de Latinoamérica se utilizó el análisis de conglomerados usando como variables los indicadores que conforman el Regional Authority Index (RAI) de HOOGHE et al. (2021). Estas variables permitirán construir una serie de clústeres para así clasificar a los países y utilizarse en el modelo de regresión por lo que $\alpha_i \approx \sum_{k=1}^N \delta_k Cluster_{k-1}$, es decir, intentaremos aproximar la heterogeneidad no observable mediante los conglomerados resultantes y aplicarla a modelos estimados por efectos aleatorios para lograr obtener una estimación consistente y eficiente. En un inicio se utilizó el algoritmo de k-medias, pero ante la presencia de valores atípicos, se re-estimó por medio del algoritmo de k-medioides. Para el modelo utilizado para el caso país (República Dominicana), dado que las pruebas de especificación de exogeneidad estricta (la prueba de Hausman robusta) arrojó que no existe correlación entre la α y $X\beta$, no se procedió a hacer uso de un análisis de conglomerados para modelizar la heterogeneidad no observada a nivel regional para el caso de República Dominicana.

Correlación contemporánea y autocorrelación serial

Dado que los datos presentan correlación contemporánea, heterocedasticidad y autocorrelación serial, se iteraron diferentes modelos con estimaciones de la varianza distinta, como lo son la de Beck y Katz (1995) y Driscoll y Kraay (1998), que permiten ajustar por problemas de heterocedasticidad, autocorrelación serial y contemporánea, corrigiendo los errores estándares⁴. Para el caso del modelo utilizado en República Dominicana, dado que pruebas de heterocedasticidad y de autocorrelación serial arrojan evidencia de este tipo de perturbaciones, se estimaron los modelos con errores estándares robustos para muestras pequeñas de Pustejovsky y Tipton, (2018)

Resultados

Descriptivos asociados al análisis de k-medias/k-medioides

Las variables de '*autogobierno*' (x1 a x5) y de '*gobierno compartido*' (x6 a x10) representan diferentes aspectos de la autonomía política y fiscal regional. Las

⁴ Dado que la estimación de Driscoll y Kraay es más eficiente y robusta que la de Beck y Katz bajo el supuesto de autocorrelación, se prefirió esta ya que la de Beck y Katz se necesita corregir la autocorrelación aparte, ya sea incluyendo una variable dependiente rezagada en el modelo o usando transformación de Prais-Winsten al modelo.

correlaciones más robustas se observaron dentro de la dimensión de 'autogobierno': el grado de desconcentración mostró una fuerte correlación positiva tanto con la autonomía política ($r = 0.702$, $p < 0.01$) como con la autonomía fiscal ($r = 0.928$, $p < 0.01$). Además, la autonomía política también se correlacionó fuertemente con la autonomía fiscal ($r = 0.839$, $p < 0.01$). Estos resultados indican una convergencia significativa dentro de los elementos que definen el autogobierno.

Por otro lado, las correlaciones entre las variables de autogobierno y gobierno compartido tendieron a ser más débiles, con la mayoría de ellas por debajo del umbral de 0.50. Esto sugiere que, aunque relacionadas conceptualmente, las dimensiones de 'autogobierno' y 'gobierno compartido' pueden operar de manera más independiente en la práctica. En conjunto, estos hallazgos delimitan un panorama donde la cohesión interna del autogobierno es más pronunciada que la sinergia entre el autogobierno y el gobierno compartido. Estas relaciones matizadas deben ser un foco clave para comprender las complejidades de la autonomía política y fiscal a nivel subnacional.

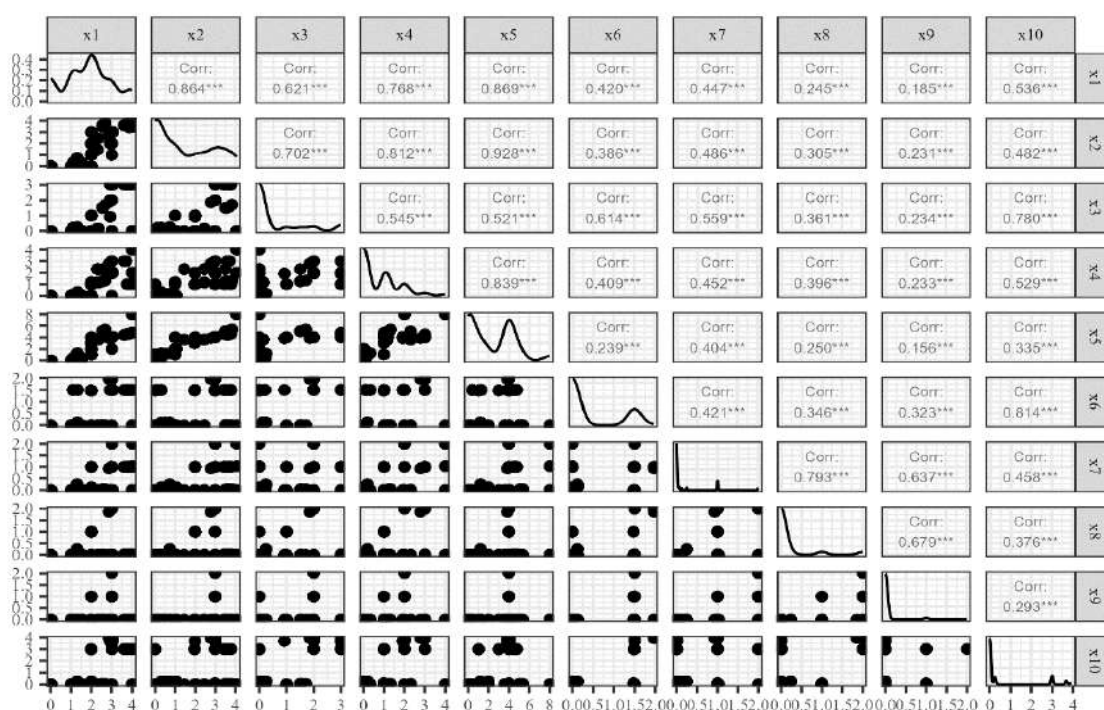


Ilustración 1 Matriz de correlación de las variables del Índice de Autoridad Regional (RAI). Fuente: Elaboración propia. Nota: x1= grado de desconcentración, x2= autonomía política, x3= autonomía fiscal, x4= capacidad de pedir préstamos, x5= capacidad representativa, x6= creación de piezas legislativas, x7= control en deliberaciones intergubernamentales, x8=Control fiscal sobre el gobierno nacional, x9=Control de endeudamiento sobre el gobierno nacional, x10=Control constitucional sobre el gobierno central.

Si bien se condujeron una serie de test que determinan que los datos están agrupados o por conglomerados, al proceder a realizar un análisis de k-medias, si bien, la validez externa del mismo y la silueta fueron aceptables, existen muchos outliers que resultan ser valores extremos. En la tabla 1 podemos observar el recuento de los valores

extremos por cada variable en la que se encontró valores extremos. También los resultados del estadístico de Hopkins arrojaron un 0.10 por lo que la data se agrupa en conglomerados.

Variable	FALSE	TRUE
x10	0	87
x3	39	0
x4	19	0
x7	16	77
x8	0	89
x9	0	18

Tabla 1. Recuento de valores extremos por variable. Fuente: Elaboración propia.

En la ilustración 2 se visualiza la suma de cuadrados total intragrupo estimado por el método del mismo nombre para ver cuál es el número de conglomerados más adecuado para el ejercicio. Esta prueba indica que las reducciones marginales más altas de la suma de cuadrados se encuentran entre 3 y 4 clústeres.

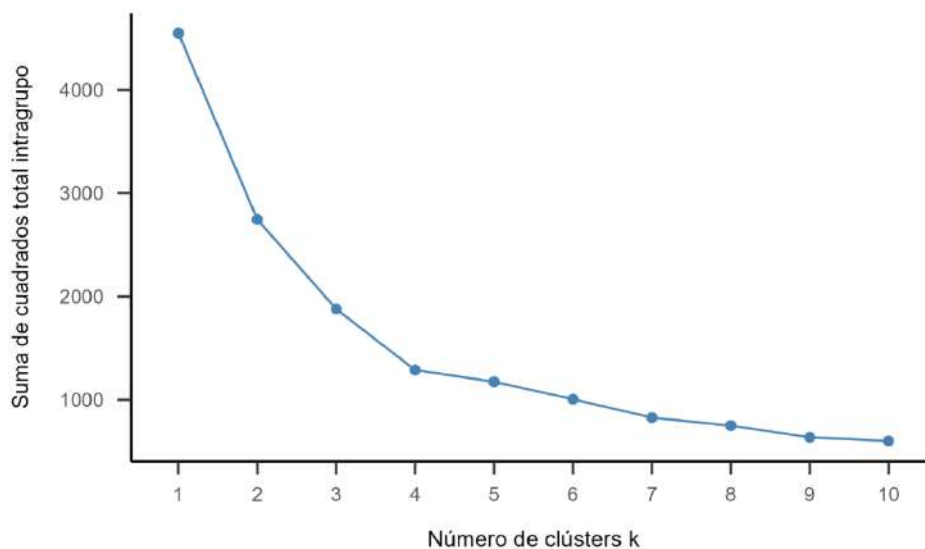


Ilustración 2. Suma de cuadrados total intragrupo y número de clústeres óptimo. Fuente: Elaboración propia.

Por otro lado, la prueba de estimar el número de clústeres basada en el estadístico Gap arroja que el número óptimo de clústeres es de 10.

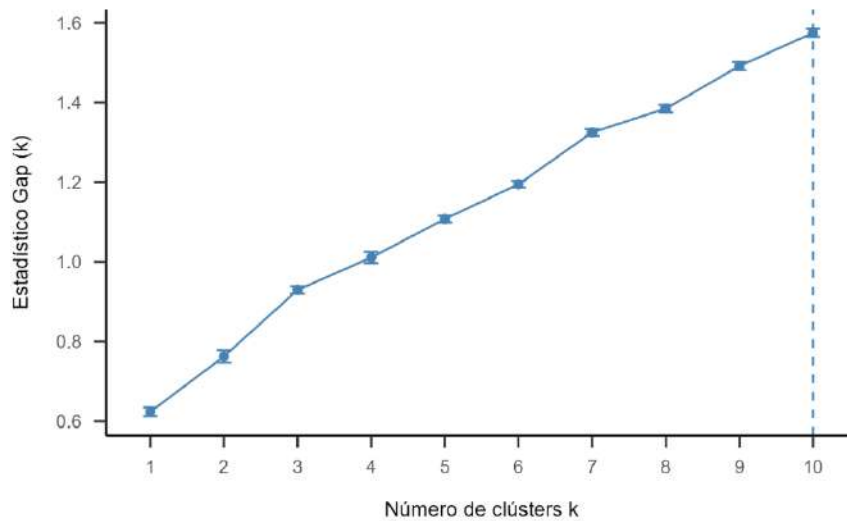


Ilustración 3. Gráfico de línea de estadístico Gap. Fuente. Elaboración propia.

Al efectuar los cálculos de los clústeres, se procedió a estimar las medidas de validación interna y externa de los conglomerados. Como variable de validación externa, se creó una dummy basada en los valores mínimos, medios y máximos del puntaje de las dimensiones de autogobierno y gobierno compartido (shared rule). En la tabla 2 muestra que el segundo análisis resultante, con 3 conglomerados, si bien tuvo una silueta promedio menor al resto, tiene mayor validez externa con un índice de Rand corregido de 0.72.

	Conglomerados	Silueta promedio	Índice de Rand
#1	4	0.58	0.70
#2	3	0.55	0.72
#3	10	0.62	0.58

Tabla 2. Medidas de validación interna y externas de los análisis de conglomerados por k-medioídes. Fuente: Elaboración propia.

En el gráfico 3 se observó que algunos países tienen un alto grado de autogobierno, pero un bajo índice de gobierno compartido, mientras que otros tienen medidas más equilibradas en ambos índices. Los países con altos valores en ambas dimensiones se encuentran más arriba y a la derecha del gráfico. Países con alto nivel de autogobierno regional y con un gobierno compartido con el gobierno nacional es Argentina.

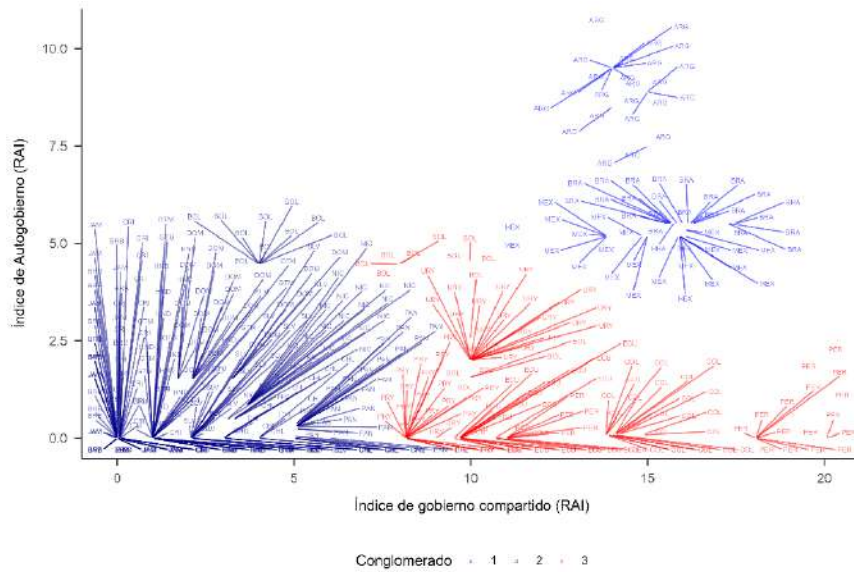


Ilustración 4. Gráfico de dispersión de etiquetas, índice de Auto gobierno y índice de Gobierno compartido segregado por conglomerado. Fuente: Elaboración propia.

A su vez, al examinar la relación entre desigualdad y el grado de autoridad compartida entre los gobiernos regionales y el gobierno nacional, existe no linealidad entre ambas variables y una asociación débil, pero denota que, a partir de cierto umbral, el grado de autoridad compartida con el gobierno nacional tiene un efecto en reducir la desigualdad.

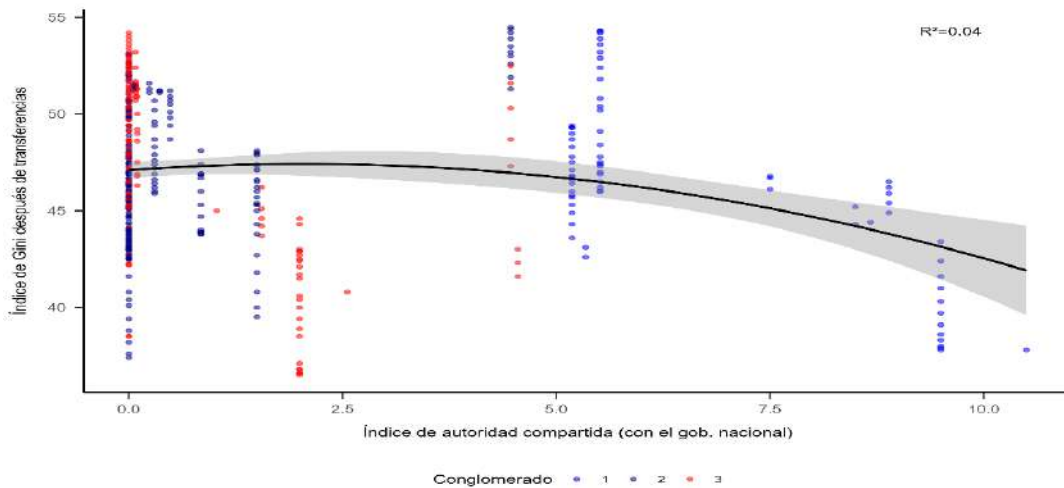


Ilustración 5. Gráfico de dispersión, Índice de Gini e índice de autoridad compartida (RAI). Fuente: Elaboración propia. Nota: El R^2 del es de un modelo que incluye la variable independiente más su término cuadrático

En la ilustración 6 se observa que existe una relación un tanto en forma de S (spline) pero muy débil ($R^2 = 0.04$) entre al grado de autogobierno de las regiones administrativas de los países analizados y la desigualdad.

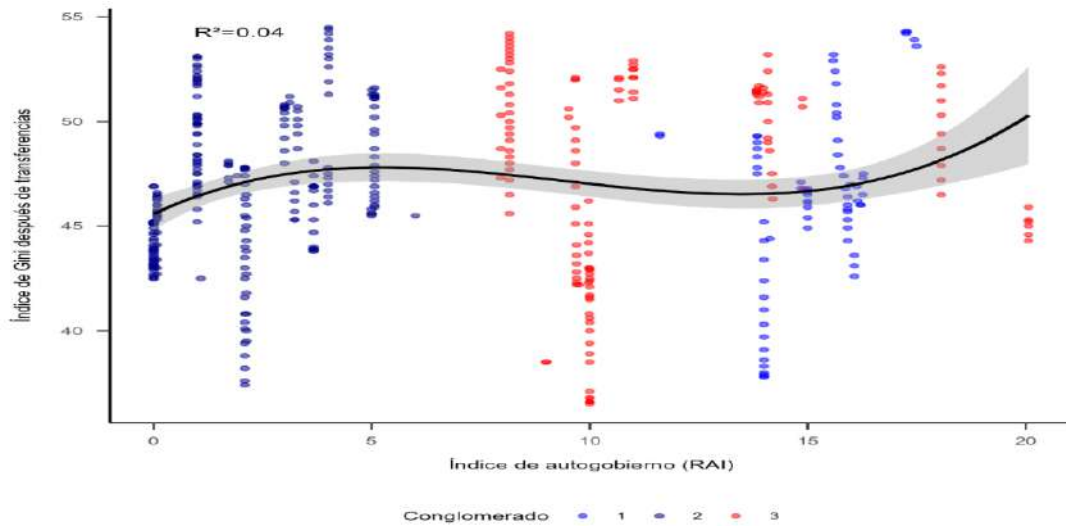


Ilustración 6. gráfico de dispersión, índice de Gini e índice de autogobierno. Fuente: Elaboración propia. Nota: El R^2 del es de un modelo que incluye la variable independiente más su término cuadrático y cúbico.

Ahora bien, en la ilustración 7 observamos un diagrama de dispersión que relaciona el índice de Gini después de transferencias y los ingresos del turismo producto de la venta de artículos de viaje en dólares. Se observa una relación negativa, pero forma de curva de Kuznets invertida, es decir que la desigualdad se reduce, pero a partir de un umbral, la desigualdad empieza a aumentar.

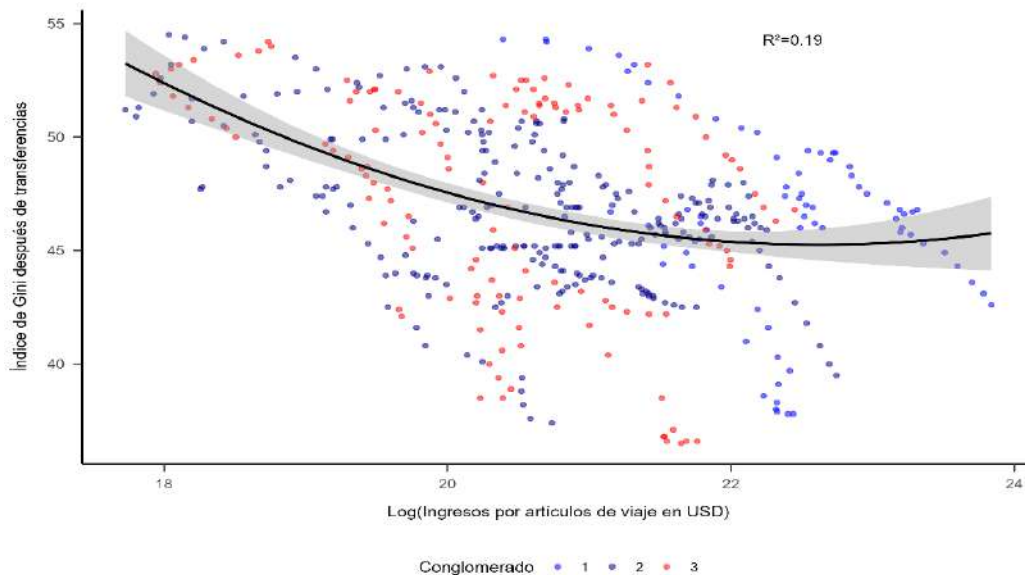


Ilustración 7. Gráfico de dispersión, Índice de Gini después de transferencias y Log(ingresos del turismo por artículos de viaje en USD). Fuente: Elaboración propia. Nota: El R^2 del es de un modelo que incluye la variable independiente más su término cuadrático.

El gráfico 8 analiza la relación entre el nivel de ingresos generados por artículos de viaje, medidos en logaritmos de dólares estadounidenses, y el Índice de Autoridad Regional. Un R^2 de 0.32 indica una correlación significativa, sugiriendo que un incremento en la autonomía regional puede estar asociado con mayores ingresos del turismo.

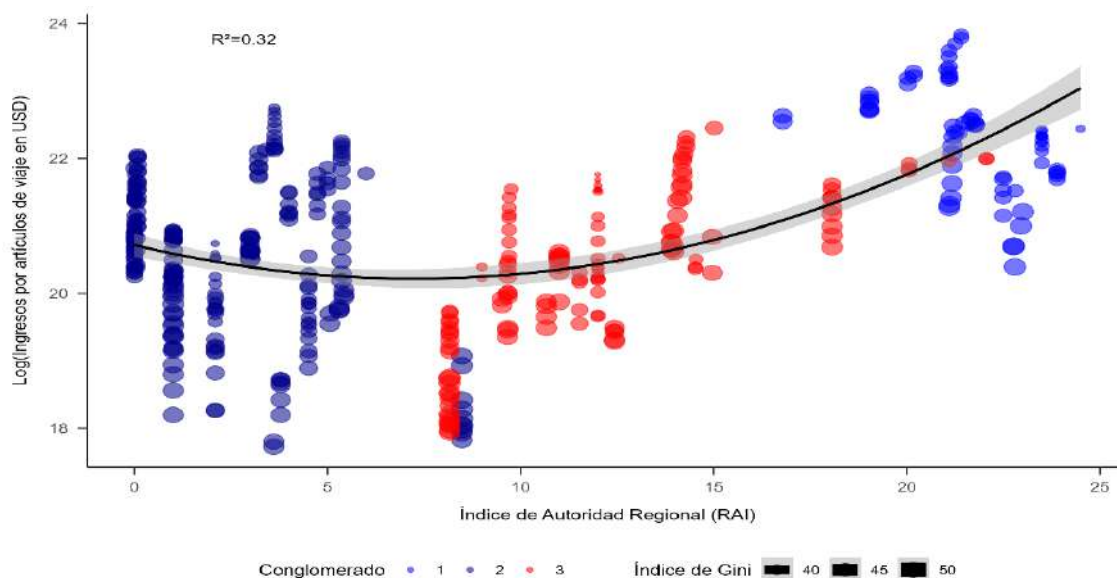


Ilustración 8. Gráfico de dispersión, Log(Ingresos del turismo por artículos de viaje en USD) por índice de Autoridad Regional. Fuente: Elaboración propia. Nota: El R^2 del es de un modelo que incluye la variable independiente más su término cuadrático. Los puntos del diagrama fueron ponderados por el índice de Gini post transferencias.

El gráfico 9 presenta la relación entre el Índice de Autoridad Compartida y los ingresos por turismo, con un R^2 de 0.17, sugiriendo una asociación positiva; a mayor autoridad compartida, tienden a incrementarse moderadamente los ingresos por turismo.

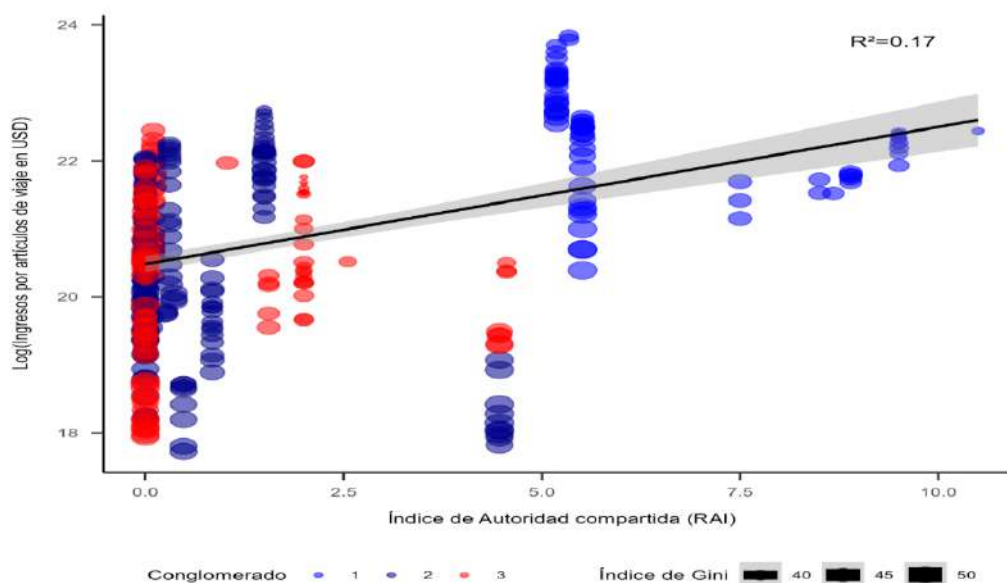


Ilustración 9. Gráfico de dispersión, Ingresos por artículos de viaje en USD e índice de autoridad compartida. Fuente: Elaboración propia. . Nota: El R^2 del es de un modelo que incluye la variable

independiente más su término cuadrático. Los puntos del diagrama fueron ponderados por el índice de Gini post transferencias.

El gráfico 10 sugiere una correlación positiva moderada entre la autoridad compartida y el ingreso turístico, con un R^2 de 0.17. La desigualdad de ingresos, representada por el índice de Gini, no muestra una tendencia clara en relación con la actividad turística condicionada por el grado de autogobierno.

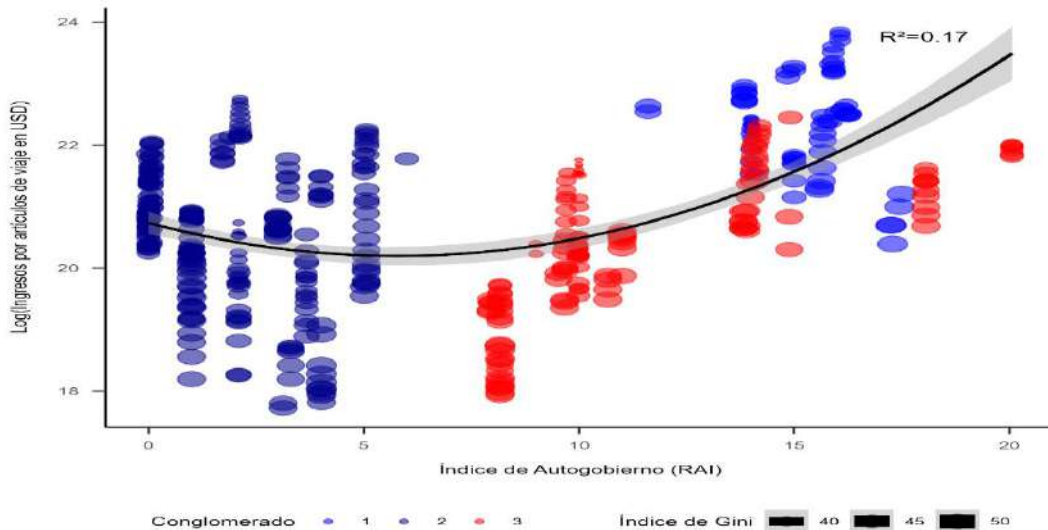


Ilustración 10. Gráfico de dispersión, Ingresos del turismo por artículos de viaje en USD e índice de Autogobierno ponderado por índice de Gini. Fuente: Elaboración propia.

Por otro lado, al graficar la relación entre la productividad (aproximada como el PIB por cada persona empleada en la economía) por el índice de corrupción política del V-DEM, arroja que existe una relación inversa y fuerte ($R^2 = 0.33$). Los países que están por encima de la línea de ajuste tienden a tener mayor autoridad compartida y poseen niveles bajos y medios de corrupción política, mientras que aquellos con poca autoridad regional propia se encuentran en las zonas de menor crecimiento económico y por debajo de la recta de regresión.

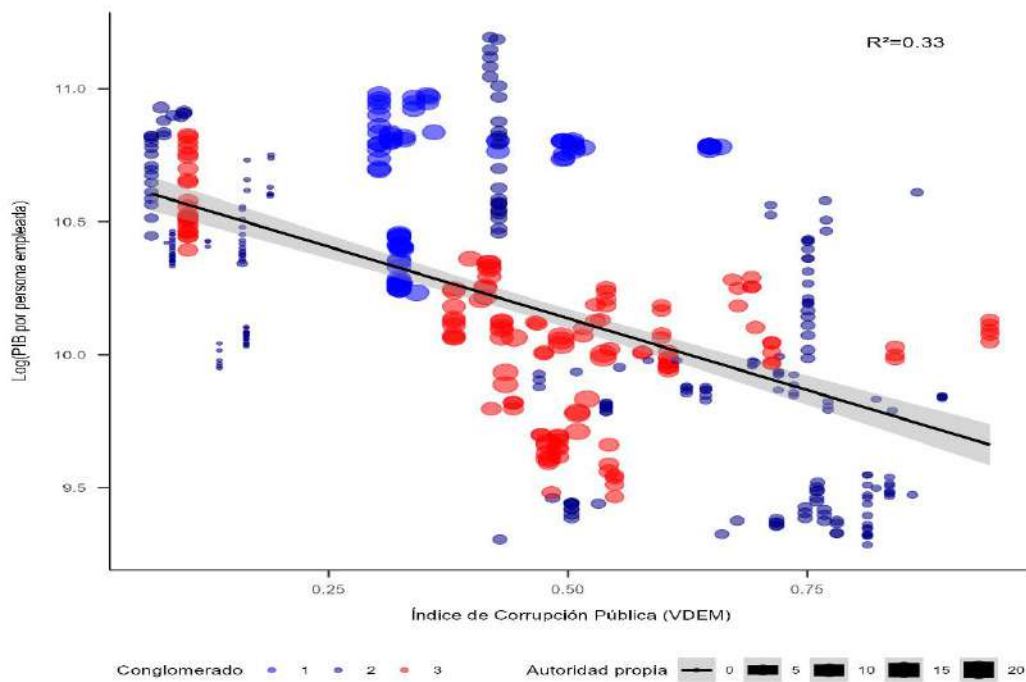


Ilustración 11. Gráfico de dispersión, Log(PIB por persona empleada) e índice de corrupción pública (VDEM). Fuente: Elaboración propia. Nota: el diagrama fue ponderado por el índice de autogobierno del Regional Authority Index (RAI).

Ahora bien, ponderando la gráfica anterior por el índice de autoridad compartida con el gobierno regional se observa cierta aglomeración de datos con un PIB por encima de la recta de regresión que poseen un alto grado de autoridad compartida y menor índice de corrupción política. Este, al igual que en el gráfico anterior, se encuentran en el conglomerado uno mientras que dentro de dicho conglomerado se observan datos por debajo de la tendencia con alto grado de corrupción política y un puntaje de 0 a 2.5 en el índice de autoridad compartida.

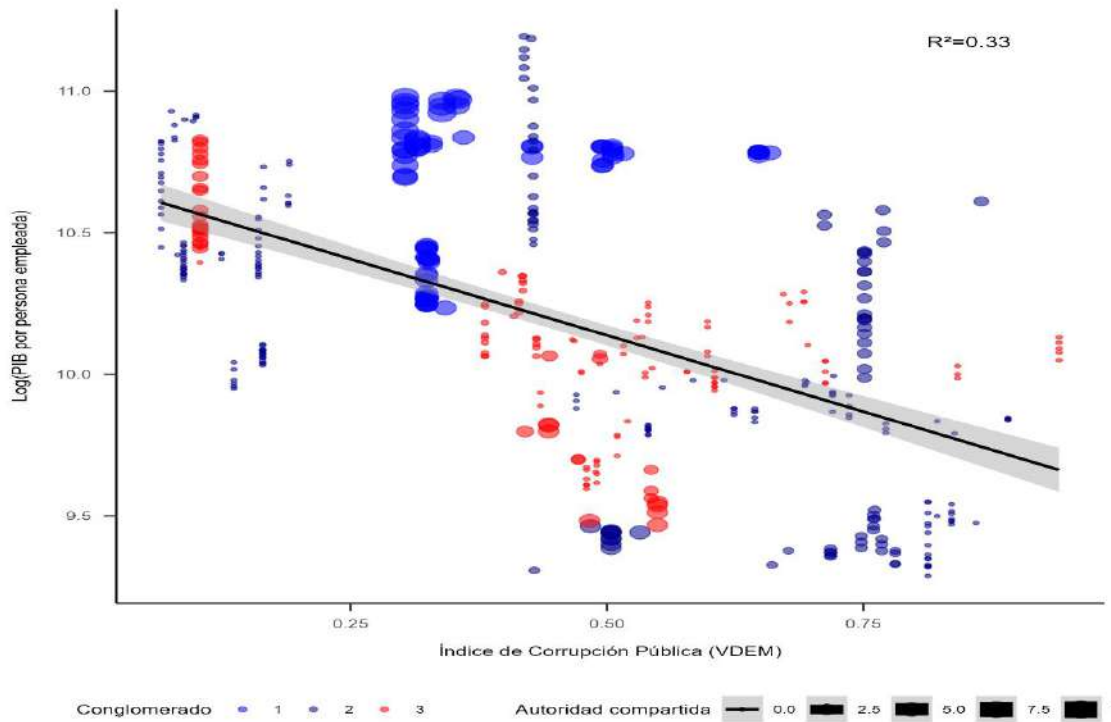


Ilustración 12. Gráfico de dispersión, Log(PIB por persona empleada) e índice de corrupción pública (VDEM). Fuente: Elaboración propia. Nota: el diagrama fue ponderado por el índice de gobierno compartido del Regional Authority Index (RAI).

Analizando en nexa entre la participación política y la autoridad regional en general, en el gráfico 13 se observa una relación cuadrática entre los componentes participativos y la autoridad regional y con un buen ajuste ($R^2=0.24$). A pesar de ello, parece no existir aglomeración de puntos con menor desigualdad

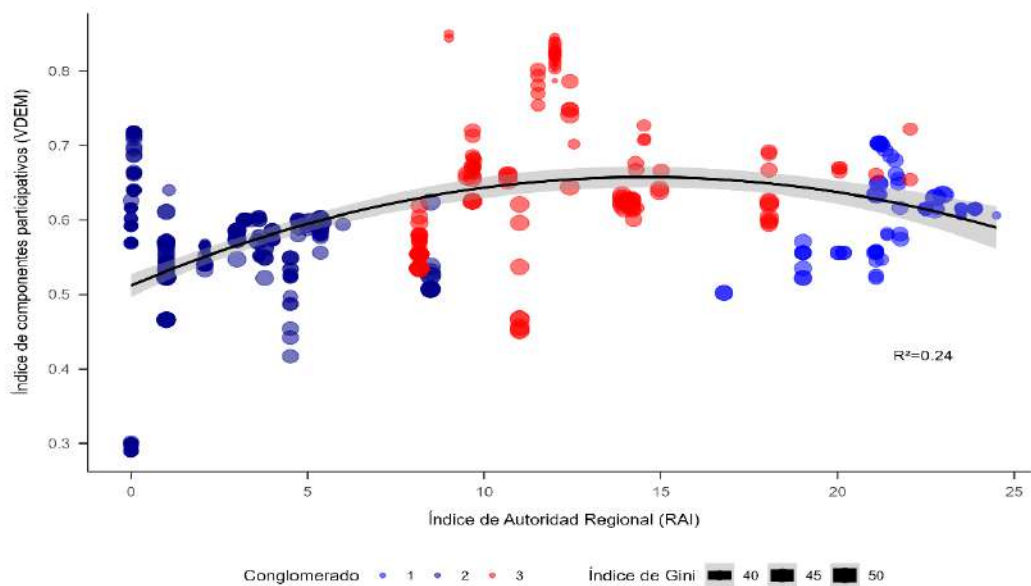


Ilustración 13. Gráfico de dispersión, Índice de componentes participativos (VDEM) e Índice de Autoridad Regional (RAI). Fuente: Elaboración propia. Nota: El diagrama fue ponderado por el índice de Gini post-transferencias.

Ahora bien, cuando examinamos la relación entre desigualdad y participación política observamos una relación cuadrática por igual, en donde la desigualdad parece crecer con la participación, pero a partir de cierto umbral, la desigualdad empieza a caer. Ahora bien, los países en promedio con baja desigualdad y con niveles relativamente altos de mecanismos de participación política efectivos son con puntajes de 15 a 20 en el índice de Autoridad Regional, indicando así que un posible canal de transmisión es las capacidades regionales.

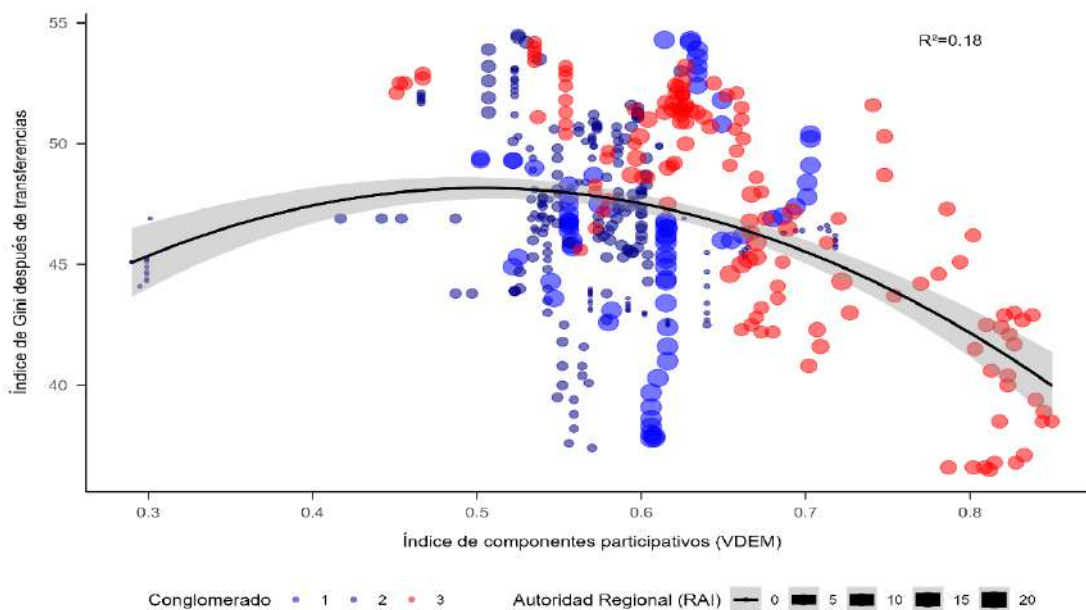


Ilustración 14. Gráfico de dispersión, Índice de Gini después de transferencias e Índice de componentes participativos. Fuente: Elaboración propia. Nota: el diagrama fue ponderado por el Índice de Autoridad Regional (RAI).

Resultado de las regresiones

En la tabla 3 se encuentran los resultados del primer grupo de regresiones. El logaritmo del gasto en transporte de artículos de viaje turísticos en USD muestra significatividad, así como su término cuadrático, arrojando evidencia de una curva de Kuznets o una relación cuadrática. En las especificaciones 2 y 3 el autogobierno y en la especificación 4 el grado de autoridad compartida muestran ser estadísticamente significativo. La participación política muestra también signos esperados respecto de la desigualdad.

	Log(Gini Disp)				
	(1)	(2)	(3)	(4)	(5)
Log(Tur(1))	0.394*** (0.041)	0.396*** (0.036)	0.416*** (0.045)	0.373*** (0.044)	0.362*** (0.047)
Log(Tur(1)) ²	-0.012*** (0.001)	-0.012*** (0.001)	-0.012*** (0.001)	-0.010*** (0.001)	-0.010*** (0.001)
Autoridad Comp.		0.009 (0.006)	0.007 (0.005)	0.008** (0.004)	
Autogobierno		-0.003** (0.001)	-0.004*** (0.001)	-0.0003 (0.002)	
Índice RAI					0.001 (0.001)
Poliarquía (VDEM)			0.137*** (0.030)	0.126*** (0.031)	0.128*** (0.031)
Corrupción (VDEM)			0.116** (0.047)	0.074 (0.045)	0.065 (0.046)
Partip. Pol. (VDEM)			-0.254*** (0.036)	-0.207*** (0.028)	-0.244*** (0.035)
Log(PIB pc PPP USD)				-0.118*** (0.009)	-0.119*** (0.009)
Log(POB)				0.021 (0.013)	0.022* (0.013)
Gob. Consumo (%)				-0.004** (0.002)	-0.004** (0.002)
Constant	0.420 (0.403)	0.408 (0.330)	0.253 (0.430)	1.035** (0.462)	1.150** (0.484)
RMSE	0.043	0.043	0.042	0.033	0.033
F-Statistic	67.8	108.83	134.16	2643.94	2294.51
Países	19	19	19	19	19
Cluster 1 FE	Si	Si	Si	Si	Si
Observations	456	456	456	456	456
R ²	0.418	0.424	0.444	0.662	0.657
Adjusted R ²	0.412	0.415	0.432	0.652	0.648

Note: *p<0.1 **p<0.5 ***p<0.01

Tabla 3. Tabla de regresión. Fuente: Elaboración propia. Nota: Errores estándares robustos por Driscoll y Kraay (1998) entre paréntesis. Se incluyó dummies temporales. Log(Tur (1))= Gasto en turismo por motivos de transporte de artículos de pasajeros en USD.

En la tabla 4, se muestran similares especificaciones a la tabla anterior, solo que esta vez la variable de turismo es el logaritmo de los ingresos por ítems de viaje en USD. En

las especificaciones 3, 4 y 5 arroja una relación cuadrática mientras que en la especificación 3 el autogobierno parece reducir la desigualdad, pero cuando se controla por factores económicos, poblacionales y por el tamaño del gobierno, el efecto desaparece. Ahora bien, no sucede así con el grado de autoridad total regional (especificación 5), en donde se muestra que está asociada a una reducción de la desigualdad. Al igual, en las especificaciones 3, 4 y 5 el grado de participación política se asocia con menor desigualdad.

	Log(Gini Disp)				
	(1)	(2)	(3)	(4)	(5)
Log(Tur(2))	0.044 (0.054)	0.038 (0.066)	0.123* (0.071)	0.107** (0.049)	0.108** (0.043)
Log(Tur(2)) ²	-0.003** (0.001)	-0.003 (0.002)	-0.005*** (0.002)	-0.004*** (0.001)	-0.004*** (0.001)
Autoridad Comp.		-0.007 (0.009)	-0.007 (0.008)	-0.004 (0.007)	
Autogobierno		-0.002 (0.001)	-0.004** (0.002)	-0.003 (0.002)	
Índice RAI					-0.003** (0.001)
Poliarquía (VDEM)			0.157*** (0.038)	0.162*** (0.038)	0.164*** (0.032)
Corrupción (VDEM)			0.098** (0.045)	0.062 (0.048)	0.061 (0.046)
Partip. Pol. (VDEM)			-0.178*** (0.035)	-0.226*** (0.046)	-0.228*** (0.052)
Log(PIB pc PPP USD)				-0.057*** (0.016)	-0.057*** (0.016)
Log(POB)				0.046*** (0.010)	0.046*** (0.009)
Gob. Consumo (%)				-0.003 (0.002)	-0.003 (0.002)
Constant	4.180*** (0.521)	4.325*** (0.680)	3.424*** (0.743)	3.030*** (0.685)	3.021*** (0.579)
RMSE	0.036	0.037	0.036	0.035	0.035
F-Statistic	51.8	194.6	333.25	1225.83	1382.3
Países	19	19	19	19	19
Cluster 2 FE	Si	Si	Si	Si	Si
Observations	456	456	456	456	456
R ²	0.525	0.575	0.596	0.624	0.624
Adjusted R ²	0.521	0.569	0.588	0.614	0.614

Note: *p<0.1 **p<0.5 ***p<0.01

Tabla 4. Tabla de regresión. Fuente: Elaboración propia. Nota: Errores estándares robustos por Driscoll y Kraay (1998) entre paréntesis. Se incluyó dummies temporales. Log(Tur(2))= Logaritmo de ingresos del turismo por artículos de viaje (USD).

Por otro lado, en la tabla 5 se muestran los resultados de otros modelos, pero tomando como proxy del desarrollo turístico, el total del gasto del turismo como porcentaje de las importaciones (en logaritmo, expresando así la velocidad del crecimiento de la tasa) Esta no muestra ser estadísticamente significativa en los modelos 4 y 5. El autogobierno esta asociada negativamente (en (2) y en (3)) a la desigualdad, pero dicho efecto desaparece una vez se introducen variables de control.

	Log(Gini Disp)				
	(1)	(2)	(3)	(4)	(5)
Log(Tur(3))	0.066*	0.090***	0.095***	0.064	0.063
	(0.039)	(0.033)	(0.030)	(0.043)	(0.040)
Log(Tur(3)) ²	-0.021	-0.022*	-0.021**	-0.022	-0.022
	(0.014)	(0.012)	(0.010)	(0.014)	(0.013)
Autoridad Comp.		-0.010	-0.015	-0.003	
		(0.010)	(0.010)	(0.007)	
Autogobierno		-0.004***	-0.007***	-0.001	
		(0.002)	(0.002)	(0.002)	
Índice RAI					-0.002
					(0.001)
Poliarquía (VDEM)			0.157***	0.153***	0.157***
			(0.036)	(0.037)	(0.031)
Corrupción (VDEM)			0.109***	0.035	0.031
			(0.037)	(0.044)	(0.045)
Partip. Pol. (VDEM)			-0.324***	-0.257***	-0.267***
			(0.046)	(0.055)	(0.059)
Log(PIB pc PPP USD)				-0.130***	-0.129***
				(0.007)	(0.007)
Log(POB)				0.029***	0.029***
				(0.008)	(0.006)
Gob. Consumo (%)				-0.003	-0.003
				(0.002)	(0.002)
Constant	3.788***	3.881***	3.978***	4.602***	4.596***
	(0.105)	(0.110)	(0.122)	(0.151)	(0.111)
RMSE	0.037	0.043	0.046	0.036	0.036
F-Statistic	2.01	9.2	120.82	1454.92	1404.12
Países	19	19	19	19	19
Cluster 2 FE	Si	Si	Si	Si	Si
Observations	456	456	456	456	456
R ²	0.065	0.094	0.163	0.606	0.604
Adjusted R ²	0.057	0.082	0.146	0.596	0.594
Note:				*p<0.1; **p<0.05; ***p<0.01	

Tabla 5. Tabla de regresión. Fuente: Elaboración propia. Nota: Errores estándares robustos por Driscoll y Kraay (1998) entre paréntesis. Se incluyó dummies temporales. Log(Tur(3))= Logaritmo de gastos de turismo internacional como % de las importaciones.

En la tabla 6, el logaritmo del gasto en turismo internacional en dólares muestra en todas las especificaciones una relación cuadrática y estadísticamente significativa mientras que en las especificaciones 2 y 3 el índice de autogobierno reduce la desigualdad y es estadísticamente significativo, al introducir las variables de control deja de ser un efecto estadísticamente relevante. La participación política en todas las especificaciones termina reduciendo la desigualdad y la corrupción política deja de ser significativa una vez se controla por el crecimiento poblacional y del PIB per cápita y el consumo del gobierno general como porcentaje del PIB.

	Log(Gini Disp)				
	(1)	(2)	(3)	(4)	(5)
Log(Tur(4))	0.113*	0.111*	0.152**	0.247***	0.242***
	(0.064)	(0.064)	(0.059)	(0.049)	(0.051)
Log(Tur(4)) ²	-0.004***	-0.004***	-0.005***	-0.006***	-0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Autoridad Comp.		-0.004	-0.007	0.001	
		(0.009)	(0.007)	(0.006)	
Autogobierno		-0.002**	-0.002*	-0.001	
		(0.001)	(0.001)	(0.002)	
Índice RAI					-0.0005
					(0.001)
Poliarquía (VDEM)			0.150***	0.156***	0.157***
			(0.034)	(0.032)	(0.029)
Corrupción (VDEM)			0.095**	0.058	0.054
			(0.043)	(0.045)	(0.045)
Partip. Pol. (VDEM)			-0.288***	-0.251***	-0.264***
			(0.040)	(0.038)	(0.044)
Log(PIB pc PPP USD)				-0.106***	-0.105***
				(0.010)	(0.011)
Log(POB)				0.039***	0.039***
				(0.010)	(0.009)
Gob. Consumo (%)				-0.004**	-0.004**
				(0.002)	(0.002)
Constant	3.395***	3.511***	3.098***	1.882***	1.943***
	(0.658)	(0.659)	(0.599)	(0.536)	(0.529)
RMSE	0.041	0.042	0.042	0.035	0.035
F-Statistic	48.7	93.97	101.07	2650.49	2928.05
Países	19	19	19	19	19
Cluster 2 FE	Si	Si	Si	Si	Si
Observations	456	456	456	456	456
R ²	0.397	0.450	0.459	0.631	0.629
Adjusted R ²	0.392	0.442	0.448	0.621	0.620

Note:

*p<0.1 **p<0.05 ***p<0.01

Tabla 6. Tabla de regresión. Fuente: Elaboración propia. Nota: Errores estándares robustos por Driscoll y Kraay (1998) entre paréntesis. Se incluyó dummies temporales. Log(Tur(4))= Logaritmo de gastos de turismo internacional (USD).

Analizando los resultados del modelo (5) de la tabla 4 se construyó un diagrama de dispersión con el Gini predicho (se aplicó el exponencial al logaritmo de las predicciones para llevar las predicciones del Gini a nivel y no en tasa de crecimiento, que es lo que matemáticamente representa el logaritmo del índice), se observa para el periodo una relación de Kuznets invertida, en donde vemos una desaceleración de la desigualdad a partir de cierto nivel de crecimiento de la actividad turística.

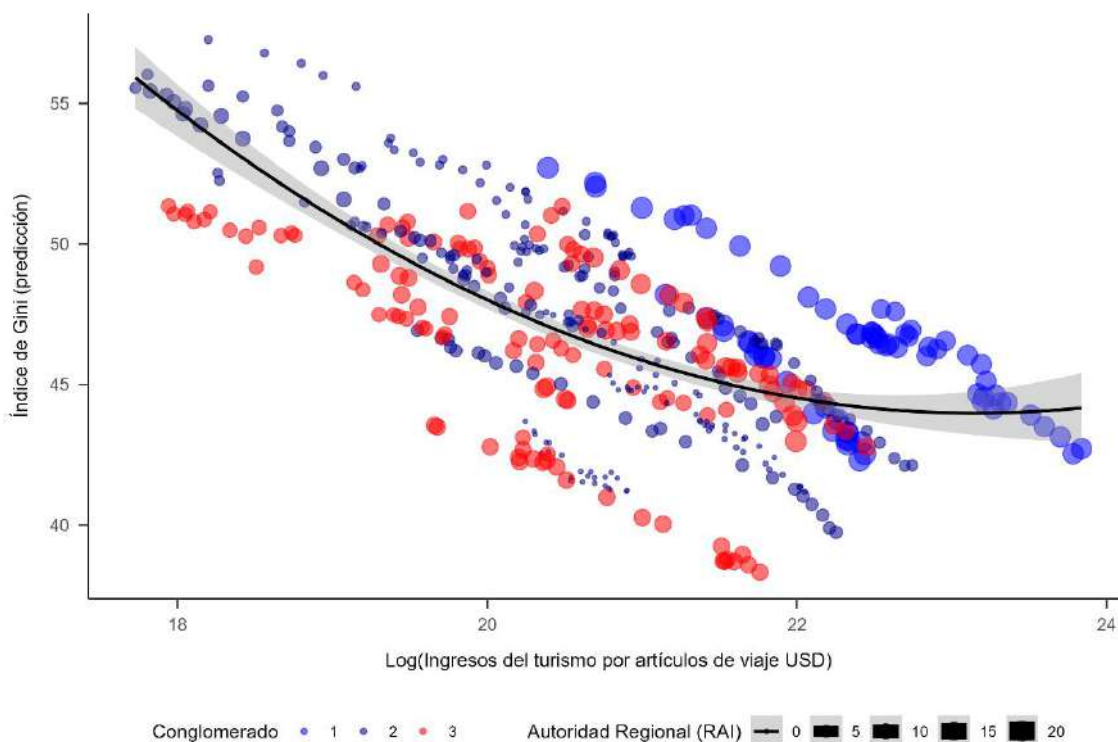


Ilustración 15. Gráfico de dispersión, Predicción del índice de Gini por el Logaritmo de los ingresos del turismo por artículos de viaje en USD. Fuente: Elaboración propia. Nota: Ponderado por el índice de autoridad regional (RAI).

Caso de estudio: República Dominicana

República Dominicana es uno de los países de la región de Latinoamérica y El Caribe con un sector turístico muy atractivo y una economía que es de las que mayor crecimiento sostenido ha mantenido en los últimos 30 años. Examinando de manera exhaustiva observamos que existen desigualdades dentro de las macro regiones del país (medida por el índice de Palma, que es más sensible a la desigualdad entre ricos y pobres, es decir, entre el 10 % más rico y el 40 % más pobre). El desempeño de la gestión administrativa de los ayuntamientos (medida por el puntaje otorgado en el Sistema de Monitoreo de la Administración Pública Municipal (SISMAP municipal)) parece no estar relacionado con la desigualdad ni con el desarrollo turístico.

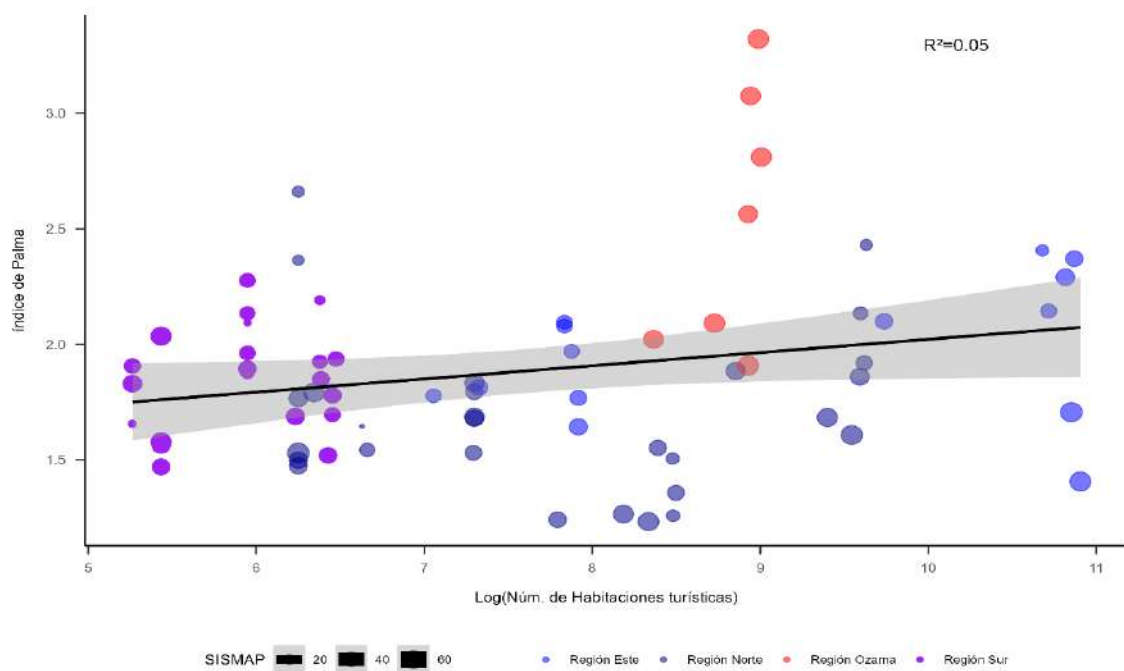


Ilustración 16. Gráfico de Dispersión, Índice de Palma y Log(Número de habitaciones para hospedaje), 2016-2022. Fuente: Elaboración propia con datos del Sistema de Indicadores Dominicano (SISDOM), datos administrativos del Ministerio de Turismo/ASONAHORES y del Sistema Municipal de la Administración Pública (SISMMap).

En el gráfico 14 se observa la relación entre la desigualdad, medida por el índice de Gini para los hogares y el logaritmo del Producto Interno Bruto per cápita regional⁵. Existe una relación negativa entre renta per cápita y desigualdad. Los puntos del diagrama de dispersión están ponderados por la ratio entre ingresos y egresos de los gobiernos locales en millones de pesos (2010=100), el cual es una variable proxy de las capacidades institucionales fiscales regionales. Al parecer ciertas regiones de planificación con mayor desigualdad tienden a tener menores ratios de ingresos y egresos

⁵ La clasificación regional usada en el panel es la establecida por la ley 345-12 de regiones únicas de planificación, las cuales son diez (10) que pueden luego agruparse en 4 macro regiones.

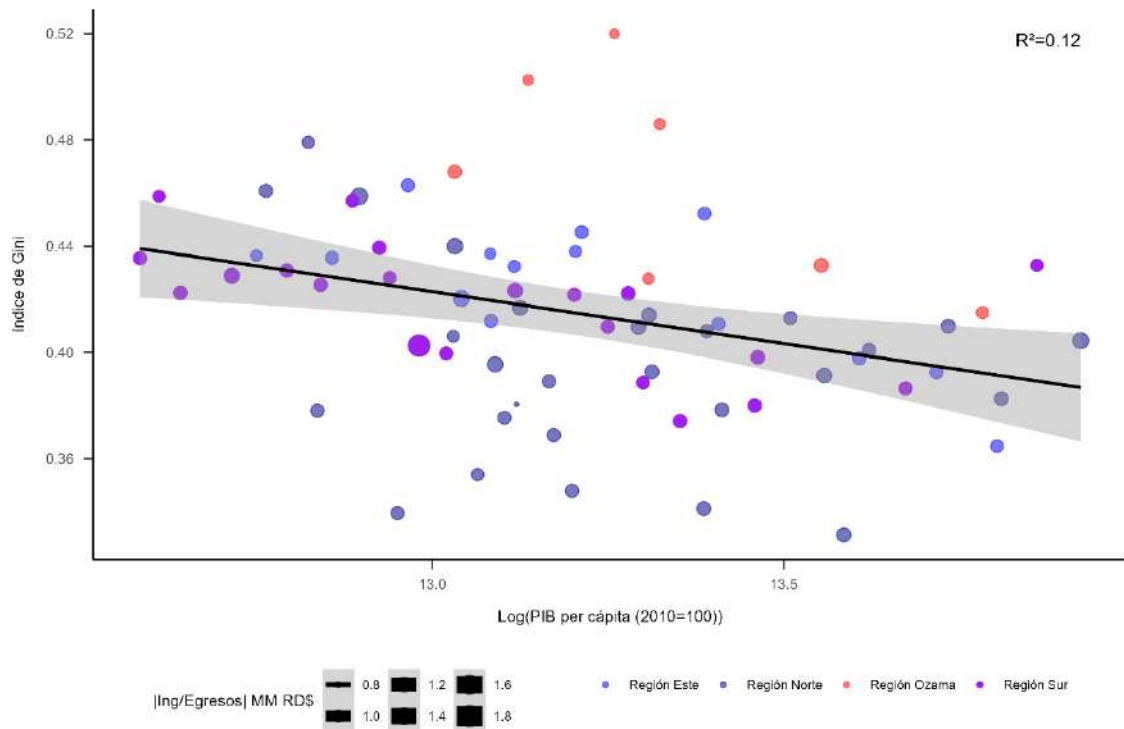


Ilustración 17. Gráfico de Dispersión, Índice de Gini para hogares y Log(PIB per cápita (2010=100)), 2016-2022. Fuente: Elaboración propia con datos del Sistema de Indicadores Dominicano (SISDOM) y Dirección General de Presupuesto (DIGEPRESS).

En el gráfico 15 dispersión de los puntos indica que hay una tendencia positiva; es decir, a medida que aumentan los ingresos del gobierno local, también lo hace el índice de Palma, aunque la relación no es muy fuerte. La variabilidad en los puntos sugiere diferencias significativas entre ayuntamientos dentro de las mismas regiones y entre regiones. Algunos ayuntamientos con ingresos similares tienen puntajes SISMAP muy diferentes, lo que podría deberse a la eficiencia en la administración de recursos, particularidades regionales o la calidad de los servicios públicos locales. La dispersión también indica que hay ayuntamientos en todas las regiones que tienen un desempeño similar en el índice SISMAP a pesar de tener diferentes niveles de ingresos gubernamentales locales.

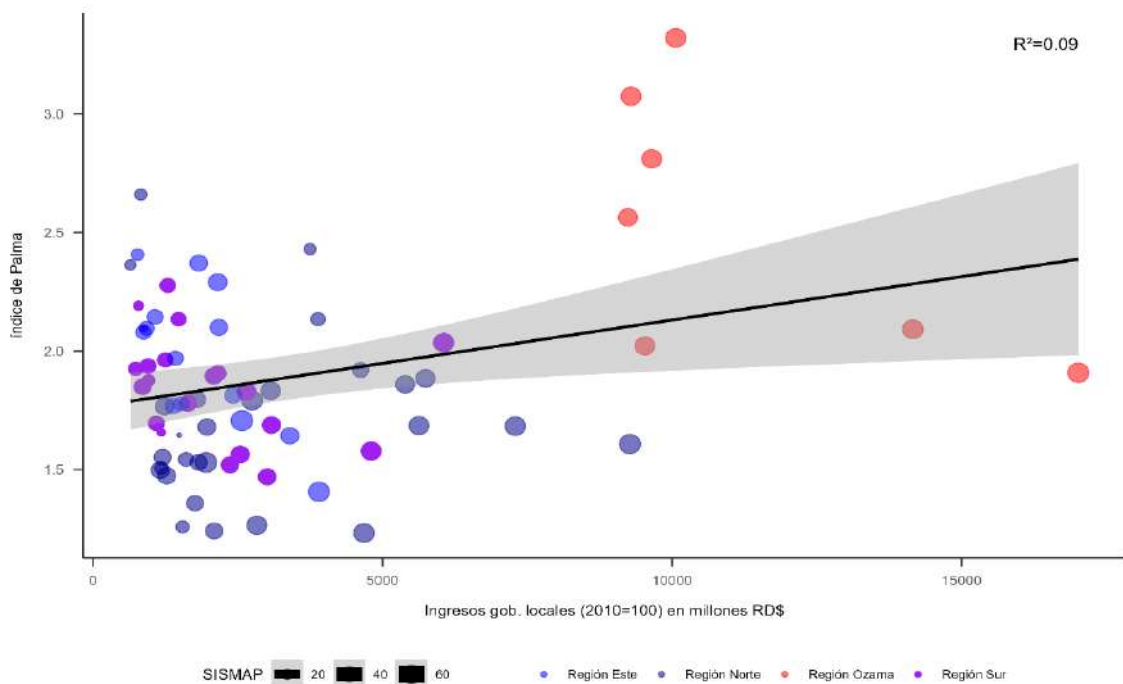


Ilustración 18. Gráfico de dispersión, Índice de Palma e Ingresos de los gobiernos locales en millones de RD\$ (2010=100), 2016-2022. Fuente: Elaboración propia con datos del Sistema de Indicadores Dominicano (SISDOM) y Dirección General de Presupuesto (DIGEPRESS).

El gráfico 16 proporciona evidencia de una asociación positiva entre el gasto de los gobiernos locales y la desigualdad de ingresos entre dos quintiles altos, aunque la relación no es muy fuerte y hay una gran dispersión en los datos. Esto implica que las políticas gubernamentales locales que conducen a un mayor gasto no necesariamente están asociadas de manera fuerte con la reducción de la desigualdad entre estos grupos de ingresos, al menos según lo que se puede inferir de este análisis.

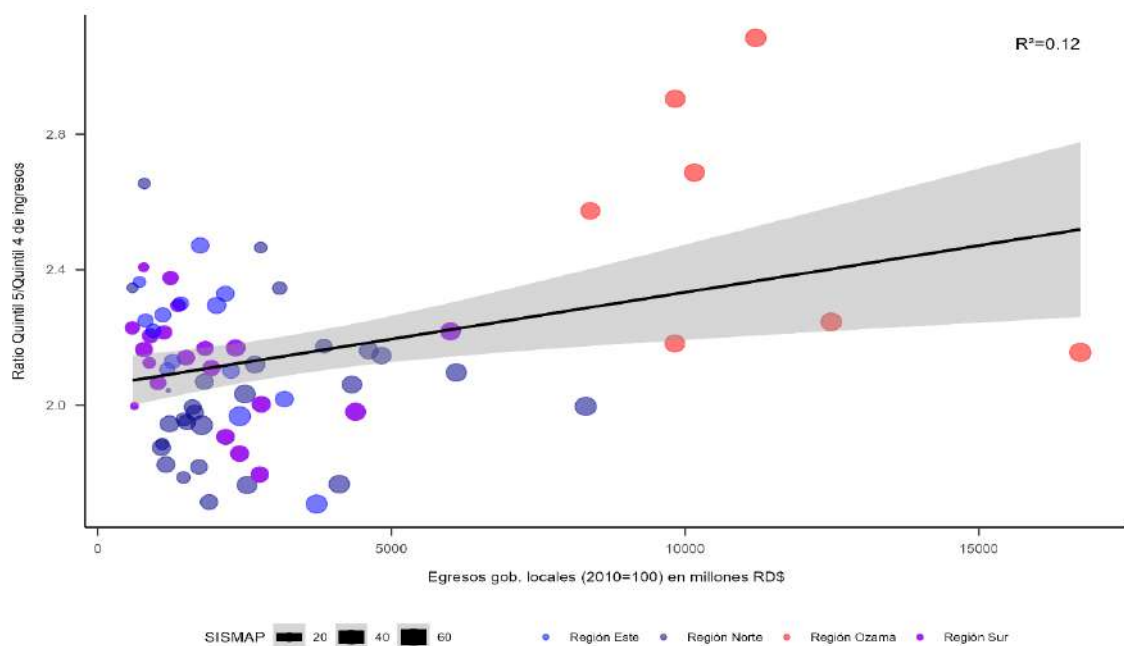


Ilustración 19. Gráfico de dispersión, Ratio entre Quintil 5 y 4 de mayores ingresos y egresos de los gobiernos locales en millones de RD\$ (2010=100), 2016-2022. Fuente: Elaboración propia con datos del Sistema de Indicadores Dominicano (SISDOM) y Dirección General de Presupuesto (DIGEPRESS).

En el contexto del coeficiente Gini, la variable LnPIB (logaritmo del PIB per cápita) es significativamente negativa ($\beta = -0.052$, $p < 0.05$), sugiriendo que un incremento en el PIB per cápita está asociado con una disminución en la desigualdad de ingresos, ajustada por otras variables en el modelo. Este resultado es consistente con la teoría económica que postula que el crecimiento económico puede conducir a una distribución más equitativa de los ingresos, aunque la significancia del efecto es modesta.

Por otro lado, la variable $|\ln g/Eg|$ Gob. Local (ratio de ingresos/egresos de los gobiernos locales) muestra un efecto negativo significativo en el coeficiente Gini ($\beta = -0.062$, $p < 0.05$), indicando que un mayor balance fiscal local se correlaciona con menores niveles de desigualdad. Este hallazgo puede reflejar la importancia de la gestión fiscal a nivel local en la promoción de la equidad económica.

La variable DA (desempleo abierto) no parece tener un efecto significativo en el coeficiente Gini, pero sí muestra una relación positiva y significativa con la ratio Q2/Q1 ($\beta = 0.006$, $p < 0.05$), lo que podría indicar que el desempleo abierto afecta desproporcionadamente a los hogares de menores ingresos.

El puntaje SISMAP Municipal (SMM) no mostró un efecto significativo en las medidas de desigualdad en este análisis, lo que sugiere que las medidas de gobernanza y administración municipal, según se capturan a través de esta variable, no están directamente asociadas con la desigualdad de ingresos de los hogares en la muestra.

estudiada. Las variables de turismo no fueron estadísticamente significativas en ninguna de las regresiones.

	Gini (1)	Palma (2)	R.Kuznets (3)	Q5/Q4 (4)	Q4/Q3 (5)	Q3/Q2 (6)	Q2/Q1 (7)
Log(Tur)	0.013 (0.014)	0.103 (0.111)	0.510 (0.528)	0.055 (0.065)	-0.015 (0.010)	0.012 (0.017)	0.035 (0.041)
Ln(PIB)	-0.052** (0.015)	-0.510* (0.238)	-1.899** (0.752)	-0.286 (0.169)	-0.079*** (0.022)	-0.057 (0.039)	-0.013 (0.073)
DA	0.002 (0.001)	0.020 (0.013)	0.101 (0.053)	0.011 (0.009)	0.002** (0.001)	0.001 (0.002)	0.006* (0.003)
SMM	-0.0004 (0.0004)	-0.004 (0.005)	-0.017 (0.017)	-0.003 (0.003)	0.0001 (0.0005)	-0.0001 (0.001)	-0.002 (0.001)
Ing/Eg Gob. Local	-0.062** (0.019)	-0.720** (0.231)	-2.724** (0.807)	-0.416* (0.160)	-0.011 (0.024)	-0.081** (0.027)	-0.142** (0.035)
Ln(PU)	0.017 (0.014)	0.256 (0.117)	0.871 (0.539)	0.164* (0.069)	0.034** (0.011)	0.010 (0.018)	0.027 (0.044)
t(2020)	-0.017* (0.009)	-0.196* (0.099)	-0.613 (0.334)	-0.148* (0.068)	-0.017 (0.020)	-0.004 (0.009)	0.030 (0.018)
t(2022)	0.004 (0.011)	0.050 (0.118)	0.279 (0.383)	-0.001 (0.075)	0.006 (0.014)	0.010 (0.018)	0.039 (0.026)
Constant	0.877*** (0.195)	5.584* (2.619)	21.928* (10.063)	3.985* (1.858)	2.090*** (0.352)	2.084*** (0.466)	1.554 (1.120)
COV(B)	CR2	CR2	CR2	CR2	CR2	CR2	CR2
Regiones	10	10	10	10	10	10	10
RMSE	0.019	0.225	0.769	0.154	0.033	0.03	0.049
Observations	70	70	70	70	70	70	70
R ²	0.467	0.417	0.446	0.385	0.355	0.209	0.260
Adjusted R ²	0.397	0.340	0.373	0.304	0.271	0.105	0.163

Note:

*p<0.1**p<0.5***p<0.01

Tabla 7. Tabla de regresiones de panel para República Dominicana (2016-2022). Fuente: Elaboración propia. Nota: CR2= Corrección de Pustejovsky y Tipton, (2018).

Conclusiones

El análisis de conglomerados forma parte de un arsenal de técnicas estadísticas dentro de lo que se conoce como aprendizaje no automatizado que pueden ser útiles al momento de hacer análisis no solo predictivos, sino también para estrategias teórico-metodológicas y analíticas para estudiar fenómenos complejos. Este trabajo de fin de master evidencia como algoritmos de k-medios sirven para estudiar los efectos del turismo y de las capacidades institucionales de los gobiernos sub-nacionales o regionales en la desigualdad económica de la región de Latinoamérica, complementando así el análisis econométrico tradicional.

Este trabajo es consecuente con la literatura institucionalista de Levitsky y Murillo (2012) y de Mella Gómez (2022), de que la inestabilidad institucional hace que los resultados en países de la región sean un poco distintos a lo que se observa en países industrializados o en democracias representativas realmente institucionalizadas. Se corroboran dos hipótesis sobre que el turismo genera desigualdad, pero a partir de cierto umbral termina aumentando o realentizando la velocidad de su reducción, corroborando un poco con los resultados de Alam y Paramati (2016), que encuentran una curva de Kuznets invertida en países en vías de desarrollo. Esto es, quizás, porque no se incorpora el papel que juegan las instituciones regionales y que en gran parte de este tipo de estudios se utilizan formas tradicionales de lidiar con la heterogeneidad, dejando de capturar especificidad en comparación con la estrategia empírica usada en este estudio. Ahora bien, este estudio no puede llegar a conclusiones taxativas al respecto de lo anteriormente dicho. Para corroborar esta hipótesis implicaría desarrollar líneas de investigación en materia econométrica y estadística, generando estudios de simulación para comprender mejor las ventajas relativas detrás de las técnicas tradicionales versus este abordaje usado en esta investigación y comprender mejor el proceso generador de datos detrás.

En países como República Dominicana, logramos entrever que la desigualdad y el desarrollo turístico no están del todo relacionados, quizás porque el modelo de turismo constituido, evidenciado por Ceara-Hatton (2008, 2015) y otros autores (Contreras, 2011; Mella Gómez, 2022; Torres y Corporan, 2021). Por otro lado, mejorar las capacidades institucionales fiscales de los gobiernos locales se asocia a mayores reducciones de la desigualdad. Esto quiere decir que quizás una vía para poder lograr que el turismo contribuya a la reducción de desigualdades económicas es por vía de mejorar las capacidades de los gobiernos locales (tanto lo administrativo como en lo fiscal). Este caso quizás pone en contraste el por qué existe una relación heterogénea en la región de Latinoamérica entre turismo, descentralización política y la desigualdad

económica. Esto puede indicar que uno de los factores que han logrado esta marcada reducción de la desigualdad en la región es la actividad económica y el crecimiento de la región pero que se hace con poca inversión en instituciones, lo cual puede comprometer no solo el crecimiento de largo plazo, sino hacer que las desigualdades del pasado persistan.

En cambio, la participación política presenta ser un factor que incide de manera significativa en la reducción de las desigualdades a nivel regional a pesar de la volatilidad institucional en lo que refiere a la descentralización política de la región, lo cual es consistente con lo expresado por Ivan Finot (2002, 2005, 2007). La persistencia de las instituciones con prácticas corruptas en lo administrativo del Estado sigue siendo un problema (aunque desde un punto de vista estadístico en las regresiones solo en algunas especificaciones resultan estadísticamente significativas) mientras que la calidad de la democracia electoral se asocia con más desigualdad, algo coherente con la idea de O'Donnell (2004) de que las democracias delegativas por lo general son marcos institucionales extractivos (en términos de Acemoglu) en donde la concentración del poder político y los fallos institucionales y el poco enforcement generado por la intermitencia de las instituciones que menciona Levitsky y Murillo (2012), son factores que generan mayor desigualdad mediante mecanismos como la captura regulatoria o de las políticas públicas, siendo este, la democracia electoral, un efecto importante en términos teóricos y estadísticos. Esto implica que es necesario investigar más a fondo los mecanismos que generan esta relación, más allá de lo evidenciado por autores como O'Donnell (2004), Levitsky y Murillo (2012), Mella Gómez (2022) y Mella Gómez y Prestol Flaquer (2023).

Referencias

- Acemoglu, D., Gallego, F. A., y Robinson, J. A. (2014). Institutions, human capital, and development. *Annu. Rev. Econ.*, 6, 875-912.
- Acosta Guzmán, J. A. (2017). Impacto económico del turismo de crucero: caso República Dominicana. *El periplo sustentable*, 33, 35-55.
- Al Tal, R., Mukheimer, T., Theodory, R., Zwayyed, Z., Almehdaw, N., Al Abdallat, D., y Saticilef, F. (2021). Tourism Industry Impacts On Spatial Inequality In The Dead Sea Region. *GEOGRAPHY, ENVIRONMENT, SUSTAINABILITY*, 14(3), 54-62. <https://doi.org/10.24057/2071-9388-2020-216>
- Alam, Md. S., y Paramati, S. R. (2016). The impact of tourism on income inequality in developing economies: Does Kuznets curve hypothesis exist? *Annals of Tourism Research*, 61, 111-126. <https://doi.org/10.1016/j.annals.2016.09.008>
- Alvarado, J., y Oddone, N. (2017). Fortalecimiento de la cadena de valor de turismo en Pedernales (República Dominicana). *Fortalecimiento de cadenas de valor rurales. Santiago: CEPAL, 2017. LC/TS. 2017/24. p. 243-292.*
- Beck, N., y Katz, J. N. (1995). What To Do (and Not to Do) with Time-Series Cross-Section Data. *American Political Science Review*, 89, 634-647. <https://doi.org/10.2307/2082979>
- Birdsall, N., Lustig, N., y McLeod, D. (2013). Declining inequality in Latin America: some economics, some politics. En *Routledge handbook of Latin American politics* (pp. 158-180). Routledge.
- Boettke, P. J., y Marciano, A. (2016). The distance between Buchanan's "An Economic Theory of Clubs" and Tiebout's "A Pure Theory of Local Public Expenditures". New insights based on an unpublished manuscript. *The European Journal of the History of Economic Thought*, 24(2), 205-237. <https://doi.org/10.1080/09672567.2016.1168464>
- Büscher, B., y Fletcher, R. (2016). Destructive creation: capital accumulation and the structural violence of tourism. *Journal of Sustainable Tourism*, 25(5), 651-667. <https://doi.org/10.1080/09669582.2016.1159214>
- Caro-González, F. J., Acosta Guzmán, J. A., Orgaz-Agüera, F., y Castellanos-Verdugo, M. (2015). Turismo, desarrollo sostenible y percepción de los stakeholders. Un estudio de caso en República Dominicana. *Revista de economía del Caribe*, 15, 153-182.
- Ceara-Hatton, M. (2005). El Desarrollo Humano y la política social en la República Dominicana. *Boletín mensual de la Oficina de Desarrollo Humano/PNUD*, 19.

- Ceara-Hatton, M. (2015). Qué significa cambiar el «modelo económico». *Foro Político Social: Crecimiento, Desarrollo, Desigualdad. Fundação Frederich Erbert stiftung. Republica Dominicana.*
- Chong, A., y Gradstein, M. (2007). Inequality and institutions. *The Review of Economics and Statistics*, 89, 454-465. <https://doi.org/10.1162/rest.89.3.454>
- Contreras, P. (2011). Expansión y agotamiento del modelo turístico dominicano. El turismo en los informes de desarrollo humano en la República Dominicana. *TURISMO PLACEBO*, 11.
- Coppedge, M., Gerring, J., Knutsen, C. H., Krusell, J., Medzihorsky, J., Pernes, J., Skaaning, S.-E., Stepanova, N., Teorell, J., Tzelgov, E., Wilson, S. L., y Lindberg, S. I. (2019). The Methodology of “Varieties of Democracy” (V-Dem)¹. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, 143(1), 107-133. <https://doi.org/10.1177/0759106319854989>
- Dahl, R. A. (2008). *Polyarchy: Participation and opposition*. Yale university press.
- Driscoll, J. C., y Kraay, A. C. (1998). Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data. *Review of Economics and Statistics*, 80, 549-560. <https://doi.org/10.1162/003465398557825>
- Dryzek, J. S. (1996). The informal logic of institutional design. *The theory of institutional design*, 103-125. <https://doi.org/10.1017/CBO9780511558320.005>
- Fang, J., Gozgor, G., Paramati, S. R., y Wu, W. (2020). The impact of tourism growth on income inequality: Evidence from developing and developed economies. *Tourism Economics*, 27(8), 1669-1691. <https://doi.org/10.1177/1354816620934908>
- Finot, I. (2001). *Descentralización en América Latina: teoría y práctica*. Cepal.
- Finot, I. (2002). Descentralización y participación en América Latina: una mirada desde la economía. *Revista de la CEPAL*.
- Finot, I. (2005). Descentralización, transferencias territoriales y desarrollo local. *Revista de la CEPAL*.
- Finot, I. (2007). Los procesos de descentralización en América Latina. *Investigaciones Regionales-Journal of Regional Research*, 173-205.
- Gasparini, L., y Lustig, N. (2011). *The Rise and Fall of Income Inequality in Latin America*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199571048.013.0027>

- Ghosh, S., y Mitra, S. K. (2021). Tourism and inequality: A relook on the Kuznets curve. *Tourism Management*, 83, 104255. <https://doi.org/10.1016/j.tourman.2020.104255>
- Hernández R, D. R., Anna Cristina et al. (2019). El turismo como herramienta para el desarrollo inclusivo y sostenible en la República Dominicana. *Boletín del Observatorio de Políticas Sociales y Desarrollo*, 2.
- Hoffman, K., y Centeno, M. A. (2003). The Lopsided Continent: Inequality in Latin America. *Annual Review of Sociology*, 29(1), 363-390. <https://doi.org/10.1146/annurev.soc.29.010202.100141>
- HOOGHE, E., Marks, G., Schakel, A. H., Niedzwiecki, S., Chapman-Osterkatz, S., y Shair-Rosenfield, S. (2021). *Regional authority index (RAI) v. 3*.
- Levitsky, S., y Murillo, M. (2005). *Argentine democracy: The politics of institutional weakness*. Penn State Press.
- Levitsky, S., y Murillo, M. (2010). Variación en la fortaleza institucional. *Revista de sociolog*, 31-56. <https://doi.org/10.5354/0719-529X.2010.14399>
- Levitsky, S., y Murillo, M. V. (2012). Construyendo instituciones sobre cimientos débiles: lecciones desde América Latina. *Politai*, 3, 17-44. <https://revistas.pucp.edu.pe/index.php/politai/article/view/14124>
- Li, X., Hipel, K. W., y Dang, Y. (2015). An improved grey relational analysis approach for panel data clustering. *Expert Systems with Applications*, 42(23), 9105-9116. <https://doi.org/10.1016/j.eswa.2015.07.066>
- Liao, N., y He, Y. (2018). Exploring the effects of influencing factors on energy efficiency in industrial sector using cluster analysis and panel regression model. *Energy*, 158, 782-795. <https://doi.org/10.1016/j.energy.2018.06.049>
- LOMBARDO, R., Falcone, M., et al. (2011). Crime and Economic Performance. A cluster analysis of panel data on Italy's NUTS 3 regions. En *WP. n. 12, 2011, Dipartimento di Economia e Statistica, Università della Calabria*.
- Mahoney, J., Thelen, K., et al. (2010). A theory of gradual institutional change. *Explaining institutional change: Ambiguity, agency, and power*, 1, 1-37.
- Martínez-Vázquez, J., Radics, A., y Pérez, B. (2017). *Descentralización en República Dominicana: desempeño actual y perspectivas de reforma*.
- Mella Gómez, L. E. (2022). Turismo, vaivenes y desigualdad: cómo repercuten las instituciones participativas y de transparencia en la actividad turística y su incidencia en

la desigualdad en Latinoamérica (2002-2019). *Ciencia, Economía y Negocios*, 6(1), 57-84. <https://doi.org/10.22206/ceyn.2022.v6i1.pp57-84>

Mella Gómez, L. E., y Prestol Flaquer, J. (2023). Corrupción, crecimiento e instituciones: ¿existe una tasa de sacrificio en América Latina? *Problemas del Desarrollo. Revista Latinoamericana de Economía*, 55(216), 3-34. <https://doi.org/10.22201/iiec.20078951e.2024.216.70055>

Nguyen, C. P. (2022). Tourism and gender (in)equality: Global evidence. *Tourism Management Perspectives*, 41, 100933. <https://doi.org/10.1016/j.tmp.2021.100933>

Nguyen, C. P., Schinckus, C., Su, T. D., y Chong, F. H. L. (2020). The Influence of Tourism on Income Inequality. *Journal of Travel Research*, 60(7), 1426-1444. <https://doi.org/10.1177/0047287520954538>

Nogueira, M. C., y Madaleno, M. (2021). Are International Indices Good Predictors of Economic Growth? Panel Data and Cluster Analysis for European Union Countries. *Sustainability*, 13(11), 6003. <https://doi.org/10.3390/su13116003>

North, D. C. (1990). *Institutions, institutional change and economic performance*. Cambridge university press.

O'Donnell, G. (2004). Accountability horizontal: la institucionalización legal de la desconfianza política. *Revista Española de Ciencia Política*, 11, 11-31.

Oates, Wallace E. (1968). The Theory of Public Finance in a Federal System. *The Canadian Journal of Economics*, 1(1), 37. <https://doi.org/10.2307/133460>

Oates, Wallace E. (2008). CHAPTER On the theory and practice of fiscal decentralization. *Inst. Found. Public Financ. Econ. Leg. Perspect*, 165, 3-61.

Prestol Flaquer, J. R. (2022). *¿Es la municipalización un diseño institucional apropiado para la descentralización del desarrollo económico? Algunas consideraciones para el caso latinoamericano*.

Przeworski, A. (1988). Democracy as a Contingent Outcome of Conflicts. En *Constitutionalism And Democracy*. Cambridge University Press.

Pustejovsky, J. E., y Tipton, E. (2018). Small-Sample Methods for Cluster-Robust Variance Estimation and Hypothesis Testing in Fixed Effects Models. *Journal of Business & Economic Statistics*, 36(4), 672–683. <https://doi.org/10.1080/07350015.2016.1247004>

- Solt, F. (2020). Measuring Income Inequality Across Countries and Over Time: The Standardized World Income Inequality Database. *Social Science Quarterly*, 101(3), 1183-1199. <https://doi.org/10.1111/ssqu.12795>
- Thelen, K. (2009). Institutional change in advanced political economies. *British Journal of Industrial Relations*, 47, 471-498. <https://doi.org/10.1111/j.1467-8543.2009.00746.x>
- Tiebout, C. M. (1956). A Pure Theory of Local Expenditures. *Journal of Political Economy*, 64(5), 416-424. <https://doi.org/10.1086/257839>
- Torche, F. (2014). Intergenerational Mobility and Inequality: The Latin American Case. *Annual Review of Sociology*, 40(1), 619-642. <https://doi.org/10.1146/annurev-soc-071811-145521>
- Torres, J., y Corporán, J. L. (2021). *Producción de pobreza en el mundo del trabajo*.
- Wallis, J. J., y Oates, W. E. (1988). Decentralization in the public sector: An empirical study of state and local government. En *Fiscal federalism: Quantitative studies* (pp. 5-32). University of Chicago Press.
- Zhao, L. (2020). Tourism, Institutions, and Poverty Alleviation: Empirical Evidence from China. *Journal of Travel Research*, 60(7), 1543-1565. <https://doi.org/10.1177/0047287520947792>

Anexos

	Especificación	Test de Hausman	Corrección GL	GL1	GL2	P-valor
1	Log(Tur (1))	2.28	CR2	7	4.38	0.21
2	Log(Tur (2))	1.27	CR2	7	4.04	0.43
3	Log(Tur (3))	3.34	CR2	7	4.65	0.11
4	Log(Tur (4))	2.84	CR2	7	4.48	0.15

Tabla 8. Prueba de Hausman Robusta. Fuente: Elaboración propia. Nota: CR2= Corrección de Pustejovsky & Tipton, (2018)

Var. Dep.	Test de Breuch-Pagan	GL	P-valor
1 gini	9.76	8.00	0.28
2 palma	12.63	8.00	0.13
3 rkuznets	17.11	8.00	0.03
4 q5q4	9.90	8.00	0.27
5 q4q3	6.42	8.00	0.60
6 q3q2	17.83	8.00	0.02
7 q2q1	15.31	8.00	0.05

Ilustración 20. Prueba de Breuch Pagan de Homocedasticidad para modelos de panel para Rep. Dom. Fuente: Elaboración propia.

Var. Dep.	Test de Breuch-Godfrey	GL1	GL2	P-valor
1 gini	11.88	2	59	0.00
2 palma	12.47	2	59	0.00
3 rkuznets	10.11	2	59	0.00
4 q5q4	12.11	2	59	0.00
5 q4q3	1.08	2	59	0.35
6 q3q2	0.03	2	59	0.97
7 q2q1	1.06	2	59	0.35

Ilustración 21. Tabla con resultados de prueba de Autocorrelación de Breuch-Godfrey/Wooldridge para datos en panel. Fuente: Elaboración propia. Nota: Se estimó para un componente autorregresivo (AR) de orden 2.

Var. Dep.	Test de Hausman	Corrección GL	GL1	GL2	P-valor
1 gini	0.35	CR2	6	0.65	0.86
2 palma	0.24	CR2	6	0.73	0.91
3 rkuznets	0.18	CR2	6	0.68	0.94
4 q5q4	0.62	CR2	6	0.75	0.76
5 q4q3	0.73	CR2	6	0.74	0.74
6 q3q2	3.75	CR2	6	0.68	0.47
7 q2q1	0.25	CR2	6	0.65	0.90

Ilustración 22. Prueba de Hausman Robusta para modelos de panel de Rep. Dom. Fuente: Elaboración propia.

Código de R utilizado para el análisis del dato:

```
##1.0 Análisis para La tesis####
library(haven)
base_final_tesis <- read_dta("G:/Mi unidad/Tesis de maestría estadística/base_final_tesis.dta")
#View(base_final_tesis)
setwd("G:/Mi unidad/Tesis de maestría estadística")

##1.1 Librerías a cargar####
library(writexl)
library(PASWR)
library(papaja)
library(rstatix)
library(GGally)
library(readr)
library(xtable)
library(clusterend)
library(factoextra)
library(NbClust)
library(fpc)
library(dplyr)
library(ggplot2)
library(lmtest)

##2.0 PCA/Cluster analysis:####

#Primer paso, vamos a renombrar y hacer una matriz con Las variables que componen el índices
# Es decir, hacer una notación para así manejanos mejor con qué significa y para simplificar
# La cantidad de Letras a poner y minimizar La cantidad de código:

n_ori<-c("n_instdepth", "n_policyautonomy", "n_fiscalautonomy", "n_borrowautonomy",
         "n_representation", "n_lawmaking", "n_execcontrol", "n_fiscalcontrol", "n_borrowcontrol",
         "n_constitutional")
#Luego creamos el vector de Los nombres nuevos:

n_nue<- paste0("x", 1:10)

#Crear Loop:

for(i in seq_along(n_ori)){
  colnames(base_final_tesis)[colnames(base_final_tesis)==n_ori[i]]<-n_nue[i]
}

##2.1 Generar variable de regiones:####

table(base_final_tesis$country_name, base_final_tesis$iso3n)
suda<-c(32, 68, 76, 152, 170, 218, 600, 604, 858) #sudamérica
caribe<-c(52, 214, 388)
base_final_tesis$region<- ifelse(base_final_tesis$iso3n %in% suda, 1, 0)

base_final_tesis<-base_final_tesis %>%
  mutate(region3= case_when(
    iso3n %in% suda ~ 1,
    iso3n %in% caribe ~ 2,
    TRUE ~ 0 ))

#Etiquetas:
base_final_tesis$region2<-ifelse(base_final_tesis$region==1, "SA", "Caribe y CA")

base_final_tesis<-base_final_tesis %>%
  mutate(regionlab=ifelse(region3==1, "Sudamérica",
                          ifelse(region3==2, "Caribe", "Centroamérica")))

# 3.0 Gráficos y supuestos####
##3.1 Gráfico 1:####

G1<-paste0("x", 1:10)
G2<-paste0("x", 1:5)
G3<-paste0("x",6:10)

ggpairs(base_final_tesis[,c(G1)],
        upper=list(continuous=wrap("cor", size=2)))+
  theme_bw()+
  theme(text=element_text(size=8, family="serif"),
        legend.title = element_text(size=8),
        legend.text=element_text(size=8))
ggsave("graf001.png", dpi=400)
ggpairs(base_final_tesis[,c(G1)],
        upper=list(continuous=wrap("cor", size=3)))+
  theme_bw()+
  theme(text=element_text(size=8),
        legend.title = element_text(size=8),
        legend.text=element_text(size=8))

##3.2 Gráfico 2: regiones y autogobierno.####
ggpairs(base_final_tesis[,c(G2)],
        upper=list(continuous=wrap("cor", size=0.05)),
        aes(col=base_final_tesis$region2))+
  theme_bw()+
  theme(text=element_text(size= 7))
#,
#Legend.title = element_text(size=8),
#Legend.text=element_text(size=8)

##3.3 Gráfico3: Regiones y gobierno compartido####
ggpairs(base_final_tesis[,c(G3)],
        upper=list(continuous=wrap("cor", size=3)),
```

```

aes(col=base_final_tesis$region2))+
theme_bw()+
theme(text=element_text(size=8),
      legend.title = element_text(size=8),
      legend.text=element_text(size=8))
## 3.4 outliers:
taboutlier<-base_final_tesis %>%
  #group_by(region) %>%
  identify_outliers(x1)

tt1<-tibble(taboutlier)
View(tt1)
library(purrr)
# Genera el vector de nombres de las columnas a las que quieres aplicar identify_outliers
variables <- paste0("x", 2:10)
# Define una función para aplicar identify_outliers y conservar el nombre de la variable
identify_outliers_by_region <- function(df, variable) {
  df %>%
  # group_by(!sym(region)) %>%
  identify_outliers(!sym(variable)) %>%
  ungroup() %>%
  # select(-one_of(region)) %>% # Eliminar la columna de región si no la deseas en los resultados
  mutate(variable = variable) # Añade una columna con el nombre de la variable
}

# Aplica la función a cada variable y combina los resultados en un solo dataframe
taboutlier <- map_dfr(variables, ~identify_outliers_by_region(base_final_tesis, .x))
tb1<-table(taboutlier$variable ,taboutlier$is.extreme)
tb1f<-xtable::xtable(tb1)
print(tb1f, type="html", file="tablaoutlier.html")
##4.0 Escalar las variables a utilizar####
stdf<-scale(base_final_tesis[,c(G1)])
##4.1 Tendencia de la agrupación (prueba de Hopkins)###
set.seed(1234)
hopkins::hopkins(stdf, m=nrow(stdf)/10)
clustertend::hopkins(stdf, n=nrow(stdf)/10)
#La data presenta agrupación:
##4.2 Estimar el número óptimo de Clusters: ###
set.seed(1234)
ncl<-NbClust(stdf,
             min.nc = 2,
             max.nc=10,
             method="kmeans",
             index="all")
##4.3 Establecer el clúster: ####
cluster<-kmeans(stdf, centers=3)
variacion<-((cluster$betweenss/cluster$totss)*100)
variacion
##4.4 Interpretación del cluster:####
library(flexclust)
cluster2<-as.kcca(cluster, stdf)
cluster2
barplot(cluster2)
##4.5 Importancia de las variables:
library(FeatureImpCluster)
importancia<-FeatureImpCluster(cluster2, as.data.table(stdf))
plot(importancia)+theme_bw()
##5.0 Validación del cluster####
##5.1 Validación interna: ####
library(cluster)
distancia<-dist(stdf)
fviz_dist(distancia, lab_size = 0.1, show_labels = F)
sk<-silhouette(cluster$cluster, distancia)
fviz_silhouette(sk)+theme_bw()
##5.2 Validez externa####
### 5.2.1 Crear una variable de agrupación del clúster, externa:
# Creemos una variable que escala si el promedio de la muestra:
# a) tiene en promedio más auto gobierno y gobierno compartido que la media
# b) Si tiene mayor puntaje de autogobierno solamente
# c) Si posee más gobierno compartido (por encima del promedio)
# d) Si tiene todos los indicadores por debajo.
base_final_tesis<-base_final_tesis %>%
  mutate(prueba=case_when(n_sharedrule>=1.48 & n_selfrule>=6.81~3,
                          n_sharedrule>=1.48 ~1,n_selfrule>=6.81 ~2, T ~0))
###5.2.1 Estimación índive de Rand###
library(fpc)
inrand<-cluster.stats(d=distancia,
                     alt.clustering=base_final_tesis$prueba,
                     clustering=as.numeric(cluster$cluster))
inrand$corrected.rand
### EL índice es de 0.726, indica buena validez.
##6.0 Cluster de K medianas:####
# Se observó que existen outliers, por lo que usar un método más
#robusto a valores atípicos es necesario.
##6.1 Estimar num. de clusters####
###6.1.1 Método gráfico:####
fviz_nbcust(stdf, pam, method="wss")+theme_apa()+
  labs(title="", y="Suma de cuadrados total intragrupo",
        x= "Número de clústers k")+theme(text=element_text(size=10))
ggsave("pam.png", dpi=300)
###6.1.2 Por método gap stat####
set.seed(1234)
gap_st<-clusGap(stdf,
                FUN=pam,
                K.max= 10,
                B=50)
fviz_gap_stat(gap_st)+theme_apa()+
  labs(title="", y="Estadístico Gap (k)",
        x= "Número de clústers k")+theme(text=element_text(size=10))

```

```

ggsave("gap.png", dpi=300)
##6.2 Estimar Los clusters:####
cluster3<-pam(stdf, 4, metric = "manhattan")
cluster4<-pam(stdf, 3, metric="manhattan")
cluster5<-pam(stdf, 10, metric="manhattan")
### En este caso estimaremos varios en base a Los métodos gráficos:
# uno de 3, 4 y 10 conglomerados.
###6.2.1Validación:####
distancia3<-dist(stdf, method = "manhattan")
fviz_dist(distancia3, lab_size = 0.1, show_labels = F)
sk3<-silhouette(cluster3$clustering, distancia3)
sk4<-silhouette(cluster4$clustering, distancia3)
sk5<-silhouette(cluster5$clustering, distancia3)
summary(sk3)$si.summary[4]
fviz_silhouette(sk3)+theme_apa()+
  labs(title="", subtitle = "Ancho promedio de la silueta: 0.58",
        y="Ancho de la silueta Si",
        x="")+theme(text=element_text(size=14),
                    legend.position = "bottom")
ggsave("silueta.png", dpi=300)
fviz_silhouette(sk4)+theme_apa()+
  labs(title="", subtitle = "Ancho promedio de la silueta: 0.55",
        y="Ancho de la silueta Si",
        x="")+theme(text=element_text(size=14),
                    legend.position = "bottom")
fviz_silhouette(sk5)+theme_apa()+
  labs(title="", subtitle = "Ancho promedio de la silueta: 0.62",
        y="Ancho de la silueta Si",
        x="")+theme(text=element_text(size=14),
                    legend.position = "bottom")
### Validación externa:
inrand2<-cluster.stats(d=distancia3,
                      alt.clustering = base_final_tesis$prueba,
                      clustering = as.numeric(cluster3$clustering))

inrand2
inrand3<-cluster.stats(d=distancia3,
                      alt.clustering = base_final_tesis$prueba,
                      clustering = as.numeric(cluster4$clustering))

inrand3$corrected.rand
inrand4<-cluster.stats(d=distancia3,
                      alt.clustering=base_final_tesis$prueba,
                      clustering=as.numeric(cluster5$clustering))

inrand4
inrand$corrected.rand
### Creación de tabla:
library(papaja)
library(xtable)
tablavalidacion<-data.frame(cluster=c("#1", "#2", "#3"),
                           num_clust=c(max(cluster3$clustering),
                                         max(cluster4$clustering),
                                         max(cluster5$clustering)),
                           silo=c(round(summary(sk3)$si.summary[4],3),
                                   round(summary(sk4)$si.summary[4],3),
                                   round(summary(sk5)$si.summary[4],3)),
                           Rand=c(round(inrand2$corrected.rand,3),
                                   round(inrand3$corrected.rand,3),
                                   round(inrand4$corrected.rand,3)))

rownames(tablavalidacion)<-NULL
v<-c("Cluster", "# de Cluster", "Silueta", "Índice de Rand")
colnames(tablavalidacion)<-v
tablavalidacion2<-xtable::xtable(tablavalidacion)
print(tablavalidacion2, type="html", file="tabla.html")
tablavalidacion
##### Incluir estos clusters en La base original:####
base_final_tesis$c11<-as.factor(cluster3$clustering)
base_final_tesis$c12<-as.factor(cluster4$clustering)
base_final_tesis$c13<-as.factor(cluster5$clustering)
#7.0 Gráficos:####
##7.1 Segundo cluster:####
library(ggplot)
ggplot(base_final_tesis[,c("n_selfrule", "n_sharedrule",
                          "c12", "CountryCode")],
       aes(n_selfrule, n_sharedrule, color=c12))+
  geom_text_repel(aes(label=CountryCode), size=2.0,
                 max.overlaps = 330,
                 box.padding = 0.35)+
  theme_apa()+labs(y="Índice de Autogobierno (RAI)",
                  x="Índice de gobierno compartido (RAI)",
                  color="Conglomerado")+
  theme(legend.position = "bottom",
        text= element_text(size=12),
        axis.title.x = element_text(size=12),
        axis.title.y=element_text(size = 12))+
  scale_color_manual(values=c( "#0000FF88",
                              "#00008B88", "#FF000088"))
ggsave("grafico1.png", dpi=300,
       width=2953, height = 2025, units = "px")
###7.1.1 Desigualdad y descentralización ####
m1<-lm(gini_disp ~ n_sharedrule+I(n_sharedrule^2), data=base_final_tesis)
r1<-summary(m1)$r.squared
summary(m1)
ggplot(base_final_tesis[,c("gini_disp", "n_sharedrule", "c12")],
       aes(n_sharedrule,gini_disp, color=c12))+geom_point()+
  theme_apa()+labs(y="Índice de Gini después de transferencias",
                  x="Índice de autoridad compartida (con el gob. nacional)",
                  color="Conglomerado")+
  geom_smooth(method = "lm", color="black", formula= y~poly(x,2))+
  theme(legend.position = "bottom",text= element_text(size=12))+
  annotate("text", x=max(base_final_tesis$n_sharedrule)*0.95,

```

```

    y=max(base_final_tesis$gini_disp),
    label=sprintf("R²=%.2f", r1),
    hjust=1, vjust=1,
    size=4, color="black", size=8)+
  scale_color_manual(values=c( "#0000FF88",
    "#00008B88", "#FF00088"))
ggsave("grafico2.png", dpi=300,
  width=2953, height = 2025, units = "px")
#### 7.1.1.1 Autogobierno####
m2<-lm(gini_disp ~ n_selfrule+ I(n_selfrule^2)+I(n_selfrule^3), base_final_tesis)
rr2<-summary(m2)$r.squared
ggplot(base_final_tesis[, c("gini_disp", "n_selfrule",
  "c12")],
  aes(n_selfrule, gini_disp, color=c12))+
  geom_point()+theme_apa()+
  labs(y="Índice de Gini después de transferencias",
    x="Índice de autogobierno (RAI)",
    color="Conglomerado")+
  geom_smooth(method="lm", color="black",
    formula= y ~poly(x, 3))+
  theme(legend.position = "bottom",
    text= element_text(size=12),
    axis.title.x = element_text(size=12),
    axis.title.y=element_text(size = 12)
  )+ annotate("text", x=max(base_final_tesis$n_selfrule)*0.15,
    y=max(base_final_tesis$gini_disp),
  label=sprintf("R²=%.2f", rr2),
  hjust=1, vjust=1, size=4, color="black")+
  scale_color_manual(values=c( "#0000FF88",
    "#00008B88", "#FF00088"))
ggsave("grafico3.png", dpi=300,
  width=2953, height = 2025, units = "px")

ggplot(base_final_tesis[,c("gini_disp", "regionlab", "c12")],
  aes(y=gini_disp, x=c12, fill=regionlab))+
  geom_violin()+theme_apa()+labs(y="Índice de Gini después de transferencias",
    x="Conglomerado",
    fill= "")+
  theme(legend.position = "bottom", text= element_text(size=10),
    axis.title.x = element_text(size=10),
    axis.title.y=element_text(size = 10))+
  scale_fill_manual(values=c( "#0000FF88",
    "#00008B88", "#FF00088"))
ggsave("grafico4.png", dpi=300,
  width=2953, height = 2025, units = "px") # Este gráfico no va.
####7.1.2 Relacionadas a turismo:####
modelo<-lm(gini_disp~log(N)+I(log(N)^2), data=base_final_tesis)
r2<-summary(modelo)$r.squared
summary(modelo)

ggplot(base_final_tesis[,c("gini_disp", "N", "c12")],
  aes(log(N),gini_disp, color=c12))+geom_point()+
  theme_apa()+labs(y="Índice de Gini después de transferencias",
    x="Log(Ingresos por artículos de viaje en USD)",
    color="Conglomerado")+
  geom_smooth(method = "lm", color="black",
    formula=y~x+I(x^2))+theme(legend.position = "bottom",
    text= element_text(size=12))+
  annotate("text", x=max(log(base_final_tesis$N))*0.95, y=max(base_final_tesis$gini_disp),
    label=sprintf("R²=%.2f", r2),
    hjust=1, vjust=1,
    size=4, color="black")+
  scale_color_manual(values=c( "#0000FF88",
    "#00008B88", "#FF00088"))

ggsave("grafico5.png", dpi=300,
  width=2953, height = 2025, units = "px")

modelo222<-lm(gini_disp~log(G)+I(log(G)^2), data=base_final_tesis)
r222<-summary(modelo222)$r.squared
summary(modelo222)

ggplot(base_final_tesis[,c("gini_disp", "G", "c12")],
  aes(log(G),gini_disp, color=c12))+geom_point()+
  theme_apa()+labs(y="Índice de Gini después de transferencias",
    x="Log(Ing. por gastos por artículos para el transporte de pasajeros)",
    color="Conglomerado")+
  geom_smooth(method = "lm", color="black",
    formula=y~x+I(x^2))+theme(legend.position = "bottom",
    text= element_text(size=10))+
  annotate("text", x=max(log(base_final_tesis$G))*0.99,
    y=max(base_final_tesis$gini_disp),
    label=sprintf("R²=%.2f", r222),
    hjust=1, vjust=1,
    size=4, color="black")+
  scale_color_manual(values=c( "#0000FF88",
    "#00008B88", "#FF00088"))

ggsave("grafico5x2.png", dpi=300,
  width=2953, height = 2025, units = "px")
#####
modelo23<-lm(gini_disp~log(F)+
  I(log(F)^2), data=base_final_tesis)
r23<-summary(modelo23)$r.squared
summary(modelo23)

ggplot(base_final_tesis[,c("gini_disp",

```

```

        "F", "c12"]],
    aes(log(F),gini_disp, color=c12))+geom_point()+
  theme_apa()+labs(y="Índice de Gini después de transferencias",
    x="Log(Gasto turístico asociado a transporte internacional, USD)",
    color="Conglomerado")+
  geom_smooth(method = "lm", color="black",
    formula=y~x+I(x^2))+theme(legend.position = "bottom",
    text=element_text(size=10))+
  annotate("text", x=max(log(base_final_tesis$F))*0.99,
    y=max(base_final_tesis$gini_disp),
    label=sprintf("R²=%.2f", r23),
    hjust=1, vjust=1,
    size=4, color="black")+
  scale_color_manual(values=c( "#0000FF88",
    "#00008B88", "#FF00088"))

ggsave("grafico5x3.png", dpi=300,
  width=1953, height = 2025, units = "px")

#### 7.1.2.1Turismo y descentralización:####
##### Autoridad Regional#####
m3<-lm(log(N)~n_RAI+I(n_RAI^2), data=base_final_tesis)
r3<-summary(m3)$r.squared
summary(m3)
ggplot(base_final_tesis[,c("N", "n_RAI", "c12", "gini_disp")],
  aes(n_RAI,log(N), color=c12, size=gini_disp))+geom_point()+
  theme_apa()+labs(
    y="Log(Ingresos por artículos de viaje en USD)",
    x= "Índice de Autoridad Regional (RAI)",
    color="Conglomerado",
    size= "Índice de Gini")+
  geom_smooth(method = "lm", color="black",
    formula=y~poly(x, 2))+theme(legend.position = "bottom",
    text=element_text(size = 12))+
  annotate("text", x=max(base_final_tesis$n_RAI)*0.15, y=max(log(base_final_tesis$N)),
    label=sprintf("R²=%.2f", r3),
    hjust=1, vjust=1,
    size=4, color="black")+
  scale_color_manual(values=c( "#0000FF88",
    "#00008B88", "#FF00088"))
ggsave("grafico6.png", dpi=300,
  width=2953, height = 2025, units = "px")

#####gobierno compartido####
m4<-lm(log(N)~ n_sharedrule, data=base_final_tesis)
r4<-summary(m4)$r.squared
summary(m4)

ggplot(base_final_tesis[,c("N", "n_sharedrule", "c12", "gini_disp")],
  aes(n_sharedrule,log(N), color=c12, size=gini_disp))+
  geom_point()+
  theme_apa()+labs(
    y="Log(Ingresos por artículos de viaje en USD)",
    x= "Índice de Autoridad compartida (RAI)",
    color="Conglomerado",
    size= "Índice de Gini")+
  geom_smooth(method = "lm", color="black",
    formula=y~x)+
  theme(legend.position = "bottom",
    text=element_text(
      size=12))+
  annotate("text", x=max(base_final_tesis$n_sharedrule)*0.95,
    y=max(log(base_final_tesis$N)),
    label=sprintf("R²=%.2f", r4),
    hjust=1, vjust=1,
    size=4, color="black")+
  scale_color_manual(values=c( "#0000FF88",
    "#00008B88", "#FF00088"))
ggsave("grafico7.png", dpi=300,
  width=2953, height = 2025, units = "px")

#####Autogobierno####
m5<-lm(log(N)~ poly(n_selfrule,2), data=base_final_tesis)
r5<-summary(m4)$r.squared
summary(m5)

ggplot(base_final_tesis[,c("N", "n_selfrule", "c12", "gini_disp")],
  aes(n_selfrule,log(N), color=c12, size=gini_disp))+geom_point()+
  theme_apa()+labs(
    y="Log(Ingresos por artículos de viaje en USD)",
    x= "Índice de Autogobierno (RAI)",
    color="Conglomerado",
    size= "Índice de Gini")+
  geom_smooth(method = "lm", color="black",
    formula=y~poly(x,2))+theme(legend.position = "bottom",
    text=element_text(
      size=12))+
  annotate("text", x=max(base_final_tesis$n_selfrule)*0.95,
    y=max(log(base_final_tesis$N)),
    label=sprintf("R²=%.2f", r5),
    hjust=1, vjust=1,
    size=4, color="black")+
  scale_color_manual(values=c( "#0000FF88",
    "#00008B88", "#FF00088"))

```

```

ggsave("grafico8.png", dpi=300,
       width=2953, height = 2025, units = "px")
####7.1.3 Democracia y descentralización:
#### Autoridad regional
m6<-lm(vdem_polyarchy~n_RAI+I(n_RAI^2), base_final_tesis)
summary(m6)
r6<-summary(m6)$r.squared
ggplot(base_final_tesis[,c("vdem_polyarchy", "n_RAI",
                          "c12", "gini_disp")],
       aes(n_RAI, vdem_polyarchy,color=c12, size=gini_disp))+
  theme_apa()+geom_point()+labs(y="Índice de democracia electoral (VDEM)",
                                x= "Índice de Autoridad Regional (RAI)",
                                color="Conglomerado",
                                size= "Índice de Gini")+
  geom_smooth(method = "lm", color="black",
             formula=y~poly(x,2))+theme(legend.position = "bottom",
                                       text=element_text(
                                         size=12))+
  annotate("text", x=max(base_final_tesis$n_RAI)*0.95,
            y=max(base_final_tesis$vdem_polyarchy)*.50,
            label=sprintf("R²=%.2f", r6),
            hjust=1, vjust=1,
            size=4, color="black")+
  scale_color_manual(values=c( "#0000FF88",
                              "#00008B88", "#FF000088"))
ggsave("grafico9.png", dpi=300,
       width=2953, height = 2025, units = "px")

####Gobierno compartido: ####

m7<-lm(vdem_polyarchy~n_sharedrule, base_final_tesis)
summary(m7)
r7<-summary(m7)$r.squared
ggplot(base_final_tesis[,c("vdem_polyarchy", "n_sharedrule",
                          "c12", "gini_disp")],
       aes(n_sharedrule, vdem_polyarchy,color=c12, size=gini_disp))+
  theme_apa()+geom_point()+labs(y="Índice de democracia electoral (VDEM)",
                                x= "Índice de Autoridad Compartida (RAI)",
                                color="Conglomerado",
                                size= "Índice de Gini")+
  geom_smooth(method = "lm", color="black",
             formula=y~poly(x,2))+theme(legend.position = "bottom",
                                       text=element_text(
                                         size=12))+
  annotate("text", x=max(base_final_tesis$n_sharedrule)*0.95,
            y=max(base_final_tesis$vdem_polyarchy)*.50,
            label=sprintf("R²=%.2f", r7),
            hjust=1, vjust=1,
            size=4, color="black")+
  scale_color_manual(values=c( "#0000FF88",
                              "#00008B88", "#FF000088"))

ggsave("grafico10.png", dpi=300,
       width=2953, height = 2025, units = "px")

#### Autogobierno####

m8<-lm(vdem_polyarchy~n_selfrule+I(n_selfrule^2), base_final_tesis)
summary(m8)
r8<-summary(m8)$r.squared
ggplot(base_final_tesis[,c("vdem_polyarchy", "n_selfrule",
                          "c12", "gini_disp")],
       aes(n_selfrule, vdem_polyarchy,color=c12, size=gini_disp))+
  theme_apa()+geom_point()+labs(y="Índice de democracia electoral (VDEM)",
                                x= "Índice de Autogobierno (RAI)",
                                color="Conglomerado",
                                size= "Índice de Gini")+
  geom_smooth(method = "lm", color="black",
             formula=y~poly(x,2))+theme(legend.position = "bottom",
                                       text=element_text(
                                         size=12))+
  annotate("text", x=max(base_final_tesis$n_selfrule)*0.95,
            y=max(base_final_tesis$vdem_polyarchy)*.50,
            label=sprintf("R²=%.2f", r8),
            hjust=1, vjust=1,
            size=4, color="black")+
  scale_color_manual(values=c( "#0000FF88",
                              "#00008B88", "#FF000088"))

ggsave("grafico11.png", dpi=300,
       width=2953, height = 2025, units = "px")

####7.1.3 Elemento participativo####
####Autoridad regional####

m9<-lm(vdem_partip~n_RAI+I(n_RAI^2), base_final_tesis)
summary(m9)
r9<-summary(m9)$r.squared
ggplot(base_final_tesis[,c("vdem_partip", "n_RAI",
                          "c12", "gini_disp")],
       aes(n_RAI, vdem_partip,color=c12, size=gini_disp))+
  theme_apa()+geom_point()+labs(y="Índice de componentes participativos (VDEM)",
                                x= "Índice de Autoridad Regional (RAI)",
                                color="Conglomerado",
                                size= "Índice de Gini")+
  geom_smooth(method = "lm", color="black",

```

```

        formula=y~poly(x,2))+theme(legend.position = "bottom",
        text=element_text(
        size=12))+
    annotate("text", x=max(base_final_tesis$n_RAI)*0.95,
    y=max(base_final_tesis$vdem_partip)*.50,
    label=sprintf("R²=%.2f", r9),
    hjust=1, vjust=1,
    size=4, color="black")+
    scale_color_manual(values=c( "#0000FF88",
    "#00008B88", "#FF000088"))

ggsave("grafico12.png", dpi=300,
    width=2953, height = 2025, units = "px")

m9x<-lm(vdem_partip~n_sharedrule+I(n_sharedrule^2), base_final_tesis)
summary(m9x)
r9x<-summary(m9x)$r.squared
ggplot(base_final_tesis[,c("vdem_partip", "n_sharedrule",
    "c12", "gini_disp")],
    aes(n_sharedrule, vdem_partip,color=c12, size=gini_disp))+
    theme_apa()+geom_point()+labs(y="Índice de componentes participativos (VDEM)",
    x= "Índice de Autoridad compartida (RAI)",
    color="Conglomerado",
    size= "Índice de Gini")+
    geom_smooth(method = "lm", color="black",
    formula=y~poly(x,2))+theme(legend.position = "bottom",
    text=element_text(
    size=12))+
    annotate("text", x=max(base_final_tesis$n_sharedrule)*0.95,
    y=max(base_final_tesis$vdem_partip)*.50,
    label=sprintf("R²=%.2f", r9x),
    hjust=1, vjust=1,
    size=4, color="black")+
    scale_color_manual(values=c( "#0000FF88",
    "#00008B88", "#FF000088"))

ggsave("grafico12x.png", dpi=300,
    width=2953, height = 2025, units = "px")

m9z<-lm(gini_disp~vdem_partip+I(vdem_partip^2), base_final_tesis)
summary(m9z)
r9z<-summary(m9z)$r.squared
ggplot(base_final_tesis[,c("vdem_partip", "n_RAI",
    "c12", "gini_disp")],
    aes(vdem_partip, gini_disp,color=c12, size=n_RAI))+
    theme_apa()+geom_point()+labs(y="Índice de Gini después de transferencias",
    x= "Índice de componentes participativos (VDEM)",
    color="Conglomerado",
    size= "Autoridad Regional (RAI)")+
    geom_smooth(method = "lm", color="black",
    formula=y~poly(x,2))+theme(legend.position = "bottom",
    text=element_text(
    size=12))+
    annotate("text", x=max(base_final_tesis$vdem_partip)*0.95,
    y=max(base_final_tesis$gini_disp),
    label=sprintf("R²=%.2f", r9z),
    hjust=1, vjust=1,
    size=4, color="black")+
    scale_color_manual(values=c( "#0000FF88",
    "#00008B88", "#FF000088"))

ggsave("grafico12x.png", dpi=300,
    width=2953, height = 2025, units = "px")

###7.2 Crecimiento económico:###

m10<-lm(log(GDPpercapitaPPPconstant20)~n_RAI+I(n_RAI^2), base_final_tesis)
summary(m10)
r10<-summary(m10)$r.squared
ggplot(base_final_tesis[,c("GDPpercapitaPPPconstant20", "n_RAI",
    "c12", "gini_disp")],
    aes(n_RAI, log(GDPpercapitaPPPconstant20),color=c12, size=gini_disp))+
    theme_apa()+geom_point()+labs(y="Log(PIB per cápita USD (t=2017=100))",
    x= "Índice de Autoridad Regional (RAI)",
    color="Conglomerado",
    size= "Índice de Gini")+
    geom_smooth(method = "lm", color="black",
    formula=y~poly(x,2))+theme(legend.position = "bottom",
    text=element_text(
    size=12))+
    annotate("text", x=max(base_final_tesis$n_RAI)*0.95,
    y=max(log(base_final_tesis$GDPpercapitaPPPconstant20)),
    label=sprintf("R²=%.2f", r10),
    hjust=1, vjust=1,
    size=4, color="black")+
    scale_color_manual(values=c( "#0000FF88",
    "#00008B88", "#FF000088"))

ggsave('grafico13.png', dpi=300,
    width=2953, height = 2025, units = "px")

summary(lm(gini_disp~poly(Agedependencyratioofworki,2), base_final_tesis))

```

```

summary(lm(gini_disp~GDPperpersonemployedconstan, base_final_tesis))

m14<-lm(gini_disp~log(N)+I(log(N)^2), base_final_tesis)
summary(m14)
r14<-summary(m14)$r.squared
ggplot(base_final_tesis[,c("N", "vdem_pubcorr",
                           "c12", "gini_disp")],
        aes(log(N),gini_disp, color=c12, size=vdem_pubcorr))+
  theme_apa()+geom_point()+labs(y="Índice de Gini después de transferencias",
                                x="Log(ingresos por artículos de viaje en USD)",
                                color="Conglomerado",
                                size="Corrupción Púb. (VDEM)")+
  geom_smooth(method = "lm", color="black",
              formula=y~poly(x,2))+theme(legend.position = "bottom",
                                         text=element_text(
                                           size=12))+
  annotate("text", x=max(log(base_final_tesis$N))*0.95,
           y=max(base_final_tesis$gini_disp),
           label=sprintf("R²=%.2f", r14),
           hjust=1, vjust=1,
           size=4, color="black")+
  scale_color_manual(values=c( "#0000FF88",
                              "#00008B88", "#FF000088"))

ggsave('grafico14.png', dpi=300,
        width=2953, height = 2325, units = "px")

m16<-lm(log(GDPperpersonemployedconstan)~vdem_pubcorr, base_final_tesis)
summary(m16)
r16<-summary(m16)$r.squared

ggplot(base_final_tesis[,c("n_sharedrule", "vdem_pubcorr",
                           "c12", "GDPperpersonemployedconstan")],
        aes(vdem_pubcorr,log(GDPperpersonemployedconstan), color=c12, size=n_sharedrule))+
  theme_apa()+geom_point()+labs(y="Log(PIB por persona empleada)",
                                x="Índice de Corrupción Pública (VDEM)",
                                color="Conglomerado",
                                size="Autoridad compartida"
                                )+
  geom_smooth(method = "lm", color="black")+theme(legend.position = "bottom",
                                         text=element_text(
                                           size=10))+
  annotate("text", x=max(base_final_tesis$vdem_pubcorr)*0.95,
           y=max(log(base_final_tesis$GDPperpersonemployedconstan)),
           label=sprintf("R²=%.2f", r16),
           hjust=1, vjust=1,
           size=4, color="black")+
  scale_color_manual(values=c( "#0000FF88",
                              "#00008B88", "#FF000088"))

ggsave('grafico16.png', dpi=300,
        width=2253, height = 2325, units = "px")

##### Productividad#####

ggplot(base_final_tesis[,c("n_selfrule", "vdem_pubcorr",
                           "c12", "GDPperpersonemployedconstan")],
        aes(vdem_pubcorr,log(GDPperpersonemployedconstan), color=c12, size=n_selfrule))+
  theme_apa()+geom_point()+labs(y="Log(PIB por persona empleada)",
                                x="Índice de Corrupción Pública (VDEM)",
                                color="Conglomerado",
                                size="Autoridad propia"
                                )+
  geom_smooth(method = "lm", color="black")+theme(legend.position = "bottom",
                                         text=element_text(
                                           size=10))+
  annotate("text", x=max(base_final_tesis$vdem_pubcorr)*0.95,
           y=max(log(base_final_tesis$GDPperpersonemployedconstan)),
           label=sprintf("R²=%.2f", r16),
           hjust=1, vjust=1,
           size=4, color="black")+
  scale_color_manual(values=c( "#0000FF88",
                              "#00008B88", "#FF000088"))

ggsave('grafico17.png', dpi=300,
        width=2253, height = 2325, units = "px")

###8.0 Regresiones de panel###
library(plm)
library(stargazer)
library(clubSandwich)
#install.packages("pder")
library(pder)
panelid<-c("iso3n", "year")
#Funcion RMSE:
library(lmtest)

pd<-pdata.frame(base_final_tesis, index = panelid)

esp1<-log(gini_disp)~log(N)+I(log(N)^2)+n_RAI+

```

```

vdem_pubcorr+vdem_polyarchy+vdem_partip+log(GDPpercapitaPPPcurrentint)+
log(PopulationtotalSPPOPTOTL)+Generalgovernmentfinalconsump

esp2<-log(gini_disp)~log(N)+I(log(N)^2)+n_RAI+
vdem_pubcorr+vdem_polyarchy+vdem_partip+log(GDPpercapitaPPPcurrentint)+
log(PopulationtotalSPPOPTOTL)+Generalgovernmentfinalconsump+as.factor(c12)

esp2f<-log(gini_disp)~log(N)+I(log(N)^2)+n_sharedrule+n_selfrule+
vdem_pubcorr+vdem_polyarchy+vdem_partip+log(GDPpercapitaPPPconstant20)+
log(PopulationtotalSPPOPTOTL)+as.factor(c12)
esp3<-log(gini_disp)~log(N)+I(log(N)^2)+n_sharedrule+n_selfrule+
vdem_pubcorr+vdem_polyarchy+vdem_partip+log(GDPpercapitaPPPconstant20)+
log(PopulationtotalSPPOPTOTL)+Generalgovernmentfinalconsump+as.factor(c12)

esp4<-log(gini_disp)~log(N)+I(log(N)^2)+n_sharedrule+n_selfrule+
vdem_pubcorr+vdem_polyarchy+vdem_partip+log(GDPpercapitaPPPconstant20)+
log(PopulationtotalSPPOPTOTL)+Generalgovernmentfinalconsump

modelo1<-plm(esp2, data= pd, index=panelid,
model= "pooling")
summary(modelo1)

modelo2<-plm(esp1, data=pd, index=panelid, model="within")
summary(modelo2)
modelo2tw<-plm(esp1, data=pd,
index=panelid, model="within", effect = "twoways")

summary(modelo2tw)
pFtest( modelo2, modelo1)

pFtest(modelo2tw, modelo1)

#Test de de multiplicador de Lagrange:
plmtest(modelo1,type="bp")

modelo3<-plm(esp2, data=pd,
model="random", index=panelid)
summary(modelo3)

modelo4<-plm(esp2, data=base_final_tesis,
model="random", index=panelid, effect="twoways")

summary(modelo4)
####Heterocedasticidad:####
bptest(modelo1, studentize = T)
bptest(modelo2, studentize = T)
bptest(modelo2tw, studentize = T)
bptest(modelo3, studentize = T)
bptest(modelo4, studentize = T)
####AutoCorrelación serial:####
pwartest(esp1, data=pd) #efectos fijos.

pwfdtest(esp1, data=pd) #Primeras diferencias:
pwfdtest(esp1, data=pd, h0="fe")
####Correlación contemporánea:####

cd1<-pcdtest(modelo1, test= "lm")
cd1

cbind(
"rho"=pcdtest(modelo1, test="rho")$statistic,
"|rho"=pcdtest(modelo1, test="absrho")$statistic)

pcdtest(modelo3)
pcdtest(modelo2tw, type="bcsc1m")
cbind(
"rho"=pcdtest(modelo2tw, test="rho")$statistic,
"|rho"=pcdtest(modelo2tw, test="absrho")$statistic)

pcdtest(modelo4, type="bcsc1m")
cbind(
"rho"=pcdtest(modelo4, test="rho")$statistic,
"|rho"=pcdtest(modelo4, test="absrho")$statistic
)

###Prueba de Hausman robusta:###

#Primero, estimar within effects:

pd<-within(pd,{
wln<-log(N)-tapply(log(N), iso3n, mean)[factor(iso3n)]
wlg<-log(G)-tapply(log(G), iso3n, mean)[factor(iso3n)]
wlint<-log(Internationaltourismexpenditu)-tapply(log(
Internationaltourismexpenditu), iso3n, mean)[factor(iso3n)]
wlf<-log(F)-tapply(log(F), iso3n, mean)[factor(iso3n)]
wsr<-n_sharedrule-tapply(n_sharedrule, iso3n, mean)[factor(iso3n)]
wsfrM0<-n_selfrule-tapply(n_selfrule, iso3n, mean)[factor(iso3n)]
wrai<-n_RAI-tapply(n_RAI, iso3n, mean)[factor(iso3n)]
wpcvd<-vdem_pubcorr-tapply(vdem_pubcorr, iso3n, mean)[factor(iso3n)]
wplvd<-vdem_polyarchy-tapply(vdem_polyarchy, iso3n, mean)[factor(iso3n)]

```

```

wlpob<-log(PopulationtotalSPPOPTOTL)-tapply(log(PopulationtotalSPPOPTOTL),
iso3n, mean)[factor(iso3n)]
wlpib<-log(GDPpercapitaPPPcurrentint)-tapply(log(GDPpercapitaPPPcurrentint),
iso3n, mean)[factor(iso3n)]
# wipcpp<-PriceLevelratioofPPPconvers-tapply(PriceLevelratioofPPPconvers,
# iso3n, mean)[factor(iso3n)]
wginil<-lag(log(gini_disp))-tapply(lag(log(gini_disp)),iso3n, mean)[factor(iso3n)]
wgc<-Generalgovernmentfinalconsump-tapply(Generalgovernmentfinalconsump, iso3n, mean)[factor(iso3n)]

})
esph1<-gini_disp ~ log(N)+I(log(N)^2) + n_RAI + vdem_pubcorr + vdem_polyarchy +
vdem_partip + log(GDPpercapitaPPPcurrentint) +
log(PopulationtotalSPPOPTOTL) +
Generalgovernmentfinalconsump+ wrai+wLN+I(wLN^2)+wpcvd+wplvd+wlpib+wgc+
as.factor(c12)

esph2<-gini_disp ~ log(G)+I(log(G)^2) + n_RAI + vdem_pubcorr + vdem_polyarchy +
vdem_partip + log(GDPpercapitaPPPcurrentint) +
log(PopulationtotalSPPOPTOTL) +
Generalgovernmentfinalconsump+ wrai+wLG+I(wLG^2)+wpcvd+wplvd+wlpib+wgc+
as.factor(c11)
esph3<-gini_disp ~ log(Internationaltourismexpenditu)+
I(log(Internationaltourismexpenditu)^2) + n_RAI +
vdem_pubcorr + vdem_polyarchy +
vdem_partip + log(GDPpercapitaPPPcurrentint) +
log(PopulationtotalSPPOPTOTL) +
Generalgovernmentfinalconsump+ wrai+wInt+I(wInt^2)+wpcvd+wplvd+wlpib+wgc+
as.factor(c12)
esph4<-gini_disp ~ log(F)+
I(log(F)^2) + n_RAI +
vdem_pubcorr + vdem_polyarchy +
vdem_partip + log(GDPpercapitaPPPcurrentint) +
log(PopulationtotalSPPOPTOTL) +
Generalgovernmentfinalconsump+ wrai+wlf+I(wlf^2)+wpcvd+wplvd+wlpib+wgc+
as.factor(c12)

# A partir de ahí, incluir dentro del Hausman rob
hausmanrobusto<-plm(esph1,
data=pd, model="random", index = panelid)
hausmanrobusto2<-plm(esph2, data=pd,
model="random", index=panelid)

hausmanrobusto3<-plm(esph3, data=pd,
model="random",
index=panelid)
hausmanrobusto4<-plm(esph4, data=pd,
model="random",
index=panelid)
#( "wrai", "poly(wLN,2)", "wpcvd", "wplvd", "wlpib", "wipcpp")
test1<-Wald_test(hausmanrobusto,
constraints = constrain_zero(c("wrai", "wLN",
"I(wLN^2)", "wpcvd", "wplvd", "wlpib", "wgc")),
vcov="CR2")
test1

test2<-Wald_test(hausmanrobusto2,
constraint=constrain_zero(c("wrai", "wLG",
"I(wLG^2)", "wpcvd", "wplvd", "wlpib", "wgc")),
vcov="CR2")

test2

test3<-Wald_test(hausmanrobusto3,
constraints =constrain_zero(c("wrai", "wInt",
"I(wInt^2)", "wpcvd", "wplvd", "wlpib", "wgc")),
vcov="CR2")

test3

test4<-Wald_test(hausmanrobusto4,
constraints =constrain_zero(c("wrai", "wlf",
"I(wlf^2)", "wpcvd", "wplvd", "wlpib", "wgc")),
vcov="CR2")

test4

tdhmr<-data.frame(Especificación=c("Log(Tur (1))",
"Log(Tur (2))",
"Log(Tur (3))",
"Log(Tur (4))"),
Test_hausman=c(test1$Fstat, test2$Fstat,
test3$Fstat,
test4$Fstat),
Corrección=c("CR2", "CR2", "CR2", "CR2"),
GL1=c(test1$df_num,
test2$df_num,
test3$df_num,
test4$df_num),
GL2=c(test1$df_denom,
test2$df_denom,
test3$df_denom,
test4$df_denom),
p_valor=c(test1$p_val,
test2$p_val,

```

```

test3$p_val,
test4$p_val))

colnames(tdhmr)<-c("Especificación", "Test de Hausman",
"Corrección GL","GL1", "GL2","P-valor")
print(xtable::xtable(tdhmr), type="html", file= "hausman.html")
# 9.0 Tabla de regresión ####

##9.1 Método de introducción de variables:

m1<-log(gini_disp)~log(G)+ I(log(G)^2) +as.factor(c11)
m2<-log(gini_disp)~log(G)+ I(log(G)^2)+ n_sharedrule+n_selfrule+as.factor(c11)
m3<-log(gini_disp)~log(G)+ I(log(G)^2)+n_sharedrule+n_selfrule+vdem_polyarchy +
vdem_pubcorr+vdem_partip + as.factor(c11)
m4<-log(gini_disp)~log(G)+ I(log(G)^2)+n_sharedrule+n_selfrule+ vdem_polyarchy +
vdem_pubcorr+ vdem_partip + log(GDPpercapitaPPPcurrentint)+ log(PopulationtotalSPPOPTOTL) +
Generalgovernmentfinalconsump+as.factor(c11)
m5<-log(gini_disp)~log(G)+ I(log(G)^2)+n_RAI+ vdem_polyarchy + vdem_pubcorr+
vdem_partip + log(GDPpercapitaPPPcurrentint)+ log(PopulationtotalSPPOPTOTL) +
Generalgovernmentfinalconsump+as.factor(c11)

#Función para calcular error cuadrático medio:
crmse<-function(model){
sqrt(mean(residuals(model)^2))
}

M<-list(m1, m2, m3, m4, m5)

mre<-list()
vcov<-list()
se<-list()
pv<-list()
RMSE<-list()
ftest<-list()
fr<-list()
for(i in seq_along(M)){
formula<-M[[i]]
mre[[i]]<-plm(formula, data=pd, model="random", index=panelid, effect =
"twoways")
vcov[[i]]<-coefest(mre[[i]], function(x) vcovSCC(x, cluster="time"))
ftest[[i]]<-pwaldtest(mre[[i]], test="F", function(x) vcovSCC(x, cluster="time"))
se[[i]]<-vcov[[i]][,2]
pv[[i]]<-vcov[[i]][,4]
fr[[i]]<-ftest[[i]]$statistic
RMSE[[i]]<-crmse(mre[[i]])
}

stargazer(mre, type="text",
se=se,
p=pv,
star.cutoffs = c(0.1, 0.05, 0.01),
dep.var.caption = "",
dep.var.labels = c("Log(Gini Disp)"),
covariate.labels = c("Log(Tur(1))",
"Log(Tur(1))^2",
"Autoridad Comp.",
"Autogobierno",
"Índice RAI",
"Poliarquía (VDEM)",
"Corrupción (VDEM)",
"Partip. Pol. (VDEM)",
"Log(PIB pc PPP USD)",
"Log(POB)",
"Gob. Consumo (%)",
add.lines = list(c("RMSE", round(RMSE[[1]],3),
round(RMSE[[2]],3),
round(RMSE[[3]],3),
round(RMSE[[4]],3),
round(RMSE[[5]],3)
),
c("F-Statistic", round(fr[[1]],2),
round(fr[[2]],2), round(fr[[3]],2),
round(fr[[4]],2), round(fr[[5]],2)),
c("Países", 19, 19, 19, 19, 19),
c("Cluster 1 FE", "Si", "Si", "Si", "Si", "Si")),
omit.stat = c("F"),
notes = "Errores estándares robustos por Driscoll y Kraay entre paréntesis")

stargazer(mre, type="html", out="tablareg.html",
se=se,
p=pv,
star.cutoffs = c(0.1, 0.05, 0.01),
dep.var.caption = "",
dep.var.labels = c("Log(Gini Disp)"),
covariate.labels = c("Log(Tur(1))",
"Log(Tur(1))^2",
"Autoridad Comp.",
"Autogobierno",
"Índice RAI",
"Poliarquía (VDEM)",
"Corrupción (VDEM)",
"Partip. Pol. (VDEM)",
"Log(PIB pc PPP USD)",
"Log(POB)",

```

```

        "Gov. Consumo (%)",
add.lines = list(c("RMSE", round(RMSE[[1]],3),
                 round(RMSE[[2]],3),
                 round(RMSE[[3]],3),
                 round(RMSE[[4]],3),
                 round(RMSE[[5]],3)
                ),
c("F-Statistic", round(fr[[1]],2),
  round(fr[[2]],2), round(fr[[3]],2),
  round(fr[[4]],2), round(fr[[5]],2)),
c("Países", 19, 19, 19, 19, 19),
c("Cluster 1 FE", "Si", "Si", "Si", "Si", "Si")),
omit.stat = c("F"),
notes = "Errores estándares robustos por Driscoll y Kraay entre paréntesis")

#### Modelo con Log ingresos turismo por travel items ####
mm1<-log(gini_disp)~log(N)+ I(log(N)^2) +as.factor(c12)
mm2<-log(gini_disp)~log(N)+ I(log(N)^2) + n_sharedrule+n_selfrule+as.factor(c12)
mm3<-log(gini_disp)~log(N)+ I(log(N)^2)+n_sharedrule+n_selfrule+vdem_polyarchy +
  vdem_pubcorr+vdem_partip + as.factor(c12)
mm4<-log(gini_disp)~log(N)+ I(log(N)^2)+n_sharedrule+n_selfrule+ vdem_polyarchy +
  vdem_pubcorr+ vdem_partip + log(GDPpercapitaPPPcurrentint)+ log(PopulationtotalSPPOPTOTL) +
  Generalgovernmentfinalconsump+as.factor(c12)
mm5<-log(gini_disp)~log(N)+ I(log(N)^2)+n_RAI+ vdem_polyarchy + vdem_pubcorr+
  vdem_partip + log(GDPpercapitaPPPcurrentint)+ log(PopulationtotalSPPOPTOTL) +
  Generalgovernmentfinalconsump+as.factor(c12)

mre2<-list()
vcov2<-list()
se2<-list()
pv2<-list()
RMSE2<-list()
ftest2<-list()
fr2<-list()
MM<-list(mm1, mm2, mm3, mm4,mm5)

for(i in seq_along(MM)){
  formula<-MM[[i]]
  mre2[[i]]<-plm(formula, data=pd, model="random", index=panelid, effect =
    "twoways")
  vcov2[[i]]<-coefest(mre2[[i]], function(x) vcovSCC(x, cluster="time"))
  ftest2[[i]]<-pwaldtest(mre2[[i]], test="F", function(x) vcovSCC(x, cluster="time"))
  fr2[[i]]<-ftest2[[i]]$statistic
  se2[[i]]<-vcov2[[i]][,2]
  pv2[[i]]<-vcov2[[i]][,4]
  RMSE2[[i]]<-crmse(mre2[[i]])
}

stargazer(mre2, type="text",
  se=se2,
  p=pv2,
  star.cutoffs = c(0.1, 0.05, 0.01),
  add.lines = list(c("RMSE", round(RMSE2[[1]],3),
                    round(RMSE2[[2]],3),
                    round(RMSE2[[3]],3),
                    round(RMSE2[[4]],3),
                    round(RMSE2[[5]],3)
                   ),
c("F-Statistic", round(fr2[[1]],2),
  round(fr2[[2]],2), round(fr2[[3]],2),
  round(fr2[[4]],2), round(fr2[[5]],2))),
omit.stat = c("F"))

## Tabla ###

stargazer(mre2, type="html", out = "tablareg2.html",
  se=se2,
  p=pv2,
  star.cutoffs = c(0.1, 0.05, 0.01),
  add.lines = list(c("RMSE", round(RMSE2[[1]],3),
                    round(RMSE2[[2]],3),
                    round(RMSE2[[3]],3),
                    round(RMSE2[[4]],3),
                    round(RMSE2[[5]],3)
                   ),
c("F-Statistic", round(fr2[[1]],2),
  round(fr2[[2]],2), round(fr2[[3]],2),
  round(fr2[[4]],2), round(fr2[[5]],2))),
omit.stat = c("F"))

stargazer(mre2, type="html", out="tablareg2.html",
  se=se2,
  p=pv2,
  star.cutoffs = c(0.1, 0.05, 0.01),
  dep.var.caption = "",
  dep.var.labels = c("Log(Gini Disp)",
                    "Log(Tur(2))",
                    "Log(Tur(2))^2",
                    "Autoridad Comp.",
                    "Autogobierno",
                    "Índice RAI",

```



```

p=pv3,
star.cutoffs = c(0.1, 0.05, 0.01),
dep.var.caption = "",
dep.var.labels = c("Log(Gini Disp)"),
covariate.labels = c("Log(Tur(3))",
                    "Log(Tur(3))^2",
                    "Autoridad Comp.",
                    "Autogobierno",
                    "Índice RAI",
                    "Poliarquía (VDEM)",
                    "Corrupción (VDEM)",
                    "Partip. Pol. (VDEM)",
                    "Log(PIB pc PPP USD)",
                    "Log(POB)",
                    "Gov. Consumo (%)" ),

add.lines = list(c("RMSE", round(RMSE3[[1]],3),
                  round(RMSE3[[2]],3),
                  round(RMSE3[[3]],3),
                  round(RMSE3[[4]],3),
                  round(RMSE3[[5]],3)
                ),
                c("F-Statistic", round(fr3[[1]],2),
                  round(fr3[[2]],2), round(fr3[[3]],2),
                  round(fr3[[4]],2), round(fr3[[5]],2)),
                c("Países", 19, 19, 19, 19, 19),
                c("Cluster 2 FE" "Si", "Si", "Si", "Si", "Si")),
omit.stat = c("F"),
notes = "Errores estándares robustos por Driscoll y Kraay entre paréntesis")

stargazer(mre3, type="html", out="tablareg3.html",
          se=se3,
          p=pv3,
          star.cutoffs = c(0.1, 0.05, 0.01),
          dep.var.caption = "",
          dep.var.labels = c("Log(Gini Disp)"),
          covariate.labels = c("Log(Tur(3))",
                              "Log(Tur(3))^2",
                              "Autoridad Comp.",
                              "Autogobierno",
                              "Índice RAI",
                              "Poliarquía (VDEM)",
                              "Corrupción (VDEM)",
                              "Partip. Pol. (VDEM)",
                              "Log(PIB pc PPP USD)",
                              "Log(POB)",
                              "Gov. Consumo (%)" ),

          add.lines = list(c("RMSE", round(RMSE3[[1]],3),
                            round(RMSE3[[2]],3),
                            round(RMSE3[[3]],3),
                            round(RMSE3[[4]],3),
                            round(RMSE3[[5]],3)
                          ),
                          c("F-Statistic", round(fr3[[1]],2),
                            round(fr3[[2]],2), round(fr3[[3]],2),
                            round(fr3[[4]],2), round(fr3[[5]],2)),
                          c("Países", 19, 19, 19, 19, 19),
                          c("Cluster 2 FE" "Si", "Si", "Si", "Si", "Si")),
                          omit.stat = c("F"),
                          notes = "Errores estándares robustos por Driscoll y Kraay entre paréntesis")

##### Modelo log gasto en turismo internacional (USD)

m1<-log(gini_disp)~log(F)+
I(log(F)^2)+as.factor(c12)
m2<-log(gini_disp)~log(F)+
I(log(F)^2)+n_sharedrule+n_selfrule+as.factor(c12)
m3<-log(gini_disp)~log(F)+
I(log(F)^2)+n_sharedrule+n_selfrule+vdem_polyarchy +
vdem_pubcorr+vdem_partip + as.factor(c12)
m4<-log(gini_disp)~log(F)+
I(log(F)^2)+n_sharedrule+n_selfrule+ vdem_polyarchy +
vdem_pubcorr+ vdem_partip + log(GDPpercapitaPPPcurrentint)+ log(PopulationtotalSPPOTOTL) +
Generalgovernmentfinalconsump+as.factor(c12)
m5<-log(gini_disp)~log(F)+
I(log(F)^2)+n_RAI+ vdem_polyarchy + vdem_pubcorr+
vdem_partip + log(GDPpercapitaPPPcurrentint)+ log(PopulationtotalSPPOTOTL) +
Generalgovernmentfinalconsump+as.factor(c12)

mre4<-list()
vcov4<-list()
se4<-list()
pv4<-list()
RMSE4<-list()
ftest4<-list()
fr4<-list()
MMMM<-list(m1, m2, m3, m4, m5)

for(i in seq_along(MMMM)){
  formula<-MMMM[[i]]
  mre4[[i]]<-plm(formula, data=pd, model="random", index=panelid, effect =
    "twoways")
  vcov4[[i]]<-coefest(mre4[[i]], function(x) vcovSCC(x, cluster="time"))
  ftest4[[i]]<-pwaldtest(mre4[[i]], test="F", function(x) vcovSCC(x, cluster="time"))
  fr4[[i]]<-ftest4[[i]]$statistic
  se4[[i]]<-vcov4[[i]][,2]
  pv4[[i]]<-vcov4[[i]][,4]
}

```

```

RMSE4[[i]]<-crmse(mre4[[i]])
}

stargazer(mre4, type="text",
  se=se4,
  p=pv4,
  star.cutoffs = c(0.1, 0.05, 0.01),
  dep.var.caption = "",
  dep.var.labels = c("Log(Gini Disp)",
  covariate.labels = c("Log(Tur(4))",
    "Log(Tur(4))^2",
    "Autoridad Comp.",
    "Autogobierno",
    "Índice RAI",
    "Poliarquía (VDEM)",
    "Corrupción (VDEM)",
    "Partip. Pol. (VDEM)",
    "Log(PIB pc PPP USD)",
    "Log(POB)",
    "Gob. Consumo (%)"
  ),
  add.lines = list(c("RMSE", round(RMSE4[[1]],3),
    round(RMSE4[[2]],3),
    round(RMSE4[[3]],3),
    round(RMSE4[[4]],3),
    round(RMSE4[[5]],3)
  ),
  c("F-Statistic", round(fr4[[1]],2),
    round(fr4[[2]],2), round(fr4[[3]],2),
    round(fr4[[4]],2), round(fr4[[5]],2)),
  c("Países", 19, 19, 19, 19, 19),
  c("Cluster 2 FE", "Si", "Si", "Si", "Si", "Si")),
  omit.stat = c("F"))

stargazer(mre4, type="html",out="tablareg4.html",
  se=se4,
  p=pv4,
  star.cutoffs = c(0.1, 0.05, 0.01),
  dep.var.caption = "",
  dep.var.labels = c("Log(Gini Disp)",
  covariate.labels = c("Log(Tur(4))",
    "Log(Tur(4))^2",
    "Autoridad Comp.",
    "Autogobierno",
    "Índice RAI",
    "Poliarquía (VDEM)",
    "Corrupción (VDEM)",
    "Partip. Pol. (VDEM)",
    "Log(PIB pc PPP USD)",
    "Log(POB)",
    "Gob. Consumo (%)"
  ),
  add.lines = list(c("RMSE", round(RMSE4[[1]],3),
    round(RMSE4[[2]],3),
    round(RMSE4[[3]],3),
    round(RMSE4[[4]],3),
    round(RMSE4[[5]],3)
  ),
  c("F-Statistic", round(fr4[[1]],2),
    round(fr4[[2]],2), round(fr4[[3]],2),
    round(fr4[[4]],2), round(fr4[[5]],2)),
  c("Países", 19, 19, 19, 19, 19),
  c("Cluster 2 FE", "Si", "Si", "Si", "Si", "Si")),
  omit.stat = c("F"))

#####
## Análisis REp. Dom###
#####

####Paso 1: Cargar base de datos####
panel <- read_dta("G:/Mi unidad/Turismo pobreza paper/basepanelfinal.dta")

#### Paso 2: Estimar modelos####
panel$dummy<-ifelse(panel$año==2020 | panel$año==2022, 1, 0)
panel$Intur<-log(panel$establecimientos)
panel$rieq<-abs(panel$ingred2/panel$egred)

####Paso 3: Hacer pruebas de especificación####
###Palma y turismo
palma1<-lm(palma~log(habitaciones), panel)
pr1<-summary(palma1)$r.squared
summary(palma1)
ggplot(panel[, c("palma", "habitaciones",
  "macroregion", "sismap")],
  aes(log(habitaciones), palma, color=macroregion, size=sismap))+
  geom_point()+theme_apa()+labs(y="índice de Palma",
  x="Log(Núm. de Habitaciones turísticas)",
  color="",
  size="SISMAP")+
  geom_smooth(method="lm", color="black", formula=y~x)+
  theme(legend.position = "bottom", text=element_text(size=10))+
  annotate("text", x=max(log(panel$habitaciones))*0.95,

```

```

        y=max(panel$palma),
        label=sprintf("R²=%.2f", pr1),
        hjust=1, vjust=1,
        size=4, color="black",
        size=8)+
    scale_color_manual(values=c( "#0000FF88",
                                "#00008B88", "#FF000088", "purple"))

ggsave("graficord1.png", dpi=300,
       width=2953, height = 2025, units = "px")

###Palma PIB
palma2<-lm(gini~lnpibpc, panel)
pr2<-summary(palma2)$r.squared
summary(palma2)
ggplot(panel[, c("gini", "lnpibpc",
                "macroregion", "rieg")],
       aes(lnpibpc, gini, color=macroregion, size=rieg))+
  geom_point()+theme_apa()+labs(y="índice de Gini",
                                x="Log(PIB per cápita (2010=100))",
                                color="",
                                size="|Ing/Egresos| MM RD$")+
  geom_smooth(method="lm", color="black", formula=y~x)+
  theme(legend.position = "bottom", text=element_text(size=10))+
  annotate("text", x=max(panel$lnpibpc),
         y=max(panel$gini),
         label=sprintf("R²=%.2f", pr2),
         hjust=1, vjust=1,
         size=4, color="black",
         size=8)+
  scale_color_manual(values=c( "#0000FF88",
                                "#00008B88", "#FF000088", "purple"))

ggsave("graficord2.png", dpi=300,
       width=2953, height = 2025, units = "px")

###

palma3<-lm(palma~(ingred2), panel)
pr3<-summary(palma3)$r.squared
summary(palma3)
ggplot(panel[, c("palma", "ingred2",
                "macroregion", "sismap")],
       aes(ingred2,palma, color=macroregion, size=sismap))+
  geom_point()+theme_apa()+labs(y="índice de Palma",
                                x="Ingresos gob. locales (2010=100) en millones RD$",
                                color="",
                                size="SISMAP")+
  geom_smooth(method="lm", color="black", formula=y~x)+
  theme(legend.position = "bottom", text=element_text(size=10))+
  annotate("text", x=max(panel$ingred2),
         y=max(panel$palma),
         label=sprintf("R²=%.2f", pr3),
         hjust=1, vjust=1,
         size=4, color="black",
         size=8)+
  scale_color_manual(values=c( "#0000FF88",
                                "#00008B88", "#FF000088", "purple"))

ggsave("graficord3.png", dpi=300,
       width=2953, height = 2025, units = "px")

###Q5/Q4

palma4<-lm(q5q4~(egred), panel)
pr4<-summary(palma4)$r.squared
summary(palma4)
ggplot(panel[, c("q5q4", "egred",
                "macroregion", "sismap")],
       aes(egred, q5q4, color=macroregion, size=sismap))+
  geom_point()+theme_apa()+labs(y="Ratio Quintil 5/Quintil 4 de ingresos",
                                x="Egresos gob. locales (2010=100) en millones RD$",
                                color="",
                                size="SISMAP")+
  geom_smooth(method="lm", color="black", formula=y~x)+
  theme(legend.position = "bottom", text=element_text(size=10))+
  annotate("text", x=max(panel$egred),
         y=max(panel$q5q4),
         label=sprintf("R²=%.2f", pr4),
         hjust=1, vjust=1,
         size=4, color="black",
         size=8)+
  scale_color_manual(values=c( "#0000FF88",
                                "#00008B88", "#FF000088", "purple"))

ggsave("graficord4.png", dpi=300,
       width=2953, height = 2025, units = "px")

###Prueba de Hausman robusta: within group means ###

panel<-within(panel,{
  mIntur<-Intur+tapPLY(Intur, idregion, mean)[factor(idregion)]
  mlnpibpc<-lnpibpc+tapPLY(lnpibpc, idregion, mean)[factor(idregion)]
  mdam<-dam+tapPLY(dam, idregion, mean)[factor(idregion)]
  msismap<-sismap+tapPLY(sismap, idregion, mean)[factor(idregion)]
  mrieg<-rieg+tapPLY(rieg, idregion, mean)[factor(idregion)]
  mlnpurbana<-lnpurbana+tapPLY(lnpurbana, idregion, mean)[factor(idregion)]
})
#Lista de coeficientes y resultados:

```

```

mlist<- list()
clist <- list()
model_stats<-list()
hausman_list<-list()
hetero_list<-list()
auto<-list()

# Función auxiliar para calcular el RMSE
calc_rmse <- function(model) {
  sqrt(mean(residuals(model)^2))
}
Y<-c("gini", "palma", "rkuznets", "q5q4", "q4q3", "q3q2", "q2q1" )
index_vars<-c("idregion", "año")

for(i in seq_along(Y)) {
  var <- Y[i]
  formula <- as.formula(paste(var, "~", "lntur+lnpibpc+dam+sismap+rieg+lmpurbana+t5+t7"))
  model <- plm(formula, data = panel, model = "random", index=index_vars)
  het<-bptest(model, studentize = T)
  coef_test <- coef_test(model, vcov = "CR2", cluster = panel$idregion, test = "saddlepoint")
  lmpg<-pbgttest(model, order=2, type="F")

  mlist[[paste0("model", i)]]<- model
  clist[[paste0("coef_test", i)]] <- coef_test
  hetero_list[[paste0("bptest", i)]]<-het
  auto[[paste0("pbgttest", i)]]<-lmpg
  n <- length(resid(model))
  reg<-length(resid(model))/7
  r2 <- summary(model)$r.squared[1]
  adj_r2 <- summary(model)$r.squared[2]
  rmse <- calc_rmse(model)

  # Almacenar estadísticas del modelo en una lista con nombres secuenciales
  model_stats[[paste0("model", i)]] <- list(
    N = n,
    Regiones=reg,
    R2 = r2,
    Adj_R2 = adj_r2,
    RMSE = rmse
  )
}
hausman_list<-list()

for(i in seq_along(Y)) {
  var <- Y[i]
  formula2<-as.formula(paste(var, "~", "0+lntur+lnpibpc+dam+sismap+
  rieg+lmpurbana+mLntur+mLnpibpc+mdam+msismap+mrieg+mLnpurbana+t5+t7"))
  modhaus<-plm(formula2, data=panel, model="random", index = index_vars)
  hausmanr<-Wald_test(modhaus, constraints = constrain_zero(c("mLntur", "mLnpibpc", "mdam", "msismap", "mrieg", "mLnpurbana")), vco
  v="CR2")

  hausman_list[[paste0("Wald_test", i)]]<-hausmanr
}

####Tabla de pruebas de Hausman: ####

tabla_hausmanr<-data.frame(espe=c(Y),
  test=c(hausman_list[[1]]$Fstat,
  hausman_list[[2]]$Fstat,
  hausman_list[[3]]$Fstat,
  hausman_list[[4]]$Fstat,
  hausman_list[[5]]$Fstat,
  hausman_list[[6]]$Fstat,
  hausman_list[[7]]$Fstat),
  corr=c("CR2",
  "CR2",
  "CR2", "CR2", "CR2",
  "CR2", "CR2"),
  g11=c(hausman_list[[1]]$df_num,
  hausman_list[[2]]$df_num,
  hausman_list[[3]]$df_num,
  hausman_list[[4]]$df_num,
  hausman_list[[5]]$df_num,
  hausman_list[[6]]$df_num,
  hausman_list[[7]]$df_num),
  g12=c(hausman_list[[1]]$df_denom,
  hausman_list[[2]]$df_denom,
  hausman_list[[3]]$df_denom,
  hausman_list[[4]]$df_denom,
  hausman_list[[5]]$df_denom,
  hausman_list[[6]]$df_denom,
  hausman_list[[7]]$df_denom),
  pv=c(hausman_list[[1]]$p_val,
  hausman_list[[2]]$p_val,
  hausman_list[[3]]$p_val,
  hausman_list[[4]]$p_val,
  hausman_list[[5]]$p_val,
  hausman_list[[6]]$p_val,
  hausman_list[[7]]$p_val)
)

```

```

colnames(tabla_hausmanr)<-c("Var. Dep.", "Test de Hausman",
                          "Corrección GL", "GL1", "GL2", "P-valor")

tabla_hausmanr1<-xtable::xtable(tabla_hausmanr)

print(tabla_hausmanr1, type="html", file="hausmanRD.html")

####Pruebas de heterocedasticidad####
tabla_homocedasticidad<-data.frame(espe=c(Y),
                                   test=c(hetero_list[[1]]$statistic,
                                           hetero_list[[2]]$statistic,
                                           hetero_list[[3]]$statistic,
                                           hetero_list[[4]]$statistic,
                                           hetero_list[[5]]$statistic,
                                           hetero_list[[6]]$statistic,
                                           hetero_list[[7]]$statistic),
                                   gl=c(hetero_list[[1]]$parameter,
                                         hetero_list[[2]]$parameter,
                                         hetero_list[[3]]$parameter,
                                         hetero_list[[4]]$parameter,
                                         hetero_list[[5]]$parameter,
                                         hetero_list[[6]]$parameter,
                                         hetero_list[[7]]$parameter),
                                   pv=c(hetero_list[[1]]$p.value,
                                         hetero_list[[2]]$p.value,
                                         hetero_list[[3]]$p.value,
                                         hetero_list[[4]]$p.value,
                                         hetero_list[[5]]$p.value,
                                         hetero_list[[6]]$p.value,
                                         hetero_list[[7]]$p.value)
)

colnames(tabla_homocedasticidad)<-c("Var. Dep.", "Test de Breuch-Pagan", "GL", "P-valor")

print(xtable::xtable(tabla_homocedasticidad), type="html", file = "homocedas.html")
####Tabla autocorrelación####
tabla_auto<-data.frame(espe=c(Y),
                      test=c(auto[[1]]$statistic,
                              auto[[2]]$statistic,
                              auto[[3]]$statistic,
                              auto[[4]]$statistic,
                              auto[[5]]$statistic,
                              auto[[6]]$statistic,
                              auto[[7]]$statistic),
                      gl=c(auto[[1]]$parameter[1],
                            auto[[2]]$parameter[1],
                            auto[[3]]$parameter[1],
                            auto[[4]]$parameter[1],
                            auto[[5]]$parameter[1],
                            auto[[6]]$parameter[1],
                            auto[[7]]$parameter[1]),
                      gl2=c(auto[[1]]$parameter[2],
                              auto[[2]]$parameter[2],
                              auto[[3]]$parameter[2],
                              auto[[4]]$parameter[2],
                              auto[[5]]$parameter[2],
                              auto[[6]]$parameter[2],
                              auto[[7]]$parameter[2]),
                      pv=c(auto[[1]]$p.value,
                            auto[[2]]$p.value,
                            auto[[3]]$p.value,
                            auto[[4]]$p.value,
                            auto[[5]]$p.value,
                            auto[[6]]$p.value,
                            auto[[7]]$p.value)
)

colnames(tabla_auto)<-c("Var. Dep.", "Test de Breuch-Godfrey", "GL1", "GL2", "P-valor")

tabla_auto

print(xtable::xtable(tabla_auto), type="html", file = "autocorr.html")
#### Tabla de regresiones final####

stats_df<-do.call(rbind, model_stats)
stats_df<-data.frame(model=names(model_stats), stats_df, stringsAsFactors = F)

se_list<-lapply(clist, function(df) df[, "SE"])
p_list<-lapply(clist, function(df) df[, "p_saddle"])
r<-lapply(stats_df$RMSE,function(df) df[])

stargazer(mlist, type="text",
          se= se_list,

```

```

p= p_list,
star.cutoffs = c(0.1, 0.05, 0.01),
covariate.labels = c("Log(Tur)", "Ln(PIB)",
                    "DA", "SMM",
                    "|Ing/Eg| Gob. Local ",
                    "Ln(PU)", "t(2020)",
                    "t(2022)"),
dep.var.caption = "",
dep.var.labels = c("Gini", "Palma", "R.Kuznets",
                  "Q5/Q4", "Q4/Q3", "Q3/Q2",
                  "Q2/Q1"),
add.lines= list(c("COV(B)", "CR2",
                 "CR2", "CR2", "CR2", "CR2", "CR2"),
               c("Regiones", reg, reg, reg, reg, reg, reg,
                 reg, reg),
               c("RMSE", round(r$model1, 3),
                 round(r$model2, 3), round(r$model3, 3),
                 round(r$model4, 3), round(r$model5, 3),
                 round(r$model6, 3), round(r$model7, 3)
                 )),
omit.stat = c("F")
)

```

```

stargazer(mlist, type="html",out="tablaregresionreptom.html",
se= se_list,
p= p_list,
star.cutoffs = c(0.1, 0.05, 0.01),
covariate.labels = c("Log(Tur)", "Ln(PIB)",
                    "DA", "SMM",
                    "|Ing/Eg| Gob. Local ",
                    "Ln(PU)", "t(2020)",
                    "t(2022)"),
dep.var.caption = "",
dep.var.labels = c("Gini", "Palma", "R.Kuznets",
                  "Q5/Q4", "Q4/Q3", "Q3/Q2",
                  "Q2/Q1"),
add.lines= list(c("COV(B)", "CR2",
                 "CR2", "CR2", "CR2", "CR2", "CR2"),
               c("Regiones", reg, reg, reg, reg, reg, reg,
                 reg, reg),
               c("RMSE", round(r$model1, 3),
                 round(r$model2, 3), round(r$model3, 3),
                 round(r$model4, 3), round(r$model5, 3),
                 round(r$model6, 3), round(r$model7, 3)
                 )),
omit.stat = c("F")
)

```

Un gráfico para evidenciar La relación de Kuznets invertida

```

pd$pred_log_gini_disp<-predict(mre2[5], newdata=pd, type="response")
base_final_tesis$pred_gini_disp<-exp(pd$pred_log_gini_disp)

ggplot(base_final_tesis[,c("N", "pred_gini_disp",
                          "c12", "n_RAI")],
        aes(log(N),pred_gini_disp, color=c12, size=n_RAI))+
  theme_apa()+geom_point()+labs(y="Índice de Gini (predicción)",
                                x= "Log(Ingresos del turismo por artículos de viaje USD)",
                                color="Conglomerado",
                                size= "Autoridad Regional (RAI)"
  )+
  geom_smooth(method = "lm", color="black", formula=y~poly(x,2))+
  theme(legend.position = "bottom",
        text=element_text(
          size=12))+
  scale_color_manual(values=c( "#0000FF88",
                               "#00008B88", "#FF000088"))

ggsave('graficofinal.png', dpi=300,
        width=2953, height = 2025, units = "px")

```