



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

---

---

FACULTAD DE CIENCIAS

CORRELACIONES ESTADÍSTICAS ENTRE GASTO  
MEDIO EN RUBROS COMPARADOS CON EL PODER  
ADQUISITIVO EN VARIOS PAÍSES

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

MATEMÁTICA

P R E S E N T A :

NORMA RAQUEL GUZMAN MENDIETA

TUTOR

DR. MARCELO DEL CASTILLO MUSSOT



CIUDAD UNIVERSITARIA, CDMX, 2024



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



*«I am enough of an artist to draw freely upon my imagination. Imagination is more important than knowledge. Knowledge is limited. Imagination encircles the world»*

Albert Einstein



# Agradecimientos

Me gustaría expresar mi más sincero agradecimiento a mis padres por su invaluable apoyo financiero y por guiarme en mi camino hacia el éxito. Su amor y dedicación han sido fundamentales para mi desarrollo personal y profesional. También quisiera reconocer a Edgar, cuyo paradero es desconocido, por creer en mi capacidad para superar los desafíos de mis primeros semestres en la universidad. Su apoyo fue crucial para ayudarme a enfrentar la depresión que experimenté durante ese tiempo. Finalmente, quisiera agradecer al asesor Dr. Marcelo del Castillo por su paciencia y comprensión durante mis episodios de ansiedad. Su guía y sabiduría fueron fundamentales para ayudarme a superar esos momentos difíciles.

Quiero agradecer a Gaby, que pese a lo lejos que nos encontramos sigue siendo una de las personas que mas me apoya.

También agradezco a todos los sinodales su apoyo. En particular, las sugerencias y correcciones de Alejandro Radillo han sido muy importantes para que esta tesis sea más clara y esté libre de algunos posibles errores.



# Índice general

Agradecimientos	v
1. Introducción	1
2. Conceptos	3
3. Metodología	7
4. Gastos en ingresos en 2018	9
5. Análisis de correlación	17
6. Análisis de regresión	31
7. Conclusiones	55
Apéndice A. Mínimos cuadrados	59
Bibliografía	66



# 1 Introducción

El estudio del consumo ha ocupado un lugar prominente en la investigación macroeconómica, manteniendo su relevancia a pesar de las variaciones teóricas. Se entiende por consumo doméstico el valor agregado de los bienes y servicios obtenidos por un conjunto de individuos que cohabitan. Este concepto engloba los costos de mercado de todas las adquisiciones realizadas para cubrir las necesidades y aspiraciones de dichos individuos, abarcando tanto bienes duraderos como no duraderos, tales como vehículos, electrodomésticos y utensilios de cocina. [22]

El desembolso realizado por las familias tiene una correlación directa con la calidad de vida, el progreso humano y el bienestar. La magnitud del gasto familiar es un indicador del grado de avance del sistema económico global. Existen múltiples elementos que influyen en la envergadura y la composición del gasto familiar, así como en su grado de satisfacción, abarcando desde las necesidades más elementales hasta las más secundarias.

Es crucial reconocer que el progreso económico no garantiza una mejora uniforme en la calidad de vida y el bienestar de los individuos. De ahí que se originó la noción de desarrollo humano enfocado en el ser humano. Este tipo de desarrollo se orienta a otorgar libertad y oportunidades propicias para que los individuos puedan llevar la vida que elijan, con dignidad y con la garantía de que sus necesidades fundamentales sean atendidas.

La atención persistente y relevante en examinar las discrepancias entre naciones es esencial para producir saberes que contribuyan a entender el estándar de vida de las colectividades. La exploración de la conducta de los sujetos en relación con las tendencias de consumo en distintos países facilita la emergencia de análisis novedosos. [22] Nos encontramos en un mundo donde la tecnología progresa a un ritmo acelerado y la brecha entre los estándares y costos de vida continúa ampliándose. Un elemento que contribuye a este escenario es el crecimiento económico global desequilibrado.

El ámbito laboral se encuentra segmentado y puede ser categorizado para diferenciar entre empleo formal e informal en diversos países emergentes. Los ingresos tienden a ser notablemente inferiores y las jornadas laborales más extensas en el sector informal. La elección de un trabajador de integrarse en uno de estos sectores se fundamenta en un análisis lógico de los costos y beneficios en comparación con el empleo formal. [4] Numerosos elementos pueden incidir en la desigualdad entre los estándares y costos de vida. Entre éstos se encuentran la brecha de clases, el índice de pobreza, la cali-

dad y accesibilidad de la vivienda; las horas laborales requeridas para satisfacer las necesidades básicas, el producto interno bruto (PIB), el acceso asequible a servicios médicos de alta calidad y la calidad y disponibilidad de la educación. Asimismo, la segmentación del mercado laboral y la distinción entre empleo formal e informal también pueden aportar a esta desigualdad.

Existen patrones de desembolso que fluctúan en función de las características sociales, psicológicas, el nivel de ingresos y el país donde se efectúa el gasto. Por ejemplo, las tendencias de consumo en una nación con una religión predominante y regida por una ley religiosa pueden diferir de las de un país con pluralidad religiosa.

Investigaciones recientes han revelado hallazgos fascinantes sobre las tendencias de consumo. Por ejemplo, se ha descubierto que los jóvenes sin ingresos formales están dispuestos a sacrificar su nutrición para economizar y costear la tarifa de un número telefónico, dado que la comunicación es de gran importancia para ellos.

Del mismo modo, los investigadores que han estudiado el comportamiento de las personas con ingresos bajos cuando estos mejoran han indicado que uno de los cambios principales es que cubren su necesidad primordial de alimentación. No obstante, si disponen de un poco más de dinero adicional, también se consideran otras prioridades como la comunicación y la infraestructura. Es decir, ya no sólo se enfocan en satisfacer completamente su necesidad primordial, sino que también procuran cubrir otras necesidades.

Un monto específico de capital puede ser adecuado para cubrir una necesidad en una nación, mientras que en otra esa misma suma podría permitir la adquisición de un mayor volumen de productos para satisfacer la misma necesidad.

Por tal motivo, es esencial disponer de los datos del gasto de consumo doméstico en una única divisa, dado que la conversión a una moneda local puede resultar en cifras extremas. Para este propósito, es preferible emplear la paridad del poder adquisitivo (*PPP*, por sus siglas en inglés), que nos permite establecer una divisa base. La *PPP* representa la cantidad de moneda local requerida para obtener un bien en nuestra divisa base.

Aplicaremos el factor de cambio de la *PPP* a nuestros datos, ya que nos facilita una mejor visualización de los mismos y permite realizar comparaciones sobre el gasto en bienes y servicios domésticos en distintas naciones.

## 2 Conceptos

### **Producto Interno Bruto (PIB)**

El producto interno bruto (PIB) es el valor de mercado de todos los bienes y servicios finales producidos dentro de las fronteras de un país en un período específico de tiempo. El PIB se calcula generalmente en base anual, aunque también puede calcularse en base trimestral. En el caso de los Estados Unidos, el gobierno publica una estimación anual del PIB para cada trimestre fiscal y para cada año. Los conjuntos de datos individuales incluyen el informe en términos reales, ajustado por cambios en los precios y la inflación neta.

### **Paridad del Poder Adquisitivo (PPP)**

La paridad del poder adquisitivo (PPP) es uno de los métodos más populares de análisis macroeconómico para comparar la productividad económica y los estándares de vida entre países. La PPP es una teoría económica que compara diferentes monedas a través de una “canasta básica” aproximada. Según este concepto, dos monedas estarán en equilibrio cuando la “canasta básica” tenga el mismo precio en ambos países, teniendo en cuenta el cambio de divisas. [26]

### **Factor de Conversión PPP**

La paridad del poder adquisitivo (PPP) convierte diferentes monedas a una moneda común y, en el proceso de conversión, ajusta el poder de compra teniendo en cuenta la diferencia en los niveles de precios entre países. Por lo general, los países con mayores ingresos tienen precios más altos, mientras que los países con menores ingresos tienen precios más bajos (efecto Balassa-Samuelson).

En términos formales, el efecto Penn es un descubrimiento económico que indica que los ingresos reales entre países de ingresos altos y bajos son sistemáticamente sobreestimados al convertir el Producto Interno Bruto (PIB) a las tasas de cambio de mercado. El modelo Balassa-Samuelson es citado como la causa principal del efecto Penn en la economía neoclásica.

Por ejemplo, la mayoría de los productos suelen ser más económicos en países de bajos ingresos que en los países más acaudalados. Un individuo proveniente de un país desarrollado que se encuentre de vacaciones en un país en desarrollo normalmente

descubrirá que su dinero tiene un mayor valor en el extranjero que en su lugar de origen. A modo de ilustración, una Big Mac tenía un costo de \$7.84 en Noruega y \$2.39 en Egipto en enero de 2013, mientras que para el año 2023, el precio de una Big Mac en Noruega es de \$6.26 y en Egipto es \$2.43, a pesar de que el producto es esencialmente el mismo. [20]

## Correlación

Una correlación es una herramienta estadística para medir la relación entre dos o más variables, es decir, el grado en el cual las variables están asociadas entre sí; de tal manera que un cambio en una está acompañado por un cambio en la otra variable.

## Coefficiente de correlación de Spearman

El coeficiente de correlación de Spearman, también conocido como  $\rho$  de Spearman, es una medida no paramétrica que evalúa la relación entre dos variables ordenadas o clasificadas. Este coeficiente mide la dependencia estadística entre los rankings de dos variables y varía entre  $-1$  y  $+1$ . Un valor de  $0$  indica que no hay correlación entre las variables, mientras que un valor de  $+1$  indica una relación monótona perfecta y positiva y un valor de  $-1$  indica una relación monótona perfecta y negativa.

## Correlación no es lo mismo que causalidad

Efectivamente, la correlación y la causalidad son conceptos distintos. La correlación alude a una relación en la que dos variables fluctúan de manera conjunta, mientras que la causalidad sugiere que una modificación en una variable provoca un cambio en otra. No obstante, es crucial recordar que la correlación no implica necesariamente causalidad. A pesar de que dos variables puedan estar correlacionadas, ello no implica que una sea la causa de la otra. Podrían existir otras variables o factores que incidan en ambas y expliquen su correlación. Para establecer una relación causal entre dos variables, se requiere la realización de estudios experimentales o de intervención que permitan controlar y manipular las variables para determinar si una modificación en una provoca un cambio en la otra.

## Regresión lineal

En estadística, la regresión lineal es un enfoque lineal para modelar la relación entre una respuesta escalar y una o más variables dependientes e independientes. En la regresión lineal, las relaciones se modelan utilizando funciones predictoras lineales cuyos parámetros desconocidos del modelo se estiman a partir de los datos. Estos modelos se denominan modelos lineales. Dado un conjunto de datos  $\{y_i, x_{i1}, \dots, x_{ip}\}$  donde  $i = 1, \dots, n$  de  $n$  unidades estadísticas, un modelo de regresión lineal asume que

la relación entre las variables dependientes y el vector de predictores  $x$  es lineal. Esta relación está modelada a través de una variable de error  $\epsilon$ , una variable no observada que agrega ‘ruido’ a la relación lineal entre la variable dependiente y los predictores. Así el modelo toma la forma siguiente:

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \epsilon_i = x_i^T \beta + \epsilon_i \quad i = 1, \dots, n$$

Donde  $T$  denota la transpuesta, de modo que  $x_i^T \beta$  es el producto interno entre los vectores  $x_i$  y  $\beta$ .

### Regresión lineal simple

La relación entre una variable de respuesta  $Y$  y una variable predictora  $X$  está postulada como un modelo lineal de la forma:

$$Y = \beta_0 + \beta_1 X + \epsilon \quad (\text{I})$$

Donde  $\beta_0$  y  $\beta_1$  son constantes llamadas coeficientes del modelo de regresión o parámetros de regresión, así mismo  $\epsilon$  es una perturbación o error aleatorio. Asumimos que, en el rango de observaciones estudiadas, la ecuación lineal (I) nos denota la variable aleatoria. En particular,  $\epsilon$  es información no sistemática para determinar  $Y$  que no está todavía capturada en  $X$ . El coeficiente  $\beta_1$ , llamado pendiente, se puede interpretar como el cambio en  $Y$  por unidad de cambio en  $X$ . El coeficiente  $\beta_0$ , también llamado constante o coeficiente de intersección, es el valor predictivo de  $Y$  cuando  $X = 0$ . De acuerdo con la ecuación (I), también podemos reescribirla de la siguiente manera:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \quad i = 1, 2, \dots, n \quad (\text{II})$$

Donde esta ecuación nos denota las observaciones.



## 3 Metodología

Para llevar a cabo el análisis de correlaciones y regresiones, así como para generar gráficos de dispersión y tablas relacionadas con estos análisis, emplearemos el software R, ampliamente reconocido entre economistas e investigadores aplicados. [15] Inicialmente, nos embarcaremos en un proceso de recolección y depuración de los conjuntos de datos, los cuales se obtienen tanto del *World Data Bank* como de los sitios web de las instituciones nacionales de estadística de cada país.

*General Authority for Statistics (Arabia Saudita) [3], Federal Statistical Office of Germany (Alemania) [10], Statistični urad RS (Eslovenia) [25], Spanish Statistical Office (España) [23], Hellenic Statistical Authority (Grecia) [11], Italian National Institute of Statistics (Italia) [12], Central Bureau of Statistics (Israel) [6], National Bureau of Statistics of the Republic of Moldova (Moldavia) [17], Office for National Statistics (Reino Unido) [19], Statistical Office of the Republic of Serbia (Serbia) [24], Federal Statistical Office (Suiza) [9]*

Para nuestro estudio, recolectamos datos correspondientes al año 2018, dado que la información de los años subsiguientes aún no estaba disponible. Elegimos aquellos países cuyos datos eran accesibles y, tras obtenerlos, realizamos una conversión de la moneda de los datos al índice de paridad de poder adquisitivo del año 2018, lo que nos permitió tener los datos en dólares y simplificar su análisis. Es importante señalar que excluimos a Estados Unidos de nuestra selección, ya que el formato de presentación de sus datos en el sitio web oficial de estadísticas del gobierno no era propicio para nuestro análisis.

En una etapa posterior, recurrimos a los paquetes estadísticos de R para extraer los resultados del análisis de correlación. Inicialmente, sometimos los datos a la prueba de Shapiro-Wilk para determinar si cumplían con los criterios necesarios para ser analizados mediante el método de Spearman, o si era preciso recurrir a otro método. Tras confirmar que el conjunto de datos no exhibía un comportamiento normal, procedimos a elaborar la matriz de correlación y los gráficos de dispersión para examinar el comportamiento de los pares de variables. En el contexto de nuestra investigación, pusimos especial énfasis en aquellos pares que presentaban un coeficiente de correlación

elevado, específicamente en los intervalos  $\rho \in (0.8, 1)$  y  $\rho \in (-1, -0.8)$ .

Posteriormente, procedimos a analizar los datos a través de un análisis de regresión lineal para identificar las relaciones que resultaban de nuestro interés, es decir, aquellas donde  $0.8 \leq \rho \leq 1$  y  $-1 \leq \rho \leq -0.8$ . A partir de este punto, derivamos la ecuación correspondiente a la línea de regresión, calculamos los residuos, el residuo estándar, las F-estadísticas y los grados de libertad asociados a nuestra regresión. Este proceso nos permitió obtener una comprensión más profunda de las relaciones entre las variables en estudio.

## 4 Gastos en ingresos en 2018

### GASTOS

Según la organización para la cooperación y el desarrollo económico (OCDE), los gastos del hogar se refieren al monto que los hogares residentes gastan en satisfacer sus necesidades diarias, como alimentos, ropa, vivienda, energía, transporte, bienes duraderos, costos de salud, ocio y servicios varios [18]. Esta variable representa alrededor del 60 % del producto interno bruto (PIB) y es esencial para el análisis económico de la demanda. Se recolectaron datos de las instituciones estadísticas de Arabia Saudita, Alemania, España, Eslovenia, Grecia, Italia, Israel, Suiza, Serbia, Moldavia y Reino Unido para analizarlos debido a la importancia de sus características económicas y sociales.

Debemos aclarar que no tomamos en cuenta a Estados Unidos, ya que los datos ofrecidos por la base de datos estadounidense no nos ofrecían los datos en el formato en que los países seleccionados lo hacían.

Se recolectaron datos referentes a los gastos del hogar en diversas categorías de consumo, que incluyen comida y bebidas no alcohólicas, bebidas alcohólicas, tabaco y narcóticos, ropa y calzado, combustibles y energía eléctrica, servicios y bienes para el hogar, salud, transporte, comunicación, recreación y cultura, educación, restaurantes y hoteles, así como una miscelánea de bienes y servicios. Para mantener la consistencia de los datos en dólares, se aplicó el factor de conversión del poder adquisitivo del *World Data Bank* correspondiente al año 2018. Los datos ya han sido ajustados por este factor y se muestran en una tabla en dólares (\$).

Insumos	Alemania	España	Grecia	Italia	Reino Unido	Suiza	Moldavia	Serbia	Eslovenia	Arabia Saudita	Israel
Comidas y bebidas no alcohólicas	3336	4223	3996	5540	3697	7004	620	2258	2817	655	711
Bebidas alcohólicas, tabaco y narcóticos	516	519	639	559	762	1100	22	319	351	13	51
Vestimenta y calzado	1464	1463	999	1427	1428	2105	151	349	1315	150	138
Combustibles, energía eléctrica	10896	9181	2438	10833	4643	16079	258	1099	2736	904	589
Servicios y bienes del hogar	1572	1372	753	1295	2483	2417	54	278	1123	262	1145
Salud	1380	1010	1296	1449	421	2668	72	293	565	56	276
Transporte	4548	3790	2248	3509	4929	8316	57	612	4184	498	847
Comunicación	852	964	712	745	1092	2063	65	344	1016	207	92
Recreación y cultura	3648	1643	813	1533	4551	6062	21	333	1776	115	213
Educación	336	461	556	192	531	691	7	99	188	90	254
Restaurantes y hoteles	2016	2948	1876	1564	3026	6504	27	205	1297	182	119
Miscelánea de bienes y servicios	1332	2298	1527	2210	2654	3464	61	389	2191	901	192
<b>Total</b>	<b>31896</b>	<b>29872</b>	<b>17352</b>	<b>30855</b>	<b>30270</b>	<b>58472</b>	<b>1416</b>	<b>6577</b>	<b>19560</b>	<b>4032</b>	<b>4627</b>

Tabla 4.1: Gastos sobre la canasta básica convertidos bajo PPP

Según los datos de la tabla anterior, en 2018, Suiza tuvo el mayor gasto del hogar, con un total de 58472 dólares, mientras que Moldavia tuvo el menor gasto del hogar, con solo 1416 dólares, lo que representaría un 2.4% del gasto de los suizos, Arabia Saudita e Israel también tuvieron gastos del hogar bajos, con 4032 y 4627 dólares respectivamente.

En lo que respecta al gasto en combustibles y energía eléctrica, Suiza, Italia y Alemania lideran con el mayor gasto anual, superando los 10,000 dólares por año. En contraparte, Arabia Saudita registró el menor gasto en esta categoría, lo cual es previsible dada su condición de uno de los principales productores de petróleo, junto con Estados Unidos. En relación a los sectores de salud y educación, se percibe una tendencia general de disminución en el gasto destinado a estos dos rubros en comparación con el gasto en restaurantes y hoteles.

En términos generales, el gasto en servicios de salud no excede los 2700 dólares al año. Suiza encabeza la lista de países con el mayor gasto en este sector, seguido por Grecia, Italia, Alemania y España. En contraste, Arabia Saudita tiene un gasto promedio de sólo 56 dólares en este sector.

En relación al sector de la educación, Arabia Saudita es la nación que asigna el mayor porcentaje de su PIB a este sector, con un 7%. No obstante, los hogares saudíes invierten una cantidad relativamente baja en educación. Por otro lado, Grecia es el país que asigna el menor porcentaje de su PIB a la educación, con solo un 4%, pero los hogares griegos invierten 556 dólares en este sector. Suiza es la nación donde los hogares invierten la mayor cantidad de dinero en educación, con un promedio de 691 dólares.

A continuación, se presenta una tabla que muestra los porcentajes del PIB destinados a la educación por cada país, el gasto de los hogares en euros y los resultados de la prueba PISA para el año 2018. [1]

Países	% PIB	Cantidad destinada (\$)	Resultados PISA
Alemania	5	336	500
Arabia Saudita	7	90	386
Eslovenia	5	188	504
España	4	461	483
Grecia	4	556	453
Italia	4	192	477
Israel	6	254	465
Moldavia	5	7	424
Reino Unido	5	531	504
Serbia	4	99	442
Suiza	5	691	498

Tabla 4.2: Resultados en la prueba PISA para el año 2018, así como el PIB destinado por cada país.

A pesar de que Arabia Saudita es la nación que destina el mayor porcentaje de su PIB a la educación, sus resultados en la prueba PISA fueron inferiores, con menos de 400 puntos. Esto podría atribuirse a la influencia de diversos factores sociales, económicos y psicológicos. En contraste, Eslovenia y el Reino Unido lograron resultados superiores a los 500 puntos en la prueba PISA, seguidos de Alemania, que apenas superó el límite de los 500 puntos. Estos países destinan el 5%, 5% y 5% de su PIB a la educación, respectivamente. Grecia es la nación que destina el menor porcentaje de su PIB a la educación, con solo un 4%, pero los hogares griegos invierten más de 500 euros en este sector. Sus resultados en la prueba PISA son ligeramente superiores a los de países como Arabia Saudita, Serbia y Moldavia.

En países donde la religión predominante no es el cristianismo, como Arabia Saudita e Israel, se pueden notar diferencias en las categorías de gasto de los hogares. En Arabia Saudita, donde el Islam es la fe mayoritaria, se asigna una gran parte del presupuesto a diversos servicios. Estos incluyen actividades como visitas a spas, tratamientos faciales, maquillaje, peluquería y terapia. Por otro lado, en Israel, independientemente de que la mayoría de la población practique el judaísmo, el transporte representa la mayor parte del gasto de los hogares. Es crucial considerar que la religión puede tener un impacto en los gastos del hogar de varias formas. Por ejemplo, puede moldear las decisiones de consumo y las prioridades de gasto de los hogares, basándose en sus creencias y prácticas religiosas. Asimismo, puede tener un papel en cómo los hogares distribuyen su presupuesto entre diferentes categorías de gasto.

No obstante, es esencial recordar que la religión es sólo uno de los múltiples factores que pueden afectar los gastos del hogar y que su efecto puede variar dependiendo del contexto y las circunstancias personales. De acuerdo con [5], uno de los elementos que contribuye al tráfico y a los costos de transporte en Israel es la infraestructura limitada de transporte ferroviario, que en 2018 sólo representaba el 2% del total. Adicionalmente, el servicio de transporte público no está disponible en días festivos religiosos, lo que incrementa la utilización de medios de transporte privados. Los gastos del hogar fluctúan en función de las condiciones socioeconómicas de los países estudiados. A continuación, se muestra el promedio de ingresos correspondiente al año 2018.

## INGRESOS

El ingreso familiar se define como la suma neta de los ingresos de todos los miembros de un hogar que superan un límite de edad determinado. [21] No es necesario que los individuos tengan algún parentesco para ser considerados miembros de la misma unidad familiar. Una estadística económica frecuentemente reportada es el ingreso familiar promedio. No obstante, dado que muchos hogares pueden estar compuestos por una sola persona, el ingreso familiar promedio puede ser inferior al ingreso promedio por individuo. Existen diversos factores que influyen en esto, entre ellos el tamaño de la familia. En general, las familias más grandes tienden a tener un ingreso total más alto, pero cuando se divide entre el número de miembros (para obtener el ingreso promedio por persona) este suele ser menor. Otros factores que influyen son la edad del jefe de familia, el número de miembros que trabajan y el ciclo de vida de la familia. Al realizar comparaciones de riqueza y calidad de vida entre distintas ciudades, estados o países, resulta útil considerar las estadísticas de ingresos familiares. La suma total de los ingresos de cada integrante de un hogar se denomina ingresos del hogar. Estos ingresos pueden derivarse de diversas fuentes, como salarios, inversiones y beneficios sociales. Para calcular los ingresos del hogar, se consideran todas las personas mayores de 18 años que residen en el mismo domicilio, sin importar si tienen algún parentesco entre sí.

El ingreso bruto se refiere a la suma total de dinero que se obtiene antes de realizar deducciones por concepto de impuestos y otros gastos. En contraste, el ingreso neto es la cantidad de dinero que se conserva después de realizar dichas deducciones del ingreso bruto. Se recolectaron datos referentes a los ingresos del año 2018 a partir de diversas bases de datos. No obstante, se excluyeron países como Israel y Moldavia, ya que sus bases de datos no contienen información categorizada de esta forma. Los datos de ingresos provienen de una base de datos que indica el riesgo de pobreza y la desigualdad de ingresos en los países miembros de la Unión Europea. Los ingresos se corresponden con el nivel educativo del individuo y el ingreso del hogar es la suma de los ingresos de los miembros del hogar que tienen 18 años o más.

## INGRESOS DE UNA PERSONA POR ESCOLARIDAD

Países	Primaria y al menos secundaria	Bachillerato y técnicos	Universidad
Serbia	1941	2950	4502
Arabia Saudita	2600	3563	5715
Grecia	6896	8524	12320
Eslovenia	11455	13455	17911
España	12833	16233	22886
Italia	15184	20040	27117
Alemania	19276	23869	32485
Reino Unido	20007	23837	32406
Suiza	38207	47931	61333

Tabla 4.3: Ingreso de una persona por escolaridad en distintos países para el año 2018.

Es relevante destacar que en Serbia, el ingreso promedio de un individuo, sin importar su nivel educativo, no sería suficiente para cubrir todos los gastos del hogar. El costo promedio de la canasta básica es de \$6677, mientras que el ingreso promedio para una persona con educación primaria es de \$1941, con bachillerato es de \$2950 y con educación universitaria es de \$4502. Por lo tanto, sería inviable para una persona con educación primaria o secundaria cubrir el gasto promedio en alimentos de un hogar con su ingreso promedio.

En lo que respecta a los ingresos de una sola persona, los datos indican que en 2018, Suiza era el país con los ingresos anuales más altos en función de la escolaridad del trabajador. En particular, el ingreso promedio de un trabajador universitario en Suiza ascendía a \$61333 al año.

La siguiente tabla contiene datos de supuestos en los que se considera que el ingreso de al menos dos personas por nivel de escolaridad debería ser el doble del monto obtenido por una sola persona. Esta suposición se hizo debido a la falta de datos precisos sobre el ingreso de dos personas en las bases de datos asociadas a cada uno de los países estudiados.

Países	Primaria y al menos secundaria	Bachillerato y tecnicos	Universidad
Serbia	3882	5900	9004
Arabia Saudita	5200	7126	11430
Grecia	13792	17048	24640
Eslovenia	22910	26910	35822
España	25666	32466	45772
Italia	30368	40080	54234
Alemania	38552	47738	64970
Reino Unido	40014	47674	64812
Suiza	76414	95862	122666

Tabla 4.4: Suposición de ingresos de al menos dos personas por nivel de escolaridad para el año 2018.

Después de observar estos datos, surge la pregunta natural: ¿Por qué ocurre esto en Serbia? Para responder a esta pregunta, es necesario considerar la historia del país. Según [13], la desigualdad en los ingresos y la dificultad para cubrir el promedio anual de gastos puede deberse a la transición posterior a los eventos que separaron a Yugoslavia. Serbia fue el último país en transitar a una economía de mercado debido al aislamiento internacional y la desestabilización política y económica que ocurrió en los años noventa.

La economía de Serbia aún sufre las consecuencias de los problemas relacionados con la destrucción de la economía en los años noventa, seguida de un empobrecimiento significativo y una erosión de la cohesión social. [14] Una política social activa comenzó después de la encuesta sobre los estándares de vida realizada por el Banco Mundial en 2002, seguida de una estrategia de reducción de la pobreza adoptada en 2003. Esta estrategia tenía como objetivo lograr tres metas: autosuficiencia, equidad y cohesión social. Los programas incluían políticas para mejorar las oportunidades laborales con una protección social eficiente, salud y educación, así como reducir la pobreza [14].

Según OCDE [18], el aumento en la proporción de personas que trabajan a tiempo parcial con contratos temporales o que trabajan por cuenta propia es uno de los principales impulsores de la desigualdad en los ingresos. Aunque el tema de la desigualdad en los países de Europa del Este podría analizarse más a fondo, para nuestros propósitos, dejaremos este tema como un punto de partida para futuros trabajos. En países como Arabia Saudita, Grecia, Eslovenia, España e Italia, se observa una

tendencia en la que los trabajadores con menor nivel de escolaridad no logran obtener ingresos suficientes para cubrir el promedio anual de gastos del hogar. Esto es preocupante y puede afectar no solo la condición física de las personas, sino también su salud mental.

# 5 Análisis de correlación

## CORRELACIÓN DE SPEARMAN

La correlación de Spearman, a veces referida como la correlación de rangos de Spearman, es una métrica no paramétrica que evalúa la correlación de rango, es decir, la dependencia estadística entre los rankings de dos variables.

Supongamos que tenemos una muestra de tamaño  $n$ , con  $n$  datos duros representados por  $X_i$  y  $Y_i$ . Estos datos serán transformados en rangos y representados por  $R(X_i)$ ,  $R(Y_i)$  y se calculará el coeficiente de correlación por  $\rho_s$ . [8] Cuya fórmula esta dada de la siguiente manera:

$$\rho_s = 1 - \frac{6 \sum d_i^2}{n(n^2-1)}$$

Una correlación indica el grado y la dirección en que dos variables están relacionadas. No necesariamente se ajusta a una línea a través de los puntos de datos; sino que simplemente calcula el coeficiente de correlación para determinar cómo cambia una variable en relación con la otra.

Cuando el valor de correlación  $\rho$  es igual a cero, significa que no hay relación entre las variables. Si  $\rho > 0$ , entonces la correlación es positiva y hay una tendencia de que una variable aumente a medida que la otra también lo hace. Por otro lado, si se presenta el caso contrario entonces la correlación es negativa ya que mientras una variable aumenta, la otra tiende a disminuir.

Al abordar el tema de las correlaciones, es crucial no interpretarlas en términos de causa y efecto. Independientemente de cómo se etiqueten las dos variables, ya sea como “dependiente” o “independiente”, si ambas cambian, el valor del coeficiente de correlación permanecerá inalterado.

A continuación, se muestran gráficos de dispersión para situaciones con coeficientes de correlación tanto altos y positivos como negativos, representando una correlación inversa. Expondremos la matriz de correlación para dos casos: el primero (caso I) relacionado con indicadores socioeconómicos, y el segundo (caso II) vinculado a los gastos de diversas categorías para distintos países. Destacaremos las correlaciones que, siendo negativas o positivas, han captado nuestra atención. Para este propósito, emplearemos los resultados derivados del análisis de correlación realizado en el software R, donde se obtuvieron matrices de correlaciones con el valor de “ $\rho$ ”. Estos datos se han ajustado utilizando el factor “*PPP*” para facilitar la interpretación de la

información presentada.

## CASO I. CORRELACIONES DE INDICADORES

En este caso recolectamos datos acerca de algunos indicadores económicos, de los diez países a estudiar, tomando en cuenta que las bases de datos de los distintos países presentan los datos que se necesitan. Para ayudar en el proceso de graficación y análisis en el software de R, optamos por indicar con letras los nombres de los indicadores. A continuación, presentamos la lista de indicadores a analizar: consumo de energías renovables ( $C_{er}$ ); crecimiento poblacional ( $C_p$ ); desempleo ( $D$ ); emisiones de  $CO_2$  ( $ECO_2$ ); empleados en agricultura ( $E_a$ ); esperanza de vida ( $E_v$ ); exportaciones de bienes y servicios ( $E_{xby}$ ); exportaciones de combustible ( $E_x$ ); gasto militar ( $G_m$ ); gasto público ( $G_p$ ); gasto en educación ( $G_e$ ); homicidios por cada 100,000 habitantes ( $H$ ); importaciones de bienes y servicios ( $I_{bys}$ ); inflación ( $I_n$ ); ingreso bruto ( $I_b$ ); ingresos fiscales ( $I_f$ ); trabajadores asalariados ( $T_a$ ); tratados comerciales ( $T_c$ ).

Los ingresos brutos se refieren a la suma total de ingresos obtenidos en la nómina antes de aplicar impuestos y deducciones. En contraste, los ingresos fiscales son aquellos fondos recaudados por el Estado para financiar actividades del sector público.

A continuación, mostraremos la tabla 5.1 con los valores de  $\rho$ . Se puede apreciar una línea de valores donde los pares exactos son iguales a uno, formando una diagonal que divide nuestra tabla en dos partes. Es importante destacar que la matriz es simétrica, con la diagonal compuesta por unidades.

Indicador	Tc	Exbys	Ea	Cer	Ibys	Ev	D	Gp	If	Gm	ECO2	Exc	H	In	Ge	Ib	Ta	Cp
Tc	1	0.73	0.56	0.69	0.71	0.06	0.03	-0.28	-0.47	-0.58	-0.58	-0.24	-0.23	-0.15	-0.19	0.37	-0.13	-0.35
Exbys	0.73	1	0.29	0.39	0.54	0.29	0.16	0.1	-0.18	-0.21	-0.25	0.2	-0.62	0.04	-0.47	0.32	-0.05	-0.08
Ea	0.56	0.29	1	0.81	0.65	-0.12	0.01	0.13	0.11	-0.46	-0.56	-0.2	0.24	-0.62	-0.51	-0.39	-0.84	-0.9
Cer	0.69	0.39	0.81	1	0.9	0.12	-0.14	-0.13	0	-0.77	-0.84	-0.55	0.2	-0.67	-0.46	-0.28	-0.66	-0.74
Ibys	0.71	0.54	0.65	0.9	1	0.1	-0.12	-0.02	0.03	-0.61	-0.84	-0.38	-0.16	-0.51	-0.42	-0.18	-0.48	-0.52
Ev	0.06	0.29	-0.12	0.12	0.1	1	0.18	0.01	0.09	-0.06	-0.08	-0.3	-0.2	-0.35	-0.56	0.12	0.01	0.07
D	0.03	0.16	0.01	-0.14	-0.12	0.18	1	0.27	0.16	0.15	0.56	0.54	-0.2	0.36	-0.48	-0.2	-0.13	-0.2
Gp	-0.28	0.1	0.13	-0.13	-0.02	0.01	0.27	1	0.84	0.26	0.25	0.44	-0.47	-0.01	-0.5	-0.5	-0.42	-0.27
If	-0.47	-0.18	0.11	0	0.03	0.09	0.16	0.84	1	0.2	0.14	0.17	-0.08	-0.3	-0.47	-0.81	-0.53	-0.36
Gm	-0.58	-0.21	-0.46	-0.77	-0.61	-0.06	0.15	0.26	0.2	1	0.7	0.67	-0.16	0.27	0.18	0.02	0.37	0.51
ECO2	-0.58	-0.25	-0.56	-0.84	-0.84	-0.08	0.56	0.25	0.14	0.7	1	0.74	-0.08	0.64	0.11	0.01	0.35	0.39
Exc	-0.24	0.2	-0.2	-0.55	-0.38	-0.3	0.54	0.44	0.17	0.67	0.74	1	-0.39	0.6	-0.1	-0.06	0.12	0.22
H	-0.23	-0.62	0.24	0.2	-0.16	-0.2	-0.2	-0.47	-0.08	-0.16	-0.08	-0.39	1	-0.43	0.2	-0.39	-0.31	-0.35
In	-0.15	0.04	-0.62	-0.67	-0.51	-0.35	0.36	-0.01	-0.3	0.27	0.64	0.6	-0.43	1	0.41	0.44	0.64	0.59
Ge	-0.19	-0.47	-0.51	-0.46	-0.42	-0.56	-0.48	-0.5	-0.47	0.18	0.11	-0.1	0.2	0.41	1	0.46	0.7	0.58
Ib	0.37	0.32	-0.39	-0.28	-0.18	0.12	-0.2	-0.5	-0.81	0.02	0.01	-0.06	-0.39	0.44	0.46	1	0.77	0.64
Ta	-0.13	-0.05	-0.84	-0.66	-0.48	0.01	-0.13	-0.42	-0.53	0.37	0.35	0.12	-0.31	0.64	0.7	0.77	1	0.93
Cp	-0.35	-0.08	-0.9	-0.74	-0.52	0.07	-0.2	-0.27	-0.36	0.51	0.39	0.22	-0.35	0.59	0.58	0.64	0.93	1

Tabla 5.1: Valores de correlación  $\rho$  para diferentes pares de variables.

En la tabla anterior, se pueden observar los coeficientes de Spearman para diferentes parejas de variables. Como se mencionó anteriormente, la diagonal en la tabla divide los valores de  $\rho$ . También podemos notar que hay pocos casos en los que los valores de correlación son de interés, ya sea porque son cercanos a uno o porque son cercanos a menos uno.

En los casos siguientes, se presentan las parejas de variables que tienen una correlación positiva. Esto significa que el coeficiente de correlación se encuentra en el intervalo  $0.8 \leq \rho \leq 1$ .

Cuando dos variables se mueven en la misma dirección simultáneamente, se dice que tienen una relación positiva. Consideremos la teoría económica como ejemplo, donde la oferta aumenta en conjunto con el precio. Este fenómeno se explica en gran medida por el hecho de que cuando el precio de un producto es alto, los vendedores encuentran beneficioso aumentar sus ventas.

## CORRELACIONES POSITIVAS

En el análisis inicial, se evaluó la relación entre el consumo de energías renovables y las importaciones de bienes y servicios. Los resultados revelaron un coeficiente de correlación ( $\rho$ ) de 0.9, indicando una asociación positiva entre estas variables. Sin embargo, es importante destacar que esta relación podría no ser perfecta, ya que podría haber variabilidad en los datos.

También podemos notar la presencia de *outliers*, que son aquellas observaciones anormales y extremas en una muestra estadística, las cuales pueden afectar potencialmente la estimación de parámetros. De la gráfica 5.1 podemos notar la presencia de éstos, siendo los datos correspondientes a Eslovenia e Italia.

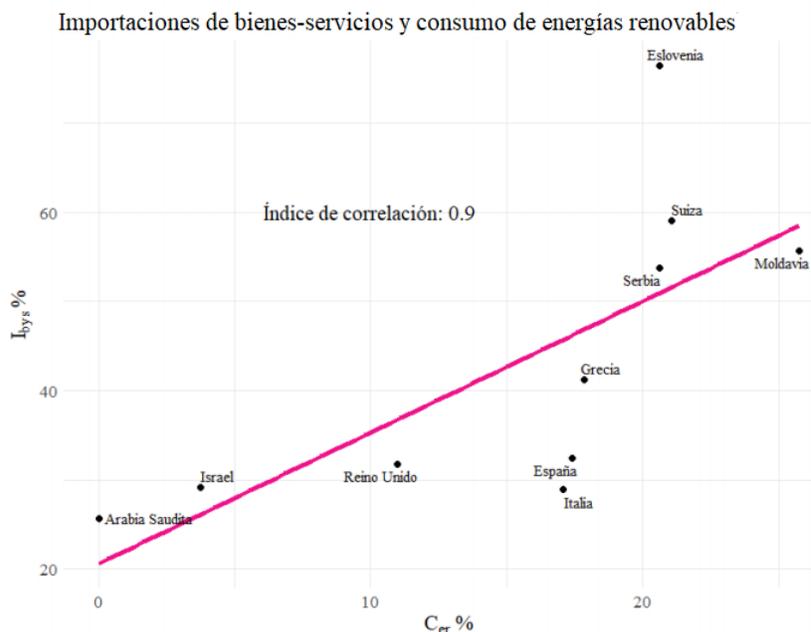


Figura 5.1: Gráfico de correlación de Spearman que muestra la relación entre consumo de energías renovables e importaciones de bienes y servicios durante el año 2018.

Para la pareja de variables conformada por empleados en agricultura y consumo de energías renovables, el análisis y los resultados obtenidos por el software R indican que el coeficiente de correlación  $\rho$  es de  $\rho = 0.81$ . La figura 5.2 muestra el gráfico de dispersión asociado a la correlación de estas dos variables.

A partir de la gráfica siguiente, podemos observar la presencia de Eslovenia y Arabia Saudita como *outliers*.

La pareja de variables formada por gasto poblacional e ingresos fiscales tiene un gráfico de dispersión que se muestra en la figura 5.3. El coeficiente de correlación

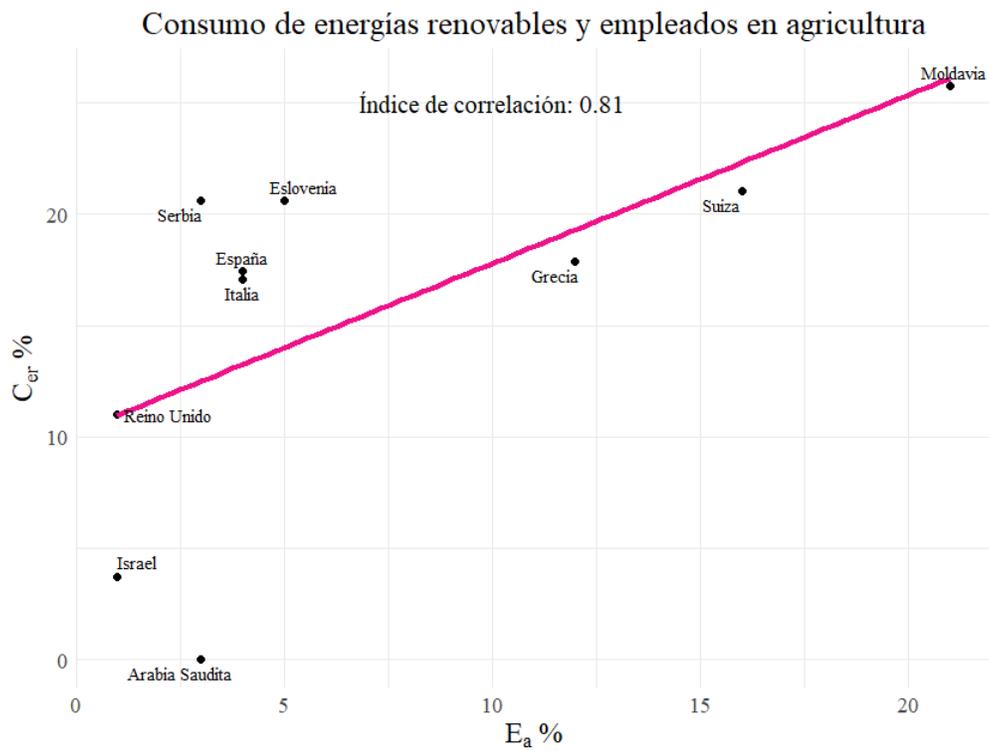


Figura 5.2: Gráfico de correlación de Spearman que muestra la relación entre empleados en agricultura y consumo de energías renovables durante el año 2018.

entre estas dos variables es  $\rho = 0.84$ , lo que indica que existe una correlación positiva entre ellas.

Observamos también la presencia de valores extremos, también conocidos como *outliers* con valores asociados a Reino Unido y Arabia Saudita.

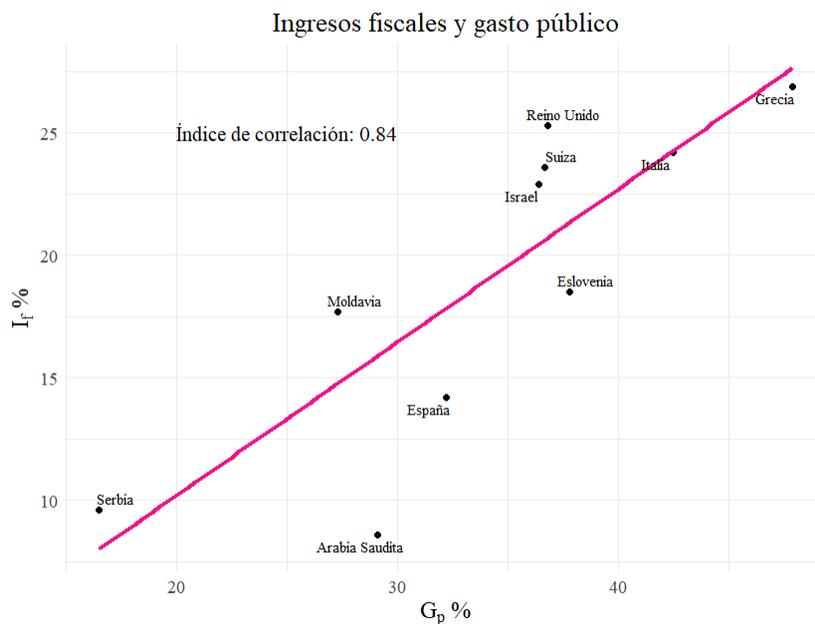


Figura 5.3: Gráfico de correlación de Spearman que muestra la relación entre gasto público e ingresos fiscales durante el año 2018.

La pareja de variables conformada por trabajadores asalariados y el crecimiento poblacional tiene un gráfico de dispersión que se muestra en la figura 5.4. El coeficiente de correlación entre estas dos variables es  $\rho = 0.93$ .

Podemos notar la presencia de *outliers* como lo son valores de Moldavia e Israel.

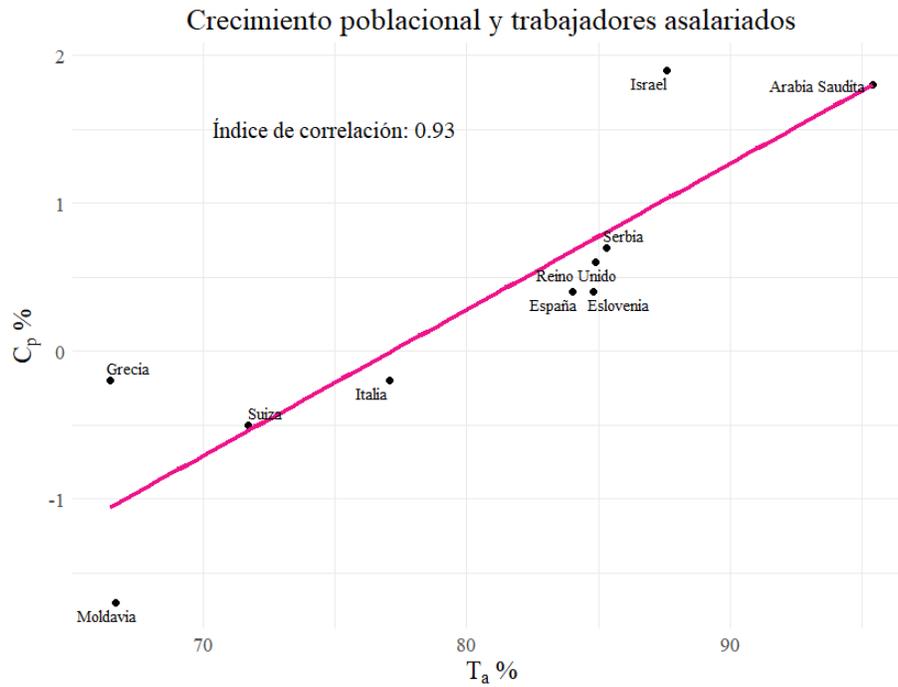


Figura 5.4: Gráfico de correlación de Spearman que muestra la relación entre trabajadores asalariados y crecimiento poblacional durante el año 2018.

### CORRELACIONES NEGATIVAS

En las siguientes páginas se presentarán casos interesantes con correlaciones que representan el comportamiento inverso, es decir que se consideran correlaciones negativas con  $\rho \in [-1, -0.8]$ .

En este caso en la figura 5.5, tenemos a la pareja de variables conformada por el consumo de energías renovables y las emisiones de  $CO_2$ , donde como se esperaba el coeficiente de correlación es de  $\rho = -0.84$ .

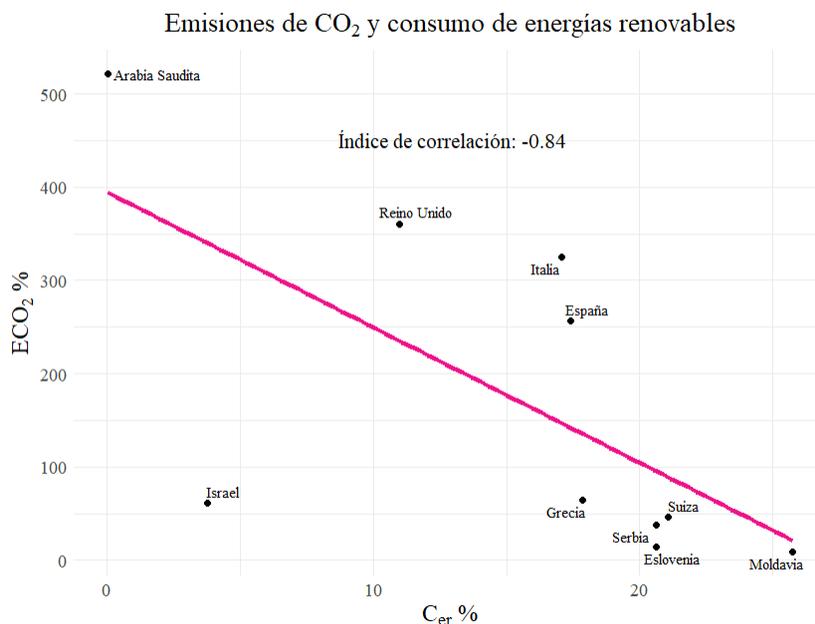


Figura 5.5: Gráfico de correlación negativa de Spearman que muestra la relación entre consumo de energías renovables y emisiones de CO<sub>2</sub> durante el año 2018.

Es decir, el comportamiento que se observa entre estas dos variables es inverso, pues si la variable consumo de energías renovables crece observaremos que el comportamiento de las emisiones de CO<sub>2</sub> decrece.

También observamos la presencia de *outliers* como Arabia Saudita, Italia e Israel.

La figura 5.6 muestra el gráfico de dispersión para las variables de empleados en agricultura y crecimiento poblacional. El coeficiente de correlación es de  $\rho = -0.9$ . También observamos la presencia de *outliers* como lo son los valores de Moldavia e Italia.

Para el gráfico de dispersión número 5.7, tenemos que la pareja de variables está conformada por los empleados en agricultura y los trabajadores asalariados, de donde el coeficiente de correlación obtenido es de  $\rho = -0.84$ .

Observamos la presencia de *outliers* como Arabia Saudita y Grecia.

La figura 5.8 muestra el gráfico de dispersión para las variables de importaciones de bienes-servicios y emisiones de CO<sub>2</sub>. El coeficiente de correlación es de  $\rho = -0.84$ , lo que indica una correlación negativa. La presencia de *outliers* se da en Arabia Saudita, Reino Unido e Israel.

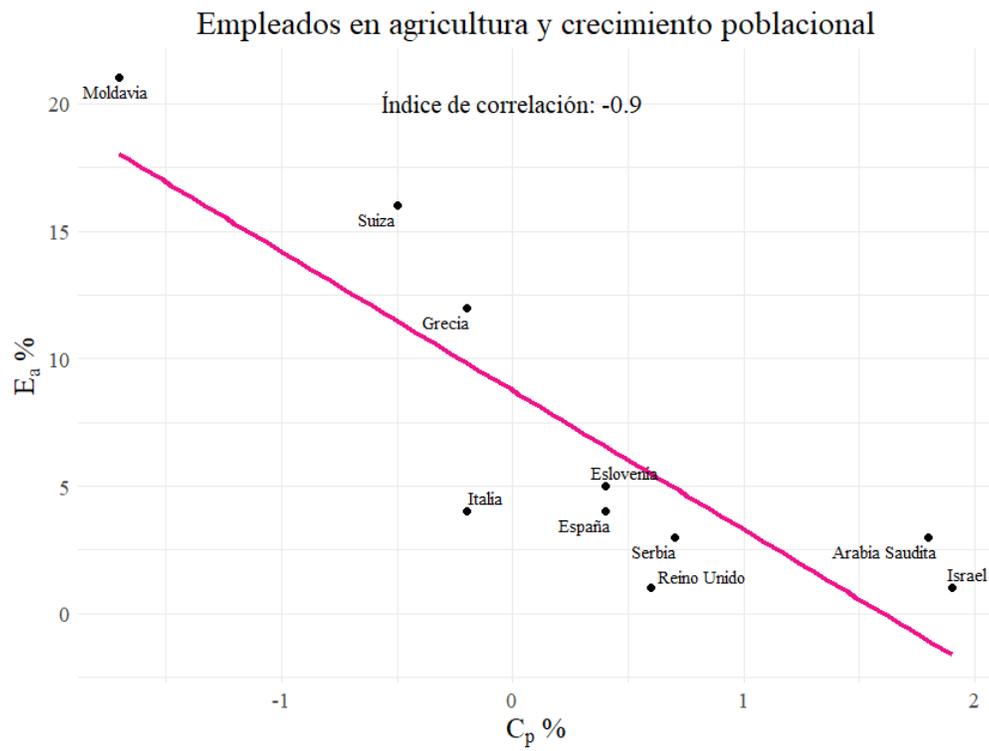


Figura 5.6: Gráfico de correlación negativa de Spearman que muestra la relación entre empleados en agricultura y crecimiento poblacional durante el año 2018.

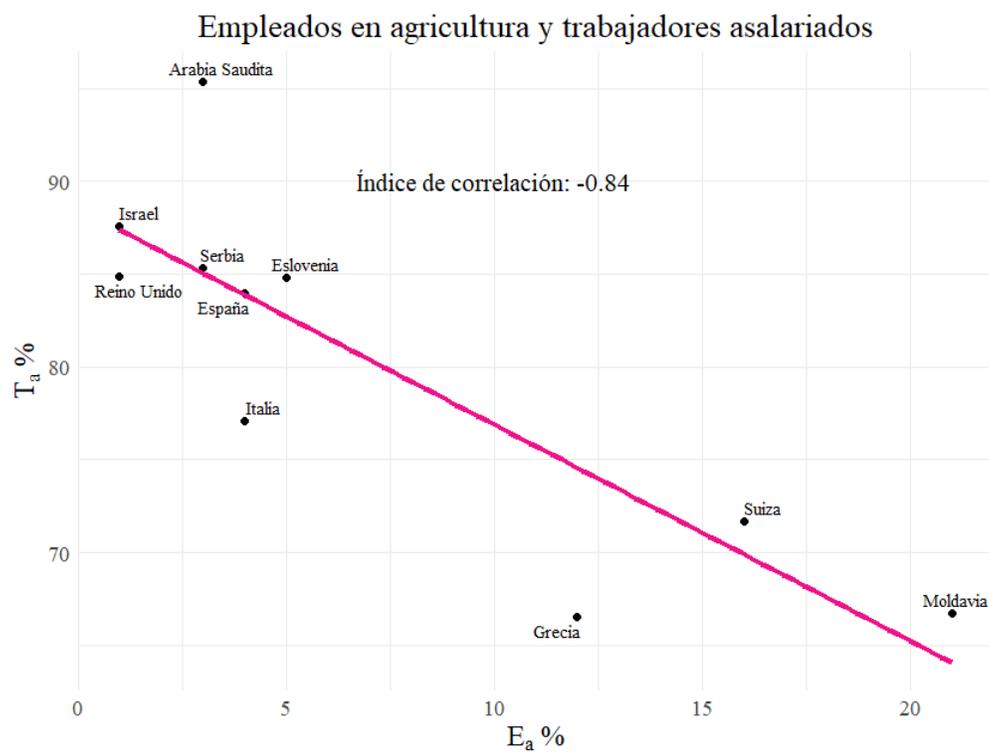


Figura 5.7: Gráfico de correlación negativa de Spearman que muestra la relación entre empleados en agricultura y trabajadores asalariados durante el año 2018.

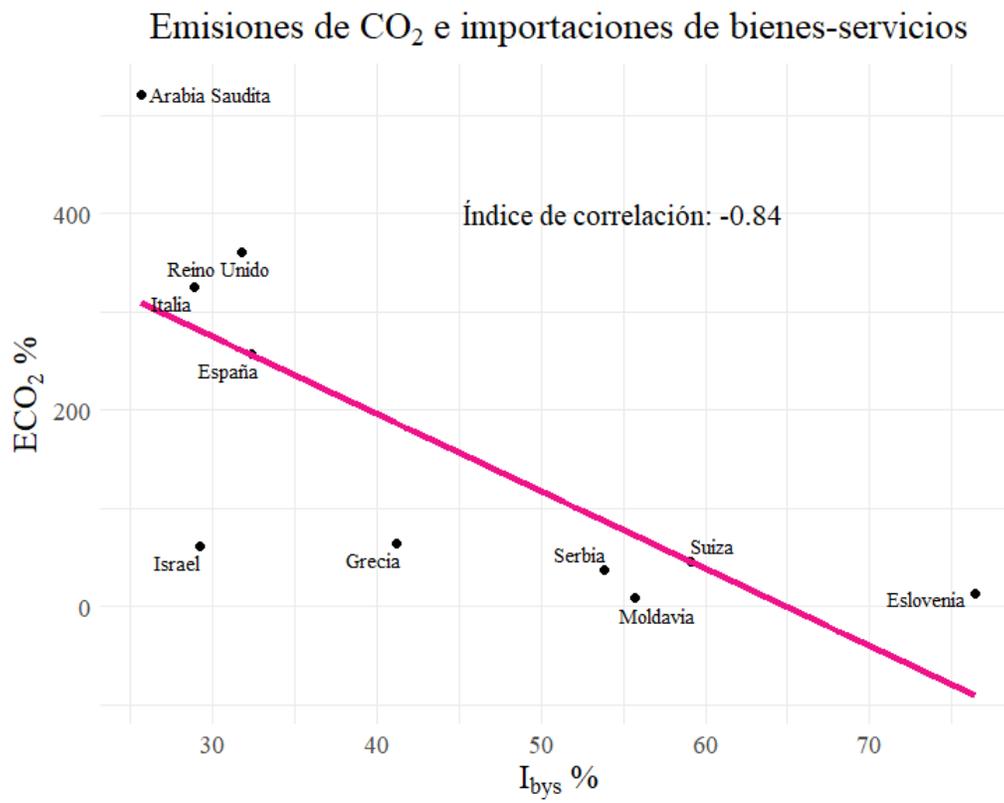


Figura 5.8: Gráfico de correlación negativa de Spearman que muestra la relación entre importaciones de bienes-servicios y emisiones de CO<sub>2</sub> durante el año 2018.

## CASO II. ANÁLISIS DE CORRELACIONES DE GASTOS DE CONSUMO DEL HOGAR

El gasto de consumo final nacional de los hogares es el gasto realizado por los hogares en bienes y servicios que se utilizan para satisfacer directamente las necesidades o carencias de los miembros del hogar.

Por lo que tenemos por objeto de análisis en este caso doce categorías que conforman el consumo del hogar en diferentes países que son: comida y bebidas no alcohólicas ( $C_b$ ); bebidas alcohólicas y tabaco ( $B_T$ ); vestido y calzado ( $V_E$ ); combustibles y energía eléctrica ( $C_E$ ); mantenimiento del hogar ( $M$ ); salud ( $S$ ); transporte ( $T$ ); comunicación ( $C$ ); recreación y cultura ( $R_C$ ); educación ( $E$ ); restaurantes y hoteles ( $R_H$ ) y la miscelánea de bienes y servicios ( $M2$ ). Los datos encontrados en las bases de datos de países como Arabia Saudita, Alemania, España, Eslovenia, Grecia, Italia, Israel, Suiza, Serbia, Moldavia y Reino Unido llamaron nuestra atención. Esto se debe a la diversidad religiosa y a la pertenencia a la Unión Europea de algunos de estos países, lo que hace que su comportamiento y datos sean de interés.

En la siguiente tabla tenemos todos los valores de  $\rho$  para cada pareja de variables que corresponde a los gastos del hogar.

Valores de R												
Categorías	M	E	T	R·c	R·h	C	M2	B·t	V·c	C·e	C·b	S
M	<b>1</b>	0.74	0.19	0.32	0.14	-0.19	0.35	-0.28	-0.45	-0.45	-0.52	-0.35
E	0.74	<b>1</b>	0.02	0.11	0.3	-0.18	0.21	-0.13	-0.5	-0.45	-0.29	0.05
T	0.19	0.02	<b>1</b>	0.37	0.42	-0.7	-0.22	-0.04	-0.42	0.04	-0.49	0.29
R·c	0.32	0.11	0.37	<b>1</b>	0.59	-0.27	0.02	0.41	-0.23	-0.12	-0.86	-0.48
R·h	0.14	0.3	0.42	0.59	<b>1</b>	-0.19	0.38	0.42	-0.15	0.13	-0.66	-0.15
C	-0.19	-0.18	-0.7	-0.27	-0.19	<b>1</b>	0.56	0.18	0.52	-0.16	0.27	-0.42
M2	0.35	0.21	-0.22	0.02	0.38	0.56	<b>1</b>	0.16	0.23	-0.02	-0.18	-0.56
B·t	-0.28	-0.13	-0.04	0.41	0.42	0.18	0.16	<b>1</b>	0.43	-0.22	-0.01	-0.03
V·c	-0.45	-0.5	-0.42	-0.23	-0.15	0.52	0.23	0.43	<b>1</b>	-0.19	0.46	0.02
C·e	-0.45	-0.45	0.04	-0.12	0.13	-0.16	-0.02	-0.22	-0.19	<b>1</b>	0.02	-0.15
C·b	-0.52	-0.29	-0.49	-0.86	-0.66	0.27	-0.18	-0.01	0.46	0.02	<b>1</b>	0.5
S	-0.35	0.05	0.29	-0.48	-0.15	-0.42	-0.56	-0.03	0.02	-0.15	0.5	<b>1</b>

Tabla 5.2: Valores de correlación  $\rho$  para diferentes pares de variables.

## CORRELACIÓN NEGATIVA

En el caso de las variables presentadas que concentran los gastos del hogar, tenemos un caso donde las correlaciones son negativas; a continuación tenemos este caso que deben ser resaltado.

Por un lado, tenemos la pareja conformada por comidas y bebidas no alcohólicas y recreación-cultura, cuyo gráfico de dispersión asociado está en la figura 5.9, y la presentaremos a continuación.

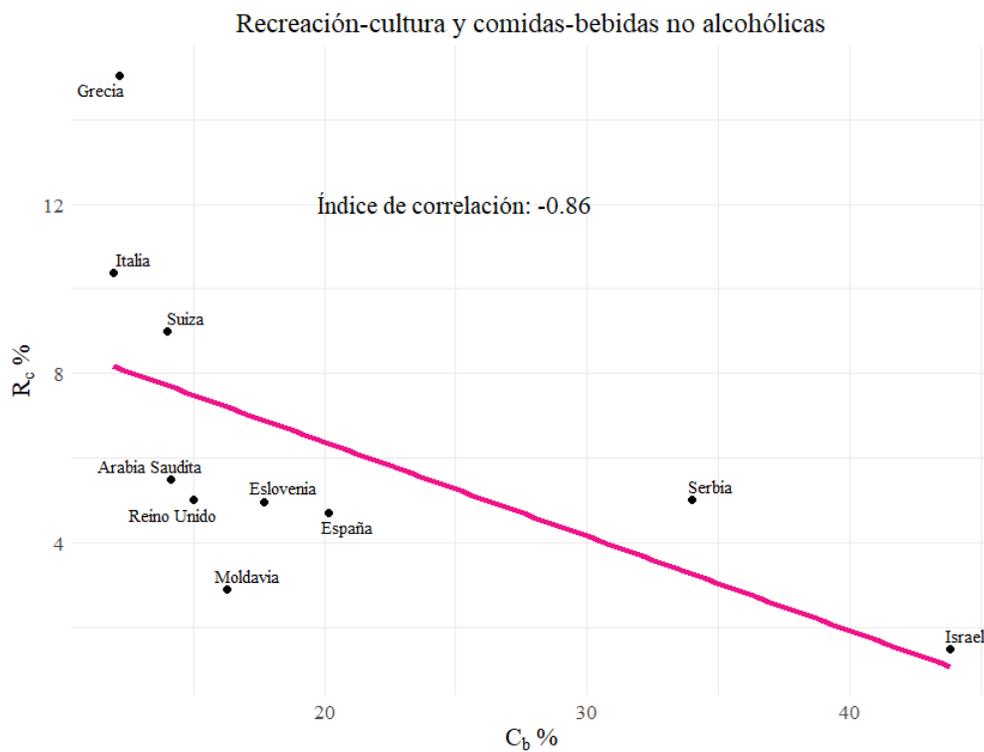


Figura 5.9: Gráfico de correlación negativa de Spearman que muestra la relación entre comidas y bebidas no alcohólicas y recreación-cultura durante el año 2018.

Notamos que tenemos que el coeficiente de correlación de estas variables es de  $\rho = -0.86$  siendo este un valor negativo.

Observamos la presencia de *outliers* como Grecia y Moldavia.

## 6 Análisis de regresión

El modelo de regresión lineal examina la relación entre una variable predictora y una variable de respuesta conocida. Este tipo de modelo se conoce como regresión lineal simple o *SLR* (por sus siglas en inglés).

El análisis de regresión es uno de los procedimientos estadísticos más empleados, que ofrece maneras sencillas de establecer una relación funcional entre diversas variables. Para examinar las relaciones entre las variables que consideramos de mayor interés, recurriremos al método de mínimos cuadrados mediante el software R, que simplifica la tarea.

En nuestro caso particular, hemos recolectado datos sobre indicadores económicos y gastos domésticos en distintas naciones. Estos datos han sido procesados para obtener una base de datos enfocada en los gastos del hogar en términos de paridad de poder adquisitivo.

MODELOS DE REGRESIÓN LINEAL APLICADO PARA INDICADORES  
ECONÓMICOS Y GASTOS DEL HOGAR

Entre este grupo de parejas de variables encontramos las siguientes: Consumo de energías renovables ( $C_{er}$ ) e importación de bienes y servicios ( $I_{bys}$ ); consumo de energías renovables ( $C_{er}$ ) y empleados en agricultura ( $E_a$ ); empleados en agricultura ( $E_a$ ) y miscelánea de bienes y servicios ( $M2$ ); gasto público ( $G_p$ ) e ingresos fiscales ( $I_f$ ); gasto público ( $G_p$ ) y transporte ( $T$ ); finalmente trabajadores asalariados ( $T_a$ ) y crecimiento poblacional ( $C_p$ ); emisiones de  $CO_2$  ( $ECO_2$ ) y consumo de energías renovables ( $C_{er}$ ); empleados en agricultura ( $E_a$ ) y crecimiento poblacional ( $C_p$ ); trabajadores asalariados ( $T_a$ ) y empleados en agricultura ( $E_a$ ); emisiones de  $CO_2$  ( $ECO_2$ ) e importaciones de bienes y servicios ( $I_{bys}$ ); recreación-cultura ( $R_c$ ) y comidas-bebidas no alcohólicas ( $C_b$ ).

Para ver la teoría que se implementará puede consultarse en el apéndice A.

### Consumo de energías renovables y empleados en agricultura

En este primer caso, examinaremos la relación entre el consumo de energías renovables y el número de empleados en agricultura. Según los datos que hemos recopilado, tenemos los siguientes cálculos que nos ayudarán a calcular los mínimos cuadrados más adelante:

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n} = 15.514 \quad \bar{x} = \frac{\sum_{i=1}^n x_i}{n} = 7$$

Por otra parte, también estimaremos los parámetros denotados por  $\hat{\beta}_0$  y  $\hat{\beta}_1$ , por lo que tenemos:

$$\hat{\beta}_1 = \frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2} = 0.756$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 10.222$$

Podemos decir entonces que la ecuación de regresión de mínimos cuadrados está dada de la siguiente manera:

$$C_{er} = 10.222 + 0.756E_a$$

### Consumo de energías renovables y empleados en agricultura

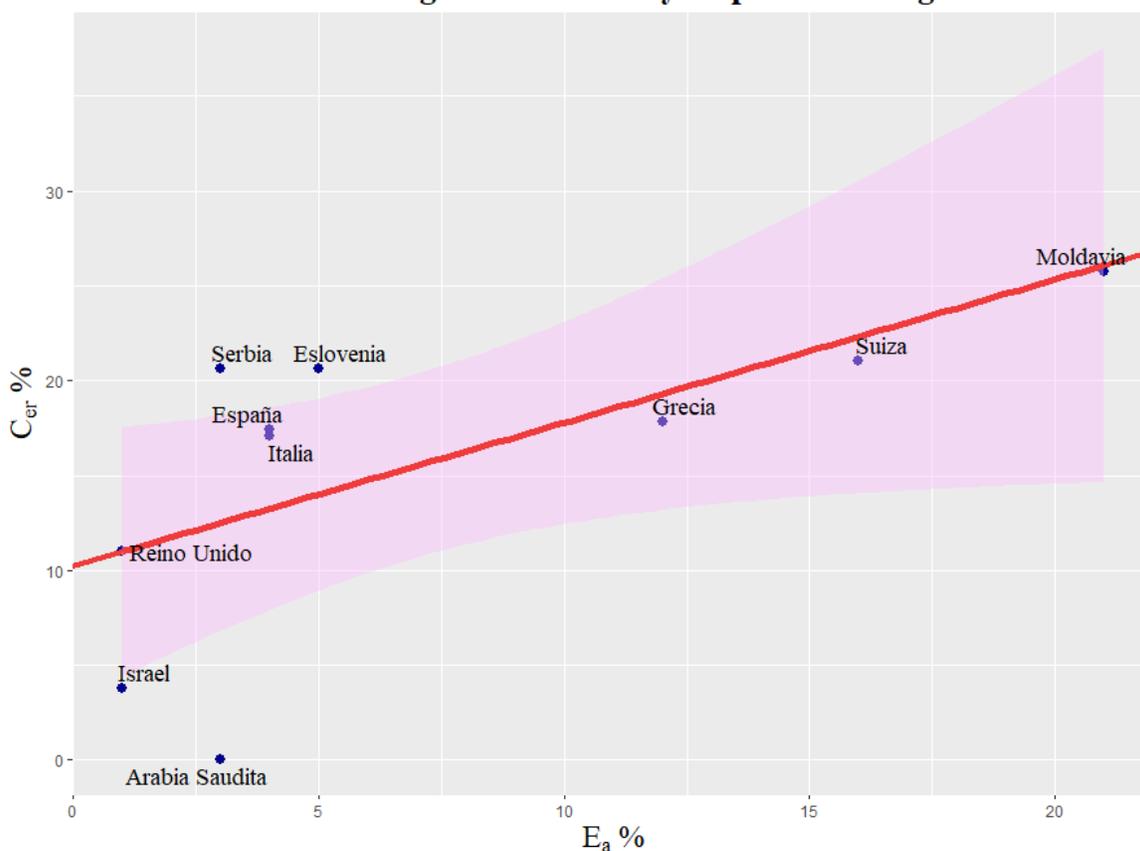


Figura 6.1: Gráfico de regresión que ilustra la relación entre el consumo de energías renovables (eje Y) y el empleo en agricultura (eje X). La línea de regresión muestra una correlación positiva entre las variables. Los datos fueron recopilados en 2018. La región sombreada indica el intervalo de confianza.

Interpretamos la ecuación de la manera siguiente, la constante representa el porcentaje estimado de consumo de energías renovables que es aproximadamente 10.222%, esto cuando se tiene el 0% de empleados en agricultura. El número de empleados en agricultura representa el incremento en éstos en función de cada valor de la incógnita de empleados en agricultura.

Donde tenemos que el consumo de energías renovables representa la variable dependiente. Por lo que tenemos que: si el promedio de empleados en agricultura es cero, entonces el porcentaje de consumo de energías renovables será de 10.222%.

En otras palabras, si se incrementa el número de trabajadores en la agricultura, se producirá un incremento en el consumo de energías renovables.

Tabla de coeficientes

	Estimado	Error estándar	p-value
Intersección	10.222	3.077	0.01
Empleados en agricultura	0.756	0.321	0.046

Tabla 6.1: Coeficientes de regresión para el modelo que predice la relación entre el consumo de energías renovables y los empleados en agricultura. Los coeficientes se estimaron utilizando un modelo de regresión lineal simple.

Tenemos que la variable predictora empleados en agricultura es un predictor estadísticamente significativo de la variable de respuesta consumo de energías renovables.

También podemos calcular los intervalos de confianza al 95%. Realizaremos los cálculos de manera que se pueda observar el procedimiento. Posteriormente, continuaremos realizando las pruebas en R. Por lo tanto, tenemos lo siguiente:

$$95\%CI = \hat{\beta} \pm t_{\alpha/2} \times SE$$

Debemos calcular primero el coeficiente de regresión estimado, seguido del valor crítico 't' y la desviación estándar para poder calcular el intervalo de confianza.

Tenemos que su valor para este caso es el siguiente:

$$\hat{\beta}_1 = \frac{\sum(y_i - \bar{y})(x_i - \bar{x})}{\sum(x_i - \bar{x})^2} = 0.756$$

Para calcular el valor crítico, primero debemos determinar los grados de libertad (df) y el nivel de confianza  $\alpha$ . En el caso de un nivel de confianza del 95%,  $\alpha$  sería 0.05. Una vez que tenemos estos valores, podemos consultar la tabla de la distribución  $t$  de Student para encontrar el valor crítico asociado. Por lo tanto, consecuentemente, tenemos:

$$95\%CI = 0.756 \pm (-2.306)0.321$$

De donde se obtiene lo siguiente:

$$95\%CI = [0.0153, 1.496]$$

Podemos afirmar con un 95% de confianza que la diferencia promedio en  $y$  entre grupos que difieren en una unidad en  $x$  está en algún lugar entre el intervalo cerrado  $[0.0153, 1.496]$ .

## Importaciones de bienes-servicios y consumo de energías renovables

Para los siguientes análisis, seguiremos el mismo proceso que en el caso anterior. Por lo tanto, omitiremos los procedimientos y presentaremos solo los resultados. En este caso, examinaremos la relación entre las importaciones de bienes-servicios y el consumo de energías renovables.

Donde tenemos que la ecuación de regresión de mínimos cuadrados esta dada por:

$$I_{bys} = 20.613 + 1.47C_{er}$$

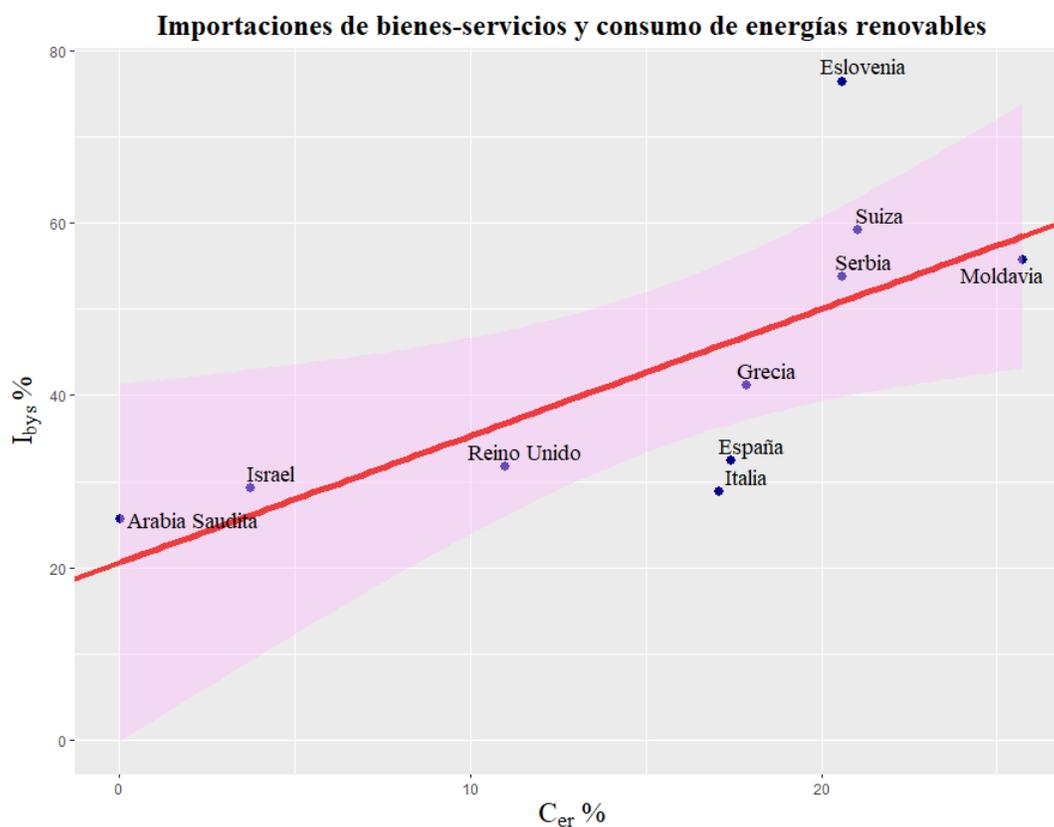


Figura 6.2: Gráfico de regresión que muestra la relación entre las importaciones de bienes y servicios (eje Y) y el consumo de energías renovables (eje X). La línea de regresión indica una correlación positiva. Los datos se recopilaron en 2018. La región sombreada indica el intervalo de confianza.

Además, tenemos la tabla (6.2) de coeficientes a continuación:

---

**Tabla de coeficientes**

---

	Estimado	Error estándar	p-value
Intersección	20.613	8.997	0.051
Consumo de energías renovables	1.47	0.519	0.022

Tabla 6.2: Coeficientes de regresión para el modelo que predice la relación entre las importaciones de bienes-servicios y el consumo de energías renovables. Los coeficientes se estimaron utilizando un modelo de regresión lineal simple.

Como tenemos que el  $p\text{-value} = 0.022$  entonces la variable predictora consumo de energías renovables es un predictor estadísticamente significativo de la variable de respuesta importaciones de bienes y servicios.

Por lo que podemos decir que estamos el 95% seguros de que la diferencia promedio en  $y$  entre grupos que difieren en una unidad en  $x$  está en el intervalo  $[0.273, 2.666]$

### Miscelánea de bienes-servicios y empleados en agricultura

En este tercer caso estamos frente a la relación entre la miscelánea de bienes-servicios y los empleados en agricultura. De donde tenemos lo siguiente:

La ecuación del modelo de regresión esta dada de la siguiente manera y también podemos observar su gráfico de dispersión asociado:

$$M2 = 3.807 + 0.685E_a$$

Podemos ver en la siguiente tabla 6.3 algunos datos que sobre la relación entre las variables donde podemos ver el error estándar y el  $p\text{-value}$ .

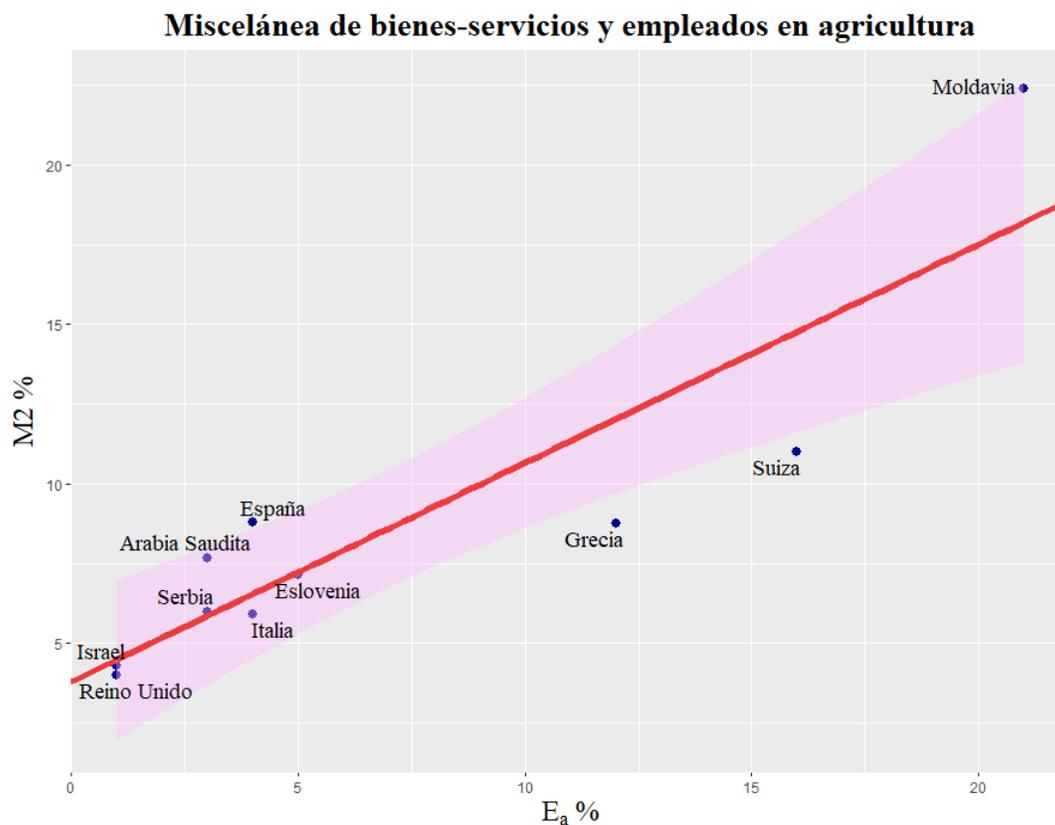


Figura 6.3: Gráfico de regresión que muestra la relación entre las miscelánea de bienes y servicios (eje Y) y los empleados en agricultura (eje X). La línea de regresión indica una correlación positiva. Los datos se recopilaron en 2018. La región sombreada indica el intervalo de confianza..

Tabla de coeficientes			
	Estimado	Error estándar	p-value
Intersección	3.807	1.176	0.011
Empleados en agricultura	0.685	0.122	0.0005

Tabla 6.3: Coeficientes de regresión para el modelo que predice la relación entre la miscelánea de bienes-servicios y los empleados en agricultura. Los coeficientes se estimaron utilizando un modelo de regresión lineal simple.

Aquí tenemos que señalar que el  $p\text{-value} = 0.0005$  es decir que rechazamos la hipótesis.

Por lo que podemos decir que estamos el 95 % seguros de que la diferencia promedio en  $y$  entre grupos que difieren en una unidad en  $x$  está en el intervalo  $[0.402, 0.968]$

### Ingresos fiscales y gasto público

En este cuarto caso tenemos la relación entre ingresos fiscales y gasto público. La ecuación de regresión de los mínimos cuadrados está dada de la siguiente forma, así mismo podemos observar su gráfico de dispersión asociado:

$$I_f = -2.3 + 0.625G_p$$

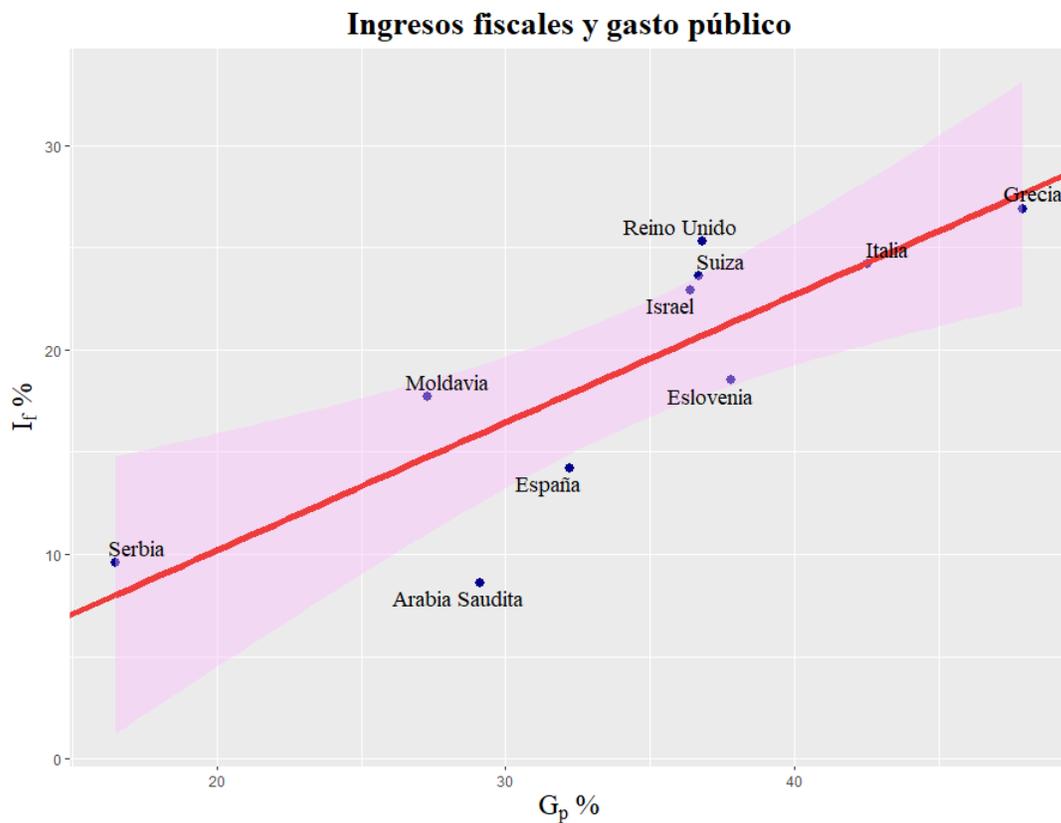


Figura 6.4: Gráfico de regresión que muestra la relación entre ingresos fiscales (eje Y) y el gasto público (eje X). La línea de regresión indica una correlación positiva. Los datos se recopilaron en 2018. La región sombreada indica el intervalo de confianza.

En la tabla 6.4 podemos ver la tabla de coeficientes asociada a esta relación.

Tabla de coeficientes			
	Estimado	Error estándar	p-value
Intersección	-2.3	5.289	0.675
Gasto público	0.625	0.149	0.0031

Tabla 6.4: Coeficientes de regresión para el modelo que predice la relación entre ingresos fiscales y el gasto público. Los coeficientes se estimaron utilizando un modelo de regresión lineal simple.

Tenemos que el valor del  $p\text{-value} = 0.003$ , lo que confirma nuestra sospecha de que podemos descartar la hipótesis nula.

Podemos decir que estamos el 95% seguros de que la diferencia promedio en  $y$  entre grupos que difieren en una unidad en  $x$  está en el intervalo cerrado  $[0.279, 0.970]$ .

### Ingresos fiscales y transporte

Para este quinto caso tenemos en primer lugar la relación entre ingresos fiscales y el transporte.

La ecuación de regresión de los mínimos cuadrados está dada de la siguiente forma, además también tenemos el gráfico de dispersión asociado a este:

$$I_f = 10.191 + 0.68T$$

Tabla de coeficientes

	Estimado	Error estándar	p-value
Intersección	10.191	5.955	0.125
Transporte	0.68	0.428	0.151

Tabla 6.5: Coeficientes de regresión para el modelo que predice la relación entre ingresos fiscales y transporte. Los coeficientes se estimaron utilizando un modelo de regresión lineal simple.

En este caso, de acuerdo a la tabla 6.5, no rechazamos la hipótesis nula dado que nuestro  $p\text{-value} = 0.151$ .

### Crecimiento poblacional y trabajadores asalariados

Para este caso donde los coeficientes de correlación fueron positivos, tenemos la relación dada entre crecimiento poblacional y el trabajadores asalariados, estos datos consisten en los porcentajes asociados a las variables antes mencionadas.

La ecuación de regresión de los mínimos cuadrados está dada de la siguiente forma y además, se tiene el gráfico de dispersión asociado:

$$C_p = -7.622 + 0.098T_a$$

En la tabla siguiente (6.6) podemos ver el valor de los coeficientes asociados a esta relación.

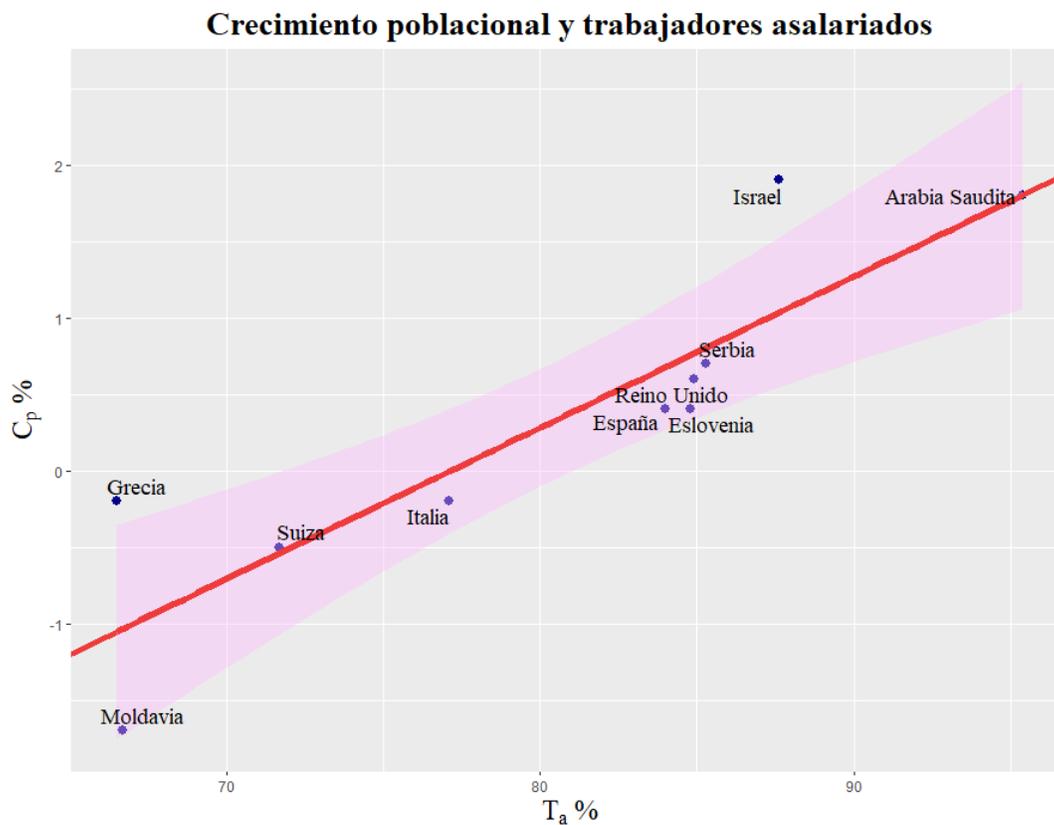


Figura 6.5: Gráfico de regresión que muestra la relación entre el crecimiento poblacional (eje Y) y los trabajadores asalariados (eje X). La línea de regresión indica una correlación positiva. Los datos se recopilaron en 2018. La región sombreada indica el intervalo de confianza.

Tabla de coeficientes

	Estimado	Error estándar	p-value
Intersección	-7.622	1.482	0.0008
Trabajadores asalariados	0.098	0.018	0.0006

Tabla 6.6: Coeficientes de regresión para el modelo que predice la relación entre crecimiento poblacional y los trabajadores asalariados. Los coeficientes se estimaron utilizando un modelo de regresión lineal simple.

Dado que  $p\text{-value} = 0.0006$ , la variable predictora trabajadores asalariados es un predictor estadísticamente significativo de la variable de respuesta crecimiento poblacional.

Por lo que podemos decir que estamos el 95 % seguros de que la diferencia promedio en  $y$  entre grupos que difieren en una unidad en  $x$  está en el intervalo  $[0.056, 0.141]$

## Consumo de energías renovables y emisiones de dióxido de carbono

En este caso tenemos la relación dada entre consumo de energías renovables y emisiones de  $CO_2$ .

La ecuación de regresión de los mínimos cuadrados está dada de la siguiente forma y además, se tiene el gráfico de dispersión asociado:

$$E_{CO_2} = 394.60 - 14.50C_{er}$$

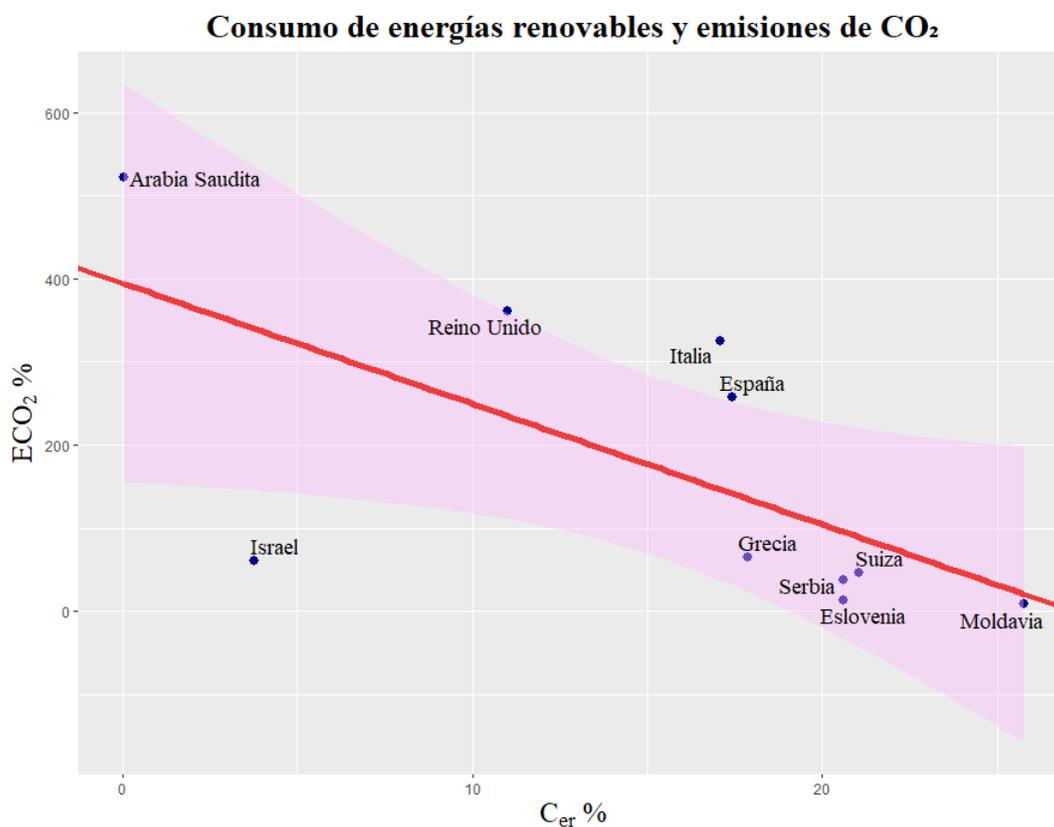


Figura 6.6: Gráfico de regresión que muestra la relación emisiones de  $CO_2$  (eje Y) y consumo de energías renovables (eje X). La línea de regresión indica una correlación negativa. Los datos se recopilaron en 2018. La región sombreada indica el intervalo de confianza.

En la tabla 6.7 podemos ver el valor de los coeficientes asociados a esta relación. Dado que el  $p\text{-value} = 0.041$ , la variable predictora consumo de energías renovables es un predictor estadísticamente significativo de la variable de respuesta emisiones de  $CO_2$ .

Tabla de coeficientes			
	Estimado	Error estándar	p-value
Intersección	394.6	103.877	0.005
Consumo de energías renovables	-14.502	5.993	0.041

Tabla 6.7: Coeficientes de regresión para el modelo que predice la relación entre emisiones de  $CO_2$  y consumo de energías renovables. Los coeficientes se estimaron utilizando un modelo de regresión lineal simple.

Por lo que podemos decir que estamos el 95% seguros de que la diferencia promedio en  $y$  entre grupos que difieren en una unidad en  $x$  está en el intervalo  $[-28.320, -0.682]$ .

### Crecimiento poblacional y empleados en agricultura

En este caso tenemos la relación dada entre crecimiento poblacional y empleados en agricultura.

La ecuación de regresión de los mínimos cuadrados está dada de la siguiente forma y además, se tiene el gráfico de dispersión asociado:

$$E_a = 8.745 - 5.452C_p$$

En la tabla siguiente (6.8) podemos ver el valor de los coeficientes asociados a esta relación.

Dado que  $p\text{-value} = 0.0022$ , la variable predictora crecimiento poblacional es un predictor estadísticamente significativo de la variable de respuesta empleados en agricultura.

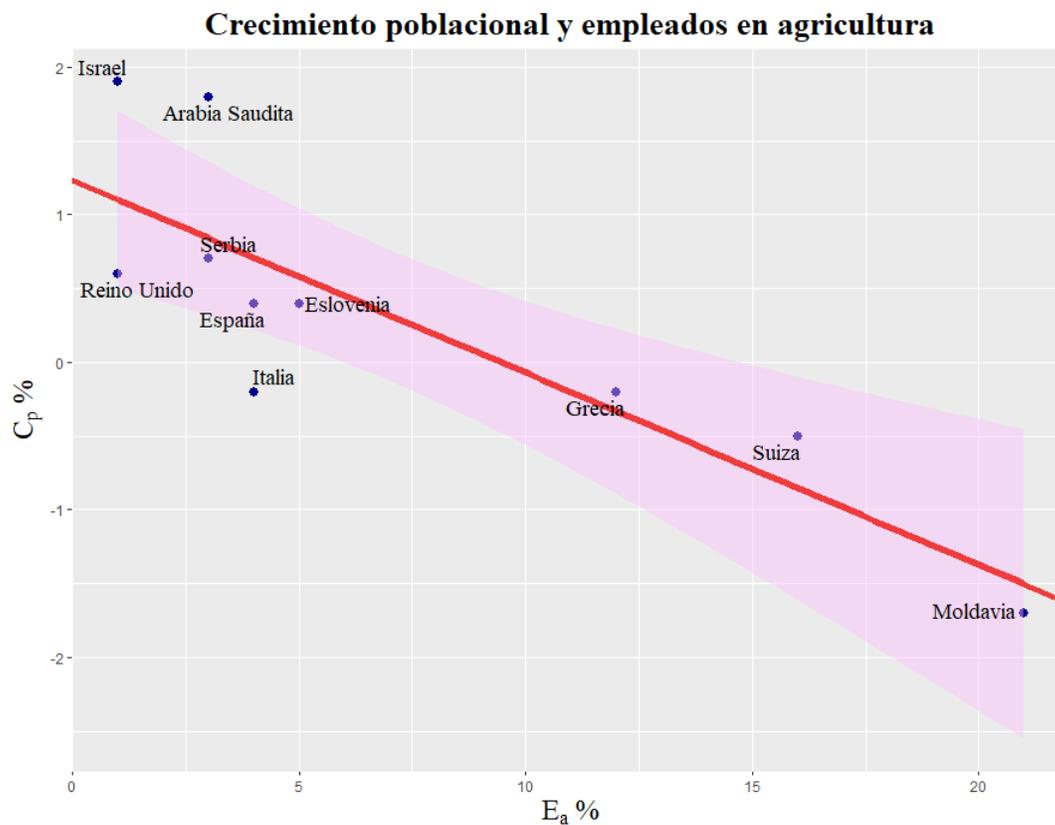


Figura 6.7: Gráfico de regresión que muestra la relación crecimiento poblacional (eje Y) y empleados en agricultura (eje X). La línea de regresión indica una correlación negativa. Los datos se recopilaron en 2018. La región sombreada indica el intervalo de confianza.

Tabla de coeficientes

	Estimado	Error estándar	p-value
Intersección	8.745	1.308	0.0001
Crecimiento poblacional	-5.452	1.233	0.0022

Tabla 6.8: Coeficientes de regresión para el modelo que predice la relación entre los empleados en agricultura y el crecimiento poblacional. Los coeficientes se estimaron utilizando un modelo de regresión lineal simple.

Por lo que podemos decir que estamos el 95 % seguros de que la diferencia promedio en  $y$  entre grupos que difieren en una unidad en  $x$  está en el intervalo  $[-8.296, -2.608]$ .

### Empleados en agricultura y trabajadores asalariados

En este caso tenemos la relación dada entre empleados en agricultura y trabajadores asalariados.

La ecuación de regresión de los mínimos cuadrados está dada de la siguiente forma y además, se tiene el gráfico de dispersión asociado:

$$T_a = 88.562 - 1.166E_a$$

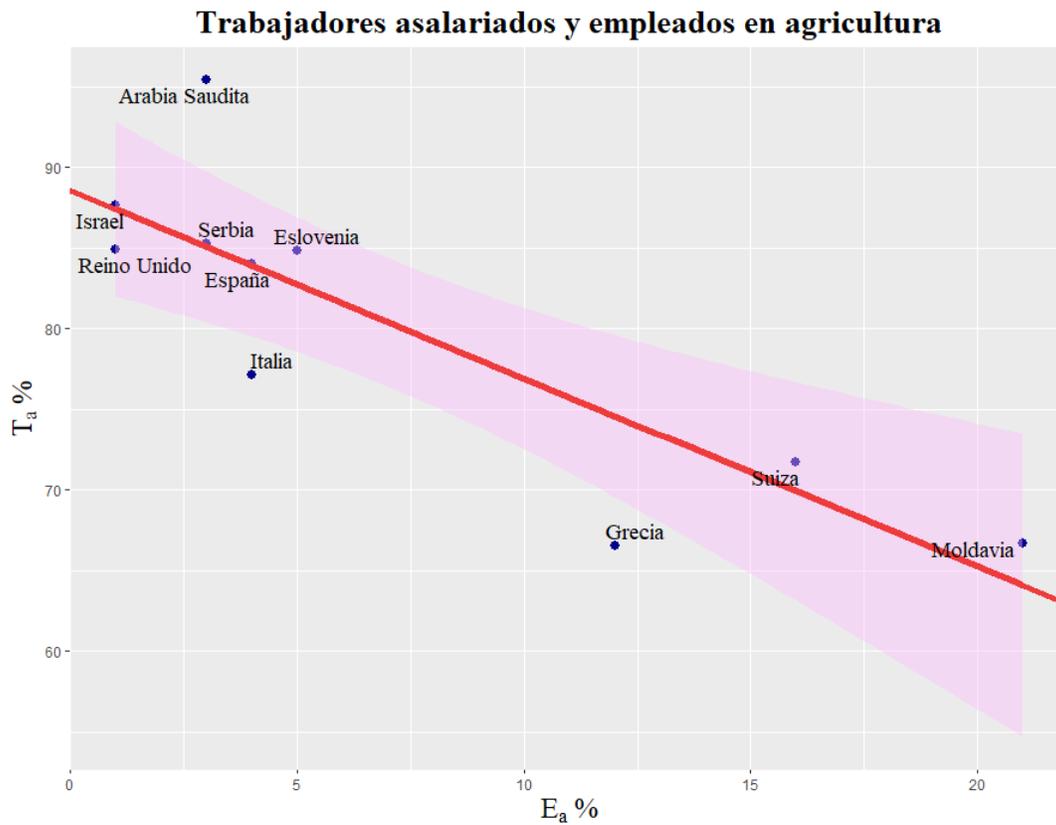


Figura 6.8: Gráfico de regresión que muestra la relación entre trabajadores asalariados (eje Y) y empleados en agricultura (eje X). La línea de regresión indica una correlación negativa. Los datos se recopilieron en 2018. La región sombreada indica el intervalo de confianza.

En la tabla siguiente (6.9) podemos ver el valor de los coeficientes asociados a esta relación.

Tabla de coeficientes			
	Estimado	Error estándar	p-value
Intersección	88.562	2.53	0.000000000486
Empleados en agricultura	-1.166	0.264	0.002

Tabla 6.9: Coeficientes de regresión para el modelo que predice la relación entre los trabajadores asalariados y los empleados en agricultura. Los coeficientes se estimaron utilizando un modelo de regresión lineal simple.

Dado que  $p\text{-value} = 0.002$ , la variable predictora empleados en agricultura es un predictor estadísticamente significativo de la variable de respuesta trabajadores asalariados.

Por lo que podemos decir que estamos el 95 % seguros de que la diferencia promedio en  $y$  entre grupos que difieren en una unidad en  $x$  está en el intervalo  $[-1.775, -0.557]$ .

### Importaciones de bienes-servicios y emisiones de dióxido de carbono

En este caso tenemos la relación dada entre importaciones de bienes-servicios y emisiones de  $CO_2$ .

La ecuación de regresión de los mínimos cuadrados está dada de la siguiente forma y además, se tiene el gráfico de dispersión asociado:

$$E_{CO_2} = 511.35 - 7.870I_{bys}$$

En la tabla 6.10 podemos ver el valor de los coeficientes asociados a esta relación.

Dado que  $p\text{-value} = 0.015$ , la variable predictora importaciones de bienes y servicios es un predictor estadísticamente significativo de la variable de respuesta emisiones de  $CO_2$ .

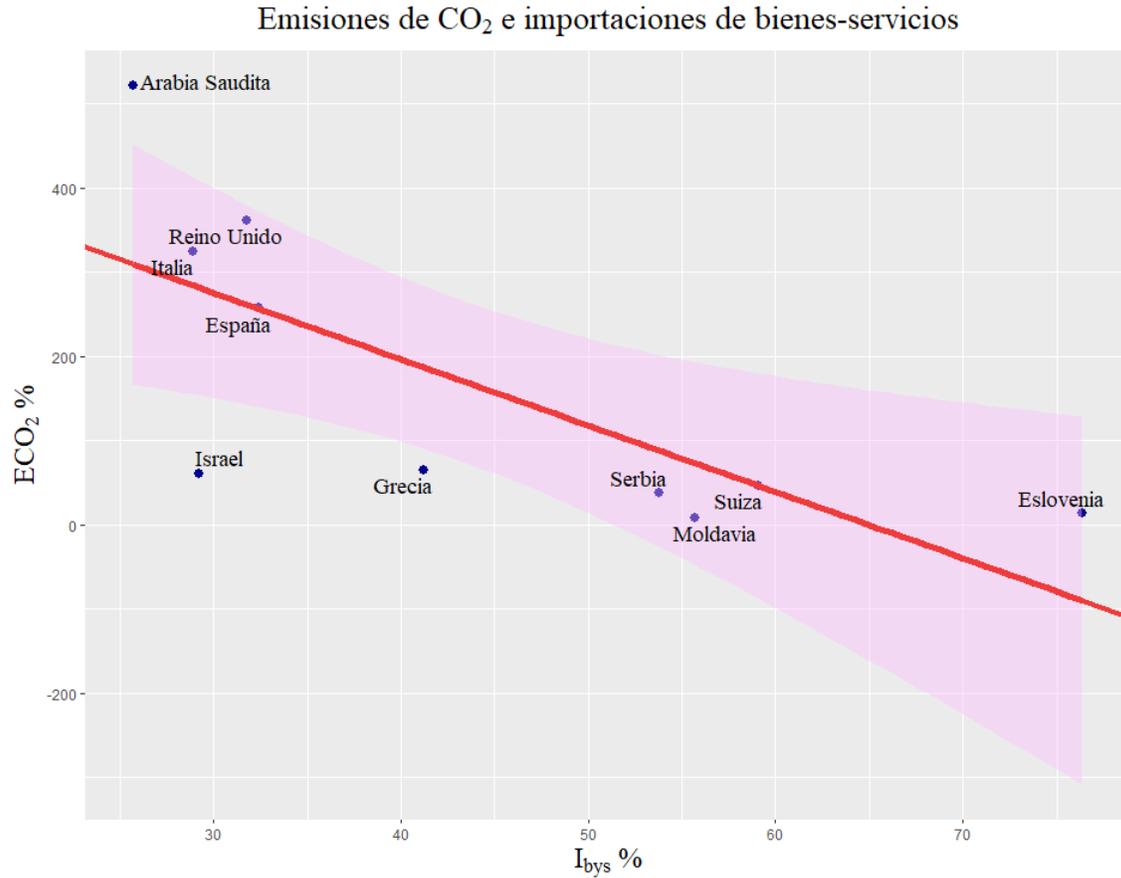


Figura 6.9: Gráfico de regresión que muestra la relación entre emisiones de  $CO_2$  (eje Y) e importaciones de bienes y servicios (eje X). La línea de regresión indica una correlación negativa. Los datos se recopilaron en 2018. La región sombreada indica el intervalo de confianza.

**Tabla de coeficientes**

	Estimado	Error estándar	p-value
Intersección	511.359	119.546	0.0027
Importaciones de bienes y servicios	-7.87	2.582	0.015

Tabla 6.10: Coeficientes de regresión para el modelo que predice la relación entre las emisiones de  $CO_2$  y las importaciones de bienes y servicios. Los coeficientes se estimaron utilizando un modelo de regresión lineal simple.

Por lo que podemos decir que estamos el 95% seguros de que la diferencia promedio en  $y$  entre grupos que difieren en una unidad en  $x$  está en el intervalo  $[-13.825, -1.915]$ .

### Recreación-cultura y comidas-bebidas no alcohólicas

En este caso tenemos la relación dada entre recreación-cultura y comidas-bebidas no alcohólicas.

La ecuación de regresión de los mínimos cuadrados está dada de la siguiente forma y además, se tiene el gráfico de dispersión asociado:

$$R_c = 10.834 - 0.222C_b$$

En la tabla siguiente (6.11) podemos ver el valor de los coeficientes asociados a esta relación.

Tabla de coeficientes

	Estimado	Error estándar	p-value
Intersección	10.834	2.407	0.002
Comidas y bebidas alcohólicas	-0.222	0.108	0.072

Tabla 6.11: Coeficientes de regresión para el modelo que predice la relación entre recreación-cultura ( $R_c$ ) y comidas-bebidas no alcohólicas ( $C_b$ ). Los coeficientes se estimaron utilizando un modelo de regresión lineal simple.

Aquí tenemos que señalar que el  $p\text{-value} = 0.07$  es decir que mantenemos la hipótesis nula.

### Sobre los modelos

Según los resultados obtenidos en la construcción de nuestros modelos de regresión, podemos observar que los cinco casos cumplen con las suposiciones de linealidad. Es decir:

- Normalidad
- Linealidad
- Homoscedasticidad

Para llevar a cabo esta tarea, recurriremos nuevamente al software R, que, mediante el empleo de métodos numéricos, nos permitirá realizar el diagnóstico de cada uno de los casos. El paquete llamado "gvlma" nos permite hacer pruebas automáticas para varios supuestos generales de modelos lineales.

En cuanto a la normalidad, nos basaremos en los resultados de asimetría y curtosis (*kurtosis*). La asimetría se refiere a una medida de la falta de simetría en una distribución de probabilidades. Mientras que la curtosis se refiere a una medida de la forma de una distribución de probabilidad, específicamente con relación a la “agudeza” o “apuntalamiento” de su pico. Estos dos nos proporcionarán información acerca de si los residuos de nuestros datos se ajustan a la distribución normal.

La linealidad de nuestro modelo estará determinada por el valor de *Global Stat*, que nos permitirá evaluar la hipótesis nula de que nuestro modelo es efectivamente una combinación lineal de los predictores. En otras palabras, *Global Stat* nos proporcionará información acerca de si nuestro modelo cumple con las suposiciones de linealidad. Por otro lado, la homoscedasticidad se refiere a la presencia de variaciones en la varianza de los residuos. Para evaluar esta característica, se realizan pruebas estadísticas que nos permiten determinar si la varianza de los residuos es relativamente constante. En caso contrario, estaríamos en presencia de heteroscedasticidad, lo cual podría afectar la validez de nuestro modelo.

Evaluación de supuestos del modelo lineal, utilizando la prueba global en 4 grados de libertad					
Nivel de significancia = 0.05					
Modelo	Linealidad	Asimetría	Curtosis	Heterocedasticidad	
Consumo de energías renovables y empleados en agricultura	p-value = 0.401	p-value = 0.400	p-value = 0.846	p-value = 0.097	
<b>Tipo de decisión</b>	<b>Aceptable</b>	<b>Aceptable</b>	<b>Aceptable</b>	<b>Aceptable</b>	
Importaciones de bienes-servicios y consumo de energías renovables	p-value = 0.4304	p-value = 0.4045	p-value = 0.8569	p-value = 0.1646	
<b>Tipo de decisión</b>	<b>Aceptable</b>	<b>Aceptable</b>	<b>Aceptable</b>	<b>Aceptable</b>	
Miscelanea de bienes-servicios y empleados en agricultura	p-value = 0.311	p-value = 0.975	p-value = 0.743	p-value = 0.454	
<b>Tipo de decisión</b>	<b>Aceptable</b>	<b>Aceptable</b>	<b>Aceptable</b>	<b>Aceptable</b>	
Ingresos fiscales y gasto público	p-value = 0.750	p-value = 0.378	p-value = 0.742	p-value = 0.332	
<b>Tipo de decisión</b>	<b>Aceptable</b>	<b>Aceptable</b>	<b>Aceptable</b>	<b>Aceptable</b>	
Crecimiento poblacional y trabajadores asalariados	p-value = 0.7971	p-value=0.2611	p-value=0.8537	p-value=0.8921	
<b>Tipo de decisión</b>	<b>Aceptable</b>	<b>Aceptable</b>	<b>Aceptable</b>	<b>Aceptable</b>	

Tabla 6.12: Resultados de la prueba de suposiciones para los modelos de regresión estudiados.

Evaluación de supuestos del modelo lineal, utilizando la prueba global en 4 grados de libertad					
Nivel de significancia = 0.05					
Modelo	Linealidad	Asimetría	Curtosis	Heterocedasticidad	
Emissiones de CO <sub>2</sub> y consumo de energías renovables	p-value = 0.939	p-value = 0.500	p-value = 0.825	p-value = 0.860	
<b>Tipo de decisión</b>	<b>Aceptable</b>	<b>Aceptable</b>	<b>Aceptable</b>	<b>Aceptable</b>	
Empleados en agricultura y crecimiento poblacional	p-value = 0.299	p-value = 0.765	p-value = 0.364	p-value = 0.765	
<b>Tipo de decisión</b>	<b>Aceptable</b>	<b>Aceptable</b>	<b>Aceptable</b>	<b>Aceptable</b>	
Trabajadores asalariados y empleados en agricultura	p-value = 0.358	p-value = 0.755	p-value = 0.929	p-value = 0.0721	
<b>Tipo de decisión</b>	<b>Aceptable</b>	<b>Aceptable</b>	<b>Aceptable</b>	<b>Aceptable</b>	
Emissiones de CO <sub>2</sub> e importaciones de bienes y servicios	p-value = 0.437	p-value=0.923	p-value=0.777	p-value=0.349	
<b>Tipo de decisión</b>	<b>Aceptable</b>	<b>Aceptable</b>	<b>Aceptable</b>	<b>Aceptable</b>	
Recreación-cultura y comida-bebidas no alcohólicas	p-value = 0.562	p-value=0.294	p-value=0.924	p-value=0.913	
<b>Tipo de decisión</b>	<b>Aceptable</b>	<b>Aceptable</b>	<b>Aceptable</b>	<b>Aceptable</b>	

Tabla 6.13: Resultados de la prueba de suposiciones para los modelos de regresión estudiados, segunda parte.

De acuerdo con nuestros análisis, podemos observar que nuestros modelos cumplen con las suposiciones de los modelos de regresión lineal. Además, a partir de los datos obtenidos, podemos obtener más información acerca de nuestro modelo. Por ejemplo, la siguiente tabla nos proporciona información adicional que nos permite profundizar en nuestro análisis.

Información estadística acerca de la cualidad del modelo					
Modelo	Error residual estándar	Múltiple raíz de R	Raíz ajustada de R	Estadística de F	p-value
Consumo de energías renovables vs empleados en agricultura	6.625 con 8 grados de libertad	0.409	0.335	5.54 en 1 y 8 Grados de libertad	0.046
Importaciones de bienes y servicios vs consumo de energías renovables	12.69 con 8 grados de libertad	0.5007	0.438	8.023 en 1 y 8 Grados de libertad	0.023
Miscelánea de bienes y servicios vs empleados en agricultura	2.54 con 8 grados de libertad	0.795	0.77	31.16 en 1 y 8 Grados de libertad	0.0005
Ingresos fiscales vs gasto público	3.903 con 8 grados de libertad	0.684	0.645	17.39 en 1 y 8 Grados de libertad	0.003

Tabla 6.14: Evaluación de la calidad de los modelos de regresión entre las distintas parejas analizadas. Los indicadores de calidad incluyen el coeficiente de determinación  $R^2$ , el error estándar de la estimación y el  $p$ -value del modelo, primera parte.

Información estadística acerca de la cualidad del modelo					
Modelo	Error residual estándar	Múltiple raíz de R	Raíz ajustada de R	Estadística de F	p-value
Emisiones de CO <sub>2</sub> vs consumo de energías renovables	146.5 con 8 grados de libertad	0.42	0.35	5.85 en 1 y 8 Grados de libertad	0.041
Empleados en agricultura vs crecimiento poblacional	3.942 con 8 grados de libertad	0.7	0.67	19.54 en 1 y 8 Grados de libertad	0.0022
Trabajadores asalariados vs empleados en agricultura	5.465 con 8 grados de libertad	0.7	0.67	19.49 en 1 y 8 Grados de libertad	0.0022
Emisiones de CO <sub>2</sub> vs importaciones de bienes y servicios	131.2 con 8 grados de libertad	0.53	0.47	9.29 en 1 y 8 Grados de libertad	0.015

Tabla 6.15: Evaluación de la calidad de los modelos de regresión entre las distintas parejas analizadas. Los indicadores de calidad incluyen el coeficiente de determinación  $R^2$ , el error estándar de la estimación y el  $p$ -value del modelo, segunda parte.



## 7 Conclusiones

A partir de los resultados derivados de nuestros modelos, podemos concluir lo siguiente: En el primer modelo, que establece una relación entre el consumo de energías renovables con el número de empleados en agricultura, se observa que el modelo explica el 40.9 % de la variabilidad en los datos de respuesta en torno a la media. En el segundo modelo, que relaciona las importaciones de bienes y servicios con el consumo de energías renovables, se observa que el modelo explica el 50 % de la variabilidad en los datos de respuesta en torno a la media. En el tercer modelo, que relaciona la miscelánea de bienes y servicios con el número de empleados en agricultura, se observa que el modelo explica el 79 % de la variabilidad en los datos de respuesta. En el cuarto modelo, que relaciona los ingresos fiscales con el gasto público, se observa una variabilidad del 68 % en los datos de respuesta.

En relación con el modelo que vincula los ingresos fiscales con el transporte, se observa uno de los menores porcentajes de variabilidad, con un 23 % en los datos de respuesta. Finalmente, en el último caso tenemos que se presenta una variabilidad del 78 % entre el crecimiento poblacional y los trabajadores asalariados.

Dada la variabilidad cercana al 80 % observada en algunos modelos, es plausible que una transformación de los datos podría resultar beneficiosa para alcanzar un mayor porcentaje de variabilidad. Si nuestros modelos satisfacen las suposiciones y se ajustan de manera adecuada, podríamos contemplar la realización de una transformación de datos para mejorar su ajuste. Dado que se cumplen los supuestos de regresión para los modelos, podríamos considerar una transformación logarítmica o de raíz cuadrada.

En lo que respecta a las relaciones entre variables, la relación entre la miscelánea de bienes-servicios y el consumo de energías renovables, así como entre el consumo de energías renovables y el número de empleados en agricultura, sugiere que podría ser de interés llevar a cabo una investigación sobre el comportamiento de estas dos variables en otros países. Esto podría proporcionar una visión más amplia y permitirnos entender si existen patrones similares en diferentes contextos. Sería interesante explorar si hay material disponible que aborde estas relaciones en otros países. Esto podría enriquecer aún más nuestro análisis y comprensión de estas variables.

En cuanto a la relación entre las energías renovables y el empleo en la agricultura, se han realizado investigaciones [2] que confirman la existencia de una relación entre ambas. [16] De acuerdo con estos estudios, aproximadamente 2.5 mil millones de personas a nivel global dependen de la agricultura para su subsistencia, lo que

posiciona a este sector como un elemento crucial para el desarrollo. No obstante, los patrones actuales de consumo energético en los sistemas agroalimentarios revelan una disparidad regional, una carencia de acceso a energía moderna y una dependencia constante de los combustibles fósiles. Estos factores representan desafíos significativos que deben ser abordados para garantizar un futuro sostenible y equitativo.

Según el nivel de desarrollo del país, se estima que cerca de un cuarto del consumo energético se destina a la fase de producción (cultivos, ganadería y pesca), el 45 % se emplea en el procesamiento y distribución de alimentos, y el 30 % se utiliza en la venta al por menor, preparación y cocción de alimentos. Se calcula que en 2019, el 35 % de la población aún recurría a la madera para cocinar. Esta práctica no sólo tiene graves repercusiones para la salud, sino que también impacta negativamente en el medio ambiente.

La energía renovable puede desempeñar un papel crucial en la satisfacción de las necesidades de electricidad, calefacción, refrigeración y transporte de los sistemas alimentarios, tanto en países desarrollados como en desarrollo. Al hacerlo, puede contribuir significativamente a los esfuerzos para erradicar el hambre, reducir las emisiones de gases de efecto invernadero y aumentar la capacidad de adaptación de los agricultores y las empresas agrícolas. Además, puede ayudar a incrementar los ingresos y a minimizar el impacto ambiental del sector alimentario. Estos beneficios subrayan la importancia de fomentar y adoptar fuentes de energías renovables en el ámbito de la agricultura y la alimentación.

La irrigación solar representa una de las aplicaciones de energías renovables que se están incorporando en las cadenas agroalimentarias. Esta técnica ha sido adoptada de manera extensa para mejorar el acceso al agua, lo cual posibilita la realización de múltiples ciclos de cultivo y fortalece la resiliencia frente a los patrones cambiantes de precipitación atribuidos al calentamiento global. El empleo de bombas de riego solares ha propiciado un incremento en los ingresos de los agricultores de hasta un 50 % o más en India, en comparación con el riego que depende exclusivamente de las lluvias. Este avance demuestra el potencial de las energías renovables para transformar la agricultura y mejorar la vida de las personas.

Para impulsar el uso de energías renovables en los sistemas agroalimentarios y promover la seguridad energética y alimentaria, así como para avanzar hacia los Objetivos de Desarrollo Sostenible y el cumplimiento del Acuerdo de París, es esencial una acción coordinada. Esta acción debe involucrar a los responsables de la toma de decisiones en el gobierno, el sector privado, organizaciones internacionales, instituciones financieras, la academia y organizaciones no gubernamentales. Solo a través de un esfuerzo conjunto y coordinado podremos lograr un futuro sostenible y resiliente.

Algunas conclusiones sobre las relaciones obtenidas:

1. La relación entre las energías renovables y las importaciones de bienes y servicios se fundamenta en el compromiso cada vez mayor del sector privado de utilizar energías renovables en el transporte de mercancías. Las opciones más comunes incluyen energía solar, eólica, hidráulica, geotérmica y sinérgica. Cabe recordar que en este caso, ambas variables presentan una correlación de  $\rho = 0.9$ , lo que sitúa a esta relación como la segunda con el coeficiente más alto en nuestro estudio. Este hallazgo subraya la creciente interdependencia entre la adopción de energías renovables y las prácticas comerciales en el sector privado.
2. La relación entre el crecimiento poblacional y los trabajadores asalariados presenta un coeficiente de correlación de  $\rho = 0.93$ , lo que representa la correlación más alta entre los pares estudiados en nuestro análisis. Tradicionalmente, un incremento en la población implica un aumento en la futura fuerza laboral, lo que se traduce en una mayor productividad y un incremento en los empleos remunerados. Sin embargo, es importante tener en cuenta que este fenómeno está sujeto al contexto social y político del país en cuestión. Este hallazgo subraya la importancia de considerar las dinámicas demográficas en el análisis del mercado laboral.
3. La relación entre el gasto público y los ingresos fiscales se manifiesta en la planificación del gasto público de diversos países. Los ingresos fiscales constituyen una parte significativa de los ingresos públicos, que sirven como fundamento para la elaboración de los presupuestos de gasto público. El gasto público contribuye a mejorar el acceso a la salud, garantizar la justicia, incrementar el empleo, fomentar el crecimiento económico y asegurar una vida digna, entre otros aspectos. En nuestro estudio, encontramos que la correlación entre estas dos variables es de  $\rho = 0.84$ , lo que las sitúa dentro del grupo que presenta un coeficiente de correlación superior a 0.8. Este hallazgo subraya la estrecha interrelación entre las políticas fiscales y el gasto público en el contexto económico.
4. La relación entre el gasto público y el transporte se evidencia principalmente en el gasto público destinado a infraestructura, que puede potenciar la conectividad y eficiencia del transporte, tanto público como privado. La construcción y mantenimiento de infraestructuras públicas, así como el uso compartido de carreteras, son ejemplos claros de esta relación. En nuestro estudio, encontramos que el coeficiente de correlación entre estas dos variables es de  $\rho = 0.79$ , lo que representa la correlación más baja entre los pares estudiados. Este hallazgo subraya la importancia de considerar el gasto público en infraestructura de transporte en el análisis económico.

Sobre la presencia de Arabia Saudita como un outlier repetitivo concluimos lo siguiente:

1. La presencia de Arabia Saudita como un valor atípico en varios análisis podría deberse a posibles inconsistencias en los datos obtenidos de su base de datos. La falta de algunos registros sobre este país podría ser el resultado de una recopilación de datos insuficiente, la ausencia de informes públicos o las políticas de privacidad. Un ejemplo notable es la falta de disponibilidad de datos sobre la tasa de incidencia de la pobreza. Este tipo de situaciones subraya la importancia de la precisión y la integridad de los datos en la realización de análisis estadísticos y la formulación de conclusiones significativas. Es crucial tener en cuenta estos factores al interpretar los resultados y considerar la posibilidad de realizar más investigaciones para validar los hallazgos.
2. La variabilidad del conjunto de datos es un aspecto crucial a considerar en cualquier análisis estadístico. En el caso de Arabia Saudita, su comportamiento atípico podría ser un indicativo de tendencias o patrones específicos en la región de Asia Occidental. Sería interesante expandir la investigación para incluir un conjunto más amplio de países de esta región. Esto permitiría comparar y contrastar los resultados, y posiblemente arrojar luz sobre si el comportamiento observado en Arabia Saudita es un fenómeno aislado o si es parte de una tendencia regional más amplia. Sin embargo, es importante tener en cuenta que cualquier análisis adicional debe realizarse con cuidado, teniendo en cuenta las posibles diferencias en la recopilación y el reporte de datos entre diferentes países.
3. Exclusión de outliers: La última hipótesis sugiere que se podría mejorar la modelización excluyendo los países que presentan una mayor incidencia como outliers, como Arabia Saudita o Israel.

# A Mínimos cuadrados

## ESTIMACIÓN DE PARÁMETROS

Según la información recopilada, nuestro objetivo es calcular los valores de los parámetros  $\beta_0$  y  $\beta_1$ . En otras palabras, buscamos encontrar la línea recta que mejor se ajuste a la representación de los puntos en el diagrama de dispersión, tomando en cuenta la variable predictora y la respuesta. [7]

Calculamos los valores de los parámetros utilizando el método de mínimos cuadrados. Este método nos proporciona la línea que reduce al mínimo la suma de los cuadrados de las distancias verticales entre cada punto y la línea. Dichas distancias verticales representan los errores en la variable de respuesta. Podemos obtener estos errores reescribiendo la ecuación de la siguiente manera:

Tenemos la ecuación de la regresión lineal dada de la siguiente forma:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad i = 1, 2, \dots, n$$

Si la reescribimos obtenemos la ecuación de la forma:

$$\varepsilon_i = y_i - \beta_0 - \beta_1 x_i \quad i = 1, 2, \dots, n$$

Así mismo la suma de los cuadrados de estas distancias puede ser escrita de forma consecutiva como:

$$S(\beta_0, \beta_1) = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2$$

Los valores de  $\hat{\beta}_0$  y  $\hat{\beta}_1$  que minimizan  $S(\beta_0, \beta_1)$  están dados por:

$$\hat{\beta}_1 = \frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2}$$

y

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Notemos que la fórmula está dada por  $\hat{\beta}_1$  que antes se encontraba en términos de  $\hat{\beta}_0$  ya que  $\hat{\beta}_0$  está escrita en términos de  $\hat{\beta}_1$ .

Las estimaciones de  $\hat{\beta}_0$  y  $\hat{\beta}_1$  son llamadas los mínimos cuadrados estimados de  $\beta_0$  y  $\beta_1$  ya que estos son la solución del método de los mínimos cuadrados. La intersección y la pendiente de la recta garantizan la menor suma posible de la suma de cuadrados de la distancia vertical para cada punto de la línea. Por esta razón la recta es llamada la línea de regresión de los mínimos cuadrados que está dada por la ecuación:

$$\widehat{Y} = \widehat{\beta}_0 + \widehat{\beta}_1 X \quad (\text{A.0.1})$$

Notemos que los mínimos cuadrados van a existir ya que siempre podemos encontrar la línea que nos dé la suma de los mínimos cuadrados de la distancia vertical. En algunos casos la línea de los mínimos cuadrados podría no ser única. Pero estos casos no son comunes en la práctica.

Para cada observación en nuestros datos podemos calcular por medio de la siguiente ecuación:

$$\widehat{y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_i \quad i = 1, 2, \dots, n \quad (\text{A.0.2})$$

Estos son llamados valores “ajustados” dependiendo de la traducción que se le dé. De modo que el  $i$ -ésimo valor adecuado,  $i$ , es un punto en la línea de regresión de los mínimos cuadrados de la ecuación (A.0.1) correspondiente al punto  $x_i$ . La distancia vertical correspondiente a la  $i$ -ésima observación es:

$$e_i = y_i - \widehat{y}_i \quad i = 1, 2, \dots, n \quad (\text{A.0.3})$$

Estas distancias verticales son llamadas residuos mínimos cuadrados ordinarios. Una de las propiedades de los residuos es que la suma es cero. Eso significa que la suma de las distancias sobre la línea es igual a la suma de las distancias por la parte debajo de la línea.

## PRUEBA DE HIPÓTESIS

Como se dijo anteriormente, la utilidad de  $X$  como predictor de  $Y$  puede ser medida de manera informal mediante un análisis del coeficiente de correlación y el correspondiente diagrama de dispersión de  $Y$  contra  $X$ . Una manera más formal de medir la utilidad de  $X$  como predictor de  $Y$  es realizar una prueba de hipótesis sobre el parámetro de regresión  $\beta_1$ . Notemos que la hipótesis  $\beta_1 = 0$  significa que no hay una relación lineal entre  $Y$  y  $X$ . La prueba de la hipótesis requiere que asumamos lo siguiente:

Para cada valor fijo de  $X$ , se supone que los errores  $\widehat{\epsilon}$  son cantidades aleatorias independientes normalmente distribuidas con media cero y varianza común  $\sigma^2$ . Con estas suposiciones,  $\widehat{\beta}_0$  y  $\widehat{\beta}_1$  son estimaciones imparciales de  $\beta_0$  y  $\beta_1$ , respectivamente. (Se dice que  $\widehat{\theta}$  es una estimación imparcial de un parámetro  $\theta$  si el valor esperado de  $\widehat{\theta}$  es igual a  $\theta$ ). Y sus varianzas están dadas por las ecuaciones siguientes:

$$\text{Var}(\widehat{\beta}_0) = \sigma^2 \left[ \frac{1}{n} + \frac{\bar{x}^2}{\sum(x_i - \bar{x})^2} \right] \quad (\text{A.0.4})$$

y

$$\text{Var}(\widehat{\beta}_1) = \frac{\sigma^2}{\sum(x_i - \bar{x})^2} \quad (\text{A.0.5})$$

Cabe destacar que las distribuciones muestrales correspondientes a los estimados de mínimos cuadrados  $\widehat{\beta}_0$  y  $\widehat{\beta}_1$  se ajustan a una distribución normal, cuyas medias son representadas por  $\beta_0$  y  $\beta_1$ , y cuyas varianzas están dadas por las expresiones (A.0.4) y (A.0.5), respectivamente.

Las varianzas de  $\widehat{\beta}_0$  y  $\widehat{\beta}_1$  dependen del parámetro desconocido denotado como  $\sigma^2$ , por lo que necesitamos estimar el valor de  $\sigma^2$  de los datos. Una estimación imparcial de sigma está dada por:

$$\sigma^2 = \frac{\sum e_i^2}{n-2} = \frac{\sum (y_i - \widehat{y}_i)^2}{n-2} = \frac{SSE}{n-2} \quad (\text{A.0.6})$$

Donde *SSE* es la suma de los errores cuadrados residuales por sus siglas en inglés. El número  $n - 2$  en el denominador de (A.0.6) también es llamado el grado de libertad (*df* por sus siglas en inglés). Que es igual al número de observaciones menos el número de los coeficientes de regresión estimados.

Remplazando  $\sigma^2$  en la ecuación (A.0.4) y (A.0.5) por  $\sigma^2$  en la ecuación (A.0.6) obtenemos las estimaciones imparciales de las varianzas  $\widehat{\beta}_0$  y  $\widehat{\beta}_1$ . La estimación de la desviación estándar se llama error estándar (*S.E*) de la estimación. Entonces los errores estándar de  $\widehat{\beta}_0$  y  $\widehat{\beta}_1$  serán:

$$S \cdot E(\widehat{\beta}_0) = \sigma \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum(x_i - \bar{x})^2}} \quad (\text{A.0.7})$$

$$S \cdot E(\widehat{\beta}_1) = \frac{\sigma}{\sqrt{\sum(x_i - \bar{x})^2}} \quad (\text{A.0.8})$$

Respectivamente donde  $\sigma$  es la raíz cuadrada de  $\sigma^2$  en (A.0.6). El error estándar de  $\widehat{\beta}_1$  es una medida precisa de como la pendiente ha sido estimada. Mientras menor sea el error estándar diremos que la estimación es más precisa. Con las distribuciones de muestreo de  $\widehat{\beta}_0$  y  $\widehat{\beta}_1$  estamos entonces en la posición de llevar a cabo un análisis estadístico sobre la utilidad de  $X$  como predictor de  $Y$ . Bajo el supuesto de la normalidad, una prueba apropiada para probar la hipótesis nula de  $H_0 : \beta_1 = 0$  contra la alternativa de  $H_1 : \beta_1 \neq 0$  es el *t*-test.

$$t_1 = \frac{\widehat{\beta}_1}{S \cdot E(\widehat{\beta}_1)} \quad (\text{A.0.9})$$

La estadística  $t_1$  esta distribuida como  $t$  de Student con  $n - 2$  grados de libertad. La prueba se lleva a cabo mediante una comparación de observaciones de valores críticos apropiados. Sea  $t_{(n-2, \frac{\alpha}{2})}$  donde  $\alpha$  es un nivel de significancia específico. Se divide  $\alpha$  por 2 ya que tenemos dos casos de hipótesis alternativas, de acuerdo con esto si  $H_0$  será rechazada por el nivel de significancia  $\alpha$  sí:

$$|t_1| \geq t_{(n-2, \frac{\alpha}{2})} \quad (\text{A.0.10})$$

Donde  $|t_1|$  denota el valor absoluto de  $t_1$ . Un criterio equivalente al anterior es comparar el  $p$ -value y  $t$ -test con  $\alpha$  donde  $H_0$  será rechazado sí:

$$p(|t_1|) \leq \alpha \quad (\text{A.0.11})$$

Donde  $p(|t_1|)$  llamado también  $p$ -value es la probabilidad de que una variable aleatoria tiene una distribución  $t$  de Student con  $n - 2$  grados de libertad sea mayor que  $|t_1|$  (el valor absoluto de la observación  $t$ -test). El  $p$ -value es la suma de los dos lados bajo la curva, el  $p$ -value es usualmente calculado y aplicado bajo los softwares estadísticos como resultado de un análisis de regresión lineal.

Por otra parte, al rechazar que  $H_0 : \beta_1 = 0$  significa que  $\beta_1$  es diferente de cero, por lo que la variable predictora  $X$  es estadísticamente significativa de la respuesta de  $Y$ . Para completar todas las pruebas de las hipótesis de los parámetros de regresión, es necesario que se prueben otros casos que pueden aplicarse en la práctica.

Para ello probaremos las otras hipótesis:

$$H_0 : \beta_1 = \beta_1^0 \quad (\text{A.0.12})$$

El  $t$ -test puede ser generalizado a una prueba más general con la hipótesis siguiente  $H_0 : \beta_1 = \beta_1^0$ , donde  $\beta_1^0$  es una constante elegida por el investigador, contra la alternativa bilateral de la hipótesis  $H_1 : \beta_1 \neq \beta_1^0$ . La prueba estadística apropiada en este caso es la prueba de  $t$  donde:

$$t_1 = \frac{\widehat{\beta}_1 - \beta_1^0}{s \cdot e(\widehat{\beta}_1)} \quad (\text{A.0.13})$$

Notemos que cuando  $\beta_1^0 = 0$ , entonces el  $t$ -test de la ecuación (A.0.12), se reduce a la prueba de la ecuación (A.0.9). La estadística de  $t_1$  en la ecuación (A.0.12) también está distribuida como  $t$  de Student con  $n - 2$  grados de libertad.

Por otro lado, con la hipótesis de que:

$$H_0 : \beta_0 = \beta_0^0 \quad (\text{A.0.14})$$

Existe la necesidad de probar esta hipótesis para el parámetro  $\beta_0$  ya que puede presentarse en la práctica. Más específicamente, suponiendo que queremos probar que;  $H_0 : \beta_0 = \beta_0^0$  contra la alternativa hipótesis que  $H_1 : \beta_0 \neq \beta_0^0$  donde  $\beta_0^0$  es una constante escogida por el investigador.

La prueba más apropiada para este caso está dada de la siguiente manera:

$$t_0 = \frac{(\hat{\beta}_0 - \beta_0^0)}{(S.E(\hat{\beta}_0))} \quad (\text{A.0.15})$$

Si  $\beta_0^0 = 0$ , un caso especial de este es obtener lo siguiente:

$$t_0 = \frac{\hat{\beta}_0}{S \cdot E(\hat{\beta}_0)} \quad (\text{A.0.16})$$

El cual prueba que  $H_0 : \beta_0 = 0$  contra la alternativa  $H_1 : \beta_0 \neq 0$

Los mínimos cuadrados estimados de los coeficientes de regresión, los errores estándar, la *t-test* para la prueba que corresponden al coeficiente es cero, y el *p-value* están comúnmente dados como resultados de los paquetes de análisis de regresión. Estos valores son usualmente dados en forma de una tabla, la cual es llamada tabla de coeficientes.

## INTERVALOS DE CONFIANZA

Para construir intervalos de confianza para los parámetros de regresión, también necesitamos asumir que las  $\varepsilon$  tienen distribución normal, lo cual nos permite concluir que en las distribuciones muestrales de  $\hat{\beta}_0$  y  $\hat{\beta}_1$  son normales. Consecuentemente, sea  $(1 - \frac{\alpha}{2}) \times 100\%$  intervalo de confianza para  $\beta_0$  dado de la siguiente manera:

$$\hat{\beta}_0 \pm t_{(n-2, \frac{\alpha}{2})} s.e(\hat{\beta}_0) \quad (\text{A.0.17})$$

Donde  $t_{(n-2, \frac{\alpha}{2})}$  representa el  $(1 - \frac{\alpha}{2})$  percentil de una distribución de *t* con  $n - 2$  grados de libertad. De manera similar, los límites del  $(1 - \frac{\alpha}{2}) \times 100\%$  intervalo de confianza para  $\beta_1$  están dados por:

$$\hat{\beta}_1 \pm t_{(n-2, \frac{\alpha}{2})} s.e(\hat{\beta}_1) \quad (\text{A.0.18})$$

El intervalo de confianza en (A.0.16) tiene una interpretación habitual, si tenemos que tomar muestras repetidas del mismo tamaño para los mismos valores de  $X$  por ejemplo; 95 % del intervalo de confianza para cada parámetro de la pendiente para cada muestra, entonces el 95 % de estos intervalos puede contener el verdadero valor de la pendiente.

## PREDICCIONES

La ecuación de regresión propiamente ajustada puede ser usada para predicciones. Distinguiremos estas entre dos tipos:

1. La predicción de que uno de los valores responde a la variable  $Y$ , el cual corresponde a cualquier valor elegido,  $x_0$  de la variable predictora.
2. La estimación de la respuesta media es  $\mu_0$  cuando  $X = x_0$

Para el primer caso, el valor predecible  $\hat{y}_0$  es:

$$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0 \quad (\text{A.0.19})$$

El error estándar de esta predicción será:

$$S.E(\hat{y}_0) = \hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}} \quad (\text{A.0.20})$$

Los límites de confianza para el valor predecible con coeficiente de confianza  $(1 - \alpha)$  están dados por:

$$\hat{y}_0 \pm t_{(n-2, \frac{\alpha}{2})} s.e(\hat{\mu}_0) \quad (\text{A.0.21})$$

Para el segundo caso, la respuesta media de  $\mu_0$  está estimada por:

$$\hat{\mu}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0 \quad (\text{A.0.22})$$

El error estándar de esta estimación es:

$$S \cdot E(\mu_0) = \hat{\sigma} \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}} \quad (\text{A.0.23})$$

Del cual sigue los límites de confianza para  $\mu_0$  con un coeficiente de confianza  $(1 - \alpha)$  dados por:

$$\mu_0 \pm t_{(n-2, \frac{\alpha}{2})} S \cdot E(\hat{\mu}_0) \quad (\text{A.0.24})$$

Cabe señalar que el valor estimado de  $\mu_0$  es idéntico a la respuesta predictora  $\hat{y}_0$ . Esta relación puede apreciarse al comparar las ecuaciones (A.0.19) y (A.0.22). No obstante, el error estándar de  $\hat{\mu}_0$  es inferior al error estándar de  $\hat{y}_0$ , lo cual puede constatar al cotejar las ecuaciones (A.0.20) y (A.0.23). Desde un punto de vista intuitivo, esto resulta coherente, dado que predecir una observación individual conlleva una mayor variabilidad que estimar la respuesta media cuando  $X = x_0$ . El promedio subyacente en la respuesta media disminuye la variabilidad y la incertidumbre asociadas con la estimación. En cuanto a los límites establecidos en las ecuaciones (A.0.21) y (A.0.24), es importante diferenciarlos. Los límites en (A.0.21) son a veces denominados límites de predicción o límites de pronóstico, mientras que los límites en (A.0.24) son conocidos como límites de confianza.

# Bibliografía

- [1] World Data Bank. <https://data.worldbank.org>, 2018.
- [2] Jan-Justus Andreas, Charlotte Burns, and Julia Touza. Renewable energy as a luxury? A qualitative comparative analysis of the role of the economy in the EU's renewable energy transitions during the 'double crisis'. *Ecological Economics*, 142:81–90, 2017.
- [3] Autoridad General de Estadísticas de Arabia Saudita. General Authority for Statistics (Arabia Saudita). <https://www.stats.gov.sa>, 2023.
- [4] Tony Avirgan, Sarah Gammage, and Josh Bivens. Good jobs, bad jobs, no jobs: Labor markets and informal work in Egypt, El Salvador, India, Russia, and South Africa. 2005.
- [5] Yasemin Bal-Gündüz. *Israel: Selected Issues and Statistical Appendix*. International Monetary Fund, 2001.
- [6] Central Bureau of Statistics. Central Bureau of Statistics (Israel). <https://www.gov.il>, 2018.
- [7] Peter Y Chen and Paula M Popovich. *Correlation: Parametric and nonparametric measures*. Number 139. Sage, 2002.
- [8] Jacob Cohen, Patricia Cohen, Stephen G West, and Leona S Aiken. *Applied multiple regression/correlation analysis for the behavioral sciences*. Routledge, 2013.
- [9] Federal Statistical Office. Federal Statistical Office (Suiza). <https://www.bfs.admin.ch>, 2019.
- [10] Federal Statistical Office of Germany. Federal Statistical Office of Germany (Alemania). <https://www.destatis.de>, 2020.
- [11] Hellenic Statistical Authority. Hellenic Statistical Authority (Grecia). <https://www.statistics.gr>, 2019.
- [12] Italian National Institute of Statistics. Italian National Institute of Statistics (Italia). <https://www.istat.it>, 2019.

- [13] Gorana Krstić. Why income inequality is so high in Serbia: Empirical evidence and a measurement of the key factors. *Economic annals*, 61(210):23–46, 2016.
- [14] Gorana Krstić. Factors contributing to income and wage inequality: Comparative evidence from Croatia, Serbia and Slovenia. *Eastern European Economics*, 59(5):423–448, 2021.
- [15] David J Lilja and Greta M Linse. *Linear regression using R: An introduction to data modeling*. University of Minnesota Libraries Publishing, 2022.
- [16] Zhou Lu, Linchuang Zhu, Chi Keung Marco Lau, Aliyu Buhari Isah, and Xiaoxian Zhu. The role of economic policy uncertainty in renewable energy-growth nexus: Evidence from the Rossi-Wang causality test. *Frontiers in Energy Research*, 9:750652, 2021.
- [17] National Bureau of Statistics of the Republic of Moldova. National Bureau of Statistics of the Republic of Moldova (Moldavia). <https://statbank.statistica.md>, 2018.
- [18] OECD. House spending (indicator). DOI: 10.1787/b5f46047-en (Accessed on February 16 2024).
- [19] Office for National Statistics. Office for National Statistics (Reino Unido). <https://www.ons.gov.uk>, 2019.
- [20] Michael R Pakko and Patricia S Pollard. Burgernomics: A big mac™ guide to purchasing power parity. *Federal Reserve Bank of St. Louis Review*, 85(November/December 2003), 2003.
- [21] Peter Saunders, Jonathan Bradshaw, and Michael Hirst. Using household expenditure to develop an income poverty line. *Social Policy & Administration*, 36(3):217–234, 2002.
- [22] Eric Seeley. Human needs and consumer economics: the implications of Maslow’s theory of motivation for consumer expenditure patterns. *The Journal of Socio-Economics*, 21(4):303–324, 1992.
- [23] Spanish Statistical Office. Spanish Statistical Office (España). <https://www.ine.es>, 2022.
- [24] Statistical Office of the Republic of Serbia . Statistical Office of the Republic of Serbia (Serbia). <https://data.stat.gov.rs>, 2019.
- [25] Statistični urad RS. Statistični urad RS (Eslovenia). <https://www.stat.si>, 2018.
- [26] Alan M Taylor and Mark P Taylor. The purchasing power parity debate. *Journal of economic perspectives*, 18(4):135–158, 2002.