



**COMILLAS**  
UNIVERSIDAD PONTIFICIA

ICAI

ICADE

CIHS

FACULTAD DE CIENCIAS EMPRESARIALES

**RELACIÓN ENTRE PREMIOS DE LA LOTERÍA DE NAVIDAD ESPAÑOLA Y  
VARIACIONES EN COMPORTAMIENTOS DEMOGRÁFICOS A NIVEL  
PROVINCIAL**

Autor: Nicolás Álvarez Escudero

5º E3 Analytics

Trabajo de Fin de Grado de Business Analytics

Tutor: D. Riccardo Ciacci

Madrid

Junio 2023

## **Resumen:**

El estudio de cómo afectan las variaciones en la riqueza en comportamientos demográficos interesa a numerosas ramas de las ciencias sociales, entre las que se encuentran la economía, la política y la sociología. Sin embargo, sus efectos son difíciles de estudiar debido a la dificultad para encontrar sucesos aleatorios aislados donde se puedan medir. En el presente estudio utilizamos los datos de la Lotería de Navidad española, debido a una serie de características singulares que la convierten en un suceso perfecto para comprender cómo afectan las variaciones repentinas en la riqueza de una población a su comportamiento. Este trabajo se inspira en la investigación realizada por Bagues y Esteve-Volart en *Politicians' Luck of the Draw: Evidence from the Spanish Christmas Lottery*, intentando aportar nuevos enfoques y resultados. Para ello se han seleccionado, las variaciones en el número de divorcios, matrimonios y nacimientos a nivel provincial como variables dependientes a estudiar, utilizando como variables independientes los premios y el gasto de cada provincia en la Lotería de Navidad, junto con otras variables económicas. A través de un análisis estadístico basado en regresiones lineales OLS, he podido constatar el importante peso de los efectos fijos de variables como las provincias en las tres variables dependientes estudiadas, además de ver la existencia de una relación positiva de limitada significancia estadística entre los premios repartidos en las provincias y la variación en el número de nacimientos en el año posterior, tendencia también observada por Cesarini en *Fortunate families? The effects of wealth on marriage and fertility*.

**Palabras clave:** *Lotería de Navidad, comportamiento demográfico, efectos fijos, Bagues, regresión lineal, mínimos cuadrados ordinarios*

## **Abstract**

The study of the impact of variations in wealth on demographic behaviour is of interest to many branches of the social sciences, including economics, politics and sociology. However, their effects are often difficult to study due to the complexity of finding random isolated events where they can be measured. In this study we use data from the Spanish Christmas Lottery, due to a series of unique characteristics that make it a perfect event to understand how income shocks on a population affect its behaviour. This paper is inspired by the research carried out by Bagues and Esteve-Volart in *Politicians' Luck of the Draw: Evidence from the Spanish Christmas Lottery*, trying to provide new approaches and results. To this end, we have selected variations in the number of divorces, marriages and births at the provincial level as dependent variables to be studied, using as independent variables the prizes and spending in each province on the Christmas Lottery and adding other economic variables. Through a statistical analysis based on OLS linear regressions, I have been able to confirm the important weight of the fixed effects of variables such as provinces in the three dependent variables studied, as well as the existence of a positive relationship of limited statistical significance between the prizes distributed in the provinces and the variation in the number of births in the following year, a trend also observed by Cesarini in *Fortunate families? The effects of wealth on marriage and fertility*.

**Keywords:** *Christmas lottery, demographic behaviour, fixed effects, Bagues, linear regression, OLS*

- 1. Introducción**
- 2. Revisión de la Literatura**
- 3. La Lotería de Navidad**
  - a. Historia
  - b. Características singulares
  - c. Características de los jugadores
- 4. Fundamentos teóricos**
- 5. Obtención de los datos**
  - a. Datos Loterías
  - b. Datos económicos
  - c. Datos de comportamiento demográfico
- 6. Tratamiento de datos**
  - a. Primer Panel
  - b. Primer Panel
  - c. Tercer panel
  - d. Cuarto panel
- 7. Análisis empírico**
  - a. Condiciones económicas y comportamientos demográficos.
  - b. Lotería de Navidad
    - i. Vinculación del premiado y la provincia
    - ii. Impacto de la Lotería en las variables económicas
    - iii. Lotería de Navidad y comportamientos demográficos
- 8. Explicación del código**
  - a. Preparación de los datos
  - b. Creaciones variables *dummies*
  - c. Definición de las variables
  - d. Creación del modelo
- 9. Conclusiones**
- 10. Bibliografía**
- 11. Anexo**

## **1. Introducción**

El presente estudio se centra en el análisis de la relación entre los premios repartidos por la Lotería de Navidad en cada provincia y las variaciones que se producen en variables de comportamiento demográfico, específicamente en el número de divorcios, matrimonios y nacimientos. Además, se considerarán diversas variables económicas con el objetivo de examinar la influencia que las variaciones repentinas en la riqueza de una población pueden tener en su comportamiento sociodemográfico. La comprensión de los factores que influyen en el comportamiento demográfico de una población es de gran importancia, ya que permite obtener conocimientos valiosos sobre las dinámicas sociales y demográficas que pueden afectar el desarrollo y el bienestar de una sociedad. En particular, el análisis de la relación entre el premio repartido en una provincia y las variables de comportamiento demográfico puede proporcionar información crucial sobre cómo los cambios económicos inesperados pueden influir en las decisiones y los patrones demográficos de una población. La relación entre la riqueza y el comportamiento demográfico es objeto de estudio en diversas disciplinas, como la economía, la sociología y la política.

El reparto de los premios de la Lotería de Navidad debido a la especial naturaleza y características de este concurso, hacen que se pueda considerar como una forma de medir de un incremento repentino en la riqueza de una población determinada, en este caso de una provincia. Pudiendo así ver si se puede entender que existe un impacto significativo en el corto plazo de este incremento, en las decisiones individuales y familiares. Los cambios en el nivel de riqueza influyen en las preferencias, las oportunidades y las restricciones que enfrentan las personas y las parejas, lo que a su vez puede afectar su comportamiento demográfico. El reto por tanto consiste en analizar si la Lotería de Navidad puede efectivamente provocar este efecto y además si puedo hacerlo en el corto plazo, que en el caso de este estudio será un año.

En este estudio, se analizarán las variaciones en el número de divorcios, matrimonios y nacimientos como indicadores clave del comportamiento demográfico. Estos eventos demográficos reflejan las decisiones y las interacciones complejas entre factores individuales, familiares, económicos y sociales. Al examinar cómo estas variables de comportamiento demográfico responden a las variaciones en el premio repartido en una provincia, se espera obtener una comprensión más profunda de las relaciones entre las variables. Además de las variables demográficas, se considerarán diversas variables económicas para poder así analizar el efecto directo de la variable de premios, capturando información sobre el contexto macroeconómico bajo el que se producen estas variaciones en las variables dependientes. Las variables económicas seleccionadas son, variación del PIB per cápita, variación de la tasa de

desempleo, variación del IPC y variación en el precio del metro cuadrado de la vivienda. Para llevar a cabo este estudio, se aplicarán modelos de regresión de mínimos cuadrados ordinarios, utilizadas para tratar datos panel, que permiten modelar la relación entre las variables independientes (premio repartido, gasto en Lotería y variables económicas) y las variables dependientes (número de divorcios, matrimonios y nacimientos) incluyendo los efectos fijos de algunas variables. Se estimarán los coeficientes de regresión y se evaluará su significancia estadística para determinar la dirección y la fuerza de las relaciones.

## **2. Revisión de la Literatura**

En este trabajo, la revisión de literatura se centrará en dos trabajos clave que han abordado el tema de la influencia de sucesos aleatorios, como lo son las loterías en decisiones o sucesos demográficos posteriores, estos trabajos son; “*¿Familias afortunadas? Los efectos de la riqueza en el matrimonio y la fecundidad*” (Cesarini, 2023) y “*La suerte de los políticos: Evidencia de la Lotería de Navidad española*” (Bagues, 2015).

El objetivo principal de esta sección es analizar críticamente las contribuciones y hallazgos de estos dos trabajos seleccionados, proporcionando así una base sólida para el desarrollo del estudio y estableciendo un marco teórico y conceptual robusto. Se realizará un análisis de las metodologías utilizadas, los resultados obtenidos y las conclusiones alcanzadas en cada uno de los trabajos examinados.

Se comenzará por analizar la metodología, resultados y fuentes del estudio realizado en Suecia “*¿Familias afortunadas? Los efectos de la riqueza en el matrimonio y la fecundidad*”. El estudio utilizó la asignación aleatoria de premios de lotería para analizar los efectos de la riqueza en los resultados familiares, como la formación y disolución del matrimonio, así como la fertilidad. La muestra de la estimación fue construida mediante el emparejamiento de tres grupos de jugadores de lotería y sus familias con registros de toda la población que incluyen información anual sobre resultados laborales y características demográficas. Se utilizó una estrategia de identificación que aprovechó el conocimiento institucional sobre la asignación de premios en cada lotería para asignar a los jugadores a grupos, donde los montos de los premios se asignaron aleatoriamente según las reglas de la lotería.

Los resultados mostraron que la riqueza de la lotería aumenta las probabilidades de matrimonio a corto y medio plazo, especialmente para los ganadores masculinos. Además, se observó una divergencia en los efectos de la riqueza en la disolución del matrimonio, teniendo las mujeres ganadoras una tasa de divorcio más alta a corto plazo, pero que se desvanece en el largo plazo. Se especula que el incremento repentino positivo de riqueza acelera la

disolución de matrimonios en proceso de ruptura. También se encontró que el riesgo de divorcio a largo plazo disminuye cuando los maridos ganan la lotería, lo cual es consistente con estudios anteriores que muestran que los ingresos o empleo del marido estabilizan los matrimonios. Además, se demuestra que los acuerdos de divorcio en Suecia tienden a favorecer al cónyuge más rico. En cuanto a la fertilidad, se observan efectos pequeños pero positivos de la riqueza en la fertilidad completada, especialmente impulsados por los ganadores masculinos y relacionados con el aumento de la formación de matrimonios para los hombres. Sin embargo, es importante tener en cuenta que los efectos positivos estimados en relación con la riqueza son probablemente el límite superior de los efectos que se esperarían en respuesta a un aumento generalizado del nivel de vida y la riqueza.

Para la realización de este estudio, los investigadores comentan que son cruciales las aportaciones previas de David Cesarini uno de los autores de la investigación ya que en su previo trabajo se ha basado la metodología de este estudio. Este investigador lleva estudiando tópicos similares como muestran sus estudios; *“Riqueza, salud y desarrollo infantil: Datos administrativos sobre jugadores suecos de lotería”* o *“El efecto de la riqueza en la oferta de trabajo individual y familiar: Datos de las loterías suecas”*.

A pesar de que la metodología empleada en el presente trabajo difiere sustancialmente de la empleada por Cesarini, las conclusiones y resultados del estudio me son de enorme utilidad a la hora de contrastar los resultados de este trabajo y analizar si las relaciones encontradas pueden respaldarse en los resultados de Cesarini.

El segundo estudio que ha servido de extrema utilidad para el trabajo es el originalmente titulado *“Politicians’ Luck of the Draw: Evidence from the Spanish Christmas Lottery”* y en español *“La suerte de los políticos: Evidencia de la Lotería de Navidad española”* realizado por Bagues y Esteve-Volart. Este estudio tiene como objetivo investigar las causas del voto económico y su impacto en el comportamiento electoral, centrándose en la Lotería de Navidad española como acontecimiento con unas características singulares para comprender la correlación entre los resultados económicos y la reelección de los políticos en ejercicio. Se abordan dos problemas comunes en la investigación empírica: la identificación de variaciones económicas independientes de las acciones de los políticos y la atribución de responsabilidad de los eventos económicos por parte de los votantes. Mediante el análisis de premios y gastos provinciales de la Lotería de Navidad durante un periodo de tiempo específico, se encuentra que las ganancias inesperadas de la lotería tienen un efecto significativo en los resultados electorales a nivel nacional, lo que respalda la existencia del voto económico. Los datos obtenidos a través de las elecciones navideñas en España revelan que, a pesar de la aleatoriedad de los premios de la Lotería de Navidad, el partido en el poder tiende a obtener

más votos en las provincias ganadoras. Cada premio de lotería se asocia con aproximadamente la mitad de un voto adicional para el partido en el poder. Además, se observa un aumento significativo en la cuota de votos del partido en el poder en la provincia ganadora media, equivalente a aproximadamente 0,63 puntos porcentuales. Estos resultados se mantienen después de realizar varias pruebas de robustez. Si bien los gobernantes de derechas parecen beneficiarse más de los premios de lotería, la diferencia no es estadísticamente significativa. Contrariamente a las explicaciones convencionales basadas en la información asimétrica o los errores de atribución, los votantes en las provincias ganadoras muestran una mayor inclinación a votar por el titular del cargo sin percibir mejoras en la situación económica o política nacional. Esto sugiere que el aumento de la riqueza puede reforzar la preferencia por el statu quo o generar mayor indulgencia hacia los gobernantes. En conjunto, este estudio destaca que la correlación entre los buenos resultados económicos y la reelección de los gobernantes puede tener un origen más complejo que la falta de información de los votantes. Para el desarrollo de este estudio los autores se refieren a el trabajo *“Estimación del efecto de los ingresos no salariales en la oferta de trabajo, los ingresos, el ahorro y el consumo: Datos procedentes de una muestra de jugadores de lotería”* además de datos electorales y de riqueza en España.

El presente estudio como ya he mencionado tiene como objetivo examinar si tal y como muestra Cesarini (2023), se puede identificar una correlación significativa entre los incrementos aleatorios en la riqueza de diferentes provincias españolas, provocados por el reparto de premios de la Lotería de Navidad, con las variaciones en comportamientos demográficos, como lo son el número de matrimonios, divorcios y nacimientos. Para llevar a cabo este estudio he utilizado como guía en la metodología, el estudio ya citado de Bagues y Esteve-Volart (2015), pero adaptándolo a los datos disponibles y a las variables dependientes seleccionadas.

Debido a la información disponible y las limitadas capacidades de investigación y análisis, este estudio se encargará de analizar si existe una correlación entre el reparto de premios de la Lotería de Navidad y variaciones a corto plazo (1 año) en los comportamientos demográficos mencionados. Se utilizarán datos desde 2014 hasta 2018 ambos años incluidos.

### **3. La Lotería de Navidad**

#### **3.1. Historia**

La Lotería de Navidad es un juego nacional de lotería llevado a cabo por el operador público de loterías y juegos de azar español, Sociedad Estatal de Loterías y Apuestas

del Estado (SELAE), el cual está adscrito al Ministerio de Hacienda. La historia de la Lotería en España se remonta al siglo XV. Sin embargo, la Lotería de Navidad, tal como se conoce hoy en día, tiene sus raíces en los sorteos navideños regionales que se celebraban en diferentes partes del país. Estos sorteos eran populares en varias localidades y solían llevarse a cabo durante las festividades de Navidad.

En 1763, el rey Carlos III creó la Lotería Real con el objetivo de recaudar fondos para los hospitales de Madrid. Este fue uno de los primeros intentos de unificar los sorteos regionales en una sola lotería nacional. Sin embargo, el sorteo de Navidad surge por primera vez en 1811 como iniciativa de Ciriaco González Carvajal, ministro del Consejo y Cámara de Indias para recaudar dinero en el contexto de la Guerra de la Independencia, como en su forma actual no se estableció hasta 1812 cuando se celebraría el 22 de diciembre. Los premios siempre los cantan niños del Colegio de San Ildefonso, esta tradición viene desde antes incluso que se creara la Lotería de Navidad, ya en 1771 los niños del colegio eran los encargados de cantar los números de los sorteos de Lotería Nacional. En concreto, la primera intervención de la historia del colegio le correspondió al niño Diego López. Según relatan desde la asociación de antiguos alumnos del Colegio.

Desde entonces su popularidad ha aumentado enormemente, consistiendo en la actualidad en una tradición más de la Navidad española que se ha celebrado todos los años superando varias crisis económicas muy graves e incluso una guerra civil, durante esos años se celebraron dos. En la actualidad es probablemente uno de los sorteos mas grandes del mundo en términos de cantidad apostada y participación, se televisa todos los 22 de diciembre y dura varias horas. Se compara con la Lotería de Año Nuevo en China, también conocida como la "Lotería de la Caja Roja" o "Lotería de la Primavera". Esta lotería se celebra durante el Festival de Primavera, que es la festividad más importante en el calendario chino. La Lotería de Año Nuevo en China es muy popular y atrae a una gran cantidad de participantes. Al igual que la Lotería de Navidad española, cuenta con premios significativos y se considera una tradición cultural importante. Durante el sorteo, se extraen una serie de números y se otorgan premios en efectivo a los poseedores de los boletos ganadores.

La Lotería de Navidad no solo es un evento que genera grandes expectativas y emoción entre los españoles, sino que también tiene un impacto significativo en la economía del país. Además de los premios ganados por los participantes, los beneficios generados por la venta de boletos se destinan a financiar proyectos y programas sociales, culturales y de bienestar.

### 3.2. Características singulares

La Lotería de Navidad española tiene dos peculiaridades que hacen de ella un fenómeno singular para analizar cambios en el comportamiento de la sociedad ante variaciones significativas en la riqueza.

La primera característica singular de esta lotería es el hecho de que es juego tremendamente sindicado. Tal y como explican Bagues y Esteve-Volart (2015) esta situación comenzó a darse cuando en 1862 se reformó la Lotería aumentando los precios de los cupones. Esto se hizo como medida social para evitar que las personas más pobres que hubiesen caído en la adicción de apostar dejarán de hacerlo ya que no podían comprar por su cuenta un solo cupón. Esto no evitó que los más pobres siguieran apostando, sino que dio pie a que se empezaran a agrupar en grupos para comprar un cupón.

Los cupones tienen 5 dígitos, el número de cupones ha ido aumentando a lo largo de los años. Hasta 2004 el número máximo que se podía jugar era el 59.999, habiendo por lo tanto 66.000 números distintos. En 2005 se reformó aumentando el número hasta 85.000 y en la lotería de 2022 se emitieron hasta 100.000. Cada número tiene un número determinado de series, que va variando con los años, por ejemplo, en 2021 fueron 172 series por número y en 2022, 180 series por número. Cada serie contiene 10 décimos cuyo valor en la última lotería de Navidad fue de 20€ por décimo. Los décimos pueden a la vez dividirse en participaciones pudiendo comprar un porcentaje proporcional de este, que suele costar entre 2 y 5 euros. Si tenemos en cuenta todo esto es evidente la cantidad de personas que se ven afectadas por la Lotería de Navidad, ya que un mismo número puede tocar a entre 1.800 y 18.000 personas.

La cifra total recaudada varía dependiendo del número de números y series emitidas, situándose en la Lotería de Navidad de 2022 en 3.600.000.000 euros, de este total se reparten en premios el 70%, 2.520.000.000 euros en 27.547.200 premios. El primer premio, llamado “el Gordo” consiste en 4.000.000 de euros repartidos entre los propietarios de un solo número, por cada euro apostado se pueden ganar 20.000 euros. Para el segundo son 6.250 euros por euro apostado, para el tercero 2.500 euros y para el cuarto 1.000 euros.

El segundo aspecto significativo de la Lotería de Navidad es la concentración geográfica de sus premios. Esto sucede porque la misma tienda suele vender todos los cupones del mismo número, tal y calcularon Bagues y Esteve-Volart (2015), 80% de los décimos de los números ganadores habían sido vendidos por una tienda cada año. Este hecho tiene sus razones, y es que al haberse convertido en una tradición

normalmente familiares, compañeros de trabajo o amigos quieren compartir el mismo número, además hace la distribución más sencilla y los ganadores se concentran siendo más reconocibles.

Tal y como expresan Bagues y Esteve-Volart (2015) hay dos fenómenos que suponen una excepción a estas características, una tienda en el pueblo de Sort de Lleida que recibe compradores de todo el territorio nacional, esta tienda ha llegado a facturar más de 100.000.000 vendiendo décimos de lotería de Navidad, o cual representa un porcentaje medianamente significativo. La venta de cupones por internet también acaba con estas características, sin embargo, el porcentaje sigue siendo muy pequeño, situándose en torno al 0.3% según el Anuario de juego de España de 2021.

### 3.3. Características de los jugadores

Otro aspecto importante a tener en cuenta a la hora de realizar este estudio son las características del jugador típico de la Lotería de Navidad, tal y como muestra el Anuario de juego de España de 2021 “*en 2020 el 70,5% de los residentes en España entre 18 y 75 años compró lotería de Navidad, lo que equivale a 24,7 millones de personas*”. La participación tan alta tiene varias razones, entre ellas el hecho de que al ser un evento tan social en el que la mayoría de las personas de un grupo participa, hace que las personas tengan miedo a ser los únicos dentro de un grupo social, ya sea trabajo, amigos o familiar, que por no haber participado no ganen si toca el premio. Según Bagues y Esteve-Volart (2015) el 87% de los jugadores de la lotería participan de forma sindicada.

La participación por provincia no depende tampoco de la riqueza, como se puede ver en la Figura 1 del anexo donde se representa la dispersión de los datos que relaciona la media del PIB per cápita del periodo seleccionado con la media de gasto por habitante en la lotería de navidad para cada provincia.

Se puede ver como aparecen provincias con PIB per cápita bajos con un gasto muy superior a otras provincias con PIB per cápita muy superiores, y viceversa. En la representación se puede ver también un *outlier*, con un gasto por habitante muy superior al resto ese dato corresponde a Lérida que como he comentado supone una excepción en las características de esta Lotería ya que en el pueblo de Sort (Lérida) compran Lotería de Navidad jugadores de toda España.

#### 4. Fundamentos teóricos

Para el desarrollo del análisis empírico de este trabajo ha sido necesaria la comprensión y asimilación de una serie de conceptos que considero importante exponer en este apartado.

El primer concepto a tener claro es el de regresión lineal de mínimos cuadrados o OLS, este es un método estadístico que busca modelar la relación entre una variable dependiente y una o más variables independientes mediante una ecuación lineal. El objetivo es encontrar los coeficientes de la ecuación que minimicen la suma de los cuadrados de las diferencias entre los valores observados y los valores predichos por el modelo.

El proceso técnico de regresión lineal de mínimos cuadrados implica calcular los coeficientes mediante métodos matemáticos, como la matriz inversa y la optimización numérica. Estos coeficientes representan la pendiente y la intersección de la línea de regresión que mejor se ajusta a los datos observados. Actualmente contamos con múltiples herramientas entre las que se encuentra Python que ya tienen integradas estas funciones en algunas de sus librerías. La regresión es muy útil a la hora de identificar relaciones entre variables ya que gracias a la regresión lineal permite analizar y cuantificar la relación entre las variables independientes y la variable dependiente. Ayuda a determinar la dirección (positiva o negativa) y la fuerza, cuanto mayor sea el valor del coeficiente en valor absoluto, mayor es la fuerza de la relación, lo que proporciona información sobre cómo los cambios en las variables independientes afectan a la variable dependiente. También es clave para el control de variables, ya que permite incluir múltiples variables independientes en el modelo, se puede controlar el efecto de otras variables en el resultado de interés. Esto permite evaluar el efecto específico de cada variable independiente en la variable dependiente, manteniendo constantes las demás variables. Inferencia estadística: La regresión lineal proporciona herramientas para realizar pruebas estadísticas sobre los coeficientes estimados. Esto permite evaluar si la relación observada entre las variables es estadísticamente significativa y proporciona información sobre la confiabilidad de los resultados.

El segundo de ellos es el concepto de *panel data* o datos panel, estos son una forma especial de datos en el campo de la econometría y la estadística. Los datos en forma de panel están estructurados en múltiples observaciones a lo largo del tiempo para un conjunto de entidades individuales. En ellos existen normalmente dos dimensiones; una temporal, que como el propio nombre indica representa diferentes puntos en el tiempo, en este caso es anual. Y otra dimensión de entidad que representa las diferentes entidades individuales observadas, en este caso son las provincias de España. Los datos en forma

de *panel data* tienen algunas peculiaridades que es importante conocer antes de realizar su análisis. A continuación voy a enumerar algunas de ellas:

- a. Efectos fijos: los efectos fijos son aquellas características únicas de cada entidad que no varían a lo largo del tiempo, los efectos fijos reflejan las diferencias constantes entre las entidades que no pueden ser explicadas por las variables independientes incluidas en el modelo. Cuando se analizan datos de panel, los efectos fijos pueden generar variaciones sistemáticas en los datos que no están relacionadas con las variables independientes consideradas en el modelo. Por lo tanto, al modelar datos de panel, es conveniente incluir los efectos fijos en el análisis para controlar su influencia. Esto se puede lograr mediante la inclusión de variables ficticias o dummies para cada entidad en el modelo. Estas variables capturan los efectos fijos específicos de cada entidad y permiten estimar los coeficientes asociados a ellos.
- b. Efectos aleatorios: estas son características de los datos que sí varían a lo largo del tiempo y entre diferentes entidades, estos datos si que pueden analizarse estadísticamente.
- c. Autocorrelación: se da porqué los datos de panel tienen una dimensión temporal, y por ello existe la posibilidad de que las observaciones consecutivas estén correlacionadas entre sí.
- d. Heterogeneidad: Los datos de panel pueden mostrar heterogeneidad tanto en la dimensión temporal como en la de entidad. Esto implica que las características y comportamientos de las entidades individuales pueden ser diferentes, y es importante tener en cuenta esta heterogeneidad al realizar el análisis.

Bagues y Esteve-Volart (2015) introducen en las regresiones lineales de su estudio los efectos fijos de la variable Año en todas sus regresiones y el de la variable Provincia en una de ellas. Como ya he mencionado supra, incluir los efectos fijos de una variable dentro de una regresión lineal de mínimos cuadrados implica incorporar variables dummy en el modelo para capturar las diferencias fijas o constantes entre las categorías o grupos de la variable en cuestión. Cuando se trabaja con datos que involucran diferentes categorías o grupos, los efectos fijos ayudan a tener en cuenta las diferencias sistemáticas

entre esas categorías. Estas diferencias pueden ser características específicas de cada categoría que no cambian a lo largo de toda la base de datos.

En el contexto de la regresión lineal de mínimos cuadrados, se utilizan variables ficticias para representar cada categoría de la variable con efectos fijos. Estas variables toman el valor de 1 si la observación pertenece a una categoría específica y 0 en caso contrario. De esta manera, se modela el efecto fijo promedio asociado con cada categoría. Al incluir los efectos fijos en el modelo, se controlan las diferencias entre las categorías en términos de su impacto constante en la variable dependiente. Esto permite estimar los coeficientes de las variables independientes y evaluar su significancia sin verse afectados por los efectos fijos.

El siguiente concepto a comentar, es la clusterización de los errores estándar por provincias. Clusterizar el error estándar en función de una variable en el contexto de una regresión lineal de mínimos cuadrados implica agrupar las observaciones en grupos basados en una variable específica y ajustar el error estándar del modelo teniendo en cuenta la dependencia o correlación dentro de cada clúster. Cuando los datos presentan una estructura de clústeres o agrupamientos naturales, como por ejemplo cuando se tienen geográficos como los de nuestra base de datos, es posible que las observaciones dentro de un mismo clúster sean más similares o estén correlacionadas entre sí. En tales casos, el supuesto de independencia de las observaciones requerido por la regresión lineal clásica puede verse violado. Para tener en cuenta la correlación dentro de los clústeres, se utiliza la clusterización del error estándar. Esto implica calcular un nuevo error estándar ajustado que refleje la dependencia entre las observaciones dentro de cada clúster. La clusterización del error estándar tiene en cuenta la variabilidad tanto dentro de los clústeres como entre los clústeres, lo que permite obtener estimaciones más precisas de los coeficientes de regresión y evaluar su significancia de manera adecuada. La clusterización del error estándar se utiliza principalmente para corregir la inferencia estadística cuando las observaciones dentro de un clúster están correlacionadas. Al ajustar el error estándar, se controla la sobreestimación o subestimación de la significancia de los coeficientes debido a la dependencia entre las observaciones dentro de un clúster. Esto proporciona inferencias más precisas y confiables en presencia de clústeres o agrupamientos en los datos.

## **5. Obtención de los datos**

Contamos con datos de tres tipos, datos sobre las ventas y premios de la lotería de navidad por provincias, datos económicos a nivel provincial y datos demográficos sobre el

matrimonios, divorcios y nacimientos a nivel provincial. A continuación, resumo el proceso de obtención y creación de la base de datos y las principales variables.

### 5.1. Datos Loterías

La obtención de los datos de la Lotería de Navidad ha sido bastante compleja, tras contactar numerosas veces con la SELAE sin obtener una respuesta satisfactoria, realicé una labor de investigación consiguiendo los datos de los Anuarios del Juego en España que prepara la Universidad Carlos III de Madrid con información facilitada por la SELAE. Este hecho a limitado significativamente el alcance del presente estudio, ya que no he podido disponer de todos los datos de los que me hubiese gustado.

Los Anuarios se encuentran en formato PDF y por tanto he tenido que ir recopilando la información de diferentes años y elaborando un fichero de Excel para posteriormente poder tratarla.

En los Anuarios solo estaba disponible el desglose por provincias desde 2014 a 2019, ambos años incluidos. Sin embargo, solo he utilizado los datos desde 2014 hasta 2018, por razones que explicaré en el apartado 5. El número de años con datos disponibles de Lotería de Navidad son en gran parte los que han limitado el número de años seleccionado para realizar el estudio.

### 5.2. Datos económicos

Todos los datos económicos se han obtenido del Instituto Nacional de Estadística (INE), excepto la variación anual del precio del metro cuadrado que la he obtenido de la base de datos de Idealista, todos los datos son a nivel provincial. Las variables económicas seleccionadas son:

- a. PIB total: el Producto Interior Bruto reportado por el INE a nivel provincial. El dato corresponde al año en el que se reparte la Lotería.
- b. Población: número total de habitantes reportado por el INE.
- c. Variación PIB per cápita: los datos se han obtenido de la base de datos del INE, se ha calculado la variación como la diferencia porcentual entre el PIB per cápita del año pasado al año de la Lotería ( $t-1$ ) y el año de la Lotería ( $t$ ).

- d. Variación en el IPC: es la inflación, calculada a partir de los datos de IPC que reporte el INE. La variación se ha calculado igual que la anterior variable.
- e. Variación tasa de empleo: tasa de empleo que reporta el INE por provincias. La variación se ha calculado igual que la anterior variable.
- f. Variación precio del metro cuadrado de vivienda: datos obtenidos de la base de datos de Idealista, proporcionan directamente la variación anual del precio del metro cuadrado por provincia.

### 5.3. Datos de comportamiento demográfico

Los datos de comportamiento demográfico pertenecen también a las bases de datos del INE, sin embargo, estos han debido de ser tratados para poder usarse, ya que no nos interesaban los valores absolutos, sino las variaciones interanuales. Todo el tratamiento de datos lo he realizado mediante Excel.

Las variables seleccionadas han sido:

- a. Variación interanual de divorcios: se ha obtenido de la base de datos del INE, sumando divorcios y separaciones anuales. Para cada año la variación resulta de hacer el cálculo de la diferencia porcentual entre los datos de ese año y los del año posterior
- b. Variación interanual de matrimonios: se ha obtenido de la base de datos del INE, sumando los tres tipos de matrimonios que desglosa el INE. La variación se ha calculado igual que para los divorcios.
- c. Variación interanual de nacimientos: se ha obtenido de la base de datos del INE directamente, la variación se ha calculado como en las dos anteriores variables

## 6. Tratamiento de datos

Tras la recopilación y limpieza de los datos he realizado un análisis descriptivo de estos para comprobar su utilidad dentro del objetivo del presente trabajo. A continuación, describiré todo el razonamiento detrás de la selección de los datos y variables. Para ello me voy a basar en la Tabla 1 del Anexo, que resume las diferentes variables mencionadas anteriormente.

En la Tabla 1 se pueden distinguir cuatro paneles:

### 6.1. Primer Panel

En este panel se recogen los datos estadísticos de la Lotería de Navidad de todas las provincias. Las dos primeras filas muestran el porcentaje que representan suma total del gasto en Lotería de Navidad y los premios repartidos por esta sobre el PIB total nacional. Vemos como de media el gasto en Lotería de Navidad representa cerca de un 0,24% del PIB y los premios repartidos, alrededor del 0,17% del PIB nacional. Estos datos tienen sentido, ya que sabemos que se reparte alrededor del 70% del total recaudado, lo cual, en nuestro caso, es en torno al 72%. Las dos siguientes filas muestra un análisis provincial, podemos ver como el gasto anual por provincia se sitúa en 0,29% del PIB provincial, dato muy similar al reportado por Bagues y Esteve-Volart (2015) 0,30%. Sabemos que la media del total de premios repartidos en cada provincia supone cerca del 0,28% del PIB provincial. Este dato parece no tener mucho sentido, pero en primer lugar hay que tener en cuenta que este dato no está teniendo en cuenta la proporción que representa cada PIB provincial sobre el total nacional (a diferencia del dato de la segunda fila del panel) y además hay que fijarse en la enorme dispersión que existe entre el dato de la provincia más alto 7,26%, que corresponde a Lérida en 2016, donde se repartieron más de 874 millones de euros y los datos más bajos en torno al 0,01% del PIB provincial. Estos resultados se dan por las características ya comentadas de la Lotería de Navidad; la mayoría de los billetes con el mismo número se suelen vender en la misma geografía, por lo tanto, cuando toca en una provincia se producen grandes entradas de capital.

Las últimas dos filas, introducen información sobre la relación de gasto y premios con el número de habitantes por provincia. Una vez más sucede lo mismo que con los datos anteriores, gasto y premios por habitante tienen valores muy similares, pero hay que tener en cuenta que no se está teniendo en cuenta que es una media de los valores de todas las provincias y por lo tanto no tiene en cuenta la proporción que representa el número de habitantes de esa provincia sobre el total nacional, además de que en los premios como es lógico el dato de la desviación típica es cinco veces mayor que en los gastos (0,66% frente a 0,13%).

### 6.2. Segundo panel

El segundo panel recoge los datos de las provincias ganadoras. El criterio para designar que provincias han sido las ganadoras, ha sido seleccionar aquellas que cumplieran la siguiente condición:

$$\frac{Premio_{s,t}}{Gasto_{s,t}} > Promedio \frac{Premio_t}{Gasto_t}$$

Donde  $s$  es una provincia y  $t$  un año determinado. Serán por tanto aquellas provincias donde su ratio Premio / Gasto sea superior a la media de esa ratio de todas las provincias para ese año. Cada año el número de provincias ganadoras varía, habiendo años en los que el número es mayor porque el premio está más repartido y otros menos porque el premio se concentró en menos provincias. Con este criterio he pretendido diferenciar más claramente que provincias han sido las ganadoras, ya que, los datos sobre el reparto de los primeros premios no son demasiado claro para algunos años, y realizando esta selección basándome en un dato numérico aseguro que selecciono aquellas que verdaderamente han ganado. Se ve claramente en las dos primeras filas de este panel, como el gasto en estas provincias como porcentaje del gasto total supone un 0,19% mientras que los premios sobre los premios totales suponen el 65%.

Hay 57 observaciones que cumplen el criterio diseñado, en las que la media de gasto como porcentaje del PIB provincial es ligeramente superior al de la otra muestra (0,30%), sin embargo, las diferencias relevantes se ven en el porcentaje que representan los premios sobre el PIB provincial (0,82%) y en los premios por habitante (199 euros de media) frente a los 64 euros de premios por habitante de la media de todas las provincias.

### 6.3. Tercer panel

El tercer panel recoge las variables dependientes a analizar, que como mencioné anteriormente son las variaciones en divorcios, matrimonios, nacimientos y constitución de sociedades. Estas variaciones se han calculado de la siguiente forma:

$$\Delta Variable_t = \frac{Numero\ de\ sucesos_{t+1} - Numero\ de\ sucesos_t}{Numero\ de\ sucesos_t} * 100$$

Tal y como mencioné en la Introducción del trabajo, pretendo en este estudio ver la influencia a corto plazo de la Lotería por eso las variaciones son interanuales y están calculadas tomando como año 0 el año en el que se produce la Lotería, por ejemplo: Para la Lotería de 2014 la variación en el número de nacimientos se obtendrá tras restar al número de nacimientos de 2015, el número de nacimientos de 2014, dividir

el resultado por el número de nacimientos de 2014 y multiplicarlo por 100 para obtener el resultado en porcentaje.

Podemos ver la existencia de mayor variabilidad en las variaciones de divorcios que en las de nacimientos y matrimonios, siendo la desviación típica del 10%, el máximo de 48,4% y el mínimo de -41%.

#### 6.4. Cuarto panel

El cuarto panel recoge la variación en las variables económicas, estas variaciones a diferencia de las de las variables de matrimonios, divorcios y nacimientos, han sido calculadas con respecto al año anterior. Es decir, para 2014, he calculado el porcentaje de variación respecto de 2013 y así sucesivamente.

### 7. Análisis empírico

En esta sección el objetivo es analizar si existe alguna relación y en su caso, de que tipo entre las condiciones económicas y las variables dependientes seleccionadas. Para ello voy a seguir la metodología de Bagues y Esteve-Volart (2015) primero analizando de forma general como afectan las variables económicas seleccionadas a las tres variables, para posteriormente analizar cómo afectan las variaciones aleatorias en la riqueza que provocan el reparto de los premios de la Lotería de Navidad.

Antes de comenzar el análisis empírico es conveniente fijar el nivel de significancia que se va a aceptar a la hora de interpretar los resultados, Bagues (2015) en su estudio fija el máximo en un 10%, sin embargo, en este trabajo el nivel máximo permitido para entender que existe una relación con significancia estadística va a ser el 15%. Hay diferentes razones que justifican esta decisión.

En primer lugar, el carácter exploratorio que tiene este trabajo de fin de grado, aumentando el nivel de significancia puede permitirme descubrir relaciones iniciales en las que se pueda indagar con mayor profundidad en futuras investigaciones. La posibilidad de desechar una hipótesis nula verdadera supone menor riesgo que aceptar una hipótesis nula falsa, ya que como he mencionado este es un trabajo que propone una primera aproximación académica y se espera una posterior indagación en profundidad de las relaciones encontradas.

#### 7.1. Condiciones económicas y comportamientos demográficos.

Siguiendo la metodología citada comienzo realizando las regresiones de las variaciones en los comportamientos demográficos con las mismas variables

económicas seleccionadas por Bagues y Esteve-Volart (2015), concretamente, la variación en el PIB per cápita, la variación en la tasa de desempleo, la inflación y la variación en el precio del metro cuadrado de la vivienda. Al haber seleccionado cuatro variables dependientes realizaré cuatro regresiones una para cada variable dependiente. Las regresiones tendrán la siguiente función:

$$\Delta Variable\ demográfica_{s,t} = \alpha_t + \beta \Delta Variables\ macroeconómicas_{s,t} + \varepsilon_{s,t}$$

Donde  $\Delta Variable\ demográfica_{s,t}$  hace referencia a la variación porcentual en el valor de la variable dependiente (divorcios, matrimonios, nacimientos y constitución de sociedades), en una provincia  $s$  entre el año de la Lotería de Navidad y el año posterior.  $\beta \Delta Variables\ macroeconómicas_{s,t}$  hace referencia a la variación de las variables económicas mencionadas en una provincia  $s$  entre el año en que se produce la Lotería de Navidad y el año anterior.

Además, se incluye la constante  $\alpha_t$  para controlar los efectos fijos anuales. Estos son aquellos factores que hayan afectado simultáneamente al comportamiento demográfico de todas las provincias en un mismo año. Realizar una regresión teniendo en cuenta los efectos fijos permite controlar la variable año, evitando así las variaciones que hayan sido provocadas por la variable año, de esta forma identificamos las variaciones entre provincias durante el mismo año. Además, para todas las regresiones clusterizamos el error estándar al nivel de provincias ya que ha sido la unidad seleccionada para realizar la recopilación de datos.

En la Tabla 2 hemos recogido los resultados de las tres regresiones de mínimos cuadrados (OLS) realizadas con Python. Como podemos ver cada variable dependiente se encuentra en una columna distinta. Los resultados de estas regresiones se encuentran en la primera columna de cada una de las variables dependientes, estas son la (1), (5) y (9).

- a. Variación interanual de divorcios: en esta regresión estamos evaluando la posible correlación entre la variación del número de divorcios entre el año de la Lotería de Navidad y el año posterior y una serie de variables económicas seleccionadas. Los resultados muestran una relación negativa entre el aumento del precio del metro cuadrado de la vivienda y el número de divorcios en el año posterior, con un coeficiente de 0,20 para un nivel de confianza superior al 85%. Esto implicaría que un aumento de un 1% en el precio del metro cuadrado de la vivienda disminuiría en un 0,29% el número de divorcios en el año siguiente. Esta relación

puede tener múltiples explicaciones, una de ellas puede ser que el alza de precios haga que parejas con la intención de divorciarse, no lo hagan en el corto plazo, debido al aumento en el coste de comprar o alquilar una casa. Es fácil de entender como el divorcio empobrece a las personas que antes eran pareja al deber duplicar muchos gastos que antes se compartían, en definitiva, se eliminan muchas sinergias, entre ellas la vivienda común. No he encontrado ninguna otra relación estadísticamente significativa entre las variaciones de otras variables económicas y la variación de divorcios en el año posterior.

- b. Variación interanual de matrimonios: en esta regresión estamos evaluando la posible correlación entre la variación del número de matrimonios entre el año de la Lotería de Navidad y el año posterior, y una serie de variables económicas seleccionadas. Los resultados no muestran la existencia de ninguna relación estadística significativa, la razón de esto puede ser porque verdaderamente no exista o por la necesidad de una base de datos mayor para poder examinar estas relaciones. Algunas de las relaciones entre las variables independientes y la variación en el número de matrimonios han sido analizadas en el pasado. Este es el caso del estudio *Inflation and Marriage in Israel* de Jona Schellekens and David Gliksberg (2012). En el que analizando datos de 20 años “*muestran que la inflación tuvo un efecto negativo sustancial en la decisión de contraer matrimonio. Más concretamente, demostramos que la alta inflación influyó en el matrimonio al crear incertidumbre económica*”.
- c. Variación interanual de nacimientos: en esta regresión estamos evaluando la posible correlación entre la variación del número de nacimientos entre el año de la Lotería de Navidad y el año posterior, y una serie de variables económicas seleccionadas. Es la misma que aparece en la regresión de los divorcios y muestra una relación positiva muy débil entre el aumento del precio de la vivienda y el número de nacimientos en el año siguiente. El coeficiente es de 0,057, lo que implicaría que un aumento en el precio de la vivienda de un 1% supondría un aumento positivo de 0,057% en la variación del número de nacimientos. Al igual que en el apartado anterior existe literatura anterior que analiza el impacto de variables económicas sobre el número de nacimientos en los años posteriores, uno de estos estudios es *Does Higher Inflation Mean Lower Fertility Rate: The Case of Turkey* de Julide Yalcinkaya Koyuncu y Yüksel Okşak (2022). En el encontraron una relación negativa estadísticamente significativa entre el IPC y la

tasa de fertilidad en Turquía. Sin embargo, es verdad que este estudio analizaba datos a lo largo de más de 20 años, quizás la relación entre la variación entre el número de nacimientos y el aumento de la inflación no sea fácil de contrastar estadísticamente en tan solo un año.

## 7.2. Lotería de Navidad

En la sección anterior hemos podido constatar las relaciones existentes entre las condiciones económicas y las variables dependientes seleccionadas. Conociendo esas relaciones vamos ahora a examinar si existe alguna relación estadísticamente significativa entre los incrementos repentinos en la riqueza de la población provocados por los premios de la Lotería de Navidad y las variaciones en el año posterior del número de divorcios, matrimonios y nacimientos.

### 7.2.1. Vinculación del premiado y la provincia

El primer paso para poder realizar este análisis debería ser comprobar que los jugadores compran los billetes de la Lotería en los lugares donde residen y que por lo tanto se puede asegurar que el premio no sale de la provincia que lo recibe. Este análisis lo llevaron a cabo Bagues y Esteve-Volart (2015) siguiendo la siguiente regresión:

$$\Delta_k y_{s,t} = \alpha + \beta \text{Premios}_{s,t} + \gamma \text{Gasto}_{s,t} + \epsilon_{s,t}$$

Donde  $\Delta_k y_{s,t}$  hace referencia a la variación en la renta por domicilio en la provincia  $s$ , entre el año  $t$  y el año  $t + k$ .  $\beta \text{Premios}_{s,t}$  es el valor de todos los premios repartidos en cada provincia durante la Navidad del año  $t$ .  $\gamma \text{Gasto}_{s,t}$  se refiere al valor total de los billetes de Lotería de Navidad vendidos en la provincia  $s$ . Las estimaciones de estas regresiones las incluyen en su trabajo y muestran que los premios son retenidos en la provincia en donde se reparten, cada euro de premio de lotería implica un incremento de 88 céntimos en la renta de la provincia donde es repartido, durante el año en el que los premios se cobran (normalmente durante el año siguiente). Además, muestran que el reparto de premios de Lotería de Navidad no tiene ninguna relación con las variaciones de capital que ocurrieron el año anterior. Debido a las limitaciones de extensión y profundidad de este trabajo, entendemos que esta característica de la Lotería de Navidad y por lo tanto, sigue sucediendo en los años en los que estamos realizando el estudio.

Manteniéndose la riqueza en la provincia en donde se compraron los billetes y por tanto donde se repartió el premio.

#### 7.2.2. Impacto de la Lotería en las variables económicas

En segundo lugar, es preciso conocer que efectos tiene el shock de capital provocado por el premio de la lotería en diferentes variables económicas que van a influir también en las variables dependientes que tenemos intención de analizar. Gracias a Bagues y Esteve-Volart (2015) sabemos lo siguiente:

- En primer lugar, que alrededor del 65% de los premios son depositados en los bancos ya que estos experimentan una subida de una cantidad similar en aquellas provincias ganadoras durante hasta casi 3 años después del reparto del premio
- En cuanto a la deuda en la provincia parece ser que no hay ningún movimiento notable, esto se explica, entendiendo que parte de los ganadores utilizarán el premio para cancelar deuda existente y otros en cambio aprovecharán este incremento de capital para poder solicitar nuevos préstamos. Ambos movimientos acaban cancelándose.
- En el estudio muestran también como existe un aumento significativo (de un valor de hasta el 5% del total del premio) en el registro de nuevos coches.
- No hay un impacto significativo en el PIB de la provincia tras ganar premios de la Lotería de Navidad, tampoco parece existir una relación entre el incremento de capital provocado por la Lotería y la población, ni el IPC.
- Por último, constataron la existencia de un leve decrecimiento en la tasa de actividad en los años siguientes y un decrecimiento en la tasa de desempleo. Estos resultados no fueron nada significativos y concuerdan con lo expuesto por Imbens et al. (2001) que reportó un decrecimiento en la tasa de actividad en aquellos ganadores de cuantías superiores a 2 millones de dólares, pero no encontrando un efecto significativo en cuantías inferiores, como lo son en la Lotería de Navidad española donde el décimo mejor premiado es de 400.000 euros.

#### 7.2.3. Lotería de Navidad y comportamientos demográficos

Una vez constatado todo lo anterior podemos realizar la regresión de cada una de las variables de comportamiento demográfico seleccionadas. La función utilizada es la siguiente:

$$\Delta Variable\ demográfica_{s,t} = \alpha_t + \beta Premios_{s,t} + \gamma Gasto_{s,t} + \varepsilon_{s,t}$$

Donde  $\beta Premios_{s,t}$  hace referencia al porcentaje sobre el PIB provincial recibido por una provincia  $s$  en el año  $t$ .  $\gamma Gasto_{s,t}$  hace referencia al capital gastado en Lotería de Navidad durante el mismo periodo y en la misma provincia. Las regresiones han sido realizadas con Python, utilizando modelos de regresión lineal, incluyendo los efectos fijos de los años y clusterizando los errores estándares por provincias.

Hemos realizado diferentes regresiones para cada variable, todos los datos se recogen en la Tabla 2 del Anexo. A continuación, analizaré para cada una de las variables dependientes las relaciones significativas que aparecen en las distintas regresiones.

- a. Variación de divorcios: la segunda regresión utilizando como variable dependiente la variación en el número de divorcios (entre  $t$  y  $t+1$ ), y como independientes el valor de los premios y del gasto en Lotería de Navidad sobre el PIB provincial. Los resultados aparecen en la columna (2) de la Tabla 2. Los resultados muestran una relación con un nivel de confianza superior al 85% entre el porcentaje de premios repartidos en cada provincia sobre el PIB provincial y la variación en el número de divorcios del siguiente año. El coeficiente es 0,944, lo que implica que un aumento de un 1% en la variable de Premios/PIB se traduce en un aumento de 0,94% en el número de divorcios respecto al año anterior. Estos resultados concuerdan con la tendencia examinada por Cesarini (2023), el estudio citado, concluye que los incrementos de capital provocados por la Lotería en una muestra de hombres y mujeres aumenta en el corto plazo el número de disoluciones matrimoniales.

La siguiente regresión pretende aislar los efectos directos que tienen los premios en la variación de la variable dependiente, por eso es idéntica a la anterior regresión, pero añadiendo como variables independientes todas las variables económicas ya comentadas (Var. PIB per cápita, tasa de desempleo, Var. IPC y Var. Precio m2 vivienda). En este caso se ha repetido la relación

negativa entre variación del precio por metro cuadrado de la vivienda y el número de divorcios, esta vez con un coeficiente un poco superior  $-0,242$  y con un nivel de confianza superior al 90%. Además, la relación positiva entre Premios / PIB y divorcios ha disminuido de  $0,931$  a  $0,824$ , manteniéndose la significancia de la relación. Los coeficientes del resto de variables económicas se han mantenido en valores muy similares en esta regresión comparándola con los valores que obtuvimos en la regresión de la columna (1) y (2).

La última regresión realizada añade a la regresión anterior los efectos fijos de la variable Provincias, esto implica que se tienen en cuenta dentro de las variables independientes los cambios que llevan implícitas las observaciones por ser o no de una provincia determinada. La inclusión de estos efectos fijos aumenta enormemente el número de variables independientes ya que se suman 51 provincias (se deja siempre una de las variables fuera para evitar la multicolinealidad perfecta, esta se da cuando existe una relación lineal exacta entre dos o más variables explicativas en un modelo de regresión. En otras palabras, una variable explicativa puede ser expresada como una combinación lineal exacta de otras variables explicativas en el modelo).

Los resultados de esta regresión aparecen en la columna (4), estos no muestran ninguna relación significativa entre las variables independientes y la variación de matrimonios, la relación que vimos en la columna (3) entre la variable Premios / PIB disminuye hasta un nivel de confianza muy poco relevante. Además, otros coeficientes sufren variaciones considerables como por ejemplo la Variación de PIB per cápita que disminuye drásticamente o el de la variación del IPC que aumenta.

- b. Variación de matrimonios: la segunda regresión utilizando como variable dependiente la variación en el número de matrimonios (entre  $t$  y  $t+1$ ), y independientes el valor de los premios y del gasto en Lotería de Navidad sobre el PIB provincial. Los resultados se muestran en la columna (6) de la Tabla 2. Los resultados no muestran ninguna relación estadísticamente significativa entre Premios/PIB y la variación en el número de matrimonios del año siguiente.

Los resultados de la siguiente regresión para la variación de matrimonios aparecen en la columna (7), en ella como menciono en el apartado superior, la regresión es igual que la de la columna anterior (6) añadiendo las variables

económicas como variables independientes. Esto hace que varíen algunos coeficientes ninguno de forma considerable.

La cuarta regresión para esta variable dependiente como he indicado anteriormente introduce los efectos fijos de la variable Provincia. Los resultados se muestran en la columna (8), al igual que en las anteriores regresiones para esta variable, no aparece ninguna relación significativa entre las variables. Pero, se producen algunas variaciones relevantes en los coeficientes, cabe destacar el cambio del coeficiente de Premios / PIB de un valor negativo a un valor positivo.

- c. Variación de nacimientos: la segunda regresión utilizando como variable dependiente la variación en el número de nacimientos (entre  $t$  y  $t+1$ ), y independientes el valor de los premios y del gasto en Lotería de Navidad sobre el PIB provincial. Los resultados se muestran en la columna (10) de la Tabla 2 y como se puede ver aparece una relación con significancia estadística. Esta es una relación positiva entre el valor de Premios / PIB y la variación de nacimientos al año siguiente, con un nivel de confianza superior al 95%. El coeficiente es de 0,503, lo que implica que un aumento de un 1% en la variable Premios / PIB supondría un aumento positivo de 0,57% en la variación del número de nacimientos. Esta relación es muy interesante y coincide con una de las conclusiones del trabajo de Cesarini (2023). En este estudio los resultados muestran un aumento en la tasa de fertilidad en el corto plazo en aquellas personas que han ganado la Lotería, aumentándose este efecto en el medio y largo plazo.

La siguiente regresión ya he comentado que añade variables independientes para aislar que efectos directos que tienen los premios sobre la variable independiente y como podemos ver en la columna (11), los coeficientes se mantienen en valores muy similares, pero aparecen algunos cambios relevantes, en concreto la disminución del coeficiente de Premios / PIB de 0,503 a 0,421, con un nivel de confianza mayor al 85%. El resto de los coeficientes varían muy residualmente, las relaciones con significancia estadísticas vistas en las columnas (9) se mantienen en la regresión de la columna (11), sin embargo, aumenta el nivel de significancia del coeficiente de la Variación de los precios por metro cuadrado de viviendas hasta un 5%. La cuarta regresión para esta variable dependiente se encuentra en la columna (12), esta añade a la regresión anterior los efectos fijos de las provincias.

Los resultados de esta regresión para la variable de Nacimientos, a diferencia de los resultados de esta regresión para las otras variables, no se producen cambios en los coeficientes muy grandes, lo cual implica que los efectos de las provincias en este caso no son tan notables. Además, es muy destacable y positivo el hecho de que exista una relación con significancia estadística. Esta se da entre la variable Premios / PIB y Variación en el número de Nacimientos. El coeficiente es igual a 0,678, y tiene una significancia estadística del 15%. Como se puede ver este coeficiente ha aumentado con respecto a la regresión anterior, en la que era de 0,421.

Como comentario general para las tres variables dependientes, he podido ver como al introducir los efectos fijos de las provincias en la cuarta regresión, cuyos resultados se muestran en las columnas (4), (8) y (12), han aparecido cambios significativos en los coeficientes. Es relevante señalar que no en todas las variables la introducción de estos efectos fijos ha tenido la misma repercusión. Cabe destacar como he mencionado anteriormente que, en la regresión con la variación de nacimientos como variable dependiente, el cambio ha sido mucho menor. Esto muestra que las provincias en esta variable no suponen una variabilidad tan importante como en las otras dos. Sin embargo, sigue existiendo una variación con respecto a las regresiones anteriores, esta tiene diferentes explicaciones. Una de las más plausibles es la heterogeneidad individual de las entidades observadas. Hay que tener en cuenta que los efectos fijos de las provincias se han incluido con la intención de ver qué características vienen definidas por las entidades observadas en los datos. Lo que quiere decir esto es que cada provincia tiene unas características singulares que modifica la variable dependiente. Esta interacción con la variable independiente si no se incluyen los efectos fijos de las provincias, no era notable y por lo tanto podían verse relaciones significativas entre las variables que en verdad no venían explicadas por esa variable dependiente sino por las características de la provincia.

## **8. Explicación del código**

Como he comentado a lo largo del trabajo, el análisis de los datos se ha realizado mediante Python, concretamente he utilizado las librerías pandas y statsmodel. Pandas sirve principalmente para la lectura de datos y el tratamiento de los datos, mientras que statsmodel se ha utilizado para la realización de regresiones.

He adjuntado el código empleado en el anexo sin embargo considero útil incluir una explicación no exhaustiva de las funciones empleadas y de como he construido el código.

8.1. Preparación de los datos: primero es necesario importar las librerías mencionadas, a continuación, seleccionar los datos con los que se quiere trabajar. En mi caso los tenía en formato Excel y he tenido primero que convertirlos a CSV con la función `data.csv()`.

8.2. Creaciones variables dummies: el primer paso para poder introducir los efectos fijos de variables en un modelo es crear las variables dummies para cada entidad dentro de esas variables. Para ello el primer paso es convertir las variables cuyos efectos fijos quieres tener en cuenta en variables categóricas. Esto lo he hecho con pandas, usando la función `pd.get_dummies ( data [ ' Provincia ' ], drop_first = True )`. La parte final de la función es muy importante añadirla porque evita que se incluyan todas las entidades como variables dummies y se de multicolinealidad perfecta. Después es importante convertir estas variables a número ya que si no dará un error al trabajar con statsmodel. Para ello utilizo la función `año_dummies.astype ( int )`.

8.3. Definición de las variables: el siguiente paso es definir la Y y las X, la variable dependiente es sencilla de definir, simplemente `y = df [ ' Variación divorcios ' ]`. Para las X habrá que mencionar todas las variables que queremos introducir en el modelo que se encuentran en la base de datos que estamos utilizando, para posteriormente concatenar las variables dummies creadas anteriormente, para así tener en cuenta los efectos fijos. Por último es imprescindible añadir a la función el intercepto o constante, esto se hace con `X = sm.add_constant ( X )`.

8.4. Creación del modelo: en este caso hemos elegido trabajar con un modelo de regresión lineal de mínimos cuadrados o Ordinary Least Squares (OLS) por lo que el modelo será el siguiente, `model = sm.OLS ( y , X )`. Lo último que falta es establecer que cree clusterice el error estándar por provincias, lo cual se hace mediante el siguiente código, `model = model.fit ( cov_type = ' cluster ' , cov_kwds = { ' groups ' : df [ ' Provincia ' ] } )`.

Esta es la base del código que ha sido empleado, adaptándolo para cada tipo de regresión realizada y variables analizada. Se incluye en el Anexo este código unificado a modo de muestra. Además, en el Anexo se incluye la base de datos utilizada.

## 9. Conclusiones

Tal y como he comentado el estudio de como afectan cambios repentinos y significativos en el capital de las personas en comportamientos posteriores de la población, es un tema de enorme interés. Sin embargo, es complicado de estudiar debido a la limitada información sobre sucesos de este tipo. La Lotería de Navidad presenta una serie de características idóneas para evaluar estos impactos. En primer lugar, la trascendencia económica es muy grande, tal y como muestro en el trabajo los españoles entre 2014 y 2018 gastaron en Lotería de Navidad cerca del 0,25% del PIB nacional. Además, el juego en esta Lotería se suele hacer de forma sindicada, es decir dividiéndose los números apostados en décimos, lo que da lugar a un mayor reparto del premio es decir a que las variaciones sustanciales en la riqueza ocurren en todo un círculo social, familiar o laboral. Por último, lo que permite que este análisis se realice atendiendo a una distribución geográfica es que los billetes con el mismo número los suele vender el mismo distribuidor, lo que permite que los premios estén concentrados geográficamente, en este análisis la observación se hace a nivel de provincia, porque es el máximo desglose disponible que existe sobre la totalidad de premios, sin embargo, podría llevarse a un plano municipal e incluso por barrios dentro de una misma ciudad.

Mediante el conocimiento de estas características de la Lotería de Navidad, hemos podido utilizar información provincial acerca del gasto y los premios repartidos y analizar si estos cambios en la riqueza repentinos tenían un impacto significativo a nivel provincial en tres variables demográficas diferentes, matrimonios, divorcios y nacimientos. Para realizar este análisis he seguido la metodología propuesta por Bagues (2015), realizando alguna modificación. La primera modificación consiste en el tipo de datos seleccionados para el análisis, Bagues contaba con observaciones a lo largo de más años y decidió filtrar los datos fijándose solo en aquellas provincias en las que hubiese tocado alguno de los tres primeros en cada año de su estudio. Yo sin embargo he seleccionado todas las provincias de cada año sin importar si habían sido premiadas o no, lo he hecho así porque entiendo que aquellas que no recibieron los premios deberían presentar una variación menor en las variables seleccionadas que en aquellas que sí, además de que no disponía de un número suficiente de observaciones como para poder filtrar como Bagues. El segundo cambio metodológico consiste en el nivel de confianza mínimo que he seleccionado, Bagues lo estableció en un 90% mientras que en este estudio lo he establecido en 85%. Las razones de este cambio ya las he comentado en el trabajo, pero básicamente es un cambio que responde a la naturaleza exploratoria e inicial de esta investigación.

El análisis de los datos sin embargo se ha realizado exactamente igual que Bagues, empleando regresiones de mínimos cuadrados, que incluyeran los efectos fijos tanto de las provincias como de los años y clusterizando los errores estándar por provincias. Los resultados los he comentado en el apartado 7. de este trabajo. Las conclusiones sobre estos resultados son diferentes para cada variable dependiente analizada.

- a. Comenzaré con la variable de variación de divorcios, al seleccionar esta variable mi objetivo era analizar si podía apreciarse una relación significativa entre un incremento sustancial en la riqueza de una provincia provocada por la Lotería y la variación en el número de divorcios en el año siguiente. A medida que he avanzado en el análisis de las regresiones he podido ver como relaciones significativas que aparentemente existían entre las variables, han ido desapareciendo al incluir nuevas variables y en especial los efectos fijos de las provincias. Este es el caso de la relación entre Premios / PIB y divorcios, los resultados muestran como al añadir los efectos fijos de las provincias esta relación deja de tener significancia. Este efecto ya ha sido comentado y se explica por la fuerte influencia que tiene en la variable dependiente las características singulares de la entidad que se está observando (la provincia). Sabiendo esto el siguiente paso para continuar con esta investigación debería ser ver que variables relevantes están siendo omitidas. Es cierto que los resultados obtenidos no coinciden con las conclusiones de Cesarini (2023), que entendieron que un aumento que la Lotería catalizaba el proceso de una pareja que estuviera pensando en divorciarse, es cierto que la metodología empleada en su estudio es totalmente diferente ya que Cesarini han sido capaz de analizar los comportamientos de individuos concretos que habían recibido la Lotería.
  
- b. Si miramos los resultados obtenidos en las regresiones con la variación en el número de matrimonios como variable dependiente, vemos como no hay en ningún momento relaciones significativas entre las variables. Detrás de estos resultados puede haber diferentes causas. La primera de ellas puede ser la limitación de los datos utilizados, esto puede ser posible si tenemos en cuenta que tan solo he contado con los datos de 5 años, por lo tanto, puede que el número de datos sea insuficiente a nivel temporal y que no se esté recogiendo toda la variabilidad. Otra razón puede ser que no haya incluido en el estudio variables que influyen significativamente en la relación entre las variables estudiadas. Es cierto que el hecho de que el análisis fuese a nivel provincial me ha impedido incluir otras variables que probablemente tengan un impacto importante como pueden ser edad, nivel de estudios o ingresos. Por último,

puede ser que la relación entre la variación en el número de matrimonios y el incremento repentino en la riqueza estén relacionadas, pero de una forma más compleja a la realizada en este análisis, y por lo tanto que no sea posible modelarla mediante regresiones lineales.

- c. La última variable analizada es la variación en el aumento de los nacimientos, una vez más mi intención ha sido examinar la existencia de la relación entre los premios repartidos por la Lotería de Navidad en una provincia y la variación en los nacimientos en el próximo año en esa provincia. Creo que este análisis sí que ha brindado resultados que pueden interpretarse y servir de punto de partida para una investigación posterior. Estos muestran una significancia estadística de la relación positiva entre los Premios / PIB de cada provincia y la variación en el número de nacimientos del año posterior. El coeficiente de la variable independiente es 0,678, lo cual implica que ante un aumento de un 1% en la variable Premios / PIB, supone un incremento del 0,678% en la variación de nacimientos del año posterior. Este efecto se ha consolidado a través de una serie de comprobaciones de robustez, al añadir los efectos fijos de los años, además de incluir variables independientes económicas y por último teniendo en cuenta los efectos fijos de las provincias. Esto efectos coinciden con los resultados de Cesarini (2023), que muestran como la tasa de fertilidad de las personas que reciben un premio de la lotería aumentaba en el corto plazo y ese efecto se mantenía en el largo plazo. Es cierto que además existe extensa literatura que analiza la relación entre riqueza y las tasas de fertilidad, *Poverty and Fertility in Less Developed Countries: A Comparative Analysis* de Aassve, Arnstein et al. (2005), en este tipo de trabajos el análisis se realiza desde una perspectiva macroeconómica y se suele concluir que el aumento de riqueza y desarrollo de un país está íntimamente relacionado con la tasa de fertilidad y que a mayor desarrollo menores tasas. Sin embargo, podemos ver como en países más desarrollados como España o Suecia, la riqueza sí que supone un límite para el aumento de los nacimientos ya que cuando aumenta la riqueza aumenta la tasa de natalidad. Yo creo que una posible explicación a este suceso es que, en sociedades más desarrolladas y ricas, al poder controlar la reproducción, el ciudadano suele elegir no tener hijos por una cuestión de falta de recursos, en cuanto estos aumentan considerablemente como sucede con la Lotería, el número de nacimientos aumenta.

Finalmente considero que a pesar de que la realización de este estudio ha contado con algunas limitaciones en cuanto a la información disponible y a mis limitadas capacidades

de investigación. He podido crear un base de datos con mucho valor para futuros análisis, además de llegar a conclusiones y aportar un pequeño grano de arena para comprender como las condiciones económicas pueden afectar a comportamientos demográficos, campo de investigación que tiene especial interés porque los cambios en variables como los matrimonios, divorcios y nacimientos configuran la sociedad en la que vivimos.

## 10. Bibliografía

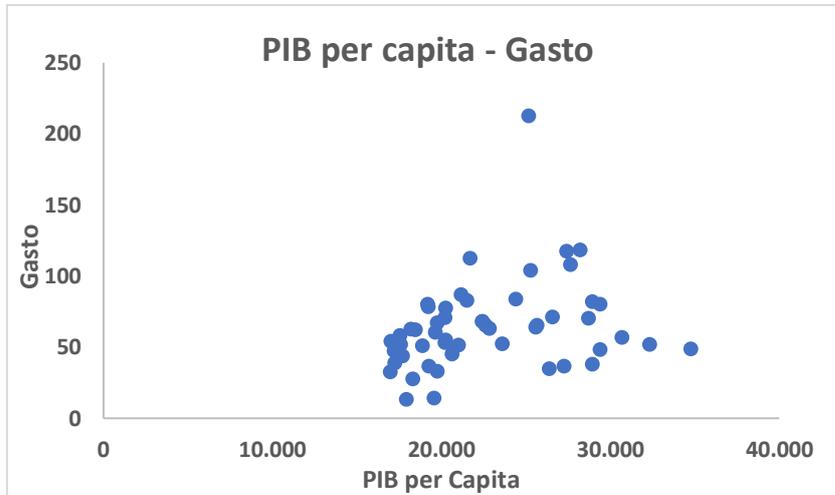
- Bagues y Esteve-Volart. 2015. “Politicians’ Luck of the Draw: Evidence from the Spanish Christmas Lottery”. Recuperado de [<https://www.bde.es/f/webpi/SES/seminars/2013/files/sie1308.pdf>]
- Cesarini. 2023. “Fortunate Families? The Effects of Wealth on Marriage and Fertility”. Recuperado de [[https://www.nber.org/system/files/working\\_papers/w31039/w31039.pdf](https://www.nber.org/system/files/working_papers/w31039/w31039.pdf)]
- Schellekens, J., y Gliksberg, D. 2012. “Inflation and Marriage in Israel”. Recuperado de [<https://journals.sagepub.com/doi/abs/10.1177/0363199012469954?journalCode=jfha>].
- Koyuncu y Okşak. 2022. “Does Higher Inflation Mean Lower Fertility Rate: The Case of Turkey”. Recuperado de [<https://dergipark.org.tr/en/download/article-file/2497073>]
- Consejo Empresarial del Juego y Universidad Carlos III de Madrid (2019). Anuario del Juego en España. Recuperado de [<https://www.azarplus.com/wp-content/uploads/2019/11/ANUARIO-JUEGO.pdf>]
- Consejo Empresarial del Juego y Universidad Carlos III de Madrid (2020). Anuario del Juego en España. Recuperado de [<https://www.infoplay.info/photos/Anuario%20del%20Juego%20en%20Espa%C3%B1a%202020.pdf>]
- Consejo Empresarial del Juego y Universidad Carlos III de Madrid (2021). Anuario del Juego en España. Recuperado de [<https://cejuego.com/wp-content/uploads/2021/11/2021-Anuario-Paginas.pdf>]

- Instituto Nacional de Estadística. (2023). Demografía y población. Recuperado de [[https://www.ine.es/dyngs/INEbase/es/categoria.htm?c=Estadistica\\_P&cid=1254734710984](https://www.ine.es/dyngs/INEbase/es/categoria.htm?c=Estadistica_P&cid=1254734710984)]
- Instituto Nacional de Estadística. (2023). Nivel y condiciones de vida. Recuperado de [[https://www.ine.es/dyngs/INEbase/es/categoria.htm?c=Estadistica\\_P&cid=1254735976604](https://www.ine.es/dyngs/INEbase/es/categoria.htm?c=Estadistica_P&cid=1254735976604)]
- Instituto Nacional de Estadística. (2023). Economía. Recuperado de [[https://www.ine.es/dyngs/INEbase/es/categoria.htm?c=Estadistica\\_P&cid=1254735570541](https://www.ine.es/dyngs/INEbase/es/categoria.htm?c=Estadistica_P&cid=1254735570541)]
- Idealista (2023). Informes de precios de venta en España de 2018, 2017, 2016, 2015 y 2014. Recuperado de [<https://www.idealista.com/sala-de-prensa/informes-precio-vivienda/venta/report/2018/>]
- Aassve, Arnstein et al. (2005). “Poverty and fertility in less developed countries: a comparative analysis, ISER Working Paper Series, No. 2005-13, University of Essex, Institute for Social and Economic Research (ISER), Colchester.
- Lindahl, Mikael. 2005. “Estimating the Effect of Income on Health and Mortality Using Lottery Prizes as an Exogenous Source of Variation in Income.” *Journal of Human Resources* 60(1): 145-68.
- Bleakley, H. and J. Ferrie (2016): “Shocking Behavior: Random Wealth in Antebellum Georgia and Human Capital Across Generations,” *Quarterly Journal of Economics*, 131, 1455–1495.
- Ross, Lee D. and Richard E. Nisbett. 1991. *The Person and the Situation*. New York: McGraw-Hill.
- Garvía, Roberto. 2007. “Syndication, Institutionalization, and Lottery Play.” *American Journal of Sociology* 113(3): 603-52.
- Casado, Julio. 2007. “Manual básico de Estadística”. Recuperado de [[https://www.ine.es/ine/oposiciones/temario\\_2021/manual\\_basico\\_estadistica.pdf](https://www.ine.es/ine/oposiciones/temario_2021/manual_basico_estadistica.pdf)]

- Imbens, Guido, Don Rubin and Bruce Sacerdote. 2001. “Estimating the Effect of Unearned Income on Labor Supply, Earnings, Savings and Consumption: Evidence from a Sample of Lottery Players.” *American Economic Review* 91(4): 778-94.

## Anexo

### Figura 1



### Tabla 1: Estadística descriptiva

Tabla Estadística Descriptiva - Información Provincial	Media	Desviación típica	Mín.	Máx.	N
1. Lotería de Navidad: Todas las provincias					
Total gasto (% del PIB nacional)	0,24%	-	-	-	-
Total premios (% del PIB nacional)	0,17%	-	-	-	-
Gasto (% del PIB provincial)	0,29%	0,13%	0,07%	0,89%	260
Premios (% del PIB provincial)	0,28%	0,66%	0,01%	7,26%	260
Gasto por habitante	64	33	12	219	260
Premio por habitante	64	163	2	2040	260
2. Lotería de Navidad - Provincias mas premiadas					
Total gasto (% del gasto Total)	19%	-	-	-	-
Total premios (% de premio Total)	65%	-	-	-	-
Gasto (% del PIB)	0,30%	0,10%	0,13%	0,55%	57
Premios (% del PIB)	0,86%	1,24%	0,08%	7,26%	57
Gasto por habitante	68	25	26	132	57
Premio por habitante	199	311	16	2040	57
3. Comportamientos demográficos					
Variación divorcios al año	-1,9%	10,1%	-41,6%	48,4%	260
Variación matrimonios al año	0%	6%	-13%	37%	260
Variación nacimientos al año	-4%	3%	-13%	8%	260
4. Variables macroeconómicas					
Población (en miles)	895	1.143	84	6.587	260
Variación PIB per capita	3%	2%	-4%	13%	260
Variación IPC	0,5%	1,0%	-1,7%	2,0%	260
Variación tasa desempleo	-10,5%	10,9%	-54,5%	22,4%	260
Variación precio m2 vivienda	-0,4%	4,4%	-17,2%	18,5%	260

**Tabla 2: Presentación de resultados**

Variables independientes	Variables dependientes											
	Δ Divorcios (%)				Δ Matrimonios (%)				Δ Nacimientos (%)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Δ PIB per cápita	0,270 (0,292)	-	0,292 (0,300)	0,0822 (0,426)	-0,130 (0,121)	-	-0,138 (0,125)	-0,170 (0,172)	-0,072 (0,097)	-	-0,052 (0,100)	-0,143 (0,126)
Δ Tasa desempleo	-0,073 (0,06)	-	-0,069 (0,064)	-0,063 (0,082)	0,050 (0,063)	-	0,048 (0,063)	0,054 (0,078)	-0,020 (0,021)	-	-0,015 (0,022)	-0,012 (0,024)
Δ IPC	2,533 (3,268)	-	2,408 (3,341)	3,092 (4,130)	-0,695 (1,157)	-	-0,654 (1,177)	-0,402 (1,437)	0,526 (0,621)	-	0,415 (0,639)	0,286 (0,889)
Δ Precio m2 vivienda	-0,205* (0,129)	-	-0,242** (0,141)	-0,323 (0,248)	0,080 (0,092)	-	0,070 (0,098)	-0,169 (0,167)	0,057* (0,036)	-	0,077*** (0,039)	-0,035 (0,076)
Premios / PIB	-	0,944*	0,824*	1,141	-	-0,220	-0,143	0,185	-	0,503***	0,421*	0,678*
Gasto / PIB	-	-2,517 (5,934)	-4,524 (6,550)	-71,263 (64,687)	-	-1,512 (2,603)	-0,793 (2,736)	16,918 (23,108)	-	0,925 (1,457)	1,491 (1,320)	-8,516 (27,275)
Efecto fijos Años	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí	Sí
Efecto fijos Provincias	No	No	No	Sí	No	No	No	Sí	No	No	No	Sí
R2 Ajustado	0,004	-0,004	0,001	-0,073	0,161	0,156	0,155	0,046	0,093	0,1	0,097	-0,009
N	260	260	260	260	260	260	260	260	260	260	260	260

Errores estándares robustos clusterizados por provincias, entre paréntesis debajo de cada coeficiente. Todas las regresiones incluyen los efectos fijos de los años. Las variables Premios/PIB y Gasto/PIB hacen referencia al total de premios repartidos en cada provincia entre el PIB provincial, habiendo tomado todas las observaciones de cada año. \*Significancia estadística al 15%; \*\*Significancia estadística al 10%; \*\*\*Significancia estadística al 5%

## Código

Calculo primeras regresiones (1), (5) y (9)

### 1. Divorcios

```
import pandas as pd
import statsmodels.api as sm
df = pd.read_csv ("data.csv")
df['Año'] = pd.Categorical(df['Año'])
año_dummies = pd.get_dummies(df['Año'], drop_first=True)
año_dummies = año_dummies.astype(int)
X = pd.concat([df[['Variación PIB per capita', 'Inflacion', 'Variación tasa desempleo', 'Variación precio m2 vivienda']], año_dummies], axis=1)
X = sm.add_constant(X)
y = df['Variación divorcios']
model = sm.OLS(y, X)
model = model.fit(cov_type='cluster', cov_kwds={'groups': df['Provincia']})
print(model.summary())
```

### 2. Matrimonios

```
import pandas as pd
import statsmodels.api as sm
df = pd.read_csv ("data.csv")
df['Año'] = pd.Categorical(df['Año'])
año_dummies = pd.get_dummies(df['Año'], drop_first=True)
año_dummies = año_dummies.astype(int)
X = pd.concat([df[['Variación PIB per capita', 'Inflacion', 'Variación tasa desempleo', 'Variación precio m2 vivienda']], año_dummies], axis=1)
X = sm.add_constant(X)
y = df['Variación matrimonios']
model = sm.OLS(y, X)
```

```

model = model.fit(cov_type='cluster', cov_kwds={'groups': df['Provincia']})
print(model.summary())

```

### 3. Nacimientos

```

import pandas as pd
import statsmodels.api as sm
df = pd.read_csv ("data.csv")
df['Año'] = pd.Categorical(df['Año'])
año_dummies = pd.get_dummies(df['Año'], drop_first=True)año_dummies =
año_dummies.astype(int)
X = pd.concat([df[['Variación PIB per capita', 'Inflacion', 'Variación tasa desempleo', 'Variación
precio m2 vivienda']], año_dummies], axis=1)
X = sm.add_constant(X)
y = df['Variación Nacimientos']
model = sm.OLS(y, X)
model = model.fit(cov_type='cluster', cov_kwds={'groups': df['Provincia']})
print(model.summary())

```

Calculo segundas regresiones (2), (6) y (10)

#### 1. Divorcios

```

import pandas as pd
import statsmodels.api as sm
df = pd.read_csv ("data.csv")
df['Año'] = pd.Categorical(df['Año'])
año_dummies = pd.get_dummies(df['Año'], drop_first=True)año_dummies =
año_dummies.astype(int)
X = pd.concat([df[['Premios/PIB', "Gasto/PIB"]], año_dummies], axis=1)
X = sm.add_constant(X)
y = df['Variación divorcios']
model = sm.OLS(y, X)
model = model.fit(cov_type='cluster', cov_kwds={'groups': df['Provincia']})
print(model.summary())

```

#### 2. Matrimonios

```

import pandas as pd
import statsmodels.api as sm
df = pd.read_csv ("data.csv")
df['Año'] = pd.Categorical(df['Año'])
año_dummies = pd.get_dummies(df['Año'], drop_first=True)año_dummies =
año_dummies.astype(int)
X = pd.concat([df[['Premios/PIB', "Gasto/PIB"]], año_dummies], axis=1)
X = sm.add_constant(X)
y = df['Variación matrimonios']
model = sm.OLS(y, X)
model = model.fit(cov_type='cluster', cov_kwds={'groups': df['Provincia']})
print(model.summary())

```

#### 3. Nacimientos

```

import pandas as pd
import statsmodels.api as sm
df = pd.read_csv ("data.csv")
df['Año'] = pd.Categorical(df['Año'])
año_dummies = pd.get_dummies(df['Año'], drop_first=True)año_dummies =
año_dummies.astype(int)
X = pd.concat([df[['Premios/PIB', "Gasto/PIB"]], año_dummies], axis=1)
X = sm.add_constant(X)
y = df['Variación Nacimientos]
model = sm.OLS(y, X)
model = model.fit(cov_type='cluster', cov_kwds={'groups': df['Provincia']})
print(model.summary())

```

Calculo terceras regresiones (3), (7) y (11)

### 1. Divorcios

```

import pandas as pd
import statsmodels.api as sm
df = pd.read_csv ("data.csv")
df['Año'] = pd.Categorical(df['Año'])
año_dummies = pd.get_dummies(df['Año'], drop_first=True)año_dummies =
año_dummies.astype(int)
X = pd.concat([df[['Premios/PIB', "Gasto/PIB", 'Variación PIB per capita', 'Inflacion', 'Variación
tasa desempleo', 'Variación precio m2 vivienda']], año_dummies], axis=1)
X = sm.add_constant(X)
y = df['Variación divorcios']
model = sm.OLS(y, X)
model = model.fit(cov_type='cluster', cov_kwds={'groups': df['Provincia']})
print(model.summary())

```

### 2. Matrimonios

```

import pandas as pd
import statsmodels.api as sm
df = pd.read_csv ("data.csv")
df['Año'] = pd.Categorical(df['Año'])
año_dummies = pd.get_dummies(df['Año'], drop_first=True)año_dummies =
año_dummies.astype(int)
X = pd.concat([df[['Premios/PIB', "Gasto/PIB", 'Variación PIB per capita', 'Inflacion', 'Variación
tasa desempleo', 'Variación precio m2 vivienda']], año_dummies], axis=1)
X = sm.add_constant(X)
y = df['Variación matrimonios]
model = sm.OLS(y, X)
model = model.fit(cov_type='cluster', cov_kwds={'groups': df['Provincia']})
print(model.summary())

```

### 3. Nacimientos

```

import pandas as pd
import statsmodels.api as sm
df = pd.read_csv ("data.csv")

```

```

df['Año'] = pd.Categorical(df['Año'])
año_dummies = pd.get_dummies(df['Año'], drop_first=True)año_dummies =
año_dummies.astype(int)
X = pd.concat([df[['Premios/PIB', "Gasto/PIB", 'Variación PIB per capita', 'Inflacion', 'Variación
tasa desempleo', 'Variación precio m2 vivienda']], año_dummies], axis=1)
X = sm.add_constant(X)
y = df['Variación Nacimientos]
model = sm.OLS(y, X)
model = model.fit(cov_type='cluster', cov_kwds={'groups': df['Provincia']})
print(model.summary())

```

Calculo cuartas regresiones (4), (8) y (12)

### 1. Divorcios

```

import pandas as pd
import statsmodels.api as sm
df = pd.read_csv ("data.csv")
df['Año'] = pd.Categorical(df['Año'])
provincias_dummies = pd.get_dummies(df['Provincia'], drop_first=True)
año_dummies = pd.get_dummies(df['Año'], drop_first=True)año_dummies =
año_dummies.astype(int)

provincias_dummies = provincias_dummies.astype(int)
X = pd.concat([df[['Premios/PIB', "Gasto/PIB", 'Variación PIB per capita', 'Inflacion', 'Variación
tasa desempleo', 'Variación precio m2 vivienda']], provincias_dummies, año_dummies], axis=1)
X = sm.add_constant(X)
y = df['Variación divorcios']
model = sm.OLS(y, X)
model = model.fit(cov_type='cluster', cov_kwds={'groups': df['Provincia']})
print(model.summary())

```

### 2. Matrimonios

```

import pandas as pd
import statsmodels.api as sm
df = pd.read_csv ("data.csv")
df['Año'] = pd.Categorical(df['Año'])
provincias_dummies = pd.get_dummies(df['Provincia'], drop_first=True)
año_dummies = pd.get_dummies(df['Año'], drop_first=True)año_dummies =
año_dummies.astype(int)

provincias_dummies = provincias_dummies.astype(int)
X = pd.concat([df[['Premios/PIB', "Gasto/PIB", 'Variación PIB per capita', 'Inflacion', 'Variación
tasa desempleo', 'Variación precio m2 vivienda']], provincias_dummies, año_dummies], axis=1)
X = sm.add_constant(X)
y = df['Variación matrimonios]
model = sm.OLS(y, X)
model = model.fit(cov_type='cluster', cov_kwds={'groups': df['Provincia']})
print(model.summary())

```

### 3. Nacimientos

```

import pandas as pd
import statsmodels.api as sm
df = pd.read_csv ("data.csv")
df['Año'] = pd.Categorical(df['Año'])
provincias_dummies = pd.get_dummies(df['Provincia'], drop_first=True)
año_dummies = pd.get_dummies(df['Año'], drop_first=True)
año_dummies = año_dummies.astype(int)

provincias_dummies = provincias_dummies.astype(int)
X = pd.concat([df[['Premios/PIB', "Gasto/PIB", 'Variación PIB per capita', 'Inflacion', 'Variación
tasa desempleo', 'Variación precio m2 vivienda']], provincias_dummies, año_dummies], axis=1)
X = sm.add_constant(X)
y = df['Variación Nacimientos']
model = sm.OLS(y, X)
model = model.fit(cov_type='cluster', cov_kwds={'groups': df['Provincia']})
print(model.summary())

```

### Imagen de la base de datos utilizada

Provincia	Premios	Comunidad	Año	Gasto	Gasto por hab	Premio	Premio por habitante	Premios / Gastos	Variación divorcios	Variación matrimonios	Variación Nacimientos
CADIZ	1	COMUNIDAD AUTONOMA DE ANDALUCIA	2014	33.023.940	26,5	94.703.350	75,9	2,9	-8%	7%	-2%
CORDOBA	2	COMUNIDAD AUTONOMA DE ANDALUCIA	2014	33.992.400	42,5	16.620.090	19,6	0,5	-5%	5%	-2%
GRANADA	2	COMUNIDAD AUTONOMA DE ANDALUCIA	2014	47.487.860	51,7	20.859.270	22,7	0,4	-1%	1%	1%
HUELVA	2	COMUNIDAD AUTONOMA DE ANDALUCIA	2014	11.847.100	22,7	5.070.940	9,7	0,4	-1%	2%	-3%
JÁEN	2	COMUNIDAD AUTONOMA DE ANDALUCIA	2014	31.045.180	47,5	14.496.720	22,2	0,5	2%	4%	-3%
MÁLAGA	2	COMUNIDAD AUTONOMA DE ANDALUCIA	2014	61.622.390	37,9	30.858.960	19,9	0,5	-1%	2%	-2%
SEVILLA	2	COMUNIDAD AUTONOMA DE ANDALUCIA	2014	61.957.720	32,0	33.652.000	17,4	0,5	-3%	4%	-3%
HUESCA	0	COMUNIDAD AUTONOMA DE ARAGON	2014	21.858.180	98,2	8.620.780	38,8	0,4	10%	-1%	8%
TERUEL	0	COMUNIDAD AUTONOMA DE ARAGON	2014	10.100.620	73,0	4.847.440	35,0	0,5	17%	-12%	-1%
ZARAGOZA	2	COMUNIDAD AUTONOMA DE ARAGON	2014	64.724.400	68,9	33.030.240	24,2	0,5	22%	6%	-3%
ASTURIAS	2	PRINCIPADO DE ASTURIAS	2014	79.668.890	75,6	29.903.940	28,4	0,4	-8%	-1%	-2%
ILLES BALEARS	2	COMUNIDAD AUTONOMA DE ILLES BALEARS	2014	34.414.780	30,7	20.410.510	18,2	0,6	0%	8%	-1%
LAS PALMAS	1	COMUNIDAD AUTONOMA DE CANARIAS	2014	29.951.440	26,8	17.536.580	15,9	0,6	-3%	4%	-3%
S. C. TENERIFE	2	COMUNIDAD AUTONOMA DE CANARIAS	2014	42.676.290	42,2	19.514.000	19,2	0,5	-8%	5%	0%
CANTABRIA	0	COMUNIDAD AUTONOMA DE CANTABRIA	2014	35.334.680	60,3	14.887.400	25,4	0,4	6%	2%	-4%
ALBACETE	1	COMUNIDAD AUTONOMA DE CASTILLA-LA MANCHA	2014	21.141.100	53,4	15.536.320	39,3	0,7	25%	3%	-3%
CIUDAD REAL	0	COMUNIDAD AUTONOMA DE CASTILLA-LA MANCHA	2014	32.981.450	63,9	12.131.620	23,5	0,4	2%	5%	2%
CUENCA	0	COMUNIDAD AUTONOMA DE CASTILLA-LA MANCHA	2014	15.754.120	75,9	16.648.440	89,9	1,2	16%	3%	-2%
GUADALAJARA	0	COMUNIDAD AUTONOMA DE CASTILLA-LA MANCHA	2014	11.356.540	44,6	5.430.380	21,6	0,5	16%	7%	0%
TOLEDO	2	COMUNIDAD AUTONOMA DE CASTILLA-LA MANCHA	2014	32.296.460	46,5	12.328.960	18,6	0,4	10%	-1%	-2%
ÁVILA	2	COMUNIDAD AUTONOMA DE CASTILLA Y LEON	2014	10.772.590	64,8	6.194.920	37,2	0,6	-6%	1%	-1%
BURGOS	2	COMUNIDAD AUTONOMA DE CASTILLA Y LEON	2014	40.323.740	110,9	17.046.160	46,9	0,4	0%	5%	-2%
LEON	1	COMUNIDAD AUTONOMA DE CASTILLA Y LEON	2014	33.153.280	69,8	25.208.820	52,3	0,8	-8%	-9%	0%
PALENCIA	2	COMUNIDAD AUTONOMA DE CASTILLA Y LEON	2014	15.323.780	92,1	4.972.590	28,9	0,3	0%	-2%	-2%
SALAMANCA	2	COMUNIDAD AUTONOMA DE CASTILLA Y LEON	2014	18.409.380	56,6	7.814.550	22,8	0,4	-4%	2%	0%
SEGOVIA	0	COMUNIDAD AUTONOMA DE CASTILLA Y LEON	2014	16.710.920	105,5	7.035.360	44,4	0,4	-1%	16%	-6%
SORIA	0	COMUNIDAD AUTONOMA DE CASTILLA Y LEON	2014	18.247.520	199,0	6.705.090	22,6	0,4	-40%	-6%	-1%
VALLADOLID	0	COMUNIDAD AUTONOMA DE CASTILLA Y LEON	2014	30.529.200	57,8	11.356.800	21,5	0,4	-7%	-2%	-3%
ZAMORA	0	COMUNIDAD AUTONOMA DE CASTILLA Y LEON	2014	13.419.380	72,4	7.687.410	41,5	0,6	-8%	1%	-6%
BARCELONA	2	COMUNIDAD AUTONOMA DE CATALUÑA	2014	257.286.580	47,4	267.817.750	49,3	1,0	-2%	3%	-2%
GERONA	2	COMUNIDAD AUTONOMA DE CATALUÑA	2014	26.746.290	36,1	9.657.590	12,9	0,4	4%	0%	2%
LLEIDA	0	COMUNIDAD AUTONOMA DE CATALUÑA	2014	56.153.760	130,2	29.147.390	67,6	0,5	-23%	-7%	-4%
TARRAGONA	0	COMUNIDAD AUTONOMA DE CATALUÑA	2014	28.713.960	36,2	12.312.020	15,5	0,4	-1%	4%	-1%
BADAJOZ	2	COMUNIDAD AUTONOMA DE EXTREMADURA	2014	22.776.880	33,1	11.522.290	16,8	0,5	-3%	4%	-4%
CACERES	1	COMUNIDAD AUTONOMA DE EXTREMADURA	2014	23.079.090	56,8	17.224.530	42,4	0,7	2%	2%	-1%
A CORUÑA	1	COMUNIDAD AUTONOMA DE GALICIA	2014	64.717.220	57,3	36.058.940	31,9	0,6	-5%	3%	2%
LUGO	1	COMUNIDAD AUTONOMA DE GALICIA	2014	19.940.520	58,6	14.312.940	42,1	0,7	3%	-1%	1%
OURENSE	0	COMUNIDAD AUTONOMA DE GALICIA	2014	12.936.940	40,4	5.162.580	16,2	0,4	2%	-2%	-1%
PONTEVEDRA	2	COMUNIDAD AUTONOMA DE GALICIA	2014	42.796.420	45,1	19.941.620	20,9	0,5	8%	3%	-5%
LA RIOJA	1	COMUNIDAD AUTONOMA DE LA RIOJA	2014	23.747.920	75,6	14.582.520	46,4	0,6	-18%	-5%	-4%