

Aplicación del modelo Heckman a las decisiones de gasto en turismo en España

by

Carlos Casado González

A thesis submitted in conformity with the requirements
for the MSc in Economics, Finance and Computer Science

University of Huelva & International University of Andalusia

uhu.es

un
i **Universidad**
Internacional
de Andalucía
A

Noviembre 2018

Aplicación del modelo Heckman a las decisiones de gasto en turismo en España

Carlos Casado González

Máster en Economía, Finanzas y Computación

Mónica Carmona Arango

Universidad de Huelva y Universidad Internacional de Andalucía

2018

Resumen.

Este trabajo examina los determinantes de las decisiones de participación y nivel de gasto en el turismo para el caso de los hogares españoles. Para ello se utilizan los microdatos de la Encuesta de Presupuesto Familiares (EPF) españoles, que elabora el Instituto Nacional de Estadística (INE), para el año 2017. Mediante un modelo Heckman se trata de evidenciar que ambas decisiones son independientes la una de la otra, y que además vienen determinadas por distintos factores, tanto económicos como sociodemográficos. En los resultados obtenidos se pone de manifiesto algunas conclusiones que ya se sabían de antemano y otras que no están tan claras en la literatura analizada para este estudio.

Palabras clave: Gasto en turismo, participación turística, modelo Heckman, hogares españoles, modelo turístico.

Abstract.

This work examines the determinants of the participation decisions and level of expenditure in tourism for the case of Spanish households (o homes). For this, the microdata from the Spanish Family Budget Survey, which elaborates the National Institute of Statistics for the year 2017, is used. Through a Heckman model, it tries to evidence that both decisions are independent from each other, and that they are also determined by different factors, both economics and sociodemographic. In the obtained results, some conclusions that were known from before and some others that weren't so clear in the analyzed literature for this study become clear.

Índice.

1. Introducción.....	4
2. Revisión bibliográfica.....	5
3. Metodología.....	7
3.1. Datos.....	7
3.2. Variables.....	10
3.3. El modelo de Heckman.....	13
4. Resultados de la estimación.....	14
5. Conclusiones.....	18
Bibliografía.....	23
Anexo.....	24

1. Introducción.

El turismo es a día de hoy una de las industrias más importantes y con mayor crecimiento en la gran mayoría de los países del mundo. El gasto de los consumidores en turismo ha aumentado constantemente y lo sigue haciendo año tras año. Según la Organización Mundial del Turismo (OMT), la expansión del sector turístico seguirá en línea ascendente, hasta el punto de que llegará a convertirse en el sector comercial más importante del mundo (OMT, 2017). Esta predisposición positiva del gasto en turismo es un estímulo para el negocio turístico en el mercado de servicios.

Investigar, por tanto, el gasto turístico es un elemento importante para las administraciones públicas y las empresas ya que podemos estudiar factores como el análisis de los retornos turísticos en un destino, al igual que se pueden establecer estrategias y políticas de marketing, para mejorar los ingresos en este campo o ser aún más competitivos.

Es por todos sabido que el sector del turismo en España juega un papel muy importante en su economía y es uno de sus pilares básicos. Nuestro país fue en el año 2017 el segundo mayor receptor de turistas en el mundo, únicamente superados por Francia, llegando a la cifra de 81,8 millones de turistas extranjeros (OMT, 2018). Según el informe anual del Consejo Mundial de Viajes y Turismo (WTTC, por sus siglas en inglés), el turismo aportó en España en este mismo año 172.900 millones a su economía, lo que supone un 14,9% del Producto Interior Bruto (WTTC, 2018).

Si nos centramos únicamente en el turismo nacional, según la Encuesta de Turismo de Residentes que elabora el Instituto Nacional de Estadística (INE), en el año 2017 los residentes en España realizaron 193,7 millones de viajes y se gastaron en ellos un total de 44.233 millones de euros. De estas cifras el 91,1% de los viajes tuvo como destino España y el 8,9% restante el extranjero (INE, 2018). Según esta encuesta en esos viajes se produjeron 802 millones de pernoctaciones y los motivos de estos viajes fueron en su gran mayoría por ocio y vacaciones o para visitar a familiares y amigos, más concretamente un 86,8%. Los viajes con destino interno acumularon el 82,4% de las pernoctaciones y el 67,7% del gasto total. Por el otro lado, los viajes al extranjero supusieron el 17,6% del gasto total.

El objetivo principal de este trabajo consiste en descubrir los determinantes del gasto en turismo de los hogares, es decir, que factores llevan a los españoles a gastar o no en turismo, y una vez tomada la decisión de gastar en turismo, cuáles son los factores que hacen que los hogares decidan gastar más o menos en esta partida. Para poder llevar a cabo este análisis utilizamos un conjunto de datos que hemos elaborado a partir de los microdatos que nos provee el INE en su Encuesta de Presupuesto Familiares (EPF). La EPF nos aporta información muy detallada acerca de todos los gastos de los hogares españoles (incluido el gasto en turismo), además de los ingresos del hogar, las características sociodemográficas, económicas y la información que tiene que ver con el trabajo de los diferentes miembros del hogar.

Para ello, este documento se estructura de la siguiente manera. Una vez introducido el tema de análisis y los objetivos, en el segundo apartado se realiza una revisión de los principales trabajos publicados acerca de las decisiones de gasto en turismo, sobre todo aquella bibliografía que realiza estudios en turismo con un modelo Heckman. En la tercera parte, se explican los datos que se han utilizado para este análisis, así como las distintas variables con las que vamos a trabajar, además de explicar de manera teórica el modelo Heckman. Con todo esto pasaremos a la cuarta parte, donde se encuentra nuestro modelo estimado y se extraen los resultados del mismo. Por último, en la quinta y última parte se extraen las conclusiones que se obtienen del modelo que hemos estimado.

2. Revisión bibliográfica.

Nos encontramos con dos ramas a la hora de estudiar los determinantes microeconómicos del gasto en turismo, por una parte, los estudios que tratan sobre el gasto turístico realizado en destinos concretos y por otra parte los que examinan el gasto en turismo como un gasto común en los presupuestos de las familias. Este trabajo se decanta por la segunda de las opciones propuestas, ya que se analiza el gasto turístico en los presupuestos de las familias españolas y no el gasto en turismo en un destino concreto.

El gasto en turismo representa una proporción creciente en el presupuesto de las familias (Melenber y Van Soest, 1996), por tanto, parece conveniente analizar el gasto turístico como una partida relevante en dichos presupuestos, ya que parece que cada vez se va haciendo más recurrente en los patrones de consumo de las familias de todo el mundo.

Los estudios realizados, bien sea recopilando información en origen directamente de los individuos o de los hogares, han utilizado encuestas específicas sobre comportamiento turístico o bien encuestas estructurales realizadas con otros fines (Mateo Erroz, 2012). Entre las primeras encontramos artículos de Mergoupis y Steuer (2003), Toivonen (2004), Eugenio-Martín y CamposSoria (2011) y, sobre todo, una que utilizaremos mucho en este estudio que es un artículo de Nicolau y Mas (2005a, b, 2009) con una encuesta específica sobre el comportamiento de los españoles ante las vacaciones realizada por el Centro de Investigaciones Sociológicas (CIS). Ahora bien, en este estudio se han utilizado datos provenientes de una encuesta estructural, mas concretamente de la Encuesta de Presupuestos Familiares de España (EPF). En esta línea la mayoría de los trabajos utilizan también encuestas de presupuestos de los hogares, como los estudios de Melenberg y Van Soest (1996) que utiliza datos provenientes de la Encuesta de Gasto en Consumo de los Países Bajos y sobre todo los estudios de Alegre y Pou (2004) y Alegre et al., (2009) con datos españoles de la Encuesta de Presupuestos Familiares. Esta última investigación representa una pieza fundamental para este trabajo puesto que trabaja con datos procedentes de la misma encuesta que este estudio.

En relación a lo que dice la literatura que trata este tema, sobre los determinantes de la participación y el nivel del gasto turístico obtenemos las siguientes conclusiones:

En primer lugar, se destaca la capacidad explicativa que tienen los ingresos corrientes. La teoría económica tiene claro que existe una relación positiva entre los ingresos corrientes y la decisión de participar (Melenberg y Van Soest, 1996; Alegre y Pou, 2004; Alegre et al., 2009; Nicolau y Mas, 2005; Eugenio-Martín y Campos-Soria, 2011), como en el nivel de gasto turístico (Melenberg y Van Soest, 1996; Nicolau y Mas, 2005).

Por otro lado, tenemos otras variables con las que existen también cierto consenso sobre su capacidad explicativa y su signo. Estas variables son el tamaño del hogar o familia, es decir el numero total de miembros, además de la tenencia de hijos o niños menores en el hogar. Para estas variables se ha estimado un efecto negativo sobre las decisiones de participación y de gasto en turismo (Melenberg y Van Soest, 1996; Alegre y Pou, 2004; Nicolau y Mas, 2005).

Con respecto al tamaño del municipio de origen, se concluye que hay una relación positiva entre este y la decisión de participar, pero que no afecta en la decisión de nivel de gasto (Alegre y Pou, 2004; Nicolau y Mas, 2005).

En educación todos los trabajos llegan a la misma conclusión, existe un efecto positivo entre el nivel educativo y la decisión de participar, además de en el gasto (Alegre y Pou, 2004; Nicolau y Mas, 2005; Eugenio-Martín y Campos-Soria, 2011).

Otra variable muy a tener en cuenta es la edad. De la literatura consultada no son muchos los estudios que tienen en cuenta la edad a excepción del artículo de Nicolau y Mas, (2005). Este estudio apunta a que la edad afecta a la decisión de cuanto gastar en turismo. Ahora bien, según el artículo la edad genera diferentes puntos de vista hacia el ocio y el consumo turístico. Por tanto, podemos esperar un efecto no lineal en la edad, que viene a decir que con las personas más jóvenes hay un aumento en el gasto en vacaciones hasta que llegan a una edad límite y comienzan a reducir su gasto.

La tenencia de hipoteca por parte de los hogares ha sido otra variable analizada, ya que una carga adicional en el hogar relacionada con la compra de la vivienda podría condicionar la posibilidad de atender a otros gastos. En la literatura consultada, el artículo de Alegre, Mateo y Pou, 2013, concluye en que no se observan diferencias entre las casas de propiedad con una hipoteca y las casas de propiedad sin una hipoteca. Esta premisa se cumple siempre que no haya un periodo de crisis, ya que según este estudio en épocas de crisis parece que esta misma afecta a los hogares con una hipoteca, no con respecto a la decisión de participación, sino en términos del nivel de gasto que se realiza. En nuestro estudio introduciremos dicha variable, aunque no nos encontremos en una época de crisis o recesión, para ver si en el año 2017 este fue un factor determinante del gasto en turismo.

Por último, tenemos la situación del consumidor o del hogar con respecto al mercado laboral que algunos estudios la incluyen únicamente como un condicionante temporal que limita el viaje vacacional (Alegre y Pou, 2004; Nicolau y Mas, 2005). Otros estudios como el de Eugenio-Martin y Campos-Soria (2011), tienen en cuenta la situación de desempleo en el entrevistado, además de la categoría profesional para aquellos que si trabajan. Sus resultados y conclusiones sugieren un impacto negativo de la situación de desempleo sobre la probabilidad de viajar, aunque bien es cierto que no es estadísticamente significativo al 10%.

3. Metodología.

3.1. Datos.

Como ya comentamos, la base de datos que vamos a utilizar es la Encuesta de Presupuestos Familiares, que elabora el Instituto Nacional de Estadística (INE). La EPF

es un estudio de carácter anual, que representa al conjunto de la población española y está diseñada para recoger de manera detallada la información de gastos e ingresos de los hogares españoles, además de información sociodemográfica y laboral relacionada con sus miembros. En cada ola se entrevista a unos 24.000 hogares aproximadamente.

El formato actual de la EPF, no es el mismo que hace unos años, si no que el utilizado ahora se lleva ejecutando desde enero de 2006. Algunos aspectos que se han cambiado a lo largo de los años son el tamaño de la muestra, el tipo de la población que se ha considerado, el diseño de los cuestionarios, el sistema de recogida de las encuestas además de la periodicidad. Gracias a esta encuesta se pueden realizar numerosos estudios desde diferentes enfoques (demográficos o socioeconómicos) y con diferentes fines, lo que la convierte en una encuesta multiobjetivo. Además, no existe, en todo el entorno de la Unión Europea, otra encuesta dirigida a los hogares que abarque un campo de observación tan amplio, lo que nos permite poder cruzar información sobre múltiples variables relacionadas con los individuos de la misma.

Además, según el INE la EPF se elabora con una serie de objetivos, los cuales se mencionan a continuación (INE, 2017):

- Obtención de estimaciones del agregado gasto de consumo anual de los hogares para el conjunto nacional y para las comunidades autónomas, así como su clasificación según diversas variables del hogar.
- Estimación del cambio interanual del agregado gasto de consumo para el conjunto nacional y para las comunidades autónomas.
- Estimación del consumo en cantidades físicas de determinados bienes alimenticios para el conjunto nacional.

Además, destacan otros dos objetivos prioritarios debido a su importancia, relacionados con necesidades concretas de diversos usuarios de la encuesta: la estimación del gasto como instrumento para la obtención del consumo privado en la Contabilidad Nacional, y la estimación de la estructura de ponderaciones a partir del gasto necesaria para el cálculo del IPC (INE, 2017).

Según el INE, el gasto en consumo final de los hogares se registra a precios de adquisición, es decir, al precio que el comprador debería pagar efectivamente por los productos en el momento de la compra y según su precio al contado. Se recoge, además

del importe real de dichos gastos en bienes o servicios, todo el gasto añadido que pudiera haber sido provocado en su compra (por ejemplo, propinas).

La clasificación que realiza la EPF de los gastos, recoge los diferentes tipos de los mismos según la nueva clasificación europea de consumo, denominada ECOICOP (European Classification of Individual Consumption by Purpose), que ofrece un mayor desglose en algunas de las parcelas de gasto y permite la comparabilidad con otras estadísticas como el Índice de Precios de Consumo (IPC).

La ECOICOP se estructura en los siguientes doce grandes grupos:

Tabla 1. Clasificación de los grupos de gasto de la EPF.

- 1. Alimentos y bebidas no alcohólicas.**
- 2. Bebidas alcohólicas y tabaco.**
- 3. Vestido y calzado.**
- 4. Vivienda, agua, electricidad, gas y otros combustibles.**
- 5. Muebles, artículos del hogar y artículos para el mantenimiento corriente del hogar.**
- 6. Sanidad.**
- 7. Transporte.**
- 8. Comunicaciones.**
- 9. Ocio y cultura.**
- 10. Enseñanza**
- 11. Restaurantes y hoteles.**
- 12. Otros bienes y servicios.**

Fuente: Elaboración propia a partir de la EPF del INE.

Para este estudio, utilizaremos un subgrupo del grupo 9 de gastos, “*Ocio cultura*”, concretamente el subgrupo de “*Paquetes turísticos*”, donde se recogen los siguientes aspectos:

- Vacaciones todo incluido o tours que prevén viaje, comida, alojamiento, guías, etc.
- Viajes de excursión de medio día o un día.
- Peregrinaciones.
- Campamentos de verano.

Además, en este subgrupo se diferencia en función de dónde se efectúan los gastos, si se realiza en gasto en España o en un país diferente.

Un factor importante a tener en cuenta es la posibilidad de analizar el comportamiento de los hogares que realizan gasto en turismo, pero también contamos con la información de aquellos hogares que no realizan ningún gasto en esta partida.

Hay que puntualizar, que para poder trabajar con los microdatos que nos ofrece el INE hay que crear una serie de diccionarios en el software que utilizaremos para este estudio, STATA15, ya que los datos son proporcionados por el INE en un fichero plano. Una vez creados dichos diccionarios se pueden extraer los datos con facilidad. En el anexo de este trabajo, se presenta el proceso para extraer y preparar los datos, con la finalidad de preparar una base de datos que permita su análisis.

3.2. Variables.

En el trabajo que aquí nos ocupa, se va a proceder a analizar las decisiones de participación y gasto turístico. Para ello se van a introducir las variables explicativas típicas de la literatura, además de información acerca de la situación laboral del sustentador principal y la tenencia o no de una hipoteca (o alquiler). Desde un punto de vista econométrico vamos a estimar un modelo Heckman. Este modelo nos va permitir separar las decisiones que nos llevan a participar o no en el gasto turístico y una vez hemos decidido participar, cuáles son los factores que nos llevan a aumentar o reducir nuestro gasto turístico.

Las variables que se van a utilizar son las que se han mencionado en el punto 2 de este trabajo, pero todas ellas toman diferentes valores según la información que aporte a nuestro modelo econométrico.

La variable dependiente será la variable “*GASTOTUR*”, la cual es una variable que contiene “missing values”, para todos aquellos hogares que no gastan nada en turismo, para aquellos que si gastan en turismo toma el valor de los gastos totales (tanto turismo nacional como internacional). La razón por la cual esta variable se compone de “missing values” es porque en STATA el modelo Heckman interpreta los “missing values” como decisiones de no participar y los “0” como decisiones de participación, por tanto, para todos aquellos hogares que no participan hay que darles un “missing value” en su variable de gasto en turismo.

La primera de las variables explicativas o independientes es la variable *“INGRESOS”*, esta variable toma como valores el logaritmo de los ingresos mensuales netos totales del hogar. Se ha tomado logaritmos para que, a la hora de interpretar dicha variable, al ser una variable continua, podamos hacerlo en términos de variación porcentuales (elasticidad).

Pasamos ahora a las variables *“EDADSP”* y *“EDADSP2”*, ambas variables recogen la edad del sustentador principal de la familia, con la única diferencia que *“EDADSP2”*, es la misma variable solo que elevada al cuadrado, esto es, para recoger el efecto no lineal que tiene la edad en la decisión de hacer turismo y la decisión de la cuantía de gasto en turismo, como se menciona en el segundo punto del trabajo.

La variable *“MUNICIPIO”*, nos da la información acerca del municipio en el que vive la familia encuestada, toma el valor “0” si el municipio cuenta con menos de 100 mil habitantes y el valor “1” si son 100 mil o mas habitantes los que residen en el mismo.

En *“HOGAR”*, recogemos el tamaño del hogar encuestado, el valor “1” es para aquellos hogares con uno o dos miembros, “2” para los hogares con tres miembros, “3”, para los que cuentan con cuatro miembros y “4” para hogares con cinco o más miembros.

La variable *“ESTUDIOS”*, toma los valores “1”, “2” y “3” según el nivel de estudios del sustentador principal, que son, estudios obligatorios (graduado escolar o menor), bachilleratos o ciclos formativos hasta grado superior y estudios universitarios como grados, másteres o doctorados, respectivamente.

“HIPOTECA”, recoge información acerca de si los hogares pagan una hipoteca o un alquiler por su casa, el valor “0” en esta variable indica que el hogar encuestado no cuenta con alquiler ni hipoteca y el valor “1” todo lo contrario.

Y para terminar con las variables explicativas tenemos la variable *“OCUPACION”*, esta variable nos indica si el sustentador principal de la familia se encontraba ocupado, es decir trabajando, o desocupado en el momento en el que se realiza la encuesta. El valor “0” indica desempleado o desocupado y el valor “1” lo contrario.

Para ver mejor y más fácilmente dichas variables explicativas y los valores que toman, toda esta información se recoge en la siguiente tabla:

Tabla 2. Variables explicativas y sus valores.

VARIABLES EXPLICATIVAS	VALORES
INGRESOS	Continua, logarítmica. Creada a partir del logaritmo de la variable “IMPEXAC”
EDADSP	Continua. Variable original de la EPF.
EDADSP2	Continua, Variable EDADSP al cuadrado.
MUNICIPIO	0 → menos de 100 mil habs. 1 → 100 mil o más habs. Creada a partir de la variable “TAMAMU”
HOGAR	1 → uno o dos miembros. 2 → tres miembros. 3 → cuatro miembros. 4 → cinco o más miembros. Creada a partir de la variable “TAMAÑO”
ESTUDIOS	1 → graduado escolar (ESO) o menor 2 → bachillerato o hasta ciclos formativos de grado superior 3 → estudios universitarios (grado, másteres, doctorado etc) Creada a partir de la variable “ESTUDREDSP”
HIPOTECA	0 → no cuenta con alquiler o hipoteca 1 → paga una hipoteca o alquiler Creada a partir de la variable “REGTEN”
OCUPACION	0 → no ocupado (desempleado) 1 → ocupado Creada a partir de la variable “OCUSP”

Fuente: Elaboración propia.

3.3. El modelo de Heckman.

El modelo de Heckman postula que las decisiones de gasto en turismo de los hogares pueden interpretarse, como un proceso de dos etapas: en primer lugar, la decisión de hacer un viaje o gastar en turismo y, en segundo lugar, cuanto gastar en él.

Sea Z_{tk} un grupo de variables k que representan las características de t individuos que determinan la decisión de hacer un viaje o irse de vacaciones, -- medida por una variable latente d_t^* -- y γ_k como los coeficientes que reflejan el efecto de estas variables en esta decisión; y siendo x_{ts} un grupo de variables s que representan las características individuales de t que determinan el nivel de gasto turístico E_t , y β_s los coeficientes que reflejan el efecto de estas variables en dicho nivel de gasto. Así el modelo de Heckman (1979) adopta la siguiente forma:

$$d_t^* = \sum_{k=1}^K \gamma_k Z_{tk} + u_t \quad (1)$$

$$E_t = \sum_{s=1}^S \beta_s X_{ts} + \varepsilon_t \quad \text{sólo si } d_t^* > 0 \quad (2)$$

donde los errores u_t y ε_t siguen una distribución normal bivalente con una media de cero, varianzas σ_u , σ_ε , respectivamente, y covarianza $\sigma_{\varepsilon u}$. Por lo tanto, definimos una variable dicotómica d_t , que toma el valor uno cuando la variable latente $d_t^* > 0$, y cero si no. De esta manera $d_t = 1$, indica la decisión de tomar vacaciones y $d_t = 0$ la de optar por no tomarse vacaciones. Esta interpretación hace que la ecuación (1) represente un modelo binomial.

Considerando la función de densidad conjunta de u_t y ε_t , la estimación de los parámetros del sistema anterior de ecuaciones por el método de máxima verosimilitud es inmediata, sin embargo, la maximización de la función de probabilidad es compleja debido a problemas de convergencia (Greene, 1999; Sigelman y Zeng, 1999), lo que nos lleva utilizar el método de estimación de dos etapas propuesto por Heckman (1979), que nos permite obtener datos consistentes en las estimaciones de los parámetros. Este método comienza con la expresión de la expectativa condicional de y , de modo que:

$$E(E_t | d^* > 0) = x\beta + \sigma_{\varepsilon u} \sigma_\varepsilon \lambda(-\gamma z) \quad (3)$$

donde λ es la relación inversa de Mills, definida como $\lambda(-\gamma z) = \phi(-\gamma z)/(1 - \Phi(-\gamma z))$; β y γ son los vectores de parámetros que miden el efecto de las variables x y z mencionadas anteriormente; ϕ y Φ son las funciones de densidad y distribución de una normal, respectivamente. La expresión condicional esperada muestra que ambas decisiones (irse de vacaciones y nivel de gasto), están relacionadas; en otras palabras, la expectativa de E_t es igual a $x\beta$ solo cuando los errores u_t y ε_t no están correlacionados, $\sigma_{\varepsilon u} = 0$; de lo contrario, la expresión esperada de E_t se ve afectada por las variables de la ecuación (1). Por tanto, la importancia del parámetro $\sigma_{\varepsilon u}$ ratifica la superioridad de este modelo sobre otros. De la ecuación (3), encontramos que:

$$E_t | d^* > 0 = E(E_t | d^* > 0) + v_t = x\beta + \sigma_{\varepsilon u} \sigma_{\varepsilon} \lambda(-\gamma z) + v_t \quad (4)$$

donde v_t es el termino de error que se distribuye $N(0, \sigma_{\varepsilon}(\lambda(\lambda - \gamma z)))$.

Comenzando con la ecuación (4), Heckman propone el siguiente procedimiento en dos etapas:

1. Estimar por máxima verosimilitud los coeficientes γ del modelo Probit representado en la ecuación (1) y calcular la relación inversa de Mills para cada observación de la muestra $\hat{\lambda}$;
2. Estimar β y $\beta_{\lambda} = \sigma_{\varepsilon u} \sigma_{\varepsilon}$ con una regresión MCO de E_t sobre x , y la estimación de $\hat{\lambda}$.

Por tanto, en nuestro contexto de elección de dos etapas, se propone la aplicación del modelo Heckman estimado por el método anterior para modelar simultáneamente las decisiones de irse de vacaciones y el nivel de gasto turístico.

4. Resultados de la estimación.

Como comentamos anteriormente los datos de la EPF han sufrido diferentes transformaciones antes de poder estimar un modelo de Heckman sobre los mismos.

Al estar proporcionada la información en ficheros planos se procede a la elaboración de los diccionarios que nos permiten almacenar todos los datos con su información correspondiente en una base de datos con formato STATA. Una vez extraídos y

fusionados, la prioridad fue quedarse con aquellos gastos que eran relevantes para nuestro estudio, es decir, aquellos gastos que tenían que ver con el turismo, para ello, se eliminaron todas aquellas observaciones que no tuviesen los códigos de gasto “09601” y “09602” correspondientes a paquetes turísticos nacionales y paquetes turísticos internacionales respectivamente. Para aquellos hogares que tenían gasto en ambas partidas se sumaron dichas partidas, ya que en nuestro estudio no se diferencia entre turismo nacional e internacional, si no que tratamos el gasto en turismo en general.

Una vez se realizó todo este trabajo, borramos todas aquellas variables que no necesitamos en nuestro trabajo y estimamos el modelo Heckman.

Tabla 3. Modelo Heckman para el turismo.

Gastos en turismo condicionados a la participación.	Coef.	Robust std. err.
Ingresos	282.6638**	136.1761
Edad	-43.47501*	22.63996
Edad al cuadrado	0.4052842**	0.2022941
Sexo		
Hombre ^a	-	-
Mujer	-186.3319**	76.82363
Constante	3684.345*	1991.404
Decisión de participación.	Coef.	Robust std. err.
Ingresos	0.3047577***	0.206698
Municipio	0.0224454	0.0233213
Tamaño del hogar		
Uno o dos miembros ^a	-	-
Tres miembros	-0.1198307***	0.030206
Cuatro miembros	-0.0712865**	0.0324344
Cinco o más miembros	-0.1746093***	0.0510754
Estudios		
Hasta graduado en ESO ^a	-	-
Hasta ciclo grado superior	0.0645824**	0.0325558
Estudios Universitarios	0.189758***	0.0288987
Hipoteca		
No tiene hipoteca o alquiler ^a	-	-
Cuenta con hipoteca o alquiler	-0.1181376***	0.0262651
Edad	0.0243196***	0.0057659
Edad al cuadrado	-0.0002303***	0.0000516
Ocupación		
Desempleado ^a	-	-
Empleado	0.0328413	0.0356735
Constante	-4.073755***	0.2034432
mills		

lambda	431.4705***	-4.58
rho	-0.79030	
sigma	2500.3755	
Heckman selection model –two-step estimates (regression model with sample selection)	Number of obs = 22,029 Censored obs = 19,048 Uncensored obs = 2,621	
	Wald chi2(4) = 20.30 Prob > chi2 = 0.0004	

^a Categoría de referencia

Fuente: Elaboración propia a partir de los datos de la EPF.

La tabla 4 presenta los resultados de la estimación del modelo de Heckman de dos ecuaciones, basado en toda nuestra muestra de hogares de la EPF. En dicha tabla, se muestran las estimaciones tanto para la ecuación de participación como para la ecuación de gasto en turismo (para aquellos hogares que realizan gasto en turismo).

Como muestra dicha tabla el modelo es estadísticamente significativo en su conjunto (p-value = 0.0004 con la prueba de Wald).

Una de las variables mas interesantes para explicar tanto la participación como el nivel de gasto turístico es el ingreso de los hogares. Esta variable (incluida en logaritmos) fue significativa en ambas ecuaciones y, en ambos casos, mostró el signo positivo que se esperaba de ella. Se puede observar que para la cantidad que se decide gastar, una vez se ha optado por irse de vacaciones un aumento de un 1% en los ingresos supone aumentar dicho gasto en 2.82 unidades monetarias. Podríamos decir que los ingresos es una de las variables mas relevantes, si no es la más, a la hora de decidir gastar y cuanto en turismo.

La edad del sustentador principal de la familia tiene un efecto significativo en el caso del gasto condicionado a la participación ya que, en ambas ecuaciones, tanto en la de participación como en la de nivel de gasto, las dos variables relacionadas con la edad han sido significativas. Estas variables al igual que la de ingresos, resultan bastantes interesantes a la hora de su interpretación, ya que como bien hemos dicho anteriormente en este trabajo, la edad tiene un efecto no lineal en las decisiones de participación y gasto en turismo. Comprobamos que, en la ecuación de participación en los primeros años, un aumento de la edad supone un incremento de la probabilidad de irte de vacaciones, hasta que llegamos a cierto punto donde cambia el sentido, y un aumento de la edad supone justo lo contrario, una disminución de la probabilidad de decidir irse de vacaciones. Sin

embargo, en la decisión de cuanto gastar en turismo la edad tiene justo el efecto contrario, un aumento de un año, cuando el sustentador principal es joven, supone una reducción de la cantidad que gastas en turismo hasta que llegas a cierto punto donde esta tendencia se da la vuelta y un incremento en los años conlleva a gastar más en turismo. Estos resultados, referentes a la edad son los que se esperaban según la literatura que se ha utilizado para este trabajo (Nicolau y Mas, 2005).

En nuestro modelo la variable sexo, la cual se refiere al sexo del sustentador principal de la familia, también ha resultado ser significativa, dicha variable vemos que afecta a la hora de decidir cuanto gastar en turismo. Según los resultados de nuestro modelo, si el sustentador principal de la familia es mujer el gasto en turismo disminuye en 186 unidades, las razones por las cuales esta variable es así, las expondremos en las conclusiones de este trabajo.

En cuanto al tamaño del municipio donde se encuentra el hogar, la variable para un municipio con más de 100 mil habitantes, ha resultado no ser significativa en nuestra ecuación de participación. Después de probar varias formas para dicha variable se observó que cuanto mayor fuese el número de habitantes que utilizamos para el punto de corte en esta variable, más significativa se volvía, es decir que, si el punto de corte fuese 200 mil habitantes en vez de 100 mil, quizás podríamos hablar de una variable significativa. El problema es que la EPF, el punto de corte más alto que tiene es el que hemos utilizado. También se puede afirmar que el signo que obtiene la variable es el esperado, ya que a medida que aumenta el tamaño del municipio donde reside el hogar, es más probable que los hogares decidan irse de vacaciones.

El número de miembros del hogar ejerce un efecto negativo en la ecuación de participación. Además, dicha variable ha sido significativa para cualquiera de los niveles que tiene. Como se puede observar a medida que aumenta el tamaño del hogar se reduce la probabilidad de irse de vacaciones, cuando el hogar está formado por uno o mas miembros la probabilidad de participación se reduce entre 7 y 17 puntos porcentuales.

Otra variable con alta capacidad explicativa es el nivel de educación del jefe de familia. Esta variable muestra un efecto positivo, como se señaló en la revisión de la literatura, y es estadísticamente significativa a un nivel del 5% y 1% para ambos niveles de estudios, comparados con el nivel básico de estudios. Estos resultados encajan con los estudios de

Alegre y Pou (2004); Nicolau y Mas (2005); y también con los estudios de Eugenio-Martin y Campos Soria (2011).

La existencia de una hipoteca o alquiler como podemos ver es una variable con un nivel de significación muy alto y, por tanto, muy relevante a la hora de tomar la decisión de irse o no de vacaciones. Además, el valor de la variable es el esperado, pues cuando un hogar tiene la obligación de pagar una hipoteca o un alquiler se reduce la probabilidad de que decida irse de vacaciones.

Por último, tenemos la variable ocupación que indica si el sustentador principal se encuentra trabajando o desempleado, pero ésta no es estadísticamente significativa en nuestro modelo, aunque tiene el valor que se espera de ella, algo que esperábamos gracias a la literatura revisada (Alegre y Pou, 2004, Nicolau y Mas, 2005). Esta variable puede que no sea significativa debido a que la información de la misma se recoge una semana antes de elaborarse la EPF, es decir, que puede haber cabezas de familia que durante el año hayan estado desempleados y una semana antes de elaborar la encuesta haya encontrado empleo y, por tanto, no estaría introduciendo la información correcta en el modelo.

Para terminar con los resultados de nuestro modelo hay que destacar que nuestro modelo es válido, y esto lo sabemos gracias al parámetro lambda, es decir, la relación inversa de Mills. Si esta variable resulta significativa, hay evidencias de que nuestro modelo efectivamente cuenta con dos etapas que es lo que pretendemos nosotros demostrar, que las decisiones de participación y gasto en turismo constituyen dos etapas diferenciadas. Por tanto, todas nuestras interpretaciones son válidas, ya que, según nuestro modelo, primero se decide si ir o no de vacaciones y más tarde el nivel de gasto.

5. Conclusiones.

En este estudio, las decisiones de gasto en turismo se analizaron utilizando microdatos extraídos de un conjunto de datos nacional, la Encuesta de Presupuestos Familiares de España. Al estimar un modelo Heckman, se puede hacer un análisis de la decisión de participación y del nivel de gasto en turismo. Los resultados econométricos rechazaron la hipótesis de que ambas decisiones se puedan analizar de forma conjunta, esto es consideramos que el proceso de decisión consta de dos etapas. Además, los resultados estimados confirman algunas de las relaciones establecidas en estudios anteriores. Se han

detectado efectos positivos para las ecuaciones de participación y gasto en el caso de los ingresos. Efecto negativo de la edad y el sexo, y efecto positivo de la edad al cuadrado en la ecuación de gasto. Por otro lado, para la ecuación de participación se detectaron efectos positivos para los estudios, la edad, el tamaño del municipio y la ocupación, aunque los dos últimos no resultaron significativos, y efectos negativos en el tamaño del hogar, la edad al cuadrado y la tenencia de hipoteca o alquiler.

Nuestro trabajo nos ratifica la lógica de que un mayor nivel de los ingresos hace que los hogares decidan irse de vacaciones y gastar más en las mismas. Como es evidente un mayor poder adquisitivo conlleva a una mejora del nivel de vida en todos los aspectos y uno de esos aspectos es el turismo, ya que, a mayores ingresos, mayor parte de los mismos podemos dedicar a otras actividades tales como el turismo.

En el ámbito de la edad podemos sacar varias conclusiones bastante interesantes. En la ecuación de participación obteníamos dos efectos para la edad y la edad al cuadrado. La edad tenía un efecto positivo sobre la decisión de irse o no de vacaciones y la edad al cuadrado un efecto negativo. En mi opinión el signo de estas variables es debido a que cuando el sustentador de la familia es joven, se tienen pocas cargas familiares u obligaciones, lo que hace que la posibilidad de viajar sea más fácil para los hogares, en cambio, la edad al cuadrado tiene un efecto negativo en la decisión de hacer turismo o no, esto puede ser debido que a medida que aumenta la edad del sustentador principal de la familia, este tendrá mayores responsabilidades y obligaciones tales como trabajo, hijos, mayores cargas familiares etc., motivos por los cuales es más difícil para los hogares poder irse de vacaciones. También tenemos la edad en la ecuación de gasto, al igual que en la de participación los efectos de la variable edad y la variable de edad al cuadrado son diferentes. En este caso los efectos son totalmente contrarios a los de la ecuación de participación, la edad tiene un efecto negativo en el gasto en turismo en los años de mayor juventud, esto es debido a que la renta de las personas jóvenes es mucho menos que la de los individuos de avanzada edad que ya cuenta con una madurez alta en el mercado laboral, además los jóvenes optan por un turismo más “low cost” que las personas mas mayores que ellas. También puede ser debido a que en los años de juventud los individuos buscan un ahorro mayor debido a que tienen expectativas tales como la compra de una vivienda, un coche u otros elementos. Por otro lado, el signo positivo de la edad en los años donde esta es más avanzada, es debido a justo lo contrario que acabamos de explicar, es decir, las personas de cierta edad cuentan con unos puestos de trabajos mas estables y

con mejores sueldos que las personas jóvenes, por tanto, tienen un mayor poder adquisitivo y por ende más posibilidades para aumentar su gasto en turismo. Además, estas personas optan por un tipo de turismo muy diferente al “low cost”, ya que optan por turismo del tipo “sol y playa” o turismo de “salud”, los cuáles son más caros que los tipos de turismo por los que optan las personas jóvenes.

El sexo del sustentador principal de la familia, también parece ser un factor importante a la hora de cuanto gastar en turismo. En nuestro modelo podemos comprobar que el género del sustentador principal afecta a la cantidad gastada por el hogar en turismo se reduce en 186 unidades monetaria. Una posible explicación podría estar relacionada con un menor nivel de ingresos, pero tendemos a pensar que las mujeres racionalizan el gasto priorizando en otras partidas.

Aunque el tamaño del municipio haya resultado no significativo individualmente en nuestro modelo, sí que tenía el signo que se esperaba. Es normal que a medida que crece el tamaño de un municipio o ciudad, aumente la probabilidad de que se opte por tomarse vacaciones, ya que a medida que crece el tamaño de los municipios la vida es más frenética y deja menos tiempos a los individuos para disfrutar y realizar actividades de ocio y, por tanto, tomarse unas vacaciones significa desconectar de dichas rutinas monótonas.

El tamaño del hogar ha sido otra variable muy significativa en nuestro modelo a la hora de explicar el porque los individuos deciden irse o no de vacaciones. A medida que aumenta el tamaño del hogar se reduce la probabilidad de irse de vacaciones. Este resultado es algo que nos esperábamos ya que a medida que aumenta el tamaño de los hogares, los gastos del mismo también lo hacen en aspectos que son esenciales para la vida de las personas (alimentación, higiene, educación), por tanto, a los hogares le quedan menos recursos para otros gastos que no son tan necesarios, como es el turismo.

En cuanto al nivel de estudios del sustentador principal de la familia, cabe esperar que a medida que aumenta el nivel de estudios del mismo, aumente también la probabilidad de tomarse unas vacaciones. Las personas que tienen un mayor nivel de estudios suelen ser personas más cultas que las que tienen un menor nivel de los mismos y, por tanto, sienten mayor curiosidad por aumentar su cultura viajando y visitando lugares con un alto nivel cultural o intelectual. Además de todo esto, aquellas personas con mayores estudios cuentan con trabajos mejor remunerados que aquellos que tienen un nivel de estudios más

bajos y, por consiguiente, sus posibilidades de viajar son más altas. Aparte de la remuneración dichos empleos suelen contar con más días de vacaciones que los empleos de aquellas personas con un nivel de estudios más bajos y, por tanto, tienen más posibilidades durante el año de poder tomarse unas vacaciones.

La tenencia de una hipoteca o alquiler, supone un gasto fijo todos los meses que es imposible eludir para los hogares, por tanto, es algo evidente que las personas que tienen que afrontar un pago de hipoteca o alquiler tengan menos probabilidades de irse de vacaciones que aquellas que no cuentan con dicho gasto, ya que el poder adquisitivo de las que sí cuentan con el mismo será menor.

La situación laboral del sustentador principal de la familia, es evidentemente un factor importantísimo a la hora de tomar la decisión de optar o no por unas vacaciones. Aunque en nuestro modelo, es cierto, que dicha variable no ha sido significativa por lo que hemos explicado en el punto anterior, pero el resultado sí ha sido el esperado en una variable de este tipo, y es que, si una persona se encuentra empleada, es más probable que decida tomarse unas vacaciones que una persona desempleada. Por norma general, las personas con trabajo tienen unos ingresos mayores que las personas en situación de desempleo (aunque no siempre es así, ya que habría que analizar la tenencia de otro tipo de ingresos no provenientes del trabajo) y, por tanto, sus posibilidades de gasto son mayores.

Por tanto, y ante toda esta información que nos ha aportado el modelo, tenemos evidencia suficiente como para decir que las decisión de tomar vacaciones y la de cuanto gastar en las mismas son dos decisiones independientes, es decir, dichas decisiones constituyen dos etapas y que en dichas etapas influyen distintas características de los individuos, tanto personales (como son los ingresos, el tamaño del hogar), como sociodemográficas (como pueden ser el sexo, la educación, la edad, etc.).

Este trabajo puede ser útil para empresas u organizaciones a la hora de, por ejemplo, determinar estrategias de marketing y promoción del turismo, ya que gracias a un modelo Heckman podemos enfocarnos al público que más convenga, ya sea tanto a la hora de la captación, donde influiría la parte del modelo que se encarga de la selección, como en la parte de gasto, donde entraría en acción la parte del modelo que se encarga de determinar cuáles son los factores que determinan el nivel de gasto en turismo.

Bibliografía.

Alegre, J., & Pou, L. L. (2004). Micro-economic determinants of the probability of tourism consumption. *Tourism Economics*, 10(2), 125-144.

Alegre, J., Mateo, S., & Pou, L. L. (2013). Tourism participation and expenditure by Spanish households: The effects of the economic crisis and unemployment. *Tourism Management*, 39(2), 37-49.

Eugenio-Martín, J. y Campos-Soria, J.A., (2011), Income and the substitution pattern between domestic and international tourism demand, *Applied Economics*, 43(20), 2519-2531.

Instituto Nacional de Estadística., (2017), Encuesta de Presupuestos Familiares, Base 2006, Ficheros de usuarios, Año 2017.

Melenberg, B., & Van Soest, A. (1996). Parametric and semi-parametric modelling of vacation expenditures. *Journal of Applied Econometrics*, 11, 59-76.

Mergoupis y Steuer, M., (2003), Holiday taking and income, *Applied Economics*, 35(3), 269-284.

Nicolau, J. L., & Mas, F. (2005b). Heckit modeling of tourist expenditure: evidence from Spain. *International Journal of Service Industry Management*, 16(3), 271-293.

Organización Mundial del Turismo., (2018), UNWTO Annual Report 2017.

Toivonen, T., (2004), Changes in the propensity to take holiday trips abroad in EU countries between 1985 and 1997, *Tourism Economics*, 10(4), 403-417.

Anexo.

Como bien se ha mencionada en el trabajo los microdatos que nos facilita el INE se encuentran en un fichero plano, es decir, que es imposible trabajar con ellos en nuestro software econométrico STATA15.

Ilustración 1: Fichero de usuario de hogar 2017, en formato plano.

```
20170000110563212012577902040330000000201010201020100000002000002010201000301010220180101 7 7 6 9 9 6 4 3 5 4 4 2 20238 6 1 1
20170000212111122006525713210220000002000020002000000001000100020002000200000170150101101010 5 5 3 7 4 2 6 6 8 80185 6 1 1
20170000313311112014205193390220000000200002000200000000010100020002000200000170150100101010 5 5 4 7 4 2 6 6 8 80265 6 1 1
20170000407465311005712384050110000000100000100000000010000010000010000000000 2 2 2 1 1 1 1 2 1 1 1-9-90147 1 1 1
20170000512164222005268378390220000000200002000200000000002000020000101000200000170150101 6 6 5 6 6 5 3 2 2 3 3 5 20155 1 1 1
20170000616264222003931538980330000000201010201020100000101000002010201000301010220180101 7 7 6 9 9 6 4 3 5 4 4 2 20138 1 1 1
20170000709565322015719091780110000000100000100010000000001000001000001000000000 2 2 2 1 1 1 1 2 1 1 1-9-90147 1 2 2 1
20170000806164211004461541480440000000400000400040000020002000003010202010301010310250202111101212 8 5 4 6 5 5 2 10150 1 1 1
2017000091056431100906910712022000000200002000200000000101000020002000200000170150101 5 5 4 5 5 3 1 2 3 3 8 80166 1 1 1
201700010121121120037603783904400000030102020202000200000200002020202000402020290230202 8 8 71010 7 5 3 5 4 4 2 20150 1 1 1
201700011057621110066373151704400000004000010301030001010002000003010202000401010310250303111101212 7 5 4 6 5 5 2 10145 6 1 1
201700012074653220045436710903300000003000030003000000001010101020102000300000240200202101010 7 7 3 7 4 3 6 6-9-90185 6 1 1
20170001307412122004173953700220000000200002000200000000002000001010101000200000170150101 6 6 5 6 6 5 3 2 2 3 3 5 50158 1 1 1
201700014016621220091776072005500000004010104010400010100010200010400050104020203602803021212101212 9 8 4 6 5 5 8 60174 1 1 1
201700015152653110016142229304400000002020202020200000101000002020202000402020270210101 8 8 71010 6 5 3 5 4 4 2 20132 1 1 1
20170001609565322009376025350440000000202020202020101000002000002020103000402020270210101 8 8 71010 7 5 3 5 4 4 5 20139 1 1 1
2017000170226532200583069976011000000010000010001000000000010000010001000000 1 1 1 2 2 1 2 1 1 1 1 8 80167 1 1 1
20170001801665312010281623590220000002000020002000000002000020000200000170150101 6 6 5 6 6 5 3 2 2 3 3 8 20162 1 1 1
20170001907411110065248697002200000002000020002000000002000020000200000170150101 6 6 5 6 6 5 3 2 2 3 3 2 20160 1 1 1
20170002014664312005568368010330000000201010201020001000002000001020102000301010220180101 7 7 6 9 9 7 4 3 5 4 4 5 50239 6 1 1
201700021146633220053420916502200000002000020002000000000200002000200000170150101 5 5 4 5 5 3 1 2 3 3 8 80171 1 1 1
201700022074111220143838534001100000001000001000100000000010000010000010000000 2 2 2 1 1 1 1 2 1 1 1-9-90149 1 1 1
20170002307412222004367575560440000000301020202020002000002000002020103000402020290230200 8 8 71010 7 5 3 5 4 4 5 20245 6 1 1
20170002409565322015996490810330000000300000300030000010200000003000300000300000240200201121210 7 7 10 8 4 3 6 6 1 10132 1 1 1
20170002504564212009100348310330000000300000300030000000101010001020003000300000240200200111110 7 7 9 4 4 3 6 6 8 60165 1 1 1
20170002615211122001344953020220000000200002000200000001000101010002000200000170150101121210 5 5 4 8 4 2 6 6 8-90185 6 1 1
```

La Ilustración 1 muestra un ejemplo de como es un fichero plano, concretamente, este es uno de los ficheros utilizados en este trabajo, para más aclaración, se trata del “fichero de usuario de hogar 2017”. Como es de esperar con un archivo de este tipo es imposible trabajar, y para ello, se elaboran los diccionarios de STATA, para poder extraer toda la información que contiene este fichero de manera lógica.

En el diccionario hay que indicar donde empieza cada variable, cual es la longitud de la misma, cual es su nombre, el formato que le queremos dar en STATA (float, byte, int, str) y si queremos una pequeña descripción de la misma. Para realizar este proceso el INE nos facilita un informe en el que viene recogida toda esta información necesaria para la creación de los diccionarios.

La Ilustración 2 muestra como sería un ejemplo de diccionario en STATA, más concretamente, muestra el diccionario de hogar que se ha utilizado para este trabajo, es decir, el diccionario utilizado para extraer los datos de la Ilustración 1.

Para este trabajo se utilizó tanto el fichero de hogar como el fichero de gastos, para ello, hubo que realizar dos diccionarios distintos y extraer dos bases de datos. Una vez tuvimos las mencionadas bases de datos hubo que fusionar ambas, para agruparlas en una sola base de datos que recogiese toda la información necesaria para empezar a trabajar.

Ilustración 2: Ejemplo de diccionario en STATA.

```
1 dictionary using "C:\Users\Usuario\Desktop\TFM Mecofin\Microdatos\microhogar.raw" {
2   _column(1) float AÑO %4.0f "Año de la encuesta"
3   _column(5) float NUMERO %5.0f "Orden del hogar"
4   _column(10) byte CCAA %2.0f "Comunidad Autonoma"
5   _column(12) byte NUTS1 %1.0f "Region"
6   _column(13) byte CAPROV %1.0f "Capital provincia 1=si 2=no"
7   _column(14) byte TAMAMU %1.0f "Tamaño municipio"
8   _column(15) byte DENSI %1.0f "Densidad de poblacion"
9   _column(16) byte CLAVE %1.0f "Clave de colaboracion efectiva del hogar"
10  _column(17) byte CLATEO %1.0f "Clave de colaboracion teorica del hogar"
11  _column(18) float FACTOR %11.6f "Factor plobacional"
12  _column(29) byte NMIEMB %2.0f "Numero miembros del hogar"
13  _column(31) byte TAMAÑO %1.0f "Tamaño del hogar"
14  _column(32) byte NMIEMSD %2.0f "Numero de miembros que pertenecen al servicio domestico"
15  _column(34) byte NMIEMHU %2.0f "Numero de miembros del hogar que son huespedes"
16  _column(36) byte NMIEMIN %2.0f "Numero de miembros del hogar que son invitados"
17  _column(38) byte NMIEM1 %2.0f "Numero de miembros del hogar de 14 o mas años"
18  _column(40) byte NMIEM2 %2.0f "Miembors menores de 14 años"
19  _column(42) byte NMIEM3 %2.0f "Miembros menores de 16 años"
20  _column(44) byte NMIEM4 %2.0f "Miembros mayores de 16 años"
21  _column(46) byte NMIEM5 %2.0f "Miembros menores de 18 años"
22  _column(48) byte NMIEM6 %2.0f "Miembros mayores de 18 años"
23  _column(50) byte NMIEM7 %2.0f "Numero de miembros de 0 a 4 años"
24  _column(52) byte NMIEM8 %2.0f "Numero de miembros de 5 a 15 años"
```

Para ello se utilizó el número de hogar como punto de unión. Una vez fusionadas hubo que depurarlas manteniendo solo aquellos gastos que eran necesarios para nuestro trabajo y asignando un gasto 0, concretamente “missing values”, a aquellos hogares que no gastaban nada en turismo. Para poder llevar acabo este cometido se utilizó la variable “CODIGO”, la cual era una variable de tipo “string”, que tomaba distintos valores en función del tipo de gasto que realizaba el hogar.

Una vez realizado todo este trabajo y habiendo modificado o creado todas las variables necesarias para nuestro estudio, se procedió a borrar todas aquellas que no necesitábamos para obtener una base de datos mas manejable y que fue la que se usó para el trabajo. Podemos, a continuación, una parte de dicha base de datos en la Ilustración 3.

Ilustración 3: Base de datos utilizada para el trabajo.

Editor de Datos (Navegación) - [BD0 - copia]

Archivo Edición Ver Datos Herramientas

EDADSP[1] 38

	EDADSP	INGRESOS	HOGAR	GASTOTUR1	ESTUDIOS	OCUPACION	SEXO	EDADSP2	MUNICIPIO	HIPOTECA				
1	38	7.716906	2	.	3	1	1	1444	0	1				
2	85	6.489205	1	.	3	0	1	7225	1	1				
3	65	8.131531	1	.	3	0	1	4225	1	0				
4	47	7.095064	1	.	1	1	0	2209	0	0				
5	55	7.313221	1	.	3	1	0	3025	0	1				
6	38	7.688913	2	.	3	1	0	1444	0	1				
7	47	6.907755	1	.	2	1	0	2209	0	1				
8	50	7.649693	3	.	3	1	0	2500	0	0				
9	66	7.134891	1	.	3	0	0	4356	0	1				
10	50	7.451645	3	.	2	1	0	2500	0	0				
11	45	7.410952	3	178	1	1	1	2025	0	1				
12	85	7.574045	2	.	1	0	1	7225	0	0				
13	58	7.46049	1	1504.05	1	1	0	3364	0	0				
14	74	7.095893	4	.	1	0	0	5476	0	1				
15	32	7.698483	3	.	3	1	0	1024	0	1				
16	39	7.149917	3	.	1	1	0	1521	0	1				
17	67	5.905362	1	.	2	0	0	4489	0	1				
18	62	6.747587	1	.	1	0	0	3844	0	1				
19	60	7.924796	1	.	2	1	0	3600	1	1				
20	39	7.549609	2	.	1	1	1	1521	0	1				
21	71	8.260732	1	.	1	0	0	5041	0	0				
22	49	7.495542	1	2721.328	3	1	0	2401	1	0				
23	45	7.789868	3	301.0564	3	1	1	2025	0	1				
24	32	7.938802	2	.	2	1	0	1024	0	1				
25	65	7.783224	2	.	1	0	0	4225	0	1				
26	85	6.594413	1	.	1	0	1	7225	1	0				

Vars: 10 Orden: Dataset Obs: 22.029