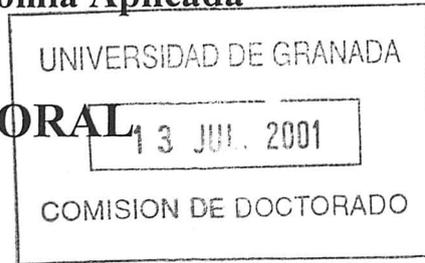




UNIVERSIDAD DE GRANADA

Facultad de Ciencias Económicas y Empresariales
Departamento de Economía Aplicada

TESIS DOCTORAL



UTILIZACIÓN DE LA
VEROSIMILITUD EMPÍRICA EN
ALGUNOS PROBLEMAS DE
INTERÉS ECONÓMICO.



M^a del Pilar Fernández Sánchez
Granada. 2.001

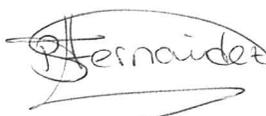


UNIVERSIDAD DE GRANADA.
Facultad de Ciencias Económicas y Empresariales.
Departamento de Economía Aplicada.

Utilización de la Verosimilitud Empírica en algunos
problemas de interés económico.

Memoria presentada para optar al grado de Doctor por la licenciada en Ciencias Económicas y Empresariales (Economía) de la Universidad de Santiago de Compostela M^a del Pilar Fernández Sánchez, que realizó y superó el programa de doctorado Métodos cuantitativos para la Economía Aplicada en el Departamento de Economía Aplicada de la Universidad de Granada durante el bienio 1.997-1.999.

Granada, 13 Julio, 2001



Vº Bº de los directores (por orden alfabético)



Dr.D. Agustín Hernández Bastida
Universidad de Granada.



Dr. D. Carlos Sánchez González
Universidad de Granada.

Memoria de Tesis Doctoral leída en el departamento de Economía Aplicada de la Universidad de Granada el día

ante el Tribunal siguiente,

--

Presidente
Prof. Dr. D.
Cargo y Universidad:
Secretario
Prof. Dr. D.
Cargo y Universidad:
Vocales
Prof. Dr. D.
Cargo y Universidad:
Prof. Dr. D.
Cargo y Universidad:
Prof. Dr. D.
Cargo y Universidad:
Prof. Dr. D.
Cargo y Universidad:
Obtuvo la calificación de

Agradecimientos.

A lo largo de la elaboración de esta Memoria he tenido la oportunidad de aprender muchas cosas relacionadas con la Estadística, la Economía, el manejo de ciertos programas informáticos,..., pero además he comprendido que no era un trabajo en solitario, sino que ha habido muchas otras personas que, de forma directa o indirecta, han colaborado en él. A todos vosotros quiero agradecer la ayuda prestada y que, aunque a veces yo no he creído en mis posibilidades, vosotros habéis seguido apostando por mí.

En primer lugar quiero dar las gracias a mis padres, a Julián y a Quique, y a mi abuela por el apoyo constante y el empuje que me han dado. Para Pablo, un recuerdo especial, por ofrecerme su cariño día a día.

A mis amigos (Cati, Mamen, Belén, Carmen M^a, María, M^a Victoria, Julia, Celia, Luis, Gregorio, Antonio ...), gracias por estar en los momentos bajos y porque vuestro interés ha sido siempre sincero. A M^a Dolores, por todos los pequeños enormes detalles que siempre tiene conmigo.

Un recuerdo también para mis compañeros de Métodos Cuantitativos, porque cuando he acudido a ellos, he obtenido, no la respuesta que esperaba, sino siempre mejor. Y al profesor Guerra Quintana, de la Universidad de Las Palmas, por su colaboración en temas informáticos.

Por último, muchas gracias a D. Agustín Hernández Bastida y a D. Carlos Sánchez González, sin los que este trabajo no hubiera sido posible. Infinitas gracias por vuestra paciencia, por las horas dedicadas, y por hacerme sentir que siempre había un momento para mí. A todos los que os interesáis por mí, GRACIAS.

ÍNDICE

página

Introducción.	XIII
Capítulo 1. Resultados técnicos.	1
1.1.- Principio de verosimilitud. Función de verosimilitud.	1
1.2.- Parámetros de perturbación.	20
1.3.- Un perfil de verosimilitud: Verosimilitud empírica.	27
1.4.- Verosimilitudes de perfil multinomial en un análisis bayesiano: Verosimilitud modificada.	37
Capítulo 2. Análisis de cantidades compradas, expresadas en unidades físicas, para algunos artículos de alimentación de los hogares españoles según la Encuesta Continua de Presupuestos Familiares.	45
2.1.- Introducción y planteamiento.	45
2.2.- Principales resultados sobre consumo de carne, aceite y tabaco.	54
Capítulo 3. Verosimilitud empírica y distribución de la renta.	77
3.1.- La renta y su distribución.	77
3.1.1.- Un poco de historia.	77
3.1.2.- Puntos de vista en la distribución de la renta.	79
3.2.- Estudios de renta.	81
3.2.1.- Modelización paramétrica de la renta.	82
3.2.2.- Enfoque no paramétrico de la distribución de la renta.	89
3.3.- Aclaraciones metodológicas.	91
3.3.1.- Datos: Ingresos o gastos.	91
3.3.2.- Individuos u hogares.	93
3.3.3.- Ingresos y ocultación.	94
3.4.- Análisis de los ingresos en unidades monetarias, según su naturaleza, a partir de la clasificación de la Encuesta Continua de Presupuestos Familiares.	96
3.4.1.- Ingreso total trimestral.	98
3.4.2.- Ingreso monetario.	102
3.4.3.- Ingreso no monetario.	104
3.4.4.- Participación de los distintos tipos de ingreso en el total.	106
3.5.- Comentarios y utilización de los resultados obtenidos.	110

Capítulo 4. Análisis de los gastos en unidades monetarias según la clasificación de la Encuesta Continua de Presupuestos Familiares.	115
4.1.- Introducción y planteamiento.	115
4.2.- Análisis del gasto total trimestral de los hogares españoles.	121
4.3.- Análisis de los grupos de gasto.	124
4.3.1.- Grupo 1. Alimentos, Bebidas y Tabacos.	124
4.3.2.- Grupo 2. Vestido y Calzado.	126
4.3.3.- Grupo 3. Vivienda, Calefacción y Alumbrado.	128
4.3.4.- Grupo 4. Muebles, Artículos de mobiliario. Artículos de Menaje y gastos de conservación corrientes de la casa.	130
4.3.5.- Grupo 5. Medicinas y otros productos farmacéuticos.	132
4.3.6.- Grupo 6. Transportes y Comunicaciones.	134
4.3.7.- Grupo 7. Esparcimiento, Espectáculo, Enseñanza y Cultura.	136
4.3.8.- Grupo 8. Otros bienes y servicios.	138
4.3.9.- Grupo 9. Otros gastos no mencionados anteriormente.	140
4.4.- Análisis de las participaciones de los grupos de gasto en el total.	142
4.4.1.- Participación del grupo 1 .	143
4.4.2.- Participación del grupo 2.	145
4.4.3.- Participación del grupo 3.	147
4.4.4.- Participación del grupo 4.	149
4.4.5.- Participación del grupo 5.	151
4.4.6.- Participación del grupo 6.	153
4.4.7.- Participación del grupo 7.	155
4.4.8.- Participación del grupo 8.	157
4.4.9.- Participación del grupo 9.	159
4.5.- Comentarios y utilización de los resultados obtenidos.	161
4.6.- Comparación entre ingresos y gastos totales. Otros comentarios.	165
Capítulo 5. Conclusiones.	171
Referencias bibliográficas.	177
Anexo 1: Resumen de la metodología de la Encuesta Continua de Presupuestos Familiares.	191
Anexo 2: Programas informáticos.	203
Anexo 3: Relación de algunas abcisas y el valor de la verosimilitud empírica para la variable “cociente entre gastos e ingresos”.	221

Figura	Índice de figuras	página
1.1.-	Probabilidad de evidencia débil frente a tamaño de muestra.	13
1.2.-	Probabilidad de evidencia engañosa frente a tamaño de muestra.	15
1.3.-	Probabilidades de error, de evidencia engañosa y de fallo frente a tamaño de muestra.	18
2.1.-	Resultados de descriptiva: carne, aceite y tabaco.	56
2.2.-	Verosimilitud empírica para carne de cerdo.	58
2.3.-	Cuadro para resultados de inferencia.	61
2.4.-	Verosimilitud empírica para carne pollo y gallina.	63
2.5.-	Resultados de inferencia para carne de pollo y gallina.	64
2.6.-	Verosimilitud empírica para carne de ternera.	65
2.7.-	Resultados de inferencia para carne de ternera.	66
2.8.-	Verosimilitud empírica para aceite de girasol.	67
2.9.-	Resultados de inferencia para aceite de girasol.	68
2.10.-	Verosimilitud empírica para aceite de oliva.	69
2.11.-	Resultados de inferencia para aceite de oliva.	70
2.12.-	Verosimilitud empírica para tabaco negro.	71
2.13.-	Resultados de inferencia para tabaco negro.	72
2.14.-	Verosimilitud empírica para tabaco rubio.	73
2.15.-	Resultados de inferencia para tabaco rubio.	74
3.1.-	Resultados de descriptiva: ingreso total, tipos de ingreso y participación de los tipos de ingreso en el total.	97
3.2.-	Verosimilitud empírica para el ingreso total trimestral.	98
3.3.-	Verosimilitud empírica para el ingreso monetario.	102
3.4.-	Resultados de inferencia para el ingreso monetario.	103
3.5.-	Verosimilitud empírica para el ingreso no monetario.	104
3.6.-	Resultados de inferencia para el ingreso no monetario.	105
3.7.-	Verosimilitud empírica para la participación del ingreso monetario dentro del ingreso total.	106
3.8.-	Resultados de inferencia para la participación del ingreso monetario dentro del ingreso total.	107
3.9.-	Verosimilitud empírica para la participación del ingreso no monetario dentro del ingreso total.	108
3.10.-	Resultados de inferencia para la participación del ingreso no monetario dentro del ingreso total.	109

4.1.- Resultados de descriptiva: gasto y grupos de gasto.	119
4.2.- Verosimilitud empírica para el gasto total trimestral.	121
4.3.- Verosimilitud empírica para el grupo 1 de gasto.	124
4.4.- Resultados de inferencia para el grupo 1 de gasto.	125
4.5.- Verosimilitud empírica para el grupo 2 de gasto.	126
4.6.- Resultados de inferencia para el grupo 2 de gasto.	127
4.7.- Verosimilitud empírica para el grupo 3 de gasto.	128
4.8.- Resultados de inferencia para el grupo 3 de gasto.	129
4.9.- Verosimilitud empírica para el grupo 4 de gasto.	130
4.10.- Resultados de inferencia para el grupo 4 de gasto.	131
4.11.- Verosimilitud empírica para el grupo 5 de gasto.	132
4.12.- Resultados de inferencia para el grupo 5 de gasto.	133
4.13.- Verosimilitud empírica para el grupo 6 de gasto.	134
4.14.- Resultados de inferencia para el grupo 6 de gasto.	135
4.15.- Verosimilitud empírica para el grupo 7 de gasto.	136
4.16.- Resultados de inferencia para el grupo 7 de gasto.	137
4.17.- Verosimilitud empírica para el grupo 8 de gasto.	138
4.18.- Resultados de inferencia para el grupo 8 de gasto.	139
4.19.- Verosimilitud empírica para el grupo 9 de gasto.	140
4.20.- Resultados de inferencia para el grupo 9 de gasto.	141
4.21.- Verosimilitud empírica para la participación en el gasto total del grupo 1.	143
4.22.- Resultados de inferencia para la participación en el gasto total del grupo 1.	144
4.23.- Verosimilitud empírica para la participación en el gasto total del grupo 2.	145
4.24.- Resultados de inferencia para la participación en el gasto total del grupo 2.	146
4.25.- Verosimilitud empírica para la participación en el gasto total del grupo 3.	147
4.26.- Resultados de inferencia para la participación en el gasto total del grupo 3.	148
4.27.- Verosimilitud empírica para la participación en el gasto total del grupo 4.	149
4.28.- Resultados de inferencia para la participación en el gasto total del grupo 4.	150
4.29.- Verosimilitud empírica para la participación en el gasto total del grupo 5.	151

4.30.- Resultados de inferencia para la participación en el gasto total del grupo 5.	152
4.31.- Verosimilitud empírica para la participación en el gasto total del grupo 6.	153
4.32.- Resultados de inferencia para la participación en el gasto total del grupo 6.	154
4.33.- Verosimilitud empírica para la participación en el gasto total del grupo 7.	155
4.34.- Resultados de inferencia para la participación en el gasto total del grupo 7.	156
4.35.- Verosimilitud empírica para la participación en el gasto total del grupo 8.	157
4.36.- Resultados de inferencia para la participación en el gasto total del grupo 8.	158
4.37.- Verosimilitud empírica para la participación en el gasto total del grupo 9.	159
4.38.- Resultados de inferencia para la participación en el gasto total del grupo 9.	160
4.39.- Verosimilitud empírica para el cociente entre gastos e ingresos totales.	166
4.40.- Resultados de inferencia para el cociente entre gastos e ingresos totales.	167

INTRODUCCIÓN.

El objeto de esta Memoria es la utilización de la noción estadística de Verosimilitud empírica en algunos problemas de interés económico. La investigación que se propone y se desarrolla en los capítulos que siguen se enmarca en un escenario de investigación más amplio que brevemente describimos a continuación. Pensamos que es interesante esta descripción porque nos permitirá una mejor comprensión del entorno.

Esencialmente lo que se hace en este trabajo es utilizar una verosimilitud de perfil multinomial; esta utilización de verosimilitudes de perfil multinomial está sustentada, al menos, en las siguientes investigaciones que enumeramos.*

Desde 1.985 se han publicado una serie de trabajos que desarrollan la noción de “verosimilitud empírica” como un modelo planteado en un escenario no paramétrico que construye un perfil multinomial . Estos artículos consideran una aproximación no paramétrica y no bayesiana y sus objetivos son habitualmente asintóticos.

* En los comentarios que siguen se evitan las referencias bibliográficas; creemos que aparecen las suficientes en los capítulos correspondientes, aquí solo enunciaremos las ideas principales de las diversas alternativas.

En la primera mitad de la década de los ochenta se publican una serie de trabajos en un escenario de análisis bayesiano con información a priori parcial. En estos trabajos se desarrollan razonablemente “verosimilitudes modificadas” para hacerlas compatibles vía teorema de Bayes con una información a priori más pobre que la necesaria en un análisis estadístico bayesiano habitual.

En 1.984 se publica un trabajo en el marco de la auditoria de cuentas en el que se desarrolla una aproximación intuitiva para utilizar el teorema de Bayes con una verosimilitud multinomial “más verosímil” (en la nomenclatura del autor).

En la primera mitad de la década de los noventa se publican una serie de artículos en los que se prueba que “la verosimilitud más verosímil” es un caso especial de “verosimilitud modificada”. Además es muy clara la extraordinaria semejanza de la “verosimilitud más verosímil” con la “verosimilitud empírica”.

En la segunda mitad de la década de los noventa se publican diversos artículos que hacen un uso extensivo de los resultados anteriores y de la verosimilitud de perfil multinomial en problemas de auditoria de cuentas y de valoración contingente en parques naturales. Estos trabajos incitan a un mayor interés por esta verosimilitud.

Desde nuestro punto de vista tenemos una función de verosimilitud de perfil multinomial que deseamos utilizar en escenarios de muestras finitas (es decir no nos planteamos resultados asintóticos) y queremos,

- utilizar la función de verosimilitud como una forma de representar evidencia,
- disponer de una función de verosimilitud, proporcionada por los datos, para un posterior análisis bayesiano.
- además pensamos que esta verosimilitud puede ser útil en diversos problemas de naturaleza económica (análogamente a lo que ha ocurrido en la auditoria de cuentas).

Por tanto hace unos años que nos planteamos la siguiente investigación,

(i) Para utilizar la verosimilitud, que ya en lo que sigue denominaré empírica para destacar su dependencia de los datos, en un análisis bayesiano son importantes los siguientes aspectos,

-estudiar la utilización de distribuciones a priori “poco informativas”; ya que no tendría sentido usar una verosimilitud empírica (“pegada” a los datos) con una distribución a priori basada en una especulación más o menos gratuita,

-estudiar las fluctuaciones de las conclusiones frente a pequeñas variaciones en la especificación de la información a priori.

En definitiva estudiar distribuciones a priori de máxima entropía y análisis de robustez bayesiana.**

(ii) Utilizar la función de verosimilitud como una forma de representar evidencia y ver su comportamiento en algunos problemas de interés económico. A una primera exploración en este camino se dedica esta Memoria. Insistimos desde este momento en que este trabajo se ubica en una aproximación bayesiana a la Estadística, aunque no hagamos un análisis bayesiano explícito; es decir, consideramos que el parámetro de interés se comporta como una variable aleatoria.

La Memoria está estructurada en cinco capítulos y tres anexos. El último capítulo es un capítulo de conclusiones donde aparece un resumen – esquema de las mismas porque con más detalle se han incorporado al texto principal. Análogamente se han incorporado al texto principal posibles y razonables líneas de investigación futuras por ello no aparece un capítulo dedicado explícitamente a ese punto.

Antes de describir someramente los distintos capítulos, un breve comentario sobre los datos utilizados. Se ha trabajado con los datos suministrados por la Encuesta continua de Presupuestos Familiares, más concretamente con los correspondientes al primer trimestre de 1.997, que es el último trimestre del que se tienen datos disponibles antes de la reforma de la encuesta. No tiene demasiada importancia los datos utilizados, ya que lo que se desea poner de manifiesto es la capacidad de la metodología para interpretar esos u otros datos. Según el estudio que pretenda realizarse, pueden solicitarse los datos adecuados, ya sean anuales, clasificados por comunidades, para diferentes países o incluso, los datos provenientes de una encuesta diseñada para conocer un fenómeno concreto.

** Este punto de la investigación es el objeto de la Tesis doctoral del profesor Guerra Quintana (Universidad de Las Palmas de G.C.) y del profesor Amor Pulido (Universidad de Jaén).

La Memoria se ha estructurado de la siguiente forma:

En el capítulo 1 se presentan las características técnicas de la metodología estadística a utilizar. En el epígrafe 1.1 se hace un breve repaso al concepto de verosimilitud y se desarrolla la idea de cómo la función de verosimilitud puede ser utilizada para representar la evidencia proporcionada por los datos.

El epígrafe 1.2 presenta el inconveniente que se produce cuando en la función de verosimilitud existe lo que hemos llamado parámetros de perturbación, y recoge algunas de las formas de solucionarlo. Una de esas posibles soluciones son los perfiles de verosimilitud, y en el apartado 1.3 destacamos la verosimilitud empírica. Recoge su definición, así como otros resultados estadísticos adicionales, algunos trabajos relacionados, usos,...

En el epígrafe 1.4 nos acercamos a las verosimilitudes de perfil multinomial, desde una óptica diferente, desde una óptica bayesiana. Se hace un resumen de los principales trabajos a este respecto y se establecen las especificaciones técnicas de lo que se ha dado en llamar verosimilitud modificada. Por último, se establece el paralelismo existente entre la verosimilitud modificada y la verosimilitud empírica.

El capítulo 2 está dedicado a la aplicación de nuestra metodología a las cantidades compradas, expresado en unidades físicas de una serie de artículos de alimentación, a partir de la Encuesta Continua de Presupuestos Familiares. El capítulo se ha dividido en dos epígrafes. En el primero de ellos, el 2.1, se establece una comparación entre la forma de trabajar con modelos econométricos y el uso de la verosimilitud empírica, destacando las posibilidades de obtención y actualización de resultados, con la ventaja de la ausencia de modelización pero el inconveniente de no poder explicar a partir de otras variables.

El apartado 2.2 contiene un resumen de las principales características descriptivas de los datos y los resultados obtenidos con esta metodología al aplicarla al análisis de la cantidad semanal comprada de carne de cerdo, pollo y ternera; aceite de girasol y de oliva; tabaco negro y rubio. Presentamos la gráfica de la verosimilitud empírica para cada artículo y un cuadro de resultados, así como algunas explicaciones del mismo.

El capítulo 3 está dedicado a la renta y su distribución, desde el punto de vista de la verosimilitud empírica.

Tras una introducción se lleva a cabo una revisión profunda de los distintos enfoques empleados en la literatura económica para analizar la distribución de la renta y otros aspectos relacionados con ella, tanto desde una perspectiva paramétrica, epígrafe 3.2.1, como no paramétrica –epígrafe 3.2.2- destacando en el primer caso los modelos más utilizados, las propiedades a cumplir por esos modelos, etc,..., y en el segundo caso, las técnicas no paramétricas más utilizadas. En ambos casos se hace también una recopilación de los trabajos realizados en ese sentido.

El siguiente epígrafe (3.3) está dedicado a comentar algunos conceptos metodológicos relacionados con el estudio de la distribución de la renta: que tipos de datos se utilizan, epígrafe 3.3.1, cuál es la unidad perceptora de rentas, epígrafe 3.3.2, y el tema de los ingresos y la ocultación (epígrafe 3.3.2).

El 3.4 recoge los resultados obtenidos al analizar los ingresos monetarios, no monetarios y totales para los hogares españoles, así como la participación media por hogar de cada tipo de ingreso en el total.

El apartado 3.5 está dedicado a los comentarios de los resultados del epígrafe anterior y a su posible utilización.

En el capítulo 4 abordamos el estudio, con la metodología descrita, de los gastos en unidades monetarias, de los hogares españoles, teniendo en cuenta la clasificación de los grupos de gastos establecidos por la Encuesta Continua de Presupuestos Familiares.

Se inicia el capítulo con unas breves notas acerca de los estudios de demanda y los trabajos de esta naturaleza, especialmente aplicados al caso español. Al mismo tiempo se realiza un análisis descriptivo de los grupos de gasto que van a ser objeto de estudio.

El epígrafe 4.2 recoge el análisis de la verosimilitud empírica para el gasto total trimestral de los hogares, de igual modo que en el 4.3 se lleva a cabo el estudio de los 9 grupos de gasto, presentando los resultados obtenidos.

En el siguiente epígrafe, 4.4, hemos querido utilizar la verosimilitud empírica para conocer la participación media por hogar de cada grupo de gasto dentro del total, para en el 4.5 comentar los resultados de los dos epígrafes anteriores.

El último epígrafe confronta los resultados obtenidos tanto para los ingresos como para los gastos totales, enlazando con el posible uso de la verosimilitud empírica para detectar la posible presencia de ocultación de rentas.

El capítulo 5 está dedicado a un breve resumen de lo realizado en esta Memoria, y de los principales resultados obtenidos.

Termina el trabajo con las referencias bibliográficas utilizadas para la elaboración del mismo y con tres anexos. El 1º de ellos está dedicado a la metodología de la encuesta que ha suministrado los datos, y los otros dos restantes a los programas informáticos utilizados y otras cuestiones relacionadas.

CAPÍTULO 1

RESULTADOS TÉCNICOS.

1.1.- Principio de verosimilitud. Función de verosimilitud.

Entre todas los criterios generales sobre el comportamiento estadístico, el Principio de verosimilitud es uno de los más sencillos y a su vez de los más enmarañados. Esencialmente establece que toda la evidencia que se obtiene de un experimento, sobre una cantidad desconocida, que notamos θ , está contenida en la función de verosimilitud de θ para los datos dados (fijos); las implicaciones de este principio son profundas.

Las medidas de precisión más habituales usadas en Estadística son preexperimentales. Esto significa que una vez que se ha decidido un método en particular para usarlo, la precisión de la evidencia proporcionada por un experimento se identifica con el comportamiento a “largo plazo” del método cuando el experimento se repite indefinidamente. Este comportamiento a largo plazo se evalúa promediando el resultado sobre todo el espacio muestral \mathfrak{N} , es decir, no solamente sobre el resultado obtenido, sino sobre todos los posibles resultados que podríamos obtener.

Frente a esta visión preexperimental, el Principio de verosimilitud establece la necesidad de utilizar un razonamiento posexperimental en el que

solamente los datos actualmente observados x son relevantes y no el resto de puntos de \aleph que podrían haber ocurrido. Existen diversas posiciones intermedias entre estos dos puntos de vista.

Veamos que se entiende por función de verosimilitud:

Para una variable aleatoria X con función de densidad $f_\theta(x)$, para todo $\theta \in \Theta$, se denomina función de verosimilitud a la función $L_x(\theta) = f_\theta(x)$; es decir, la densidad considerada como una función de θ para la observación dada x (una realización de X).

Como ya hemos comentado, el Principio de verosimilitud establece que toda la información sobre θ que se puede obtener de un experimento está contenida en la función de verosimilitud para la observación actual x . Dos funciones de verosimilitud para θ (del mismo o diferentes experimentos) que sean proporcionales - con factor de proporcionalidad que no dependa del parámetro - contienen la misma información sobre θ .

Se conoce al menos desde Fisher (1.925; 1.934) que la función de verosimilitud "aleatoria" $L_x(\theta)$ es un estadístico minimal suficiente para θ y por tanto contiene toda la información sobre θ desde un punto de vista clásico. Sin embargo, el Principio de verosimilitud va más allá manteniendo que solamente la función de verosimilitud, $L_x(\theta)$, es relevante dados los datos actuales.

El término "verosimilitud" apareció por primera vez en Fisher (1.921), que hizo diversas consideraciones en Fisher (1.925; 1.934; 1956) pero sin hacer comentarios sobre el Principio de verosimilitud.

En una serie de trabajos, Barnard promovió y desarrolló versiones del Principio de verosimilitud (ver Barnard (1.947a; 1.947b; 1.949)).

El Principio de verosimilitud recibió un notable empuje en 1.962 gracias a los trabajos de Barnard, Jenkins y Winsten (1.962) y Birnbaum (1.962).^[1]

^[1] El lector interesado en discusiones generales sobre el Principio de verosimilitud puede ver Pratt (1.965), Cox y Hinkley (1.974); Dawid (1.981), Barnett (1.982) o Basu (1.975). Para una historia del concepto de verosimilitud ver Edward (1.974).

Los trabajos anteriores se refieren a una breve historia del Principio de verosimilitud desde un punto de vista no bayesiano. En la aproximación bayesiana a la Estadística, el Principio de verosimilitud está implícito; en efecto, si $\pi(\theta)$ es la densidad a priori para θ , entonces la densidad a posteriori viene dada por,

$$\pi(\theta/x) = \frac{L_x(\theta) \cdot \pi(\theta)}{m(x)},$$

(suponiendo que $m(x) = E^\pi L_x(\theta) = \int_{\Theta} L_x(\theta) \pi(\theta) d\theta > 0$), que depende del experimento solamente a través de $L_x(\theta)$, (suponiendo que la elección de π es independiente del experimento y de x).

Supongamos una función de verosimilitud para un θ , $L_x^{(1)}(\theta)$, y una distribución a priori sobre θ , $\pi(\theta)$. La distribución a posteriori, $\pi_1(\theta/x)$, vendría dada por la expresión que dimos anteriormente.

Si tuviésemos otra función de verosimilitud proporcional a la anterior, por ejemplo, $L_x^{(2)}(\theta) = c(x)L_x^{(1)}(\theta)$, la distribución a posteriori sería:

$$\pi_2(\theta/x) = \frac{c(x)L_x^{(1)}(\theta)}{c(x) \int_{\Theta} L_x^{(1)}(\theta) \pi(\theta) d\theta} = \pi_1(\theta/x).$$

Puesto que toda la inferencia bayesiana se deduce de la distribución a posteriori, el Principio de verosimilitud es una consecuencia del paradigma bayesiano.

El trabajo de Birnbaum en 1.962 contiene numerosos ejemplos a favor del Principio de verosimilitud y estableció el desarrollo axiomático del Principio a partir de conceptos más sencillos y plausibles como son la suficiencia y la condicionalidad.

Veamos cuales son los grandes pasos del desarrollo de Birnbaum para el caso discreto:

Por un experimento E entendemos una terna $(X, \theta, \{f_\theta\})$, donde X es la variable aleatoria observable (que toma valores en \aleph) y tiene densidad $f_\theta(x)$ para algún θ en Θ .

En principio todas las metodologías estadísticas trabajan exclusivamente con la información anterior (en lo que se refiere al experimento).

Se lleva a cabo el experimento y se obtiene una realización $X = x$; y , a partir de E y x inferimos o concluimos alguna cosa sobre θ o sobre algo relacionado con θ . A esa inferencia o conclusión la denominamos "evidencia sobre θ obtenida a partir de E y de x " y la notamos $Ev(E, x)$.

Hay que observar que no se presupone nada sobre lo que esta evidencia es y tampoco se prejuzga sobre la posibilidad de que la evidencia $Ev(E, x)$ pueda depender de otras informaciones, tales como la información a priori sobre θ o una función de pérdida en un problema de decisión. Lo que realmente interesa es la manera en que esa evidencia $Ev(E, x)$ dependerá del experimento y de los datos que se han observado.

Comentemos a continuación los principios a partir de los cuales Birnbaum desarrolló el Principio de verosimilitud: el Principio de condicionalidad y el Principio de suficiencia.

Principio de condicionalidad débil.^[2]

Supongamos que existen dos experimentos $E_1 = (X_1, \theta, \{f_\theta^1\})$ y $E_2 = (X_2, \theta, \{f_\theta^2\})$, cada uno con una variable aleatoria observable con su función de densidad. Solamente el parámetro desconocido θ es común.

^[2] Esencialmente el principio de condicionalidad dice que si se selecciona un experimento por algún mecanismo aleatorio independiente de θ , entonces solo es relevante el experimento que realmente se ha realizado. La idea general de este principio es debida a Cox (1.958) y la definición que se da aquí es una versión más débil debida a Basu (1.975).

Consideremos el experimento mixto E^* , en el que se observa $J = 1$ ó 2 , cada uno con probabilidad $1/2$ (independiente de θ , X_1, X_2) y a continuación se realiza el experimento E_j . Formalmente, $E^* = (X^*, \theta, \{f_\theta^*\})$ donde $X^* = (J, X_j)$ y $f_\theta^*((j, x_j)) = \frac{1}{2} f_\theta^j(x_j)$. Entonces,

$$Ev(E^*, (j, x_j)) = Ev(E_j, x_j),$$

es decir, la evidencia sobre θ obtenida a partir de E^* es la evidencia obtenida a partir del experimento que se ha realizado.

Principio débil de suficiencia.^[3]

Consideremos un experimento $E = (X, \theta, \{f_\theta\})$ y supongamos que $T(X)$ es un estadístico suficiente para θ . Si $T(x_1) = T(x_2)$, se verifica que $Ev(E, x_1) = Ev(E, x_2)$.

Es decir, si dos muestras presentan el mismo valor para un estadístico suficiente, las dos muestras proporcionan la misma evidencia.

El Principio de verosimilitud se deduce directamente a partir de los principios débil de condicionalidad y débil de suficiencia.^[4]

El desarrollo formal del Principio de verosimilitud puede verse en Berger y Wolpert (1.984), tanto para el caso del espacio muestral discreto, que es el que aparece en Birnbaum (1.962), como para el caso más general de espacio continuo.

^[3] Esta formulación es debida a Dawid (1.977).

^[4] Pueden formularse una variedad de principios alternativos que también conducen al principio de verosimilitud, ver Basu (1.975), Dawid (1.977), Barndoff-Nielsen (1.978) y Berger (1.984).

En las notas anteriores se ha expuesto, muy breve y muy someramente, el Principio de verosimilitud y se ha enlazado con otros principios más familiares en Estadística. Obsérvese que en el desarrollo anterior no se precisa en ningún momento el significado exacto de la evidencia, y tampoco se ha dado ninguna indicación precisa sobre la forma de implementar el Principio de verosimilitud.

El objetivo de esta parte de la Memoria no es plantear y desarrollar una discusión sobre fundamentos en Estadística, referida al Principio de verosimilitud y su implementación^[5]. Más bien, el objetivo es destacar la importancia de la función de verosimilitud y su posible utilización.

Las notas que siguen a continuación discuten y amplían la idea de que la función de verosimilitud puede ser utilizada como una forma de representar la evidencia proporcionada por unos datos.

En Estadística se necesita un concepto cuantitativo explícito de evidencia y algunos procedimientos para medir y controlar la frecuencia con que los experimentos producen evidencia débil o engañosa.

Al no disponer de una expresión directa para la evidencia en la teoría habitual de contraste de hipótesis, las cantidades que están presentes - las probabilidades α y β de error - son habitualmente interpretadas como medidas de intensidad de la evidencia; así cuando se rechaza H_1 a favor de H_2 en un contraste con α y β pequeños^[6], se entiende que hay una fuerte evidencia que soporta a la hipótesis H_2 frente a la hipótesis H_1 ; o cuando se pide un p-valor pequeño se considera equivalente a una petición de evidencia fuerte.

La ley de verosimilitud incluye ese concepto necesario de evidencia, y cuando ésta se mide por la razón de verosimilitudes tenemos los medios para especificar los objetivos con precisión, en términos de la evidencia generada

^[5] El lector interesado puede consultar el trabajo citado de Berger y Wolpert (1.984) y la gran cantidad de referencias bibliográficas que aparecen en dicho trabajo.

^[6] H_1 : hipótesis nula; H_2 : hipótesis alternativa.

α : probabilidad de cometer error tipo I; β : probabilidad de cometer error tipo II.

p-valor: valor de la probabilidad que limita el nivel de significación a partir del cual se rechaza la hipótesis.

por el experimento, y para controlar las probabilidades de resultados no satisfactorios.

Consideremos el mismo modelo para la teoría de Neyman-Pearson:

Tenemos dos hipótesis H_1 y H_2 y estas hipótesis implican funciones de densidad de probabilidad f_1 y f_2 , respectivamente, para una variable aleatoria X . Para obtener evidencia sobre H_1 frente a H_2 , disponemos de observaciones (x_1, x_2, \dots, x_n) . Interpretaremos los datos de acuerdo a la ley de verosimilitud, es decir, la razón

$$\frac{L_2}{L_1} = \prod_{i=1}^n \frac{f_2(x_i)}{f_1(x_i)}$$

mide la evidencia de H_2 frente a la hipótesis H_1 , y el tamaño de esta razón mide la fortaleza de la evidencia. Así, una razón menor que 1 significa evidencia a favor de H_1 , mientras que el caso contrario indica evidencia a favor de H_2 .

El propósito de la toma de observaciones es obtener una evidencia fuerte en un sentido o en otro; es decir, una evidencia fuerte que soporte a H_2 frente a H_1 o al revés. Los datos serán inútiles si solamente constituyen una evidencia débil respecto al par de hipótesis, ya que no podríamos concluir nada.

Sin embargo, una evidencia débil no es el único tipo de resultado no satisfactorio que los datos pueden producir. Será todavía peor que los datos produzcan una evidencia fuerte que soporte a H_2 cuando H_1 es cierta o al revés; es decir, la evidencia débil es mala pero la evidencia fuerte que es engañosa es todavía peor.

Para ver la diferente óptica que presentan, comparemos esta formulación con la teoría de Neyman-Pearson.

Formulación de Neyman-Pearson:

Los datos se utilizan para elegir entre dos alternativas H_1 y H_2 , y no para representar o interpretar las observaciones como evidencia.

Antes de que se observe la variable se selecciona un posible conjunto de valores, que es lo que se denomina región crítica. Se observa $X = x$. Si la observación cae en la región crítica se elige H_2 , y en caso contrario se escoge H_1 .

Podemos concluir erróneamente de una de las dos siguientes formas:

- Al elegir H_2 cuando H_1 es cierta – es lo que se conoce como error tipo I- ó,
- Al elegir H_1 cuando H_2 es cierta - error tipo II -.

Queremos asegurarnos, con probabilidades respectivas de al menos $1 - \alpha$ y $1 - \beta$, que no cometemos ninguno de los dos errores.

Formulación de la evidencia:

Este enfoque está basado en el Principio de verosimilitud y soporta una teoría para representar los datos como evidencia y para medir la fuerza de esa evidencia.

Los datos se utilizan para generar evidencia sobre H_1 frente a H_2 . Se puede producir un resultado no satisfactorio que puede tomar una de las dos siguientes formas:

- evidencia fuerte que soporta la hipótesis falsa frente a la verdadera, lo que se conoce como evidencia engañosa, o
- evidencia débil.

Lo que queremos es asegurarnos, con probabilidades respectivas de al menos $1-M$ y $1-W$, que los datos no producirán un resultado no satisfactorio de cualquiera de los dos tipos.

Teniendo en cuenta que pueden ser un poco más desconocidos, comentemos un poco estos conceptos:

La evidencia fuerte se representa por una razón de verosimilitud de al menos algún valor umbral k . Las observaciones:

- dan una evidencia fuerte a favor de H_2 si $\frac{L_2}{L_1} \geq k$;
- dan evidencia fuerte a favor de H_1 si $\frac{L_2}{L_1} \leq \frac{1}{k}$, y
- dan evidencia débil si $\frac{1}{k} < \frac{L_2}{L_1} < k$.

Aquí se plantea una cuestión análoga a la de los contrastes de hipótesis, y es la siguiente: ¿qué valor de k es apropiado?

En los contrastes se pregunta: ¿qué nivel de significación es adecuado?, considerando por convención el 5%.

En la misma línea de especificar un convenio útil podemos considerar una razón de verosimilitud de $8^{[7]}$ para identificar una evidencia fuerte y una razón de $32^{[8]}$ para una evidencia muy fuerte.

^[7] Puede ser útil el siguiente ejemplo, muy sencillo, para interpretar las razones de verosimilitudes.

Supongamos que se dispone de dos urnas idénticas, una de ellas contiene solamente bolas blancas y la otra contiene igual número de bolas blancas que negras. Se selecciona una de las urnas y se extraen bolas con reemplazamiento. Tenemos dos hipótesis sobre el contenido de la urna seleccionada “todas las bolas son blancas” y “la mitad de las bolas son blancas”; las observaciones dan evidencia.

Supongamos que se extraen tres bolas y que las tres resultan ser blancas. Estas tres observaciones las consideramos una evidencia fuerte a favor de la hipótesis “todas las bolas son blancas” frente a “la mitad de las bolas son blancas”. La razón de verosimilitudes es $2^3 = 8$.

En general, si se observan b bolas blancas seguidas, la razón de verosimilitudes a favor de “todas blancas” frente a “la mitad son blancas” es $\frac{1}{\left(\frac{1}{2}\right)^b} = 2^b$.

^[8] Con el ejemplo de la nota a pie de página anterior correspondería a extraer 5 bolas consecutivas de color blanco.

Posteriormente, utilizaremos la noción de intervalo de verosimilitud. Este concepto recogerá la idea intuitiva siguiente:

Seleccionaremos un conjunto de puntos del espacio paramétrico donde la función de verosimilitud toma un “valor grande”. El término grande lo precisamos de acuerdo al valor del máximo de la función de verosimilitud.

Concretando, entenderemos por intervalo de verosimilitud de fuerza 8 (por ejemplo), a un intervalo (v_1, v_2) , tal que:

$$\begin{aligned} \forall v \in (v_1, v_2), \frac{L(a)}{L(v)} < 8; \\ \forall v \notin (v_1, v_2), \frac{L(a)}{L(v)} > 8, \end{aligned}$$

siendo a la abscisa del máximo de $L(\cdot)$.

Análogamente se obtiene el intervalo de verosimilitud para una fuerza de evidencia 32, o de cualquier otro valor.

Desde luego, puede ser apropiado especificar valores diferentes k_1 y k_2 para caracterizar:

- La evidencia “suficientemente fuerte” a favor de H_2 , si $\frac{L_2}{L_1} \geq k_2$;
- y evidencia “suficientemente fuerte” a favor de H_1 si $\frac{L_2}{L_1} \leq \frac{1}{k_1}$.

Una vez elegido k_1 y k_2 , podemos calcular la probabilidad de fallar al obtener evidencia suficientemente fuerte, - o lo que es lo mismo, probabilidad de obtener evidencia débil- la cual viene dada por:

$$\Pr \left[\frac{1}{k_1} < \frac{L_2}{L_1} < k_2 \right] \text{ bajo } H_1 \text{ y bajo } H_2.$$

Veamos a continuación un ejemplo que ilustra las probabilidades de evidencia débil y engañosa en un caso con la distribución Normal.

Ejemplo.(Royall (1.997), pág 90 y sgtes).

Consideremos una distribución $N(\theta, \sigma^2)$, con σ^2 conocido, y supongamos que nos planteamos el siguiente contraste de hipótesis con hipótesis nula simple y alternativa también simple

$$\begin{aligned} H_1 : \theta = \theta_1 \\ H_2 : \theta = \theta_1 + \delta, \delta > 0 \end{aligned}$$

En este caso la razón de verosimilitudes es igual a,

$$\frac{L_2}{L_1} = \exp \left\{ \left[\bar{x} - \left(\theta_1 + \frac{\delta}{2} \right) \right] n \frac{\delta}{\sigma^2} \right\}.$$

Por tanto, las observaciones que dan una razón de verosimilitudes intermedia

$$\frac{1}{k_1} < \frac{L_2}{L_1} < k_2,$$

(lo que significa que las observaciones no han producido una evidencia fuerte), serán aquellas tales que la media muestral \bar{x} pertenece a un intervalo que notamos $(\bar{x}_L; \bar{x}_U)$, donde

$$\begin{aligned} \bar{x}_L &= \theta_1 + \frac{\delta}{2} - \sigma^2 \frac{\log k_1}{n\delta}, \quad y \\ \bar{x}_U &= \theta_1 + \frac{\delta}{2} + \sigma^2 \frac{\log k_2}{n\delta} \end{aligned}$$

Si notamos las probabilidades de obtener evidencia débil bajo H_1 y H_2 por W_1 y W_2 respectivamente; podríamos calcularlas de la siguiente forma:

$$W_1 = \Pr[\bar{x}_L < \bar{x} < \bar{x}_U / H_1].$$

Tipificando bajo la hipótesis H_1 , se obtiene:

$$W_1 = \Pr \left[\frac{\frac{\delta}{2} - \frac{\sigma^2 \log k_1}{n\delta}}{\sigma/\sqrt{n}} < Z < \frac{\frac{\delta}{2} + \frac{\sigma^2 \log k_2}{n\delta}}{\sigma/\sqrt{n}} \right];$$

$$W_1 = \Phi \left(\frac{\delta\sqrt{n}}{2\sigma} + \frac{\sigma \log k_2}{\delta\sqrt{n}} \right) - \Phi \left(\frac{\delta\sqrt{n}}{2\sigma} - \frac{\sigma \log k_1}{\delta\sqrt{n}} \right),$$

donde Φ representa la función de distribución de una $N(0,1)$.

De forma análoga calcularíamos la probabilidad de evidencia débil bajo H_2 .

Si definimos la función $W(k_1, k_2)$ por ,

$$W(k_1, k_2) = \Phi \left(\frac{\sqrt{n}\delta}{2\sigma} + \frac{\sigma \log k_1}{\sqrt{n}\delta} \right) - \Phi \left(\frac{\sqrt{n}\delta}{2\sigma} - \frac{\sigma \log k_2}{\sqrt{n}\delta} \right),$$

entonces, podemos escribir las probabilidades anteriores W_1 y W_2 , de la siguiente forma,

$$W_1 = \Pr(\bar{x}_L < \bar{x} < \bar{x}_U / H_1) = W(k_2, k_1),$$

$$W_2 = \Pr(\bar{x}_L < \bar{x} < \bar{x}_U / H_2) = W(k_1, k_2).^{[9]}$$

Para los mismos valores de k_1 y k_2 las dos probabilidades de obtener evidencia débil W_1 y W_2 son iguales.

El siguiente gráfico muestra la probabilidad común $W(8,8)$, ($k_1 = k_2 = 8$) y pone de manifiesto como decrece cuando n crece; los casos dibujados corresponden a las situaciones en que las dos medias de hipótesis difieren en 0,25; 0,5 y 1 veces la desviación típica.

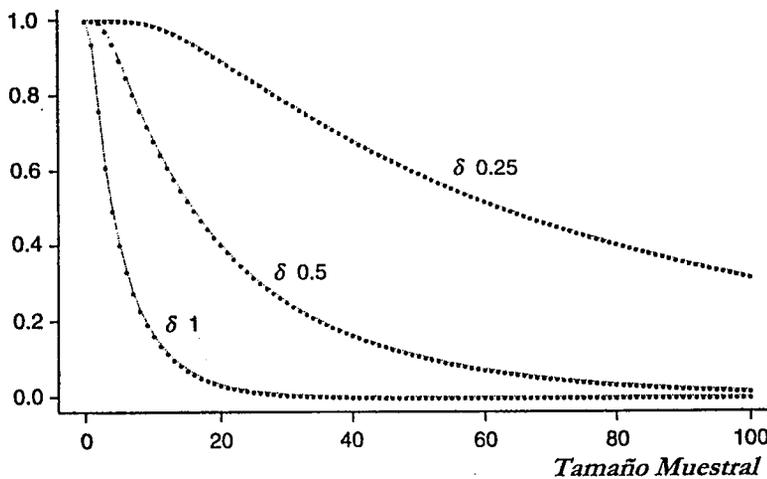


Figura 1.1.

^[9] Cuando $n \rightarrow \infty$, claramente W_1 y W_2 convergen a 0, por tanto se puede fijar el riesgo de fallar al dar evidencia suficientemente fuerte, en cualquier valor $W \in (0,1)$ y encontrar un n suficientemente grande para asegurarnos que $\max(W_1, W_2) \leq W$.

Para tener una probabilidad de evidencia débil menor que 0,05 se necesitan 20 observaciones, cuando las medias difieren en una vez la desviación típica; sobre 60 cuando difieren en 0,5, y alrededor de 100 cuando difieren en 0,25.

Consideremos ahora el resultado indeseable más serio, que es encontrar evidencia fuerte engañosa.

Si H_1 es cierta, habrá evidencia fuerte engañosa cuando observamos una muestra cuya media sea mayor que \bar{x}_U .

Esta probabilidad es $M_1 = \Pr[\bar{x} > \bar{x}_U / H_1]$. Tipificando bajo la hipótesis H_1 , se obtiene,

$$M_1 = \Pr \left[Z > \frac{\frac{\delta}{2} + \frac{\sigma^2 \log k_2}{n\delta}}{\sigma/\sqrt{n}} \right] = \Phi \left(-\frac{\delta\sqrt{n}}{2\sigma} - \frac{\sigma \log k_2}{\delta\sqrt{n}} \right)$$

Si H_2 es cierta, ocurrirá evidencia engañosa cuando la media muestral sea menor que \bar{x}_L , y la probabilidad de este suceso es $M_2 = \Pr[\bar{x} < \bar{x}_L / H_2]$. Se calcularía de forma análoga a como hemos calculado M_1 .

$$\text{Si notamos } M(k) = \Phi \left(-\frac{\sqrt{n} \delta}{2\sigma} - \frac{\sigma \log k}{\sqrt{n} \delta} \right),$$

estas probabilidades serán:

$$M_1 = M(k_2);$$

$$M_2 = M(k_1).$$

Se puede probar que $M(k)$ no puede ser mayor que $\Phi(-\sqrt{2 \log k})$ ^[10] para cualquier elección de n ; por tanto,

para $k = 8$, $M(k) \leq 0,021$;
 para $k = 32$, $M(k) \leq 0,004$.

Para cualquier $k > 1$ y $\frac{\delta}{\sigma} \neq 0$, el comportamiento de $M(k)$, probabilidad de evidencia engañosa, sigue el mismo patrón:

Comienza en 0 cuando $n = 0$, crece con n hasta $n = 2(\log k) \left(\frac{\sigma}{\delta}\right)^2$, donde alcanza el mayor valor posible que es $\Phi(-\sqrt{2 \log k})$, y a continuación decrece. Esto se ilustra en el siguiente gráfico:

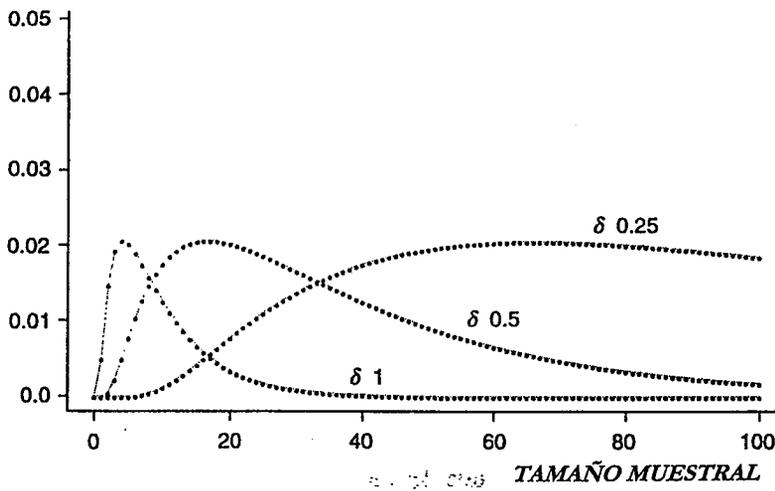


Figura 1.2.

[10] Ver ejercicio 1.5 de Royall (1.997).

Para $k = 8$ el gráfico indica como la probabilidad de evidencia engañosa $M_1 = M_2 = M(8)$ varía con el tamaño de la muestra, de nuevo cuando las medias difieren en 1, 0,5 y 0,25 veces la desviación típica. Para cada curva el máximo se alcanza en $\Phi(-\sqrt{2 \log 8}) = 0,021$.

De la Figura 1 vemos que cuando $\frac{\delta}{\sigma} = 0,25$ y el tamaño muestral es pequeño (por ejemplo menor que 10), la probabilidad de evidencia débil W está próxima a 1; es decir, existe una posibilidad pequeña de encontrar evidencia moderadamente fuerte ($k \geq 8$) a favor de cualquiera de las hipótesis cuando las medias difieren en solamente 1/4 de la desviación típica. Desde luego, esto también significa (ver Figura 2) que existe una muy pequeña posibilidad de evidencia fuerte engañosa. Esto es debido a que tendremos evidencia fuerte a favor de H_2 solamente si $\bar{x} \geq x_U$.

Por ejemplo, para $n = 10$; $\frac{\delta}{\sigma} = 0,25$; y $k_1 = 8$, la probabilidad de este suceso bajo H_2 es $M_2 = 1 - \Phi(3,026) = 0,001$.

La probabilidad de evidencia engañosa, M_1 , es en cierto sentido análoga a la probabilidad α de los tests de Neyman-Pearson, de rechazar H_1 cuando H_1 es cierta, porque con frecuencia el rechazo de H_1 se entiende que significa que se ha observado una fuerte evidencia contra H_1 .

Con el objeto de comparar, consideremos la suma $M_2 + W_2$, es decir la probabilidad, cuando H_2 es cierta, de tener evidencia fuerte engañosa o evidencia débil.

Esta es la probabilidad de fallar porque no hemos encontrado evidencia fuerte que soporte H_2 cuando H_2 es cierta; es análoga a la probabilidad β de un error tipo II, (la probabilidad de rechazar H_1 a favor de H_2 cuando H_2 es cierta).

Pero la aproximación que estamos considerando aquí reconoce que la probabilidad de fallar porque no hemos encontrado evidencia fuerte a favor de H_2 frente a H_1 , cuando H_2 es cierta, consiste en dos componentes diferentes:

- La probabilidad M_2 de encontrar evidencia engañosa a favor de H_1 , y
- La probabilidad W_2 de encontrar solamente evidencia débil.

Existe una distinción empírica importante entre evidencia débil y engañosa. Cuando se han observado los datos conocemos si se ha producido evidencia débil. Es decir, sabemos si ha ocurrido o no el suceso evidencia débil, cuyas probabilidades bajo H_1 y H_2 son respectivamente W_1 y W_2 . Por el contrario, cuando obtenemos evidencia fuerte no podemos saber si es engañosa o no.

Para resaltar la diferencia entre las probabilidades de error (α, β) y las probabilidades $(M_1, M_2 + W_2)$ de evidencia engañosa a favor de H_2 y de fallo al no encontrar evidencia fuerte a favor de H_2 cuando H_2 es cierta, dibujamos estas cantidades en la siguiente figura con respecto al tamaño muestral n .

Elegimos una desviación típica $\sigma = 1$; la diferencia entre las medias de las hipótesis $\delta = \frac{1}{2}$; y $k_1 = k_2 = 8$, para representar la evidencia fuerte y $\alpha = 0,05$.

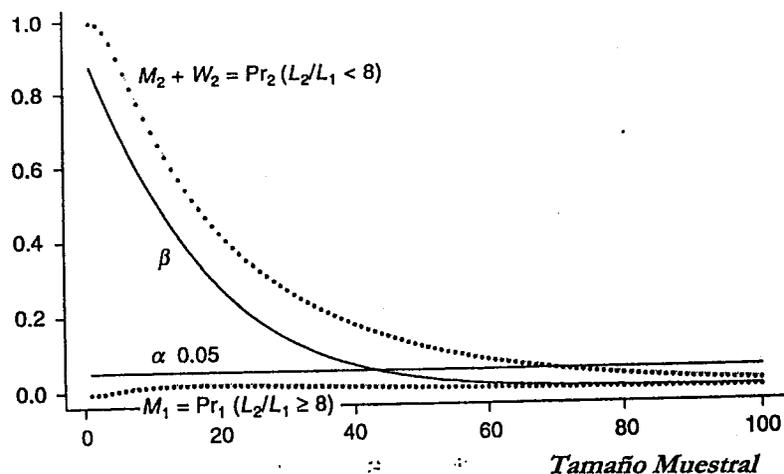


Figura 1.3.

Las funciones que se representan son:

$$\alpha = 0,05;$$

$$M_1 = \Pr \left[\frac{L_2}{L_1} \geq 8 / H_1 \right] = M(k_2) = \Phi \left(\frac{-\sqrt{n}}{4} - 2 \frac{\log 8}{\sqrt{n}} \right);$$

$$\beta = \Pr [\text{no rechazar } H_1 / H_2] = \Phi \left(1,645 - \frac{\sqrt{n}}{2} \right);$$

$$M_2 + W_2 = \Pr \left[\frac{L_2}{L_1} < 8 / H_2 \right] = M(k_1) + W(k_1, k_2) = \Phi \left(\frac{-\sqrt{n}}{4} + 2 \frac{\log 8}{\sqrt{n}} \right).$$

Aquí,

α es la probabilidad de elegir H_2 cuando H_1 es cierta;

M_1 es la probabilidad de encontrar evidencia fuerte a favor de H_2 cuando H_1 es cierta;

β es la probabilidad de fallar porque no elegimos H_2 cuando H_2 es cierta, y

$M_2 + W_2$ es la probabilidad de fallar porque no se ha encontrado evidencia fuerte a favor de H_2 cuando H_2 es cierta.

Observemos que con la aproximación que se está considerando, el resultado no es una decisión, no se sigue una acción de ellos, sino que tenemos unos datos y los interpretamos correctamente.

Si observamos $X = x$ interpretaremos estos datos como la evidencia que soporta a H_2 frente a H_1 a través del factor:

$$\frac{L_2}{L_1} = \exp \left\{ \left[\bar{x} - \left(\theta_1 + \frac{\delta}{2} \right) \right] n \frac{\delta}{\sigma^2} \right\},$$

- Si este número está entre $\frac{1}{k_1}$ y k_2 , nuestro intento es fallido porque la evidencia no es tan fuerte como buscábamos,

- Si es mayor que k_2 o menor que $\frac{1}{k_1}$ estaremos satisfechos porque hemos encontrado evidencia fuerte (que soporta a H_2 frente a H_1 o al revés) suficiente para nuestros objetivos.

Si la media muestral observada es ligeramente mayor que $\theta_1 + \frac{\delta}{2} + \sigma^2 \frac{\log 8}{n\delta}$, entonces tendremos fuerte evidencia que soporta a H_2 frente a H_1 . No tenemos posibilidad para saber si la evidencia es engañosa; conocemos que nuestra interpretación es correcta, es decir tenemos evidencia fuerte a favor de H_2 frente a H_1 y también sabemos que el procedimiento seguido no produce con frecuencia evidencia fuerte engañosa.

La diferencia crítica entre la aproximación seguida para generar evidencia y los contrastes de significación es que en la nueva aproximación se distingue entre fortaleza o intensidad de la evidencia, la cual se mide por el

cociente de verosimilitudes, y las probabilidades de que esa evidencia (de alguna intensidad especificada) sea generada bajo diversas condiciones.

Estas probabilidades^[11] pueden ser importantes para planificar un estudio, o para decidir si se debe de seguir adelante con otro, pero después de haber obtenido las observaciones ninguna de dichas probabilidades tiene nada que decir sobre la interpretación del conjunto de datos como evidencia.

La nueva aproximación y la antigua son completamente distintas en ese sentido, ya que en la antigua, la interpretación como evidencia, de un conjunto de datos dado, se hace en términos de las probabilidades de error, mientras que en ésta no es así.

En definitiva, probabilidades tales como la de obtener evidencia débil o evidencia engañosa son importantes en una fase de planificación, pero no deben de utilizarse para interpretar los resultados. Las medidas apropiadas son las razones de verosimilitudes y son irrelevantes las probabilidades de lo que ha sucedido o de lo que podía haber sucedido si las observaciones hubieran sido diferentes.

1.2. - Parámetros de perturbación.

De las notas anteriores se deduce que podemos dibujar la función de verosimilitud para un modelo cuya distribución está etiquetada por un parámetro unidimensional θ y literalmente vemos lo que dicen los datos. Se pueden observar hechos especiales tales como el valor en que la función alcanza un máximo, los límites del intervalo donde el cociente de verosimilitudes es superior a $1/8$, etc; en general, lo mejor para ver lo que los datos dicen es sencillamente examinar la función de verosimilitud entera.

^[11] Que en la discusión anterior hemos notado por M_1, M_2, W_1, W_2 .

Cuando el parámetro es multidimensional no es tan fácil examinar la función de verosimilitud. El significado es el mismo: la distribución identificada por el valor del parámetro θ_1 está mejor soportada que la distribución identificada por θ_2 por el factor

$$\frac{L(\theta_1)}{L(\theta_2)},$$

con independencia de la dimensión de θ . Pero cuando la dimensión aumenta, se hace más difícil visualizar y entender la función. Esto es especialmente complicado cuando originariamente se está interesado en un aspecto del modelo representado por un componente o por una función unidimensional del vector de parámetros.

A continuación dedicamos algunos comentarios a este punto y para ello es cómodo escribir el parámetro de la forma (θ, γ) , donde θ es la parte en la que actualmente estamos interesados y γ es el resto. Nuestro problema es representar e interpretar la evidencia sobre θ únicamente, pero la función de verosimilitud $L(\theta, \gamma)$, depende simultáneamente de las dos partes: θ y un parámetro de perturbación γ .

La razón de verosimilitudes

$$\frac{L(\theta_1, \gamma)}{L(\theta_2, \gamma)}$$

mide el soporte relativo para las dos distribuciones de probabilidad identificadas por (θ_1, γ) y (θ_2, γ) , o, equivalentemente, el soporte para el

valor del parámetro (θ_1, γ) frente al valor (θ_2, γ) . Para un valor fijo de γ esta razón mide el soporte para θ_1 frente a θ_2 , pero depende del valor de γ .

Necesitamos una función de verosimilitud para θ solamente, pero en general no podemos eliminar el parámetro γ . Esto se debe a que es el vector (θ, γ) y no solamente θ el que determina una distribución de probabilidad para la variable aleatoria observada y lo que la razón de verosimilitudes mide es el soporte para una distribución frente a otra. Un valor de θ no determina una distribución sino una familia de distribuciones (definida cuando γ varía).

No existe una solución general para este problema de parámetros de perturbación, no existe un camino general para eliminarlos. Pero si existen algunas soluciones parciales satisfactorias que brevemente enunciamos a continuación.

Parámetros ortogonales.

Cuando la función de verosimilitud tiene una forma de producto

$$L(\theta, \gamma) = L_1(\theta) L_2(\gamma),$$

los parámetros θ y γ se denominan ortogonales (Anscombe (1.964)) y el factor $L_1(\theta)$ es la función de verosimilitud para θ .

Algunas veces los parámetros no son ortogonales pero se puede alcanzar la ortogonalidad por una reparametrización del modelo, es decir, se puede ortogonalizar.

Verosimilitudes marginales.

Cuando una reparametrización ortogonal no es posible, hay otras soluciones para encontrar representaciones útiles de la evidencia sobre un parámetro de interés. Por ejemplo, consideremos un modelo $N(\mu, \sigma^2)$ y estamos interesados en la evidencia sobre σ^2 proporcionada por la muestra (x_1, \dots, x_n) . Los parámetros μ y σ^2 no son ortogonales y no es posible una reparametrización ortogonal exacta. Pero el estadístico

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

tiene una distribución de probabilidad marginal que depende de σ^2 y no depende de μ . Por lo tanto la observación $S^2 = s^2$ da evidencia sobre σ^2 , que es independiente de μ , y esa evidencia viene representada por la función de verosimilitud marginal,

$$L_M(\sigma^2) \propto (\sigma^2)^{-(n-1)/2} e^{-\frac{(n-1)s^2}{2\sigma^2}}.$$

Esta no es toda la evidencia sobre σ^2 en la muestra (x_1, \dots, x_n) pero en muchos sentidos es toda la evidencia que podemos desligar del parámetro μ sin hipótesis adicionales.

La función de verosimilitud marginal no es la función de verosimilitud para σ^2 en el sentido del factor de la verosimilitud ortogonal pero representa el mejor y el más útil resumen unidimensional de la evidencia sobre σ^2 en la muestra.

Pero no en todos los problemas en que una reparametrización ortogonal es imposible se dispone de verosimilitudes marginales; incluso cuando existen verosimilitudes marginales también hay otros candidatos a considerar para representar la evidencia sobre un parámetro simple de interés en presencia de parámetros de perturbación.

A continuación vemos algunos de ellos:

Verosimilitudes condicionadas.

En el caso anterior de la varianza de una distribución Normal hemos visto que la distribución de probabilidad marginal de una función de los datos adecuadamente elegida puede dar una evidencia neta sobre un parámetro simple de interés; es decir, evidencia que puede representarse e interpretarse independientemente de los parámetros de perturbación.

Algunas veces se pueden obtener resultados análogos a partir de modelos de probabilidad condicionados. Por ejemplo, supongamos que X e Y son variables aleatorias independientes con distribución Binomial;

$$X \rightarrow Bi(m; p_x); Y \rightarrow Bi(n; p_y)$$

donde m y n son conocidos.

Entonces la distribución de X condicionada a $X + Y$ depende sólo de la razón

$$\Psi = \frac{p_x(1-p_y)}{p_y(1-p_x)},$$

(ver Edwards (1.972), pág 192, para datos de esta naturaleza).

Verosimilitudes estimadas.

Otras dos técnicas para eliminar parámetros de perturbación de una función de verosimilitud consisten en sustituirlos por estimadores.

La primera de estas técnicas considera como objetivo a la función de verosimilitud con el parámetro de perturbación fijo en su verdadero valor,

$$L(\theta, \gamma_0);$$

y se estima esta función por

$$L(\theta, \hat{\gamma})$$

donde $\hat{\gamma}$ es un estimador consistente de γ_0 , como por ejemplo el estimador de máxima verosimilitud.

Con esto tenemos una verosimilitud estimada para θ ,

$$L_E(\theta) = L(\theta, \hat{\gamma})$$

La segunda técnica que también reemplaza al parámetro γ consiste en lo siguiente:

En vez de usar un estimador del verdadero valor, γ_0 , se usa para cada valor de θ el valor $\hat{\gamma}(\theta)$ que maximiza a $L(\theta, \gamma)$, (ver Kalbfleisch y Sprott (1.970)).

Esto da un “perfil” de verosimilitud,

$$L_p(\theta) = L(\theta, \hat{\gamma}(\theta)) = \max_{\gamma} \{L(\theta, \gamma)\}.$$

Un ejemplo importante de perfil de verosimilitud se presenta en el modelo $N(\theta, \sigma^2)$, cuando σ^2 es el parámetro de perturbación. En este caso, para un valor fijo de θ la verosimilitud alcanza un máximo en

$$\hat{\sigma}^2(\theta) = \frac{\sum_{i=1}^n (x_i - \theta)^2}{n}$$

y el perfil de verosimilitud es

$$L_p(\theta) = L(\theta, \hat{\sigma}^2(\theta)) \propto \left[\sum_{i=1}^n (x_i - \theta)^2 \right]^{-n/2}.$$

Una observación a favor del perfil de verosimilitud es que, si existe una transformación del parámetro que ortogonaliza, y por tanto existe una verosimilitud ortogonal para el parámetro de interés, entonces el perfil de verosimilitud es esa verosimilitud ortogonal. (ver Tsou (1.991)).

Verosimilitudes integradas (análisis bayesiano).

En esta aproximación la función de verosimilitud es proporcional a la densidad condicionada de los datos, dados los valores de los parámetros aleatorios (θ, γ) ; es decir, es proporcional a $f(x|\theta, \gamma)$. Entonces la función de verosimilitud de θ únicamente es proporcional a la densidad condicionada a θ solamente. Dentro del modelo bayesiano esto se obtiene simplemente integrando en el parámetro de perturbación; es decir, integrando

$$f(x, \gamma|\theta) = f(x|\theta, \gamma)g(\gamma|\theta),$$

donde $g(\gamma/\theta)$ es la densidad de probabilidad a priori de γ dado θ , sobre el rango de γ .

Desde un punto de vista frecuentista esto es equivalente a usar en vez del modelo original con dos parámetros $f(x/\theta, \gamma)$, el nuevo modelo de un parámetro $f(x/\theta)$, para la variable aleatoria observable X .

Observemos que sería apropiado escribir $f_g(x/\theta)$ para señalar la dependencia de la distribución a priori $g(\gamma/\theta)$. Además, el parámetro de perturbación se ha eliminado de la función de verosimilitud añadiendo otro elemento al modelo, la distribución de probabilidad a priori.

Posiblemente de todas las aproximaciones descritas la más prometedora es la de los perfiles de verosimilitud. Referencias fundamentales en esta aproximación son Kalbleisch y Sprott (1.970), Cox y Reid (1.987), Barndorff-Nielsen (1.986), McCullagh y Tibshirani (1.990)...

1.3. - Un perfil de verosimilitud: Verosimilitud empírica.

La noción de verosimilitud empírica que desarrollamos sucintamente más adelante, fue introducida por Owen (1.988; 1.990), quien en la introducción a sus trabajos realiza una excelente aproximación intuitiva que exponemos a continuación.

Sean x_1, \dots, x_n observaciones independientes de una función de distribución F_0 .

Es muy conocido que la función de distribución empírica

$$F_n = \frac{1}{n} \sum_{i=1}^n \delta_{x_i} \quad [12]$$

[12] Aquí δ_x denota la distribución degenerada en x , es decir en un punto de masa en x .

es un estimador de máxima verosimilitud^[13] no paramétrico de F_0 , basado en x_1, \dots, x_n , ya que maximiza la función de verosimilitud

$$L(F) = \prod_{i=1}^n F\{x_i\},$$

donde $F\{x_i\}$ es la probabilidad del conjunto $\{x_i\}$ bajo F , x_i es el valor observado de X_i y F es cualquier medida de probabilidad.

Se define la función razón de verosimilitud empírica como:

$$R(F) = \frac{L(F)}{L(F_n)}.$$

Supongamos que estamos interesados en $T(F_0)$, donde T es un funcional estadístico. El estimador de máxima verosimilitud no paramétrico de $T(F_0)$ es $T(F_n)$. Bajo algunas condiciones, conjuntos de la forma:

$$C = \{T(F) / R(F) \geq r\}^{[14]}, \text{ con } r \text{ constante,}$$

[13] Que F_n es un estimador de máxima verosimilitud de F_0 ya era bien conocido por Kiefer y Wolfowitz (1.956) o por Kaplan y Meier (1.958). El lector interesado puede ver también Johansen (1.978) o más recientemente Bailey (1.984) o Vardi (1.985).

[14] El primer uso de una función razón de verosimilitud empírica para construir intervalos de confianza aparece en Thomas y Grunkemeier (1.975), que utilizaron argumentos heurísticos para justificar la probabilidad de cobertura asintótica

pueden usarse para construir regiones de confianza y contrastes no paramétricos para $T(F_0)$.

Bajo tales condiciones un contraste para $T(F_0)=t$ rechazará cuando $t \notin C$, es decir, cuando ninguna distribución con $T(F)=t$ tiene una verosimilitud $L(F) \geq r \cdot L(F_n)$.

Claramente^[15]son necesarias algunas restricciones sobre F , porque si no para $r < 1$ se obtiene $C = \mathfrak{R}^p$. Este problema se resuelve restringiendo nuestra atención a distribuciones con soporte en la muestra, es decir a distribuciones $F \ll F_n$.^[16]

Para distribuciones $F \ll F_n$, la razón de verosimilitud $R(F)$ es la de una distribución Multinomial^[17] sobre la muestra. Puede comprobarse que la condición $F \ll F_n$ no es tan restrictiva como podría pensarse a primera vista.

La verosimilitud empírica es una técnica que consiste fundamentalmente en calcular el perfil de verosimilitud de una distribución general Multinomial cuya masa de probabilidad se encuentra localizada en datos puntuales.^[18]

Fue introducida por Owen (1.988) para construir intervalos de confianza, subrayando la posibilidad de que esta metodología fuese aplicada en numerosos contextos tales como los modelos aleatorios, modelos

^[15] Aquí solo exponemos los grandes pasos de la argumentación, para detalles ver Owen (1.990) pág 91 y 92.

^[16] Distribuciones absolutamente continuas respecto a F_n .

^[17] Algunos resultados sobre verosimilitudes Multinomiales pueden verse en Hoeffding (1.965), Tusnady (1.977) o Berk y Jones (1.979).

^[18] La verosimilitud empírica tiene paralelos en la literatura bootstrap. Ver por ejemplo, el bootstrap bayesiano de Rubin (1.981), (genera distribuciones empíricas ponderadas $\sum g_i \delta_{x_i}$,

donde los g_i son variables aleatorias positivas con suma la unidad) o el bootstrap no paramétrico de Efron (1.981), que usa la misma familia de distribuciones Multinomiales como verosimilitud empírica.

autorregresivos, modelos de regresión,...Owen presentó la verosimilitud empírica como una alternativa a los métodos bootstrap de verosimilitud propuestos por Hall (1.987).

La verosimilitud empírica presenta una serie de ventajas en relación con los métodos clásicos de estimación multiparamétrica. Un problema de los métodos clásicos es el conjunto de hipótesis que se requieren acerca de ,

- la distribución de la población de la que se obtiene la muestra, y
- la distribución muestral planteada.

Estas hipótesis pueden ser difíciles de justificar.

La verosimilitud empírica sugiere un modelo para un conjunto de parámetros en el que no se requieren hipótesis de partida sobre la distribución de la población de la que procede la muestra, ni de la forma de la función de verosimilitud ni de la correlación entre variables. La verosimilitud y sus regiones son una función sólo de los datos muestrales.

Veamos que entendemos por la verosimilitud empírica.

Para ello nos centraremos en regiones de confianza para parámetros que se pueden expresar como funciones suaves de la media, es decir, funciones dependientes de la media que son funciones derivables y con derivada continua. Esta aproximación nos permitirá comentarios posteriores.

Consideramos la siguiente nomenclatura:

- (i) Sea $\tilde{\theta}$ alguna característica de la población, por ejemplo la media, sobre la que deseamos realizar alguna inferencia estadística, por ejemplo una región de confianza.
- (ii) Sea $\vec{X}_1, \dots, \vec{X}_n$ una muestra aleatoria simple de tamaño n , que puede ser vectores; independientes y con función de distribución común F , por supuesto desconocida.

- (iii) Sea $\vec{p} = (p_1, \dots, p_n)$, un vector n-dimensional tal que, $p_i \geq 0$ para $i = 1, \dots, n$; y $\sum_{i=1}^n p_i = 1$.
- (iv) Sea $\vec{\theta}(\vec{p})$ el valor supuesto para el parámetro $\vec{\theta}$ cuando la población es discreta con probabilidad p_i en el punto \vec{X}_i . Por ejemplo, si $\vec{\theta}$ es la media poblacional, entonces
- $$\vec{\theta}(\vec{p}) = \sum_{i=1}^n p_i \vec{X}_i$$

Podemos suponer, en general, que $\vec{\theta}$ tiene dimensión r y los \vec{X}_i son vectores de dimensión s.

Podemos ahora definir la verosimilitud empírica de la siguiente forma:

Definición 1. (Verosimilitud empírica).

La verosimilitud empírica para $\vec{\theta}(\vec{p})$, evaluada en un valor concreto $\vec{\theta} = \vec{\theta}_1$, se define como el máximo en los p_i del producto de los p_i sujeto a las restricciones que se indican, es decir,

$$L(\vec{\theta}_1) = \text{Max}_{\vec{p}} \prod_{i=1}^n p_i \text{ sujeto a } \left[0 \leq p_i \leq 1, i = 1, \dots, n; \sum_{i=1}^n p_i = 1; \vec{\theta}(\vec{p}) = \vec{\theta}_1 \right].$$

Por tanto, la verosimilitud empírica es una verosimilitud de perfil multinomial y el número de parámetros es uno menos que el tamaño muestral. El objetivo es obtener el valor de la verosimilitud empírica del parámetro de interés.

Como un caso particular observemos que si en la definición de la verosimilitud empírica se considera sólo la restricciones

$$0 \leq p_i \leq 1 \text{ para } i = 1, \dots, n; \text{ y } \sum_{i=1}^n p_i = 1,$$

entonces la función objetivo se maximiza cuando

$$p_i = \frac{1}{n}, i = 1, \dots, n;$$

y el máximo que define la verosimilitud empírica sería:

$$L(\bar{\theta}) = \frac{1}{n^n}, \text{ que notaremos por } L(\hat{\bar{\theta}}).$$

Otras definiciones útiles son las siguientes.

Definición 2. (Razón de verosimilitud empírica).

La razón de verosimilitud empírica para $\bar{\theta}(\bar{p})$, evaluada en $\bar{\theta} = \bar{\theta}_1$, es el cociente entre la verosimilitud empírica evaluada en $\bar{\theta}_1$ y el máximo de la verosimilitud Multinomial sin restricciones, es decir,

$$R(\bar{\theta}_1) = \frac{L(\bar{\theta}_1)}{L(\hat{\theta})} = \frac{L(\bar{\theta}_1)}{n^{-n}} = \underset{\bar{p}}{\text{Max}} \prod_{i=1}^n np_i$$

sujeto a $\left[0 \leq p_i \leq 1, i = 1, \dots, n; \sum_{i=1}^n p_i = 1; \bar{\theta}(\bar{p}) = \bar{\theta}_1 \right]$.

Definición 3. (Logaritmo de la razón de verosimilitud empírica).

El logaritmo de la razón de verosimilitud empírica viene dado por:

$$l_E(\bar{\theta}_1) = -2 \log [R(\bar{\theta}_1)] = -2 \log \left\{ \underset{\bar{p}}{\text{Max}} \prod_{i=1}^n np_i \right\}$$

sujeto a $\left[0 \leq p_i \leq 1, i = 1, \dots, n; \sum_{i=1}^n p_i = 1; \bar{\theta}(\bar{p}) = \bar{\theta}_1 \right]$.

A continuación, veremos algunos resultados importantes en el escenario de la verosimilitud empírica.

Lema 1. (sobre el problema de optimización en la verosimilitud empírica).

Los valores de $p_i, i = 1, \dots, n$, en los que se alcanza el máximo que aparece en la definición de la función de verosimilitud empírica de la media evaluada en $\bar{\theta} = \bar{\theta}_1$ son:

$$p_i = \frac{1}{n(1 + \bar{t}^{nr} (\bar{X}_i - \bar{\theta}_1))}, \quad i = 1, \dots, n;$$

donde \bar{t} es un vector que depende de $\bar{\theta}_1$ y que satisface,

$$\sum_{i=1}^n p_i (\bar{X}_i - \bar{\theta}_1) = \sum_{i=1}^n \frac{(\bar{X}_i - \bar{\theta}_1)}{n(1 + \bar{t}^r (\bar{X}_i - \bar{\theta}_1))} = \bar{\theta} \quad (c1)$$

□

Por tanto, la función de verosimilitud empírica es,

$$L(\bar{\theta}_1) = \frac{1}{n^n} \prod_{i=1}^n \frac{1}{1 + \bar{t}^r (\bar{X}_i - \bar{\theta}_1)},$$

donde \bar{t} satisface la condición (c1).

El logaritmo de la razón de verosimilitud empírica queda:

$$l_E(\bar{\theta}_1) = 2 \sum_{i=1}^n \log(1 + \bar{t}^r (\bar{X}_i - \bar{\theta}_1));$$

donde \bar{t} satisface la condición (c1).

En Estadística cuando se utiliza la razón de verosimilitud usual, es decir, asociada a una verosimilitud paramétrica, es necesario determinar su distribución para que pueda ser utilizada en la construcción de intervalos de confianza y en el diseño de contrastes de hipótesis. En diversos casos es posible encontrar la distribución exacta de dicha razón de verosimilitudes, por ejemplo, en problemas asociados a la distribución Normal. Sin embargo, en otros muchos casos no es posible esa determinación exacta de la distribución de probabilidad de la razón de verosimilitudes y es necesario acudir a su distribución asintótica para utilizarla como una aproximación.

Para la razón de verosimilitudes habitual este propuesto fue resuelto por el Teorema de Wilks, probando que bajo amplias condiciones la distribución asintótica es una distribución χ^2 de Pearson. (ver Wilks(1.938) o Wilks (1.962), pág 419 y sgtes).

En el caso de la verosimilitud empírica también es posible enunciar un resultado sobre la distribución asintótica de la razón de verosimilitud empírica.

De nuevo se obtiene como distribución asintótica la distribución χ^2 de Pearson (ver Hall y La Scala (1.990)).

Este resultado permite construir regiones de confianza con un nivel de confianza aproximado. Sobre este aspecto DiCiccio, Hall y Romano (1.989) probaron que el error de cobertura^[19] es de orden $\frac{1}{n}$, siendo n el tamaño de la muestra; además, es posible efectuar correcciones^[20] sencillas que reducen el error de cobertura a órdenes $\frac{1}{n^2}$.

En definitiva, la versión para verosimilitud empírica del teorema de Wilks, demostrado por Hall y la Scala (1.990), establece que bajo un conjunto amplio de condiciones se verifica que,

$$[21] l_E(\bar{\theta}_0) = -2 \log [R(\bar{\theta}_0)] = -2 \log \frac{L(\bar{\theta}_0)}{L(\hat{\theta})} \xrightarrow{n \uparrow \infty} \chi^2(t).$$

Es decir, dado un número c positivo y menor que 1, se verifica que:

[19] Diferencia entre el nivel de confianza nominal y el real.

[20] Lo que se denomina una corrección de Bartlett; correcciones análogas para verosimilitud paramétrica puede verse en Bandorff-Nielsen y Cox (1.984).

[21] No es importante para nosotros especificar minuciosamente el conjunto de condiciones bajo las que este resultado es válido. Tampoco necesitamos ser más explícitos sobre la determinación del número de grados de libertad, t ; para más detalles ver el trabajo citado de Hall y La Scala (1.990)

$$\lim_{n \rightarrow \infty} \text{Prob} \left[l_E(\bar{\theta}_0) \leq c \right] = \text{Prob} \left[\chi^2(t) \leq c \right];$$

por tanto, si c es un valor tal que:

$$\text{Prob} \left[\chi^2(t) \leq c \right] = 1 - \alpha, \text{ con } \alpha \in [0, 1]$$

y

$$RC_c = \left\{ \bar{\theta} : l_E(\bar{\theta}) \leq c \right\}$$

entonces, RC_c es una región de confianza de nivel de confianza asintótica $1 - \alpha$:

$$\lim_{n \rightarrow \infty} \left(\text{Prob} \left[\bar{\theta}_0 \in RC_c \right] \right) = 1 - \alpha.$$

Sobre el resultado anterior Owen (1.989) presentó versiones previas para casos simplificados y Owen (1.990) presentó resultados adicionales.

En un escenario de Estadística clásica (o frecuentista) y no paramétrica son muchos y muy diversos los trabajos que se han publicado sobre la noción de verosimilitud empírica. No es objeto de esta Memoria una recopilación completa y exhaustiva de los mismos dada la diferente utilización que haremos de la misma, pero referenciamos a continuación algunos de estos trabajos para el lector interesado.

DiCiccio, Hall y Romano (1.989) intentan comparar las funciones de verosimilitud empírica y paramétrica, donde el parámetro de interés es una función suave del vector de medias. La respuesta de los autores es que la superficie de verosimilitud obtenida por verosimilitud empírica, aproxima en cierto sentido una superficie de verosimilitud paramétrica, y que esta familia paramétrica artificial construida por ellos no contiene más información, en el sentido de Kullback- Leibler, que la disponible en el escenario no paramétrico.

Chen y Hall (1.993) estudian la construcción de regiones de confianza para cuantiles, analizando la precisión de cobertura.

Chen y Qin (1.993) estudian el caso de poblaciones finitas y el uso de información auxiliar.

Zhang (1.996) construye intervalos de confianza para una función de distribución en presencia de información auxiliar.

Owen (1.991) y Chen (1.993) utilizan la noción de verosimilitud empírica en el modelo lineal.

Monti (1.997) utiliza la noción de verosimilitud empírica en modelos de series temporales.

1.4. - Verosimilitudes de perfil multinomial en un análisis bayesiano: Verosimilitud modificada.

En el apartado anterior se ha desarrollado brevemente una verosimilitud de perfil multinomial, denominada verosimilitud empírica, como una metodología no paramétrica para realizar inferencias estadísticas, fundamentalmente regiones de confianza y contrastes de hipótesis. El punto de vista adoptado ha sido fundamentalmente frecuentista y preexperimental, es decir, la bondad de un procedimiento se ha evaluado examinando el comportamiento del mismo frente a todos los posibles resultados muestrales. Las herramientas principales han sido las diversas modificaciones del teorema de Wilks que nos asegura una distribución asintótica χ^2 de Pearson para el estadístico de interés. Estos resultados junto a los obtenidos sobre precisión de

cobertura permiten la obtención de conclusiones aproximadas (aunque muy afinadas) para situación de muestras necesariamente finitas.

En este apartado queremos adoptar un punto de vista distinto. Nos planteamos la utilización de verosimilitudes de perfil multinomial en un problema de análisis bayesiano; por tanto, nuestro planteamiento será condicional y postexperimental.

Ya en el apartado anterior citamos alguna utilización de verosimilitudes de perfil multinomial en análisis bayesiano (por ejemplo Rubin (1.981)), pero ahora vamos a desarrollar el tema que nos ocupa comenzando en el punto en que surgió nuestro interés por el mismo; concretando una utilización de métodos estadísticos en un problema económico-empresarial como es la Auditoría de cuentas.

En el problema de determinación de cotas (regiones de confianza unilaterales) para el error de contabilidad en una investigación de auditoría es ya antigua la utilización de la distribución Multinomial. Desde un punto de vista no bayesiano el uso de la Multinomial se inició con Fienberg, Neter y Leitch (1.977); Neter, Leitch y Fienberg (1.978) y con Leitch, Neter, Plante y Sinha (1.982).

En un planteamiento bayesiano la distribución Multinomial se ha utilizado en Auditoría de cuentas en trabajos como los de Tsui, Matsumura y Tsui (1.985); Matsumura, Tsui y Wong (1.987) o Byekwaso (1.994).

En estos trabajos se plantea un problema con la distribución a priori, porque la alta dimensión del parámetro hace muy difícil, en la práctica, la especificación de la información relevante para el auditor en la forma de una distribución de probabilidad. Como consecuencia se hace una elección automatizada de la distribución de Dirichlet^[22] como distribución a priori.

Exponemos a continuación otra posibilidad que se ha desarrollado en la literatura para la utilización de la distribución Multinomial, intentando evitar el problema de especificar una distribución a priori de tan alta dimensión, y que es la alternativa que nos interesa.

[22] La distribución conjugada natural de la verosimilitud Multinomial.

Se considera una verosimilitud Multinomial, por tanto dependiente de un parámetro k -dimensional, y se intenta aprovechar la información relevante a priori en la forma en que está disponible; a saber, el experto tiene información sobre el parámetro “cantidad de error” que es una combinación lineal de los parámetros de la Multinomial (es decir, el parámetro sobre el que realmente se tiene información es un subparámetro).

El problema surge en que la verosimilitud Multinomial y la distribución a priori relevante no son combinables a través del teorema de Bayes. Para ello es necesario “modificar” la verosimilitud de una manera coherente de tal forma que la “verosimilitud modificada” sea compatible, vía teorema de Bayes, con la información a priori relevante disponible y se pueda obtener la distribución a posteriori (objetivo de un análisis bayesiano).

La nomenclatura verosimilitud modificada que hemos usado en el párrafo anterior ha surgido por diversos caminos, en principio independientes:

En un contexto de información a priori parcial definida sobre un sigma álgebra fue introducida y estudiada en Cano Sánchez, Hernández Bastida y Moreno Bas (1.985; 1.987 (a), (b) y (c)).

En una aproximación heurística directamente dirigida al problema de Auditoría de cuentas se inició en McCray (1.984), posteriormente ampliado en McCray (1.986; 1998) o en McCray, Steele y Hernández Bastida (1.999).

La unión de las líneas anteriores y su utilización extensiva en problemas de Auditoría de cuentas se produce en los trabajos Hernández Bastida y Vázquez Polo (1.997); Hernández Bastida, Moreno Carretero y Vázquez Polo (1.997) y Hernández Bastida, Martel Escobar y Vázquez Polo (1.998).

A continuación se describe la idea de los trabajos anteriores. La descripción es deliberadamente intuitiva y omitiendo desarrollos técnicos que están suficientemente tratados en los artículos referenciados. Además consideramos discreto al espacio paramétrico porque ello será suficiente para las aplicaciones prácticas que nos interesarán posteriormente aunque muchos de los resultados técnicos implícitos son válidos para espacios continuos.

Consideremos una situación en la que pueden establecerse un número determinado de categorías T_1, T_2, \dots, T_k ; con probabilidades asociadas p_1, \dots, p_k desconocidas y verificando

$$p_i \in [0, 1], i = 1, \dots, k; \sum_{i=1}^k p_i = 1.$$

Notaremos por $X_i, i = 1, \dots, k$, al representante de la categoría i -ésima y por $\vec{p} = (p_1, \dots, p_k)$ al vector de probabilidades, en el que cada uno de los p_i representa la probabilidad de que una realización concreta pertenezca a la i -ésima categoría. Claramente \vec{p} es el parámetro, de dimensión k , del problema y un análisis bayesiano directo exige una distribución a priori para \vec{p} .

Supongamos que estamos interesados en una magnitud, que denominamos estado de la naturaleza y notamos por θ , que pertenece a un conjunto discreto de posibles estados de la naturaleza y que dicha magnitud es una combinación lineal de los parámetros del problema,^[23] a saber,

$$\theta = \sum_{i=1}^k X_i p_i.$$

La magnitud de interés (sobre la que deseamos realizar inferencias) y además la magnitud intuitiva para el usuario es θ . Por ello, lo natural es que el

[23] En el problema de estimación del error de una contabilidad en Auditoría, cada categoría T_i es una categoría de fallo (taint) en un muestreo de unidades monetarias, que viene caracterizada por el número de céntimos erróneos en una unidad monetaria. Por tanto, para errores de sobrevaloración y de subvaloración los X_i son los i céntimos erróneos, con $i = -100, -99, \dots, 0, 1, \dots, 99, 100$, y la cantidad de error (magnitud de interés en el problema e intuitiva para el auditor) es $RBV \sum_{i=-100}^{100} \frac{ip_i}{100}$ donde RBV es el valor registrado total.

usuario tenga información relevante sobre θ , y sobre θ podrá especificar una distribución a priori que notamos por $\xi(\theta)$.

Si se toma una muestra de tamaño n y notamos por n_i el número de elementos que pertenecen a la categoría i -ésima, donde

$$0 \leq n_i \leq n; i = 1, \dots, k; \sum_{i=1}^k n_i = n,$$

la verosimilitud es Multinomial:

$$L(\vec{p}) \equiv L(p_1, \dots, p_k) = \frac{n!}{\prod_{i=1}^k n_i!} \prod_{i=1}^k p_i^{n_i}.$$

Ya hemos comentado antes que la información a priori disponible es $\xi(\theta)$ y que por razones prácticas no es posible disponer de una distribución a priori de dimensión k para \vec{p} ^[24].

Una solución ante la imposibilidad de combinar directamente la verosimilitud con la distribución a priori, consiste en modificar coherentemente la verosimilitud original del problema, de manera que la nueva verosimilitud modificada sea compatible con la distribución a priori $\xi(\theta)$.

^[24] Observemos que la información a priori disponible es más pobre que la necesaria para aplicar directamente el teorema de Bayes; por ello anteriormente se hablaba de información a priori parcial.

Definición 4. ^[25] (Verosimilitud modificada).

La función definida por:

$$L(\theta) = \underset{\vec{p}}{\text{Max}} \prod_{i=1}^k p_i, \text{ sujeto a } \left[\begin{array}{l} p_i \in [0,1], i=1, \dots, k; \\ \sum_{i=1}^k p_i = 1 \\ \sum_{i=1}^k X_i p_i = \theta \end{array} \right]$$

se denomina verosimilitud modificada. ^[26]

Intuitivamente, la verosimilitud modificada hace indistinguibles, es decir, identifica entre sí, a todas aquellas k -uplas (p_1, \dots, p_k) que producen un mismo valor de θ , porque la distribución a priori no distingue por debajo de θ . La identificación se hace tomando el superior porque no es posible promediar integrando (sin más formalismo teórico) ya que no se dispone de una medida que permita esa integración.

^[25] Exponemos en esta definición la versión que más nos interesa para nuestros objetivos. Comparando con versiones desarrolladas en los trabajos referenciados ésta es una versión particularizada.

^[26] En los distintos trabajos citados esta función $L(\theta)$ ha recibido diversos nombres, en particular el profesor McCray suele denominarla “verosimilitud más probable” o nombres parecidos.

Definición 5. (Distribución a posteriori con la verosimilitud modificada).

Se denomina distribución a posteriori con la verosimilitud modificada a la expresión habitual, es decir,

$$\xi(\theta / \text{datos}) = \frac{L(\theta)\xi(\theta)}{\int_{\Theta} L(\theta) d\xi(\theta)}.$$

Si comparamos las dos definiciones anteriores con la definición de verosimilitud empírica expuesta en el epígrafe anterior se observa lo siguiente:

- 1.- La definición de verosimilitud modificada es idéntica a la de verosimilitud empírica.
- 2.- Si en la definición de distribución a posteriori con la verosimilitud modificada se considera una distribución a posteriori Uniforme, se obtiene una normalización de la verosimilitud modificada que proporciona una distribución de probabilidad para el parámetro θ .

Por tanto, aunque los puntos de vista sean distintos, en el epígrafe anterior y en éste estamos considerando exactamente la misma expresión formal para la verosimilitud de perfil multinomial.

Esta es la metodología estadística que va a ser utilizada en los siguientes capítulos. Haremos uso de esta técnica para analizar el consumo en unidades físicas, así como los ingresos y gastos de los hogares españoles, expresados en unidades monetarias, y otros estudios.

CAPÍTULO 2

ANÁLISIS DE CANTIDADES COMPRADAS, EXPRESADAS EN UNIDADES FÍSICAS, PARA ALGUNOS ARTÍCULOS DE ALIMENTACIÓN DE LOS HOGARES ESPAÑOLES, SEGÚN LA ENCUESTA CONTINUA DE PRESUPUESTOS FAMILIARES.

2.1. – Introducción y planteamiento.

Es bastante usual pedir a los economistas su opinión acerca de una gran variedad de temas, tales como previsiones de desempleo, un proyecto de inversión, la privatización de una parte de la industria, la eficiencia de la sanidad pública...

Para ayudar a resolver problemas, los economistas deben de ser capaces de comprender como funciona la economía. Así, al menos parcialmente, es una disciplina empírica basada en la observación del mundo real. Pero también es una disciplina teórica. Los economistas desarrollan teorías sobre el comportamiento fundadas en algunos supuestos. Por ejemplo, sabemos que los consumidores tienden a comprar menos cantidades de bienes

y servicios cuando sus precios aumentan o cuando la renta de los consumidores disminuye.

Los economistas desarrollan estas teorías y construyen modelos persiguiendo principalmente un objetivo: conocer lo mejor posible el fenómeno en cuestión.

Una manera de estudiar y conocer la realidad ha sido mediante el uso de modelos. Un modelo no es más que una representación simplificada de la realidad.

Puesto que no es sencillo conocer los fenómenos económicos en su totalidad tratan de desentrañarlos, dividiéndolos para que resulte más sencilla la tarea.

¿Qué se suele hacer? Veamos un ejemplo.

Supongamos que estamos interesados en un fenómeno concreto como puede ser el consumo. Esta no es una variable observable directamente, así que acudiremos a otras.

Cada vez más en los últimos años han ido apareciendo distintas fuentes de suministro de datos a nivel microeconómico, fundamentalmente gracias a distintas encuestas. Vamos a utilizar una de esas encuestas, la Encuesta Continua de Presupuestos Familiares.

Elegida la variable que se quiere estudiar, por ejemplo, la cantidad comprada de un bien en unidades físicas o el gasto en dicho bien, hay que buscar datos referidos a ella. A continuación, habría que escoger, utilizando los conocimientos que nos proporciona la Teoría Económica, las variables que pueden ser relevantes y que pueden influir en nuestra variable que hemos denominado “cantidad comprada del artículo x” o “gasto en el artículo”, como una posible aproximación al consumo.

Este es un tema muy importante porque no siempre resulta sencilla la selección de variables. Podría pensarse en una primera aproximación en variables como:

- a) La renta de la persona que adquiere el bien;

- b) El precio del artículo considerado;
- c) El precio de bienes sustitutivos del anterior,

u otras muchas variables, como cantidad comprada en un período anterior, precio o cantidad de bienes complementarios, período del año en que nos encontramos...

Es posible que nos planteemos que puede influir en la cantidad comprada de un bien el hecho de vivir en una región concreta, el sexo de la persona que lo adquiere, la edad..., y nos encontremos en la situación de tener que “cuantificar” variables que en principio no lo son, mediante el uso de variables ficticias, o que no exista una variable que recoja exactamente lo que pretendemos medir y haya que acudir a otras lo más semejantes posible.

Por tanto en los primeros estadios hay que decidir la variables o variables a explicar y cuales van a ser las explicativas^[27]. No debemos olvidar que el modelo no debe de tener un número excesivamente grande de variables explicativas (lo cual necesariamente nos exige un esfuerzo de elección) y que en numerosas ocasiones la selección definitiva de variables no depende sólo de la que establece la Teoría Económica, sino que viene determinada por la disponibilidad de datos. Es necesario tener una muestra suficientemente grande para todas las variables.

El siguiente paso importante es especificar la relación que liga las variables que hemos escogido. Es el momento de ver que forma funcional es la más adecuada, el número de ecuaciones que va a tener el modelo.... De nuevo en esto nos ayuda la Teoría Económica, pero sin perder de vista que los parámetros han de ser estimados, estimación que puede resultar compleja en diversas situaciones.

Tampoco hay que pasar por alto que todos los modelos llevan aparejados un término de perturbación aleatorio, sobre el que hay que especificar una serie de hipótesis que van desde cómo se introduce en el modelo (aditiva o multiplicativamente) a sus principales características estocásticas, que a su vez influirán en la técnica de estimación que se utilice.

[27] Aunque sería más correcto la denominación de variables endógenas y predeterminadas, vamos a utilizar, en principio, los términos explicada y explicativa.

Escogido el método se hace indispensable un programa informático que nos ayude en la estimación de los parámetros de interés.

Una vez obtenidas las estimaciones se puede comprobar si las variables son o no importantes a la hora de explicar la variable objeto de estudio, si se corresponden los signos con la Teoría Económica, o bien utilizar el modelo para elegir entre alternativas,...., necesitando la mayoría de estas posibilidades de supuestos adicionales relacionados con la distribución que siguen determinados estadísticos. También se podrá predecir la cantidad comprada futura, o el gasto futuro en ese bien sabiendo los valores futuros de las variables explicativas, predicciones que van unidas a algunas hipótesis que también hay que tener en cuenta.

En grandes pinceladas estos son los pasos principales desarrollados por la Econometría para encontrar y explotar un modelo que nos explique lo mejor posible los factores que influyen en nuestro fenómeno, y nos permite además “jugar” con el modelo para ver como se vería afectada la variable ante modificaciones en las variables explicativas.

Estos pasos suponen esfuerzo y toda modelización es susceptible de ser criticada en función de las hipótesis especificadas, y de ser mejorada en el sentido de encontrar otras variables o ensayar formas funcionales distintas que se ajusten mejor a las observaciones, con el propósito de lograr lo que al principio comentábamos: comprender y conocer mejor el fenómeno.

Y ese es también el objetivo que nosotros vamos a perseguir, representar lo mejor posible el fenómeno en cuestión. La metodología que proponemos ha sido comentada minuciosamente en el capítulo anterior, y nos va a ayudar a obtener una información precisa y muy interesante que nos permita representar qué está ocurriendo.

Se podrán extraer conclusiones a partir de los resultados obtenidos y las comparaciones que se puedan establecer y que arrojan luz sobre el fenómeno. Y todo ello sin variables explicativas, ni escoger formas funcionales o métodos de estimación.

Es una técnica sencilla y muy versátil, ya que puede aplicarse en contextos diversos, con una gran economía de medios técnicos y conceptuales. Lo único que necesita son datos referidos al fenómeno que vamos a

representar y un programa informático válido para cualquier ordenador personal.

Los datos han sido los suministrados por la ECPF. Son datos desagregados sobre muchas variables relacionadas con los hogares españoles. La ECPF recoge trimestralmente los gastos de las familias, en alimentación y en otros tipos de gastos, y además proporciona información sobre la cantidad comprada en unidades físicas de algunos artículos de alimentación. En el momento de la elaboración de este trabajo, sólo se disponía de datos referentes a 32 artículos,^[28] pero tras la reforma de la ECPF ese número ha sido ampliado a todos los alimentos para los que se conoce su gasto.

De los 32 artículos se han seleccionado tres grupos por considerarlos significativos o porque podían arrojar resultados muy interesantes sobre su consumo:

- .- El grupo de las carnes, que incluye carne de cerdo, de pollo y de ternera;
- .- El grupo de aceites: de girasol y de oliva;
- .- El grupo de tabaco: negro y rubio en la modalidad de cigarrillos.

Se dispone por tanto de observaciones sobre la cantidad comprada semanalmente por los hogares de estos artículos, medidas en unidades físicas: kilos para las carnes, litros para los aceites y cajetillas para el tabaco.

Existe en la literatura abundantes referencias al tratamiento de los microdatos (datos referidos a familias o individuos) en el tema del consumo y de la demanda (ver Blundell (1.988); Blundell, Pashardes y Weber (1.993); López Nicolás (1.995); Ramajo Hernández (1.995)). Estos autores ponen de manifiesto una característica que presentan estas encuestas y que es la presencia generalizada de muchos ceros en las respuestas de las familias. Son debidos fundamentalmente a tres causas:

- .- Que no se haya comprado ese bien en el período en que se lleva a cabo la encuesta.
- .- Que no se consuma ese bien.

^[28] En el Anexo 1 pueden consultarse los artículos que recoge la ECPF.

- Que el consumidor, dada su restricción presupuestaria, maximice su utilidad no comprando dicho bien.

La presencia de tantos ceros invalida los métodos de estimación clásicos (mínimos cuadrados ordinarios, método de regresiones aparentemente no relacionadas) provocando estimadores sesgados e inconsistentes (Maddala (1.983); Amemiya (1.984)).

Se han hecho grandes esfuerzos para llevar a cabo el tratamiento econométrico adecuado a estas situaciones. Deaton e Irish (1.984), Blundell y Meghir (1.987), Atkinson y otros (1.989), Labeaga (1.991), García y Labeaga (1.991), Jones (1.989), Pudney (1.989), y Meghir y Robin (1.992), proporcionan formas para contrastar las distintas causas.

Nuestra metodología no trata de conocer cual es el origen de esos ceros, limitándonos a utilizar los datos tal y como nos han sido suministrados, sin incurrir en tratamientos diferenciados en función de esas posibles causas.

Como se ha comentado anteriormente, esta técnica no utiliza una modelización econométrica con variables explicativas. Indudablemente, no se va a poder dar una explicación del fenómeno a partir de otras magnitudes, ni llevar a cabo simulación de políticas alternativas, o estimar la elasticidad-precio o elasticidad-renta..., si bien supone la ventaja de no tener que especificarlas, obtener datos sobre ellas, ni buscar la forma funcional más adecuada.

Es cierto que no podemos obtener los resultados propios de una modelización econométrica, pero eso no hace menos todo el conocimiento que sobre el fenómeno que nos preocupa, podemos obtener con esta metodología.

¿Qué podemos conocer con esta metodología acerca del fenómeno cantidad comprada en unidades físicas de un artículo?

(1) Podemos conocer cuál va a ser la **cantidad media más probable** que se compre de carne de cerdo, de pollo o de ternera. Sabremos, por ejemplo, que lo más probable que se compre de carne de cerdo es 0,69 kilos por semana, o que en cuanto a pollo, lo más probable es 1,15 kilos por semana.

Al igual podemos hacer con los aceites. 1,1 litros por semana de aceite de oliva es la cantidad media más probable de compra, mientras que para el aceite de girasol, sólo será 0,5 litros.

(2) Podemos obtener un **rango de valores** en los que existe una probabilidad, tan grande como queramos, de que se encuentre la cantidad media a comprar del artículo. Así el número medio de cajetillas de tabaco rubio compradas por los hogares españoles está comprendido entre 2,42 y 2,78 con una probabilidad del 95%, o hay una probabilidad del 80% de que se compre en tabaco negro entre 1,43 y 1,62 cajetillas.

Para la carne de ternera hablamos de una probabilidad del 90% de que se compre entre 0,48 y 0,54 kilos por semana.

(3) Podemos ofrecer **otro rango de valores**, también con una probabilidad elevada, que presentan una evidencia grande de compra.

Hay una fuerte evidencia de compra para los valores comprendidos entre 1,09 y 1,21 kilos por semana, para la carne de pollo. Y además podemos saber cuantas veces es más verosímil que se presente un valor que otro.

Hay casi 2 veces más evidencia a favor de comprar 1,12 kilos de pollo que 1,19. Es 518 veces más verosímil que se compre 1,13 que 1,24 kilos, o 3.230 veces más que se compre 1,16 kilos que 1,25. Nos da un instrumento que nos permite comparar entre valores para saber cual de ellos es más verosímil que se produzca.

Para conocer otros resultados de las posibilidades que hemos comentado para otros artículos, se pueden consultar los cálculos que se adjuntan un poco más adelante referidos a los mencionados artículos.

(4) Esta metodología también facilita un mecanismo que nos permite **escoger entre alternativas**, rechazando la que tiene menor probabilidad. No sería muy difícil elegir entre que la compra media de aceite de oliva estuviese entre 0,75 litros y 1,5 o entre 1,75 y 2,25 litros, si sabemos que una es más probable que otra.

(5) Nos permite también, de una forma sencilla, ver la **evolución** que un artículo determinado ha ido experimentando a lo largo del tiempo. Si se dispone de datos sobre la cantidad de carne de ternera podrá verse si el valor

más probable ha ido o no cambiando, si ha crecido mucho o poco, y en definitiva saber que ha ocurrido con la adquisición de ese artículo. Y todo eso sin la complejidad de modelos de series de tiempo o uso de series cronológicas.

(6) Establece una forma muy útil de **comparación** entre las cantidades compradas de ese artículo **según** las **características sociodemográficas** que puedan presentar los hogares.

Podríamos saber si en Andalucía se compra más carne de pollo que en Cataluña, o si los hogares de la provincia de Madrid con hijos en edad escolar compran menos aceite de girasol que ese tipo de hogares en Salamanca.

¿Son más fumadores en Palencia que en Córdoba? ¿Prefieren el pollo los hogares con mayoría de presencia femenina en Extremadura?. Estas y otras preguntas se pueden responder fácilmente aportando además otros muchos datos comparativos.

Lo único que es necesario son los datos que pueda suministrar una encuesta oficial o bien la realización de una específica para el caso concreto.

(7) Posibilita la **comparación** ente **artículos diferentes** que nos permite hacernos una idea de cómo se compra en los hogares españoles, cuáles pueden ser los artículos preferidos o al menos, más consumidos.

De los tres tipos de carnes hay una que se muestra como la preferida por los hogares españoles: la carne de pollo. Su adquisición es muy superior a la carne de cerdo o de ternera, comprándose más del doble que de carne de ternera, y un 66% más que de carne de cerdo.

Si comparamos la carne de cerdo con la de ternera se comprueba que hay un 32% más de compra de carne de cerdo.

Los hogares que más adquieren carne de vacuno lo hacen un 6% menos que los hogares que menos carne de cerdo compran.

Sin embargo, los que menos carne de pollo compran lo hacen en un 43% más que los que más compran carne de cerdo.

En cuanto al aceite los resultados ponen de manifiesto que los hogares españoles siguen apostando por el aceite de oliva. La compra de aceite de oliva es más del doble que la de aceite de girasol.

Los que más aceite de girasol compran lo hacen en un 45% menos que los que menos aceite de oliva compran.

Se compra más tabaco rubio que negro y en un 68% más. Los que menos tabaco rubio compran lo hacen en un 42% más que los que más tabaco negro adquieren.

(8) **Actualización** relativamente rápida de los resultados. Dado que la Encuesta es continua y se puede disponer en poco tiempo de los nuevos datos, este mecanismo se puede ir alimentando e inmediatamente disponer de los resultados de “última hora”.

Estas son algunas de las muchas y variadas posibilidades que se pueden obtener a partir del estudio del fenómeno cantidad comprada en unidades físicas del artículo considerado.

Esta técnica es por tanto una manera de representar formalmente un fenómeno, y lo importante no es sólo la gran cantidad de comparaciones y posibilidad de estudios que ofrece, y que pueden usarse para distintas finalidades, sino la sencillez y rapidez de cálculo con la que se obtienen los resultados.

No hay que hacer hipótesis de partida acerca de la distribución de la que provienen los datos, se evita la problemática de la elección de variables y formas funcionales, no son necesarias hipótesis sobre la introducción en el modelo de la componente aleatoria....

Aquí, repetimos tan sólo se necesitan datos relativos la fenómeno, que pueden estar ya disponibles. La rapidez en la obtención de resultados, así como la sencillez en el método y la ausencia de hipótesis adicionales, pensamos equilibran los inconvenientes citados y motiva el interés por explorar esta aproximación.

2.2.- Principales resultados sobre consumo de carne, aceite y tabaco.

En este apartado vamos a presentar los principales resultados obtenidos usando la metodología que hemos descrito anteriormente.

Se ha aplicado a los datos suministrados por la Encuesta Continua de Presupuestos Familiares, elaborada por el INE, correspondiente al primer trimestre del año 1.997, último para el que había datos disponibles antes de la reforma de la ECPF.

Hemos trabajado con los datos correspondientes a cantidades físicas compradas por los hogares en una semana. Somos conscientes de que posiblemente hubiera sido más adecuado utilizar datos provenientes del Panel de Consumo Alimentario que solía elaborar el Ministerio de Agricultura, Pesca y Alimentación, - parte de cuya finalidad era conocer la demanda de productos alimenticios - pero dejaron de hacerse públicos los datos a partir del 1.992.

Los artículos que se han utilizado para ilustrar la metodología son:

Carne de tres tipos: cerdo, pollo y ternera;
Aceite de girasol y de oliva, y
Tabaco negro y rubio.

Las unidades en que vienen medidos dichos artículos, conforme a la Encuesta, son:

Kilos para sólidos;
Litros para líquidos, y
Cajetillas para tabaco.

En primer lugar expondremos un cuadro resumen con las principales medidas descriptivas de las observaciones, que nos permita hacernos una idea de los datos con los que se está trabajando.

A continuación se han calculado las funciones de verosimilitud empírica para la media. Los gráficos muestran la función de verosimilitud empírica obtenida a partir de las observaciones de que se dispone. El eje de abscisas recoge los valores del parámetro que estamos considerando, en este caso cantidad comprada media semanal, y el eje de ordenada los valores de la verosimilitud empírica.

Comentaremos los intervalos de verosimilitud, para una fortaleza de la evidencia de 8 y de 32.

Además se llevará a cabo una estimación puntual y por intervalo para el valor del parámetro. La estimación puntual vendrá dada por el estimador de máxima verosimilitud. Si además, como comentamos en el apartado de la metodología, utilizamos una distribución a priori Uniforme junto con lo que dimos en llamar verosimilitud modificada, obtenemos una normalización de la verosimilitud que proporciona una distribución de probabilidad para el parámetro. Así es sencillo obtener probabilidades e intervalos de confianza, intervalos que podrán reinterpretarse como evidencia.

1.- DESCRIPTIVA

Comentemos alguno de los resultados obtenidos:

Para los grupos considerados vamos a hacer referencia a las siguientes medidas:

- .- Posición central: valor de la media y de la mediana;
- .- Posición: valor mínimo y máximo, y los percentiles 10, 25, 75 y 90;
- .- Dispersión: rango, desviación típica, coeficiente de variación y amplitud del intervalo que agrupa el 80 % de las observaciones centrales;
- .- Forma: coeficiente de asimetría de Pearson y coeficiente de curtosis de Pearson.

Todos los artículos presentan asimetría positiva y un apuntamiento mayor que la Normal, siendo muy grande el valor que toma el coeficiente de curtosis en el caso del artículo carne de cerdo.

Presentamos a continuación un cuadro que resume las características que antes comentábamos.

Figura 2.1.

POSICIÓN CENTRAL	POSICIÓN		DISPERSIÓN	FORMA
Media Mediana	m M C10	C25 C75 C90	R Desv.Tip Coef.Var C90-C10	Asimetría Curtosis
Carne de cerdo				
0,68 0	0 57,02 0	0 1 1,89	57,02 2,27 3,31 1,89	16,33 335,48
Carne de pollo				
1,15 0,75	0 12 0	0 2 3	12 1,39 1,21 3	1,76 5,21
Carne de ternera				
0,51 0	0 11,5 0	0 0,8 1,5	11,5 0,87 1,69 1,5	3,1 18,41
Aceite de girasol				
0,49 0	0 24,01 0	0 0 2	24,01 1,57 3,21 2	6,39 55,39

Aceite de oliva				
1,12	0	0	50,01	8,99
0	50,01	1	3,26	108,62
	0	3	2,9	
			3	
Tabaco negro				
1,57	0	0	50,01	3,77
0	50,01	0	4,11	20,77
	0	7	2,63	
			7	
Tabaco rubio				
2,63	0	0	55,02	2,81
0	55,02	4	5,1	11,26
	0	9	1,94	
			9	

Para el grupo de tabaco el más homogéneo es el tabaco rubio, mientras que en las carnes la ternera presenta el valor más bajo, resultando el menos homogéneo la carne de cerdo.

Si nos fijamos en los aceites resulta más homogéneo el aceite de oliva, con un coeficiente de variación de 2,9 frente a 3,21 en el caso del aceite de girasol.

2. RESULTADOS OBTENIDOS A PARTIR DE LA VEROSIMILITUD EMPÍRICA.

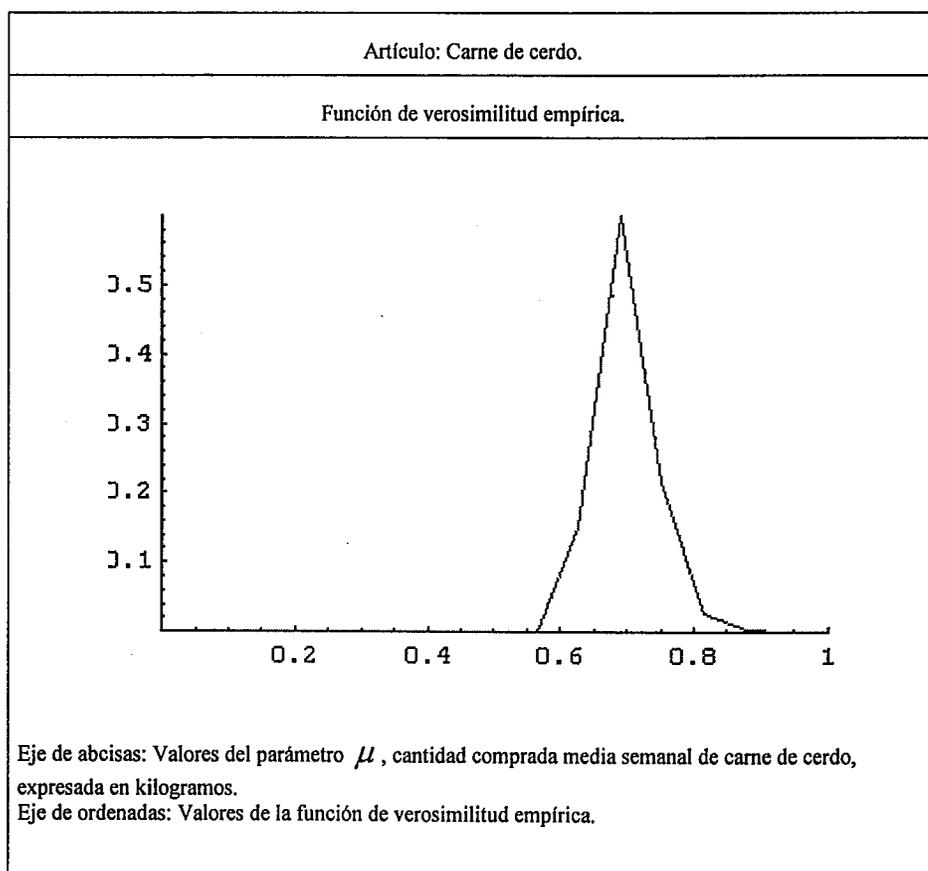
El orden de los artículos va a ser: carne, aceite y tabaco.

Carne de cerdo.

Se disponía de 3.112 observaciones, y con ellas obtuvimos la función de verosimilitud empírica para la compra media. El máximo se alcanza en 0,69 kilos por semana.

Presentamos a continuación la función de verosimilitud empírica obtenida.

Figura 2.2.



Los intervalos de verosimilitud vienen dados por:

I. Verosimilitud para una fuerza de evidencia 8....(0,63 ; 0,76)

I. Verosimilitud para una fuerza de evidencia 32....(0,63 ; 0,81)

Si notamos por “a” el valor de mayor verosimilitud y por $L(a)$ la verosimilitud correspondiente a ese valor, obtenemos los siguientes cocientes, que nos dan una idea de la fortaleza de la evidencia.

$$\begin{aligned} L(a) / L(0,63) &= 3,97; \\ L(a) / L(0,75) &= 2,72; \\ L(a) / L(0,81) &= 21,57; \\ L(a) / L(0,56) &= 5.124,42; \\ L(a) / L(0,88) &= 278,96. \end{aligned}$$

Estos cocientes nos dan información sobre cuantas veces es más verosímil el valor 0,69 en relación a los puntos más cercanos en la gráfica. A medida que nos alejamos, la fuerza de la evidencia, medida por el cociente de verosimilitudes, es mucho mayor. Así, el valor del máximo es casi 4 veces más verosímil que 0,63; tiene una verosimilitud 21,57 mayor la del valor 0,81 y 5.124, veces mayor que la de 0,56.

Si utilizamos una distribución a priori Uniforme junto con la verosimilitud modificada, obtendríamos una verosimilitud normalizada, sobre la que se pueden calcular probabilidades, y por tanto, intervalos de confianza.

Exponemos a continuación una tabla con algunos percentiles de la verosimilitud normalizada:

Orden del percentil	0,05	0,25	0,5	0,75	0,95	0,99
Valor del percentil	0,5	0,52	0,53	0,55	0,58	0,59

Estimación puntual y regiones de confianza para la cantidad comprada media semanal de carne de cerdo.

Para la estimación puntual vamos a utilizar el estimador de máxima verosimilitud. En este caso, la cantidad comprada media estimada es 0,69 kilos por semana.

Algunas regiones de confianza se pueden obtener con los percentiles de la verosimilitud normalizada.

I. Confianza (80%).....(0,58 ; 0,72)
I. Confianza (90%).....(0,58 ; 0,75)
I. Confianza (95%).....(0,57 ; 0,76)

Podemos decir que la cantidad comprada media se encuentra entre estos valores con una probabilidad del 80, 90 y 95 por ciento respectivamente.

También podríamos calcular el nivel de confianza para cada intervalo de verosimilitud.

Así, el intervalo de verosimilitud obtenido para una fuerza de evidencia 8 es (0,63; 0,76), que tiene un nivel de confianza del 82,5%.

Por otro lado, el nivel de confianza el intervalo de verosimilitud para una fuerza de evidencia 32, (0,63; 0,81) es del 84%.

Estos intervalos de confianza pueden interpretarse en términos de la evidencia, presentando gran fuerza en el sentido de grandes cocientes de verosimilitudes en relación a valores de fuera del intervalo.

Por ejemplo,

$$L(0,63) / L(0,56) = 1290,81;$$

$$L(0,63) / L(0,81) = 5,43;$$

$$L(0,63) / L(0,88) = 70,27;$$

Para el resto de los artículos que hemos analizado, vamos a suministrar la gráfica de la función de verosimilitud empírica y el cuadro que exponemos a continuación que resuma los principales resultados obtenidos, y que tendrían una interpretación análoga a la realizada con el artículo carne de cerdo.

Figura 2.3.

Artículo:		Unidades:				
Intervalo para una fuerza de evidencia 8.		(;)				
Intervalo para una fuerza de evidencia 32.		(;)				
Cocientes entre el valor de la función de verosimilitud empírica en el máximo y el valor de la función de verosimilitud empírica para otros valores cercanos en la gráfica.		$L(a)/L() =$				
Orden del Percentil.	0,05	0,25	0,5	0,75	0,95	0,99
Valor del Percentil.						
Estimación puntual.						
80%		(;)				
Intervalos de confianza 90%		(;)				
95%		(;)				
Nivel de confianza de los intervalos de verosimilitud expuestos al principio.						
Algunos cocientes del valor de la función de verosimilitud empírica en los valores de los intervalos anteriores y el valor de la función de verosimilitud empírica en valores de fuera de esos intervalos.		$L()/L() =$				

Este cuadro resume los principales resultados que se obtienen para cada artículo.

En la primera fila encontraremos el nombre del artículo así como la unidad en que se mide.

A continuación, leyendo de arriba hacia abajo, encontraremos un bloque de tres filas. Las dos primeras indican los intervalos de verosimilitud para una fuerza de evidencia 8 y 32. La tercera recoge algunos cocientes entre el valor de la función de verosimilitud empírica en el máximo y el valor de la función de verosimilitud empírica para otros valores cercanos en la gráfica, indicando la intensidad de esa evidencia.

El siguiente bloque recoge algunas cantidades relacionadas con la verosimilitud normalizada. Así encontramos dos filas donde aparecen el orden del percentil y el valor correspondiente. La siguiente casilla indica el valor del estimador de máxima verosimilitud, y por último los intervalos de confianza para el parámetro para un nivel de confianza del 80,90 y 95%.

El bloque final tiene dos casillas. La primera presenta en dos filas el nivel de confianza de los intervalos de verosimilitud expuestos al principio, y en la segunda aparecerán algunos cocientes entre el valor de la función de verosimilitud empírica en los valores de los intervalos anteriores y el valor de la función de verosimilitud empírica en valores de fuera de esos intervalos, que ponen de manifiesto la intensidad de la evidencia.

Este cuadro será utilizado en los capítulos siguientes para recoger los resultados obtenidos para los ingresos, gastos y otras variables objeto de estudio.

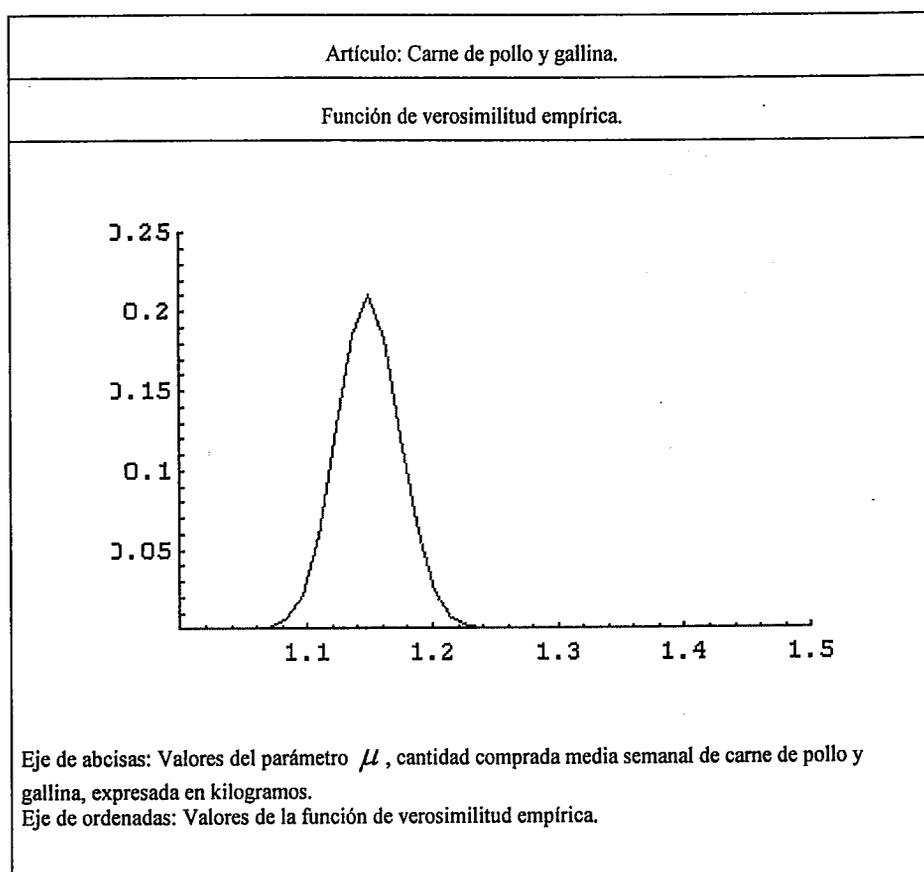
Veamos los restantes artículos:

Carne de pollo.

Se disponía de 3.114 observaciones. El valor más grande de la función de verosimilitud empírica se obtiene para el valor 1,15 kilos por semana.

Esta es su función de verosimilitud empírica:

Figura 2.4.



A continuación presentamos su cuadro con los principales resultados:

Figura 2.5.

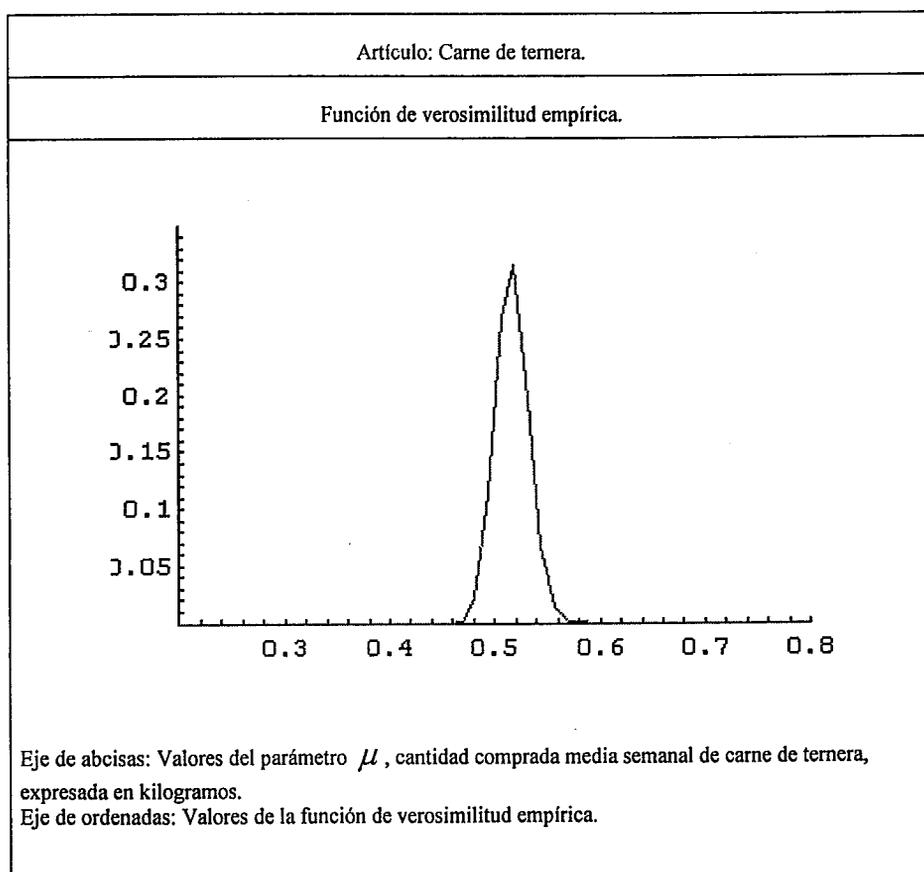
Artículo: Carne de pollo y gallina. Unidad: Kilo.						
Intervalo para una fuerza de evidencia 8.		(1,11 ; 1,18)				
Intervalo para una fuerza de evidencia 32.		(1,09 ; 1,21)				
Cocientes entre el valor de la función de verosimilitud empírica en el máximo y el valor de la función de verosimilitud empírica para otros valores cercanos en la gráfica.		$L(a) / L(1,08) = 38,76;$ $L(a) / L(1,09) = 10,06;$ $L(a) / L(1,11) = 3,59;$ $L(a) / L(1,21) = 28,46;$ $L(a) / L(1,23) = 116,18;$				
Orden del Percentil.	0,05	0,25	0,5	0,75	0,95	0,99
Valor del Percentil.	1,10	1,12	1,14	1,16	1,18	1,20
Estimación puntual.			1,15.			
Intervalos de confianza		80%	(1,10 ; 1,17)			
		90%	(1,10 ; 1,18)			
		95%	(1,09 ; 1,19)			
Nivel de confianza de los intervalos de verosimilitud expuestos al principio.		(1,11 ; 1,18).....85%				
		(1,09 ; 1,21).....96,5%				
Algunos cocientes del valor de la función de verosimilitud empírica en los valores de los intervalos anteriores y el valor de la función de verosimilitud empírica en valores de fuera de esos intervalos.		$L(1,12) / L(1,19) = 1,99;$ $L(1,13) / L(1,08) = 33,92;$ $L(1,13) / L(1,24) = 518,66;$ $L(1,16) / L(1,25) = 3.230,37;$ $L(1,19) / L(1,23) = 33,49;$ $L(1,9) / L(1,07) = 59,73;$				

Carne de ternera.

Se disponía de 3.112 observaciones, ya que había un par de cantidades que tuvieron que ser suprimidas ante el valor tan anómalo que mostraban. El valor que presenta mayor verosimilitud es 0,52 kilos por semana.

Su función de verosimilitud empírica:

Figura 2.6.



Principales resultados:

Figura 2.7.

Artículo: Carne de ternera. Unidad: Kilo.						
Intervalo para una fuerza de evidencia 8.		(0,49 ; 0,54)				
Intervalo para una fuerza de evidencia 32.		(0,48 ; 0,56)				
Cocientes entre el valor de la función de verosimilitud empírica en el máximo y el valor de la función de verosimilitud empírica para otros valores cercanos en la gráfica.		$L(a) / L(0,47) = 167,13;$ $L(a) / L(0,48) = 14,38;$ $L(a) / L(0,54) = 4,63;$ $L(a) / L(0,56) = 22,45;$ $L(a) / L(0,57) = 171,37;$				
Orden del Percentil.	0,05	0,25	0,5	0,75	0,95	0,99
Valor del Percentil.	0,48	0,5	0,51	0,52	0,54	0,55
Estimación puntual.			0,52.			
Intervalos de confianza		80%	(0,48 ; 0,53)			
		90%	(0,48 ; 0,54)			
		95%	(0,48 ; 0,54)			
Nivel de confianza de los intervalos de verosimilitud expuestos al principio.		(0,49 ; 0,54).....87,5%				
		(0,48 ; 0,56).....96,5%				
Algunos cocientes del valor de la función de verosimilitud empírica en los valores de los intervalos anteriores y el valor de la función de verosimilitud empírica en valores de fuera de esos intervalos.		$L(0,51) / L(0,47) = 141,43;$ $L(0,51) / L(0,56) = 19;$ $L(0,53) / L(0,45) = 2.933,94;$ $L(0,53) / L(0,57) = 106,54;$ $L(0,54) / L(0,44) = 74.751,68;$ $L(0,54) / L(0,58) = 36,99;$				

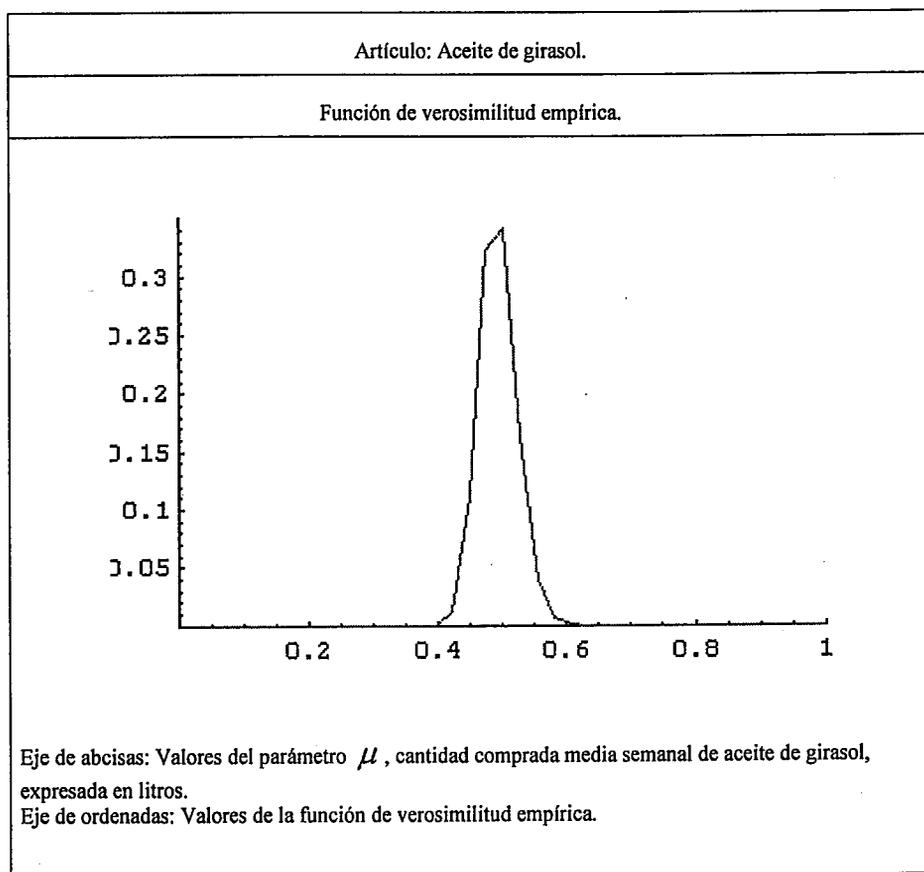
Veamos ahora otro par de artículos: aceite de girasol y de oliva.

Aceite de girasol.

Se disponía de 3.114 observaciones. El máximo de la verosimilitud se alcanza en 0,5 litros por semana.

Función de verosimilitud empírica:

Figura 2.8.



Para este artículo los resultados obtenidos han sido los siguientes:

Figura 2.9.

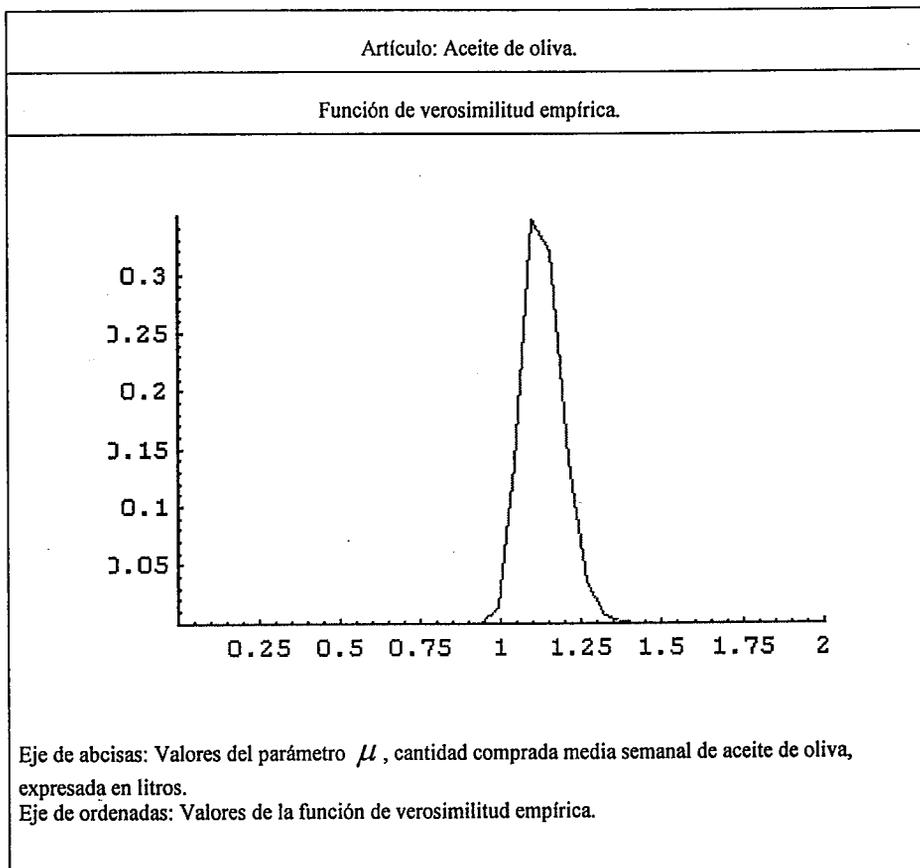
Artículo: Aceite de girasol. Unidad: Litro.						
Intervalo para una fuerza de evidencia 8.			(0,45 ; 0,53)			
Intervalo para una fuerza de evidencia 32.			(0,42 ; 0,55)			
Cocientes entre el valor de la función de verosimilitud empírica en el máximo y el valor de la función de verosimilitud empírica para otros valores cercanos en la gráfica.			$L(a)/L(0,4) = 1.307,61;$ $L(a)/L(0,45) = 3,05;$ $L(a)/L(0,53) = 2,09;$ $L(a)/L(0,55) = 8,33;$ $L(a)/L(0,58) = 55,4;5$			
Orden del Percentil.	0,05	0,25	0,5	0,75	0,95	0,99
Valor del Percentil.	0,43	0,46	0,48	0,5	0,52	0,55
Estimación puntual.			0,5.			
Intervalos de confianza			80%	(0,43 ; 0,51)		
			90%	(0,43 ; 0,53)		
			95%	(0,42 ; 0,55)		
Nivel de confianza de los intervalos de verosimilitud expuestos al principio.			(0,45 ; 0,53).....80%			
			(0,42 ; 0,55).....96,5%			
Algunos cocientes del valor de la función de verosimilitud empírica en los valores de los intervalos anteriores y el valor de la función de verosimilitud empírica en valores de fuera de esos intervalos			$L(0,45) / L(0,42) = 9,74;$ $L(0,45) / L(0,4) = 429;$ $L(0,45) / L(0,58) = 18,19,$ $L(0,53) / L(0,42) = 14,19;$ $L(0,53) / L(0,39) = 624,99;$ $L(0,53) / L(0,58) = 26,5;$			

Aceite de oliva.

Se disponía de 3.114 observaciones. El máximo de la verosimilitud se alcanza para el valor 1,1 litros por semana.

Función de verosimilitud empírica:

Figura 2.10.



Principales resultados:

Figura 2.11.

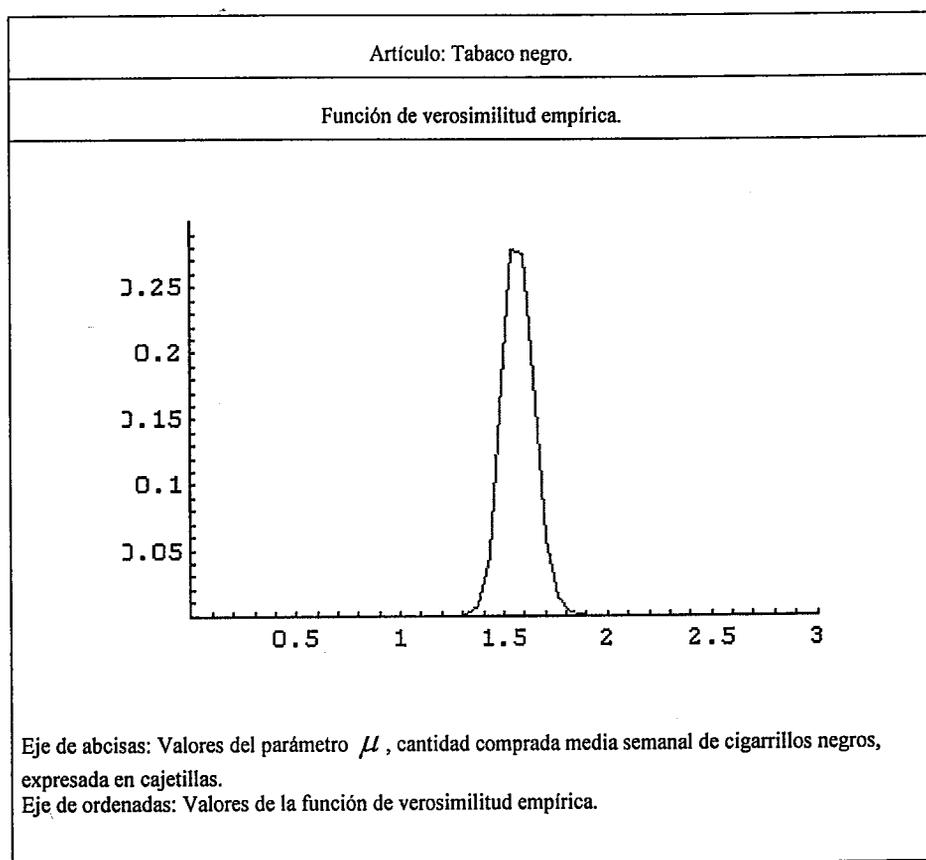
Artículo: Aceite de oliva. Unidad: Litro.						
Intervalo para una fuerza de evidencia 8.		(1,04 ; 1,21)				
Intervalo para una fuerza de evidencia 32.		(0,99 ; 1,26)				
Cocientes entre el valor de la función de verosimilitud empírica en el máximo y el valor de la función de verosimilitud empírica para otros valores cercanos en la gráfica.		$L(a)/ L(0,93) = 1.479,16;$ $L(a)/ L(0,99) = 24,93;$ $L(a)/ L(1,04) = 2,53;$ $L(a)/ L(1,26) = 9,74;$ $L(a)/ L(1,32) = 58,07;$				
Orden del Percentil.	0,05	0,25	0,5	0,75	0,95	0,99
Valor del Percentil.	1	1,06	1,1	1,14	1,2	1,26
Estimación puntual.			1,1.			
Intervalos de confianza		80%	(1 ; 1,17)			
		90%	(1 ; 1,21)			
		95%	(0,99 ; 1,24)			
Nivel de confianza de los intervalos de verosimilitud expuestos al principio.		(1,04 ; 1,21).....80%				
		(0,99 ; 1,26).....96,5%				
Algunos cocientes del valor de la función de verosimilitud empírica en los valores de los intervalos anteriores y el valor de la función de verosimilitud empírica en valores de fuera de esos intervalos.		$L(1,04) / L(1,26) = 3,84;$ $L(1,04) / L(0,93) = 583,42;$ $L(1,1) / L(0,93) = 1479,16;$ $L(1,1) / L(1,32) = 58,07;$ $L(1,21) / L(1,32) = 23,57,$ $L(1,21) / L(1,37) = 191,73;$				

Tabaco negro.

Se disponía de 3.114 observaciones. El máximo se alcanza para el valor 1,54 cajetillas por semana.

Función de verosimilitud empírica:

Figura 2.12.



Principales resultados:

Figura 2.13.

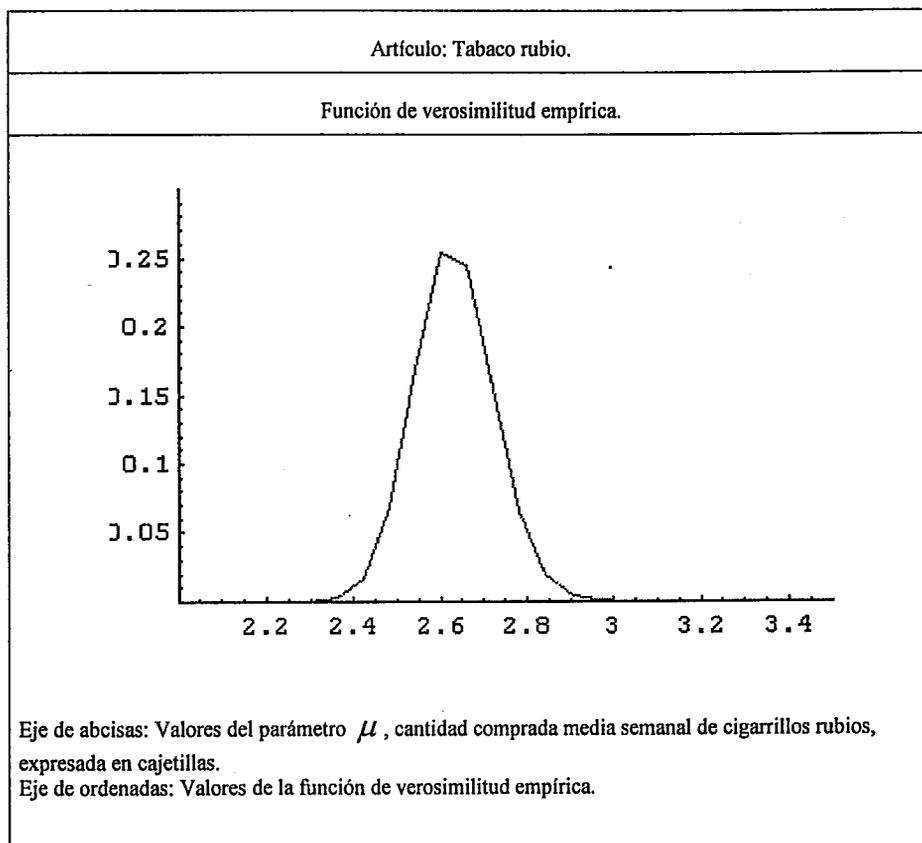
Artículo: Tabaco negro. Unidad: Cajetilla.						
Intervalo para una fuerza de evidencia 8.		(1,43 ; 1,70)				
Intervalo para una fuerza de evidencia 32.		(1,43 ; 1,76)				
Cocientes entre el valor de la función de verosimilitud empírica en el máximo y el valor de la función de verosimilitud empírica para otros valores cercanos en la gráfica.		$L(a)/L(1,37) = 41,9;$ $L(a)/L(1,43) = 6,04;$ $L(a)/L(1,48) = 1,78;$ $L(a)/L(1,7) = 4,78;$ $L(a)/L(1,81) = 124,96;$				
Orden del Percentil.	0,05	0,25	0,5	0,75	0,95	0,99
Valor del Percentil.	1,43	1,49	1,54	1,59	1,67	1,73
Estimación puntual.			1,54.			
Intervalos de confianza		80%	(1,43 ; 1,62)			
		90%	(1,43 ; 1,67)			
		95%	(1,4 ; 1,7)			
Nivel de confianza de los intervalos de verosimilitud expuestos al principio.		(1,43 ; 1,70).....92,5%				
		(1,43 ; 1,76).....94%				
Algunos cocientes del valor de la función de verosimilitud empírica en los valores de los intervalos anteriores y el valor de la función de verosimilitud empírica en valores de fuera de esos intervalos.		$L(1,43) / L(1,37) = 6,93;$ $L(1,43) / L(1,81) = 20,67;$ $L(1,48) / L(1,87) = 603,57;$ $L(1,59) / L(1,32) = 631,57;$ $L(1,65) / L(1,92) = 7.079,06;$ $L(1,7) / L(1,26) = 4.831,97;$				

Tabaco rubio.

Se disponía de 3.114 observaciones. El máximo de la verosimilitud se alcanza en 2,6 cajetillas por semana.

Función de verosimilitud empírica:

Figura 2.14.



Principales resultados:

Figura 2.15.

Artículo: Tabaco rubio. Unidad: Cajetilla.						
Intervalo para una fuerza de evidencia 8.		(2,48 ; 2,78)				
Intervalo para una fuerza de evidencia 32.		(2,42 ; 2,84)				
Cocientes entre el valor de la función de verosimilitud empírica en el máximo y el valor de la función de verosimilitud empírica para otros valores cercanos en la gráfica.		$L(a)/ L(2,36) = 107,41;$ $L(a)/ L(2,42) = 15,17;$ $L(a)/ L(2,78) = 3,85;$ $L(a)/ L(2,84) = 13,04;$ $L(a)/ L(2,90) = 62,05;$				
Orden del Percentil.	0,05	0,25	0,5	0,75	0,95	0,99
Valor del Percentil.	2,45	2,54	2,6	2,66	2,76	2,83
Estimación puntual.			2,6.			
Intervalos de confianza		80%	(2,45 ; 2,7)			
		90%	(2,45 ; 2,76)			
		95%	(2,42 ; 2,78)			
Nivel de confianza de los intervalos de verosimilitud expuestos al principio.		(2,48 ; 2,78).....87,5%				
		(2,42 ; 2,84).....96,5%				
Algunos cocientes del valor de la función de verosimilitud empírica en los valores de los intervalos anteriores y el valor de la función de verosimilitud empírica en valores de fuera de esos intervalos.		$L(2,48) / L(2,36) = 28,99;$ $L(2,54) / L(2,9) = 41,02;$ $L(2,6) / L(2,29) = 1.364,29,$ $L(2,66) / L(2,9) = 59,8;$ $L(2,72) / L(2,96) = 246,15;$ $L(2,78) / L(2,29) = 354,14;$				

Como se ha visto, se han seleccionado para esta Memoria unos cuantos artículos de los 32 para los que la ECPF disponía de información, pero somos conscientes de que resultaría igualmente interesante abordar el estudio de otros artículos, tales como por ejemplo el grupo de los pescados (y compararlos con las carnes); las frutas y las verduras; los distintos tipos de leche o los refrescos y otros tipos de bebidas. Estos 32 artículos fueron seleccionados por el INE para un estudio de nutrición y tradicionalmente se mantuvieron en la ECPF. Actualmente, tras la reforma realizada, también es posible conocer las cantidades compradas para todos los artículos de alimentación para los que se conoce su gasto, lo cual facilita y amplía considerablemente las posibilidades de estudio y comparación.

Capítulo 3

VEROSIMILITUD EMPÍRICA Y DISTRIBUCIÓN DE LA RENTA.

3.1. – La renta y su distribución.

3.1.1.- Un poco de historia.

A la hora de abordar el tema de la renta y cómo se distribuye, el primer problema que se encuentra es saber qué entendemos por renta y desde qué óptica vamos a estudiar su distribución.

Históricamente, el concepto más utilizado por los economistas es el de Renta Nacional, que mide el volumen de recursos generados a lo largo de un año económico y que puede ser usado para el consumo o la inversión, tanto pública como privada, teniendo en cuenta el saldo que se obtiene de las operaciones con el exterior.

La Renta Nacional y su evolución está muy relacionada con los ciclos económicos de un país, es decir, con las etapas de bonanza o depresión que tiene su economía. Si es importante analizar como evoluciona la renta aún lo es más el estudio de su distribución, ya que influye en su propio crecimiento,

necesario para el desarrollo de un país, y tiene consecuencias sobre las tensiones sociales y el bienestar.

Los efectos que sobre la economía y la sociedad tiene la distribución de la renta han sido una preocupación constante de los economistas desde los tiempos de Adam Smith. La teoría de la distribución se centraba fundamentalmente en lo que hoy se llama distribución factorial: los salarios, las rentas, los beneficios, los intereses... Con la Revolución Francesa comenzó el interés por ver cómo se distribuían los ingresos desde una perspectiva personal. Saint-Simon, Fourier y Owen fueron autores que trataron estos temas. Ya en nuestro siglo, Keynes orientó la política económica a lograr mayores niveles de renta, lo que acabaría con la insuficiencia de las rentas personales y haría innecesaria la igualdad distributiva, que se trasladaría al concepto de equidad.

Tras la segunda Guerra Mundial, la Renta Nacional como indicador de los recursos generados en un año por una economía, ha manifestado un crecimiento importante en la mayoría de los países, especialmente en los llamados de nueva industrialización.

Poco a poco se ha puesto de manifiesto la importancia del análisis de la renta, y prueba de ello es la cantidad de estudios sobre sus componentes, su distribución, sus usos..., especialmente porque puede utilizarse para obtener indicadores diversos que midan el bienestar material de una nación.

Una equitativa distribución de la riqueza y de la renta, desde cualquier punto de vista, es necesaria para reducir las tensiones que aparecen en una sociedad. En la medida en que se da un tratamiento adecuado a los problemas de la política de distribución, la conflictividad social se reduce, mejorando el nivel de bienestar social. Por eso en muchos países el modelo distributivo no es ajeno a la intervención del Estado por medio de políticas que tratan de corregir las desigualdades de renta entre los sectores, los factores, las regiones y las familias a través de la imposición fiscal y el gasto público.

Esas políticas no pueden ser llevadas a cabo aisladas del esquema global de la política económica sino en coordinación, para evitar que repercutan negativamente en los fines generales de aquella.

Con lo comentado hasta ahora puede entenderse la importancia que para un país tiene el estudio de la renta y su distribución, y porqué han proliferado los trabajos sobre ella, desde diversos ángulos.

3.1.2. - Puntos de vista en la distribución de la renta.

Cuatro son los puntos de vista que pueden utilizarse para analizar el esquema de distribución de la renta: sectorial, factorial, espacial y personal. Pasemos a describir brevemente en que consisten.

La distribución **sectorial** de la renta estudia cómo se reparte la renta entre los sectores productivos, agricultura, industria, construcción y servicios.

La distribución **factorial** o funcional de la renta muestra cómo se reparte entre todos y cada uno de los factores productivos la renta obtenida. Así se puede hablar de:

- Rentas del trabajo : englobaría las rentas salariales y las cotizaciones sociales a cuenta del trabajador pagadas por el empresario.
- Rentas de capital : incluye el ahorro empresarial (beneficios no distribuidos), intereses y dividendos, y las rentas de alquileres.
- Rentas mixtas : formadas por rentas agrarias, de profesionales, de servicios, así como la de empresarios y trabajadores autónomos.

La distribución **personal** de la renta es sin duda el aspecto de mayor interés, al menos desde el punto de vista sociológico. Trata de mostrar la estructura de distribución de la renta, clasificada por tramos de mayor a menor dimensión. Permite identificar las disparidades de renta entre las distintas clases y su grado de concentración.

Desde la óptica **espacial** de la distribución de la renta, se trata de establecer si existen diferencias relativas de renta entre los distintos espacios en que puede dividirse una determinada área geográfica. Permite estudiar la evolución y situación de los desequilibrios de una región dentro de una comunidad o de un país. El problema que plantea el análisis de la distribución

espacial de la renta no es tanto confirmar si existen desigualdades como detectar su tendencia y cuantificar sus diferencias, para ver si son tolerables por la sociedad o suponen desequilibrios inaceptables que dañan a una parte del territorio y a su población. Se busca también identificar sus causas para darle una solución adecuada.

Teniendo en cuenta las diversas perspectivas desde las que puede analizarse la renta y su distribución, el enfoque de la distribución personal es el que resulta más interesante desde el punto de vista social, y por ello es objeto de numerosos trabajos. La distribución sectorial y factorial tiene más contenido económico y explica su relación con el crecimiento y desarrollo, pero el punto de vista personal nos da una idea de cómo la política económica ha contribuido a lograr una mayor equidad social, corrigiendo las diferencias relativas de renta que el modelo económico genera.

El problema derivado de la distribución personal de la renta y su equidad es fuente de preocupación para los políticos, sociólogos y economistas, por la repercusión que tiene en el equilibrio y armonía de los estratos sociales.

La idea de bienestar económico y sus interrelaciones con la justicia y equidad distributiva de la renta lleva al planteamiento de un problema fundamental: el conocimiento de la distribución de la renta.

Aunque éste es un aspecto esencial de la distribución personal de la renta y numerosos son los trabajos que han tratado de darle respuesta, hay otros muchos temas que son objeto de estudio y que podemos enmarcar en dos grandes áreas:

- Análisis del nivel de desigualdad en una distribución de renta dada, y su redistribución;
- Análisis de la pobreza.

Para ello, en primer lugar sintetizaremos los principales trabajos a este respecto.

3.2. - Estudios de renta.

Son varias las áreas de investigación que podemos considerar relacionadas con la distribución de la renta.

La primera que vamos a comentar es la del estudio de la desigualdad de la distribución. Los trabajos se centran en presentar índices distintos que permiten conocer el nivel de desigualdad. Destacan los de Lorenz (1.905), Gini (1.936), Atkinson (1.970), Cowell (1.980) y Shorrocks (1.980;1.983). También hay abundantes referencias de trabajos en los que se analiza la evolución de la desigualdad. Aplicado al caso español destacamos Ruiz Castillo (1.987), López Menéndez (1.992), De Haro García (1.995), Fernández y otros (1.995)....

La redistribución, es decir, el análisis de cómo determinadas acciones llevadas a cabo por el estado pueden influir en la distribución personal de la renta es otro de los temas que despiertan gran atención. Los primeros trabajos de esta naturaleza son los de Jakobson (1.976) y Kakwani (1.977), en los que estudian la relación entre la curva de concentración y la imposición.

Muy en relación con los estudios de desigualdad se encuentra el análisis de la pobreza, que fija su interés en aquel segmento de la población que menos renta tiene. Sen (1.976) establece las bases por las que se considera si un individuo es o no pobre. Muchos son los trabajos orientados a la obtención de indicadores de pobreza o a estudiar su evolución y su distribución.

Otro aspecto al que se le ha dedicado mucho esfuerzo es al conocimiento de la distribución personal de la renta, bien sea mediante una modelización paramétrica o bien con un enfoque no paramétrico. Dado que nuestro trabajo se puede enmarcar dentro de este último aspecto, comentaremos los principales trabajos en este campo.

3.2.1. – Modelización paramétrica de la renta.

En la modelización paramétrica se parte de un modelo de probabilidad, el cual queda perfectamente definido por su correspondiente familia de funciones de densidad,

$$\{f_{\theta}(x), \theta \in \Theta\},$$

y que se propone a priori, con la intención de describir la población de la que se extraen los datos. Surge de forma natural la pregunta: ¿qué es lo que determina la elección de un modelo u otro?. Parece adecuado suponer que la elección del modelo vendría determinada por las características específicas del fenómeno a estudiar. Sin embargo, en muchas ocasiones no parece muy justificada la formulación de una familia, prestando más atención al mayor o menor ajuste de los datos que a una estimación bajo esas condiciones.

Lo que caracteriza a la modelización paramétrica es el suponer conocida la función de densidad excepto por un vector θ de parámetros, desconocido. Por tanto, una cuestión esencial será la estimación de dicho vector de parámetros.

Una vez estimado el modelo es importante conocer la bondad del ajuste realizado.

Podríamos decir que en modelización paramétrica de la renta hay tres fases fundamentales: la especificación del modelo, su estimación y el análisis de la bondad del ajuste.

Por especificación entendemos la elección de un modelo de probabilidad perfectamente conocido a excepción del parámetro o parámetros que posea. Así, decimos que la variable renta sigue una distribución de Pareto, una Gamma, una Lognormal,.... Sin embargo, como Bassman y otros (1.990) plantean no podemos hablar de que exista un modelo único o una forma funcional universalmente aceptada. Esto nos lleva al problema de la existencia de una gran cantidad de formas funcionales posibles, y parecería adecuado establecer algún criterio que nos permita escoger entre ellas.

Si se desea explicar o conocer una realidad, parece razonable pensar que una postura podría ser estudiar ese fenómeno y ver cuáles son sus

principales características. Así podríamos tener una idea de si hay algunos modelos que respondan mejor a esas características que otros. En 1.898 Edgeworth fue el primero en proponer algunas propiedades deseables a ser cumplidas por los modelos de distribución de la renta. Cantelli en 1.921, Fréchet (1.939), Aitchison y Brown (1.957), Mandelbrot (1.960), Dagum (1.977) y otros autores han desarrollado trabajos en este sentido.

Como De Haro García (1.993), Callealta y otros (1.996), y Prieto Alaiz (1.998) señalan, las principales propiedades a cumplir por un modelo serían:

1.- Tener un fundamento teórico empírico, es decir, debería derivarse a partir de supuestos que recogieran las características de regularidad de la distribución de la renta observada, a lo largo del tiempo y del espacio.

Así, hay formas funcionales que han sido derivadas a partir de características propias de la renta recogidas por medio de un proceso estocástico, como por ejemplo la distribución de Pareto, Lognormal, Gram Charlier, Champernowne, Fisk, Singh-Maddala,.....

Otras, se obtienen a partir del planteamiento de una serie de ecuaciones diferenciales, como por ejemplo la distribuciones de Dagum, las derivadas a partir del sistema de Pearson...

El resto de los modelos son modelos *ad hoc*, que se proponen con el objetivo de ajustar modelos de distribución de la renta. El punto de vista sugerido por Pretorius en 1.930 de que “la superioridad de una función sobre otra depende más del éxito con el que esa función puede aplicarse a los datos, que a la forma como se ha originado” es apoyado por numerosos autores con posterioridad, como Weibull, pero se puede comprobar que los modelos que proporcionan buenos ajustes presentan en común ciertas propiedades como la que acabamos de señalar.

2.- Converger asintóticamente a la Ley de Pareto. Mandelbrot (1.960) establece lo que se conoce como *Ley débil de Pareto*, que es universalmente aceptada, de manera que se considera una propiedad deseable por cualquier modelo probabilístico.

Sea X la variable que recoge el nivel de renta de la población, con función de distribución $F(x)$; un nivel mínimo de renta x_0 y α como el parámetro de Pareto. La Ley débil establece que :

$$\lim_{x \rightarrow \infty} \frac{1 - F(x)}{\left(\frac{x}{x_0}\right)^{-\alpha}} = 1 \quad \alpha \geq 0, \quad 0 < x_0 \leq x.$$

Esta propiedad implica un débil grado de contacto con el eje de abcisas. Las distribuciones que cumplen esta propiedad, como Fisk y Pareto, se les denomina “de cola pesada”, y se caracterizan porque su función de densidad converge lentamente a cero cuando la variable tiende a infinito.

No obstante, algunos de los modelos que más se han utilizado históricamente, como la Lognormal o la distribución Gamma, no cumplen esta propiedad.

3.- Significado económico de los parámetros. Si el objetivo es describir lo mejor posible la realidad económica, parece lógico que los parámetros de los modelos tengan significado económico. Así, se pueden distinguir entre parámetros de escala, expresados en la unidad monetaria de medida de la variable y parámetros de desigualdad, que reflejan el grado de desigualdad de la distribución de renta.

El hecho de que un modelo tenga pocos parámetros representa una ventaja desde el punto de vista de su simplicidad, pero puede describir pobremente el comportamiento de la variable. En general, las distribuciones con tres o cuatro parámetros son preferidas a las distribuciones con un menor número de parámetros. Esto ha sido confirmado empíricamente por los trabajos de Singh-Maddala (1.996), McDonald (1.984), Dagum (1.977)... Sin embargo, otros autores afirman que la mejora en la precisión del ajuste no siempre es tan espectacular como para que merezca la pena la complejidad que presentan los modelos con muchos parámetros (ver De Haro García, (1.993)).

Atendiendo a la sencillez, a igualdad en la bondad de ajuste se prefiere menos parámetros.

4.- Presentar asimetría a la derecha. Es la característica más representativa de la distribución de la renta observada, a lo largo del tiempo y el espacio y suficiente para que un modelo sea rechazado si no la recoge.

Además sería conveniente que los modelos presentasen flexibilidad para ajustar distribuciones tanto ceromodales como unimodales, y fueran capaces de adaptarse a los posibles cambios que se den a lo largo del tiempo.

5.- Buen ajuste del modelo. Es deseable que el modelo explique lo mejor posible el comportamiento de la población. Una vez fijado el modelo se procede a la estimación de los parámetros y a la comprobación de cuanto difiere la distribución observada de la estimada, por diversos criterios.

Muchos han sido los modelos utilizados para describir la distribución personal de la renta. Presentamos a continuación un cuadro que recoge los principales modelos utilizados y los trabajos en los que se hace referencia a ellos. Para una revisión de estos modelos ver Pena y otros (1.996), Esteban y otros (2.000).

Modelo de distribución	Algunos trabajos que utilizan este modelo de distribución
Pareto	De Haro García (1.993)
Gamma	March-Ammon (1.898); Amoroso (1.925); D'Addario (1.936); Salem-Mount (1.974); Singh-Maddala (1.976); Bartels (1.977); McDonald-Ransom (1.979); Dagum (1.983); Mc Donald (1.984); De Haro García (1.993); Mc Donald- Xu (1.995); Cowell (1.996); Prieto Alaiz (1.998);
Lognormal	Kapteyn (1.903); Edgeworth (1.924); Rutherford (1.955); Aitchison- Brown (1.969); Salem-Mount (1.974); Singh-Maddala (1.976);

	Dagum (1.983); Mc Donald (1.984) Brachmann-Stich- Trede (1.995); Cowell, (1.996); Prieto Alaiz(1.998);
Gamma Generalizada	Amoroso (1.925); Kloeck-Van Dijk (1.978); Atoda-Naosumi-Suruga-Tachibanaci (1.980); Taille, (1.980); Mc Donald, (1.984); Esteban García (1.986); Brachmann-Stich-Trede (1.995);
Weibull	Mc Donald (1.984); Brachmann-Stich-Trede (1.995);
Champernowne	Fisk (1.961);
Fisk	Mc Donald, (1.984); De Haro García (1.993); Brachmann-Stich- Trede (1.995); Prieto (1.998);
Beta	Thurow (1.970); Mc Donald- Ransom (1.979); Mc Donald- Xu (1.995);
Beta 1	Mc Donald (1.984); Mc Donald- Xu (1.995); Brachmann-Stich-Trede (1.995);
Beta 2	Mc Donald (1.984); Mc Donald- Xu (1.995); Brachmann-Stich-Trede (1.995);
Singh-Maddala	Mc Donald- Ransom (1.979); Dagum (1.983); Mc Donald (1.984) Brachmann-Stich-Trede (1.995); Prieto Alaiz (1.998);
Dagum	Victoria Feser (1.995) Prieto Alaiz (1.998);
Beta Generalizada 1	Mc Donald (1.984); Mc Donald- Xu (1.992); Mc Donald- Xu (1.995); Brachmann-Stich-Trede (1.995);
Beta Generalizada 2	Mc Donald (1.984); Mc Donald- Xu (1.992); Mc Donald- Xu (1.995); Brachmann-Stich-Trede (1.995);

De las propiedades que ha de cumplir un modelo, es a la bondad del ajuste a la que se ha prestado más atención. De hecho, una gran parte de la literatura se ha dedicado, y aún lo sigue haciendo, a proponer modelos y compararlos con otros existentes, jerarquizándolos desde el punto de vista del mejor o peor ajuste a los datos.

Así Salem y Mount (1.974), comparan la distribución Lognormal con la Gamma, mostrando un mejor ajuste a los datos de ésta última, aunque exagera la asimetría.

Dos años más tarde, Singh y Maddala obtienen un nuevo modelo que se comporta mejor que la distribuciones de Pareto y Lognormal, ajustando mejor que la distribución Gamma usada por Salem y Mount.

En 1.983 Dagum compara la distribución Lognormal, Gamma, Singh-Maddala y Dagum tipos I y III, obteniendo mejores resultados con las dos últimas distribuciones.

Mc Donald (1.984), De Haro García (1.993), Mc Donald -Xu, (1.995)... son otros trabajos que comparan distintos ajustes de distribuciones.

En éstos y en otros muchos trabajos la bondad del ajuste se evalúa mediante criterios que de alguna manera miden la distancia entre las probabilidades observadas y las estimadas. Se mira que modelo proporciona un coeficiente de determinación más elevado, una suma de cuadrados de los errores más baja,..., o se utilizan distintos contrastes, como los basados en la distribución chi-cuadrado, especialmente el estadístico de Pearson o el estadístico neperiano de la razón de verosimilitud, o los contrastes basados en la función de distribución empírica, siendo el más utilizado el estadístico de Kolmogorov-Smirnov.

También hay otros criterios para comprobar la bondad del ajuste mediante determinados gráficos que muestran las discrepancias entre la función de distribución estimada y la empírica, o entre la función de densidad estimada y el histograma de frecuencias.

Para datos agrupados se utiliza un test basado en los límites del ratio de concentración de Lorenz, establecidos por Gastwirth en 1.973. Si la distribución escogida es un modelo adecuado para describir la renta, la medida de Lorenz obtenida a partir de los parámetros estimados debe caer entre esos

límites. La distribución Lognormal no pasa dicho test, a diferencia de la distribución Gamma o la Singh-Maddala.

Así como la especificación y la bondad del ajuste son importantes, también lo es la estimación del modelo escogido. Sin embargo, en los trabajos de modelización de la distribución de la variable aleatoria renta, no se le ha prestado demasiada atención. El método a utilizar era algo más o menos establecido a priori, en el sentido de que se usan aquellos métodos que proporcionan buenas propiedades, fundamentalmente la eficiencia asintótica.

Tan sólo McDonald y Ransom en 1.979, realizaron un trabajo en el que mostraban la importancia, no sólo de la elección de la forma funcional, sino también de la técnica de estimación utilizada.

Generalmente el método de estimación depende del tipo de datos de que se dispone y de las propiedades de los estimadores.

Si los datos son agrupados, los métodos utilizados son:

- Método de máxima verosimilitud para datos agrupados;
- Método de estimación de mínima chi-cuadrado de Pearson;
- Método de los mínimos cuadrados.

Para datos individuales se utiliza fundamentalmente el método de máxima verosimilitud.

Sin embargo, también hay algunos trabajos Victoria Feser (1.993; 1.994), Cowell (1.996), Prieto Alaiz (1.998), que utilizan métodos de estimación robustos como alternativa a los estimadores clásicos. Utilizan M-estimadores, y dentro de ellos, el estimador óptimo s-robusto.

La ventaja de la modelización es que simplifica mucho si se quiere realizar análisis relacionados con la distribución de la renta. Viendo la evolución de los parámetros se puede tener una idea de la variación de la distribución; se usan a su vez para obtener índices de pobreza o de desigualdad, o para comprobar el efecto que determinadas políticas del Gobierno han podido tener...

Pero frente a todas esas ventajas no debemos olvidar que la validez de todos los estudios que se realicen dependen del modelo utilizado. ¿Qué ocurre

si el modelo no es el “adecuado”?. Hay que escoger el modelo, que debe cumplir unas propiedades bastantes restrictivas, elegir el método de estimación y que sea el adecuado, mirar la bondad que proporciona el ajuste.

Junto a esta modelización paramétrica de la renta existen otras posibilidades que han tenido mucho auge en los últimos tiempos, y que pasamos a comentar brevemente a continuación.

3.2.2. – Enfoque no paramétrico de la distribución de la renta.

Hemos comentado como la modelización paramétrica asume que los datos provienen de una de las familias paraméricas de distribución, por ejemplo Lognormal, Beta,..., y por tanto, la función de densidad queda totalmente establecida cuando se estiman los parámetros de su expresión correspondiente.

Hay otra posibilidad, que es contemplar la estimación de la distribución de la renta desde el punto de vista no paramétrico. En este caso no se hacen hipótesis sobre la distribución de los datos observados. Aunque se asume que la distribución tiene una función de densidad, se deja a los datos hablar por sí solos a la hora de estimar esa función, más que imponer que la función pertenezca a una familia determinada. En los últimos años, los avances en la técnica informática han dado un impulso importante a estos métodos.

La ventaja de esta técnica es que permite conseguir mayor flexibilidad en cuanto a la posible forma de la distribución y ajuste a los datos, detectando comportamientos de la variable renta que no se observan con la formulación clásica.

Los métodos más utilizados en la estimación no paramétrica de la densidad en el caso de la renta son el histograma, el estimador “naive” y la estimación por núcleo.

Comentaremos brevemente en que consiste el método del núcleo puesto que es el más utilizado para el caso de la renta. Para más información sobre técnicas de estimación no paramétricas puede consultarse Parzen (1.962), o Silverman (1.986).

El estimador núcleo viene dado por:

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K\left(\frac{x - X_i}{h}\right)$$

donde K es la función núcleo y h es un parámetro de suavizamiento, n es el tamaño de la muestra y X_i es la observación i -ésima.

Puede entenderse el estimador como una suma de gibas (“bumps”) situados en las observaciones. La función núcleo determina la forma de los “bultos” y el parámetro de suavizamiento, también llamado ventana, determina su amplitud.

Directamente de la definición del núcleo se deduce que:

1. Las funciones núcleo son simétricas alrededor de 0 e integran la unidad.
2. Como la función núcleo es una función de densidad, la estimación por este método también lo es:

$$\int_{-\infty}^{\infty} k(x) dx = 1 \quad \Rightarrow \quad \int_{-\infty}^{\infty} \hat{f}(x) dx = 1.$$

3. La propiedad de suavidad de la función núcleo es heredada por la estimación. Si k es n veces continuamente diferenciable, también lo es la estimación.
4. Las estimaciones núcleo no dependen de ninguna elección de orígenes. Especificado el núcleo y la amplitud, la estimación es única para un conjunto de datos dados.
5. La mayoría de las funciones núcleo son positivas para asegurar que la estimación sea una función de densidad.

Hay muchas funciones núcleo. En la distribución de la renta se usa la función núcleo Epanechnikov, donde

$$K(x) = \begin{cases} \frac{3}{4}(1-x^2) & \text{si } |x| \leq 1; \\ 0 & \text{otro caso.} \end{cases}$$

Tanto Wand y otros (1.991), como Marron y Schmitz (1.992) utilizan la estimación núcleo de la función de densidad para estudiar la renta en el Reino Unido. estiman la densidad por el método del núcleo para datos de renta alemanes. Deaton (1.988), aplica un análisis no paramétrico a la distribución de la renta en Tailandia; Gerfin (1.994) estudia la distribución de la renta y su desigualdad para Suiza; Gouveia y Tavares (1.995) para la renta y el gasto en Portugal, y DiNardo y otros (1.996) llevan a cabo un estudio del mercado de trabajo y la distribución de los salarios. Brachmann y otros (1.995) y Cowell y otros (1.998), utilizan métodos paramétricos y no paramétricos en sendos trabajos aplicados a Alemania y a Brasil. Utilizan funciones núcleo y algunos modelos tradicionales para estudiar la distribución de la renta.

Aplicado al caso español encontramos trabajos en los que se modeliza la distribución de la renta usando estos métodos: Herrerías y otros (1.996), Del Oro y Presedo (1.992), Cao y otros (1.997), y trabajos en los que se obtienen curvas de Lorenz mediante esta técnica, Del Oro y Presedo (1.992), Herrerías y García (2.000); y otros estudios relacionados con la desigualdad, Herrerías y otros (1.998), Bolancé y Guillén (2.000).

3.3. – Aclaraciones metodológicas.

Vamos a centrarnos en el conocimiento sobre la renta de los hogares españoles y se hacen necesarios algunos comentarios acerca de determinados conceptos.

3.3.1. – Datos: Ingresos o gastos.

En toda metodología estadística es necesario disponer de datos. En el caso de la renta son muchas y variadas las observaciones que pueden responder al nombre de renta. La consideración de estas observaciones como

datos de renta es debida a las diferentes definiciones que hay de la misma, así como el período de tiempo o la unidad de análisis considerada.

El hecho de escoger unas definiciones u otras puede afectar a los resultados obtenidos. Si se quiere estudiar la desigualdad no es lo mismo utilizar datos de rentas anuales que datos correspondientes a períodos menores, ya que la renta anual fluctúa menos.

Utilizar un tipo de datos u otros puede influir en la modelización de la renta, en la obtención de estimaciones, al afectar tanto a la elección de la forma funcional como al método de estimación.

En este aspecto la verosimilitud empírica presenta una ventaja porque al no llevar a cabo una modelización no hay que estar tan preocupados por la elección del tipo de datos, aunque por supuesto, dependiendo de para lo que vaya a utilizarse posteriormente puede resultar más adecuado un tipo que otro.

La elección concreta de las observaciones de renta depende en gran parte del objetivo del estudio. Por ejemplo, Atkinson (1.983) sugiere que la renta media anual puede ser utilizada para analizar la desigualdad en la distribución personal de la renta, pero no resulta adecuada para analizar la pobreza.

Una discusión clásica y recogida sobradamente en la literatura económica es la referida a la elección de ingresos o gastos para representar la renta.

Muchos son los trabajos que se han apoyado en las distribuciones del gasto deducidas de las Encuestas de Hogares, esgrimiendo dos razones fundamentales: Por una parte defienden que el gasto refleja mejor la renta permanente de un hogar, y por otro lado, aunque el ingreso parece el concepto más significativo, está sujeto en muchos casos a ocultación y ocasionalidad.

Así Ruiz Castillo (1.987) y Blundell y Preston (1.994), sostienen que los ingresos pueden exagerar la verdadera posición económica de los individuos en un momento determinado. Es posible que las características de ese individuo (edad, estudios,...) hagan suponer que tendrá distintos ingresos en el futuro, con lo que debería considerarse como renta lo que obtiene a lo largo de su ciclo vital, y no en un momento dado.

Esto supone una serie de problemas, de tipo teórico y práctico como Cowell (1.977) apunta. Se pueden evitar utilizando el gasto, suponiendo que los individuos realizan su gasto no en función de la renta presente, sino de la que pretenden obtener.

Sin embargo, Blundell y Preston (1.991) consideran esta aproximación válida solamente si las necesidades y preferencias de los individuos permanecen constantes a lo largo del ciclo vital, lo cual es demasiado irreal.

Frente a este punto de vista anterior, Atkinson (1.983) mantiene que los ingresos aproximan mejor la posición económica porque reflejan su capacidad de compra, que pueden ejercitar en el mismo período o trasladar al futuro.

Pena y otros (1.996) consideran que la renta y el gasto son conceptos distintos aunque están relacionados. En los tramos bajos de renta, donde la propensión marginal a consumir está cerca de 1, es aceptable aproximar la renta mediante el gasto. Pero no es así a medida que aumenta el nivel de renta, separándose cada vez más del gasto y teniendo ambas variables distribuciones distintas.

3.3.2. - Individuos u hogares.

Otra elección es la que se produce entre considerar la renta de los hogares o la de los individuos. Aparentemente es preferible utilizar la renta individual, pero no hay que perder de vista que la mayoría de las encuestas que pueden aportar información sobre este tema se refieren al hogar. Un ejemplo son las Encuestas de Presupuestos Familiares.

Además, en una gran mayoría de las familias parte del flujo rentas es fruto de la aportación de los distintos miembros .

Si se opta por utilizar el hogar como unidad preceptora de rentas pueden considerarse distintas posibilidades:

- .- Utilizar la renta total del hogar, o
- .- Tener en cuenta el tamaño y composición del hogar. Así una vía habitual es dividir los ingresos o los gastos por el número de miembros que constituyen el hogar, dando lugar a lo que se denomina renta per cápita. Otra

posibilidad, que tiene en cuenta la existencia de economías de escala en el consumo del hogar y que las necesidades de los individuos son diferentes en función de la edad y otras características, consiste en ponderar los miembros de la familia dándoles distinto peso. Son las llamadas escalas de equivalencia. La más conocida es la difundida por la OCDE, que asigna 1 al primer adulto, 0,7 al resto de adultos y 0,5 a los niños. Esto da lugar al uso del término renta equivalizada.

3.3.3. – Ingresos y ocultación.

Nosotros vamos a estar interesados en representar formalmente la estructura de ingresos de los hogares españoles. Para ello vamos a utilizar la información suministrada por la ECPF. Se ha podido obtener información referente al total de ingresos trimestrales, según su naturaleza clasificados en monetarios y no monetarios.

Los ingresos monetarios son los que resultan una vez deducidas las cantidades satisfechas en concepto de impuestos y otros pagos asimilados (Seguridad Social, derechos pasivos, mutualidades...).

La partida de ingresos totales es básicamente equiparable a la renta disponible del hogar, concepto utilizado en la Contabilidad Nacional para referirse al saldo de la cuenta de renta del sector Hogares.

Por tanto, según la información aportada por la Encuesta podemos diferenciar entre:

1.- Ingresos monetarios. Se dividen según su origen en:

- Ingresos por trabajo por cuenta ajena;
- Ingresos por trabajo por cuenta propia;
- Rentas de capital y propiedad;
- Pensiones;
- Prestaciones de desempleo;
- Otras transferencias regulares;
- Otros ingresos monetarios.

2.- Ingresos no monetarios. Se entienden los obtenidos:

- Por trabajo por cuenta ajena;
- Por trabajo por cuenta propia;
- De rentas de capital y propiedad;
- Por otros ingresos no monetarios.

La mayor parte de los autores coinciden en que el objetivo de las Encuestas de Presupuestos no es fundamentalmente conocer los ingresos de los hogares, sino que el diseño de la encuesta está orientado a conocer los gastos de consumo. Esto hace que la obtención y seguimiento de los datos de gasto sea más minucioso que el de los ingresos. De cometerse un error (olvido, cantidades equivocadas), es más probable éste sea en los ingresos y no en los gastos.

Un problema generalizado, en el sentido de que ocurre en este tipo de encuestas, es lo que se ha dado en llamar subdeclaración. Por este término hemos de entender que los ingresos declarados por los hogares son inferiores a los que deberían haberse comunicado. Esto se pone de manifiesto en la mayoría de los casos porque muchos hogares (más de los que cabría esperar), declaran tener más gastos que ingresos para el período de tiempo considerado, y porque al contrastar las partidas de ingresos totales con los datos agregados de la Contabilidad Nacional, se observa que hay una gran discrepancia a la baja, por parte de los ingresos declarados (ver Pena y otros (1.996)).

Alcaide y Alcaide (1.974; 1.977; 1.983) estudian en algunos trabajos el tema de la ocultación de rentas, observando la relación entre los datos de la Encuesta de Presupuestos Familiares y los datos de la Contabilidad Nacional.

Recogiendo esta idea, Pena y otros (1.996) desarrollan un método para corregir esta ocultación, ya que como Prieto y Pena (2.000) señalan, puede influir en los estudios de desigualdad y otros análisis que se lleven a cabo.

La metodología propuesta para corregir esta ocultación consiste en modificar los datos originales de ingresos de cada uno de los hogares entrevistados en la Encuesta Básica de Presupuestos Familiares, para que fuesen consistentes con lo obtenido de la Contabilidad Nacional. Primero se hace una corrección en los niveles de ingreso, (ver Merediz y otros (1.996)), y

posteriormente una corrección en la distribución (ver Callealta y otros (1.996)).

En esta línea, Ortega y otros (2.001) estudian el tratamiento de la ocultación mediante métodos robustos.

La mayor parte de los autores aceptan que no todos los ingresos presentan el mismo grado de subdeclaración. Los ingresos no monetarios no se ven tan afectados por este hecho, y para los monetarios la subdeclaración es diferente según los tramos de renta. La ocultación recae más en las clases de mayores ingresos porque estas clases acceden más fácilmente a bienes generadores de renta, y es precisamente en las rentas de capital y de la propiedad donde se produce una mayor ocultación.

En esta Memoria no se ha llevado a cabo corrección alguna de los datos suministrados por la encuesta. No obstante, en el epígrafe 4.6 del capítulo correspondiente al gasto, se hace referencia a la posible presencia de ocultación y cómo esto es recogido y puesto de manifiesto por nuestra metodología.

3.4. – Análisis de los ingresos en unidades monetarias, según su naturaleza, a partir de la clasificación de la Encuesta Continua de Presupuestos Familiares.

Antes de entrar en el estudio de los distintos tipos de ingresos, así como de las participaciones de las clases de ingresos en el total, y al igual que se hace en otros capítulos, exponemos un cuadro resumen las principales medidas descriptivas para las tres variables suministradas por la ECPF y de las otras dos construídas a partir de la misma.

Figura 3.1.

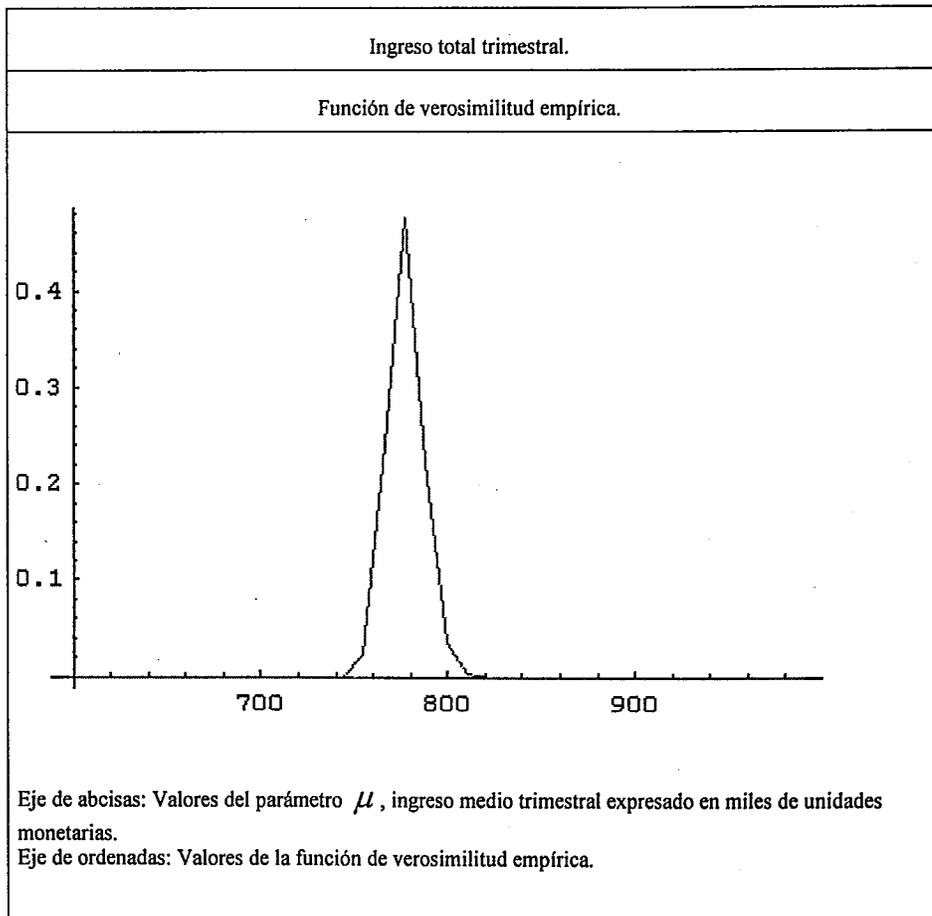
POSICIÓN CENTRAL	POSICIÓN		DISPERSIÓN	FORMA
Media Mediana	m M C10	C25 C75 C90	R Desv.Tip Coef.Var C90-C10	Asimetría Curtosis
Ingreso total				
776.772,4 671.782,4	40.000 10.202.000 302.053,4	456.996,5 974.208,6 1.346.210,7	10.162.000 517.369,3 0,67 1.044.157,2	4,71 59,28
Ingresos monetarios				
643.485,8 555.000,6	0 9.994.000 223.376,5	352.001,1 812.750,5 1.150.013,1	9.994.000 470.840,5 0,73 926.636,5	5,51 78,67
Ingresos no monetarios				
133.286,6 118.934,5	0 1.800.000 0	75.000 172.500 244.916,9	1.800.000 107.052,1 0,80 244.916,9	2,93 24,56
Participación de los ingresos monetarios en el total				
0,812 0,826	0 1 0,667	0,75 0,891 1	1 0,13 0,16 0,333	-1,639 6,95
Participación de los ingresos no monetarios en el total				
0,188 0,174	0 1 0	0,109 0,250 0,333	1 0,13 0,694 0,333	1,639 6,951

3.4.1. – Ingreso total trimestral.

Esta variable se ha obtenido a partir de las observaciones correspondientes a los ingresos monetarios y no monetarios, mediante agregación.

Obtuvimos un total de 3.114 observaciones y con ellas calculamos la función de verosimilitud empírica. El máximo de la verosimilitud se alcanza para el valor del parámetro de 770,09, expresado en miles de unidades monetarias.

Figura 3.2.



Los intervalos de verosimilitud vienen dados por:

I. Verosimilitud para una fuerza de evidencia 8(765,91 ; 788,27)
I. Verosimilitud para una fuerza de evidencia 32(754,72 ; 799,46)

Si notamos por “a” el valor de mayor verosimilitud y por L(a) la verosimilitud correspondiente a ese valor, obtenemos los siguientes cocientes:

$$\begin{aligned}L(a) / L(743,54) &= 1587,34. \\L(a) / L(754,72) &= 21,29. \\L(a) / L(765,91) &= 1,97. \\L(a) / L(788,27) &= 2,12. \\L(a) / L(799,46) &= 13,57. \\L(a) / L(810,64) &= 181,21.\end{aligned}$$

Estos cocientes nos indican, por ejemplo, que el valor para el que se alcanza el máximo es 1,96 veces más verosímil que 765,91 ó 2,12 veces más que 788,27. Su verosimilitud es 13 veces y media mayor que la del valor 799,46...

Estas relaciones aumentan a medida que nos salimos de los valores de los intervalos. Así, la verosimilitud del máximo es 21,29 veces mayor que la de 754,72, ó 181 veces ó 1587 veces,..., haciéndose esa fuerza de la evidencia mucho mayor conforme nos alejamos.

Si utilizamos una distribución a priori Uniforme junto con la verosimilitud modificada, obtendríamos una verosimilitud normalizada, sobre la que se pueden calcular probabilidades, y por tanto, intervalos de confianza.

Exponemos a continuación una tabla con algunos percentiles de la verosimilitud normalizada:

Orden del percentil	0,05	0,25	0,5	0,75	0,95	0,99
Valor del percentil	755,99	765,27	771,47	777,65	787,66	797,14

Estimación puntual y regiones de confianza para el ingreso medio trimestral.

Para la estimación puntual vamos a utilizar el estimador de máxima verosimilitud. En este caso, el ingreso medio estimado es 777,09 (miles de unidades monetarias).

Una estimación por intervalo usando los percentiles de la verosimilitud normalizada vendría dada por:

I. Confianza (80%).....(755,99 ; 782,66)
I. Confianza (90%).....(755,99 ; 787,66)
I. Confianza (95%).....(754,84 ; 792,35)

Podemos decir que el ingreso medio se encuentra entre estos valores con una confianza del 80, 90 y 95 por ciento respectivamente.

También podríamos conocer el nivel de confianza de los intervalos de verosimilitud.

Así, el intervalo de verosimilitud para una fuerza de evidencia 8 obtenido, (765,91; 788,27), es aproximadamente el intervalo de confianza del 65%.

Por otro lado, el intervalo de verosimilitud para una fuerza de evidencia 32, (754,72; 799,46) es el intervalo de confianza del 96,5%.

Estos intervalos de confianza pueden interpretarse en términos de la evidencia, presentando gran fuerza en el sentido de grandes cocientes de verosimilitudes en relación a valores de fuera del intervalo.

Por ejemplo,

$$L(765,91) / L(754,72) = 10,82.$$

$$L(765,91) / L(743,54) = 806,27.$$

$$L(765,91) / L(810,64) = 92,04.$$

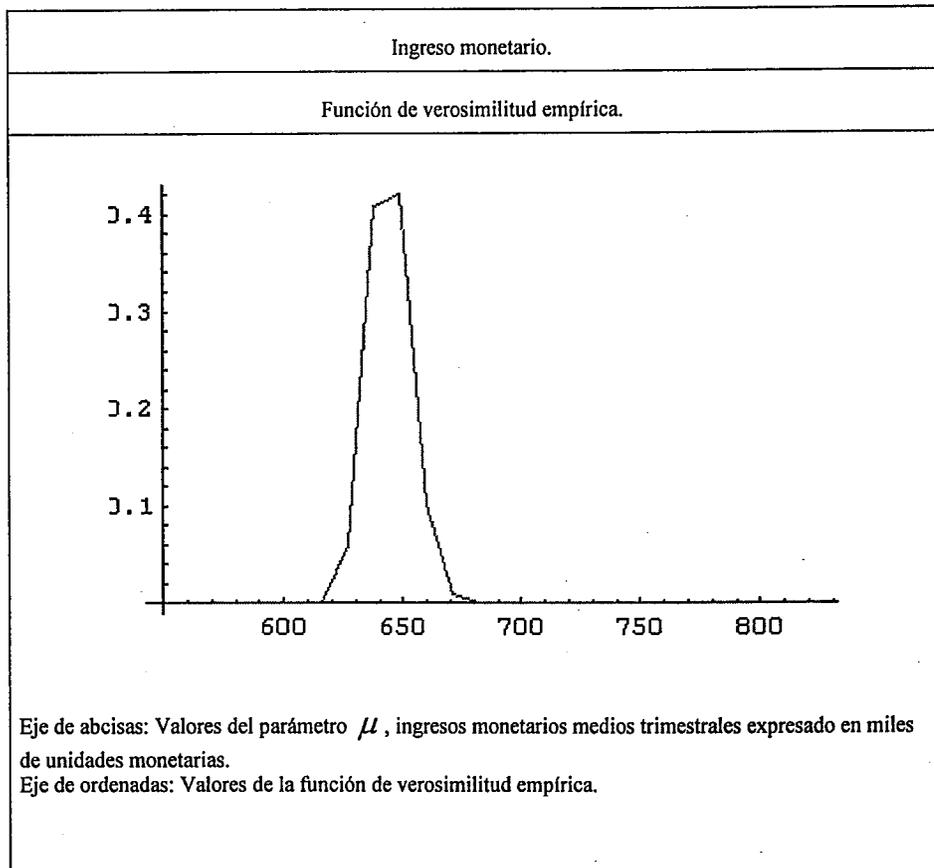
$$L(788,91) / L(743,54) = 746,82.$$

$$L(788,91) / L(754,72) = 10,02.$$

3.4.2. - Ingreso monetario.

Obtenida la función de verosimilitud empírica se observa que el valor que presenta una mayor verosimilitud es 648,61 en miles de unidades monetarias.

Figura 3.3.



Los principales resultados obtenidos conforme a la tabla que presentamos en un capítulo anterior son los siguientes:

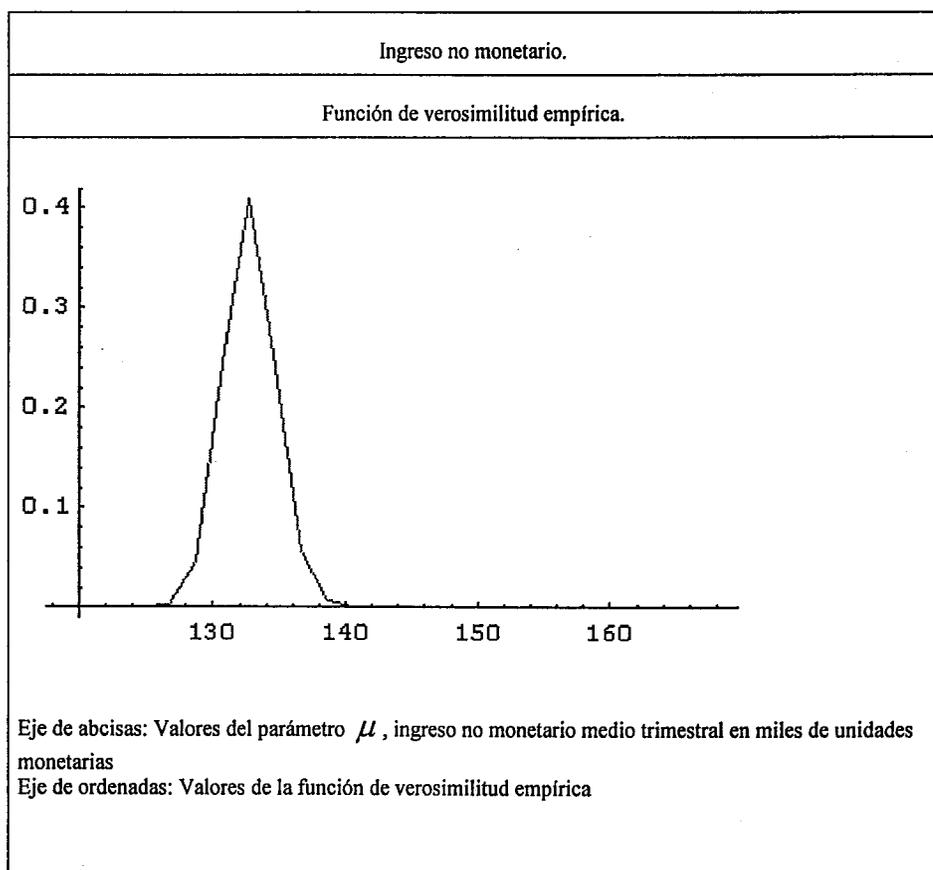
Figura 3.4.

Ingreso monetario en miles de unidades monetarias.						
Intervalo para una fuerza de evidencia 8.		(626,62 ; 659,60)				
Intervalo para una fuerza de evidencia 32.		(626,62 ; 659,60)				
Cocientes entre el valor de la función de verosimilitud empírica en el máximo y el valor de la función de verosimilitud empírica para otros valores cercanos en la gráfica.		$L(a)/ L(615,63) = 465,75;$ $L(a)/ L(626,62) = 7,14;$ $L(a)/ L(637,62) = 1,03;$ $L(a)/ L(659,60) = 4,18;$ $L(a)/ L(670,60) = 45,83;$				
Orden del percentil	0,05	0,25	0,5	0,75	0,95	0,99
Valor del percentil	624,77	631,75	638,46	644,97	655,21	659,27
Estimación puntual.		648,61.				
Intervalos de confianza	80%	(624,77 ; 647,58)				
	90%	(624,77 ; 655,21)				
	95%	(620,11 ; 657,94)				
Nivel de confianza de los intervalos de verosimilitud expuestos al principio.		(626,62 ; 659,60).....94%				
		(626,62 ; 659,60).....94%				
Algunos cocientes del valor de la función de verosimilitud empírica en los valores de los intervalos anteriores y el valor de la función de verosimilitud empírica en valores de fuera de esos intervalos.		$L(626,62) / L(615,63) = 65,26;$ $L(637,62) / L(604,64) = 371.286,78;$ $L(648,61) / L(670,60) = 45,83;$ $L(648,61) / L(615,63) = 465,75;$ $L(659,60) / L(681,59) = 204,66;$ $L(637,17) / L(681,59) = 827,29;$				

3.4.3. - Ingreso no monetario.

Función de verosimilitud empírica:

Figura 3.5.



Principales resultados:

Figura 3.6.

Ingreso no monetario, en miles de unidades monetarias.						
Intervalo para una fuerza de evidencia 8		(130,68 ; 136,62)				
Intervalo para una fuerza de evidencia 32		(128,70 ; 136,62)				
Cocientes entre el valor de la función de verosimilitud empírica en el máximo y el valor de la función de verosimilitud empírica para otros valores cercanos en la gráfica.		$L(a)/L(126,72) = 177,70;$ $L(a)/L(128,7) = 9,18;$ $L(a)/L(134,64) = 1,71;$ $L(a)/L(136,62) = 7,45;$ $L(a)/L(138,6) = 69,80;$				
Orden del percentil	0,05	0,25	0,5	0,75	0,95	0,99
Valor del percentil	128,72	130,35	131,69	133,08	135,04	136,48
Estimación puntual.			132,66.			
Intervalos de confianza		80%	(128,72 ; 133,90)			
		90%	(128,72 ; 135,04)			
		95%	(127,72 ; 135,94)			
Nivel de confianza de los intervalos de verosimilitud expuestos al principio.		(130,68 ; 136,62).....69%				
		(128,70 ; 136,62).....94%				
Algunos cocientes del valor de la función de verosimilitud empírica en los valores de los intervalos anteriores y el valor de la función de verosimilitud empírica en valores de fuera de esos intervalos.		$L(130,68) / L(126,72) = 105,30;$ $L(130,68) / L(138,6) = 41,36;$ $L(132,66) / L(138,6) = 69,80;$ $L(134,64) / L(124,74) = 7.815,17;$ $L(136,62) / L(140,58) = 152,69;$				

3.4.4. – Participación de los distintos tipos de ingreso en el total.

Ya que nuestra intención es conocer el fenómeno de los ingresos en los hogares españoles, parece interesante la posibilidad de analizar de que manera participan los dos tipos de ingresos que hemos considerado en el ingreso total.

Para ello construimos unas nuevas variables que reflejaban para cada hogar la participación que en el ingreso total tenían los dos tipos de ingresos. Se trabajó sólo con dos cifras decimales a la hora de obtener la verosimilitud empírica. Las estimaciones de las participaciones suman en este caso 1,02.

Figura 3.7.

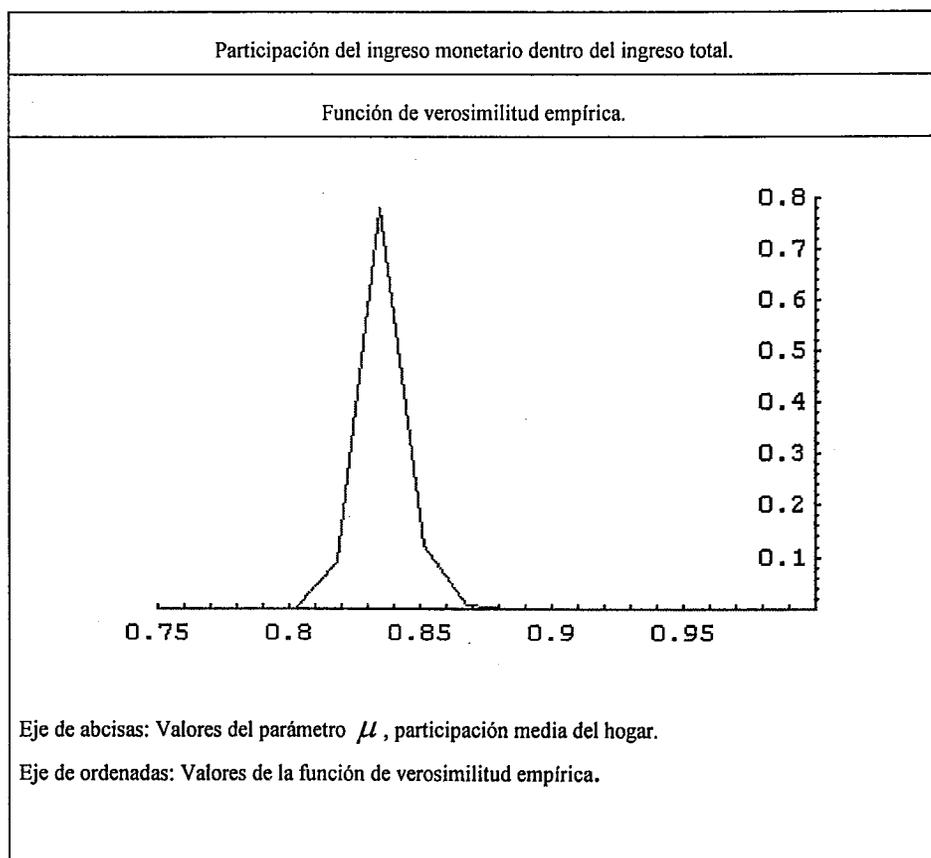


Figura 3.8.

Participación del ingreso monetario dentro del ingreso total.						
Intervalo para una fuerza de evidencia 8.		(0,835 ; 0,851)				
Intervalo para una fuerza de evidencia 32.		(0,818 ; 0,851)				
Cocientes entre el valor de la función de verosimilitud empírica en el máximo y el valor de la función de verosimilitud empírica para valores cercanos en la gráfica.		$L(a)/L(0,802) = 2,02 \times 10^9$; $L(a)/L(0,818) = 8,56$; $L(a)/L(0,851) = 6,38$; $L(a)/L(0,868) = 108,72$;				
Orden del percentil	0,05	0,25	0,5	0,75	0,95	0,99
Valor del percentil	0,811	0,822	0,827	0,832	0,846	0,851
Estimación puntual.			0,835.			
Intervalos de confianza		80%	(0,811 ; 0,834)			
		90%	(0,811 ; 0,846)			
		95%	(0,806 ; 0,849)			
Nivel de confianza de los intervalos de verosimilitud expuestos al principio.		(0,835 ; 0,851)		19%		
		(0,818 ; 0,851)		89%		
Algunos cocientes del valor de la función de verosimilitud empírica en valores de los intervalos anteriores y el valor de la función de verosimilitud empírica en valores de fuera de esos intervalos.		$L(0,835)/L(0,802) = 2,02 \times 10^9$; $L(0,835)/L(0,868) = 108,72$; $L(0,835)/L(0,884) = 2.765,98$; $L(0,851)/L(0,802) = 3,16 \times 10^8$; $L(0,851)/L(0,868) = 17,04$; $L(0,851)/L(0,884) = 433,59$;				

Figura 3.9.

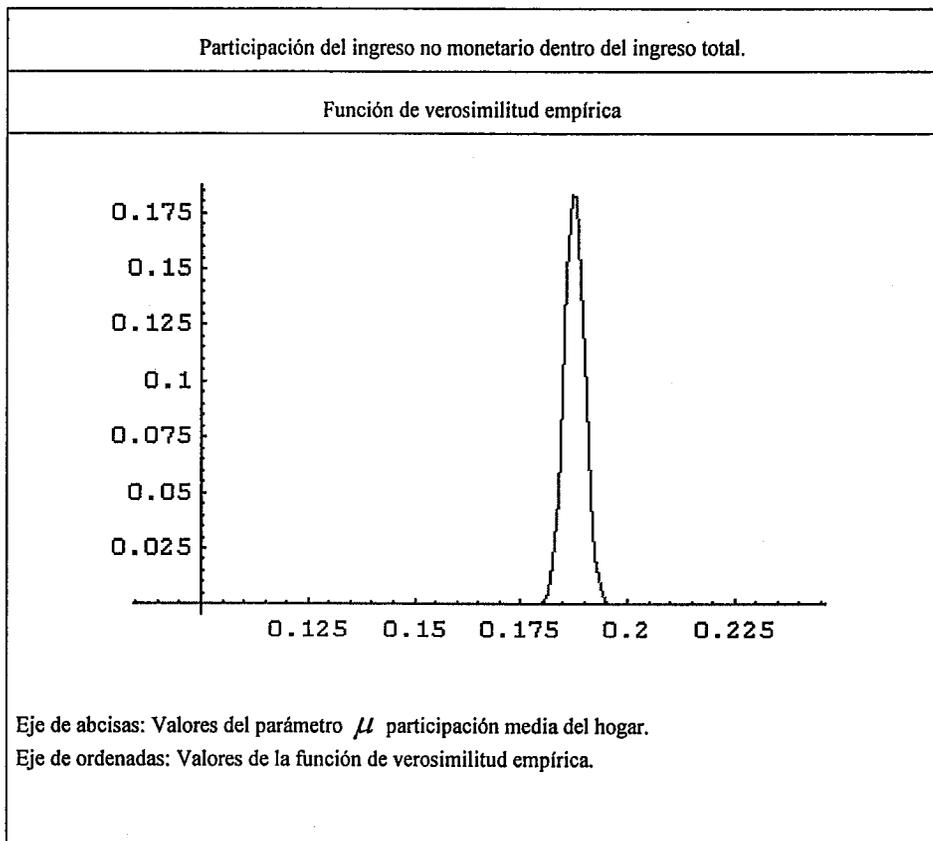


Figura 3.10.

Participación del ingreso no monetario dentro del ingreso total.						
Intervalo para una fuerza de evidencia 8.		(0,183 ; 0,191)				
Intervalo para una fuerza de evidencia 32.		(0,181 ; 0,194)				
Cocientes entre el valor de la función de verosimilitud empírica en el máximo y el valor de la función de verosimilitud empírica para otros valores cercanos en la gráfica.		L(a)/ L(0,184) = 3,85; L(a)/ L(0,185) = 1,94; L(a)/ L(0,190) = 1,96; L(a)/ L(0,191) = 3,74; L(a)/ L(0,192) = 8,71;				
Orden del percentil	0,05	0,25	0,5	0,75	0,95	0,99
Valor del percentil	0,183	0,185	0,187	0,188	0,191	0,192
Estimación puntual.			0,187.			
Intervalos de confianza	80%	(0,183 ; 0,189)				
	90%	(0,183 ; 0,191)				
	95%	(0,182 ; 0,192)				
Nivel de confianza de los intervalos de verosimilitud expuestos al principio		(0,183 ; 0,191).....90%				
		(0,181 ; 0,194).....98%				
Algunos cocientes del valor de la función de verosimilitud empírica en los valores de los intervalos anteriores y el valor de la función de verosimilitud empírica en valores de fuera de esos intervalos		L(0,185) / L(0,180) = 67,33; L(0,187) / L(0,179) = 703,78; L(0,189) / L(0,195) = 65,78; L(0,191) / L(0,187) = 1.312,78; L(0,192) / L(0,196) = 38,67; L(0,194) / L(0,178) = 200,211;				

3.5. – Comentarios y utilización de los resultados obtenidos.

Se ha utilizado la metodología de la verosimilitud empírica con datos de renta para los hogares españoles, suministrados por la ECPF.

Como técnica no paramétrica tiene la ventaja de que no impone un modelo de partida, permitiendo una mayor flexibilidad para que los datos reflejen la realidad. Supone un ahorro en el sentido de que todo el trabajo de selección de un modelo, la estimación y el análisis de la bondad del ajuste no se lleva a cabo.

Al mismo tiempo también se diferencia de la estimación no paramétrica de la función de densidad. Aquí se obtiene la función de verosimilitud empírica, la cuál directamente nos va a servir para conocer:

- .- La distribución de ingresos medios por hogar;
- .- La distribución de ingresos monetarios medios por hogar,
- .- La distribución de ingresos no monetarios medios por hogar;
- .- La composición de los ingresos por hogar.

Comentemos brevemente cada uno de estos puntos:

a) Hemos obtenido que 770,09 es el ingreso medio por hogar más probable, medido en miles de unidades monetarias. De aquí en adelante, y de igual modo para el capítulo de los gastos, esas van a ser las unidades de medida en que vengán expresados los resultados, a excepción de las participaciones de ingresos y gastos en el total.

También conocemos entre que valores se van a encontrar esos ingresos medios con una elevada probabilidad. La media de los ingresos por hogar está entre 754,84 y 792,35 con una probabilidad del 95%.

Por otro lado se dispone de información sobre cuales son los valores de los ingresos medios que presentan mayor verosimilitud. Se ha comentado en el epígrafe anterior como deben interpretarse esas comparaciones entre valores. Algunos tienen mayor verosimilitud de que se produzcan y dentro de éstos los

valores son comparables dos a dos, quedando claramente establecido cual de ellos presenta mayor evidencia de realización.

Conocer el ingreso medio por hogar tiene importancia no sólo en sí mismo, sino que permite además otros estudios y comparaciones que son muy interesantes.

En primer lugar nos referimos a poder conocer la evolución de esos ingresos a lo largo del tiempo. Comparando las cantidades respecto a un mismo año base podemos establecer el comportamiento de los ingresos medios por hogar, y evaluar si ha habido crecimiento o no, a que ritmo,....

Un segundo aspecto a destacar es la posibilidad de comparar los ingresos medios de zonas geográficas (países, comunidades o provincias) diferentes, lo cual nos permite establecer una clasificación de zonas por nivel de ingresos. Puede realizarse ese estudio simplemente utilizando encuestas que estén disponibles y que dispongan de datos para los hogares de cada zona.

En esta línea de intentar conocer el mayor o menor grado de pobreza de una zona, además de poder comparar el ingreso más probable de diferentes regiones, está la posibilidad de utilizar los percentiles de la distribución. Estos valores nos dan el porcentaje de hogares cuyos ingresos medios son inferiores a una determinada cantidad. No es lo mismo que el 25% de los hogares tengan unos ingresos medios inferiores a 75.000 unidades monetarias a que sean el 75% de los hogares.

Si se establece lo que se conoce como línea de pobreza, cantidad mínima por debajo de la cual se es considerado como pobre, es posible conocer el porcentaje de hogares que no alcanzan dicho nivel, y al mismo tiempo se pueden establecer comparaciones entre hogares correspondientes a provincias o comunidades distintas, que ponga de manifiesto si existen diferencias regionales por niveles de pobreza.

También es posible usar esta metodología para ver si algunas políticas han tenido o no efectos en la reducción de los niveles de pobreza. Dado que se dispone de información cada muy poco tiempo, los resultados se actualizan rápidamente y se puede analizar si hay cambios o cuanto tiempo han tardado en notarse los posibles efectos.

Otro posible uso es la elección entre afirmaciones alternativas acerca del valor que tiene los ingresos medios de los hogares. Con este método es sencillo conocer la probabilidad de ambas hipótesis y descartar por tanto la afirmación que tiene menos probabilidad.

b) También es posible obtener resultados usando los datos suministrados por la ECPF acerca de los ingresos medios de tipo monetario. Se obtiene la función de verosimilitud empírica y se observa que el valor del ingreso medio monetario por hogar más probable es de 648,61.

En este caso es posible, al igual que los ingresos totales, realizar un análisis comparativo entre hogares de distintas zonas, ver la evolución temporal de la magnitud considerada o escoger entre alternativas.

Sabemos que la media de estos ingresos monetarios tiene una probabilidad del 90% de encontrarse entre 624,77 y 655,21.

Por otro lado sabemos que los valores que presentan una verosimilitud que podemos llamar fuerte son los comprendidos entre 626,62 y 659,60. Al mismo tiempo hay una probabilidad muy alta, casi el 95% de que los ingresos monetarios medios están entre esos valores.

Otra posibilidad interesante que merece la pena señalar es que dada la clasificación de los ingresos según su origen, se puede obtener la verosimilitud empírica para las prestaciones de desempleo. También es posible conocer la participación que estas prestaciones representan dentro de los ingresos totales, y eso puede aportar información indirecta sobre el desempleo de esa región. Por supuesto también podría utilizarse con fines comparativos y supone una manera alternativa de comprobar el efecto de políticas encaminadas a reducir el nivel de desempleo.

c) De igual modo es fácil obtener la distribución de los ingresos medios no monetarios. Como era de esperar el valor medio más probable de esos ingresos es muy reducido en comparación con los monetarios. El valor medio por hogar más probable es 132,66. La probabilidad de que este valor medio se encuentre entre 128,72 y 135,04 es del 90%.

Por otro lado sabemos que el valor 133,66 es casi 178 veces más verosímil que el valor 126,72; es poco más de 9 veces más verosímil que

128,7 o que tiene una verosimilitud casi dos veces más grande que el valor 134,64.

d) Esta técnica también nos permite conocer cual es la proporción media que representa una categoría de ingresos dentro del total, lo cual nos puede dar una idea de la estructura de ingresos de los hogares españoles. De nuevo hay muchas posibilidades de comparación, dependiendo del tipo de ingresos y de hogares que se utilicen para el estudio.

En este caso se ha optado por analizar la proporción media de los dos tipos de ingresos considerados. A partir de los datos originales es sencillo diseñar una nueva variable que podríamos llamar "proporción de ingresos monetarios en el total" u otra, "proporción de ingresos no monetarios en el total". De nuevo disponemos de 3.114 observaciones, una para cada hogar, y se calcula la verosimilitud empírica para la proporción media.

Parece obvio que el peso que tendrá los ingresos monetarios será mayor que el de los ingresos no monetarios. Pero, ¿cuánto?. La proporción media por hogar para los ingresos monetarios es del 83%, mientras que en los no monetarios supone un 18%.

Otra posibilidad es calcular el porcentaje que representa el ingreso monetario en el total para toda la población. Sumando para todos los hogares los ingresos monetarios y dividiéndolos por la suma de los ingresos de todos los hogares, obtendríamos para la población ese porcentaje. Disponiendo de esa información para varios períodos, sin más que acudir a la ECPF, tendríamos una serie de observaciones a las que podríamos aplicar esta metodología y obtener la proporción media para la población.

Estas son algunas de las utilidades que presenta la verosimilitud empírica relacionadas con el tema de la renta. Todos los estudios de los que hemos hablado podrían extenderse a comparaciones entre países sin más que disponer de datos a ese respecto.

Realmente, lo que hace digna de interés a esta metodología es su enorme versatilidad, ya que puede ser aplicada en cualquiera de las parcelas que hemos comentado anteriormente, así como el aportar la ausencia de hipótesis, que el caso de la renta son muy restrictivas y encorsetan un poco los resultados. Es un método que incorpora la ventaja de las técnicas no paramétricas, la flexibilidad, pero al mismo tiempo es de gran sencillez en los

conceptos, proporciona facilidad y rapidez en la obtención de resultados y presenta una gran capacidad para ir incorporando nuevos datos y reajustarse.

CAPÍTULO 4

ANÁLISIS DE LOS GASTOS EN UNIDADES MONETARIAS SEGÚN LA CLASIFICACIÓN DE LA ENCUESTA CONTINUA DE PRESUPUESTOS FAMILIARES.

4.1. – Introducción y planteamiento.

Desde que en 1.857, Engel inicia sus trabajos relacionados con el presupuesto familiar, dando lugar a las llamadas Leyes de Engel, una de las ramas de la Economía que más interés ha despertado es el análisis de los presupuestos familiares. En la mayor parte de los países desarrollados existen encuestas periódicas específicas que recogen la distribución del gasto familiar entre los distintos bienes de consumo. En España, la Encuesta de Presupuestos Familiares cubre todos los gastos realizados por los hogares encuestados durante un período determinado, especificando completamente la naturaleza del gasto.

Las Curvas de Engel tratan de establecer modelos teóricos que reflejen el comportamiento o conducta de los consumidores (en general de una categoría de gastos) frente a los ingresos o gasto total. Su aplicación principal es la obtención de elasticidades de demanda. Esto obliga a disponer de curvas

teóricas que modelicen correctamente el comportamiento de estas variables o bien funciones de elasticidad.

Se suelen presentar las curvas de demanda como la relación entre la cantidad demandada del bien y el precio de ese bien, así como el precio de otros bienes y la renta. La elección de esa función que relaciona unas variables con otras condiciona los resultados obtenidos.

Históricamente, la selección de la curva dependía de que se ajustase bien, fuese flexible para detectar cambios, estimable con los datos de que se disponía...

Estas curvas básicamente son una estimación de una función de regresión con datos atemporales, que obedece al siguiente esquema:

$$Y=f(x)+\varepsilon$$

donde Y es el nivel de gasto, x son los ingresos, y ε es un término de perturbación aleatoria.

Desde el punto de vista metodológico hay trabajos que tratan la estimación de la regresión individual, en donde las modelizaciones más utilizadas son:

- a) La función lineal: $Y = a + bx + \varepsilon$;
- b) La función potencial: $Y = ax^b + \varepsilon$;
- c) La función semilogarítmica: $Y = a + bLn x + \varepsilon$;
- d) La función hiperbólica: $Y = a + \frac{b}{x} + \varepsilon$;
- e) La función logarítmica inversa: $\log\left(\frac{1}{Y}\right) = a + \frac{b}{x} + \varepsilon$;
- f) La función inversa: $\frac{1}{Y} = a + \frac{b}{x} + \varepsilon$;

El lector interesado puede consultar Leser (1.963).

No obstante, ninguna de estas modelizaciones es universalmente aceptada porque salvo la segunda forma funcional, en todas ellas disminuye la

elasticidad-renta al aumentar la renta. Esta situación es factible para bienes inferiores pero no se da en otro tipo de bienes.

También existen extensiones de los modelos anteriores en trabajos aplicados tratando de ajustar de un modo más efectivo los datos manejados. Así aparece la especificación y estimación de sistemas completos de demanda. Ejemplos de ellos son el sistema de demanda cuadrático, QES, (ver Pollack y Wales (1.978), Howe y otros (1.979)), o el sistema de demanda Casi Ideal (AIDS), propuesto por Deaton y Muellbauer (1.980).

En el caso español tiene especial relevancia este último sistema, especialmente aplicado al consumo de alimentos. Destacan los trabajos de Molina (1.993); García y Molina (1.995); Gracia y Albisu (1.994; 1.995); Ramajo Hernández (1.996); Laajimi y Albisu (1.997); Ben Kaabia y Gil (2.000).

En algunas de estas aplicaciones se incorporan variables de tipo socioeconómico para agrupar los datos y posteriormente ajustar en cada grupo una curva de Engel.

Los distintos modelos responden a un enfoque paramétrico cuyo objetivo fundamental es obtener información sobre los diferentes condicionantes de los gastos de consumo total, así como estimar las elasticidades-renta o elasticidad-gasto de los distintos grupos de artículos.

Otros trabajos, Lewbel (1.991); Del Oro y Riobóo (1.997); y Del Oro; Riobóo y Rodríguez (2.000), persiguen objetivos similares pero desde una óptica no paramétrica, al considerar que la teoría económica no ofrece suficiente información sobre el modelo paramétrico adecuado para representar la curva de Engel, obtando por la flexibilidad que supone el hecho de no tener que imponer que pertenezca a una determinada familia de funciones.

En esta Memoria presentamos una vía alternativa para obtener información sobre la estructura de los gastos de los hogares españoles. Aquí no se estima ninguna función de regresión y por tanto directamente no vamos a obtener conclusiones en base a variables explicativas, dado que nuestro enfoque no las utiliza, ni podremos conocer las elasticidades de demanda.

En cambio, si conoceremos la estructura de gasto de los hogares españoles, el volumen medio de gasto en los distintos grupos y las

participaciones de gasto de cada grupo en el total, además de otras posibilidades que comentaremos más adelante.

Conocer la estructura del gasto, compararla con la de diferentes países, es el objeto de diversos trabajos (ver Parras y Torres (1.996); Alonso y Donoso (1.986); Gracia y Albisu (1.986); y Guillén y Campa (1.994)). Al mismo tiempo, también puede utilizarse la estructura de gasto para ver su repercusión en otros temas, como por ejemplo en el endeudamiento de los hogares, (ver De la Cuesta González (1.993))

Nosotros vamos a estudiar la estructura de gasto de los hogares españoles utilizando la verosimilitud empírica con datos obtenidos de la ECPF. La encuesta establece una clasificación en 9 grupos de gasto que son:

- Grupo 1: Alimentos, Bebidas y Tabacos
- Grupo 2: Vestido y Calzado.
- Grupo 3: Vivienda, Calefacción, Alumbrado.
- Grupo 4: Muebles, Artículos de mobiliario. Artículos de menaje para el hogar y gastos de conservación corrientes de la casa.
- Grupo 5: Medicinas y otros productos farmacéuticos.
- Grupo 6: Transportes y Comunicaciones.
- Grupo 7: Esparcimiento, Espectáculos, Enseñanza y Cultura.
- Grupo 8: Otros bienes y servicios.
- Grupo 9: Otros gastos no mencionados anteriormente.

Las partidas que componen estos grupos pueden consultarse en el Anexol.

Se dispone de información para 3.114 hogares. El gasto total se ha obtenido por agregación de las distintas partidas.

En primer lugar exponemos un cuadro resumen con las principales características descriptivas, tanto para el gasto total como para cada uno de los grupos.

A continuación se ha obtenido la verosimilitud empírica para esos mismos grupos, así como para las participaciones de cada grupo de gasto dentro del total. Se acompaña también un cuadro resumen con los principales resultados obtenidos.

Figura 4.1.

POSICIÓN CENTRAL	POSICIÓN		DISPERSIÓN	FORMA
Media Mediana	m M C10	C25 C75 C90	R Desv.Tip Coef.Var C90-C10	Asimetría Curtosis
Gasto total				
751.502,3 638.673,5	17.899,8 4.613.783,4 285.586,4	431.651,4 938.061,8 1.298.124,4	4.595.883,6 499.089,5 0,66 1.012.538,4	2,37 9,52
Grupo 1: Alimentos, Bebidas y Tabacos				
180.580,8 158.814,5	0 3.647.579 60.595,6	100.850,7 227.054,7 314.632,7	3.647.579 136.708,5 0,76 254.037,1	7,81 160,06
Grupo 2: Vestido y Calzado				
61.741 36.150	0 967.500 0	10.500 84.562,5 154.920	967.500 77.269,4 1,25 154.920	3,11 18,57
Grupo 3: Vivienda, Calefacción y Alumbrado				
187.931,7 157.385,4	5.803,8 2.529.427,3 69.913,8	105.633,9 227.015,4 320.549,1	2.523.623,4 145.981,2 0,78 250.635,2	4,5 42,1
Grupo 4: Muebles, Artículos de mobiliario. Menaje y conservación				
45.311,3 17.988,5	0 2.279.155 0	5.850 47.239 105.826,8	2.279,155 103.474,8 2,28 105.826,8	10,94 189,4

Grupo 5: Medicinas y otros productos farmacéuticos				
20.526,6	0	0	1.060.965	7,26
2.340	1.060.965	21.907,5	48.845,4	94,22
	0	56.461,2	2,38	
			56.461,2	
Grupo 6: Transportes y comunicaciones				
92.473,1	0	15.120,2	3.380.857	9,16
50.086,2	3.380.857	102.698,6	205.577,4	106,21
	5.756,2	188.188,1	2,22	
			182.431,8	
Grupo 7: Esparcimiento, Espectáculos, Enseñanza y Cultura				
53.879,5	0	3.196,2	915.137	3,32
25.372,3	915.137	70.287,2	80.668,3	17,25
	0	141.447,8	1,5	
			141.447,8	
Grupo 8: Otros bienes y servicios				
85.086,4	0	18.136,5	1.119.863	3,05
52.861,5	1.119.863	116.200	101.834,4	16,74
	4.429,5	205.904,6	1,20	
			201.475,1	
Grupo 9: Otros gastos no mencionados anteriormente				
23.971,7	0	0	1.883.008	10,52
2.600	1.883.008	19.494	73.665,1	182,72
	0	58.293,2	3,07	
			58.293,2	

Entre todos los grupos, el más homogéneo respecto al gasto, medido por el coeficiente de variación de Pearson, es el grupo de los Alimentos, Bebidas y Tabacos, (grupo 1), y el de menor homogeneidad el grupo 9, ya que se dan valores de 0,76 y 3,07 respectivamente.

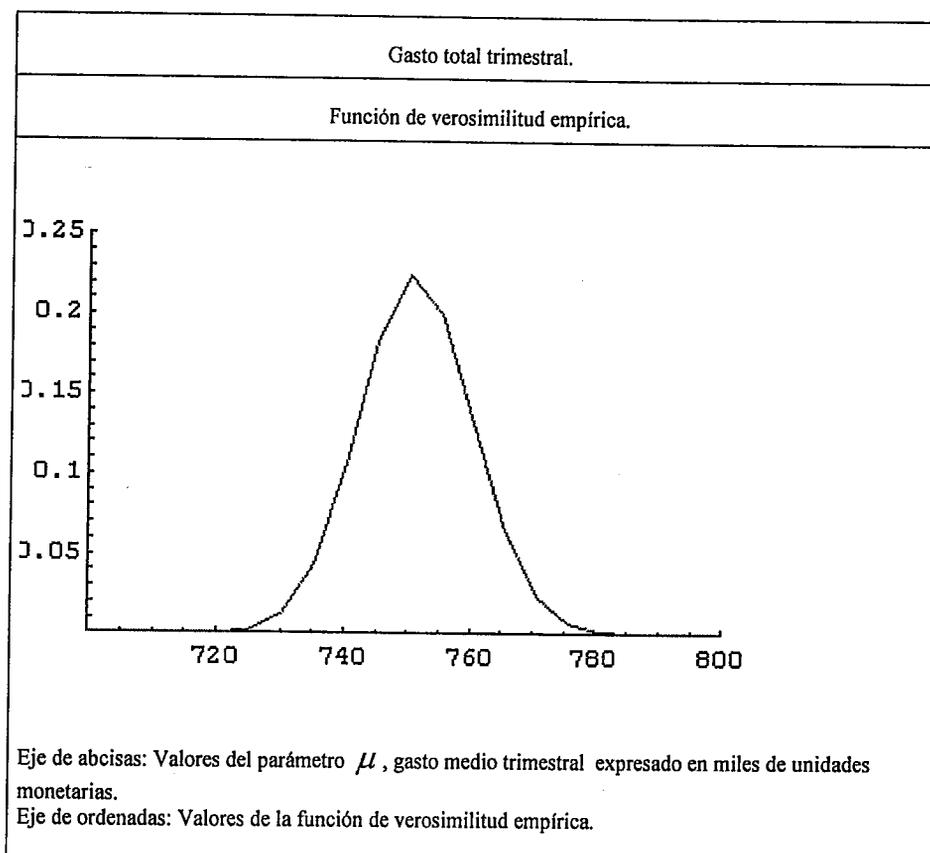
En todos los grupos se observa asimetría a la derecha, siendo el grupo 4, Muebles y otros, y el grupo 8, Otros Bienes y Servicios, los que tienen mayor y menor valor respectivamente, 10,94 y 3,05. Todos los grupos

presentan un apuntamiento mayor que la Normal, oscilando entre el valor 16,74 del grupo 8 y el 189,4 del grupo 4.

4.2. - Análisis del gasto total trimestral de los hogares españoles.

Se disponía de 3.114 observaciones y se calcula la función de verosimilitud empírica. La abcisa del máximo es 750,31. En este caso nos vamos a detener un poco en las explicaciones. En el resto de los análisis nos limitaremos a exponer un cuadro con los resultados obtenidos, análogo al utilizado en capítulos anteriores.

Figura 4.2.



Los intervalos de verosimilitud vienen dados por:

I. Verosimilitud para una fuerza de evidencia 8(735,14 ; 765,48)
I. Verosimilitud para una fuerza de evidencia 32(730,08 ; 770,54)

Si notamos por “a” el valor de mayor verosimilitud y por L(a) la verosimilitud correspondiente a ese valor, obtenemos los siguientes cocientes:

$$L(a) / L(725,02) = 96,26.$$

$$L(a) / L(730,08) = 18,48.$$

$$L(a) / L(740,19) = 2,12.$$

$$L(a) / L(765,48) = 3,48.$$

$$L(a) / L(770,54) = 9,35.$$

$$L(a) / L(775,59) = 32,67.$$

De nuevo estos cocientes nos dan información sobre cuantas veces es más verosímil el valor 750,31 en relación a los puntos más cercanos en la gráfica. A medida que nos alejamos, la fuerza de la evidencia, medida por el cociente de verosimilitudes, es mucho mayor.

Normalizando la verosimilitud es sencillo obtener probabilidades y los intervalos de confianza.

Orden del percentil	0,05	0,25	0,5	0,75	0,95	0,99
Valor del percentil	734,21	742,58	748,71	754,83	764,12	770,27

Estimación puntual y regiones de confianza para el gasto medio trimestral.

Para la estimación puntual vamos a utilizar el estimador de máxima verosimilitud. En este caso, el gasto medio estimado es 750,31 (miles de unidades monetarias).

Algunas regiones de confianza se pueden obtener a partir de la verosimilitud normalizada

I. Confianza (80%)(734,21 ; 758,40)
I. Confianza (90%)(734,21 ; 764,13)
I. Confianza (95%)(731,27 ; 767,11)

Podemos decir que el gasto medio se encuentra entre estos valores con una confianza del 80, 90 y 95 por ciento respectivamente.

¿Cuál sería el nivel de confianza que correspondería a los intervalos de verosimilitud?. Si el intervalo que tenemos es el de una fuerza de evidencia 8, (735,14; 765,48), corresponde al nivel de confianza del 90%.

Al intervalo de verosimilitud obtenido para una fuerza de evidencia 32, (730,079 ; 770,538) le corresponde un nivel de confianza del 96,5%,

Es posible ver que los valores de estos intervalos dan grandes cocientes de verosimilitudes.

$$L(735,14) / L(730,08) = 3,55.$$

$$L(735,14) / L(775,59) = 6,27$$

$$L(740,19) / L(725,02) = 45,35.$$

$$L(760,42) / L(775,59) = 19,13.$$

$$L(760,42) / L(780,65) = 85,56.$$

Estos cocientes nos indican cuantas veces es más verosímil un valor frente al otro.

4.3. – Análisis de los grupos de gasto.

4.3.1. – Grupo 1. Alimentos, Bebidas y Tabacos.

Figura 4.3.

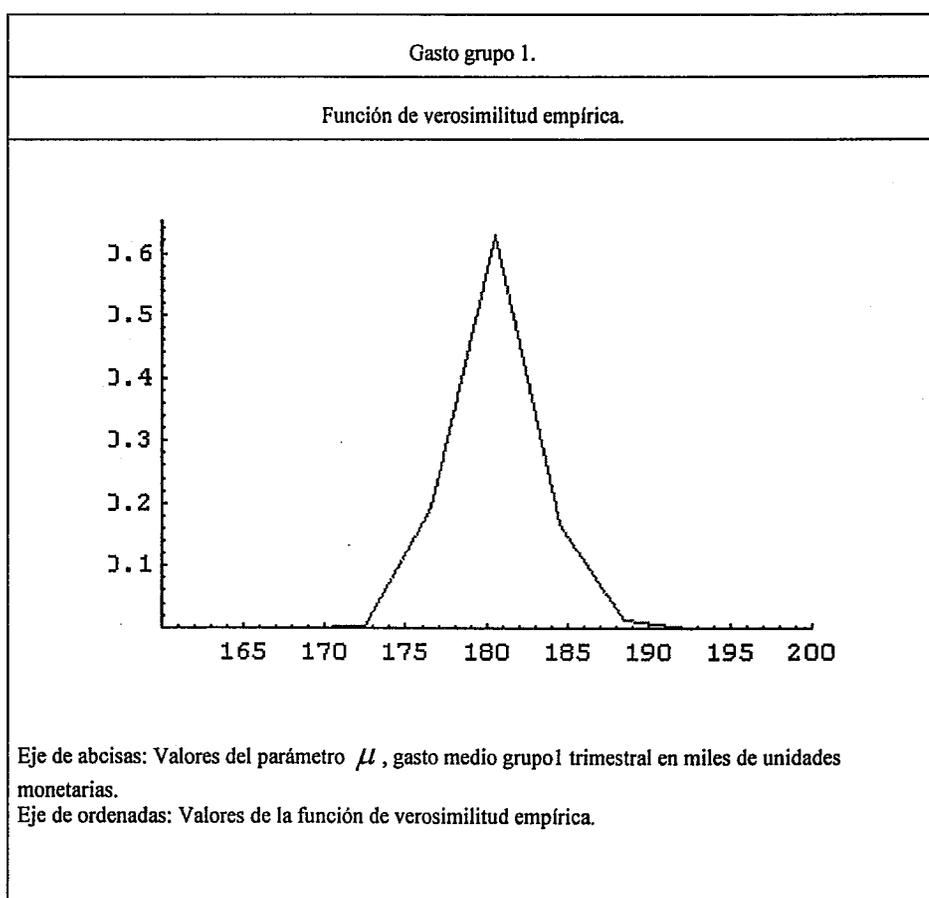


Figura 4.4.

Gasto grupo 1 en miles de unidades monetarias.						
Intervalo para una fuerza de evidencia 8.			(176,51 ; 184,54)			
Intervalo para una fuerza de evidencia 32.			(176,51 ; 184,54)			
Cocientes entre el valor de la función de verosimilitud empírica en el máximo y el valor de la función de verosimilitud empírica para otros valores cercanos en la gráfica.			$L(a)/L(172,50) = 367,64;$ $L(a)/L(176,51) = 3,26;$ $L(a)/L(184,54) = 3,78;$ $L(a)/L(188,55) = 49,80;$			
Orden del percentil	0,05	0,25	0,5	0,75	0,95	0,99
Valor del percentil	173,51	176,87	178,47	180,07	183,65	185,56
Estimación puntual.			180,53.			
Intervalos de confianza			80%	(173,51 ; 181,23)		
			90%	(173,51 ; 183,65)		
			95%	(172,99 ; 184,25)		
Nivel de confianza de los intervalos de verosimilitud expuestos al principio.			(176,51 ; 184,54).....79%			
			(176,51 ; 184,54).....79%			
Algunos cocientes del valor de la función de verosimilitud empírica en los valores de los intervalos anteriores y el valor de la función de verosimilitud empírica en valores de fuera de esos intervalos.			$L(176,51)/L(168,49) = 886.688,80;$ $L(176,51)/L(196,57) = 8.248,28;$ $L(180,52)/L(200,58) = 711.311,2;$ $L(180,52)/L(192,56) = 1.051,32;$ $L(184,54)/L(172,50) = 97,37;$ $L(184,54)/L(188,55) = 13,18;$			

4.3.2. – Grupo 2. Vestido y Calzado.

Figura 4.5.

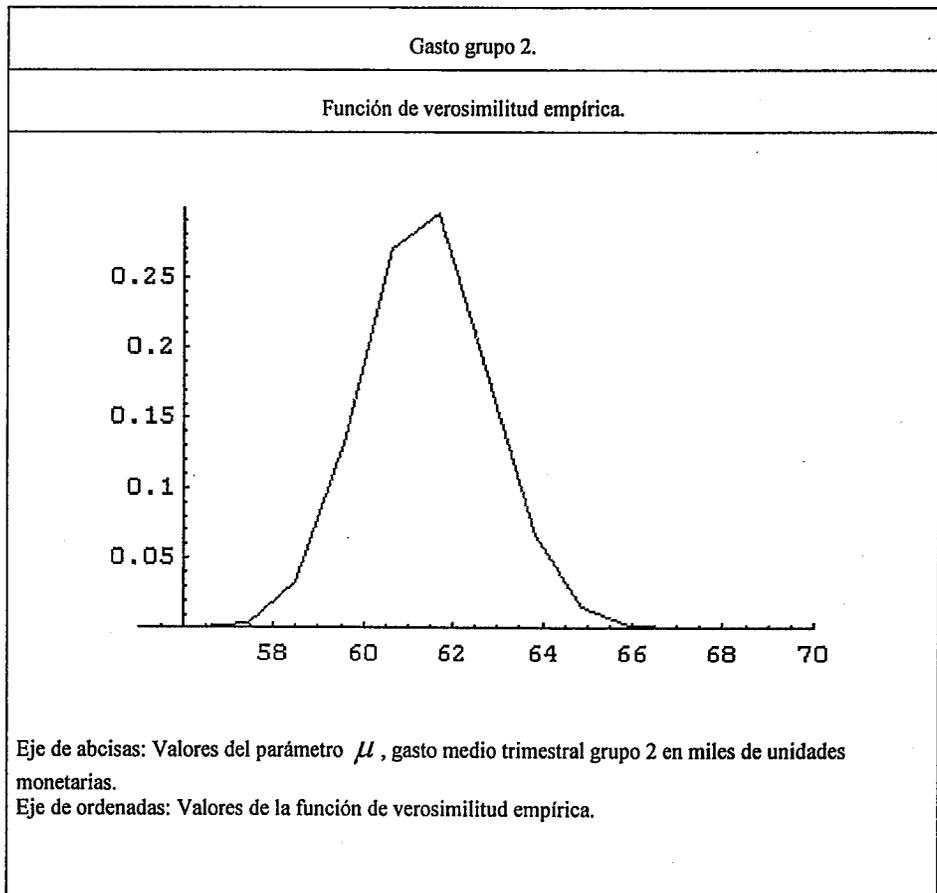


Figura 4.6.

Gasto grupo 2 en miles de unidades monetarias.						
Intervalo para una fuerza de evidencia 8.			(59,57 ; 63,82)			
Intervalo para una fuerza de evidencia 32.			(58,50 ; 64,88)			
Cocientes entre el valor de la función de verosimilitud empírica en el máximo y el valor de la función de verosimilitud empírica para otros valores cercanos en la gráfica.			$L(a) / L(58,50) = 8,95;$ $L(a) / L(59,57) = 2,23;$ $L(a) / L(60,63) = 1,09;$ $L(a) / L(62,76) = 1,62;$ $L(a) / L(63,82) = 4,43;$ $L(a) / L(64,88) = 19,37;$			
Orden del percentil	0,05	0,25	0,5	0,75	0,95	0,99
Valor del percentil	58,11	59,88	60,85	61,79	62,67	63,31
Estimación puntual.			61,69.			
Intervalos de confianza		80%	(58,61 ; 62,37)			
		90%	(58,61 ; 63,31)			
		95%	(58,11 ; 63,71)			
Nivel de confianza de los intervalos de verosimilitud expuestos al principio.			(59,57 ; 63,82).....77,5%			
			(58,50 ; 64,88).....94%			
Algunos cocientes del valor de la función de verosimilitud empírica en los valores de los intervalos anteriores y el valor de la función de verosimilitud empírica en valores de fuera de esos intervalos.			$L(58,50) / L(56,38) = 149,84;$ $L(59,57) / L(67,01) = 547,99;$ $L(60,63) / L(57,44) = 67,81;$ $L(62,76) / L(65,95) = 79,22;$ $L(63,82) / L(67,01) = 276,05;$ $L(64,88) / L(57,44) = 3,83;$			

4.3.3. – Grupo 3: Vivienda, Calefacción, y Alumbrado.

Figura 4.7.

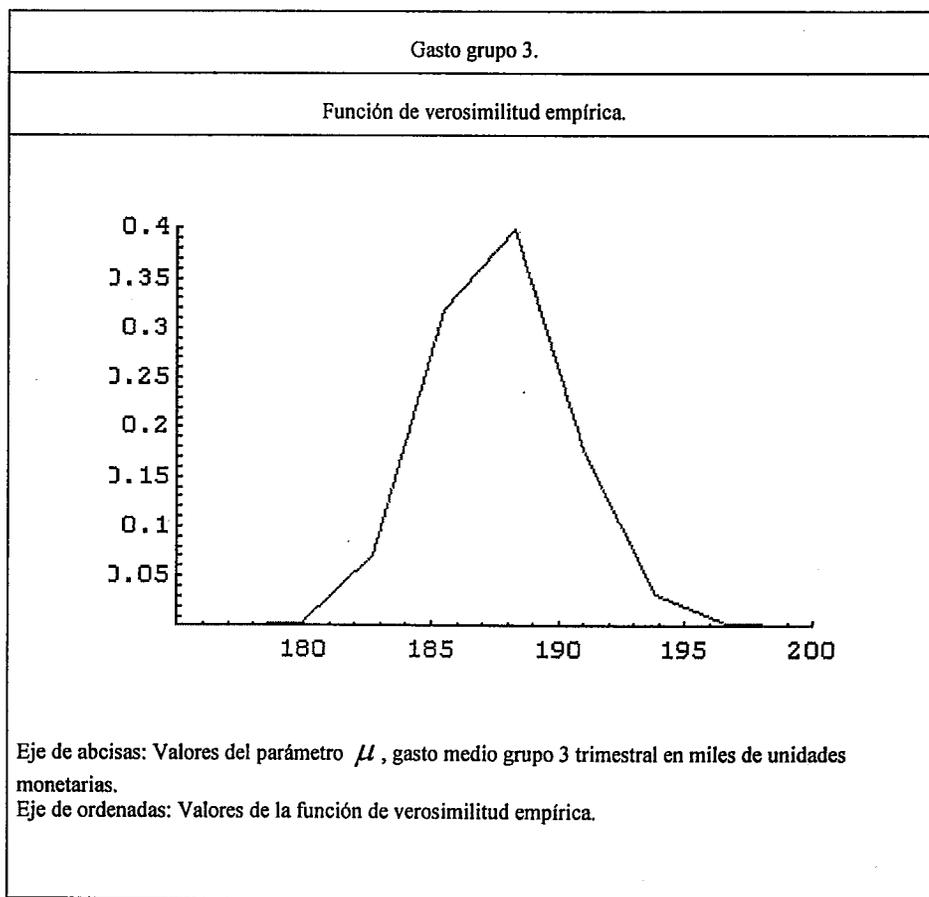


Figura 4.8.

Gasto grupo 3 en miles de unidades monetarias.						
Intervalo para una fuerza de evidencia 8.			(182,72 ; 191,05)			
Intervalo para una fuerza de evidencia 32.			(182,72 ; 193,83)			
Cocientes entre el valor de la función de verosimilitud empírica en el máximo y el valor de la función de verosimilitud empírica para otros valores cercanos en la gráfica.			$L(a) / L(179,94) = 104,27;$ $L(a) / L(182,72) = 5,55;$ $L(a) / L(185,49) = 1,25;$ $L(a) / L(191,05) = 2,31;$ $L(a) / L(193,83) = 12,74;$			
Orden del percentil	0,05	0,25	0,5	0,75	0,95	0,99
Valor del percentil	181,72	184,24	186,24	187,98	190,80	193,21
Estimación puntual.			188,27.			
Intervalos de confianza		80%	(181,72 ; 189,19)			
		90%	(181,72 ; 190,80)			
		95%	(180,76 ; 191,89)			
Nivel de confianza de los intervalos de verosimilitud expuestos al principio.		(182,72 ; 191,05).....87,5%				
		(182,72 ; 193,83).....89%				
Algunos cocientes del valor de la función de verosimilitud empírica en los valores de los intervalos anteriores y el valor de la función de verosimilitud empírica en valores de fuera de esos intervalos.			$L(185,49) / L(177,17) = 7,986;$ $L(182,72) / L(196,61) = 24,54;$ $L(188,27) / L(179,94) = 104,27;$ $L(191,05) / L(199,38) = 999,06;$ $L(193,83) / L(177,17) = 786,69;$ $L(196,61) / L(199,38) = 16,97;$			

4.3.4. – Grupo 4. Muebles, Artículos de mobiliario. Artículos de menaje para el hogar y gastos de conservación corrientes de la casa.

Figura 4.9.

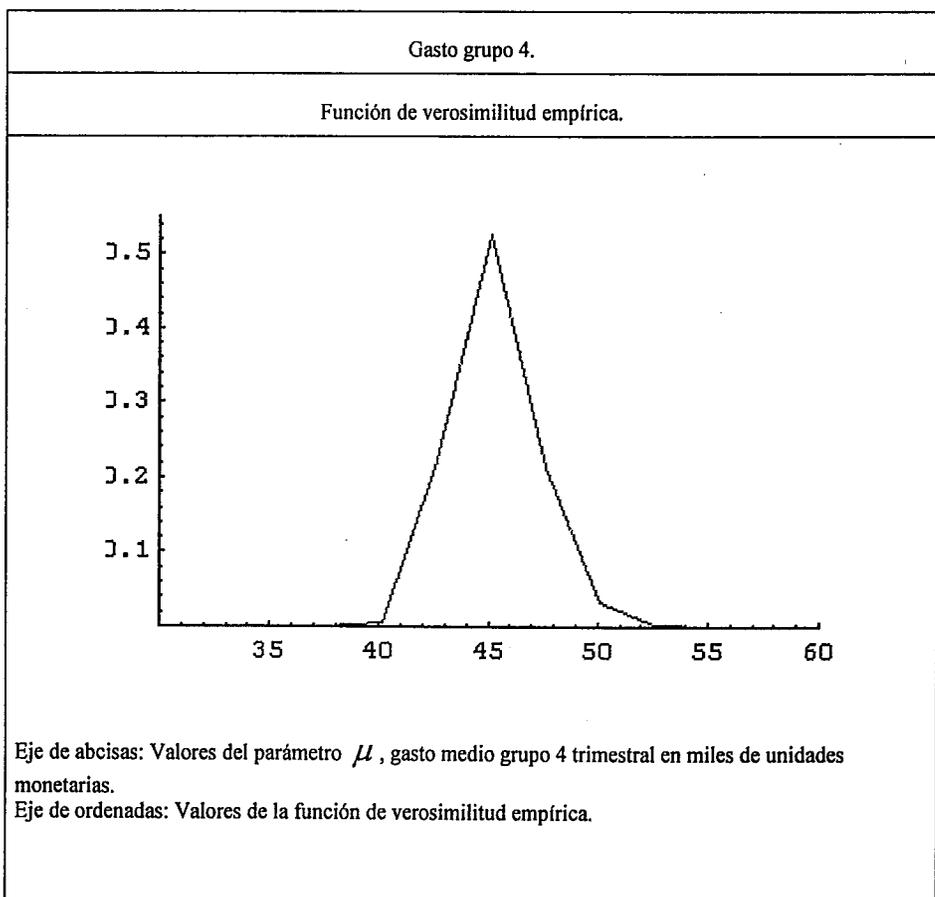


Figura 4.10.

Gasto grupo 4 en miles de unidades monetarias.						
Intervalo para una fuerza de evidencia 8.			(42,62; 47,63)			
Intervalo para una fuerza de evidencia 32.			(42,62 ; 50,14)			
Cocientes entre el valor de la función de verosimilitud empírica en el máximo y el valor de la función de verosimilitud empírica para otros valores cercanos en la gráfica.			$L(a) / L(40,11) = 96,29;$ $L(a) / L(42,62) = 2,37;$ $L(a) / L(47,63) = 2,46;$ $L(a) / L(50,14) = 16,39;$ $L(a) / L(52,64) = 189,07;$			
Orden del percentil	0,05	0,25	0,5	0,75	0,95	0,99
Valor del percentil	40,61	42,72	43,92	45,12	46,87	49,59
Estimación puntual.			45,12.			
Intervalos de confianza		80%	(40,61 ; 46,28)			
		90%	(40,61 ; 47,45)			
		95%	(40,33 ; 48,41)			
Nivel de confianza de los intervalos de verosimilitud expuestos al principio.			(42,62 ; 47,63).....70%			
			(42,62 ; 50,14).....74%			
Algunos cocientes del valor de la función de verosimilitud empírica en los valores de los intervalos anteriores y el valor de la función de verosimilitud empírica en valores de fuera de esos intervalos.			$L(42,62) / L(35,10) = 108.267,9;$ $L(45,12) / L(37,60) = 256.638,6;$ $L(47,63) / L(40,11) = 39,17;$ $L(50,14) / L(55,15) = 182,10;$ $L(42,62) / L(57,63) = 24.125,8;$ $L(45,12) / L(52,64) = 189,07;$			

4.3.5. – Grupo 5. Medicinas y otros productos farmacéuticos.

Figura 4.11.

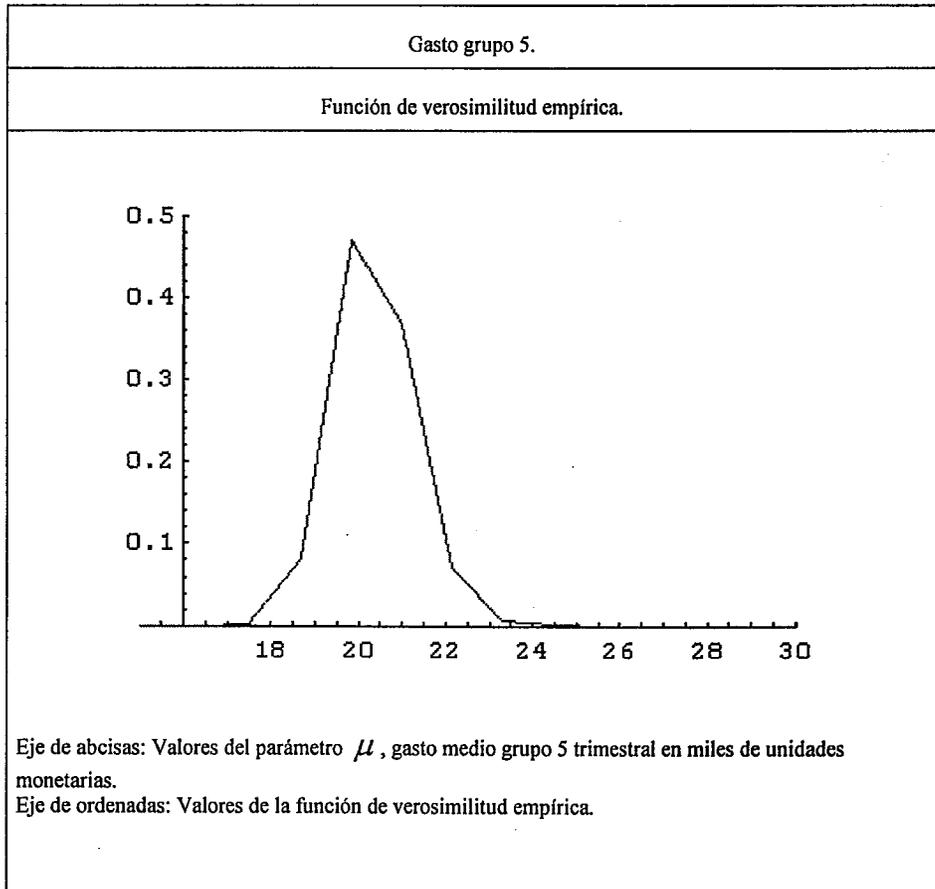


Figura 4.12.

Gasto grupo 5 en miles de unidades monetarias.						
Intervalo para una fuerza de evidencia 8.			(18,66 ; 22,15)			
Intervalo para una fuerza de evidencia 32.			(18,66 ; 22,15)			
Cocientes entre el valor de la función de verosimilitud empírica en el máximo y el valor de la función de verosimilitud empírica para otros valores cercanos en la gráfica.			$L(a) / L(17,49) = 501,56;$ $L(a) / L(18,66) = 5,80;$ $L(a) / L(20,99) = 1,27;$ $L(a) / L(22,15) = 6,54;$ $L(a) / L(23,32) = 77,40;$			
Orden del percentil	0,05	0,25	0,5	0,75	0,95	0,99
Valor del percentil	18,19	19,07	19,69	20,44	20,92	22,09
Estimación puntual.			19,82.			
Intervalos de confianza		80%	(18,19 ; 20,76)			
		90%	(18,19 ; 21,45)			
		95%	(17,83 ; 21,85)			
Nivel de confianza de los intervalos de verosimilitud expuestos al principio.			(18,66 ; 22,15).....89%			
			(18,66 ; 22,15).....89%			
Algunos cocientes del valor de la función de verosimilitud empírica en los valores de los intervalos anteriores y el valor de la función de verosimilitud empírica en valores de fuera de esos intervalos.			$L(18,66) / L(23,32) = 13,34;$ $L(19,82) / L(17,49) = 501,56;$ $L(20,99) / L(23,32) = 60,58;$ $L(22,15) / L(24,49) = 214,42;$ $L(18,66) / L(16,32) = 256.732,6;$ $L(20,99) / L(17,49) = 392,58;$			

4.3.6. – Grupo 6. Transportes y Comunicaciones.

Figura 4.13.

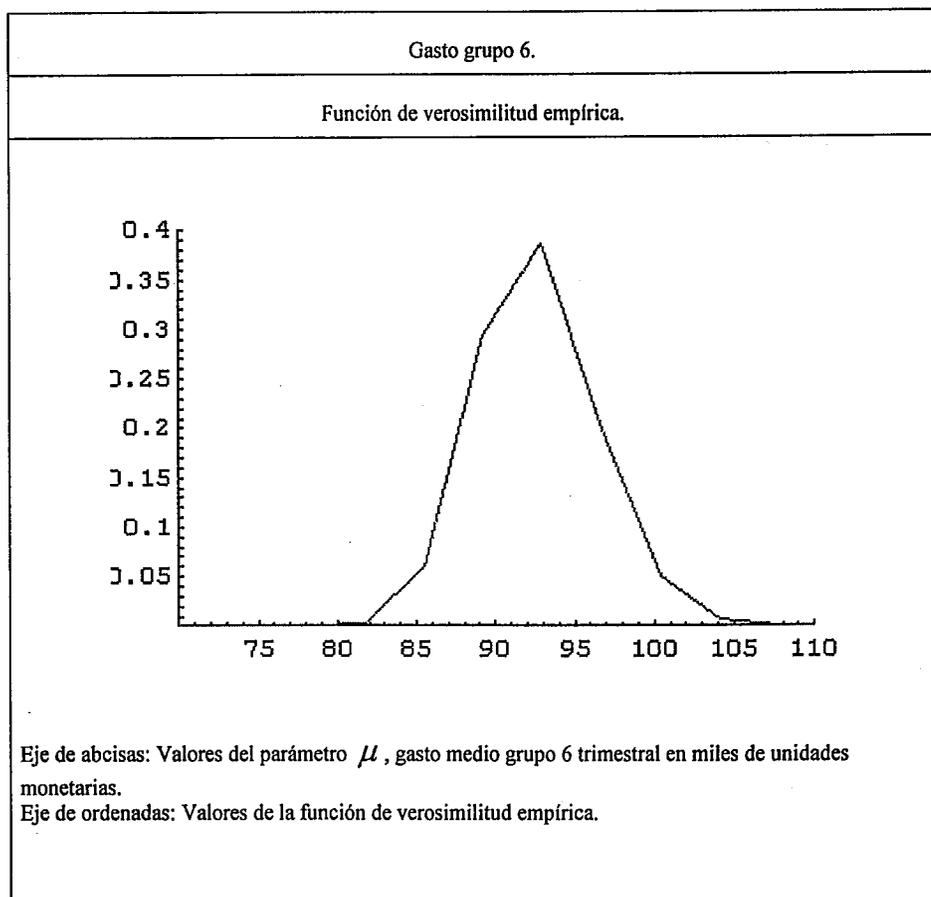


Figura 4.14.

Gasto grupo 6 en miles de unidades monetarias.						
Intervalo para una fuerza de evidencia 8.		(85,51 ; 100,39)				
Intervalo para una fuerza de evidencia 32.		(85,51 ; 100,39)				
Cocientes entre el valor de la función de verosimilitud empírica en el máximo y el valor de la función de verosimilitud empírica para otros valores cercanos en la gráfica.		$L(a)/L(81,80) = 202,25;$ $L(a)/L(85,51) = 6,48;$ $L(a)/L(89,23) = 1,32;$ $L(a)/L(96,67) = 1,95;$ $L(a)/L(100,39) = 7,58;$				
Orden del percentil	0,05	0,25	0,5	0,75	0,95	0,99
Valor del percentil	84,79	87,88	90,62	93,09	95,91	100,30
Estimación puntual.			92,95.			
Intervalos de confianza		80%	(84,79 ; 94,97)			
		90%	(84,79 ; 97,39)			
		95%	(83,23 ; 99,21)			
Nivel de confianza de los intervalos de verosimilitud expuestos al principio.		(85,51 ; 100,39).....89%				
		(85,51 ; 100,39).....89%				
Algunos cocientes del valor de la función de verosimilitud empírica en los valores de los intervalos anteriores y el valor de la función de verosimilitud empírica en valores de fuera de esos intervalos.		$L(85,51) / L(107,82) = 69,78;$ $L(89,23) / L(81,80) = 153,64;$ $L(92,95) / L(107,82) = 452,36;$ $L(96,67) / L(78,08) = 42.020,9;$ $L(100,39) / L(104,104) = 6,41;$ $L(85,51) / L(81,80) = 31,20;$				

4.3.7. – Grupo 7. Esparcimiento, Espectáculos, Enseñanza y Cultura.

Figura 4.15.

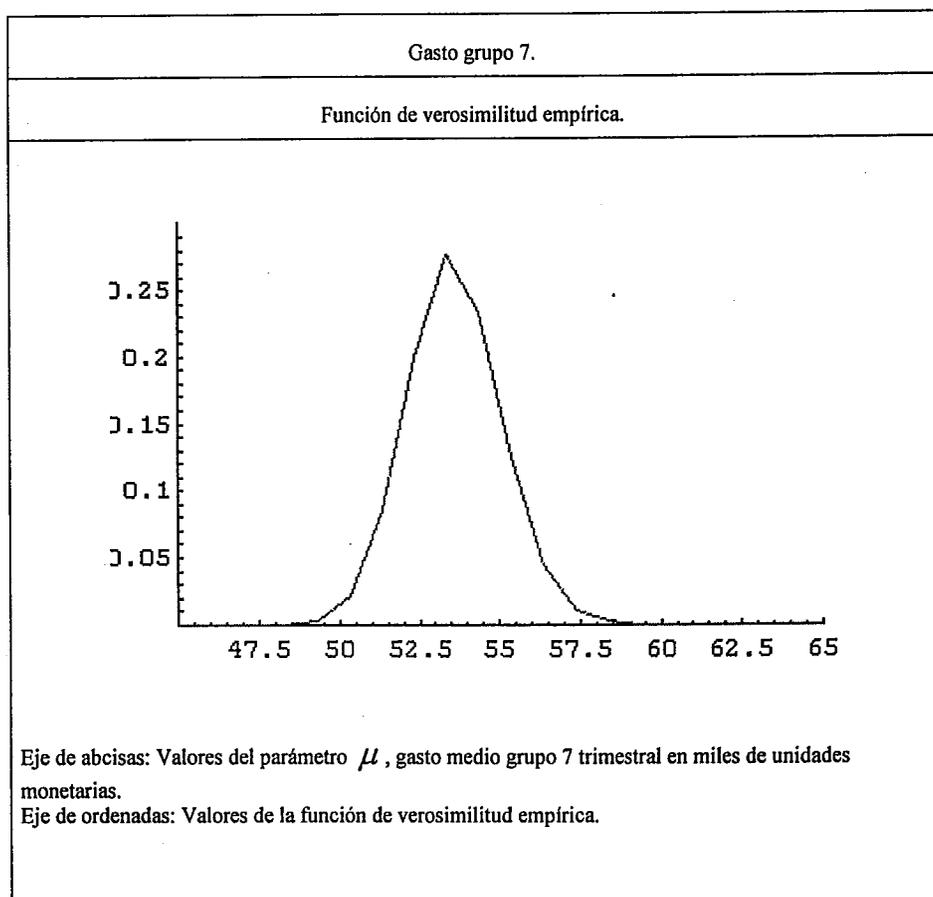


Figura 4.16.

Gasto grupo 7 en miles de unidades monetarias.						
Intervalo para una fuerza de evidencia 8.		(51,33 ; 56,36)				
Intervalo para una fuerza de evidencia 32.		(50,32 ; 57,37)				
Cocientes entre el valor de la función de verosimilitud empírica en el máximo y el valor de la función de verosimilitud empírica para otros valores cercanos en la gráfica.		$L(a)/L(50,32) = 13,60;$ $L(a)/L(51,33) = 3,25;$ $L(a)/L(55,36) = 2,20;$ $L(a)/L(56,36) = 6,21;$ $L(a)/L(57,37) = 25,46;$				
Orden del percentil	0,05	0,25	0,5	0,75	0,95	0,99
Valor del percentil	50,64	52,04	53,03	54,06	55,52	56,64
Estimación puntual.			53,34.			
Intervalos de confianza		80%	(50,64 ; 54,61)			
		90%	(50,64 ; 55,52)			
		95%	(50,35 ; 56,09)			
Nivel de confianza de los intervalos de verosimilitud expuestos al principio.		(51,33 ; 56,36).....87,5%				
		(50,32 ; 57,37).....96,5%				
Algunos cocientes del valor de la función de verosimilitud empírica en los valores de los intervalos anteriores y el valor de la función de verosimilitud empírica en valores de fuera de esos intervalos.		$L(50,32) / L(59,38) = 84,69;$ $L(51,33) / L(48,31) = 499,57;$ $L(52,33) / L(49,32) = 76,92;$ $L(53,34) / L(58,38) = 146,92;$ $L(55,36) / L(49,32) = 48,14;$ $L(56,36) / L(48,31) = 261,30;$				

4.3.8. – Grupo 8. Otros bienes y servicios.

Figura 4.17.

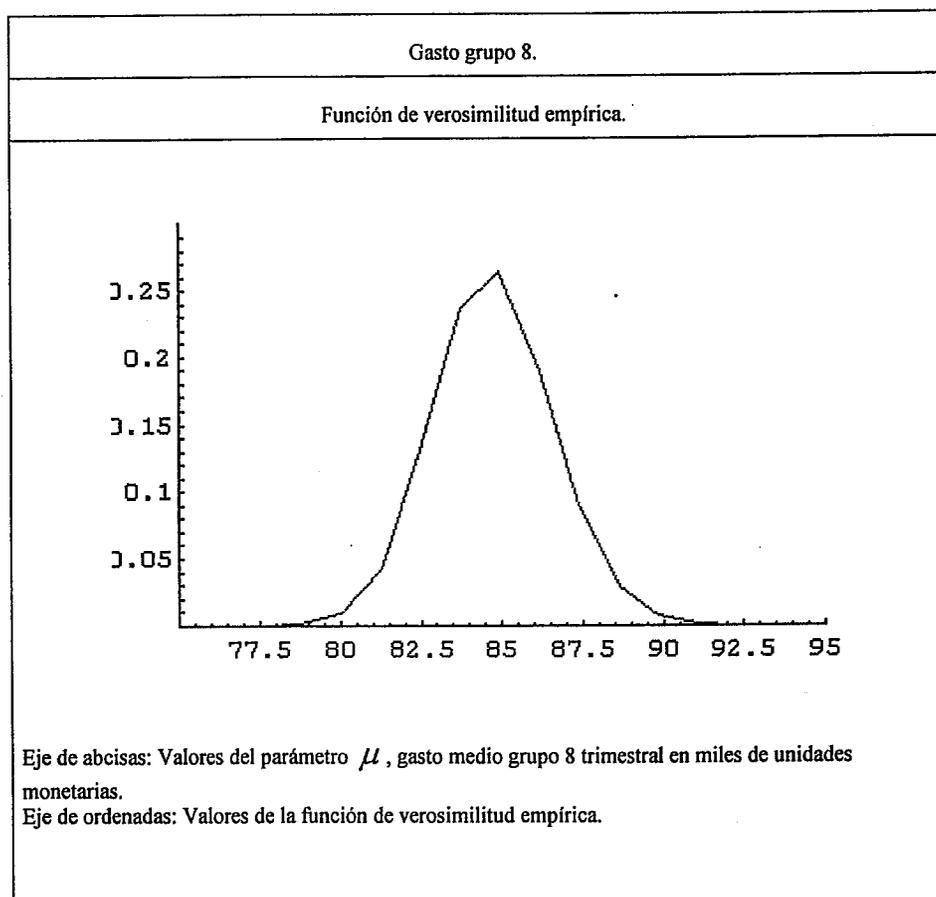


Figura 4.18.

Gasto grupo 8 en miles de unidades monetarias.						
Intervalo para una fuerza de evidencia 8.			(81,24 ; 87,39)			
Intervalo para una fuerza de evidencia 32.			(81,24 ; 88,62)			
Cocientes entre el valor de la función de verosimilitud empírica en el máximo y el valor de la función de verosimilitud empírica para otros valores cercanos en la gráfica.			$L(a)/L(81,24) = 6,16;$ $L(a)/L(82,47) = 2,03;$ $L(a)/L(83,70) = 1,12;$ $L(a)/L(86,16) = 1,39;$ $L(a)/L(87,39) = 2,94;$			
Orden del percentil	0,05	0,25	0,5	0,75	0,95	0,99
Valor del percentil	81,18	82,82	84,08	85,37	86,53	88,54
Estimación puntual.			84,93			
Intervalos de confianza		80%	(81,18 ; 86,01)			
		90%	(81,18 ; 87,22)			
		95%	(80,46 ; 87,91)			
Nivel de confianza de los intervalos de verosimilitud expuestos al principio.			(81,24 ; 87,39).....90%			
			(81,24 ; 88,62).....94%			
Algunos cocientes del valor de la función de verosimilitud empírica en los valores de los intervalos anteriores y el valor de la función de verosimilitud empírica en valores de fuera de esos intervalos.			$L(81,24)/L(78,78) = 49,47;$ $L(82,47)/L(89,85) = 19,48;$ $L(83,70)/L(80,01) = 28,85;$ $L(86,16)/L(78,78) = 218,46;$ $L(87,39)/L(89,85) = 13,46;$ $L(83,70)/L(91,09) = 211,10;$			

4.3.9. – Grupo 9. Otros gastos no mencionados anteriormente.

Figura 4.19.

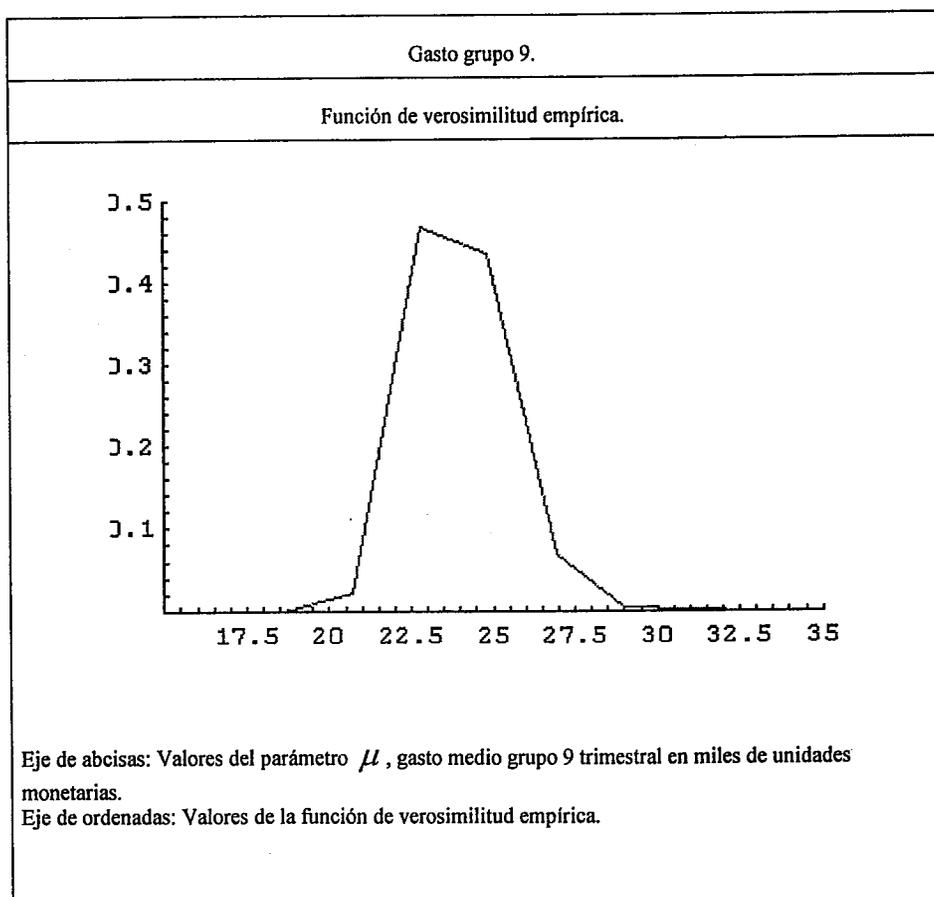


Figura 4.20.

Gasto grupo 9 en miles de unidades monetarias.						
Intervalo para una fuerza de evidencia 8.		(22,78 ; 26,93)				
Intervalo para una fuerza de evidencia 32.		(20,71 ; 26,93)				
Cocientes entre el valor de la función de verosimilitud empírica en el máximo y el valor de la función de verosimilitud empírica para otros valores cercanos en la gráfica.		$L(a) / L(20,71) = 22,92;$ $L(a) / L(24,85) = 1,07;$ $L(a) / L(26,93) = 6,73;$ $L(a) / L(29) = 94,25;$ $L(a) / L(31,06) = 1.853,89;$				
Orden del percentil	0,05	0,25	0,5	0,75	0,95	0,99
Valor del percentil	20,84	21,72	22,83	24,02	24,73	26,78
Estimación puntual.			22,78			
Intervalos de confianza	80%	(20,84 ; 24,50)				
	90%	(20,84 ; 25,60)				
	95%	(20,73 ; 26,34)				
Nivel de confianza de los intervalos de verosimilitud expuestos al principio.		(22,78 ; 26,93).....49%				
		(20,71 ; 26,93).....96,5%				
Algunos cocientes del valor de la función de verosimilitud empírica en los valores de los intervalos anteriores y el valor de la función de verosimilitud empírica en valores de fuera de esos intervalos.		$L(20,71) / L(29) = 4,11;$ $L(20,71) / L(31,06) = 80,88;$ $L(22,78) / L(29) = 94,26;$ $L(22,78) / L(31,06) = 1.853,8;$ $L(24,85) / L(33,14) = 39.616,1;$ $L(26,93) / L(35,21) = 150.230,4;$				

4.4. – Análisis de las participaciones de los grupos de gasto en el total.

En este apartado se va a estudiar el porcentaje que cada tipo de gasto representa dentro del gasto total.

Para ello, en primer lugar se obtiene la participación de cada grupo de gasto dentro del total para los 3.114 hogares que constituyen la muestra. Basta con dividir el gasto de cada hogar entre el gasto total de dicho hogar, y ya se dispone de una nueva variable. Por no trabajar con demasiados decimales se redondeó a dos decimales cada participación de los distintos grupos de gasto, y a continuación se calculó la verosimilitud empírica en cada caso y se llevó a cabo la inferencia para la participación media.

Al igual que en el resto de la Memoria pueden encontrarse los resultados en los cuadros que resumen la inferencia realizada, así como las distintas funciones de verosimilitud. Si nos fijamos en la estimación de máxima verosimilitud de cada participación media de tipo de gasto y las sumamos nos da una cantidad de 0,999, lo cual es prácticamente la unidad.

Pasemos pues a presentar los principales resultados obtenidos al analizar las diferentes participaciones de los grupos de gasto en el gasto total.

4.4.1. – Participación del grupo 1.

Figura 4.21.

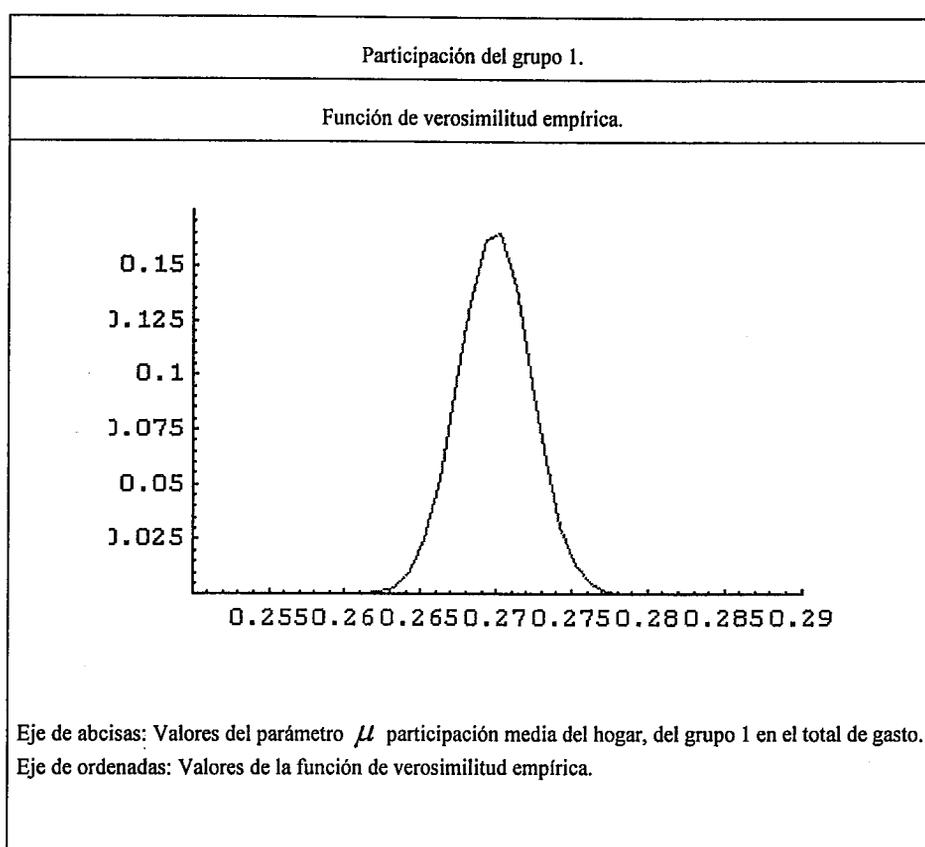


Figura 4.22.

Participación del grupo 1 en el gasto total.						
Intervalo para una fuerza de evidencia 8.			(0,265 ; 0,274)			
Intervalo para una fuerza de evidencia 32.			(0,264 ; 0,275)			
Cocientes entre el valor de la función de verosimilitud empírica en el máximo y el valor de la función de verosimilitud empírica para otros valores cercanos en la gráfica.			$L(a)/L(0,267) = 1,80;$ $L(a)/L(0,268) = 1,24;$ $L(a)/L(0,269) = 1,02;$ $L(a)/L(0,273) = 2,69;$ $L(a)/L(0,274) = 5,27;$			
Orden del percentil	0,05	0,25	0,5	0,75	0,95	0,99
Valor del percentil	0,265	0,268	0,269	0,271	0,273	0,275
Estimación puntual.			0,270.			
Intervalos de confianza		80%	(0,265 ; 0,272)			
		90%	(0,265 ; 0,273)			
		95%	(0,265 ; 0,274)			
Nivel de confianza de los intervalos de verosimilitud expuestos al principio				(0,265 ; 0,274).....95%		
				(0,264 ; 0,275).....98%		
Algunos cocientes del valor de la función de verosimilitud empírica en los valores de los intervalos anteriores y el valor de la función de verosimilitud empírica en valores de fuera de esos intervalos			$L(0,265) / L(0,262) = 27,22;$ $L(0,266) / L(0,276) = 10,72;$ $L(0,267) / L(0,261) = 430;$ $L(0,271) / L(0,277) = 92,28;$ $L(0,272) / L(0,276) = 20,54;$ $L(0,273) / L(0,261) = 286,97;$			

4.4.2. – Participación del grupo 2.

Figura 4.23.

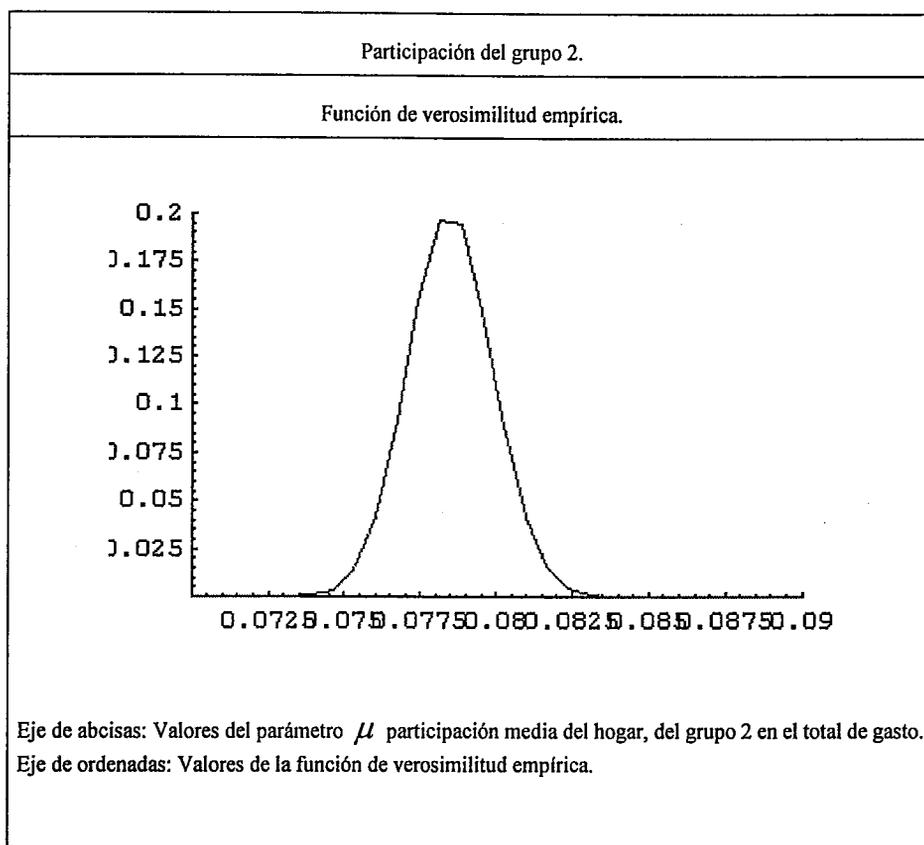


Figura 4.24.

Participación del grupo 2 en el gasto total.						
Intervalo para una fuerza de evidencia 8.		(0,076 ; 0,081)				
Intervalo para una fuerza de evidencia 32.		(0,076 ; 0,082)				
Cocientes entre el valor de la función de verosimilitud empírica en el máximo y el valor de la función de verosimilitud empírica para otros valores cercanos en la gráfica.		$L(a)/ L(0,076) = 4,72;$ $L(a)/ L(0,077) = 2,14;$ $L(a)/ L(0,078) = 1,28;$ $L(a)/ L(0,079) = 1,01;$ $L(a)/ L(0,080) = 2,20;$				
Orden del percentil	0,05	0,25	0,5	0,75	0,95	0,99
Valor del percentil	0,076	0,077	0,078	0,079	0,080	0,081
Estimación puntual.			0,078.			
Intervalos de confianza	80%	(0,076 ; 0,079)				
	90%	(0,076 ; 0,080)				
	95%	(0,075 ; 0,081)				
Nivel de confianza de los intervalos de verosimilitud expuestos al principio		(0,076 ; 0,081).....92,5%				
		(0,076 ; 0,082).....94%				
Algunos cocientes del valor de la función de verosimilitud empírica en los valores de los intervalos anteriores y el valor de la función de verosimilitud empírica en valores de fuera de esos intervalos		$L(0,076) / L(0,074) = 60,03;$ $L(0,077) / L(0,082) = 19,43;$ $L(0,077) / L(0,075) = 42,05;$ $L(0,079) / L(0,083) = 170,14;$ $L(0,079) / L(0,074) = 214,06;$ $L(0,080) / L(0,082) = 18,85;$				

4.4.3. – Participación del grupo 3.

Figura 4.25.

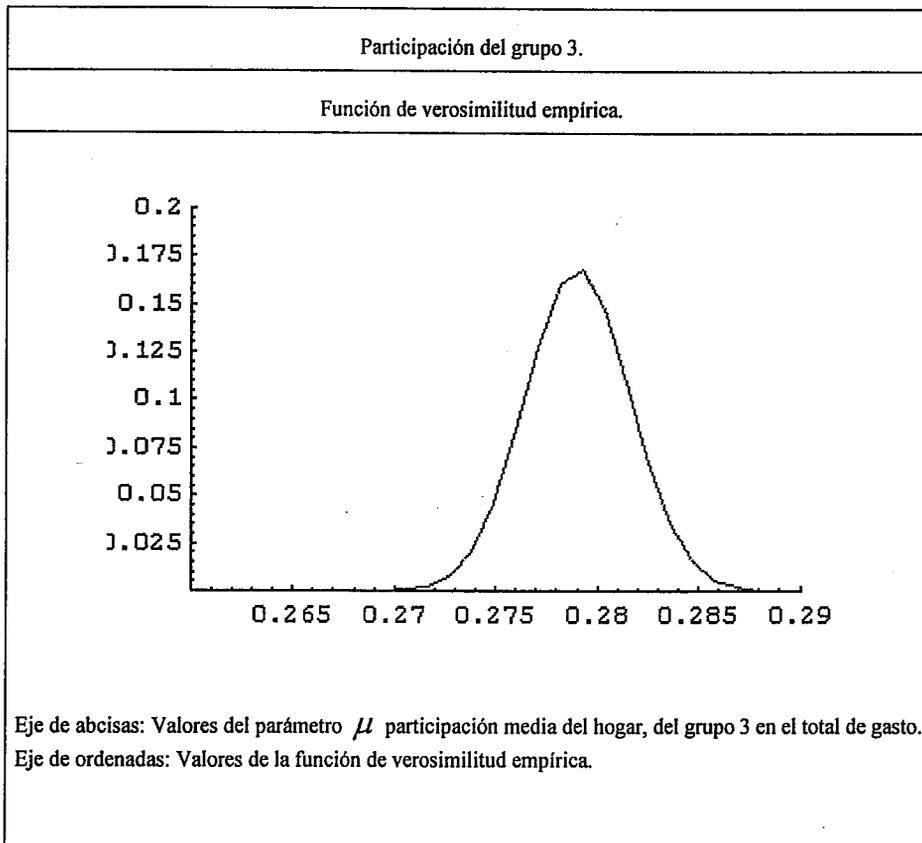


Figura 4.26.

Participación del grupo 3 en el gasto total.						
Intervalo para una fuerza de evidencia 8.		(0,275 ; 0,283)				
Intervalo para una fuerza de evidencia 32.		(0,273 ; 0,286)				
Cocientes entre el valor de la función de verosimilitud empírica en el máximo y el valor de la función de verosimilitud empírica para otros valores cercanos en la gráfica.		$L(a)/L(0,276) = 2;$ $L(a)/L(0,277) = 1,32;$ $L(a)/L(0,278) = 1,05;$ $L(a)/L(0,281) = 1,55;$ $L(a)/L(0,282) = 2,50;$				
Orden del percentil	0,05	0,25	0,5	0,75	0,95	0,99
Valor del percentil	0,274	0,277	0,278	0,280	0,283	0,284
Estimación puntual.			0,279.			
Intervalos de confianza		80%	(0,274 ; 0,281)			
		90%	(0,274 ; 0,283)			
		95%	(0,273 ; 0,283)			
Nivel de confianza de los intervalos de verosimilitud expuestos al principio		(0,275 ; 0,283).....87,5%				
		(0,273 ; 0,28696,5%				
Algunos cocientes del valor de la función de verosimilitud empírica en los valores de los intervalos anteriores y el valor de la función de verosimilitud empírica en valores de fuera de esos intervalos		$L(0,276) / L(0,272) = 35,44;$ $L(0,277) / L(0,288) = 247,81;$ $L(0,278) / L(0,288) = 311,85;$ $L(0,280) / L(0,270) = 248,19;$ $L(0,281) / L(0,272) = 45,77;$ $L(0,282) / L(0,287) = 35,85;$				

4.4.4. – Participación del grupo 4.

Figura 4.27.

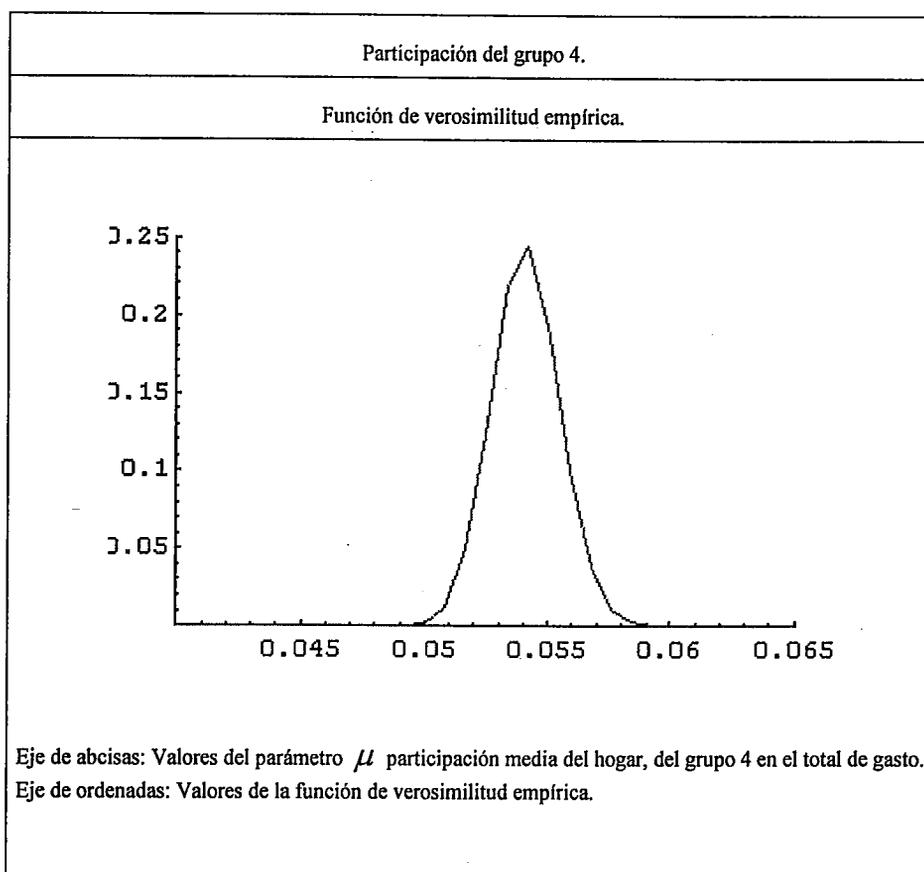


Figura 4.28.

Participación del grupo 4 en el gasto total.						
Intervalo para una fuerza de evidencia 8.		(0,052 ; 0,057)				
Intervalo para una fuerza de evidencia 32.		(0,051 ; 0,057)				
Cocientes entre el valor de la función de verosimilitud empírica en el máximo y el valor de la función de verosimilitud empírica para otros valores cercanos en la gráfica.		$L(a)/L(0,051) = 4,84;$ $L(a)/L(0,052) = 1,86;$ $L(a)/L(0,053) = 1,11;$ $L(a)/L(0,056) = 2,44;$ $L(a)/L(0,057) = 6,27;$				
Orden del percentil	0,05	0,25	0,5	0,75	0,95	0,99
Valor del percentil	0,051	0,053	0,054	0,055	0,056	0,057
Estimación puntual.			0,054.			
Intervalos de confianza		80%	(0,051 ; 0,055)			
		90%	(0,051 ; 0,056)			
		95%	(0,051 ; 0,057)			
Nivel de confianza de los intervalos de verosimilitud expuestos al principio		(0,052 ; 0,057).....84%				
		(0,051 ; 0,057).....95%				
Algunos cocientes del valor de la función de verosimilitud empírica en los valores de los intervalos anteriores y el valor de la función de verosimilitud empírica en valores de fuera de esos intervalos		$L(0,051) / L(0,049) = 351,05;$ $L(0,052) / L(0,058) = 52,82;$ $L(0,053) / L(0,059) = 514,80;$ $L(0,055) / L(0,050) = 107,60;$ $L(0,056) / L(0,049) = 696,75;$ $L(0,057) / L(0,058) = 15,68;$				

4.4.5. – Participación del grupo 5.

Figura 4.29.

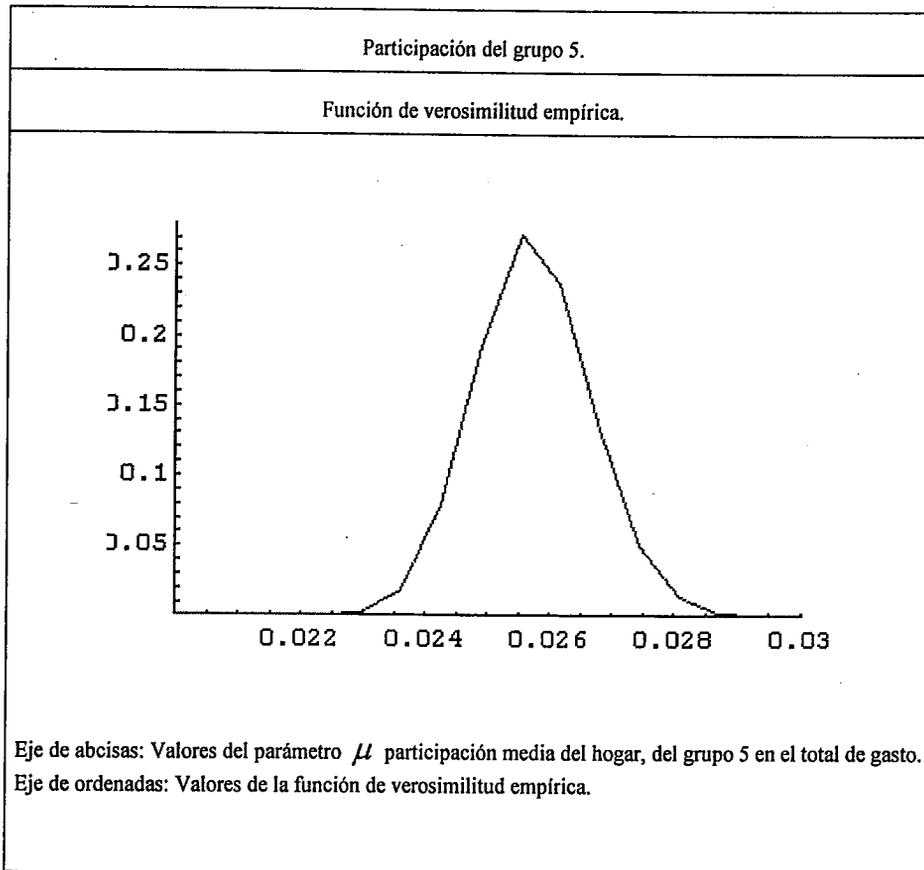


Figura 4.30.

Participación del grupo 5 en el gasto total.						
Intervalo para una fuerza de evidencia 8.		(0,024 ; 0,027)				
Intervalo para una fuerza de evidencia 32.		(0,028 ; 0,028)				
Cocientes entre el valor de la función de verosimilitud empírica en el máximo y el valor de la función de verosimilitud empírica para otros valores cercanos en la gráfica.		$L(a) / L(0,024) = 3,43;$ $L(a) / L(0,025) = 1,41;$ $L(a) / L(0,026) = 1,15;$ $L(a) / L(0,027) = 2,04;$ $L(a) / L(0,028) = 5,41;$				
Orden del percentil	0,05	0,25	0,5	0,75	0,95	0,99
Valor del percentil	0,023	0,024	0,025	0,026	0,027	0,028
Estimación puntual.			0,025.			
Intervalos de confianza		80%	(0,024 ; 0,026)			
		90%	(0,024 ; 0,027)			
		95%	(0,023 ; 0,027)			
Nivel de confianza de los intervalos de verosimilitud expuestos al principio		(0,024 ; 0,027).....90%				
		(0,028 ; 0,028).....96,5%				
Algunos cocientes del valor de la función de verosimilitud empírica en los valores de los intervalos anteriores y el valor de la función de verosimilitud empírica en valores de fuera de esos intervalos		$L(0,023) / L(0,022) = 624,40;$ $L(0,024) / L(0,029) = 75,05;$ $L(0,025) / L(0,023) = 125,42;$ $L(0,026) / L(0,029) = 638,56;$ $L(0,027) / L(0,022) = 61,33;$ $L(0,028) / L(0,029) = 135,40;$				

4.4.6. – Participación del grupo 6.

Figura 4.31.

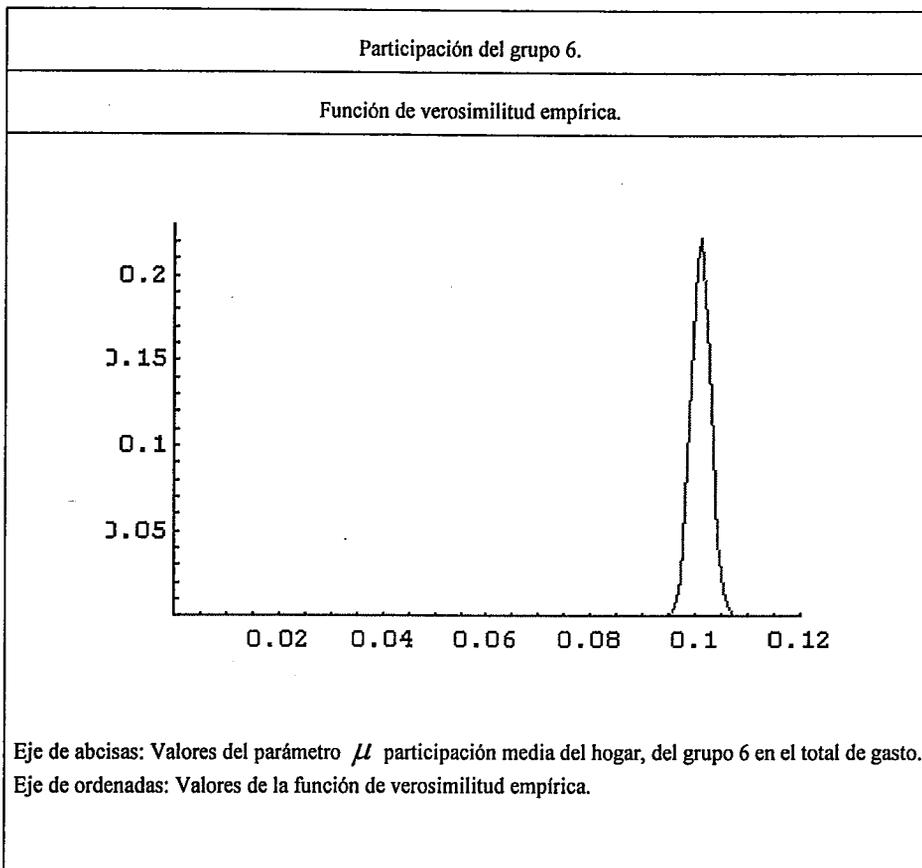


Figura 4.32.

Participación del grupo 6 en el gasto total.						
Intervalo para una fuerza de evidencia 8.		(0,098 ; 0,104)				
Intervalo para una fuerza de evidencia 32.		(0,097 ; 0,105)				
Cocientes entre el valor de la función de verosimilitud empírica en el máximo y el valor de la función de verosimilitud empírica para otros valores cercanos en la gráfica.		$L(a)/ L(0,098) = 3,34;$ $L(a)/ L(0,099) = 1,58;$ $L(a)/ L(0,100) = 1,07;$ $L(a)/ L(0,102) = 1,28;$ $L(a)/ L(0,103) = 2,21;$				
Orden del percentil	0,05	0,25	0,5	0,75	0,95	0,99
Valor del percentil	0,097	0,099	0,100	0,102	0,104	0,105
Estimación puntual.			0,101.			
Intervalos de confianza		80%	(0,098 ; 0,104)			
		90%	(0,098 ; 0,104)			
		95%	(0,098 ; 0,104)			
Nivel de confianza de los intervalos de verosimilitud expuestos al principio		(0,098 ; 0,104).....94%				
		(0,097 ; 0,105).....96,5%				
Algunos cocientes del valor de la función de verosimilitud empírica en los valores de los intervalos anteriores y el valor de la función de verosimilitud empírica en valores de fuera de esos intervalos		$L(0,098) / L(0,096) = 13,51;$ $L(0,099) / L(0,095) = 189,23;$ $L(0,100) / L(0,102) = 1,20;$ $L(0,101) / L(0,103) = 2,22;$ $L(0,102) / L(0,096) = 1,73;$ $L(0,103) / L(0,095) = 135,61;$				

4.4.7. – Participación del grupo 7.

Figura 4.33.

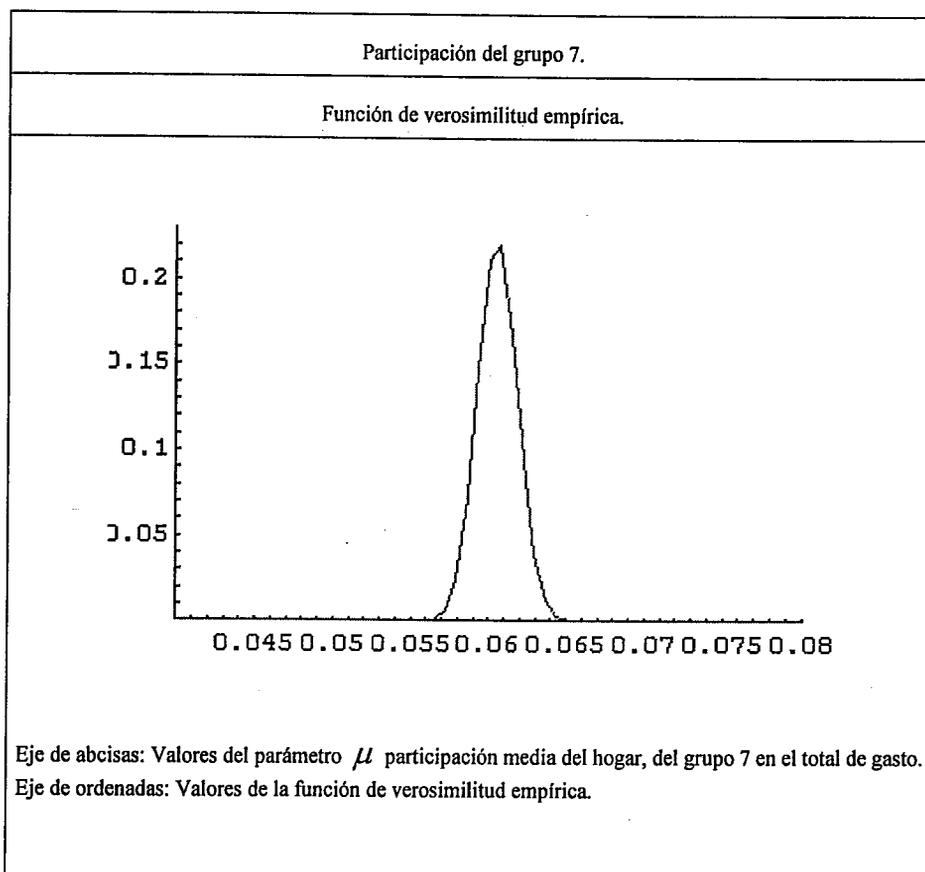


Figura 4.34.

Participación del grupo 7 en el gasto total.						
Intervalo para una fuerza de evidencia 8.		(0,059 ; 0,063)				
Intervalo para una fuerza de evidencia 32.		(0,058 ; 0,064)				
Cocientes entre el valor de la función de verosimilitud empírica en el máximo y el valor de la función de verosimilitud empírica para otros valores cercanos en la gráfica.		$L(a) / L(0,058) = 3,04;$ $L(a) / L(0,059) = 1,50;$ $L(a) / L(0,060) = 1,04;$ $L(a) / L(0,062) = 2,34;$ $L(a) / L(0,063) = 5,57;$				
Orden del percentil	0,05	0,25	0,5	0,75	0,95	0,99
Valor del percentil	0,058	0,059	0,060	0,061	0,062	0,063
Estimación puntual.			0,061.			
Intervalos de confianza		80%	(0,058 ; 0,061)			
		90%	(0,058 ; 0,062)			
		95%	(0,057 ; 0,063)			
Nivel de confianza de los intervalos de verosimilitud expuestos al principio.		(0,059 ; 0,063).....79%				
		(0,058 ; 0,064).....94%				
Algunos cocientes del valor de la función de verosimilitud empírica en los valores de los intervalos anteriores y el valor de la función de verosimilitud empírica en valores de fuera de esos intervalos		$L(0,058) / L(0,056) = 148,41;$ $L(0,069) / L(0,057) = 35,16;$ $L(0,060) / L(0,064) = 71,27;$ $L(0,061) / L(0,065) = 283,81;$ $L(0,062) / L(0,057) = 15,60;$ $L(0,063) / L(0,064) = 12,79;$				

4.4.8. – Participación del grupo 8.

Figura 4.35.

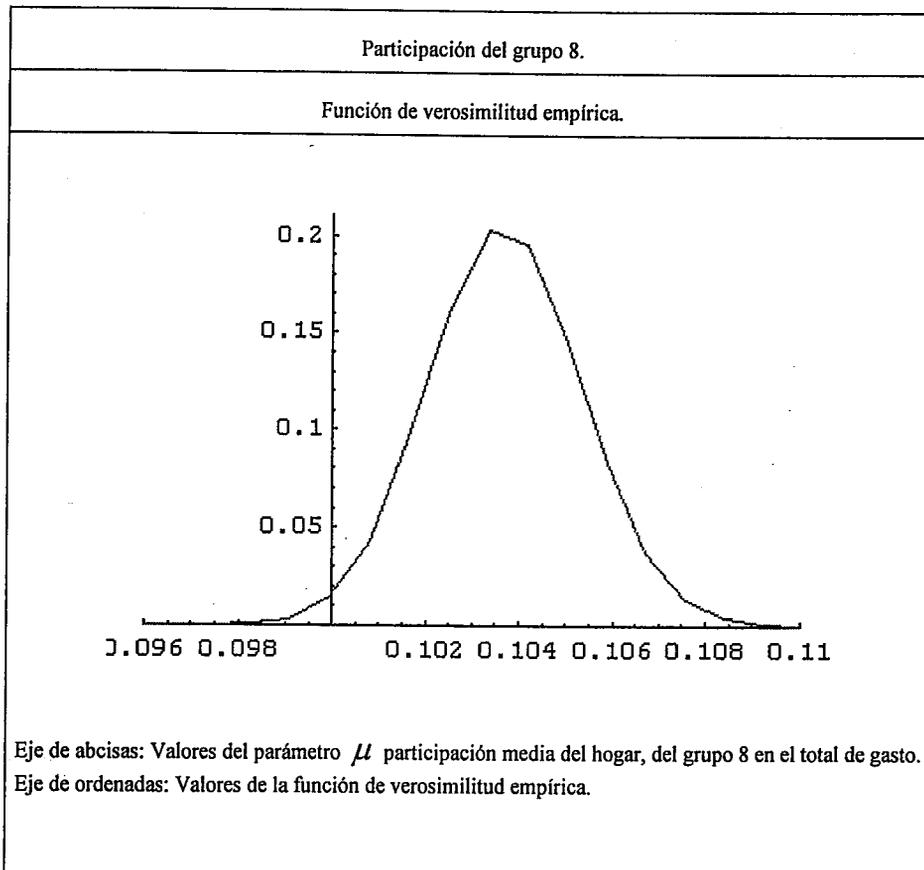


Figura 4.36.

Participación del grupo 8 en el gasto total.						
Intervalo para una fuerza de evidencia 8.		(0,101 ; 0,107)				
Intervalo para una fuerza de evidencia 32.		(0,099 ; 0,107)				
Cocientes entre el valor de la función de verosimilitud empírica en el máximo y el valor de la función de verosimilitud empírica para otros valores cercanos en la gráfica.		$L(a) / L(0,100) = 4,71;$ $L(a) / L(0,101) = 2,12;$ $L(a) / L(0,102) = 1,26;$ $L(a) / L(0,105) = 1,39;$ $L(a) / L(0,106) = 2,42;$				
Orden del percentil	0,05	0,25	0,5	0,75	0,95	0,99
Valor del percentil	0,100	0,102	0,103	0,104	0,106	0,107
Estimación puntual.			0,103.			
Intervalos de confianza		80%	(0,100 ; 0,105)			
		90%	(0,100 ; 0,106)			
		95%	(0,100 ; 0,107)			
Nivel de confianza de los intervalos de verosimilitud expuestos al principio		(0,101 ; 0,107).....89%				
		(0,099 ; 0,107).....98%				
Algunos cocientes del valor de la función de verosimilitud empírica en los valores de los intervalos anteriores y el valor de la función de verosimilitud empírica en valores de fuera de esos intervalos		$L(0,101) / L(0,098) = 149,84;$ $L(0,102) / L(0,107) = 11,95;$ $L(0,103) / L(0,099) = 57,25;$ $L(0,104) / L(0,108) = 51,631;$ $L(0,105) / L(0,098) = 227,35;$ $L(0,106) / L(0,099) = 23,65;$				

4.4.9. – Participación del grupo 9.

Figura 4.37.

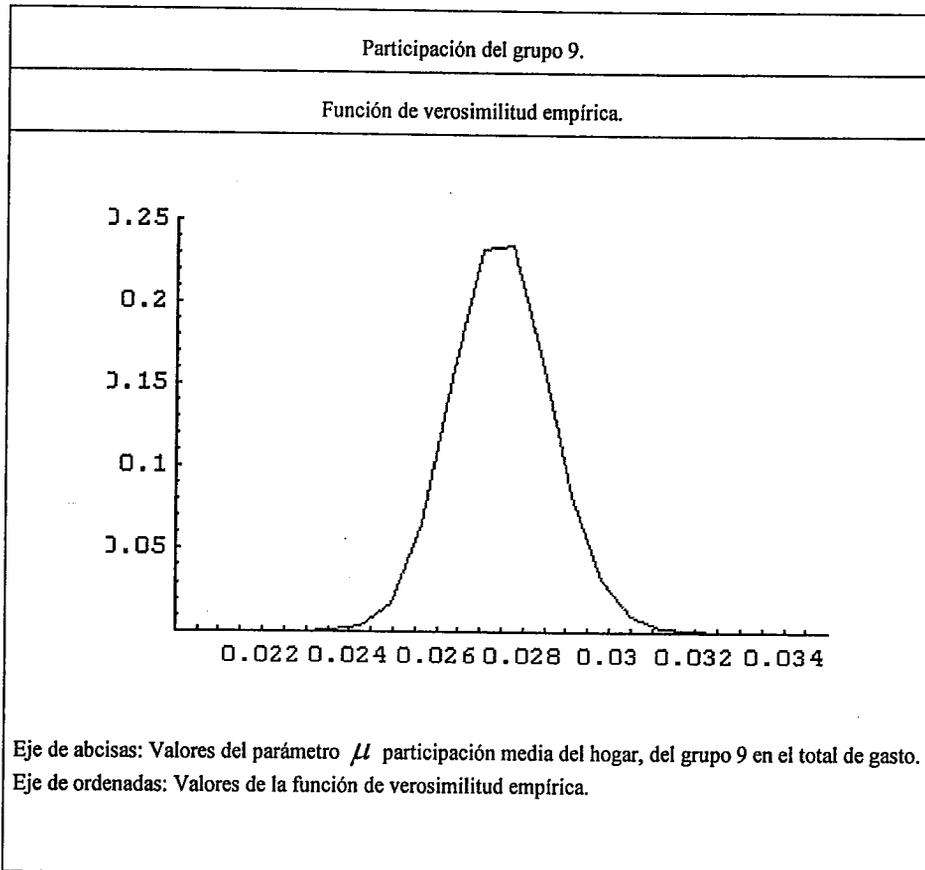


Figura 4.38.

Participación del grupo 9 en el gasto total.						
Intervalo para una fuerza de evidencia 8.		(0,026 ; 0,030)				
Intervalo para una fuerza de evidencia 32.		(0,025 ; 0,030)				
Cocientes entre el valor de la función de verosimilitud empírica en el máximo y el valor de la función de verosimilitud empírica para otros valores cercanos en la gráfica.		$L(a)/ L(0,025) = 3,57;$ $L(a)/ L(0,026) = 1,52;$ $L(a)/ L(0,027) = 1,01;$ $L(a)/ L(0,028) = 1,42;$ $L(a)/ L(0,029) = 2,81;$				
Orden del percentil	0,05	0,25	0,5	0,75	0,95	0,99
Valor del percentil	0,025	0,026	0,027	0,028	0,029	0,030
Estimación puntual.			0,028.			
Intervalos de confianza		80%	(0,025 ; 0,028)			
		90%	(0,025 ; 0,029)			
		95%	(0,025 ; 0,030)			
Nivel de confianza de los intervalos de verosimilitud expuestos al principio		(0,026 ; 0,030).....84%				
		(0,025 ; 0,030).....95%				
Algunos cocientes del valor de la función de verosimilitud empírica en los valores de los intervalos anteriores y el valor de la función de verosimilitud empírica en valores de fuera de esos intervalos		$L(0,026) / L(0,024) = 59,19;$ $L(0,027) / L(0,031) = 115,30;$ $L(0,027) / L(0,023) = 1.080,05;$ $L(0,028) / L(0,032) = 455,31;$ $L(0,029) / L(0,023) = 383,93;$ $L(0,030) / L(0,024) = 12,04;$				

4.5. – Comentarios y utilización de los resultados obtenidos.

La verosimilitud empírica nos permite representar formalmente el fenómeno que nos interesa, en este caso, el gasto medio de los hogares españoles, y el gasto medio por hogar para cada grupo de gasto.

El estudio de cada uno de ellos por separado ya es de por sí interesante, y nos va a dar una visión de cuál es la estructura de gasto que hay en España.

En conjunto, el gasto medio por hogar más probable es 750,31.

¿A qué tipo de preguntas se puede contestar?. Algunas son las siguientes:

. - ¿Ha crecido este gasto en los últimos años? ¿Se gasta más o menos que antes?. Basta con acudir a la ECPF y utilizar los datos disponibles para obtener la función de verosimilitud empírica de cada momento que estemos considerando. Deflactando las cantidades podemos saber como ha sido esa evolución.

. - ¿Quiénes son los que más gastan?. Usando los datos disponibles a este respecto, podemos comparar el gasto medio en el norte y en el sur de la península, o por comunidades o por provincias, obteniendo una idea de la cantidad de gasto en cada zona. De igual forma, lo que se ha comentado de la evolución del gasto medio total es aplicable a esta situación, viendo que comunidad ha tenido un ritmo de crecimiento mayor, cuál ha experimentado más cambios...

A su vez, esta comparación puede realizarse atendiendo a características de la familia como el número de miembros, el sexo de los componentes del hogar, el número de hijos..., según el nivel de desagregación que pueda dar la encuesta que se esté utilizando.

¿Puede utilizarse el gasto como una medida del bienestar y de la pobreza?. Ya se comentó en el capítulo anterior que hay autores que consideran más apropiado emplear cantidades de gasto para medir la riqueza de los individuos. Para los que opinen así, los datos obtenidos pueden servir

para conocer la situación de los hogares españoles y también para establecer posibles comparaciones. Los resultados ponen de manifiesto que el 50% de los hogares tienen un gasto inferior a la cantidad que hemos considerado el gasto medio más probable.

. - ¿Hay más información que podamos tener sobre el gasto medio?. Se sabe que hay una probabilidad del 90% de que ese gasto se encuentre entre 734,21 y 764,13. También es posible conocer otros valores entre los que se puede encontrar ese gasto y su probabilidad asociada, así como la probabilidad de que el gasto medio sea menor que una determinada cantidad. De igual modo conocemos que hay valores que son más verosímiles que otros. Un gasto medio de 735,14 es 6,27 veces más verosímil que un gasto de 775,59, pero un valor de 740,19 es 45 veces más verosímil que 725,35. Existe una fuerte evidencia de que el valor medio se encuentre entre 730,08 y 770,54.

. - ¿Podemos saber cuánto se gasta y en qué?. La encuesta está especialmente diseñada para conocer el tipo de gasto de los hogares españoles, con lo que se puede responder a esa pregunta fácilmente. Analizando el gasto en los distintos grupos extraemos algunas conclusiones muy interesantes.

En primer lugar podemos ordenar los grupos de gasto de mayor a menor, según el gasto medio más probable:

Grupo 3: Vivienda, calefacción y alumbrado (188,27).

Grupo 1: Alimentación, bebidas y tabaco (180,53).

Grupo 6: Transportes y comunicaciones (92,95).

Grupo 8: Otros bienes y servicios (84,93).

Grupo 2: Vestido y calzado (61,69).

Grupo 7: Esparcimiento, espectáculo, enseñanza y cultura (53,34).

Grupo 4: Muebles, artículos de mobiliario, menaje y gastos de conservación del hogar (45,12).

Grupo 9: Otros gastos no mencionados anteriormente (22,78).

Grupo 5: Medicinas y otros productos farmacéuticos (19,18).

Los mayores gastos corresponden a los grupos Vivienda y Alimentación, siendo el grupo 3 el que presenta un gasto medio mayor con un 4,28% más que el de Alimentación.

Un poco más de la mitad del gasto del grupo 1 es lo que se dedica a Transporte y comunicaciones, con un gasto medio por hogar de 92,95.

Casi 85.000 unidades monetarias es lo que se gasta en otros bienes y servicios, un 37% más de lo que se destina a Vestido y calzado, que es el 5º grupo en el que más se gasta.

En Esparcimiento, enseñanza, espectáculo y cultura, el valor medio más probable es de 53,34, tres veces menos que el gasto en Alimentación, y un 18% más que el gasto dedicado a Muebles y menaje del hogar.

El grupo que menos gasto tiene es el correspondiente a Medicinas y otros productos farmacéuticos, con un gasto medio de 19,18.

Sería muy interesante conocer si este orden de gasto se mantiene, es decir, si es el mismo por comunidades. Se podría mirar ya no sólo si es similar la cantidad que dedican a un mismo grupo de gasto, sino si se encuentra en idéntica posición en la lista de gastos.

Se podría comprobar si las distintas comunidades se encuentran por encima o por debajo del gasto medio de cada grupo.

Por supuesto, al igual que se comentó anteriormente, es posible estudiar la evolución de cada grupo, así como establecer posibles comparaciones en base a las características del hogar.

En este tema las posibilidades de análisis y comparación son muy numerosas. La ECPF está diseñada para conocer los gastos de los hogares, así que en principio, se dispone de información acerca del gasto realizado en cada uno de los componentes que constituyen cada grupo. Sería fácil ver el gasto sólo en alimentación, en ropa de hombre, en medicinas.... Se puede ir descendiendo en el análisis hasta el nivel que se desee, llevando a cabo todo tipo de comparaciones, según en lo que estemos interesados.

También parece de interés analizar la participación de cada grupo de gasto en el total. Así, de forma análoga a la participación de los ingresos en el total, se calcula para cada uno de los 3.114 hogares, el porcentaje que ese gasto concreto representaba en el total.

Así hemos obtenido la participación media por hogar para cada grupo de gasto. Los resultados obtenidos ponen de manifiesto que:

- . - Más de la mitad del gasto, casi el 55%, se dedica a los grupos Vivienda y Alimentación.
- . - El 20,4% se dedica al grupo de Transportes y Otros bienes y servicios en partes prácticamente iguales.
- . - No se gasta ni un 3% en Medicinas y productos farmacéuticos.
- . - Al grupo Esparcimiento y espectáculo, se dedica un 6%.
- . - Un 7,8% se destina a Vestido y calzado.

De nuevo, pueden establecerse todo tipo de comparaciones entre provincias, por grupos, y estudiar la participación de gasto de un subgrupo dentro de su total. Por ejemplo, podemos analizar la parte que se destina a Alimentos dentro del grupo 1, y si se quiere descender más pues se puede ver el porcentaje que se destina a carnes, o a pastas,...De igual forma se puede proceder en cualquiera de los otros grupos, suministrando información sobre un gran número de situaciones.

Estudiar la evolución de la participación de un subgrupo en su grupo o de un grupo en el total, no es sólo para saber cual ha sido el comportamiento. También puede usarse esta metodología para ver cuales son los grupos de gasto que más “sufren”. Supongamos que comparamos 3 momentos distintos del tiempo con gastos medios totales, aunque podría hacerse con cualquiera de los otros grupos, claramente diferentes. Obtengamos la participación por hogar de cada grupo. ¿Cuál es el grupo que más se ve afectado? ¿Por dónde se recortan gastos?. Es posible pensar que lo que se dedica a alimentación o a vivienda sea más o menos lo mismo, o que posiblemente el gasto dedicado a esparcimiento se reduzca considerablemente. Éstas y otras hipótesis pueden ser contrastadas. Puede ser una forma indirecta de ver que grupos presentan mayor variabilidad, y cuales mantienen inalterada su participación en el total. Una manera diferente de ver que artículos son más tendentes al cambio.

Recordamos que la obtención y actualización de resultados es rápida, que no se necesitan complejos modelos de ningún tipo para obtener

conclusiones, que pueden extenderse a todo tipo de comparaciones. Sencillez, rapidez, actualización, y múltiples posibilidades de análisis es lo que esta metodología aporta en este campo.

4.6. – Comparación entre ingresos y gastos totales. Otros comentarios.

En el capítulo anterior analizamos los ingresos atendiendo a su diferente naturaleza, y en éste hemos estudiado la estructura de gasto de dichos hogares. Nos queda pues examinarlos conjuntamente y establecer algunas conclusiones.

Se ha obtenido que el valor más probable de ingreso medio trimestral es 770,09. Por otro lado, el gasto medio trimestral más probable es 751,50. Por tanto, los hogares españoles sólo tienen un 2,4% más de ingresos que de gastos.

Observando los percentiles de la distribución de ingresos y de gastos, se obtiene que el 25% de la población tiene unos ingresos inferiores a 765,27 y unos gastos inferiores a 742,71. Un 50% ingresa más de 771,47 y gasta más de 754,83.

Llama la atención la poca diferencia que existe entre los gastos y los ingresos medios. Eso da una idea de que las familias españolas viven muy al día, en el sentido de que prácticamente gastan todo lo que ingresan. Esto significa que pueden tener un problema de liquidez para hacer frente a determinados imprevistos, junto con el hecho de que, como se pone de manifiesto, la capacidad de ahorro es muy pequeña.

Por otra parte, hay una cuestión que merece la pena resaltar. Analizando los datos suministrados por la ECPF, se observa que hay un número de familias suficientemente grande, que declara tener unos mayores gastos que ingresos, ocurriendo esto en el 44% de los casos.

La metodología de la verosimilitud empírica, ¿es capaz de recoger esta situación?, ¿qué información puede aportar al respecto?.

A partir de los datos observados se pueden construir diferentes variables que traten de recoger este efecto. Se trabajó con la diferencia y el cociente entre ingresos y gastos, pero finalmente sólo adjuntamos los resultados correspondientes al cociente.

Se dividió para cada hogar, el gasto total entre el ingreso total, obteniendo un número que podía ser menor o mayor que 1, dependiendo de si el numerador era o no mayor que el denominador.

Se obtiene la verosimilitud empírica para esta nueva variable y a continuación exponemos los principales resultados:

Figura 4.39.

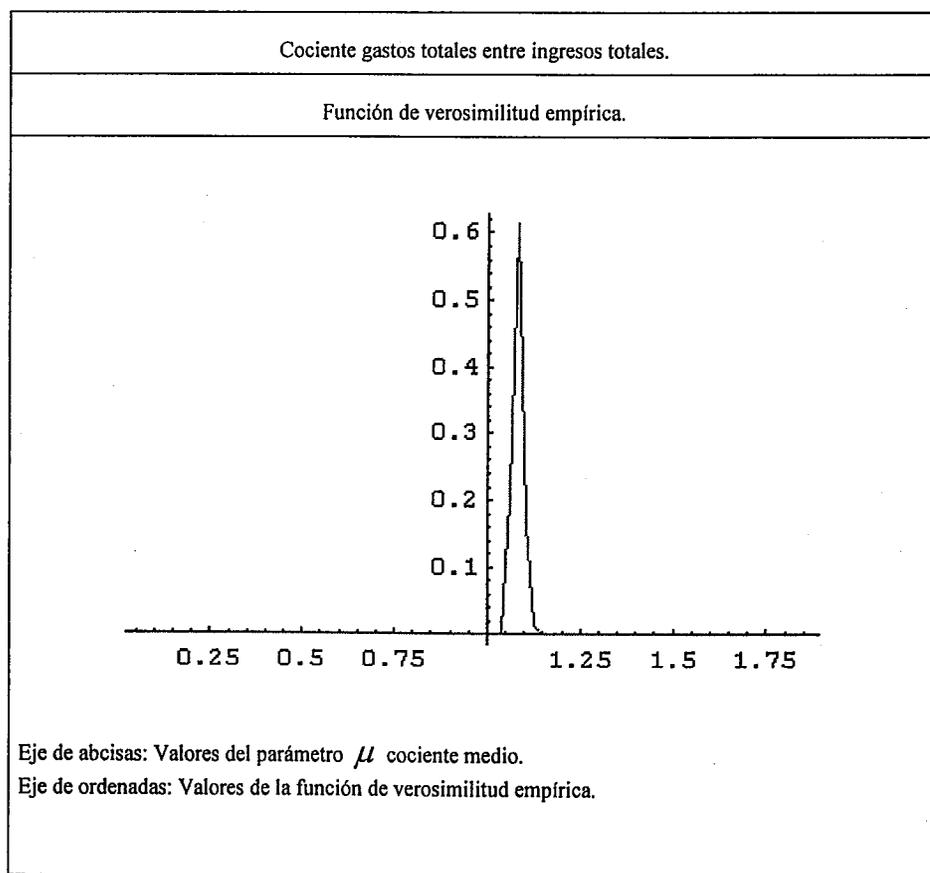


Figura 4.40.

Cociente gastos totales entre ingresos totales.						
Intervalo para una fuerza de evidencia 8.		(1,06 ; 1,10)				
Intervalo para una fuerza de evidencia 32.		(1,06 ; 1,10)				
Cocientes entre el valor de la función de verosimilitud empírica en el máximo y el valor de la función de verosimilitud empírica para otros valores cercanos en la gráfica.		$L(a)/L(0,025) = 321,16;$ $L(a)/L(0,026) = 3,03;$ $L(a)/L(0,027) = 3,69;$ $L(a)/L(0,028) = 47,43;$				
Orden del percentil	0,05	0,25	0,5	0,75	0,95	0,99
Valor del percentil	1,05	1,06	1,07	1,08	1,10	1,11
Estimación puntual.			1,08.			
Intervalos de confianza		80%	(1,05 ; 1,09)			
		90%	(1,05 ; 1,10)			
		95%	(1,04 ; 1,10)			
Nivel de confianza de los intervalos de verosimilitud expuestos al principio		(1,06 ; 1,10).....77,5%				
		(1,06 ; 1,10).....77,5%				
Algunos cocientes del valor de la función de verosimilitud empírica en los valores de los intervalos anteriores y el valor de la función de verosimilitud empírica en valores de fuera de esos intervalos		$L(0,026) / L(0,024) = 105,94;$ $L(0,027) / L(0,031) = 47,43;$ $L(0,027) / L(0,023) = 270,10;$ $L(0,028) / L(0,032) = 15,64;$ $L(0,029) / L(0,023) = 998,07;$ $L(0,030) / L(0,024) = 86,91;$				

Antes de comentar algunos de estos resultados hay que decir que también tiene mucha importancia, el rango de valores que toma el cociente medio. Ese rango oscila entre 0,07 y 19,21. En el Anexo 3 aparecen los 100 primeros valores medios calculados así como su verosimilitud, para quien esté interesado. En los primeros 45 puntos en los que se calcula le verosimilitud, ese cociente es inferior a uno y luego es mayor.

Esto subraya lo que habíamos comentado y lo recoge perfectamente: hay hogares para los que los gastos declarados son mayores que los ingresos declarados, y otros en los que ocurre exactamente lo contrario.

Pero una situación en la que los hogares gastan más que ingresan no es lo más saneado en primer lugar, y en segundo lugar no es sostenible. Una posibilidad que explique esa situación es que los hogares acuden al endeudamiento. La otra, y que no excluye la anterior, es que haya un problema de subvaloración. ¿Qué entendemos por eso?. Pues la situación en la que hay una declaración inferior a lo realmente ingresado o gastado. En la mayoría de los trabajos en los que se hace referencia a esta situación, se está de acuerdo básicamente en una cosa: esa subvaloración proviene del lado de los ingresos. Esto se puede justificar en primer lugar por la reticencia de los individuos en general a dar información sobre rentas e ingresos, especialmente por el tema fiscal, y en segundo lugar, por la fuente de la que provienen los datos. Las encuestas de presupuestos están especialmente diseñadas para conocer la estructura de gasto de los hogares, preguntando por los distintos tipos de gastos. Es en los ingresos donde es más sencillo que se produzca una ocultación de los verdaderos ingresos. Eso podría ser una respuesta a porqué hay tantos hogares que declaran algo que en principio no parece lo más esperado.

Como hemos comentado, nuestra metodología recoge esta situación y además nos da cierta información adicional:

El valor más probable de ese cociente entre gastos e ingresos es 1,08. Es decir, lo más probable es que los hogares españoles tengan unos gastos muy similares a los ingresos, cuando no un poco superiores. Además, hay una probabilidad del 95% de que ese cociente se encuentre entre 1,04 y 1,10. Aquí

también se pone en evidencia que la situación más probable es la de que esos ingresos y gastos sean muy similares. Como explicación, las que apuntamos anteriormente.

CAPÍTULO 5 CONCLUSIONES.

En este capítulo vamos a resumir a grandes rasgos que es lo que se ha realizado en esta Memoria.

Se inicia con una revisión del concepto de verosimilitud y de otros conceptos necesarios para sentar las bases de lo que vamos a utilizar, la verosimilitud empírica. En este trabajo, la verosimilitud considerada es la del parámetro valor medio, según la variable utilizada.

Los siguientes capítulos están dedicados a la aplicación de la noción de verosimilitud empírica a diversos problemas de interés económico. Los datos empleados provienen de la ECPF, y son datos referentes a:

- .- Cantidades compradas por los hogares españoles en unidades físicas;
- .- Datos de ingresos de los hogares españoles;
- .- Datos de gastos de los hogares españoles.

La metodología proporciona resultados de inferencia acerca del parámetro desconocido, fijando nuestra atención en la moda de la verosimilitud; los intervalos de

verosimilitud para una fuerza de evidencia determinada, y diversas regiones de confianza.

Se obtienen conclusiones relacionadas con:

- .- Cantidades compradas semanalmente de artículos de alimentación;
- .- Ingresos totales;
- .- Ingresos monetarios;
- .- Ingresos no monetarios;
- .- La participación de cada tipo de ingreso dentro del total.
- .- Gasto total;
- .- Cada uno de los grupos de gasto;
- .- La participación de cada grupo de gasto en el total;
- .- La comparación entre ingresos y gastos totales.

Así, en relación a las cantidades compradas conocemos cuál va a ser la cantidad media más probable que se compra de carne de cerdo, de pollo o de ternera: 0,69; 1,15 y 0,52 kilos por semana.

Obtenemos un rango de valores en los que existe una probabilidad, tan grande como queramos, de que se encuentre la cantidad media a comprar del artículo.

Proporciona otro rango de valores, también con una probabilidad elevada, que presenta una evidencia grande de compra.

Esta metodología también facilita un mecanismo que nos permite escoger entre alternativas, rechazando la que tiene menor probabilidad.

Facilita el estudio de la evolución que un artículo determinado ha ido experimentando a lo largo del tiempo, y establece una forma muy útil de comparación entre las cantidades compradas de ese artículo según las características sociodemográficas que puedan presentar los hogares.

Favorece la comparación ente artículos diferentes, permitiendo hacernos una idea de cómo se compra en los hogares españoles, y cuáles pueden ser los artículos preferidos o al menos, los más consumidos.

Referido a los ingresos, hemos obtenido que 770,09 es el ingreso medio por hogar más probable, medido en miles de unidades monetarias. También conocemos entre que valores se van a encontrar esos ingresos medios con una elevada probabilidad. Por ejemplo, la media de los ingresos por hogar está entre 754,84 y 792,35 con una probabilidad del 95%. Para los ingresos monetarios y no monetarios, la moda se cifra en 648,61 y 132,66 miles de unidades monetarias, respectivamente.

La metodología nos permite conocer el ingreso medio por hogar, que tiene importancia en sí mismo, y además desarrollar otros estudios y comparaciones que son muy interesantes:

En primer lugar nos referimos a poder conocer la evolución de esos ingresos a lo largo del tiempo. Comparando las cantidades respecto a un mismo año base podemos establecer el comportamiento de los ingresos medios por hogar, y evaluar si ha habido crecimiento o no, a que ritmo,....

Un segundo aspecto a destacar es la posibilidad de comparar los ingresos medios de zonas geográficas (países, comunidades o provincias) diferentes, lo cual nos permite establecer una clasificación de zonas por nivel de ingresos. Puede realizarse ese estudio simplemente utilizando encuestas que estén disponibles y que dispongan de datos para los hogares de cada zona.

También es posible usar esta metodología para ver si algunas políticas han tenido o no efectos en la reducción de los niveles de pobreza. Dado que se dispone de información cada muy poco tiempo, los resultados se actualizan rápidamente y se puede analizar si hay cambios o cuanto tiempo han tardado en notarse los posibles efectos.

Otra posibilidad interesante que merece la pena señalar es que, dada la clasificación de los ingresos según su origen, se puede obtener la verosimilitud empírica para las prestaciones de desempleo. También es posible conocer la participación que estas prestaciones representan dentro de los ingresos totales, y eso puede aportar información indirecta sobre el desempleo de esa región. Por supuesto también podría utilizarse con fines comparativos y supone una manera alternativa de comprobar el efecto de políticas encaminadas a reducir el nivel de desempleo.

Es posible calcular el porcentaje que representa el ingreso monetario en el total para toda la población. Sumando para todos los hogares los ingresos monetarios y dividiéndolos por la suma de los ingresos de todos los hogares, obtendríamos para la población ese porcentaje. Disponiendo de esa información para varios períodos, sin más que acudir a la ECPF, tendríamos una serie de observaciones a las que podríamos aplicar esta metodología y obtener la proporción media para la población. Esto puede realizarse para el esto de artículos considerados.

Por último, permite la actualización relativamente rápida de los resultados. Dado que la Encuesta es continua y se puede disponer en poco tiempo de los nuevos datos, este mecanismo se puede ir alimentando e inmediatamente disponer de los resultados de “última hora”.

En cuanto al gasto medio existe una probabilidad del 90% de que dicho gasto se encuentre entre 734,21 y 764,13 en miles de unidades monetarias. El gasto medio más probable obtenido para cada grupo de gasto es el siguiente:

Grupo 1: Alimentación, bebidas y tabaco (180,53).

Grupo 2: Vestido y calzado (61,69).

Grupo 3: Vivienda, calefacción y alumbrado (188,27).

Grupo 4: Muebles, artículos de mobiliario, menaje y gastos de conservación del hogar (45,12).

Grupo 5: Medicinas y otros productos farmacéuticos (19,82).

Grupo 6: Transportes y comunicaciones (92,95).

Grupo 7: Esparcimiento, enseñanza, espectáculo y cultura (53,34)

Grupo 8: Otros bienes y servicios (84,93).

Grupo 9: Otros gastos no mencionados anteriormente (22,78).

Los mayores gastos corresponden a los grupos Vivienda y Alimentación, siendo el grupo 3 el que presenta un gasto medio mayor con un 4,28% más que el de Alimentación.

Un poco más de la mitad del gasto del grupo 1 es lo que se dedica a Transporte y comunicaciones, con un gasto medio por hogar de 92,95.

Casi 85.000 unidades monetarias es lo que se gasta en otros bienes y servicios, un 37% más de lo que se destina a Vestido y calzado, que es el 4º grupo en el que más se gasta.

En Esparcimiento, enseñanza, espectáculo y cultura, el valor medio más probable es de 53,34, tres veces menos que el gasto en Alimentación, y un 18% más que el gasto dedicado a Muebles y menaje del hogar.

El grupo que menos gasto tiene es el correspondiente a Medicinas y otros productos farmacéuticos, con un gasto medio de 19,18.

También hemos obtenido la participación media por hogar para cada grupo de gasto, poniendo de manifiesto que:

- . - Más de la mitad del gasto (54,9%) se dedica a los grupos Vivienda y Alimentación.
- . - El 20,4% se dedica al grupo de Transportes y Otros bienes y servicios en partes prácticamente iguales.
- . - No se gasta ni un 3% en Medicinas y productos farmacéuticos.
- . - Al grupo Esparcimiento y espectáculo, se dedica un 6%.
- . - Un 7,8% se destina a Vestido y calzado.

De nuevo, puede establecerse un estudio pormenorizado de los gastos, distinguiendo por regiones, montante, destino el mismo,...etc, y estudiar la participación de gasto de un subgrupo dentro de su total. Por ejemplo, podemos analizar la parte que se destina a Alimentos dentro del grupo 1, y si se quiere descender más pues se puede ver el porcentaje que se destina a carnes, o a pastas,...De igual forma se puede proceder en cualquiera de los otros grupos, suministrando información sobre un gran número de situaciones.

Llama la atención la poca diferencia que existe entre los gastos y los ingresos medios. Eso da una idea de que las familias españolas viven muy al día, en el sentido de que prácticamente gastan todo lo que ingresan. Esto significa que pueden tener un problema de liquidez para hacer frente a determinados imprevistos, junto con el hecho de que, como se pone de manifiesto, la capacidad de ahorro es muy pequeña.

Combinando el análisis de los gastos e ingresos destaca el estudio de la ocultación de rentas y la posibilidad de conocer los tipos y las participaciones de gasto para distintos niveles de renta.

Esta técnica es, por tanto, una manera de representar formalmente un fenómeno, y lo importante no es sólo la gran cantidad de comparaciones y posibilidad de estudios que ofrece, y que pueden usarse para distintas finalidades, sino la sencillez y rapidez de cálculo con la que se obtienen los resultados.

No hay que hacer hipótesis de partida acerca de la distribución de la que provienen los datos, se evita la problemática de la elección de variables y formas funcionales, ni son necesarias hipótesis sobre la introducción en el modelo de la componente aleatoria.

REFERENCIAS BIBLIOGRÁFICAS.

1. AITCHISON, J. y BROWN, J.A.C. (1.957). "The Lognormal distribution". Cambridge University Press.
2. ALCAIDE INCHAUSTI, A. y ALCAIDE INCHAUSTI, J. (1.974). "Metodología para la estimación de la distribución personal de la renta en España en 1.970". Hacienda Pública Española, 26, pág 55-63.
3. ALCAIDE INCHAUSTI, A. y ALCAIDE INCHAUSTI, J. (1.977). "Distribución personal de la renta en España y en los países de la OCDE". Hacienda Pública Española, 47, pág 17-57.
4. ALCAIDE INCHAUSTI, A. y ALCAIDE INCHAUSTI, J. (1.984). "Distribución personal de la renta española en 1.980". Hacienda Pública Española, 85, pág 485-509.
5. ALONSO, J.A. y DONOSO, V. (1.986). "El consumo en España tras la adhesión a la CEE". Estudios sobre Consumo, 7, pág 13-36.
6. AMEMIYA, T. (1.984). "Tobit models: a survey". Journal of Econometrics, 24, pág 3-61.
7. AMOROSO, L. (1.925). "Ricerche intorno alla curva dei redditi". Annali di Matematica Pura ed applicata, 4-21, pág 113-157.
8. ANSCOMBE, F.J. (1.974). "Normal likelihood functions". Ann. Inst. Statist. Math, 26, pág 1-19.
9. ATKINSON, A.B. (1.970). "On the measurement of income inequality". Journal of Economic Theory, 2, pág 244-263.
10. ATKINSON, A.B.; GOMULKA, J. y STERN, N. (1.989). "Spending on alcohol: evidence from the Family Expenditure Survey 1.970-1.983". Discussion paper of the London School of Economics.

11. ATODA, NAOSUMI, SURUGA y TACHIBANAKI. (1.980). "Statistical inference of functional forms for income distribution". Working Papers, Universidad de Kyoto.
12. BARNARD. G.A. (1.947a). "A review of "Sequential Analysis" by Abraham Wald". J.A.S.A., 42, pág 658-669.
13. BARNARD. G.A. (1.947b). "The meaning of significance level". Biometrika, 34, pág 179-182.
14. BARNARD. G.A. (1.949). "Statistical inference (with discussion)". J.R.S.S., B-11, pág 115-139.
15. BARNARD, G.A; JENKINS, G.M. y WINSTEN, C.B. (1.962). "Likelihood inference and time series". J.R.S.S., A-125, pág321-372.
16. BARNDORFF-NIELSEN, O.E. (1.978). "Information and exponential families in statistical theory". Ed. Wiley, New York.
17. BARNDORFF-NIELSEN, O.E. (1.986). "Inference on full or partial parameters based on the standaralized log likelihood ratio". Biometrika, 73, pág 307-322.
18. BARNDORFF-NIELSEN, O.E. y COX, D.R. (1.984). "Bartlett adjustment to the likelihood ratio statistic and the distribution of the maximum likelihood estimator". J.R.S.S., B 46, pág 483-495.
19. BARNETT, V. (1.975). "Comparative statistical inference". Ed. Wiley, New York.
20. BARTELS, C.P.A. (1.977). "Economic aspect of regional welfare, income distribution and unemployment". Studies in Applied Regional Science.
21. BASSMAN, R.L; HAYES, K.J.; SLOTJE, D.J. y JHONSON, J.D. (1.990). "A general functional form for aproximating the Lorenz curve". Journal of Econometrics, 43, pág 77-90.
22. BASU, D. (1.975). "Statistical information and likelihood (with discussion)". Sankhya, A-37, pág 1-71.
23. BAYLEY, K.R. (1.984). "Asymptotic equivalence between the cox estimator and the general ML estimators of regression and survival parameters in the cox model". Ann. Stat., 12, pág, 730-736.
24. BEN KAABIA, M. y GIL, J.M. (2.000). "Estimación y contrastación de sistemas aids cointegrados: una aplicación a la demanda de carnes en España". Actas XIV Reunión Asepelt, Oviedo.
25. BERAN, R. (1.987). "Prepivoting to reduce level error of confidence sets". Biometrika, 74, pág 457-468.

-
26. BERGER, J. (1.984). "In defence of the likelihood principle: axiomatics and coherence". Bayesian Statistics II. Eds Bernardo, De Groot, Lindley y Smith.
 27. BERGER, J. y WOLPERT, R.L. (1.984) "The likelihood principle". Lecture Notes-Monograph Series. Vol 6. Institute of Mathematical Statistics. Hayward, California.
 28. BERK, R.H y JONES, D.H. (1.979) "Goodness-of-fit statistics that dominate the kolmogorov statistics". Z. Wahrsch. Verw. Gebiete, 47, pág 47-59.
 29. BIRNBAUM, A. (1.962). "On the foundations of statistical inference (with discussion)". J.A.S.A., 57, pág 269-306.
 30. BLUNDELL, R. (1.988). "Consumer behaviour: theory and empirical evidence. A survey". Economic Journal, 98, pág 16-65.
 31. BLUNDELL, R.; PASHARDES, P. y WEBER, G. (1.993). "What do we learn about consumer demand patterns from micro-data?". American Economic Review, 83, pág 570-597.
 32. BOLANCÉ, C. y GUILLÉN, M. " Métodos no paramétricos en el análisis de las desigualdades salariales". Manuscrito no publicado.
 33. BRACHMANN, K.; STICH, A. y TREDE, M. (1.995). "Evaluating parametric income distribution models". Discussion Papers in Statistics and Econometrics, 6.
 34. BYEKWASO, S. (1.994). "Bayesian sequential inference for error rates and error amounts in accounting data". Tesis doctoral. Faculty of the College of Arts and Sciencies. American University, Washington.
 35. CALLEALTA BARROSO, J.; CASAS SÁNCHEZ, J.M. y NÚÑEZ VELÁZQUEZ, J. (1.996). "Distribución de la renta per cápita disponible es España: descripción, desigualdad y modelización" en Pena y otros (1.996). Distribución personal de la renta en España. Ed Pirámide. Madrid. Cap 5.
 36. CANO SÁNCHEZ, J.A.; HERNÁNDEZ BASTIDA, A. y MORENO BAS, E. (1.986). "Posterior measure under partial prior information". Statistica, anno XLVI, pág 219-230. Italia.
 37. CANO SÁNCHEZ, J.A.; HERNÁNDEZ BASTIDA, A. y MORENO BAS, E. (1.987a). "On minimum risk parameter estimation under partial prior information". Publ. Inst. Stat.Univ XXXII, fasc.4, pág 7-22. París.
 38. CANO SÁNCHEZ, J.A.; HERNÁNDEZ BASTIDA, A. y MORENO BAS, E. (1.987b). "A short note on the J-minimax and mean-max risk function". Publ. Inst. Stat.Univ XXXII, fasc.3, pág 25-32. París.
 39. CANO SÁNCHEZ, J.A.; HERNÁNDEZ BASTIDA, A. y MORENO BAS, E. (1.987c). "Conditioning with upper integral, upper sufficiency". Publ. Inst. Stat.Univ XXXII, fasc.3, pág 15-24. París.

40. CAO, R.; DELGADO, M.A. y GONZÁLEZ MANTEIGA, W. (1.997). "Nomparametric curve estimation: an overview". *Investigaciones Económicas*, 21, pág 209-252.
41. CANTELLI, F.P. (1.921). *Metron*, 1, 3, pág 83-91.
42. CHEN, S.X. (1.993). "On the accuracy of empirical likelihood confidence regions for linear regression model". *Ann. Inst. Statist. Math.*, 45-4, pág 621-637.
43. CHEN, S.X. y HALL, P. (1.993). "Smoothed empirical likelihood confidence intervals for quantiles". *Ann. Stat.*, 21-3, pág 1.166-1.181.
44. CHEN, J. y QIN, J. (1.993). "Empirical likelihood estimation for finite populations and the effective usage of auxiliary information". *Biometrika*, 80, 1, pág 107-116.
45. COWELL, F.A. (1.980). "On the measure of inequality measures". *Review of Economics Studies*, 47, pág 521-531.
46. COWELL, F.A. y VICTORIA FESER, M.P. (1.996). "Robustness properties of inequality measure". *Econometrica*, 64, 1, pág 77-101.
47. COWELL, F.A.; FERREIRA, F. y LITCHFIELD, J. (1.998). "Income distribution in Brazil 1.981-1.990. Parametric and no parametric approaches". *Journal of Income Distribution*, 8, 1, pág 63-76.
48. COX, D.R. (1.958). "Some problems connected with statistical inference". *Ann. Math. Stat.*, 29, pág 357-372.
49. COX, D.R. y HINKLEY, D.V. (1.974). "Theoretical Statistics". Ed Chapman and Hall, Londres.
50. COX, D.R. y REID, N. (1.987). "Parameter orthogonality and approximate conditional inference (with discussion)". *J.R.S.S., B*, 49, pág 1-39.
51. D'ADDARIO, R. (1.949). *G.Econ. Ann. Econ*, 8, pág 91-114.
52. DAGUM, C. (1.977). "A new model of personal income distribution: specification and estimation". *Economie Appliquée*, 30, pág 413-436.
53. DAGUM, C. (1.983). *Encyclopaedia of Statistical Sciences*, 4, 5, pág 27-32.
54. DAWID, A.P. (1.977). "Conformity of inference patterns". *Recent Developments in Statistics*, Eds Barra et al, North Holland, Amsterdam.
55. DAWID, A.P. (1.981). "Statistical inference" *Encyclopedia of Statistical Sciences*. Kotz-Johnson. Ed Wiley, New York.

56. DEATON, A. (1988). "Rice price and income distribution in Thailand: a non-parametric analysis". Conference papers, Royal Economic Society, Oxford.
57. DEATON, A. y MUELLBAUER, J. (1980). "An almost ideal demand system". *American Economic Review*, 70, pág 312-326.
58. DEATON, A. y IRISH, M. (1984). "Statistical models for zero expenditures in household surveys". *Journal of Public Economics*, 23, pág 59-80.
59. DE HARO GARCÍA, J. (1993). "Modelos de distribución de renta biparamétricos". Papeles de trabajo. Cuaderno de Ciencias Económicas y Empresariales, 11, pág 6-40.
60. DE HARO GARCÍA, J. (1995). "Medición de la desigualdad y el bienestar social. Análisis de la distribución de la renta en España". Tesis doctoral, Universidad de Málaga.
61. DE LA CUESTA GONZÁLEZ, M. (1993). "Evolución reciente del crédito al consumo y su repercusión en el endeudamiento de los hogares". *Estudios sobre Consumo*, 26, pág 37-60.
62. DEL ORO SÁEZ, C.P. y PRESEDO QUINDIMIL, M.A. (1992). "Una aproximación no paramétrica a la distribución de la renta". Actas VI reunión Asepelt, Granada.
63. DEL ORO SÁEZ, C.P.; RIOBOO ALMANZOR, J.M. (1997). "Análisis de diagnóstico en regresión no paramétrica: una aplicación al cálculo de elasticidades". Actas XI Reunión Asepelt, Bilbao.
64. DEL ORO SÁEZ, C.P.; RIOBOO ALMANZOR, J.M. y RODRÍGUEZ REY, M. (2000). "Estimación de curvas de Engel: un enfoque no paramétrico y su aplicación al caso gallego". *Estudios de Economía Aplicada*, 16, pág 37-61.
65. DICICCIO, T.J.; HALL, P. y ROMANO, J.P. (1989). "Comparison of parametric and empirical likelihood functions". *Biometrika*, 76, pág 465-476.
66. DiNARDO, J.; FORTIN, N.M. y LEMIEUX, T. (1996). "Labor market institutions and the distribution of wages, 1973-1992: a semiparametric approach". *Econometrica*, 64, pág 1.001-1.044.
67. EDGEWORTH, F.Y. (1898). "On the representation of statistics by mathematical formulae". *J.R.S.S.*, 1, pág 670-700.
68. EDWARD, A.W.F. (1974) "The history of likelihood". *International Statistical Review*, 42, pág 9-15.
69. EDWARDS, A.W.F. (1972). "Likelihood". Cambridge University Press. Londres.
70. EFRON, B. (1981). "Non parametric standard errors and confidence intervals (with discussion)". *Canad. J. Statist*, 9, pág 139-172.

71. ESTEBAN GARCÍA, J.; LÓPEZ RODRÍGUEZ, M^ªI. y RUIZ PONCE, F. (2000). "Una revisión de los sistemas generadores y modelos de probabilidad descriptivos de la distribución de la renta". *Estudios de Economía Aplicada*, 14, pág 47-72.
72. FIENBERG, S.E.; NETER, J. y LEITCH, R.A. (1977). "Estimating the total overstatement error in accounting populations". *J.R.S.S.*, 72, 358, pág 295-302.
73. FERNÁNDEZ MORALES, A.; DE HARO GARCÍA, J. y MARTÍN REYES, G. (1996). "Medición de la desigualdad y el bienestar social. Análisis de la distribución de la renta en España.(1.981-1.991)". *Revista de Estudios Regionales*, 45, pág 15-42.
74. FISHER, R.A. (1921). "On the "probable error" of a coefficient of correlation deduced from a small sample". *Metron* I, 4, pág 3-32.
75. FISHER, R.A. (1925). "Theory of statistical estimation". *Proc. Cambridge.Phil. Soc.*22,pág 700-725.
76. FISHER, R.A. (1934). "Two new properties of mathematical likelihood". *Proc. Royal Soc.A* 144, pág 285-307.
77. FISHER, R.A. (1956). "Statistical Methods and Scientific Inference" Oliver and Boyd. Edimburgo.
78. FISK, P.R. (1961). "The graduation of income distribution". *Econometrica*, 29, pág 171-185.
79. FRÉCHET, M. (1939). "Sur les formules de repartition des Revenus.". *Revue de l'Institut International de statistique*, 7, pág 32-38.
80. GARCÍA, J. y LABEAGA, J.M. (1991). "A double hurdle dependent model for the demand for tobacco in Spain". Mimeo.
81. GARCÍA, I. y MOLINA, J.A. (1995). "La estructura de la demanda de alimentos en España: capacidad explicativa y predictiva". *Investigación Agraria- Economía*, 10, pág 5-26.
82. GASTWIRTH, J.L. (1973). "The estimation of the Lorenz curve and Gini index". *Review Economics and Estatistics*, 54, pág 306-316.
83. GERFIN, M. (1994). "Income distribution, income inequality and life cycle effects: a nonparametric analysis for Switzerland". *Diskussionsbeiträge des volkswirtschaftlichen Instituts*,pág 94-103.
84. GINI, C. (1936). "On the measure of concentration with especial reference to income and wealth". Cowless Comission.
85. GOUVEIA, M. y TAVARES, J. (1995). "The distribution of household income and expenditure in Portugal: 1.980 and 1990". *Review of Income and Wealth*, 41, pág 1-17.

-
86. GRACIA, A. y ALBISU, L.M. (1.994). "La dieta española en transición". *El Boletín, MAPA*, 18, pág 28-35.
87. GRACIA, A. y ALBISU, L.M. (1.994). "La demanda de bebidas alcohólicas en España". *Investigación Agraria- Economía*, 9, 1, pág 113-126.
88. GRACIA, A. y ALBISU, L.M. (1.995). "La demanda de productos cárnicos y pescados en España: aplicación de un sistema casi ideal". *Investigación Agraria- Economía*, 10, 2, pág 233-254.
89. GUILLÉN, M. y CAMPA, J.M. (1.987). "El consumo familiar en la Comunidad Europea: un análisis gráfico". *Estudios sobre Consumo*, 11, pág 71-83.
90. HALL, P. (1.986). "On the bootstrap and confidence intervals". *Ann. Stat.*, 14, pág 1431-1452.
91. HALL, P. (1.987). "On the bootstrap and likelihood based confidence regions". *Biometrika*, 74, pág 481-494.
92. HALL, P. y LA SCALA, B. (1.990). "Methodology and algorithms of the empirical likelihood". *International Statistical Review*, 58, 2, pág, 109-127.
93. HERNÁNDEZ BASTIDA, A. y VÁZQUEZ POLO, J.M. (1.997). "A note on the quasi-bayesian audit risk model for dollar-unit sampling". *The European Accounting Review*, 6-3, pág 501-507.
94. HERNÁNDEZ BASTIDA, A.; MARTEL ESCOBAR, M.C. y VÁZQUEZ POLO, J.M. (1.998). "On maximum entropy priors and a most likely likelihood in auditing". *Questiío*, 22, 2, pág 231-242.
95. HERNÁNDEZ BASTIDA, A.; MORENO CARRETERO, M.C. y VÁZQUEZ POLO, J.M. (1.997). "Cotas para el error total de una contabilidad: aproximaciones bayesianas basadas en la distribución multinomial". *Estudios de Economía Aplicada*, 7, pág 17-38.
96. HOEFFDING, W. (1.965). "Asymptotically optimal tests for multinomial disributions". *Ann. Stat. Math.*, 36, pág 369-401.
97. HOWE, H.; POLLACK, R.A. y WALES, T.J. (1.979). "Theory and time series estimation of the quadratic expenditure system". *Econometrica*, 47, pág 1.231-1.247.
98. HERRERÍAS PLEGUEZUELO, R. y GARCÍA FERNÁNDEZ, R. (2000). "Análisis de la desigualdad de la renta en Granada, a partir de los datos de la EPF. y diferentes estimaciones de la curva de Lorenz". *Actas XIV Reunión Asepelt, Oviedo*.
99. HERRERÍAS PLEGUEZUELO, R.; PALACIOS GONZÁLEZ, F. y RAMOS RODRÍGUEZ, A. (1.996). "Una metodología flexible para la modelización de la distribución de la renta." *Actas X Reunión Asepelt, Albacete*.

100. HERRERÍAS PLEGUEZUELO, R.; PALACIOS GONZÁLEZ, F.; CALLEJÓN CÉSPEDES, J. y PÉREZ RODRÍGUEZ, E. (1.998). "Desigualdad de la distribución de la renta en las provincias de Castilla y León: un estudio realizado con métodos no paramétricos". Actas del 6º Congreso de Economía Regional de Castilla y León, Zamora.
101. INSTITUTO NACIONAL DE ESTADÍSTICA. (1.992). "Encuesta Continua de Presupuestos Familiares. Metodología".
102. JAKOBSON, U. (1.976). "On the measurement of the degree of progression". *Journal of Public Economics*, 5, pág 161- 168.
103. JOHANSEN, S. (1.978). "The product limit estimator as maximum likelihood estimator". *Scand J. Statist. S*, pág 195-199.
104. JONES, A. (1.989). "A double hurdle model of cigarette consumption". *Journal of Applied Econometrics*, 4, pág 23-39.
105. KAKWANI, W.C. (1.977). "Applications of Lorenz curves in Economic analysis". *Econometrica*, 87, pág 71-80.
106. KALBFLEISCH, J.D. y SPOTT, D.A. (1.970). "Application of likelihood methods to models involving large numbers of parameters (with discussion)". *J. R.S.S., B*, 32, pág 175-208.
107. KAPLAN, E.L. y MEIER, P. (1.958). "Non parametric estimation from incomplete observations". *J.A.S.A.*, 53, pág, 457-481.
108. KIEFER, J y WOLFOWITZ, J. (1.956). "Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters". *Ann. Math. Stat.*, 27, pág 887-906.
109. KOECK, T.H.K. y VAN DIJK (1.978). "Efficient estimation of income distribution parameters". *Journal of Econometrics*, 8, pág 61-74.
110. LAAJIMI, A. y ALBISU, L.M. (1.997). "Análisis del comportamiento de la demanda de aceites y grasas en España". *Investigación Agraria -Economía*, 12, (1,2,3), pág 31-48.
111. LABEAGA, J.M. (1.991). "A dynamic panel data model with limited dependent variables: an application to the demand for tobacco". Documento de trabajo nº 9.201, UNED.
112. LEITCH, R.A.; NETER, J.; PLANTE, R. y SINHA, O. (1.982). "Modified multinomial bounds for large numbers of error in audits". *The Accounting Review*, 57, pág 384-400.
113. LEWBEL, A. (1.991). "The rank of demand systems: theory and nonparametric estimation". *Econometrica*, 59, pág 711-730.
114. LESER, C.E.V. (1.963). "Forms of Engel functions". *Econometrica*, 31, pág 697- 703.

115. LÓPEZ MENÉNDEZ, A.J. (1.991). "Desigualdad de renta y pobreza: una aproximación conceptual y cuantitativa". Tesis doctoral, Universidad de Oviedo.
116. LÓPEZ NICOLÁS, A. (1.995). "Algunas consideraciones sobre el uso de datos microeconómicos en el análisis de la demanda. El caso de la Encuesta Continua de Presupuestos Familiares". Hacienda Pública Española, 132,pág 133-141.
117. LORENZ, M.O. (1.905). "Methods for measuring concentration of wealth". J.A.S.A., 9, pág 209-219.
118. MADDALA, G.S.; (1.983). "Limited-dependent and qualitative variables in econometrics" Cambridge. Cambridge University Press.
119. MANDELBROT, B. (1.960). "The Pareto-Levy law and distribution of income". International Economic Review, 1, pág 79-106.
120. MARCH, L y AMMON. (1.898). "Quelques exemples de distribution salaries". Journal of the Societé Statistique de Paris, pág 517-543.
121. MAS COLELL, A.; WHINSTON, M.D. y GREEN, J.R. (1.995). "Microeconomic theory". Ed. Oxford, pág 50.
122. MATSUMURA, E.M.; TSUI, K. y WONG, W. (1.987). "An extended multinomial Dirichlet model for error bounds for dollar-unit sampling". Contemporary Accounting Research, 6, 2, pág 485-500.
123. McCRAY, J.H. (1.984). "A quasi-bayesian audit risk model for dollar-unit sampling". The Accounting Review, 59, 1, pág 35-51.
124. McCRAY, J.H. (1.986). "A general bayesian risk model for dollar-unit sampling and multiple populations". Manuscrito no publicado. The College of William and Mary. School of Business Administration, Virginia. U.S.A.
125. McCRAY, J.H. (1.993). "Most likely probability curves. Description of software and working paper". Manuscrito no publicado. The College of William and Mary. School of Business Administration, Virginia. U.S.A.
126. McCRAY, J.H. (1.994). "Multivariate most likely probabilities discussion notes". Manuscrito no publicado. The College of William and Mary. School of Business Administration, Virginia. U.S.A.
127. McCRAY, J.H. (1.996). "Practical probabilities for estimation and hipotesis testing". Research Notes. The College of William and Mary. Univ. Williamsburg. Virginia. U.S.A.
128. McCRAY, J.H. (1.998). "Practical multiparameter multivariate likelihood inference". A Working Paper. The College of William and Mary. Univ. Williamsburg. Virginia. U.S.A.

129. McCRAY, J.H; STEELE, A; y HERNÁNDEZ-BASTIDA, A. (1.999). "Likelihood functions of joint parameters for inference". A Working Paper. The College of William and Mary. Univ. Williamsburg. Virginia. U.S.A.
130. McCULLAGH, P. y TIBSHIRANI, R. (1.990). "A simple method for the adjustment of profile likelihoods". *J.R.S.S., B*, 52, pág 325-344.
131. McDONALD, J.B. (1.984). "Some generalized function for the size distribution of income". *Econometrica*, 52, 647-664.
132. McDONALD, J.B. y RANSOM, R.R. (1.979). "Functional forms, estimation techniques and the distribution of income". *Econometrica*, 79, pág 1.513-1.525.
133. McDONALD, J.B. y XU, Y.J. (1.992). "A generalization of the beta of the first kind and second kind with an application". *American Statistical Association: Proceedings of the Business and Economic Statistics Section*, pág 155-160.
134. McDONALD, J.B. y XU, Y.J. (1.995). "A generalization of the beta distribution with application". *Journal of Econometrics*, 66, pág 133-152.
135. MEGHIR, C. y ROBIN, J. M. (1.992). "Frequency of purchase and the estimation of demand systems". *Journal of Econometrics*, 53, pág 53-85.
136. MEREDIZ MONTERO, A. y PENA TRAPERO, J.B. (1.996). "La cuenta de los hogares en 1.973, 1.980 y 1.990, por autonomías, categorías socioprofesionales y clases de hábitat" en Pena y otros (1.996). *Distribución personal de la renta en España*. Ed Pirámide. Madrid. Cap 3.
137. MOLINA, J.A. (1.993). "Evolución de la demanda de productos alimenticios en los países mediterráneos: estimaciones del sistema de demanda casi ideal". *Investigación Agraria-Economía*, 8, 3, pág 331-348.
138. MONTI (1.997). "Empirical likelihood confidence regions in time series models". *Biometrika*, 84, 2, pág 395-405.
139. NETER, J.; LEITCH, R.A. y FIENBERG, S.E. (1.978). "Dollar-unit sampling: Multinomial bounds for total overstatement and understatement errors". *The Accounting Review*, 53, 1, pág 77-93.
140. ORTEGA DATO, J.F. y CALLEALTA BARROSO, J. (2.001). "Estimación robusta de la ocultación de los ingresos personales en España". *Actas XV Reunión Asepelt, La Coruña*.
141. OWEN, A.B. (1.988). "Empirical likelihood ratio confidence intervals for a single functional". *Biometrika*, 75, pág 237-249.
142. OWEN, A.B. (1.990). "Empirical likelihood ratio confidence regions". *Ann. Stat.*, 18-1, pág, 90-120.

143. OWEN, A.B. (1.991). "Empirical likelihood for linear models". *Ann. Stat.*, 19-4, pág 1.725-1.747.
144. PARETO W. (1.895). *G. Degli. Econ.*, 11, pág 59-68.
145. PARRAS ROSA, M. y TORRES RUIZ, F. (1.996). "La estructura del gasto en consumo privado en España y la Unión Europea". *Estudios sobre Consumo*, 36, pág 55-67.
146. PARZEN, E. (1.962). "On estimation of a probability density function and mode". *The Annals of Statistics*, 33, pág 1.065-1.076.
147. PENA, J.B.; CALLEALTA BARROSO, J.; CASAS SÁNCHEZ, J.M.; MEREDIZ MONTERO, A. y NÚÑEZ VELÁZQUEZ, J. (1.996). "Distribución personal de la renta en España". Ed. Pirámide.Madrid.
148. POLLACK, R.A.; WALES, T.J. (1.978). "Estimation of complete demand systems from household budget data: the linear and quadratic expenditure systems". *American Economic Review*, 68, pág 349-539.
149. PRATT, J.W. (1.965) "Bayesian interpretation of standard inferenc statements (with discussion)". *J.R.S.S.*, B-27, pág 169-203.
150. PRIETO ALAIZ, M. (1.998). "Modelización paramétrica de la distribución personal de la renta para España mediante métodos robustos". Tesis doctoral, Universidad de Valladolid.
151. PRIETO ALAIZ, M. y PENA TRAPERO J.B. (2000). "Repercusiones de la ocultación de renta sobre la medición de la desigualdad". *Estudios de Economía Aplicada*, 14, pág 153-172.
152. PUDNEY, S. (1.989). "Modelling individual choice. The econometrics of corners, kinks and holes." Basil Blackwell, Oxford.
153. ROYALL, R.M. (1.997) "Statistical evidence: a likelihood paradigm". Ed. Chapman y Hall.
154. RAMAJO HERNÁNDEZ, J. (1.995) "Análisis de los determinantes de la demanda en Extremadura". *Actas IX Reunión Asepelt*, 1, pág 397-407.
155. RAMAJO HERNÁNDEZ, J. (1.996), "La demanda de productos alimenticios, bebidas y tabaco en Extremadura: un enfoque de regresiones censuradas". *Investigación Agraria-Economía*, 11, 3, pág 470-498.
156. RUBIN, D.B. (1.981). "The bayesian bootstrap". *Ann. Stat.*, 9, pág 130-134.
157. RUIZ CASTILLO, J. (1.987). "La medición de la pobreza y de la desigualdad en España". *Estudios económicos*, 42, Banco de España. Madrid.

158. RUTHERFORD, R.S.G. (1.955). "Income distribution: a new model". *Econometrica*, 23, pág 194-277.
159. SALEM, S.K. y Mount, G.S. (1.974). "A convenient descriptive model of income distribution". *Econometrica*, 42, pág 115-127.
160. SÁNCHEZ GONZÁLEZ, C. y FERNÁNDEZ SÁNCHEZ, M^a P. (1.998). "Empirical likelihood and equilibrium". Ponencia presentada en el 18 International Symposium on forecasting, Edimburgo, 4 junio 1.998.
161. SEN (1.976). "Poverty: An ordinal approach to measurement". *Econometrica*, 44, pág 219-231.
162. SHORROCKS, A.F. (1.980). "The class of additively Descomposable Inequality measures". *Econometrica*, 48, pág 613-625.
163. SHORROCKS, A.F. (1.983). "Ranking income distribution". *Economica*, 50, pág 1-17.
164. SILVERMAN, B.W. (1.986). "Density estimation for statistics and data analysis". Ed Chapman and Hall, Londres.
165. SINGH, S.K. y MADDALA, G.S. (1.976). "A function for the size distribution of income". *Econometrica*, 44, pág 963- 970.
166. TAILLE, C. (1.980). "Lorenz ordering within the generalized Gamma family of income distribution". *Statistical distributions Scientific Work*, 6, pág 181-192.
167. THOMAS, D.R, y GRUNKEMEIER, G.L. (1.975). "Confidence interval estimation of survival probabilities for censored data". *J.A.S.A.*, 70, pág 865-871.
168. THURLOW, L.C. (1.970). "Analysing the american distribution". *American Economic Review*, 60, pág 261-269.
169. TSON, T.S. (1.991). "Robust likelihoods". PhD dissertation, Universidad Johns Hopkins, Baltimore, Maryland.
170. TSUI, K.; MATSUMURA, E.M. y TSUI, K. (1.985). "Multinomial-Dirichlet bounds for dollar-unit sampling in auditing". *The Accounting Review*, 40, 1, pág 76-96.
171. TUSNADY, G. (1.977). "On asymptotically optimal tests". *Ann. Stat*, 5, pág 385-393.
172. VARDI, Y. (1.985). "Empirical distributions in selection bias models". *Ann. Stat.*, 13, pág 178-203.
173. VICTORIA FESER, M.P. (1.993). "Robust methods for personal income distribution models". Tesis doctoral nº 384, Universidad de Ginebra, Suiza.

174. VICTORIA FESER, M.P. y RONCHETTI, E. (1994). "Robust methods for personal income distribution models". *The Canadian Journal of Statistics*, 22, pág 247-258.

175. VICTORIA FESER, M.P. (1995). "Robust methods for personal income distribution models with application to Dagums models". *Research on Economic Inequality*, Jai Press, pág 225-239.

176. WAND, M.P.; MARRON, J.S. y RUPPERT, D. (1991). "Transformations in density estimation". *J.A.S.A.*, 86, pág 343-353.

177. WILKS, S.S. (1938). "The large-sample distribution of the likelihood ratio for testing composite hypotheses". *Ann. Math. Stat.*, 9, pág 60-62.

178. WILKS, S.S. (1962). "Mathematical Statistics". Ed. Wiley, pág 419.

179. ZHANG (1996). "Confidence intervals for a distribution function in the presence of auxiliary information". *Computational Statistics and Data Analysis*, 21, pág 327-342.

ANEXO 1 .

Encuesta Continua de Presupuestos Familiares: Resumen de la metodología. ^[1]

La Encuesta Continua de Presupuestos Familiares (ECPF)^[2] pretende conocer el origen y cuantía de las rentas familiares y su materialización en gastos de consumo. Los objetivos concretos de la investigación son proporcionar las siguientes estimaciones:

- (i) Gastos de consumo trimestrales para el conjunto nacional;
- (ii) Gastos de consumo anuales y su estructura para el conjunto nacional;
- (iii) Gastos de consumo anuales según diversas variables de clasificación;
- (iv) Ingresos trimestrales monetarios y no monetarios para el conjunto nacional;
- (v) Ingresos anuales según su origen y naturaleza para el conjunto nacional;
- (vi) Ingresos anuales según diversas variables de clasificación;
- (vii) Consumo de alimentos y bebidas en cantidades físicas para el conjunto nacional.

En esta Monografía se utilizan los datos de (i), (iv) y (vii).

Las unidades de análisis de la ECPF son los hogares privados que residen en viviendas familiares principales. Se define el hogar como la persona o conjunto de personas que ocupan en común una vivienda familiar principal o parte de ella, y consumen y comparten alimentos y otros bienes con cargo a un mismo presupuesto.

Se utiliza como marco de muestreo el censo de viviendas de 1.981 actualizado, siendo la unidad primaria de muestreo la sección censal y la unidad última de muestreo la vivienda familiar principal.

^[1] Para una exposición detallada de la metodología y de los datos obtenidos de la encuesta, ver " Encuesta Continua de Presupuestos Familiares. Metodología". Instituto Nacional de Estadística. Madrid. 1.992.

^[2] Los precedentes de la ECPF están en la Encuesta Permanente de Consumo, realizada por el INE desde 1.977 hasta 1.983, y en las Encuestas de Presupuestos Familiares que se inicia en 1.958 y finaliza en 1.990.

El ámbito geográfico lo constituye todo el territorio nacional español incluyendo Ceuta y Melilla. El ámbito temporal sometido a estudio es cada ciclo trimestral del año.

Como se ha indicado anteriormente la ECPF estudia fundamentalmente los gastos de consumo y los ingresos de los hogares. Los gastos de consumo de los hogares se clasifican según su naturaleza siguiendo la clasificación del consumo final de las familias, PROCOME, del Sistema Europeo de Cuentas.

Los ingresos de los hogares se clasifican según su naturaleza en monetarios y no monetarios, diferenciándose en cada caso según su origen (trabajo por cuenta ajena, por cuenta propia,...).

Se denomina período de estudio al intervalo de tiempo al cual van referidos los datos. En la ECPF el período de estudio es, en primer lugar, el trimestre y finalmente el año.

Se denomina período muestral al tiempo durante el cual colabora realmente el hogar seleccionado. En la ECPF este período es de una semana.

La semana es un intervalo de tiempo excesivamente breve para abarcar la adquisición de toda la gama de bienes y servicios de consumo. Por ello, durante el período muestral, en la entrevista se pide a los hogares información sobre los pagos realizados en determinados períodos de tiempo por el consumo de aquellos bienes y servicios cuya frecuencia se realiza con periodicidad mayor que la semana. A estos períodos de tiempo se les denomina períodos de referencia. En líneas generales en la ECPF se establecen los siguientes períodos de referencia:

- La semana, que coincide con el período muestral, para la totalidad de alimentos y bebidas consumidos en el hogar y algunos otros bienes y servicios.
- El mes, para la mayor parte del resto de bienes y servicios.
- El trimestre, para los ingresos monetarios y para algunos bienes y servicios.

Las anteriores consideraciones llevan a la utilización de los denominados factores de elevación temporal y que son,^[3]

- 13 para los gastos de referencia semanal,
- 3 para los gastos de referencia mensual,
- 1'083 (= 13/12) para los gastos de referencia trimestral^[4],
- 1 para los ingresos monetarios.

Una de las principales características objeto de estudio, como ya se ha indicado, es los gastos de consumo del hogar. En este punto es conveniente precisar con qué criterio se ha llevado a cabo la evaluación de los gastos de consumo. Para medir el consumo de un bien o

^[3] Teniendo en cuenta que el período de estudio es 1 trimestre = 3 meses = 13 semanas.

^[4] Por necesidades de planificación, para los gastos de consumo se considera como un trimestre las doce semanas anteriores al último día de colaboración, incluyendo por tanto la semana muestral. En cambio, para los ingresos monetarios se consideran los tres meses anteriores al inicio de las entrevistas.

servicio pueden adoptarse distintos criterios en cuanto al momento de su realización, que fundamentalmente son tres, consistentes en obtener el valor de los bienes y servicios:

- (a) Consumidos durante el período de referencia.
- (b) Adquiridos en dicho período con independencia de que hayan sido abonados o no en su totalidad durante el mismo.
- (c) Pagados en este período.

De estos tres criterios en la ECPF se ha adoptado el del pago, es decir, se recogerán los pagos efectuados durante los respectivos períodos de referencia, independientemente de que se hayan consumido realmente los bienes y servicios adquiridos.

Para finalizar este apartado realizamos un breve comentario sobre el tipo de muestreo. El tipo de muestreo corresponde a un muestreo bietápico con estratificación y subestratificación de las unidades de primera etapa, diseñándose en cada comunidad autónoma una muestra independiente.

Las unidades de primera etapa coinciden con las secciones censales existentes en el momento de la encuesta. El tamaño muestral de secciones es 584, seleccionadas entre las 32.000 aproximadamente, en que está dividido el territorio nacional.

Como unidades de segunda etapa, se han tomado las viviendas familiares principales y dentro de éstas no se hace submuestreo alguno, tomándose a todos los hogares y personas que tengan su residencia habitual en las mismas. El tamaño muestral de viviendas es de 3.200.

La clasificación de las secciones en estratos se ha realizado dentro de cada comunidad autónoma según la importancia demográfica del municipio al que pertenecen las secciones, distinguiéndose 4 estratos.^[5]

Para cada estrato se realiza una subestratificación de secciones, clasificando las secciones según la proporción de población activa de tres grupos.^[6]

Teniendo en cuenta los objetivos de la Encuesta pudimos obtener información sobre los gastos e ingresos que se tienen en cuenta en las familias, así como de los artículos de los que se suministran cantidades físicas. Dicha información es la que hemos usado en esta Monografía.

Nosotros hemos trabajado con datos sobre ingresos y gastos de consumo de los hogares, y cantidades físicas de determinados alimentos consumidos por los mismos.

La clasificación de los ingresos por hogar que hace la Encuesta es la siguiente:

^[5] Municipios menores de 10.000 habitantes; de 10.000 a 50.000 excepto capitales de provincia; más de 50.000 habitantes y capitales de provincia; Madrid y Barcelona.

^[6] Para detalles ver "Encuesta Continua de Presupuestos Familiares. Metodología" (1.992), pág.35.

1.- Ingresos monetarios:

- Por trabajo por cuenta ajena;
- Por trabajo por cuenta propia;
- Por rentas de capital y propiedad;
- Pensiones;
- Prestaciones de desempleo;
- Otras transferencias regulares,
- Otros ingresos monetarios.

2.- Ingresos no monetarios:

- Por trabajo por cuenta ajena;
- Por trabajo por cuenta propia;
- Rentas de capital y propiedad;
- Otros ingresos no monetarios.

Para los gastos de consumo del hogar, se dispone de información sobre 226 artículos, que presentamos a continuación, agrupados según la clasificación llevada a cabo por la Encuesta:^[7]

Grupo1: Alimentos, Bebidas y Tabacos.

- 1011- Arroz.
- 1020- Harinas y cereales poco transformados.
- 1041- Otros tipos de pan.
- 1050- Pastelería , pasteles y bizcochos.
- 1062- Pastas alimenticias.
- 1071- Otros productos a base de cereales y féculas.
- 1110- Carne de ovino.
- 1131- Otras aves frescas, refrigeradas o congeladas.
- 1140- Jamón serrano.
- 1152- Jamón de York.
- 1161- Otros productos de charcutería, no enlatados.
- 1170- Carnes preparadas y otros productos que contienen carnes.
- 1182- Otras carnes y casquería.
- 1230- Pescados secos o ahumados.
- 1242- Pescados en conserva y preparados a base de pescado.
- 1251- Crustáceos y moluscos frescos, congelados, refrigerados o cocidos.
- 1281- Otras leches frescas.
- 1290- Leche condensada, concentrada y evaporada.

^[7] El número de cuatro dígitos que precede al artículo es el código que dicho artículo recibe en la Encuesta. Para más detalle ver la Metodología.

- 1302- Otras leches en conserva.
- 1311- Yogur.
- 1320- Otros productos lácteos.
- 1332- Queso manchego o tipo manchego.
- 1341- Otros quesos.
- 1362- Otros productos de huevo.
- 1371- Mantequilla.
- 1380- Margarina.
- 1410- Otros aceites comestibles.
- 1422- Otras grasas animales y vegetales.
- 1491- Frutos secos.
- 1500- Frutas congeladas, en conserva y en zumo.
- 1551- Alubias secas.
- 1560- Garbanzos.
- 1572- Lentejas.
- 1581- Otras legumbres y hortalizas secas.
- 1590- Legumbres y hortalizas congeladas.
- 1602- Legumbres y hortalizas en conserva y preparadas.
- 1620- Mandioca y otros tubérculos.
- 1641- Café natural o tostado.
- 1650- Café soluble.
- 1662- Sucedáneos de café, contengan o no café.
- 1671- Te.
- 1680- Cacao.
- 1692- Confituras, mermelada, miel y jarabes.
- 1701- Chocolate y otros productos de chocolate.
- 1710- Preparación en polvo que contengan cacao.
- 1722- Confitería.
- 1731- Helados.
- 1740- Condimentos, especias y otros productos alimenticios.
- 1752- Aguas minerales.
- 1782- Licores.
- 1812- Otras bebidas alcohólicas.
- 1821- Gastos no desglosables en alimentos y bebidas.
- 1851- Cigarros puros y pequeños cigarros.
- 1860- Otras formas de tabacos y accesorios.

Grupo2: Vestido y Calzado.

- 2013- Prendas exteriores de abrigo y para la lluvia, de hombre.
- 2025- Trajes completos de hombre.
- 2034- Otras prendas exteriores de hombre.
- 2043- Prendas deportivas, de trabajo, de dormir, y de casa, de hombre.
- 2055- Ropa interior de hombre.
- 2064- Prendas exteriores de abrigo y para la lluvia, de mujer.
- 2073- Traje sastre y vestidos de mujer.

- 2085- Otras prendas exteriores de mujer.
- 2094- Prendas deportivas, de trabajo, de dormir y de casa, de mujer.
- 2103- Ropa interior de mujer.
- 2115- Prendas exteriores de niño y niña para abrigo y lluvia.
- 2124- Trajes y vestidos completos de niño y niña.
- 2133- Otras prendas exteriores de niño y niña.
- 2145- Prendas deportivas, de colegio, de dormir y de casa, de niño y niña.
- 2154- Ropa interior de niño y niña.
- 2163- Ropa de bebé (menos de 3 años).
- 2175- Complementos de vestir.
- 2184- Reparación de prendas y artículos de vestir.
- 2193- Gastos no desglosables en vestido.
- 2205- Calzado de hombre.
- 2214- Calzado de mujer.
- 2223- Calzado de niño y niña (hasta los 13 años).
- 2235- Reparación de calzado.
- 2244- Gastos no desglosables en calzado.

Grupo 3: Vivienda, Calefacción, Alumbrado.

- 3018- Alquiler neto de la vivienda principal.
- 3027- Alquiler de la vivienda secundaria.
- 3036- Otros alquileres.
- 3045- Alquiler imputado a la vivienda principal en propiedad.
- 3054- Alquiler imputado a la vivienda secundaria en propiedad.
- 3066- Gastos comunitarios netos de la vivienda principal ocupada por su propietario.
- 3078- Gastos de reparación, entretenimiento y mejora de la vivienda principal ocupada por su propietario.
- 3087- Gastos comunitarios de la vivienda secundaria ocupada por su propietario.
- 3096- Gastos de reparación, entretenimiento y mejora de la vivienda secundaria ocupada por su propietario.
- 3108- Gastos de reparación, acondicionamiento y conservación de la vivienda, realizados por el inquilino.
- 3117- Contribución urbana, arbitrios municipales, recogida de basuras y tasas de alcantarillado, en vivienda principal.
- 3126- Contribución urbana, arbitrios municipales, recogida de basuras y tasas de alcantarillado, en vivienda secundaria.
- 3138- Gastos de distribución de agua (incluido gastos de contrato) en vivienda principal.
- 3147- Gastos de distribución de agua (incluido gastos de contrato) en vivienda secundaria.
- 3156- Energía eléctrica (incluido los gastos de contrato) en vivienda principal.
- 3186- Energía eléctrica (incluido los gastos de contrato) en vivienda secundaria.
- 3195- GLP (incluido gastos de contrato) en vivienda principal.
- 3204- GLP (incluido gastos de contrato) en vivienda secundaria.
- 3213- Combustibles líquidos (incluido gastos de contrato) en vivienda principal.
- 3225- Combustibles líquidos (incluido gastos de contrato) en vivienda secundaria.
- 3234- Gastos de carbón, coque y otros combustibles sólidos en vivienda principal.

- 3246- Gastos de calefacción colectiva en vivienda principal.
- 3255- Gastos de carbón, coque y otros combustibles sólidos en vivienda secundaria.
- 3267- Gastos de calefacción colectiva en vivienda secundaria.

Grupo 4: Muebles, Artículos de mobiliario. Artículos de menaje para el hogar y gastos de conservación corrientes de la casa.

- 4017- Muebles.
- 4026- Otros muebles y accesorios de amueblamiento.
- 4038- Alfombras y revestimientos de suelo.
- 4044- Reparación de muebles y accesorios fijos, alfombras y otros revestimientos de suelos.
- 4053- Artículos textiles y otros artículos de amueblamiento del hogar.
- 4065- Reparación de artículos textiles y otros artículos de amueblamiento del hogar.
- 4074- Gastos no desglosables en textiles para el hogar y otros accesorios de amueblamiento.
- 4086- Frigoríficos, congeladores y combinados.
- 4098- Máquinas lavadoras, escurridoras, secadoras y planchadoras.
- 4107- Lavavajillas.
- 4116- Cocinas e infiernillos.
- 4128- Aparatos de calefacción, eléctricos o no.
- 4137- Aparatos de limpieza y conservación.
- 4146- Máquinas de coser y tricotar.
- 4158- Otros aparatos para el menaje.
- 4164- Reparación de aparatos de calefacción y cocina, frigoríficos, lavadoras, y otras grandes máquinas para el hogar.
- 4173- Cristalería y vajilla.
- 4185- Cuberterías.
- 4194- Utensilios de cocina y menaje, de todos los materiales, no eléctricos.
- 4203- Utensilios de jardín.
- 4215- Material eléctrico usado en el hogar.
- 4224- Reparaciones de cristalería, vajilla y utensilios de menaje y jardín.
- 4233- Gastos no desglosables en cristalería, vajilla y utensilios del hogar.
- 4242- Artículos de limpieza y conservación del hogar.
- 4251- Otros productos no duraderos para el hogar.
- 4263- Tintorería, lavandería y zurcidos.
- 4275- Otros servicios de la vivienda.
- 4284- Servicio doméstico.

Grupo 5: Medicinas y otros productos farmacéuticos.

- 5010- Medicamentos adquiridos con subvención.
- 5022- Medicamentos adquiridos sin subvención.
- 5034- Otros productos farmacéuticos y de higiene.
- 5046- Aparatos y material terapéutico con subvención y reparaciones.
- 5058- Aparatos y material terapéutico sin subvención y reparaciones.

- 5064- Servicios médicos.
- 5073- servicios de dentista.
- 5085- Servicios de enfermeras y otros sanitarios.
- 5097- Cuidados en hospitales y asimilados.
- 5106- Remuneración de los servicios de seguros contra accidentes y de enfermedad.
- 5115- Gastos no desglosables en servicios médicos y conservación de la salud.

Grupo 6: Transportes y Comunicaciones.

- 6018- Automóviles nuevos.
- 6027- Automóviles usados.
- 6036- Otros medios de transporte personal.
- 6045- Neumáticos, piezas de recambio.
- 6054- Gastos de conservación y reparación efectuados por un taller.
- 6060- Carburantes.
- 6072- Lubricantes.
- 6087- Otros gastos.
- 6093- Gastos no desglosables en la utilización de vehículos.
- 6102- Transporte urbano colectivo.
- 6111- Taxis y gran turismo urbanos y otros medios de transporte locales.
- 6123- Transportes por carretera y ferrocarriles.
- 6135- Otros servicios de transporte.
- 6141- Otros gastos ligados al servicio de transportes.
- 6150- Servicios de correos.
- 6162- Telégrafos.
- 6171- Teléfonos.
- 6177- Teléfonos, recibos.

Grupo 7: Esparcimiento, Espectáculos, Enseñanza y Cultura.

- 7017- Aparatos de radio.
- 7026- Televisores.
- 7038- Aparatos electrodomésticos.
- 7047- Aparatos y material de fotografía y cine y de proyección.
- 7056- Instrumentos de música, embarcaciones y otros bienes.
- 7065- Discos, cintas, cassettes magnéticas y otros accesorios.
- 7074- Artículos y accesorios de deporte.
- 7083- Juegos y juguetes, instrumentos pequeños de música, sellos y álbumes para filatelia y otros.
- 7086- Ordenadores personales, vídeos y juegos electrónicos.
- 7095- Películas, otros artículos fotográficos y álbumes para fotos.
- 7101- Plantas y flores.
- 7110- Animales de compañía y recreo.
- 7125- Piezas, accesorios y reparaciones de los artículos recreativos.

- 7131- Gastos de cine, teatros y conciertos.
- 7140- Gastos en otros espectáculos.
- 7152- Entradas a salas de fiestas, discotecas, cafés-teatros, tablaos, etc.
- 7164- Otros gastos para el ejercicio de actividades deportivas y recreativas.
- 7173- Impuestos y alquileres por el uso de aparatos recreativos.
- 7185- Otros servicios recreativos.
- 7194- Libros no de texto.
- 7200- Diarios, revistas y otros impresos.
- 7215- Enseñanzas anteriores a EGB. Recibos.
- 7218- Enseñanzas anteriores a EGB. Matrículas.
- 7224- EGB. Recibos.
- 7227- EGB. Matrículas.
- 7233- Enseñanzas de BUP, COU y FP. Recibos.
- 7236- Enseñanzas de BUP, COU y FP. Matrícula y otros gastos.
- 7248- Transporte escolar.
- 7257- Comedor escolar.
- 7260- Material escolar.
- 7275- Enseñanza universitaria. Recibos.
- 7278- Enseñanza universitaria. Matrículas y otros gastos.
- 7284- Otras enseñanzas. Recibos.
- 7287- Otras enseñanzas. Matrículas y otros gastos.
- 7296- Gastos de alojamiento en régimen de internado.

Grupo 8: Otros bienes y servicios.

- 8013- Servicios de peluquería, institutos de belleza y análogos para caballeros.
- 8022- Servicios de peluquería, institutos de belleza y análogos para señoras.
- 8034- Artículos duraderos para el cuidado personal.
- 8040- Artículos no duraderos para el cuidado personal.
- 8055- Joyas, relojes, sortijas y piedras preciosas.
- 8064- Artículos de fumador.
- 8073- Artículos de viaje y otros.
- 8082- Material y objetos de papelería y dibujo.
- 8091- Comidas y cenas fuera del hogar.
- 8097- Banquetes de ceremonias y celebraciones, fuera del hogar.
- 8100- Comidas y cenas gratuitas, bonificadas o semigratuitas en el lugar de trabajo.
- 8112- Otros servicios de restauración.
- 8121- Desayunos fuera del hogar.
- 8130- Otros gastos en cafés.
- 8148- Gastos en hoteles y servicios de alojamiento análogos.
- 8157- Viajes organizados, todo incluido.
- 8163- Servicios financieros no declarados en otra parte.
- 8172- Servicios de referencia semanal, no incluidos en otra parte.
- 8175- Servicios de referencia mensual, no incluidos en otra parte.
- 8184- Gastos no desglosables correspondientes a este grupo.

Grupo 9: Otros gastos no mencionados anteriormente.

- 9018- Impuestos sobre animales; permiso de caza, de pesca y otras tasas.
- 9027- Primas brutas de seguros.
- 9030- Dinero de bolsillo a los niños menores de 14 años.
- 9042- Juegos de apuestas de azar.
- 9051- Transferencias ocasionales a otros hogares.
- 9063- Remesas a miembros del hogar de 14 y más años, no residentes en la vivienda.
- 9078- Artículos funerarios.
- 9084- Transferencias regulares a otros hogares.
- 9093- Contribuciones voluntarias, cotizaciones .etc.
- 9105- Transferencias a instituciones sin fines de lucro.

Acerca del consumo en cantidades físicas de alimentos se dispone de información sobre 32 artículos que detallamos a continuación.

- 1032- Pan corriente de trigo.
- 1080- Carne de vaca, fresca, congelada y refrigerada.
- 1092- Carne de ternera, fresca, congelada y refrigerada.
- 1101- Carne de cerdo, fresca, congelada y refrigerada.
- 1122- Pollo y gallina frescos, congelados y refrigerados.
- 1191- Merluza fresca.
- 1200- Pescadilla fresca.
- 1212- Merluza y pescadilla congelada.
- 1221- Otros pescados frescos o congelados de mar o de río.
- 1260- Leche fresca o pasteurizada de vaca.
- 1272- Leche esterilizada de vaca.
- 1350- Huevos frescos.
- 1392- Aceite de oliva.
- 1401- Aceite de girasol.
- 1431- Naranjas.
- 1440- Otros cítricos.
- 1452- Plátanos.
- 1461- Manzanas.
- 1470- Peras.
- 1482- Otras frutas frescas.
- 1512- Coliflores y coles.
- 1521- Tomates.
- 1530- Judías verdes.
- 1542- Otras legumbres y hortalizas frescas.
- 1611- Patatas.
- 1632- Azúcar.
- 1761- Gaseosas sin sabor.
- 1770- Refrescos con sabor.

- 1791- Vinos de mesa.
- 1800- Cervezas.
- 1830- Cigarrillos negros.
- 1842- Cigarrillos rubios.

De todos estos ingresos, gastos y consumo en cantidades físicas disponemos de observaciones correspondientes a 3.114 hogares, pero en esta Memoria se han utilizado sólo los que se han comentado en cada capítulo.

ANEXO 2

Programas informáticos.

En las páginas siguientes se encuentran impresos los programas informáticos utilizados para los cálculos de inferencia que aparecen en los diferentes capítulos de la Memoria. Han sido realizados con el programa Matemática, versión 3.0.

Son dos programas muy similares. Uno de ellos se ha utilizado para todos los cálculos correspondientes al capítulo de las cantidades compradas en unidades físicas, así como para los resultados de inferencia de cocientes o participaciones. El otro se ha empleado para los ingresos y los gastos, expresados en unidades monetarias.

A la hora de las representaciones gráficas de la verosimilitud empírica que aparecen en los distintos capítulos, hemos escogido, para una mejor visualización, sólo el valor de la moda y otros cercanos a él. La gráfica de la verosimilitud empírica recuerda a una función de densidad, pero no lo es. La función de densidad tiene como dominio el espacio muestral y la verosimilitud empírica, el espacio paramétrico. Así, en el eje de abscisas encontraremos los valores que puede tomar la magnitud desconocida valor medio, aunque aquí sólo los valores que presentan una verosimilitud empírica relativamente alta. No obstante, el programa permite conocer todos los valores del parámetro y su verosimilitud correspondiente.


```

<<Graphics`Colors`
k=3114; (*FIJAR AQUI EL NUMERO DE ERRORES DISTINTOS DETECTADOS*)
Array[x,k];
Array[x2,k];
Array[n,k];
Array[p,k];
ptos=1001; (*FIJAR EL NUMERO DE PUNTOS *)
entera[x_]=If[x<0,Ceiling[x],Floor[x]];
Array[like,ptos-1];
Array[thetas,ptos];
Array[ppostacum,ptos-1];
Array[ppostnor,ptos-1];
Array[percentil,25];
Array[prob,25];
Array[priori,ptos-1];
Array[antes,{5,ptos-1}];
Array[daspues,{5,ptos-1}];

fichdatos=OpenRead["c:\\nombre del fichero.txt"];
i=1;
salida=0;
datomaximo=0;
While[salida==0,
  {dato[i]=Read[fichdatos,RecordSeparators->{"\n"}];
  (*Print[i," ",dato[i]];*)
  (*Input[];*)
  If[datomaximo<dato[i],datomaximo=dato[i]];
  If[dato[i]==EndOfFile,{Print["final"];salida=1;});
  i=i+1;
  };
Close[fichdatos]
nobs=i-2;
Print["Numero de observaciones= ",nobs];

(*Do[Print["dato["i,"]=",dato[i]],{i,1,nobs}];*)

```

```

Do[{x[i] = 0; n[i] = 0}, {i, 1, k}];
nk = 1;
x[1] = dato[1];
n[1] = 1;
Do[{Do[{
  If[dato[i] < x[j],
    {Do[{
      n[li] = n[li - 1];
      x[li] = x[li - 1]
    }, {li, nk + 1, j + 1, -1}];
    nk++;
    n[j] = 1;
    x[j] = dato[i];
    Break[];
  },
  {If[dato[i] == x[j],
    {n[j]++;
    Break[];
  }];
  }
  ] (* cierra If[dato[i]<x[j] *)
  }, {j, 1, nk}];
If[dato[i] > x[nk],
  {n[nk + 1] = 1;
  x[nk + 1] = dato[i];
  nk++;
  }
]; (* cierra If[dato[i]>x[nk] *)
}, {i, 2, nobs}]
k = nk;

(* Do[Print[kk,"-> ",x[kk]," ",n[kk]],{kk,1,nk}];*)
a = x[1];
b = (110 / 100) datomaximo;
xj = b;

```

```

(* CALCULOS INICIALES *)

nt=Sum[n[i],{i,1,k}];
cptest=Sum[n[i]*x[i],{i,1,k}];
cptest=cptest/nt;
scalef=entera[N[-Sum[n[i]*Log[n[i]/nt],{i,1,k}]]];
desvtip=Sqrt[Sum[n[i]*(x[i]-cptest)^2,{i,1,k}]/nt];
xj=(110/100)datomaximo;
xmax=x[1];
xmin=x[1];
Do[{xmax=Max[xmax,x[i+1]],
    xmin=Min[xmin,x[i+1]]},
    {i,1,k-1}];
Print[xmin," ",xmax];
Do[x2[i]=x[i]-cptest,{i,1,k}];
xmin2=xmin-cptest;
xmax2=xmax-cptest;
imax=1;
Do[If[n[i]<n[i+1],imax=i+1],{i,1,k-1}];
moda=x[imax];

(* PUNTOS EQUIDISTANTES DE theta *)
(* CALCULO DE LA PRIORI EN ESOS PUNTOS *)
(* thetas[i] y priori[i] *)

pasada=1;
sumprior=0;
(*ftheta=OpenWrite["C:\\theta.dat"]*)
(*fprior=OpenWrite["C:\\prior.dat"]*)
theta=a+(b-a)/(ptos-1);
incretheta=(b-a)/(ptos-1);
Do[{{(* Write[ftheta,N[theta,9]]*)
    thetas[i]=theta;
    theta=incretheta+theta;
    priori[i]=1;
    If[Mod[i,50]==0,Print[i]];
    (* Write[fprior,N[paso,9]]*)
    }
    ,{i,1,ptos-1}];
(*Close[ftheta]*)
(*Close[fprior]*)

(* NORMALIZACION DE LA PRIORI, ARRAYS *)

sumprior=Sum[priori[i],{i,1,ptos-1}];
Do[priori[i]=priori[i]/sumprior,{i,1,ptos-1}];
Sum[priori[i],{i,1,ptos-1}]

(* CALCULO DE LA VEROSIMILITUD EN LOS VALORES DE theta
CALCULADOS ANTES Y LA POSTERIORI EN ESOS PUNTOS *)
(* BUCLE VARIANDO theta PARA CALCULAR EL MAXIMO DE LA
VEROSIMILITUD, xj ESSEA FIJO EN 100, *)

jj=1;
f[r_]:=theta*nt/(1+theta*r)-Sum[n[i]*x2[i]/(1+r*x2[i]),
    {i,1,k}];

```

```

fprima[r_]:= -nt*theta^2/(1+theta*r)^2+
Sum[n[i]*x2[i]^2/(1+r*x2[i])^2,{i,1,k}];

(* VIENDO r COMO r[theta], LA PENDIENTE (DERIVADA) DE r
CON RESPECTO A theta es acel[r_], QUE SE UTILIZARA
PARA IR ACTUALIZANDO A CADA PASO EL VALOR INICIAL DE r*)

acel[r_]:=(-nt/(1+theta*r)^2)/fprima[r];

(* AHORA theta SE LEEVA DEL FICHERO theta.dat *)

(*ftheta=OpenRead["C:\\theta.dat"]*)
(*fprior=OpenRead["C:\\prioror.dat"]*)
(*flike=OpenWrite["c:\\like.dat"]*)
(*fpost=OpenWrite["c:\\post.dat"]*)
sumpost=0;
Tolr=.00001;
epsilon=.00000001;
r1:=-1/xmax2+epsilon;
r2:=-1/xmin2-epsilon;
rop=r2-epsilon;
r1aux=r1;
r2aux=r2;

Do[
(*Print["-----"];*)
(*Print["theta= ",theta];*)
(*theta=Read[ftheta,Expression];*)
(*prior=Read[fprior,Expression];*)

theta=thetas[jj];
prior=priori[jj];
(*Print[N[thetas[jj]]," ",priori[jj];*)
a2=nt/(theta-xj);
a1=-xj*a2;
Do[p[i]=n[i]/(a2*(x[i]-xj)),{i,1,k}];
pj=1-Sum[p[i],{i,1,k}];

(* SI pj<0, ES IMPOSIBLE LUEGO pj=0 Y SOLO INTERVIENEN LOS
p_i MUESTRALES *)

If[pj<0,
{
theta=theta-ctest;
contador1=0;
(*Print["r1= ",r1," r2=",r2];*)
While[Abs[f[rop]]>Tolr && contador1<20,
{(*Print["contador1=",contador1];*)
inc= -f[rop]/fprima[rop];
increaux=inc;
contador1++;
ropaux=rop;
rop=rop+inc;
contador2=1;
While[(rop>r2 || rop<r1) && contador2<4,
{inc=increaux/contador2;
rop=ropaux+inc;
(*Print["ctdor2=",contador2];*)
contador2++}}];

```

```

If[Abs[f[rop]]<Abs[f[ropaux]],
  {rop=Re[rop];
   },
  {contador1=20;
   }
];
(*Print[" contador1=",contador1," rop= ",rop,
  " f[rop]= ",f[rop]];*)
}>(*cierza while*);

(* BISECCION EN EL CASO DE QUE NEWTON NO CONVERJA *)
bandera=0;
(*Plot[f[x],{x,r1,x2}];*)
(*Print["imagenes", " ",r1," ",f[r1]];*)
(*Print["imagenes", " ",r2," ",f[r2]];*)
(*Input[];*)
If[Abs[f[rop]]>Tolr || rop>x2 || rop<r1,
  {bandera=1;
   r1aux=r1;
   r2aux=r2;
   rop=(r1aux+r2aux)/2;
  (*Print["imagenes", " ",f[r1aux]," ",f[r2aux]];*)
  (*Print["r1aux y r2aux ",r1aux," ",r2aux];*)
  i=0;
  While[Abs[f[rop]]>Tolr,
    {
     rop=(r1aux+r2aux)/2;
     (*Print["rop=",rop," f[rop],f[r1]=",f[rop]," ",
      f[r1aux]];*)
     If[f[rop]*f[r1aux]>0,r1aux=rop,r2aux=rop];
     (*Print[i," ",r1aux," ",r2aux];*)
     (*Print[i," ",f[r1aux]," ",f[r2aux]];*)
     i=i+1;
    }
  ];
  priorr=rop;
  Lam=nt/(1+theta*rop);
  mu=lam*rop;

  (*Print["bandera=",bandera];*)
  (*Print["rop= ",rop," f[rop]= ",f[rop]];*)
  If[pasada==1,incre=incretheta,
    {If[jj<=ptos1,incre=incre1;
     If[ptos1<jj<=ptos1+ptosmedia,incre=incre2];
     If[jj>ptos1+ptosmedia,incre=incre3];
    }
  ];
  paso=5*(piorr+acal[piorr]*incre);
  If[r1<=paso<=r2,rop=paso];
  Do[{p[i]=n[i]/(lam+mu*x2[i]),{i,1,k}};

  theta=theta-opts;

  ];

pasolike=Re[Exp[Sum[n[i]*Log[p[i]],{i,1,k}]]];
(*Print[N[theta], " like[" ,jj,"]=",N[pasolike]];*)
like[jj]=N[pasolike.12];
(*Write[File N[pasolike.10]];*)
sumpost=sumpost+N[prior*pasolike,10];
(*Write[File N[prior*pasolike,10]];*)
ppostor[jj]=N[prior*pasolike,10];

```

```

If[Mod[jj,50]==0,Print[jj]];

},{jj,4,ptos-4};
Do[{like[jj]=0.0;ppostnor[ptos-jj]=0.0;
  like[ptos-jj]=0.0;ppostnor[jj]=0.0},{jj,1,3}]

(*Close[flike]*)
(*Close[ftheta]*)
(*Close[fprior]*)
(*Close[fpost]*)

(* NORMALIZACION DE LA POSTERIORI solo para ARRAYS
  Y CALCULO DE LA POSTERIORI ACUMULADA *)

sumpost=Sum[ppostnor[i],{i,1,ptos-1}];
Do[ppostnor[i]=ppostnor[i]/sumpost,{i,1,ptos-1}]
ppostacum[1]=ppostnor[1];
Do[ppostacum[i]=ppostacum[i-1]+ppostnor[i],{i,2,ptos-1}]
Sum[ppostnor[i],{i,1,ptos-1}]

Do[Print[N[thetas[jj]], " like[" , jj, "]" = ",
  like[jj]], {jj, 1, 1000, 100}]

(* SUMO LAS VEROSIMILITUDES PARA REESCALARLAS Y
  REPRESENTARLAS CON LA PRIORI *)

sumlike=Sum[like[i],{i,1,ptos-1}];
numptos=ptos-1;

(* GRAFICA DE LA VEROSIMILITUD *)

tablalike:=Table[{thetas[i],N[like[i]/sumlike]},{i,1,numptos}];
glike=ListPlot[tablalike,PlotJoined->True,
  PlotStyle->{RGBColor[.2,0.2,0.8]},PlotRange->All]
tablalikeparcial:=Table[{thetas[i],N[like[i]/sumlike]},{i,1,350}];
glikeparcial=ListPlot[tablalikeparcial,
  PlotStyle->{RGBColor[.2,0.2,0.8]},PlotJoined->True,PlotRange->All]

(*****)

(* CALCULO DE LOS PERCENTILES, MEDIA Y DESVIACION
  TIPICA A POSTERIORI *)

prob[1] = .01;
prob[2] = .025;
prob[22] = .975;
prob[23] = .99;
prob[24] = 1.0;
Do[prob[i] = (i - 2) * .05, {i, 3, 21}];
tol3 = .00001;

```

```

sumpr = 0;
i = 1;
While[sumpr < .025,
  {sumpr = sumpr + priori[i];
  i = i + 1;
  }];
q025priori = N[10000 * (thetas[i] + thetas[i - 1]) / 2, 7];
While[sumpr < .975,
  {sumpr = sumpr + priori[i];
  i = i + 1;
  }];
q975priori = N[10000 * (thetas[i] + thetas[i - 1]) / 2, 7];

i = 1;
j = 1;
Do[{While[ppostacum[j] < prob[i], j = j + 1];
  IF[j == 1 && ppostacum[1] > prob[i],
    {m = ppostacum[1] / (thetas[1] - a);
    IF[Abs[m] < tol3, percentil[i] = thetas[1],
      percentil[i] = thetas[1] +
      (prob[i] - ppostacum[1]) / m;
    }];
  (* Print["j= ",j," i= ",i," prob[i]=",prob[i]];*)
  (*Print[" m=",m," ppostacum[1]=",ppostacum[1],
    " percentil[i]=",percentil[i]];*)
  }
];
IF[j > 1,
  {
  m = (ppostacum[j] - ppostacum[j - 1]) /
    (thetas[j] - thetas[j - 1]);
  (* SI m ES MUY PEQUEÑO PUEDE DAR PROBLEMAS
    AL DIVIDIR POR 0 (m=0) *)
  IF[Abs[m] < tol3, {percentil[i] = thetas[j]; j = j - 1,
    percentil[i] = thetas[j] + ((prob[i] - ppostacum[j]) / m)};
  percentil[i] = N[percentil[i], 7];
  }
  ]
}, {i, 1, 23}];

media = 0;
Do[media = media + thetas[i] * ppostnor[i], {i, 1, ptos - 1}];
varianza = 0;
Do[varianza = varianza + ((thetas[i] - media) ^ 2) * ppostnor[i],
  {i, 1, ptos - 1}];
desvtip = Sqrt[varianza];
mediapost = media;
sigma = desvtip;
mediapost
Do[If[ppostnor[j - 1] < ppostnor[j],
  {mediapost = N[thetas[j]],
  (*Print[j," ",mediapost];*)
  }];
];

```

```

, {j, 2, ptos - 1}]

Print["Media a posteriori=", " ", media];
Print["Desviación típica=", " ", desvtip];
Do[Print[prob[i], " ", percentil[i]];
, {i, 1, 23}]
Print["Media muestral= ", N[cpstat]];
Print["Media a posteriori=", " ", media];
Print["Moda a posteriori=", " ", modapost];

(*****Cálculo cociente de verosimilitudes***)

ptos

maxverosim = like[1];
Do[If[maxverosim < like[i],
      {maxverosim = like[i];
       imaxverosim = i}];
, {i, 2, ptos - 1}];
Print["El maximo de la verosimilitud se alcanza en:"];
Print["theta[" , imaxverosim, "]=", thetas[imaxverosim],
      " , like[" , imaxverosim, "]=", like[imaxverosim]];
Print["-----"];
divisor = Input["Valor para el cociente de la verosimilitud="];
maxver = like[imaxverosim];

i = 1;
cociente = 50000;
While[i < ptos && cociente > divisor,
      {i++;
       If [like[i] != 0, cociente = maxver / like[i]];
       (*Print[i, " cociente=", cociente];*)
      }];
extinf = i;
While[i < ptos && cociente < divisor,
      {i++;
       If [like[i] != 0, cociente = maxver / like[i]];
       (*Print[i, " cociente=", cociente];*)
      }];
i--;
extsup = i;
Print["Constante para el cociente de verosimilitudes=", divisor];
Print["Intervalos [punto ", extinf, " , punto ", extsup, " )
      = (", thetas[extinf], " , ", thetas[extsup], " )"];

Do[Print[jj, " , cociente= ", maxver / like[jj]], {jj, extinf - 1, extsup + 1}]

```

```

(*<<Graphics Colors*)
k=4000; (*FIJAR AQUI EL NUMERO DE ERRORES DISTINTOS DETECTADOS*)
Array[x,k];
Array[x2,k];
Array[n,k];
(*Array[p,k];*)
ptos=1001; (*FIJAR EL NUMERO DE PUNTOS *)
Array[like,ptos-1];
Array[khatas,ptos];
Array[ppostacua,ptos-1];
Array[ppostnor,ptos-1];
Array[percentil,25];
Array[prob,25];
Array[priori,ptos-1];
Array[antes,{5,ptos-1}];
Array[daspues,{5,ptos-1}];
entera[x_]:=IF[x<0,Ceiling[x],Floor[x]]*1.0;

fichdatos=OpenRead["c:\nombre del fichero.txt"];
i=1;
salida=0;
datomaximo=0;
While[salida==0,
  {dato[i]=Read[fichdatos RecordSeparators->{"\n"}];
  (*Print[i," ",dato[i]];*)
  (*Input[];*)
  IF[datomaximo<dato[i],datomaximo=dato[i]];
  IF[dato[i]==EndOfFile,{Print["final"];salida=1;};
  i=i+1;
  };
Close[fichdatos];
nobs=i-2;
Print["Numero de observaciones= ",nobs];

(* DATOS EMBOSCADOS Y REDONDADOS A ENTEROS*)

Do[dato[i]=entera[dato[i]/1000],{i,1,nobs}];
datomaximo=entera[datomaximo/1000];

(* ORDENACION DE LOS DATOS *)

Sum[dato[i],{i,1,nobs}];
valores=Table[0,{i,1,nobs}];
Do[valores, {i, {dato[i], {i,1,nobs}}];
valoresordenados=Sort[valores];
Do[dato[i]=valoresordenados[[i]],{i,1,nobs}];
Sum[dato[i],{i,1,nobs}];

```

```

(* LISTA DE VALORES *)
Do[Print["dato[" , i, "]" = ", dato[i]], {i, 1, nobs}];

(* CALCULO DE LCS x[i] Y SUS FRECUENCIAS n[i] *)

Do[{x[i] = 0; n[i] = 0}, {i, 1, k}];
nk = 1;
x[1] = dato[1];
n[1] = 1;
rep = 0;
Do[(*Print[i, " ", dato[i], " ", i - rep - 1, " ", x[i - rep - 1]];*)
  If[Mod[i, 100] == 0,
    Print[i, " ", i - rep - 1, " ", x[" , i - rep - 1, "]" = ", x[i - rep - 1]];
    If[dato[i] == x[i - rep - 1],
      {n[i - rep - 1]++;
       rep++;
      },
      {(*Print ["Nuevo item x[i]"];*)
       x[i - rep] = dato[i];
       n[i - rep] = 1;
       (* Print[i - rep, " ", x[i - rep]];*)
      }
    ],
  ], {i, 2, nobs}];
i = i - 1;
Print[i, " ", i - rep - 1, " ", x[" , i - rep - 1, "]" = ", x[i - rep - 1]];
k = nobs - rep;
nk = k;

a = x[1];
b = (110 / 100) datomaximo;
xj = b;

a
b
x[1]
x[nk]
rep
nk
nk + rep

```

```

(* CALCULOS INICIALES *)

nt=Sum[n[i],{i,1,k}];
cptest=Sum[n[i]*x[i],{i,1,k}];
cptest=cptest/nt;
scalef=entera[N[-Sum[n[i]*Log[n[i]/nt],{i,1,k}]]];
desvtip=Sqrt[Sum[n[i]*(x[i]-cptest)^2,{i,1,k}]/nt];
xj=(110/100)datomaximo;
xmax=x[1];
xmin=x[1];
Do[{xmax=Max[xmax,x[i+1]],
    xmin=Min[xmin,x[i+1]]},
    {i,1,k-1}];
Print[xmin," ",xmax];
Do[x2[i]=x[i]-cptest,{i,1,k}];
xmin2=xmin-cptest;
xmax2=xmax-cptest;
imax=1;
Do[If[n[i]<n[i+1],imax=i+1],{i,1,k-1}];
moda=x[imax];

N[cptest]

(* PUNTOS EQUIDISTANTES DE theta *)
(* CALCULO DE LA PRIORI EN ESOS PUNTOS *)
(* thetas[i] y priori[i] *)

pasada=1;
sumprior=0;
(*ftheta=OpenWrite["C:\\theta.dat"]*)
(*fprior=OpenWrite["C:\\prior.dat"]*)
theta=N[a+(b-a)/(ptos-1)];
incretheta=N[(b-a)/(ptos-1)];
Do[(* Write[ftheta,N[theta,9]】*)
    thetas[i]=theta;
    theta=incretheta+theta;
    priori[i]=1;
    If[Mod[i,50]==0,Print[i]];
    (* Write[fprior,N[paso,9]】*)
    ],{i,1,ptos-1}];
(*Close[ftheta]*)
(*Close[fprior]*)

(* NORMALIZACION DE LA PRIORI, ARRAYS !!! *)

sumprior=Sum[priori[i],{i,1,ptos-1}];
Do[priori[i]=priori[i]/sumprior,{i,1,ptos-1}];
Sum[priori[i],{i,1,ptos-1}]

(* CALCULO DE LA VEROSIMILITUD EN LOS VALORES DE theta
CALCULADOS ANTES Y LA POSTERIORI EN ESOS PUNTOS *)
(* BUCLE VARIANDO theta PARA CALCULAR EL MAXIMO DE LA
VEROSIMILITUD, xj ESTA FIJO EN 100, *)

jj=1;

```

```

f[r_]:=theta*nt/(1+theta*r)-Sum[n[i]*x2[i]/(1+r*x2[i]),
  {i,1,k}];
fprima[r_]:=-nt*theta^2/(1+theta*r)^2+
  Sum[n[i]*x2[i]^2/(1+r*x2[i])^2,{i,1,k}];

(* VIENDO r COMO r[theta], LA PENDIENTE (DERIVADA) DE r
  CON RESPECTO A theta es acel[r_], QUE SE UTILIZARA
  PARA IR ACTUALIZANDO A CADA PASO EL VALOR INICIAL DE r*)

acel[r_]:=(-nt/(1+theta*r)^2)/fprima[r];

(* AHORA theta SE LEERA DEL FICHERO theta.dat *)

(*ftheta=OpenRead["C:\\theta.dat"]*)
(*fprior=OpenRead["C:\\prioror.dat"]*)
(*flike=OpenWrite["c:\\like.dat"]*)
(*fpost=OpenWrite["c:\\post.dat"]*)
sumpost=0;
Tolr=.00001;
Tolr001=.001;
epsilon=.00000001;
r1:=-1/xmax2+epsilon;
r2:=-1/xmin2-epsilon;
rop=r2-epsilon;
r1aux=r1;
r2aux=r2;
cocientepaso=Table[N[n[i]/(x[i]-xj)],{i,1,nk}];
p=Table[0,{i,1,nk}];
sumacocientepaso=Sum[cocientepaso[[i]],{i,1,nk}];
Print["sumcoc=",sumacocientepaso];
Do[(*Print["jj==",jj];*)
  (*Print["-----"];*)
  (*Print["theta= ",theta];*)
  (*theta=Read[ftheta.Expression];*)
  (*prior=Read[fprior.Expression];*)

  theta=thetas[jj];
  prior=priori[jj];
  (*Print["j, " ",N[thetas[jj]]];*)
  a2=nt/(theta-xj);
  a1=-xj*a2;
  p=(1/a2)*cocientepaso;
  (*Do[p[[i]]=cocientepaso[[i]]/a2,{i,1,k}];*)
  pj=1-sumacocientepaso/a2;

  (* SI pj<0, ES IMPOSIBLE LUEGO pj=0 Y SOLO INTERVIENEN LOS
  p_A MUESTRALES *)

  Print[jj " . . .oj=",N[pj]];
  If[pj<0 (*1 If*)
    {(*Print["pj<0"];*)
     theta=thetas-qpbest;
     contador1=0;
     (*Print["r1= ",r1," r2=",r2];*)
     (* WHILE A *)
     While[Abs[p[rop]]>Tolr && contador1<20,
       (*Print["contador1=",contador1];*)
       incre=-f[rop]/fprima[rop];
       incre=Max[incre,
         contador1++];

```

```

ropaux=rop;
rop=rop+inca;
contador2=1;
While[{rop}>=2 || rop<r1] && contador2<4, (*WHILE 2*)
  {inca=inca+inca/contador2;
  rop=ropaux+inca;
  {*Print[" contador2=", contador2];*}
  contador2++;
  }; (* CIERRA WHILE 2 *)
If[Abs[f[rop]]<Abs[f[ropaux]],
  {rop=Re[rop];},
  {contador1=20;},
  ];
{*Print[" contador1=", contador1, " rop= ", rop,
  " f[rop]= ", f[rop];*}
]; (*cierra while 1 *);

(* BISECCION EN EL CASO DE QUE NEWTON NO CONVERJA *)
bandera=0;
(*Plot[f[r], {r,-1/xmax2,-1/xmin2}];*)
(*Print["imagenes", " ", r1, " ", f[r1];*)
(*Print["imagenes", " ", r2, " ", f[r2];*)
(*Input[];*)

If[Abs[f[rop]]>=Eps || rop>r2 || rop<r1,
  {bandera=1;
  r1aux=r1; r2aux=r2; paso=2;
  (* While Para quedarnos en el intervalo [r1,r2]*)
  While[{f[r1aux]}*f[r2aux]>0,
    {r1aux=r1/aux2+epsilon/(paso^5);
    r2aux=r2/aux2-epsilon/(paso^5);
    paso++;
    {*Print["paso^5=", paso^5, " ", f[r1aux], " ", f[r2aux]]*}
    };
  (*Print["Salida del While[f[r1aux]*f[r2aux]>0 \n
  imagenes, " ", f[r1aux], " ", f[r2aux];*}
  rop=(r1aux+r2aux)/2;
  i=0;
  While[Abs[f[rop]]>=Eps || i<31,
    {(*Acilador de la biseccion *)
    long=r1-r2aux;
    r1=r1+long;
    If[{f[r1]*f[r1aux+long/6]<0,
    {*Print["Acilador ", i; r2aux=r1aux+long/6},
    If[{f[r1aux+5*long/6]*f[r2aux]<0,
    {*Print["Acilador ", i; r1aux=r1aux+5*long/6}}];
    (* Fin del acilador *)
    rop=(r1+r2aux)/2;
    (*Print["rop=", rop, " f[rop], f[r1]=", f[rop], " ", f[r1aux];*}
    If[{f[rop]<Eps}, {r1aux=rop, r2aux=rop};
    (*Print[" ", r1aux, " ", r2aux];*}
    (*Print[" ", f[r1aux], " ", f[r2aux];*}
    i=i+1;
    };
  Print["Salida de biseccion con ", i,
  " iteraciones. rop=", rop, " f[rop]=", f[rop]];
  ]; (*CIERRA LA BISECCION *)

primera;
lam=mu/(1+Contar[rop]);
mu=lam*rop;

```

```

(*Print["Bandera", bandera];*)
(*Print["rop=" , rop = " , f[rop]= " , f[rop]];*)
(*If[pasota==1, incre=incretheta,
  (If[jj<=ptos1, incre=incre1];
   If[ptos1<jj<=ptos1+ptosmedia, incre=incre2];
   If[jj>ptos1+ptosmedia, incre=incre3];
  )
];
*)
paso=Re[priorr+accol[priorr]*incre];
If[x1<=paso==r2, rop=paso];
Do[{p[[i]]=(1)/(1+mu*x2[i])}, {i,1,k}];

theta=thetaoptest;

]]; (*Cierra 1 IF *)

pasolike=Re[Sum[p[[i]]*Log[p[[i]]], {i,1,k}]];

(*Print["Bandera", " like[" , jj, "=" , N[pasolike]];*)
like[jj]=N[pasolike,12];
(*Write["like.N[pasolike,10]];*)
sumpost=sumpost+N[prior*pasolike,10];
(*Write["post.N[prior*pasolike,10]];*)
ppostnor[jj]=N[prior*pasolike,10];

If[Mod[jj,10]==0, Print["]];

), {jj,10,ptos-10];

Do[{like[jj]=0, ppostnor[ptos-jj]=0.0;
  like[ptos-jj]=0.0, ppostnor[jj]=0.0}, {jj,1,9}];

(*Close["like"]*)
(*Close["theta"]*)
(*Close["paso"]*)
(*Close["post"]*)

(* NORMALIZACIÓN DE LA POSTERIORI solo para ARRAYS
Y CALCULO DE LA POSTERIORI ACUMULADA *)

(*sumpost=Sum[ppostnor[i], {i,1,ptos-1}];*)
sumpost=Array[ppostnor, {ptos-1}, 1, Plus]
Do[ppostnor[i]=ppostnor[i]/sumpost, {i,1,ptos-1}]
ppostacum[i]=ppostnor[i];
Do[ppostacum[i]=ppostacum[i-1]+ppostnor[i], {i,2,ptos-1}]
Sum[ppostnor[i], {i,1,ptos-1}]

Do[Print["theta[" , jj, "=" , like[jj]], {jj, 1, 1000, 100}]

```

```
(* GRAFICA DE LA VEROSIMILITUD *)

(* SUMO LAS VEROSIMILITUDES PARA REESCALARLAS Y
REPRESENTARLAS CON LA PRIORI *)

sumlike=Sum[like[i], {i,1,ptos-1}];
numptos=ptos-1;
tablalike:=Table[{thetas[i],N[like[i]/sumlike]}, {i,1,numptos}];
glike=ListPlot[tablalike,PlotJoined->True,
PlotStyle->{RGBColor[.2,0.2,0.8]},PlotRange->All]
tablalikeparcial:=Table[{thetas[i],N[like[i]/sumlike]}, {i,50,85}];
glikeparcial:=ListPlot[tablalikeparcial,
PlotStyle->{RGBColor[.2,0.2,0.8]},PlotJoined->True,PlotRange->All]

(*****)

(* CALCULO DE LOS PARAMETROS, MEDIA Y DESVIACION
TIPICA A POSTERIORI *)

prob[1] = .01;
prob[2] = .025;
prob[22] = .975;
prob[23] = .99;
prob[24] = 1.0;
Do[prob[i] = (1 - 2^i * .05) / (5, 3, 21)];
tol3 = .00001;
sumpr = 0;
i = 1;
While[sumpr < .025,
{sumpr = sumpr + priori[i],
i = i + 1;
}];
q025priori = N[10000 * (thetas[i] + thetas[i - 1]) / 2, 7];
While[sumpr < .975,
{sumpr = sumpr + priori[i],
i = i + 1;
}];
q975priori = N[10000 * (thetas[i] + thetas[i - 1]) / 2, 7];

i = 1;
j = 1;
Do[{While[ppostacua[j] < prob[i], j = j + 1];
If[j == 2 && ppostacua[1] > prob[i],
{m = ppostacua[1] / (thetas[1] - a);
If[Abs[m] < tol3, percentil[i] = thetas[1],
percentil[i] = thetas[1] +
(prob[i] - ppostacua[1]) / m;
}];
}];
(* Print["i=", i, " / n=", n, " prob[i]=", prob[i]];*)
```

```

(*Print[" m=",m," ppostacum[1]=",ppostacum[1],
      " percentil[i]=",percentil[i]];*)
}
];
IF[j > 1,
{
m = (ppostacum[j] - ppostacum[j - 1]) /
  (thetas[j] - thetas[j - 1]);
(* SI m ES MUY PEQUEÑO PUEDE DAR PROBLEMAS
  AL DIVIDIR POR 0 (m=0) *)
IF[Abs[m] < 10E-3, {percentil[i] = thetas[j]; j = j - 1},
  percentil[i] = thetas[j] + ((prob[i] - ppostacum[j]) / m)];
percentil[i] = N[percentil[i], 7];
}
]
, {i, 1, 23}]

media = 0;
Do[media = media + thetas[i] * ppostnor[i], {i, 1, ptos - 1}];
varianza = 0;
Do[varianza = varianza + ((thetas[i] - media) ^ 2) * ppostnor[i],
  {i, 1, ptos - 1}];
desvtip = Sqrt[varianza];
mediapost = media;
sigma = desvtip;
mediapost
Do[IF[ppostnor[j - 1] < ppostnor[j],
  {modapost = N[thetas[j]]},
  (*Print[" ", modapost]*)
]]
, {j, 2, ptos - 1}]

Print["Media a posteriori:", " ", media];
Print["Desviación típica:", " ", desvtip];
Do[Print[prob[i], " ", percentil[i]];
  , {i, 1, 23}]
Print["Media muestral:", " ", N[ptest]];
Print["Media a posteriori:", " ", media];
Print["Moda a posteriori:", " ", modapost];

(****Cálculo cociente de verosimilitudes****)

ptos

```

```

maxverosim = like[1];
Do[IF[maxverosim < like[i],
  {maxverosim = like[i];
   imaxverosim = i}
 , {i, 2, ptos - 1}];
Print["El maximo de la verosimilitud se alcanza en:"];
Print["theta[" , imaxverosim, "]" = ", thetas[imaxverosim],
  "; like[" , imaxverosim, "]" = ", like[imaxverosim]];
Print["-----"];
divisor = Input["Valor para el cociente de la verosimilitud="];
maxver = like[imaxverosim];

i = 1;
cociente = 50000;
While[i < ptos && cociente > divisor,
  {i++;
   If [like[i] != 0, cociente = maxver / like[i]];
   (*Print[i, " cociente=", cociente];*)
  }];
extinf = i;
While[i < ptos && cociente < divisor,
  {i++;
   If [like[i] != 0, cociente = maxver / like[i]];
   (*Print[i, " cociente=", cociente];*)
  }];
i--;
extsup = i;
Print["Constante para el cociente de verosimilitudes=", divisor];
Print["Intervalo=(punto=", extinf,
  ", punto ", extsup, ")={", thetas[extinf], ", ", thetas[extsup], "}"];

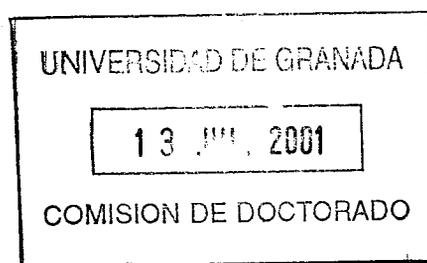
Do[Print[jj, " cociente= ", maxver / like[jj]], {jj, extinf - 1, extsup + 1}]

```

ANEXO 3

Relación de algunas de las abcisas y el valor de la verosimilitud empírica correspondiente para la variable “cociente entre gastos e ingresos”.

Obtenida la variable cociente entre gastos e ingresos de cada hogar, se calcula su verosimilitud empírica para la media. Estas páginas recogen algunos de los valores que toma esa media, así como el valor de la verosimilitud empírica correspondiente a cada punto. Los valores de la verosimilitud empírica están reescalados para que puedan representarse gráficamente, dado que lo que interesa realmente es el máximo de dicha función. Se ha calculado la verosimilitud para 1.000 puntos pero sólo reflejamos unos cuantos, suficiente para poner de manifiesto lo que se señaló en el capítulo 4, esto es, que el cociente gastos-ingresos toma valores menores y mayores que la unidad, lo cual implica hogares con ingresos declarados inferiores a los gastos.



0.0710898 like[1]=0.
0.0921796 like[2]=0.
0.113269 like[3]=0.
0.134359 like[4]= $8.0823462120 \times 10^{4595}$
0.155449 like[5]= $3.7573829685 \times 10^{4897}$
0.176539 like[6]= $1.1275809513 \times 10^{5144}$
0.197629 like[7]= $2.7738577995 \times 10^{5352}$
0.218718 like[8]= $9.0594276107 \times 10^{5532}$
0.239808 like[9]= $1.5116767292 \times 10^{5692}$
0.260898 like[10]= $4.0412341133 \times 10^{5834}$
0.281988 like[11]= $2.7881594216 \times 10^{5963}$
0.303078 like[12]= $1.1583010552 \times 10^{6081}$
0.324167 like[13]= $1.8207759215 \times 10^{6189}$
0.345257 like[14]= $2.7064304003 \times 10^{6289}$
0.366347 like[15]= $4.8784852491 \times 10^{6382}$
0.387437 like[16]= $8.3396559231 \times 10^{6469}$
0.408527 like[17]= $7.2704616226 \times 10^{6551}$
0.429616 like[18]= $1.3022090306 \times 10^{6629}$
0.450706 like[19]= $1.5395228126 \times 10^{6702}$
0.471796 like[20]= $3.2245076392 \times 10^{6771}$
0.492886 like[21]= $2.7787326662 \times 10^{6837}$
0.513976 like[22]= $2.0338195532 \times 10^{6900}$
0.535065 like[23]= $2.3690061992 \times 10^{6960}$
0.556155 like[24]= $7.5925932045 \times 10^{7017}$
0.577245 like[25]= $1.08223441030 \times 10^{7073}$
0.598335 like[26]= $1.04762282885 \times 10^{7126}$
0.619425 like[27]= $1.00200513691 \times 10^{7177}$
0.640514 like[28]= $1.3216265804 \times 10^{7226}$
0.661604 like[29]= $3.2365989919 \times 10^{7273}$
0.682694 like[30]= $1.9198954487 \times 10^{7319}$
0.703784 like[31]= $3.4987037008 \times 10^{7363}$
0.724874 like[32]= $2.4205917858 \times 10^{7406}$
0.745963 like[33]= $7.6585908372 \times 10^{7447}$
0.767053 like[34]= $1.2968410771 \times 10^{7488}$

0.788143 like[35]=1.3186092526×10⁷⁵²⁷
0.809233 like[36]=8.2381730120×10⁷⁵⁶⁴
0.830323 like[37]=2.3626259663×10⁷⁶⁰¹
0.851412 like[38]=7.0910751527×10⁷⁶³⁵
0.872502 like[39]=1.3008153423×10⁷⁶⁶⁷
0.893592 like[40]=2.8977348900×10⁷⁶⁹⁴
0.914682 like[41]=5.9053006440×10⁷⁷¹⁷
0.935772 like[42]=1.4119870314×10⁷⁷³⁷
0.956861 like[43]=6.5434116308×10⁷⁷⁵²
0.977951 like[44]=1.12653832712×10⁷⁷⁶⁵
0.999041 like[45]=1.5211177553×10⁷⁷⁷⁴
1.02013 like[46]=3.6353564508×10⁷⁷⁸⁰
1.04122 like[47]=3.6546023055×10⁷⁷⁸⁴
1.06231 like[48]=3.8716086247×10⁷⁷⁸⁶
1.0834 like[49]=1.17372931380×10⁷⁷⁸⁷
1.10449 like[50]=3.1764226754×10⁷⁷⁸⁶
1.12558 like[51]=2.4747225683×10⁷⁷⁸⁵
1.14667 like[52]=1.17599853258×10⁷⁷⁸⁴
1.16776 like[53]=4.5615685640×10⁷⁷⁸²
1.18885 like[54]=1.6993324318×10⁷⁷⁸¹
1.20994 like[55]=6.3085752628×10⁷⁷⁷⁹
1.23103 like[56]=2.3338333150×10⁷⁷⁷⁸
1.25212 like[57]=8.6038092805×10⁷⁷⁷⁶
1.27321 like[58]=3.1607553948×10⁷⁷⁷⁵
1.2943 like[59]=1.15708947519×10⁷⁷⁷⁴
1.31539 like[60]=4.2210033835×10⁷⁷⁷²
1.33648 like[61]=1.5343834570×10⁷⁷⁷¹
1.35757 like[62]=5.5579969387×10⁷⁷⁶⁹
1.37866 like[63]=2.0061602237×10⁷⁷⁶⁸
1.39975 like[64]=7.2156018505×10⁷⁷⁶⁶
1.42084 like[65]=2.5860436306×10⁷⁷⁶⁵
1.44193 like[66]=9.2353247459×10⁷⁷⁶³
1.46302 like[67]=3.2863833181×10⁷⁷⁶²
1.48411 like[68]=1.16528098880×10⁷⁷⁶¹

1.5052 like[69]=4.1170502404×10⁷⁷⁵⁹
1.52629 like[70]=1.4493770752×10⁷⁷⁵⁸
1.54738 like[71]=5.0840869220×10⁷⁷⁵⁶
1.56847 like[72]=1.7769595278×10⁷⁷⁵⁵
1.58956 like[73]=6.1883050910×10⁷⁷⁵³
1.61065 like[74]=2.1472970050×10⁷⁷⁵²
1.63173 like[75]=7.4239551609×10⁷⁷⁵⁰
1.65282 like[76]=2.5573959333×10⁷⁷⁴⁹
1.67391 like[77]=8.7776161605×10⁷⁷⁴⁷
1.695 like[78]=3.0017032249×10⁷⁷⁴⁶
1.71609 like[79]=1.02274655146×10⁷⁷⁴⁵
1.73718 like[80]=3.4719538588×10⁷⁷⁴³
1.75827 like[81]=1.17430807476×10⁷⁷⁴²
1.77936 like[82]=3.9572077549×10⁷⁷⁴⁰
1.80045 like[83]=1.3285896573×10⁷⁷³⁹
1.82154 like[84]=4.4441078769×10⁷⁷³⁷
1.84263 like[85]=1.4810390643×10⁷⁷³⁶
1.86372 like[86]=4.9173724755×10⁷⁷³⁴
1.88481 like[87]=1.6266002422×10⁷⁷³³
1.9059 like[88]=5.3605104320×10⁷⁷³¹
1.92699 like[89]=1.7599708904×10⁷⁷³⁰
1.94808 like[90]=5.7567222841×10⁷⁷²⁸
1.96917 like[91]=1.8759097646×10⁷⁷²⁷
1.99026 like[92]=6.0899241775×10⁷⁷²⁵
2.01135 like[93]=1.9695703419×10⁷⁷²⁴
2.03244 like[94]=6.3458113395×10⁷⁷²²
2.05353 like[95]=2.0368321239×10⁷⁷²¹
2.07462 like[96]=6.5128652245×10⁷⁷¹⁹
2.09571 like[97]=2.0745985359×10⁷⁷¹⁸
2.1168 like[98]=6.5832067774×10⁷⁷¹⁶
2.13789 like[99]=2.0810317183×10⁷⁷¹⁵
2.15898 like[100]=6.5532084847×10⁷⁷¹³