DISTRIBUCIÓN DE LA RENTA EN LA COMUNIDAD DE CASTILLA Y LEÓN: DOS MÉTODOS DE ESTIMACIÓN.

Rafael HERRERÍAS PLEGUEZUELO
Federico PALACIOS GONZÁLEZ
José CALLEJÓN CÉSPEDES

Facultad de CC. Económicas y Empresariales
Dpto. de Economía Aplicada
Campus de Cartuja s/n
18071 - Granada

1 INTRODUCCIÓN

Son numerosos los estudios que sobre la modelización de la distribución de la renta se han realizado y se siguen realizando. La información que esta modelización proporciona sobre la riqueza o pobreza de un país, región, provincia etc., es necesaria a la hora de planificar las distintas actuaciones tanto impositivas como redistributivas.

Camilo Dagum (1981) clasifica la práctica totalidad de los modelos hasta entonces existentes en tres sistemas básicos: sistema de D'Addario, sistema de Pearson y el sistema propuesto por el propio autor. En época más reciente se han realizado otras aportaciones entre las que destacan los estudios de McDonald (1984) y McDonald y Xu (1992), donde se propone la utilización de distribuciones beta tetraparamétricas y pentaparamétricas, respectivamente.

Herrerías, Palacios y Ramos (1996), utilizan una metodología que permite aprovechar la flexibilidad de los métodos no paramétricos y disponer de una forma funcional que modelice la distribución observada. Se obtiene la función de densidad a partir de una función generadora de probabilidad. A partir de ella, mediante integración numérica se obtiene la distribución correspondiente.

Recientemente, Herrerías, Palacios y Callejón (1996), utilizando los datos muestrales que proporciona la EPF (1990-1991), han realizado una modelización de la distribución de la renta en la provincia de Valladolid. El mejor resultado obtenido corresponde a la distribución generada por una función polinómica de grado tres.

En este estudio, tomando como valores empíricos de renta, para la Comunidad Autónoma de Castilla y León, los recogidos en la EPF (1990 - 1991) del INE, se ajusta un modelo probabilístico para tal distribución, empleando dos procedimientos diferentes aunque relacionados.

El primero de los métodos utilizados supone que la función de densidad de la distribución verifica la ecuación diferencial (sistema de Pearson):

$$\frac{f'(x)}{f(x)} = \frac{x - a}{b_0 + b_1 x + b_2 x^2} \tag{1}$$

a partir de la cual se obtiene un sistema de cuatro ecuaciones haciendo m=0, 1,2, 3, en:

$$\alpha_m a - m\,\alpha_{m-1} b_0 - (m+1)\,\alpha_m b_1 - (m+2)\,\alpha_{m+1} b_2 = \alpha_{m+1} \tag{2}$$

donde α_i representa al momento de la distribución, con respecto al origen, de orden i. (Elderton W.P. y Johnson N.L. (1969), Johnson N.L. y Kotz S (1970)). Sustituyendo los momentos poblacionales por los correspondientes muestrales se obtiene que la función de densidad corresponde a una distribución beta.

El segundo método utilizado para la modelización de la distribución supone que la función generadora de probabilidad, es decir el cociente $g(x) = \frac{f'(x)}{f(x)}$, es un polinomio de **grado impar** *r*:

$g(x) = \sum_{i=0}^{r} a_i x^i$. La estimación de los parámetros la realizamos también a través del método de los momentos, sin más que resolver el sistema de *r+1* ecuaciones, que se obtiene al hacer $m = 0,1,2,...r$ en la expresión:

$$\sum_{i=0}^{r} a_i \alpha_{i+m} + m\alpha_{m-1} = 0 \tag{3}$$

Este segundo método posee dos propiedades que lo distinguen y lo hacen preferible al primero:

a) Los estimadores obtenidos son de máxima verosimilitud, puesto que cuando la función generadora de probabilidad es un polinomio, los métodos de máxima verosimilitud y de los momentos coinciden (Kendall and Stuart (1979), Callejón y Santos (1994)).

b) Se realizan ajustes para distintos grados de la función polinómica, permitiendo así una cierta flexibilidad. En este trabajo se han obtenido las funciones de densidad para los casos en que g(x) es un polinomio de grado uno, de grado tres, de grado cinco y de grado siete; la elección entre las distintas posibilidades se realiza utilizando los conocidos métodos de bondad de ajuste. Como prueba de esta flexibilidad, se puede observar que el grado de la función generadora, a partir de la que obtenemos la distribución que mejor se ajusta a los datos muestrales, varía de una provincia a otra.

La metodología utilizada en este trabajo coincide con el realizado por los mismos autores [Herrerías-Palacios-Callejón (1996)], para la provincia de Valladolid, si bien en sus conclusiones podemos observar una diferencia en el grado de la función polinómica, generadora de la distribución. Este resultado hace que se incorpore al trabajo la modelización de la renta para cada una de las provincias que componen esta Comunidad Autónoma.

Las primeras modelizaciones corresponden a la distribución de la renta en el conjunto de la Comunidad Autónoma de Castilla y León. Posteriormente se realizan las modelizaciones en las distintas provincias que la constituyen.

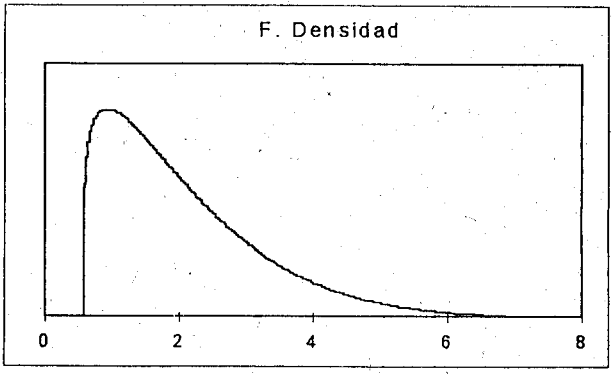
2 MODELIZACIÓN DE LA DISTRIBUCIÓN DE LA RENTA EN LA COMUNIDAD AUTÓNOMA DE CASTILLA Y LEÓN.

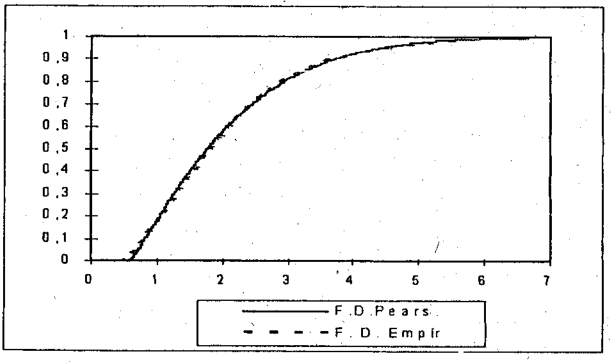
2.1 Obtención de una distribución perteneciente al sistema de Pearson.

A partir de los datos muestrales buscamos la distribución cuya función de densidad verifique la ecuación diferencial (1). Para ello calculamos los cuatro primeros momentos con respecto al origen. Sustituimos estos valores en el sistema de ecuaciones (2) y resolvemos. La distribución estimada resulta ser una beta de recorrido comprendido entre 0,573368 y 11,002447 millones de pesetas y cuyos parámetros **p** y **q** tienen los valores 1,2287 y 7,3953 respectivamente:

$$f(x) = K(x - 0{,}5733)^{0{,}2287}(11{,}0024 - x)^{6{,}3953} \qquad 0{,}5733 < x < 11{,}0024$$

Podemos tener una idea de la bondad de la estimación realizada por este primer método, comparando las gráficas de las funciones de distribución empírica y estimada. La siguiente gráfica muestra las mencionadas distribuciones.





Aunque están muy próximas las funciones de distribución empírica y estimada, sin embargo, tenemos el inconveniente de que estas distribución beta obtenida no admite rentas inferiores a 573368 pesetas, ni superiores a 11002447 pesetas, hecho que contradice los datos muestrales.

2.2 Obtención de una distribución cuya función generadora de probabilidad es polinómica.

Se entiende por función generadora de probabilidad a la derivada del logaritmo de la función de densidad, $g(x) = \dfrac{f'(x)}{f(x)}$ para los puntos en que la derivada existe y la densidad no es cero (Callejón (1995). En este trabajo se han estudiado los casos en que *g(x)* es un polinomio. Dentro de esta familia de funciones polinómicas se eligen sólo las de grado impar porque son las que dan lugar a una función de densidad que delimita con el eje de abcisas un recinto de área finita.

La relación recurrente, establecida en (3), de momentos respecto al origen de la distribución, (puede verse con mayor detalle en Herrerías-Palacios-Ramos (1996)), permite utilizar tantas ecuaciones como sean necesarias para obtener los coeficientes del polinomio.

a) Distribución normal: función polinómica de primer grado.

La primera estimación que realizamos por este método corresponde al caso en que la función generadora es un polinomio de grado 1, $g(x) = a_0 + a_1 x$. A partir de los momentos muestrales y sustituyendo en el sistema (3), para $m = 0$ y 1 se obtiene la función generadora $g(x) = 1{,}47172 - 0{,}7133x$ que corresponde a la distribución normal de media 2,063255 y de varianza 1,401934. (Véase Callejón (1995))

No parece ser una distribución aceptable porque, en principio, la experiencia dice que las distribuciones de la renta no son simétricas, lo que se corrobora al estudiar la adherencia de la distribución.

b) Función generadora polinómica de tercer grado.

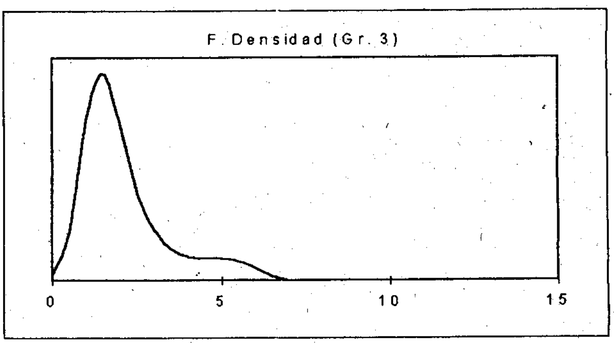
A partir de la función $g(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3$ y sustituyendo en el sistema (3) los momentos poblacionales por los muestrales, la solución es

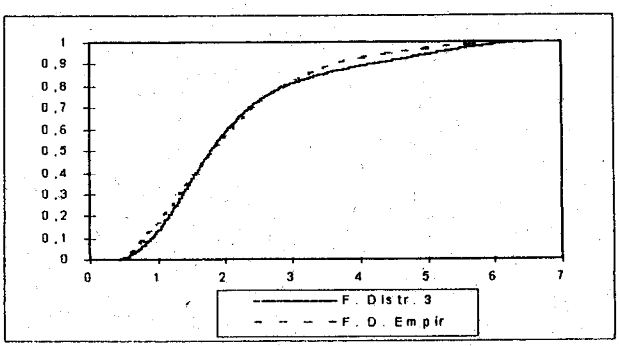
$$g(x) = 6{,}92099 - 7{,}7437x + 2{,}3290x^2 - 0{,}2116x^3$$

$f(x) = K\, e^{\int g(x)dx}$ es la densidad generada por la función polinómica, donde la constante de integración se calcula, como en cualquier densidad, bien analíticamente o por métodos numéricos, como en este caso se ha hecho, resultando así que la función de densidad de la distribución es

$$f(x) = \frac{1}{112{,}565954}\, e^{6{,}92099x - 3{,}87185x^2 + 0{,}77634x^3 - 0{,}0529x^4}$$

En el primero de los siguientes gráficos se muestra la función de densidad estimada. El segundo gráfico muestra, simultáneamente, la función de distribución empírica y la función de distribución estimada a partir de una función generadora polinómica de grado tres.



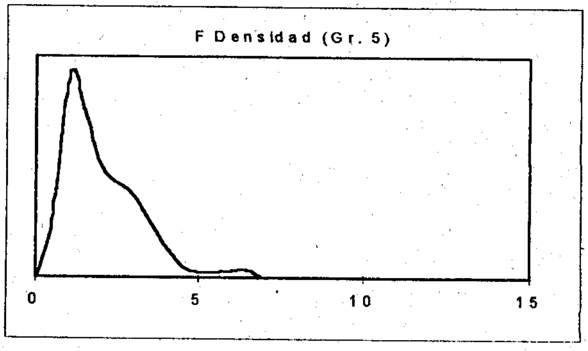


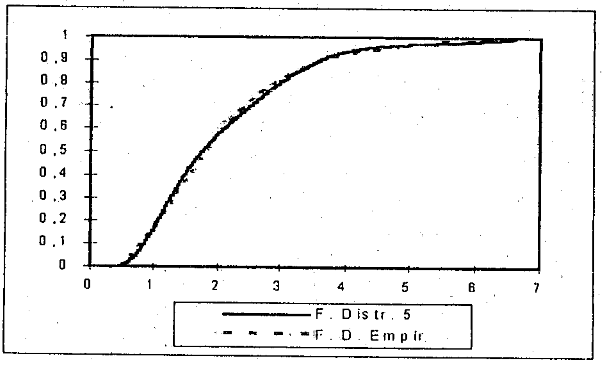
c) Función generadora polinómica de grado cinco.

El resultado obtenido de la función generadora y la función de densidad, así como las gráficas de esta última y de las funciones de distribución, empírica y estimada, han sido:

$$g(x) = 21{,}0202 - 42{,}41x + 30{,}3678x^2 - 9{,}9271x^3 + 1{,}48967x^4 - 0{,}083x^5$$

$$f(x) = \frac{1}{4.638{,}85475} e^{21{,}0202x - 21{,}2005x^2 + 10{,}1226x^3 - 2{,}4817x^4 + 0{,}2979x^5 - 0{,}0138x^6}$$



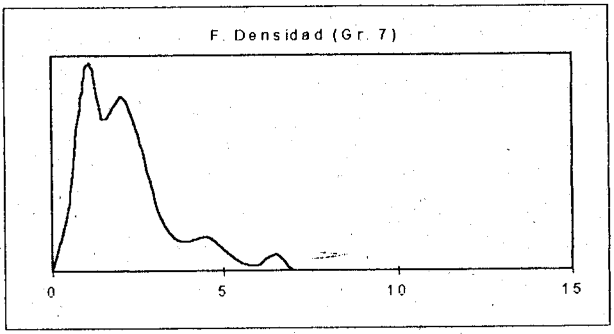


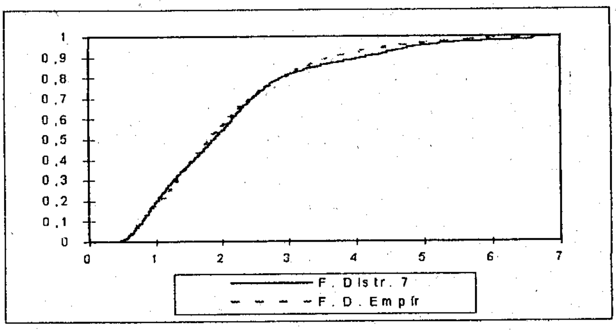
d) Función generadora polinómica de grado siete.

Los resultados obtenidos para la modelización a partir de una función polinómica de grado siete han sido:

$$g(x) = 72{,}4894 - 227{,}9x + 276{,}699x^2 - 169{,}74x^3 + 57{,}3278x^4 - 10{,}778x^5 + 1{,}05506x^6 - 0{,}0419x^7$$

$$f(x) = \frac{1}{178.216.009} e^{72{,}4894x - 113{,}95x^2 + 92{,}23x^3 - 4243x^4 + 11{,}4655x^5 - 1{,}7963x^6 + 0{,}1507x^7 - 0{,}00523x^8}$$





Puede observarse la proximidad entre las gráficas de la función de distribución empírica y la estimada. De entre las cinco modelizaciones que se presentan, la perteneciente al sistema de Pearson y las obtenidas a partir de funciones generadoras polinómicas de grados uno, tres, cinco y siete, esta última es el caso cuya suma de cuadrados de distancias entre función de distribución empírica y teórica, medidas sobre las ordenadas de cada uno de los valores de la muestra, es menor.

3 MODELIZACIÓN DE LA DISTRIBUCIÓN DE LA RENTA EN CADA UNA DE LAS PROVINCIAS DE LA CASTILLA Y LEÓN.

Para cada una de las provincias que componen esta Comunidad Autónoma se han obtenido, utilizando el método de los momentos, las modelizaciones de la distribución de la renta, a partir de las funciones generadoras polinómicas de grados uno, tres, cinco y siete. En todos los casos las funciones de distribución se han calculado mediante integración numérica sobre cada uno de los valores muestrales.

El procedimiento que nos permite elegir la distribución, que mejor modelización proporciona para cada una de las provincias, compara la función de distribución empírica de la provincia en estudio con cada una de las funciones de distribución obtenidas (a partir de un polinomio de grado uno, tres, cinco y siete) para dicha provincia.

Una primera idea sobre la bondad de los ajustes podemos obtenerla mediante una sencilla inspección visual de las gráficas. En una segunda etapa, para una selección de mayor rigor, se ha calculado el máximo del valor absoluto de la diferencia entre las funciones de distribución empírica y estimada (Kolmogorov-Smirnov), dentro de la misma provincia, para cada uno de los grados. El tercer criterio consiste en calcular la suma de los cuadrados de las distancias, sobre la ordenada, entre la función de distribución empírica y la función de distribución estimada para cada uno de los grados.

Para cada una de las provincias en estudio los tres criterios han coincidido en el grado de la distribución elegida, sin embargo, el grado del polinomio elegido en cada provincia no ha sido siempre el mismo: siete para las provincias de Avila, Burgos, Palencia, Salamanca, Soria y Zamora, cinco para las provincias de León y Segovia y tres para la provincia de Valladolid (Herrerías-Palacios-Callejón 1996).

El hecho de que una provincias se modelicen mejor a partir de un grado que de otro, no hace más que demostrar la flexibilidad del método empleado.

En la siguiente tabla se recogen los coeficientes de la función generadora polinómica, $g(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + a_5 x^5 + a_6 x^6 + a_7 x^7$, elegida como preferible a las demás, entre todas las de la misma provincia, así como la constante, K, de normalización de la función de densidad correspondiente $f(x) = K \; e^{\int g(x)dx}$.

Provincia	a_0	a_1	a_2	a_3	a_4	a_5	a_6	a_7	K^{-1}
Avila	128,68	-509,8	785,75	-618,11	270,49	-66,44	8,55	-0,447	2,287 E 11
Burgos	38,63	-101,83	104,11	-53,46	14,93	-2,29	0,18	-0,0057	315063,46
León	15,58	-26,09	16,16	-4,75	0,66	-0,03	*	*	2831,62
Palencia	62,57	-192,84	240,08	-157,04	58,46	-12,45	1,41	-0,0657	43618718,1
Salamanca	142,33	-529,58	771,28	-576,3	240,54	-56,56	6,99	-0,3528	2,528 E 134
Segovia	21,906	-43,45	31,05	-10,225	1,5533	-0,0879	*	*	8418,3741
Soria	175,78	-566,79	718,26	-465,86	170,63	-35,14	3,815	-0,1695	1,99257 E 19
Valladolid	6,31087	-6,4102	1,7788	-0,1515	*	*	*	*	129,923
Zamora	66,57	-222,03	278,95	-178,06	62,72	-12,29	1,253	-0,0516	33716849,39
Castilla y León	72,48	-227,9	276,69	-169,74	57,32	-10,77	1,0550	-0,0419	178216008,7

El anexo I muestra las gráficas de las funciones de densidad obtenidas a partir de las funciones generadoras correspondientes a la tabla anterior.

El anexo II muestra las gráficas de las funciones de distribuición obtenidas, por integración numérica, a partir de cada una de las funciones de densidad anteriores. En cada caso, en el mismo gráfico se representa también la función de distribución empírica de cada provincia

4 MEDIDA DE BONDAD DE AJUSTE.

Tal y como hemos expuesto en el apartado anterior, al objeto de comparar las estimaciones obtenidas, hemos realizado el contraste de Kolmogorov-Smirnov para cada provincia. Los resultados se muestran en la tabla siguiente:

	Distribuciones polinómicas: Estadístico de Kolmogorov				Valores críticos
	Grado 1	Grado 3	Grado 5	Grado 7	$\alpha = 0,01$
Avila	0.1360	0,0801	0,0607	0,0555	0,0841
Burgos	0,1731	0.0476	0,0640	0,0425	0,0834
León	0,1468	0,0314	0,0284	0,0327	0,0707
Palencia	0,1319	0,0673	0,0291	0,0280	0,0822
Salamanca	0,1799	0,0831	0,0687	0,0336	0,0871
Segovia	0,1496	0,0481	0,0271	0,0387	0,0827
Soria	0,1377	0,0592	0,0652	0,0382	0,0876
Valladolid	0,1700	0,0535	0,0556	0,0596	0,0792
Zamora	0,1831	0,0673	0,0462	0,0439	0,0837
Castilla y León	0,1389	0,0489	0,0372	0,0328	0,0272

5 CONCLUSIONES.

La estimación de una distribución perteneciente al sistema de Pearson resultó ser una beta, y por tanto la variable ha de estar comprendida entre dos valores. La renta anual, estimada por este sistema para toda la Comunidad, debe estar comprendida entre 573.368 pesetas y 11.002.447 pesetas. Las distribuciones obtenidas a partir de funciones generadoras polinómicas no tienen estas limitaciones.

La estimación a partir de una función generadora polinónica parece tener, además, otras ventajas: son estimaciones máximo verosímiles y permite una cierta flexibilidad puesto que elegimos entre los distintos grados de los polinomios.

A partir de la distribución polinómica seleccionada se puede obtener la curva de Lorenz, (Casas J.M., Herrerías R. y Núñez J. (1990)), el índice de concentración de Gini, o cualquier otra medida de concentración estimada (Casas J.M. y Núñez J. (1991)). Estos resultados pueden contrastarse con los obtenidos a partir de los datos muestrales.

El gráfico correspondiente al conjunto de la población de la Comunidad de Castilla y León refleja en su forma la mezcla de todas las heterogeneidades observadas entre los distintos grupos de provincias, incluso hace pensar que sería conveniente la utilización de una función generadora de grado superior.

REFERENCIAS BIBLIOGRÁFICAS.

CALLEJÓN, J. Y SANTOS, M. (1994). Comparación de dos métodos de estimación: Estimación máximo verosímil y método de los momentos. VIII Reunión Anual de ASEPELT-España, volumen I. pág. 245 - 249. Universitat de las Illes Balears.

CALLEJÓN, J. (1995). Un nuevo método para generar distribuciones de probabilidad. Problemas asociados y Aplicaciones. Tesis Doctoral. Universidad de Granada. ETD Micropublicaciones S.L.

CASAS J.M., HERRERÍAS R. Y NÚÑEZ J. (1990). Familias de formas funcionales para estimar la curva de Lorenz. IV Reunión Anual ASEPELT-ESPAÑA. Murcia.

CASAS J.M. Y NÚÑEZ J. (1991). Sobre la medición de la desigualdad y conceptos afines. Actas de la V Reunión Anual ASEPELT-ESPAÑA. Las Palmas.

DAGUM, C. (1981). Sistemas Generadores de distribución del Ingreso y la Ley de Pareto. Journal of the Inter-American Statistical Institute 125, pág. 143 - 183.

ELDERTON, W.P. Y JOHNSON, N.L. (1969). System of frequency curves. Cambridge University Press.

HERRERÍAS R., PALACIOS F. Y RAMOS, A. (1996). Una metodología flexible para la modelización de la distribución de la Renta. Ponencia presentada a la X Reunión Anual de la Asociación de Economía Aplicada ASPELT-ESPAÑA. Albacete. Pendiente de publicación.

HERRERÍAS R., PALACIOS F. Y CALLEJÓN J. (1996). Distribución de la renta en la provincia de Valladolid: dos métodos de estimación. Ponencia presentada en el "Congreso Valladolid: hoy y mañana". Valladolid. Pendiente de publicación.

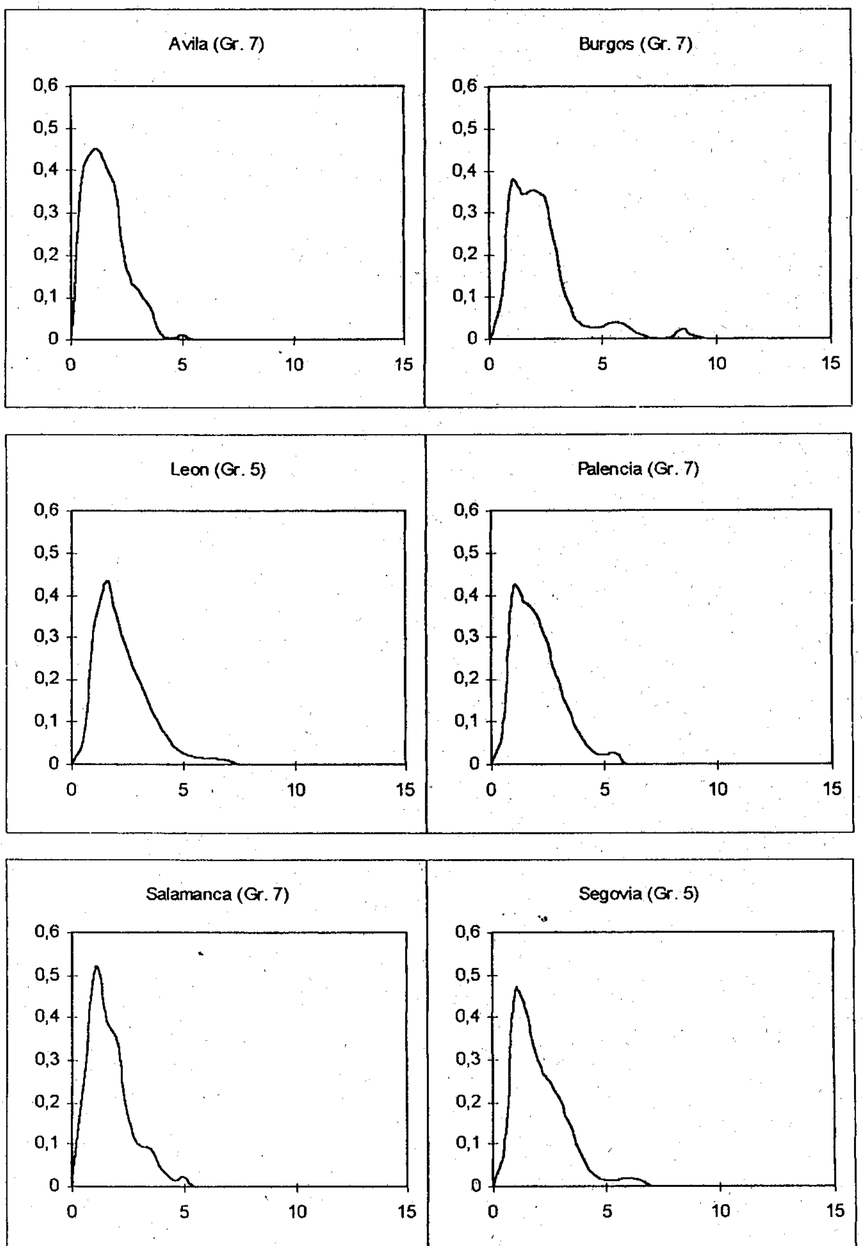
JOHNSON, N.L. Y KOTZ, S (1970). Distributios in Statistic: Continuous univariate distributions-1. John Wiley & Sons, New York.

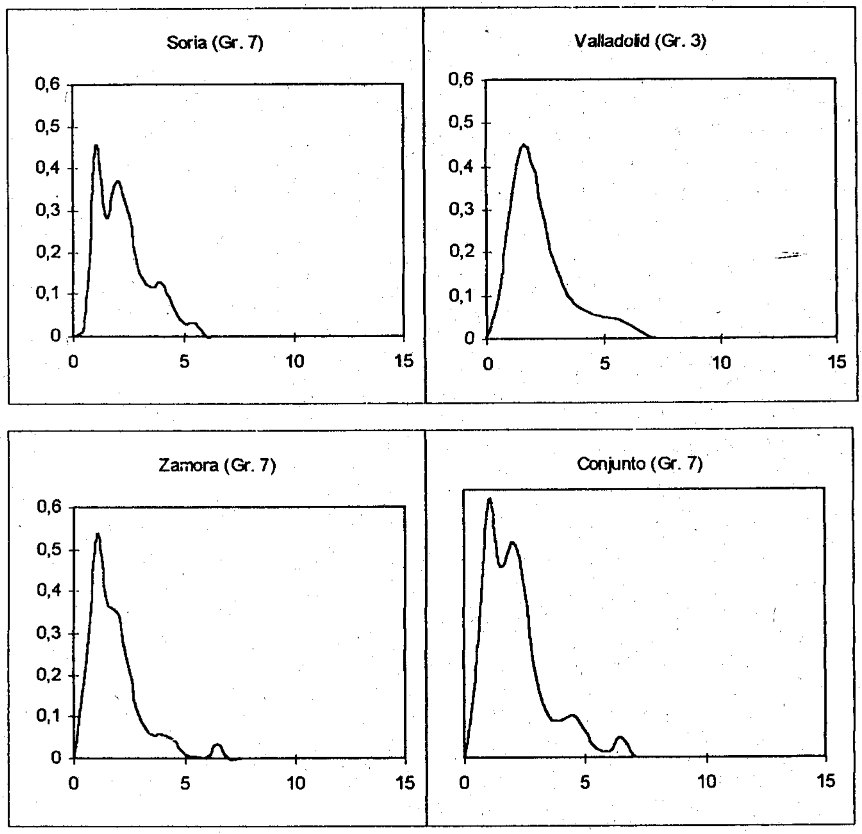
KENDALL, S.M. AND STUART, A. (1979). The advanced theory of Statistics. Griffin. London.

McDONALD, J.B. (1984). Some Generalized Functions for the Size Distribution of Income. Econometrica 52, pág. 663-674

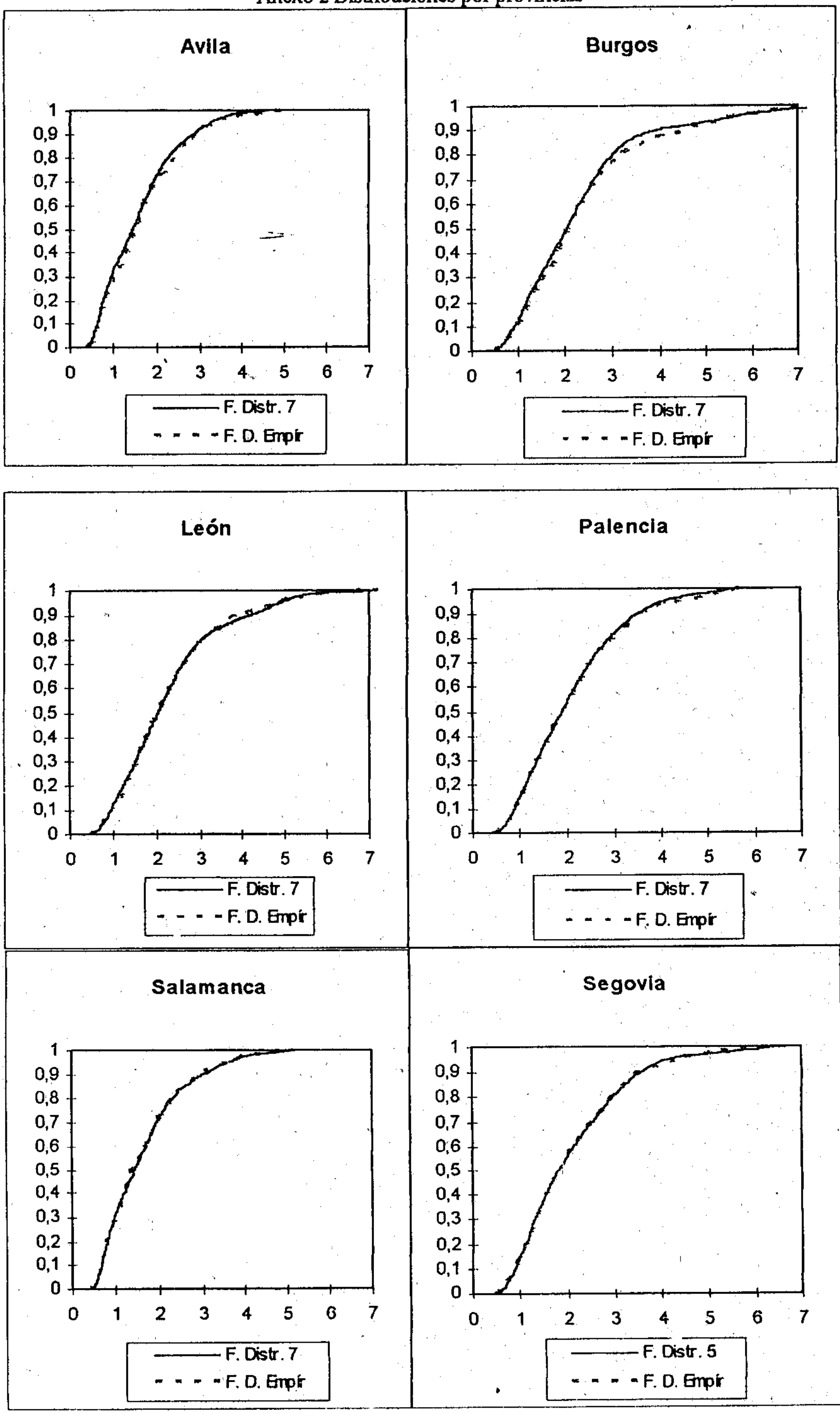
McDONALD, J.B. y XU Y.J. (1992). A generalization of the beta of the First and Second Kind with an Application. American Statistical Association: Proceedings of the Business and Economic Statistics Section, pág. 155-160

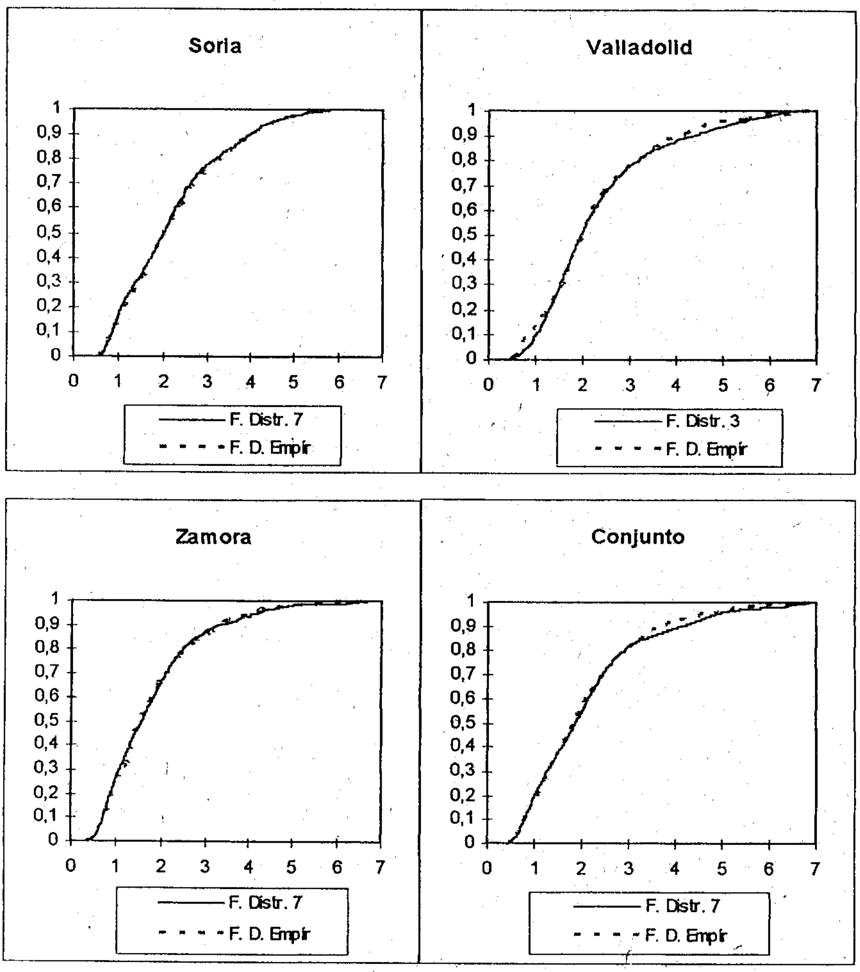
Anexo 1. Funciones de densidad por provincias




Soria (Gr. 7)
0,6
0,5
0,4
0,3
0,2
0,1
0
0 5 10 15
Valladolid (Gr. 3)
0,6
0,5
0,4
0,3
0,2
0,1
0
0 5 10 15
Zamora (Gr. 7)
0,6
0,5
0,4
0,3
0,2
0,1
0
0 5 10 15
Conjunto (Gr. 7)
0 5 10 15

Anexo 2 Distribuciones por provincias




Soria
1
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
0 1 2 3 4 5 6 7
F. Distr. 7
F. D. Empír
Valladolid
F. Distr. 3
F. D. Empír
Zamora
F. Distr. 7
F. D. Empír
Conjunto
F. Distr. 7
F. D. Empír