

REVISTA DE ESTUDIOS REGIONALES

I.S.S.N.: 0213-7585

2ª EPOCA Septiembre-Diciembre 2013



98

SUMARIO

Artículos

Leopoldo José Cabrera Rodríguez. Desigualdad social, rendimiento y logros educativos en España (1990-2012): Los desequilibrios regionales aumentan

Aitziber Etxezarreta Etxarri, Gala Cano Fuentes, Joris Hoekstra y Kees Dol. Análisis multiescalar de la burbuja inmobiliaria y los desahucios: la Comunidad Autónoma de Euskadi en el contexto estatal y europeo

Inmaculada Caravaca Barroso y Gema González-Romero. Políticas de promoción económica y de ordenación territorial de la economía del conocimiento en la aglomeración metropolitana de Sevilla

Macarena Marchante-Lara y Carlos G. Benavides-Chicón. A comparative analysis of the Provision of Urban Public Transport: Special Reference to Malaga City

Segundo Abrahán Sanabria Gómez. Asimetrías tecnológicas y desequilibrios económicos regionales: una aproximación teórica

Luis Copano Ortiz y Jesús Ventura Fernández. La organización del territorio submunicipal en Andalucía. Criterios administrativos para su delimitación

Estefanía Villar Cheda, M^a Esther Calvo Ocampo, M^a Esther López Vizcaino, Carlos L. Iglesias Patiño, Solmary Silveira Calviño y M^a Isolina Santiago Pérez. Clasificación de los municipios gallegos según su grado de urbanización

Resensiones y reseñas bibliográficas

Texto

Clasificación de los municipios gallegos según su grado de urbanización

Classification of Galician municipalities by degree of urbanization

Estefanía Villar Cheda

M^a Esther Calvo Ocampo

M^a Esther López Vizcaíno

Carlos L. Iglesias Patiño

Solmary Silveira Calviño

Instituto Galego de Estatística

M^a Isolina Santiago Pérez

Consellería de Sanidade

Recibido, Marzo de 2013; Versión final aceptada, Octubre de 2013.

PALABRAS CLAVE: Galicia, Grado de urbanización, Urbano, Rural, Clasificación, Regresión multinomial.

KEYWORDS: Galicia, Degree of urbanisation, Urban, Rural, Classification, Multinomial regression.

Clasificación JEL: O18,R11,R12

RESUMEN

En el marco de la estadística oficial y dentro de las clasificaciones que emplean un número reducido de variables geodemográficas, el Instituto Galego de Estatística (IGE) publicó en julio del 2011 una clasificación del grado de urbanización de los municipios gallegos en seis zonas, basada en un estándar europeo. Teniendo en cuenta análisis posteriores (Calvo *et al*, 2012) se optó por dividir a Galicia en tres zonas, rural, semiurbana y urbana. El objetivo de este trabajo es comprobar cómo esta agregación en 3 clases, se asemeja a la obtenida con un modelo que emplea características sociodemográficas, sociolaborales y económicas que, a priori, se vinculan con el proceso de urbanización.

ABSTRACT

Provide information at the provincial level in Galicia does not always solve the problems associated with designing and implementing public policies, especially those that aim to reduce disparities in the level of development achieved. This does not solve, either, the demands of private information.

However, there is broad agreement that the habitat can be determinant in socioeconomic status. Therefore it is interesting to have information relates to the urban and rural in Galicia or its provinces. A priori, the difference between the urban areas of two provinces seems lower than bet-

ween the urban and rural areas of the same province. It would be useful to contrast this assumption with empirical evidence.

In this framework and within the official statistics classifications which employ a small number of geo-demographic variables, the Galician Statistics Institute (IGE) published in July 2011 a classification of the degree of urbanization of Galician municipalities into six types of areas, based on a European standard. Taking into account subsequent reviews (Calvo et al, 2012) we divide Galicia into three areas, rural, semi-urban and urban. The aim of this study is to see how this aggregation into 3 classes is similar to that obtained with a statistical model that uses sociodemographic, labour and economic characteristics that, a priori, are linked to the urbanisation process.

In this case the units of analysis are the 315 municipalities of Galicia and the dependent variable is the classification of municipalities in rural, semi-urban and urban areas. To analyze its relationship with other sociodemographic, labor and economic variables, we adjust a multinomial logistic regression model which is an extension of the classical logistic regression to the case where the response variable is more than two categories (1-rural classification, 2- semi-urban, 3- urban). In this multinomial logistic regression model, the category of urban is considered as the reference, and each of the remaining categories are compared to it.

For the model variable selection we take into account the results of the three phases of the exploratory analysis: box plots, univariate multinomial logistic models and correlation analysis. We select 12 variables for use as potential explanatory variables in the multinomial model fit, for the first step. After that, we readjust the model with the variables that were significant in the first step and we repeat again what was done in the previous step. We finish the model fit in the third step, in which all the variables included are significant: the average age, the percentage of people registered in Social Security System in services, gross disposable income per capita (GDI) and the social benefits in percentage of GDI.

We calculate also the Odds ratio (OR) that is the relative probability that a municipality is classified in the category k of the response variable Y over it is classified in the reference category (third, in this case) for an element with a value of $X = 1$ compared with one that has a value of $X = 0$. If the variable X is ordinal or continuous, the OR compares the relative probability associated with a unit change in the variable X . In this case exists interaction between variables, then the estimate of the OR associated with a variable depend on the value of the variable that is interacting with it. This is the case of the social benefits in percentage of GDI, which depends of the OR of the registered in Social Security System in services.

For example, in the rural model, the OR for one unit change in the social benefits in percentage of GDI depends on the value of the OR of the people registered in Social Security System in services through the expression $\exp(1.895) \exp(-0.04 * \text{PAFIL_SERV})$, so, in a municipality with a percentage equal to 30.2% of registered in Social Security System (minimum value) the OR of benefits is 2.011 and for a municipality with an percentage of registered in services equal 84.1% (maximum) is 0.233. The unit increase in the social benefits in percentage of GDI increases the probability of rural versus urban ($\text{OR} > 1$) in municipalities with low values of the percentage of registered in Social Security System in services and decrease ($\text{OR} < 1$) in municipalities with higher values of this variable.

The fit of the model permit evaluate if the initial classification is similar to that predicted by the model. It was considered the maximum (p_1, p_2, p_3) as a criterion to classify municipalities. The estimated model has an ability to correctly classify 85.4% of the municipalities, but the model classifies better rural and semi-urban. In fact, the model is consistent with the classification of IGE in 93.2% of rural municipalities, 73.9% of semi-urban and 72.2% of urban. It is convenient, however, a more detailed analysis of the 46 municipalities assigned by the model that differ from the initial categories.

The characterization of the degree of urbanization of Galicia in three areas (urban, rural and transitional), from the aggregation of a classification in six degrees, that mainly uses the population

density, is adequately explained by a multinomial logistic model that uses other socioeconomic variables. With this, it has achieved the goal of the work in its various facets. First, the urban space in Galicia is broader than the ZPD original, enlarged with the ZIP high. Second, a classification that is based only in basic considerations geo-demographic is consistent with another that employs a richer socioeconomic data. Third and last, we have used a new methodology based on a statistical technique different from those previously used to build a clearer and a more parsimonious model.

The methodology used in this work allows to obtain conclusions directly from the variables considered in the model, thus avoiding the use of techniques to reduce the dimension (which are detrimental to the interpretability of the model) as has been done in other studies of similar characteristics (Iglesias et al, 2000; Rua et al, 2003; Pena et al, 2008).

The model also allows us to conclude, as in other studies, that the presence of the tertiary sector and high levels of income, linked to the demographic dynamics, appear to be the determinants of intra-regional disparities in Galicia (Pena et al, 2008).

1. INTRODUCCIÓN

Ofrecer información a nivel provincial en Galicia no siempre resuelve los problemas asociados a diseñar e implementar políticas públicas, sobre todo, aquellas orientadas a reducir disparidades en el grado de desarrollo alcanzado. Ni tampoco resuelve las demandas de información privadas.

Por otra parte, la falta de homogeneidad de los municipios gallegos tanto en superficie, número de habitantes o concentración de unidades de producción agrava la escasez de fuentes a este nivel de desagregación geográfica, por ejemplo, la mayoría de las encuestas por muestreo de la estadística pública no permiten ofrecer estimaciones directas a nivel inferior a la provincia.

No obstante, existe un acuerdo amplio sobre que el hábitat puede ser una característica relevante en la situación socioeconómica. Por ello, interesa disponer de información que se refiera a lo urbano y a lo rural dentro de Galicia o de sus provincias. A priori, la diferencia entre las zonas urbanas de dos provincias parece menor que la zona urbana y la rural de la misma provincia. Sería conveniente contrastar este supuesto con evidencia empírica.

En este marco y dentro de las clasificaciones que emplean un número reducido de variables geodemográficas, el Instituto Galego de Estatística (IGE) publicó en julio del 2011 una clasificación del grado de urbanización de los municipios y las parroquias de Galicia basada en un estándar europeo (IGE, 2011).

La clasificación del grado de urbanización elaborada por el IGE trata de reflejar las peculiaridades del asentamiento poblacional en Galicia y de preservar la armonización con la clasificación a nivel estatal y europeo que ya definió Eurostat (EUROSTAT, 2007) y que aplica el Instituto Nacional de Estadística (INE) para clasificar a los municipios españoles en 3 zonas: zona densamente poblada (ZDP), zona intermedia (ZIP) y zona poco poblada (ZPP).

Para el caso de Galicia esta clasificación es bastante gruesa y no permite diferenciar aglomeraciones importantes de población con cierto grado de urbanización: las denominadas villas, pequeñas ciudades (Rodríguez, 1994 y Precedo *et al*, 2008) o cabeceras comarcales (Torre y Lois, 1995). Para solventar este problema, la clasificación del IGE se definió a partir de la clasificación de Eurostat-INE de manera jerárquica, dividiendo las categorías ZIP y ZPP en subcategorías, lo que dio lugar a seis grupos: ZDP, ZIP alta, ZIP baja, ZPP alta, ZPP intermedia y ZPP baja.

Para delimitar las seis zonas de Galicia según el grado de urbanización, el IGE sólo empleó las variables población y densidad de población acompañadas del criterio de adyacencia. Por otra parte, en Calvo *et al* (2013) se describen las características demográficas, socioeconómicas y sanitarias predominantes en las seis zonas y, como conclusión, se propone para diferenciar el espacio rural y el urbano identificar a la ZPP baja como la Galicia más rural, a la ZDP y a la ZIP alta con la Galicia más urbana y a las restantes como espacio de transición (semiurbana).

Desde una óptica cuantitativa y con diferentes objetivos en mente se han realizado diversas clasificaciones de unidades territoriales de Galicia por debajo del nivel de provincia (Iglesias *et al*, 2000; Rúa *et al*, 2003; Pena *et al*, 2008, sin ánimo de ser exhaustivos).

Muchas de estas aportaciones cuantitativas emplean técnicas de reducción de la dimensionalidad por lo que puede ser deseable emplear otra técnica estadística más enfocada a explicitar el proceso de modelado que redundará en una mejor interpretación del modelo resultante.

Teniendo en cuenta lo anterior, el objetivo de este trabajo es comprobar como esa agregación en 3 clases de la clasificación de 6 zonas elaborada por el IGE se asemeja a la obtenida con un modelo logístico multinomial que emplease características sociodemográficas, sociolaborales y económicas disponibles a nivel municipal en la estadística oficial que, a priori, se vinculan con el proceso de urbanización.

Para ello, lo que resta de trabajo se estructura de la siguiente manera: en la sección siguiente se hace una descripción de las variables empleadas, se continúa con la descripción de la metodología empleada en el trabajo y, a continuación, se exponen los resultados; se finaliza con las conclusiones del estudio.

2. DATOS Y FUENTES

2.1 Descripción de las variables

Se partió de un listado de variables sociodemográficas, laborales y económicas de las que se disponía de datos a nivel municipal, para poder analizar su relación con el grado de urbanización. Las variables se presentan en el Cuadro 1 y se describen de forma más detallada en el anexo.

CUADRO 1
LISTADO DE VARIABLES INICIALES (*)

	Nombre variable	Fuente	Definición y período de referencia
VARIABLES SOCIODEMOGRÁFICAS			
1	ISF	IGE	Índice sintético de fecundidad. Año 2010
2	IND_ENV	IGE	Índice de envejecimiento. Año 2011
3	EDAD_MED	IGE	Edad media de la población. Año 2011
4	PNC	IGE, INE	Pensiones no contributivas por 1.000 habitantes. Año 2010
VARIABLES LABORALES			
5	PR_TOT	IGE, INE	Paro registrado por 1.000 habitantes entre 15 y 65 años. Año 2010
6	PR_25	IGE, INE	Paro registrado de menores de 25 años por 1.000 habitantes entre 15 y 24 años. Año 2010
7	PAFIL_AGR	IGE	Porcentaje de afiliaciones en el sector primario, diciembre 2010
8	PAFIL_INDCON	IGE	Porcentaje de afiliaciones en industria y construcción, diciembre 2010
9	PAFIL_SERV	IGE	Porcentaje de afiliaciones en servicios, diciembre 2010
10	CONT_INI	IGE	Contratos iniciales registrados (por 1.000 habitantes). Año 2010
VARIABLES ECONÓMICAS DE RENTAS			
11	RBDph	IGE	Renta Bruta Disponible (RBD) (euros por habitante). Año 2009
12	Prest_RBD	IGE	Porcentaje de las prestaciones sobre la RBD. Año 2009
13	SRP_hab	IGE	Saldo de las Rentas Primarias bruto (euros por habitante). Año 2009
14	RA_RBD	IGE	Porcentaje de la Remuneración de Asalariados sobre la RBD. Año 2009
15	RendMedio_IRPF	IGE	Rendimiento medio declarado en el IRPF (euros). Año 2009
VARIABLES ECONÓMICAS DE EMPRESAS			
16	Empr_Hab_CA	IGE	Empresas con asalariados por 1.000 habitantes. Año 2010
17	Empr_Hab_SA	IGE	Empresas sin asalariados por 1.000 habitantes. Año 2010
18	Est_habO	IGE	Establecimientos en otros sectores diferentes del sector servicios por 1.000 habitantes. Año 2010
19	Est_habS	IGE	Establecimientos en los servicios por 1.000 habitantes. Año 2010

(*) Los años de referencia de las variables utilizadas están dentro del período 2009-2011. En todas ellas se tomó el último año disponible en el momento de la elaboración de este trabajo. Fuente: páginas web del IGE y del INE (<http://www.ige.eu> y <http://www.ine.es>)

A priori, las variables sociodemográficas seleccionadas, ISF, IND_ENV, EDAD_MED y PNC, puede parecer que tengan comportamientos diferentes según el grado de urbanización; en las zonas urbanas se presentarían poblaciones más jóvenes que en las zonas rurales. En cuanto a la variable PNC, se supone mayor porcentaje de pensiones no contributivas en las zonas rurales que en las urbanas.

De las variables laborales, PR_TOT, PR_25; PAFIL_AGR, PAFIL_INDCON, PAFIL_SERV y CONT_INI, se puede pensar que las zonas rurales tendrán mayor nivel de actividad del sector primario, mientras que en las zonas urbanas las tasas de desempleo, tanto total como de jóvenes, serían superiores, y la actividad se desarrollaría principalmente en los sectores secundario y terciario. También podría suponerse una relación directa entre la variable CONT_INI y la mayor dinámica laboral en las zonas urbanas.

En el grupo de variables económicas, RBDph, Prest_RBD, SRP_hab, RA_RBD, RendMedio_IRPF, Empr_Hab_CA, Empr_Hab_SA, Est_habO, Est_habS, un primer acercamiento a la realidad económica puede llevar a pensar que las rentas generadas en el ámbito urbano son superiores que en el rural, y en cuanto a la estructura del mapa empresarial, al igual que en el grupo de variables anterior, las empresas del sector primario estarían fundamentalmente en zonas rurales y las dedicadas a industria, construcción y servicios en las zonas urbanas.

2.2 Análisis descriptivo

En primer lugar se aplicó el test de Shapiro-Wilks para analizar la normalidad de cada una de las variables. Se rechazó la hipótesis nula de provenir de una normal para todas las variables excepto para la ratio de paro registrado por 1.000 habitantes de entre 15 y 24 años ($p=0,678$), el porcentaje de afiliaciones en el sector servicios ($p=0,216$) y la ratio de empresas con asalariados por 1.000 habitantes ($p=0,204$).

En segundo lugar, para cada una de las variables se contrastó la igualdad de medias en los tres grupos de municipios (rural, semiurbano y urbano) mediante el test de Kruskal-Wallis, rechazándose en todos los casos la hipótesis nula de igualdad de medias.

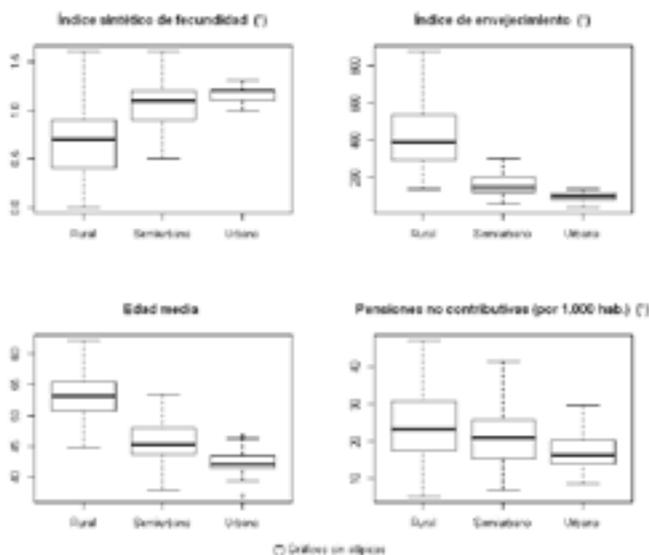
En tercer lugar, para todas las variables se realizaron diagramas de caja que muestran la distribución de cada una de ellas según el tipo de municipio, rural, semiurbano y urbano.

Variables sociodemográficas

En la Figura 1 se muestra la distribución de cada una de las variables sociodemográficas según el tipo de municipio (rural, semiurbano y urbano). En este gráfico se observa cómo, a medida que aumenta la urbanización de un municipio, aumenta

el número medio de hijos por mujer y disminuye el índice de envejecimiento, la edad media y las pensiones no contributivas.

FIGURA 1
**DIAGRAMAS DE CAJAS DE LAS VARIABLES SOCIODEMOGRÁFICAS
SEGÚN LA URBANIZACIÓN DEL MUNICIPIO**



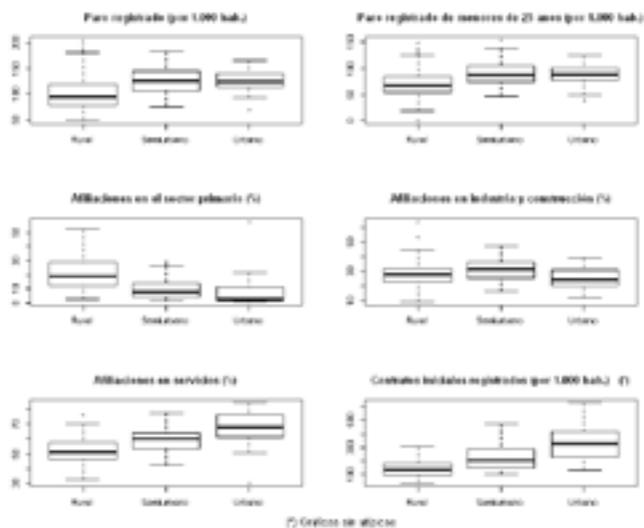
Fuente: IGE. Indicadores demográficos, <http://www.ige.eu>

Los municipios urbanos cuentan con una población más joven, con una edad media de 42 años, en contraposición a los municipios más rurales en donde la media de edad se sitúa en los 53 años (11 años mayor). Esto también se refleja en el índice de envejecimiento que, para los municipios urbanos, está en torno a 103 -es decir por cada 100 personas menores de 20 años hay 103 personas mayores de 65-, mientras que en los rurales se sitúa en 430. El índice sintético de fecundidad no discrimina mucho entre los municipios urbanos y semiurbanos, aunque su valor medio es ligeramente superior en los municipios urbanos (1,16 frente a 1,06). A pesar de que las pensiones no contributivas por 1.000 habitantes presentan cierta asociación con el grado de urbanización no logran discriminar entre rural y semiurbano y tampoco entre semiurbano y urbano.

Variables sociolaborales

Las variables sociolaborales también presentan diferencias según el grado de urbanización del municipio. En términos generales, hay una relación directa o inversa entre la urbanización y la variable de análisis, de modo que al aumentar en urbanización la variable incrementa o disminuye su valor (Figura 2). Una excepción al caso anterior se encuentra en el porcentaje de las afiliaciones en la industria y la construcción, variable que presenta valores similares en los municipios rurales y urbanos, y sensiblemente mayores en los municipios semiurbanos. Sin embargo, los otros dos porcentajes de afiliaciones, el de los servicios y el del sector primario, sí siguen la pauta general y así, a medida que aumenta la urbanización se incrementa el porcentaje de afiliaciones en los servicios y disminuye en el sector primario. Algo similar sucede con los contratos iniciales registrados por 1.000 habitantes, que se incrementan con el grado de urbanización. Las variables de paro registrado presentan una distribución similar en los municipios urbanos y semiurbanos, con valores más altos que en el grupo de los municipios rurales.

FIGURA 2
DIAGRAMAS DE CAJAS DE LAS VARIABLES SOCIOLABORALES SEGÚN LA URBANIZACIÓN DEL MUNICIPIO



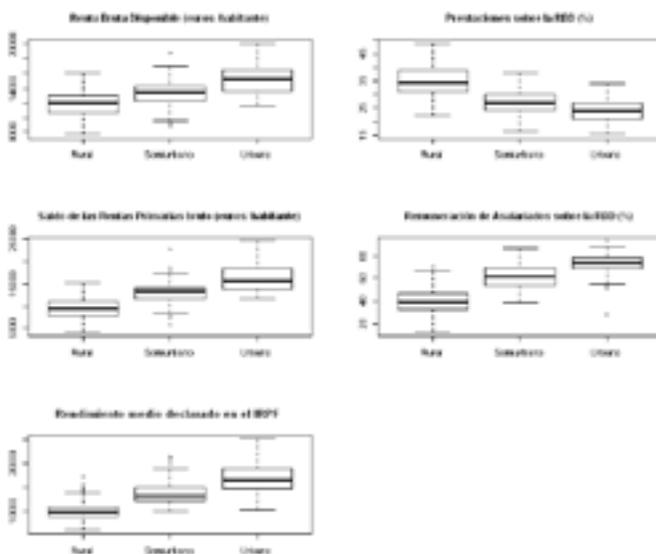
Fuente: IGE. <http://www.ige.eu>

La variable contratos iniciales presenta mucha variabilidad, sobre todo en los municipios semiurbanos, debido a la presencia de gran cantidad de datos atípicos. Por este motivo, para analizar la relación entre el número de contratos iniciales por 1.000 habitantes y la urbanización, se transformó la variable en dicotómica, asignando el valor 0 a aquellos valores situados por debajo de la mediana (164,1) y 1 en caso contrario.

Variables económicas de rentas

En las variables económicas de rentas las diferencias entre municipios rurales, semiurbanos y urbanos son considerables, tal y como se puede observar en la Figura 3. A medida que el grado de urbanización crece también lo hace el Saldo de Rentas Primarias, la Renta Bruta Disponible, el porcentaje de remuneración de asalariados sobre la RBD y el Rendimiento medio declarado. Por el contrario, el grado de urbanización influye en sentido inverso sobre el porcentaje de prestaciones sobre la RBD.

FIGURA 3
**DIAGRAMAS DE CAJAS DE LAS VARIABLES ECONÓMICAS DE RENTAS
 SEGÚN LA URBANIZACIÓN DEL MUNICIPIO**



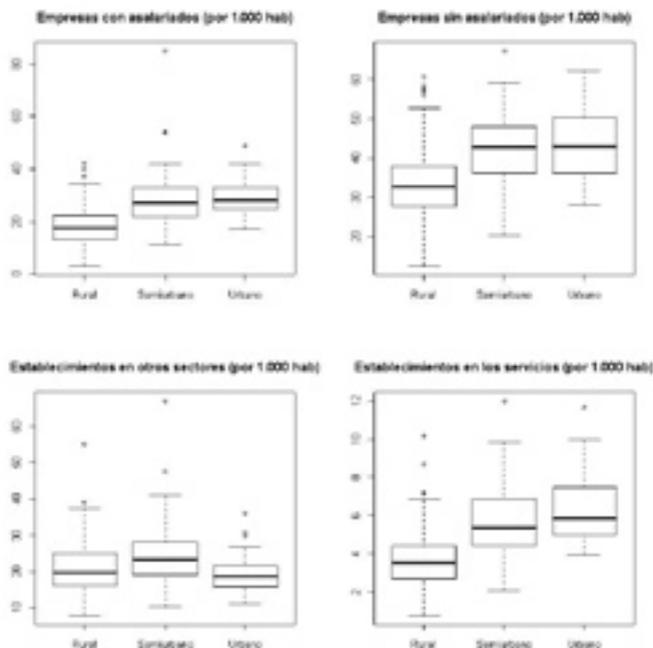
Fuente: IGE. Renta municipal de los hogares, <http://www.ige.eu>

Variables económicas de empresas

Tal y como se puede observar en la Figura 4 no parece que los establecimientos del sector servicios ni las empresas logren discriminar entre urbano y semiurbano a pesar de la existencia de asociación entre el grado de urbanización y estas tres variables cuando se consideran las tres categorías. Así, existe una gran similitud en los valores medios de cada una de las tres variables cuando se comparan las categorías semiurbano y urbano.

Los establecimientos en otros sectores distintos a los servicios por 1.000 habitantes presentan sin embargo valores más altos en los municipios rurales y en los semiurbanos que en los urbanos.

FIGURA 4
DIAGRAMAS DE CAJAS DE LAS VARIABLES ECONÓMICAS DE EMPRESAS SEGÚN LA URBANIZACIÓN DEL MUNICIPIO



Fuente: IGE. Directorio de empresas y unidades locales, <http://www.ige.eu>

METODOLOGÍA

Las unidades de análisis son los 315 municipios de Galicia y la variable dependiente es la clasificación de los municipios en rural, semiurbano y urbano. Para analizar su relación con las variables, sociodemográficas, sociolaborales y económicas, se ajustaron modelos de regresión logística multinomial (Hosmer y Lemeshow, 2000; Faraday, 2006) que consiste en una extensión de la regresión logística clásica al caso en que la variable respuesta tiene más de dos categorías.

En un modelo de regresión logística multinomial, una de las categorías de la variable respuesta se considera como referencia, y cada una de las restantes categorías se compara con ella. En el caso concreto de que la variable respuesta tenga tres categorías, como la clasificación en 1-rural, 2-semiurbano y 3-urbano, hay dos comparaciones posibles en forma del logaritmo de la razón de probabilidades (logit), que se expresa como una función lineal de las variables explicativas:

$$\text{Ln} \left[\frac{p_k}{p_3} \right] = \sum_{j=0}^m X_j \beta_{kj} = X\beta_k, \quad k=1,2$$

O, equivalentemente:

$$p_k = \frac{\exp(X\beta_k)}{1 + \exp(X\beta_1) + \exp(X\beta_2)}, \quad k=1,2$$

donde:

$$p_k = \text{Pr}(Y_k = 1 | X), \quad k=1,2 \quad \text{y} \quad p_3 = 1 - (p_1 + p_2)$$

$Y=(Y_1, Y_2, Y_3)$ es un vector tridimensional con distribución Multinomial $(1, p_1, p_2, p_3)$,

X es una matriz con $m+1$ columnas, la primera con valor uno para todos los casos ($X_0=1$) y el resto con las m variables explicativas.

β_k es el vector de $m+1$ parámetros de la regresión para la k -ésima categoría de respuesta.

Cuando sólo hay una variable explicativa dicotómica ($X=0/1$), la exponencial del coeficiente de regresión de esa variable, para la categoría k de respuesta, es el odds ratio (OR):

$$\text{OR}_k(X) = \frac{p_k(X=1)/p_3(X=1)}{p_k(X=0)/p_3(X=0)} = \frac{\exp(\beta_{k0} + \beta_{k1})}{\exp(\beta_{k0})} = \exp(\beta_{k1}), \quad k=1,2$$

El OR es la probabilidad relativa de clasificarse en la categoría k de la variable respuesta Y frente a hacerlo en la categoría de referencia (tercera, en este caso) para un elemento con un valor de $X=1$ comparado con uno que tenga un valor $X=0$. Si la variable X es ordinal o continua, el OR compara la probabilidad relativa asociada a un cambio unitario en la variable X . En general, en un modelo con más de una variable explicativa, el coeficiente de cualquier variable tiene esta misma interpretación considerando fijos los valores del resto de variables. En ese caso, el OR se dice que está ajustado por dichas variables.

Cuando se introduce en el modelo un término de interacción entre dos variables (el producto de ellas), por ejemplo X_1 y X_2 , debe tenerse en cuenta que el OR de cada una ya no es la exponencial de su coeficiente, sino que depende de los valores de la otra variable:

$$OR_k(X_1) = \exp(\beta_{k1} + \beta_{k12} X_2)$$

$$OR_k(X_2) = \exp(\beta_{k2} + \beta_{k12} X_1)$$

En este trabajo, la selección de las variables que finalmente entraron en el modelo multivariante se hizo en dos etapas. En una primera etapa se realizó un análisis ajustando un modelo simple de regresión logística multinomial para cada una de las variables explicativas, y se consideraron para su posible inclusión en el modelo aquellas con un coeficiente significativo ($p < 0,05$) en al menos una de las categorías (rural o semiurbano). Se calcularon también las correlaciones entre las variables, para realizar una valoración conjunta de los coeficientes del modelo y de las correlaciones, con el fin de eliminar variables muy correlacionadas con el resto.

En la segunda etapa del proceso de identificación del modelo final se optó por el procedimiento de las significaciones sucesivas. Este procedimiento, simple y directo, consiste en lo siguiente:

- Paso 1: se ajusta el modelo con todas las variables seleccionadas en la primera etapa.
- Paso 2: se identifican aquellas variables cuyos coeficientes sean significativamente distintos de 0 ($p < 0,05$). Para ello se usa el test de Wald en cada paso. Si todas ellas están en ese caso, se concluye el proceso. De lo contrario se va al paso 3.
- Paso 3: se ajusta el modelo con aquellas variables que resultaron significativas en el paso anterior y se vuelve al paso 2.

Los parámetros del modelo se estimaron usando el método de máxima verosimilitud y su significación se valoró con el test de Wald. Todos los cálculos se hicieron con la librería `mlogit` del paquete estadístico R (Croissant, 2011).

4. RESULTADOS

4.1 Selección de variables para el modelo

Para conocer en qué medida cada una de las variables influye en el grado de urbanización, se ajustó un modelo simple de regresión logística multinomial con cada una de ellas (Cuadro 2), y el valor de la OR para cada variable independiente se utilizó para valorar su papel y decidir si debe incluirse o no en el modelo múltiple de regresión multinomial. Para poder realizar una interpretación más sencilla de los OR se reescalaron las variables económicas de rentas RBDph, saldo de rentas primarias y rendimiento medio declarado del IRPF, dividiéndolas entre 1.000.

Tal y como se observa en el Cuadro 2, todas las variables presentan algún tipo de asociación con la variable grado de urbanización, aunque en algún caso no discriminan entre las categorías semiurbano y urbano. Conviene, por tanto, resaltar aquellas variables que no presentan una asociación significativa en la comparación semiurbano-urbano, ya que en el caso rural-urbano todas son significativas a excepción de la variable número de establecimientos en otros sectores distintos a los servicios por 1.000 habitantes.

Así, el índice sintético de fecundidad, los dos indicadores de paro, el porcentaje de afiliaciones en la agricultura y las otras tres variables económicas de empresas no permiten diferenciar entre las categorías semiurbano y urbano. El intervalo de confianza para el OR de cada una de estas variables en el caso semiurbano contiene siempre al 1. Esta similitud en el comportamiento de este conjunto de variables en los casos urbano y semiurbano se comprueba también en los diagramas de cajas de estas variables en la sección 2.2.

El estudio anterior se complementó con el análisis de las correlaciones entre las variables mediante el coeficiente de correlación de Pearson, con el fin de evitar la inclusión en el modelo de variables muy correlacionadas entre sí que puedan dar lugar a estimaciones inestables de los OR. En este paso no se pretende descartar variables, puesto que el propio método utilizado ya selecciona las más explicativas. Por tanto, se optó por seleccionar un valor de compromiso para la correlación (0,75) que permitiese, por un lado, la eliminación de variables altamente correladas y por otro, no redujese demasiado el conjunto de variables seleccionadas.

En el grupo de variables sociodemográficas, la edad media y el índice de envejecimiento presentan una correlación muy alta entre ellas (0,93). Se decidió incluir en la estimación del modelo sólo la primera de ellas, ya que las dos facilitan básicamente la misma información. La decisión de incluir la edad media en lugar del índice de envejecimiento está motivada por el hecho de que para el cálculo de la primera se utiliza información de toda la población de las unidades de análisis.

CUADRO 2
MODELO LOGIT DEL GRADO DE URBANIZACIÓN SOBRE CADA VARIABLE

Variables sociodemográficas	Coefficiente	Pr(> t)	OR	IC 95%		
Rural: ISF	-5,487	0,000	0,004	0,001	0,020	*
Semiurbano: ISF	-1,144	0,097	0,319	0,083	1,230	
Rural: IND_ENV	0,058	0,000	1,060	1,044	1,077	*
Semiurbano: IND_ENV	0,035	0,000	1,036	1,021	1,051	*
Rural: EDAD_MED	1,160	0,000	3,189	2,456	4,139	*
Semiurbano: EDAD_MED	0,496	0,000	1,642	1,339	2,014	*
Rural: PNC	0,123	0,000	1,131	1,069	1,197	*
Semiurbano: PNC	0,082	0,006	1,086	1,024	1,151	*
Variables sociolaborales						
Rural: PR_TOT	-0,031	0,000	0,969	0,956	0,982	*
Semiurbano: PR_TOT	-0,001	0,833	0,999	0,986	1,012	
Rural: PR_25	-0,035	0,000	0,966	0,950	0,981	*
Semiurbano: PR_25	0,001	0,934	1,001	0,984	1,017	
Rural: PAFIL_AGR	0,182	0,000	1,199	1,124	1,279	*
Semiurbano: PAFIL_AGR	0,047	0,161	1,048	0,982	1,119	
Rural: PAFIL_INDCON	0,058	0,032	1,060	1,005	1,118	*
Semiurbano: PAFIL_INDCON	0,116	0,000	1,123	1,059	1,191	*
Rural: PAFIL_SERV	-0,229	0,000	0,795	0,750	0,843	*
Semiurbano: PAFIL_SERV	-0,106	0,000	0,899	0,854	0,946	*
Rural: CONT_INI_C	-3,131	0,000	0,044	0,013	0,148	*
Semiurbano: CONT_INI_C	-1,583	0,014	0,205	0,058	0,728	*
Variables económicas de rentas						
Rural: RBDph	-1,016	0,000	0,362	0,278	0,473	*
Semiurbano: RBDph	-0,506	0,000	0,603	0,476	0,764	*
Rural: Prest_RBD	0,475	0,000	1,608	1,423	1,817	*
Semiurbano: Prest_RBD	0,203	0,000	1,225	1,099	1,366	*
Rural: SRP_hab	-1,063	0,000	0,345	0,272	0,437	*
Semiurbano: SRP_hab	-0,427	0,000	0,653	0,544	0,783	*
Rural: RA_RBD	-0,241	0,000	0,786	0,744	0,830	*
Semiurbano: RA_RBD	-0,090	0,000	0,914	0,876	0,955	*
Rural: RendMedio_IRPF	-1,259	0,000	0,284	0,218	0,370	*
Semiurbano: RendMedio_IRPF	-0,342	0,000	0,710	0,609	0,829	*
Variables económicas de empresas						
Rural: Empr_Hab_CA	-0,202	0,000	0,817	0,795	0,839	*
Semiurbano: Empr_Hab_CA	-0,010	0,622	0,990	0,968	1,013	
Rural: Empr_Hab_SA	-0,126	0,000	0,882	0,844	0,921	*
Semiurbano: Empr_Hab_SA	-0,016	0,449	0,984	0,944	1,026	
Rural: Est_habO	0,038	0,216	1,038	0,849	1,269	
Semiurbano: Est_habO	0,097	0,003	1,102	0,934	1,299	*
Rural: Est_habS	-0,997	0,000	0,369	0,285	0,478	*
Semiurbano: Est_habS	-0,169	0,101	0,844	0,689	1,034	

Nota: el * señala que el coeficiente es significativamente distinto de cero ($p < 0,05$)

Fuente: Elaboración propia

Por otro lado, en las variables sociolaborales se presenta un caso de multicolinealidad exacta entre las tres variables de afiliación, por lo que si se optase por incluir las tres en el modelo, el ajuste sería imposible. A lo anterior hay que añadir el hecho de que en este grupo de variables el único caso de correlación superior a 0,75 en valor absoluto es el que afecta a las variables porcentaje de afiliados en los servicios y porcentaje de afiliaciones en el sector primario. Se optó por eliminar esta última ya que del análisis del modelo simple multinomial para esta variable se desprende que no es significativa para discriminar semiurbano con respecto a urbano.

En las variables económicas relativas a la renta los únicos casos en los que las correlaciones superan 0,75 en valor absoluto son: las del saldo de rentas primarias con todas las demás variables económicas de rentas, y la del rendimiento medio declarado en el IRPF con la remuneración de asalariados sobre la RBD. Se eliminaron, por tanto, del análisis las variables saldo de rentas primarias y rendimiento medio declarado en el IRPF, ya que de este modo las correlaciones que quedan entre todas las variables de este grupo son inferiores a 0,75.

Para finalizar, en las variables económicas de empresas, los establecimientos en el sector servicios por 1.000 habitantes presenta correlaciones superiores a 0,75 con las variables relativas a las empresas con y sin asalariados. Esto último, junto al hecho de que los modelos simples multinomiales para cada una de estas variables no discriminan en el caso semiurbano, tal y como se refleja también en sus diagramas de caja, condujo a la selección de una única variable de este grupo, los establecimientos en el sector servicios por 1.000 habitantes (Est_HabS). La selección de esta variable se realizó comparando los OR obtenidos en las estimaciones de los modelos logit univariantes para esas tres variables de empresas, ya que los establecimientos en el sector servicios por 1.000 habitantes es la variable que presenta mayores diferencias entre los OR de rural y semiurbano.

Por tanto, el análisis conjunto de los modelos logit univariados y de las correlaciones condujo a la eliminación de seis variables: índice de envejecimiento, porcentaje de afiliaciones en la agricultura, saldo de rentas primarias, rendimiento medio declarado en el IRPF, empresas con asalariados por 1.000 habitantes y empresas sin asalariados por 1.000 habitantes. Las variables que se emplearon como explicativas en el primer ajuste del modelo múltiple de regresión logística multinomial son, consecuentemente, las que figuran en el Cuadro 3.

CUADRO 3
LISTADO DE VARIABLES A INTRODUCIR EN EL MODELO DE REGRESIÓN

	Nombre variable	Definición y período de referencia
1	ISF	Índice sintético de fecundidad. Año 2010
3	EDAD_MED	Edad media de la población. Año 2011
4	PNC	Pensiones no contributivas por 1.000 habitantes. Año 2010
5	PR_TOT	Paro registrado por 1.000 habitantes entre 15 y 65 años. Año 2010
6	PR_25	Paro registrado de menores de 25 años por 1.000 habitantes entre 15 y 24 años. Año 2010
8	PAFIL_INDCON	Porcentaje de afiliaciones en industria y construcción, diciembre 2010
9	PAFIL_SERV	Porcentaje de afiliaciones en servicios, diciembre 2010
10	CONT_INI_C	1 si CONT_INI >164,1 0 en otro caso
11	RBDph	Renta Bruta Disponible (euros por habitante). Año 2009
12	Prest_RBD	Porcentaje de las prestaciones sobre la RBD. Año 2009
14	RA_RBD	Porcentaje de la Remuneración de Asalariados sobre la RBD. Año 2009
18	Est_habO	Establecimientos en otros sectores diferentes al sector servicios por 1.000 habitantes. Año 2010
19	Est_habS	Establecimientos en los servicios por 1.000 habitantes. Año 2010

Fuente: Elaboración propia.

4.2 Aplicación del modelo logit multinomial

Se aplicó el procedimiento de las significaciones sucesivas para determinar las variables que formarán parte del modelo. En un primer paso, se ajustó un modelo con las 13 variables resultantes del estudio exploratorio y se seleccionaron entre ellas las que mantenían la significación estadística. Se ajustó de nuevo el modelo con las variables que resultaron significativas en este primer paso y se repitió de nuevo lo hecho en el paso anterior. Se finalizó el ajuste del modelo en el tercer paso, en el que todas las variables incluidas resultaron significativas: la edad media, el porcentaje de afiliaciones en los servicios, la renta disponible bruta por habitante y el porcentaje de las prestaciones sobre la RBD (Cuadro 4).

CUADRO 4
RESULTADOS DEL AJUSTE DEL MODELO LOGIT MULTINOMIAL CON 4
VARIABLES

	Coefficiente	Error estándar	t-valor	Pr(> t)	OR	
Rural:(intercept)	-34,790	7,344	-4,737	0,000	-	*
Semiurbano:(intercept)	-15,830	5,596	-2,829	0,005	-	*
Rural:EDAD_MED	1,596	0,221	7,232	0,000	4,933	*
Semiurbano:EDAD_MED	0,866	0,187	4,628	0,000	2,376	*
Rural:PAFIL_SERV	-0,130	0,046	-2,806	0,005	0,878	*
Semiurbano:PAFIL_SERV	-0,045	0,036	-1,245	0,213	0,956	
Rural:RBDph	-1,261	0,291	-4,337	0,000	0,283	*
Semiurbano:RBDph	-0,770	0,221	-3,481	0,000	0,463	*
Rural:Prest_RBD	-0,466	0,135	-3,456	0,001	0,627	*
Semiurbano:Prest_RBD	-0,288	0,108	-2,667	0,008	0,750	*

Fuente: Elaboración propia

En el modelo obtenido a partir de los coeficientes del Cuadro 4 las prestaciones sobre la RBD tienen unos OR menores que 1, cuando en el estudio exploratorio los OR eran mayores que uno y evidenciaban que a medida que aumentan las prestaciones sobre la RBD aumentaba la ruralidad del municipio, por tanto, es posible que exista un problema de interacción entre variables. Para resolverlo, se ajustó el modelo incluyendo todas las interacciones posibles entre las variables anteriores. Las únicas interacciones significativas son las que hay entre las variables PAFIL_SERV con RBDph y PAFIL_SERV con Prest_RBD. Así pues, la asociación entre el porcentaje de prestaciones sobre la renta y el grado de urbanización (o entre la RBD y el grado de urbanización) es diferente según el valor que tenga el porcentaje de afiliaciones al sector servicios. Se introdujeron estas interacciones en el modelo y se obtuvieron los resultados del Cuadro 5.

CUADRO 5
RESULTADOS DEL AJUSTE DEL MODELO LOGIT MULTINOMIAL CON 4
VARIABLES Y LAS INTERACCIONES SIGNIFICATIVAS

Logit	Variable	Coficiente	Error estándar	t-valor	Pr(> t)
Rural	(intercept)	-154,995	47,020	-3,296	0,001
	EDAD_MED	1,720	0,244	7,037	0,000
	Prest_RBD	1,895	0,827	2,292	0,022
	RBDph	3,012	2,033	1,481	0,139
	PAFIL_SERV	1,830	0,747	2,451	0,014
	Prest_RBD:PAFIL_SERV	-0,040	0,014	-2,879	0,004
	RBDph:PAFIL_SERV	-0,073	0,034	-2,171	0,030
Semiurbano	(intercept)	-73,607	36,838	-1,998	0,046
	EDAD_MED	0,978	0,213	4,600	0,000
	Prest_RBD	1,317	0,705	1,868	0,062
	RBDph	0,374	1,477	0,253	0,800
	PAFIL_SERV	0,802	0,547	1,465	0,143
	Prest_RBD:PAFIL_SERV	-0,026	0,011	-2,275	0,023
	RBDph:PAFIL_SERV	-0,018	0,023	-0,776	0,438

Fuente: Elaboración propia

En el modelo obtenido a partir de los coeficientes del Cuadro 5 la variable RBDph no tiene coeficientes significativamente distintos de cero, pero sí lo tiene la interacción con el porcentaje de afiliados en los servicios para el modelo rural. La variable RBDph se mantuvo en el modelo porque en este trabajo se adopta el criterio de que si una interacción es significativa, se mantienen en el modelo las variables que forman parte de esa interacción. Igual ocurre con las prestaciones sobre la RBD y el porcentaje de afiliados en los servicios, que pierden la significación en el modelo semiurbano, aunque sí tiene significación la interacción entre ambas.

Es interesante comprobar que, dado que en el modelo se han introducido dos parámetros más, la cantidad de información aportada es significativa. Para ello se compara la deviance del modelo reducido con la deviance del modelo considerado en el Cuadro 5 y se observa que la diferencia es significativa ($p=0,017$).

Por tanto, a continuación se muestran las ecuaciones del modelo final:

$$\ln\left[\frac{p_1}{p_3}\right] = -154,995 + 1,720 * EDAD_MED + 1,895 * Prest_RBD + 3,012 * RBDph + 1,830 * PAFIL_SERV - 0,040 * Prest_RBD * PAFIL_SERV - 0,073 * RBDph * PAFIL_SERV$$

$$\ln\left[\frac{p_2}{p_3}\right] = -73,607 + 0,978 * EDAD_MED + 1,317 * Prest_RBD + 0,374 * RBDph + 0,802 * PAFIL_SERV - 0,026 * Prest_RBD * PAFIL_SERV - 0,018 * RBDph * PAFIL_SERV$$

En el Cuadro 6 se muestra la expresión de los OR que se deducen del modelo para cada una de las variables.

**CUADRO 6
OR PARA LOS MODELOS RURAL Y SEMIURBANO**

	Variable	OR
Rural	EDAD_MED	5,585
	Prest_RBD	$\exp(1,895)\exp(-0,04*PAFIL_SERV)$
	PAFIL_SERV	$\exp(1,83)\exp(-0,04*Prest_RBD)\exp(-0,073*RBDph)$
	RBDph	$\exp(3,012)\exp(-0,073*PAFIL_SERV)$
Semiurbano	EDAD_MED	2,660
	Prest_RBD	$\exp(1,317)\exp(-0,026*PAFIL_SERV)$
	PAFIL_SERV	$\exp(0,802)\exp(-0,026*Prest_RBD)\exp(-0,018*RBDph)$
	RBDph	$\exp(0,374)\exp(-0,018*PAFIL_SERV)$

Fuente: Elaboración propia

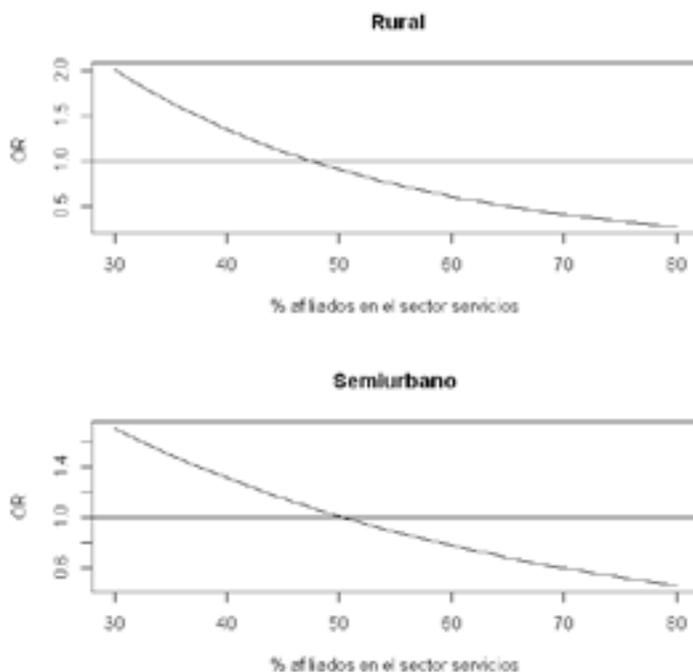
Al existir interacción entre dos variables, la estimación de los OR asociados a una variable depende del valor de la variable que está interaccionando con ella. Este es el caso de las prestaciones sobre la RBD, cuyo OR depende de los valores del porcentaje de afiliaciones en los servicios.

En el modelo para rural, el OR por una unidad de cambio en las prestaciones sobre la RBD depende del valor del porcentaje de afiliaciones en el sector servicios a través de la expresión $\exp(1,895)\exp(-0,04*PAFIL_SERV)$; así, en un municipio con un porcentaje de afiliaciones igual a 30,2% (valor mínimo) el OR de las prestaciones es de 2,011 y para un municipio con un porcentaje de afiliados en los servicios igual a 84,1% (máximo) es de 0,233 (Figura 5, superior). El incremento de una unidad en las prestaciones sobre la renta aumenta la probabilidad de rural frente a urbano ($OR > 1$) en los municipios con valores bajos del porcentaje de afiliaciones en los servicios y la disminuye ($OR < 1$) en aquellos municipios con mayores valores de esta variable.

En el modelo para semiurbano, el OR de la variable prestaciones sobre RBD también es función del porcentaje de afiliaciones en los servicios (Figura 5, inferior).

El incremento de una unidad en las prestaciones sobre la RBD provoca un aumento en la probabilidad de ser semiurbano frente a urbano ($OR > 1$) en los municipios con valores bajos del porcentaje de afiliaciones en los servicios y la disminuye ($OR < 1$) en aquellos municipios con mayores valores de esta variable.

FIGURA 5
OR DE LAS PRESTACIONES SOCIALES SOBRE LA RBDPH SEGÚN EL VALOR DEL PORCENTAJE DE AFILIADOS EN EL SECTOR SERVICIOS

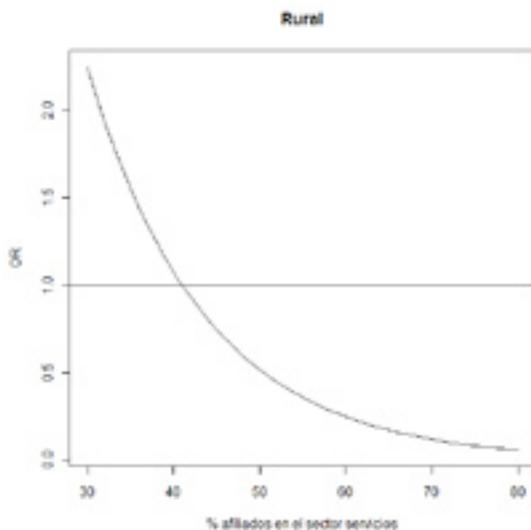


Fuente: Elaboración propia

En el caso de la RBDph, el OR depende también del porcentaje de afiliaciones en el sector servicios. No obstante, en el modelo semiurbano no es significativa ni la variable RBDph ni la interacción de ésta con cualquier otra por lo que en la Figura 6 se muestra únicamente el OR de la RBDph según el valor del porcentaje de afiliaciones en el sector servicios para el modelo rural.

En este último, el OR por una unidad de cambio en la RBDph depende del valor del porcentaje de afiliaciones en el sector servicios a través de la expresión $\exp(3,012)\exp(-0,073*PAFIL_SERV)$; así, en un municipio con un porcentaje de afiliaciones igual a 30,2% (valor mínimo) el OR de la RBDph es de 2,242 y para un municipio con un porcentaje de afiliados en los servicios igual a 84,1% (máximo) es de 0,044 (Figura 6). El incremento de una unidad en la RBDph aumenta la probabilidad de rural frente a urbano (OR>1) en los municipios con valores bajos del porcentaje de afiliaciones en los servicios y la disminuye (OR<1) en aquellos municipios con mayores valores de esta variable.

FIGURA 6
OR DE LA RBDPH SEGÚN EL VALOR DEL PORCENTAJE DE AFILIADOS EN EL SECTOR SERVICIOS PARA EL MODELO RURAL



Fuente: Elaboración propia

En el caso de las afiliaciones del sector servicios, al ser significativas las interacciones con las prestaciones sobre la RBD y con la RBDph, el OR será función de ambas.

Como el objetivo de este trabajo es comprobar cómo la clasificación en 3 clases (rural, semiurbana y urbana) se asemeja a la obtenida con un modelo que emplee características sociodemográficas, sociolaborales y económicas disponibles

a nivel municipal en la estadística pública que, a priori, se vinculan con el proceso de urbanización, es necesario ver como clasifica el modelo propuesto. La Tabla 7 permite evaluar en qué medida la clasificación inicial se asemeja a la pronosticada por el modelo. Se ha considerado el máximo de la probabilidad de (p_1 , p_2 , p_3) como criterio para clasificar a los municipios.

El modelo estimado en el Cuadro 5, junto con el criterio anterior, tiene una capacidad de clasificar correctamente el 85,4% de los municipios, aunque clasifica en mayor porcentaje a los municipios rurales que los urbanos y semiurbanos. De hecho, el modelo concuerda con la clasificación del IGE en el 93,2% de los municipios rurales, el 73,9% de los semiurbanos y el 72,2% de los urbanos (Cuadro 7).

CUADRO 7
CLASIFICACIÓN PRONOSTICADA POR EL MODELO

Pronosticado	Observado		
	Rural	Semiurbano	Urbano
Rural	93,2	15,9	0,0
Semiurbano	6,8	73,9	27,8
Urbano	0,0	10,2	72,2

Fuente: Elaboración propia

Es conveniente, de todas formas, realizar un análisis más pormenorizado de los 46 municipios en los que difieren las categorías iniciales y las asignadas por el modelo.

Municipios urbanos

Se parte de 36 municipios urbanos de los cuales el modelo clasifica como urbanos a 26 y como semiurbanos a 10 (Cuadro 8). De estos, Sada y Cambados tienen unas probabilidades muy similares de pertenecer a urbano o semiurbano. Betanzos se clasifica como semiurbano por poseer una edad media de 44 años, mientras que el resto de municipios presentan una edad media más baja, en torno a 42 años; en estos la variable determinante es el porcentaje de afiliaciones en el sector servicios (con valores mucho más bajos que la media).

CUADRO 8
MUNICIPIOS URBANOS CLASIFICADOS COMO SEMIURBANOS

Código	Nombre municipio	Clasificación IGE	P ₁	P ₂	P ₃
15009	Betanzos	ZIP alta	0,002	0,730	0,267
15011	Boiro	ZIP alta	0,015	0,646	0,338
15073	Ribeira	ZIP alta	0,014	0,677	0,309
15075	Sada	ZIP alta	0,000	0,522	0,478
32008	Barbadás	ZIP alta	0,030	0,644	0,326
36004	Bueu	ZIP alta	0,014	0,773	0,213
36006	Cambados	ZIP alta	0,011	0,527	0,462
36022	Grove (O)	ZIP alta	0,010	0,665	0,325
36029	Moaña	ZDP	0,010	0,648	0,342
36061	Vilanova de Arousa	ZIP alta	0,060	0,848	0,092

Fuente: Elaboración propia

Municipios semiurbanos

En la categoría semiurbana, formada por 88 municipios, el modelo asigna 9 municipios como urbanos y 14 como rurales.

CUADRO 9
MUNICIPIOS SEMIURBANOS CLASIFICADOS COMO URBANOS

Código	Nombre municipio	Clasificación IGE	P ₁	P ₂	P ₃
15060	Oroso	ZPP intermedia	0,002	0,303	0,696
15082	Teo	ZIP baixa	0,000	0,278	0,722
27902	Burela	ZPP alta	0,001	0,217	0,783
32009	O Barco de Valdeorras	ZPP alta	0,002	0,352	0,646
36003	Baiona	ZIP baixa	0,001	0,302	0,697
36021	Gondomar	ZIP baixa	0,003	0,386	0,611
36044	Ponteceures	ZIP baixa	0,003	0,391	0,606
36049	Salceda de Caselas	ZIP baixa	0,000	0,122	0,877
36053	Soutomaior	ZIP baixa	0,000	0,110	0,889

Fuente: Elaboración propia

Excepto Burela y O Barco de Valdeorras que constituyen unidades aisladas, el resto de municipios están en las proximidades de municipios urbanos (Pontece-

sures junto a Padrón, Oroso y Teo junto a Santiago y Baiona, Gondomar, Salceda de Caselas y Soutomaior en la zona Pontevedra-Vigo) por lo que el modelo está ampliando el área urbana con municipios de sus alrededores, como se puede observar en el Mapa 1.

CUADRO 10
MUNICIPIOS SEMIURBANOS CLASIFICADOS COMO RURALES

Código	Nombre municipio	Clasificación IGE	P_1	P_2	P_3
15006	Arzúa	ZPP intermedia	0,724	0,275	0,000
15043	Malpica de Bergantiños	ZPP intermedia	0,944	0,056	0,000
15062	Outes	ZPP alta	0,905	0,095	0,000
15901	Cariño	ZPP intermedia	0,753	0,247	0,000
27016	Chantada	ZPP intermedia	0,526	0,474	0,001
27030	Mondoñedo	ZPP intermedia	0,899	0,101	0,000
27032	Monterroso	ZPP intermedia	0,650	0,350	0,000
27050	Quiroga	ZPP intermedia	0,791	0,209	0,000
32024	Celanova	ZPP intermedia	0,539	0,460	0,001
32045	Maside	ZPP alta	0,900	0,100	0,000
32088	Vilamartín de Valdeorras	ZPP alta	0,980	0,020	0,000
36001	Arbo	ZPP intermedia	0,921	0,079	0,000
36014	Crecente	ZPP intermedia	0,985	0,015	0,000
36030	Mondariz	ZPP intermedia	0,528	0,470	0,002

Fuente: Elaboración propia

Los municipios semiurbanos clasificados por el modelo en rurales (Cuadro 10) están todos en la categoría ZPP, intermedia o alta. Es decir, inicialmente estaban clasificados en la ZPP y con la clasificación del IGE se diferenciaron del resto, aun así el modelo los clasifica como rurales. La edad media de los municipios semiurbanos que el método clasifica como urbanos está por debajo de los 42 años, distante del valor medio de esta variable en el resto de municipios semiurbanos clasificados como tal, que es de 45,5 años. Análogamente, la edad media en los municipios semiurbanos que el método clasifica como rurales es superior a 48 años. Parece que esa variable es la que determina que dichas unidades tengan unas probabilidades de pertenecer a urbano o a rural superiores a la probabilidad de ser semiurbanos.

Municipios rurales

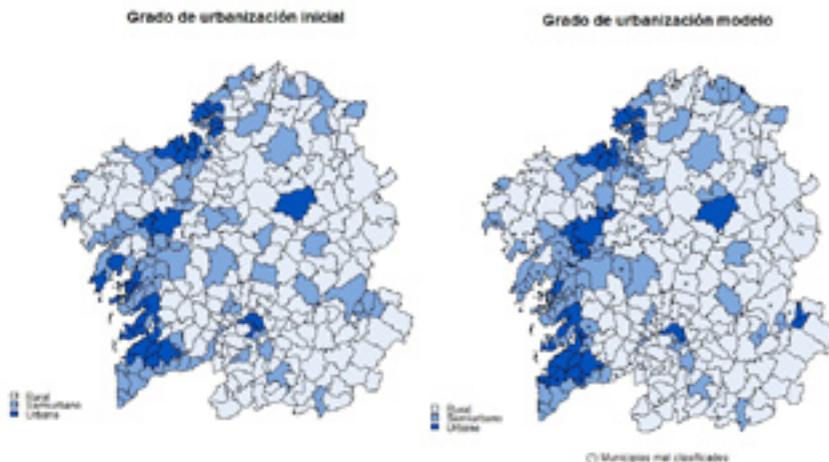
De los 191 municipios clasificados como rurales, el método clasifica como semiurbanos a sólo 13 de ellos, de los que 6 tienen unas probabilidades muy similares entre rural y semiurbano (Cuadro 11): Curtis, Oza dos Ríos, Rois, Pobra de Trives, Ponte Caldelas y Silleda. Los restantes se caracterizan por tener una edad media más baja (47 años) que los municipios que el modelo clasifica como rurales (53 años). Estos municipios también destacan por tener un porcentaje de afiliados en el sector servicios alrededor del 60%, superior a la de los municipios clasificados como rurales (51%).

CUADRO 11
MUNICIPIOS RURALES CLASIFICADOS COMO SEMIURBANOS

Código	Nombre municipio	Clasificación IGE	P ₁	P ₂	P ₃
15012	Boqueixón	ZPP baixa	0,235	0,753	0,012
15027	Coirós	ZPP baixa	0,327	0,665	0,008
15032	Curtis	ZPP baixa	0,433	0,565	0,002
15033	Dodro	ZPP baixa	0,328	0,669	0,003
15063	Oza dos Ríos	ZPP baixa	0,472	0,527	0,001
15074	Rois	ZPP baixa	0,447	0,552	0,002
27025	Xove	ZPP baixa	0,290	0,708	0,003
27029	Meira	ZPP baixa	0,280	0,716	0,003
27039	Outeiro de Rei	ZPP baixa	0,372	0,625	0,003
32063	Pobra de Trives	ZPP baixa	0,458	0,542	0,001
36037	Pazos de Borbén	ZPP baixa	0,097	0,834	0,069
36043	Ponte Caldelas	ZPP baixa	0,468	0,524	0,008
36052	Silleda	ZPP baixa	0,460	0,538	0,002

Fuente: Elaboración propia

MAPA 1
MUNICIPIOS DE GALICIA CON EL GRADO DE URBANIZACIÓN INICIAL (IZQUIERDA) Y SEGÚN EL MODELO LOGIT MULTINOMIAL (DERECHA)



Fuente: Elaboración propia

5. CONCLUSIONES

La caracterización de la urbanización de Galicia en tres espacios (urbano, rural y de transición), a partir de la agregación de una clasificación en seis grados que fundamentalmente emplea la densidad de población, viene explicada adecuadamente por un modelo logístico multinomial que emplea otro tipo de variables socioeconómicas.

Con esto se ha conseguido alcanzar el objetivo del trabajo en sus diversas facetas. Primero, el espacio urbano en Galicia es más amplio que la ZDP original, acrecentándola con la ZIP alta. Segundo, una clasificación que solo se basa en meras consideraciones geodemográficas básicas está en consonancia con otra que emplea una información socioeconómica más rica. Tercero y último, se ha utilizado una nueva metodología basada en una técnica estadística diferente de las anteriormente empleadas que construye, de una forma más clara, un modelo más escueto.

La agregación en tres clases –zonas rural, semiurbana y urbana– a partir de las 6 de la clasificación del IGE establecida para los municipios gallegos, se asemeja a la obtenida tras aplicar el modelo multinomial logístico obtenido.

Teniendo en cuenta el objetivo, una vez ajustado el modelo se puede observar que concuerda en el 85,4% de los municipios gallegos y que su concordancia es mayor con respecto a los municipios rurales que a los urbanos y semiurbanos. Esta situación viene condicionada por las variables seleccionadas. Ya en el análisis exploratorio se plantearon bastantes dificultades para seleccionar variables que discriminaran bien entre los municipios urbanos y semiurbanos, motivadas porque los primeros se corresponden con los municipios de las grandes ciudades y los segundos con los que rodean a los anteriores. La similitud entre las características de estos dos grupos en las variables disponibles hace que el punto de corte que permite diferenciar bien a un tipo de municipios del otro no esté lo suficientemente claro.

A pesar de todo, se considera que el porcentaje de bien clasificados es aceptable y, por tanto, con ello se muestra que la agregación aportada inicialmente que divide al territorio gallego en 3 zonas y que se obtuvo a partir de las variables de población, densidad de población y un criterio de adyacencia, no es sólo una clasificación geodemográfica, sino que incluye otros muchos aspectos económicos, sociolaborales y sociodemográficos que permite llegar a conclusiones similares a la de otros estudios de la literatura (Iglesias *et al*, 2000; Rúa *et al*, 2003; Pena *et al*, 2008)

La ventaja de plantear un modelo explícito es que también permite establecer asociaciones entre las características sociodemográficas, sociolaborales y económicas que forman parte del modelo (en forma de variables explicativas) y la variable respuesta, en este caso la probabilidad de pertenecer a rural, semiurbano o urbano. Así, se puede concluir que un aumento de un año en la edad media de un municipio incrementa 5 veces la probabilidad de pertenecer a la zona rural en relación con la zona urbana o que en los municipios con valores muy bajos de los afiliados en el sector servicios, un incremento en las prestaciones sobre la RBD hace que la probabilidad de ser rural aumente mientras que la de ser semiurbano disminuya.

La metodología empleada en este trabajo permite extraer conclusiones a partir de las variables directamente consideradas en el modelo, evitando así el uso de técnicas de reducción de la dimensión (que van en detrimento de la interpretabilidad del modelo) como se ha hecho en otros estudios de características similares (Iglesias *et al*, 2000; Rúa *et al*, 2003; Pena *et al*, 2008).

El modelo también permite concluir, al igual que en otros estudios, que la presencia de actividades del sector terciario y los altos niveles de renta, unidos al dinamismo demográfico, parecen ser los elementos determinantes de las disparidades intrarregionales de Galicia (Pena *et al*, 2008).

Por otra parte, teniendo en cuenta los problemas que evidencian las variables utilizadas en el modelo multinomial para diferenciar bien las categorías de urbano y semiurbano, se plantea como una posible mejora de este trabajo el uso de distintas variables para modelar cada una de las categorías multinomiales (rural y semiurbano).

REFERENCIAS BIBLIOGRÁFICAS

- CALVO, M.E., IGLESIAS, C.L., LÓPEZ, M.E., SANTIAGO, M.I., SILVEIRA, S., VILLAR, E. (2013). "El grado de urbanización en Galicia: diferencias socioeconómicas entre las distintas zonas". *Revista Galega de Economía*, vol. 22, núm. 2, pp. 57-78.
- CROISSANT, Y. (2011). *mlogit: multinomiallogitmodel*. R package version 0.2-1. <http://CRAN.R-project.org/package=mlogit>
- EUROSTAT (2007). "Task Force on Core Social Variables". http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-RA-07-006/EN/KS-RA-07-006-EN.PDF
- FARADAY, J.J. (2006). *Extending the Linear Model with R. Generalized Linear, Mixed Effects and Nonparametric Regression Models*. Chapman & Hall/CRC
- HOSMER D.W., LEMESHOW S. (2000). *Applied Logistic Regression (second edition)*. Wiley Series in Probability and Statistics.
- IGLESIAS, C., LÓPEZ, M.E. e SÁNCHEZ, P. (2000). "Dimensionalidad de la capacidad económica en las comarcas gallegas". *Revista Galega de Economía*, vol. 9, núm. 2, pp. 67-90.
- IGE (2011). "Clasificación del grado de urbanización de las parroquias y de los municipios gallegos". http://www.ige.eu/web/mostrar_paxina.jsp?paxina=003003&idioma=gl
- PRECEDO, A. MÍGUEZ, A. e FERNÁNDEZ, M. I. (2008). "Galicia: el tránsito hacia una sociedad urbana en el contexto de la unión europea". *Revista Galega de Economía*, 17, num. extraord.
- PENA, J. A. e SÁNCHEZ, J.M. (2008). "Disparidades económicas intrarregionales a escala municipal: Evidencia empírica para el caso gallego", *Revista de Estudios Regionales*, núm. 81, pp. 15-43.
- RODRÍGUEZ, R. (1994). "El desarrollo de las villas y pequeñas ciudades en los espacios periféricos agrarios. El caso gallego". *Geographica*, 31, 177-196.
- RÚA, A., REDONDO, R. e CAMPO, C. del (2003). "Distribución municipal de la realidad socioeconómica gallega", *Revista Galega de Economía*, vol. 12, núm. 2, pp. 243-262.
- TORRE, M.P. e LOIS, R. (1995). "Claves para la interpretación del mundo urbano gallego". *Anales de Geografía de la Universidad Complutense*, 15, 731-740.

ANEXO

VARIABLES SOCIODEMGRÁFICAS

Índice sintético de fecundidad o número medio de hijos por mujer: representa el número esperado de hijos por mujer a lo largo de su vida fértil (de 15 a 49 años). Se obtiene como suma de las tasas específicas de fecundidad por grupos de edad.

Índice de envejecimiento: es el cociente entre la población de 65 y más años y la población menor de 20 años, expresado en tanto por cien.

Edad media de la población: es la media aritmética de la variable edad agrupada en edades simples.

Pensiones no contributivas por 1.000 habitantes: es el cociente entre el número de beneficiarios de pensiones no contributivas de la Seguridad Social, de ayudas de emergencia social o de la renta de integración social de Galicia y la población, expresado en tanto por mil.

VARIABLES LABORALES

Paro registrado por 1.000 habitantes: es el cociente entre el número de personas inscritas en las oficinas públicas de empleo como parados y la población de 15 a 64 años, expresado en tanto por mil.

Paro registrado de menores de 25 años por 1.000 habitantes: es el cociente entre el número de personas menores de 25 años inscritas en las oficinas públicas de empleo como parados y la población de 15 a 24 años, expresado en tanto por mil.

Porcentaje de afiliaciones en el sector primario, en la industria y construcción y en los servicios: es el porcentaje que suponen las afiliaciones a la Seguridad Social en la agricultura, ganadería y pesca, en la industria y construcción y en los servicios, respectivamente, sobre el total facilitado por la Tesorería General al IGE. Los datos están referidos a 31 de diciembre.

Contratos iniciales registrados por 1.000 habitantes: es el cociente entre el número de contratos que inician la relación laboral y la población correspondiente, expresado en tanto por mil.

VARIABLES ECONÓMICAS DE RENTAS

Renta bruta disponible (RBD): macromagnitud que mide los ingresos disponibles por los hogares a lo largo de un año para destinar al consumo o al ahorro.

Prestaciones sobre la RBD: razón entre las prestaciones y la RBD, expresada en tanto por cien. Las prestaciones son las transferencias corrientes a los hogares con intervención de un tercero que son objeto de una asignación personal que tiene como fin cubrir las cargas que para los hogares se derivan de la aparición o existencia de ciertos riesgos o necesidades, y sin que exista una contrapartida equivalente y simultánea del beneficiario en la RBD.

Saldo de las rentas primarias bruto por habitante: es el saldo de la Cuenta de asignación de la renta primaria dividido por la población correspondiente. La renta primaria es la que reciben las unidades residentes en virtud de su participación directa en el proceso de producción y la renta a cobrar por el propietario de un activo financiero o de un activo material no producido por ponerlos a disposición de otra unidad institucional. Es un indicador de la capacidad de los hogares residentes de generar renta como empresarios, asalariados o perceptores de rentas de la propiedad.

Remuneración de asalariados sobre la RBD: razón entre la macromagnitud remuneración de asalariados y la renta bruta disponible, expresada en tanto por cien.

Rendimiento medio declarado en el Impuesto de la Renta de las Personas Físicas (IRPF): es el resultado de sumar, en términos generales, los rendimientos reducidos netos del trabajo, del capital mobiliario, capital inmobiliario y de las actividades económicas del modelo 100 de la declaración del IRPF, dividido por el número total de declaraciones presentadas en el ejercicio al que hace referencia la información.

Variables económicas de empresas

Empresas con asalariados por 1.000 habitantes: es el cociente entre el número de unidades jurídicas que proporcionan cobertura legal a las actividades que se realizan en uno o varios locales con uno o más asalariados y la población correspondiente, expresado en tanto por mil. Las unidades jurídicas pueden ser personas físicas o jurídicas con existencia legal independiente de la de sus propietarios. No están incluidas las explotaciones del sector primario.

Empresas sin asalariados por 1.000 habitantes: es el cociente entre las unidades jurídicas antes mencionadas que no disponen de asalariados entre la población correspondiente, expresado en tanto por mil. No están incluidas las explotaciones del sector primario.

Establecimientos en los sectores diferente de los servicios por 1.000 habitantes: es el cociente entre las unidades locales de producción que pertenecen a los sectores de producción diferentes del sector servicios y la población correspondiente, expresado en tanto por mil. No están incluidas las unidades locales del sector primario.

Establecimientos en los servicios por 1.000 habitantes: es el cociente entre las unidades locales de producción de este sector y la población correspondiente, expresado en tanto por mil.

MATRIZ DE CORRELACIONES

1	1	-0,62	-0,65	-0,16	0,32	0,27	-0,38	0,10	0,37	0,23	0,33	-0,52	0,49	0,60	0,43	0,34	0,09	0,40	0,51
2	-0,62	1	0,93 *	0,19	-0,35	-0,43	0,41	-0,13	-0,38	-0,32	-0,47	0,79 *	-0,69	-0,79 *	-0,60	-0,47	-0,15	-0,57	-0,67
3	-0,65	0,93 *	1	0,25	-0,39	-0,42	0,47	-0,13	-0,45	-0,35	-0,50	0,81 *	-0,74	-0,87 *	-0,62	-0,48	-0,13	-0,59	-0,75 *
4	-0,16	0,19	0,25	1	0,23	0,09	0,10	0,06	-0,16	-0,23	-0,44	0,28	-0,46	-0,37	-0,27	-0,23	-0,12	-0,30	-0,38
5	0,32	-0,35	-0,39	0,23	1	0,67	-0,50	0,37	0,30	0,06	-0,15	-0,02	0,01	0,33	0,12	0,06	0,00	0,08	0,16
6	0,27	-0,43	-0,42	0,09	0,67	1	-0,32	0,30	0,15	0,14	-0,12	-0,11	0,06	0,37	0,15	0,03	0,03	0,09	0,20
7	-0,38	0,41	0,47	0,10	-0,50	-0,32	1	-0,52	-0,77 *	-0,30	-0,16	0,25	-0,33	-0,66	-0,37	-0,29	-0,23	-0,28	-0,61
8	0,10	-0,13	-0,13	0,06	0,37	0,30	-0,52	1	-0,14	0,07	-0,21	0,05	-0,10	0,24	0,11	-0,01	0,44	-0,16	0,05
9	0,37	-0,38	-0,45	-0,16	0,30	0,15	-0,77 *	-0,14	1	0,29	0,34	-0,32	0,45	0,58	0,35	0,35	-0,06	0,44	0,67
10	0,23	-0,32	-0,35	-0,23	0,06	0,14	-0,30	0,07	0,29	1	0,28	-0,31	0,35	0,38	0,48	0,31	0,29	0,46	0,40
11	0,33	-0,47	-0,50	-0,44	-0,15	-0,12	-0,16	-0,21	0,34	0,28	1	-0,71	0,93 *	0,52	0,56	0,61	0,14	0,68	0,70
12	-0,52	0,79 *	0,81 *	0,28	-0,02	-0,11	0,25	0,05	-0,32	-0,31	-0,71	1	-0,84 *	-0,67	-0,67	-0,63	-0,24	-0,66	-0,67
13	0,49	-0,69	-0,74	-0,46	0,01	0,06	-0,33	-0,10	0,45	0,35	0,93 *	-0,84 *	1	0,76 *	0,65	0,63	0,16	0,71	0,85 *
14	0,60	-0,79 *	-0,87 *	-0,87	0,33	0,37	-0,66	0,24	0,58	0,38	0,52	-0,67	0,76 *	1	0,56	0,41	0,13	0,49	0,88 *
15	0,43	-0,60	-0,62	-0,27	0,12	0,15	-0,37	0,11	0,35	0,48	0,56	-0,67	0,65	0,56	1	0,73	0,61	0,84 *	0,52
16	0,34	-0,47	-0,48	-0,23	0,06	0,03	-0,29	-0,01	0,35	0,31	0,61	-0,63	0,63	0,41	0,73	1	0,57	0,81 *	0,43
17	0,09	-0,15	-0,13	-0,12	0,00	0,03	-0,23	0,44	-0,06	0,29	0,14	-0,24	0,16	0,13	0,61	0,57	1	0,33	0,05
18	0,40	-0,57	-0,59	-0,30	0,08	0,09	-0,28	-0,16	0,44	0,46	0,68	-0,66	0,71	0,49	0,84 *	0,81 *	0,33	1	0,54
19	0,51	-0,67	-0,75 *	-0,38	0,16	0,20	-0,61	0,05	0,67	0,40	0,70	-0,67	0,85 *	0,88 *	0,52	0,43	0,05	0,54	1

Nota: el número de la variable se corresponde con el número de orden del Cuadro 1

