

Variables determinantes de la hipoteca en España

Ana Tian Villanueva (100405817), Francisco Ortiz Kuefler (100405935)

INTRODUCCIÓN Y OBJETIVOS

Este documento tiene como objetivo estudiar la influencia de la población, la riqueza, el número de compras y ventas de viviendas, y el hecho de tener o no costa de las provincias españolas, en el número y valor medio de las hipotecas, y su capacidad explicativa sobre éstos. A priori es de esperar que el número de hipotecas sea casi completamente explicado por estas variables, y el valor medio de las hipotecas por la riqueza y la demanda, representada aquí por la población.

Para ello se emplean datos del Instituto Nacional de Estadística sobre las hipotecas, el PIB, los traspasos de viviendas y la población. Ceuta y Melilla se han excluido de este análisis por ser ciudades sin capacidad de expansión y que, por lo tanto, presentan datos completamente atípicos en comparación con el resto de provincias.

A partir de los datos obtenidos del INE, se han calculado el capital hipotecario medio, la densidad poblacional y el PIB per cápita mediante los correspondientes cocientes. El capital hipotecario medio y el número de hipotecas serán las variables principales. Se espera que la densidad poblacional ayude a explicar la presencia de las grandes ciudades, y que el acceso a la costa lo haga para la población costera y las viviendas de verano.

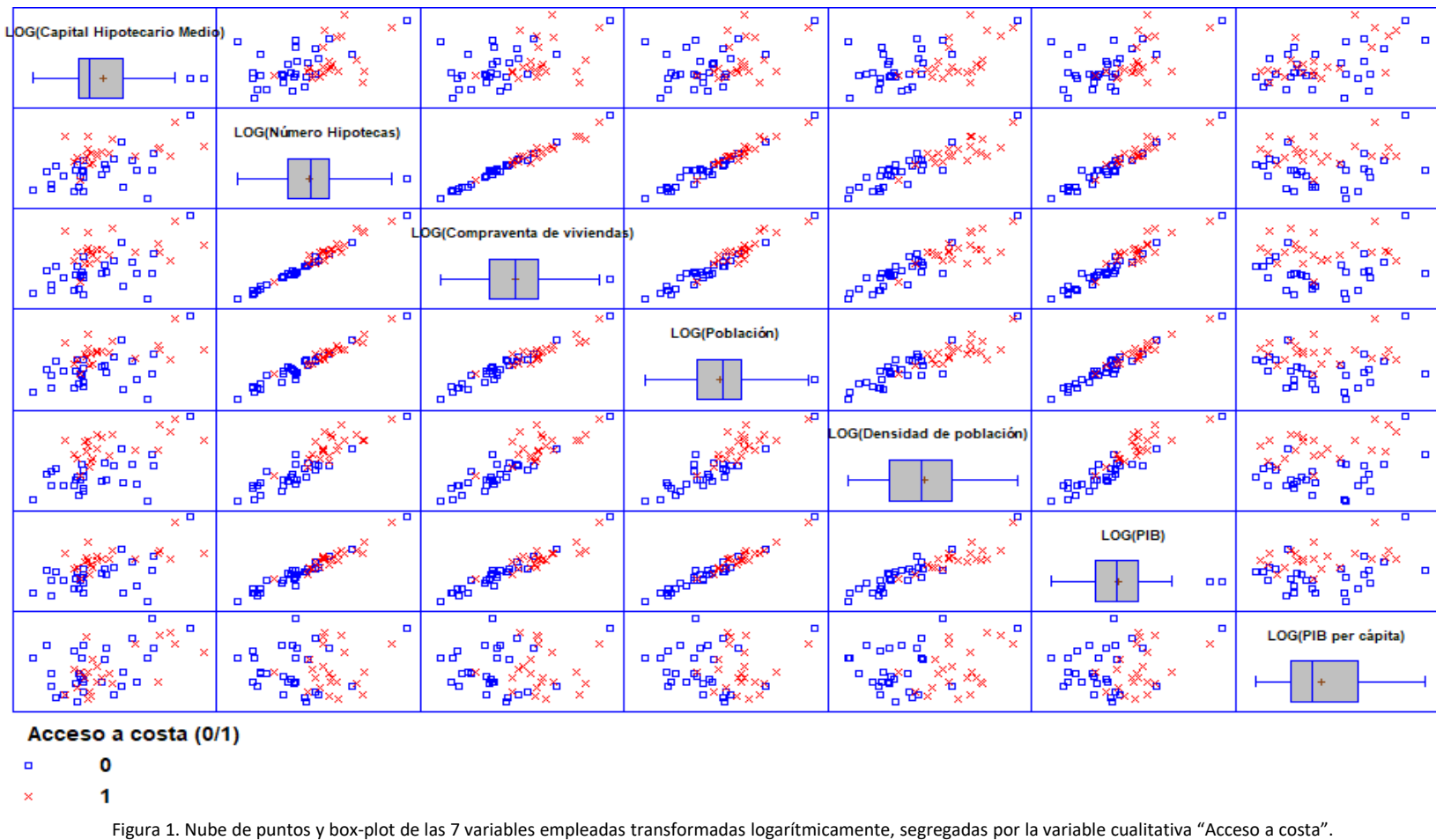
Nota: en todos los análisis se empleará un nivel de significación del 5%, i.e. $\alpha = 0.05$.

ESTUDIO DESCRIPTIVO

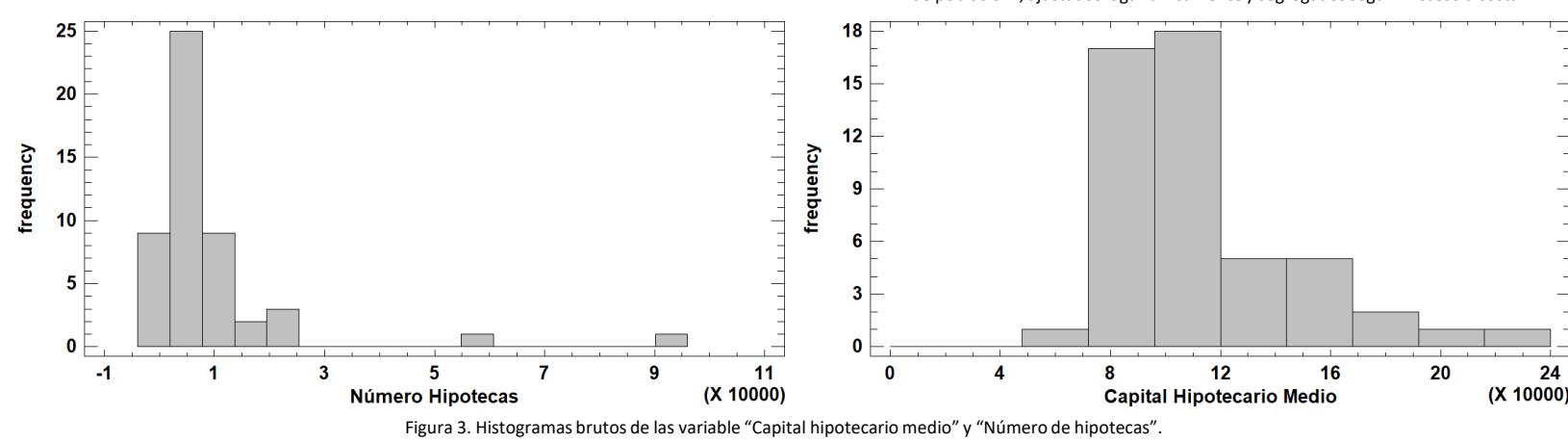
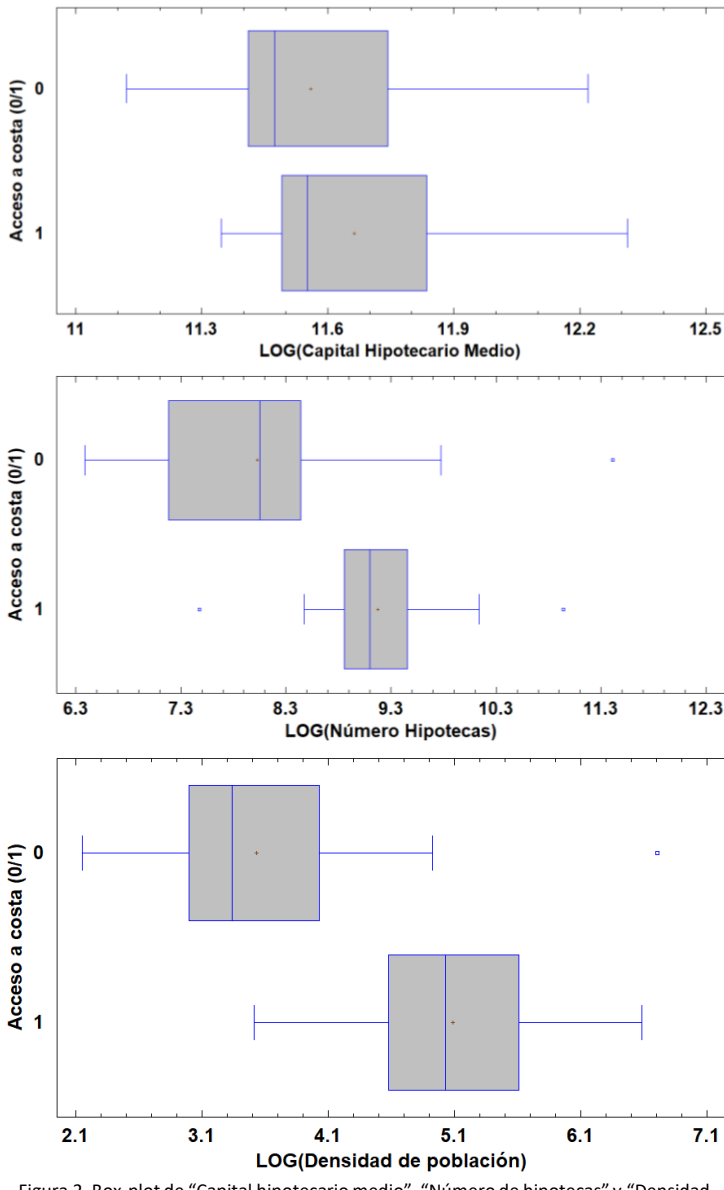
La Tabla 1 muestra las medidas resumen de las variables principales elegidas, “Capital hipotecario medio por provincias” y “Número de hipotecas por provincias”.

Las características más notables de estas variables son el amplio rango de valores y una asimetría ligera en la primera y muy notable en la segunda. Esta asimetría es compartida con todas las otras variables numéricas empleadas y, como se ve en el apartado *Ajuste de distribuciones*, se puede corregir mayoritariamente mediante una transformación logarítmica.

Para mejorar la visualización, se ha aplicado ya dicha transformación en el análisis multivariante resumido en la Figura 1, mostrada a continuación:



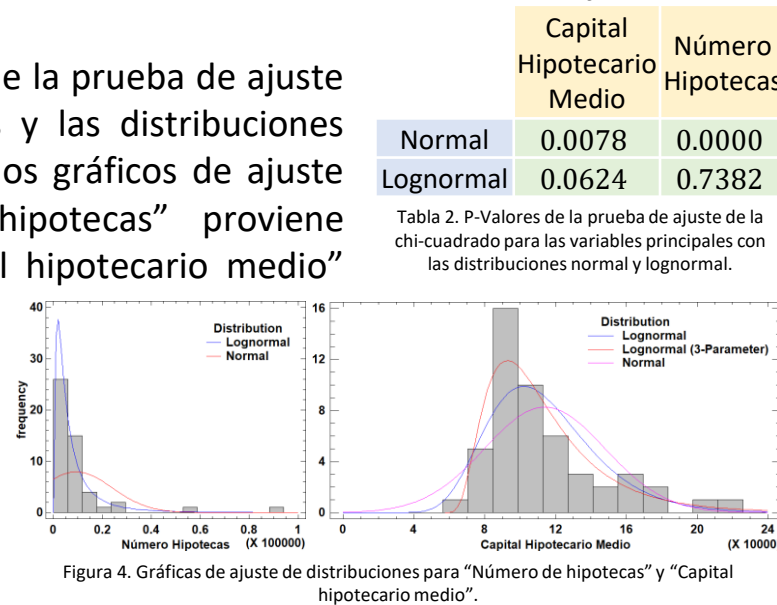
Así se ve claramente que el número de hipotecas guarda una relación lineal ajustada clara con las variables explicativas, a excepción del PIB per cápita, mientras que en el caso del capital hipotecario medio la relación no es tan evidente. También se observa en los box-plots que, a pesar de la corrección de la asimetría, existen uno o dos datos atípicos en cada variable, correspondientes a las provincias de Madrid y Barcelona, porque sus capitales son ciudades muy grandes a nivel incluso Europeo y su entorno urbano es muy distinto al del resto de provincias españolas. La segregación de provincias con acceso a costa también resulta ser un factor importante para casi todas las variables relacionadas, con las provincias costeras consiguiendo valores generalmente más elevados, con la notable excepción de Madrid. En la Figura 2 se observa este efecto de manera individual para el capital medio hipotecario, el número de hipotecas y la densidad de población. Cabe destacar, sin embargo, que esta división no tiene una consecuencia notable en la distribución general de la población, como se ve en la Figura 3. Este efecto se estudia en el apartado *Intervalos de confianza y contrastes de hipótesis*.



AJUSTE DE DISTRIBUCIONES

Como se muestra en el análisis descriptivo, los datos tienen una alta asimetría positiva. Debido a ello, es probable que sean muy poco normales y será necesario otro tipo de distribución para mejorar el ajuste y así, compararlos en igualdad de condiciones. En este caso, la distribución más ajustada es la distribución lognormal.

En la Tabla 2 se muestran los p-valores resultantes de la prueba de ajuste por la chi-cuadrado entre las variables principales y las distribuciones normal y lognormal, y en la Figura 4 se muestran los gráficos de ajuste correspondientes. Mientras que “Número de hipotecas” proviene claramente de una distribución lognormal, “Capital hipotecario medio” tiene un ajuste menos significativo. Utilizando una distribución lognormal ajustada por constante (“triparámetro”) se alcanza un p-valor mucho mayor, de 0.3435, pero por simplicidad se utilizará la distribución lognormal ordinaria.



INTERVALOS DE CONFIANZA Y CONTRASTE DE HIPÓTESIS

Con el objetivo de ver si realmente el acceso a costa es un factor diferenciador para las poblaciones se hace un contraste de hipótesis. En primer lugar, la Tabla 3 muestra los intervalos de confianza para la media y la desviación típica en las variables principales normalizadas, íntegras y separadas según acceso a costa.

A continuación se realiza un contraste de hipótesis para cada población, comparando la media y la desviación típica de las muestras con y sin costa. La hipótesis planteada desde el análisis descriptivo es que el acceso a costa aumenta los valores de la muestra. Los resultados de este contraste se muestran en la Tabla 4.

Capital Hipotecario Medio	Acceso a costa	Media		Desviación típica	
		$\hat{\mu}$	$IC_{95\%}$	$\hat{\sigma}$	$IC_{95\%}$
0	0	11.6058	(11.5293, 11.6822)	0.2690	(0.2247, 0.3353)
	1	11.4576	(11.4576, 11.6630)	0.2093	(0.2093, 0.3604)
Número Hipotecas	0	8.5315	(8.2308, 8.8322)	1.0581	(0.8838, 1.3185)
	1	8.5315	(8.2308, 8.8322)	1.0581	(0.8838, 1.3185)

Tabla 3. Intervalos de confianza para la media y la desviación típica de "log(Capital Hipotecario medio)" y "log(Número de hipotecas)" al 95% de confianza, con separación por "Acceso a costa".

Contrastes de Hipótesis de "Acceso a costa (0/1)"			
Capital Hipotecario Medio	Número Hipotecas	$H_0: \mu_0 = \mu_1$ $H_1: \mu_0 < \mu_1$	$H_0: \sigma_0 = \sigma_1$ $H_1: \sigma_0 \neq \sigma_1$

Tabla 4. Planteamiento y resultado del contraste de hipótesis de "Acceso a costa" para las variables principales normalizadas.

En ambas variables el contraste de las desviaciones estándares indica que no hay suficiente información como para rechazar la igualdad, por lo que se justifica la suposición de varianzas iguales en el contraste de medias. En el caso del número de hipotecas, tanto los intervalos de confianza como el contraste de las medias confirman la hipótesis inicial, que el acceso a costa produce un aumento de valores, puesto que los intervalos tienen intersección nula y el p-valor del contraste es cero. Sin embargo, en el caso del capital hipotecario medio, en contra de la hipótesis y siguiendo lo observado en la Figura 1, los intervalos de confianza de la media resultan prácticamente idénticos y el p-valor del contraste nos indica que hay mayor evidencia estadística a favor de la igualdad de medias. Esto implica que el acceso a costa posiblemente no tenga capacidad explicativa sobre el valor medio de las hipotecas.

REGRESIÓN MÚLTIPLE

En la Tabla 5 se muestran los resultados de una regresión múltiple empleando todas las variables explicativas, normalizadas, para explicar las dos variables principales, también normalizadas. “PIB per cápita” no se utiliza porque es una combinación lineal de “PIB” y “Población”, y sustituir “PIB” por “PIB per cápita” resulta en el mismo nivel de significación por la inclusión del dato de “Población”.

Parámetro	Número de hipotecas, $R_{adj}^2 = 98.58\%$				Capital Hipotecario Medio, $R_{adj}^2 = 38.96\%$			
	Estimación	Error estándar	Estadístico T	P-Valor	Estimación	Error estándar	Estadístico T	P-Valor
CONSTANTE	-3.7183	1.4437	-2.5755	0.0134	5.0920	2.4107	2.1122	0.0404
Compraventa de viviendas	0.6305	0.0670	9.4040	0.0000	0.0766	0.1120	0.6842	0.4974
Población	0.2353	0.1139	2.0650	0.0449	-0.5857	0.1903	-3.0782	0.0036
Densidad de población	0.0902	0.0464	1.9434	0.0584	0.0688	0.0775	0.8877	0.3795
PIB	0.1421	0.1094	1.2992	0.2007	0.5723	0.1826	3.1340	0.0031
Acceso a costa (0/1)	-0.1211	0.0545	-2.2238	0.0313	-0.0766	0.0909	-0.8426	0.4040

Tabla 5. Resultados de la regresión múltiple para las variables principales, con "Compraventa de viviendas", "Población", "Densidad de población", "PIB" y "Acceso a costa" como variables explicativas. Todas las variables están normalizadas, excepto "Acceso a costa".

En el caso del número de hipotecas se logra explicar el 99% de la varianza, por lo que el ajuste es muy bueno. La variable menos significativa, con un p-valor de 0.2, es el PIB, lo cual puede deberse a que la población más adinerada tiene más facilidades para adquirir viviendas y, por lo tanto, no requiere hipotecar dicha compra. Eliminando el PIB deja a todo el resto de variables explicativas con un p-valor muy inferior al 5% (Tabla 6), indicando que tienen un poder explicativo real. Una nota interesante es que eliminando o la densidad poblacional o el acceso a costa hace que la otra variable pierda su significación estadística, con p-valores de 0.07 y 0.15, respectivamente. Proponemos que esto se deba a que la característica importante que explican es la presencia de grandes ciudades costeras, con alta demanda de viviendas, en vez de tan sólo el acceso a costa o la densidad media de la población, y que en las provincias sin acceso a costa, que con la excepción de Madrid suelen tener poblaciones menores, la densidad de población es un factor de menor relevancia.

El caso del capital hipotecario medio es muy distinto. Las capacidad explicativa del modelo es tan sólo el 39% de la varianza de la variable respuesta, y tres de las cinco variables empleadas tienen p-valores muy altos, indicando su falta de capacidad explicativa. En este caso, no existe el mismo efecto entre la densidad de población y el acceso a costa que en el número de hipotecas, por lo que el valor medio de las hipotecas podría no ser afectado por la presencia de las grandes ciudades costeras. También pierde capacidad explicativa la compraventa de viviendas, indicando que, quizás, el número de transacciones no refleje realmente la demanda, suponiendo que esta es un factor principal en el precio de las viviendas. Puede ser que la capacidad económica de los que adquieren hipotecas es más relevante que el valor de las viviendas que hipotecan. En todo caso, eliminando estas tres variables deja un modelo con una capacidad explicativa ligeramente mayor, como se ve en la Tabla 6.

Parámetro	Número de hipotecas, $R_{adj}^2 = 98.56\%$				Capital Hipotecario Medio, $R_{adj}^2 = 41.09\%$			
	Estimación	Error estándar	Estadístico T	P-Valor	Estimación	Error estándar	Estadístico T	P-Valor
CONSTANTE	-1.9912	0.5674	-3.5094	0.0010	3.0981	1.5613	1.9843	0.0531
Compraventa de viviendas	0.6383	0.0673	9.4862	0.0000				
Población	0.3408	0.0805	4.2310	0.0001	-0.5495	0.1541	-3.5658	0.0008
Densidad de población	0.1226	0.0395	3.1066	0.0033				
PIB					0.6764	0.1497	4.5175	0.0000
Acceso a costa (0/1)	-0.1474	0.0509	-2.8937	0.0059				

Tabla 6. Resultados de la regresión múltiple simplificada para las variables principales, con "Compraventa de viviendas", "Población", "Densidad de población", "PIB" y "Acceso a costa" como variables explicativas. Las celdas coloreadas en naranja/marón corresponden a las variables desestimadas para los modelos. Todas las variables están normalizadas, excepto "Acceso a costa".

Los modelos de regresión finales quedan como sigue:

$$\begin{aligned} \text{Número de hipotecas: } Y &= \exp(0.6383 \cdot \log CV + 0.3408 \cdot \log P + 0.1226 \cdot \log DP - 0.1474 \cdot AC - 1.9912) \\ \text{Capital hipotecario medio: } Y &= \exp(0.6764 \cdot \log PIB - 0.5495 \cdot \log P + 3.0981) \end{aligned}$$

donde CV es “Compraventa de viviendas”, P es “Población”, DP es “Densidad de población”, PIB es “PIB” y AC es “Acceso a costa”.

Las gráficas de residuos no muestran patrones claros, indicando que la varianza residual podría ser simplemente ruido. No obstante, hay algunos residuos son muy atípicos, indicando que quizás haya factores que falten del modelo.

A continuación se muestran predicciones sobre el valor de las variables principales en una provincia ficticia con $CV = 20,000$, $P = 1,500,000$, $DP = 400$, $PIB = 4 \cdot 10^{10}$ y $AC = 1$:

$$\begin{aligned} \text{Número de hipotecas: } Y &= 17,393.29, \text{ error estándar } 2,455.66. \\ \text{Capital Hipotecario Medio: } Y &= 132,574.50, \text{ error estándar } 31,190.87. \end{aligned}$$

Estos resultados son consistentes en comparación con una provincia de características similares, como Vizcaya.

CONCLUSIONES

En este estudio se han analizado el valor medio de las hipotecas y el número de hipotecas adquiridas en las provincias españolas, y el poder explicativo de la población, la riqueza, el número de adquisiciones de vivienda y el acceso a costa sobre estas variables. Una hipótesis inicial era que el número de hipotecas está casi completamente explicado por estas variables, y la conclusión es que así es, exceptuando la riqueza, que aparentemente no juega un papel relevante. Otra era que el valor medio de las hipotecas dependería mayoritariamente de la riqueza y de la población, y se concluye que entre las variables elegidas así es, pero hay una parte importante de su varianza sin explicar, posiblemente por factores de inmigración u otros. Finalmente, se esperaba que la densidad poblacional ayudaría a explicar la presencia de las grandes ciudades, y se concluye que lo hace mucho mejor cuando se aporta información sobre el acceso a costa, aludiendo a la mayor concentración de la población en esas zonas.