

¿Influye la despoblación sobre el rendimiento académico? Datos de PISA 2018¹

El pasado diciembre, la **Unión Europea**, en el reglamento de los Fondos Europeos de Desarrollo Regional (FEDER) para el periodo 2021-2027, **modificó la definición de zona despoblada**, analizando, por primera vez, esta cuestión demográfica por provincias y municipios o agrupaciones de municipios. Hasta entonces, la despoblación se medía a nivel de NUTS 2, región equivalente en España a comunidad autónoma. Este hecho perjudicaba a los municipios despoblados españoles, pues el crecimiento de las ciudades de una comunidad solía compensar la reducción demográfica que están viviendo muchos enclaves rurales.

En resumen, la **nueva definición considera como áreas escasamente pobladas** aquellas entidades locales que cumplan al menos uno de estos **dos requisitos**:

- A. Que, de media, haya perdido más de un 1 % anual de población entre 2007 y 2017.
- B. Que tenga una densidad de población inferior a 12,5 hab/km².

Aprovechando este cambio de definición, en este trabajo se propone **analizar los posibles efectos de la despoblación en el rendimiento académico del alumnado que viva en zonas consideradas despobladas según la anterior definición**. Para ello, por un lado, se han empleado los **datos de rendimiento por centro educativo en lectura, matemáticas, ciencias y competencia global**, obtenidos en la última edición de la evaluación **PISA (2018)**. Por otro lado, los datos para calcular la condición de despoblada o no de cada entidad local han sido obtenidos de la página web del Instituto Nacional de Estadística (INE), que cuenta con las cifras del Padrón Continuo, y del Registro de Entidades Locales, gestionado por el Ministerio de Asuntos Económicos y Transformación Digital, con información sobre la superficie de las más de 8000 entidades locales españolas. Los análisis se han llevado a cabo haciendo uso del software de estadística espacial [ArcGIS](#).



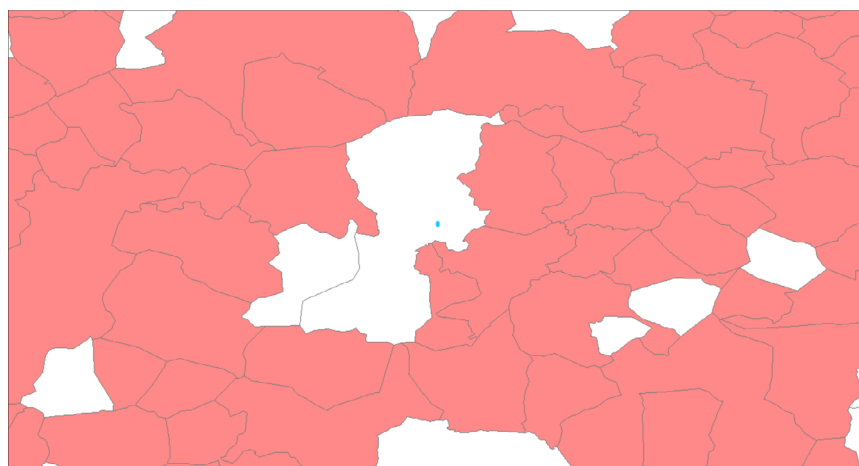
Primeros hallazgos a partir de los requisitos definidos

El análisis inicial realizado con los dos requisitos indicados previamente ha dado como resultado que al superponer la capa de centros a la de España con datos de despoblación, **se ha observado que la gran mayoría de institutos estaban localizados en municipios no despoblados** (FIGURA 1). La pequeña muestra de centros en contextos despoblados tenía como consecuencia la no significatividad estadística de la variable *despoblación* en nuestros análisis.

¹ Autores: Alejandro Díaz-Guerra Romero y Daniel Reinhardt Hervás

² La densidad de población que se ha utilizado ha sido la de 2017

FIGURA 1. Municipio no despoblado rodeado de despoblados



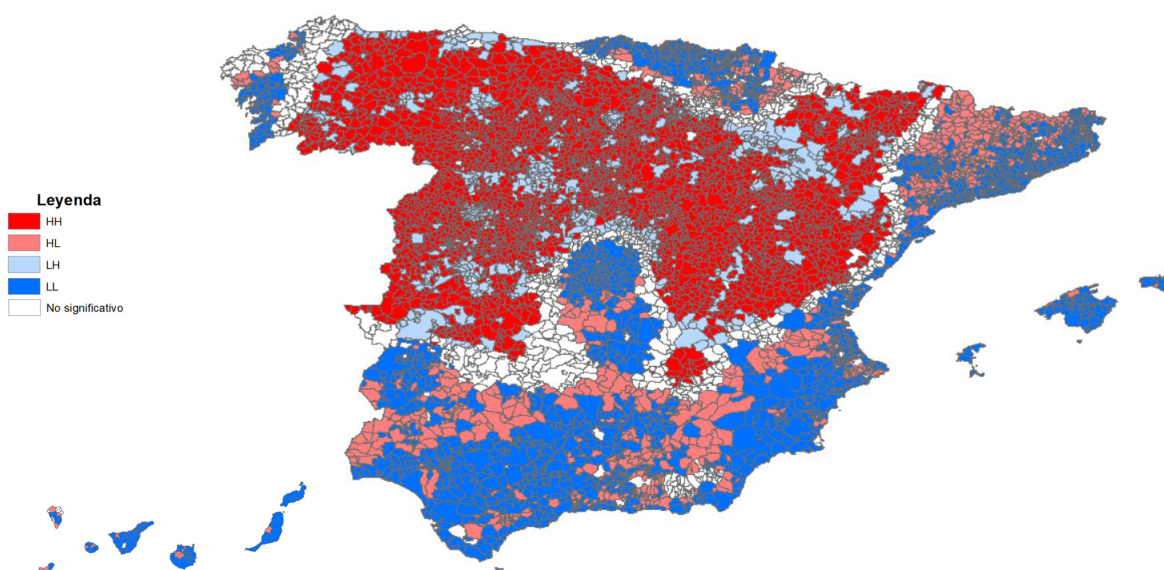
Por otra parte, se encontraban muchos casos como el de la figura anterior: centros (puntos azules) situados en municipios no despoblados (en blanco) rodeados de municipios despoblados (en rojo). Entonces, con el fin de aumentar la muestra de institutos en área despoblada, se ha planteado la siguiente **doble hipótesis**: 1) dichos centros hacen de foco de atracción a familias con hijos de municipios cercanos, contribuyendo a la no despoblación del municipio donde se encuentra el instituto, y a la despoblación de los municipios de alrededor, y 2) dichos centros, muy probablemente, cuenten con alumnos de municipios colindantes que sí se encuentran despoblados.

Por tanto, **con el fin de aumentar la muestra de centros enclavados en zonas despobladas**, se estableció una tercera condición **para poder considerar a una entidad local despoblada**:

C. Aquellos municipios no despoblados, según las dos condiciones anteriores, pero que están rodeados de municipios despoblados, son considerados también despoblados.

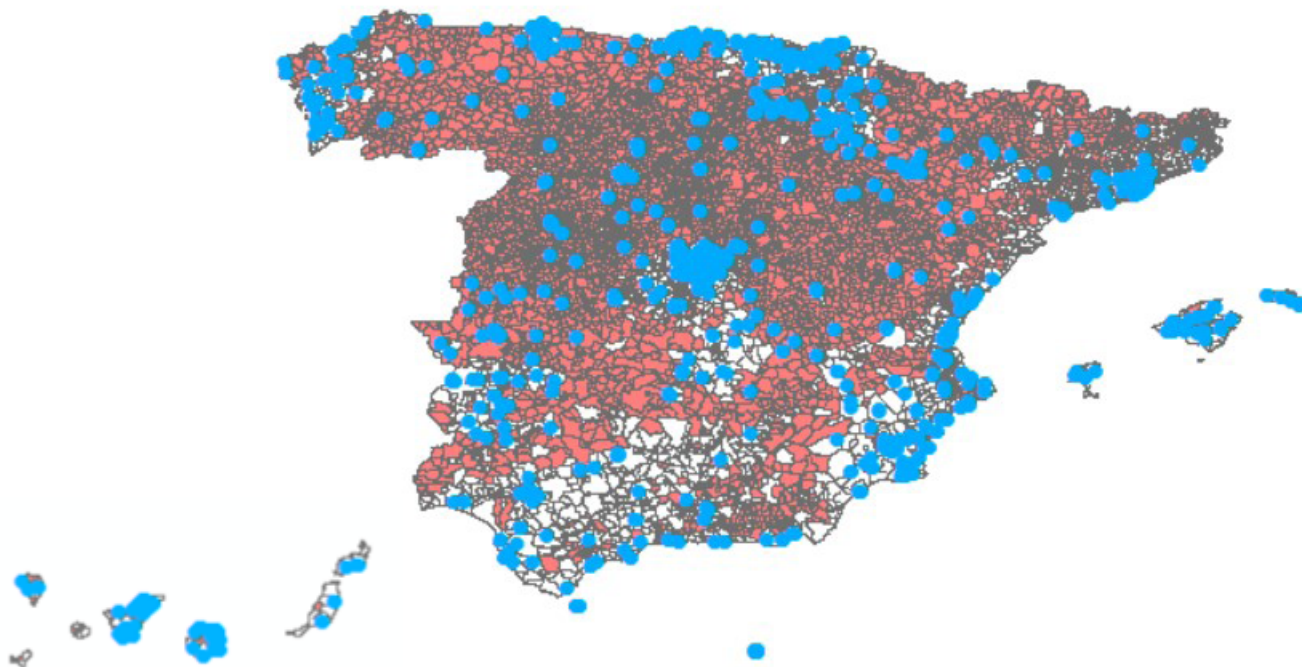
Para poder determinar con rigurosidad estadística qué municipios cumplen la anterior condición, se ha llevado a cabo un análisis de conglomerados y *outliers* (valores atípicos) espaciales. Al estar codificada la variable *despoblación* de la siguiente manera: 1 en caso de despoblado, 0 en caso de no despoblado, los municipios que cumplen el requisito tres son considerados *outliers Low-High* (LH, un valor bajo rodeado de altos) (FIGURA 2).

FIGURA 2. Análisis de conglomerados y *outliers* espaciales



Tras actualizar la variable *despoblación* con los anteriores resultados, los análisis estadísticos pudieron realizarse sin problemas. Las **capas finales sobre las que se llevaron a cabo dichos análisis** se pueden ver en la siguiente figura:

FIGURA 3. Capa de España con las entidades locales consideradas despobladas (en rojo), y capa de institutos que aplicaron PISA en 2018 (puntos azules)



¿Afecta estudiar en un municipio despoblado al rendimiento del alumnado?

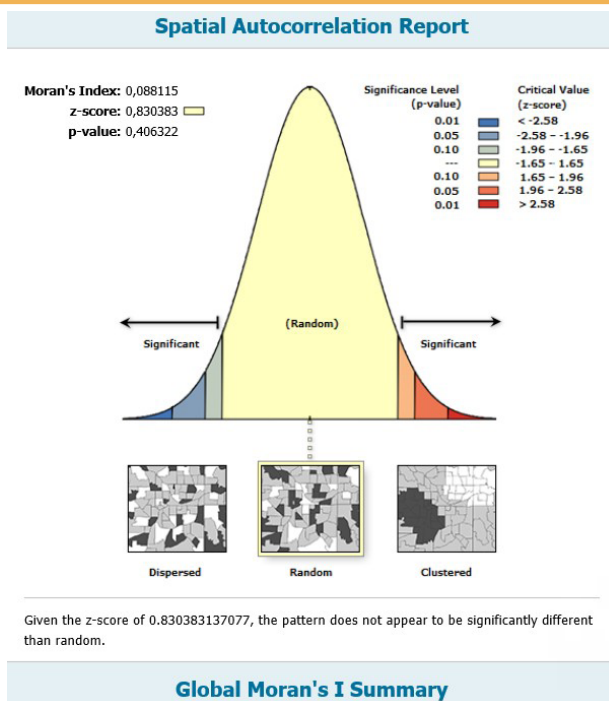
Para lograr dar respuesta a la pregunta planteada, se han realizado **modelos regresión utilizando como variables dependientes las variables de rendimiento en PISA** (una regresión por cada rendimiento) y, **como variables explicativas o independientes**, principalmente: la **despoblación** (variable *DESP_FIN*), el **nivel socioeconómico y cultural del alumno** (*ESCS*), el **tamaño del centro** (*SCHSIZE*) y el **abandono escolar** (*STUDENT_TR*).

En primer lugar, se han llevado a cabo regresiones por Mínimos Cuadrados Ordinarios (*OLS*, por sus siglas en inglés). Al realizar una *OLS*, *ArcGIS* aporta información sobre una serie de contrastes estadísticos, como el test de Koenker. Si este test es estadísticamente significativo (*p*-valor inferior a los niveles típicos de significación: 10 %, 5 % y 1 %) comporta que pueden existir problemas causados por la existencia de heterocedasticidad (varianza desigual) o de autocorrelación espacial.

La autocorrelación espacial implica que las cosas espacialmente más cercanas tienen más relación que con las más lejanas - algo que puede parecer de sentido común, pero que necesita ser probado estadísticamente. Lo anterior es importante, pues la existencia de autocorrelación espacial es un problema a la hora de llevar a cabo una regresión, tipo de análisis que supone la independencia de las observaciones. Si hay autocorrelación espacial, las observaciones no son independientes. La falta de independencia lleva a que los estimadores obtenidos por *OLS* no sean eficientes, es decir, no tengan varianza mínima. La solución pasaría entonces por solucionar el problema de autocorrelación o usar métodos robustos. Si el test de Koenker es significativo, debe realizarse un test de Moran global sobre los residuos estandarizados para ver si, efectivamente, existen problemas de autocorrelación espacial.

Los resultados de las OLS para matemáticas, ciencias y competencia global arrojaron contrastes de Koenker significativos, por lo que se llevaron a cabo los análisis de autocorrelación espacial anteriormente citados. **En ningún caso se descubrió la existencia de autocorrelación espacial estadísticamente significativa**, como podemos observar para el caso del rendimiento en competencia global:

FIGURA 4. Análisis de autocorrelación espacial de los errores estandarizados de la OLS final para el rendimiento en competencia global



No obstante, como **el test de Koenker resultó ser significativo para esos tres rendimientos**, es necesario atender a los p-valores robustos para **comprobar la significancia de los coeficientes de las distintas variables explicativas**. Si hubiera habido problemas de autocorrelación espacial, una posible solución habría sido realizar una Regresión Geográficamente Ponderada (GWR, también por sus siglas en inglés). La GWR tiene en cuenta las coordenadas de las observaciones, realizando una regresión por cada observación, en vez de una sola regresión para toda la muestra. De esta forma, se consigue un mejor ajuste global, pudiendo ver en qué zonas ajusta mejor y en cuáles peor el modelo, además de solucionar el problema de autocorrelación espacial. **No se encontró este último problema**, por ello, **se presentan** a continuación los **resultados de los últimos modelos de regresión estimados para cada variable de rendimiento**:

FIGURA 5. Resultados de las OLS finales para los distintos rendimientos

LECTURA								
Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	482,921962	2,888390	167,194151	0,000000*	2,772946	174,154859	0,000000*	-----
SCHSIZE	0,005536	0,002279	2,429283	0,015283*	0,002191	2,527123	0,011637*	1,145157
STUDENT_TR	-4,267653	1,317345	-3,239586	0,001250*	1,247245	-3,421662	0,000662*	1,348026
ESCS	39,456544	2,184642	18,060875	0,000000*	2,297466	17,173938	0,000000*	1,478536
DESP_FIN	11,934084	2,349008	5,080477	0,000001*	2,108306	5,660509	0,000000*	1,007530

MATEMÁTICAS								
Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	495,657088	2,235147	221,755934	0,000000*	2,249544	220,336703	0,000000*	-----
SCHSIZE	0,005333	0,001764	3,024135	0,002566*	0,001716	3,108260	0,001946*	1,145157
STUDENT_TR	-6,026343	1,019412	-5,911586	0,000000*	1,003958	-6,002583	0,000000*	1,348026
ESCS	44,052796	1,690559	26,058117	0,000000*	1,769242	24,899243	0,000000*	1,478536
DESP_FIN	9,982929	1,817753	5,491908	0,000000*	1,691429	5,902068	0,000000*	1,007530

CIENCIAS

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	494,135954	2,361190	209,274094	0,000000*	2,361150	209,277651	0,000000*	-----
SCHSIZE	0,003661	0,001863	1,964870	0,049689*	0,001796	2,037922	0,041800*	1,145157
STUDENT_TR	-4,927428	1,076898	-4,575573	0,000007*	1,062215	-4,638824	0,000006*	1,348026
ESCS	38,848211	1,785893	21,752824	0,000000*	1,894474	20,506065	0,000000*	1,478536
DESP_FIN	10,565403	1,920258	5,502073	0,000000*	1,748286	6,043293	0,000000*	1,007530

COMPETENCIA GLOBAL

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	521,898311	2,620468	199,162241	0,000000*	2,582357	202,101573	0,000000*	-----
SCHSIZE	0,006039	0,002068	2,920988	0,003572*	0,002024	2,983721	0,002924*	1,145157
STUDENT_TR	-4,792662	1,195151	-4,010090	0,000073*	1,145454	-4,184071	0,000036*	1,348026
ESCS	42,786680	1,981998	21,587648	0,000000*	2,065925	20,710667	0,000000*	1,478536
DESP_FIN	10,252267	2,131118	4,810745	0,000003*	1,910607	5,365974	0,000000*	1,007530

Se puede observar que, en todos los casos, el signo de los coeficientes es el mismo y su valor similar. **Todas las variables son significativas:** la columna *Robust_Pr[b]* muestra p-valores robustos menores a 0,05 en todos los casos.

Como se esperaba, el valor del coeficiente del ESCS es positivo y elevado, y el del abandono escolar es negativo. Sin embargo, **en contra de lo que se podría creer, el hecho de estar despoblado el municipio en el que se encuentra el centro contribuye positivamente al rendimiento del alumno.** Lo mismo ocurre con el **tamaño del centro: a mayor tamaño, mejor rendimiento, aunque su efecto sea mínimo.**

Como la variable *despoblación* es dicotómica y los municipios despoblados tienen un valor de 1 y los no despoblados, de 0, por ejemplo, para el caso de matemáticas, **el rendimiento medio del centro aumenta en casi 10 puntos si se trata de un instituto de un municipio despoblado.**

La variable *tamaño del centro*, al igual que las variables *ratio alumnos-profesor* y *tamaño medio de la clase* – estas dos últimas no resultaron ser, en ningún caso, estadísticamente significativas al introducirlas en los análisis – contribuyen positiva, pero de forma mínima, al rendimiento del centro. Por ello, no es posible explicar el coeficiente positivo que presenta la variable *despoblación* como consecuencia de que, en general, **las clases en las áreas rurales son más pequeñas, porque tienen un menor número de estudiantes y, por tanto, puede haber una dedicación al estudiante más personalizada.**

También es reseñable que la variable *despoblación* no deja de ser significativa pese a controlar por el nivel socioeconómico y cultural, por lo que estas dos variables no tienen excesiva relación entre ellas. Algo que también podemos ver al atender a los valores VIF: ninguno de ellos es superior a 10 por lo que no existen problemas de multicolinealidad.

Asimismo, el valor del R^2 ajustado oscila entre un 0,37, en el caso del modelo para el rendimiento en lectura, y 0,56, en el caso del rendimiento en matemáticas. El principal contribuyente a dicho ajuste es el ESCS (en los modelos en los que el ESCS es el único regresor, el R^2 ajustado varía entre 0,35, para el caso de lectura y 0,53, para el caso de matemáticas). Esto es un indicativo de que el resto de variables, aunque estadísticamente significativas, explican un escaso porcentaje de varianza.

Con todo, **los resultados obtenidos parecen contradecir la creencia común de que la despoblación afecta negativamente al rendimiento académico.** Quizá lo obtenido pueda ser consecuencia de que la variable *despoblación* se haya creado de manera dicotómica, habiendo posibles grados de despoblación que se pierden al realizar esto. De todas formas, **este análisis muestra la potencialidad de hacer uso de herramientas y análisis de estadística espacial para estudiar las evaluaciones educativas.**

Descargo de responsabilidad: las opiniones expresadas y los argumentos utilizados en este documento no reflejan necesariamente las opiniones oficiales del INEE.



MINISTERIO DE EDUCACIÓN Y FORMACIÓN PROFESIONAL

SECRETARÍA DE ESTADO DE EDUCACIÓN

DIRECCIÓN GENERAL DE EVALUACIÓN Y COOPERACIÓN TERRITORIAL

inee Instituto Nacional de Evaluación Educativa

Instituto Nacional de Evaluación Educativa

Ministerio de Educación y Formación Profesional

Paseo del Prado, 28 • 28014 Madrid • España

INEE en Blog: <http://blog.intef.es/inee/> | INEE en Twitter: @educalINEE

NIPO línea: 847-19-061-4 NIPO IBD: 847-19-060-9

Más información

