



COVID Y CCAA LA COMUNIDAD DE MADRID UN CASO EXTREMO

Vicente Pérez Quintana
Madrid, febrero 2021

Contenido

1.- Mortalidad debida a COVID-19. Comparativa por CCAA y provincias.....	2
2.- Impacto de la COVID-19. Estudio sero-epidemiológico	6
3.- Sobre las causas de las diferencias territoriales	13
3.1.- Hipótesis	13
3.2.- Resultados	15

1.- MORTALIDAD DEBIDA A COVID-19: COMPARATIVA POR CCAA Y PROVINCIAS

Entre marzo y mayo de 2020, ambos inclusive, el total de fallecimientos en España debidos directamente a la COVID-19 fue de 32.652, a los que se añaden 13.032 de los que se sospecha que la causa también fue el coronavirus. Suman 45.684 muertes, a razón de 96,3 por cada 100.000 habitantes¹. Representan un 30% del total de defunciones habidas en el trimestre.

Los números de la Comunidad de Madrid sobresalen, con un total de 14.718 fallecimientos y una tasa de 217,1 por cada 100.000 habitantes. La CM ocupa el primer lugar en cifras absolutas, muy por delante de Cataluña, y el segundo en cifras relativas, superada por Castilla-La Mancha. La ratio madrileña más que duplica la nacional. Es oportuno retener sendos datos pues señalan el comportamiento singular de la región en esta crisis.

Muertes y tasa por 100.000 habitantes por CCAA, período enero-mayo 2020						
	Todas las causas	COVID19	COVID19 >64 años	Tasa total	Tasa COVID	Tasa COVID (<64 años)
Total nacional	231.104	45.684	42.688	487,1	96,3	463,1
Andalucía	33.028	1.940	1.729	390,4	22,9	117,6
Aragón	7.000	1.145	1.102	526,8	86,2	381,8
Asturias	6.217	521	496	610,3	51,1	186,1
Baleares, Illes	3.620	248	224	309,1	21,2	122,3
Canarias	7.352	239	207	338,1	11,0	59,0
Cantabria	2.918	278	259	500,7	47,7	200,1
Castilla y León	17.130	4.697	4.506	715,8	196,3	734,4
Castilla-La Mancha	13.804	4.897	4.542	675,2	239,5	1.164,0
Cataluña	39.665	10.754	10.079	509,9	138,3	686,5
Comun. Valenciana	21.155	1.767	1.622	418,5	35,0	165,2
Extremadura	5.868	762	723	551,7	71,6	323,9
Galicia	14.357	719	674	531,7	26,6	98,0
Madrid, Comun.	37.309	14.718	13.677	550,4	217,1	1.131,5
Murcia, Región	5.222	166	153	345,6	11,0	64,3
Navarra, C. Foral	3.209	728	702	485,6	110,2	537,3
País Vasco	10.930	1.687	1.598	492,4	76,0	320,0
Rioja, La	1.722	408	388	538,7	127,6	576,5
Ceuta	278	7	5	331,6	8,3	49,3
Melilla	230	3	2	264,1	3,4	21,7

Fuente: INE. Las defunciones por COVID incluyen los casos sospechosos

En el mapa siguiente se representa la tasa de mortandad a causa de la COVID por 100.000 habitantes desagregada por provincias. Los estratos se han definido de modo que la media nacional se halle en el centro del estrato intermedio.

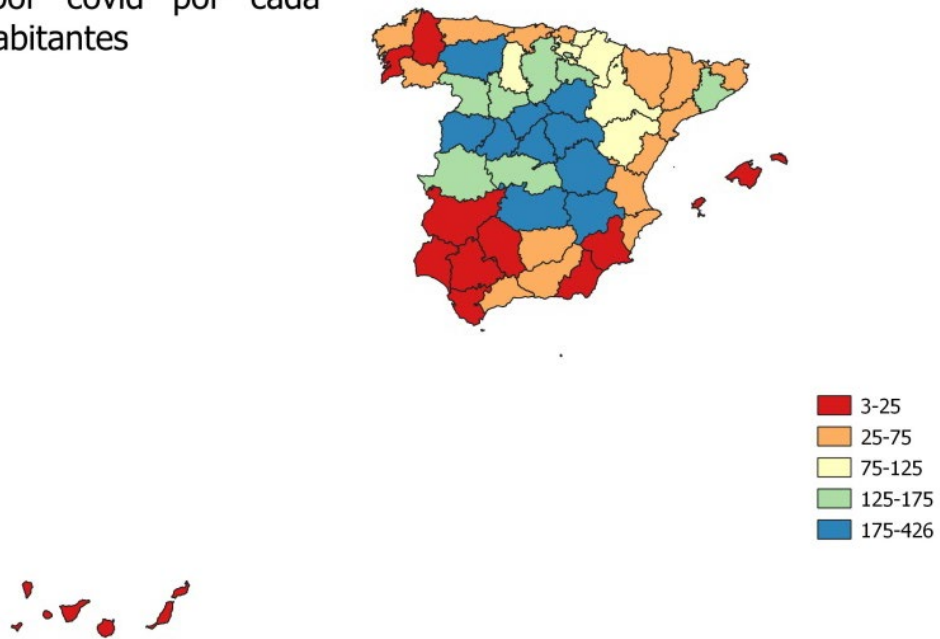
¹ La tasa está calculada sobre la población estimada en el Padrón de Habitantes (INE) a 01/01/2020. El mismo criterio se aplica a lo largo del texto.

La imagen presenta un patrón, esto es, la distribución de las provincias según la intensidad de los fallecimientos no es aleatoria, sino que define rasgos más o menos precisos.

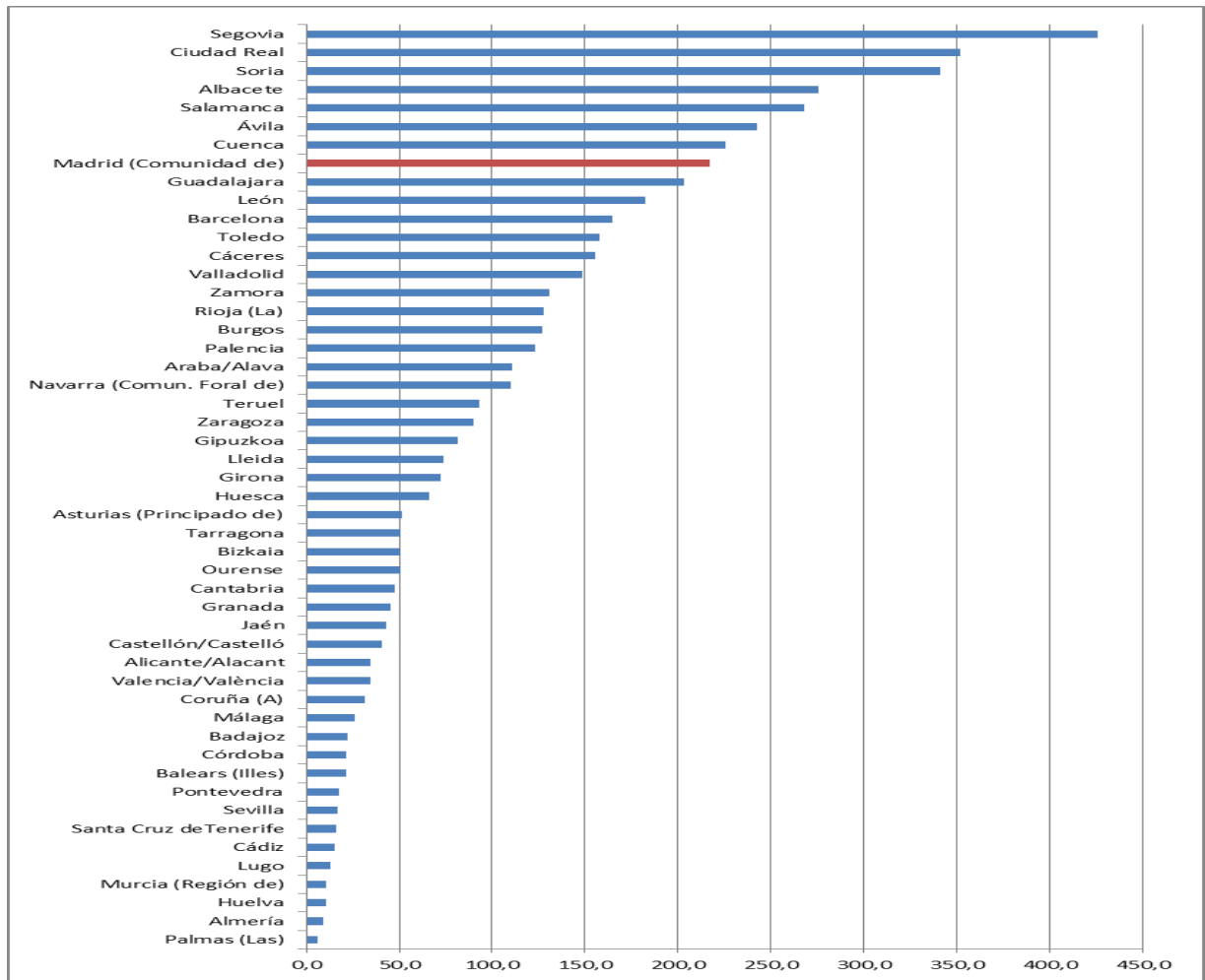
Resalta la concentración del estrato más alto en la meseta peninsular, en agudo contraste con la España costera. Madrid y las provincias limítrofes arrojan los mayores valores, frente a las tasas de los dos archipiélagos y las de Galicia, Andalucía, la costa cantábrica y la costa mediterránea. En ésta, la única excepción viene dada por Barcelona. Con valores arriba/abajo cercanos a la media se encuentran Aragón, Navarra y País Vasco. La tasa de la CM es 36 veces más alta que la de la provincia con la tasa más baja.

A su vez, las provincias limítrofes o cercanas (Segovia, Soria y Ciudad Real) presentan intensidades atípicas, muy altas.

Muertes por covid por cada
100.000 habitantes



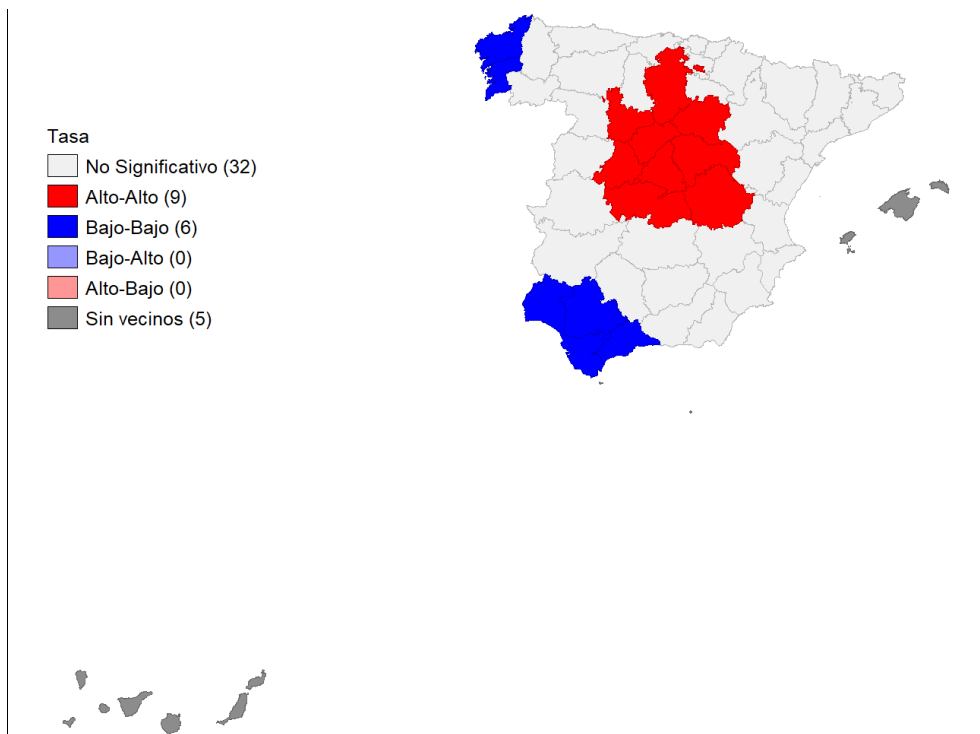
Fuente: INE



Fuente: INE

El índice univariante global de Moran mide la correlación espacial, en este caso, de la mortalidad por provincias debida a la COVID expresada por la tasa cada 100.000 habitantes. Asocia la tasa de cada provincia con la tasa promedio de las provincias vecinas, definidas éstas como aquellas con las que existe frontera. El valor observado de 0.42 es significativo y moderadamente alto. Expresa la pendiente de la recta de ajuste recogida en el diagrama de dispersión. Nótese que la mayoría de las provincias se alinean en los cuadrantes II (provincias con tasas por encima de la media emparejadas con provincias vecinas con tasas también altas) y III (idem, pero referido a tasas por debajo de la media). En cambio, son pocas las provincias con combinaciones alto/bajo (cuadrante I) o bajo/alto (cuadrante IV). En el mapa, a su vez, se recoge el índice local de Moran. Define dos agrupamientos: gran parte de la España interior en la que se concentran las provincias con valores altos de la tasa acompañadas por otras provincias también con valores altos, frente a Galicia y Andalucía occidental donde ocurre lo contrario.

Fuente: *Elaboración propia*



Fuente: *Elaboración propia*

La concentración de la mayor intensidad del impacto de la COVID en la Comunidad de Madrid y en las dos Castillas suscita un interrogante más que obvio: ¿a qué se debe?

2.- IMPACTO DE LA COVID-19: ESTUDIO SERO-EPIDEMIOLÓGICO

Antes de responder al interrogante, merece la pena examinar el diferencial de seropositividad por CCAA y provincias. El Instituto de Salud Carlos III realiza el ENE-COVID *“un amplio estudio longitudinal sero-epidemiológico, de base poblacional, cuyos objetivos son estimar la prevalencia de infección por SARS-CoV-2 mediante la determinación de anticuerpos frente al virus en España, en cada una de las Comunidades y Ciudades Autónomas y en cada provincia, proporcionando información por edad y sexo, y evaluar los cambios de la prevalencia a lo largo del tiempo”*. Hasta la fecha se han corrido cuatro rondas, tres antes del verano, durante la primera ola de la infección, y la cuarta en el mes de noviembre, durante la segunda ola.

El informe correspondiente a la cuarta ronda ofrece información detallada sobre la *“prevalencia actual”* y sobre la *“prevalencia global”*. En el presente trabajo nos centramos en esta última. Por ella se entiende *“la prevalencia acumulada en ENE-COVID para el conjunto del país, por comunidad y por provincia respectivamente. (...) Este estimador representa el porcentaje de población española no institucionalizada que, de acuerdo a los datos obtenidos con el test rápido, ha tenido contacto con SARS-CoV-2 desde el inicio de la pandemia. Para su cálculo nos hemos basado en los positivos en cualquiera de las rondas dentro de los participantes en la ronda 4”*.

Tal como reza el propio informe, el estudio ofrece el porcentaje de población española que ha sido infectada, excluido el segmento de la población institucionalizada, esto es, personas mayores y dependientes que viven en residencias y otras instituciones similares. Tampoco incluye a las personas que han fallecido ni a las hospitalizadas. Por otra parte, la muestra incorpora a personas de colectivos especiales (personal sanitario, emergencias, policía...), pero en número insuficiente. Asimismo, hay que considerar que los test no son 100% fiables. Por último, al tratarse de un estudio longitudinal, se trabaja con una muestra fija de participantes que va mermando por diferentes causas (negativa a responder, enfermedad...). Todas estas circunstancias introducen sesgos en los resultados, por lo que éstos siempre se han de tomar con la debida cautela. No obstante, dichos sesgos serán reducidos en relación a las cifras globales y serán tanto más significativos a medida que se trabaje con colectivos específicos, respaldados por submuestras pequeñas.

A diciembre de 2020, según el ENE-COVID, un 9,9% de la población española había sido infectada, variando la proporción entre un 9,6% y un 10,1%. La cifra equivale en absolutos a unos 4,7 millones de personas.

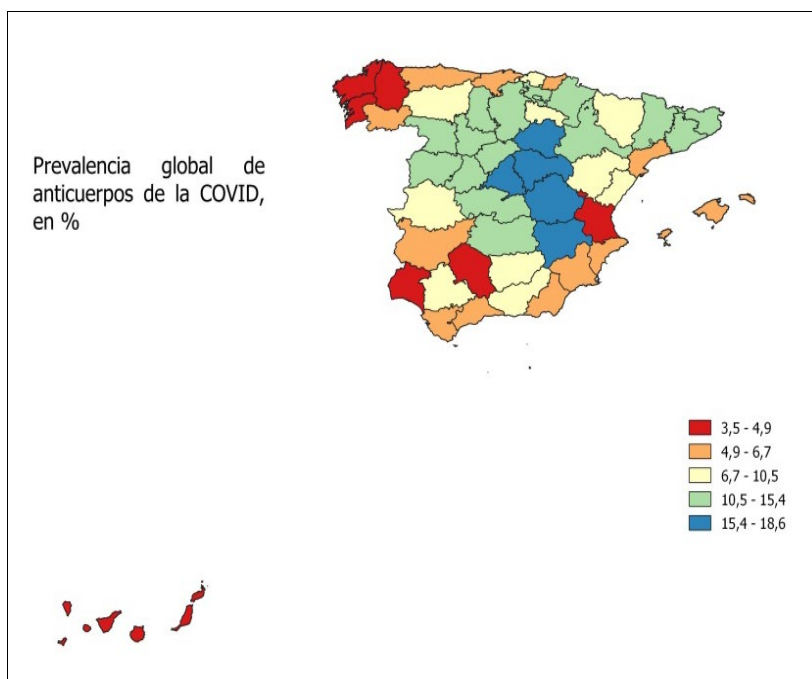
La tasa varía enormemente entre unas CCAA y otras y entre unas provincias y otras. La Comunidad de Madrid repite el primer puesto en el ranking, con una tasa del 18,6%, casi duplicando la media nacional y más que doblando la de CCAA como Canarias, Andalucía, Galicia o País Vasco. Las dos Castillas, junto con Navarra, además de la CM, señalan el foco principal. Se produce, como era de esperar, la dualidad entre la España interior y la España costera, si bien Cataluña presenta, también, una tasa elevada.

Tabla 5. Prevalencia global (Rondas 1-4) de anticuerpos IgG anti SARS-CoV-2 por CCAA

	Total			Hombres			Mujeres		
	Nº	%	IC 95%	Nº	%	IC 95%	Nº	%	IC 95%
Total Nacional	51409	9,9	9,4 - 10,4	24208	9,6	9,0 - 10,2	27201	10,1	9,5 - 10,7
Comunidades Autónomas									
Andalucía	7652	7,1	6,2 - 8,1	3558	7,1	6,0 - 8,4	4094	7,1	6,1 - 8,3
Aragón	2336	11,7	9,7 - 14,0	1109	11,5	9,0 - 14,6	1227	11,9	9,6 - 14,6
Principado de Asturias	1616	6,1	4,7 - 7,7	730	6,6	5,0 - 8,6	886	5,6	4,1 - 7,7
Islas Baleares	891	6,3	4,8 - 8,2	428	6,5	4,5 - 9,4	463	6,1	4,2 - 8,7
Canarias	1973	3,8	2,7 - 5,3	904	3,9	2,6 - 5,8	1069	3,7	2,5 - 5,5
Cantabria	1371	6,3	4,8 - 8,4	643	6,3	4,4 - 9,0	728	6,4	4,5 - 9,0
Castilla y León	6359	12,6	11,2 - 14,1	3105	12,8	11,2 - 14,6	3254	12,4	10,9 - 14,1
Castilla-La Mancha	3967	16,1	14,3 - 18,2	1888	15,4	13,2 - 17,8	2079	16,8	14,4 - 19,6
Cataluña	4817	11,6	9,9 - 13,7	2248	11,3	9,4 - 13,5	2569	12,0	10,1 - 14,2
Comunitat Valenciana	3842	5,7	4,8 - 6,8	1817	5,6	4,6 - 6,8	2025	5,8	4,6 - 7,3
Extremadura	2375	8,0	6,4 - 9,9	1153	8,6	6,6 - 11,0	1222	7,4	5,9 - 9,3
Galicia	3103	4,5	3,6 - 5,5	1445	4,6	3,5 - 6,1	1658	4,3	3,3 - 5,6
Comunidad de Madrid	2734	18,6	16,7 - 20,6	1265	17,1	14,9 - 19,4	1469	19,9	17,6 - 22,4
Región de Murcia	1389	6,1	4,3 - 8,6	637	6,7	4,5 - 9,8	752	5,6	3,8 - 8,1
Comunidad Foral de Navarra	1422	14,3	11,7 - 17,2	699	14,0	10,6 - 18,2	723	14,5	11,9 - 17,6
País Vasco	2628	8,2	6,9 - 9,8	1247	8,4	6,8 - 10,4	1381	8,0	6,5 - 10,0
La Rioja	1127	8,2	6,3 - 10,5	544	7,5	5,6 - 9,9	583	8,9	6,2 - 12,5
Ceuta	893	9,2	6,7 - 12,6	374	8,8	5,8 - 13,1	519	9,6	6,5 - 14,0
Melilla	914	10,5	7,2 - 15,1	414	9,2	5,5 - 15,0	500	11,8	8,4 - 16,4

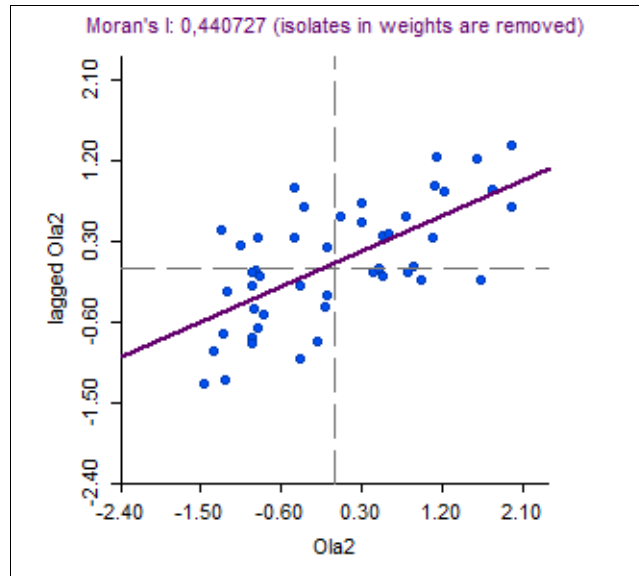
Fuente: ENE-COVID, IS Carlos III

El mapa siguiente recoge la distribución provincial de la tasa global. Repite, con mayor grado de detalle, la foto que se entresaca de la observación de las tasas autonómicas. Madrid y las provincias limítrofes y cercanas más que triplican las tasas de las provincias menos castigadas por la pandemia.

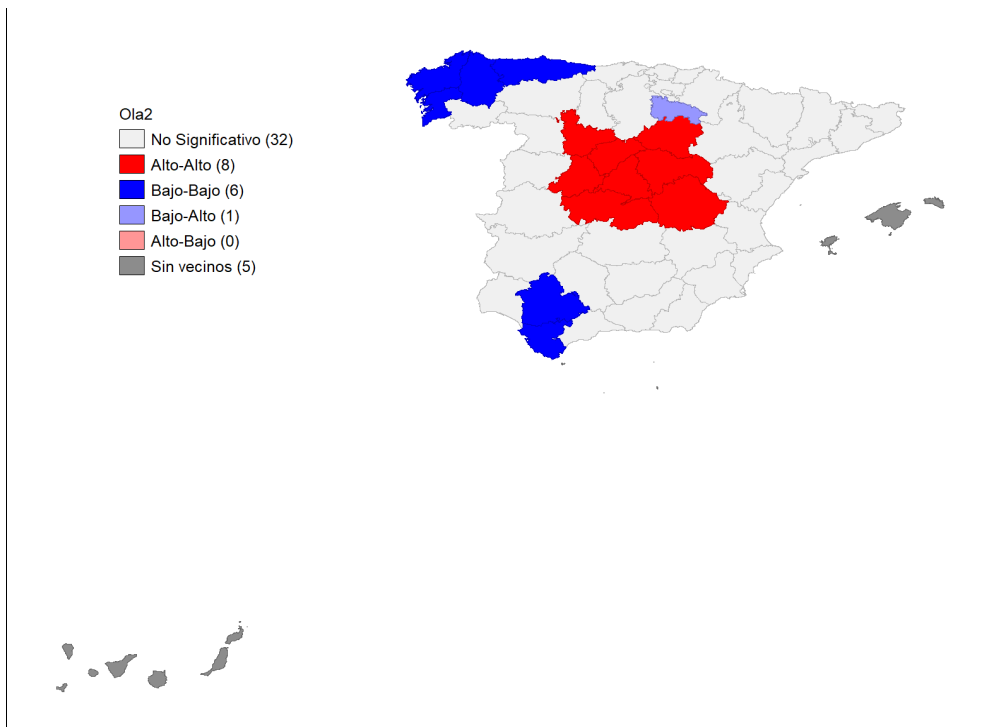


Fuente: ENE-COVID, IS Carlos III

El índice de Moran (0,44) muestra una correlación espacial significativa y moderadamente positiva, es decir, las provincias con tasas de prevalencia altas (o bajas) con frecuencia hacen frontera (según el criterio de la reina) con provincias con tasas altas (o bajas). De ahí que la mayoría de las provincias se localicen en los cuadrantes I y III del diagrama de dispersión. Las agrupaciones alto-alto y bajo-bajo aparecen bien representadas en el mapa de frío-calor adjunto.



Fuente: Elaboración propia



Fuente: Elaboración propia

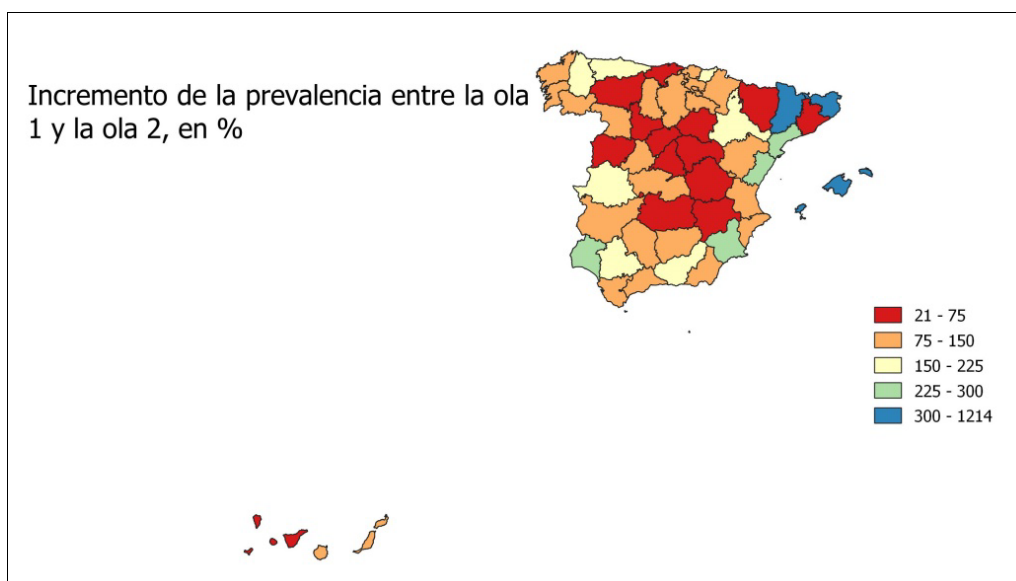
El estudio aporta un segundo dato trascendente. Entre la primera ola y la segunda la tasa española de seropositividad se duplicó, al pasar del 5% en junio al 9,9% en noviembre. Pues bien, se observa que los incrementos relativos más acusados de las tasas respectivas se producen en las CCAA y provincias con valores por debajo de la media en la primera ola. El siguiente cuadro es muy ilustrativo. El él se dicotomizan los porcentajes de aumento de las tasas entre la ola 1 y la ola 2 y las tasas de la ola 1, según que estén por encima o por debajo de las medias españolas respectivas. La mayoría de las provincias con intensidades inferiores en junio registran incrementos fuertes en noviembre y, a su vez, la mayoría de las provincias con intensidades superiores experimentan crecimientos más moderados.

Entre la primera y la segunda ola tiene lugar una aproximación de las tasas extremas, se estrechan las diferencias: el coeficiente de variación (desviación típica/media aritmética sin ponderar) de la distribución de las tasas de la primera ola es 68%, frente a 44% el de la distribución de las tasas de la segunda ola. En particular, los ascensos más acusados se dan en la costa mediterránea, mientras que los más moderados se dan en la meseta. La tasa de Madrid, p.ej., aumenta un 59%, frente a un 90% el promedio nacional. En cambio, los aumentos relativos de Ceuta, Melilla, Baleares, Gerona, Lérida... se disparan, alcanzando valores por encima y muy por encima del 300%. La asociación de sendas variables se ajusta bien a un modelo de regresión potencial, con un coeficiente de determinación o varianza explicada (R^2) del 0.58 (58%).

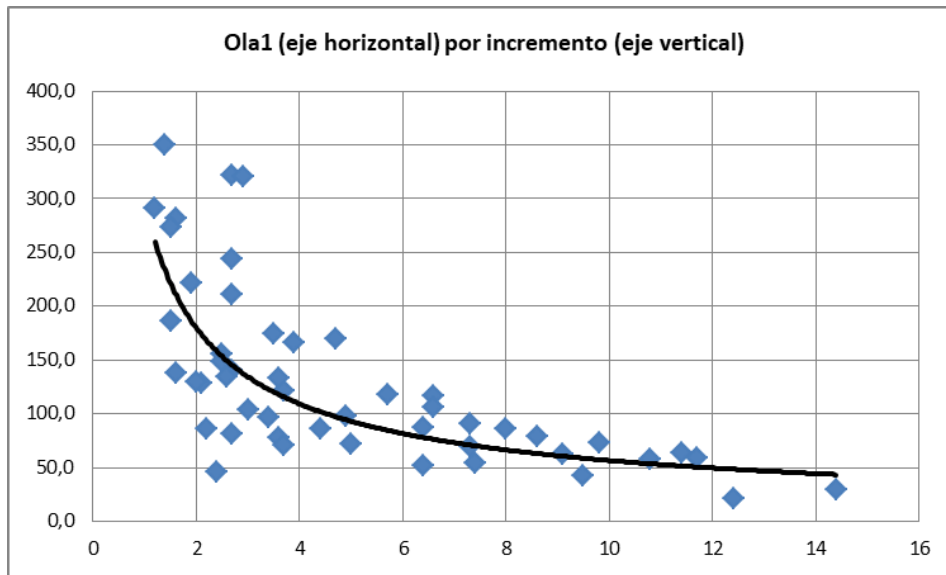
Provincias según aumento de la tasa de prevalencia por valor en ola 1

	Tasa de incremento		Total
	<	>	
Tasa ola 1	7	25	32
	14	4	18
Total	21	29	50

Fuente: Elaboración propia

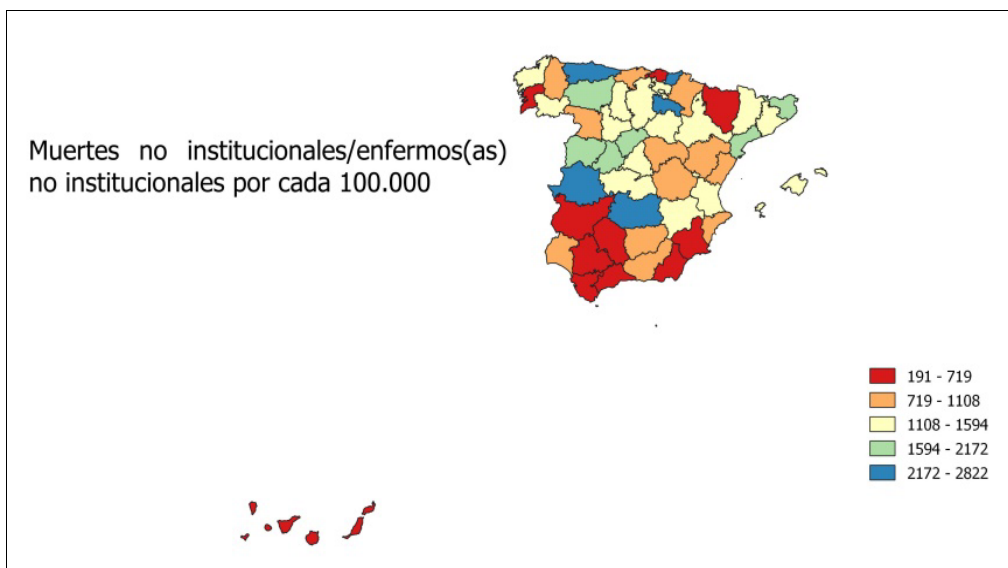


Fuente: Elaboración propia



Fuente: Elaboración propia. Línea continua: ajuste potencial

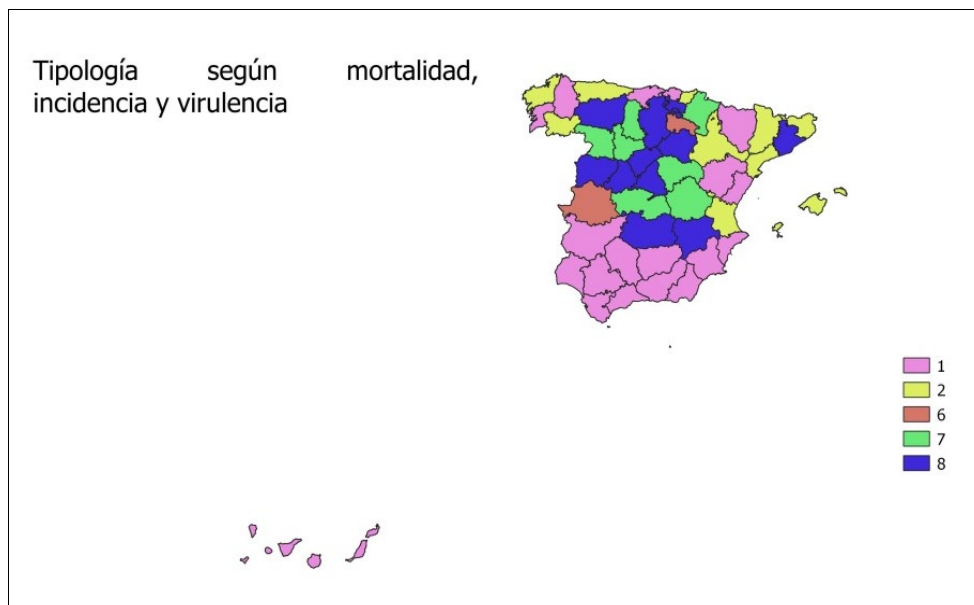
Es interesante estudiar la relación entre la prevalencia por provincias y la cifra de defunciones debidas a la COVID. Como éstas, según las ofrece el INE, se refieren al periodo marzo-junio, tomamos los casos de infección según la ENE-COVID a junio (ola 1). Solo se cuentan las defunciones ocurridas fuera de las residencias de la tercera edad y similares, toda vez que la cifra de personas afectadas incluye solo a las no institucionalizadas. El índice muertes/afectados(as) por provincias expresa la diferencial virulencia de la enfermedad. La media española es de 1.311 fallecimientos por cada 100.000 enfermos/as. Los valores más altos se dan en la España interior, con Segovia, Salamanca, Ciudad Real y Ávila, más algunas provincias cantábricas, encabezando el ranking. La CM arroja un índice similar a la media, igual que Barcelona. De nuevo, las provincias periféricas -especialmente Canarias y Andalucía- se sitúan en el polo contrario.



Fuente: Elaboración propia

El diferencial de mortalidad por provincias que se ha descrito en el epígrafe anterior puede deberse al diferencial de porcentajes de población afectada por la epidemia, al diferencial del grado de destrucción del virus o a la interrelación de ambos. Dicho de otro modo, las provincias con mayores tasas de mortalidad a causa de la COVID en el periodo marzo-junio pueden serlo porque en ellas la enfermedad tocó a más personas en la ola 1 (tasa de incidencia) y/o porque en ellas el virus fue más dañino (grado de virulencia). Cabe reunir las tres variables y generar una tipología de provincias según la combinación entre ellas. A tal fin se han dicotomizado las tres en valores más altos/más bajos que el promedio nacional. El cruce da lugar al esquema adjunto. En el mapa se recoge el detalle. La categoría 3,4 y 5 están vacías. A su vez, la categoría 6 es residual, con tan solo dos provincias. El grueso de las provincias se concentra en las categorías 1, 2 y 8, con la 7 que hace de línea de separación. Hay que subrayar que las categorías delimitan áreas contiguas, grandes manchas de color con interrupciones excepcionales. Veámoslo.

Tipología			
Mortalidad	Incidencia	Virulencia	
		Por debajo	Por encima
Por debajo	Por debajo	1	2
	Por encima	3	4
Por encima	Por debajo	5	6
	Por encima	7	8



Fuente: Elaboración propia

- Tipo 1. En él confluyen las provincias que arrojan valores por debajo de las medias respectivas en las tres variables: menos mortalidad, menos incidencia y menos virulencia. Es el tipo que más unidades suma. Incluye toda Andalucía y buena parte de la franja costera.
- Tipo 2. Abarca las provincias con menor mortalidad e incidencia, como el tipo 1, pero con una virulencia más acusada. Completa el litoral, al norte de la línea imaginaria horizontal que divide la península en dos.

- Tipo 8. Aquí se aglomeran las provincias que arrojan valores más altos que los promedios nacionales respectivos en las tres variables. Por tanto, define el polo opuesto al anterior. Abarca parte de la meseta, con la Comunidad de Madrid y la mayoría de las provincias de Castilla-León, más Ciudad Real y Albacete. También incluye a Barcelona.
- Tipo 7. Reúne las provincias con mortalidad e incidencia altas, como el tipo 8, pero con una tasa de virulencia más suave. De hecho, completa el tipo 8, formando una especie de anillo alrededor de éste. Se interpone entre el tipo 8 y los tipos 1 y 2.
- Tipo 6. Es un tipo residual, formado por Cáceres y La Rioja. Presenta una mortalidad alta, debida a un grado de virulencia elevado, mientras que la incidencia de la enfermedad es claramente inferior a la media.

En conclusión, la tipología afina la dicotomía que se ha observado en las páginas anteriores entre la España costera (incluidos los dos archipiélagos) y la meseta. En una, la mortalidad y la incidencia son inferiores a la media, diferenciándose la zona meridional por presentar una virulencia menor que la septentrional. En la otra, la mortalidad y la incidencia son superiores, siendo la virulencia más acusada en las provincias más interiores y más suave en las limítrofes con los tipos 1 y 2.

Por otra parte, las tres variables correlacionan entre sí. En un análisis de regresión múltiple, tomando la tasa de mortalidad debida a la COVID como dependiente y las otras dos como independientes, se obtiene un coeficiente de determinación (R^2) del 0.93 sobre un máximo de 1². Los coeficientes beta (que miden el efecto directo de cada variable independiente sobre la dependiente) son 0.825 y 0.323, asociados respectivamente a la incidencia y la virulencia. Esto es, la primera variable tiene un peso muy destacado sobre la segunda.

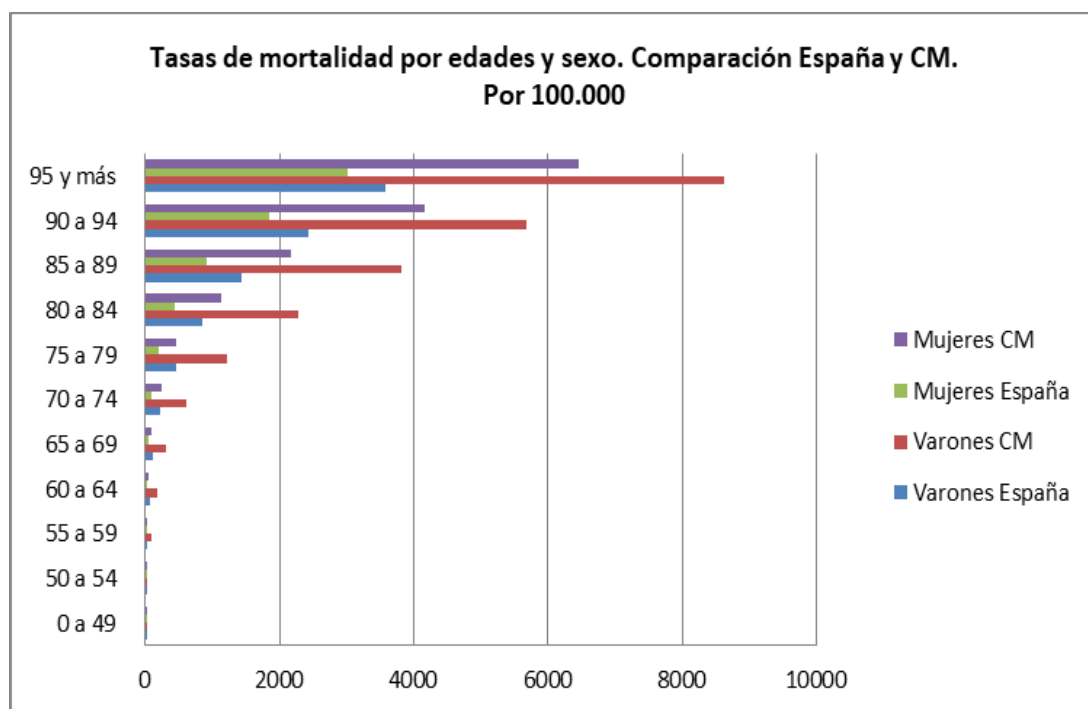
² Para que el modelo supere los contrastes de normalidad, colinealidad y heteroscedasticidad habría que hacer transformaciones en los valores de las variables (tomar logaritmos). Para los objetivos del presente trabajo, y dado que no se seguirá trabajando con esta correlación en lo que sigue, no merece la pena detenerse en la tarea.

3.- SOBRE LAS CAUSAS DE LAS DIFERENCIAS TERRITORIALES

3.1.- HIPÓTESIS

La conclusión que se desprende de los epígrafes precedentes es que la tasa diferencial de mortalidad debida a la COVID está determinada tanto por la incidencia de la enfermedad (proporción de población afectada) como por la capacidad destructora del virus (grado de virulencia). ¿Por qué ambos determinantes alcanzan valores superiores a la media en la Comunidad de Madrid? A primera vista cabría sospechar que los motivos tienen relación bien con la estructura de edades de la población, bien con el grado de urbanización, bien con ambos aspectos.

En el siguiente gráfico se representan las tasas específicas de mortalidad debida a la COVID por edad y sexo comparadas. Las madrileñas superan en todos los casos a las españolas. Aquéllas sobre las segundas vienen siendo entre 2,2 y 2,7 veces más altas. En realidad, si las tasas específicas de la CM fueran iguales a las del total nacional, entonces las muertes contabilizadas en la región en marzo-junio hubieran sido 5.950, en lugar de 14.718, es decir, 2,5 veces menos que las observadas.



Las causas de las diferencias territoriales que se han examinado hasta aquí son múltiples. Su indagación llevará, a buen seguro, mucho tiempo y requerirá muchos estudios y un amplio abanico de enfoques. En las páginas que siguen tratamos de hacer una modesta primera aproximación a la cuestión.

En primer lugar, se establece de manera razonada, a modo de hipótesis, una relación de factores que pueden ser explicativos. En segundo lugar, se somete a verificación las hipótesis mediante un abordaje estadístico multivariado.

Cabe esperar que la mortalidad por COVID será mayor en los lugares en los que la enfermedad es más frecuente. Por tanto, es previsible que la tasa de mortalidad se vea influenciada por los factores que contribuyen a la propagación del virus, tales como las variables de contacto interpersonal. Entre ellas, por ejemplo, estarán la movilidad o la densidad urbana. También se puede sospechar de la incidencia de los contactos con determinados colectivos en los que, por alguna razón, la afección es mayor. Por otra parte, la tasa de mortalidad probablemente se vea empujada hacia arriba en las poblaciones en las que los colectivos vulnerables pesan más, como, por ejemplo, el segmento de edades por encima de 65-70 años. El clima es otro factor a tener en cuenta, bien sea por su relación con el virus, bien sea por su efecto sobre la salud y la resistencia de las personas. Tratándose de una enfermedad es lógico pensar que las variables relativas al sistema sanitario tengan una importancia determinante.

Otro grupo de factores que seguramente actúa como causa es el de las medidas adoptadas por las CCAA para combatir la COVID. Lamentablemente no es fácil seleccionar a este respecto algún indicador, dada la falta de homogeneidad de las medidas y de los datos y considerando el grado de opacidad e incluso de manipulación que sufre la información. Habría sido oportuno contar con el número de rastreadores(as) por habitante, pero es una empresa imposible.

3.2.- RESULTADOS

Para operacionalizar las dimensiones mencionadas en el párrafo anterior se ha elaborado el cuadro de indicadores que se recoge a continuación. Posteriormente se analizará su relación con la tasa de mortalidad debida a la COVID.

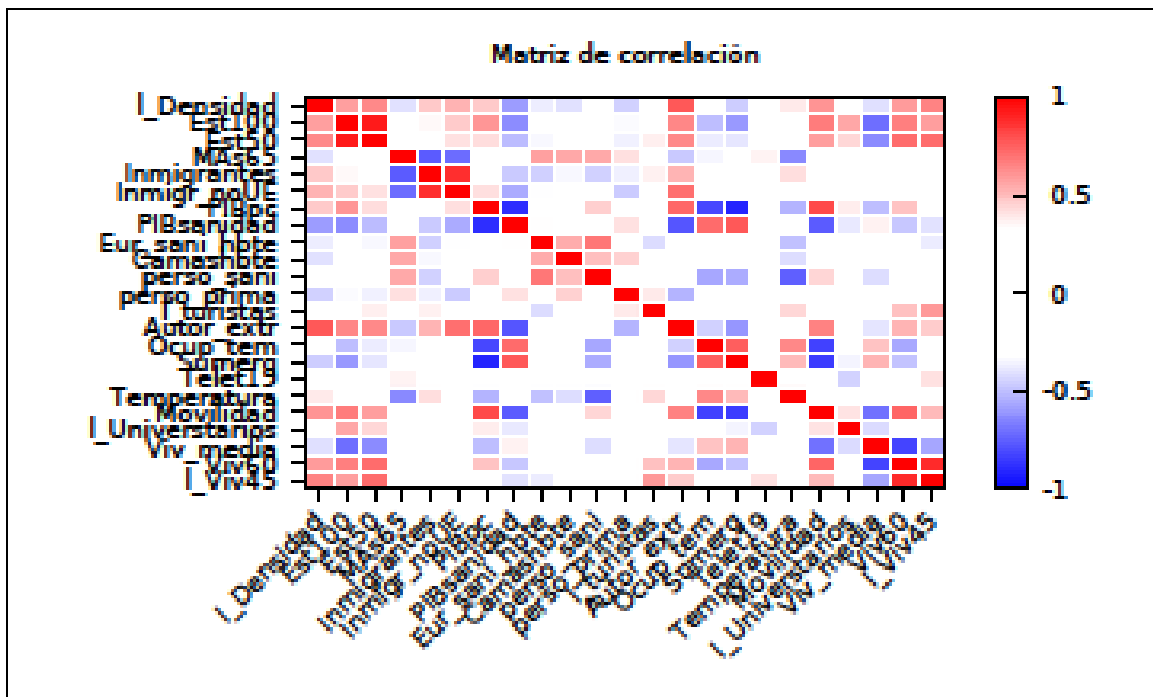
Nombre	Índice	Fuente de los datos
Temperatura	Temperatura media por CCAA marzo de 2020	El Confidencial, AEMET
Edad>65	% > 64 años sobre total población a 01/01/2020	INE, Padrón de Habitantes
Urba>50	% de población residente en municipios de más de 50.000 habitantes sobre el total de población a 01/01/2020	INE, Padrón de Habitantes
Urba>100	Ídem, municipios de más de 100.000 habitantes	INE; Padrón de Habitantes
Densidad	Habitantes/superficie a 01/01/2020	INE
Inmigrantes	%de población extranjera residente sobre total de población a 01/01/2020	INE; Padrón de Habitantes
Inmigrantes_no UE	% de población extranjera (no UE) sobre total de población	INE, Padrón de Habitantes
PIBpc	PIB per cápita en 2018	INE, Contabilidad Regional de España
PIBsanidad	Gasto sanitario público sobre el PIB en 2018	Ministerio de Sanidad
eur_sani_hbte	Gasto sanitario público por habitante en 2018	Ministerio de Sanidad
camas_hbte	Camas por 1.000 habitante en 2018	Ministerio de Sanidad
perso_sani	Personal sanitario por 1.000 habitantes en 2018	Ministerio de Sanidad
perso_prima	Personal de asistencia primaria por 1.000 habitantes en 2018	Ministerio de Sanidad
Turistas	Turistas (enero-marzo de 2020)/población a 01/01/2020, en %	INE
Autor_extr	Autorizaciones de trabajo concedidas por 1.000 hbtes, en 2016	Ministerio de Trabajo
Ocup_temp	Ocupados(as) temporales/total ocupados(as)	EPA, INE
Sumerg	Economía sumergida en % del PIB	FUNCAS
Telt19	% trabajadores(as) que teletrabajan	INE/RANSTAD
Tele20_t3	Ídem tercer trimestre de 2020	INE, EPA
Mortalidad	Tasa de mortalidad debida a la COVID	INE, IS Carlos III
Prevalencia1	% de personas infectadas por la COVID, primera oleada	IS Carlos III
Prevalencia2	Ídem, segunda oleada	IS Carlos III
Virulencia	Muertes/enfermos, en la primera oleada	Elab. propia
Movilidad	% de población residente en un área que se movió a otra área en noviembre de 2019	INE
Universitarios	Matrículas universitarias por 1.000 hbtes, curso 2019-20	Ministerio de Educación
Viv_media	Tamaño medio de las viviendas principales en 2011	Censo de 2011
Viv<60	5 de viviendas principales con menos de 60 m ²	Censo de 2011
Viv<45	Ídem, menos de 45 m ²	Censo de 2011

Se toma como unidad de observación las comunidades autónomas. Hubiera sido preferible manejar la unidad territorial provincial, como en los apartados anteriores. No es posible, toda vez que buena parte de los índices que se pretende analizar no está disponible por provincias.

Para abordar un análisis multivariable se ha de trabajar con variables con distribuciones normales. Por tanto, se ha sometido el listado del cuadro a los contrastes de Shapiro-Wilk y Jarque-Bera, dando prioridad a los resultados del primero.

Los indicadores Densidad, Turistas, Tele20_t2, Mortalidad, Prevalencia1, Universitarios y Viv<45 no superan el test. Sí lo hace la transformación logarítmica de los mismos, salvo en el caso de Tele20_t2. En este caso tampoco lo hace la transformación cuadrática, por lo que se excluye la variable del análisis.

En la figura adjunta se representa la tabla de correlaciones entre las diferentes variables. Como es sabido, el coeficiente oscila entre -1 y +1, siendo 0 y los valores comprendidos entre -0.5 y +0.5 poco o nada significativos.



Fuente: Elaboración propia

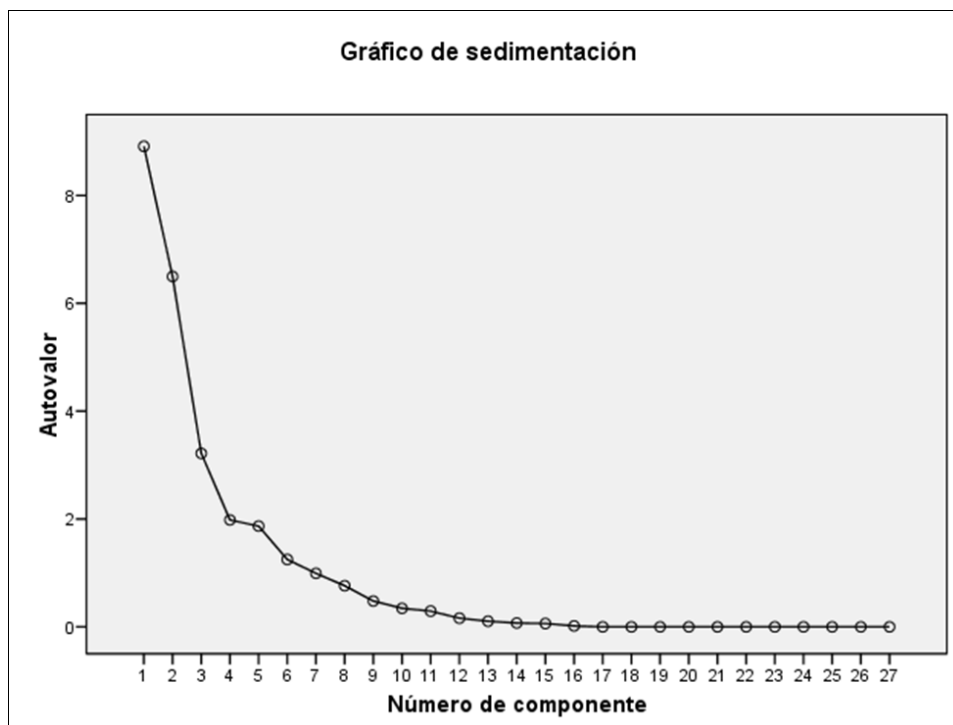
La figura muestra diversas aglomeraciones de variables, tales como la interrelación entre de los índices sanitarios o los de grado de urbanización.

Para mostrar con mayor claridad este aspecto se ha sometido al listado de variables a un análisis factorial, por el método de los componentes principales, con rotación varimax de los factores. Esta técnica permite simplificar un número alto de variables en un número reducido de dimensiones o factores. Cada uno de ellos es homogéneo y se forma con las variables originales que correlacionan mucho entre sí. Hemos extraído los cinco factores que más varianza explican: el 83%. Cada uno de ellos se caracteriza o *etiqueta* en función de las variables que arrojan mayor puntuación-saturación en él. La información correspondiente se recoge en el cuadro adjunto, en el que se destacan los valores absolutos más elevados, diferenciando según sea el signo: positivos con fondo rojo y negativos con fondo amarillo.

- Componente 1. Es el más importante, explica el 29% de la varianza. Está caracterizado por los índices de urbanización (de forma que segrega las CCAA en las que un alto porcentaje de la población reside en las grandes ciudades), el indicador de PIBpc (separando las CCAA ricas del resto), la proporción de movilidad

(a más riqueza y urbanización más se mueve la gente), vivienda pequeña (alta concentración de las mismas en las ciudades grandes) y precariedad (a más desarrollo económico menos economía sumergida y menor peso de la contratación temporal). Nótese que las CCAA que más puntúan en este componente destinan menos dinero a sanidad en términos de la participación sobre el PIB.

- Componente 2. Explica un porcentaje elevado de la varianza, el 21%. Está caracterizado por dos abanicos de variables: las relacionadas con la atención sanitaria (separa las CCAA con elevados índices de atención sanitaria) y con la composición de la población (tasa alta de mayores y poco peso de la inmigración). También se observa que la temperatura alcanza una puntuación alta con signo menos, mientras que la virulencia (proporción de muertes sobre el total de personas enfermas por la COVID en la primera oleada) es acusada: a más temperatura menos virulencia.
- Componente 3. Tiene una cuota significativa en la explicación de la varianza, 17%. Este factor es bastante puro, pues en él destacan especialmente las tasas de prevalencia en la primera y segunda oleadas y la tasa de mortalidad. Se repite la correlación negativa con la temperatura.
- Componente 4. Explica el 8% de la varianza, de forma que se aparta notablemente de los anteriores. En él destaca la densidad turística en el primer trimestre de 2020: separa a las CCAA con más turistas en relación al tamaño demográfico.
- Componente 5. Como el anterior, explica otro 8% de la varianza. Aquí la variable que caracteriza el factor es la importancia del teletrabajo en 2019, con signo menos, junto al peso de la matrícula universitaria, con signo más.



Fuente: Elaboración propia

	Matriz de componentes rotados				
	Componente				
	1	2	3	4	5
I-densidad	0,747	-0,483	-0,190	-0,006	-0,077
Est 100	0,891	-0,080	0,002	0,026	0,147
Est 50	0,831	-0,182	-0,022	0,181	-0,058
Más 65	0,001	0,909	0,082	-0,067	-0,215
Inmigrantes	0,260	-0,773	-0,047	-0,017	0,229
Inmigr_no UE	0,386	-0,673	0,017	-0,231	0,417
PIBpc	0,727	-0,003	0,488	-0,307	0,200
PIBSanidad	-0,706	0,322	-0,471	0,111	-0,154
Eur_sani_hbte	-0,054	0,732	-0,130	-0,429	0,151
Camas_hbte	0,046	0,721	0,045	0,021	0,237
Perso_sani	0,256	0,657	0,404	-0,328	0,127
Perso_prima	-0,297	0,547	-0,061	0,606	0,343
I_turistas	0,327	-0,227	-0,014	0,826	0,134
Autor_extr	0,735	-0,472	0,224	-0,128	0,158
Ocp_tem	-0,646	-0,251	-0,580	0,146	-0,006
Sumerg	-0,724	-0,071	-0,352	0,389	-0,168
Telet19	0,145	0,177	-0,109	-0,114	-0,876
Prevalencia2	0,103	-0,050	0,932	-0,054	0,072
I_prevalencia1	0,100	0,010	0,969	0,132	0,048
Virulencia	0,050	0,662	0,358	-0,003	0,046
Temperatura	-0,079	-0,580	-0,643	0,424	-0,005
Movilidad	0,860	0,091	0,319	-0,029	0,093
Viv_media	-0,819	-0,441	0,082	-0,191	-0,059
Viv60	0,846	0,123	0,070	0,346	-0,291
Imortalidad	0,046	0,355	0,904	-0,009	0,087
I_universitarios	0,384	0,188	0,133	0,042	0,667
Ivi45	0,716	-0,170	-0,067	0,461	-0,443

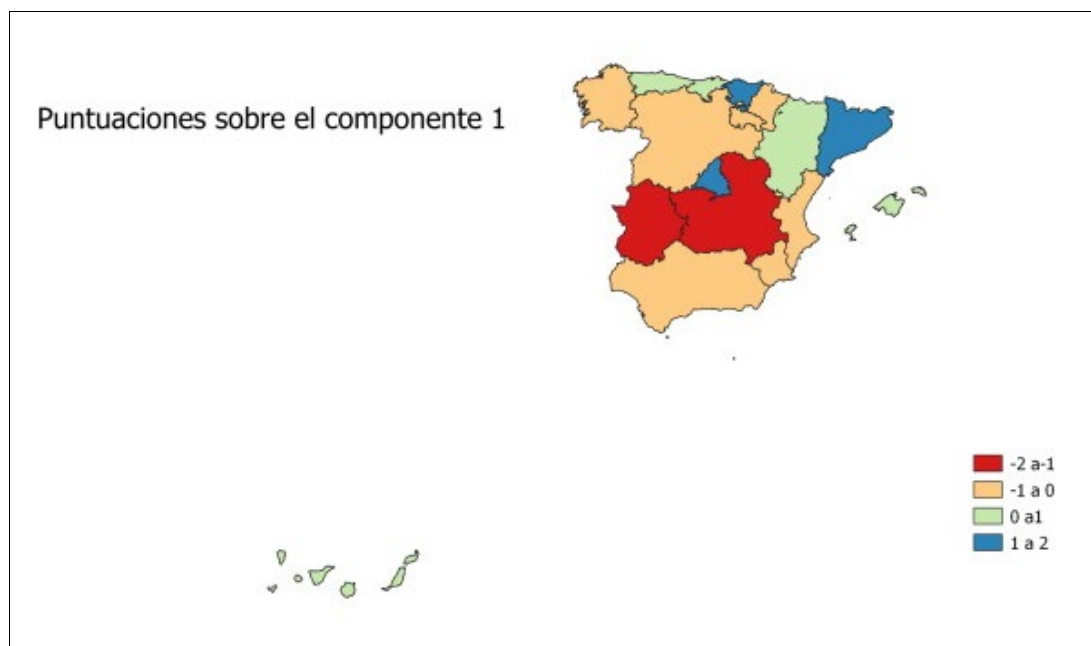
Fuente: Elaboración propia

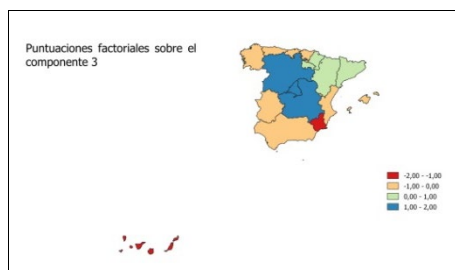
Una vez obtenida la solución factorial, es interesante estimar las puntuaciones factoriales de cada comunidad autónoma en las cinco dimensiones o componentes seleccionados. Los rankings respectivos se reproducen en el cuadro adjunto, así como en los 5 mapas siguientes. Hacer un análisis detallado de los resultados nos apartaría del objeto de la presente investigación. Valga, no obstante, resaltar la posición de la Comunidad de Madrid, al menos en los primeros componentes, los que explican casi el 70 de la varianza. Así, arroja la puntuación más elevada en el factor 1, junto con el País Vasco o Cataluña. Son CCAA con valores altos en cuanto a urbanización, actividad económica, movilidad y concentración de viviendas pequeñas, a la vez que en ellas pesa menos la economía sumergida y la precariedad laboral. En cambio, ocupa uno de los últimos lugares en el factor 2, junto a Andalucía o Baleares. En tal situación se hallan las CCAA con las ratios de atención sanitaria más bajas y, por otra parte, con poblaciones menos envejecidas y con proporciones significativas de inmigrantes. En varias de ellas el grado de virulencia es también destacado. En el componente 3, la CM ocupa el segundo lugar, detrás de Castilla-La Mancha y por delante de Castilla-León. Son las CCAA en las que el coronavirus ha sembrado más estragos. Tanto en el cuarto como en el quinto factor, la región se sitúa en posiciones más centradas, lejanas de los extremos: con un peso intermedio del turismo -variable en la que Canarias arroja un valor muy alto- y del teletrabajo en 2019 –si bien, en 2020, una vez explotó la pandemia, se ha colocado en la cabeza de las CCAA por el peso del teletrabajo-.

Ordenación de las CCAA según las puntuaciones factoriales en cada componente

Comp. 1		Comp. 2		Comp. 3		Comp. 4		Comp. 5	
Madrid	2,15493	Asturias	1,8754	C. Mancha	1,88074	Canarias	2,66181	Rioja	2,04508
P. Vasco	1,1398	C. León	1,1737	Madrid	1,70821	Andalucía	1,04474	Murcia	1.31008
Cataluña	1,01702	Extremadura	0,87415	C. León	1,01721	C. León	0,76598	Navarra	1,1638
Asturias	0,84531	P- Vasco	0,82054	Navarra	0,98946	Cataluña	0,6191	Cataluña	0,74667
Aragón	0,51673	Aragón	0,66097	Cataluña	0,63228	Extremadura	0,35138	Canarias	0,44608
Canarias	0,27878	Rioja	0,58794	Aragón	0,43785	Madrid	0,20899	Aragón	0,32197
Cantabria	0,13529	Galicia	0,58492	Rioja	0,14002	Rioja	0,13362	P. Vasco	0,31286
Balears, I.	0,07101	Cantabria	0,56166	P. Vasco	-0,1901	Aragón	0,11976	C. Valenciana	0,27669
Rioja	-0,00502	Navarra	0,07967	Cantabria	-0,21158	Cantabria	0,02397	C. León	0,22781
Murcia	-0,2208	Cataluña	-0,11174	Balears, I.	-0,30794	C. Mancha	-0,0062	Extremadura	-0,28691
C. Valencia	-0,24377	C. Mancha	-0,64756	Extremadura	-0,35604	Asturias	-0,16107	Cantabria	-0,43693
Navarra	-0,39852	C. Valenciana	-0,7291	Andalucía	-0,52364	C. Valenciana	-0,26798	Andalucía	-0,52922
Galicia	-0,46203	Canarias	-0,84584	C. Valenciana	-0,55199	Galicia	-0,54223	Madrid	-0,69102
Andalucía	-0,54382	Murcia	-0,91418	Galicia	-0,66275	P. Vasco	-1,03476	C. Mancha	-0,74155
C. León	-0,61601	Madrid	-1,03131	Asturias	-0,97696	Navarra	-1,19049	Galicia	-1,02475
Extremadura	-1,66995	Andalucía	-1,06732	Canarias	-1,33424	Balears, I.	-1,19554	Balears, I.	-1,34923
C. Mancha	-1,99894	Balears, I.	-1,87191	Murcia	-1,69153	Murcia	-1,53109	Asturias	-1,79142

Fuente: Elaboración propia





Otro aspecto del análisis de componentes principales que interesa resaltar es el cruce de las puntuaciones factoriales de las CCAA en los dos primeros ejes respecto del tercero. Éste, como se ha repetido más de una vez atrás, es un componente muy puro, en el que los indicadores que más saturan son la tasa de mortalidad y las tasas de población infectada en la primera y segunda oleada. La temperatura acompaña a las tres variables: a más temperatura menor incidencia de la enfermedad y de las muertes. A este respecto se ha generado una tipología *ad hoc* tomando en cada componente las puntuaciones por encima y por debajo de cero, tal como se indica en la tabla siguiente.

El resultado se recoge en el mapa.

Tipología		Componentes	
Componente 1	Componente 2	<0	>0
<0	<0	1	2
	>0	3	4
>0	<0	5	6
	>0	7	8

- Tipo 6. Lo forman la CM y Cataluña. Las dos CCAA más urbanizadas y con un producto económico por habitante alto. En consonancia con sendas características se produce una elevada movilidad de la población y los parques residenciales contienen proporciones significativas de vivienda pequeña. A la vez, presentan unos índices mediocres de atención sanitaria, menor envejecimiento de la estructura de edades y peso alto de la inmigración extranjera. El impacto de la COVID es acusado.
- Tipo 1. Lo integran Andalucía, Murcia y la Comunidad Valenciana, que ocupa buena parte del litoral mediterráneo. Ofrece un contrapunto al tipo anterior: menor grado de urbanización y desarrollo económico, unidos a unos índices de atención sanitaria bajos, pero con tasas de mortalidad y de contagio reducidas. El grupo de mayores de 65 años pesa poco.
- Tipo 7. Definido por la vertiente cantábrica: Asturias, Cantabria y País Vasco. Ofrece una foto polar respecto de los dos tipos anteriores. Elevada urbanización y desarrollo

económico, más mejores índices de atención sanitaria y menor impacto de la COVID. La inmigración es más reducida que la media y el segmento de los mayores es más abultado.

- Tipo 5. Incorpora ambos archipiélagos. Están en el lado positivo del factor 1 (urbanización, desarrollo económico y movilidad), con unos índices pobres de atención sanitaria y baja o muy baja incidencia de la COVID. En la medida en que su posición en el factor 1 está muy centrada, en realidad, este tipo se asemeja al tipo 1. Completan entre ambos la España litoral más meridional.
- Tipo 3. Incluye a Extremadura y Galicia. Son dos CCAA que en los tres componentes se mueven en valores centrados. Se caracterizan por estar menos urbanizadas y tener un menor desarrollo económico, junto con unos mejores índices de atención sanitaria y una menor mortalidad y extensión de la enfermedad.
- Tipo 4. Confluyen aquí Castilla-León, Navarra y La Rioja. También ofrecen menor urbanización y desarrollo económico, pero mejores indicadores de atención sanitaria y, sin embargo, más muertes y extensión de la enfermedad. El peso de los mayores de 65 años es alto.
- Tipo 7. Acoge a Castilla-La Mancha en solitario. De manera notoria está en valores extremos en los tres componentes. Menor urbanización y actividad económica, peores índices de atención sanitaria y más muertes y enfermedad.
- Tipo 8. También está formado por una sola comunidad, Aragón. Ofrece puntuaciones positivas en los tres componentes, pero centradas, por lo que se asemeja en diversos aspectos al tipo 4.



Fuente: Elaboración propia

Nótese que los tipos pares (2, 4, 6 y 8) abarcan la España interior, mientras que los impares (1, 3, 5 y 7) delimitan la España costera, con la excepción de Cataluña. En la primera, la COVID ha castigado notablemente más que en la segunda. Nótese, asimismo, que en la primera, las temperaturas invernales son menos benignas que en la segunda. La temperatura alcanza saturaciones significativas, aunque no especialmente destacadas, en los componentes 2 y 3. En el componente 2 coincide, además de con los índices de nivel de la atención sanitaria, con el indicador de virulencia, que relaciona las muertes debidas a la COVID con la población que ha sufrido la enfermedad. Los signos son, respectivamente, negativo y positivo: a más temperatura menos virulencia. En el componente 3, la temperatura tiene signo negativo, de forma que se puede leer: a más temperatura menor mortalidad y menor extensión de la enfermedad.

Este último enfoque nos coloca en la perspectiva del análisis causal propiamente dicho. ¿Qué determina las mayores-menores tasas de mortalidad debida a la COVID? Se observa que esta variable (medida por el logaritmo neperiano) presenta correlaciones bivariadas significativas únicamente con la temperatura (excepción hecha de los índices de prevalencia, como es obvio): a más temperatura menos muertes y viceversa. Con el resto de las variables, la asociación, medida por la correlación de Pearson, o es muy débil o no existe. Así, volviendo sobre los análisis precedentes, se tiene que hay CCAA con índices altos de urbanización y desarrollo económico y, a su vez, con tasas altas y con tasas bajas de mortalidad. Lo mismo sucede a propósito de las CCAA con pirámides de edad envejecidas o de las CCAA con buenos indicadores de atención sanitaria. La tipología de CCAA que se acaba de extraer es bien ilustrativa al respecto. Esta constatación es de suma importancia, toda vez que apunta a que la interrelación de las variables no es lineal.

Para andar por esta senda se ha abordado un análisis de regresión múltiple. Se toma el logaritmo de la tasa de mortalidad como variable dependiente. La aplicación de la técnica presenta una severa dificultad: el número de casos (CCAA) es pequeño. Hemos ensayado diversos modelos de agrupación de las variables explicativas aplicando el método “hacia atrás” de selección de variables. Uno de los modelos que mejores (en el sentido de que cumple los requerimientos de la estimación mínimo-cuadrática³) resultados arroja es el que se sintetiza en el recuadro adjunto. Muestra un muy alto valor predictivo de la tasa de mortalidad. La R^2 (proporción de la varianza explicada de la variable dependiente) es 0,95 y la R^2 corregida, aunque se ve algo penalizada por el volumen de variables empleadas en relación al número de casos, es también muy alta, 0.88.

Modelo 1: MCO, usando las observaciones 1-17 Variable dependiente: I_mortalidad					
	Coeficiente	Desv. Típica	Estadístico t	valor p	
Const	14.5396	4.15929	3.496	0.0101	**
Est50	0.0224803	0.00935804	2.402	0.0473	**
MA65	-0.256706	0.0646212	-3.972	0.0054	***
PIBpc	0.000113456	5.99332e-05	1.893	0.1002	
Ocup_tem	-0.234241	0.0560977	-4.176	0.0042	***
Sumerg	0.310490	0.0673332	4.611	0.0025	***
Telet19	-0.614701	0.163920	-3.750	0.0072	***
Temperatura	-0.239034	0.0804584	-2.971	0.0208	**
I_turistas	-0.668188	0.366721	-1.822	0.1112	
Inmigr_noUE	-0.335467	0.0822159	-4.080	0.0047	***

³ El modelo satisface la prueba de significancia general de los coeficientes, pues el valor p de F es 0,001. Asimismo, no presenta un problema grave de colinealidad: solo un indicador (el de PIBpc) supera ligeramente el valor crítico 10 de FIV (factor de inflación de la varianza). No hay heteroscedasticidad, según la prueba de Breusch-Pagan-Godfrey: se obtiene un valor de 6,73, inferior al crítico de 16,92. Los residuos son normales, sobre los contrastes de Jarque-Bera y de Shapiro-Wilk. Por último, no se detectan casos atípicos ni influyentes, las distancias de Cook son inferiores a 1 en todos los casos, las distancias de Mahalanobis tienen probabilidades altas, ninguna observación arroja un residuo tipificado superior a 3 veces la desviación y, en fin, el DFFITS está por debajo del valor crítico 2,61, salvo Baleares que lo supera ligeramente.

Media de la vble. dep.	4.076025	D.T. de la vble. dep.	0.988238
Suma de cuad. residuos	0.820939	D.T. de la regresión	0.342457
R-cuadrado	0.947463	R-cuadrado corregido	0.879915
F(9, 7)	14.02653	Valor p (de F)	0.001070
Log-verosimilitud	1.637464	Criterio de Akaike	16.72507
Criterio de Schwarz	25.05721	Crit. de Hannan-Quinn	17.55330

Fuente: Elaboración propia

Es interesante fijar la atención en los coeficientes beta. Indican cuánto cambio estandarizado se produce en la variable explicada -la tasa de mortalidad- por un cambio estandarizado en cada una de las variables explicativas al controlar en relación con las otras. Al estandarizar todas las variables quedan expresadas en una misma base, de tal modo que los coeficientes beta miden el sentido y la importancia de cada variable independiente sobre la dependiente.

En el modelo entran 9 variables. De ellas, tres (PIBpc, Sumerg y Est50) tienen signo positivo, es decir, la mortalidad debida a la COVID durante la primera oleada sube al hacerlo el grado de urbanización, al hacerlo el producto bruto por habitante y al hacerlo el peso de la economía sumergida. Cataluña y la Comunidad de Madrid ejemplifican lo que sucede con las dos primeras, mientras que las dos Castillas ilustran lo que sucede con la tercera. A su vez, 6 variables presentan signo negativo, es decir, la tasa de mortalidad es más baja al aumentar el peso de la población mayor de 65 años, la proporción de inmigrantes no comunitarios, la contratación eventual, el teletrabajo, la temperatura y la densidad del turismo durante el primer trimestre de 2020.

	Beta
Est50	0,317280207
Más65	-0,845968373
Inmigr_noUE	-0,935739903
PIBpc	0,573444272
Ocup_tem	-1,030247929
Telet19	-0,461331235
Temperatura	-0,665938912
I_turistas	-0,281102396
Sumerg	1,143221867

Fuente: Elaboración propia

El hecho de que en el modelo no se hallen otras variables, como, por ejemplo, las relativas a la atención sanitaria, no significa que no tengan importancia o que no incidan sobre el comportamiento de la mortalidad. La ausencia se puede deber a que su influencia está recogida por una o varias de las variables del modelo. En particular, tenemos que los indicadores de atención sanitaria (en concreto los referidos al número de camas por 1.000 habitantes, personal sanitario por 1.000 habitantes y gasto sanitario por habitante) tienen -en otros modelos alternativos- signo positivo, de forma que a pesar de disponer de una dotación sanitaria mejor (al menos, sobre la base de los indicadores elegidos) son varias las CCAA con tasas altas de mortalidad. Tal es el caso de Castilla-León, Navarra, La Rioja o Aragón. Otro grupo de variables que no aparece en el modelo es uno indicativo de las condiciones de vida: el tamaño de la vivienda.

También tienen signo positivo. Igual sucede con el índice de densidad de la matrícula universitaria sobre la población.

Más atrás se ha visto que la tasa de mortalidad es una función de la prevalencia y de la virulencia. Hay más muertes cuando hay más personas infectadas y cuando, con independencia del número de enfermos, el virus es más dañino y/o las personas son menos resistentes. Para profundizar en estas cuestiones se ha repetido el análisis de regresión múltiple tomando sucesivamente como variables dependientes el logaritmo de la prevalencia en la primera oleada, la prevalencia en la segunda oleada y la virulencia. Los modelos, en lo fundamental, reiteran las observaciones anteriores. Los tres tienen una gran capacidad predictiva. Los coeficientes de determinación son altos, incluso después de la corrección en relación al número de variables. Aparece por primera vez el indicador de movilidad: a más movimientos de la población, mayor extensión de la enfermedad y más virulencia.

Modelo 2: MCO, usando las observaciones 1-17				
Variable dependiente: I_Prevalencia1				
	Coeficiente	Desv. Típica	Estadístico t	valor p
const	13.0649	1.82602	7.155	<0.0001 ***
MAs65	-0.241645	0.0484765	-4.985	0.0005 ***
Inmigr_noUE	-0.210820	0.0510605	-4.129	0.0020 ***
Ocup_tem	-0.175707	0.0344650	-5.098	0.0005 ***
Sumerg	0.0842637	0.0322144	2.616	0.0258 **
Telet19	-0.244040	0.107251	-2.275	0.0461 **
Temperatura	-0.133146	0.0424442	-3.137	0.0106 **
Media de la vble. dep.	1.287159	D.T. de la vble. dep.	0.597204	
Suma de cuad. Residuos	0.784180	D.T. de la regresión	0.280032	
R-cuadrado	0.862580	R-cuadrado corregido	0.780128	
F(6, 10)	10.46158	Valor p (de F)	0.000807	
Log-verosimilitud	2.026854	Criterio de Akaike	9.946293	
Criterio de Schwarz	15.77879	Crit. de Hannan-Quinn	10.52605	

Fuente: Elaboración propia

Modelo 3: MCO, usando las observaciones 1-17				
Variable dependiente: Prevalencia2				
	Coeficiente	Desv. Típica	Estadístico t	valor p
const	28.8697	16.0823	1.795	0.1001
MAs65	-1.73718	0.423769	-4.099	0.0018 ***
Inmigr_noUE	-0.765268	0.406185	-1.884	0.0862 *
Sumerg	0.892679	0.405977	2.199	0.0502 *
Temperatura	-2.15224	0.403497	-5.334	0.0002 ***
Movilidad	0.909257	0.336899	2.699	0.0207 **
Media de la vble. dep.	9.129412	D.T. de la vble. dep.	4.278254	
Suma de cuad. Residuos	67.37658	D.T. de la regresión	2.474903	
R-cuadrado	0.769932	R-cuadrado corregido	0.665356	
F(5, 11)	7.362398	Valor p (de F)	0.002944	
Log-verosimilitud	-35.82717	Criterio de Akaike	83.65434	
Criterio de Schwarz	88.65362	Crit. de Hannan-Quinn	84.15128	

Modelo 4: MCO, usando las observaciones 1-17 Variable dependiente: Virulencia				
	Coeficiente	Desv. Típica	Estadístico t	valor p
const	-2621.20	2330.19	-1.125	0.2810
Sumerg	154.751	55.6350	2.782	0.0156 **
Temperatura	-190.456	41.6689	-4.571	0.0005 ***
Movilidad	93.6109	45.2486	2.069	0.0591 *
Media de la vble. dep.	1322.343	D.T. de la vble. dep.		569.1082
Suma de cuad. Residuos	1920432	D.T. de la regresión		384.3508
R-cuadrado	0.629414	R-cuadrado corregido		0.543894
F(3, 13)	7.359852	Valor p (de F)		0.003946
Log-verosimilitud	-123.0182	Criterio de Akaike		254.0363
Criterio de Schwarz	257.3692	Crit. de Hannan-Quinn		254.3676

Fuente: Elaboración propia

Beta			
	Prevalencia 1	Prevalencia 2	Virulencia
Más 65	-1,31775624	-1,322381783	
Inmigr_NoUE	-0,97309769	-0,493076544	
Ocup_tem	-1,27881575		
Sumerg	0,513408889	0,759231017	0,98942565
Telet19	-0,30307383		
Temperatura	-0,61382109	-1,385032339	-0,92137429
Movilidad		-0,865042789	0,66949777

Fuente: Elaboración propia

Resumiendo, la regresión múltiple señala:

- Ni la inmigración, ni el turismo parecen haber incidido de manera positiva en la tasa de mortalidad. Sendas variables tienen signo negativo.
- Sí lo hace, con una intensidad baja, la matriculación universitaria.
- La temperatura es un factor destacado: a más temperatura menos enfermos, menos virulencia y menos muertes.
- El grado de urbanización tira de la mortalidad hacia arriba. Hay más contactos, en las ciudades grandes hay más viviendas pequeñas y la población se mueve más de unas áreas a otras, especialmente por motivo trabajo. También hay más contaminación.
- En la medida en que más urbanización (más movilidad, más viviendas pequeñas...) va asociada a más desarrollo económico, resulta que el PIBpc también influye empujando a la tasa de mortalidad. El aporte de Cataluña y la Comunidad de Madrid es claro.
- El envejecimiento de la pirámide de edades no parece ser un factor que incida positivamente sobre la tasa. Al contrario, el indicador aparece en los cuatro modelos con coeficientes beta altos y con signo negativo. Sucede que varias de las CCAA más envejecidas (Asturias, Galicia, País Vasco, Cantabria...) tienen tasas de mortalidad bajas, mientras que otras, como la CM o Cataluña, con menos envejecimiento tienen tasas de mortalidad altas.

- El peso de la economía sumergida también tira de la mortalidad. Revela más informalidad, menos organización social, mayor representación de sectores de actividad como la agricultura y determinados servicios personales. Tienen un peso elevado en las dos Castillas, Extremadura, aunque también en CCAA con baja mortalidad como Canarias y Andalucía.
- Los indicadores referidos a la atención sanitaria no aparecen en los modelos. Por la observación de otros modelos alternativos y considerando las correlaciones de dichos indicadores con el resto se tiene que hay más mortalidad en varias de las CCAA que ofrecen más atención sanitaria. Sin duda, el resultado así expuesto es contraintuitivo y llevaría a una conclusión absurda.

Esta última apreciación, unida a las numerosas excepciones que aparecen en las regresiones anteriores, hay que ponerla en la perspectiva de una de las conclusiones principales que se extrae del análisis factorial: la causalidad de la enfermedad y la mortalidad debida a la COVID no obedece a un modelo estrictamente lineal.

Tenemos, por tanto, que la mortalidad debida a la COVID es impulsada por el grado de urbanización. Ésta va asociada a una probabilidad mayor de relaciones interpersonales, más proximidad de unos con otros... y, en consecuencia, una probabilidad superior de ocurrencia de contactos en los que es posible la transmisión de la enfermedad. Hay más movilidad de la población dentro de la propia comunidad y entre comunidades, sea por motivo trabajo, sea por otros motivos. Aunque se ha de huir de la demonización del transporte público, cabe sospechar de la seguridad del mismo durante las grandes afluencias de viajeros/as en las horas punta. Es probable, en todo caso, que haya más riesgo en los locales de ocio nocturno que en los vagones del metro o del tren, sobre todo si éstos disponen de ventilación de aire, desinfección diaria y es obligatorio el uso de la mascarilla⁴. También en las ciudades, especialmente en las grandes y en las muy grandes, hay más contaminación atmosférica. Diversos estudios⁵ -aun no concluyentes- sugieren que ésta puede ser un factor de propagación y/o de empeoramiento de los efectos del coronavirus: difusión del virus en las partículas contaminantes en suspensión y/o mayor riesgo cardio respiratorio de las personas expuestas a altas dosis de contaminación. El hacinamiento residencial -viviendas pequeñas o elevado número de ocupantes por metro cuadrado de las residencias- es un hecho característico de los barrios pobres de las ciudades.

El comportamiento de determinados colectivos poblacionales puede tener una relación con la extensión del virus. Lamentablemente esta vía de análisis se ha visto entorpecida por el uso torticero que de ella han pretendido determinados discursos racistas y xenófobos. Ni el análisis factorial (ACP) ni el de regresión han encontrado asociación entre la alta mortalidad y el alto grado de prevalencia con la densidad de turistas por habitante en el primer trienio de 2020 o con la proporción de inmigrantes extranjeros sobre la población total, sea que se tome la totalidad de los inmigrantes, sea que se tome los originarios de fuera de la UE. Al contrario, el signo es negativo⁶. Sí se observa

⁴https://www.consalud.es/pacientes/especial-coronavirus/transporte-publico-covid-19-cuales-posibilidades-reales-contagio_85537_102.html

⁵<https://www.isciii.es/InformacionCiudadanos/DivulgacionCulturaCientifica/DivulgacionISCIII/Paginas/Divulgacion/ContaminacionCoronavirusDivulgacion.aspx>

⁶ *Aquí es necesario hacer una matización relevante. La regresión múltiple y el ACP no muestran que las CCAA con más inmigrantes sufran más la enfermedad o arrojen una tasa de mortalidad más alta, que es la variable dependiente en el presente estudio. Cuestión distinta es que el segmento de los inmigrantes presenta una propensión mayor a contagiarse. En el estudio del IS Carlos III tenemos que la prevalencia, según nacionalidad, en la cuarta ronda (diciembre de 2020)*

alguna incidencia en la propagación de la enfermedad con el volumen de universitarios/as relativo a la población. La posible incidencia de los/as universitarios/as parece deberse no a la asistencia a clase, sino a la participación en fiestas y actividades de ocio sin guardar las debidas medidas de seguridad. También se observa una influencia positiva en la relación con la economía sumergida, indicativa de la presencia de colectivos excluidos de los mercados formales de trabajo, el peso de actividades agrícolas y de servicios personales y, tal vez, asociada a una estructura social menos organizada, menos vertebrada.

La temperatura aparece como un factor trascendente en todos los modelos: a más temperatura menos incidencia de la enfermedad y de la mortalidad. El rigor de las temperaturas invernales parece haber tirado de las consecuencias negativas de la COVID en la España interior, frente a las temperaturas más benignas de la España costera, tanto más en la mitad meridional. Los estudios que se han realizado sobre el asunto no son concluyentes: puede que el virus se propague menos en climas húmedos y cálidos. También puede ocurrir que, en la medida que las temperaturas más altas favorecen que la gente esté en la calle, no en sitios cerrados, aquéllas determinen de manera indirecta una menor tasa de contagios y de muertes⁷.

Obviamente una mejor dotación sanitaria tiene que contribuir a reducir la mortalidad debida a la COVID, si bien no necesariamente tiene influencia negativa en el número de contagios. El hecho de que en el análisis de regresión múltiple no salgan los indicadores de atención sanitaria o que, en los modelos en que sí lo hacen, lo hagan con signo positivo, constituye un desafío a la hora de buscar la interpretación del dato. Incluso se observa una correlación bivariada significativa (0,53) entre el logaritmo de la mortalidad y la densidad del personal sanitario sobre la población. Esto último tal vez está asociado al hecho de que el personal sanitario es uno de los colectivos más castigados por la enfermedad.

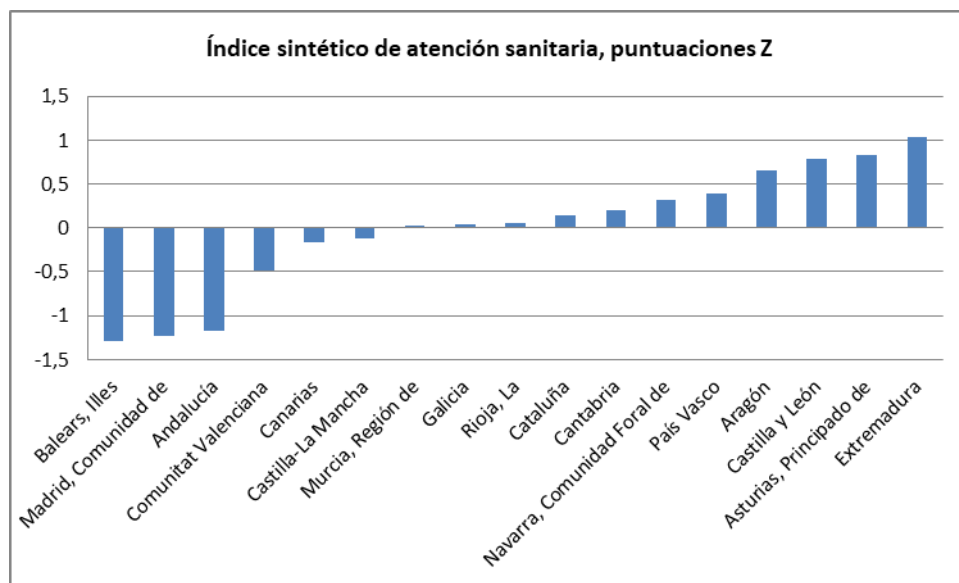
Hemos estimado dos índices sintéticos de atención sanitaria. El primero viene dado por el promedio de las puntuaciones Z de los cinco indicadores de atención que se incluye en la lista de variables más arriba descrita⁸. El segundo resulta de promediar la homogenización de los cinco indicadores⁹. Ambos índices, en realidad, ofrecen la misma foto. Se observa que hay CCAA con mortalidad alta y un índice sintético alto, como Castilla-León, y hay CCAA con mortalidad alta y un índice bajo (negativo), como la Comunidad de Madrid. La viceversa también es cierta: CCAA con mortalidad baja y atención sanitaria alta, como Asturias, y CCAA con mortalidad baja y atención sanitaria con signo menos, como Andalucía o Baleares.

es un 9,7% para los españoles/as y un 13,1% para las personas con otras nacionalidades. La población inmigrante sufre -por término medio- peores condiciones de vida que la hacen más susceptible de contagio: más hacinamiento residencial, más movilidad en transporte público, peor acceso a los recursos sanitarios, más inserción en la economía sumergida... Dicho de otro modo, no se contagian más porque sean inmigrantes, sino porque sufren unas condiciones de vida más precarias.

⁷ <https://www.isciii.es/InformacionCiudadanos/DivulgacionCulturaCientifica/DivulgacionISCIII/Paginas/Divulgacion/InformeClimayCoronavirus.aspx>

⁸ $Z = (\text{Valor-promedio}) / \text{desviación típica}$.

⁹ Índice=valor de cada CA/valor de la CA más alto.



Fuente: Elaboración propia

La inclusión del índice en el análisis de regresión en lugar de los cinco indicadores originales reitera la situación. Aquél no aparece en los modelos y, en los pocos que sí lo hace, apunta que a mejor índice sintético más mortalidad. ¿Qué está sucediendo?

- Puede ocurrir que los indicadores seleccionados no sean los idóneos o no sean suficientes. En particular, el hecho de que una CA, por ejemplo, destine más dinero por habitante a la sanidad pública que otra CA no implica que la atención sanitaria sea mejor. Podría suceder que la segunda gaste con más eficiencia desde el punto de vista sanitario. Por otra parte, cabe suponer que el nivel tecnológico y de conocimientos del personal de, por ejemplo, la red de la Comunidad de Madrid sea superior e incluso muy superior al nivel de Extremadura. Tal vez habría que considerar en la medición de los indicadores dotacionales la aportación de la red privada, si bien lo cierto es que ésta se ha *borrado* en lo que a la COVID se refiere, salvo para contratar-colonizar la prestación de servicios a la red pública.
- Parece bastante verosímil la hipótesis de que el sistema sanitario tiene poca incidencia en la propagación y en la virulencia de la enfermedad. La gente llega a los centros de salud y a los hospitales cuando ya está enferma. Así las cosas, el sistema influye en los datos de mortalidad debida a la COVID de la primera oleada en la medida en que cura a los/as pacientes. Es probable que no haya diferencias apreciables en el rendimiento de los diferentes sistemas de las CCAA frente a la COVID, tanto menos cuanto la tasa de mortalidad que se mide está referida a sucesos que tienen lugar en un periodo corto (marzo a junio de 2020).

Tanto en el epígrafe 1 como en el 2 se ha constatado que la Comunidad de Madrid está en el centro de la gran mancha de provincias con tasas de mortalidad debida a la COVID y de prevalencia altas. Este dato sugiere que la CM puede tirar de la tasa de las CCAA vecinas y/o que éstas tiran de la tasa madrileña. Se producen flujos voluminosos de intercambios poblacionales con ambas Castillas. Muchos/as residentes en ellas vienen a trabajar diariamente a Madrid. A su vez, muchos/as residentes en la CM se desplazan los fines de semana a las segundas residencias. Incluso, un volumen indeterminado de familias de la CM durante los primeros meses de la pandemia trasladó provisional o

definitivamente su domicilio a municipios castellanos. Algo parecido -aunque con bastante menos intensidad- ocurre en la interrelación de Cataluña con las comunidades vecinas. A fin de evaluar esta cuestión se ha generado una variable artificial (distancia): vale 1 para Cataluña, la CM y las CCAA fronterizas con ellas, vale 2 para el resto, salvo Canarias y Baleares, a las que se asigna el valor 3.

El análisis de regresión muestra los resultados del modelo 5 adjunto¹⁰. La nueva variable, en efecto, aparece en la ecuación con signo menos: a más distancia (medida en los términos de la variable artificial) menos mortalidad. La cercanía a la CM y a Cataluña produce espacios de difusión del virus. Entre cada comunidad y las vecinas tiene lugar un fenómeno de exportación-importación. Obviamente la relación es bilateral: el virus va y viene.

Más aun, el promedio (no ponderado) de los logaritmos de la mortalidad en las CCAA no fronterizas es 3,6 y 5,1 el promedio de los logaritmos de las CCAA fronterizas (incluidas la CM y Cataluña). La diferencia de medias es significativa, con una F (razón entre la varianza intergrupos y la varianza intragrupos) igual a 13,8 (sig = 0,002).

Modelo 5: MCO, usando las observaciones 1-17				
Variable dependiente: I mortalidad				
	Coefficiente	Desv. Típica	Estadístico t	valor p
Const	8.75135	2.89958	3.018	0.0295 **
Est50	0.0121661	0.00581873	2.091	0.0908 *
MAs65	-0.255301	0.0385787	-6.618	0.0012 ***
Inmigr_noUE	-0.315856	0.0465518	-6.785	0.0011 ***
PIBpc	0.000129316	3.38130e-05	3.824	0.0123 **
Ocup_tem	-0.142584	0.0385958	-3.694	0.0141 **
Sumerg	0.332541	0.0500430	6.645	0.0012 ***
Telet19	-0.484856	0.0989211	-4.901	0.0045 ***
Temperatura	-0.252460	0.0527513	-4.786	0.0049 ***
Movilidad	0.0924295	0.0405460	2.280	0.0716 *
I_turistas	-0.608415	0.213525	-2.849	0.0358 **
Distancia	-0.291587	0.108437	-2.689	0.0434 **
Media de la vble. dep.	4.076025	D.T. de la vble. dep.		0.988238
Suma de cuad. Residuos	0.183404	D.T. de la regresión		0.191523
R-cuadrado	0.988263	R-cuadrado corregido		0.962441
F(11, 5)	38.27220	Valor p (de F)		0.000416
Log-verosimilitud	14.37688	Criterio de Akaike		-4.753770
Criterio de Schwarz	5.244790	Crit. de Hannan-Quinn		-3.759893

Fuente: Elaboración propia

¹⁰ El modelo tiene colinealidad, al tiempo que es homoscedástico y los residuos superan la prueba de normalidad.



MADRID