

Modelo econométrico sobre la incidencia del coronavirus en Andalucía frente a Granada

Econometría II

Álvaro Carbonero Román

Juan Antonio Molina Diaz



Índice de contenido

1. Introducción.....	3
1.1. Resumen.....	3
1.2. Objetivos del trabajo.....	3
1.3. Motivos de elección del tema.....	3
1.4. Estudios previos.....	3
1.5. Hipótesis de partida.....	3
1.6. Fuentes empleadas para la obtención de los datos.....	3
1.7. Definición de las variables empleadas en los modelos y su forma de medida.....	5
2. Análisis.....	6
2.1. Modelo Gompertz.....	6
2.2. Definición de los parámetros.....	6
2.3. Análisis de Andalucía.....	7
2.4. Análisis de Granada.....	8
2.5. Comparativa.....	10
3. Conclusiones.....	12
3.1. Aportaciones personales.....	12
4. Bibliografía.....	13

Introducción

Resumen

En este documento hemos tratado de analizar la situación del coronavirus durante la llamada “segunda ola” mediante los datos ofrecidos por la Junta de Andalucía a través de su página web.

Objetivos del trabajo

El objetivo del trabajo será comparar los casos por cada 100.000 habitantes entre Granada y Andalucía, y ver si existe una relación directa entre unos y otros. También trataremos de averiguar cuándo se produjo el pico de la segunda ola en Granada, ya que los medios informaron de que fue el 10 de noviembre, un día después del pico en Andalucía. (González, 2020).

Motivos de elección del tema

Hemos escogido este tema porque creemos que es lo suficientemente interesante y actual. Además, de ser el resultado que Granada tiene una incidencia por encima de lo normal, podríamos intentar darle una explicación.

Estudios previos

Los datos recogidos de la página web de la Junta de Andalucía vienen dados en casos diarios. El modelo con el que vamos a trabajar (modelo de Gompertz), sin embargo, los utiliza de forma acumulada. Por ello hemos recopilado una tabla con todos los datos de Andalucía y Granada. Son datos cuantitativos de corte temporal.

Hipótesis de partida

Teóricamente, Granada debería tener una incidencia proporcional a su población de la misma forma que la tienen el resto de provincias, y en última instancia, Andalucía.

Fuentes empleadas para la obtención de datos

En el siguiente Excel se ha recogido los datos proporcionados por la Junta de Andalucía de la evolución de contagios por coronavirus tanto de Andalucía como de Granada. Los datos están separados por distintas columnas, significando cada uno de ellos lo siguiente:

1. Columna A. Fecha de la recogida de datos comprendida entre el 1 de octubre y el 30 de noviembre del año 2020
2. Columna B. Contagios diarios registrados en Andalucía
3. Columna C. Contagios diarios registrados en la provincia de Granada
4. Columna E. Cantidad de días que duró la conocida “Segunda Ola”
5. Columna F. Contagios acumulados en Andalucía. En esta columna es el resultado de la suma de los contagios de un día en cuestión y los días previos.
6. Columna G. Contagios acumulados en Granada.
7. Columna H. Incidencia acumulada de casos registrados por cada 100.000 habitantes en Andalucía
8. Columna I. Incidencia acumulada de casos registrados por cada 100.000 habitantes en Granada

	A	B	C	D	E	F	G	H	I
1	01/10/2020	1687	370		1	1687	370	19,93051	40,25379
2	02/10/2020	1767	315		2	3454	685	40,80615	74,52392
3	03/10/2020	1252	306		3	4706	991	55,59749	107,8149
4	04/10/2020	968	129		4	5674	1120	67,03361	121,8493
5	05/10/2020	1248	251		5	6922	1371	81,77769	149,1566
6	06/10/2020	1941	537		6	8863	1908	104,709	207,579
7	07/10/2020	2046	520		7	10909	2428	128,8808	264,1519
8	08/10/2020	2413	517		8	13322	2945	157,3884	320,3984
9	09/10/2020	2616	593		9	15938	3538	188,2943	384,9133
10	10/10/2020	1884	436		10	17822	3974	210,5522	432,3475
11	11/10/2020	1102	162		11	18924	4136	223,5714	449,9721
12	12/10/2020	1022	197		12	19946	4333	235,6455	471,4046
13	13/10/2020	1801	431		13	21747	4764	256,9228	518,2948
14	14/10/2020	2945	720		14	24692	5484	291,7155	596,6265
15	15/10/2020	3466	940		15	28158	6424	332,6634	698,8929
16	16/10/2020	3248	878		16	31406	7302	371,0359	794,4141
17	17/10/2020	2529	607		17	33935	7909	400,9139	860,4521
18	18/10/2020	1280	188		18	35215	8097	416,036	880,9053
19	19/10/2020	2693	816		19	37908	8913	447,8516	969,6813
20	20/10/2020	3758	1036		20	41666	9949	492,2493	1082,392
21	21/10/2020	4543	1196		21	46209	11145	545,921	1212,51
22	22/10/2020	4622	1181		22	50831	12326	600,5261	1340,995
23	23/10/2020	4688	1067		23	55519	13393	655,911	1457,079
24	24/10/2020	3165	858		24	58684	14251	693,3028	1550,424
25	25/10/2020	1671	255		25	60355	14506	713,0443	1578,166
26	26/10/2020	3569	921		26	63924	15427	755,2091	1678,366
27	27/10/2020	4483	1069		27	68407	16496	808,172	1794,666
28	28/10/2020	5033	1277		28	73440	17773	867,6327	1933,596
29	29/10/2020	4949	1205		29	78389	18978	926,1011	2064,693
30	30/10/2020	4961	1091		30	83350	20069	984,7112	2183,388
31	31/10/2020	3135	648		31	86485	20717	1021,749	2253,886

32	01/11/2020	2094	239			32	88579	20956	1046,487	2279,888
33	02/11/2020	1732	260			33	90311	21216	1066,95	2308,174
34	03/11/2020	3915	971			34	94226	22187	1113,202	2413,813
35	04/11/2020	4770	1126			35	98996	23313	1169,556	2536,315
36	05/11/2020	4743	1046			36	103739	24359	1225,59	2650,114
37	06/11/2020	5030	1163			37	108769	25522	1285,016	2776,641
38	07/11/2020	3190	482			38	111959	26004	1322,703	2829,08
39	08/11/2020	1899	196			39	113858	26200	1345,138	2850,404
40	09/11/2020	3756	688			40	117614	26888	1389,512	2925,254
41	10/11/2020	4361	937			41	121975	27825	1441,034	3027,194
42	11/11/2020	4423	940			42	126398	28765	1493,288	3129,461
43	12/11/2020	4285	815			43	130683	29580	1543,911	3218,128
44	13/11/2020	4163	809			44	134846	30389	1593,094	3306,142
45	14/11/2020	2306	367			45	137152	30756	1620,337	3346,069
46	15/11/2020	1365	134			46	138517	30890	1636,464	3360,648
47	16/11/2020	2466	449			47	140983	31339	1665,597	3409,496
48	17/11/2020	3385	602			48	144368	31941	1705,588	3474,99
49	18/11/2020	3207	487			49	147575	32428	1743,476	3527,973
50	19/11/2020	2922	471			50	150497	32899	1777,997	3579,215
51	20/11/2020	2742	504			51	153239	33403	1810,392	3634,047
52	21/11/2020	1600	100			52	154839	33503	1829,294	3644,927
53	22/11/2020	796	59			53	155635	33562	1838,699	3651,346
54	23/11/2020	1471	294			54	157106	33856	1856,077	3683,331
55	24/11/2020	2151	314			55	159257	34170	1881,489	3717,492
56	25/11/2020	1856	279			56	161113	34449	1903,417	3747,846
57	26/11/2020	1750	202			57	162863	34651	1924,091	3769,822
58	27/11/2020	1842	205			58	164705	34856	1945,853	3792,125
59	28/11/2020	943	41			59	165648	34897	1956,994	3796,586
60	29/11/2020	467	44			60	166115	34941	1962,511	3801,373
61	30/11/2020	986	161			61	167101	35102	1974,16	3818,888

Tabla 1. Elaboración propia a partir de los datos proporcionados por la Junta de Andalucía

Definición de las variables empleadas en los modelos y su forma de medida

El modelo de Gompertz funciona con datos acumulados, así que a partir de los de la Junta de Andalucía los hemos acumulado día a día.

Por otro lado, para poder comparar los dos modelos (Andalucía y Granada), se han calculado los casos por cada 100.000 habitantes, que además es la cifra a la que estamos acostumbrados por la televisión.

Análisis

Modelo Gompertz

El modelo elegido para nuestro trabajo ha sido el modelo de Gompertz, que ha sido el más utilizado en las previsiones del futuro de coronavirus en países como China, y que suelen tener resultados correctos y correspondientes a la realidad. La curva de Gompertz es una modificación de la logística, pero esta última asume que la fase inicial y final se producen a la misma velocidad y el de Gompertz no.

El modelo de Gompertz sigue la siguiente función (expresada en la notación usada en el programa informático Gretl):

$$C = b_1 * \exp(-\exp(-b_2 * (t - b_3)))$$

Expresada de forma más “natural”, nos quedaría de la siguiente forma:

$$C = \beta_1 e^{(-e^{-\beta_2(t-\beta_3)})}$$

Donde C es el número de casos por cada 100.000 habitantes.

Definición de los parámetros

En este modelo se ha tenido en cuenta el uso de 3 parámetros, cuya definición es la siguiente:

- β_1 : Máximo natural de casos por cada 100.000
- β_2 : Tasa de crecimiento de la enfermedad.
- β_3 : Día en el que se alcanza el máximo crecimiento.

Todos los parámetros son significativos tanto global como individualmente.

Análisis de Andalucía

Modelo 1: MC. no lineales, usando las observaciones 1-61

$$C = b1 * \exp(-\exp(-b2 * (t - b3)))$$

	Estimación	Desv. típica	Estadístico t	valor p
b1	2416.93	23.4772	102.9	2.28e-067 ***
b2	0.0522804	0.000813845	64.24	1.36e-055 ***
b3	28.3614	0.239636	118.4	7.28e-071 ***
Media de la vble. dep.	1002.685	D.T. de la vble. dep.	665.9184	
Suma de cuad. residuos	21451.39	D.T. de la regresión	19.23152	
R-cuadrado no centrado	0.999194	R-cuadrado centrado	-0.001962	
Log-verosimilitud	-265.3667	Criterio de Akaike	536.7334	
Criterio de Schwarz	543.0660	Crit. de Hannan-Quinn	539.2152	

GNR: R-squared = 0, max |t| = 3.70836e-008

La convergencia parece estar razonablemente completa

VARIABLE	COEFICIENTE	INTERVALO DE CONFIANZA 95%	
b1	2416.93	2369.94	2463.93
b2	0.0522804	0.0506513	0.0539095
b3	28.3614	27.8818	28.8411

En el test de significación global ($34951,1489 > 3,93$ para una confianza del 95%) descartamos la hipótesis nula de no significación y aceptamos la alternativa.

$$\begin{cases} H_0 \equiv \beta_1 = \beta_2 = \beta_3 = 0 \equiv \text{No significativa globalmente} \\ H_1 \equiv \beta_1 \neq 0; \beta_2 \neq 0; \beta_3 \neq 0 \equiv \text{Sí significativa globalmente} \end{cases}$$

$$F_{exp} = \frac{\frac{R^2}{k-1}}{\frac{1-R^2}{n-k}} = \frac{\frac{0,999194}{3-1}}{\frac{1-0,999194}{61-3}} = 34951,1489 > F_{2,58,0,975} = 3,93 \Rightarrow \text{Rechazo } H_0$$

Individualmente, todos los parámetros son significativos (102.9, 64.24, 118,4 > 2 para una confianza del 95%), por lo que podemos empezar con el análisis del modelo.

$$\begin{cases} H_0 \equiv \beta_i = 0 \equiv \text{No significativa individualmente} \\ H_1 \equiv \beta_i \neq 0 \equiv \text{Sí significativa individualmente} \end{cases}$$

$$t_{exp} = \frac{\hat{\beta}_i}{\sqrt{\text{Var } \hat{\beta}_i}} > t_{n-k; 1-\frac{\alpha}{2}} \Rightarrow \text{Rechazo } H_0$$

$$\widehat{\beta}_1 \Rightarrow t_{exp} = \frac{2416,93}{23,4772} = 102,9 > t_{58;0,975} = 2 \Rightarrow \text{Rechazo } H_0$$

$$\widehat{\beta}_2 \Rightarrow t_{exp} = \frac{0,0522804}{0,000813845} = 64,24 > t_{58;0,975} = 2 \Rightarrow \text{Rechazo } H_0$$

$$\widehat{\beta}_3 \Rightarrow t_{exp} = \frac{28,3514}{0,239636} = 118,4 > t_{58;0,975} = 2 \Rightarrow \text{Rechazo } H_0$$

El primer parámetro estima que el máximo de casos por cada 100.000 en Andalucía será de 2417 aproximadamente. Comparándolo con los datos ofrecidos por la junta, se quedan lejos de lo alcanzado, donde es tan sólo 1974,16.

El segundo parámetro da como valor estimado de la infectividad 0.05, o porcentualmente hablando, un 5% de infectividad. La infectividad es la velocidad con la que se expande la enfermedad.

El tercer parámetro estima que a partir del día n=28, 28 de octubre, los contagios por cada 100.000 crecerán de una forma menos acelerada. Esto quiere decir que el máximo de la función se encuentra en ese punto.

Análisis de Granada

Modelo 1: MC. no lineales, usando las observaciones 1-61
 $C = b1 * \exp(-\exp(-b2 * (t - b3)))$

	Estimación	Desv. típica	Estadístico t	valor p
b1	4314.70	35.2099	122.5	9.76e-072 ***
b2	0.0631092	0.00106107	59.48	1.11e-053 ***
b3	24.5572	0.185063	132.7	9.80e-074 ***
Media de la vble. dep.	2073.452	D.T. de la vble. dep.	1326.132	
Suma de cuad. residuos	122098.6	D.T. de la regresión	45.88190	
R-cuadrado no centrado	0.998843	R-cuadrado centrado	-0.005090	
Log-verosimilitud	-318.4074	Criterio de Akaike	642.8148	
Criterio de Schwarz	649.1475	Crit. de Hannan-Quinn	645.2967	

GNR: R-squared = 5.55112e-016, max |t| = 1.71199e-007
 La convergencia parece estar razonablemente completa

VARIABLE	COEFICIENTE	INTERVALO DE CONFIANZA 95%	
b1	4314.70	4244.22	4385.18
b2	0.0631092	0.0609853	0.0652332
b3	24.5572	24.1868	24.9277

De nuevo, procedemos a explicar el nuevo modelo aplicado a Granada de forma similar a como hemos hecho para explicar el modelo de la evolución del coronavirus durante la segunda ola en Andalucía. Aquí podemos ver como también se descarta la hipótesis nula que supone la no existencia tanto de significación individual de los parámetros como la significación global de todo el modelo, con un nivel de significación del 5%.

En el caso de la significación individual, el valor experimental de los betas son, respectivamente: 122,5, 59,48 y 132,7, que, comparándolos con una t de Student teórica con grados de libertad 58 y 0,975 cuyo resultado es aproximadamente 2, podemos decir que todos los valores experimentales son mayor que el valor teórico, por lo que se rechaza la hipótesis nula y podemos decir que los tres parámetros son individualmente significativos con un nivel de significación del 5%.

$$\begin{cases} H_0 \equiv \beta_i = 0 \equiv \text{No significativa individualmente} \\ H_1 \equiv \beta_i \neq 0 \equiv \text{Sí significativa individualmente} \end{cases}$$

$$t_{exp} = \frac{\hat{\beta}_i}{\sqrt{Var \hat{\beta}_i}} > t_{n-k; 1-\frac{\alpha}{2}} \Rightarrow \text{Rechazo } H_0$$

$$\hat{\beta}_1 \Rightarrow t_{exp} = \frac{4314,70}{35,2099} = 122,5 > t_{58;0,975} = 2 \Rightarrow \text{Rechazo } H_0$$

$$\hat{\beta}_2 \Rightarrow t_{exp} = \frac{0,0631092}{0,00106197} = 59,48 > t_{58;0,975} = 2 \Rightarrow \text{Rechazo } H_0$$

$$\hat{\beta}_3 \Rightarrow t_{exp} = \frac{24,5572}{0,185063} = 132,7 > t_{58;0,975} = 2 \Rightarrow \text{Rechazo } H_0$$

En cuanto a la significación global del modelo, debemos obtener el valor F experimental siendo su resultado 25035,82282. El valor teórico en este caso de una F de Snedecor con grados de libertad 2, 58 y 0,975 tiene un valor aproximado de 3,93, siendo este muchísimo más pequeño que el valor experimental, por lo que se rechaza la hipótesis nula y podemos afirmar que el modelo es globalmente significativo con un nivel de significación del 5% nuevamente.

$$\begin{cases} H_0 \equiv \beta_1 = \beta_2 = \beta_3 = 0 \equiv \text{No significativa globalmente} \\ H_1 \equiv \beta_1 \neq 0; \beta_2 \neq 0; \beta_3 \neq 0 \equiv \text{Sí significativa globalmente} \end{cases}$$

$$F_{exp} = \frac{\frac{R^2}{k-1}}{\frac{1-R^2}{n-k}} = \frac{\frac{0,998843}{3-1}}{\frac{1-0,998843}{61-3}} = 25035,82282 > F_{2,58,0,975} = 3,93 \Rightarrow \text{Rechazo } H_0$$

Otra vez, como en el análisis de Andalucía, vamos a pasar a explicar la interpretación de los parámetros, esta vez del modelo para Granada. El primer parámetro establece la siguiente estimación: el máximo de casos por cada 100.000 habitantes será aproximadamente 4315, estimación que, de nuevo, se queda lejos de los datos alcanzados registrados por la Junta de Andalucía, siendo la suma de 3819 aproximadamente.

Si hablamos del segundo parámetro, este está relacionado con el nivel de infectividad, ofreciendo este un valor estimado del 0,0631092. Si hablamos en términos porcentuales, la infectividad tendría una tasa del 6%.

Por último, el tercer parámetro estima que a partir del día 24 de octubre (n=24), los contagios por cada 100.000 crecerán de forma menos acelerada, es decir, el máximo de la función se encontraría en este punto.

Comparativa

Al comparar los datos obtenidos sobre Granada y Andalucía, vemos que en Granada el coronavirus ha afectado de forma mucho más violenta que en Andalucía, prácticamente duplicando la incidencia acumulada. Además de esto, en Granada la enfermedad se expande un 1% más rápido en la población. Esto va acorde con que haya más infectados.

Por último, se aprecia que en Granada se alcanzaría el máximo natural 4 días antes que en Andalucía como conjunto.

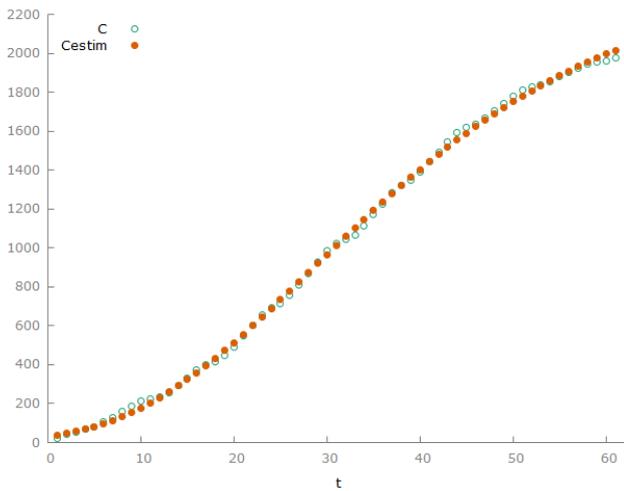


Gráfico 1. Elaboración propia a partir de los datos obtenidos en Gretl

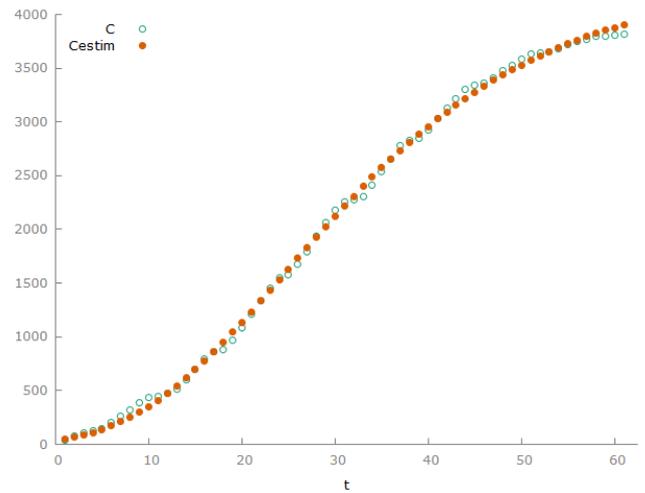


Gráfico 2. Elaboración propia a partir de los datos obtenidos en Gretl

El gráfico de la izquierda pertenece a Andalucía y el de la derecha a Granada. Los puntos verdes representan el crecimiento real y los naranjas el crecimiento natural.

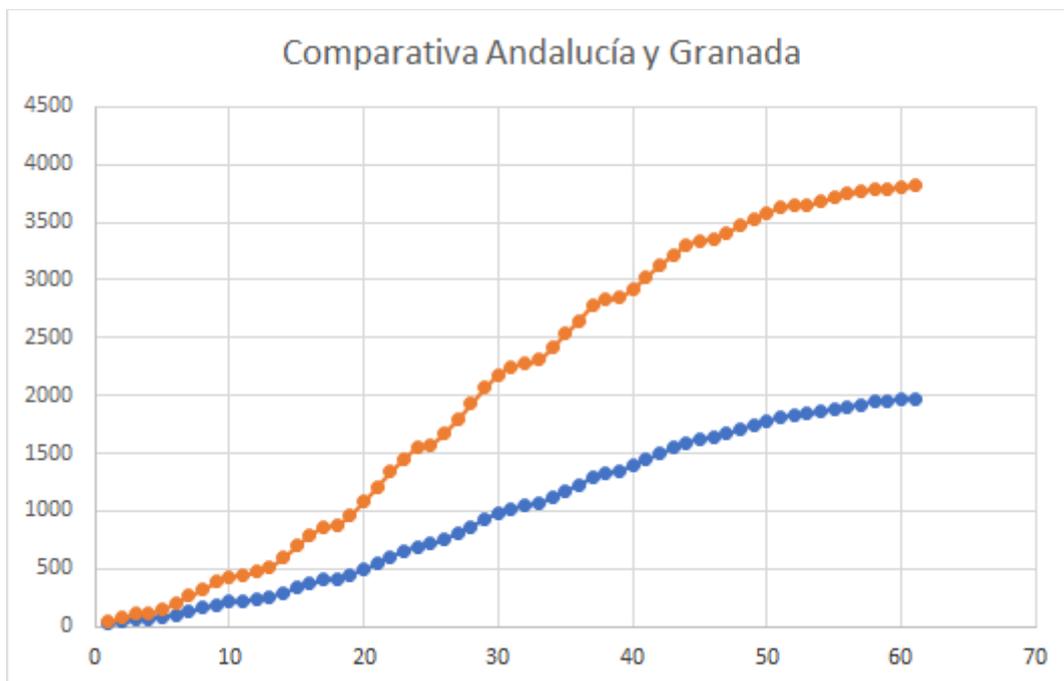


Gráfico 3. Elaboración propia a partir de los datos recogidos por la Junta de Andalucía en Excel

La curva naranja representa la incidencia acumulada en Granada, y la azul en Andalucía. Este gráfico es muy útil para estudiar la comparativa entre la evolución de la pandemia durante la segunda ola en Andalucía y en Granada.

Conclusiones

Las previsiones del modelo se cumplen, ya que la realidad es que, aunque la incidencia haya sido menor de la estimada, sí ocurre que la de Granada es casi el doble que la de Andalucía. La información aportada por los telediarios no es exactamente igual que la recogida por nosotros, pero sí podemos responsabilizar de la diferencia al factor humano.

Aportaciones personales

La mayor aportación que podemos hacer es intentar darle una explicación a la diferencia entre la realidad y las estimaciones. En términos generales, la situación no ha sido como hubiese sido naturalmente ya que los gobiernos y los individuos toman precauciones para no contagiarse. Granada es una ciudad universitaria, y por ello tiene un gran movimiento de personas a lo largo del año. Esto hace que sea más fácil trasladar la infección de un foco a otro.

Además, los estudiantes que se mantuvieron en Granada es probable que buscasen hacer actividades recreativas que incluyan a otros estudiantes, aumentando la probabilidad de contagio.

Por último, las medidas tomadas por el gobierno autonómico sobre el cierre de provincias desde el 30 de octubre hasta el 9 de noviembre no parecen tener ningún efecto hasta los días 11 o 12 de noviembre, queriendo decir que las medidas tienen cierto desfase y no actúan de forma instantánea. (Ideal, 2020).

Bibliografía

- Hueso, S. G. (2020, 25 noviembre). El pico de la segunda ola en Granada fue el 10 de noviembre, un día después que en Andalucía. *Ideal*. Recuperado de <https://www.ideal.es/granada/pico-segunda-granada-20201125185534-nt.html>
- Junta de Andalucía (2021). Informe de la Junta de Andalucía de los datos de la crisis pandémica registrados desde el 1 de octubre hasta el 30 de noviembre del 2020. *Junta de Andalucía*. Recuperado de <https://www.juntadeandalucia.es/institutodeestadisticaycartografia/salud/index.html>
- Martín Barroso, V. (27 de abril del 2020). Modelos logísticos de crecimiento aplicados a la evolución de la Covid-19 en España. *Centro de Estudios de Economía de Madrid. Universidad Rey Juan Carlos*. Recuperado de https://urjc.es/images/ceem/Documento%205_CCEM.pdf
- López-Blázquez, F. y Salamanca-Miño, B. (22 de abril del 2020). Mínimos cuadrados y la enfermedad Covid-19 (II). *Blog del Instituto de Matemáticas de la Universidad de Sevilla*. Recuperado de <https://institucional.us.es/blogimus/2020/04/minimos-cuadrados-y-la-enfermedad-covid-19-y-ii/>
- Ideal (29 de octubre del 2020). Todas las medidas vigentes de la Junta de Andalucía para frenar la pandemia. *Ideal*. Recuperado de <https://www.ideal.es/andalucia/medidas-vigentes-junta-frenar-pandemia-20201028211528-nt.html>