Mathematical Model and Data Analysis to Determine the Number of Confirmed Infections Due to Covid-19 in Spain

Gastón Sanglier[1], Sonia Cesteros[1], Eduardo J. López[1] & Roberto A. Gónzález[1]

[1]Architecture and Design Department. Engineering Area, Escuela Politécnica Superior, Universidad CEU San Pablo, Montepríncipe Campus, 28668, Boadilla Del Monte, Madrid, Spain

Correspondence: Gastón Sanglier, Engineering Area. Escuela Politécnica Superior, Universidad CEU San Pablo, Montepríncipe Campus, 28668, Boadilla Del Monte, Madrid, Spain.

Abstract

Covid-19 initially started in China, although cases of infection by this virus are currently being identified in Europe since January and February of this year camouflaged within a strong outbreak of influenza that had not been identified before. What is certain is that in about a hundred days it has spread around the world threatening humanity. There seems to be a great need to find a rapid response to the speed at which the virus is spreading. In this work, different mathematical models are studied to accurately determine the speed of propagation or infection of people infected by Covid-19 based on data collected from the evolution of the pandemic in Spain. Several mathematical models are proposed and analyzed, but the model proposed as the most suitable is a fourth degree polynomial regression adjustment that presents an R-square statistic of 99.72% which gives a great adjustment of the model for the calculation of the number of infected confirmed by this virus in Spain. Knowing these data is of vital importance to be able to take and undertake the most urgent health and social measures in an effective and orderly manner. This will have a great repercussion in being able to avoid a high number of possible infections.

Keywords: infection, epidemiology, virus, data analysis, mathematical models, Covid-19

1. Introduction

In 1927, biochemist William Ogilvy Kermack and epidemiologist Anderson Gray McKendrick (Kermack & McKendrich, 1932; Kermack & McKendrich, 1923), both Scottish, published a paper that is still used to model epidemics of infectious diseases. The problem they studied was and still is one of the leading causes of death worldwide.

Just think that the 1918 influenza pandemic, also known as the Spanish flu, killed between 50 and 100 million people, while the death toll from World War I in the previous four years was less than 20 million.

Kermack and McKendrick developed the so-called SIR model, where the population is divided into "S" for susceptible, "I" for infected and "R" for recovered. In the 'S' of susceptible are all the people who are not vaccinated - which in the case of covid-19 is the entire population - and who may become ill. In the 'I' of infected, whose curve must try not to rise above the health capacity of the country, because they are those who may require hospital care, and finally in the 'R' of recovered, which are those who neither infect nor can be infected, where the dead are always counted (Dahari et al., 2005; Dee & Shuler, 1997; Diekmann & Heesterbeek, 2000; Ellner et al.,1998). The sum of "S" plus "I" plus "R" is the total number of the population. However, these models also have their limitations. The simplest CRS models make basic assumptions, for example, that everyone has the same chance of getting the virus from an infected person because the population is perfectly mixed and that people with the disease are equally infectious until they die or recover. More advanced models subdivide people into smaller groups (by age, sex, health status, employment, number of contacts, etc.) to establish who meets whom, when and where (Brand, 1957; Brauer et al., 2008; Cadarso & González, 2007; Canini & Perelson, 2014; Cañas, García & Andérica, 2003; Checkoway, Pearce, Crawford, 1989; Chen & Bokka, 2005; Clapham et al.,2016; Haerdle, 1993; Lofgren, 1993; Schafter & Kot, 1985).

The latest data on people diagnosed with Covid-19 coronavirus by PCR in the world is 3,525,116 people, in Europe it is 1,554,703 people and in Spain it is 220,325. The number of people infected in Spain represents 6.25%

of those infected worldwide and 14.17% of those infected in Europe.The communities most affected in Spain by infected persons have been Madrid, Cataluña, Castilla La Mancha, Castilla and Leon and the País vasco.

It may sound strange, but actually counting the dead as "recovered" is part of the mathematical model that is at the base of most simulators used to show how the disease caused by the new coronavirus spreads around the world. For example, it is the model on which the interactive map of the pandemic of the Johns Hopkins University in the United States is based, an institution that has positioned itself as one of the maximum statistical references in this health crisis (CNE, 2005; Sanglier, Robas & Jiménez, 2020a).

In this paper we will analyze and present different mathematical models focused on adjusting to the total number of confirmed infections in Spain with data taken from the Center for Coordination of Health Alerts and Emergencies belonging to the Ministry of Health of Spain (MS, 2020).

2. Methods

A comparison of regression models will be made using the Statgraphics Centurion program. The objectives will be to calculate the covariance and Pearson's linear correlation coefficient between two variables, to perform a linear regression analysis on the data, to determine the existence of a simple model that best fits the data by checking the type of transformation performed, to determine the best statistically significant polynomial that best fits the data, and finally, to analyze the normality of the residues of the best fitting model (Armitage & Bery, 1994; Drake, 1998; Martínez-González, 2004; Montesinos & Hernández, 2007; Velasco, 2007; Canini & Perelson, 2014; Clapham et al., 2016; Sanglier, Robas & Jiménez, 2020b).

3. Results

The numerical data available have been extracted from the Health Alert and Emergency Coordination Centre of the Spanish Ministry of Health. A table of data for the study is attached (Table 1).

Table 1. Table of data on the number of confirmed infected, deceased and recovered in Spain

Date	Days	Infected	Deceased	Recovered	New
12/02/20	0	2	0	0	0
27/02/20	15	17	0	0	1
16/03/20	34	9191	309	N/A	N/A
17/03/20	35	11178	491	N/A	N/A
23/03/20	40	33089	2182	3355	4517
24/03/20	41	39673	2696	3794	6584
26/03/20	43	56188	4089	7015	8578
27/03/20	44	64059	4858	9357	7871
01/04/20	50	102136	9053	22647	7719
07/04/20	56	140510	13798	43208	5478
08/04/20	57	146690	14555	48021	6180
09/04/20	58	152446	15238	52165	5756
10/04/20	59	157022	15843	55668	4576
11/04/20	60	161852	16353	59109	4830
14/04/20	63	172541	18056	67504	3045
19/04/20	68	195944	20453	77357	4218
20/04/20	69	200210	20852	80587	4266
21/04/20	70	204178	21282	82514	3968
22/04/20	71	208389	21717	85915	3401
23/04/20	72	213024	22157	89250	3335
24/04/20	73	202990	22524	92355	3105
25/04/20	74	205905	22902	95708	2944

Initially, a multivariate analysis of the evolution of the number of people infected between the dates shown in the table will be carried out, from February 12 to April 25, 2020, the period in which there has been the greatest evolution in the number of people infected by Covid-19. The statistical summary obtained is set out below.

Table 2. Statistical data summary.

	Time (days)	Number of Infected
Recount	22	22
Average	52,3636	121692
Estandar deviation	19,4729	79423
Coefficient of variation	37,19%	65,27%
Minimum	0	2
Maximo	74	213024
Range	74	213022
Standardized bias	-2,21144	-0,762785
Standardized Kurtosis	1,162	-1,45575

Table 2 shows the statistical summary for each of the selected variables. It includes measures of central tendency, of variability, and of shape. Of interest are standardized bias and standardized kurtosis, which can be used to determine whether the sample is from a normal distribution. Values of these statistics outside the range of -2 to +2 indicate significant deviations from normality, which would tend to invalidate many of the statistical procedures usually applied to these data. In this case, the time variable shows values of standardized kurtosis outside the expected range.

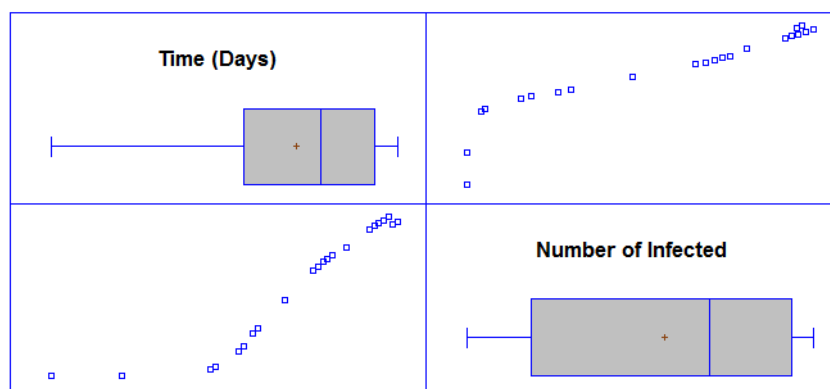In the box and whisker graph, the relationship between the two variables analyzed is shown.

Figure 1. Box chart and whisker of the variable

The correlation matrix attached below tells us if there is any kind of linear relationship between the variables.

Table 3. Variable correlation table

	Time (Days)	Number of infected
Time (Days)		0,9342
		(22)
		0,0000
Number of infected	0,9342	
	(22)	
	0,0000	

Table 3 shows the correlations between each pair of variables. These correlation coefficients range from -1 to +1, and measure the strength of the linear relationship between the variables. The number of data pairs used to calculate each coefficient is also shown in parentheses. The third number in each block of the table is a P-value that proves the statistical significance of the estimated correlations. P-values below 0.05 indicate correlations

significantly different from zero, with a 95.0% confidence level. The following pairs of variables have P-values below 0.05 and therefore have a good correlation

To determine the type of relationship between the two variables, we will start by analyzing the simple linear regression model (Figure 2).
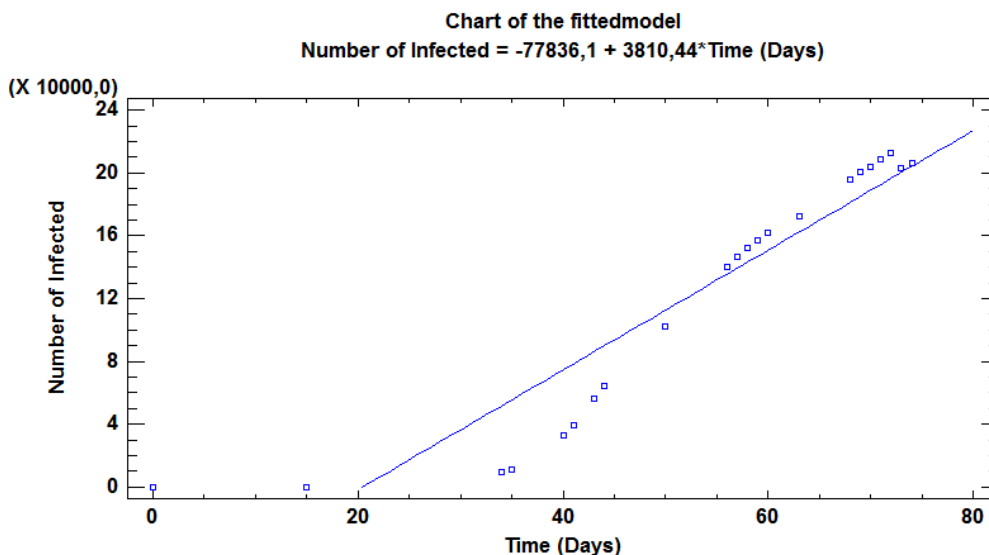

Figure 2. Simple linear regression graph

It is observed that the calculated regression line does not completely fit the data obtained, although if we analyze the analysis of variance table we have a model that is statistically significant, since its P-value is less than 0.05 and with coefficients of constant and slope of the line ($Y = a + b X$) also statistically significant since its P-values are less than 0.05 as shown in Table 4 below.

The ANOVA table presents the values of variability between and within groups. The sum of squares between groups measures the variability between the means of the factor groups. The sum of intra-group squares measures the variability within each factor group. The sum of total squares measures the variability of all data with respect to the mean. The F-ratio is the value of the mean of the inter-group squares divided by the value of the mean of the intra-group squares. The P-value indicates the level of significance (it is the area to the right of F). For small values (less than 0.05) it indicates that the sample/variable measurements are significantly different.

Table 4. Statistical results for linear model fit

Coefficients				
	Minimum Squares	Standard	Statistician	
Parameter	Estimated	Error	T	Value-P
Intercept	-77836.1	18120.9	-4.29537	0.0004
Pending	3810.44	325.257	11.7152	0.0000

Analysis of Variance					
Source	Sum of Squares	Gl	Average Square	Reason-F	Value-P
Model	1.15620E+11	1	1.15620E+11	137.25	0.0000
Residue	1.68486E+10	20	8.42430E+08		
Total (Corr.)	1.32468E+11	21			

Correlation Coefficient:	0.934243
R-square (%):	87.281
R-square (adjusted for g.l.) (%):	86.6451
Standard error of the estimate:	29024.6
Average absolute error:	21202.7
Durbin-Watson Statistic:	0.479173 (P=0.0000)
Autocorrelation of delayed residues 1:	0.58502

The output (Table 4) shows the results of fitting a linear model to describe the relationship between Number of Infected and Time (Days). The equation for the adjusted model is as follows:

$$\text{Number of Infected} = -77836.1 + 3810.44 * \text{Time (Days)} \tag{1}$$

Since the P-value in the ANOVA table is less than 0.05, there is a statistically significant relationship between Number of Infected and Time (Days) with a 95.0% confidence level.

You can choose other models from the alternative model comparison table below (Table 5):

Table 5. Comparison of alternative models

Model	Correlation	R-Square (%)
Square of X	0.9798	96
Double Square	0.9726	94.59
X-square-root X-square-root	0.9585	91.88
Logarithm-Y Square Root-X	0.9550	91.20
Linear	0.9342	87.28
Exponential	0.9012	81.22
Y-square	0.8848	78.30
Square root of X	0.8045	64.72
Log-Y X-Square	0.7589	57.59
Y-Square-Root X-Square	0.7303	53.33
Reverse-Y SquareX	-0.4507	20.31
Y-square root	unadjusted	
Reverse Y	unadjusted	
Double Square Root	unadjusted	
Reverse-Y Square Root-X	unadjusted	
Logarithm of X	unadjusted	
Y-Square Root Log-X	unadjusted	
Multiplicative	unadjusted	
Reverse-Y Log-X	unadjusted	
Y-Square-Log-X	unadjusted	
Reverse X	unadjusted	
Reverse Y-square root of X	unadjusted	
S-curve	unadjusted	
Double reverse	unadjusted	
X-Reverse Y-Square	unadjusted	
Logistics	unadjusted	
Log probit	unadjusted	

These are the models that can have a square R higher than the linear model that has been studied (R-square = 87.285).

Table 5 shows the results of fitting several curvilinear models to the data. Of the adjusted models, the X-square model is the one that gives the highest R-square value with 95.9984%. This is 8.71737% higher than the selected linear model.

Next, the X-square model that has the best R-square of all the models will be tested and the graph below is obtained. It is observed that the new model fits much better to the data than the linear model initially proposed.

Chart of the fittedmodel
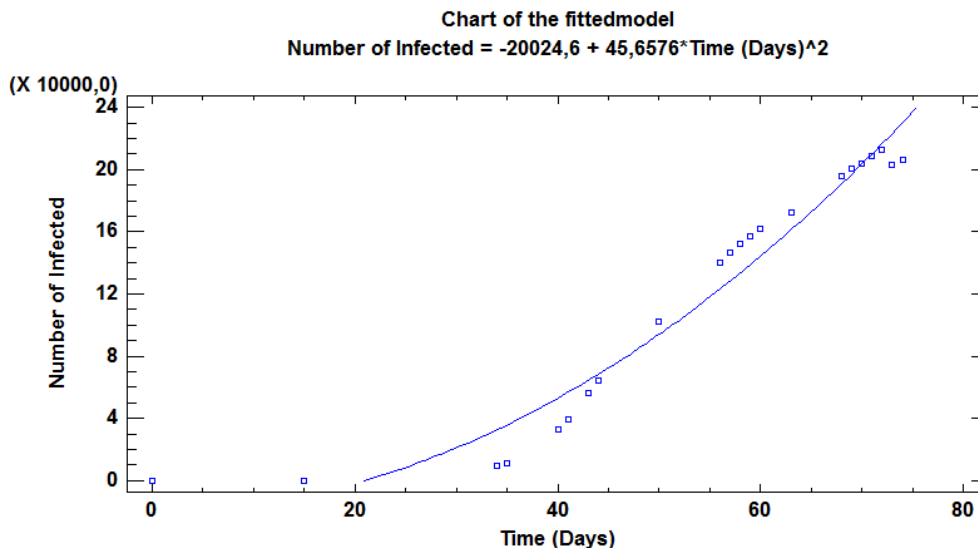Number of Infected = -20024,6 + 45,6576*Time (Days)^2

Figure 3. Fitting curve for the X-Square model

Table 6 of analysis of variance, it can be seen that in the table of the T statistician it is a statistically significant model, reaching a R-square of 95.9984% (Table 6).

Table 6. Statistical results for X-Square model

Coeffcients				
	Minimum Squares	**Standard**	**Statistician**	
Parameter	**Estimated**	**Error**	**T**	**Value-P**
Intercept	-200024	7342.05	-2.72738	0.0130
Pending	45.6576	2.0844	21.9044	0.0000

Analysis of Variance					
Source	**Sum of Squares**	**Gl**	**Average Square**	**Reason-F**	**Value-P**
Model	1.27168E11	1	1.27168E11	479.80	0.0000
Residue	5.30083E9	20	2.655042E8		
Total (Corr.)	1.32468E11	21			

Correlation Coefficient:	0.979788
R-square (%):	95.9984
R-square (adjusted for g.l.) (%):	95.7983
Standard error of the estimate:	16280.1
Average absolute error:	13415.9
Durbin-Watson Statistic:	0.372369 (P=0.0000)
Autocorrelation of delayed residues 1:	0.721239

The output (Table 6) shows the results of fitting an X-square model to describe the relationship between Number of Infected and Time (Days). The equation of the adjusted model is:

$$\text{Number of Infected} = -20024.6 + 45.6576*\text{Time (Days)}^2 \tag{2}$$

Since the P-value in the ANOVA table is less than 0.05, there is a statistically significant relationship between Number of Infected and Time (Days) with a 95.0% confidence level.

The R-Square statistic indicates that the adjusted model explains 95.9984% of the variability in the Number of Infected. The correlation coefficient is equal to 0.979788, indicating a relatively strong relationship between the variables. The standard error of the estimate indicates that the standard deviation of the residues is 16280.1. This value can be used to construct predictive limits for further observations.

The mean absolute error (MAE) of 13415.9 is the average value of the residues. The Durbin-Watson (DW) statistician examines the residues to determine if there is any significant correlation based on the order in which they are presented in the data file. Since the P-value is less than 0.05, there is an indication of a possible serial correlation at a 95.0% confidence level. Plot the residuals versus the row number to see if there is any pattern that can be detected.

You can try another grade 5 polynomial regression model. The data obtained from the analysis and the fitted model are as follows (Figure 4).
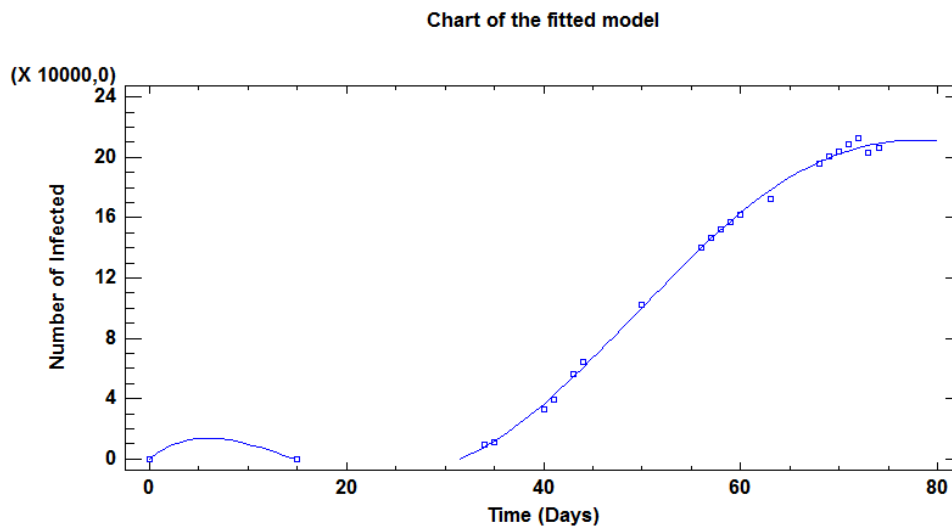

Figure 4. Fitting curve for a grade five polynomial model

Table 7. Statistical results for a grade five polynomial model

Coefficients				
	Minimum Squares	**Standard**	**Statistician**	
Parameter	**Estimated**	**Error**	**T**	**Value-P**
Constant	57.8127	3279.61	0.0176279	0.0862
Time (Days)	5370,97	1291.56	4.15851	0.0007
Time (Days)^2	-636.583	123.811	-5.14158	0.0001
Time (Days)^3	22.373	4.33493	5.16109	0.0001
Time (Days)^4	-0.276762	0.0636149	-4.35058	0.0005
Time (Days)^5	0.00114041	0.000332248	3.4324	0.0034

Analysis of Variance					
Source	**Sum of Squares**	**Gl**	**Average Square**	**Reason-F**	**Value-P**
Model	1.32296E11	5	2.64593E10	2459.01	0.0000
Residue	1.72162E8	16	1.07601E7		
Total (Corr.)	1.32468E11	21			

R-square (%):	99.87
R-square (adjusted for g.l.) (%):	99.8294
Standard error of the estimate:	3280.26
Average absolute error:	2129.26
Durbin-Watson Statistic:	1.34562 (P=0.0046)
Autocorrelation of delayed residues 1:	0.29787

It can be observed that in the table of the T statistic there are some coefficients such as the Time (Days)^5 coefficient that has a P-value > 0.05, so it can be said that it is not a statistically significant coefficient and should be eliminated from the model. If this is done, we are left with the following data (Table 8).

Table 8. Statistical results for a grade four polynomial model

Coefficients				
Parameter	Minimum Squares Estimated	Standard Error	Statistician T	Value-P
Constant	890,974	4181,02	0,2131	0,8338
Time (Days)	1657,24	901,663	1,83799	0,0836
Time (Days)^2	-232,98	49,5537	-4,70157	0,0002
Time (Days)^3	7,71493	0,951989	8,10401	0,0000
Time (Days)^4	-0,0590073	0,00601278	-9,81365	0,0000

Analysis of Variance					
Source	Sum of Squares	Gl	Average Square	Reason-F	Value-P
Model	1,32169E11	4	3,30424E10	1879,09	0,0000
Residue	2,98931E8	17	1,75842E7		
Total (Corr.)	1,32468E11	21			

R-square (%):	99,7743
R-square (adjusted for g.l.) (%):	99,7212
Standard error of the estimate:	4193,35
Average absolute error:	3170,41
Durbin-Watson Statistic:	1,0424 (P=0,0006)
Autocorrelation of delayed residues 1:	0,474525

By adjusting the polynomial to degree four, it can be seen that the model is statistically significant and all coefficients have a P-value below 0.05.

The equation of the adjusted model is:

Number of Infected = 890,974 + 1657,24*Time (Days)-232,98*Time (Days)^2

$$+ 7,71493*Time (Days)^3-0,0590073*Time (Days)^4 \qquad (3)$$

The model sought is that presented by equation 3 because it presents the best R-square of all.

Figure 5 following graph shows the grade four polynomial regression model that would best fit the studied data of the number of infected as a function of time.
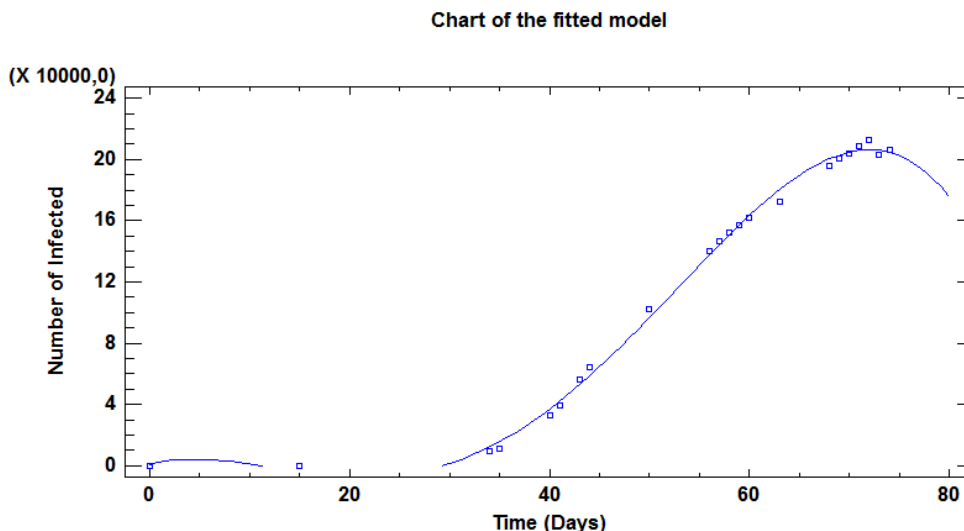
Figure 5. Fitting curve for a grade four polynomial model

The R-Square statistic indicates that the model thus adjusted explains 99.7743% of the variability in Number of Infected. The R-squared adjusted statistic, which is more appropriate for comparing models with different numbers of independent variables, is 99.7212%. The standard error of the estimate shows that the standard deviation of the residues is 4193.35. This value can be used to construct limits for new observations by selecting the Reports option from the text menu. The mean absolute error (MAE) of 3170.41 is the average value of the residue. The Durbin-Watson (DW) statistician examines the residues to determine if there is any significant correlation based on the order in which they are presented in the data file. Since the P-value is less than 0.05, there is an indication of possible serial correlation at a 95% confidence level. Plot the residuals versus the row number to see if there is any pattern that can be detected.

To determine if the order of the polynomial is appropriate, look at the P-value in the higher order term as 2.0384E-8. Since the P-value is less than 0.05, the higher order term is statistically significant at a 95% confidence level. Because of this, no lower order model is considered for testing.

We will now analyze the errors in the model selected as the best fit. We will check whether the residuals follow a normal distribution. We are going to test the model for normality and goodness-of-fit.

The following data are obtained from the normality tests (Table 9).

Table 9.　Normality tests for residuals

Testing	Statistics	Value-P
Chi-Cuadrado	5,18182	0,878706
Statistics W of Shapiro-WilK	0,975726	0,82297
Z value for asimetry	0,0655302	0,947746
Value Z for shorts	-0,813545	0,415904

It is observed from the results obtained to determine the residues that these can be adequately modelled with a normal distribution. The chi-square test divides the range of residues into 13 equally likely classes and compares the number of observations in each class with the expected number of observations. The Shapiro-Wilk test is based on the comparison of the quartiles of the normal distribution adjusted to the data. The standardized bias test looks for lack of symmetry in the data. The standardized kurtosis test looks for whether the shape of the distribution is flatter or more pointed than the normal distribution.

Because the smallest P-value of the tests performed is greater than or equal to 0.05, one cannot reject the idea that the residues come from a normal distribution with 95% confidence.

If we now look at the goodness-of-fit tests, we get the following data (Table 10).

Table 10. Goodness-of-fit testing for residuals

Chi-Square Test					
	Lower limit	Upper limit	Observed frequency	Expected frequency	Chi-Square
minor or equal		-3848,66	4	3,38	0,11
	-3848,66	-1895,52	4	3,38	0,11
	-1895,52	-364,312	3	3,38	0,04
	-364,312	1106,91	1	3,38	1,68
	1106,91	2778,06	4	3,38	0,11
major	2778,06		6	5,08	0,17

Chi-Square=2,22727 with 3 g.l.   Value-P=0,526593

Kolmogorov-Smirnov Test		Kuiper V Test		Cramer-Von Mises W^2 Test	
	Normal		Normal		Normal
DMAS	0,11034				
DMENOS	0,102861	V	0,213201	W^2	0,0408941
DN	0,11034	Modified form	1,04396	Modified form	0,0250407
Value-P	0,951605	Value-P	>=0,10	Value-P	>=0,10

Watson U^2 Test		Anderson-Darling A^2 Test		
	Normal		Normal	
U^2	0,0408936	D^2	0,246309	
Modified form	0,037884	Modified form	0,246309	
Value-P	>=0,10	Value-P	>=0,10	

The results of the various tests carried out to determine whether the residues can be adequately modelled with a normal distribution have been presented.   The chi-square test divides the range of residues into non-overlapping intervals and compares the number of observations in each class with the expected number based on the adjusted distribution.   The Kolmogorov-Smirnov test calculates the maximum distance between the cumulative residue distribution and the FDA of the adjusted normal distribution.   In this case, the maximum distance is 0.11034.   The other statisticians compare the empirical distribution function with the adjusted FDA, in different ways. It is noted that in some tests the P-value could not be determined accurately.

Because the smallest P-value of the tests performed is greater than or equal to 0.05, the idea that the residues come from a normal distribution with 95% confidence cannot be rejected.

4. Discussion

Although the percentage of people infected by Covid-19 in the world is approximately 0.0447%, considering the world population at 7.88E9 people, the coronavirus pandemic has managed to generally alert humanity. The speed of its spread around the world is what has put us on alert (Munayco, 2009; Al-Rousan, 2020; WHO, 1994).To carry out this study, the Statgraphics Centurion program has been used to calculate the regression model that best fits two variables, in this case the number of people infected by Covid-19 in Spain over time.The results obtained by simple linear regression have been compared with linear models with transformation of variables and polynomial regression model. The normality of the model residues has been analyzed.

We started with a multivariate analysis of the two variables where it was found that one of the analysed variables presented standardised kurtosis values out of range. However, it is demonstrated that there is a good correlation between the variables.

The first model proposed has been a simple linear regression model presenting an R-square of 87.281%. The model is statistically representative as the P values are less than 0.05. Equation 1 has been obtained for this model.

This model has been compared with other 27 alternative models where it has been determined that the best model was the X-square that presents a R-square of 96%, this meant an improvement of 8.71% with respect to the initial model. Equation 2 has been obtained for this model.

It was then tested with a polynomial regression model of order five, where after analyzing the statistical data it was determined that the grade five coefficient had a P value above 0.05. The order was lowered to four and a new grade four polynomial model was retested. In this case the model presented the best R-square, obtaining a value of 99.72%. It was determined as the best model and equation 3 was obtained as the final result.

Finally, a study of the model residues was made to see if they could be modeled as a normal distribution. The results obtained confirmed this assumption.

In this research article, different regression models have been highlighted in order to study the model that provides the best adjustment to the variables studied, such as the number of infected people in Spain during a period of maximum infection, depending on their evolution over time.

After discussing and analyzing the models obtained, it has been found that the model that best fits the number of confirmed infected persons in Spain is a fourth-order polynomial regression model. The study of possible models that can calculate the number of deaths and recoveries, two very important parameters in studies on the evolution of the virus in the case of pandemics of this type, is left for another occasion.

The development of this type of mathematical model is very necessary in order to help the different governments in the rapid adoption of prevention measures, both of a clinical and social nature, with the possible reduction of the number of infected people. This would lead to a greater decongestion of hospitals and health systems in general.

References

AL_Rousan, N. (2020). Data Analysis of Coronavirus CoVID-19 Epìdemic in South Korea Based on Recovered and Death Cases. *Journal of Medical Virology*. https://doi.org/10.1002/jmv25850. https://doi.org/10.1002/jmv.25850

Armitage, P. & Berry, G. (1994). *Statistical Methods in Medical Research* (3rd edition). Blackwell.

Brand, L. (1957). The Pi Theorem of Dimensional Analysis. *Arch. Rational Mech. Anal., 1*, 35-45. https://doi.org/10.1007/BF00297994

Brauer, F., Driessche, P.V. & Wu, J. (1945). *Mathematical epidemiology. Mathematical biosciences subseries*. Berlin: Springer; p. 408.

Cadarso-Suárez, C. & González-Manteiga, W. (2007). Statistics in biomedical research. Arbor. *CLXXXIII*, (725), 353-361.

Canini, L. & Perelson, A. S. (2014). Viral kinetic modeling: state of the art. *J. Pharmacokinet. Pharmacodyn. 41*, 431–443. https:/doi.org/10.1007/s10928-014-9363-3.

Cañas, E.J., García-León, F.J. & Andérica-Frías, G. (2003). Epidemiología de la infección VIH y SIDA. En Pachón J, Pujol E, Rivero A (Eds) La Infección por el VIH: Guía Práctica. 2ª ed. Sociedad Andaluza de Enfermedades Infecciosas. Sevilla, España. 614.

Checkoway, H. & Pearce, N.E. (1989). *Crawford-Brown DJ. Research methods in occupational epidemiology*. Nueva York: Oxford University Press; 117- 119.

Chen, W.Y. & Bokka, S. (2005). Stochastic modelling of nonlinear epidemiology. *J. Theor. Biol., 234*, 455-470. https://doi.org/10.1016/j.jtbi.2004.11.033

Clapham, H. E., Quyen, T. H., Kien D. T., Dorigatti I., Simmons C. P. & Ferguson N. M. (2016). Modelling virus and antibody dynamics during dengue virus infection suggests a role for antibody in virus clearance. *PLoS Comput. Biol., 12*, e1004951. https://doi.org/10.1371/journal.pcbi.1004951.

CNE (Centro Nacional de Estadística). (2005). Valoración de la epidemia de VIH en España a partir de los sistemas de notificación de casos de las Comunidades Autónomas. Centro Nacional de Epidemiología. Madrid, España. 6

Dahari, H., Major, M., Zhang, X., Mihalik, K., Rice, C. M. & Perelson, A. S. (2005). Mathematical modeling of primary hepatitis C infection: noncytolytic clearance and early blockage of virion production. *Gastroenterology, 128*, 1056–1066. https://doi.org/10.1053/j.gastro.2005.01.049.

Dee, K. U. & Shuler, M. L. (1997). A mathematical model of the trafficking of acid-dependent enveloped viruses: application to the binding, uptake, and nuclear accumulation of baculovirus. *Biotechnol. Bioeng, 54*, 468–490. https://doi.org/10.1002/(SICI)1097-0290(19970605)54:5<468:AID-BIT7>3.0.CO;2-C.

Diekmann, O. & Heesterbeek, J. A. P. (2000). *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation*. Hoboken, New Jersey: John Wiley & Sons.

Drake, AW. (1988). Fundamentals of Applied Probability Theory. McGraw-Hill. Nueva York, NY, EEUU. 256.

Elandt-Johnson, R.C. (1975). Definition of rates: some remarks on their use and misuse. *Am J Epidemiol, 102*, 267-271. https://doi.org/10.1093/oxfordjournals.aje.a112160

Ellner, S.P., Bailey, B.A, Bobashev, G.V., Gallant, A.R., Grenfell, B.T. & Nychka, N.S. (1998). Noise and Nonlinearity in Measles Epidemics: Combining Mechanistic and Statistical Approaches to Population Modeling. *American Naturalist; 151*(5), 425-40. https://doi.org/10.1086/286130

Haerdle, W. (1993). *Applied Nonparametric Regression*. CambridgeUniversity Press.

Kermack, W.O. & McKendrick, A.G. (1927) *A contribution to the mathematical theory of epidemics Proceeding of the Royal Society of London, 115(*772), 700-721.

Kermack, W.O. & McKendrick, A.G. (1932). A contribution to the mathematical theory of epidemics. *The problem of Endemicity Proceeding of the Royal Society of London, 138*(834), 55-83. https://doi.org/10.1098/rspa.1932.0171

Lofgren, E., Fefferman, N.H., Naumov, Y.N., Gorski, J. & Naumova, E.N. (2007). Influenza seasonality: underlying causes and modeling theories. *J Virol, 81*, 5429–5436.

Martínez-González, M.A. (2004). *Bioestadística Amigable*. España: Díaz de Santos.

Montesinos-López, O.A. & Hernández-Suarez, C.M. (2007). Modelos matemáticos para enfermedades Infecciosas. *Salud Publica Mex., 49*, 21826. https://doi.org/10.1590/S0036-36342007000300007

MS_Ministerio de Sanidad. (2020). https://www.mscbs.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov-China/situacionActual.htm.

Munayco, E.C. (2009). Conceptos fundamentales del modelamiento epidémico. *Bol Epidemiol (Lima), 18*(27), 528-9.

Sanglier, G., Robas, M. & Jiménez, P.A. (2020a). Application of Forecasting Methods to a Mathematical Models for Adjusting the Determination of the number of people infected by Covid-19 in the Community of Madrid. *Contemporary Engineering Sciences, 2020, 13*(1), 89-101. HIKARI Ltd. https://doi.org/10.12988/ces.2020.91453.

Sanglier, G., Robas, M. & Jiménez, P.A. (2020b). Estimation of the Epidemiological Evolution Through a Modelling Analysis of the Covid-19 Outbreak. *Acta Scientific Microbiology, 13*(4), april 2020, 152-158.

Schaffer, W.M. & Kot, M. (1985). Nearly one dimensional dynamics in an epidemic. *J Theor Biol. Jan 21; 112*(2), 403-27. https://doi.org/10.1016/S0022-5193(85)80294-0

Velasco Hernández, J.X. (2007). Modelos matemáticos en epidemiología: enfoques y alcances. Miscelánea Matemática; 44, 1127.

WHO - World Health Organization. (1994). Expanded Programme on Immunization: Accelerated measles strategies. *Wkly Epidemiol Rec, 69*(31), 229–234.

Copyrights