



[Science Archives \(ISSN:2582-6697\)](http://www.sciencearchives.org)

Journal homepage: www.sciencearchives.org



<https://doi.org/10.47587/SA.2023.4212>

Research Article



Detection of the most important factors affecting the increase in the number of deaths as a result of infection with Covid-19 using the multiple linear regression equation

Asmaa S. Qaddoori 

Department of Mathematics, College of Education for women, University, Tikrit, Iraq

Received: May 12, 2023/ Revised: June 9, 2023/Accepted: June 9, 2023

(✉) Corresponding Author: asmaa.salih@tu.edu.iq.

Abstract

In this research, a multiple linear regression analysis was conducted to identify the parameters of a model using data from individuals who passed away due to COVID-19. The data was collected from various hospitals in Salah al-Din Governorate, including Tikrit Hospital, Baiji Hospital, Shirqat General Hospital, and Samarra Hospital, between April 1st, 2020, and December 30th, 2020. The primary aim of the study was to investigate the primary factors that contribute to the rise in deaths caused by the COVID-19 pandemic, using the multiple linear regression analysis method. The findings indicated that independent variables such as age, place of death, and the deceased person's chronic diseases were the most significant factors in the increase of COVID-19-related deaths, while other variables examined did not have a clear impact.

Keywords: Multiple linear regression, COVID-19, Independent variables, Chronic diseases

Introduction

COVID-19 is a novel type of coronavirus that can lead to illnesses similar to the common cold, as well as more severe respiratory diseases like SARS and MERS. It is caused by the SARS-Cov-2 virus (The Arab Organization for Education, 2020). Even though COVID-19 has spread widely, there is still a lot that remains unknown about its characteristics, modes of transmission, and prevention. It's possible that the virus has mutated, leading to a new strain that can infect the kidneys - a trait not present in previous strains. COVID-19 is believed to have originated as an animal virus and adapted to infect humans, affecting various body systems, including the heart and kidneys (Al-Samadouni, & Al-Sayed, 2021). Early in 2019, an outbreak of the disease was first reported in Wuhan, People's Republic of China. On January 30, 2019, the World Health Organization declared that this outbreak of the virus represents a serious threat to global public health. More than 595 million cases of COVID-19 have been reported in more than 188 countries and regions as of August 20, 2022, including more than 6 million deaths. The United States is

considered the country most affected by the pandemic, as it recorded more than a quarter of the above-mentioned number of confirmed infections (The Arab Organization for Education, 2020 & He et al., 2020).

A person infected with the virus may spread it to others when they expel tiny droplets of fluids via their lips or nose when they cough, sneeze, talk, sing, or breathe. The particles range in size from the massive respiratory droplets to micro-droplets. Infection can be contracted by breathing in close proximity to a person with COVID-19 disease or when touching a surface contaminated with the virus. Keeping the virus from spreading further indoors and in crowded settings requires good respiratory hygiene, such as coughing into the crook of one's elbow, and self-isolation once symptoms appear (Al-Masoodi et al., 2020).

The symptoms of the disease range from two to 14 days after infection with the virus, and the symptoms range from mild symptoms that resemble the symptoms of influenza, such as fever, coughing, and high temperature with loss of sense of

smell and taste, to more serious symptoms such as severe headache, diarrhea, vomiting, dry cough, Muscular pain, shortness of breath, sinusitis, and the disease may develop into acute pneumonia and may cause disorder and kidney failure, leading to death (The Arab Organization for Education, 2020).

The risk of severe complications is higher among anybody over the age of 60 or who has a serious medical condition like diabetes asthma, difficulties with the heart, lungs, blood sugar, and circulation, obesity, and excess weight, chronic kidney disease, liver disease, a weakened immune system, cancer, etc. That is a disease. However, COVID-19 poses a significant risk of serious disease and death to people of all ages (Al-Samadouni, & Al-Sayed, 2021).

Knowledge of the disease's features and the virus's transmission mode remains the greatest method for preventing and slowing the development of COVID-19. Keep at least one meter away from people, use a well-fitting mask, and routinely wash your hands or rub them with an alcohol-based disinfectant to prevent the spread of disease. As well as taking

The linear equation in multiple linear regression is:

$$Y_i = b_0 + b_1X_{1i} + b_2X_{2i} + \dots + b_m X_{mi} + e_i$$

- Y_i: It is one of the values taken by the dependent variable y.
- b₀: The y segment, is after the point at which the regression line cuts.
- b₁: The slope of the y regression on the first independent variable.
- b₂: The slope of the y regression on the second independent variable

- X₁: The first independent variable.
- X₂: The second independent variable.
- e_i: Random error.

If certain requirements are satisfied, it is possible to employ multiple linear regression

1. The independent variable matrix X is established and provided, meaning it is obtained through precise measurement without any errors.
2. The absence of a full or nearly full linear association among the independent variables, i.e., is a prerequisite: (Rank(X) = k+1).
3. Third, the data from each experiment may be analyzed separately. (X_{i1}, X_{i2},X_{ik}) and the random error e_i, meaning that the columns of the matrix X are linearly independent from the random errors vector U, and this is mathematically expressed as follows: Cov.

the vaccine and following the instructions of the local authorities related to the health ban and staying at home for a certain period (He et al., 2020).

Multiple Linear Regression

One of the most sophisticated statistical techniques, multiple linear regression helps researchers identify a causal link between occurrences by making the most of the data at their disposal (Kazem, 2009). In essence, multiple linear regression involves developing a mathematical equation that relates two variables and can be utilized to estimate past occurrences as well as forecast future outcomes. It entails a regression analysis of the dependent variable (Y) on numerous independent variables (x₁, x₂, ... x_k), enabling the prediction of changes in the dependent variable affected by multiple independent variables that are typically continuous. The underlying concept of multiple linear regression depends on semantic relationships utilizing dispersion or diffusion (Al-Khathami, 2018).

$$Cov(X, U) = E(X, U) - [E(X)][E(U)] = 0$$

4. Random error = i=1, 2,...,n, e_i it has a normal distribution with an average of zero, and its variance σ² is constant from one observation to the next, that is, e_i ~ N(0, σ²) and it is assumed that the errors, (e₁, e₂.....,e_n) are statistically independent, and this is expressed mathematically" as follows:

$$E(e_i) = 0$$

$$Cov(e_i, e_i) = E(e_i, e_i) = \{\sigma^2 \text{ if } i = j \text{ and } 0 \text{ if } i \neq j\}$$

In other words, the error vector U has a variance matrix, which is U ~ N_n(0, Σ), and it follows multiple normal distributions with a zero mean.

The matrix Σ is a symmetric matrix of degree (n × n) and is expressed as follows (Al-Rawi, 1987).

$$\Sigma = \begin{bmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

There are several ways to estimate the parameters of the multiple linear regression model. The least squares method, which reduces the sum of the deviations of the estimated values from the observed values of the dependent variable y, is one of the best of these methods, as it works to make the sum of the squares of the error or the residual as low as possible (Kazem & Al-Dulaimi, 1988). Once the outcomes by means of the regression formula are known, we must state if

statistically, these coefficients are of sufficient quality to be considered reliable, keeping in mind that each coefficient's significance must be determined independently.

Using the SPSS application, we use the T-test and the accompanying probability level to determine the significance of the regression coefficients. Additionally, statistics like (R) (R^2) that are used to determine the model's general morale will be acquired. After controlling for the impact of any other factors, the R statistic displays the simple correlation coefficient, which assesses the strength of the association between two or more variables, X and Y (Tuwaiti & Wail, 2014).

R^2 is known as the coefficient of determination, and it is used to calculate the explanatory power of the estimated model (estimated equation). It is the ratio of the sum of squares owing to regression to the sum of squares surrounding Y .

$$R^2 = \frac{SSR}{SST}$$

The sum of squares of the regression:

$$SSR = \sum_{i=1}^n (\hat{y} - \bar{y})^2$$

The grand total of squares:

$$SST = \sum_{i=1}^n (y - \bar{y})^2$$

Ranging from zero to one: R^2 is a value: $0 \leq R^2 \leq 1$

The model is identical to the data, the value of R^2 approaches one, meaning that the observed values y and the expected values \hat{y} are very close (Al-Jubouri, 1990).

As for R^{-2} , it is used to interpret the explanatory power of the multiple linear regression model (because it takes into account the number of independent variables), and therefore it is called the corrector because it is originally derived from R^2 , because the value of the determination coefficient is affected by the number of explanatory variables, so the value of R^2 increases whenever we add a new explanatory variable. Therefore, its value cannot be decreased, and in order to avoid this shortcoming, we must correct its value so that it is not affected by the number of explanatory variables, by taking the degrees of freedom into account when calculating the coefficient of determination, and the formula for the corrected coefficient of determination R^{-2} becomes as follows (Kazem, & Al-Dulaimi, 1988).

$$R^{-2} = 1 - \frac{\sum_{i=1}^n \epsilon_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \cdot \frac{n-1}{n-k}$$

By comparing the explained variance with (k-1) degrees of freedom to the unexplained variance with (n-k) degrees of freedom, where n is the number of observations and k is the number of estimated features, the F test can also be used to assess the significance of the estimated model as a whole at a given level of significance (Tuwaiti, & Wail, 2014).

The formulating test hypothesis

null hypothesis: $H_0 = \beta_1 = \beta_2 \dots = \beta_k = 0$

The alternative hypothesis: $H_1: \exists \beta_1 \neq 0$

The value of the F statistic can be calculated by the following relation:

$$F_{(k-1, n-k, \alpha)} = \frac{\sum_{t=1}^n (\hat{y}_t - \bar{y})^2 / (k-1)}{\sum_{t=1}^n \epsilon_t^2 / (n-k)}$$

The computed F value ($F_{cal.}$) is then compared to the tabular F value ($F_{tab.}$). If the computed F value is larger than the tabular F value at a significant level = 0.05 or 0.01, the null hypothesis H_0 is rejected, indicating the multiple regression model's relevances. However, if the computed F value is less than the tabular F value, the null hypothesis is accepted, indicating that the multiple regression model is insignificant and that the variables included in the research do not adequately explain the occurrence of the issue (Al-Jubouri, 1990).

Practical application via the SPSS statistical program

The SPSS statistical program was utilized to analyze data collected from a sample of 150 patients selected randomly from various hospitals in Salah al-Din Governorate between May 1st, 2020, and December 31st, 2020. The study examined the relationship between the dependent variable, which was the patients' Y test results, and several independent variables that were potential influencing factors leading to patient death. These factors included age (X_1), sex (X_2), date of death (X_3), place of death (X_4), place of receiving treatment (X_5), mode of infection (X_6), duration between illness and death (X_7), injury symptoms (X_8), chronic diseases (X_9), whether the patient received necessary care at home (X_{10}), and whether the patient received appropriate treatment (X_{11}). The analysis of the data using the SPSS program produced numerous results.

Upon statistical analysis of the data, it was discovered that the mean age of the deceased patients was 51.78 years. The youngest age of a deceased patient was 28 years old, while the oldest age was 78 years old.

Table 1. Patient sex ratio

Sex	Frequency	Percent	Valid Percent	Cumulative Percent
Male	114	76.0	76.0	76.0
Female	36	24.0	24.0	100.0
Total	150	100.0	100.0	

According to the table above, the number of men was 114, with a percentage of 76%, while the number of females was 36, with a percentage of 24% of the entire sample size.

Table 2. Place of death place

Place	Frequency	Percent	Valid Percent	Cumulative Percent
Hospital	84	56.0	56.0	56.0
House	66	44.0	44.0	100.0
Total	150	100.0	100.0	

It is evident from the above table that 84 individuals, or 56% of those who passed away, did so in a hospital, while 66 people, or 44% of those who passed away at home, were counted.

Table 3. Symptoms of infection

Symptoms	Frequency	Percent	Valid Percent	Cumulative Percent
High fever and cough	15	10.0	10.0	10.0
Heat, cough, and impaired Kidney function	30	20.0	20.0	30.0
Heat, cough, and heart muscle fatigue	15	10.0	10.0	40.0
Heat, cough, loss of smell and taste, shortness of breath, and general tendency	90	60.0	60.0	100.0
Total	150	100.0	100.0	

The Table above displays the adverse effects of the illness, which were characterized by loss of taste, smell, and coughing. Shortness of breath, a sudden increase in body temperature, and a general change in the body's composition accompany these symptoms. the elderly and those with chronic illnesses.

While 60% of those affected had symptoms such as loss of taste and smell, coughing, high temperatures, and shortness of breath along with general sluggishness, 20% of those affected had impaired kidney function along with coughing and high temperatures, and 30% of those injured and 10% of those affected had symptoms such as fatigue in the heart muscle, coughing, and high temperatures, while 10% of those infected had symptoms such as loss of taste and smell.

Table 6. The period between infection and death

The period	Frequency	Percent	Valid Percent	Cumulative Percent
One week	39	26.0	26.0	26.0
Ten days	78	52.0	52.0	78.0
Two week	33	22.0	22.0	100.0
Total	150	100.0	100.0	

The above-mentioned data reveals that 26% of those who passed away within a week of illness totaled 39 infected,

Table 4. Chronic diseases suffered by the patient

Chronic diseases	Frequency	Percent	Valid Percent	Cumulative Percent
There is no	12	14.0	14.0	14.0
Hypertension	24	16.0	16.0	30.0
Hypertension and sugar	33	22.0	22.0	52.0
Heart disease	42	28.0	28.0	80.0
Respiratory disease	30	20.0	20.0	100.0
Total	150	100.0	100.0	

According to the above table, the percentage of those who do not suffer from chronic diseases was 14%, with 21 patients, and the percentage of those suffering from pressure disease was 16%, with 24 patients, while the percentage of those suffering from diabetes and pressure was 22%, with 33 patients. The percentage of persons with pressure-related respiratory disorders was 20%, with a total of 30. The number of persons suffering from heart disease rose to 28%, with 42 cases.

Table 5. Method of infection

Infection	Frequency	Percent	Valid Percent	Cumulative Percent
Contact with an infected person	60	40.0	40.0	40.0
Entering a hospital	24	16.0	16.0	56.0
Patient care	66	44.0	44.0	100.0
Total	150	100.0	100.0	

The frequencies and mode of infection transmission are shown in the table above. It reached 40% of the infected people, who numbered 60, as a result of contact with a patient, and 16% as a result of hospitalization, while the percentage of infected people as a result of caring for a patient amounted to 44%, with 66 infected.

whereas 74% of those who passed away between 10 days and two weeks after infection totaled 111 infected.

Table 7. Self-association

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.73	.612	.68	12.47362	.472
a. Dependent Variable: personals name. b. Predictors: (Constant), Did the patient receive adequate care? Where did the patient receive treatment?, How did the infection happen? Symptoms associated with the disease, The Patient gender., The .duration between injury and death, Chronic diseases, The Patients age, Death place					

It is obvious from the following table that the R^2 value corresponds to 61% of the factors that were looked at and corresponded to the patient's condition, which is what caused the death.

Table 8. Shows an analysis variance model of regression

Model	Sum of Squares	Df	Mean Square	F	Sig.
Regression	3877.671	7	553.953	3.560	.004 ^b
Residual	6534.829	42	155.591		
Total	10412.500	49			
a. Dependent Variable: personals name					
b. Predictors: (Constant), Did the patient receive adequate care?, Where did the patient receive treatment?, How did the infection happen?, Symptoms associated with the disease, The Patients. Gender, Duration between injury and death, Chronic diseases, The Patients. Age, Death place					

It is evident from the preceding Table that the computed F statistic value, which is equivalent to 3.560 at a significance level of 0.05, is higher than the tabular value. Therefore, the alternative hypothesis is accepted, and we draw the conclusion that there are many factors that contributed to the death of a person infected with the coronavirus, the most significant of which are the patient's age, the presence of chronic diseases, and the location of death. The null hypothesis, which states that there is no effect on the state of death, is rejected.

Table 9. Shows a regression model of represents the data

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	45.840	18.806		2.438	.018
The Patients. Age.	-.184-	.195	-.142-	-.942-	.035
Death place	- 23.350-	4.224	-.576-	- 2.028-	.047
Chronic diseases	1.985	1.985	.141	1.000	.012
a. Dependent Variable: personals name					

The multiple regression model that depicts the research data and the independent variables that have an impact on a death event can be found in the [Table 9](#).

Y: represents the death condition.

X_1 : The independent variable representing the age of the patient.

X_2 : The independent variable representing the place of death of the injured person.

X_3 : It is the independent variable that represents the patient's chronic illnesses.

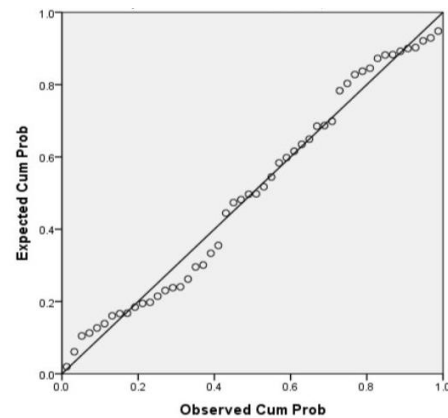


Fig. 2 Normal distribution of random errors

To test the hypothesis that random mistakes are homogeneous, the y values are shown on the horizontal axis and the random errors on the vertical axis. [Fig. 2](#) indicates that the points are equally distributed around zero, demonstrating that the model is not affected by variable degrees of random error variance, commonly known as heterogeneity.

Analyze the results obtained from SPSS

From the above outputs, we conclude the following:

It is evident from the first table that the regression technique employed is the Enter method, as the software has entered all of the independent variables in the multiple linear regression equation.

The simple correlation coefficient R reached a value of 0.73 in the second table, the Both the raw coefficient of determination R^2 (0.612) and the adjusted R^{-2} (0.68) were significant. This suggests that (0.61) of the observed changes in (death) may be

ascribed to variations in the independent explanatory variables, while the remainder (0.39) can be attributed to differences between groups.

The analysis of variance table contains the values of the analysis of variance, which may be used to determine overall explanation ability of the model using the *F* statistic. According to the analysis of the variance table, the significant

significance of the *F* test ($P = 0,004$) indicates the multiple linear regression model's excellent statistical explanatory ability.

The final table shows the constant and regression coefficient values, as well as their statistical significance for the independent factors on the dependent variable. The following is a summary of the [Table \(2-9\)](#).

Table (2-9). Showing the coefficients of the regression model with the test values

Model	Unstandardized Coefficients		Standardized Coefficients	T	Sig.
	B	Std. Error	Beta		
(Constant)	45.840	18.806		2.438	.018
The Patients. gender.	10.557	5.164	.243	2.044	.145
The Patients. age.	-.184	.195	-.142-	-.942-	.035
Death place	-23.350	11.514	-.576-	-2.028-	.047
Where did the patient receive treatment?	27.756	11.394	.687	2.436	.218
How did the infection happen?	-1.163-	3.110	-.053-	-.374-	.710
Duration between injury and death	6.346	2.874	.271	2.208	.051
Symptoms associated with the disease	.003	2.420	.000	.001	.999
Chronic diseases	1.985	1.985	.141	1.000	.012
Did the patient receive adequate care?	-14.670-	13.872	-.121-	-1.058-	.295

a. Dependent Variable: personals name

The Table shows that the independent variables (age, chronic diseases, and place of treatment) were statistically significant according to the t-test (at a significant level of $p= 0.05$), whereas the rest of the variables had no effect in the multiple regression model.

Conclusions

According to the research, the primary determinants contributing to mortality are chronic illnesses that worsen the patient's condition and exacerbate symptoms. The likelihood of death is inversely proportional to the age of the injured individual, meaning that younger individuals have a lower probability of dying compared to older individuals. Therefore, as age increases, the rate of mortality also increases. Also, the study indicated that the location where a patient undergoes treatment can influence their health results and the duration between their injury and death. Despite receiving careful attention at home, it does not match the level of medical equipment, facilities, and proficiency of the staff present in hospitals. Finally, about 38% of factors not examined in the study have an influence on the rise of mortality rates.

Recommendations

The Ministry of Health and its medical institutions, such as hospitals and healthcare centers, should increase their efforts to raise awareness among community members regarding the

nature of COVID-19, its modes of transmission, and the importance of adhering to preventative measures to lower infection rates. This can be achieved through educational materials like brochures, seminars, workshops, and courses in schools, universities, and other venues. Regular health check-ups for individuals over the age of 50 are important to detect and prevent chronic illnesses like hypertension and diabetes, as well as to promote heart health. Delayed treatment of these diseases can lead to severe consequences such as heart attacks, strokes, and kidney failure, which may increase the likelihood of complications if one contracts COVID-19. The primary mode of disease transmission is through contact with an infected individual; therefore, it is crucial to practice social distancing and adhere to safety protocols. It is important to wear a mask and practice proper hand and surface sanitization, particularly in busy and densely populated areas such as shopping malls and restaurants. Using the authorized vaccine doses can decrease the occurrence of infections. Dedication to a nourishing diet and enhancing the immune system to combat illnesses and attain well-being. Adhere to the guidelines provided by the Ministry of Health on complying with health restrictions and measures for public safety.

Conflict of Interest

The author hereby declares no conflict of interest.

Consent for publication

The author declares that the work has consent for publication.

References

- Al-Jubouri, S. H. (1990), multiple regression and analysis of variance, Ministry of Higher Education and Scientific Research, Al-Mustansiriya University.
- Al-Khathami, A. S. (2018), Violation of the assumption that the data are free of outliers and its impact on the use of multiple linear regression analysis in prediction, the comprehensive *multi-knowledge electronic journal for the publication of scientific and educational research* (MECSJ).
- Al-Masoodi, H. A., Al-Kawaz, S. M., & Abbas, A. A. (2020). Accounting Readings during the Time of COVID-19. *International Journal of Multicultural and Multireligious Understanding*, 7(5), 158-166.
- Al-Rawi, K. M. (1987), Introduction to Regression Analysis, University of Mosul, Ministry of Higher Education and Scientific Research.

- Al-Samadouni, H. F. and Al-Sayed, A. M. A. (2021), The impact of the Corona covid-19 pandemic on the level of voluntary disclosure in financial reports with a field study in the Egyptian business environment, Tanta University.
- He, H., & Harris, L. (2020). The impact of Covid-19 pandemic on corporate social responsibility and marketing philosophy. *Journal of business research*, 116, 176-182.
- Kazem, A. H. and Al-Dulaimi, M. M. A. (1988), Introduction to Linear Regression Analysis, Directorate of the Books, Printing and Publishing House, University of Mosul.
- Kazem, S. K. (2009). Comparison between estimates of the parameters of the multiple linear regression model using the OLS method and the linear goals programming method, *Journal of Administration and Economics*, Issue 77.
- The Arab Organization for Education, Culture and Science, (2020), the Corona pandemic-covid-19 - and its repercussions on the 2030 Sustainable Development Goals, ALECSO Scientific Publication, No. 2.
- Tuwaiti, M. and Wail, M. (2014), Methods of Designing and Numbering Field Studies - A Statistical Perspective.

How to cite this article

Qaddoori, A. S. (2023). Detection of the most important factors affecting the increase in the number of deaths as a result of infection with Covid-19 using the multiple linear regression equation. *Science Archives*, Vol. 4(2), 147-153. <https://doi.org/10.47587/SA.2023.4212>

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)



Publisher's Note: MD International Publishing stays neutral about jurisdictional claims in published maps and institutional affiliations.