# Using Linear Regression Analysis to Study the Recovery Cases of COVID 19 in Erbil, Kurdistan Region

**Article** · August 2021

CITATIONS
0

READS
17

**2 authors:**

Salisu Ibrahim
Yusuf Maitama Sule University
**14** PUBLICATIONS **19** CITATIONS
SEE PROFILE

Mowafaq Muhammed Tawfeeq Al-kassab
Tishk International University
**43** PUBLICATIONS **19** CITATIONS
SEE PROFILE

Drugs and Cell Therapies in Hematology (ISSN: 2281-4876)
Volume 10 Issue 1 (2021)

# Using Linear Regression Analysis to Study the Recovery Cases of COVID 19 in Erbil, Kurdistan Region

**Salisu Ibrahim[1*], Mowafaq Muhammed Al-Kassab [1]**

[1] *Department of Mathematics Education, Tishk International University, 44001 Erbil, Iraq*
[*] *salisu.ibrahim@tiu.edu.iq, mowafaq.muhammed@tiu.edu.iq*

***Article History***

*Article Received: 9/04/2021*

*Article Revised 10/05/2021*

*Article Accepted: 15/06/2021*

***Abstract.***

*This paper studies the statistical analysis of COVID 19 in Erbil, Kurdistan region by considering multiple regression model. we present the details analysis of COVID 19 in Erbil, Kurdistan using statistical model (multiple regression analysis). The regression model is used to predict the recovery cases of COVID 19 in Erbil, Kurdistan. Based on our findings from the regression model, we observed a significant difference in the regression coefficient of the model. Also, the higher the number of test patient leads to the higher number of recovery patient, moreover, there is a lesser effect in the recovery rate compared with the fatality cases. Lastly, we came up with some suggestions that will curb the spreading of COVID 19, increase the number of recovery cases, control the higher number of death cases, and prevent the future occurrence of COVID 19 in Erbil, Kurdistan region.*

## 1. Introduction

It was on Thursday morning, 24 February 2020, in Iraq, everybody was busy with their daily routine, when the Ministry of Health confirmed the first positive case of COVID 19 disease that was known for other countries now has become true and real in Iraq. The first confirmed patient was someone who came back from abroad to Iraq. The patient shows some symptoms of COVID 19 on February 24. From February 24 to March 30, 2020, the Ministry of Health declared by the President of Iraq announced a country-wide lockdown taking effect on 14 March 2020. The rapid spreading of COVID 19 has raised the total number of confirmed cases to 234934, of which 176602 have recovered and 7042 have died by the end of August 2020. The cases were increased day by day. To date, Iraq has officially confirmed to have high level cases, with 1.4 million active confirmed cases, 1.31 million recovered, and 17, 548 deaths due to COVID 19. As the virologists are focusing their attention on developing a vaccine, mathematicians rely on modeling techniques to produce multi-scenario models that could be utilized to foresee the future, see [1]. Multiple regression generally explains the relationship between multiple independent or predictor variables and one dependent in [2]. The authors in [3] studied the modeling fatality rate of COVID – 19 in Nigeria using multiple linear regression analysis. General guide associated with COVID 19 is given [4] while the obtained data and some useful information regarding this research was extracted from [5]. Some mathematical preliminaries were obtained from [6, 7].

The main aim of this paper is to develop a regression analysis model that will study the recovery cases of COVID 19 in Kurdistan region, we obtained the data regarding COVID 19 from https://gov.krd/coronavirus-en/dashboard/ and analysis/model the data using regression analysis. The

regression model is used to predict the recovery cases of COVID 19 in Erbil, Kurdistan region. Based on our findings from the regression model, we will come up with some suggestions that will curb the spreading of COVID 19, increase the number of recovery cases, control the higher number of death cases, and prevent the future occurrence of COVID 19 in Erbil, Kurdistan region. The paper is arranged as follows: Section 2 briefly explain mathematical preliminaries and statistical terms. Section 3 present the data collection and statistical analysis of COVID 19 in Erbil, Kurdistan region. method, results, and discussion are provided in Section 4. And conclusion is given in Section 5.

## 2. Mathematical Preliminaries and Statistical Definition

i. The arithmetic mean can be computed as an arithmetic formula is a group of numbers with a common difference. Arithmetic formulas are sometimes called arithmetic series or arithmetic sequences. This lesson will define an arithmetic formula and give some examples.

The formula of the Arithmetic is Arithmetic mean = $\frac{1}{n}\sum_{i=1}^{n} x_i$

ii. Geometric in Math means the average of which indicates the typical value of a set of numbers by using the product of their values (as opposed to the arithmetic mean, which uses their sum). The geometric is defined as the nth root of the product of n numbers, i.e., for a set of numbers $x_1$, $x_2$, ..., $x_n$

The formula of the geometric is Geometric mean = $(\prod_{i=1}^{n} x_i)^{\frac{1}{n}}$

iii. The harmonic: -could be such a mathematical typical. It is determined by isolating the quantity of perceptions by the correlative of each number inside the plan. Along these lines, the consonant brutal is the relating of the calculating merciless of the reciprocals.

The formula of the harmonic is harmonic mean =$\left(\frac{\sum_{i=1}^{n}\frac{1}{xi}}{n}\right)^{-1}$

iv. A standard deviation is a measurement that actions the scattering of a dataset comparative with its mean and is determined as the square base of the change.

The formula of the standard deviation= $\sqrt{\frac{\sum_{i=1}^{n}(x-\bar{x})^2}{n-1}}$ .

v. Skewness refers to a contortion or unevenness that goes amiss from the balanced ringer bend, or typical dispersion, in a bunch of information. On the off chance that the bend is moved to one side or to one side, it is supposed to be slanted. Skewness can be evaluated as a portrayal of the degree to which a given dissemination shifts from a typical appropriation. A typical conveyance has a slant of nothing, while a lognormal appropriation, for instance, would show some level of right-slant.

The formula of the skewness = $\frac{\frac{1}{n}\sum_{i=1}^{n}(x-\bar{x})^3}{\left(\frac{1}{n-1}\sum_{i=1}^{n}(x-\bar{x})^2\right)^{\frac{3}{2}}}$.

vi. The term variance refers to a factual estimation of the spread between numbers in an informational index. More explicitly, fluctuation estimates how far each number in the set is from the mean and along these lines from each and every other number in the set. Difference is regularly portrayed by this image: $\sigma^2$. It is utilized by the two investigators and brokers to decide instability and market security. The square foundation of the difference is the standard deviation (σ), which decides the consistency of a venture's profits throughout some stretch of time.

The formula of the variance is = $\frac{1}{n}\sum_{i=1}^{n}(x-\bar{x})^2$

vii. Covariance measures the directional connection between the profits on two resources. A positive covariance implies that resource returns move together while a negative covariance implies, they move contrarily.

The formula of the Covariance is = $\frac{\sum_{i=1}^{n}(xi-\bar{x})(yi-\bar{y})}{n-1}$.

viii. Pearson's correlation (also called Pearson's R) is a relationship coefficient regularly utilized in direct relapse. In case you are beginning in measurements, you will presumably find out about Pearson's R first. Indeed, when anybody alludes to the connection coefficient, they are generally discussing Pearson's. The formula of the Pearson correlation is = $\frac{\sum_{i=1}^{n}(xi-\bar{x})(yi-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2}}$.

## 3. Data Collection and Statistical Analysis of COVID 19 in Kurdistan

This section presents the data collected regarding COVID 19 from Erbil hospital for the months of October 2020, November 2020, and December 2020. The data constitutes the test results, positive results, recovery results, and dead results. The collected data are used to interpret and illustrate the histogram. pie chart, and line chart for the test results, positive results, recovery results, and dead results. The collected data and the histogram. pie chart, and line chart for the test results, positive results, recovery results, and dead results are respectively presented in the table and figures below.

Table 1: Daily data of patient for the months of October, November, and December 2020.

| The data of October 2020. | | | | | The data of November 2020. | | | | | The data of December 2020. | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Day | test | positive | recover | dead | Day | test | positive | recover | dead | Day | test | positive | recover | Dead |
| 1 | 1792 | 229 | 186 | 2 | 1 | 2176 | 398 | 119 | 5 | 1 | 1307 | 91 | 117 | 5 |
| 2 | 1640 | 192 | 187 | 2 | 2 | 2522 | 524 | 116 | 7 | 2 | 1390 | 91 | 119 | 5 |
| 3 | 841 | 103 | 158 | 3 | 3 | 2404 | 453 | 114 | 7 | 3 | 1532 | 97 | 107 | 4 |
| 4 | 2182 | 276 | 119 | 2 | 4 | 2378 | 392 | 115 | 5 | 4 | 692 | 78 | 103 | 2 |
| 5 | 1867 | 337 | 118 | 7 | 5 | 2396 | 433 | 124 | 5 | 5 | 1401 | 17 | 116 | 4 |
| 6 | 1591 | 297 | 104 | 4 | 6 | 2073 | 342 | 104 | 7 | 6 | 1548 | 65 | 113 | 3 |
| 7 | 2233 | 358 | 93 | 2 | 7 | 1331 | 241 | 123 | 4 | 7 | 1266 | 110 | 105 | 2 |
| 8 | 2153 | 281 | 105 | 4 | 8 | 1938 | 296 | 118 | 6 | 8 | 1099 | 89 | 115 | 2 |
| 9 | 1633 | 206 | 106 | 1 | 9 | 2078 | 303 | 106 | 2 | 9 | 1021 | 64 | 113 | 3 |
| 10 | 1467 | 217 | 134 | 3 | 10 | 2043 | 346 | 95 | 6 | 10 | 905 | 53 | 116 | 3 |
| 11 | 1774 | 233 | 110 | 6 | 11 | 828 | 77 | 94 | 3 | 11 | 126 | 41 | 117 | 2 |
| 12 | 2023 | 291 | 106 | 2 | 12 | 3298 | 489 | 106 | 3 | 12 | 1414 | 4 | 101 | 3 |
| 13 | 1917 | 317 | 104 | 4 | 13 | 2034 | 193 | 110 | 10 | 13 | 818 | 52 | 105 | 3 |
| 14 | 1914 | 235 | 115 | 5 | 14 | 1497 | 161 | 109 | 7 | 14 | 986 | 54 | 114 | 1 |
| 15 | 2035 | 322 | 136 | 10 | 15 | 2144 | 231 | 103 | 5 | 15 | 1856 | 43 | 119 | 0 |
| 16 | 2227 | 275 | 116 | 5 | 16 | 2495 | 223 | 112 | 4 | 16 | 1860 | 44 | 118 | 0 |
| 17 | 1170 | 156 | 117 | 3 | 17 | 1687 | 156 | 105 | 9 | 17 | 1514 | 35 | 122 | 2 |
| 18 | 2021 | 274 | 126 | 8 | 18 | 2143 | 193 | 115 | 5 | 18 | 662 | 47 | 114 | 2 |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 19 | 2102 | 326 | 114 | 6 | 19 | 2111 | 154 | 121 | 3 | 19 | 2040 | 3 | 108 | 0 |
| 20 | 1937 | 305 | 119 | 6 | 20 | 1992 | 188 | 105 | 8 | 20 | 1667 | 32 | 163 | 1 |
| 21 | 2594 | 422 | 143 | 5 | 21 | 1506 | 66 | 104 | 4 | 21 | 1558 | 44 | 176 | 1 |
| 22 | 2192 | 424 | 167 | 2 | 22 | 2636 | 193 | 116 | 3 | 22 | 1822 | 35 | 156 | 2 |
| 23 | 2289 | 461 | 134 | 4 | 23 | 1638 | 152 | 114 | 6 | 23 | 1550 | 37 | 134 | 1 |
| 24 | 1402 | 330 | 133 | 4 | 24 | 1668 | 135 | 115 | 5 | 24 | 1321 | 29 | 124 | 0 |
| 25 | 2141 | 331 | 128 | 8 | 25 | 1416 | 144 | 117 | 3 | 25 | 1730 | 30 | 134 | 0 |
| 26 | 2141 | 486 | 146 | 6 | 26 | 1905 | 123 | 117 | 2 | 26 | 977 | 14 | 122 | 2 |
| 27 | 2711 | 703 | 105 | 2 | 27 | 1292 | 105 | 128 | 7 | 27 | 1730 | 39 | 113 | 1 |
| 28 | 2718 | 596 | 117 | 6 | 28 | 827 | 39 | 102 | 4 | 28 | 977 | 22 | 121 | 1 |
| 29 | 2717 | 580 | 114 | 2 | 29 | 1237 | 98 | 119 | 3 | 29 | 1111 | 19 | 102 | 0 |
| 30 | 2250 | 424 | 105 | 7 | 30 | 1437 | 111 | 114 | 1 | 30 | 851 | 16 | 108 | 1 |
| 31 | 1310 | 214 | 117 | 5 | | | | | | 31 | 1043 | 41 | 233 | 0 |

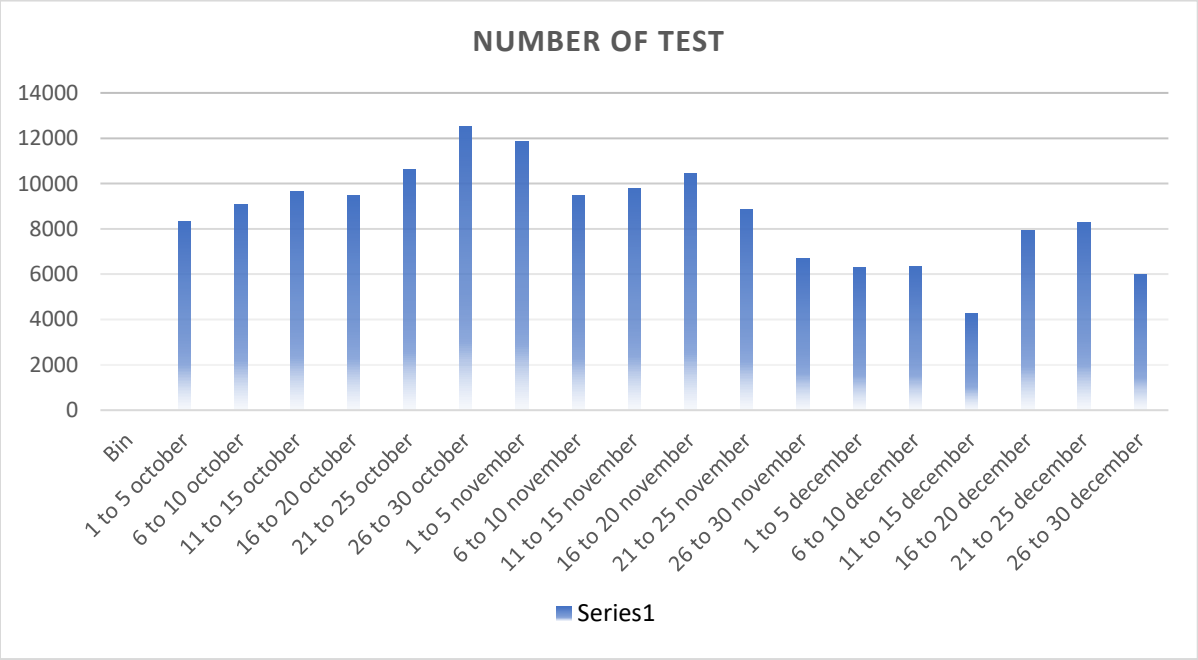


Figure 3.1: Histogram of test result for the months of October, November, and December.

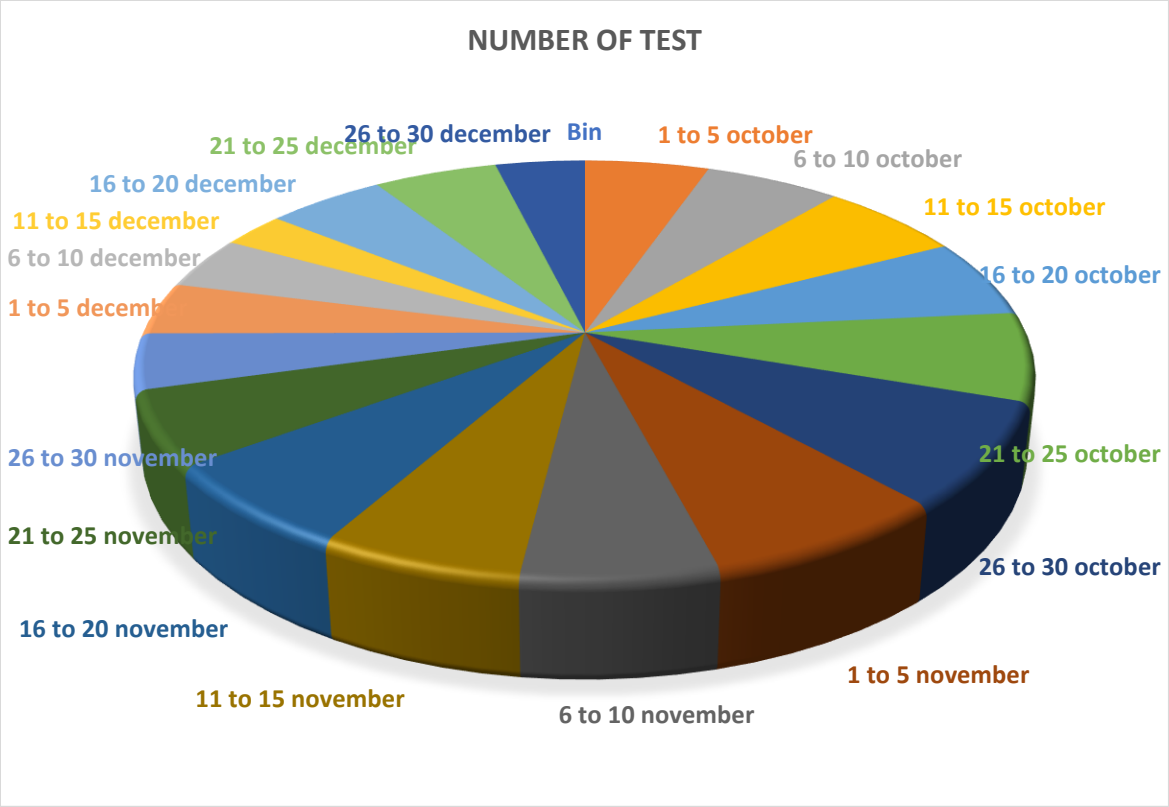


Figure 3.2: Pie chart of test result for the months of October, November, and December.

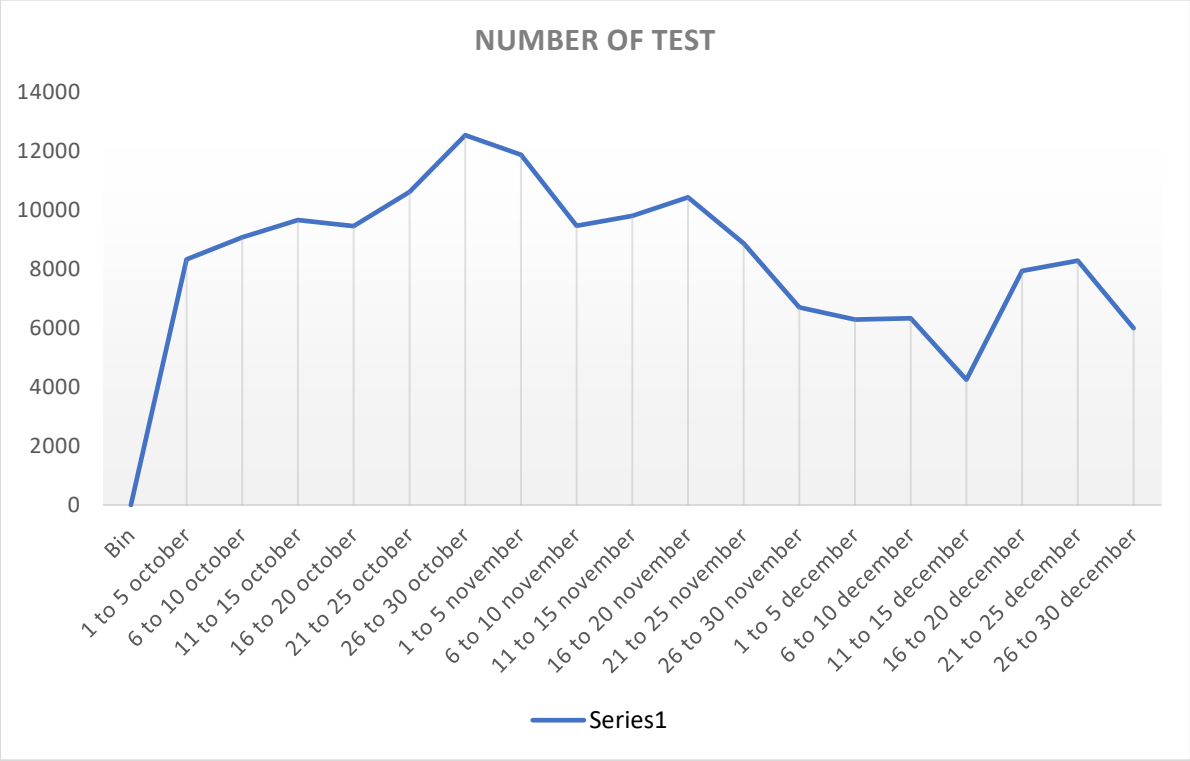

Figure 3.3: Line chart of test result for the months of October, November, and December.

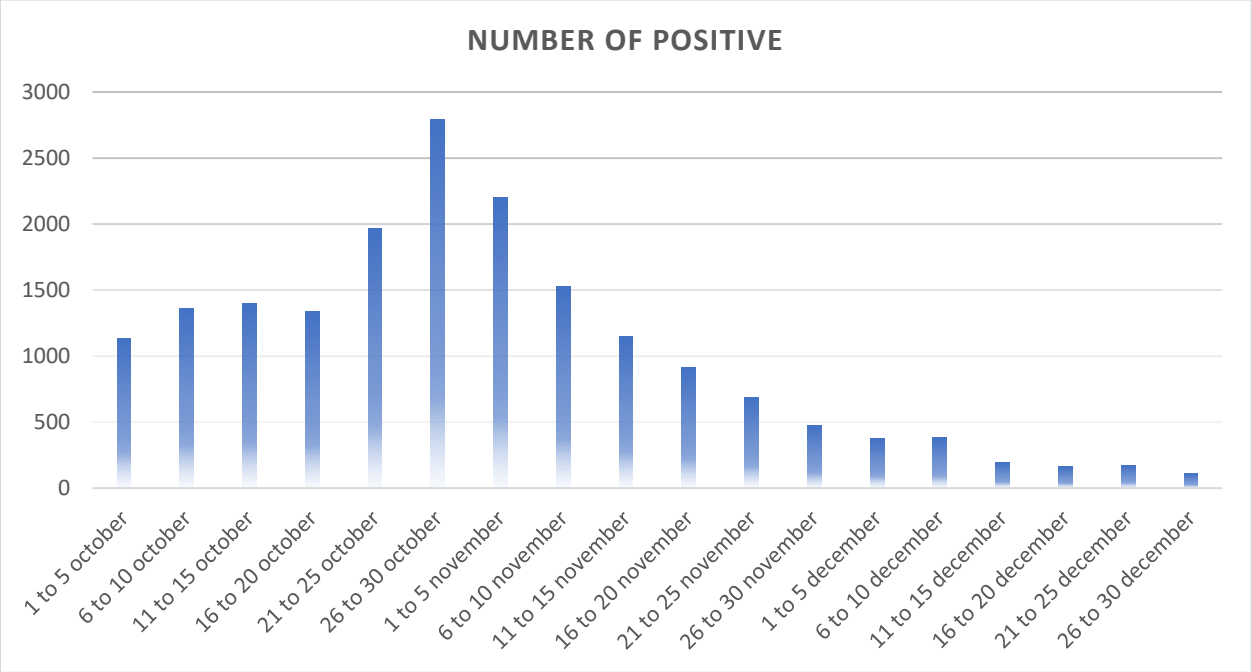


Figure 3.4: Histogram of positive result for the months of October, November, and December.

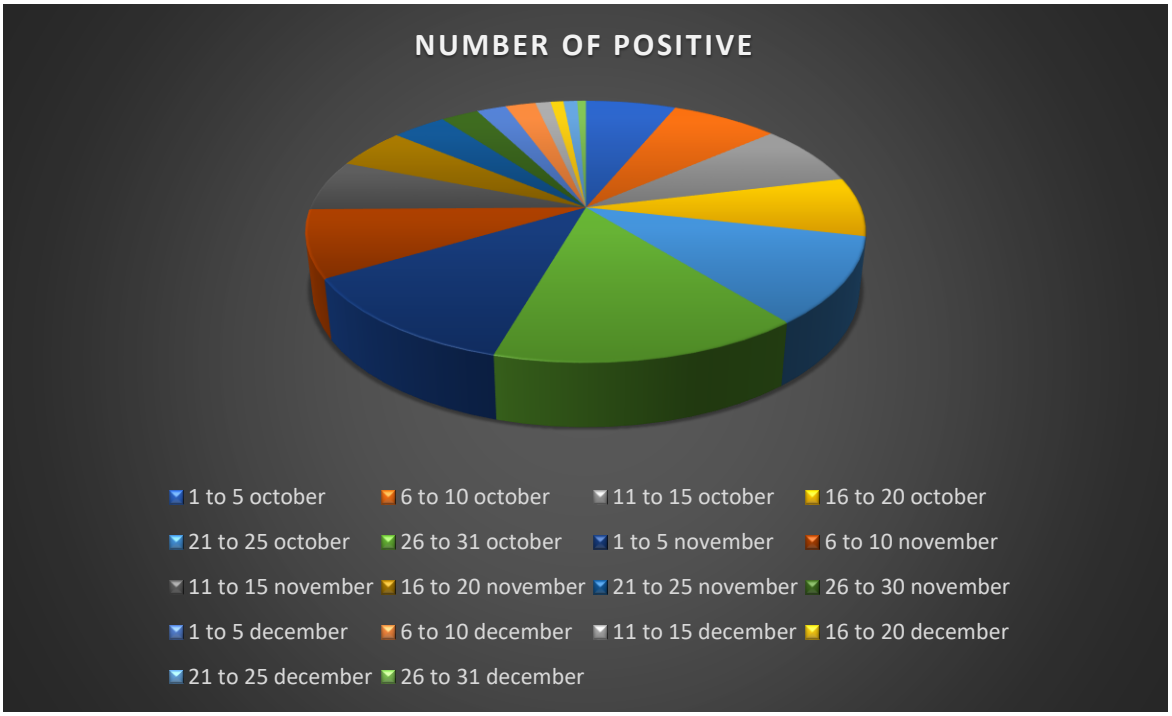


Figure 3.5: Pie chart of positive result for the months of October, November, and December.

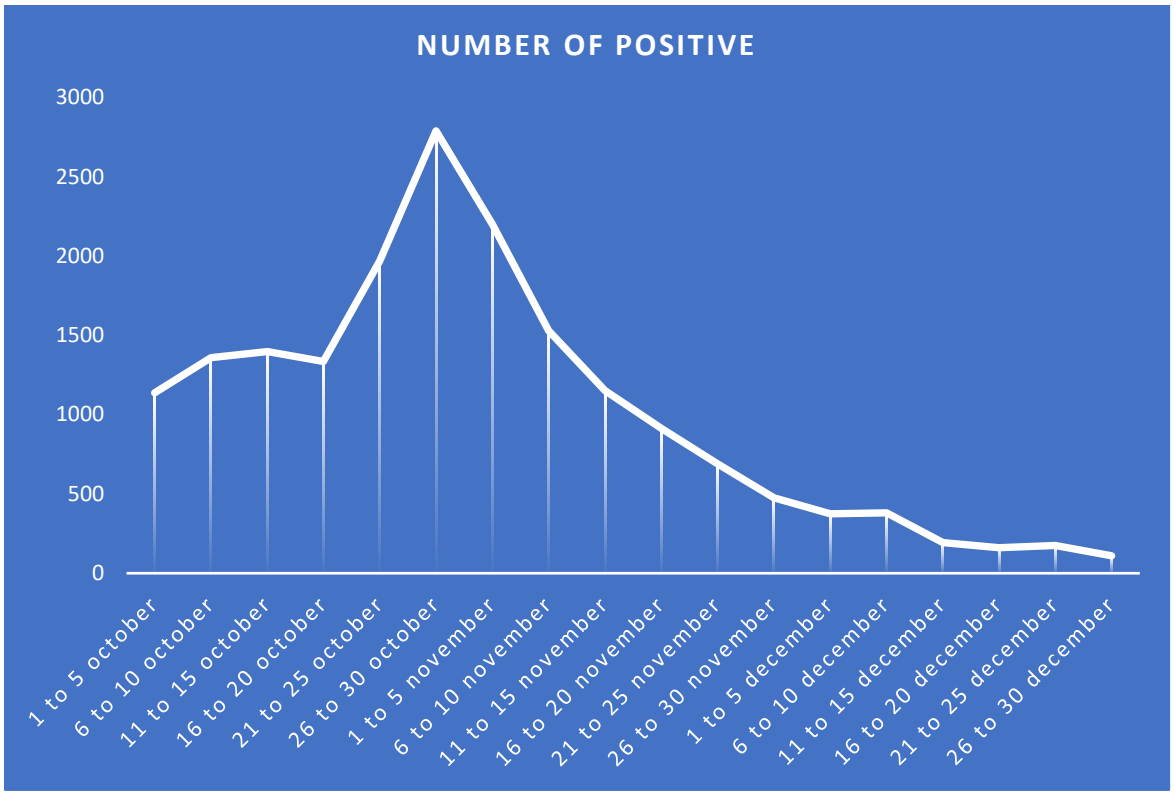


Figure 3.6: Line chart of positive result for the months of October, November, and December

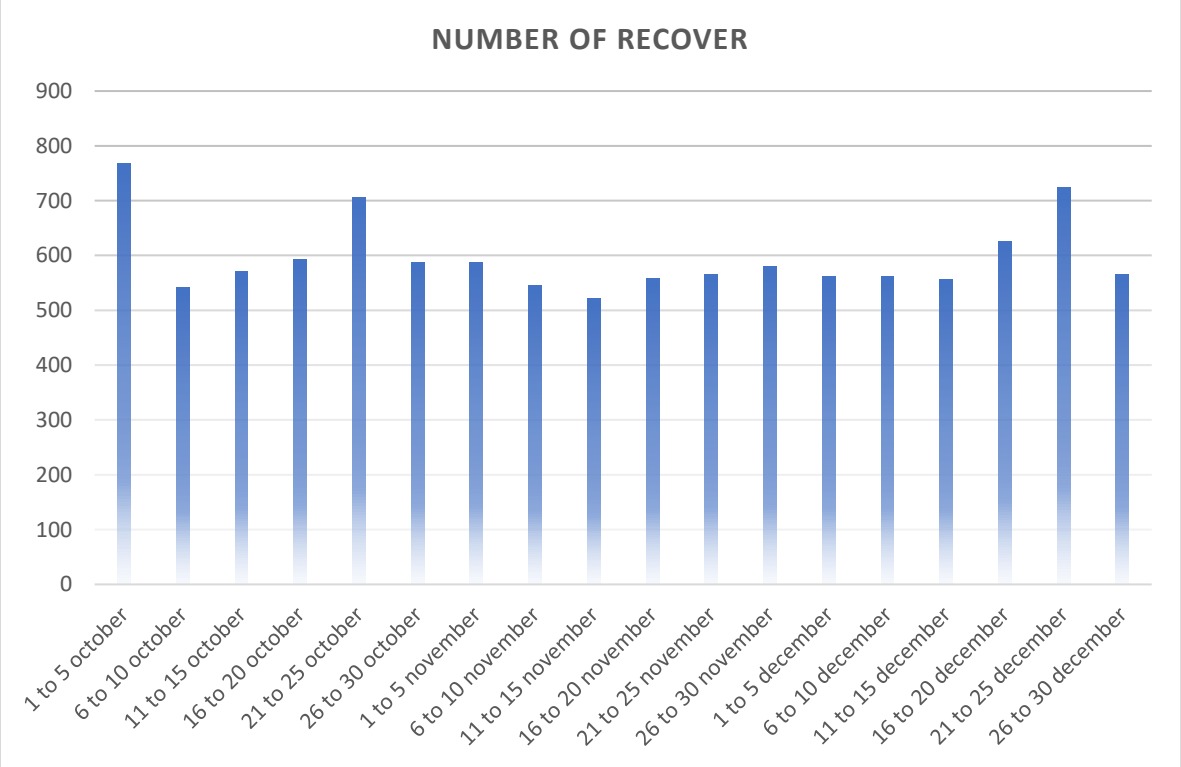


Figure 3.7: Histogram of recover result for the months of October, November, and December.

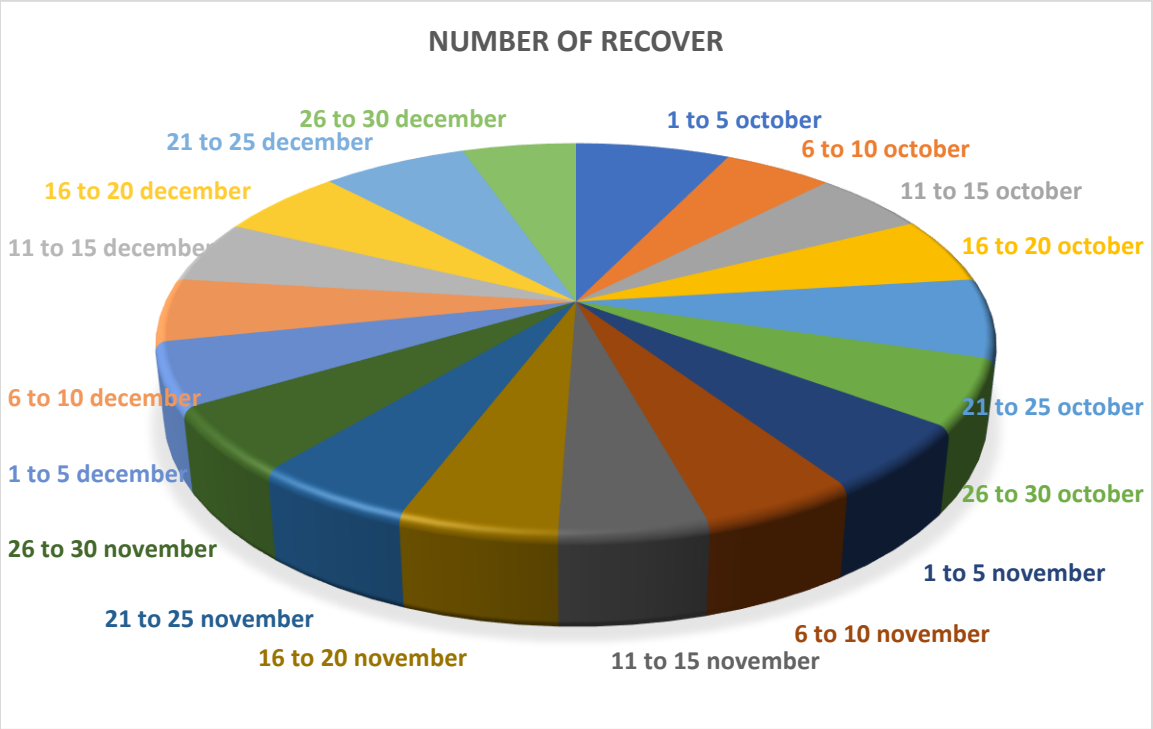


Figure 3.8: Pie chart of recover result for the months of October, November, and December.

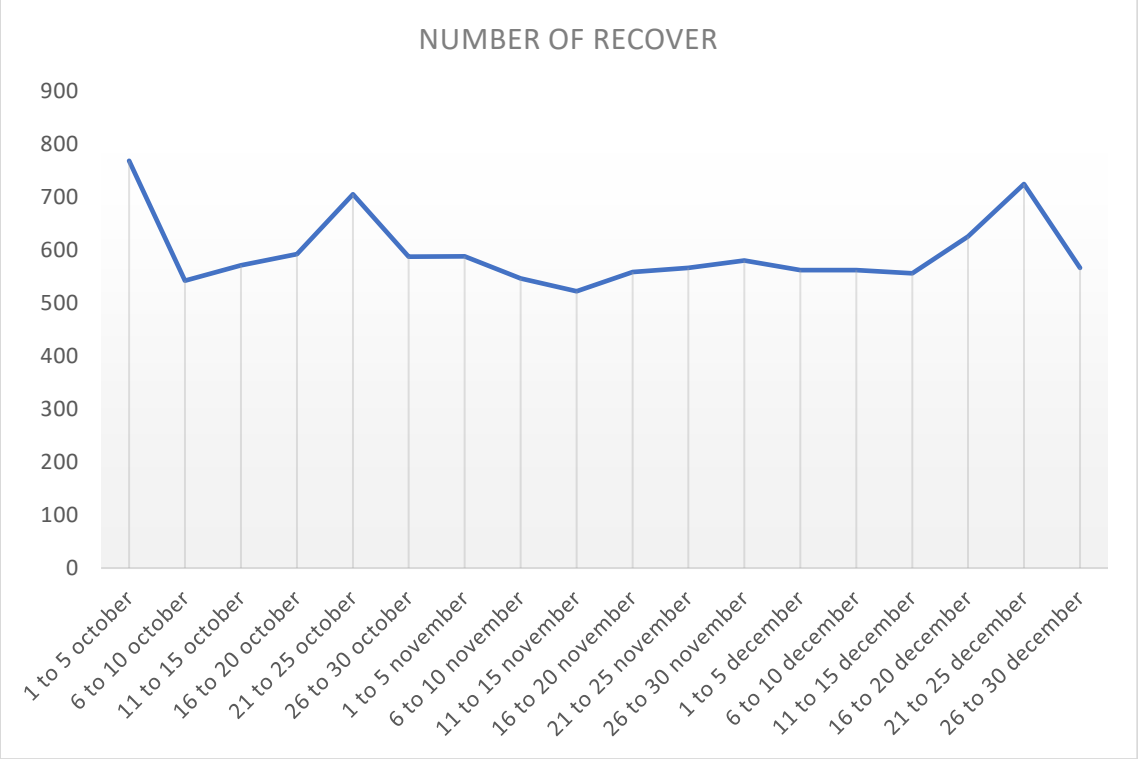


Figure 3.9: Line chart of recover result for the months of October, November, and December.

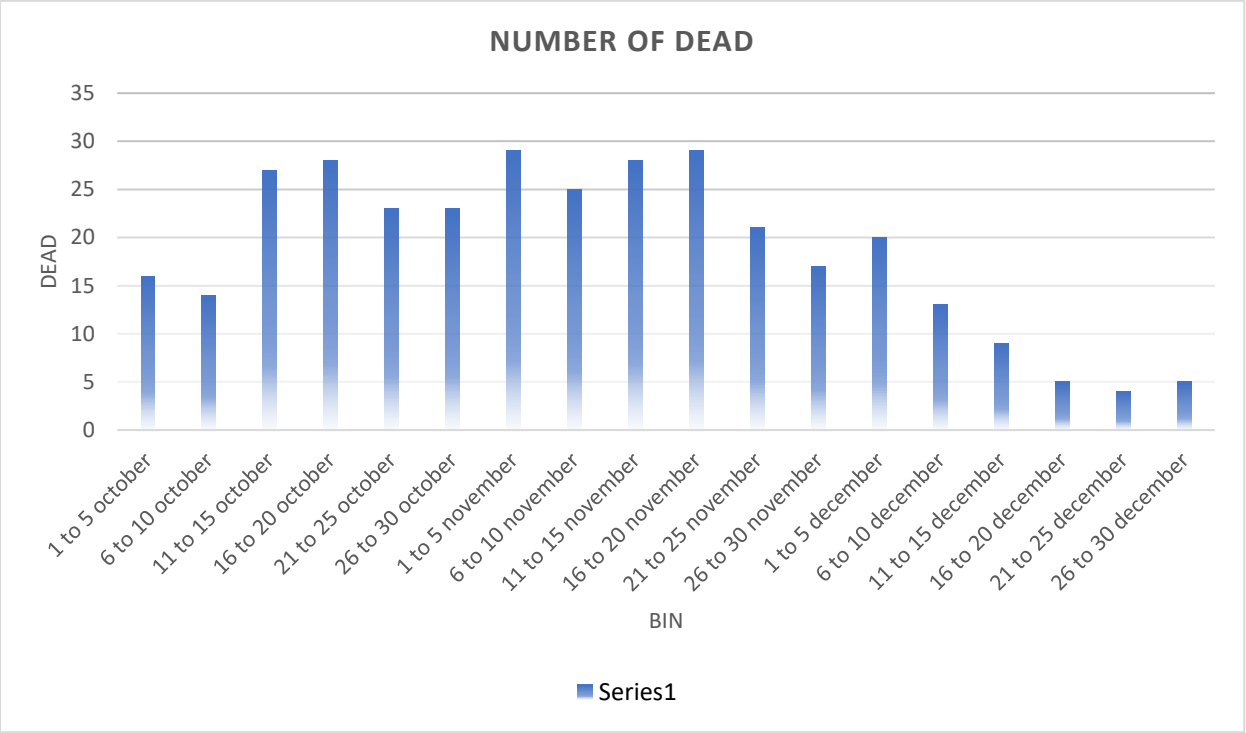


Figure 3.10: Histogram of dead result for the months of October, November, and December.

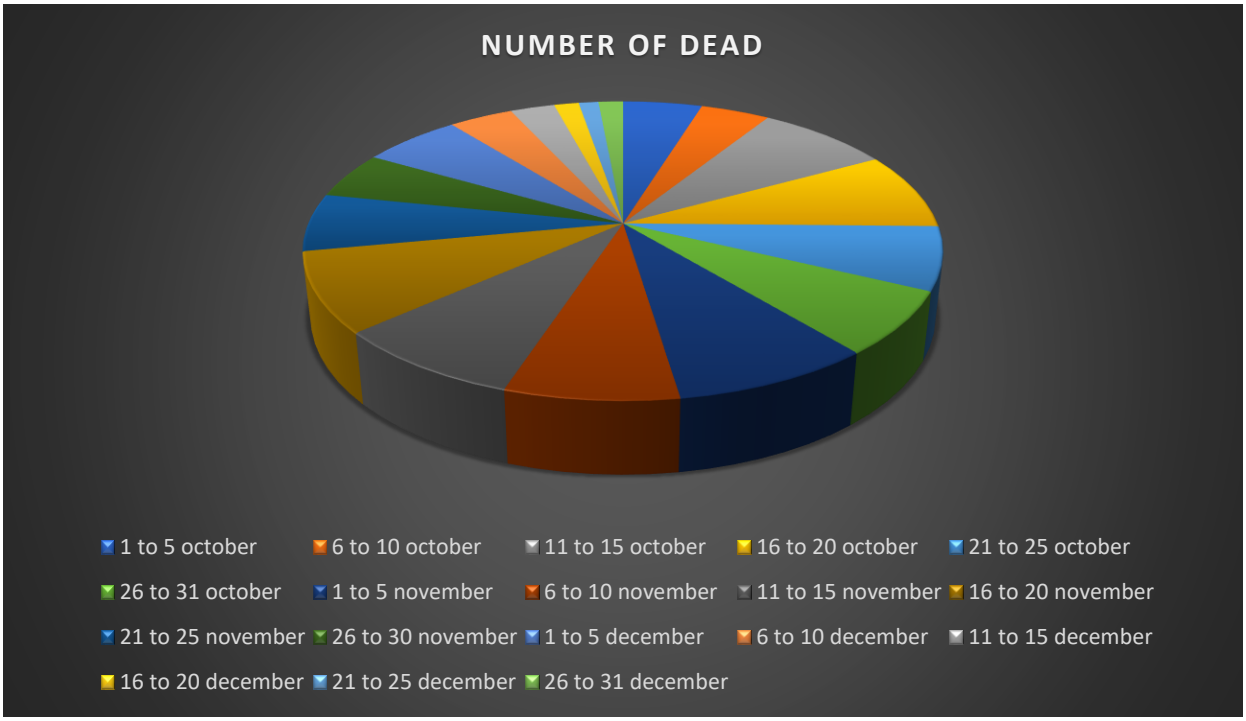


Figure 3.11: Pie chart of dead result for the months of October, November, and December.

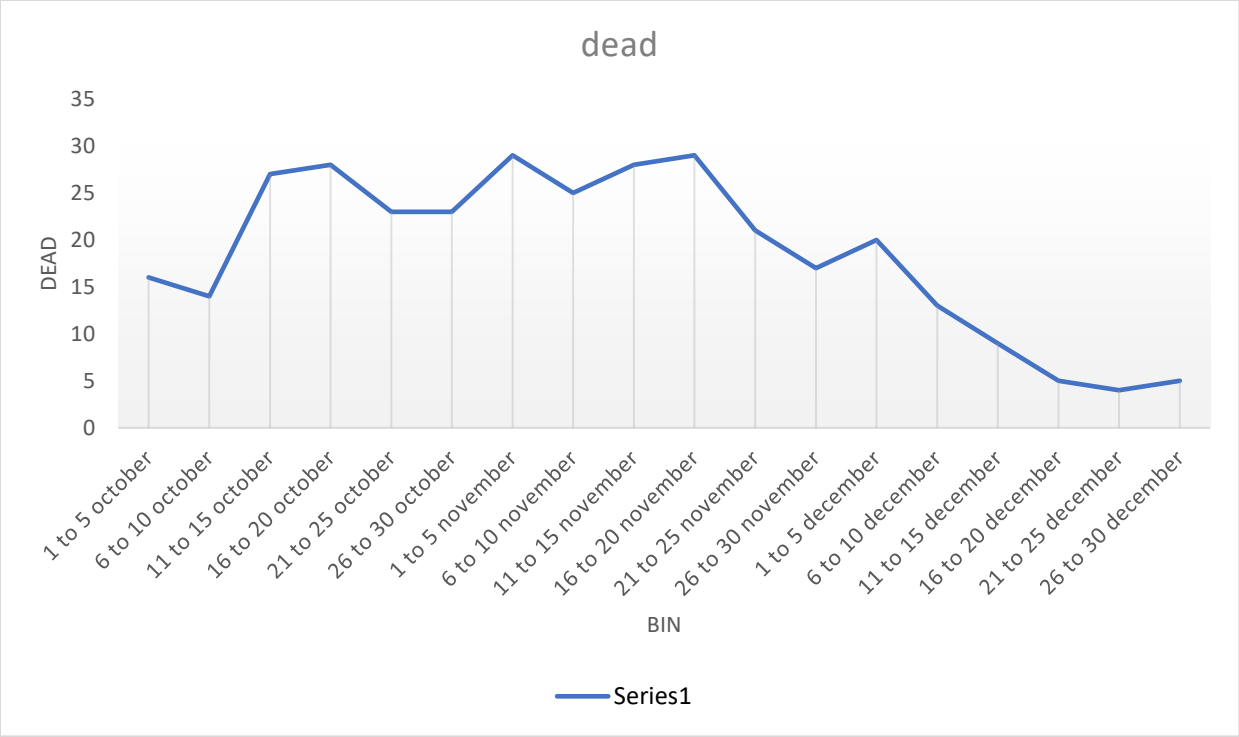


Figure 3.12: Line chart of dead result for the months of October, November, and December.

**4. Method, Results and Discussion.**

The main aim of this section is to provide the multiple linear regression analysis and results in details of the collected data from Erbil, Kurdistan region. The multiple linear regression analysis is used to investigates the recovery cases of COVID 19 in Erbil, Kurdistan region. The main goal is to come up with an estimate outcome supporting the set of variables. This case can be achieved by using the multiple linear regression analysis to find a prediction model. We want to obtain a model that fits best of the given data and demonstrate the response variable. Moreover, the model can be used to correlate the independent variables, transformation, and interaction. We are going to consider the R square ability to investigate the best model.

Consider the simple linear regression model as

$$y_i = \beta_0 + \beta_1 x_i + e_i,$$

where $\beta_0$, $\beta_1$ are the unknown constants of linear regression, $x_i$ are the independent variables, y is the dependent variable and $e_i$ is the error terms in the given data. Regarding our work, we consider the following multiple linear regression model.

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + e_i,$$

where $\beta_0$, $\beta_1$, $\beta_2$, and $\beta_3$ are the unknown coefficients of the multiple linear regression, the independent variables constitutes, $x_1$ the test cases, $x_2$ the positive cases, $x_3$ the dead cases, while y the recovery cases

which correspond to the dependent variable and $e_i$ is the error terms in the given data.

**Table 4.1: Descriptive Statistics of COVID 19.**

| Variable | N | Mean | SE Mean | Coef Var | Minimum | Q1 | Median | Q3 | Maximum | Mode | N for Mode |
|---|---|---|---|---|---|---|---|---|---|---|---|
| y | 92 | 146.40 | 8.20 | 53.73 | 32.00 | 74.25 | 122.00 | 216.50 | 315.00 | 93, 122 | 3 |
| $x_1$ | 92 | 1202.6 | 42.6 | 34.00 | 305.0 | 842.3 | 1220.5 | 1546.8 | 1932.0 | 1080,1175, 1740, 1912 | 2 |
| $x_2$ | 92 | 181.61 | 9.72 | 51.32 | 24.00 | 104.50 | 192.00 | 228.25 | 442.00 | 73 | 3 |
| $x_3$ | 92 | 5.967 | 0.385 | 61.93 | 0.000 | 3.000 | 6.000 | 8.000 | 15.000 | 4 | 14 |

**Table 4.2: Method and Correlations**

Method

| | |
|---|---|
| Correlation type | Pearson |
| Rows used | 92 |

Correlations

| | y | $x_1$ | $x_2$ |
|---|---|---|---|
| $x_1$ | 0.585 | | |
| $x_2$ | 0.285 | 0.682 | |
| $x_3$ | 0.389 | 0.419 | 0.460 |

A multiple linear regression analysis was employed to investigates the recovery cases of COVID 19 in Erbil, Kurdistan region. The cumulative number of COVID 19 test cases ( $x_1$), the confirmed positive cases ($x_2$), and the dead cases ($x_3$) are treated as the independent variables while the total number of recovery cases (y) was treated as the dependent variable. The results obtained that guaranty the model that best fits the data is presented in table 4.3.

**Table 4.3: Regression Analysis for y versus $x_1$, $x_2$, $x_3$.**

| Term | Coefficients | Std Error | T-Value | P-Value | VIF |
|---|---|---|---|---|---|
| Constant | 2.6 | 20.2 | 0.13 | 0.899 | |
| $\beta_1$ | 0.1318 | 0.0218 | 6.04 | 0.000 | 1.92 |
| $\beta_2$ | -0.2456 | 0.0980 | -2.51 | 0.014 | 2.01 |
| $\beta_3$ | 5.02 | 1.99 | 2.52 | 0.013 | 1.30 |

The model is given as

Total number of recovery cases = 2.6 + 0.1318Test Cases - 0.2456Positive Cases +5.02Dead Cases

The model recommends that to any unit increment in test cases of coronavirus, there will be 0.1318 increment in the number of recovery cases, observe that the P-value is 0.000, this indicates that the

regression coefficient is significant to the fitted model. Also, to any every unit decrement in positive cases of coronavirus, there will be 0.2456 decrement in the number of recovery cases, observe that the P-value is 0.014, this indicates that the regression coefficient is significant to the fitted model. Moreover, for any unit increment in dead cases of coronavirus, there will be 5.02 increment in the number of recovery cases, observe that the P-value is 0.013, this indicates that the regression coefficient is significant to the fitted model. Furthermore, the constant 2.6 indicate an increment in the recovery cases which is due to obeying the rules and regulations surrounding the quarantine and self-isolation.

**Table: 4.4: Model Summary**

| S | R-sq | R-sq(adj) | R-sq(pred) |
|---|---|---|---|
| 61.4472 | 40.99% | 38.97% | 36.21% |

The R-square is 0.62.

**Table: 4.5: Analysis of Variance Table and Overall Significant of the Model**

| Model | DF | Adj SS | Adj MS | F-Value | P-Value |
|---|---|---|---|---|---|
| Regression | 3 | 230767 | 76922 | 20.37 | 0.000 |
| $\beta_1$ | 1 | 137653 | 137653 | 36.46 | 0.000 |
| $\beta_2$ | 1 | 23735 | 23735 | 6.29 | 0.014 |
| $\beta_3$ | 1 | 24030 | 24030 | 6.36 | 0.013 |
| Error | 88 | 332267 | 3776 | | |
| Total | 91 | 563034 | | | |

**Table 4.6: Fits and Diagnostics for Unusual Observations**

| Obs | y | Fit | Resid | Std Resid | | |
|---|---|---|---|---|---|---|
| 43 | 93.0 | 94.5 | -1.5 | -0.03 | | X |
| 57 | 280.0 | 137.1 | 142.9 | 2.36 | R | |
| 58 | 296.0 | 157.9 | 138.1 | 2.29 | R | |
| 83 | 315.0 | 166.5 | 148.5 | 2.45 | R | |

R means large residual and X means unusual.

The overall significant indicates that there is significant difference in the regression coefficients of the model as shown in table 4.5. Which shows that the regression model best fit for the data.

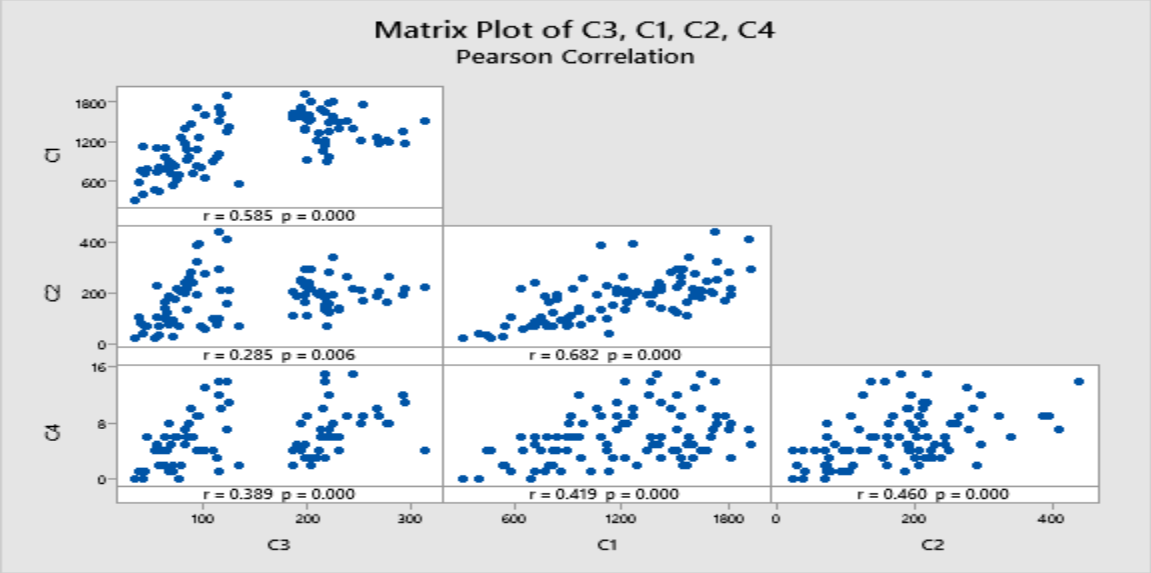


Figure 4.1: Pearson Correlation Plot for $y, x_1, x_2, x_3$,
where C3 stand for the dependent variable y, C1, C2, C4, stand for the independent variables $x_1, x_2, x_3$ respectively.

**5. Conclusion**

This paper model the recovery cases of COVID 19 in Erbil, Kurdistan region using multiple regression analysis. Based on the estimated model outcomes, we observed that, the higher the number of test patient leads to the higher number of recovery patient, moreover, there is a lesser effect in the recovery rate compared with the fatality cases. Among the few problems that leads to the higher rate of the fatality rates are the patients with underlined health challenges associated with COVID 19, COVID 19 patients with age above sixty years and the health personal which are in the front line. Our suggestions are, more attention and precautions should be given to the front-line health personal, covid patient with age above 60, patients with underlined health challenges associated with COVID 19 and lastly, government should make sure they imposed all the necessary action that will make everybody obeys the rules and regulations surrounding the quarantine, self-isolation and COVID 19 in general. Perhaps this may reduce the number of positive cases, the number of death cases, and increase the number of recovery cases.

.

**Funding**: No funding.

**Available data**: https://gov.krd/coronavirus-en/dashboard/

**Conflict of Interest**: There is no conflict of interest regarding this article.

**Acknowledgement**: The authors would like to thank all the peer reviewers for their valuable contributions.

**References**

1. A. Atangan and S. Igret Araz. (2020, May) Mathematical model of COVID 19 spread in Turkey and South Africa: Theory, methods and applications. https://www.researchgate.net/publication/341360578. DOI: 10.1101/2020.05.08.20095588.

2. Chang, S. L., Harding, N., Zachreson, C., Cliff, O. M., & Prokopenko, M. (2020). Modelling transmission and control of the COVID-19 pandemic in Australia. Nature Communication, 11(1), 1-13
3. Adejumo, T. J., Akomolafe, A. A., Owolabi, A. T., 2Okegbade, A. I., Oladapo, O. J., Idowu, J. I. and Gbolagade S. D. 2020, August. Modeling Fatality Rate of COVID – 19 in Nigeria Using Multiple Linear Regression Analysis. 8(8), 439-448.
4. World Health Organization. (2020). WHO statement regarding cluster of pneumonia cases in Wuhan, China. Beijing: WHO, 9vv, online:https://www.who.int/china/news/detail/09-01-2020-who-statementregarding-cluster-of-pneumonia-cases-in-wuhan-china.
5. https://gov.krd/coronavirus-en/dashboard/
6. Ibrahim, S. (2020). Numerical Approximation Method for Solving Differential Equations. Eurasian Journal of Science & Engineering, 6(2), 157-168.
7. Rababah, A., & Ibrahim, S. (2016b). Weighted $G^1$-Multi-Degree Reduction of Bézier Curves. International Journal of Advanced Computer Science and Applications, 7(2), 540-545. https://thesai.org/Publications/ViewPaper?Volume=7&Issue=2&Code=ijacsa&SerialNo=70