Al Kut Journal of Economics and Administrative Sciences /ISSN: 1999 -558X /ISSN Online 2707-4560/ Vol (15) Issue: 49-2023 (December)

AL KUT JOURNAL OF ECONOMIC AND ADMINISTRATIVE SCIENCES
Publisher: College of Economics and Management - Wasit University

A Predictive Model to Assess the Severity of COVID-19 Infection and the Effected Factors for Nineveh Health Workers

https://doi.org/10.29124/kjeas.1549.23

Dr.Elham Abdulkreem Hussain(1) **Dr. Ahmad Najim Sheet Al-shallawi (2)**

ABSTRACT

Covid-19 is a disease that affects the respiratory system, and causes severe symptoms that are sometimes fatal, and it is transmitted easily from one person to another, unlike other viruses. The infection with the Corona virus and the severity of infection with it are among the important issues in our lives because of the disasters it caused to humanity in all areas of life, including the economic field.

Predictive models are used in different fields such as cost , risk functions, and others that depend on different forms of data.

Prediction models use various forms of data, such as population demographics, past case numbers, and mobility data, to make predictions about the future spread of the virus. Some models focus on short-term predictions, while others aim to make long-term forecasts. The aim of this research was to develop a prediction model to predict the severity of Covid-19. The study utilized a sample of the staff working in the Nineveh Health Department who were infected while working. The ordinal logistic model was used to fit the best model. The optimal model was selected based on the Pearson's Chi-squared test.

The sample includes 536 staff with 29 predictors. It was split into a training set and a testing set, with the 70% training set used to fit the model and the 30% testing set that is used to

[1] Lecturer, e.a.hussain@ntu.edu.iq, Administrative Technical College / Mosul - Northern Technical University.

[2]. Lecturer, a.n.s.al-shallawi@ntu.edu.iq; Administrative Technical College / Mosul - Northern Technical University.

evaluate its performance Pseudo R-Square, the results showed that a few of the predictors were statistically significant in predicting the severity of Covid-19, therefore all the non-significant predictors were excluded, The confusion matrix of the training set shows a few misclassifications but the mean square error is high, the testing set results show few misclassifications, but the mean square error is lower

The final model was able to predict severity with an accuracy of 80%. The model also identified the important predictors for the outcome e.g. extreme fatigue and runny noise.

Key words: Logestic Ordinal, Prediction, Pseudo R-Square,VIF.

1. INTRODUCTION

Iraq, like many countries around the world, has been severely affected by the COVID-19 pandemic. The virus first reached Iraq in February 2020 and since then, it has spread rapidly, infecting and killing many Iraqis(Alatrany et al., 2022). Despite the efforts of the government and healthcare professionals, the number of confirmed cases and deaths continue to rise. The pandemic has had a significant impact on Iraq's healthcare system. The situation has been further exacerbated by the need of medical supplies, equipment, and infrastructure. The severity of the infection with COVID-19 in Iraq underscores the importance of taking measures to control the spread of the virus and provide support to those affected by the pandemic (Omowonuola, 2019).

Ordinal logistic regression is a widely used statistical method for analysing data that involves an ordinal response variable (Harrell et al., 1996) (Harrell & Lee, 1984). This type of regression is used when the outcome variable has more than two categories that have a natural order or hierarchy, but the differences between categories are not equal. The use of ordinal logistic regression has been applied in various fields, including health and social sciences, marketing, and economics. In the health and social sciences, ordinal logistic regression has been used to predict the severity of a disease or disorder based on patient characteristics. In marketing, it has been used to model consumer behaviour and predict purchase intent based on demographic and psychographic variables. In economics, it has been used to model customer satisfaction and loyalty.

Ordinal logistic regression prediction models can provide valuable insights into the severity of Covid-19 infections and help identify high-risk patients who may require more intensive treatment or monitoring. By analysing various risk factors and their relationship to the severity of Covid-19 infections, these models can provide a more accurate assessment of the potential outcomes for individual patients. This can help healthcare providers prioritize their resources and take a more targeted approach to treatment and care. Additionally, these models can help public health officials identify at-risk populations and implement measures to mitigate the spread of the virus. Overall, the use of ordinal logistic regression prediction models can be a valuable tool in the fight against Covid-19. Therefore, the research's aims to use the OLR to develop a model to determine the main variable contributors in explaining the Covid-19 severity.

2. MATERIALS AND METHODS

2.1 Ordinal logistic regression model

Ordinal logistic regression is a statistical model used to analyse the relationship between a set of independent variables and an ordinal dependent variable, which is a variable with three or more ordered categories (Harrell et al., 1996). The model estimates the coefficients of the independent variables and calculates the probability of each category of the dependent variable based on these coefficients (Steyerberg, 2009).

The ordinal logistic regression model is defined by the following equation:

$$ln\big(p(Y \leq j)\big) = \alpha_j + \beta_1 \mathrm{x}_1 + \beta_2 x_2 + \ldots + \beta_p x_p \qquad (1)$$

where:

- **p(Y≤j)** is the cumulative probability of the dependent variable being in category j or lower.
- $\boldsymbol{\alpha_j}$ is the intercept parameter for category j, which represents the log-odds of the dependent variable being in category j or lower when all independent variables are zero.
- $\boldsymbol{\beta_1, \beta_2, ..., \beta_p}$ are the coefficients of the independent variables $X_1, X_2, ..., X_p$, which represent the change in log-odds of the dependent variable being in category j or lower for a one-unit increase in the corresponding independent variable.

To interpret the coefficients of the independent variables, we exponential them to obtain the odds ratio. The odds ratio is defined as the ratio of the odds of the dependent variable being in category j or lower for a one-unit increase in the corresponding independent variable to the odds of the dependent variable being in category j or lower when the independent variable is held constant.

The odds ratio for the k^{th} which is coded as OR^k independent variable is calculated as follows:

$$OR^k = exp(\beta k) \qquad (2)$$

A value of OR^k greater than 1 indicates that a one-unit increase in the k^{th} independent variable is associated with an increased odds of the dependent variable being in a higher category, while a value of OR^k less than 1 indicates a decreased odd. It is also possible to calculate the marginal effects of the independent variables on the probabilities of the dependent variable being in each category. The marginal effect of the k^{th} which is coded as MEk independent variable on the probability of the dependent variable being in category j is defined as:

$$MEkj = \left(ORk^{j-1}\right)P(Y=j)\left(1-P(Y=j)\right) \qquad (3)$$

where P(Y=j) is the predicted probability of the dependent variable being in category j, and (1-P(Y=j)) is the predicted probability of the dependent variable being in a category higher than j. Overall, ordinal logistic regression provides a powerful tool for analysing the relationship between independent variables and an ordinal dependent variable. By estimating the coefficients of the independent variables and calculating the corresponding probabilities, it enables us to identify the factors that are most strongly associated with the outcome variable and to make predictions about the probabilities of different categories of the dependent variable.

2.2 Data Analysis

In this analysis, an ordinal logistic regression model was used to predict the severity of Covid-19 based on several predictors, including demographic information and laboratory results. Then, the data was split into a training set (70%) and a testing set (30%). The multi-collinearity diagnostic was undertaken to check the hypotheses that related to predictors, as it can be seen in [table 1]. The variable selection style has induced to determine the factors that have a significant effect on the severity of Covid-19. and the model was trained on the training set and evaluated on the testing set using the confusion matrix.

Table(1) The values of VIF to the variables

Series	Variance inflation factor (VIF)	Series	Variance inflation factor (VIF)
x1	13.559	x15	1.399790
x2	13.196	x16	1.319276
x3	1.225854	x17	1.177107
x4	1.088	x18	1.177107
x5	1.186	x19	1.591357
x6	1.1084	x20	1.570912
x7	1.3240	x21	1.358919
x8	1.4090	x22	1.283
x9	1.335870	x23	1.225746
x10	1.208	x24	2.046038
x11	1.2052	x25	1.185150
x12	1.434337	x26	1.879299
x13	1.4325	x27	1.209608
x14	1.5209	x28	1.363340

2.3 Training Set Results:

The results of the ordinal logistic regression model on the training set showed that only a few of the predictors were statistically significant in predicting the severity of Covid-19. The predictors were found to have p-values less than 0.05, indicating that they are significantly associated with Covid-19 severity. As it can be seen in [table2]

Table(2) The results of the ordinal logistic regression model

Series	Estimated (β)	Std. Error	T-Value	P-Value
x1	0.042304449	0.04762135	1.1394397	0.25451982
x2	-0.842609571	0.44831827	-2.2139473	**0.02683241**
x3	-0.060909866	0.28569182	-1.2144027	0.22459398
x4	0.030086365	0.05206364	0.9860431	0.32411196
x5	0.004752315	0.33674016	-0.8585968	0.39056303
x6	-0.222074739	0.33221644	-0.7238204	0.46917599
x7	-0.349489181	0.27660259	-0.7982601	0.42471957
x8	-0.142083358	0.33671789	0.3557715	0.72201167
x9	0.507773642	0.27966492	0.7799423	0.43542485
x10	-0.139303152	0.26851395	-0.5800696	0.56186766
x11	-1.155239143	0.31465301	-1.9017179 0.	05720805
x12	-0.272368289	0.29052883	-1.9908328	**0.04649927**
x13	0.378571385	0.29984696	1.0593134	0.28945709
x14	-0.121950871	0.30148804	-1.0312656	0.30241626
x15	-0.066399065	0.36753671	0.1695142	0.86539219
x16	-0.056094241	0.47421721	-0.3623411	0.71709716
x17	0.297899897	0.27754540	0.3985980	0.69018942
x18	0.347554511	0.26088296	0.9620345	0.33603226
x19	0.759724025	0.34332780	1.6799035	0.09297610
x20	-0.577585718	0.36673710	-0.6034143	0.54623314

x21	0.246528247	0.28528562	1.9336119	0.05316087
x22	0.913932864	1.16737820	0.9133848 0.	0.36104020
x23	-0.846489471	0.60249724	-0.2447976	0.80661317
x24	0.745415710	0.39807771	2.2153229	**0.02673791**
x25	-1.570709065	0.73052029	-2.3696719	**0.01780387**
x26	0.472115381	0.37307842	0.6434142	0.51995538
x27	-0.075766908	0.57141925	0.3836356	0.70124858
x28	0.532081260	0.44066021	0.8702091	0.38418616
1|2	-2.089153072	1.24651645	-2.1562068	**0.03106752**
2|3	2.711715891	1.24069231	1.7404854	0.08177381

The next step was excluded all the non-significant predictors and redevelop the model, the results as in [table 3]

Table(3) non-significant predictors

Series	**Estimated (β)**	**Std. Error**	**T-Value**	**P-Value**
X2	-0.46358457	0.1264061	-3.66742197	2.450083e-04
X6	0.01478213	-0.3220147	-0.04590512	9.633859e-01
X11	-0.65917351	0.2932582	-2.24775826	2.459161e-02
X25	-1.42276636	0.6342900	-2.24308493	2.489134e-02
1|2	-3.36309490	0.5088316	-	3.857618e-11

			6.60944565	
2|3	1.06781188	0.4590126	2.32632391	2.000128e-02

The confusion matrix of the training set showed that the model was able to correctly predict the severity of Covid-19 for most of the observations, with a misclassification rate **of 0.183**. However, the model was less accurate in predicting the most severe form of Covid-19, with no correct predictions for this category, as in [table 4]

Table (4)

Actual / Predict	Poor	Moderate	Severe
Poor	23	3	0
Moderate	39	269	3
Severe	0	0	22

2.4 Testing Set Results:

The results of the model on the testing set were consistent with the results on the training set. The confusion matrix showed that the model was again able to correctly predict the severity of Covid-19for most of the observations, with a slightly lower misclassification rate of **0.191**. Again, the model was less accurate in predicting the most severe form of Covid-19, with no correct predictions for this category.

Table (5)

Actual / Predict	Poor	Moderate	Severe
Poor	4	3	0
Moderate	17	116	8
Severe	0	1	3

2.5 Discussion

The results of the ordinal logistic regression model indicate that some of the independent variables have a significant effect on the dependent variable. The p-values associated with each coefficient in the training set results show that only a few variables have a significant effect on the dependent variable at the 5% level of significance. Specifically, the coefficients for **x2, x6, x11**, and **x25** are significant at this level, suggesting that these variables are important predictors of the dependent variable.

The confusion matrix of the training set shows that the model was able to correctly predict most of the observations, with only a few misclassifications. However, the mean square error of misclassification is relatively high, indicating that the model still has room for improvement.

The testing set results show that the model performed similarly on the test set as on the training set, with relatively few misclassifications. The mean square error of misclassification is slightly lower in the testing set than in the training set, which is a positive sign, but the difference is small.

Overall, these results suggest that the ordinal logistic regression model is a useful tool for predicting the dependent variable based on the independent variables, but that there is still room for improvement in the model. Further analysis could be done to identify additional variables that may be important predictors, or to refine the model to better fit the data. Additionally, other models could be tested to compare their performance against the ordinal logistic regression model.

3. Conclusion:

Overall, the ordinal logistic regression model showed some success in predicting the severity of covid-19 based on the selected predictors. The model was able to correctly predict the severity of covid-19 for most of the observations, but was less accurate in predicting the most severe form of covid-19. Further research may be needed to identify additional predictors that can improve the accuracy of the model, particularly for the most severe form of Covid-19.

It is also important to note that the model was based on a relatively small sample size and limited set of predictors, and that the results may not be generalizable to other populations or datasets. Nevertheless, the results provide some insight into the potential usefulness of an ordinal logistic regression model for predicting the severity of covid-19.

References

Alatrany, S. S. J., Ogden, R., Falaiyah, A. M., ALdrraji, H. A. S., & Alatrany, A. S. S. (2022). The passage of time in Iraq during the covid-19 pandemic. *PLoS ONE*, *17*(4 April), 1–15. https://doi.org/10.1371/journal.pone.0266877

Harrell, F. E., & Lee, K. L. (1984). *REGRESSION MODELLING STRATEGIES FOR IMPROVED PROGNOSTIC PREDICTION*. *3*(April 1983), 143–152.

Harrell, F. E., Lee, K. L., & Mark, D. B. (1996). Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine*, *15*(4), 361–387. https://doi.org/10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4

Omowonuola, A. (2019). Pr ep rin t n ot pe er re v iew pe er re v Pr ep rin t n ot iew. *Ijnrd*, *2*(3), 54–59. https://doi.org/10.37200/IJPR/V24I10/PR300236

Steyerberg, E. W. (2009). Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. In *Book* (Vol. 19). https://doi.org/10.1080/10543400903244270