https://doi.org/10.26682/sjuod.2023.26.2.6
Journal of University of Duhok., Vol. 26, No.2 (Pure and Engineering Sciences), Pp 47-56, 2023

USING MULTINOMIAL LOGISTIC REGRESSION TO IDENTIFY FACTORS AFFECTING PLATELET

SHERZAD MOHAMMAD AJEEL*, JIYAN ALI HAJI** and BANAZ HAMZA JAHWAR*
Dept. of Mathematics, College of Science, University of Duhok, Kurdistan Region-Iraq
Dept. of Medical Laboratory, College of Health Science, University of Cihan / Duhok, Kurdistan Region-Iraq

(Received: March 12, 2023; Accepted for Publication: August 8, 2023)

ABSTRACT

The objective of this work is to find an application for the Multinomial Logistic Regression (MLR) model, which is one of the essential methods for categorical data analysis. The focus of this paradigm is a single nominal or ordinal response variable with more than two categories. Data analysis using this method has been conducted in various disciplines, including health, social sciences, behavioral studies, and education.

To practically identify the model's application, we utilized real data from Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) patients. Five explanatory variables were included in creating the main multinomial logistic regression model. A series of statistical tests were performed to confirm the model's suitability for the data. Furthermore, the model was put to the test by randomly selecting two observations from the data to forecast their categorization based on the explanatory variable values used.

Our conclusion is that the multinomial logistic regression model enables us to effectively characterize the link between the explanatory variable set and the response variable, identify the impact of each variable, and forecast the classification of any given case.

KEYWORDS: Multinomial logistic regression model; Odds ratio; categorical data analysis; maximum likelihood method; Binary variable.

1. INTRODUCTION

Regression models have grown in significance as statistics has progressed over time, and they are now widely used in the investigation of various phenomena. Regression models originally appeared as linear and non-linear parametric models. These models are distinguished by the presumption that the sample being examined is selected from a population with a known distribution, such as a normal distribution or any other distribution. Then, using the technique of maximum likelihood, the method of determination, or other methods of estimation, the parameters of these models are estimated [1].

One of the qualitative parametric models, the logistic regression model, which we commonly refer to as (L-R), uses a descriptive variable with two or more responses as the dependent variable. It plays a significant role in the study of social phenomena. Regression models were still being developed when nonparametric and subsequently semi-parametric models first appeared. These models were discovered to be a compromise between parametric and nonparametric models, as their underlying assumptions are more robust than those of nonparametric models and simpler than those pertaining to the analysis of the connection between the models (Cox). Parametric models include the survival periods of explanatory variables in the regression model [2].

The logistic regression model is considered to be the most adaptable model among traditional regression models because it does not require the interconnection within the independent or dependent elements to be linear, nor does it assume that other explanatory elements have a normal distribution or be of the continuous or discontinuous type [3].

Regarding the Search Problem:

The techniques of linear regression serve the purpose of scientific investigation, making them an essential and fundamental part of any data analysis to describe and explain the relationship between the dependent variable and the explanatory factors. However, examining the link between independent factors and dependent variables that are binary response variables is not viable, given the prevalence of such dependent variables in the study of many phenomena.

To support this type of research, additional regression techniques are required, such as logistic regression.

While the Research Objective Include:

The logistic regression model is a method for computing statistics and determining measurement accuracy. In order to obtain more precise estimates and if it is known that the dependent variables in health studies are frequently of a qualitative nature or represent the length of patients' survival, it is important to use the logistic regression model, one of the parametric models in study and interpretation of the relationship between explanatory variables and the two-response dependent variable.

References Review

The use of the logistic regression model has been a significant topic addressed by researchers since the beginning of the twentieth century.

In 2007, Ying Liu [4] presented his doctoral thesis titled "On Goodness of Fit of Logistic." The researcher proposed the use of "Logistic Regression" to test the goodness of conformity with the model. The results showed that this method provides better results than many well-known tests. Additionally, the researcher applied it in tests of good conformance to other linear models such as the "Log-Linear Model."

In 2011, Abbasi [5], a researcher from the Department of Biological and Population Statistics at the Institute of Statistical Studies and Research at Cairo University, presented a research paper entitled "Regression," which reviewed applications in social sciences using binary and multiple logistic regression and discussed how to calculate transaction values. The coefficients of the normal linear regression model and the logistic regression model were estimated for the same data using SPSS, and the results confirmed that the logistic regression model is preferred in its analysis for binary data.

Following that, in 2012, researcher Fathy [6] presented research titled "Using Methods of Information Criteria and Model Diagnostic Methods for Choosing the Best Multiple Linear Regression Model with Application on Children with Thalassemia Patients in Mosul." The results obtained by the researcher using Akaike's Information Criteria (AIC) and Bayesian Information Criteria (BIC) showed that the use of information standards methods, specifically Adjusted R-Square (ADJRSQ), is better than the model diagnostic method, Schwarz Bayesian Criteria (SBC), for choosing the best multiple linear regression model. The researcher recommended using New Information Criteria (NIC) standards and comparing them to make the selection.

In 2016, YR Gel, V Lyubchich, LL Ramirez Ramirez [7] presented research titled "Fast Patchwork Bootstrap for Quantifying Estimation Uncertainties." The researchers proposed a new bootstrap method called "Sparse Random Networks" to address non-parametric instances when estimates for vast, random networks are ambiguous. They then designed a technique to infer from the network degree distribution function under the assumption that the degree distribution of the grid and grid system are unknown.

Limitation

This study aims to provide local data on certain hematological parameters and evaluate the correlation of platelets with these parameters, as well as with some other risk factors such as viral load, age, and gender, in cases of COVID-19. The study includes 100 patients with positive results for Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), while negative results are excluded.

2. METHOD

2.1 Multinomial Logistic Regression Concepts

The dependent variable (y), which is the response variable we are interested in studying, is a binary variable that follows the Bernoulli distribution. It takes the value (1) with probability (p) and the value (0) with probability q=(1-p). In mathematical terms, the probability mass function (PMF) of the Bernoulli distribution is given by:

$$P(Y = y) = p^y (1 - p)^{1-y}$$

where:

- Y is the random variable with values 0 or 1.
- $P(Y = y)$ is the probability that Y takes the value y (either 0 or 1).
- p is the probability of success ($P(Y = 1)$).
- $(1 - p)$ is the probability of failure ($P(Y = 0)$).

For the value 1, the probability is simply p, and for the value 0, the probability is q = (1 - p), as stated in the sentence. So, the Bernoulli distribution, as described in the sentence, is indeed given by P(Y = 1) = p and P(Y = 0) = q =

(1 - p). That is, it refers to the occurrence and non-occurrence of the response. This forms the basic premise of the logistic regression model. While both the independent and dependent variables in linear regression involve continuous values, the following model links the variables: y = b0 + b1x + e.

2.2 Multinomial Logistic Regressionis Model

The multinomial logistic regression is a fundamental extension of binary logistic regression, designed to predict a nominal dependent variable based on one or more independent variables. It can be seen as an expansion of binomial logistic regression, allowing for a dependent variable with more than two categories. Like other regression methods, multinomial logistic regression can use both nominal and continuous independent variables, as well as interactions between independent factors, to predict the dependent variable [8].

The dependent variable in L-R model, is a logistic alteration of the odds, also for the logit [9].

$$\log(odds) = logit(p) = ln\left(\frac{P}{1-P}\right) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots \quad (1)$$

Or

$$P = \frac{\exp(\alpha+\beta_1 x_1+\beta_2 x_2+\beta_3 x_3+\cdots)}{1+\exp(\alpha+\beta_1 x_1+\beta_2 x_2+\beta_3 x_3+\cdots)} \quad (2)$$

For analyzing above equation, we have

P = Is the probability that a display case in the particular category.

Exp = The exponential function (approx. 2.72)

α = Constant.

β = The predictor's or independent variables' coefficient.

2.3 Hypothesis study of Coefficients in L-R model

In our model, hypotheses appeal to:

- The alternative supposition that the model being examined is exact.
- The alternative hypothesis significantly differs from the null hypothesis (N) of zero, providing a much better prediction than chance or random occurrences. This occurs when all coefficients in the regression problem are non-zero [10].

2.4 Evaluation of Hypothesis

Next, we calculate the likelihood of each of these hypotheses resulting in the observed facts. Typically, the outcome is a very small number, so the natural logarithm is used to obtain a log probability (LL), making it easier to handle. Since probabilities are never greater than 1, LL's are always negative. Tests of a logistic model are based on log likelihood [11].

2.5 Likelihood Ratio Test

The likelihood ratio test is based on the -2LL ratio. The researcher compares the likelihood ratio (-2LL) for the model with predictors (also known as the model chi-square) to the likelihood ratio for the model with only the constant (all "b" coefficients being zero). This test determines if the researcher's model with predictors is significantly different from the one with the constant alone, at the significance level of 0.05 or lower [12].

The test measures how much more closely the explanatory variables match the data compared to the null model. Chi-square is used to estimate the significance of this correlation average, as seen in the Model Fitting Information in the SPSS output.

- H0: There is no contract within null model and final model.
- H1: There is contract between null (N) model and last model.

2.6 Some Assumptions on Logistic Regressions (L-R) Model

Logistic regression does not require a variety of the basic assumptions of linear regression and other general linear models, which depend on conventional least squares algorithms. These assumptions include linearity, normality, homoscedasticity, and measurement level [11,12].

The key points about logistic regression are as follows: [You can now present the important points about logistic regression here:

1. The measurement of the dependent variable at the nominal level is often considered the best approach.
2. One or more continuous, ordinal, or nominal independent variables (including dichotomous variables) can be used. However, ordinal independent variables need to be handled either as continuous or categorical, depending on the context and specific requirements.
3. The dependent variable should have mutually exclusive and complete categories, and the observations should be independent.
4. Multicollinearity should be avoided in logistic regression. Multicollinearity occurs when there is a strong correlation between two or more independent variables, making it challenging to determine which variable contributes to explaining the dependent variable. It also complicates the calculations in multinomial logistic regression. Therefore, an essential step

in multinomial logistic regression is to assess whether multicollinearity exists and take appropriate measures to address it.

5. Continuous independent variables in logistic regression must have a linear relationship with the logit transformation of the dependent variable.

6. No outliers, significant leverage values, or influential points should be present in logistic regression.

2.7 Factors Affecting Platelet:

There are indications from several studies that measurements are collected from healthy blood donors to investigate the physiological fluctuations of hemoglobin and platelet (PLT) concentration levels due to changes in altitude, location, seasonal variations, age, race, and sex [13].

2.7.1 Platelets and Cycle Threshold

Reverse Transcription Polymerase Chain Reaction (RT-PCR) and Cycle Threshold (CT) results are used to obtain a comprehensive picture of viral dynamics in COVID-19 patients. A higher CT value indicates a lower viral Ribonucleic Acid (RNA) burden [14]. The significance of coagulation issues and platelets in coronavirus immunopathology has led to the recent proposal of platelet count as a COVID-19 severity-associated biomarker [15]. COVID-19 causes platelet hyperactivation by increasing blood levels of platelet factor 4, soluble P-selection, and thrombopoietin [16].

2.7.2 Platelets and Age

An increased partial thromboplastin time and thrombocytopenia (low platelet count) might lead to a 28-day mortality from Corona virus infection, per a flow chart made by Zhu et al. as a help for severely unwell elderly patients [17].

2.7.3 Platelets and Neutrophil

The discovery of Neutrophil-Platelet Aggregation (NPA) by [18] further clarified the roles of leukocytes in the activation of platelets and the propagation of the coagulation process, which eventually leads to the formation of fibrin and D-dimers in COVID-19 individuals who are severely unwell. Since platelet-leukocyte aggregates may be largely responsible for the incidence of micro thrombotic sequelae and arterial/venous thrombosis in severe infections, this result provided fresh insight into the coagulopathy linked to COVID-19 [19].

2.7.4 Platelets and Lymphocyte

The Platelet Lymphocyte Ratio (PLR), an independent prognostic indicator for patients who have spent a lot of time in the hospital, increases as a result of a greater cytokine storm activating more platelets [20]. SARS-CoV-2 invades and kills lymphocytes through the Angiotensin Converting Enzyme 2 (ACE2) receptor on their membrane, resulting in lymphopenia [21]. However, 7 to 14 days following the onset of the primary symptoms, a broad rise in inflammatory mediators (cytokine storm) results in severe lymphopenia and platelet hyperactivation, which immediately results in thrombocytopenia. The lymphocyte count is normal or little lowered during the COVID-19 incubation period and early phase [22]. Additionally, a massive cytokine storm that includes interleukins and tumor necrosis factor-alpha is contributing to an increase in lymphocyte mortality (9-10) [23, 24].

2.7.5 Platelets and Gender

Age, sex, heart rate, headache, Mean Platelet Volume (MPV), and the percentage of lung involvement in pulmonary computed tomography, which were assessed during the first week of the disease, were found to be correlated with the progression to severe types of COVID-19 in 2022, according to Quispe et al. The onset of severe forms during the first week of the disease was not correlated with measures of White Blood Cells (WBC), lymphocytes, Lactate Dehydrogenase (LDH), and platelet total count; however, these variables changed in patients who had already manifested severe sickness [25].

2.8 Data Description

A study include total of hundred positive patients with SARS-CoV-19 infection, their Complete Blood Count (CBC) have been obtained with their CT value from the laboratory that they had been diagnosed in (ZANEN lab) with SARS-COVID-19 . The mean age for both sexes was 46.28 years. All patients were from Zakho city, Duhok govemate, Kurdistan region, Iraq, and According to the Guidelines on the Novel Coronavirus Infected Pneumonia Diagnosis and Treatment, they were clinically categorized. Patients' respiratory tract samples (throat swabs) were obtained in accordance with the guidelines of the Nucleic Acid Detection Kit for the 2019 Novel Coronavirus (SARS-CoV-2) for nucleic acid testing. By doing a regular blood examination, the platelet count, lymphocyte count, and neutrophil count of the patients were assessed.

In coagulation, hemostasis, thrombosis, immunomodulatory processes, and inflammation, platelets are essential players.

Changes in platelet function are brought on by pathogens and platelets interacting [26].The hematological system and hemostasis are significantly impacted by the systemic illness of COVID-19. While lymphopenia serves as a diagnostic for the degree of illness, neutrophilia is an early sign of SARS-CoV-19 infection [27].

3. ANALYZING DATA BY USING SPSS

Consider whether the additional variables statistically and substantially enhance the model when compared to the intercept alone as a first step in determining the goodness of fit (i.e., with no variables added). The "Sig." column of Table (1) shows that p = 0.445 (actually, p > 0.05), indicating that the complete model does not statistically significantly outperform the intercept-only model in terms of predicting the dependent variable.

Table (1): Model Fitting Information

Model	Model Fitting Criteria	Likelihood Ratio Tests		
	-2Log Likelihood	Chi-Square	Df	Sig.
Intercept Only	156.938			
Final	146.987	9.951	10	0.445

As seen below, the Goodness-of-Fit table offers two metrics that may be used to gauge how well the model matches the data. The Pearson chi-square statistic is shown in the top row, labeled as "Pearson." If the result is statistically significant (p < 0.05), the model does not adequately represent the data. The range of the p-value is 0.291 (from the "Sig." column of the table below), indicating that it is not statistically significant. This metric shows that the model successfully matches the data... The Pearson chi-squared statistic for testing $H_0$is

$$\chi^2 = \sum \frac{\left(n_{ij} - m_{ij}\right)^2}{m_{ij}}$$

The proposal was made in 1900 by Karl Pearson, a British statistician well-known for various contributions, including the Pearson product-moment correlation estimate. This statistic takes its minimum value of zero when all $\mathrm{n_{ij}} = \mathrm{m_{ij}}$. For a fixed sample size, greater differences $\{\mathrm{n_{ij}} - \mathrm{m_{ij}}\}$ produce larger χ^2 values and stronger evidence against $\mathrm{H_0}$.

Since larger χ^2values are more contradictory to H_0, the P-value is the null probability that $\chi^2$is at least as large as the observed value. The χ^2statistic has approximately a chi-squared distribution, for large n. The P-value is the chi-squared right-tail probability above the observed χ^2value. The chi-squared approximation improves as $\{m_{ij}\}$ increase, and $\{m_{ij} \geq 5\}$ is usually sufficient for a decent approximation[28].

The second statistic is "Deviance" the deviance formula for logistic regression, i.e. data with binary response. We have $(x_1, y_1), \ldots, (x_k, y_k)$, where $x_i \in R^p$ and $y_i \in \{0,1\}$. As usual y_i denotes the response variable and x_i being the variables we are using to explain or predict the response. Recall that deviance is :

$$D_M = -2(\log L_M - \log L_S)$$

Where L_M denotes the maximum achievable likelihood under our model while L_S denotes the likelihood under the “saturated mode”. Our model treats the x_i’s as fixed, and $y_i = 1$ with probability p_i, where p_i is a function of x_i.

Let’s compute L_M first. If , given x_i, our model predicts the success probability to $\hat{p}_i$, then the likelihood associated with this data point is

$$\hat{p}_i^{\,y_i}(1 - \hat{p}_i)^{1-y_i} .$$

Since the data points are assumed to be i.i.d., we have

$$L_M = \prod_{i=1}^{K} \hat{p}_i^{\,y_i}(1 - \hat{p}_i)^{1-y_i}.$$

$$\log L_M = \sum_{i=1}^{K} y_i \log \hat{p}_i + (1 - y_i)\log(1 - \hat{p}_i)$$

(**Note:** If our sole objective is to fit the model using maximum likelihood, we can conclude at this point, as the subsequent calculations pertain to a term that does not involve model parameters.) Next, let's compute L_S. In the saturated model, the success probability for data point i is simply y_i so

$$L_S = \prod_{i=1}^{K} {y_i}^{y_i}(1-y_i)^{1-y_i}$$

$$\log L_S = \sum_{i=1}^{K} y_i \log y_i + (1-y_i)\log(1-y_i).$$

We assume that the model accurately represents the data if the test yields no significant results (i.e., p-value > 0.05) [28].

Table (2): Goodness of Fit

	Chi-Square	Df	Sig.
Pearson	196.126	186	0.291
Deviance	146.987	186	0.984

Pseudo R-square:

SPSS can generate three tables of pseudo-R-square values for the logistic regression analysis Table (3). Unlike in OLS regression, where R^2 represents the coefficient of determination, pseudo R-square is utilized in various application areas. However, it does not provide the same interpretation as R^2 in Ordinary Least Squares (OLS) regression models.

In OLS regression, R^2 summarizes the percentage of variation in the dependent variable explained by the explanatory factors. On the other hand, pseudo-R-square in logistic regression does not have an equivalent interpretation as R^2. While the metrics suggest that the model with the highest pseudo-R-square statistic is the best, it is important to note that classification coefficients, which indicate the overall influence size, are recommended over pseudo-R-square metrics [29].

There is no metric in logistic regression data analysis that directly compares to R-squared in Ordinary Least Squares (OLS) regression. Maximum likelihood estimates are obtained using an iterative approach to create model estimates in logistic regression. The OLS technique for assessing goodness-of-fit is not applicable here since it is not computed to reduce variance. However, various "pseudo" R-squared measures have been developed to assess the goodness-of-fit of logistic models.

Although these "pseudo" R-squared values are comparable to R-squared in that they range from 0 to 1 (though some may never reach 0 or 1), they cannot be interpreted in the same way as an OLS R-squared, and different pseudo R-squared measures might produce quite different values. Higher values of pseudo R-squared suggest a better fit of the model. Due to the more frequent occurrence of floating-point precision issues with raw likelihoods, it should be noted that the majority of software programs represent the likelihood as a natural logarithm.

Table (3): Pseudo R-square

Cox and Snell	0.095
Nagelkerke	0.120
McFadden	0.063

You can identify the statistically significant independent variables in Table (4) through the Likelihood Ratio Tests. Based on the results, CT (the "CT" row) is not statistically significant since its p-value is 0.428 (from the "Sig." column). Similarly, Age (p = 0.258), Lymphocytes (p = 0.753), Neutrophils (p = 0.145), and Gender (p = 0.682) variables (the "Age," "Lymphocytes," "Neutrophils," and "Gender" rows) are also not statistically significant. The model intercept is usually of interest (i.e., the "Intercept" row).

Table (4): Likelihood Ratio Tests

	Model Fitting Criteria	Likelihood Ratio Tests		
	-2Log Likelihood	Chi-Square	Df	Sig.
Effect				
Intercept	1.470E2[a]	0.000	0	.
CT	148.683	1.695	2	0.428
Age	149.700	2.712	2	0.258
Lymphocytes	147.554	0.567	2	0.753
Neutrophils	150.846	3.859	2	0.145
Gender	147.754	0.766	2	0.682

The chi-square statistic quantifies the difference in -2 log-likelihoods between the final model and a reduced model. It represents the impact of the final model after subtracting the reduced model. The null hypothesis for each effect's parameter is a value of 0.

a. Due to the fact that eliminating the effect does not increase the degrees of freedom, the reduced model is comparable to the whole model.

The Likelihood Ratio Tests table is particularly relevant for nominal independent variables because it is the only table that addresses them, unlike the Parameter Estimates table (as shown below).

able (5): Parameter Estimates

Platelets [a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp(B)	
								Lower Bound	Upper Bound
Adequate	Intercept	3.000	2.491	1.450	1	0.229			
	CT	-0.041	0.076	0.297	1	0.585	0.960	0.827	1.113
	Age	0.030	0.028	1.170	1	0.279	1.030	0.976	1.087
	Lymphocytes	-0.042	0.592	0.005	1	0.943	0.959	0.301	3.059
	Neutrophils	-0.473	0.477	0.987	1	0.321	0.623	0.245	1.585
	[Gender=1]	-0.509	0.832	0.375	1	0.540	0.601	0.118	3.067
	[Gender=2]	0[b]	.	.	0	.	.	.	.
decrease	Intercept	1.662	2.744	0.367	1	0.545			
	CT	0.021	0.084	0.065	1	0.799	1.021	0.867	1.203
	Age	0.010	0.030	0.109	1	0.741	1.010	0.953	1.070
	Lymphocytes	0.206	0.628	0.108	1	0.742	1.229	0.359	4.205
	Neutrophils	-1.118	0.616	3.292	1	0.070	0.327	0.098	1.094
	[Gender=1]	-0.158	0.901	0.031	1	0.861	0.854	0.146	4.994
	[Gender=2]	0[b]	.	.	0	.	.	.	.

a. The reference category is: increased.
b. This parameter is set to zero because it is redundant.

The table has two rows, as you can see. This is so that these factors can contrast different result category pairings. The first row of the table, titled "Adequate," contrasts this category with the "increased" category since we designated the second category (2 = increased) as our reference category. This category is being compared to the "increased" category in the second row of the table with the label "decreased."

The primary concept behind binary logistic regression remains the same, as it involves comparing two categories. The coefficients in the logistic regression equation represent log-odds units and are used to predict the dependent variable from the independent variable, similar to OLS regression.

The odds ratio, denoted by β coefficient, is exponentiated as $\mathrm{Exp}(\beta)$. Providing the odds ratio is helpful as it may be easier to understand than the coefficient expressed in log-odds units.

Interpreting the 95% Confidence Interval for an Odds Ratio or Risk Ratio, if an odds ratio $\mathrm{Exp}(\beta)$ has a confidence interval of a to b, it suggests that there is a 95% probability that the true odds ratio would likely fall within the range $a - b$, assuming there is no bias or confounding.

The parameter estimates, also known as model coefficients, are displayed in the previous table, which is referred to as Table (5). Each variable has a corresponding coefficient, as evident from the table. However, no overall statistical significance level has been determined for these coefficients. The significance information was previously presented in Table (4).

As you can observe, there are two sets of logistic regression coefficients because there were three types of dependent variables, sometimes referred to as two logits.

"Adequate" row coefficients are found for the first set (representing the comparison of Increased category to the reference category , Increased). The "decreased" row contains the second set of coefficients (this time representing the comparison of decreased category, increased). You can see that CT for both sets of coefficients is not statistically significant ($p = 0.586,\ p = 0.799$), Age for both sets of coefficients is not statistically significant ($p = 0.279,\ p = 0.741$), Lymphocytes for both sets of coefficients is not statistically significant ($p = 0.943,\ \ p = 0.742$), and Neutrophils for both sets of coefficients are not statistically significant ($p = 0.943, p = 0.742$).

The relative log chances of having adequate vs enhanced platelets fall by 0.041 for every unit increase in the variable CT. However, a one-unit increase in the variable CT decreases the likelihood of enough platelets by 0.960 compared to an increase.

If the gender is male, the relative log chances of being in Adequate vs in Increased will fall by 0.509. In contrast, if the gender is male, the likelihood of having enough platelets decreases by 0.601 when the gender is female, compared to an increase.

A one-unit increase in the variable Age is associated with 0.030 increase in the measurement log odds of being in Adequate versus increased in platelet. On the other hand if one-unit increase in the variable Age, it increases the probability of platelets being Adequate by 1.030 comparison by increased.

The relative log probabilities of having adequate vs enhanced platelets fall by 0.042 for every unit increase in the variable lymphocytes. On the other hand if one-unit increase in the variable Lymphocytes, it decreases the probability of platelets being Adequate by 0.959 comparison by increased.

The relative log probabilities of being in adequate vs higher platelet levels fall by 0.473 for every unit rise in the variable neutrophils. On the other hand if one-unit increase in the variable Neutrophils, it reduces the probability of platelets being Adequate by 0.623. Neutrophils comparison by increased.

The relative log chances of having reduced vs increased platelets rise by 0.021 for every unit increase in the variable CT. On the other hand if one-unit increase in the variable CT , it increases the probability of platelets being decreased by 1.021 comparison by increased.

If the gender is male, the relative log chances of being in decreased versus increased will fall by 0.158. In contrast, if the gender is male, the likelihood of having enough platelets decreases by 0.854 when the gender is female, compared to an increase.

The relative log probabilities of having reduced vs increased platelets rise by 0.010 for every unit that the variable Age increases. On the other hand, if the variable Age increases by one unit, the likelihood that platelets would fall increases by 1.010 in compared to an increase.

A one-unit increase in the variable Lymphocytes is associated with a 0.206 increase in the relative log chances of having low platelets compared to high platelets. However, compared to an increase, a one-unit rise in the variable lymphocytes increases the risk that platelets will decline by 1.229.

The relative log odds of having reduced vs increased platelets fall by 1.118 when the variable Neutrophils is increased by one unit. On the other side, if the variable Neutrophils grew by one unit, it lowered the likelihood platelets would fall by 0.327 in compared to increasing.

4. CONCLUSION

To ensure the statistical validity of the model fit for the data, we conducted various tests and thoroughly examined the model's output, focusing on the odds ratio scale for parameter estimations. The likelihood ratio tests indicated that all explanatory factors were significant. However, each variable contributed differently to the model's explanation, leading us to order them based on their impact. The most influential variable was "Four," followed by "CT," "Age," "Lymphocytes," "Neutrophils," and finally "Gender."

Furthermore, the results showed that the chi-square test for the model's likelihood was not statistically significant at a level higher than 0.05 (i.e., $p > 0.05$). As a result, we cannot reject the null hypothesis, indicating that there is no statistically significant relationship between the independent variable and the dependent variable.

To evaluate the predictive capability of the model, we randomly selected two data examples and used the model to forecast their categorization based on the response variable. The model's prediction ability proved to be successful for one categorization.

Considering these significant points, we can draw the main conclusion:

1. The use of the Multinomial Logistic Regression (MLR) model enables us to handle a response categorical variable with more than two levels and a diverse set of explanatory variables.
2. When utilized in concurrent analysis, MLR showcases the individual influence of each explanatory variable as well as its cumulative effect. This is precisely what we aim to achieve in our research using this model.
3. In summary, Multinomial Logistic Regression (MLR) allows us to design a statistical model for a qualitative response variable with multiple categories, capturing intricate and linked interactions. The model equations accurately measure the influence of each explanatory variable and exclude any that are not statistically significant. This helps in building a precise and relevant model that explains the relationships among the variables and the multiple response categories.
4. The model will assist researchers examining the topic of health by providing them with insights into the relevance and effects of the variables. Additionally, researchers can compare the effects computed from different models when using similar variables.
5. When dealing with a response variable with more than two categories, the logistic regression model is suitable for various types of data. The Multinomial Logistic Regression (MLR) model, commonly used in categorical data analysis, offers extensive applications in social, educational, health, behavioral, and scientific research fields. It allows for the analysis of complex relationships without imposing limitations on the explanatory variables.

5. ACKNOWLEDGEMENTS

The authors wish to thank the college of science /Duhok university for their help and encouragement, Special thanks to Prof. dr. Aqella. Atroshy. for her technical information.

REFERENCES

Y. R. Gel, V. Lyubchic, & L.L. Ramirez, Fast Patchwork Bootstrap for Quantifying Estimation Uncertainties in Sparse Random Networks, USA, (2016).

C.R. Bilder, T.M. Loughin, Analysis of Categorical Data with R. Boca Raton, FL: Chapman & Hall/CRC, (2015).

M. H. Joseph, Practical Guide to Logistic Regression, Taylor & Francis Group, LLC, (2015).

Y. Liu, On Goodness of Fit of Logistic Regression Model, PhD. Thesis, Kansas State University, Manhattan, Kansas. (2007).

A. Abbasi, J. Altmann, L. Hossain, Identifying The Effects of Co-Authorship Networks on The Performance of Scholars: A Correlation and Regression Analysis of Performance Measures and Social Network Analysis Measures, Volume 5, Issue 4, Pages 594-607, (2011).

I. Fathy,Use of information criteria and detection model methods to select the best linear regression model with application on thalassemia children in Mosul , journal of education and science 25(2):189-200, June (2012).

Y. Gel, V. Lyubchich, L. Ramirez, Sparse Random Networks, Scientific Reports journal, (2016).

C.J. Peng, K.L. Lee & G.M. Ingersoll, An Introduction to Logistic Regression Analysis and Reporting, The Journal of Educational Research, 96(1), 3-14, (2002).

S. Aggarwal, S. Gollapudi, S. Gupta, Increased TNF-Alpha-Induced Apoptosis in Lymphocytes from Aged Humans: Changes in TNF-Alpha Receptor Expression and Activation of caspases. J Immunol, 162, 2154-2161, (1999).

D.Kleinbaum, M. Klein, Logistic Regression(Statistics for Biology and Health)

(3rd ed.), New York, NY: Springer-Verlag New York Inc, (2010).

A. Assinger, Platelets and Infection - An Emerging Role of Platelets in Viral Infection, Front Immunol, 5: 649, (2014).

A. Agresti. An Introduction to Categorical Data Analysis, New York, NY: Wiley & Sons, (1996).

H. Go , H. Ohto, K. Nollet, N. Kashiwabara, M. Chishiki, M. Hoshino, K. Ogasawara, Y. Kawasaki, N. Momoi, M. Hosoya , Perinatal Factors Affecting Platelet Parameters in Late Preterm and Term Neonates, Plos One, (2020).

K. Skok, E. Stelzl, M. Trauner, H .H. Kessler, S.F. Lax, Post-Mortem Viral Dynamics and Tropism in COVID-19 Patients in Correlation with Organ Damage. Virchows Arch.**,** 478**,** 343–353, (2021).

D. Giannis, I.A. Ziogas, P. Gianni, Coagulation Disorders in Coronavirus Infected Patients: COVID-19, SARS-CoV-1, MERS-CoV and lessons from the past. J. Clin. Virol., 127, 104362, (2020).

Y. R. Gel, V. Lyubchic, & L.L. Ramirez, Fast Patchwork Bootstrap for Quantifying Estimation Uncertainties in Sparse Random Networks, USA, (2016).

Y. Zhu, J. Zhang, Y. Li, F. Liu, Q. Zhou, Z. Peng, D. Cox, Association between Thrombocytopenia and 180-Day Prognosis of COVID-19 Patients in Intensivecare Units: a Two-Center Observational Study, Plos ONE 16 (3) e0248671, (2021) .

A. Le, L. Biard, M. Vautier, H. Bugaut, A. Mekinian, G. Maalouf, M. Vieira, A. Marcelin, M. Rosenzwajg, D. Klatzmann, JC. Corvol, O. Paccoud, A. Carillion, JE. Salem, P. Cacoub, Y. Boulaftali, D. Saadoun, Neutrophil-Platelet and Monocyte-Platelet Aggregates in COVID-19 Patients, Thromb. Haemost, 120 (12) 1733–1735, (2020).

A. Zarbock, K. Singbartl, K. Ley, Complete Reversal of Acid-Induced Acute Lung Injury By Blocking Of Platelet-Neutrophil Aggregation,J. Clin. Investig**,** 116 (12) 3211–3219,(2006).

M. Gunaldi, S. Goksu, D. Erdem, S. Gunduz, Y. Okuturlar, E. Tiken, S. Kahraman, Y. Inan, M. Yildirim, Prognostic Impact of Platelet/Lymphocyte and Neutrophil/Lymphocyte Ratios in Patients With Gastric Cancer: A Multicenter Study, Int J ClinExp Med, 8, 5937-5942, (2015).

H. Xu, L. Zhong, J. Deng, J. Peng, H. Dan, X. Zeng, Q. Chen, High Expression Of ACE2 Receptor of 2019-Ncov on the Epithelial Cells of Oral Mucosa. Int J Oral Sci; 12: 1-5, (2020).

T. Li, H. Lu, W. Zhang, Clinical Observation and Management of COVID-19 Patients. Emerg Microbes Infect; 9: 687-690, (2020).

V. Saraswat, Indian Society of Gastroenterology, Indian J Gastroenterol, 32, 1-132. 15, (2013).

S. M. Ajeel, H. Hashem, Comparison Some Robust Regularization Methods in Linear Regression via Simulation Study , Academic Journal of Nawroz University (AJNU), Vol.9, No.2, Jan, (2020).

J.F. Quispe-Pari, J.A. Gonzales-Zamora, J. Munive-Dionisio, C. Castro-Contreras, A. Villar-Astete, C. Kong-Paravicino, P. Vilcapoma-Balbin & J. Hurtado-Alegre, Mean Platelet Volume as a Predictor of COVID-19 Severity: A Prospective Cohort Study in the Highlands of Peru. Diseases, 10(2), p.22, (2022).

A. Assinger, Platelets and Infection - An Emerging Role of Platelets in Viral Infection, Front Immunol, 5: 649, (2014).

R. Angioni et al., Age-Severity Matched Cytokine Profiling Reveals Specific Signatures in Covid-19 Patients, Cell Death & Disease, vol. 11, no. 11, Available: 10.1038/s41419-020-03151-z, (2020).

A. Agresti, An Introduction to Categorical Data Analysis, 2nd edn , John Wiley & Sons, Inc., Hoboken, New Jersey, (2007).

D. Garson, Logistic Regression with SPSS, North Carolina State University, Public administration Program, (2009).