

UTILIZING MULTINOMIAL LOGISTIC REGRESSION FOR DETERMINING THE FACTORS INFLUENCING BLOOD PRESSURE

^aAzad A. Shareef, ^bSherzad M. Ajeel, *^cHussein A. Hashem

^aDept. of Statistics, College of Administration and Economics, University of Duhok, Kurdistan Region, Iraq - azada@uod.ac

^bDept. of Mathematics, College of Science, University of Duhok, Kurdistan Region, Iraq - sherzad.ajeel@uod.ac

^cDept. of Mathematics, College of Science, University of Duhok, Kurdistan Region, Iraq - hussain.hashem@uod.ac

Received: 4 Jun., 2024 / Accepted: 23 Jul., 2024 / Published: 15 Aug., 2024.

<https://doi.org/10.25271/sjuoz.2024.12.3.1322>

ABSTRACT:

The aim of this study is to investigate the practical application of the Multinomial (many explanatory variables and many categories) Logistic Regression (MLR) model, which is a fundamental tool for analyzing not only for scale data but also for categorical data with many explanatory variables. This method is primarily used when there is a single nominal or ordinal response variable with multiple categories or levels. MLR analysis has various applications across disciplines such as education, social sciences, healthcare, behavioural research, and some other fields.

We utilized real data from the Azadi Heart Center at the Duhok Hospital in the Duhok Governorate to assess the practical applicability of the model. The main multinomial logistic regression model was used with five explanatory variables. Extensive statistical tests were performed to confirm the suitability of this model for the dataset. Furthermore, the model underwent a validation process wherein two observations were randomly selected from the dataset, and their categorization was predicted based on the values of the explanatory variables utilized.

Our results suggest that the multinomial logistic regression model provides a useful method for distinguishing between the response variable and the set of explanatory factors that makes it easier to determine the exact influence of each variable and enables predictions about how a particular instance will be classified.

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

1. INTRODUCTION

Regression models have become more and more important as statistics have advanced, and they are now a standard in the study of many different phenomena. Parametric linear and non-linear models were the precursors of regression models. These models operate on the implicit assumption that the sample under investigation is drawn from a population with a known distribution, such as a normal distribution or any other previously established distribution. Then, utilizing techniques like the method of determination, maximum likelihood, or other estimating procedures, the parameters of these models are calculated [1,2].

The logistic regression model, often known as (LR), is a prominent qualitative parametric model that uses a dependent variable that is a descriptive variable with two or more answers. It is incredibly helpful in the analysis of social phenomena. As time went on, regression models evolved into nonparametric and, finally, semi-parametric models. These models were demonstrated to be a reasonable compromise between parametric and nonparametric models, offering more dependable underlying assumptions than nonparametric models while requiring fewer assumptions than those required to examine the relationship between the models (Cox). Parametric models contain factors such as the survival periods of the explanatory variables in the regression model [4].

Multinomial logistic regression is used when the dependent variable in question is nominal (equivalently categorical, meaning that it falls into any one of a set of categories that cannot be ordered in any meaningful way) and for which there are more than two categories.

Among the conventional regression models, the logistic regression model is thought to be the most adaptable. It is notable for its adaptability since it does not presume a normal distribution or continuity of other explanatory factors, nor does it need

linearity in the interactions between independent and dependent variables [4].

Scholars have been interested in the use of the logistic regression model since the early 20th century.

Ying Liu [5] made the case for the use of logistic regression in a 2007 PhD thesis titled "On Goodness of Fit of Logistic," which assessed the model's goodness of fit. The outcomes demonstrated that our approach outperformed several standard tests. Liu has employed this method to evaluate conformance with additional linear models, such as the "Log-Linear Model." Researcher Abbasi [6], who works at Cairo University's Institute of Statistical Studies and Research's Department of Biological and Population Statistics, released a study in 2011 with the title "Regression." This study discussed the computation of transaction values and delves into applications in the social sciences, with a particular emphasis on binary and multiple logistic regression approaches. Abbasi used SPSS to analyze the logistic regression model's and the normal linear regression model's coefficients for the same dataset. The results showed the effectiveness of the logistic regression model in binary data analysis.

In 2012, author Fathy [7] completed a study titled "Using Methods of Information Criteria and Model Diagnostic Methods for Choosing the Best Multiple Linear Regression Model with Application on Children with Thalassaemia Patients in Mosul." This study set out to identify the optimal multiple linear regression model through the use of information criteria techniques and model diagnostics. Applying information criteria approaches, namely Adjusted R-Square, yielded better results than the model diagnostic method Schwarz Bayesian Criteria (SBC) for selecting the optimal multiple linear regression model. LL Ramirez Ramirez, V Lyubchich, and YR Gel [8] released a research in 2016 with the title "Quantifying Estimation Uncertainties Using Fast Patchwork Bootstrap." They introduced "Sparse Random Networks," a novel bootstrap method intended

* Corresponding author

This is an open access under a CC BY-NC-SA 4.0 license (<https://creativecommons.org/licenses/by-nc-sa/4.0/>)

to handle non-parametric scenarios when estimates for large-scale random networks are not certain. They proposed a method to infer from the network degree distribution function, working under the assumption that the degree distribution of the grid and grid system is still unknown.

Sherzad Ajeel, Jian Haje, and Banaz Jahwar [9] presented a study in 2023 titled "

using multinomial logistic regression to identify factors affecting platelet." With the help of the multinomial logistic regression model, the researchers were able to predict the categorization of each individual instance, determine the influence of each variable, and adequately define the relationship between the explanatory variable set and the response variable.

The rest of the paper is organized as follows: The second section includes theoretical about material and method. The third section contains result and discussion. Moreover, the fourth section includes a description of the data. Furthermore, the last section includes a conclusion.

Concerning the Search Problem:

In scientific research, linear regression techniques are crucial instruments that provide a basic way to explain and clarify the connection between a dependent variable and explanatory factors. But they fall short when examining the relationship between independent variables and dependent variables that appear as binary response variables, which is a typical situation in the investigation of diverse phenomena. Logistic regression and other additional regression techniques are essential to meet this demand.

The objectives of the study:

The research objective encompasses the utilization of the logistic regression model as a statistical computation method to enhance accuracy in measurement. Particularly in health studies where dependent variables often exhibit qualitative characteristics or denote patients' survival durations, employing the logistic regression model becomes crucial. An essential tool for researching and analyzing the link between explanatory factors and dependent variables with binary responses is this parametric model.

Restrictions

In addition to evaluating the link between the variables and the parameters, this study intends to give local data on blood pressure and its associated factors, including age, gender, and smoking. The 100 participants in the trial were chosen at random from the Azadi Heart Hospital.

2. MATERIAL and METHOD

2.1 Concepts of Multinomial Logistic Regression:

The response variable being examined in multinomial logistic regression is represented by the dependent variable (y). The variable in question is binary and is subject to the Bernoulli distribution. Specifically, it can take two values: (1) with probability (p) and (0) with probability $q = (1 - p)$. The Bernoulli distribution's probability mass function (PMF) may be mathematically stated as follows:

$$P(Y = y) = p^y(1 - p)^{1-y} \quad y = 0, 1 \quad (1)$$

Where:

The random variable, Y , has two possible values: 0 and 1. The likelihood that Y will take the value y (either 0 or 1) is expressed as $P(Y = y)$. $P(Y = 1)$ is the probability of success, denoted by p . The chance of failure ($P(Y = 0)$) is $(1 - p)$.

As stated in the statement, the probability of an event of 1 is simply expressed as p , whereas the likelihood of an occurrence of 0 is expressed as $q = (1 - p)$. This means that the clarified Bernoulli distribution follows $P(Y = 1) = p$ and $P(Y = 0) = q = (1 - p)$. This establishes the response's occurrences and

non-occurrences, providing the basic framework for the logistic regression model. Although both the independent and dependent variables in linear regression have continuous values, the following model shows how they are related: $y = \alpha + \beta x_i + \varepsilon_i$ [10].

2.2 Multinomial Logistic Regression Model:

An important development of binary logistic regression is multinomial logistic regression, which predicts a nominal dependent variable using one or more independent variables. Because it can handle dependent variables with more than two categories, it functions as a more generalized version of binomial logistic regression. Like earlier regression techniques, multinomial logistic regression predicts the dependent variable by taking into account both nominal and continuous independent variables, along with the interactions among them [10].

The dependent variable in the LR model is a logistic modification of the chances, likewise for the logit [11].

$$\log(\text{odds}) = \text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots \quad (2)$$

Or

$$P = \frac{\exp(\alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots)}{1 + \exp(\alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots)} \quad (3)$$

We have P : Is the chance that a showcase in the specific category after studying the preceding equation.

Exp stands for the exponential function, or roughly 2.72.

α : stands for constant.

β : The coefficient of the predictor or independent variables.

2.3 Examining Coefficients in the Hypothesis of the Logistic Regression Model

Hypotheses

In our model, make reference to:

- The alternative assumption asserts the accuracy of the model being studied.
- When compared to chance or random occurrences, In terms of prediction, the alternative hypothesis performs noticeably better than the null hypothesis (N) of zero. This happens when each coefficient in the regression issue is not zero [12].

2.4 Assessment of the Hypothesis

Next, we determine how likely it is that the observed facts will be produced by each of these ideas. The result is usually a very small value; thus, managing it is made easier by applying the natural logarithm to give the log probability (LL). LLs are always negative since probabilities are always less than or equal to 1. A logistic model is assessed using log-likelihood [13].

2.5 The Likelihood Ratio Test

For the likelihood ratio test, the -2LL ratio is essential. The likelihood ratio (-2LL) of the model with predictors (sometimes referred to as the model chi-square) is compared to the model with just the constant (i.e., all "b" coefficients are zero) by the researcher. This study establishes if the researcher's model with predictors differs substantially from the model that just contains the constant at a significance level of 0.05 or lower [14].

The test determines the extent to which the explanatory variables explain the data better than the null model. Chi-square is used to determine the significance of the correlation average, as indicated in the Model Fitting Information in the SPSS report.

H_0 : The final model and the null model are not compared.

H_1 : The final model and the null model are compared.

2.6 The Logistic Regression Model's Assumptions

Logistic regression diverges from traditional linear regression and other general linear models, as it doesn't necessitate several fundamental assumptions such as linearity, normality, homoscedasticity, and measurement level, which are contingent on conventional least squares algorithms [13,14].

Key Points about Logistic Regression:

1. Utilizing a nominal level measurement for the dependent variable is often deemed the most suitable approach.
2. One may use one or more independent variables (including dichotomous variables) that are continuous, ordinal, or nominal. However, ordinal independent variables must be treated as continuous or categorical, depending on the situation and requirements.
3. To maintain the independence of observations, the dependent variable must have extensive categories that are mutually exclusive.
4. Logistic regression should avoid multicollinearity. When two or more independent variables exhibit a significant correlation, it can be difficult to identify which variable best explains the dependent variable. This phenomenon is known as multicollinearity. It further complicates multinomial logistic regression computations. Therefore, determining if multicollinearity exists and taking appropriate action to reduce it is an important step in multinomial logistic regression.
5. In logistic regression, continuous independent variables should show a linear correlation with the dependent variable's logit transformation.
6. There shouldn't be any influential points, substantial leverage values, or outliers in a logistic regression.

3. DATA DESCRIPTION

In this study, Multinomial Logistic Regression was applied to a medical dataset, and the SPSS program version 25 was used to analyze the data. A sample of data was collected randomly for male and female at the Azadi heart center at the Duhok hospital in the Duhok Governorate in Kurdistan Region of Iraq and contained 100 observations that included a dependent variable Y representing the Blood Pressure (BP) and many independent variables that include fasting blood sugar, age, sugar status, smoking, and gender.

3.1 Factors Affecting Blood Pressure

Several variables, both controllable and uncontrollable, affect blood pressure. About 95% of hypertension cases result from a mix of risk factors, while only 5% have a distinct, treatable cause. Blood pressure responses can vary even among those with similar risk profiles. Managing lifestyle choices, stress, and regular doctor visits can help lower the risk of hypertension and maintain healthy blood pressure [15]. It is shown in the following bar chart.

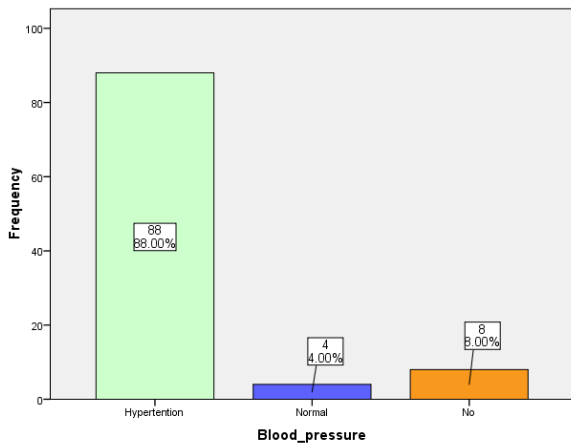


Figure 1: Bar chart for blood pressure according to the dataset.

3.1.1 Blood Pressure and Fasting Blood Sugar

Many studies have examined the link between hypertension and type 2 diabetes, but few explore the association between hypertension and fasting blood sugar in non-diabetic individuals. Fasting blood sugar levels, measured after an 8-12 hour fast, typically range from 70 to 100 mg/dL. Levels above this can indicate prediabetes or diabetes. Elevated fasting blood sugar can lead to complications like type 2 diabetes [16]. Both blood pressure and fasting blood sugar are crucial indicators of cardiovascular and metabolic health. Their relationship highlights the importance of holistic health management, as high fasting blood sugar may increase the risk of hypertension and vice versa [17].

3.1.2 Blood Pressure and Age

As individuals age, blood pressure typically increases as blood vessels undergo natural thickening and stiffening, heightening the likelihood of hypertension. Nevertheless, there's a concerning trend of rising high blood pressure among children and teenagers, potentially linked to the increasing prevalence of overweight or obesity in this study. High blood pressure frequently exhibits familial patterns, with much of our understanding derived from genetic investigations. Numerous genes are associated with slight elevations in the risk of high blood pressure. Studies indicate that certain DNA alterations during fetal development may also predispose individuals to high blood pressure later in life. Furthermore, certain individuals possess a heightened sensitivity to dietary salt intake, a factor implicated in high blood pressure, and this sensitivity often displays familial clustering [18], which can be shown in the following graph.

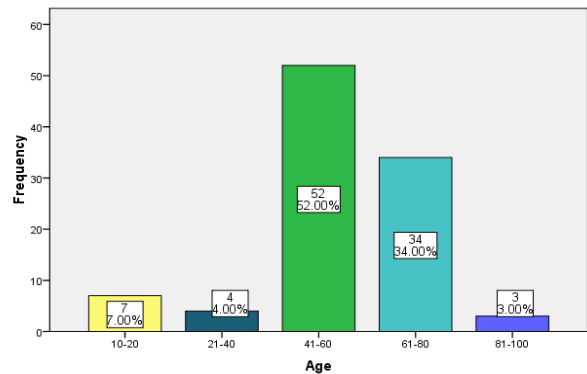


Figure 2: Bar chart Age according to the dataset.

3.1.3 Blood Pressure and Sugar Status

Blood pressure and blood sugar levels are closely linked, highlighting the need for comprehensive health management strategies focused on cardiovascular and metabolic health. Regular monitoring, lifestyle changes, and medical interventions are crucial for maintaining optimal levels [19]. Blood sugar, the concentration of glucose in the blood, comes from foods like fruits and dairy and is often added to foods for sweetness. The mean total sugar intake was determined using two 24-hour dietary recalls, and sugar-sweetened beverage consumption was assessed over the past year, which is illustrated in the chart below.

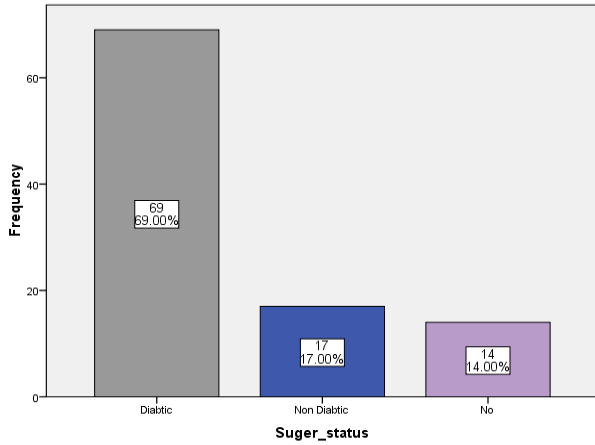


Figure 3: Bar chart for sugar status according to the dataset.

3.1.4 Blood Pressure and Smoking

Smoking is linked to severe hypertension and causes a sudden increase in heart rate and blood pressure. Nicotine, an adrenergic agonist, releases catecholamines and may boost vasopressin production. Interestingly, epidemiological studies indicate that smokers' blood pressure is often the same or lower than nonsmokers. However, 24-hour ambulatory blood pressure monitoring revealed that smokers have higher mean diurnal systolic blood pressure (SBP) than nonsmokers. Since smokers typically refrain from smoking during office visits, office BP measurements might not accurately reflect their average BP [20]. The relationship between smoking and BP was assessed using logistic and linear regression analysis. Using data from the Health Survey for England (HSE), we investigated BP values in smokers and nonsmokers [21]. The findings are presented in the following graph.

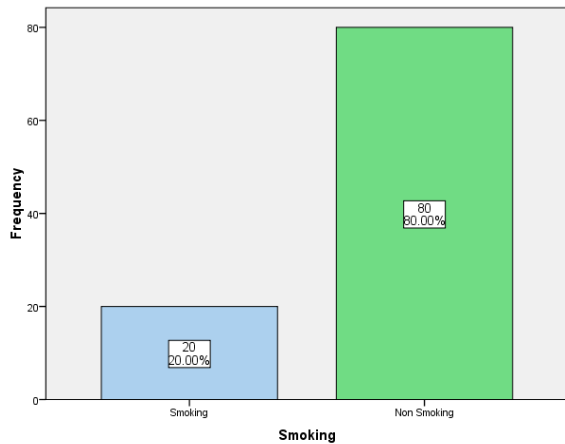


Figure 4: Bar chart for smoking according to the dataset.

3.1.5 Blood Pressure and Gender

High blood pressure (HBP) affects about 30% of adult men and women or roughly 600 million people worldwide. Its prevalence has doubled over the past three decades. Although the number of people with well-controlled blood pressure has increased, they remain a minority among those with HBP [22]. Accurate blood pressure measurements are essential for patients with hypertension and related medical conditions. Blood pressure can vary between genders due to factors such as hormonal influences, body composition, kidney function, lifestyle, and social and cultural factors. Both men and women are at risk of hypertension and related complications like heart disease and stroke. Regular monitoring, lifestyle changes, and appropriate medical management are vital for maintaining cardiovascular health in both genders. The following chart illustrates these findings [23]. It is shown in the following bar chart.

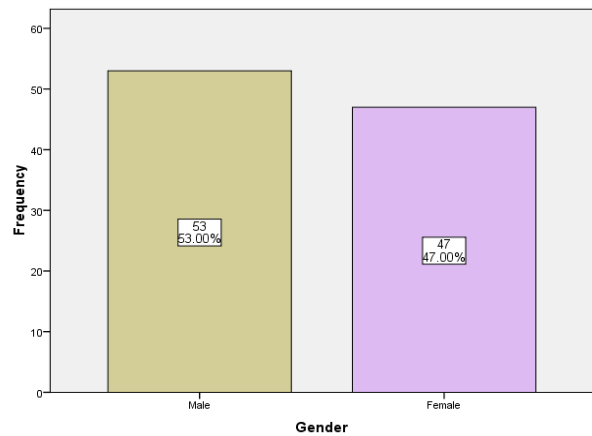


Figure 5: Bar chart for gender according to the dataset.

4. RESULT and DISCUSSION

First, assess if the inclusion of additional variables substantially improves the model compared to solely using the intercept. This initial step aids in gauging the goodness of fit. In Table (1), the "Sig." column displays that $p < 0.05$, suggesting that the comprehensive model significantly surpasses the intercept-only model in predicting the dependent variable.

- H₀: There is no effect of all factors on blood pressure
- H₁: There is an effective of all factors on blood pressure

Table 1: Model Fitting Information

Model	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	88.661			
Final	.000	88.661	10	.000

As may be seen below, the Goodness-of-Fit table provides two metrics to assess how well the model matches the data. The Pearson chi-square statistic, which is displayed in the top row under "Pearson." (relating to the "Sig." column of the table), indicates its statistical significance. This metric demonstrates how well the data fit the model. The Pearson chi-squared statistic is used to test the following equation.

$$(4) \quad \chi^2 = \sum \frac{(n_{ij} - m_{ij})^2}{m_{ij}}$$

The British statistician Karl Pearson, who is renowned for several achievements, notably the Pearson product-moment correlation estimate, came up with the concept in 1900. This statistic takes its smallest value of zero when all of n_{ij} equals m_{ij} . For a fixed sample size, larger disparities $\{n_{ij} - m_{ij}\}$ lead to larger χ^2 values and stronger evidence to reject H₀.

The P-value is the null probability that is at least as big as the observed value because greater χ^2 values are more contradictory to H_0 . For big n, the χ^2 statistic roughly follows a chi-squared distribution. The chi-squared right-tail probability above the observed χ^2 value is known as the P-value. When $\{m_{ij}\}$ increases, the chi-squared approximation gets better, and $\{m_{ij} \geq 5\}$ is often adequate for a good approximation [24].

The formula for deviation in logistic regression, or data with a binary answer, is the second statistic, or "Deviance". $(x_1, y_1), \dots, (x_k, y_k)$ are the values we have, where $y_i \in \{0,1\}$ and $x_i \in R^p$. As is customary, x_i represents the variables we are utilizing to explain or predict the answer, while y_i indicates the response variable. Remember that deviation is:

$$D_M = -2(\log L_M - \log L_S) \tag{5}$$

Where L_S represents the likelihood under the "saturated mode", and L_M represents the maximum possible likelihood under our model. In our model, $y_i=1$ with probability p_i , where p_i is a function of x_i , and the x_i values are treated as fixed. First, let's compute L_M . Should our model predict the success probability to be $p \leq i$ given x_i , then the probability corresponding to this data point is

$$\hat{p}_i^{y_i} (1 - \hat{p}_i)^{1-y_i} \tag{6}$$

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

$$L_M = \prod_{i=1}^K \hat{p}_i^{y_i} (1 - \hat{p}_i)^{1-y_i} \tag{7}$$

$$\log L_M = \sum_{i=1}^K y_i \log \hat{p}_i + (1 - y_i) \log (1 - \hat{p}_i) \tag{8}$$

(Note: The following calculations relate to a term that does not require model parameters; thus, if our only goal is to fit the model using maximum likelihood, we may stop here.) Let's compute L_S next. The success probability for a given data point i in the saturated model is just y_i .

Note: If our only goal is to fit the model with maximum probability, we may stop here because the calculations that follow include a term that has nothing to do with the model parameters.) Let us proceed to compute L_S . The success probability for data point i in the saturated model is only y_i .

$$L_S = \prod_{i=1}^K y_i^{y_i} (1 - y_i)^{1-y_i} \tag{9}$$

$$\log L_S = \sum_{i=1}^K y_i \log y_i + (1 - y_i) \log (1 - y_i) \tag{10}$$

If the test produces no significant findings (i.e., p-value > 0.05), we can decide that the model does not properly describe the data. [24].

Table 2: Goodness of Fit

	Chi-Square	df	Sig.
Pearson	.000	64	1.000
Deviance	.000	64	1.000

Three tables of pseudo-R-square values for logistic regression analysis may be obtained from SPSS, as indicated in Table (3). Unlike Ordinary Least Squares (OLS) regression, where R-squared indicates the coefficient of determination, pseudo R-square is employed in a variety of scenarios. However, it has a distinct meaning from the R-squared in OLS regression models.

R-squared gives a summary of how much of the variance in the dependent variable in an OLS regression can be explained by the explanatory factors. However, pseudo-R-square in logistic regression does not signify the same thing as R-squared. It's crucial to keep in mind that, despite the possibility that higher pseudo-R-square values indicate a better model fit, classification coefficients, which display overall influence size, are preferred above these metrics.

R-squared in OLS regression is not a data analysis metric that is directly equivalent to logistic regression. In logistic regression, model estimates are produced iteratively using maximum likelihood estimates. Since it isn't computed to minimize variance, the OLS method for evaluating goodness-of-fit isn't appropriate in this situation. Nonetheless, a number of "pseudo" R-squared metrics have been created to assess the logistic models' goodness-of-fit.

These "pseudo" R-squared values span from 0 to 1, much like R-squared, even if some of them might not reach 0 or 1. They cannot, however, be interpreted in the same way as an OLS R-squared, and the outcomes of various pseudo R-squared measurements may differ. Higher pseudo R-squared values often imply a better fit for the model. Note that while floating-point accuracy difficulties with raw likelihoods are widespread, most software portrays probability as a natural logarithm.

Table 3: Pseudo R-square

Cox and Snell	.588
Nagelkerke	1.000
McFadden	1.000

Through Likelihood Ratio Tests, statistically significant independent variables in Table (4) can be determined. In this analysis, "Fasting Blood Sugar" stands out as statistically significant, as indicated by its p-value being less than 0.05 (from the "Sig." column).

Furthermore, the p-values for the variables "Age," "Sugar Status," "Smoking," and "Gender" are all less than 0.05, indicating that they are statistically significant. It's important to remember that the model intercept, or the "Intercept" row, may also be taken into account.

Table 4: Likelihood Ratio Tests

Effect	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	.000 ^a	.000	0	.
Fasting Blood Sugar	22.181	22.181	2	.000
Age	16.535 ^b	16.535	2	.000
Sugar Status	17.282	17.282	2	.000
Smoking	13.612	13.612	2	.001
Gender	34.446	34.446	2	.000

The chi-square statistic shows the difference in -2 log-likelihoods between the final model and a reduced model. To create this simplified model, an effect is removed from the final model. The null hypothesis states that every parameter of the effect is 0.

a. The reduced model is considered equivalent to the final model because the omission of the effect doesn't elevate the degrees of freedom.

b. unexpected singularities in the Hessian matrix indicate that adjustments are needed, such as removing specific predictor variables or combining categories.

Table 5: Parameter Estimates according to hypertension category on the blood pressure

Blood Pressure ^a		B	Std. Error	Wald	df	Exp(B)
Hypertension	Intercept	9465.774	121479.035	.006	1	
	Fasting Blood Sugar	-11.330	102.062	.012	1	.000
	Age	-.037	142.393	.000	1	.000
	Sugar Status	-1728.447	19845.413	.008	1	.000
	Smoking	-1568.009	28711.861	.003	1	.000
	[Gender=1]	-701.174	25339.868	.001	1	.000
	[Gender=2]	0 ^b	.	.	0	.

a. The reference category is No.

b. This parameter is set to zero because it is redundant.

The Table (5) to contrast various pairs of outcome categories. We chose the second category (2 = No) as our reference. Coefficients for the first set are given in the "Hypertension" row, indicating how the Hypertension category compares to the reference category "No."

For every unit increase in fasting blood sugar, the relative log odds of being in the Adequate blood pressure group versus the No category decrease by 11.330. This means higher fasting blood sugar is associated with a significantly lower likelihood of being in the Adequate group.

If the gender is male, the relative log odds of being in the Hypertension group compared to the No category decrease by 701.174. This suggests that males are far less likely to fall into the Hypertension group, though the large value may indicate an error or an extreme effect.

For every unit increase in age, the relative log odds of being in the Hypertension group versus the No category decrease by 0.037. This indicates that as age increases, the likelihood of being in the Hypertension group slightly decreases.

For every unit increase in Sugar Status, the relative log odds of being in the Hypertension category compared to the No category decrease by 1728.447. This drastic decrease might suggest a data error or an exceptionally strong effect of Sugar Status.

For every unit increase in smoking, the relative log odds of being in the Hypertension category compared to the No category decrease by 1568.009. Similar to the previous point, this large decrease might indicate a possible data error or a very strong influence of smoking on blood pressure categories.

Table 6: Parameter estimates according to normal category on the blood pressure

Blood Pressure ^a		B	Std. Error	Wald	df	Exp(B)
Normal	Intercept	9449.650	124074.203	.006	1	
	Fasting Blood Sugar	-10.999	109.892	.010	1	.000
	Age	-32.787	4145.905	.000	1	.000
	Sugar Status	-1754.982	20984.627	.007	1	.000
	Smoking	-1530.268	30719.470	.002	1	.000
	[Gender=1]	-768.823	27049.914	.001	1	.000
	[Gender=2]	0 ^b	.	.	0	.

a. The reference category is No.

b. This parameter is set to zero because it is redundant.

The Table (6) is shown to contrast various pairs of outcome categories. We chose the second category (2 = No) as our reference. Coefficients for the second set are given in the "Normal" row, indicating how the Normal category compares to the reference category "No."

For every unit increase in fasting blood sugar, the relative log chances of being in the Normal blood pressure group versus the No category decrease by 10.999. This indicates that higher fasting blood sugar is significantly associated with a lower likelihood of having normal blood pressure.

If the gender is male, the relative log chances of being in the Normal blood pressure group compared to the No category

decrease by 768.823. This suggests that males are far less likely to have normal blood pressure, though the large value may indicate an error or an extreme effect.

For every unit increase in age, the relative log chances of having blood pressure in the Normal group versus the No category decreased by 32.787. This means that as age increases, the likelihood of having normal blood pressure significantly decreases.

For every unit increase in sugar status, the relative log chances of having blood pressure in the Normal group versus the No category decrease by 1754.982. This drastic decrease might

suggest a data error or an exceptionally strong effect of sugar status.

For every unit increase in smoking, the relative log chances of having normal blood pressure in the Normal group compared to the No category decreased by 1530.2678. This large decrease suggests a very strong negative influence of smoking on normal blood pressure, or it could indicate a possible data error. $\text{Exp}(\beta)$ is the exponentiation of the odds ratio, which is represented by the β coefficient. It is advantageous to present the odds ratio as it may be simpler to understand than the coefficient written in log-odds units.

The previous table, referred to as Table (5) and Table (6), displays the parameter estimations, which are also called model coefficients. The table shows that each variable has a matching coefficient. Nevertheless, these coefficients lack an unambiguous overall statistical significance level. Table (4) included the significant information earlier.

The comparison of two categories is the underlying idea that underpins binary logistic regression. Similar to OLS regression, the logistic regression equation uses coefficients to forecast the dependent variable from the independent variable. These coefficients represent log-odds units.

CONCLUSION

We ran a number of tests and closely evaluated the results, paying particular attention to parameter estimations on the odds ratio scale, to make sure the model suited the data statistically. Based on the likelihood ratio tests, every explanatory factor was found to be significant. We had to order the variables according to their influence, though, because each one contributed in a different way to the explanation of the model. Interestingly, "Four" turned out to be the most important variable, with "Fasting Blood Sugar," "Age," "Sugar Status," "Smoking," and "Gender" coming in order of importance.

Furthermore, the results indicated that the likelihood chi-square test for the model was statistically significant at a level less than 0.05 ($p < 0.05$). There was a statistically significant connection between the independent and dependent variables, as evidenced by the null hypothesis being rejected.

We used the model to predict the classification of two data instances based on the response variable after selecting them at random. This allowed us to evaluate the predictive ability of the model. The prediction power of the model worked well for one classification.

In light of these important details, the primary conclusion may be summed up as follows:

1. When dealing with answer categorical variables that have more than two levels and a variety of explanatory factors, the Multinomial Logistic Regression (MLR) model has been shown to be an effective tool.
2. When used in a contemporaneous analysis, MLR clarifies each explanatory variable's individual impact as well as their combined effect, which is exactly in line with the goals of the study.
3. To put it simply, MLR makes it easier to build a statistical model that captures complex and interrelated interactions for qualitative response variables that have several categories. Each explanatory variable's impact is efficiently quantified by the model equations, which exclude variables with little statistical significance. As a result, a precise and pertinent model that clarifies the links between various answer categories and variables is developed.
4. The model provides insights into the significance and consequences of many factors, making it a useful tool for academics looking into health-related problems. Researchers can also contrast the results of several models using comparable variables.

5. The logistic regression model, particularly the Multinomial Logistic Regression (MLR) variation, is a flexible tool appropriate for various forms of data analysis when response variables have more than two categories. In the fields of social, educational, health, behavioural, and scientific research, MLR finds wide applications that allow the investigation of intricate interactions without limiting the explanatory factors.

REFERENCES

- Y. R. Gel, V. Lyubchic, & L.L. Ramirez, Fast Patchwork Bootstrap for Quantifying Estimation Uncertainties in Sparse Random Networks, USA, (2016).
- 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).
- 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).
- S. M. Ajeel, J. A. Haji, B. H. Jahwar. Using Multinomial Logistic Regression to Identify Factors Affecting Platelet, Journal of University of Duhok., Vol. 62, No.2, (2023).
- 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).
- Johansson, J.K., Niiranen, T.J., Puukka, P.J. and Jula, A.M., 2010. Factors affecting the variability of home-measured blood pressure and heart rate: the Finn-home study. Journal of hypertension, 28(9), pp.1836-1845.
- Kuwabara, M. and Hisatome, I., 2019. The relationship between fasting blood glucose and hypertension. American Journal of Hypertension, 32(12), pp.1143-1145.
- Lv, Y., Yao, Y., Ye, J., Guo, X., Dou, J., Shen, L., Zhang, A., Xue, Z., Yu, Y. and Jin, L., 2018. Association of blood pressure with fasting blood glucose levels in Northeast China: a cross-sectional study. Scientific reports, 8(1), p.7917.
- Leung, H., Wang, J.J., Rochtchina, E., Tan, A.G., Wong, T.Y., Klein, R., Hubbard, L.D. and Mitchell, P., 2003. Relationships between age, blood pressure, and retinal

- vessel diameters in an older population. *Investigative ophthalmology & visual science*, 44(7), pp.2900-2904.
- Murphy SP, Johnson RK. The scientific basis of recent US guidance on sugars intake. *Am J Clin Nutr*. 2003;78:827S–833S.
- Primatesta, P., Falaschetti, E., Gupta, S., Marmot, M.G. and Poulter, N.R., 2001. Association between smoking and blood pressure: evidence from the health survey for England. *Hypertension*, 37(2), pp.187-193.
- Mann, S.J., James, G.D., Wang, R.S. and Pickering, T.G., 1991. Elevation of ambulatory systolic blood pressure in hypertensive smokers: a case-control study. *Jama*, 265(17), pp.2226-2228.
- Zhou, B., Carrillo-Larco, R.M., Danaei, G., Riley, L.M., Paciorek, C.J., Stevens, G.A., Gregg, E.W., Bennett, J.E., Solomon, B., Singleton, R.K. and Sophiea, M.K., 2021. Worldwide trends in hypertension prevalence and progress in treatment and control from 1990 to 2019: a pooled analysis of 1201 population-representative studies with 104 million participants. *The Lancet*, 398(10304), pp.957-980.
- Song, J.J., Ma, Z., Wang, J., Chen, L.X. and Zhong, J.C., 2020. Gender differences in hypertension. *Journal of cardiovascular translational research*, 13, pp.47-54.
- A. Agresti, *An Introduction to Categorical Data Analysis*, 2nd edn, John Wiley & Sons, Inc., Hoboken, New Jersey, (2007).