pijmath
Pattimura International Journal of Mathematics (PIJMath)
Volume 02 issue 01, pages 33-42, May 2023
E-ISSN: 2830-6791
RESEARCH ARTICLE ◙ OPEN ACCESS
DOI: https://doi.org/10.30598/pijmathvol2iss1pp33-42

Ordinal Logistic Regression Analysis of Factors Affecting the Blood Sugar Levels of Diabetes Mellitus Patients

Mayawi [1], **Nurhayati** [2], **Taufan Talib** [3*], **Ariestha W Bustan** [4], **Novita S Laamena** [5]

[1] Computer Engineering Study Program, Wiratama Science and Technology Polytechnic of North Maluku
[2] Mathematics Education Study Program, Almuslim University
[3] Mathematics Education Study Program, Pattimura University
[4] Mathematics Study Program, Morotai Pacific University
[5] Statistics Study Program, Pattimura University
Jl. Ir. M. Putuhena, Ambon, 97233, Indonesia

Corresponding author's e-mail: * taufan.talib@gmail.com

ABSTRACT

Article History

Received : 7th April 2023
Accepted: 29th April 2023
Published: 30th April 2023

Keywords
Diabetes Mellitus;
Blood Sugar Level;
Ordinal Logistic Regression;

This study aimed to analyze the effect of risk factors on blood sugar levels in patients with diabetes mellitus using ordinal logistic regression analysis. Risk factors used as independent variables are age, gender, Body Mass Index (BMI), blood pressure, Cholesterol Level (TC), Low-Density Lipoprotein (LDL), High-Density Lipoprotein (HDL), Thyrocalcitonin Hormone (TCH) and Loss Triglyceride (LTG). The data used in this study were obtained from https://hastie.su.domains/Papers/LARS/diabetes.data. The number of samples taken was 100 respondents who had been diagnosed with diabetes mellitus. The results showed that risk factors such as age, Body Mass Index (BMI), Cholesterol Level (TC), Low-Density Lipoprotein (LDL), High-Density Lipoprotein (HDL), and serum Thyrocalcitonin Hormone (TCH) type had a significant effect on blood sugar levels in patients with diabetes mellitus. The best logit model for ordinal logistic regression is Logit 1, namely $\boldsymbol{g(x_1) = -2.721 - 0.079\,X_1 + 2.813\,X_3 + 0.100\,X_5 - 0.099\,X_6 - 0.119\,X_7 - 0.989\,X_8}$ *and Logit 2 is* $\boldsymbol{g(x_2) = -8.571 - 0.079\,X_1 + 2.813\,X_3 + 0.100\,X_5 - 0.099\,X_6 - 0.119\,X_7 - 0.989\,X_8}$. *It is concluded that ordinal logistic regression analysis can be used to identify factors that influence blood sugar levels in patients with diabetes mellitus and help develop more effective management and intervention strategies.*

[1] ***How to cite this article:***
Mayawi, Nurhayati, T. Talib, A. W. Bustan, and N. S. Laamena, "ORDINAL LOGISTIC REGRESSION ANALYSIS OF FACTORS AFFECTING THE BLOOD SUGAR LEVELS OF DIABETES MELLITUS PATIENTS", *Pattimura Int. J. Math. (PIJMATH)*, vol. 02, iss. 01, pp. 33-42, May 2023.

1. Introduction

Diabetes mellitus (DM) is a chronic disease affecting millions of people's health worldwide. One health problem that often occurs in people with DM is an uncontrolled increase in blood sugar levels. Uncontrolled blood sugar levels can cause serious complications, such as organ damage and death [1].

The International Diabetes Federation (IFD) organization [2] estimates that the number of people with diabetes will continue to increase and reach more than 700 million by 2045. It shows the importance of handling and controlling diabetes mellitus as a global health problem. In Indonesia, people with diabetes mellitus are increasing from year to year. According to data from the Ministry of Health [3], the number of people with diabetes in Indonesia reached more than 10.6 million in 2020. It indicates that diabetes mellitus is a health problem that needs serious attention from the government, medical personnel, and the wider community.

Diabetes mellitus also has risk factors or precipitating factors that contribute to the incidence of the disease. Risk factors that are considered to affect people with diabetes mellitus are age, gender, body mass index, blood pressure, glucose levels, and cholesterol levels. Age is a major risk factor in the occurrence of diabetes mellitus. As age increases, the risk of developing diabetes mellitus increases. In addition, gender also affects the occurrence of diabetes mellitus. Women have a higher risk than men of developing diabetes mellitus. Body mass index (BMI) also affects the occurrence of diabetes mellitus. A person with a high BMI or overweight has a higher risk of developing diabetes mellitus. High blood pressure or hypertension also contributes to diabetes mellitus [4].

High blood glucose and cholesterol levels are also important risk factors for diabetes mellitus. High glucose levels in the blood can damage blood vessels and nerves and increase the risk of organ complications. High cholesterol levels in the blood can also increase the risk of heart disease and stroke in people with diabetes mellitus [4]. These factors can be used as independent variables in ordinal logistic regression analysis to analyze their effect on the dependent variable, namely blood sugar levels in diabetes mellitus [5].

In this study, factors that are thought to affect blood sugar levels in patients with diabetes mellitus were analyzed using the ordinal logistic regression method. The purpose of this study was to obtain clearer information about the factors affecting blood sugar levels in patients with diabetes mellitus and provide useful information for the development of interventions to control blood sugar levels in patients with diabetes mellitus.

1.1 Ordinal Logistic Regression

Ordinal logistic regression is an analytical technique used to analyze the relationship between ordinal or polychotomous dependent variables and one or more independent variables. This technique uses a logistic regression model to predict the possibility of an event occurring in each category of the dependent variable. The ordinal logistic regression model has the same principle as the binary logistic regression model, which uses a logit function to relate the dependent variable to the independent variable. However, in ordinal logistic regression, the dependent variable has more than two categories, so the logit function is applied to each category of the dependent variable. This study uses ordinal logistic regression with a dependent variable that has an ordinal scale with three categories that can provide more detailed information about the relationship between the independent variable and the dependent variable compared to binary logistic regression, which can only predict the occurrence or non-occurrence of an event [6] [7].

In trichotomous logistic regression, the dependent variable with an ordinal scale with three categories can be coded into 0, 1, and 2 to simplify the analysis. However, before starting the analysis, it is necessary to determine which outcome category is used as the reference category to compare other outcome categories. Once the reference category is determined, the dependent variable Y in trichotomous logistic regression is parameterized into two logit functions. Therefore, in forming the logit function, Y = 1 and Y = 2 are compared against Y = 0. The logistic regression model for the dependent variable Y with p predictor variables can be written as follows:

$$\pi(x) = \frac{e^{(\beta_0+\beta_i x_1+\cdots+\beta_p x_p)}}{1+e^{(\beta_0+\beta_i x_1+\cdots+\beta_p x_p)}} \quad (1)$$

By using logit transformation, two logit functions were obtained. The first logit function $g_1(x)$ is the transformation of the ratio between the probability of occurrence in the outcome category Y = 1 and the probability of occurrence in the comparison outcome category Y = 0. This logit function can be calculated as follows:

$$g_1(x) = \ln\left(\frac{P(Y=1|x)}{P(Y=0|x)}\right) = \beta_{10} + \beta_{11}x_1 + \cdots + \beta_{1p}x_p = x'\beta_1 \quad (2)$$

The second logit function $g_2(x)$ is the transformation of the ratio between the probability of occurrence in the outcome category Y = 2 and the probability of occurrence in the comparison outcome category Y = 0. This logit function can be calculated as follows:

$$g_2(x) = \ln\left(\frac{P(Y=2|x)}{P(Y=0|x)}\right) = \beta_{20} + \beta_{22}x_1 + \cdots + \beta_{2p}x_p = x'\beta_2 \quad (3)$$

Based on the two logit functions, the trichotomous logistic regression model can be written as follows:

$$P(Y = 0|x) = \frac{1}{1+e^{g_1(x)}+e^{g_2(x)}} \tag{4}$$

$$P(Y = 1|x) = \frac{e^{g_1(x)}}{1+e^{\beta_1(x)}+e^{\beta_2(x)}} \tag{5}$$

and

$$P(Y = 2|x) = \frac{e^{g_2(x)}}{1+e^{\beta_1(x)}+e^{\beta_2(x)}} \tag{6}$$

Thus, the trichotomous logistic regression model can predict the probability of occurrence in each outcome category based on the given values of the predictor variables.

Following the rules of the binary logistic model, it was assumed that $P(Y = j|x) = \pi_j(x)$ for $j = 0, 1, 2$ for each function of the vector $2(p + 1)$ with parameters $\beta^T = (\beta_1^T \beta_2^T)$. The general statement for the conditional probability in the three-category model is:

$$P(Y = j|x) = \frac{e^{g_j(x)}}{\sum_{k=0}^{2} e^{g_k(x)}}, j = 0,1,2 \tag{7}$$

with vector $\beta 0 = 0$ so that $g0(x) = 0$.

In **Equation (7)**, one of the categories (in this case, category 0) was used as a reference or comparison base to measure the effect of the other categories on the dependent variable. In trichotomous ordinal logistic regression, category 0 was used as the reference category and as the basis for comparing the effect of category one and category two on the dependent variable. Therefore, in the given trichotomous logistic regression model equation, the value of $\beta_0^{(1)}$ was set as 0 for the reference category (i.e., Y = 0), and $\beta_1^{(1)}$ and $\beta_2^{(1)}$ denoted the effect of the predictor variables on the respective logit functions (i.e., $g_1(x)$ and $g_2(x)$) compared to the reference category.

1.2 Parameter Estimation

[8] suggests that model parameter testing is carried out to examine the role of independent variables in the model. Parameter testing was done partially and overall. The partial test aimed to determine the effect of each independent variable on the dependent variable separately. In contrast, the *overall* test was used to evaluate the effect of the independent variables on the dependent variable and the feasibility of the regression model.

In a logistic regression model, the conditional *likelihood function* for a sample of *n* observations can be expressed as the product of the probabilities of each individual in the sample, i.e.:

$$l(\beta) = \prod_{i=1}^{n} [\pi_0(x_i)^{y_{0i}} \pi_1(x_i)^{y_{1i}} \pi_0(x_i)^{y_{0i}}] \tag{8}$$

The log-likelihood function of the logistic regression model for a sample of *n* observations with predictor variable *x* and response variable *y* can be written as follows:

$$L(\beta) = \sum_{i=1}^{n} y_i g_1(x_i) + y_{2i} g_2(x_i) - \ln(1 + e^{g_1(x_i)} + e^{g_2(x_i)}) \tag{9}$$

To get the value β that maximizes $L(\beta)$, differentiation was performed on $L(\beta)$, with the condition that $\frac{\partial L}{\partial \beta} = 0$ and $\frac{\partial^2 L}{\partial^2 \beta} < 0$.

1.3 Model goodness-of-fit test

It is necessary to conduct a model fit test to find out whether the model with the dependent variable is a suitable model. *Akaike Information Criterion* (AIC) and *Bayesian Information Criterion* (BIC) are information criteria measurement methods used to evaluate the goodness of the model. Both methods consider the likelihood value, the number of parameters used in the model, and the data obtained. The lower the AIC or BIC value, the better the model built [9] [10] [11] [12].

The goodness of fit test in logistic regression was conducted using the Pearson test statistic, which was used to test the difference between the distribution of observations and model predictions. The hypothesis tested in the goodness of fit test is as follows:

H_0 : The logistic regression model is appropriate (there is no significant difference between the observations and the model predictions).

H_1 : The logistic regression model does not fit (there is a significant difference between the observations and the model predictions).

Suppose the Pearson test value exceeds the critical value at the specified significance level. In that case, the null hypothesis is rejected, so it can be concluded that the model is unsuitable. Conversely, if the Pearson test value is smaller than the critical value, the null hypothesis is accepted, so it can be concluded that the model is appropriate.

1.4 Coefficient of Determination

The coefficient of determination (R-Square) in logistic regression models is indicated by several metrics, including McFadden's R-Square, Cox and Snell's R-Square, and Nagelkerke's R-Square. The three metrics estimate how much variation in the blood sugar level data of people with diabetes mellitus can be explained by the regression model built. According to [13], a model is said to be good if the *Nagelkerke* coefficient is more than 70%, which means that the independent variable created by the model affects 70% of the dependent variable.

1.5 *Odd Ratio*

According to [8], the Odds Ratio (OR) measures the strength of association between dependent and independent variables in a logistic regression model. Alternatively, it indicates how much of a risk or protective factor an independent variable provides to the dependent variable. An OR value > 1 indicates that the probability of the event measured by the dependent variable increases when the independent variable increases. Conversely, an OR value < 1 indicates that the likelihood of the event decreases as the independent variable increases.

2. Research Methods

The method used was ordinal logistic regression analysis. This method allows to evaluate of the relationship between an ordinal dependent variable (in this case, blood sugar level) and one or more independent variables (e.g., age, gender, Body Mass Index (BMI), blood pressure, Cholesterol Level (TC), Low-Density Lipoprotein (LDL), High-Density Lipoprotein (HDL), Thyrocalcitonin Hormone (TCH) and Loss Triglyceride (LTG)) [4] [14]. Independent and dependent variables can be seen in **Table 1**.

Table 1. Independent and Dependent Variables

Variables	Variable Name	Measurement Scale	Description
Dependent	Blood Sugar Level (Y)	Ordinal	1 = Low (<100 mg/dl) 2 = Normal (100-140 mg/dl) 3 = High (>140 mg/dl)
Independent	Age $(X)_1$	Ratio Scale	-
	Gender $(X)_2$	Ordinal	1 = Male 2 = Female
	Body Mass Index $(X)_3$	Ordinal	1 = Skinny (<18.5) 2 = Ideal (18.5 - 24.9) 3 = Fat (>24.9)
	Blood Pressure $(X)_4$	Ordinal	1 = Low (<100 mmHg) 2 = Normal (100-120 mmHg) 3 = High (>120 mmHg)
	Cholesterol Level $(X)_5$	Ratio Scale	-
	Low Density Lipoprotein $(X)_6$	Ratio Scale	-
	High Density Lipoprotein $(X)_7$	Ratio Scale	-
	Thyrocalcitonin Hormone $(X)_8$	Ratio Scale	-
	Triglyceride Loss $(X)_9$	Ratio Scale	-

Data were obtained from Stanford University research on patients with diabetes mellitus (https://hastie.su.domains/Papers/LARS/diabetes.data). In this study, the data were grouped according to the specified categories by taking a sample of 100 patients to be analyzed using ordinal logistic regression.

According to [15], the steps of testing multinomial logistic regression analysis with ordinal data-dependent variables are as follows:

a. The dependent variable is categorized into 1, 2, and 3, namely low, normal, and high blood sugar levels.
b. Conduct a Multicollinearity Test to determine whether it contains multicollinearity by looking at the amount of intercorrelation among independent variables. It can be seen from the amount of Tolerance Value and Variance Inflation Factor (VIF) by looking at the Tolerance Value ≥ 0.10 or the same as the VIF value ≤ 10.
c. Conduct a *likelihood ratio* or simultaneous test to test the entire model using all independent variables. This simultaneous test aims to determine whether the independent variable significantly affects the dependent variable

as a whole. From the equation $g(x_i) = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik}$ it is obtained the hypothesis to be tested as follows:

$H_0: \beta_0 = \beta_1 = \cdots = \beta_0 = 0$, which means there is no significant influence between the independent and dependent variables simultaneously.

$H_1: \exists \beta_j \neq 0$ means that at least one independent variable significantly affects the model.

If H_0 is rejected, then at least one independent variable significantly affects the dependent variable.

d. Conduct a model parameter test using a partial test (Wald test) to test each independent variable on the dependent variable. This partial test aimed to determine the role of each independent variable in the model individually. The hypothesis used is:

$H_0: \beta_j = 0$ means no influence between the jth independent and dependent variables.

$H_1: \exists \beta_j \neq 0$ means an influence between the jth independent variable and the dependent variable.

If H_0 is rejected, there is a significant influence between the jth independent and dependent variables.

e. The logistic regression model goodness test is carried out to test whether the resulting model is feasible. The hypothesis used is:

H_0 : There is no significant difference between the observations and the possible results or predictions of the model (model fit)

H_1 : There is a significant difference between the observations and the possible results or predictions of the model (the model does not fit).

If H_0 is accepted, then the model is appropriate.

f. The odd ratio is a measure to determine the risk of the tendency of one category to another.

g. Opportunity model of logistic regression equation

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik})}{1 + \exp(\beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik})} \quad (10)$$

With the logit transformation model for the model:

$$g(x) = \frac{[\ln \pi(x)]}{[1 - \pi(x)]} = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} \quad (11)$$

3. Results And Discussion

3.1 Multicollinearity Test

The results of the multicollinearity test can be seen in **Table 2**.

Table 2. Multicollinearity Testing

Coefficients								
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	2.859	1.045		2.737	.008		
	Age	7.216E-5	.004	.002	.019	.985	.589	1.697
	Gender	-.013	.083	-.010	-.157	.876	.838	1.193
	IMT	.093	.142	.075	.658	.512	.283	3.531
	BP	.015	.066	.015	.230	.819	.803	1.246
	TC	.000	.002	-.019	-.183	.855	.336	2.977
	LDL	.000	.002	.019	.180	.858	.336	2.974
	HDL	.002	.004	.034	.422	.674	.568	1.762
	TCH	.009	.048	.016	.181	.857	.464	2.154
	LTG	-.007	.099	-.005	-.068	.946	.697	1.434
a. Dependent Variable: Glu								

To identify the presence of multicollinearity, the Tolerance value and Variance Inflation Factor (VIF) of each independent variable can be tested. The tolerance value is the opposite of VIF. It is used to evaluate how much variation in one independent variable cannot be explained by other independent variables in the model. A low Tolerance value (<0.1) and a high VIF (>10) indicate multicollinearity. The test results in **Table 2** shows that all independent variables in the study have a Tolerance value ≥ 0.10 and a VIF value ≤ 10. It can be concluded that there is no multicollinearity in the regression model.

3.2 Parameter Estimation

The results of parameter estimation using the *Maximum Likelihood Estimation* method are presented in **Table 3**.

Table 3. Parameter Estimation Results

Parameter Estimates						
		Estimate	**Std. Error**	**Wald**	**df**	**Sig.**
Threshold	[Glu=1]	2.721	4.948	.303	1	.582
	[Glu=2]	8.571	5.140	2.781	1	.095
Location	Age	.079	.027	8.522	1	.004
	TC	-.100	.025	15.537	1	.000
	LDL	.099	.028	12.886	1	.000
	HDL	.119	.040	8.878	1	.003
	TCH	.989	.435	5.179	1	.023
	LTG	1.021	.890	1.318	1	.251
	[Gender=1]	-.248	.634	.153	1	.695
	[Gender=2]	0[a]	.	.	0	.
	[IMT=2]	-2.813	.798	12.413	1	.000
	[IMT=3]	0[a]	.	.	0	.
	[BP=1]	-1.895	1.294	2.144	1	.143
	[BP=2]	-2.403	1.365	3.101	1	.078
	[BP=3]	0[a]	.	.	0	.
Link function: Logit.						
a. This parameter is set to zero because it is redundant.						

Based on **Table 3**, with a significance level of $\alpha = 0{,}05$, it is obtained that the independent variables that affect the dependent variable are variables whose Sig. Value is less than α. The influential variables are age, Body Mass Index (BMI), Cholesterol Level (TC), *Low-Density Lipoprotein (*LDL), *High-Density Lipoprotein* (HDL), and serum *Thyrocalcitonin Hormone (*TCH). Thus, logit one and logit two functions can be formed as follows:

$$g(x_1) = -2.721 - 0.079\, X_1 + 2.813\, X_3 + 0.100\, X_5 - 0.099\, X_6 - 0.119\, X_7 - 0.989\, X_8 \qquad (12)$$

$$g(x_2) = -8.571 - 0.079\, X_1 + 2.813\, X_3 + 0.100\, X_5 - 0.099\, X_6 - 0.119\, X_7 - 0.989\, X_8 \qquad (13)$$

3.3 Parameter Testing

In this parameter test, two tests were used, namely the simultaneous test or overall test (G test) and the partial test. The parameter test results are as follows.

a. Simultaneous Test

Simultaneous Test Results can be seen in **Table 4**.

Table 4. Simultaneous Test Results

Model Fitting Information				
Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	183.119			
Final	81.173	101.946	10	.000
Link function: Logit.				

Table 4 shows that the significance value is less than $\alpha = 0{,}05$, so it can be concluded that the model is significant, and the next test can be done, namely the Partial test.

b. Partial Test

A partial test is used to determine whether there is an influence of each independent variable on the dependent variable. Partial Test Results can be seen in **Table 5**.

Table 5. Partial Test Results for All Variables

Effect	Model Fitting Criteria -2 log Likelihood of Reduced Model	Likelihood Ratio Tests			
		Chi Square	df	Sig.	Sig
Intercept	71.646	11.580	2	0.003	Reject H_0
Age	64.686	4.620	2	0.099	Accept H_0
Gender	61.223	1.156	2	0.561	Accept H_0
IMT	67.942	7.875	2	0.019	Reject H_0
BP	60.983	0.916	2	0.632	Accept H_0
TC	92.770	32.704	2	0.000	Reject H_0
LDL	85.497	25.430	2	0.000	Reject H_0
HDL	69.395	9.329	2	0.009	Reject H_0
TCH	70.987	10.921	2	0.004	Reject H_0
LTG	60.107	0.040	2	0.980	Accept H_0

The Likelihood ratio test value in **Table 5** compares whether the model created with certain independent variables better explains the data than a model that only uses a constant as an independent variable. In this study, the test results show that the independent variables of BMI, TC, serum type of LDL, HDL, and serum type of TCH have a significance of less than α=0.05, which means that these variables have a significant influence in forming a logistic regression model and are better used than using only constants as independent variables.

The partial test of the five influential factors was carried out again, and the results were obtained as in **Table 6**.

Table 6. Partial Test Results for Affected Variables

Effect	Model Fitting Criteria -2 log Likelihood of Reduced Model	Likelihood Ratio Tests			
		Chi-Square	df	Sig.	Conclusion
Intercept	100.146	21.995	2	0.000	Reject H_0
IMT	92.704	14.552	2	0.001	Reject H_0
TC	115.975	37.824	2	0.000	Reject H_0
LDL	111.479	33.327	2	0.000	Reject H_0
HDL	92.501	14.349	2	0.001	Reject H_0
TCH	88.830	10.679	2	0.000	Reject H_0

It shows that all the independent variables in the logistic regression model affect the dependent variable significantly, and no variable can be removed without reducing the model prediction quality. Therefore, using all independent variables in the model is still important and relevant to form the best model.

3.4 Model goodness-of-fit test

To find out whether all independent variables have an influence on the dependent variable, it is necessary to test the *goodness of fit* with the Pearson Test, and the results can be seen in **Table 7**.

Table 7. Model goodness-of-fit test results

Goodness-of-Fit

	Chi-Square	df	Sig.
Pearson	102.140	188	1.000
Deviance	81.173	188	1.000

Link function: Logit.

Based on **Table 7**, it is known that the *p-value* of *Pearson* is greater than 1.000 $\alpha = 0{,}05$, so H_0 is accepted, and the model is feasible to use, or the model is suitable.

3.5 Coefficient of Determination

The coefficient of determination can be determined from the Mc Fadden, Cox and Snell, and Nagelkerke values as in **Table 8**.

Table 8. Results of the Coefficient of Determination

Pseudo R-Square

Cox and Snell	.639
Nagelkerke	.761
McFadden	.557

Link function: Logit.

Based on **Table 8**, the Nagelkerke coefficient of determination of 0.761 indicates that the logistic regression model built can explain variations in the dependent variable (blood sugar levels) by 76.1%. It indicates that the independent variables used in the model significantly influence blood sugar levels in patients with diabetes mellitus. The factors that have a significant effect are tested again, and the results are shown in **Table 9**.

Table 9. Results of the Coefficient of Determination Test for Influential Variables

Pseudo R-Square	
Cox and Snell	0.650
Nagelkerke	0.774
McFadden	0.573

Link function: Logit.

From **Table 9**, the Nagelkerke R-Square value obtained shows that the regression model built can explain 77.4% of the variation in the dependent variable with the independent variables used. The Cox and Snell R-Square and McFadden R-Square values also show good model performance, although with slightly different values. Thus, the regression model built in this study is good enough to explain the relationship between the independent variable and the dependent variable. Based on **Table 8** and **Table 9**, using all variables provided a greater coefficient of determination than using only the influential variables.

3.6 Modeling Blood Sugar Levels with Ordinal Logistic Regression

After conducting these tests, the best logit model of ordinal logistic regression in diabetes mellitus cases was formed. **Logit 1** is the log of the ratio between the chances of low and high blood sugar levels in diabetes mellitus. Similarly, **Logit 2** compares the chances of normal and high blood sugar levels in diabetes mellitus. Therefore, the following interpretation is obtained.

a. The regression output results show that the regression coefficient for the age variable is 0.079, with a significance level of $p < 0.05$. Each additional year of age will increase blood sugar levels by 0.079 mmHg if the other variables (BMI, TC, LDL, HDL, and TCH) are fixed.

b. Every kilogram increase in Body Mass Index (BMI) will increase blood sugar levels by 0.697 mmHg if age, Cholesterol Level (TC), Low-Density Lipoprotein (LDL), High-Density Lipoprotein (HDL), and serum Thyrocalcitonin Hormone (TCH) are fixed.

c. Every additional one mmol/L of Cholesterol Level (TC) will reduce blood sugar levels by 0.100 mmol/L if age, Body Mass Index (BMI), Low-Density Lipoprotein (LDL), High-Density Lipoprotein (HDL), and serum Thyrocalcitonin Hormone (TCH) types are fixed.

d. Every additional one mmol/L of Low-Density Lipoprotein (LDL) will increase blood sugar levels by 0.099 mmHg if age, Body Mass Index (BMI), Cholesterol Level (TC), High-Density Lipoprotein (HDL), and serum Thyrocalcitonin Hormone (TCH) levels remain constant.

e. Every additional one mmol/L of High-Density Lipoprotein (HDL) will increase blood sugar levels by 0.119 mmHg if age, Body Mass Index (BMI), Cholesterol Level (TC), Low-Density Lipoprotein (LDL), and serum Thyrocalcitonin Hormone (TCH) are fixed.

f. Every additional one mmol/L of Thyrocalcitonin Hormone (TCH) will increase blood sugar levels by 0.989 mmHg if age, Body Mass Index (BMI), Cholesterol Level (TC), Low-Density Lipoprotein (LDL), and High-Density Lipoprotein (HDL) are fixed.

g. If age, Body Mass Index (BMI), Cholesterol Level (TC), Low-Density Lipoprotein (LDL), High-Density Lipoprotein (HDL), and Thyrocalcitonin Hormone (TCH) are equal to 0, then a low blood sugar level of 2,271 and a normal blood sugar level of 8,571 remain.

3.7 Model Interpretation

If the model has been tested and the model results are good, and the significance is real, the data can be interpreted using the *odds ratio* test, as shown in **Table 10** below.

Table 10. Odds Ratio Test Results

Glucose Level	Variables	Sig	Odds Ratio = Exp(β)
Low	IMT	0.001	0.183
	TC	0.019	1.330
	LDL	0.012	0.731
	HDL	0.155	0.783
	TCH	0.410	0.162
Normal	IMT	0.001	0.026
	TC	0.001	1.023
	LDL	0.005	0.911
	HDL	0.005	0.809
	TCH	0.007	0.092

The following is the interpretation results based on the **Table 10**, above.

The odd ratio measures the association between two events, such as blood sugar levels and other risk factors (BMI, TC, LDL, HDL, TCH). The odd ratio is the ratio of the odds of a patient developing diabetes mellitus with certain risk factors compared to the odds of a patient not developing diabetes mellitus with the same risk factors.

For BMI, the odds ratio shows that as body weight increases in patients with low blood sugar levels, the chance of having diabetes mellitus is 0.183 times smaller than in those with high blood sugar levels. Whereas in people with normal blood sugar levels, the chance of someone suffering from diabetes mellitus is smaller by 0.026 times than in people with high blood sugar levels. It shows that BMI has an inverse relationship with the risk of developing diabetes mellitus.

For TC, the odds ratio shows that the higher the TC level in people with low blood sugar levels, the greater the chance of someone suffering from diabetes mellitus by 1.330 times compared to people with high blood sugar levels. Whereas in people with normal blood sugar levels, the chance of someone suffering from diabetes mellitus is greater by 1.023 times than in people with high blood sugar levels. It shows that TC has a positive relationship with the risk of developing diabetes mellitus.

For LDL and HDL, the odds ratio shows that the higher the LDL and HDL levels in people with low blood sugar levels, the less likely a person has diabetes mellitus than people with high blood sugar levels. Likewise, in patients with normal blood sugar levels, the chances of suffering from diabetes mellitus are smaller than in those with high blood sugar levels. It shows that LDL and HDL have an inverse relationship with the risk of developing diabetes mellitus.

For TCH, the odds ratio shows that the higher the TCH level in patients with low blood sugar levels, the chance of someone suffering from diabetes mellitus is smaller by 0.162 times than those with high blood sugar levels. While in patients with normal blood sugar levels, the chance of someone suffering from diabetes mellitus is smaller by 0.251 times than in patients with high blood sugar levels. It shows that TCH has an inverse relationship with the risk of developing diabetes mellitus.

4. Conclusions

Based on the above discussion, the following conclusions can be drawn:

a. Ordinal logistic regression analysis is a type of non-linear regression used not only for models with dependent variables in the form of ordinal-scale data.

b. The influential independent variables are age, Body Mass Index (BMI), Cholesterol Level (TC), Low-Density Lipoprotein (LDL), High-Density Lipoprotein (HDL), and serum Thyrocalcitonin Hormone (TCH) type.

c. The best logit model for ordinal logistic regression is
Logit 1
$$g(x_1) = 2.721 + 0.079\,X_1 - 2.813\,X_3 - 0.100\,X_5 + 0.099\,X_6 + 0.119\,X_7 + 0.989\,X_8$$
Logit 2
$$g(x_2) = 8.571 + 0.079\,X_1 - 2.813\,X_3 - 0.100\,X_5 + 0.099\,X_6 + 0.119\,X_7 + 0.989\,X_8$$

References

[1] World Health Organization, "Global report on diabetes," World Health Organization, 2016. [Online]. Available: https://www.who.int/publications/i/item/9789241565257

[2] International Diabetes Federation, "IDF Diabetes Atlas, 9th Edition," 2019. [Online]. Available: https://www.diabetesatlas.org.

[3] Ministry of Health, "Situasi dan analisis diabetes"["Diabetes situation and analysis"], Data and Information Center, Ministry of Health, 2020. [Online]. Available: https://pusdatin.kemkes.go.id/article/view/20111700001/situasi-dan-analisis-diabetes.html. [Accessed: March. 5, 2023].

[4] American Diabetes Association, "Risk Factors for Type 2 Diabetes," 2021. [Online]. Available: https://www.diabetes.org/diabetes-risk.

[5] A. Misnadiarly, "Faktor-faktor yang berhubungan dengan Kejadian Diabetes Mellitus Tipe 2"["Factors associated with the Incidence of Type 2 Diabetes Mellitus"] Journal of Public Health, vol. 1, no. 1, pp. 38-45, 2006. doi: 10.20961/jkm.v1i1.12158.

[6] R. Williams, "Generalized ordered logit/partial proportional odds models for ordinal dependent variables," Stata Journal, vol. 6, no. 1, pp. 58-82, 2006.

[7] A. Yudisasanta and M. Ratna, "Analisis Faktor-Faktor yang Mempengaruhi Kepuasan Pasien BPJS Kesehatan di RSUD Majalengka Menggunakan Regresi Logistik Ordinal" ["Analysis of Factors Affecting BPJS Health Patient Satisfaction at

Majalengka Regional Hospital Using Ordinal Logistic Regression"], Journal of Economics and Business Education, vol. 2, no. 2, pp. 53-61, 2012.

[8] D.W. Hosmer and S. Lemeshow, Applied Logistic Regression. John Wiley & Sons Inc, New York, 1989.

[9] H. Akaike, "A new look at the statistical model identification," IEEE Transactions on Automatic Control, vol. 19, no. 6, pp. 716-723, 1974. doi 10.1109/TAC.1974.1100705.

[10] I. M. Tirta, Analisis Regresi dengan R [Regression Analysis with R]. Jember: University of Jember, 2009.

[11] G. Schwarz, "Estimating the dimension of a model," The Annals of Statistics, vol. 6, no. 2, pp. 461-464, 1978. doi: 10.1214/aos/1176344136.

[12] H. Pardede, "Penerapan Metode Regresi Logistik Ordinal dalam Identifikasi Faktor Risiko yang Berhubungan dengan Kejadian Stunting pada Anak Balita di Indonesia" ["Application of Ordinal Logistic Regression Method in Identifying Risk Factors Associated with the Incidence of Stunting in Toddlers in Indonesia"] Journal of Mathematics, Statistics and Computing, vol. 9, no. 1, pp. 24-33, 2013.

[13] R. Rizki, "Analisis Regresi Logistik Dalam Menganalisis Faktor-Faktor Yang Berhubungan Dengan Kecenderungan Anak Mengalami Stunting" ["Logistic Regression Analysis in Analyzing Factors Associated with the Tendency of Children to Experience Stunting"] Journal of Public Health Sciences, vol. 7, no. 1, pp. 52-60, 2016.

[14] R. H. B. Christensen, "Ordinal Regression Models for Ordinal Data," R package version 2019.12-10, 2019. [Online]. Available: https://cran.r-project.org/web/packages/ordinal/vignettes/clm_article.pdf.

[15] F. Nikie, "Analisis Regresi Logistik Ordinal pada Faktor-Faktor yang Mempengaruhi Kadar Gula Darah Penderita Diabetes Mellitus" ["Ordinal Logistic Regression Analysis on Factors Affecting Blood Sugar Levels of Diabetes Mellitus Patients"] Thesis, Gadjah Mada University, Yogyakarta, 2016.