International Journal of Applied Mathematics and Computing
Vol. 1, No.3 October 2024
e-ISSN: 3047-146X; Page 31-40
DOI ; https://doi.org/10.62951/ijamc.v1i3.27
Available online at: https://international.arimsi.or.id/index.php/IJAMC

Factors That Influence Diabetes Disease
(Case Study: Pima Indians)

Ni Made Deviani Prisilia[1*], Adelia Yuniarti[2], Citra Annisa Rahmania[3], Made Ayu Asri Oktarini Putri[4], Made Susilawati[5]
[1-5]study Program Mathematics , Faculty Of Mathematics And Natural Sciences,Udayana University,Indonesia
deviani.prisilia@gmail.com[1] ,adeliays11@gmail.com[2], annisacitra934@gmail.com[3], ayuoktarini29@gmail.com[4], mdsusilawati@unud.ac.id[5]*

Corresponding Author: deviani.prisilia@gmail.com[*]

Abstract.*Diabetes is one of the non-communicable diseases that is considered dangerous due to its susceptibility to complications. This disease is caused by high blood sugar levels in a person's body, which makes the blood more alkaline and slows down the metabolic process. In this study, we observed 8 variables that are considered influential in diabetes and will build a regression model that can predict the response variable (y) through Logistic Regression Analysis. Logistic Regression Analysis is a statistical analysis method used to describe the relationship between a dependent variable with two or more categories and one or more independent variables that are categorical or continuous. Based on the results, the logistic regression model for factors influencing diabetes in the Indian Pima tribe includes variables such as number of pregnancies, glucose level, blood pressure, body mass index, and diabetes pedigree function.*

Keywords *: Regression logistics , Diabetes, Likelihood ratio test, Wald , Hosmer and Lameshow*

1. INTRODUCTION

Diabetes is a non-communicable disease that has a high risk of various health complications . This disease is the result of increased blood glucose levels in a person's body so that the blood becomes more alkaline and slows down the metabolic process. The death rate caused by this disease is increasing and is predicted to continue to increase every year. The increasing number of diabetes sufferers significantly contributes to the increase in the death rate from this disease. *The International Diabetes Federation* (IDF) projects an increase of 227 million people with diabetes from 415 million in 2015 to 642 million in the next 25 years.

Diabetes consists of several types, including type 1 diabetes which generally appears in children and requires insulin injections, type 2 diabetes which occurs more often in adults, and gestational diabetes which is experienced by pregnant women.

Many factors influence a person to become a diabetes sufferer. In this study, we observed 8 independent variables. namely , the number of pregnancies during life, glucose, blood pressure, skin thickness, body mass index, insulin levels, age, and diabetes lineage function which are considered to have an influence on diabetes and a regression model will be formed which can predict the response variable (y) through Logistic Regression Analysis.

Received: September 13,2024; Revised: September 27,2024; Accepted: September 30,2024; Online Available: October 02, 2024

Analysis is a statistical technique used to model the relationship between a categorical dependent variable and one or more independent variables, either categorical or continuous

2. RESEARCH METHODS

Data Sources

On study This using secondary data obtained from *National Institute of Diabetes and Digestive and Kidney Diseases* on the number of patients with or without diabetes who were of Pima Indian descent (a subgroup of Native Americans) and women aged at least 21 years. The study This using Pima Indian data because ethnic group This own level the highest prevalence of diabetes in the world. By Because that , tribe the Lots become subject study For understand diabetes disease . Many diabetes factors found on ethnic group this is wrong the only one is factor *Beta3-Adrenergic Receptor* (ADRB3) so Trp64Arg missense mutation in this gene Lots found on Pima Indians and associate with obesity , so that allegedly can increase opportunity obesity (Indra, 2006) .

Research Variables

Variables used in study can seen in Table 1.

Table 1. Research Variables and Measurement Scales

Variables	Label	Unit	Scale Measurement	Category
Y	Results	-	Nominal	0 : No Diabetes 1: Diabetes
X_1	Amount Pregnancy During Life	-	Interval	-
X_2	Glucose	$mmol/L$	Ratio	-
X_3	Pressure Blood	$mmhg$	Ratio	-
X_4	Skin Thickness	mm	Ratio	-
X_5	Body Mass Index	kg/m^2	Ratio	-
X_6	Insulin Levels	$\mu U/mL$	Ratio	-
X_7	Age	$tahun$	Ratio	-
X_8	Function Diabetes Genealogy	-	Ratio	-

Steps in Data analysis

Data analysis in study This done with use device soft IBM SPSS *Statistics 25* statistics . The following are the steps taken:

1. Do analysis descriptive For see description general data.
2. Research data analysis done with using regression model logistics . Stage beginning analysis involving testing significance simultaneous model parameters using statistics G test .

Testing significance simultaneous model parameter analysis is carried out with use test ratio Likelihood Ratio Test or G test . This aiming For evaluate significance influence combination all over variable free to variable bound with statistics test used defined as following :

$$G = -2\ln\left[\frac{L_0}{L_1}\right]$$

WhereL_0 is *likelihood* of the model only consists of from constant andL_1 is *likelihood* of the model it consists of from overall variables . Statistics G test follows distribution *chi-square* with the decision obtained from do comparison with mark $x^2 tabel$.

Do test significance of parameters in partial with Wald test .

Test significance of parameters in partial done with use Wald test . Test This used For know worthy or whether or not a variable free For enter to in the model. Statistics test used defined as following :

$$W = \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)}$$

$$SE(\hat{\beta}_i) = \sqrt{(\sigma^2(\hat{\beta}_i))}$$

Where $SE(\hat{\beta}_i)$is the estimated standard error for the coefficient β_i and $\hat{\beta}_i$is the estimated value for the parameter (β_i).

Testing the model's suitability (*Goodness of Fit*) using the test *Hosmer and Lameshow* .

Test Goodness *of Fit* is carried out with use test *Hosmer and Lameshow* . Test This used For know conformity between the model and the data. Test This see what is the regression model logistics obtained worthy For used . Statistics test used defined as following :

$$\hat{C} = \sum_{k=1}^{g} \frac{(O_k - n'_k\bar{\pi}_k)^2}{n'_k\bar{\pi}_k(1-\bar{\pi}_k)}$$

Statistics test *Hosmer and Lameshow* follow distribution *chi-square* with degrees free $g-1$where the decision is obtained from do comparison with mark $x^2 tabel$.

1. Forming a regression model logistics .
2. Calculating the accuracy of classification of diabetes incidence data using the *Apparent Error Rate* ($APER$).
3. Interpreting regression models logistics .

3. RESULTS AND DISCUSSION

Descriptive Analysis

Statistics descriptive from patient data Woman Pima Indians must be at least 21 years old shown on Table 2.

Table 2. Descriptive Statistics of Pima Indian Female Patient Data

	Frequency	Percentage
No Diabetes	500	65,1
Diabetes	268	34.9
Total	768	100

Source : Processed data , 2023

Based on Table 2 shows the data obtained as much as 768data from those $65,1\%$without diabetes and $34,9\%$those with diabetes.

Logistic Regression

In this step, several tests were carried out on data from female Pima Indian patients using the logistic regression method.

Simultaneous Parameter Significance Test

The simultaneous parameter significance test conducted is the G test. This test is used to determine the overall influence of independent variables on dependent variables. The G test statistic follows the distribution χ^2. The hypotheses used are:

$H_0 = \beta_i = 0$, with $i = 1,2,\cdots,8$(no significant influence of the independent variables $i$simultaneously on diabetes)

H_1 =there is at least one $\beta_i \neq 0$, with $i = 1,2,\cdots,8$(there is a significant influence of the independent variable $i$simultaneously on diabetes).

Results test significance of parameters in simultaneous displayed on Table 3. below .

Table 3. Results of Simultaneous Parameter Significance Test

Step	$-2\ Log\ likelihood$	*Nails R Square*
1	723,445	0,408

Source : Processed data , 2023

Based on Table 3. shows that the G test value $= 723{,}445 \geq \chi^2_{0{,}05{,}8} = 15{,}507$is H_0rejected, which means that there is at least one influence of independent variables simultaneously on diabetes. In addition, it can be seen that the determination coefficient of logistic regression is as large as $0{,}408$which means that the influence of independent variables on the dependent variable is $40{,}8\%$.

Partial Parameter Significance Test

The partial parameter significance test conducted is the Wald test. This test is used to determine whether or not an independent variable is suitable to enter the model. The hypotheses used are:

- $H_0 = \beta_i = 0$, with $i = 1{,}2, \cdots, 8$(no significant influence of the independent variable $i$on diabetes)
- $H_0 = \beta_i \neq 0$, with $i = 1{,}2, \cdots, 8$(there is a significant influence of the independent variable $i$on diabetes).

Results test significance of parameters in partial displayed on Table 4. below.

Table 4. Results of Partial Parameter Significance Test

Variables	Wald	Sig.
X_1	14,747	0,000
X_2	89,897	0,000
X_3	6,454	0,011
X_4	0,008	0,929
X_5	1,749	0,186
X_6	35,347	0,000
X_7	9,983	0,002
X_8	2,537	0,111
Constants	137,546	0,000

Source : Processed data , 2023

Based on Table 4. shows that the Wald test value $X_1, X_2, X_3, X_6, X_7$is greater than the value $\chi^2_{0{,}05{,}1}$ $(3{,}841)$ so that the decision $H_0$is rejected. Thus, the variables $X_1, X_2, X_3, X_6,$ andX_7 there is significant influence to diabetes disease .

Model Suitability Test (*Goodness of Fit*)

Model feasibility test (*Goodness of Fit*) was carried out For know conformity predicted value by model with mark from the data. This test looks at whether the regression model logistics obtained give accurate predictions . The test statistics used are the *Hosmer and Lameshow* test . The hypotheses used are:

- H_0= there is no difference between the model and the data so the model can be said to be suitable
- H_1= there is a difference between the model and the data so that the model can be said to be feasible.

Results test model feasibility is shown in Table 5. below.

Table 5. Results of the Model Suitability Test (*Goodness of Fit*)

Step	Chi-square	df	Sig.
1	8,323	8	0,403

Source : Processed data , 2023

Based on Table 5 shows that the values $\chi^2_{HL} = 8{,}323 < \chi^2_{0{,}05{,}6} = 12.5$and $Sig. = 0{,}403 > \alpha = 0{,}05$are H_0accepted, which means that there is no difference between the model and the data, so the model is said to be suitable.

Formation of Logistic Regression Model

Selection of logistic regression variables using the Wald coefficient which is greater thanχ^2 And mark significance not enough from αso that the independent variables that are required in forming the logistic regression model are shown in Table 6 below.

Table 6. Estimation of Logistic Regression Model Parameters

Variables	B	Exp(B)
X_1	0,123	1,131
X_2	0.035	1,036
X_3	$-0{,}013$	0,987
X_4	0,001	1,001
X_5	$-0{,}001$	0,999
X_6	0,090	1,094
X_7	0,945	2,573
X_8	0,015	1,015
Constants	$-8{,}405$	0,000

Source: Processed data, 2023

Based on Table 6. The logistic regression model formed is as follows.

$$g(x) = \ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = -8.405 + 0.123X_1 + 0.035X_2 - 0.013X_3 + 0.090X_6 + 0.945X_7 (3.1)$$

with,

X_1= number of pregnancies during life

X_2= glucose

X_3= blood pressure

X_6= body mass index

X_7= function diabetes pedigree

Accuracy of Logistic Regression Classification

Results classification of incident data diabetes disease using test *APER* is obtained in Table 7 below.

Table 7. Results of Accuracy of Classification of Diabetes Disease Incidents

Observation	Prediction		Amount
	No Diabetes	Diabetes	
No Diabetes	445	55	500
Diabetes	112	156	268
Amount	557	211	768

Source : Processed data , 2023

Based on Table 7, the level of classification error can be calculated using the *APER test*, namely:

$$APER = \frac{112 + 55}{445 + 55 + 112 + 156} = 0{,}217$$

The classification accuracy of the data is obtained as follows:

$$(1 - 0{,}217) = 0{,}783 = 78{,}3\%$$

So that the percentage of overall classification accuracy is obtained $78{,}3\%$with a classification error of $21{,}7\%$. It can be concluded that the accuracy of the classification of diabetes is quite high.

Interpretation of Logistic Regression Model

Next, the model interpretation will be carried out using *the Odds Ratio.* The interpretation of the logistic regression model in equation (3.1) is as follows:

a. Amount Pregnancy During Life

$Exp\ (\beta) = 1.131$meaning if amount pregnancy Pima Indian tribe increases One unit so will increase risk caught diabetes disease of 1.131if other factors are constant.

Research conducted by (Dabelea et al., 2008) to put forward that amount high pregnancy (four pregnancy or more) is associated with improvement risk of type 2 diabetes in women in

later day . This is because of every pregnancy generally followed with addition weight . If a woman own Lots pregnancy , addition heavy body repetitive This can increase risk obesity . By Because that , has Lots pregnancy can increase risk have diabetes.

b. Glucose

$Exp\ (\beta) = 1.036$meaning if glucose Pima Indian tribe increases One $mmol/L$ unit so will increase risk caught diabetes disease of 1.036if other factors are constant. Supported by study (Guo et al., 2023) to explain that Wrong One important aspect of managing diabetes is detecting blood glucose levels frequently and accurately. One of tool measurements that can be used For diagnosing diabetes is *fasting blood glucose* (FBG).

This matter explain that factor glucose influential significant against diabetes so that monitoring level glucose can become component important in help treatment And choose style the right life (Kong et al., 2023).

c. Blood pressure

$Exp\ (\beta) = 0.987$meaning if the pressure blood Pima Indian tribe increases One $mmhg$then it will reduce the risk caught diabetes disease of 0.987 if other factors are of value constant .

This statement is supported by research (Amira et al., 2014)that explains that increased blood pressure is a common complication in people with diabetes mellitus. This increase in blood pressure significantly increases the risk of morbidity and mortality due to various cardiovascular and microvascular complications, such as diabetic nephropathy .

d. Body Mass Index

$Exp\ (\beta) = 1.094$ meaning if index mass body Pima Indian tribe increases One kg/m^2 so will increase risk caught diabetes disease of 1.094if other factors are constant. This statement is supported by research results (Komariah & Rahayu, 2020)showing that people with type 2 diabetes mellitus have the highest body mass index $\geq 25{,}00$ (59%).

e. Diabetes Genealogy Function

$Exp\ (\beta) = 2.573$This means that if the Pima Indian 2.573diabetes genealogy function indicator increases by one unit, it will increase the risk of developing diabetes by the amount if other factors are held constant.

According to study (Zhang et al., 2018) about prevalence And risk factors for diabetes and *impaired fasting glucose* (IFG) in China East Sea explain that risk of diabetes occurs on women in the countryside And those who have family history of diabetes.

On Study this is also explained possibility somebody suffering from diabetes will more tall compared to those who don't own history of diabetes in family . This is can happen Because diabetes disease involves factor genetics And factor environment , such as nutrition obtained as well as style adapted life from member family . However , due to matter This give impact positive on level awareness And more treatment tall (Moonesinghe et al., 2018).

4. CONCLUSION AND SUGGESTIONS

Conclusion

Based on results And discussion , a regression model was obtained logistics For factors that influence diabetes disease in Pima Indians as following .

$$g(x) = \ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = -8.405 + 0.123X_1 + 0.035X_2 - 0.013X_3 + 0.090X_6 + 0.945X_7$$

Influential factors significant to diabetes disease in Pima Indians among them , the number pregnancy during life , glucose , pressure blood , index mass body , and function diabetes pedigree .

Suggestions

In study Next , the author can consider variable others , such as habit smoke And history hypertension . In addition that , it is recommended for society that has history family with diabetes for quick do examination in order to be able to done prevention And treatment since early .

5. BIBLIOGRAPHY

Amira, N., Pandelaki, K., & Palar, S. (2014). Relationship pressure blood and long time suffering from diabetes with rate glomerular filtration in subject diabetes mellitus type 2. E-Clinic (ECl), 2(1), 1–6.

Dabelea, D., Mayer-Davis, E. J., Lamichhane, A. P., D'Agostino, R. B., Liese, A. D., Vehik, K. S., Venkat Narayan, K. M., Zeitler, P., & Hamman, R. F. (2008). Association of intrauterine exposure to maternal diabetes and obesity with type 2 diabetes in youth: The SEARCH case-control study. Diabetes Care, 31(7), 1422–1426. https://doi.org/10.2337/dc07-2417

Guo, K., Li, Y., & Bian, H. (2023). Development of an electrochemical sensor for simultaneous determination of glucose and insulin: Application for accurate classification of diabetes mellitus. International Journal of Electrochemical Science, 18, 1–7. https://doi.org/10.1016/j.ijoes.2023.100212

Hana, F. M. (2020). Classification of diabetes patients using decision tree algorithm. SISKOM-KB Journal (Computer Systems and Artificial Intelligence).

Hendayana, R. (2013). Application of logistic regression methods in analyzing agricultural technology adoption. Agricultural Informatics.

Hidayati, N., Sukarsa, G., & Nilakusmawati, E. (2020). Comparison of discriminant analysis and logistic regression to classify the eligibility of Bidikmisi applicants' visitation. E-Journal of Mathematics.

Indra, M. R. (2006). Genetic basis of visceral obesity. Brawijaya Medical Journal, 21(4), 19.

Komariah, & Rahayu, S. (2020). Relationship between age, gender, and body mass index with fasting blood sugar levels in type 2 diabetes mellitus patients at the Proklamasi Outpatient Clinic, Depok, West Java. Kusuma Husada Health Journal.

Kong, A. P. S., Lim, S., Yoo, S. H., Ji, L., Chen, L., Bao, Y., Yeoh, E., Chan, S. P., Wang, C. Y., Mohan, V., Cohen, N., McGill, M. J., & Twigg, S. M. (2023). Asia-Pacific consensus recommendations for application of continuous glucose monitoring in diabetes management. Diabetes Research and Clinical Practice, 201, 1–16. https://doi.org/10.1016/j.diabres.2023.110718

Moonesinghe, R., Beckles, G. L. A., Liu, T., & Khoury, M. J. (2018). The contribution of family history to the burden of diagnosed diabetes, undiagnosed diabetes, and prediabetes in the United States: Analysis of the National Health and Nutrition Examination Survey, 2009–2014. Genetics in Medicine, 20(10), 1159–1166. https://doi.org/10.1038/gim.2017.238

Ruslie, R. H., & Darmadi. (2012). Logistic regression analysis for factors affecting adolescent nutritional status. Andalas Medical Journal.

Tampil, Y. K. (2017). Logistic regression analysis to determine factors affecting cumulative achievement index (IPK) of students of FMIPA Sam Ratulangi University Manado. d'CARTESIAN: Journal of Mathematics and Applications.

Zhang, F. L., Xing, Y. Q., Guo, Z. N., Wu, Y. H., Liu, H. Y., & Yang, Y. (2018). Prevalence and risk factors for diabetes and impaired fasting glucose in Northeast China: Results from the 2016 China National Stroke Screening Survey. Diabetes Research and Clinical Practice, 144, 302–313. https://doi.org/10.1016/j.diabres.2018.09.005