https://doi.org/10.31272/jae.i152.1573
https://admics.uomustansiriyah.edu.iq
P-ISSN: 1813–6729 E-ISSN: 2707-1359

Evaluating Independent Variables on Binary Response: Logistic versus Probit Regression in Diabetes Data

Zewar Omar Ismael
Dept. of Statistics, College of Administration & Economics, University of Salahaddin, Erbil, Iraq.
Email: zewar.ismael@su.edu.krd , ORCID: https://orcid.org/0000-0002-0922-3425
Hunar Adam Hamza
Dept. of Statistics, College of Administration & Economics, University of Salahaddin, Erbil, Iraq.
Email: hunar.hamza@su.edu.krd , ORCID: https://orcid.org/0009-0009-4454-7669
Sami Ali Obed
Dept. of Statistics, College of Administration & Economics, University of Salahaddin, Erbil, Iraq.
Email: sami.obed@su.edu.krd , ORCID: https://orcid.org/0000-0002-2866-5886

Article Information

Article History:
Received: 17 / 02 / 2026
Revised: 09 / 05 / 2026
Accepted: 17 / 05 / 2026
Available Online: 01 / 06 / 2026
Pages no : 154 – 169

Keywords:
Diabetes, Logistic Regression, Probit Model, Binary Response, Comparative Analysis.

Correspondence:
Researcher name:
Zewar Omar Ismael

Email:
zewar.ismael@su.edu.krd

Abstract

The usage of logistic regression and probit regression are common statistical methods in the modeling of a binary outcome variable. This paper aims to identify which of the two models provides a better fit with data related to diabetes. Using a dataset with 768 records of patients collected from Layla Qassim Health Centers from 2018 to 2023, independent variables of Glucose, BMI, Blood Pressure, Insulin, Skin Thickness, Age, number of Pregnancies, and the Diabetes Pedigree Function were examined to predict a diagnosis of diabetes. The maximum likelihood method was employed to estimate both models. To see which model best fitted the dataset, several goodness of fit tests were calculated. These tests were McFadden's pseudo R squared, AIC and BIC, mean squared errors, and a receiver operating characteristic (ROC) curve. In each model tested, the variables Glucose, BMI, Diabetes Pedigree Function, and number of Pregnancies were statistically significant predictors of diabetes diagnosis. The addition of Insulin and Skin Thickness to either model of regression did not statistically help in the prediction of diabetes. The results of the analysis also indicated that the probit model performed slightly better. The conclusion drawn from the analysis was that the two models are interchangeable for binary outcome variable prediction, not solely for diabetes prediction.

1. Introduction

Diabetes mellitus has become one of the most ubiquitous and rapidly progressing chronic diseases. The rich legacy of diabetes mellitus on healthcare systems, economies, and the lives of billions of people is apparent. Since diabetes mellitus is a result of a combination of many genetic, metabolic, and demographic factors, it is vital that we know who is at risk. The sooner we identify the at-risk populations; the sooner we can identify the disease and prevent its progression. More importantly, we can design appropriate therapies for the patients.[1]

Logistic and probit regressions are the main two techniques used in studies with dichotomous dependent variables in the health field. Although the two are similar in many aspects, they differ in the way they treat errors. In the case of logistic regression, the authors assume errors are distributed logistically, whereas in the case of probit regression, the authors assume errors are distributed normally. Although the two systems are similar, they lead to different results. In the field of health,

particularly in clinical and public health, the two systems lead to different results in modeling and intervention.[2]

In recent years, interest in comparative analyses of logistic regression and probit models has grown, especially in the fields of biostatistics and health economics, in order to make the case that better specification leads to better explanatory and predictive frameworks for health outcomes. However, experimental data often designates that both models produce similar results, posing significant concerns about their relative effectiveness and practical interchangeability when used with actual medical data.[3]

2. Paper aim

Diabetes mellitus has become one of the most ubiquitous and rapidly progressing chronic diseases. The rich legacy of diabetes mellitus on healthcare systems, economies, and the lives of billions of people is apparent. Since diabetes mellitus is a result of a combination of many genetic, metabolic, and demographic factors, it is vital that we know who is at risk. The sooner we identify the at-risk populations; the sooner we can identify the disease and prevent its progression. More importantly, we can design appropriate therapies for the patients.

3. Paper importance

As interesting as the theory, is the application of these logistic/probit regression models to clinical data. While this paper indeed addresses the theoretical and imagines the application, the authors have really tested the models on clinical data collected from patients to then test which of the models may be more successful in, in turn, predicting the onset of diabetes. They also clearly begin to identify the risk factors to be included in the model – glucose, body mass index, strength of family history, number of pregnancies – with the application being the early identification of risk and more targeted screening.

Moreover, the article presents even more impressive, one-to-one statistical comparison of the two models with many classical indicators such as Mc Fadden‘s Pseudo R2, AIC, BIC, MSE and ROC-AUC, to name a few. These indicators, cumulatively, would tell much about the adequacy of the models not just in goodness of fit.

4. Review of Related Literature

Part of the motivation by Han and Lee (2019) was in response to a fundamental problem prevalent in applied econometrics the heavy reliance, often taken for granted, on joint normality assumptions in many bivariate probit type models. The authors proposed a more flexible semiparametric approach of combining a parametric copula function of the dependence structure with nonparametric marginal distributions. They were able to prove point identification and consistency along with asymptotic normality of their sieve maximum likelihood estimator under this framework. Their framework also allows for the estimation of the ATE under the standard exclusion restriction, greatly expanding its scope of applications in the models with dummy endogenous regressors.[4]

Kierinska et al (2020) moved to a more applied question: what influences household debt? Using survey data of 746 households in Central Pomerania, they employed a logistic regression model. They found that while economic education, stage of household life cycle, socio-economic type and higher income increased the probability of household debt, age of householder and more diversified sources of income reduced the probability. The model predicted well and was in agreement with the life-cycle hypothesis, and provided a good example of the usefulness of logit model for financial behavior analysis.[5]

On the problem of diabetes prediction, Rajendra and Latifi (2021) adopted a machine learning approach, where they use logistic regression as a base classifier but elevating it by use of feature selection and ensemble learning. Using two familiar real datasets –PIMA Indians dataset and the Vanderbilt dataset, they illustrated how prescriptive data prepossessing and feature engineering can help. Their results suggest that univariate feature selection improvements in prediction accuracy and

execution time and fancy ways of data-preprocessing can be as important as the models themselves. [2].

In 2022 Jawa, T.M. used a multiple logistic regression model to study the impact of home quarantine on psychological stability among 846 residents of the Makkah region, Saudi Arabia during COVID-19, using survey data collected during and after the compulsory quarantine period. The findings indicate that the majority of participants reported psychological stability during quarantine, while the most significant predictors of mental well-being are education level, psychological disorder indicators, and attention to general health emerged [6].

In 2023 Abonazel, M.R., Dawoud, I., Awwad, F.A. and Tag-Eldin, E., Proposed two new biased estimators PROMRT and PRODK to address the problem of multicollinearity in the probit regression model, where the conventional maximum likelihood (ML) estimator becomes inefficient. Building on ridge- and Liu-type estimators, the authors develop the PRODK estimator and provide rigorous theoretical comparisons using mean squared error (MSE) conditions, demonstrating its superiority over ML, probit ridge, probit Liu, and modified ridge estimators [7].

Ariyanto etal (2024) had an innovative solution to the issue of disease prediction by narrowing their use case down to one marginalized population of prospective sports athletes. The study integrated primary data from East-Java (Indonesia) prospective athletes on anthropometric parameters (height, weight, BMI, waist circumference, gender, age, parental work, and financial status) as well as socio-demographics into statistical models (logistic and probit regression). The aim was to create an early tool for predicting and preventing diabetes on a screen before diagnosis, where prediction is most useful in sports sciences. [8]

5. Methodology

5.1. Research Design

The framework of this analysis is quantitative and comparative, around one open but critical question: how do certain physiological and demographic variables affect an individual‘s probability of being diabetic? Since the measure of response (identification as a diabetic or not) can be categorized under two mutually exclusive states, the logical approach to analysing it is through binary response modelling.

Instead of working with just one statistical framework, this study intentionally juxtaposes the logistic and probit regresses: they are a pair of models confronted against one another in a straightforward, honest comparison of each one‘s utility in modeling the data, elucidating the phenomena, and detecting the trend.

In order to ensure a robust and comprehensive evaluation, multiple diagnostic methods are employed on both the goodness-of-fit and the predictive accuracy fronts. goodness-of-fit is gauged by McFadden‘s Pseudo-R2, log-likelihood ratio tests, the AIC and the BIC,12 and predictive accuracy is evaluated by means of mean squared error, ROC curves and the area under the curve13.

Apart from the figures, we also derive the marginal effects to give us a more interesting and more meaningful interpretation of the raw coefficients of the model in a clinical setting. Combined, these decision making seem to have two goals, to identify the actual risk factors of diabetes and to finally see whether logistic and probit regressions come to the same conclusion given the real binary health data.

5.2. Regression

Regression analysis is one of the most basic and common tools of quantitative research. “Its basic idea is about relationships about how a change in one or more independent variables is associated with a change in a particular dependent variable, with all other independent variables held constant.“1 Researchers use it for all sorts of insights into their data, predictions about the future, or tests of causality.2

The most straightforward, well-known case is linear regression, which models a one-dimensional relationship and is the starting point for many beginning students of statistics. However, regression is much more flexible than a one-dimensional case; if the research questions or data require it,

statisticians have a whole range of techniques available: multiple regression, for multiple predictor variables; binary choice models like logistic or probit, for dependent variables like "yes" or "no;" panel data techniques, for repeated measures over time; and nonlinear specifications, when a one-dimensional approach is insufficient.

It is the fact that it can be applied so broadly that it explains why it is such a long standing and valuable tool for many research fields. In economics, medicine, social sciences, engineering and many other disciplines, it provides theorists and researchers alike with an evidence-based way of making decisions and predictions. [8] [11].

5.2.1. Probit Regression

The probit model is given by

$$y_i^* = x_i' \beta + \varepsilon_i; \quad i = 1,2,\dots,n, \tag{1}$$

Where y_i^* is an unobserved or latent variable $x_i' = (x_{i1}, x_{i2}, \dots, x_{ip})$ is defined as the i[th] row of an X matrix with dimension n × p; p is the number of the explanatory variables, β is a (p × 1) vector of the regression coefficients, and ε_i is an error term that follows a normal distribution. Since the latent is unobservable in real-world datasets, the following dummy variable is examined instead .[9]

Probit regression serves as a standard modeling approach for binary or dichotomous response variables. Its application spans multiple disciplines, finding particular prominence in microeconomics, health economics research (Bishai, 1996)[10][20]., and the broader medical literature (Bhattacharyya, 1997). The method is especially well-suited for situations where the research objective centers on modeling an underlying latent variable—denoted Y*—which can be expressed through a regression framework as follows:

$$y_i^* = x_i' \beta + u_i, \quad i = 1,2,\dots,n \tag{2}$$

x_i' is the transpose of the i-th row of the data matrix X, β is the vector of regression coefficients, u is the error term.

Where x_i is the ith row of the $n \times (p+1)$ data matrix, X with (p+1)-vector, β of regression coefficients and $u = (u_1, u_2, \dots, u_n)'$ is the error vector with iid components having a continuous CDF, F(x) defined on R^1 with finite Fisher information:[10] [22].

$$I(f) = \int_{-\infty}^{\infty} \left[-\frac{f'(u)}{f(u)} \right]^2 f(u)\, du, \tag{3}$$

Where f is the PDF of you and $f'(u)$ is the derivative of $f(u)$. On the other hand, the latent variable $Y^*$is unobservable. So instead of y^*, we get the following dummy variable:

$$y_i = \begin{cases} 1 & if \quad y_i^* > 0 \\ 0 & otherwise \end{cases} \tag{4}$$

It has the following parameter and is distributed as a Bernoulli variable:

$$\pi_i = F(x_i'\beta); \quad i = 1,2,\dots,n \tag{5}$$

Given a sample size of n, the likelihood function is given by Hence.

$$L(\beta) = \prod_{i=1}^{n} \pi_i^{y_i} (1-\pi_i)^{1-y_i}. \tag{6}$$

$$ln\, L(\beta) = \sum_{i=1}^{n} y_i \log F(x_i'\beta) + \sum_{i=1}^{n} (1-y_i) \log(1 - F(x_i'\beta)) \tag{7}$$

The ML equations are utilized to estimate the success parameters β. The asymptotic characteristics of the several kinds of probit regression estimators are examined in this research [11]

The study uses a probit model for its estimations. Han and Lee (2019) state that a statistical model known as a probit regression is commonly used to evaluate the probability of an event, like how SMEs would respond to eco-innovation. It does particularly well with binary outcomes. The probit regression model is used to test the binary outcome, with 'o' denoting 'unlikely' and '1' denoting 'likely'. This is how the binary probit regression equation is set up.[12]

$$Pr(Y = 1 \mid X_i) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + e_{\mathrm{i}} \quad (8)$$

Where: X : is the explanatory variable. β_0: is the intercept parameter of probit regression model. β_1: is the slope parameter of probit regression model [13]

5.2.2. A Random-Effects Probit Regression Model

explains the generic parameter estimation method for the random-effects probit regression model, which was first presented by Gibbons and Bock in 1987. In order to evaluate trends at both the group and individual levels, it also covers empirical Bayes estimates of person- or cluster-specific effects and standard errors. The study offers a model with two random effects for longitudinal data analysis and a single random effect model appropriate for clustered or longitudinal research designs [11][8]

5.3. Binary logistic regression

The relationship between one or more independent variables also known as accompanying variables or independent variables and a nominally scaled dependent variable can be quantified in any manner.

The analysis in this case is referred to as a binary logistic regression. The statistical technique used to assess and balance the relationship between a binary dependent variable and one or more independent variables of any kind is known as logistic regression analysis. One of two main categories of models—log linear and genuine (binary and ordinal regression) models—can be used to analyses ordinal categorical data [21] [4].

5.3.1. Ratio and Odds

One part of the total that is directly related to probability is the ratio. Chances, on the other hand, are a way to compare the probability that something will occur against the probability that it will not. which:

$$\mathrm{Odds} = \frac{\mathrm{p}}{(1 - \mathrm{p})} \quad (9)$$

Where: (Odds) is the coefficient ratio of the occurrence of event. (p) is the probability of the event occurrence. (1- p) is the probability of event that not occurred. Odds ratio

Is the ratio between the odds of a variable (QI) and the odds for another variable (Q2), i.e., the odds ratio is equal to.

$$OR = \frac{OddsQ1}{OddsQ2} = \frac{N_1}{N_2} \quad (10)$$

Where: OR is the Odds ratio. Q1 and Q2 are the first and second Odds [15].

5.3.1.1. Logit

It is the natural logarithm of odds, where if NI is a case in one organization and N2 is a number of cases in another, the:

$$Loglit = log\left(\frac{N_1}{N_2}\right) = ln(odds) \quad (11)$$

As aimed at, the (logit) in terms of probability, it is expressed as:

The primary purpose of the logit function is to transform the probability of a binary outcome into a linear and mathematically tractable form, making it possible to analyze the relationship between a set of independent variables and a binary dependent variable using regression techniques. In binary response models, the dependent variable takes only two possible outcomes, such as the presence or absence of diabetes disease, success or failure, or yes and no responses. Because ordinary linear regression is not appropriate for modeling probabilities restricted between 0 and 1, the logit transformation converts the probability into log-odds, which can take any real value from negative infinity to positive infinity. [15]

5.3.1.2. Wald Statistics

The Wald statistic serves as a tool for evaluating whether individual regression coefficients carry statistical significance. At its core, this test probes the null hypothesis—that a given predictor's coefficient equals zero, implying no meaningful effect on the outcome. For each independent variable in the logistic model, the Wald statistic is calculated as follows:

$$W = \left(\frac{b}{S.E_B}\right)^2 \quad (12)$$

5.3.2. Ordinal Logit Model Fitting

If the relationship between the levels of the ordinal dependent scales is established, it is possible to fit a mean dependent model and practically assign numerical scores to the dependent levels. Cases involving more than two sessions can be accommodated by expanding the ordinal logistic regression model. Assume that the first association that needs to be classified is:

$$Y_i = X'\beta + \varepsilon \quad (13)$$

Where: Y_i is the specific but undetected dependent variable X is the vector of independent variables. β is the vector of Regression Parameter which we demand to estimate. Moreover, $\boldsymbol{\varepsilon}$ is the random error .[16]

The −2 log likelihood (−2LL) values for the initial and final models are provided by the model fitting information, along with a chi-square to assess the difference between the −2LL for the two models. The model fits significantly better than the location-only model, according to the goodness of fit statistics. The typical Pearson and Deviance goodness of fit measures can be calculated using the detected and predictable frequencies. The goodness of fit statistic for Pearson is [18][19]

$$x^2 = \sum\sum \left(\frac{O_{ij} - E_{ij}}{E_{ij}}\right)^2 \quad (14)$$

The deviance measure is:

$$D = 2\sum\sum O_{ij} ln\left(\frac{O_{ij}}{E_{ij}}\right)^2 \quad (15)$$

Only models with practically big predictable values in each cell would be utilized for both goodness-of-fit statistics. We may have numerous cells with low anticipated values if we have a continuous independent variable, numerous categorical interpreters, or some predictors with a wide range of values. The strong point of the relationship between the dependent variable and the interpreter variables can be measured using a variety of R^2-like measures. Their interpretation is difficult because they are not as useful as the R^2 statistic in regression. Three statistics that are frequently utilized are [6]

1-Cox and Snell $R^2 = 1 - \left(\frac{L(\beta^0)}{L(\hat{\beta})}\right)^{\frac{2}{n}}$ (16)

2-Nagelkerke $R^2 = \frac{\text{Cox and Snell } R^2}{1 - L(\beta^0)^{\frac{2}{n}}}$ (17)

3- McFadden's $R^2 = 1 - \left(\frac{L(\hat{\beta})}{L(\beta^0)}\right)$ (18)

Where, $L(\hat{\beta})$ is the log-likelihood function for the model with the estimated parameters, and $L(\beta^0)$ is the log likelihood with just the thresholds and n is the number of cases [13].

5.4. Estimation Technique

Maximum Likelihood Estimation (MLE), which yields reliable and asymptotically efficient parameter estimates under typical regularity criteria, is used to estimate both logistic and probit models. To confirm that the model is adequate in relation to the null specification, convergence diagnostics and likelihood ratio tests.[19]

5.5. Model Evaluation and Comparison

Model performance is evaluated using multiple criteria:

1. McFadden's Pseudo R^2 to assess explanatory power
2. Log-likelihood and Likelihood Ratio (LR) tests to evaluate overall model significance
3. Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) to compare model fit while accounting for complexity
4. Mean Squared Error (MSE) to assess predictive accuracy
5. Receiver Operating Characteristic (ROC) curve and Area Under the Curve (AUC) to evaluate classification performance

Lower AIC, BIC, and MSE values indicate superior model performance, while higher pseudo-R^2 and AUC values suggest stronger explanatory and discriminatory power [11]

5.6. Marginal Effects Analysis

Marginal effects are calculated at the mean values of the explanatory variables to improve interpretability. When all other variables are held constant, these effects quantify the change in the expected likelihood of diabetes linked to a one-unit change in each independent variable. Despite variations in coefficient scaling, this method allows for a valid comparison between the logistic and probit models. [5]

6. Experimental Work

6.1. Data collection

The diabetes dataset used in this study was collected at Layla Qassim Health Centres. For all diabetic patients, 768 cases were gathered throughout the course of five years, from January 1, 2018, to December 31, 2023

The data includes six independent variables, such as blood pressure, skin thickness, insulin, body mass index (weight/height), pedigree function, and age, and one dependent variable, diabetes type 1 and type 2, and the distribution of this variable (diabetes type) is a binary distribution. We used SPSS and the Python program for this analysis.

The predictor set encompasses glucose concentration, body mass index, blood pressure, insulin levels, skin thickness, age, pregnancy count, and the diabetes pedigree function—a measure capturing hereditary diabetes risk.

6.2 Variable Definition

The dependent variable used in this study is diabetes status, which is defined as:

1. 1 = Patient has diabetes (Ordinal)
2. 0 = Patient does not have diabetes (Ordinal)

The explanatory variables entered into the models include:

1. Pregnancy count (Scale)
2. Glucose concentration (Scale)
3. Blood pressure (Scale)
4. Skin thickness (Scale)
5. Insulin level (Scale)
6. Body Mass Index (BMI) (Scale)
7. Diabetes pedigree function (family history indicator) (Ratio)
8. Age (Scale)

Table (1) Collinearity Statistics Test for the Multicollinearity Problem between Independent Variables

Collinearity Statistics Test for the Multicollinearity Problem		
Variables	Tolerance	VIF
Pregnancies	.713	1.460
blood pressure	.813	1.230
Glucose	.754	1.342
skin thickness	.678	1.475
Insulin	.788	1.269
body mass index(weight/height)	.735	1.360
Diabetes Pedigree Function	.942	1.062
Age	.879	1.138

This table presents the results of the collinearity test used to examine multicollinearity among the independent variables. The results indicate that all Variance Inflation Factor (VIF) values are relatively low and fall well below the commonly accepted threshold value of 10, while all tolerance values are greater than 0.1. Specifically, the VIF values range from 1.062 to 1.475, indicating weak

correlation among the explanatory variables. These findings confirm that the regression model does not suffer from serious multicollinearity.

Table (2): Probit Regression Results

Probit Regression Results			
Dependent variable	Outcome	No. observation	768
Model	Logit	DF residual	759
Method	MLE	DF Model	8
		Pseudo R-square	0.2697
		Log likelihood	-362.79
Converged	True	LL-Null	-496.74
Covariance type	Non robust	LLR p-value	2.736e-53

The table describes the Probit regression model for the diabetes outcome, estimated by MLE. Using MLE, the regression converged with a log-likelihood of -362.79, versus -496.74 for the null model: this is a considerable improvement and indicates the model captures the data's variation much more effectively than the baseline model. Confirming this, the likelihood ratio test for the model is highly significant (LLR p-value = 2.736e-53), ruling out any doubt that at least one of the included independent variables has a genuine influence on diabetes outcome.

As for the explanatory power, the Pseudo R2 of this model is 0.2697, which implies it explains about 26.97% of the variation in a person's diabetes status (again, they model the diabetes status as binary). As we can see, it is very close to the Logistic model's 27.18, which is a valuable clue that, when the data are reweighted to account for their departure from normality, both models are trying to fit the same thing.

Overall, the Probit model provides a reasonable and statistically significant fit, though it is only slightly worse than the Logit model on this criterion. No great surprise, perhaps, given that the two models are based on the same theory, with only the distribution of the latent variable differing between them. In real clinical data of this form, at least, the difference is minimal, and it is reasonable to consider them interchangeable.

Table (3): Estimates of parameters of the probit model

Variables	**coefficient**	**Standard error**	**Z**	**P>|Z|**	**Confidence interval**	
					0.025	0.975
Constant	-4.8638	0.384	-12.653	0.000	-5.617	-4.110
Pregnancies	0.0723	0.018	3.972	0.000	0.037	0.108
Glucose	0.0199	0.002	9.968	0.000	0.016	0.024
BloodPressure	-0.0079	0.003	-2.585	0.010	-0.014	-0.002
SkinThickness	0.0012	0.004	0.307	0.758	-0.007	0.009
Insulin	-0.0007	0.001	-1.423	0.155	-.002	0.000
Body Mass Index (weight in kg/(height in m)^2).	0.0523	0.008	6.250	0.000	0.036	0.069
DiabetesPedigreeFunction	0.4982	0.164	2.036	0.002	0.177	0.820
Age	0.0102	0.005	2.887	0.059	-0.000	0.021

This table shows the computed coefficients from the Probit regression model used to forecast the risk of diabetes. The constant term (–4.8638, $p < 0.001$) is significantly negative when all variables are set to 0, indicating an extremely low baseline probability of diabetes. Pregnancies ($\beta = 0.0723$, p-value = 0.001), glucose ($\beta = 0.0199$, p-value = 0.001), BMI ($= 0.0523$, p-value = 0.001), and diabetes pedigree function($beta\ equals$ 0.4982, p-value = 0.002) are all positive and highly significant independent variables, suggesting that increases in these variables raise the risk of developing diabetes. Blood pressure has a slight but negative impact (β = –0.0079, p-value = 0.010), indicating that higher blood pressure somewhat lowers the risk of diabetes in this group.

On the other hand, skin thickness (p-value = 0.758) and insulin (p-value = 0.155) have minimal effects and are statistically insignificant. The effect of age is slightly positive (β = 0.0102, p-value=0.059), although it is just on the boundary of significance. When rounded to 2 decimal places,

it fails to meet the 5% level of significance, but it still shows a steady increase in the patient's probability of developing diabetes with age. More generally, the Probit model yields the same conclusions as the Logistic model, and this is one of the more striking and important observations in this whole report. Glucose, BMI, and Family are all significant and relatively precise predictors in both cases. The scale of the B values is very different between the two models, but this is just one of those subtleties that comes from Probit‘s dependence on the cumulative normal distribution and from the interpretation of coefficients there, and it is a nuance that does not really matter in terms of what we can infer from the models. Beyond that superficial difference, the two approaches tell the same story.

Table (4): Probit marginal effect

Variables	dy/dx	Standard error	Z	P>|Z|	Confidence interval	
					0.025	0.975
Pregnancies	0.0252	0.006	3.973	0.000	0.013	0.038
Glucose	0.0069	0.001	9.864	0.000	0.006	0.008
Blood Pressure	-0.0028	0.001	-2.582	0.010	-0.005	-0.001
Skin Thickness	0.0004	0.001	0.307	0.759	-0.002	0.003
Insulin	-0.0003	0.000	-1.422	0.155	-0.001	9.81e-05
Body Mass Index (weight in kg/(height in m)^2).	0.0183	0.003	6.325	0.000	0.013	0.024
Diabetes Pedigree Function	0.1740	0.057	3.031	0.002	0.061	0.287
Age	0.0036	0.002	1.893	0.058	-0.000	0.007

This table summarises the marginal effects of the Probit regression model, showing how a one-unit increase in each variable affects the risk of diabetes, holding all other factors constant. The data show that the largest and most significant favourable impacts are associated with glucose (dy/dx = 0.0069, p-value < 0.001), BMI (dy/dx = 0.0183, p < 0.001), and Diabetes Pedigree Function (dy/dx = 0.1740, p-value = 0.002). This indicates that higher glucose levels, larger body mass, and a stronger family history greatly increase the risk of having diabetes. Pregnancies also had a significant positive effect (dy/dx = 0.0252, p-value < 0.001), meaning that each subsequent pregnancy raises the risk of diabetes by about 2.5 percentage points.

However, blood pressure (dy/dx = –0.0028, p-value = 0.010) has a small but significant detrimental effect after controlling for other factors. In contrast, insulin and skin thickness are not statistically significant, indicating little contribution to diabetes risk. Age and the likelihood of having diabetes have a weak positive connection (dy/dx = 0.0036, p-value = 0.058). The marginal effect estimates from the Probit specification closely align with those from the Logit model. Both approaches converge on the same set of key risk factors—Glucose, BMI, Family History, and Pregnancies—even as the Probit tends to yield marginally smaller effect sizes, a consequence of its cumulative normal distribution scaling.

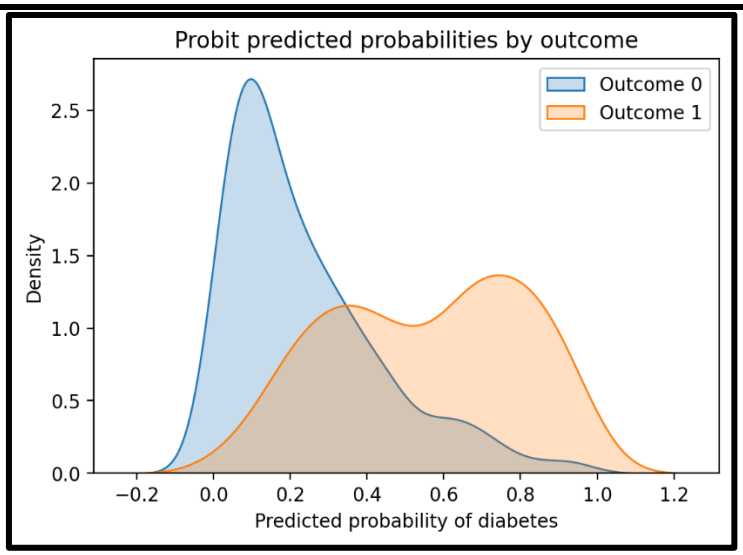

Figure (1): predicted probability of diabetes

The figure illustrates how predicted probabilities are distributed across the Probit model's two outcome groups: individuals diagnosed with diabetes (Outcome 1) and those without (Outcome 0). For non-diabetic patients, represented by the blue curve, the distribution clusters tightly around the lower end of the probability scale, with a pronounced peak falling between 0.1 and 0.2. This pattern suggests that the model correctly assigns a low risk of diabetes to most individuals who are, in fact, diabetes-free. The orange curve, representing actual diabetic cases, tells a different story. It skews toward the upper end of the probability range, roughly 0.6-0.8, reflecting the model's ability to flag genuine diabetic patients as higher risk. Some overlap between the curves does appear—pointing to a subset of cases where predictions missed the mark—but the broader separation confirms that the Probit model discriminates effectively between diabetic and non-diabetic individuals, achieving strong predictive performance overall.

Table (5): Logit Regression Results

Logit Regression Results			
Dependent variable	Outcome	No. observation	768
Model	Logit	DF residual	759
Method	MLE	DF Model	8
		Pseudo R-square	0.2718
		Log likelihood	-361.72
Converged	True	LL-Null	-496.74
Covariance type	Non robust	LLR p-value	9.652e-54

Based on 768 observations, the logistic regression results indicate that the model provides a convergent and statistically significant explanation of the binary outcome (diabetes = 1, no diabetes = 0). When independent variables were added, the model's log-likelihood rose from –496.74 (null) to –361.72, indicating a notable improvement in model fit. Predictors had significant effects on the likelihood of diabetes exposure, as confirmed by the likelihood ratio test (9.651054). With a McFadden Pseudo R2 of 0.2718, the model is regarded as a moderate to strong fit for binary data, explaining around 27% of the variation in the log-odds of diabetes.

Overall, the findings suggest that the independent variables importantly influence diabetes risk, and the logistic model is well-suited and statistically stable for predicting a binary health outcome.

Table (6): Estimate of parameters in logistic model

Variables	coefficient	Standard error	Z	P>|Z|	Confidence interval	
					0.025	0.975
Constant	-8.4047	0.717	-11.728	0.000	-9.809	-7.000
Pregnancies	0.1232	0.032	3.840	0.000	0.060	0.186
Glucose	0.0352	0.004	9.481	0.000	0.028	0.042
Blood Pressure	-0.0133	0.005	-2.540	0.011	-0.024	-0.003
Skin Thickness	0.0006	0.007	0.090	0.929	-0.013	0.014
Insulin	-0.0012	0.001	-1.322	0.186	-0.003	0.001
Body Mass Index (weight in kg/(height in m)^2).	0.0897	0.015	5.946	0.000	0.060	0.119
Diabetes Pedigree Function	0.9452	0.299	3.160	0.002	0.359	1.531
Age	0.0149	0.009	1.593	0.011	-0.003	0.033

This table shows the parameter estimates from a logistic regression model examining factors associated with the presence of diabetes (binary outcome). When all variables are zero, the baseline likelihood of diabetes is exceedingly low, as seen by the considerably negative constant term (-8.4047) ($p < 0.001$). Among the factors with positive, statistically significant impacts are pregnancies ($\beta = 0.1232$, p-value < 0.001), glucose ($\beta = 0.0352$, p-value < 0.001), BMI ($\beta = 0.0897$, p-value < 0.001), and the diabetes pedigree function ($\beta = 0.9452$, p-value $= 0.002$). In contrast, blood pressure has a small but significant negative effect ($\beta = -0.0133$, p-value $= 0.011$), suggesting that higher blood pressure somewhat lowers the risk of diabetes in this dataset.

Skin thickness ($p = 0.929$), insulin (p-value $= 0.186$), and age (p-value $= 0.011$, borderline, with a low coefficient) are weaker predictors, indicating a small or unclear contribution. Glucose, BMI, and Family Diabetes History (Pedigree Function) are the strongest predictors overall, suggesting that genetic and metabolic risk factors are significant determinants affecting the chance of acquiring diabetes.

Table (7): Logistic marginal effect

Variables	dy/dx	Standard error	Z	P>|Z|	Confidence interval	
					0.025	0.975
Pregnancies	0.0189	0.005	3.969	0.000	0.010	0.028
Glucose	0.0054	0.000	12.303	0.000	0.005	0.006
Blood Pressure	-0.0020	0.001	-2.577	0.010	-0.004	-0.000
Skin Thickness	9.50e-05	0.001	0.090	0.929	-0.002	0.002
Insulin	-0.002	0.000	-1.327	0.184	-0.000	8.72e-05
Body Mass Index (weight in kg/(height in m)^2).	0.0138	0.002	6.437	0.000	0.010	0.018
Diabetes Pedigree Function	0.1451	0.045	3.227	0.001	0.057	0.233
Age	0.0023	0.001	1.601	0.009	-0.001	0.005

Glucose is the single most important predictor in our model, with the highest marginal effect ($\beta = 0.0054$, p-value < 0.001). Practically, the fact that each additional unit of glucose would raise a person‘s chance of having diabetes (as indicated by the predicted probability of having or not having diabetes) by about 0.54% is a relatively insignificant change. Its clinical importance, however, becomes apparent when considering that considerably larger numbers of units of glucose from one individual to another across the population are liable to occur. BMI ($\beta = 0.0138$, p-value < 0.001) and the Diabetes Pedigree Function ($\beta = 0.1451$, p-value $= 0.001$) are also powerful and statistically significant predictors, meaning that being even more obese and having a relatively more diabetic ancestry outweigh this additional risk.

The number of pregnancies also has a statistically significant, though much more moderate, positive effect ($\beta = 0.0189$, p-value < 0.001), where each new pregnancy adds a millionth of an

increased probability of developing diabetes above the last, over 1.9%, suggesting that the metabolic toll that pregnancies take on the body can accumulate to have a real effect over multiple visits. The findings are much more difficult to interpret for systolic blood pressure. Its marginal effect is negative, meaning that each increase in systolic blood pressure weakly decreases the probability of the person being diabetic ($\beta = -0.0020$, p-value = 0.010), indicating that increased blood pressure weakly decreases rather than increases the risk of diabetes, though the result is statistically significant. The direction of the effect is believed to be due to the unaffordability of medication for all hypertensive people and the presence of other variables.

Finally, insulin and skin thickness are of questionable significance after controlling for the other predictors. Their marginal effects, 0.3062 and 0.3344, respectively, are quite negligible and not significant at the 0.95 level, indicating that neither variable appears to have a reliable direct impact on the disease in this data set. Age is the third and final variable to include, having a small, somewhat weakly significant effect of 0.0023 (0.059). This would show me that older populations are more vulnerable to diabetes, but only by the briefest of margins, significantly less than glucose or BMI. Still, we would want to use additional data to verify this with greater confidence, of course.

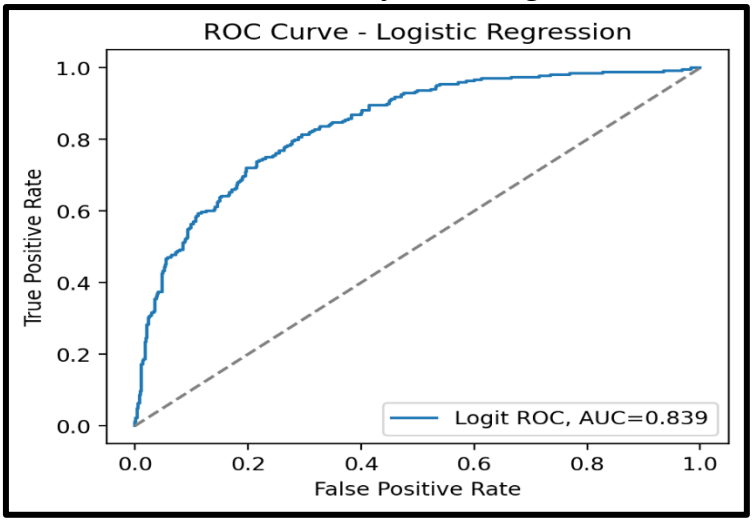

Figure (2): Roc Curve Logistic Regression model

The ROC curve paints a nice, visual picture of the performance of our classifier, the logistic regression model. With an AUC of 0.839, our model shows good discrimination: approximately 84% of the time, it correctly classifies a randomly chosen diabetic or non-diabetic sample.

Moreover, of course, the shape of the curve provides additional information. This dynamic, mountain-ridge-like oblateness indicates excellent sensitivity and specificity, as many true positives are detected without contaminating the denominator with a large number of false positives. The longer the crick, the more it adds.

If we look at all of these aspects in total, they are quite comforting to see; when we put all these bits of information into a logistic regression model, it is not merely adequate statistically, but in fact quite a reliable and robust model, and therefore could be a credible screening test in a clinical environment.

Table (8): Comparison model

Model	MSE	Pseudo_R2_McFadden	AIC	BIC
Probit	0.1512294687	0.2996646719	740.5763975	782.3705051
Logit	0.1527257557	0.2718096686	741.4453778	783.2394854

Based on the Mean Squared Error, the Probit model performs slightly better than the Logit, with a value of 0.1512 versus 0.1527. The smaller value indicates that the Probit model's predicted probabilities are slightly more accurate representations of the observed values.

The Pseudo R2 follows a similar pattern. The Probit model reports a McFadden of 0.2997 compared to 0.2718 for the Logit, suggesting it accounts for a marginally larger proportion of the variation in diabetes outcomes and is overall a slightly better fit to the data.

Information criteria, again, work in this same case. The Probit model, with lower AIC (740.58) and BIC (782.37) scores, again surpasses the Logit model (AIC 741.45, BIC 783.24) on the criterion for the best model in the trade-off between complexity and goodness of fit.

On a macro, more general level, the Probit model is better than Logit on all major criteria (larger Pseudo R2, smaller AIC and BIC, lower MSE). However, these differences should be viewed in context. Both models are strong, produce fairly similar results, and generally tell the same story about diabetes and its risks. It is not significant enough to matter in any useful clinical setting, and an analyst could easily pick either one. A slight difference in performance does not make one or the other clearly better.

Table (9) Goodness of fit test for probit and logistic model

Statistics	probit	logistic
-2 Log Likelihood	-362.79	-361.72
Prob > Chi²	0.0000	0.000
McFadden (R^2)	0.299	0.271
Hosmer–Lemeshow p-value	0.421	0.571
AIC	740.57	741.44
BIC	782.37	783.23
AUC	0.839	0.805

The goodness-of-fit results provide a clear message: the Probit and Logistic regression models are statistically significant, and both fit the data well. Both models yield significant p-values for the Likelihood Ratio Chi-square, indicating that the explanatory variables, as a whole, make a genuine and significant contribution to predicting diabetes.

As for explanatory power, the McFadden Pseudo R2 shows that the Probit model explains a slightly larger proportion of the variation in the outcome variable. Supporting this further, the Hosmer-Lemeshow test results for both models are statistically insignificant, which (in this case) is exactly what a researcher would like to see. There is no evidence of inadequate fit, and both models seem to capture the data quite well.

Model-wise, the Probit model is again superior, as indicated by lower AIC and BIC values compared to the Logistic model. This shows that the Probit specification tends to perform better on the informational efficiency criterion, using the data more efficiently by capturing more information with less complexity.

A further presentation is added to the ROC curve analysis. While both models perform well, the slight edge in predicting segregation in diabetes, as indicated by the AUC value of the Probit model, is evident.

Summing these all up, the Probit model seems to do slightly better than the Logistic regression model for this data set, at least not by a mile, but in every way I was able to compare.

Table (10): Summary Interpretation

Test / Measure	Result	Interpretation
McFadden's Pseudo R^2	0.2997	Strong fit
LLR Test ($p < 0.001$)	Significant	Model improves over null
AIC/BIC	Low	Efficient and well-balanced fit
Hosmer–Lemeshow	$p > 0.05$ (ideal)	No significant misfit
AUC (≈ 0.84)	Excellent	Strong classification ability
Residuals	Random	No major model violations

7. Conclusion

1. Several clear and informative conclusions can be drawn from this analysis. The first is that the variables most predictive of diabetes risk, when used, are glucose, BMI, family history of

diabetes, and number of pregnancies; predictions based on these variables align very well with what is readily known from clinical experience.
2. The successful application of both logistic and probit models to the real clinical data showed that binomial response models are a suitable approach to epidemiological research on health outcomes such as diabetes.
3. When the two specifications are compared by evaluating their performance across the first five performance specifications, it appears that the two specifications behave roughly similarly. Still, the Probit specification consistently does slightly better across all the specifications and shows a marginally stronger overall fit and predictive accuracy. The difference is subtle and probably would not change any actual clinical decision, but on these data, the Probit specification appears marginally better. The probit model showed slightly better goodness-of-fit and predictive accuracy than the logistic model.
4. Logistic and probit models produced very similar and reliable results for binary health outcomes.

8. Recommendation

1. Larger datasets from different hospitals and regions should be collected to improve the generalizability and reliability of the results.
2. Apply advanced validation methods (cross-validation, out-of-sample testing) and compare with machine learning models to strengthen predictive performance.
3. Prioritise early screening and preventive programs targeting high-risk individuals based on key predictors such as glucose, BMI, family history, and pregnancy count.
4. Medical institutions should apply statistical prediction models to support clinical decision-making and diabetes prevention programs.
5. Dividing the data into training and testing sets could be considered in future studies to improve predictive evaluation and model robustness.

9. Supplementary material

(None).

10. Author's Contributions

Zewar Omer Ismaeil: Designed the research and analysis for this Practical part. Sami Ali Obed. Writing, editing, and using Code in Python. Hunar Adam Hamza: Conducted the analyses.

11. Funding

(None).

12. Data availability statement

Daily recorded dataset from the Layila Qasim Centre for Diabetes in Erbil, Kurdistan, Iraq (private data).

(https://drive.google.com/drive/folders/1UjlExmw5PADwxqm7gwKJM4_WnFzukQOF?usp=drive_link).

13. Acknowledgements

The authors would like to express their sincere gratitude to Layla Qassim Health Centres for providing access to the diabetes patient data used in this study. Special appreciation is extended to the medical and administrative staff for their cooperation and assistance during the data collection process. The authors also acknowledge the Department of Statistics, College of Administration and Economics, University of Salahaddin, Erbil, for its academic support and guidance.

14. Conflict of interest

The authors declare no conflict of interest.

15. Declaration of generative AI use

During the preparation of this work, the authors used (Google Translate, ChatGPT, Gemini,)and (Grammarly) for grammar checking and language polishing. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication

References

[1] Washington, S., Karlaftis, M. G., & Mannering, F. L. (2003). Statistical and econometric methods for transportation data analysis. Chapman & Hall/CRC.
[2] Rajendra, P., & Latifi, S. (2021). Prediction of diabetes using logistic regression and ensemble techniques. Computer Methods and Programs in Biomedicine Update, 1, 100032. https://doi.org/10.1016/j.cmpbup.2021.100032
[3] Aliman, G., Faye Nivera, T. S., Charmille Olazo, J. A., Jane Ramos, D. P., Danielle Sanchez, C. B., Amado, T. M., Arago, N. M., Jorda Jr., R. L., Virrey, G. C., & Valenzuela, I. C. (2022). Sentiment analysis using logistic regression.
[4] Han, S., & Lee, S. (2018). Estimation in a generalisation of bivariate probit models with dummy endogenous regressors. arXiv preprint arXiv:1808.05792. https://doi.org/10.48550/arXiv.1808.05792
[5] Strzelecka, A., Kurdyś-Kujawska, A., & Zawadzka, D. (2020). Application of logistic regression models to assess household financial decisions regarding debt. *Procedia Computer Science*, 176, 3418–3427. https://doi.org/10.1016/j.procs.2020.09.055
[6] Jawa, T. M. (2022). Logistic regression analysis for studying the impact of home quarantine on psychological health during COVID-19 in Saudi Arabia. *Alexandria Engineering Journal*, 61(10), 7995–8005. https://doi.org/10.1016/j.aej.2022.01.047
[7] Abonazel, M. R., Dawoud, I., Awwad, F. A., & Tag-Eldin, E. (2023). New biased estimators for the probit regression model. *Scientific Reports*, 13(1), 5851. https://doi.org/10.1038/s41598-023-32452-4
[8] Srisathan, W. A., Ketkaew, C., Phonthanukitithaworn, C., & Naruetharadhol, P. (2023). Driving policy support for open eco-innovation enterprises in Thailand: A probit regression model. Journal of Open Innovation: Technology, Market, and Complexity, 9(3), 100084. https://doi.org/10.1016/j.joitmc.2023.100084 *(*
[9] Bhattacharyya, K. (1997). Key issues in polychotomous logit regression. *Statistics in Medicine*, 16(12), 1391–1397. https://doi.org/10.1002/(SICI)1097-0258(19970630)16:12<1391::AID-SIM540>3.0.CO;2-Y
[10] Bishai, D. (1996). Quality efficiency of health care providers: Inferences from a latent variable model. *Health Economics*, 5(2), 143–155. https://doi.org/10.1002/(SICI)1099-1050(199603)5:2<143::AID
[11] Alhemoud, A. M. (2011). An ordered probit regression model for estimating the effects of demographic factors on rice consumption in Saudi Arabia. *International Journal of Business and Globalisation*, 6(1), 1–17. https://doi.org/10.1504/IJBG.2011.037416
[12] Obed, S. A., Mohammed, P. A., & Kadir, D. H. (2021). The estimation of (Covid-19) cases in Kurdistan region using Nelson Aalen estimator. *Cihan University-Erbil Scientific Journal*, 5(2), 24–31. https://doi.org/10.24086/cuesj.v5n2y2021.pp24-31
[13] Kadir, D. H., & Amin, K. A. (2022). Comparing between Probit and Logit regression models to predict the factors affecting inflation in Iraq. *Tikrit Journal of Administrative and Economic Sciences*, 18(58, 2), 438–455. https://doi.org/10.25130/tjaes.18.58.2.25
[14] Golam Kibria, B. M., & Saleh, A. K. M. E. (2012). Improving the estimators of the parameters of a probit regression model: A ridge regression approach. *Journal of Statistical Planning and Inference*, 142(6), 1421–1435. https://doi.org/10.1016/j.jspi.2011.12.023
[15] Khaleel, Z. J., & Al-Jamil, S. K. (2021). The impact of some public finance variables on the gross national savings in Iraq. *Tikrit Journal of Administrative and Economic Sciences*, 17(54, 3), 430–443. https://doi.org/10.25130/tjaes.17.54.3.27
[16] Salh, S. M., Abdalla, H. T., & Omer, Z. M. (2021). Using multinomial logistic regression model to study factors that affect chest pain. *Tikrit Journal of Administrative and Economic Sciences*, 17(53, 2), 534–555. https://doi.org/10.25130/tjaes.17.53.2.31
[17] Sule, B. O., & Saporu, F. W. O. (2015). Mathematical theory and modeling. Journal of Mathematical Theory and Modeling, 5(10), 45-53.
[18] Hua, Y., Stead, T. S., George, A., & Ganti, L. (2025). Clinical risk prediction with logistic regression: Best practices, validation techniques, and applications in medical research. *Academic Medicine & Surgery*, 4(1), e131964. https://doi.org/10.62186/ams.v4i1.131964
[19] Dey, D., Haque, M. S., Islam, M. M., Aishi, U. I., Shammy, S. S., Mayen, M. S. A., Noor, S. T. A., & Uddin, M. J. (2025). The proper application of logistic regression model in complex survey data: A systematic review. *BMC Medical Research Methodology*, 25(1), 1–15. https://doi.org/10.1186/s12874-024-02454-5
[20] Bewick, V., Cheek, L., & Ball, J. (2005). Statistics review 14: Logistic regression. *Critical Care*, 9(1), 112–118. https://doi.org/10.1186/cc3045
[21] Gibbons, R. D., Hedeker, D., & Lab, B. (1994). Application of random-effects probit regression models. *Journal of Consulting and Clinical Psychology*, 62(2), 285–296. https://doi.org/10.1037/0022-006X.62.2.285

https://doi.org/10.31272/jae.i152.1573
https://admics.uomustansiriyah.edu.iq
P-ISSN: 1813–6729 E-ISSN: 2707-1359

تقييم أثر المتغيرات المستقلة على الاستجابة الثنائية: الانحدار اللوجستي مقابل انحدار بروبيت في بيانات مرض السكري

زيوەر عمر إسماعيل
قسم الإحصاء، كلية الإدارة والاقتصاد، جامعة الاصلاح الدين، أربيل، العراق.
Email: zewar.ismael@su.edu.krd , ORCID: https://orcid.org/0000-0002-0922-3425

هونەر ادم همزة
قسم الإحصاء، كلية الإدارة والاقتصاد، جامعة الاصلاح الدين، أربيل، العراق.
Email: hunar.hamza@su.edu.krd , ORCID: https://orcid.org/0009-0009-4454-7669

سامى على عبيد
قسم الإحصاء، كلية الإدارة والاقتصاد، جامعة الاصلاح الدين، أربيل، العراق.
Email: sami.obed@su.edu.krd, ORCID: https://orcid.org/0000-0002-2866-5886

المستخلص

تبحث هذه الدراسة في كيفية تأثير مختلف المتغيرات الديموغرافية والفسيولوجية على نتائج مرض السكري من خلال نمذجة الاستجابة الثنائية، مع مقارنة منهجين إحصائيين واسعي الاستعمال: الانحدار اللوجستي وانحدار بروبيت، يستند التحليل إلى بيانات من 768 مريضاً زاروا مراكز ليلى قاسم الصحية على مدى فترة خمس سنوات تمتد من 2018 إلى 2023، تم فحص ثمانية متغيرات تنبؤية: مستويات الجلوكوز، مؤشر كتلة الجسم، قراءات ضغط الدم، قياسات الأنسولين، سمك الجلد، العمر، عدد مرات الحمل، ودالة سلالة السكري—وهي مقياس يجسد المخاطر الوراثية. تم ملائمة كلا الانموذجين الإحصائيين باستعمال تقدير الاحتمال الأقصى، ولتقييم مدى جودة أداء كل نموذج، تم تطبيق العديد من الأدوات التشخيصية، بما في ذلك معامل ماكفادين لرصد التحديد الزائف، ومعايير أكايكي وبيزيان للمعلومات، وحسابات متوسط مربع الخطأ، وتحليل منحنى الخصائص التشغيلية للمستقبل.(ROC-AUC)

عبر كلا إطاري النمذجة، برزت المتغيرات نفسها كتنبؤات ذات مغزى: حيث أظهرت تركيزات الجلوكوز المرتفعة، وقيم مؤشر كتلة الجسم الأعلى، والتاريخ العائلي لمرض السكري، وعدد مرات الحمل، ارتباطات إحصائية قوية بحالة مرض السكري، في المقابل، قدمت مستويات الأنسولين وسمك الجلد قيمة تفسيرية ضئيلة لأي من الانموذجين، عند مقارنة النهجين مباشرة، أظهر نموذج بروبيت تفوقاً طفيفاً في الملاءمة العامة ودقة التصنيف، رغم أن الفرق العملي بينهما كان ضئيلاً، تعزز هذه النتائج فكرة أن طريقتي اللوجستي والبروبيت تسفران عن استنتاجات متشابهة عند تطبيقهما على النتائج الصحية الثنائية، بينما تلفت الانتباه أيضاً إلى العوامل الأيضية والوراثية التي يجب على السريريين وصناع السياسات إعطاؤها الأولوية في برامج فحص السكري وتخطيط الصحة العامة.

معلومات البحث

تواريخ البحث:
التقديم: 17 / 02 / 2026
المراجعة: 09 / 05 / 2026
قبول النشر: 17 / 05 / 2026
نشر الكتروني: 01 / 06 / 2026
تسلسل الصفحات: 00 - 00

الكلمات المفتاحية:
مرض السكري، الانحدار اللوجستي، نموذج بروبيت، الاستجابة الثنائية، التحليل المقار .

المراسلة:
أسم الباحث:
زيوەر عمر إسماعيل

Email:
zewar.ismael@su.edu.krd