Cihan University-Erbil Scientific Journal (CUESJ)

Research Article

The Use of Tobit and Logistic Regression Models to Study Factors that Affect Blood Pressure in Cardiac Patients

Bekhal S. Sedeeq, Banaz W. Y. Meran

Department of Statistics, College of Administration and Economics, Salahaddin University-Erbil, Kurdistan Region - F.R. Iraq

ABSTRACT

This research studies the factors that affect blood pressure in cardiac patients using the Tobit and logistic regression models. The data have been collected, from 500 patients with heart disease in hospital – heart center – Erbil. The two levels of blood pressure, low and high blood pressures, were taken from the patients as dependent variables plus other, independent variables (gender, age, urea, cholesterol, creatinine, and weight). The research shows that the median of blood pressure by means of arterial pressure (MAP) equation contains each of high and low blood pressures differently. This is due to the threshold value of 99.33, equal to,12/8 mmHg, which represent a normal level of a human blood pressure. The aim of this research is to explain the main concepts and processes of Tobit regression analysis (censored and truncated) and logistic regression analysis, which are used for predicting the factors of independent variables that have more effect on the response variables, for example, blood pressureand to compare the outcomes of the two models (Tobit regression and Logistic regression) in order to determine which of the models best fits our data in which AIC and BIC are used. The data analysis of this research shows that the logistic regression model best fits our data, as compared with the Tobit regression model. The data analysis has been achieved using statistical packages in R programming, MATLAB and Statistical Package for the Social Sciences (SPSS) V.26.

Keywords: Akaike information criterion, Bayesian information criterion, censoring, logistic regression model, Tobit regression model, truncation

INTRODUCTION

The health sector is one of the most vital sectors that undertake the task of providing health services to all members of society through health institutions to protect and improve society and achieve the well-being of its members. In fact, one of the components of building health in society is to ward off all diseases, and at the lead of these diseases is cardiovascular disease, which is one of the problems that challenge medicine.

Cardiovascular disease, the leading cause of death worldwide, is greatly exacerbated by high blood pressure. Around,54% of strokes-and 47% of coronary heart disease occur worldwide as a result of high blood pressure. High blood pressure is common medical issue that becomes more prevalent as people become older.[1]

There have been several statistical studies of individual data that include observations in which a dependent variable is equal to 0, for a digit of observations in the dataset. This behavior is known as censored or truncated data.[2]

Since James Tobin's work, the Tobit regression has received a huge and diverse amount of theoretical interest (1958). Its use in practical applications has been developed in fields such as economics, biology, finance, and medicine. Tobit regression can be seen as a linear regression model where only the data, on the dependent variable are incompletely, observed.[3]

Logistic regression is a statistical, method for examining the relationship among a dependent variable of a nominal level and one or other independent variables so that those independent variables are from any type of measurement level.[4]

Logistic regression is investigation of the best method in the case of the binary dependent variable, according to Dayton[5] logistic regression, which is a categorical data statistical modeling method. It is the usages of the same logic as ordinary least squares regression.

This study, aims at:

1. Utilizing both the Tobit and the logistic regression models.

Corresponding Author:
Banaz W. Y. Meran, Department of Statistics, College of Administration and Economics, Salahaddin University-Erbil, Kurdistan Region - F.R. Iraq.
E-mail: banaz.yaqoob@su.edu.krd

Received: July 18, 2022
Accepted: August 12, 2022
Published: November 20, 2022

DOI: 10.24086/cuesj.v6n2y2022.pp133-140

2. Knowing which factors of independent variables is more effecter on the dependent variable (blood pressure) using both models.
3. Comparing the results of the two models to determine which one best fits our data that are using Akaike information criterion (AIC) and Bayesian information criterion (BIC).

LITERATURE REVIEWS

Shirafkan *et al.*[6] estimated the potential of Tobit regression, as a method to study time to onset of cytomegalovirus in renal recipient transplants. The results of this study revealed that the age of patients was influenced by the time of onset of disease. Consequently, the bigger the age, the shorter time required to-start infection.

Karim *et al.,*[7] Tobit model and the traditional regression model were compared on data taken from patients with kidney disease, and through using some statistical measures, it was concluded that Tobit's model is preferable than the traditional regression model for this type of data.

Taleb *et al.*[8] used a binary logistic regression model, estimating the coefficients of this model the least squares method for people with heart disease. The aim of the investigation was to compare the actual reasons of death with the estimated causes of death. The binary logistic regression model concluded that smoking was the leading cause of death.

Rambeli *et al.*[9] used the binary logistic model for a study the aim of which was to identify the factors that influence a teacher's decision to remain in their professional life. The result showed that the income was considered as a key factor in teacher remaining committed to the profession.

METHODOLOGY

This research contains three parts: The first part of this research included the introduction, methodology, aim of the study, and review of the literature. The second part of the research included the basic concepts of the two models and the last part contains data applications, implementation of two models, and interpretation of the results.

Tobit Regression Model

Regression analysis is one of the most important ways to find out the significant effect between the dependent variable and explanatory, variables but sometimes, the dependent variable is constrained to the threshold point. In this situation, the use of the traditional regression model is biased. With this kind of data, using the Tobit regression model is the best option. Logistic-regression, it is used as in this, study. The analysis of the Tobit regression means the, statistical method used to examine the relationship between the limited dependent variable and explanatory variables of any type. The analysis in this situation is named a Tobit regression. A limited dependent variable is the one whose range of potential values is constrained in some significant way.[10] Limited dependent variable models contain: a – Censoring, in which some data are lost but other data are present for certain persons in a data collection and b – truncation, in which some individuals are purposefully removed from observation.[11]

Tobin used a regression model based on household expenditures that particularly took into account the fact that (his regression model's dependent variable) cannot be negative. Tobin coined the phrase "model of constrained dependent variables" to describe his approach, a term invented by Goldberger (1964) for the reason that they resemble Probit models. These models are frequently denoted to as truncated or censored models. If the observations are lost outside of a carefully defined range and are censored, the model is called a truncated model.[12] The structural Tobit model

$$y^* = X_i\beta + \mathrm{e}_i$$

Where, $e_i \sim N(0, \sigma^2)$. y* is a latent variable that is observed y, shown by the following equation $y_i = \begin{cases} y^* & if \quad y^* > \tau \\ \tau_y & if \quad y^* \leq \tau \end{cases}$

When assume that $\tau = 0$ in the standard Tobit model, data are censored next to a value of 0. For use there is:[13]

$$y_i = \begin{cases} y^* & if \quad y^* > 0 \\ 0 & if \quad y^* \leq 0 \end{cases}$$

As it has been shown, the probability function is censored

$$L = \prod_i^N \left[\frac{1}{\sigma}\phi\left(\frac{y-\mu}{\sigma}\right)\right]^{di}\left[1-\Phi\left(\frac{\mu-\tau}{\sigma}\right)\right]^{1-di}$$

In the conventional Tobit model, one groups $\tau = 0$ and parameterize μ as $(X_i\beta)$. That is Tobit model's likelihood function:

$$L = \prod_i^N \left[\frac{1}{\sigma}\phi\left(\frac{y_i - X_i\beta}{\sigma}\right)\right]^{di}\left[1-\Phi\left(\frac{X_i\beta}{\sigma}\right)\right]^{1-di}$$

The Tobit model's log likelihood function is

$$\ln L = \sum_{i=1}^N \left\{ \begin{array}{l} d_i\left(-\ln\sigma + \ln\phi\left(\frac{y_i - X_i\beta}{\sigma}\right)\right) + \\ (1-d_i)\ln\left(1-\Phi\left(\frac{X_i\beta}{\sigma}\right)\right) \end{array} \right\}$$

The overall log, likelihood is split into two components. The first component corresponds to a traditional regression for an uncensored observation, while the, second component represents the associated likelihood of censoring the test.[14]

Standard Tobit model

Censored regression method has been very widespread as a standard Tobit model since Tobin (1958) first presented it to evaluate the relationship between household expenditure and household income.

$$y^* = x_i'\beta + u_i \qquad i = 1, 2, \ldots . n$$

$$y = y^* \qquad \text{if} \qquad y^* > n$$

$$y = 0 \qquad \text{if} \qquad y^* \leq n$$

Where, u_i is identical independent distribution (iid). Drawings from $N(0, \sigma_u)$.

y_i and X_i are observed in the sample, but the y_i^* is unobserved if $y_{2i}^* < 0$. The likelihood function for this model is:[12]

Table 1: Five hundred patients sample

ID	Y: BP	X1: Gender	X2: Age	X3: Urea	X4: Cholesterol	X5: Creatinine	X6: Weight
1	88 (0)	1	70	68	127	1.26	89
2	76 (0)	2	40	41	117	0.97	96
3	99.33	1	44	34	201	0.59	74
4	85 (0)	2	43	33	221	0.67	107
5	117	1	55	54	226	1.12	88
6	92 (0)	1	67	31	173	0.72	76
7	96.67	1	47	43	165	0.78	65
497	107.67	1	54	31	160	0.87	85
498	93.67	2	63	16	109	0.6	69
499	106.67	1	58	25	100	0.73	67
500	118.33	1	69	38	134	1.11	65

BP: Blood pressure

$$L=\prod_{0}\left[1-\Phi\left(\frac{x_i'\beta}{\sigma}\right)\right]\prod_{1}\frac{1}{\sigma}\Phi\left[\frac{y_i-x_i'\beta}{\sigma}\right]$$

Censoring

A regression, model is said to be-censored when the recorded, data on the dependent variable (the response) cutoff outside a certain-range with multiple observations at the-endpoints of that range. When the data are censored, variation in the observed dependent variable will underestimate the impact of the regression on the actual dependent variable. As a result, coefficient estimates from standard ordinary least squares regression employing censored data are often biased toward zero.[15]

Truncation

Truncated data, or missing data, are discovered when an observation is not reported despite of whether it is below or above a specific level that differs. In actuality, these are known as left and right truncation, respectively. The truncation effect can also happen when only a small portion of a larger population is represented in the sample data. In addition, the response variable in the model is truncated if observations are not possible while taking values inside a particular range. Consequently, when the dependent variable is within that range, neither the dependent nor the independent variables are observed.[2]

Logistic Regression Model

What distinguishes a logistic regression model from the, linear regression model, in logistic regression, could be the result variable which is either binary or, dichotomous. This difference among logistic and linear regression is reflected both in the choice of a parametric model and it assumptions, whereas the methods used in a logistic, regression study follow the same fundamental principles, as linear regression.[16,17]

Logistic regression model circuitously models the response variable created on probabilities linked with the digits of the dependent variable *y*. We will use *P*(*X*) to represent the possibility of a response when $y = 1$. Furthermore, we will define, 1–*P*(*X*) which represents the possibility of a response when $y = 0$. These probabilities are written as follows:

$$P(X)=P(Y=1|X_1,X_2,\ldots X_k)$$

$$1-P(X)=P(Y=0|X_1,X_2,\ldots X_k)$$

The logistic regression equation and the regression equation with a straight line, equation ($Y=\alpha+\beta_1X_1+\beta_2X_{2+}\ldots.+\beta_kX_k$) are related by the formula below. The logistic regression equation form is rewritten as such:

$$Logit\,P(X)=\log_e\left[\frac{P(X)}{(1-P(X))}\right]=\ \alpha+\beta_1X_1+\beta_2X_{2+}\ldots.+\beta_kX_k$$

The logistic regression model can be calculated using the formula below.[17,18]

$$P(X)=\frac{e^{(\alpha+\beta_1X_1+\beta_2X_2+\ldots+\beta_kX_k)}}{1+e^{(\alpha+\beta_1X_1+\beta_2X_2+\ldots.+\beta_kX_k)}}$$

Estimate the Parameters of Logistic Regression Model

The approach of maximum likelihood estimation will be used. The log likelihood is given as:[19-21]

$$L(\alpha,\ \beta)=\prod_{i=1}^{n}P(X)^{y_i}\,[1-P(X)]^{1-y_i}$$

We'll use the log likelihood method for estimation:

$$logL(\alpha\ \ ,\beta)=\sum_{i=1}^{n}y_i logP(X)+\left[\left(n-\sum_{i=1}^{n}y_i\right)\log(1-P(X)\right]$$

$$U(P)=\frac{\partial L(\alpha,\ \beta)}{\partial P}=\sum_{i=1}^{n}y_i/P(X)-\left[\left(n-\sum_{i=1}^{n}y_i\right)/(1-P(X))\right]$$

Cox and Snell R^2 statistic

In the logistic regression model, the determination coefficient R^2 used to determine the fittingness of the proposed regression,

models for the study data is changed by the R^2 Nagelkerke, Cox, and Snell R^2 computation statistics.

Can be calculated as: $R^2 \text{Cox and Snell} = 1 - \left[\frac{L_0}{L_1}\right]^{(2/n)}$

L_0: Maximum likelihood for constant in the model.
L_1: Maximum likelihood independent variables in the model
n: Sample size[22]

And, can be calculated as:

$$R^2 Nagelkerke = \frac{R^2 cox \ \& snell}{1 - L_0^{(2/n)}}$$

The Hosmer-Lemeshow test

The Hosmer and Lemeshow test is a widely-used test for determining a model's quality of fit. It accepts any number of explanatory-variables, which can be continuous or categorical. It is used to test the hypothesis:

H_0: The model is adequate for data
H_1: The model is not adequate for data

The model will be a useful model if (Hosmer and Lemeshow) static is >0.5.[22]

$$HL\chi^2_{k-2} = \sum_{i=1}^{n} \frac{(o_i - E_i)^2}{N_i \pi_i (1 - \pi_i)}$$

Wald statistic

The Wald test is used to determine whether or not the effect of the logistic regression coefficient on the independent variables.[23]

The Wald statistic is calculated according to the following formula:

$$W^2 = \left[\frac{\beta}{S.E_\beta}\right]^2$$

The classification table

The use of classification tables is one of the ways to check the quality of matching the model with the data. This method depends on creating tables that put the number of cases that have the desired trait or the cases that do not have the desired trait and that were categorized correctly or incorrectly. The concept behind using the analysis is it leading to better results if the model is data compatible.[24]

The AIC

AIC as a model selection criterion for assessing actual data, it has played a key role in solving issues in a wide range of fields, and the model by the lowermost AIC is chosen as the best model.[25]

$$AIC = -2 \log L + 2 * K$$

The BIC

In statistical model selection, the BIC is unique of the best well-known and commonly used tools. BIC is calculated for each of the models, and the model with the lower BIC value is selected as the best model.[26]

$$BIC = -2 \log L + 2 * \log N * K$$

DATA ANALYSIS

In this study, the two models of Tobit and logistic regression models have been applied on a sample taken from 500 patients with heart disease and two levels of blood pressure, high and low blood pressure, in hospital – heart center – Erbil. Blood pressure is taken from the patients as response variables and some independent-variables (gender, urea, age, cholesterol, creatinine, and weight). The study found that the average of blood pressure by means of arterial pressure (MAP) equation contains each high and low blood pressure differently because the threshold point was determined to be 99.33. To take the best model for our data in the study, two statistical measures (AIC and BIC) were used.

Note: All assumptions and tests related to Tobit and logistic regression models have been applied before we started the data analyses in this study.

Data Description for Tobit Regression Analysis

In this study, the data are gathered from 500 patients with heart diseases, and the two levels of blood pressure; high and low blood pressure were taken from patients as dependent variables and the variables: Gender, age urea, cholesterol, creatinine, and weight as the independent variables (Table 1). The researcher set that the medium of blood pressure by MAP equation contains each highest and lowest blood pressure differently because the threshold point was determined to be 93.33,

In regards, the dependent variable in this study has been defined as:

$Y=Y^*$ $Y^*>93.33$

Appropriately, the limitation threshold point Y = 93.33 was regarded the limitation threshold point y = 0, in accordance with the stated theoretical presentation, and the model will be referred to as the following model:

$Y=Y^*$ $Y^*>93.33$

$Y=0$ $Y^*<=93.33$

The explanatory variables in this study are the follows:
X_1: Gender (male and female)
X_2: Age measured by year
X_3: Urea (mg/dL)
X_4: Cholesterol (mg/dL)
X_5: Creatinine (mg/dL)
X_6: Weight measured by kg

The explanation of the variables is presented in Table 2 in which contains independent variables and dependent variable and the table views the maximum, minimum, mean, and standard deviation of data.

Application of Tobit regression model analysis (Censored and Truncated)

Because of the researcher initially checked all the necessary assumptions that must be present in the data before starting to analyze the data, and also set a unified standard for the data in this study, it becomes clear for us that there is no problem in terms of our data and we can use the data for analyzing.

Censored regression model

First, let's start with censored-regression model are: Censored (formula = Y~ X, -left = 0, right = Inf., data= data 1) Total (n = 500 observation, left censored = 132 observation, Uncensored = 368 observation, left censored (Y < 99.3 then Y* = 0: Observation).

Table 3 presents the results of the exact regression model. The coefficients of the independent variables sex, age, cholesterol, creatinine, and weight are positive because the variables have a positive relationship with the dependent variable (blood pressure), while the coefficient of the independent variable urea is negative because the variable has a negative relationship with the dependent variable (blood pressure). According to the findings in Table 3, the variables age, cholesterol, and creatinine all significantly affect blood pressure. The summarized fit to the censored regression model is log-likelihood = −2125,549, AIC = 4265.1, BIC = 4294.6. The score for the best model is determined by the lowest value for AIC and BIC.

Truncated regression model

In this case, the number of observations turns to 368 due to a truncation.

Table 4 shows the results from truncated regression model. Those coefficients of the independent variables such as gender, age, cholesterol, urea, and weight are positive for the reason that the variables take a positive relationship with the dependent variable (blood pressure) whereas the coefficient of the independent variable creatinine is negative because has a negative connection with the dependent variable (blood pressure). It is through the conclusion in Table 4 that only the urea variable has the effect on blood pressure.

Logistic Regression Analyses

In this part, we use binary logistic regression since the dependent variable in this study is blood pressure and the researcher has taken the average of blood pressure by MAP equation which contains each of the high and low blood pressure differently because of the threshold points which was determined to be 93.33. The patient whose blood pressure is >93.33 is considered to be infected and takes the worth of one while the patient whose blood pressure is ≤93.33 is considered to be uninfected and takes the value of 0, and the other variables are gender, age, cholesterol, urea, creatinine, and weight.

Y: The dependent variable has a binary response code:

$$Y = \begin{cases} 1 & \textit{Infected} \\ 0 & \textit{Uninfected} \end{cases}$$

The independent variables are the same as the variables written above.

Application of binary logistic regression analysis

Let's start with the outcome of the classification table starting with the zero stage in which the model is free of independent variables (only the constant).

Table 5 represents the baseline model, which is a model without our explanatory variables. The overall right percentage was 73.6% which refers the model's overall explanatory strength. The initial log likelihood function (-2 log likelihood function)= 577.2.

Omnibus test of logistic model coefficients

Based on the model coefficients in omnibus tests, we find that the Chi-square tests are to illustrate if there is an important variance between the factors of the nil model and the current model.

Table 2: Descriptive statistics of variable

Variables	Minimum	Maximum	Mean	SD
BP	70.33	141.00	101.23	10.97
Gender	1	2	1.33	0.47
Age	17	86	57.73	11.24
Urea	11	198	40.11	20.53
Cholesterol	63	326	167.36	45.02
Creatinine	0.11	10.70	1.58	1.195
Weight	48	135	78.83	13.47

BP: Blood pressure, SD: Standard deviation

Table 3: Censored regression model

Coefficients	Estimate	SE	*t*	Pr (>*t*)
Constant	−127.94562	25.77314	−4.964	6.89e-07***
Gender	3.57944	5.59007	0.640	0.522
Age	1.91136	0.24742	7.725	1.12e-14***
Cholesterol	0.25688	0.05688	4.516	6.30e-06***
Urea	−0.11037	0.12724	−0.867	0.386
Creatinine	12.21152	2.09200	5.837	5.31e-09***
Weight	0.31290	0.19800	1.580	0.114

*** mean P ≤ 0.001

Table 4: Truncated regression model

Coefficients	Estimate	SE	*t*	Pr (>t)
Constant	98.7995674	4.7143913	20.9570	<2e-16***
Gender	0.6703624	0.9215182	0.7275	0.46695
Age	0.0645897	0.0448466	1.4402	0.14980
Cholesterol	0.0015114	0.0094141	0.1605	0.87245
Urea	0.0520044	0.0202811	2.5642	0.01034*
Creatinine	−0.2198352	0.3154896	−0.6968	0.48592
Weight	0.0040456	0.0340886	0.1187	0.90553

SE: Standard error; * mean P ≤ 0.05 and *** mean P ≤ 0.001

Table 5: Classification table shows that the model has a constant bound zero step

Observed	Predicted		
	Y		Percentage correct
	Uninfected	Infected	
Step 0			
Y			
Uninfected	0	132	0.0
Infected	0	368	100.0
Overall percentage			73.6

Table 6 displays that the current model is meaningfully more suitable than the null model. Model coefficients omnibus test offers significant reduction in the −2 log likelihood value = 426.572= as compared to −2 log likelihood value =577.2 of the null model. This indicates that the Chi-square values for step, block, and model are all the same. $P < 0.05$ illustrating that the model's accuracy increases when explanatory variables are included. Furthermore, the Chi-square is significant ($\chi^2 = 150.628$, df = 6, $P < 0.05$). As a result, our new model is bested.

Hosmer and Lemeshow test

Being a goodness of fit test for logistic regression, it fits the data model.

Table 7 explains since $P = 0.226$, it is larger than the level of significance at 5%. We may well conclude that the data are will be suitable to the model.

Cox and Snell R^2, Nagelkerke's R^2

The Cox and Snell R^2, Nagelkerke's R^2 values are utilized to estimate the model's fit to the data.

As the Nagelkerke's R^2, Cox and Snell R^2 values given in Table 8 are examined, the ratios of interpretation of the independent variables over the dependent variable are shown. The value of Nagelkerke R^2 is the modified form for the Cox and Snell R^2 coefficient. Giving to the outcomes shown in Table 8, it is seen that the dependent variables determine 26% of the variance in the independent variables according to the value of Cox and Snell R^2, and 38% according to the value of Nagelkerke R^2. The amount of the −2 log likelihood statistic is 426.572 in model summary for the whole model.

Table 9 is corresponding to Table 5 but it is based on the model that contains our explanatory variables. The total percentage of correct was 79.2% which replicates the model's overall explanatory strength. Classification table contains the constant term and the rest of the predictors, that is, 91.6% were correct for the infected of blood pressure and correctly classified. This table shows how many events were correctly predicted (59 cases were observed to be uninfected and were correctly predicted to be uninfected; 337 cases were observed to be infected and were correctly predicted to be infected) and how many were not (73 cases are observed to be uninfected but are predicted to be infected; 31 cases are observed to be infected but are predicted to be uninfected).

Table 6: Omnibus test

Step 1	χ^2	df	Significant
Step	150.628	6	0.000
Block	150.628	6	0.000
Model	150.628	6	0.000

Table 7: Hosmer and Lemeshow test

Step	χ^2	df	Significant
1	10.595	8	0.226

Table 8: Summary of logistic regression

Step	−2 Log likelihood	Cox and Snell R^2	Nagelkerke R^2
1	426.572a	0.260	0.380

Variables in the equation logistic-regression model

The variables in the equation logistic regression are the most essential of all the outputs. This table must be studied carefully since it contains the answers to our questions about the common relationship between all variables.

Table 10 shows the values of Wald test that represents the parameter of the test value of the model and it appears that the variables (age, urea cholesterol, and creatinine) represent the significant variables in the research. It is by associating the *P*-value with the level of significant (0.05) that the p-value represents the significance of the effect of the variable on the patient condition, that is, it is significant when $P < 0.05$ was considered for the variable under test. It shows that the variables (gender and weight) are not significant variables in the study, and it is through comparing the *P*-value with the level of significant (0.05) that the *P*-value represents the non-significance of the effect of the variables on the patient condition, that is, it is no significant when $P > 0.05$ was considered.

As far as Table 10 is concerned, the logistic computation coefficients that may be utilized to build a predictive equation could be:

$$Y = \log\left(\frac{p(x)}{1-p(x)}\right) = \alpha + \beta_1 x_1 + \beta_2 x_2 \ldots + \beta_k x_k$$

The effect factors for blood pressure in cardiac patients can be ranked as follows based on the value of the odds ratio. Likewise, we can write the logistic regression computation with just significant variables:

Logistic equation (Model): $Y = -7.689 + 0.080$ Age - 0.016 Urea + 0.013 Cholesterol + 1.451 Creatinine

Table 10 shows, $\text{Exp}(\beta) = e^b$ represents the ratio changed in the odds of the event of importance for a one unit variation in the predictor.

The value of $\text{Exp}(\beta)$ for the variable gender indicates that when the gender changes from the value 0 (female) to the value 1 (male), the probability of disease blood pressure in patients with heart disease increases because the value of $\text{Exp}(\beta)$ is >1. Such incomes indicate that the blood pressure of males is higher than females.

Table 9: Classification tables

Observed	Predicted		
	Y		Percentage correct
	Uninfected	Infected	
Step 1			
Y			
Uninfected	59	73	44.7
Infected	31	337	91.6
Overall percentage			79.2

Table 10: Variables in the equation

Step 1	β	SE	Wald	df	Significant	Exp(β)	95% CI for Exp(β)	
							Lower	Upper
Gender	0.108	0.264	0.168	1	0.682	1.114	0.664	1.868
Age	0.080	0.012	41.674	1	0.000***	1.083	1.057	1.110
Urea	−0.016	0.007	4.692	1	0.030*	0.984	0.970	0.998
Cholesterol	0.013	0.003	21.265	1	0.000***	1.013	1.007	1.019
Creatinine	1.451	0.276	27.595	1	0.000***	4.267	2.483	7.333
Weight	0.011	0.009	1.327	1	0.249	1.011	0.992	1.030
Constant	−7.689	1.294	35.304	1	0.000	0.000		

CI: Confidence interval, SE: Standard error, * mean P ≤ 0.05 and *** mean P ≤ 0.001

Table 11: Comparison

Tobit regression model (censored)	Truncated regression model	Logistic regression model
AIC=4265.1	AIC=2574.6	AIC=591.2
BIC=4294.6	BIC=2602.0	BIC=620.7

AIC: Akaike information criterion, BIC: Bayesian information criterion

The odds ratio for the variable age is >1, Exp(β) = 1.083. This means that each additional rise of 1 year in age is in touch with the increase in the odds infected of blood pressure in cardiac patients.

The odds ratio for the variable urea is <1. This means that each additional increase of one unit in urea is associated with decrease in the odds infection of blood pressure in cardiac patients with 0.984 times.

The odds ratio for the variable cholesterol is >1. This means that each additional increase of one unit in cholesterol is related to the increase in the odds of blood pressure in cardiac patients with 1.013 times.

The odds ratio for the variable creatinine is >1. This means that each additional increase of one unit in creatinine is related to the increase in the odds infection of blood pressure in cardiac patients with 4.267 times.

The odds ratio for the variable weight is >1. This means that each additional increase of one unit in weight is associated-with the increase in the odds infection of blood pressure in cardiac patients with 1.011 times.

DISCUSSION

There are different techniques for comparing the analysis of two or more models; however, the AIC and BIC criteria are two that may be worth considering.

Table 11 shows a comparison between three regression models censored, truncated, and logistic, for choosing the most fit model to our data of blood pressure in cardiac patients, the AIC and BIC values with the least values are chosen. The results display that the logistic regression model is better and more suitable rather than truncated regression and censored regression for our data, because it's AIC equal to 591.2 and BIC = 620.7 are the lowest values contrast with Tobit models (censored and truncated).

CONCLUSION

It has been concluded the following:

1. In the censored, regression model, the explanatory, variables (age, cholesterol, and urea) significantly impacted on blood pressure.
2. The results, from truncated regression model show, that only the urea variable has the effect on blood pressure.
3. According, to the results classification table, logistic model is, correctly classifying the consequences for 79.2% of the cases, compared to 73.6% in the null, model.
4. According to the Hosmer-Lemeshow test, our data fit the logistic regression based on a χ^2 = 10.595 and *P*-value greater, than a significant level.
5. Wald's test showed that the variables of age, creatinine, cholesterol, and urea, respectively, contributed significantly to the prediction, depending on the *P*-value (0.000 < 0.005). The variables that do not have a significant, effect are weight and gender.
6. It was concluded, that the logistic regression model for the sample under study or for our data is better, than the censored regression model and the truncated regression model after comparing, their AIC and BIC values.

REFERENCES

1. C. Y. Wu, H. Y. Hu, Y. J. Chou, N. Huang, Y. Chou and C. Li. High blood pressure and all-cause and cardiovascular disease mortalities in community-dwelling older adults. *Medicine(Baltimore)*, vol. 94, no. 47, p. e2160, 2015.
2. W. H. Greene. Censored Data and Truncated Distributions, *SSRN Electron. J,* 2005.
3. M. H. Odah, B. K. Mohammed and A. S. M. Bager. Tobit regression model to determine the dividend yield in Iraq. *LUMEN Proceedings*, vol. 3, pp. 347-354, 2018.
4. J. S. Cramer. *The Origins of Logistic Regression*: Tinbergen Institute Discussion Papers, 2002.
5. C. M. Dayton. Logistic regression analysis. *Stat*, vol. 474, p. 574, 1992.
6. H. Shirafkan, J. Yazdani-Charati, S. A. Mozaffarpur, S. Khafri, R. Akbari and A. A. Pasha. Application of tobit model in time until *Cytomegalovirus* infection in kidney transplant recipients. *Acta Medica*, vol. 32, p. 1237, 2016.
7. R. M. H. Karim and S. M. Salh. Using tobit model for studying factors affecting blood pressure in patients with renal failure. *UHD Journal of Science and Technology*, vol. 4, no. 2, pp. 1-9, 2020.
8. H. R. Talib and S. A. Mazloum. The use of binary logistic

regression method to analyze the factors affecting heart disease deaths: An applied study on a sample of patients in Dhi Qar Governorate. *Journal of Al-Rafidain University College College for Sciences*, 2020, no. 46, 2020.
9. N. Rambeli, E. Hashim, F. C. Leh, N. S. Hudin, M. F. Ramli, M. C. Mustafa, *et al.* Decision to leave or remain in the career as early childhood educator: A binary logistic regression model. *Review of International Geographical Education Online*, vol. 11, no. 5, pp. 450-456, 2021.
10. G. S. Maddala. *Limited-Dependent and Qualitative Variables in Econometrics*. New York, NY: Cambridge University, 1983.
11. N. M. Ahmed. Limited dependent variable modelling (truncated and censored regression models) with application. *The Scientific Journal of Cihan University Sulaimanyia*, vol. 2, no. 2, pp. 82-96, 2018.
12. T. Amemiya. Tobit models: A survey. *Journal of Econometrics*, vol. 24, no. (1-2), pp. 3-61, 1984.
13. J. S. Long, Regression Models for Categorical and Limited Dependent Variables. Thousand Oaks, CA: Saga Publication, 1997.
14. A. Flaih, J. Guardiola, H. Elsalloukh and C. Akmyradov. Statistical inference on the ESEP tobit regression model. *Journal of Statistics Applications and Probability Letters,* vol. 6, no. 1, pp. 1-9, 2019.
15. K. Y. Chay and J. L. Powell. Semiparametric censored regression models. *Journal of Economic Perspectives*, vol. 15, no. 4, pp. 29-42, 2001.
16. D. Hosmer Jr., S. Lemeshow and R. Sturdivant. Applied Logistic Regression. vol. 398. New York: John Wiley and Sons, 2013.
17. A. M. Khudhur and D. H. Kadir. An application of logistic regression modeling to predict risk factors for bypass graft diagnosis in Erbil. *Cihan University-Erbil Scientific Journal*, vol. 6, no. 1, pp. 57-63, 2022.
18. N. M. M. Abd Elsalam. Binary logistic regression to identify the risk factors of eye glaucoma. *International Journal of Sciences Basic and Applied Research*, vol. 23, no. 1, pp. 366-376, 2015.
19. N. S. K. Barznj. Using logistic regression analysis and linear discriminant analysis to identify the risk factors of diabetes. *Zanco Journal of Humanity Sciences*, vol. 22, no. 6, pp. 248-268, 2018.
20. S. Menard. *Applied Logistic Regression Analysis.* vol. 106, New York: Sage, 2002.
21. N. H. Mahmood, R. O. Yahya and S. J. Aziz. Apply binary logistic regression model to recognize the risk factors of diabetes through measuring glycated hemoglobin levels. *Cihan University-Erbil Scientific Journal*, vol. 6, no. 1, pp. 7-11, 2022.
22. V. Bewick, L. Cheek and J. Ball. Statistics review 14: Logistic regression. *Critical Care*, vol. 9, no. 1, pp.112-118, 2005.
23. D. Hosmer and S. Lemeshow. Applied Logistic Regression. New York: Johnson Wiley and Sons, 2000.
24. I. R. Soderstrom and D. W. Leitner. *The effects of base rate, selection ratio, sample size, and reliability of predictors on predictive efficiency indices associated with logistic regression models*, 1997.
25. S. Konishi and G. Kitagawa. *Information Criteria and Statistical Modeling*. Berlin: Springer Science and Business Media, 2008.
26. K. I. Mawlood. Using logistic regression and cox regression models to studying the most prognostic factors for leukemia patients. *Qalaai Zanist Scientific Journal*, vol. 4, no. 3, pp. 705-724, 2019.