Arab American University

Faculty of Graduate Studies

Predicting the risk of Osteoporosis in patients with Asthma and Chronic Obstructive Pulmonary disease in Palestine

By
Asala Jawad Fau'd Aljammal

Supervisor
Dr. Saleh Afaneh

This Thesis was submitted in Partial Fulfillment of the Requirements for the Master's Degree in Applied Mathematics

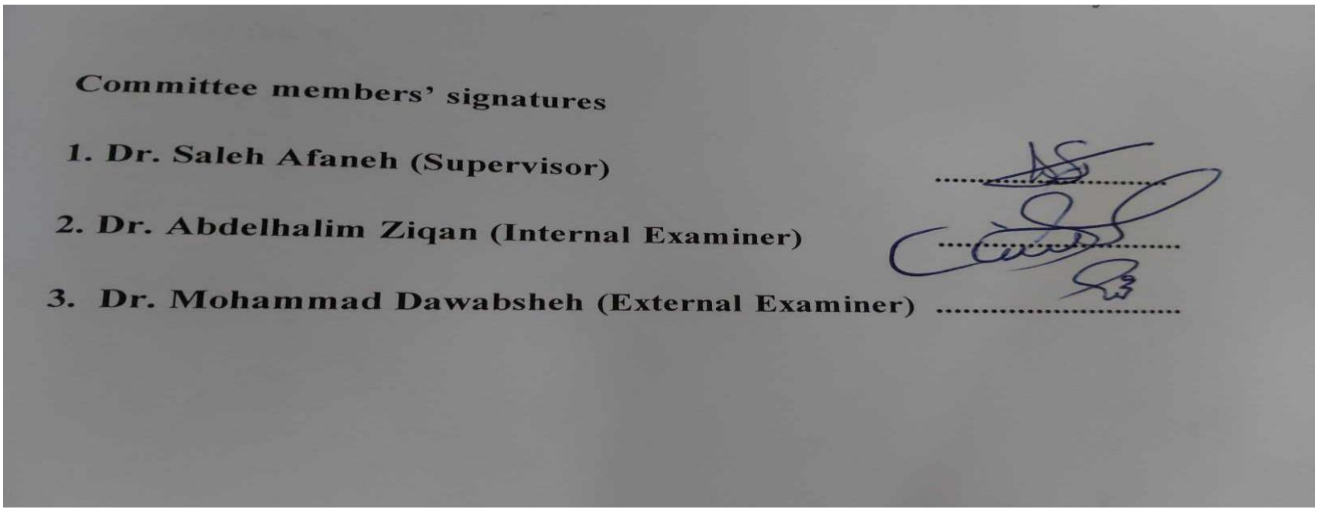
February/2023

Predicting the risk of Osteoporosis in patients with Asthma and Chronic Obstructive Pulmonary disease in Palestine

By

Asala Jawad Fau'd Aljammal

This thesis was defended successfully on February 2023 and approved by:



Committee members' signatures

1. Dr. Saleh Afaneh (Supervisor) ....
2. Dr. Abdelhalim Ziqan (Internal Examiner) ....
3. Dr. Mohammad Dawabsheh (External Examiner) ....

II

Declaration

I, **Asala Jawad Fau'd Aljammal**, one of the students of the Faculty of Graduate Studies at the Arab American University hereby declare that this thesisentitled **"Predicting the risk of Osteoporosis in patients with Asthma and Chronic Obstructive Pulmonary disease in Palestine"**, is all by my work and the resources that are used in this thesis(including the internet resources) have been referred to andproperly acknowledged as required.

I declare that I have fully understood the concept of plagiarism and I acknowledge that my thesis will be immediately rejected in case of including any type of plagiarism.

Asala Jawad Fau'd Aljammal 201312827

Signature: Asala

Date: 18/9/2023

Dedication

My husband is the main supporter in completing this thesis, my father who encouraged me to complete postgraduate studies and the prayers of my mother and brothers and sisters who supported me throughout my studies and Dr. Saleh Afaneh who gave a helping hand.

Acknowledgements

In the first place, I thank God who enabled me to accomplish this work. Many thanks to my father, who encouraged me to study for a master's degree, and to my dear husband, who encouraged and support me throughout my study and for the great help in completing this work.

I will never forgot, my supervisor, Dr.Saleh Afaneh for his generosity, guidance and valuable that enabled me to accomplish this thesis.

Abstract

In this thesis we investigated the risk of osteoporosis in patients in Jenin city, and its relationship with some parameters of patients intected with asthma and chronic obstructive pulmonary disease. For that purpose, we analyzed the data using the SPSS program and applied different methods to study the different outcomes and the characteristics of the variables.

Table of Contents

List of Tables

List of Figures

Introduction

Mathematics is of great importance in many important areas, and among these areas in which it contributed significantly is human life and health, as through its equation and laws we can predict the occurrence of disease and thus limit its spread. One of the most important topics in biostatistics is logistic regression because it predicts the probability of an event occurring. Which can be used to analyze available data to explain the various factors and outcomes of an event.

We have noticed the prevalence of osteoporosis, especially in patients with asthma and chronic obstructive pulmonary disease, so in this research, we will study the probability of developing osteoporosis in patients with asthma and chronic obstructive pulmonary disease in Nablus using logistic regression.

For this purpose, real data were collected from a thoracic clinic, which may have an impact on the subject of the study, such as age, gender, and other factors, and they were extracted and analyzed using SPSS, and reach to the results.

1. Background and related work

Logistic regression is used to obtain the Odds ratio for the presence of more than one independent variable. It is similar to multiple linear regression, except that the dependent variable in binary logistic regression is a binary variable that carries success, failure, injury, or lack thereof. (Sperande, 2013)

It is therefore difficult to use linear regression (simple or multiple) in health studies Social and economic because it requires that the dependent variable be continuous and quantitative, while the dependent variable in those studies is often descriptive separate. Therefore, logistic regression is used in such studies. (Abbas, 2012)

An example of using the binary logistic regression method was used to study the most important financial indicators influencing the performance of stocks in the Kuwait Stock Exchange, and the results showed the importance of the binary logistic regression model in predicting stock performance.

This model is characterized by the fact that it does not require that the relationship between it and the dependent variable be linear, nor does it require that the independent variables follow a normal distribution (Farhud, 2014).

Logistic regression is also used in medical research because it can predict outcomes and measure correlation. Logistic regression is a powerful method for analyzing the effect of a group of independent variables on a dependent binary variable by identifying and studying the effect of each independent variable (Stoltzfus, 2011).

Abbas 2012 mentioned that the use of logistic regression model in forecasting is important for countries with dependent qualitative economic variables, because the method of the least squares cannot be used in estimating regression models with dependent qualitative variables.

The objective of his dissertation was the practical application of the logistic regression model qualitative dependent economic variables, and the study of the suitability of the logistic regression model.

However, logistic regression is commonly used in health and medicine by identifying factors that increase susceptibility to disease. Logistic regression is used to predict the probability of an event occurring. The model uses several independent variables and only one dependent variable. The dependent variable is a nominal variable for binary if binary logistic regression, or a classification variable for multiple logistic regression (Sari,Dash, 2017).

logistic regression model assumes that the dependent variable is qualitative and follows a Bernoulli distribution with a value of 1 (event p occurring) and a value of 0 (non-occurrence of the 1-p event). least squares method cannot be used in the case of logistic regression, so the most likely method is to estimate the regression coefficients Logistics to find equations and solve them numerically (Kazina, Ashwash, 2017).

Abu Doma, Othman, and Muhammad (2018) used the logistic model to discuss the problem of heart disease by identifying the influencing factors using the model. Determines whether a person suffers from heart disease or does not suffer from heart disease. He used several independent factors, including: age, sex, smoking and blood pressure to predict the risk of heart disease. The result was that there is a strong relationship between blood pressure and heart disease.

Haider (2011) stated that logistic regression is more common for predicting disease susceptibility than using logistic regression to classify and predict obesity. The result was a binary variable: obesity or normality. In this case, statistical significance was determined by the P-value using the Wald test.

The researcher Babiker 2020 used binary logistic regression in his research to identify the most important economic and social factors that lead to the death of women during pregnancy and childbirth in Khartoum in 2018. It was found that logistic regression has a higher classification ability and great importance in distinguishing between deceased and non-deceased women based on the death factors that were studied to get it.

Al-Afifi 2010 also mentioned that the logistic regression model is a special case of the generalized linear model (GLM).

One study showed the prevalence of osteoporosis among chronic obstructive pulmonary patients. The prevalence of osteoporosis using cortisone therapy is associated with osteoporosis in this population. Osteoporosis was also assessed through bone density estimates and x-rays. 68% of the participants had osteoporosis, but cortisone use alone could not explain the increased prevalence of osteoporosis. (Jørgensen and others,2006) By studying vertebral fractures in steroid-dependent asthma and osteoporosis, reduced bone mass leads to vertebral fractures. The study showed that vertebral fractures occur in patients with asthma who are treated with steroids continuously. (Luengo and others,1991)

One study indicated that asthma is related to osteoporosis, as it divided asthma patients into three groups:

Patients who did not require treatment (controlled), Patients with continued treatment. Patients who did not receive treatment despite the appearance of symptoms. The study used logistic regression. prevalence of osteoporosis was higher in asthmatic patients than in the control group. study showed that asthma was associated with osteoporosis in the adult Korean population.(Hye Wee and others,2020)

2. Main objectives of the thesis

- Application of the concept of logistic regression to the data of asthma and Chronic obstructive pulmonary disease (COPD).

- Studying the effect of some selected and assumed (independent) variables on the dependent variable.

- Access to a model that helps in predicting the incidence of disease and knowing the suitability of the model for prediction.

- The study hypotheses are:

$H_0 = 0$: There is no effect of the independent variables.

$H_0 \neq 0$: There is an effect of independent variables.

3. Thesis abstract

There are three main chapters in this research, The first chapter presents generalized linear regression with a brief definition, then in the second chapter we talk about the concept of binary and multiple logistic regression and the logistic regression model and test the suitability of the model where we will expand on the binary logistic regression as the subject of the study, in addition to testing the importance of the model parameters, as well as the definitions and concepts that have been They are used in this research and any other mathematical laws that can be used in analyzing data, such as hypothesis testing.

The third chapter explains the practical (applied) aspect of this research through the data that we rely on in the analysis and the results and conclusions that were reached by analyzing the data collected through the thoracic doctor's clinic. We also identify all the variables that were included in the data file, and then display the data analysis results including SPSS tables and properties of the different variables.

Finally, we summarize the study and its main conclusions and suggest future work to which this thesis may be associated.

Chapter 1

Parametric regression model

1.1 Regression

Definition: Regression analysis includes statistical methods to infer the relationship between variables.

It is an analysis in which the dependent variable depends on the independent variable to reach a mathematical relationship linking the independent and dependent variable. The linear regression model is the simplest and most common modelIts methods are the most important and the basis for any analysis, but it fails to analyze the data when the dependent variable is binary. (Babtain,2010)

1.2 Justifications for not using linear regression.

Since we want to predict the dependent variable that takes the value 0 or 1 and not the dependent variable itself, i.e., we want the probability of it being zero or one, the dependent variable here differs from the dependent variable used in linear regression, and therefore the use of logistic regression is better in the case of binary variables. In addition, if least squares regression is used in this case, it may lead to unexpected values, as we know that the value of the probabilities ranges between zero and one, and if it is used with data with a binary dependent variable, it may make the probability exceed one or less than zero. This is rejected and not compatible with the definition of possibilities. (Hassan,2010)

1.3 Parametric regression model

These models are among the first models for regression analysis, both linear and nonlinear. It is assumed that the study sample is drawn from a measurable normal distribution, and then its parameters are estimated using the method of the greatest probability or another method. These models are used to describe the relationship between the dependent variable and the explanatory variables.It has quantitative and qualitative types. The quantity is both linear and non-linear, and the qualitative one in which the dependent variable is bi-response.

Logistic regression is one of these models, in which the dependent variable is binary or multiple and has many very important uses. (Nasrawi,2017)

Chapter 2

Logistic regression

2.1 Logistic regression

Mathematics is of great importance in human life and health because it contains predictive models used in various medical fields, such as predicting the occurrence of a disease. An example of such models is logistic regression, which is used to predict the probability of an event or disease occurring based on the given data (Dreiseitl, Ohno-Machado, 2003).

It is known that linear regression is suitable for data with a linear relationship, especially in data with continuous outcomes (0, 1) that do not carry a binary result, so logistic regression is a natural and suitable alternative for such data that carries with a binary dependent variable. (Kumarek, 2004)

Therefore, in this chapter, we are interested in studying the relationship between the important variables in the study, namely: age, gender, exercise, history of asthma, chronic obstructive pulmonary disease, family history, healthy food, soft drinks, and exercise. And using this relationship to predict the values of the dependent variable (osteoporosis). Then we arrive at a mathematical model that shows the relationship between the dependent variable and the independent variables. (Mabrouk Saeed 2018).

The dependent binary variable is distributed by one outcome, which is P (Y = 0) for failure (nothing happens or no disease) and P(Y = 1) = success (thing happens or the disease).

The relationship between the dependent variable and the explanatory variables is nonlinear and is converted to linear as follows:

Logit Transformation=ln$\frac{P(x_i)}{1-P(x_i)}$ $=\beta_0 + \beta_i x_i$, i= 1,2,3, …

p (x_i) =$\frac{e^{\beta_0 + \beta_i x_i}}{1+e^{\beta_0 + \beta_i x_i}}$, i=1,2,3, …

β_0: The fixed limit of the logistic regression model.

β_i:Logistic regression parameters to be estimated.

x_i:independent variables. (Nasrawi,2017)

This model is called logistic regression model, it is a special case of Multinomial Logistic Regression, the result has a binomial distribution. (Al-Afifi,2010)

The logistic curve is a general linear model that takes the form of a logistic function as shown in the following figure:

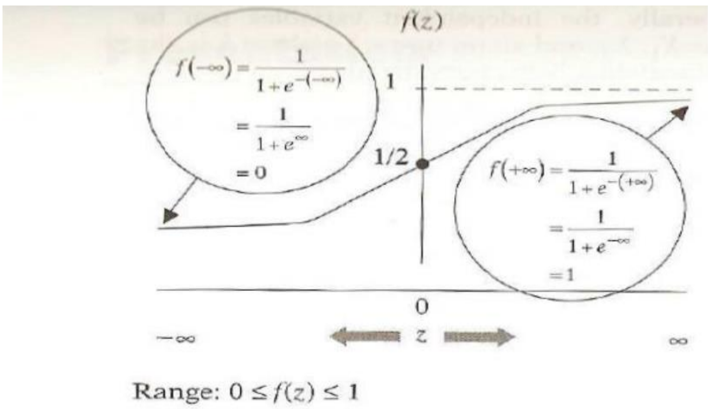


Figure 1 The logistic function varies according to the effect of the independent variable x.

The equation of the logistical form in the above figure is:

P(x) =$\frac{e^{a+bx}}{1+e^{a+bx}}$ =$\frac{1}{1+e^{-(a+bx)}}$

α: Threshold.

β: regression coefficients (Abbas,2012)

- The importance of logistic regression

Logistic regression is a powerful statistical tool because it:

- Explains accurately, clearly and effectively the effect of the independent variable on the dependent variable.
- Provides a test of the sentiment of transactions.
- In addition, he can arrange the independent variables in terms of the strength of their influence on the dependent variable (Abbas, 2012).

- Logistic regression assumptions:

1- The values of the dependent variable are nominal

2- The explanatory variables not necessary continuous.

3- There should be non-linear relationship between the independent and dependent variables

4- It is not necessary that the independent variable follows a normal distribution. (Nasrawi,2017)

Types of logistic regression:

- **Multiple logistic regression**

This regression is used when one qualitative response variable has more than two levels.It has been used in the fields of medicine, social, educational, and other fields.

The probability function is used in estimating the parameters of the model, so it is preferred to use it since it is less restrictive in its hypotheses than the normal regression, which depends on the method of least squares, which has conditions on its hypotheses and requires that the dependent variable be a quantitative variable.(Alfifi,2010)

The polynomial logistic regression model is a simple extension of the binary logistic regression model. This form is used when the dependent variable has more than two nominal or unordered classes.In other words, the dependent variable in the polynomial logarithm model extends to more than two states.

There are several methods for estimating the parameters of the polynomial logistic model. The most common estimation method is the maximum-likelihood method, and there are other methods: the Maximum Posteriori Probability Estimates, the Ridge estimation for symmetric-side MLR models. Constraints, neural networks, and polynomial logic, Generalized moments method for MLR model estimation.(Shehada,2015)

- **Ordinal logistic regression**

$\text{Log}(\frac{P(y\le j)}{1-P(y\le j)})=\alpha-(\beta_1 x_1+\beta_2 x_2+\ldots+\beta_n x_n)$

α: Threshold.

β: regression coefficients

$\text{Log}(\frac{P(y\le j)}{1-P(y\le j)})$:Link function

$x_1,\ldots,x_n$: Explanatory variables

$P(y \le j)$: Cumulative probability

Characteristics of ordinal logistic regression:

1- No assumptions are required on the explanatory variables.

2- The dependent variable is ordinal.

3- Suitable when the dependent variable is multiple.(Suleiman and Dibba,2022)

- **Binary logistic regression**

The model is called binary logistic regression model if the dependent variable is binary ,meaning it takes one of the two values 0 which means nothing happened and the value 1 if something happened. An independent variable is a quantitative or qualitative variable.

We assumed the dependent variable y and the independent variable x, the following figure shows the relationship between the dependent and independent variable (logistic regression function model). (hassan,2014)

It is necessary to ensure that these conditions are met, and then do the bilateral logistic analysis:

1- The dependent variable is binary nominal and takes one of the values 1or 0.

Table 1: dependent variable table

original value	internal value
No	0
Yes	1

2- That the model contains one or more independent variables, and the independent variable is either nominal, categorical, relative or ordinal.

3- That there are no outliers in the independent variables (mahalanobis test).

4- It must be ensured that there is no problem of multi-linearity between the variables.(Zoe and others,2021)

2.2 Logistic regression transformations

There are several transformations of logistic regression that contribute to solving some challenges it faces because the binary dependent variable carries either a value of zero or one, namely:

2.2.1 Odds

It is the ratio of the probability of something happening to the probability of it not happening.

Odds of (Event) $=\frac{p}{1-p}$

P:The probability of something happening.

1-p:The probability that something will not happen.

The difference between it and probability is that if oddfor a team's victory is equal to$\frac{3}{4}$,This means that it is 3 times success to 4 times failure, but in the case of probability, it is $\frac{3}{7}$ because it means 3 success times out of 7 attempts.

It is expressed by a number. For example, when we say that Odd is equal to 5, this means that the chance of something happening is 5 times the chance of its non-occurrence.(Babtain,2009)

Through it, we can know the effect of the independent variables, because the logistic regression shows the probability of something happening rather than not happening.

p= $\alpha+\beta x$

Because the value of x changes, it makes the probability not fall between 0 and 1, so we will convert the probabilities through the natural logarithm.

Logit(y)=ln(odds)=ln($\frac{p}{1-p}$)=$\alpha+\beta$

p: The probability of the event.

X: independent variable

α, β: logistic regression parameters

This equation indicates that through logistic regression the probabilities are modeled as a linear function of the variable X. This is a simple logistic model because it has only one independent variable.

By taking the inverse of the logarithm, the equation becomes:

P=P(Y=interested outcome/X=x,a specific value)= $\frac{e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}}=\frac{1}{1+e^{-(\alpha+\beta x)}}$

This is the prediction equation that indicates the probability of obtaining the occurrence of the event, i.e. success.

But if the independent variable is multiple, the logistic regression becomes complex:

Logit(y)=ln(odds)=ln($\frac{p}{1-p}$) $=$ α+$\beta_1 X_1$+……+ α+$\beta_k X_k$

P=P(Y=interested outcome/ X_1 = x_1 ,…, X_k = x_k)= $\frac{e^{\alpha+\beta_1 X_1+\cdots\ldots+\ \alpha+\beta_k X_k}}{1+e^{\alpha+\beta_1 X_1+\cdots\ldots+\ \alpha+\beta_k X_k}}$

=$\frac{1}{1+e^{-(e^{\alpha+\beta_1 X_1+\cdots\ldots+\ \alpha+\beta_k X_k})}}$

(manahi , kamr,2018)

2.2.2 Odds ratio

Determines the effectiveness of the association between two events. It measures the probabilities of exposure to the event in comparison with the probability of the absence of this exposure.(manahi, kamr,2018)

Babtain 2010 mentioned that it is a ratio between the odd1 for a variable and the odd2 for another variable, that is, it is equal to:

Odd Ratio= $\frac{odd1}{odd2}$

Where the significance of logistic regression is that exp(β_1) is equal to the odd ratio.

Chapter 3

Methods for estimating model parameters

The importance of estimating the parameters of any regression model lies in the interpretation of the relationship between the response variable (dependent) and the explanatory variable (independent) in an approximate mathematical formulaleast squares, etc.) (Taha,2005)

The maximum likelihood estimation method is used to estimate the logistic regression coefficients. As for the least squares method, which we use in estimating the parameters of the normal regression, we cannot use it because it is not suitable in the case of logistic regression.

3.1 The maximum likelihood estimation method

The maximum likelihood estimation method is an iterative method that relies on repeating the calculations several times until the best estimates of the coefficients are reached to explain the observed data.

This method is distinguished in that it does not require any conditions on the independent variables, as it is valid for estimation, whether the variables are nominal, categorical, or ordinal.(Qazaini,2016)

It is one of the most famous statistical methods in estimation, as it measures the observed probabilities of a number of independent variables that enter the sample, as it represents the product of multiplying these probabilities the maximum likelihood estimation method.(Ghanem, Al Jaouni,2011)

Furthermore, it allows Square-Chi to be used to judge the fit of the model to the data, which is equivalent to the least squares method in linear regression, but seeks to maximize the logarithm of probabilities (knowing that the observed values can be the predictor of the dependent variable through the independent variables) .(Bolad,2013)

The likelihood function(L) is :

$$L(\beta,y)=\prod_{I=1}^{N}\frac{n!}{y!(n!-y!)}P(x_i)^{y_i}(1-p(x_i))^{n_i-y_i}$$

Sort the equation, so it becomes

$$L(\beta,y)=\prod_{I=1}^{N}(\frac{p(x_i)}{1-P(x_i)})^{y_i}(1-P(x_i))^{n_i}$$

But we know he is:

$$Ln\left(\frac{p(x_i)}{1-P(x_i)}\right)=\beta_0+\beta_i x_i$$

by taking the logarithm, so it becomes

$$\frac{p(x_i)}{1-P(x_i)}=e^{\beta_0+\beta_i x_i}$$

We substitute it into the previous equation, and it becomes

$$L(\beta,y)=\prod_{I=1}^{N}(e^{\beta_0+\beta_i x_i})^{y_i}(1-\frac{e^{\beta_0+\beta_i x_i}}{1+e^{\beta_0+\beta_i x_i}})^{n_i}$$

Then we take the logarithm

$$Ln(L(\beta,y))=\sum_{i=1}^{N}y_i(\beta_0+\beta_i x_i)-n_i\, Ln(1+e^{\beta_0+\beta_i x_i})$$

L: The likelihood function

P:The probability of something appearing

N:The total number

To get values, we derive the equation, set it equal to zero, and then solve it .(Nasrawi,2017)

3.2 Interpretation of logistic model transactions

3.2.1 Interpretation of coefficients in terms of logit coefficient

This case is somewhat like the method of linear regression, but the difference is that in this method a change in the independent variable will be followed by a change in the logit and not in the dependent variable itself.

3.2.2 Interpretation of coefficients in terms of odd ratio

Provides an interpretation of the logistic regression coefficients by finding a relationship between odd for an independent variable at a certain level and odd for the same variable when adding a unit to the previous level.

3.3.3 Interpretation of coefficients in terms of probabilities

Increasing the independent variable by one unit will increase the logiteln $(\frac{p}{1-p})$.(Malika,2020)

Chapter 4

Model suitability evaluation

Check the suitability of the model as a whole:

4.1 Classification table

A table showing the number of cases that have a positive trait and the number of cases that have a negative trait versus the number of expected cases that have a trait and the number that don't have that trait and the number of cases that have been correctly and incorrectly classified and it is said that the model matches the data if the data is expected to be classified correctly.

Table 2: classification table

<table>
<tr><td colspan="2" rowspan="2">Classification</td><td colspan="2">Expected</td><td rowspan="2">Total</td></tr>
<tr><td>Positive</td><td>Negative</td></tr>
<tr><td rowspan="2">Viewer</td><td>Positive
P</td><td>correct positive
TP</td><td>correct negative
FN</td><td>P</td></tr>
<tr><td>negative
N</td><td>False positive
FP</td><td>False negative
TN</td><td>P'</td></tr>
<tr><td colspan="2">Total</td><td>Q</td><td>Q'</td><td>1</td></tr>
</table>

The classification table statistics are:

1-Sensitivity(SE) :The value of the probability that the expected rating is positive for the case that is actually positive.

SE=$\frac{TP}{(TP+FN)}$=$\frac{TP}{P}$

2-Specifity(SP): The value of the probability that the expected rating will be

negative for a case that is actually negative.

SP=$\frac{TN}{(FP+TN)}$ = $\frac{TN}{P\prime}$

3-Hit ratio (EF): It is the correct classification ratio or also called the efficiency

ratio.

EF=TP+TN

Hit ratio=$\frac{EF}{Total}$ =$\frac{(TP+TN)}{(P+P')}$ =$\frac{(TP+TN)}{(Q+Q')}$

It is used to measure the accuracy of the statistical model and its ability to predict the dependent variable. However, it is not used to measure the extent to which the observed data matches the expected model.

(Qazania,2017)

4.2 ROC curve (receiver operating characteristic curve)

is a curve of cut-off points for sensitivity versus accuracy.

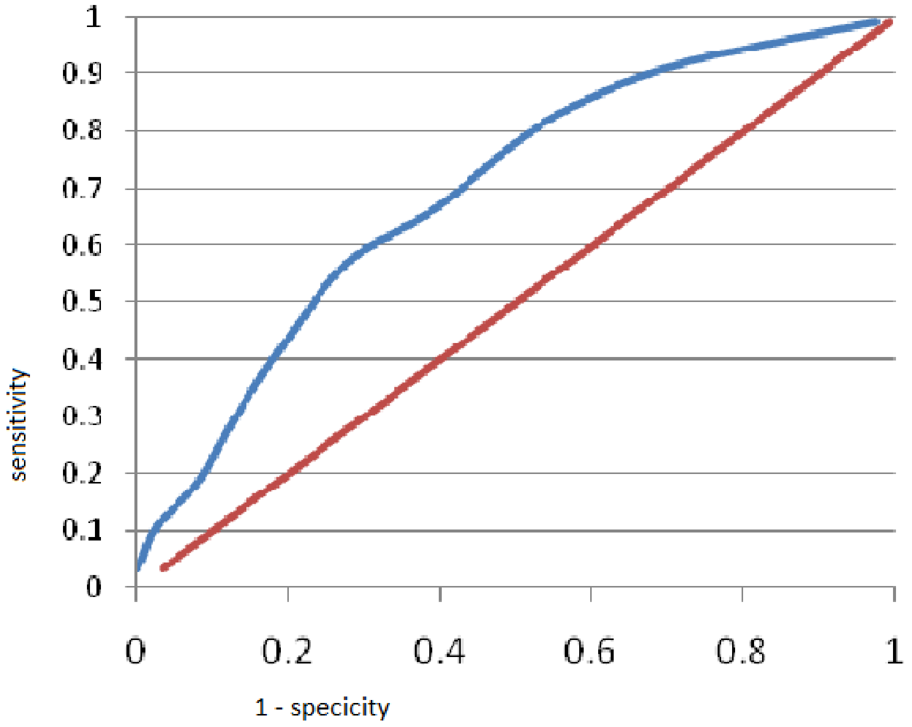


Figure 2 Curve ROC representation of the data expected from the logistic model . (Babtain,2010)

The curve begins with the coordinates (0, 0) corresponding to the cut-off points that go to all negative cases. As for the second side, it (1, 1) corresponds to the cut-off point of the decision, which goes to the positive cases. positive and negative. (Babtain,2009)

The higher the curve is in the upper left direction of the diameter of the shell, the higher the discriminatory ability of the model, and when the curve is higher than the diameter of the shell, the model has the ability to discriminate and classify cases . The area under the curve values are interpreted as follows:

0.9 ≤ROC Superior discriminatory ability.

0.8≤ROC≤0.9 Excellent discrimination ability .

$0.7 \leq ROC \leq 0.8$ acceptable discriminatory ability .

ROC=0.5 Form has no differentiating ability than chance.(Qazania,2017)

4.3 Goodness-of-fit test

4.3.1 Chi-square test

It is denoted by the symbolX^2, and its formula:

$$X^2 = \sum_{i=1}^{n} \frac{(Xi - Yi)^2}{Y_i}$$

X_i : Viewed values.

Y_i:Expected values.

n:Views.

The model is compatible with the data (observed values = expected values) in the event that the statistic value of the test is greater than a specific level of significance, that is, it is not significant,Who suggested this test is Pearson , The distribution of this test approximates the chi-square distribution, The degree of freedom is equal to the number of observations minus the number of parameters in the model P.(abodoma,2019)

Hypothesis:

H_0 :There are no significant differences between the observed frequencies and the expected frequencies for the random variable X.

H_1:There are significant differences between the observed frequencies and the expected frequencies for the random variable X.

The null hypothesis is rejected if the calculated chi-square value is greater or equal to the tabular chi-square value, and this means that there are significant differences between the observed and expected frequencies.(mahdi,2014)

This test depends on the residuals, which are known as:

$$r = \frac{y_{i-y'_i}}{\sqrt{y'_{i(1-y'_{i)}}}}$$

y_i: It is the dependent variable whose value is either 0 or 1 .

y'_i: The corresponding predicted variable based on a logistic regression model.

$y' = \frac{\exp(\alpha+\beta_1 x_{i1}+\cdots+\beta_k x_{ik})}{1+\exp(\alpha+\beta_1 x_{i1}+\cdots+\beta_k x_{ik}}$(manahi , kamr,2016)

4.3.2 Hosmer-lemeshow test

Hosmer and Lemeshow (1980) built a model of goodness-of-fit test, which is the Hosmer and Lemeshow test, which is a widely used test, especially in medical studies, for its simplicity.

This test evaluates differences in the number of observed and expected events from the data collected based on the fit values from the model.

To calculate the test statistics, we divide the data into groups by making the data approximately equal in size and the values within each group are similar.(Surjanovic and others ,2020)

This test is commonly used to goodness of fit of logistic regression.The Hosmer–Lemeshow test is a chi-square test,It is performed by sorting the n records in the data set according to the probability of succe. And evaluating the Hosmer–Lemeshow C statistic:

$$\widehat{C_g} = \sum_{i=1}^{g} \left(\frac{(O_{s,i} - E_{s,i})^2}{E_{s,i}} + \frac{(O_{f,i} - E_{f,i})^2}{E_{f,i}} \right)$$

$O_{s,i}$, $O_{f,i}$: the observed number of failures and successes

$E_{s,i}$, $E_{f,i}$:the expected number of successes and faulures

Thus, the p-value for the Hosmer–Lemeshow test is:

$$P = \int_{\widehat{C_g}}^{\infty} X^2{}_{g-2}(x)dx$$

$X^2{}_{g-2}(x)$:is the probability density function of the X^2 distribution with g-2 degrees of freedom evaluated at x.

(Paul and others,2013)

4.4 Pseudo R^2($R^2{}_L$)

the improvement in model likelihood over a null model. (decrease in variance)

$$R^2{}_L = \frac{G_M}{D_0} = \frac{G_M}{(G_M + D_M)}$$

D_0 = Regression that includes only the fixed term without the independent variables.

D_M = *Regression that includes the fixed term in addition to the independent variables.*

G_M=*the difference between*D_0 *and* D_M.(*Babtain,2010)*

The statisticR^2 is used to test the strength of the logistic regression (Contribution of the influencing factors in the model to the dependent variable),and that is through :

4.4.1 Nagelkerke R square (${R_N}^2$)

This statistic is characterized by its ability to reach its maximum value, unlikeCox and Snell R square.

$${R_N}^2 = \frac{{R_{CS}}^2}{1-e^{(\frac{2LL(baseline)}{n})}}$$

$${R_N}^2 = \frac{{R_{CS}}^2}{1-(e^{\frac{-baseline}{2}})^{\frac{2}{n}}}$$

4.4.2 Cox and Snell R square(${R_{CS}}^2$)

$${R_{CS}}^2 = \text{1-}e^{(-\frac{2}{n}(LL(new)[LL(baseline))}$$

$${R_{CS}}^2 = \text{1-}e^{\frac{1}{2}(new)-(baseline))^{\frac{2}{n}}}$$

n:Sample volume.

LL(new):log likelihood for a model

LL(baseline):log likelihood for reference original form

It cannot take the value of one integer, that is, it cannot reach its maximum value of 1, so we useNagelkerke R square.(abodoma,2019)

4.5 Wald test

It is used to determine the contribution of the independent variables. It follows a distributionX^2 .This statistic represents the ratio of the square of the regression coefficient to the square of the standard error of the coefficient. The Wald statistic is compared with a distribution X^2at one degree of freedom.(manahi, kamr,2018)

Wald's statistic is used to calculate the significance of the estimated parameters using logistic regression for each of the regression coefficients corresponding to each independent variable, in order to test the null hypothesis (the effect of Logit coefficient is zero)The significance of the parameters for each logistic regression coefficient corresponding to each independent variable must be less than 0.05 in order to reject the null hypothesis and accept the alternative hypothesis, which means that the independent variable has an impact on predicting the value of Y (that is, the effect of the coefficient is not equal to zero is significant at the level of morale 0.05).(Faraj, Osman and Muftah ,2022)

This test is known as the Wald's statistic that follows a probability distribution, which is a chi-square. It tests whether the value of B for the predictive variables in the model is different from 0. If it is, then we can say that the predictive variables contribute to predicting the value of y.

An important way to know the importance of explanatory variables in the model.

the null hypothesis:

H_0:β=0

H_1:β≠0

W=$(\frac{\hat{\beta}}{S(\hat{\beta})}^2)$

W:Wald's statistic

$\hat{\beta}$:parameter amount.

S($\hat{\beta}$):Standard deviation of the parameter estimator.

Tests the hypothesis that the logistic regression coefficient associated with the independent variable is zero,The larger the value of w, the greater the proportion of the contribution of the corresponding variable.

W^2distribution tracking X^2,And if you countwinstead of W^2 It trace distributionZ ,And The equation becomes:

W=$(\frac{\beta}{SE_\beta})$(abodoma,2019)

4.6 Confidence interval estimation

A confidence interval for parameter β_j can be created as follows:

$\beta_j \pm z_{1-\frac{\alpha}{2}}$ S.E(β_j)

The confidence interval for the odd ratio for the variable is:

Exp($\beta_j \pm z_{1-\frac{\alpha}{2}}$ S.E(β_j)

(ahmad,2014)

Chapter 5

Data analysis and empirical results

This part of the study presents a general description of the variables of the study, as well as the impact of asthma and pulmonary embolism on the risk of osteoporosis for Palestinian citizens. The study is based on an econometric analysis using a logistic regression model, where the dependent variable is the incidence of osteoporosis, and the independent variables are the incidence of asthma and the incidence of pulmonary embolism. The model also neutralizes the effect of a some of other factors, such as gender, age group, healthy food, family medical history, soft drinks, and practicing sports.

5.1 Descriptive statistics

The study sample consisted of three main groups; each group consisted of 20 individuals, so that the first group consisted of patients with asthma, the second group of patients with pulmonary embolism, and the third group of healthy individuals, which was considered as a reference group.

Table 1 presents the descriptive analysis of the study variables and the characteristics of each of the three groups. It is clear from the table that the rate of osteoporosis among patients with asthma was about 40%, while it was about 50% among patients with pulmonary embolism, but it was only 10% among the reference group.

For the controlling variables, the proportion of males was about 70% in the group of asthma patients, 60% among the pulmonary embolism group, and 40% among the control group. In terms of age, the group of asthma patients consisted of about 65% of the elderly, while the percentage of elderly people was about 70% in the group of patients with pulmonary embolism, and about 30% in the reference group.With regard to the patients' family history, it appears from Table 1 that about 60% of asthma patients have a family history, compared to 25% for pulmonary embolism patients, and only 15% for the reference group.On the other hand, about 55% of asthma patients consume soft drinks, compared to 65% for patients with pulmonary embolism, and 35% for the control group.In addition, about 25% of patients with asthma practice sports continuously, compared to 20% of patients with pulmonary embolism, and 45% of the control group.Likewise, about 35% of individuals in the group of patients with asthma deliberately eat healthy foods, compared to 20% for patients with pulmonary embolism, and 50% for individuals in the control group.

Table 3: Descriptive statistics for study variables.

Variable	Asthma group	Pulmonary embolism	Reference Group
Osteoporosis rate	40%	50%	10%
Male (% of total)	70%	60%	40%
Elderly people (% of total)	65%	70%	30%
Family history (% of total)	60%	25%	15%
Healthy food (% of total)	35%	20%	50%
Soft drink (% of total)	55%	65%	35%
Sports (% of total)	25%	20%	45%

Table 2 below indicates that the highest rate for osteoporosis is among the subgroup of females with asthma with rate of about 66.7%, followed by the subgroup of elderly people with pulmonary embolism with rate of about 64.3%. In the other hand the lowest rate of osteoporosis is among the males in the reference group and sport individuals with asthma with zero rate, followed by the subgroup of no family history in the reference group with rate 5.9%. In general we deduce that there is relatively high rate of osteoporosis compared with others between females, elderly people, with family history, soft drink, no sport, and no healthy food. The next section of this chapter clarify if those differences between different categories according to osteoporosis are statistically significant or not.

Table 4: Osteoporosis rate by groups and control variables

Variable		Asthma group	Pulmonary embolism	Reference Group	All groups
Gender	Male	28.6%	50.0%	0%	29.4%
	Females	66.7%	50.0%	16.7%	38.5%
Age	Elderly people	42.9%	64.3%	16.7%	45.5%
	Young people	38.5%	16.67%	7.14%	18.5%
Family history	Yes	41.7%	80%	33.3%	50%
	No	37.5%	40%	5.9%	25%
Soft drink	Yes	54.6%	53.8%	14.3%	45.2%
	No	22.2%	42.9%	7.7%	20.7%
Sports	Yes	0.0%	50.0%	11.1%	16.7%
	No	53.3%	50.0%	9.1%	40.5%
Healthy food	Yes	14.3%	25.0%	0%	9.5%
	No	53.9%	56.3%	20%	46.2%

5.2 Econometric analysis

This section of the study presents the results of the logistic regression model analysis of the effect of asthma and obstructive pulmonary disease, in addition to other controlling variables, on the risk of osteoporosis.

The logistic regression model is used when the dependent variable is a binary variable, so the least squares method cannot be used if the dependent variable is of this type. One of the values of the binary variable (the occurrence of the event) takes the value (1), with a probability of (p), while the other, represented by the non-occurrence of the event, takes the value (0) with a probability of (1-p). It is worth noting that there are no restrictions on the independent variables in this model, as they may be continuous variables, categorical variables, or a mixture between them (Park, Hyeoun-Ae 2013). The logistic regression model if there is one independent variable is represented by the following:

equation: $Ln\left(\frac{p}{1-p}\right) = \hat{\beta}_0 + \hat{\beta}_1 X$

In other words, the previous equation can be written as follows:

$$\left(\frac{p}{1-p}\right) = e^{\hat{\beta}_0 + \hat{\beta}_1 X}$$

$$p = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 X}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 X}}$$

Where:

p: The probability of the event of interest occurring (probability of success).

(1-p): The probability of the event occurring is not of concern (probability of failure).

$\left(\frac{p}{1-p}\right)$:The odds ratio of the event of interest.

Based on the above, the logistic regression equation can be written in the case of (k) independent variables, as follows:

$$Ln\left(\frac{p}{1-p}\right) = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \cdots + \hat{\beta}_k X_k$$

The coefficients of the logistic regression model are estimated using the Maximum Likelihood method, which is considered one of the well-known statistical estimation methods, where the maximum likelihood function measures the probabilities observed for the number of independent variables that are studied (p1p2… pk), and this function is the product of these probabilities (Park, Hyeoun-Ae 2013).

The logistic regression model has several advantages that make it suitable for use in such cases, as it provides an idea of the amount of the effect of the independent variable on the probability of the occurrence of the dependent variable (the binary variable), and this model also arranges the independent variables according to the most influencing the dependent variable. This model is also distinguished by the fact that it is not required that the independent variables follow a normal distribution (such as the least squares method), nor that the relationship between the independent variables and the dependent variable be linear. Likewise, the suitability of the logistic regression model can be verified by checking the suitability of the model as a whole, as well as by testing the statistical significance of each variable separately.

In this study, the logistic regression function will be estimated so that the dependent variable is the osteoporosis, which takes two values represented by patient with osteoporosis, which takes the value (1), and the second value, which iswithout osteoporosis, takes the value (0). The independent variables are as follows:

- **Asthma:** The value (1) is for asthma paiients, and the value (0) is for non-asthma, meaning that non-asthma group are the reference group in the models.

- **Pulmonary embolism:** The value (1) is for pulmonary embolism paients, and the value (0) is for pulmonary embolism, meaning that non-asthma group are the reference group in the models.

- **Gender** The value (1) is for males, and the value (0) is for females, meaning that females are the reference group in the models.

- **Age**: The value (1) is for elderly people, and the value (0) is for young, meaning that young are the reference group in the models.

- **Family history**: The value (1) is for family history, and the value (0) is for non family history, meaning that “no family history” are the reference group in the models.

- **Healthy food**: The value (1) is for healthy food persons, and the value (0) is for non healthy food, meaning that “non healthy food” are the reference group in the models.

- **Soft drink**: The value (1) is for soft drink, and the value (0) is for no soft drink, meaning that “no soft drink” are the reference group in the models.

- **Sport**: The value (1) is for practicing sports, and the value (0) is for non priacticing sports, meaning that “no sport” are the reference group in the models.

5.3 The results of logistic regression

Table 3 indicates the results of the logistic regression model for estimating the variables that affect the risk ofosteoporosis in Palestine , and it is evident from this model that some of independent variables are statistically significant at the level of 5% which are astham, pulmonary embolism, age, and health food. While the variables of gender, family history, soft drink, and sport are not statisticallt significant at 5% level of significance. The following is an explanation of the effect of each significant variable on osteoporosis:

- **Asthma:** The results of the model indicate that paients with asthma are more susceptible to osteoporosis, as theodds ratio is about 7.45 (They are more susceptible to osteoporosis by about 7.5 than healthy persons), holding other variables constant.

- **Pulmonary embolism:** The results of the model indicate that paients with Pulmonary embolism are more susceptible to osteoporosis, as theodds ratio is about 8.45 (They are more susceptible to osteoporosis by about 8.5 than healthy persons), holding other variables constant.

- **Age:** The results of the model indicate that elderly people are more susceptible to osteoporosis, as theodds ratio is about 8.99 (They are more susceptible to osteoporosis by about 8.99 than young people), holding other variables constant.

- **Helthy food:** The results of the model indicate that healthy food people are less susceptible to osteoporosis, as theodds ratio is about 0.04 (They are less susceptible to osteoporosis by about 0.04 than other people), holding other variables constant.

Table 5: Results of logistic regression of osteoporosis function

Independent Variables	Dependent Variable: Osteoporosis				
	Odds Ratio	Z	p-value	Confidence interval	
Male	0.16	-1.72	0.085	0.02	1.29
Asthma	7.45	1.98	0.047	1.02	54.11
Pulmonary embolism	8.45	2.44	0.015	1.52	47.05
Age	8.99	2.10	0.036	1.16	69.61
Family	1.15	0.17	0.863	0.23	5.83
Healthy Food	0.04	-2.80	0.005	0.00	0.38
Soft Drinks	0.26	-1.49	0.137	0.04	1.54
Sport	0.68	-0.45	0.656	0.12	3.76
Constant	0.38	-0.98	0.325	0.05	2.64

5.4 Diagnosis tests

It is necessary to clarify the suitability of the model as a whole in estimating the factors affecting the likelihood of osteoporosis, and there are several measures to test this. From them we will use goodness of fit, , model classification efficiency test, and ROC curve analysis.

1. Chi-square test and (R^2) goodness of fit

The chi-squared test indicates that the logistic regression model has significant effect in estimating the probability of osteoporosis, as the value of chi-squared test is about 17.06 as shown in table 4, and in a significant sense (0.0297), which indicates that the independent variables have importnat influence and contribution statistically significant in the probability of osteoporosis. In addition the value of Pseudo R^2 indicates that the independent variables explain about 32% of the variation in the dependent variable.

Table 6: Goodness-of-fit test for Logistic model forosteoporosis

Observations	LR chi2(8)	Prob > chi2	Pseudo R2
60	17.06	0.0297	0.3207

2. Model classification efficiency test

This test is one of the methods of examining the efficiency of the model, as it is clear from Table 4 that the estimated model achieved a correct overall rating rate of 80% (the number of correct predictions divided by the total number of the study sample), and this percentage is somewhat high and evidence of the model's efficiency.

Table 7: Efficiency of the classification of a logistic regression model for estimating osteoporosis.

Classification		**Predicted**		**Correctly classified %**
		+	**-**	
Observed	+	14	6	70%
	-	6	34	85%
Aggregate correctly classified %				**80%**

3. ROC curve

The ROC curve is one of the most important methods for testing the efficiency of the logistic model, as the null hypothesis states that the area under the ROC curve resulting from the logistic regression analysis does not differ from the chance ratio (0.5), and it is clear from Figure 1 that the ROC curve resulting from the logistic regression function of osteoporosis is far from the shell factor, as it clearly moves away from the diameter of the shell, which confines 50% of the area under it, and thus gives a larger area than the shell gives. As it can be seen from the figure that the area under the curve is about 86%, which is a good percentage, and it enables us to reject the null hypothesis, and therefore the model helps to predict the classification of the states of the dependent variable.

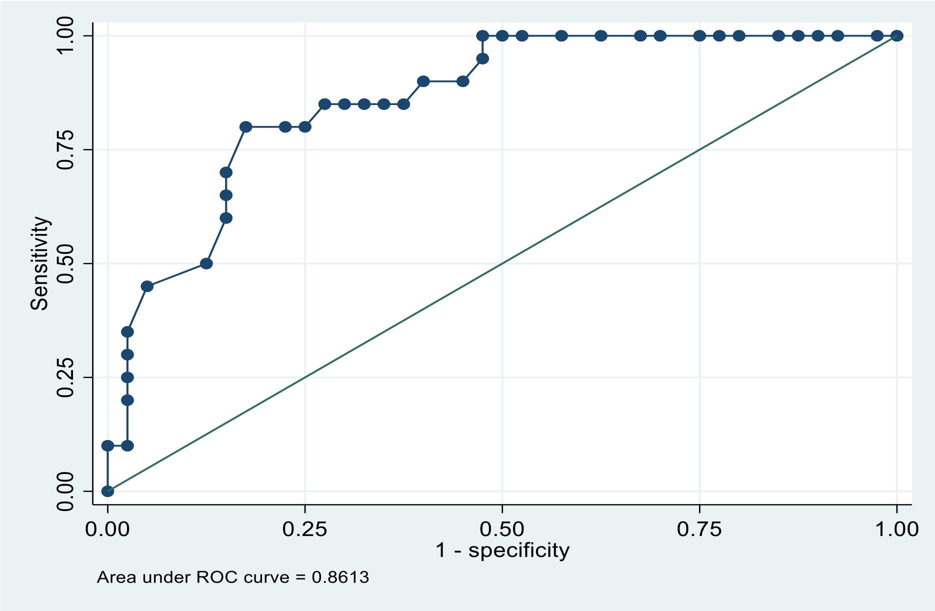


Figure 3: ROC curve, for a logistic regression model for estimating osteoporosis

Conclusions and future work

1. The use of logistic regression in building prediction models for data with two-valued dependent variables because of its high explanatory power.

2. Increasing the health awareness of patients with asthma and chronic obstructive pulmonary disease through the periodic review of the doctor and conducting the necessary analysis.

3. Recommending the study of multiple and ordinal logistic regression in future studies.

4. Comparison of logistic regression with other statistical methods.

References

- A,Ghanem, F. Al Jaouni**:The use of the two-response logistic regression technique in studying the most important economic and social determinants of family income adequacy "An applied study on a random sample of families in Damascus Governorate "**,Damascus University, Journal of Economic and Legal Sciences , Damascus University, Syria, 2011.
- A.Abbas:**Using the logistic regression model to predict countries with qualitative dependent economic variables**,Tikrit University, Kirkuk University Journal of Administrative and Economic Sciences,Iraq,2012.
- A.Abdsalam:**Analytical Study to Detect the Effect of Skewnes on Chi- Square Test Statistics**, Sudan, 2014.
- A.abdsalam:**Analytical Study to Detect the Effect of Skewnes on Chi- Square Test Statistics**,Sudan, 2014.
- A.Babkir:**Using binary logistic regression to determine the most important economic and social factors that lead to women's deaths during pregnancy and childbirth - a field study on the state of Khartoum - Sudan during the year 2018**,2018,Sudan.
- A.Babtain:**Logistic regression and how to use it to build prediction models for data with two-valued dependent variable,**Saudi Arabia ,2009.
- A.Faraj,R. Osman ,A. Muftah:**Using the two-response logistic regression technique to determine the most important economic and social factors affecting household income**, ,Liby, 2022.

- A.manahi ,S. kamr:**Financial indicators affecting the performance of stocks in the Iraqi Stock Exchange for the year 2016 using the binary logistic regression model**, Iraqi,*2018*.
- A.Taha**:Some traditional estimator methods and Bayesian estimator for parameters of the general linear regression model(A comparative study with an application in a medical field)**,University of Mosul, Iraq,2005.
- B,Malika**: Building a scoring ladder to predict the failure of institutions using logistic regression,**Algeria, ,2020.
- D.Guffey**:Hosmer-Lemeshow goodness-of-fit test: Translations to the Cox Proportional Hazards Model**,University of Washington,American Seattle,2012.
- E.Zoe, R.AbuSherida,N.madi,H.hamrosh**: Using the binary logistic regression model to study the most important factors affecting the persuasion of the Libyan citizen of the existence of the Corona virus**,The comprehensive multidisciplinary electronic magazine,Libya.,2021.
- F,ahmad:**Using the logistic model to determine the factors affecting the incidence of anemia (anemia) in children 2009-2013 AD**, ,Sudan ,2014.
- F.hassan:**Using Logistic Model to Determine the Factors Affecting AnemiaIncidence in Children 2009 -2013**, Sudan ,2014.

- F.hassan:**Using Logistic Model to Determine the Factors Affecting AnemiaIncidence in Children 2009 -2013**,Sudan ,2014.

- F.namk:**Using the two-response logistic regression technique and the multiple linear regression technique to diagnose the factors affecting the high repetition rates in higher education** ,Iraq ,2015.
- F.namk:**Using the two-response logistic regression technique and the multiple linear regression technique to diagnose the factors affecting the high repetition rates in higher education** ,Iraq ,2015.
- H.Abodoma and H.Osman and T. Muhammed:**Estimation and analysis of factors affecting the odds of developing and predicting heart disease using a logistic regression model** ,Sudan,2018.
- H.Abodoma,H.Osman and T. Muhammed:**Estimation and analysis of factors affecting the odds of developing and predicting heart disease using a logistic regression model** ,Sudan,2018.
- H.Abodoma: **Using Logistic Regression and Discriminant Analysis Techniques for Factors Affecting Heart Diseases Infection**,sudan ,2019.
- H.abodoma: **Using Logistic Regression and Discriminant AnalysisTechniques for Factors Affecting Heart Diseases Infection**,sudan ,2019.
- I.Mabrouk,A.Saeed**:The use of logistic regression in estimating unemployment rates in the Arab countries 2017**,Mohamed Boudiaf University of Messil a,Algeria,2018.
- J.Bartlett: **The Hosmer-Lemeshow goodness of fit test for logistic regression**,2014.

- J.Stoltzfus**:Logistic Regression: A Brief Primer**, the Society for Academic Emergency Medicine,2011.

- J.Stoltzfus**:Logistic Regression: A Brief Primer**, the Society for Academic Emergency Medicine,2011.

- Jee Hye Wee, Chanyang Min, Min Woo Park, Soo Hwan Byun, Hyo-Jeong Lee, Bumjung Park, Hyo GeunChoi:**The association of asthma and its subgroups with osteoporosis: a cross-sectional study using KoGES HEXA data**,BMC journal, Korea, 2020.
- L.Qazaini**: Using the logistic regression model to predict the failure of economic institutions**, University of Larbi Ben M'hidi, Oum El Bouaghi, Algeria.
- L.qazaynia, R. ashoosh:**Using a logistic regression model to predict the failure of economic enterprises**, , Algeria,2017.

- L.qazaynia, R. ashoosh:**Using a logistic regression model to predict the failure of economic enterprises**, , Algeria,*2017.*

- M,Hassan:**Using two-response logistic regression to identify the most important economic factors Social and demographic affecting HIV infection**, Gezira State, Sudan,2010.
- M.sari,M.dash:**Logistic Regression Model : Concept, characteristics, applications"With an example of abinary logistic regression in SPSS",**Algeria,2017.

- M.sari,M.dash:**Logistic Regression Model : Concept, characteristics, applications"With an example of abinary logistic regression in SPSS",**Algeria,2017.
- M.Shehada**:Statistical Inference on The Multinomial Logistic Regression Coefficients With Application** ,Gaza,2015.
- M.Shehada**:Statistical Inference on The Multinomial LogisticRegression Coefficients WithApplication**,Gaza,2015.
- MaiteLuengo, Cesar Picado, Luis Del Rio, Nuria Guafiabens, Josep M Montserrat,JorgeSetoain:**Vertebral fractures in steroid dependent asthma and involutional osteoporosis: a comparative study**, Spain ,1991.
- N,Nasrawi:**Using the bootstrap method in analyzing parametric and semi-parametric models and comparing them**, Iraq ,2017.
- N.R. Jørgensen, P. Schwarz, I. Holme, B.M. Henriksen,L.J. Petersen, V. Backer**:The prevalence of osteoporosis in patients with chronic obstructive pulmonary disease—A cross sectional study**,Elsevier,2006.
- N.Surjanovic,R.Lockhart,T.Loughin :**A Generalized Hosmer-Lemeshow Goodness-of-Fit Test for a Family of Generalized Linear Models** ,Canada ,2020.
- P,Hyeoun-Ae :**An Introduction to Logistic Regression: From Basic Concepts to Interpretation with Particular Attention to Nursing Domain,**College of Nursing and System Biomedical Informatics National Core Research Center, Seoul National University, Seoul, Korea,2013.

- P. Paul, M.Pennellb,S. Lemeshowb:**Standardizing the power of the Hosmer–Lemeshow goodness of fit test in large data sets**,Wiley Online Library, 2013.
- P.Komarek: **Logistic Regression for Data Mining and High-Dimensional Classification**,Carnegie Mellon University,USA ,2004.
- R,Bolad:**Application of the two-response logistic regression model to patients Diabetes mellitus in Gezira State - Sudan 2012**, sudan,2013.
- R.lfifi :**The Use of Multinomial Logistic Regression Model on Physical Violence Data**, Al- Azhar University , Gaza, ,2010.
- S.Dreiseitl , L.Ohno-Machado :**Logistic regression and artificial neural network classification models: a methodology review**,USA,2003.
- S.Farhud:**Using logistic regression to study the factors affecting stock performance (Applied study on the Kuwait Stock Exchange)** ,Gaza,2014.
- S.Farhud:**Using logistic regression to study the factors affecting stock performance (Applied study on the Kuwait Stock Exchange)** ,Gaza,2014.
- S.Heydar**:Comparison of Artificial Neural Networks with Logistic Regression for Detection of Obesity**,J Med Syst,2011.
- S.Sperande:**Understanding logistic regressionanalysis**,Brazil ,2013.
- S.Sperandei:**Understanding logistic regression analysis**,Brazil ,2013.

الملخص

قمنا في هذه الأطروحة بدراسة خطورة الإصابة بهشاشة العظام لدى مرضى مدينة جنين ، وعلاقتها ببعض مؤشرات المرضى المصابين بالربو ومرض الانسداد الرئوي المزمن، ولهذا الغرض قمنا بتحليل البيانات باستخدام برنامج احصائي وقمنا بتطبيق طرق مختلفة لدراسة النتائج المختلفة وخصائص المتغيرات.