



## **Using Logistic Regression and Cox Regression Models to Studying the Most Prognostic Factors for Leukemia patients**

***Ast. Prof. Dr. Kurdistan Ibrahim Mawlood***

Statistics Department / College of Administration and Economics/Salahaddin University – Erbil

Email: [kurdistan.mawlood@su.edu.krd](mailto:kurdistan.mawlood@su.edu.krd)

---

### **ARTICLE INFO**

***Article History:***

Received: 21/12/2018

Accepted: 27/1/2019

Published: Summer 2019

Doi:

**10.25212/lfu.qzj.4.3.20**

---

***Keywords:***

*Survival Analysis, logistic regression, Cox regression model, Hazard Function, Leukemia.*

### **ABSTRACT**

The basic idea of this study focused on the using of two advanced statistical methods for studding the most important factors affecting the leukemia in Erbil city. The logistic regression was chosen and Cox regression as being applicable on this studies. The results indicated that, in spite of the different regression coefficients in somewhat to logistic regression and Cox regression, have not reached to the same variables that have an impact on the phenomenon. Moreover the results indicated that the surgery is the most important factor affecting the leukemia survival patients in both methods. validation was done by calculating two model selecting criterion; Akaike Information Criterion (AIC) and Bayesian information criterion (BIC) of each models are compared the smaller values of them.

The data set of this study was obtained from Nanakali Hospital in the period from 1st January 2013 to 31st December 2018. The results obtained by utilizing the statistical packages ( Mat-lab and SPSS).

**Key words:** Survival Analysis, logistic regression, Cox regression model, Hazard Function, Leukemia..

## **INTRODUCTION**

Cancer is an abnormal cell proliferation and organized in the body to become abnormal and multiply without control outside the natural growth line, the modern theories say that the reason is a simple genetic immune system cannot notice this bug with time cause the cell's exit from the control and then the emergence of genetic medicine, to study such as those specifications and characteristics must be based on scientific analysis which depends on efficient statistical methods with quantitative and scientific measurements and criteria. This paper aims to determine the important variables that influence the leukemia disease applying logistic regression and cox regression models.

## **2. Background Information:**

In this section, essential information that is related to this research will be describes. It consists of several sub-sections. The first sub-section conclude Logistic regression model, Cox Regression, Similarities and differences between logistic regression model and Cox regression and Measures of Model Selection, the second sub-section shows the result, and the Last sub-section shows conclusion.

### **2.1 Logistics Regression Model**

Analysis of logistic regression is one of the important statistical methods that can be used in many areas of life, when there is a binary response variable or classified nature the relationship takes the formula of the logistic distribution function model and the predictor variables can be quantitative, qualitative, mixed, ordinal and binary. (mawlood 2000).

The reasons for the importance of logistical regression are: ease of use, requires a few assumptions as a model using, the logistic formulas result from a wide variety of basic assumptions about explanatory variables, and its ability to estimate regardless of the method of sampling, whether the Prospect or Retrospective.

Let the response variable  $Y$  be the binary variable, assuming that  $P(Y = 1)$  is dependent on a vector of predictor variables  $\bar{x}$ . The goal is to model

$$p(\bar{x}) \equiv P(Y = 1 | \bar{x})$$

Because  $Y$  is binary variable, modeling  $p(\bar{x})$  means modeling  $E(Y | \bar{x})$ , if model  $p(\bar{x})$  modeled as a linear function of explanatory variables,  $\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$

Then the model may be result in estimated probability value which are over a restricted value [0,1]. Then to avoid this the logistic function can be used, it is assume that

$$p(\bar{x}) = \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}$$

Where  $x_1, \dots, x_p$  represent explanatory variables,  $p(\bar{x})$  lies between [0,1] or satisfy the probability condition, and by making transformation for  $p(\bar{x})$  we obtain another function: (David, W & Hosmer, JR. 2013)

$$\dots(1) \quad \log\left(\frac{p(\bar{x})}{1 - p(\bar{x})}\right) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$$

This model is called the logistic regression model.

### **2.1.1 Maximum Likelihood Estimation Method for Parameters**

Estimation of parameters in logistic regression (the coefficients,  $\beta_0, \beta_1, \dots, \beta_p$ ) can be estimated by maximum likelihood method, (albayaty 2012)

$$\prod_{i=1}^n \left\{ p(\bar{x}_i)^{y_i} [1 - p(\bar{x}_i)]^{1-y_i} \right\}$$

$$\dots(2) \quad \log_e L(\beta) = \sum_{i=1}^n Y_i (\mathbf{X}_i' \beta) - \sum_{i=1}^n \log_e [1 + \exp(\mathbf{X}_i' \beta)]$$

(The mles are determined numerically, by maximizing the log likelihood.)

by taking the first derivative of the log maximum likelihood equation (2 - 31) and their equivalents by zero we get the equations are nonlinear in the parameters the solution can be estimated numerically and therefore resort to the use of Newton Rafson iterative method and after a few cycles of succession produced appropriate estimates of the parameters.

### 2.1.2 Evaluating the Performance of the models

in linear regression analysis ordinary least square used to fit a model, t, F tests, and residuals are used to the coefficient's and the model. The situation is different with in logistic regression the situation is different, the approximate chi-square and z tests and likelihood ratio test are used .

To test the hypothesis:

$$H_0 : \beta_0 = \dots = \beta_p = 0$$

at Least two of them are not equal zero.  $H_1$ :

chi-square test is used which is based on difference between the estimated log likelihoods corresponding to the two models, the test statistics is given by:

$$\dots(3) \quad LR(p) = -2[LnL(\alpha) - LnL(\alpha, \beta)]$$

Where:

$LnL(\alpha)$  represents the logarithm of likelihood function of the Reduced Model , which contains only the Intercept parameter.

represents the logarithm of likelihood of the final model  $LnL(\alpha, \beta)$

The LR is distributed as chi-square P, where P represents the number of explanatory variables in the model. (Archer, J. & Lemeshow, S. 2006)

To test the hypothesis:

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0$$

the test statistics is:

$$Z = \frac{\hat{\beta}_j}{\sqrt{\hat{V}ar(\hat{\beta}_j)}}$$

Where  $Z \sim_{approx}^{H_0} N(0,1)$

to testing the model is adequate for data or is not :

$H_0$ : The model is adequate for data

$H_1$ : The model is not adequate for data

the approximate chi-square test (Hosmer and Lemeshow for goodness of test) based on likelihood ratio is used.

### **2.1.3 Q test**

Q test proposed by Liu & Dyer (1988) for the logistic regression model is similar to the coefficient of determination  $R^2$  in the Linear Regression Model and is calculated as follows:

After estimating the model, the observations are arranged in descending order according to the prior possibilities of response ( $Y=1$ ), and Q can be expressed as follows:

$$Q = \frac{2(\bar{R}_1 - \bar{R}_0)}{n}$$

Where:

Average of  $n_1$  ranks of response ( $Y=1$ )  $\bar{R}_1$  :

Average of  $n - n_1$  ranks of response ( $Y=0$ )  $\bar{R}_0$  :

The value of Q is lie in the range [1-0], and it is greatly affected by anomalies with low predictive values, especially when the number of cases ( $Y=1$ ) is relatively small (Archer, J. & Lemeshow, S. 2006).

## 2.2 Survival analysis

Survival analysis is usually deals with the analysis of data in times of events in the history of individual life. The survival analysis and modeling the time it takes events occur; this typical event is death, which is derived from the name ' survival ' analysis.

The most interesting survival modeling deals with the relationship between survival time and one or more variables, commonly called covariates. The Cox proportional hazard model (introduced by Cox, 1972) widely applicable and the method most frequently used in survival analysis.

Let  $T$  be a random variable represents survival time of an event, with probability density function  $f(t)$  and cumulative function  $F(t)=\Pr(T \leq t)$ , the survival function  $S(t)$  is defined as (John, F. 2014).:

$$S(t) = \Pr(T > t) = 1 - F(t)$$

The hazard function is the function that symbolized as  $h(t)$  and gives the failure rate for the survival time, which is defined as the probability of a failure during a small period of time (conditional failure rate) assuming that the individual might have remained alive until the beginning of the period, as well as the individual fail in the so small time per unit time given that the individual have remained alive until time (t);

$$h_{(t)} = \lim_{\Delta t \rightarrow 0} \frac{pr(t \leq T \leq t + \Delta t | T \geq t)}{\Delta t} \quad (4)$$

also known as immediate risk, or conditional failure rate and death rate of a certain age, the hazard function gives the risk of failure for each unit of time during operation, which play an important role in The data remain, in practice when there is no censored observations the risk function is the percentage of patients

who die from per unit time, knowing they had remained alive until the beginning of the period or (Peter, D. W. 1998):

$$h(t) = \frac{\text{number of patients dying per unit time in the interval}}{\text{number of patients surviving at } t}$$

### 2.2.1 Cox regression model

The Cox proportional hazards regression model is the most convenient way to build regression models for survival data, time to-event outcome, based upon the values of given covariates.

The corresponding survival functions are related as follows:

$$S_{(t|x)} = S_{o(t)} \exp(\sum_{i=1}^p B_i X_i) \quad \dots(5)$$

One subject's hazard is a multiplicative replica of another's; comparing subject j to subject m, the model is stated as:

$$\frac{h(t|x_j)}{h(t|x_m)} = \frac{\exp(x_j B_x)}{\exp(x_m B_x)} \quad \dots(6)$$

A parametric regression model based on the exponential distribution (John, F. 2014):

$$\log h_{i(t)} = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}$$

Or equivalent:

$$\begin{aligned} h_{i(t)} &= \exp(\alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}) \\ &= e^\alpha + e^{\beta_1 x_{i1}} + e^{\beta_2 x_{i2}} + \dots + e^{\beta_k x_{ik}} \end{aligned}$$

Where

- i indexes subjects;
- $x_{i1}, x_{i2}, \dots, x_{ik}$  Are the values of the covariates for the i<sup>th</sup> subject.

Let  $h(t|x_t)$  denote the hazard rate at time for an individual have covariate value  $x_t$

$$h(t|x_t) = h_0(t) * \exp(\beta'x)$$

Here  $x_t = (x_{1t}, x_{2t}, \dots, x_{kt})$ ,  $\beta = (\beta_1, \beta_2, \dots, \beta_k)$

k is the total number of the covariates.

$\beta_t$  Is the constant Proportional effect of treatment.

$h_0(t)$  is called the baseline hazard; it is the hazard for the respective individual when all independent variable values are equal to zero (Peter, D. W. 1998).

As indicated by Hosmer and Lemeshow (1999); in Cox regression the measure that is analogous to  $R^2$  in multiple regression is:

$$R_p^2 = 1 - \exp\left[\frac{2}{n}(L_0 - L_p)\right]$$

Where:

$L_0$  is the log likelihood of the model with no covariates.

$L_p$  is the log likelihood of the model that includes the covariates.

$n$  is the number of observations (censored or not).

## **2.3 Similarities and differences between logistic regression model and Cox regression**

### **2.3.1 The similarities between the two models**

1. Nature of the model: The two models are considered multiple regression methods.
2. Independent variables: The nature and quality of the independent variables in the two models is not specified

The variables are quantitative, descriptive, or mixed between the two.

. Odds Ratio (OR) =  $\exp(\beta)$  & Hazard Ratio (HR) =  $\exp(\beta)$ : Explanation Tool-3

4. The nature of the interpretation: Logistic regression One of its objectives is the OR estimation. Cox regression is one of its objectives. The HR estimate and the two ratios explain the effect of independent variables on the dependent variable(AltIbani 2011).



5. Likelihood Function: The Likelihood function in conditional logistic regression is the same as the Likelihood function in Cox regression.
6. Goodness of fit testing of the overall model: The Chi-square test is use to goodness of the overall model's fit in the both models.
7. Regression coefficient significant test: The test of the significance of each independent variable is Wald Test.
8. Time: Both models study the relationship between the time of the event and the independent variables.
9. Time period: If the follow-up period is short for less than five years or time measures, the results in both models are the same or the estimated regression coefficients in the two methods are identical in terms of sign and amount (Hosmer, D. and Lemeshow, S. 2007 & Wienke, A. 2011).

### **2.3.2 Differences between the two models**

1. The dependent variable: In the logistic regression, the dependent variable is descriptive binary, while in Cox regression the dependent variable (descriptive binary + time until event occurs).
2. Censoring: The logistic regression model does not deal with censored data while Cox regression model deals with it.
3. Method of estimation of regression coefficients: Logistic regression coefficients are estimated using the Maximum Likelihood Estimation method, while Cox regression coefficients are estimated using the Partial Likelihood Estimation
4. Time: In the logistic regression, the probability of an event occurring at one point in time is described the time of the interview is the instantaneous time of the event, ie the time is constant (zero), while in the Cox regression the probability of the event being described at any time point prior to the occurrence of the event means that time is changing (Altibani S.2011 & Abbas A. K., 2012)

5. Time period: If the follow-up period is longer than five years, the results from the use of both models will be different and this is what agreed by both: (Hosmer, D. and Lemeshow, S. 2007 )

where they agreed that:

- As the follow-up period increases, the logistic regression coefficients increase in amount other than coefficients Cox gradient which remains as it is.
- As the follow-up period increases, the standard error estimate in the Cox regression model is reduced comparing with a logistic regression model.

## **2.4 Measures of the Model Selection**

In this study two measures for selecting the best model was used by comparing the accuracy and performance of methods for. Comparing models simply involves calculating the measures for each model; the model with the lowest value is chosen as the best model (Lee, T. & Wang, W. 2003).

### **2..4.1 Akaike’s Information Criterion**

The Akaike Information Criterion (AIC) compares the quality of a set of statistical models to each other. For example, Akaike’s Information Criterion is calculated as follows:

$$\text{AIC} = -2\log\text{-likelihood} + 2K \quad \dots(7)$$

Where:

K: is the number of model parameters (the number of variables in the model plus the intercept).

Log-likelihood is a measure of model fit, this is usually obtained from statistical output (Moore, 2016).

### **2..4.2 The Bayesian Information Criterion**

The Bayesian information criterion (BIC) is one of the most widely known and pervasively used equipment in statistical model selection. BIC is computed for each of the models corresponding to the minimum value of BIC is selected.

$$\text{BIC} = -2\ln L + 2 * \ln N * k \quad \dots(8)$$

Where L is the value of the likelihood, N is the number of recorded measurements, and k is the number of estimated parameters.

Comparing models with the Bayesian information criterion simply involves calculating the BIC for each model; the model with the lowest BIC is chosen as the best model (Lee, T. & Wang, W. 2003).

### **3. Results and Discussions:**

Is in this section two models in parametric methods binary logistic regression and Cox regression (were used for survival analysis data. Also; all the corresponding results have been given and a comparison between the two models has been done. Two statistical measures (AIC and BIC) were used to evaluate the best model in our data. The following programs were used to analyze digital medical image processing:

1. Mat-lab.
2. SPSS (V: 23).

#### **3.1 Data Collection**

The data set for this study about cancer disease was collected in **Nanakali Hospital**. The data consisted of **1475** cases have been collected during **5** years period; beginning from **1<sup>st</sup> January 2013** through **31<sup>st</sup> December 2017** on all leukemia patients admitted to the hospital with follow up period until **1<sup>st</sup> April 2018**. Out of those patients **160** died during the study and **1315** survived or under censored. The survival time was measured in months and this data was left censored; defined as the period between the diagnosis date of chest cancer and the occurrence of the event of interest (death) or until the end of the study.

**Table (1) The Response Variables Measured for these Data at Diagnosis:**

Name Variables	Description
----------------	-------------

Age	Age of patient at diagnosis leukemia
Gender	Male = (1), Female =(2)
Event status	Died = (1), Alive = (2)
Surgery	Made surgery = (1), Does not make surgery = (2)
Radio	Took Radiotherapy= (1), Does not take Radiotherapy = (2)
Chemo	injected Chemotherapy = (1), Does not inject Chemotherapy = (2)
Hormone	Used hormone = (1), Does not use hormone = (2)

### 3.2 Application of Logistic Regression Model

In applying binary logistic regression, we reached to these results and interpretations:

The statistical hypothesis of goodness of fit test for logistic regression model define as:

H0: the model adequate

H1: the model not adequate

**Table (2) Hosmer and Lemeshow Test for goodness of fit**

Chi-square	df	Sig.
6.275	8	.617

If the lack of significance or P.value in Hosmer and Lemeshow test statistics for Goodness-of-Fit is 0.05 or less, we reject the null hypothesis that there is no difference between the observed and predicted values of the dependent; if it is

greater, as we want, we fail to reject the null hypothesis that there is no difference, implying that the model's estimates fit the data at an acceptable level. As here, table (2) shows that the lack of significance of the Chi-Squared test indicates that the model is a good fit, and we notes that the value of L-H statistics is equal to 6.275.

Cox and snells  $R^2_{C\&S}$  is based on calculating the proportion of unexplained variance that is reduced by adding variables to the model the Cox and Snell  $R^2_{C\&S}$  value is an alternative index of goodness of fit related to the R square value in the linear regression, Nagelkerke R square is a modification of  $R^2_{C\&S}$ , more appropriate for logistic regression. The value of Cox and snells  $R^2_{C\&S}$  and Nagelkerke R square coefficients are (27%) and (33%) which is an acceptable.

**Table (3) Variables in the Equation**

Variables	B	S.E.	Wald	df	Sig.	Exp(B)
gender	.128	.171	.558	1	.455	1.137
age	-.003	.004	.549	1	.459	.997
surgey	.083	.040	4.344	1	.037	1.086
radio	-.010	.036	.082	1	.775	.990
chemo	-.005	.049	.012	1	.912	.995
hormon	.038	.049	.589	1	.443	1.038
Constant	1.794	.398	20.336	1	.000	6.012

In Table (3) the Wald statistic and the corresponding significance level test the significance of each of the independents in the model. The ratio of the logistic coefficient B to its standard error S.E., squared, equals the Wald statistic. If the Wald statistic is significant (i.e., less than 0.05) then the parameter is significant in

the model. Of the independent variable, surgery is significant but (gender, age, radio, chemo and hormone) are not.

The "Exp(B)" column is label for the odds ratio of the row independent with the dependent variable (status). It is the predicted change in odds for a unit increase in the corresponding independent variable. Odds ratios less than 1 correspond to decreases and odds ratios more than 1.0 correspond to increases in odds. Odds ratios close to 1.0 indicate that unit changes in that independent variable do not affect the dependent variable.

**The prediction equation is:**

Log (odds of a reaction to drug) = 1.947 + 0.128 (gender) -0.003 (age) + 0.083 (surgery) - 0.01 (radio) - 0.005 (chemo) + 0.038 (hormon).

As with any regression, the positive coefficients indicate a positive relationship with the dependent variable.

The prediction equation with significant factores can be written is:

$$\hat{y} = 1.947 + 0.083 (\text{surgery}).$$

**3.3 Application of Cox Regression Model**

In applying Cox regression, we reached to these results and interpretations:

**Table (4) Case Processing Summary in Cox-regression available in analysis**

Case Processing Summary			
		N	Percent
Cases available in analysis	Event	160	10.8%
	Censored	1315	89.2%
	Total	1475	100.0%

The status variable identifies whether the event has occurred for a given case. If the event has not occurred, the case is said to be censored. The case processing summary Table (4) shows that 160 cases are event data and 1315 cases are censored; these are patients who have not died.

The Omnibus Tests of Model Coefficients Table (5) is uses chi-square tests to see if there is a significant difference between the factors of the baseline model and the new model, if the new model has a significantly reduction in the baseline then it suggests that the new model is explaining more of the variance in the outcome and is an improvement, the test hypotheses is:

$H_0$ : Explanatory variables are included in the model.

$H_1$ : Explanatory variables are not included in the model.

**Table (5) Omnibus Tests of Model Coefficients**

<b>Omnibus Tests of Model Coefficients</b>									
Step	-2 Log Likelihood	Overall (score)			Change From Previous Step			Change From Block	
		Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square	
1 <sup>a</sup>	2057.611	9.345	1	.002	10.792	1	.001	10.792	
2 <sup>b</sup>	2052.375	14.919	2	.001	5.236	1	.022	16.028	
a. Variable(s) Entered at Step Number 1: surgery									
b. Variable(s) Entered at Step Number 2: chemo									

The chi-square change from previous step is the difference between the -2 log-likelihood of the model at the previous step and the current step. If the step was to add a variable, the inclusion makes sense if the significance of the change is less than 0.05. If the step was to remove a variable, the exclusion makes sense if the significance of the change is greater than 0.10. In the first step and second, *surgery* , *chemo* are added to the model.

**Table (6) Variables in the Equation in survival model**

Variables in the Equation								
X	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
gender	-.166	.161	1.057	1	.210	.847	.617	1.162
Age	.003	.004	.420	1	.308	1.003	.995	1.011
surgey	-.142	.039	13.496	1	.000	.868	.805	.936
Radio	.016	.033	.246	1	.739	1.016	.953	1.084
chemo	.082	.044	3.424	1	.017	1.086	.995	1.185
hormon	-.042	.042	.999	1	.301	.959	.883	1.041

The model-building process takes place in six treatments (gender, age, surgery, radio, chemo and hormone).

Table (6) shows model fitting and parameter estimation of Cox-PH model.

In order to understand the effects of individual predictors, we look at column Exp(B).

- ❖ The value of Exp(B) for surgery means that the leukemia hazard for all patients that make a surgery are 0.868 months.



- ❖ chemo factor is one of the affecting factors to the risk or death in leukemia diseases.  $\text{Exp}(0.082) = 1.086$  means increases in the risk of the death for patient with chemo factor or who injected Chemotherapy. However, the p-value of chemo factor equal to  $0.017 \leq \alpha = 0.05$  is statistically significant and the 95% confidence interval for the hazard ratio included.
- ❖ The remaining four prognostic factors which are (gender, age, radiology and hormone) are not statistically significant in the model or have not effects the survival of the leukemia patients in this study.

When  $x$  is the vector of all the fixed covariates (gender, age, surgery, radiology, chemo and hormone)  $\beta$  is the corresponding vector of the regression coefficient for the fixed covariates, the model will be:

$$h_i(t) = h_0(t) * \exp(B'x)$$

$$h_i(t) = h_0(t) \exp(-.166 \text{ gender} + 0.03 \text{ age} - 0.142 \text{ surgery} + 0.16 \text{ radiology} + +0.82 \text{ chemo} - 0.042 \text{ hormone})$$

The Cox regression model with significant factor, the factors that the score statistics with significance values less than 0.05, which is two factors (surgery and chemo), will be as follows:

$$h_i(t) = h_0(t) \exp(- 0.142 \text{ surgery} - 0.042 \text{ hormone})$$

**Table (7) Measures of the Model Selection (AIC and BIC)**

Models	AIC	BIC
logistic Regression model	1006.607	35.403
Cox Regression model	2092.375	43.53

Table (7), determines which of the two models is more suitable in our data based on two measures; the AIC and BIC were computed; the minimum value of AIC and BIC are selected. The results shows that the logistic Regression model is the best model because its AIC and BIC values are lower in comparison with AIC and BIC in Cox Regression model.

#### 4. Conclusions

From the results obtained from the applied chapter, we can infer that the most important conclusions reached by the study are:

1. The Cox-PH and AFT model does not identify the same prognostic factors that influence in leukemia survival time.
2. The Significance variables are estimated in logistic regression model is:  
$$\hat{y} = 1.947 + 0.083 (\text{surgery}).$$
3. The results of cox regression model of this study illustrated that the most common factors that effecting on the leukemia in using data set are (surgery and hormone).
4. Comparing the results of the logistic regression model with the Cox regression model based on the AIC and BIC criterions it is concluded that logistic regression model has smallest value of AIC and BIC criterions or is the most suitable model for the data set used in this study.

#### REFERENCES

1. التلباني, شادي (2011): "دراسة مقارنة بين نموذج الانحدار اللوجستي ونموذج انحدار كوكس لدراسة أهم العوامل الاقتصادية والديموغرافية المؤثرة على معرفة واتجاهات الشباب نحو قضايا الصحة الإنجابية" رسالة دكتوراه غير منشوره , جامعة أبو بكر بلقايد , تلمسان , الجزائر.
2. عباس, علي خظر(2012): "استخدام نموذج الانحدار اللوجستي في التنبؤ بالدوال ذات المتغيرات الاقتصادية التابعة النوعية" مجلة جامعة كركوك للعلوم الادارية والاقتصادية, المجلد (2) العدد(2).
3. مولود, كوردستان ابراهيم (2000): "استخدام التحليل المميز لتشخيص اهم العوامل المؤثرة في التصنيف السريري لمرض القلب" رسالة ماجستير علوم في الاحصاء, جامعة صلاح الدين / اربيل, كلية الادارة والاقتصاد.
4. Archer, J. & Lemeshow, S. (2006) "**Goodness-of-fit test for a logistic regression model fitted using survey sample data**". The Stata Journal, Vol. 6, NO. 1, PP. 97–105.

5. David, W & Hosmer, JR. (2013) "**Applied Logistic Regression**". Third Edition, John Wiley & Sons, Inc., Hoboken, New Jersey, Canada.
6. John, F. (2014). "**Introduction to Survival analysis**". new work: sociology 761.
7. Hosmer, D. & Lemeshow, S. (2000). "**Applied Logistic Regression**". Second Edition, New York: Johnson Wiley & Sons, Inc.
8. Hosmer, D. & Lemeshow, S. & May, R. (2007). "**Applied Survival Analysis: Regression Modeling of Time to Event Data**". Second Edition, Wiley, New York, USA .
9. Lee, T. & Wang, W. (2003). "**Statistical Methods for Survival Data Analysis**". Second Edition. Wiley, New York.
10. Menard, Scott. (2002): "**Applied Logistic Regression Analysis**" Second Edition, Sage Publication, Inc.
11. Moore D. F. (2016) "**Applied Survival Analysis Using R**". Springer International Publishing Switzerland.
12. Peter, D. W. (1998). "**Predicting Recidivism Using Survival Models**". London: Springer-Verlag.
13. Wienke, A. (2011). "**Frailty Models in Survival Analysis**". london: Chapman & Hall/CRC.

### پوخته

ئامانجی ئەم توێژینە وە یە بریتی بوو لە هەلسەنگاندنی کاریگەری چەند فاکتەرێکی گرنگ لەسەر مانەوای شێرپەنجەیی خوین، بۆیە ئەم توێژینە وە یە خویندەوای کردوو بۆ شیکردنەوای بەکارهێنان و جێ بەجێ کردنی دوو مۆدیل بەمەبەستی شی کردنەوای زانستی. یەکەم مۆدیل بریتیه لە (مۆدیلی لۆجستی دووانی) هەر وەها مۆدیلی دووهم بریتیه لە (مۆدیلی کۆکس)، فاکتەرکانی کاریگەری شێرپەنجەیی خوین لە هەر دوو مۆدیلە کە وەکو یەك دەرئەچوون. وەدەهەرکەوت کە گۆراوی (نەشتەرگەری) بە گۆراوی کاریگەر هەژمار دەکری (مۆدیلی لۆجستی دووانی) هەر وەها (مۆدیلی کۆکس) بۆ گۆراوی پشت پێ بەستراو کە ئەویش توش بوانی نەخۆشی شێرپەنجەیی خوین.

بۆ هەلبژاردنی باشترین مۆدیل بۆ داتا کەمان دوو پێوهری ئاماری گرنگ بەکارهێنران کە ئەوانیش (AIC) و (BIC) بە بەراوردکردنی بچوکتەری نرخ دەبیتە باشترین مۆدیل. ئەم داتایەمان وەرگرت لە نەخۆشخانەی نانەکەلی تایبەت بە نەخۆشییەکانی شێرپەنجە، داتایە کەمان لە 1475 نەخۆش کۆکردەو



# QALAAI ZANIST SCIENTIFIC JOURNAL

*A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq*

Vol. ( 4 ), Issue ( 2 ), Spring 2019

ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

---

لهسەر ئه و نهخۆشانهی که توشی شپیرهنجهی خوئن بوون له ماوهی نیوان 1 ی مانگی کانونی دووهم له 2013 تا کو 31 ی مانگی کانونی یهکههه له 2017.

ئهنجامهکانمان دۆزییهوه به بهکارهینانی دوو پرۆگرامی ئاماری (SPSS) و (Mat-Lab).

کلیلی وشهکان: شیکردنهوهی مانهوه, مۆدیلی کۆکس, مۆدیلی لۆجستی دووانی, نهخشهی مهترسی, شپیرهنجهی خوئن.