Survival Analysis and Cox Regression Model

Ali Saleh Mahdi ALtarish [1]   ,   Prof. Dr. Kamil Alakus[2]

Abstract

Through this research, we discussed the Survival analysis, and knowing the variables affecting the survival function, we reviewed the Kaplan-Meier method to find out the differences between the two groups of treatment for prostate cancer patients.

The sample taken was 38, The survival probability of the patients was estimated according to the type of treatment, and significant differences were found between them.

And, then, we took other data to predict the regression line equation for the survival time through the Cox regression method, data were processed and results were extracted by the SPSS program. A sample of size 30 was taken and the regression line equation was estimated with the presence of two variables, They are the method of treatment and marital status (married - unmarried), and the extent of their influence on the equation was concluded.

Keywords: Survival Data, Cox Regression, Survival Analysis, The Censoring, Hazard Function

Affiliation of Authors

[1,2] Institute of Graduate Studies Department of Statistics, Ondokuz Mayis University, Turkey, Samsun, 55000-55999

[1]ali_altarish@yahoo.com

[2]kamilalakus@gmail.com

[1] Corresponding Author

تحليل البقاء على قيد الحياة ونموذج انحدار كوكس

علي صالح مهدي الطارش [1] ،   الاستاذ الدكتور كامل الكوش [2]

الخلاصة

من خلال هذا البحث ٫ناقشنا تحليل البقاء على قيد الحياة ومعرفة المتغيرات التي تؤثر على دالة البقاء على قيد الحياة ٫قمنا بمراجعة طريقة كابلان ماير لمعرفة الاختلافات بين مجموعتي علاج مرضى سرطان البروستات.

العينة الماخوذة كانت 38 ٫تم تقدير احتمالية البقاء على قيد الحياة للمرضى حسب نوع العلاج ٫حيث وجدت فروق ذات دلالة احصائية بينهم.

ثم اخذنا بيانات اخرى للتنبؤء بمعادلة خط الانحدار لوقت البقاء من خلال طريقة انحدار كوكس ٫ وتمت معالجة البيانات واستخراج النتائج بواسطة برنامج SPSS. تم اخذ عينة بحجم 30 وتم تقدير معادلة خط الانحدار بوجود متغيرين وهما طريقة العلاج والحالة الاجتماعية (متزوج-غبر متزوج) وتم استنتاج مدى تاثيرهما على المعادلة .

الكلمات المفتاحية : بيانات البقاء ٫ انحدار كوكس ٫تحليل البقاء على قيد الحياة٫ الاختفاء٫دالة المخاطر

انتساب الباحثين

[1,2] معهد الدراسات العليا قسم الاحصاء، جامعة اوندكوز مايس، تركيا، سامسون، -55000 55999

[1]ali_altarish@yahoo.com

[2]kamilalakus@gmail.com

[1] المؤلف المراسل

Introduction

Survival Analysis

Survival Analysis is the term used scientifically or by statisticians to describe the analysis of data that corresponds to time from a well-defined origin of time until the occurrence of a specific event or endpoint [1].

What is between the beginning and the end of the event and it is possible that the end of the event at the time of the occurrence of the event under study is death or birth, and either the failure to determine the final state of the singular (the time of failure) [2].

For example, to find out the effect of two or more treatments on the survival of patients, in such a case, the event of death is the meaning of the study, and with that, it is possible to obtain data of a similar form when the end is not fatal, such as pain relief or the time of the return of symptoms, and it is also possible. The application of this methodology in the analysis in other applied fields such as animal survival time in a specific study or seed storage times, … etc.

Survival analysis has two characteristics. The first is that the survival time is calculated after determining the condition under study that was a specific disease condition. One of the statistical methods concerned with studying multiple response phenomena is multiple regression, and

because of the special nature of survival data such as containing incomplete data

(Data is incomplete) In many species, it is not appropriate to use multiple regression, and the second characteristic is that most traditional methods such as regression models, analysis of variance, and (t) testing, become inappropriate in the case of the data being torsional, which means that the majority of the data is focused. on the right side of the distribution [3].

Because of these two characteristics force us not to use multiple regression models and other statistical models, this ultimately led to the conclusion of the Cox regression model, and in order to better understand the concept of survival analysis, we must know some of the concepts of survival analysis, including the basic survival functions, and the observational data [4].

Basic Survival Functions

Assuming that (T) represents a positive continuous random variable (survival time) which has a probability density function (f(t)) and a cumulative function (F (t)). To clarify the concept of the basic survival function, the following concepts must be understood:

Survivor Function

It is formulated as the probability of survival beyond time (t) (or at least until time t) and can be found in: -

$$S(t) = p(T \geq t) = 1 - F(t) = \int_t^\infty f(x)\,dx, \qquad t \geq 0 \dots \dots (1)$$

$$f(t) = \frac{\partial F(t)}{\partial t} = \frac{\partial}{\partial t}\big(1 - S(t)\big) = -\frac{\partial S(t)}{\partial t} \dots \dots \dots \dots \dots \dots \dots (2)$$

$$f(t) = \frac{-\partial S(t)}{\partial t} = f(x) \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots (3)$$

where f(t), probability density function, and F(t) also cumulative distribution function which is

written as followss:

$$F(t) = 1 - S(t) \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots . (4)$$

Hazard Function

İt is also called the risk rate or the instantaneous current rate of occurrence of an event, it is denoted as h (t) and can be found by the following:

$$h(t) = \lim_{\Delta t \to 0} \frac{p\left(t \le T < t + \Delta t /_{T} \ge t\right)}{\Delta t} = \frac{f(t)}{S(t)} = \frac{-\bar{S}(t)}{S(t)} \dots\dots\dots (5)$$

The risk function is also known as the immediate failure rate or death rate (conditional probability). The risk function gives the probability of failure risk per unit over the period, which has an important role in analyzing the survival time data.

Cumulative Hazard Function

It is the sum of the risks that occurs up to the arrival of time (t) and can be found by:

$$H(t) = \int_0^t f(x)dx = \int_0^t \frac{f(x)}{S(x)} dx \dots\dots\dots\dots\dots\dots\dots . (6)$$

$$= \int_0^t \frac{1}{S(x)} \left[\frac{\partial S(x)}{\partial x}\right] dx = -LnS(t) \dots\dots\dots\dots\dots\dots (7)$$

$$\therefore S(t) = exp\left(-H(t)\right) \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (8)$$

$$and \ F(t) = h(t) * exp\left(-H(t)\right) \dots\dots\dots\dots\dots\dots\dots . (9)$$

The Expectation of Life

Suppose μ represents the mean or predicted value of t, where:

$$\mu = \int_0^\infty t * f(t) \, dt \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots . \dots (10)$$

$$\mu = E(T) = \int_0^\infty S(t) \, dt \quad , \qquad S(0) = 1, \ \ S(\infty) = 0 \dots\dots\dots\dots\dots . (11)$$

where S (t) represents the probability that the item will survive at time t, while E(t) represents the life expectancy of the item.

The Censoring

The concept of invisibility, observation, or observation is often repeated in semi-parametric

models, especially in the Cox model, and here we have to clarify this concept, albeit briefly. There is vocabulary that records the occurrence of an event, while others cannot record the time when the event has been occurred.

Type the Censoring

There are several types of Censoring which can be summarized in the following points:

Censorship of the first type: This type of disappearance is related to the number of vocabulary that occurred in the event, as we have a random variable (number of vocabulary). As for the duration of the study, it is fixed and also called static Censoring, as the study stops according to a period that the researcher pre-determines before the start of the study. And it is divided into:

Censorship on the right: In some cases of survival studies, the specified study period ends, but some items have exceeded the study period, in the words another length of stay of the item with the length of time of the study.

Censorship on the left: Some items were included in the study, but the time of the occurrence of the event was not known to the researcher, for example, knowing the time of death. For people who have a specific disease, the time of infection is unknown.

Censorship of the second type: In this type of observation, a certain (fixed) number of items are specified for the study, but the time for completion of the study is unknown.

There is a third type of Censoring, but it is less frequent in studies in general and it is called Random Censoring, which is this type.

Cox Regression Model

Cox David (1972) formulated the Cox model as one of the relative risk models. It is based in its basic concept on the importance of the pre-study time as a fundamental factor, in other words, the main feature of this.

The method is to study the relationship between the time prior to the occurrence of the event under the study with one variable or another variable in its independent capacity. The nature of the data is not required, whether it is quantitative, descriptive, or mixed, and one of the most prominent models used in this type of studies is the Cox regression model [6].

Several books offer in-depth coverage of Cox's decline. These books assume basic knowledge of Statistical theory, especially with regression analysis. Collett (1994) provides a comprehensive introduction to

Themes. Hosmer and Lemeshow (1999) are almost entirely dedicated to this topic. Therneau and Grambsch (2000) presented a complete and recent discussion for this topic. We found their discussion of the rest was very useful analysis Klein and Mosheberger (1997) present a very readable calculation of survival analysis in General and includes a clear calculation of Cox's regression[7].

Survival analysis can be defined as a statistical method for analyzing data if the dependent variable is the previous time.

For the occurrence of the event under the study, and the time may be in the form of hours, days, weeks, months, and years from the beginning of the study until a period. The occurrence of an event is death, To begin with, it is important to point out the terms that can be given for survival analysis in different studies and researches according to the overlap.

Other sciences with statistics, and depending on the nature of the data, in health and life statistics and life applications it is defined.

Analysis Hazard, but in engineering and mechanical studies, it is known as (Failure Analysis Time), or (Analysis Reliability), and in (psychological studies), it is known as history analysis (Analysis History Event) or (Analysis Duration), as well as a name in economics (Analysis Transition). … etc, or anything else [8].

Properties Cox Regression Model

• semi-parametric model.

• There is no need to choose a specific probability model to represent the number of dwellings.

• The possibility of dealing with two types of separate and continuous measurements of event times.

• It is concerned with the effect of variables on risk.

• The vocabulary in which the values of the independent variables are equal has the same risk function.

Here, we will use two examples once using survival analysis by using the Kaplan-Meier method and the second example by using Cox regression analysis.

We will comment on the results to see the differences between the two methods.

Kaplan & Meier Method

This method was proposed by the two scholars (Meier & Kaplan)in 1958 (AD) and it is also called the Kaplan-Meier estimation or the estimator of the marginal product, and it is defined as the probability of survival for the individual for a certain period taking into account the fact that time has multiple periods, as it is used to analyze survival data individually For each of the study variables with the dependent variable, which represents the time if the data contain more than one variable to give us a broader understanding of the extent of the difference between the survival functions of two or more groups of the peripheral variable, and all the terms of the study, whether controlled or not, contribute to their assessment, and this is what distinguishes them from other methods of analyzing.

Other data survival data, such as life tables and the Cox regression model, use independent variables together to analyze the data. To measure the part of the injured person's life and for a certain period after treatment, the Kaplan-Meier estimate for control data can be calculated by the following formula:

$$\widehat{S}(t_i) = \prod_{i=1}^{n} \left(1 - \frac{1}{ni}\right) \dots \dots \dots \dots \dots \dots (12)$$

Suppose that we have n experimental units arranged according to the viewing times of these units in order $(t_1, t_2, \cdots, t_n)$ than the Kaplan-Meier estimate can be calculated with the following formula:

$$\widehat{S}(t_i) = \prod_{i:T_i \leq t} \left(\frac{n-i}{n_i - i + 1}\right)^{di} for\ t \leq Tn \dots \dots \dots \dots \dots \dots (13)$$

where (di) represents the rank of observations in time (t_i).

Through the above, we will, through the following example, we will use a survival analysis by using Kaplan-Meier to compare two treatments that were tested on prostate cancer patients and take the second example, we find out the variables affecting the survival function by using Cox regression on the same example, and by using the SPSS program[12].

Items of Research

The most important thing that made this model popularly used in life analysis is the possibility of benefiting from it by estimating a portion of its mediators only to be on a semi-parametric model, which is useful for knowing the degree and shape of the effect for each independent variable on risk. The formula for the Cox relative risk regression model is given as follows [9]:

$$h(t, X) = h_0(t) * exp\left(\sum_{i=1}^{p} \beta_i \ X_i\right) \dots \dots \dots \dots \dots \dots \dots \dots (14)$$

where $X = (X_1, X_2, \cdots, X_p)$ the variables predictors $\beta = (\beta_i, \beta_2, \beta_3 \dots \dots \beta_p)$ parameter of the model. $h_0(t)$ The fundamental risk function (it is an intermediate function that takes the form of a non-negative probability distribution).

Here, it must be said that $(exp(\sum_{i=1}^{p} \beta_i \ X_i))$ is the amount of relative risk, which does not depend on

time (t), meaning that the change in the explanatory variables (Xi) of increase or decrease does not depend on a specific point in time.

Also, the ratio between the two risk rates is fixed, It does not depend on time and it is as follows:

$$\frac{h(t/x_1)}{h(t/x_2)} = \frac{h_0 exp(\beta_1 X_1)}{h_0 exp(\beta_2 X_2)} = e^{B(X_1 - X_2)} \ \dots \dots \dots \dots \dots \dots (15)$$

This use of logarithmic transformation can provide a contribution for solving the difficulties encountered in the interpretation of Cox's regression model, by taking the logarithm of the risk function in the Cox regression model to read: To become, the risk function as in the equation by using a logarithmic transformation :

A linear synthesis of the basic variables, noting that it compresses the outliers values of the data,

which helps to solve many problems, including the presence of extreme observations, the heterogeneity of the variance, and the skewness of the distribution. The change in the independent variables is offset by a change in the logarithm of the risk function and not the dependent variable itself.

$$Log \ h(t) = (B_1 X_1 + B_2 X_2 + \cdots \cdots \cdots + B_p X_p) \dots \dots \dots \dots \dots (16)$$

To become, the risk function as in the equation by using a logarithmic transformation:

A linear synthesis of the basic variables, noting that it compresses the outliers values of the data, helps to solve many problems, including the presence of extreme observations, the heterogeneity of the variance, and the skewness of the distribution. The change in the independent variables is offset by a change in the logarithm of the risk function and not the dependent variable itself.

Method Estimating the Cox Regression Model

In 1975, the scientist Cox proposed the method Likelihood Partial to estimate the parameters of the model[10].

Cox regression, which is still used for this purpose till now, as this method highlights the likelihood of predicting or predicting the values of the variable dependent through the independent variables in the absence of knowledge of the nature of the underlying risk function (h0(t)).

The partial function is taken for the following form:

$$l_p(\beta) = \prod_{i=1}^{m} \frac{e^{xi\beta}}{\sum_{j \in R(ti)} e^{xi\beta}} \ldots \ldots \ldots \ldots \ldots \ldots (17)$$

Where, (m) is the number of failures, and (X_i) represents the variable values of the item whose time of slowing down (t_i) and the risk group

(R_{ti}), meaning that all the terms at the risk of failure are:

$$R_{(t_i)} = \{j = t_j \geq t_i\} \ldots \ldots \ldots \ldots \ldots \ldots \ldots (18)$$

And by taking the natural logarithm of the equivalent becomes

$$log\left(l_p(\beta)\right) = \sum_{i=1}^{m} \left\{ x_i\beta - log \sum_{j \in R(ti)} e^{xi\beta} \right\} \ldots \ldots \ldots \ldots \ldots \ldots \ldots .. (19)$$

After differentiating them and setting them equal to zero, we obtain the values of the parameters of the model.

Choosing The Best Independent Variables to Estimate the Model

Including most study variables in the regression model estimation which leads to more comprehensive and general results about proximal phenomena

Introducing a large number of variables into the analysis process is costly in terms of time and effort, and not most of them.

Variables are important or some of them have no effect in the relationship with the dependent variable, and some variables have a great internal correlation.

The most important method used for this is the regression method The gradient depends on two

methods, The Forward Selection Method and The Backward Deletion Method.

The Forward Selection Method

In this way, the regression equation begins, which is free of the independent variables (explanatory), in the first step (Step1), and in the next step (Step2), only one explanatory variable with the lowest value (value-P) is chosen from among the eating of the independent variables and in the next step. (Step3) A second variable is inserted into the model, which corresponds to the lowest value (value-P) among the remaining variables and above it for the variable in the previous step, and so we continue to enter the variables to insert a variable in each step.

Until the ideal model is reached, provided that all the variables included in the model have a value (value-P) less than the level of the specified significance ($\propto$).

The Backward Deletion Method

In contrast to the previous method (backward deletion), as this method begins with the assumption that the model contains all independent variables in the first step (Step1), then in the next step (Step2) deletes the corresponding variable greater than -P (value), whose value is greater than the level of significance The determinant ($\propto$), and thus the independent variables will continue to be deleted once every step until the optimum model is reached that includes only the significant independent variables from them, i.e. those that have a value (P-value) less than the specified level of significance ($\propto$).

Interpretation of Cox Regression Model Parameters

The amount of change in the dependent variable (Y_i) as a result of changing the independent variable in one unit represents the interpretation of the parameters of the linear regression model as it is known, while the interpretation of the regression coefficients of the Cox model takes another form or formula as a result of the difference in the regression equation, and it can be expressed as the amount of change in the value of risk Modified logarithm$\left(h(t/x)\right)$ when the variable changes in one unit[1].

If the value of the Cox regression coefficient is negative and when the value of the independent variable increases by one unit, this leads to a decrease in the risk ratio, which means that the situation is heading towards the better, while if the value is relative to Cox, the regression coefficient is positive and when the value of the independent variable increases by one unit One, this increases the risk, which means the situation gets worse.

Evaluate the Estimated Cox Regression Model

After reconciling the model, it is necessary to know how well the model represents the data, which means that we can rely on the model estimated to describe the original data, and there are two ways to do this :[1],[11]

Morality Test of the Estimated Model

It is determined here whether the independent variables have a role in explaining the behavior of the dependent variable, where the fit of the model as a whole is known as computing the value of the coefficient of determination, which we can express here with the following formula:

where R^2: is the value of the coefficient of determination.

$Log \, L_f$ represents the logarithm of the partial weighting of the model that contains all the variables.

$Log \, L_0$: represents the logarithm of the partial weighting of the model that does not contain all the variables.

n: The total number of observations.

$$R^2 = 1 - exp\left[\frac{2}{n}\left(Log \, L_0 - Log \, L_f\right)\right] \dots \dots \dots \dots \dots \dots \dots (20)$$

Test the Significantly of Estimated Parameters

The second method for verifying the significance of the estimated model is by evaluating the significance of each parameter of the estimated model separately. Its statistic is usually used as a parent (Statistic Wald), knowing that its parent statistic follows the chi-square distribution with one degree of freedom, and is calculated by:

$$W^2 = \left[\frac{\widehat{\beta}_j}{S.E_{\widehat{\beta}_j}}\right]^2 \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots .(21)$$

Where, $S.E_{\widehat{\beta}_j}$: represents the standard error of the model parameters.

$\widehat{\beta}_j$: Represents the estimated parameter of the model.

Example [1]: The data in this example is a randomized controlled clinical trial comparing prostate cancer treatments in 1967 by the Collaborative Research Group for the Urology Department. The treatments used in the study were a placebo and (1.0) mg of (diethylstilbestrol). The time frame of the study was the date on which the patient was randomly selected for the study and the final point was the death of the patient. The data also contained variables of age, hemoglobin level, and the size of the primary tumor in square cm.

The table gives the data for (38) patients, where the survival times are given in months. The survival times are considered for patients who died from other causes or were lost during the monitoring process. A variable related to the disease status at the end of the study. Because of cancer, zero if he is properly controlled, and the individual treated with DES takes a value of 2 and takes a value of 1 if he is treated with a placebo.

Example [2]: Two treatment methods will be taken with a sample of 30 people who are monitored weekly, taking into account whether they are married or not. Married will be expressed as 1 and unmarried with 2.

We will use the SPSS statistical program to find out the statistical indicators and to analyze the data so that we can know which of the statistical indicators are of significance and statistical importance that can be relied upon regarding the study.

We entered the data and chose the Kaplan-Meier method to measure the differences between the two treatment methods, and the results are as follows in a table (1):

Table (1): Case Processing Summary

Treatment	Total N	N of Events	Censored	
			N	**Percent**
The placebo	18	5	13	72.2%
des	20	1	19	95.0%
Overall	38	6	32	84.2%

In this table, it becomes clear to us that (18) of those treated with placebo died (5) and lost (13) while those treated with (DES) (20) died of them (1).

From the life table, we note the following, See Table (2):

Table (2): Survival Table

	1	15	2.000	censored	.	.	0	17
	2	7	14.000	died	.941	.057	1	16
	3	37	23.000	censored	.	.	1	15
	4	24	24.000	censored	.	.	1	14
	5	28	26.000	died	.874	.084	2	13
	6	34	36.000	died	.807	.101	3	12
	7	29	42.000	died	.739	.113	4	11
	8	8	43.000	censored	.	.	4	10
The placebo	9	6	51.000	censored	.	.	4	9
	10	10	52.000	censored	.	.	4	8
	11	4	58.000	censored	.	.	4	7
	12	11	59.000	censored	.	.	4	6
	13	27	61.000	censored	.	.	4	5
	14	38	62.000	censored	.	.	4	4
	15	1	65.000	censored	.	.	4	3
	16	16	67.000	censored	.	.	4	2
	17	22	67.000	censored	.	.	4	1
	18	21	69.000	died	.000	.000	5	0
	1	32	5.000	censored	.	.	0	19
	2	9	16.000	censored	.	.	0	18
des	3	19	28.000	censored	.	.	0	17
	4	25	45.000	censored	.	.	0	16
	5	20	50.000	died	.938	.061	1	15
	6	5	51.000	censored	.	.	1	14

7	14	51.000	censored	.	.	1	13
8	33	54.000	censored	.	.	1	12
9	12	55.000	censored	.	.	1	11
10	30	57.000	censored	.	.	1	10
11	3	60.000	censored	.	.	1	9
12	2	61.000	censored	.	.	1	8
13	26	64.000	censored	.	.	1	7
14	23	65.000	censored	.	.	1	6
15	17	66.000	censored	.	.	1	5
16	18	66.000	censored	.	.	1	4
17	36	67.000	censored	.	.	1	3
18	13	68.000	censored	.	.	1	2
19	31	70.000	censored	.	.	1	1
20	35	70.000	censored	.	.	1	0

We note in this life table, that patient No. 15 lost or dropped out of the study after two months from the start of the study, patient No. 7 reached death or died after 14 months, and that the probability of survival is (94%), and so for the rest of the placebo treatment, we see that patient No. 21 died after 69 months, so we notice that the probability of surviving after 69 months is (0).

As for those treated with treatment (DES), patient No. 32 was lost from the study after 5 months, and patient No. 20 died after 50 months, and the probability of survival is (93%).

In Table (3), we note that the probability of survival for those treated with placebo treatment is 58 months, and the probability of survival for those treated with DES equals 68 months. look table (3):

Table (3): Means and Medians for Survival Time

Treatment	Mean[a]				Median			
			95% Confidence Interval				95% Confidence Interval	
	Estimate	Std. Error	Lower Bound	Upper Bound	Estimate	Std. Error	Lower Bound	Upper Bound
The placebo	58.840	5.067	48.909	68.772	69.000	.000	.	.
des	68.750	1.210	66.378	71.122	.	.	.	.
Overall	64.203	2.375	59.549	68.858	.	.	.	.

a. Estimation is limited to the largest survival time if it is censored.

From it, we note that the probability of survival for treatment (DES) is greater than the placebo treatment

:

From the following table, we note the level of importance between the two treatments, see table (4)

Table (4): Overall Comparisons

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	4.421	1	.035
Breslow (Generalized Wilcoxon)	3.416	1	.065
Tarone-Ware	3.659	1	.056

Test of equality of survival distributions for the different levels of Treatment.

We note that the value of (P_value=0.035) for the test (log-rank) is less than (0.05), meaning that there are statistically significant differences in favor of the DES treatment.

As for the other two tests, the value of (P_value) is greater than (0.05), and this means that there are no statistically significant differences between the two treatments.

In this Figure, through the graph of the survival function, the survival time for therapists is calculated by using the two methods. See Figure (1):
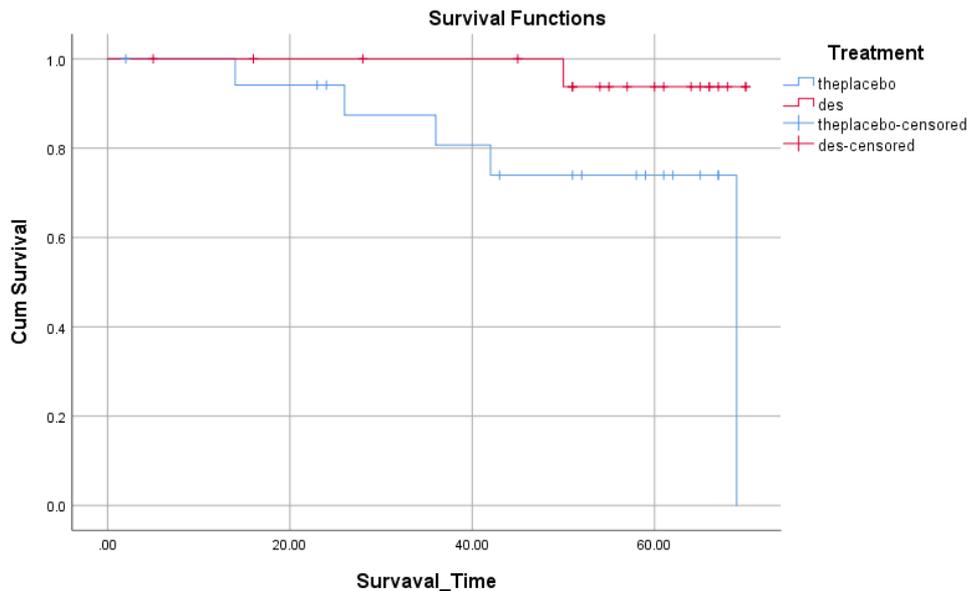

Figure (1) Survival Functions

In Example (2), the data will be defined on the spss program in the same way as before.

The general method will be taken in the program, and neither forward progress nor reverse deletion will be determined. We will get the results and interpret them as follows, Look at the table (5):

Table (5): Case Processing Summary

		N	Percent
Cases available in the analysis	**Event**	19	63.3%
	Censored	10	33.3%
	Total	29	96.7%
Cases dropped	**Cases with missing values**	0	0.0%
	Cases with negative time	0	0.0%
	Censored cases before the earliest event in a stratum	1	3.3%
	Total	1	3.3%
Total		30	100.0%

a. Dependent Variable: TIME

From this table, we note that those who reached death (19), where their percentage is (63%), and those who lost (10) and (33%), and we notice that there is one case lost before the first death.

In this table, the efficiency of the prediction model is tested, see table (6):

Table (6): Coefficient

Log-Likelihood	Overall (score)			Change From Previous Step			Change From Previous Block		
	Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square	df	Sig.
58.952	23.602	2	.000	23.281	2	.000	23.281	2	.000

a. Beginning Block Number 1. Method = Enter

Since the value of (p_value=0.00) is less than (0.05), the current model is highly efficient.

In this table, the deviation coefficients are calculated for the predictive equation, and through the value of (h), see table (7):

Table (7): Variables in the Equation

	B	SE	Wald	df	Sig.	TREATM ENT
TREATMENT	3.794	1.145	10.975	1	.001	nts44.451
SOCIL_STATUS	-1.529	.759	4.060	1	.044	.217

We can say that the variables (treatment methods and marital status) can predict the time model through which survival is carried out.

In this Figure, through the graph of the survival function, the survival time for therapists is calculated by using the two methods. in this drawing, we know which is more likely to survive, as we note that married people, which represents the color blue, are more likely to survive than unmarried people. See Figure (2):
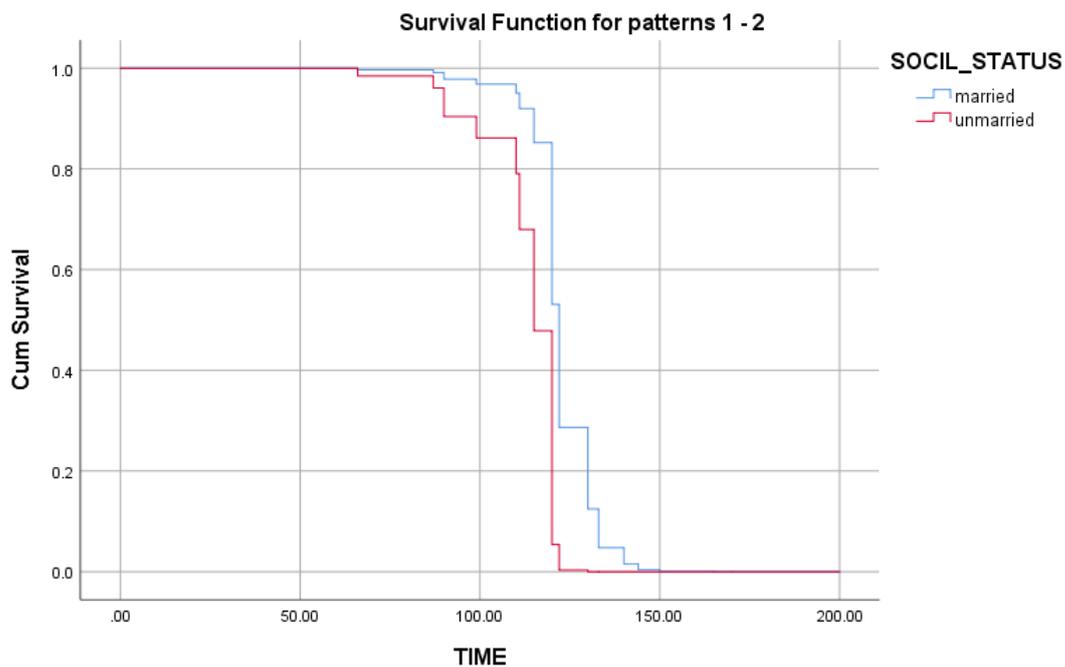

Figure (2) Survival Functions

Conclusions

- Through this research, it turns out that the Kaplan- Meier method is one of the important ways to show the differences between groups to show the knowledge of the highest probability of survival.

- Cox regression analysis is one of the best predictive methods for semi-parametric data.

- The larger the data, the better the result, especially in the field of medical research.
- The variable of marital status affects the likelihood of survival of patients, as well as treatment methods.

Recommendations

Through the research, we recommend using the Kaplan- Meier method to know the effect of different methods of treatment on the survival time of the dependent variable.

As for knowing the variables that affect the regression line equation, the Cox method is considered one of the best methods for predicting the estimated equation and knowing the independent variables that affect the survival time

References

[1] Collett, D. (2015). Modeling survival data in medical research. CRC press.

[2] Abd ElHafeez, S., Torino, C., D'Arrigo, G., Bolignano, D., Provenzano, F., Mattace-Raso, F., & Tripepi, G. (2012). An overview on standard statistical methods for assessing exposure-outcome link in survival analysis (Part II): the Kaplan-Meier analysis and the Cox regression method. Aging clinical and experimental research, 24(3), 203-206.

[3] Murtada Farhan Huwai,2017, Using Cox Regression Model (COX) and The Kaplan-Meier Method to Study Survival Data

[4] Collett, D. (2015). Modeling survival data in medical research. CRC press

[5] Sen, P. K. (1995). Censoring in theory and practice: Statistical perspectives and controversies. Lecture Notes-Monograph Series, 177-192.

[6] Lin, D. Y. (1994). Cox regression analysis of multivariate failure time data: the marginal approach. Statistics in medicine, 13(21), 2233-2247.

[7] Statistical Software | Sample Size Software | NCSS, COX regression

[8] Vivekanandan, T., & Narayanan, S. J. (2019). A Hybrid Risk Assessment Model for Cardiovascular Disease Using Cox Regression Analysis and a 2-means clustering algorithm. Computers in biology and medicine, 113, 103400.

[9] (PDF) Finding the Least Possible Hazards in Cox Regression Model (researchgate.net)

[10] Cox, D. R , "Partial likelihood" , Biometrika , Vol-62 ,No-3, P(269- 276) , 1975.

[11] Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. Cancer Chemother Rep, 50(3), 163-170.

[12] Bian, H. (2015). Survival analysis using SPSS. East Carolina University.