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)

Study and Analysis of the Chest Cancer Data Using Survival Models

Asst. Prof. Dr. Kurdistan Ibrahim Mawlood

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

kurdistan.mawlood@su.edu.krd

Researcher *Sami Ali Obed*

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

Sami.Ali.Obed@gmail.com

ARTICLE INFO

Article History:
Received: 24/12/2018
Accepted: 10/1/2019
Published: Spring 2019
Doi:
10.25212/lfu.qzj.4.2.24

Keywords:

Survival Analysis, Cox-proportional hazard, Accelerated Failure

ABSTRACT

This research aimed to estimate the effects of prognostic factors on chest cancer survival, the research studied two models in survival analysis; the Cox-Proportional Hazard (PH) model is most usable method in present time of survival data in the occurrence covariate or prognosticates aspects, and the Accelerated Failure Time (AFT) model is another substitute way for analysis of survival data. Kaplan-Meier method has been applied to survival function and hazard function for estimation, the log-rank test was used to test the differences in the survival analysis. The data was obtained from Nanakali Hospital in the period from 1st January 2013 to 31st

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)

Time, Kaplan Meier**,** Log-Rank test**,** Chest Cancer *a.*

December 2017 with follow up period until 1st April 2018. The results for Kaplan-Meier and log-rank test showed the significant difference in survival or death by chest cancer for all presented related prognostic factors. The Cox-PH and AFT model does not identify the same prognostic factors that influenced in chest cancer survival.

The Cox Proportional Hazards model displays a significant lack of fit while the accelerated failure time model describes the data well. AFT with Weibull distribution was chosen to be the best model for our data by using Tow model selection criterion; Akaike Information Criterion (AIC) and Bayesian information criterion (BIC). Also, the results performed by the statistical package in Mat-lab, Stat-graphic and SPSS, which was used to analyze the data.

Key words: Survival Analysis, Cox-proportional hazard, Accelerated Failure Time, Kaplan Meier**,** Log-Rank test**,** Chest Cancer.

.

INTRODUCTION

In scientific and organic studies, the analysis of event time data or survival statistics aimed to describe the hazard (risk) function of event times in population. Survival evaluation is a branch of statistical that targeted on studying data where the outcome variable is the time until the occurrence of an event of interest; in medical research the event is often thought of as "death" (Biost, 2004).

In these cases, this typical start time is when the patient enters the hospital, and the end point is when the patient died or living (censored), an overview of

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)

survival analysis is discussed along with important models that are relevant to the present study (Alhasawi, 2015).

We present two tests and two models in survival analysis which are

Kaplan Meier estimation is best statistical method adapted in survival analysis to investigate the data, and the Log-Rank test applied to find the similarity and differences between the two samples such as treatment and control groups (Qi, 2009).

And exploration and description of two parametric models which are the Cox-PH model is currently the best broadly utilized for the investigation of survival analysis in the occurrence of covariates or prognostic factors, and Accelerated failure time models is choice to Cox-PH show the basic principles of AFT model are (Weibull AFT, Exponential AFT, Log-normal AFT, Gamma AFT, Log-Logistic AFT). Also Akaike's Information Criterion and Bayesian information criterion are used to select the best model between two models (Cox- Proportional Hazard and Accelerated Failure Time Models) (Wienke, 2011).

2: Methodology

This section studied some basic concepts of survival analysis; survival function, hazard function and some tests and methods used to analysis survival data.

2.1: Survival Analysis

Survival analysis is normally defined as a set of ways for examining data where the result variable is the time until the event of an occasion of premium (Biost, 2004).

The occasion can be death, event of an ailment, marriage, separate, and so on. The time-to-event or survival time can be estimated in days, weeks, months, years, and so on.

It is implemented to analyze data in order to the time-to-event result is gained. The response variable is the time-to-event and is usually called a failure time, survival time, or event time (Biost, 2004).

2.2: Functions Related to Survival Analysis

2.2.1: Survival Function

The survival function S (t) produces the survival probability approximately to time t.

Let (T) be a non-negative random variable substitution the time until the point that our occasion of intrigue happens.

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)

The survival function is the probability that the survival time, T, is greater than the specific time t; then is characterized as:

$$S(t) = Pr(T > t) = 1 - F(t) = \int_t^{\infty} f_{(u)}\, d_{(u)} \quad ...(1)$$

Where u is treatment, S (t) is survival function, and is assumed as T is a continuous random variable through Probability Density Function (P.D.F) f(t) (John, 2014).

2.2.2: Hazard Function

The hazard function h $_{(t)}$ of survival time T gives the conditional disappointment rate; this is characterized as the likelihood of disappointment amid a little time-to-interval, adopting that the separate has made due to the start of the interval, equation is (Ekman, 2017).

$$h(t) = \frac{number\ of\ patients\ dying\ per\ unit\ time\ in\ the\ interval}{number\ of\ patients\ surviving\ at\ t} \quad ...(2)$$

2.2.3: Cumulative Hazard Function

Expect that survival time is completely continuous, in which case the estimation of the cumulative hazard function might be communicated, utilizing techniques for calculation t as:

$$\begin{aligned} H(t,x,\beta) &= \int_0^t h(u,x,\beta)du \\ &= r(x,\beta)\int_0^t h_0(u)du \\ &= r(x,\beta)H_0(t) \quad ...(3) \end{aligned}$$

Where $r(x,\beta)$ characterizes how the hazard function changes as a function of subject covariates (Ekman, 2017).

2.3: The Nature of Survival Data (Censoring data)

There are several mechanisms that can lead to censored data; there are different kinds of censoring.

2.3.1: Type I Censoring the study ends at a certain time point or, if the subjects are put on test at different time points, after a certain time has elapsed.

2.3.2: Type II Censoring when this study ends there are a pre-specified number of events.

At the point when the estimation of a perception or estimation is just mostly referred to this perception as being censored.

2.3.2.1: Right censoring a subject is right censored in the event that it is realized that failure happens at some point after the recorded follow-up time (Peter, 1998).

2.3.2.2: Left censoring is defined as a subject if it is known that the failure takes place sometime before the recorded follow-up time (Heagerty, 2005).

2.3.2.3: Interval censoring is defined as a subject is period censored if it is known that the event comes about between two times, however the exact time of failure isn't known (Heagerty, 2005).

2.3.2.4: Independent Censoring

Independent censoring has been assumed, the essence of this assumption is that after adjustment for covariates, future event risk for a censored subject does not differ from the risk among other subjects who remain in follow-up and have the same covariate values (Ekman, 2017).

2.4: Statistical Testing

Have used two survival tests in this study which are:

2.4.1: Kaplan Meier Test

Kaplan Meier (KM) in 1958 made a cooperative trail and issued a paper on how to deal with time-to-event data. That is why, they hosted the KM estimator "The Kaplan Meier estimator also called product-limit estimator" that works as a tool for calculating the frequency or the number of patients enduring medical treatment. Later on, the KM curves and estimates of survival data have become a better way of analyzing data in cohort study (Wienke, 2011).

KM is non-parametric estimator of survival function this is typically used to designate survivorship of a study people and to evaluate dual study populations (Sulaiman, 2017), (Wienke, 2011).

The KM estimator of the survival function given as the equation

$$\hat{S}_{(t_i)} = \prod_{j=1}^{i}\left(1 - \frac{d_i}{R_j}\right) \qquad \ldots(4)$$

Where

R_j : The total number of individuals alive at the start of the interval.

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)

d_i: The number of individuals who died.

2.4.2: The Log-Rank Test

The Log-rank test is a non–parametric method for testing the null hypothesis that the groups being compared are illustrations from a similar population as regards survival experience.

The log-rank test is the most normally used test for comparing survival distributions. It is relevant to data where there is dynamic censoring and provides the same weight to initial and overdue failures (Vittinghoff, 2004).

A statistic for the equivalence of the death rates in the two groups is

$$\chi^2 = \frac{(O_A - E_A)^2}{V_A} \qquad \text{...(5)}$$

Where O_A observation of failure time and E_A expected of failure time and Var_A is the variance.

Which is approximately a χ^2, the log rank statistic approaches to chi–square distribution in a single degree of freedom; hazard ratio sampling variability is given by (Abbas, 2012).

2.5: Models in Survival Analysis

There are some models in survival analysis, they are as follows:

2.5.1: Parametric Survival Models

Parametric methods assume that the basic distribution of survival times follows certainly known probability distributions. Popular ones include the AFT model, cox-PH model, exponential, Weibull, and lognormal distributions (Vittinghoff, 2004).

2.5.1.1: Accelerated Failure Time

It is another popular regression model, often, used to analyze survival data, also, AFT model relate the lifetime distribution to the explanatory variable (stress, covariate). This distribution can be defined by the survival, cumulative distribution, or probability density functions, is best seen if they are formulated in terms of the hazard rate function (Emmanuel, 2017).

Regarding Ti as a random variable representing the (possibly unobserved) survival time of the ith unit, Since Ti must be non-negative value, and it should be considered modeling its logarithm using a customary linear model:

$$\log Ti = x_i{'}\beta + \varepsilon_i \qquad \text{...(6)}$$

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)

In case, ε_i is advisable error term and x_i is covariate factor, Ti is survival time.

The distribution of survival time to be specified (exponential, weibull, log-logistic, log-normal and gamma AFT model) (Cleves, 2010).

In this work we used Weibull AFT model.

Weibull AFT model

The Weibull distribution is called by a scale parameter λ and shape parameter p. If p is a smaller amount than 1 immediate hazard monotonically decreases with time, if p equals 1 instantaneous hazard is constant over time (equivalent to the exponential distribution) and if p is greater than 1 instantaneous hazard increases with time.

$h(t) = h_0 \, exp^{x\beta}$ ...(7) where $h_0 = \lambda p t^{p-1}$

Where β includes an intercept term β_0 (Dhillon, 2000).

Probability density function $f(t) = \lambda \upsilon t^{\upsilon-1} e^{-\lambda t^{\upsilon}} \ (\lambda > 0, \upsilon > 0)$

Survival function $S(t) = e^{-\lambda t^{\upsilon}}$

Hazard function $\mu(t) = \lambda \upsilon t^{\upsilon-1}$

Cumulative hazard function F (t) = $1 - e^{-\left(\frac{t}{\lambda}\right)^{r}} \quad 0 < t < \infty$

Expectation $E(t) = \lambda^{-\frac{1}{\upsilon}} \Gamma(1 + \frac{1}{\upsilon})$

Variance $V(T) = \lambda^{\frac{-2}{\upsilon}} (\Gamma(1 + \frac{2}{\upsilon}) - \Gamma(1 + \frac{1}{\upsilon})^2)$

Where Γ is the gamma function with $\Gamma(k) = \int_0^{\infty} s^{k-1} e^{-s} d_s \ (k > 0)$ (Wienke, 2011) .

2.5.1.2: Parametric Proportional Hazards Models (Cox-PH model)

It is a flexible tool for measuring the connection of multiple predictors to a censored data, time-to-event result.

The Cox proportional hazard method is beneficial for modeling the time to distinct event, based upon the values of given covariates (Alhasawi, 2015).

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)

The corresponding survival functions are related as follows:

$$S_{(t|x)} = S_{o(t)}{}^{\exp\left(\sum_{i=1}^{p} BiXi\right)} \qquad ...(8)$$

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

$$\frac{h_{(t|x_j)}}{h_{(t|x_m)}} = \frac{\exp(x_j B_x)}{\exp(x_m B_x)}$$

This parametric regression model constructed on the exponential distribution:

$$h\ (t) = h_0\ exp^{x\beta} \qquad ...(9)$$

in the equation above

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

Or else equivalent to:

$$h_{i(t)} = \exp(\alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik})$$
$$= e^{\alpha} + e^{\beta_1 x_{i1}} + e^{\beta_2 x_{i2}} + \cdots + e^{\beta_k x_{ik}} \qquad ...(10)$$

Where

- i indexes subjects;
- $x_{i1}, x_{i2}, \ldots, x_{ik}$ are the values of the covariates for the i th subject (Ekman, 2017).

2.5.2.1: Partial Likelihood Estimate for Cox-PH model

Suitable to the Cox-PH model, the assessment of the bias line (h_0 (t)) and β is wanted to try to enlarge the sameness function for the detected data concurrently considering h_0 (t). In the same way, extra populaces approach is presented via Cox-PH that the partial likelihood function that does not depend on h_0 (t) is acquired. This partial likelihood is a strategy progressed to make interpretation about the regression parameters in the occurrence of nuisance parameters h_0 (t) in the Cox-PH model, the partial likelihood function will be constructed based on the proportional hazards model (John, 2014).

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)

$$h_t dt = \frac{F(t)dt}{S(t)} \quad \ldots(11)$$

Where S (t) is the survival function up to a period t and F (t) is cumulative survival function.

2.5.2: Non- Parametric Survival Analysis

In the implementation of survival analysis, it is consistently a virtuous idea to acquaint numerical or graphical outcome of the survival times for the participants. Normally, survival data are usefully summarized via approximations of the survival function and hazard function. There are three non-parametric methods for describing time to event data which are (Kaplan –Meier method, Nelson Aalen method, Life table method) (Qi, 2009).

2.5.3: Semi-Parametric Models of Survival Analysis

On a regular basis one is concerned in relating two or more groups of times-to-event. If the groups are similar, except for the treatment under study, then, the nonparametric methods are used directly. Not only the subjects in the groups have some additional characteristics but also affect their result.

Such variables may be applied as covariates (explanatory variables, confounders, risk factors, independent variables) in clearing up the response (dependent) variable. After adjustment for these possible explanatory variables, the assessment of survival times between groups should be less biased and more exact than a simple comparison (Cleves, 2010).

2.6: Measures of the Model Selection

There are some measures for selecting the best model by comparing the accuracy and performance of several estimation methods for any data set:

2.6.1: Akaike's Information Criterion

A better way of looking at the model search procedure is to compute a quantity known as the Akaike Information Criterion (AIC) examines the state of a set of statistical models together, for instance; you might be interested in what

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)

variables contribute to low socioeconomic status and how the variables contribute to that status.

However, the AIC will be chosen as the best model from a set, or, when all of the used models are poor, the best of a terrible bundle will be chosen. That is why; the selection of the best model is considered running a hypothesis test to understand the relationship between the variables in the used model and the required result.

Akaike's Information Criterion is as follows:

$$AIC = -2(Log - likelihood) + 2k \qquad ...(12)$$

wherever:

K: is the number of model parameters i.e. 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).

2.6.2: The Bayesian Information Criterion

The Bayesian statistics criterion (BIC) is one of the most widely recognized and pervasively used equipment in statistical model selection. Its reputation derived from its computational simplicity and effective performance in lots of modeling frameworks in practice (Moore, 2016).

The Bayesian Information Criterion is as follows:

$$BIC=-2*\ln L+2*\ln N*k \qquad ...(13)$$

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

Contrasting models with the Bayesian statistics criterion simply includes calculating the BIC for each model; the model with the lowest BIC is chosen as the best model (Ibrahim, 2001).

3: Application

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)

This section includes a statistical study about the analysis of the data, we show here some of the techniques of survival analysis for cancer data especially chest cancer. This is done by using two tests; Kaplan Meier estimator was used to estimate the mean and median survival time data, and Log-rank test to comparing the levels of treatment, and two models in parametric models for survival analysis data which are (Cox PH model and the AFT model). All the corresponding effects and comparisons in the main methods are provided including Cox PH and AFT model.

We used two the statistical measures (AIC and BIC) for evaluating the best survival model in the data, and hence selecting the best model is used. The result of each method was performed by statistical package in (Mat-lab), (Stat-graphic) and (SPSS).

3.1: Data Collection

The data set for this study was collected in **Nanakali Hospital,** it is for "Cancer" disease. The data consisted of **590** cases have been collected during **5** years periods beginning from **1st** January **2013** through **31st** December **2017** on all chest cancer patients admitted to hospital with follow up period until **1st** April **2018** of those patients **502** died during the study and **88** survived or under censored. The survival time was measured in months and defined as the period between the diagnosis date of chest cancer and the occurrence of the event of interest (death from chest cancer) or until the end of the study.

The response variables measured for these data at diagnosis are

Age	Age of patient at diagnosis chest cancer
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

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)

Hormone	Used hormone = (1) , Does not use hormone = (2)
Immune	Took immune system = (1) , Does not take immune system = (2)

3.2: Kaplan Meier Test

KM is non-parametric way of survival function that is usually adapted to illustrate the survival of study population and to compare a couple of cases.

The most commonly summary statistic used in survival analysis is the mean and median survival analysis. The mean time to event requires that all times to events are known, the mean admission time will allow us to estimate how many months are needed to patient until death with given admissions incidence.

Table (1)
The Means and Medians for Survival Time for (Surgery) in each group

Surgery	Mean				Median			
	Estimate	S.E.	95% C.I.		Estimate	.E.	95% C.I.	
			Lower Bound	Upper Bound			Lower Bound	Upper Bound
Yes	14.439	1.000	12.478	16.400	11.000	580	9.864	12.136
No	14.956	1.282	12.443	17.468	11.000	535	9.952	12.048
Unknown	14.985	.995	13.035	16.935	16.000	483	15.053	16.947
Overall	14.603	.682	13.265	15.940	12.000	571	10.881	13.119

Table (1) gives the results of KM test for surgery factor applied to a data set of size 590 patients, this table shows that the Estimated mean time for patients who made a surgery, is 14.439 months while who does not make the surgery is 14.956 months with the confidence interval (12.47, 16.4) for made a surgery and

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)

(12.44, 17.46) for does not make a surgery under probability 95%.In contrast, the median survival time for both made surgery and does not make surgery groups are equal to 11 months.

Table (2)
The Means and Medians for Survival Time for (gender) in each group

Gender	Mean				Median			
	Estimate	S.E.	95% C.I.		Estimate	S.E.	95% C.I.	
			Lower Bound	Upper Bound			Lower Bound	Upper Bound
Male	14.198	.902	12.430	15.967	12.000	.778	10.475	13.525
Female	14.826	.955	12.954	16.699	12.000	.715	10.598	13.402
Overall	14.603	.682	13.265	15.940	12.000	.571	10.881	13.119

Table (2) explain the estimated mean time until death for males is 14.198 months while for females is 14.82 months with the confidence interval (12.43, 15.96) for male and (12.95, 16.99) for female under probability 95%.In contrast, the median estimated time between chest cancer and death for both male and female groups are equal to 12 months.

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)

3.3: The Log-Rank Test

The log-rank test examines the observed and expected number of happenings for each group using the Chi-square although the estimations for the expected frequencies are specific.

Table (3)
The Log-rank test for (Surgery)

	Chi Square	**d.f.**	**Sig.**
Log Rank	1.450	2	**.0484**
Test of equality of survival distributions for levels of surgery.			

Table (3) shows that the p-value **0.0484** $\leq$ **0.05** which points out that there is a significant difference between the pair of groups (made surgery and does not make surgery) the evaluated time until death is 14.439 months for made surgery and 14.956 months for doesn't make surgery, i.e. the patients who does not make surgery have an increased chance of survival.

Table (4)

The Log-rank test for (Gender)

	Chi Square	**d.f.**	**Sig.**
Log Rank	2.254	1	**.041**
Test of equality of survival distributions for levels of gender.			

Table (4) explain that the p-value is **0.041** $\leq$ **0.05** which indicates that there is a significant difference between the two groups (male and female) on having a short time to event. The estimated time until death is 14.198 months for male and 14.82 months for female, i.e. female have an increased chance of survival of two samples.

3.4: Cox Proportional Hazard Model

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)

The Cox-PH model is a well-identified statistical procedure for discovering the relation among the survival of a patient and few illustrative factors. A Cox-PH model states s an estimation of the treatment effect on survival after the amendment for the other descriptive variables.

The model-building process takes place in six treatments (Surgery, Radio, Chemo, Hormone, Immune, Age and Gender).

Table (5)
Results of fitting a Cox PH model

Variables in the Equation								
							95.0% for Exp(B)	
x	**B**	**SE**	**Wald**	**df**	**Sig.**	**Exp(B)**	**Lower**	**pp**
Surgery	.081	.023	12.911	1	**.000**	.923	.883	64
Radio	-.022	.021	1.070	1	**.031**	.978	.938	020
Chemo	.015	.033	.219	1	.640	1.015	.952	083
Hormone	.064	.033	3.711	1	**.044**	1.066	.999	137
Immune	.036	.032	1.245	1	.265	1.036	.973	103
Gender	-.061	.093	.432	1	**.011**	.941	.784	129
Age(binned)	-.018	.048	.144	1	**.047**	.982	.894	079

Table (5) model fitting and parameter estimation of Cox PH model, the sign of the regression coefficients is positive sign means that the hazard (risk or death) is higher, but if the sign is negative it means that hazard is lower.

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)

To understand the effects of each patient, Exp (B) is the expected change in the hazard for a minimum risk.

- The value of Exp (B) for surgery gives the meaning of the chest cancer hazard for all patients that made a surgery are 0.923 months.
- Gender is one of the effected factor to the risk or death in chest cancer diseases decease by Exp(-0.061) = 0.941 which is decrease in the risk of the death for patient with (male or female), the P-value is equal to 0.011 it means that there is a greater risk of death in chest cancer in both sex.
- Age is one of the effected factor to the risk in chest cancer diseases decease by Exp(-0.018) = 0.982 which is decrease in the risk of the death for patient with (age binned).The significant value is 0.047<= α = 0.05 so there is significant effect on chest cancer.
- The estimated hazard in the Radio group is, Exp (-0.022) = 0. 978, which is a 97% drop in the risk after adjustment for the other explanatory variables in the model of the death for patient that took radiotherapy. However, the p-value of 0.031 is statistically significant and the 95% confidence interval for the hazard ratio contains, and the estimation of hazard increases by Exp (0.015) = 1.0155 for Injected Chemotherapy factor.

- The estimate of hazard in the hormone factor, exp (0.064) = 1. 066, which is decrease in the risk of the death for patient with used hormone factor, and the estimation of hazard increases by exp (0.036) = 1.036 for took immune factor.

Here we can see that the (chemotherapy and immune system) are not significant but for patients with chemotherapy and immune system are at higher risk than the other factors.

In case, x is the vector of the entire fixed covariates (surgery, radiology, chemo, hormone, sex and immune) and β is the vector of the regression coefficient leading to the fixed covariates.

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

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)

$$h_i(t) = h_0(t) \exp(0.08\, surgery - 0.022\, radio + 0.015 chemo + 0.064 hormone + 0.0$$
$$+ 0.018\, age - 0.061 gender$$

Two variables are not accepted by the above model because the score statistics with the values of greater than 0.05 which are two factors (chemotherapy and immune system).

As shows in the table above column (Wald) test for significant of coefficients of Cox (PH) model, if the maximum value in Wald column it is significant factor , surgery is one of the significant factor in our study because has a greater value in Wald test column is (12.91) with significant value is (0.000<=0.05).

And the results showed that there are significant differences for Gender, Radiotherapy, Age, and Hormone.

3.5: Accelerated Failure Time Model

The AFT model is used to show the terms of difference between treatments in survival time. The collected data fitted using (exponential, Weibull, log-logistic, log-normal and gamma AFT model).

To selection the best model is broadly depended on the value of Akaike's Information Criterion (AIC) and likelihood ratio test (LR). AIC and LR are applied in order to choose one model from our models in AFT.

Table (6)
The (LR) and (AIC) tests, for comparing AFT Model

Distribution	No. of paramet	LR	AIC
Exponential	1	-510.8978	1025.78
Weibull	2	**-390.5901**	**785.18**
Gamma	2	-480.358	964.7

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)

Log normal	2	-753.9048	1511.85
Log logistic	2	-639.8797	1283.74

Table (6) compared AFT models by statistical criteria log-likelihood ratio test (LR) and Akaike information criterion (AIC).

The smaller LR and AIC is the better, both the likelihood ratio test and the AIC are tools for choosing between two or more models and both are based on the log-likelihood calculations, and explained that the webuil AFT model is better model according to AIC=785.18 and LR=-390.59 compared with two models.

However, it is somewhat better than Gamma model, it is also noted that the log-normal and log-logistic models are sufficient enough.

Table (7)
Estimate the Coefficient of Accelerated Failure Time Model

Treatment	**B**	**Std. Error**	**Sig.**
Surgery	.062	.017	**.012**
Radio	-.034	.019	.073
Chemo	.007	.029	.798
Hormone	.058	.029	**.042**
Immune	.044	.027	.881
Age	.037	.034	**.035**
Gender	-.007	.075	.921

Table (7) presents the coefficients of AFT model to determine the relative importance of the significant factors, a factor which has a low p-value is similarly to be a meaningful outcome to the model because the changes of the component values are associated with changes in the response variables.

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)

The significant column, (p-value $\leq \alpha$) means that(Surgery, *Hormone, and Age*) factor has a significant coefficient for chest cancer disease in the model, and the rest four remained are (Radio , Chemo , Immune and Gender) are not statistical significant in Weibull AFT model.

Then the Weibull AFT model can be written as follows:

$$\log T_i = xi'\beta + \varepsilon_i$$

$$\log(time) = 0.062\, surgery - 0.034\, radio + 0.007 chemo + 0.058 hormone + 0.044 im$$
$$+ 0.037\, age - 0.007 gender$$

3.6: Comparing Models

There are several measures to compare survival functions between two or among models, in this study two measures; Akaike's information criterion and Bayesian information criterion are used for Comparing, the Weibull AFT model with Cox PH model and as follows:

Table (8)

Comparing two models Cox PH model and Weibull AFT Model by AIC and BIC

Models	NO.Parameter	Log -likelihood ratio test	AIC	BIC
Cox-PH	3	-5428.851	10863.7	55.49
Weibull AFT	2	-390.5901	785.18	31.48

Practically the objective of table (8), determines which of the two models is more suitable in our data (Cox PH model or Weibull AFT Model).

Comparing models with the Akaike's information criterion and Bayesian information criterion is made by calculating each measure for both models.

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)

The model with the lowest AIC and BIC are considered the best model, in our results shows that the Weibull AFT is the best model because AIC=785.18 is lowest value and BIC=31.48 is lowest value compares with AIC and BIC in Cox (PH) model.

In order to detect the significant factors on cancer disease after getting these results we find that (surgery, hormone, age) are the most significant factors in Weibull AFT model after comparing the value of significant with P-value.

4: Conclusion

During conducting the survival data and according to the results from the practical part the following conclusions have been drawn:

1. The comparison tests by Kaplan Meier method show that there are a statistically significant difference in (survival times) between age and gender of patients made surgery and took radiotherapy.
2. There are significant differences between levels of treatment (Surgery, Radio, Age, and Gender) by Log rank test.
3. Cox proportional hazard is popular method to analyze survival data. The Cox PH model may be used for many applications because of the relationship between the risks of an event over time.
4. Depending on Cox (PH) model the results of this study indicated that the most common factors that effected on the chest cancer are (Surgery, Radio, Hormone and Gender).
5. Provided that on accelerated failure time model the results of this paper illustrating that the most common factors that effected on the chest cancer are(surgery, hormone and age)
6. The distribution of the data was Weibull distribution by testing the data in the Stat graphic program, and the Weibull AFT model fits better and describes the data best.
7. Comparing the Cox PH model with the AFT model based on the AIC and BIC it is concluded that (Weibull AFT model) is the most suitable model for our data set that was used in this study.

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)

REFERENCES

1. Abbas, D. N.–D. (2012). Analysis of Breast Cancer Data using Kaplan–Meier Survival Analysis. Journal of Kufa for Mathematics and Computer, 7-14.

2. Alhasawi, E. (2015). Survival Analysis Approaches for Prostate Cancer. Sudbury, Ontario, Canada: Laurentian University.

3.Biost, A. (2004). Introduction to Survival Analysis. new yourk: A Stata Press Publication.

4.Cleves, M. &. (2010). An Introduction to Survival Analysis Using Stata. Texas: A Stata Press Publication.

5. Dhillon, B. S. (2000). Design Reliability Fundemential and Application. Ottawa, Ontario, Canada: Library of Congress Cataloging-in-Publication Data.

6.Ekman, A. (2017). Cox Proportional Hazard Model. Umea: John Wiley & Sons, Inc.

7.Emmanuel, B. K. (2017). Survival Analysis with Interval-Censored *Data.* new york : Chapman & Hall/Crc.

8.Heagerty, P. (2005). Survival Analysis. Va/Uw Summer.

9.Ibrahim, J. G.-H. (2001). Bayesian Survival Analysis. New York: Springer Science+Business Media, LLC.

10.John, F. (2014). Introduction to Survival analysis. new work: sociology 761.

11.Moore, D. F. (2016). Applied Survival Analysis Using R. Piscataway, NJ,: © Springer International Publishing Switzerland.

12.Qi, J. (2009). Comparison of Proportional Hazards and Accelerated Failure Time Models. Saskatchewan, 1-89.

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)

13.Sulaiman, E. I. (2017). The Kaplan Meier Estimate in Survival Analysis. Biometrics & Biostatistics International Journal, 1-5.

14.Vittinghoff, E. &. (2004). Statistics for Biology and Health ,Second edition. New York: © Springer Science+Business Media, LLC 2004, 2012.

15.Wienke, A. (2011). Frailty Models in Survival Analysis. london: Chapman & Hall/CRC.

پوختە

ئامانجی ئەم توێژینەوەیە بریتی بوو لە هەڵسەنگاندنی کاریگەری چەند فاکتەرێکی گرنگ لەسەر مانەوەی شێرپەنجەی سنگ, بۆیە ئەم توێژینەوەیە خوێندنەوەی کردووە بۆ شیکردنەوەی دوو مۆدێل بەمەبەستی شی کردنەوەی زانستی. یەکەم مۆدیل بریتیە لە کۆکس (Cox-Proportional Hazard) کە باوترین ڕێبازە لە ئێستادا بۆ شیکردنەوەی داتایی مانەوە لە ژیاندا لە بارەی ئەو گۆڕاوانەی کە هەن, هەروەها مودێلی دووەم بریتیە لە (Accelerated Failure Time)کەجێگرەوەی ئەوەی یەکەمە و بەکاردێت بۆ شیکردنەوەی هەمان داتا.

ڕێگای کابلان مایر (Kaplan-Meier) بەکارهێنرا بۆ خەملاندن لە لایەك لەسەر نەخشەی مانەوە لە ژیان و هەڵسەنگاندنی مەترسی ,لەلایەکی ترەوە تێستی(Log-Rank test) بەکار هێنرا بە مەبەستی زانینی جیاوازی نێوان ئاستەکانی شی کردنەوەی مانەوە لە ژیان.

ئەنجامەکانی هەریەکە لە (Kaplan-Meier و (Log-Rank test جیاوازیەکی بایەخداریان پێشاندا لەسەر مانەوە یاخود مردن بە هۆی نەخۆشی شێرپەنجەی سنگ لەبارەی سەرجەم فاکتەرەکانی پێشبینی کردن کە وا پەیوەستن بەم نەخۆشیە, هەر بۆیە لە هەردوو مۆدیلی (Cox-PH) و (AFT) فاکتەرکانی کاریگەری شێرپەنجەی سنگ وەکو یەك دەرنەچوون .

ئەم داتایەمان وەرگرت لە نەخۆشخانەی نانەکەلی تایبەت بە نەخۆشییەکانی شێرپەنجە, داتایەکەمان لە 590 نەخۆش کۆکردەوە لەسەر ئەو نەخۆشانەی کە توشی شێرپەنجەی سنگ بوون لە ماوەی نێوان 1 مانگی کانونی دووەم لە 2013 تاکو 31 مانگی کانونی یەکەم لە 2017 لە ماوەی چاودێری تاکو 1 نیسانی 2018.

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)

وا دەردەکەوێت کە مودێلی کۆکس (Cox-PH) بایەخداری کەم بکات لە گونجاندن لەگەل ئەم جۆرە داتایە کاتێك دابەش بونی ئاسایی نەبێت, لەم کاتەدا (AFT) بەشێوەیەکی باشتر دەگونجی بۆ داتاکەمان. هەڵبژاردنی (AFT) لەگەڵ دابەش بونی ویبل بە باشترین مۆدێل بۆ داتاکەمان ئەویش بە بەکارهێنانی دوو پێوەری ئاماری گرنگ (AIC)و ,(BIC)ئەنجامەکانمان دۆزییەوە بە بەکارهێنانی دوو پرۆگرامی ئاماری (SPSS) و(Mat-Lab) و-Stat) graphic)

الملخص

يهدف هذا البحث إلى تقدير آثار العوامل المؤثرة على بقاء مرضى سرطان الصدر ، ودرس البحث نموذجين في تحليل البقاء وهما ، نموذج كوكس للاخطار النسبية (Cox-Proportional Hazard) هو الأسلوب الأكثر استخدامًا في الوقت الحالي لبيانات البقاء على قيد الحياة في جوانب حدوث المتغيرات أو التنبؤات ، ويعتبر نموذج وقت الفشل المعجل (Accelerated Failure Time(AFT طريقة بديلة أخرى لتحليل بيانات البقاء. تم تطبيق طريقة كابلان- مايرKaplan-Meier لتقدير دالة البقاء على قيد الحياة و دالة المخاطرة ، تم استخدام اختبار رتبة اللوغارتمي -log rank test لاختبار الاختلافات في تحليل البقاء. وأظهرت النتائج لكابلان ماير واختبار رتبة اللوغارتمي فرق كبير في البقاء أو الموت لمرضى سرطان الصدر لجميع العوامل المؤثرة المدروسة. لم يحدد النموذجين Cox-PH و AFT نفس العوامل المعنوية التي أثرت في بقاء سرطان الصدر.

تم الحصول على البيانات من مستشفى نانةكةلي الخاص بمرض "السرطان" وقد تم جمع البيانات من 590 مريض تم جمعها على جميع مرضى سرطان الصدر في الفترة من 1 يناير 2013 إلى 31 ديسمبر 2017 مع فترة متابعة حتى 1 أبريل 2018.

وتم استخدام معيارين للمفاضلة لإختيار أفضل نموذج للبيانات وهي معيار معلومات أكايكي (AIC) Akaike Information Criterion ومعيار معلومات بيز (Bayesian information criterion (BIC. أيضا ، تم الحصول على النتائج بواسطة الحزمة الإحصائية Mat-lab و Stat-graphic, SPSS، والتي تم استخدامها لتحليل البيانات.