*Ministry of Higher Educationand Scientific Research*

*University of Al- Qadisiyah College of Computer Science and Mathematic*

*Department Of Mathematics*

***Survival Function for breast cancer patients in Al-forat Al-awsat region***

*A thesis submitted*

*To The Council Of The College Of Computer Science And Mathematics University Of Al- Qadisiyah As A Partial Fulfillment Of The Requirements For The Degree Of Master Of Science In Mathematics*

*By*

***Abdulla Ahmed Hamzah***

*Supervised by*

***Prof .Dr. Ihsan Kadom Shareef Al- Qurashi***

***Al-Kenani Jawad Kadom Asst .Prof .Dr*. *Ali***

2018.A.C 1439 A.H

وزارة التعليم العالي والبحث العلمي

جامعة القادسية

كلية علوم الحاسوب والرياضيات

قسم الرياضيات

**دالة البقاء لمريضات سرطان الثدي في منطقة الفرات الاوسط**

**رسالة تقدم بها**

**عبد الله احمد حمزة**

**الى مجلس كلية علوم الحاسوب والرياضيات– جامعة القادسية**

**وهي جزء من متطلبات نيل شهادة الماجستير في علوم الرياضيات**

**إشراف**

**أ.د. أحسان كاظم شريف القرشي**

**أ. م. د. علي جواد كاظم الكناني**

1439هـ 2018م

**بسم الله الرحمن الرحيم**

**{ وَلوَلاَ فضل الله عَلَيكَ وَرَحَمتُهُ لَهَمَت طآئفَةُ مِنهُم أَن يُضلوكَ وَمايُضلونَ إلآ أنَفُسَهُم ومايَضُرُونَكَ مِن شيء وَأنَزلَ الله عَليكَ الكِتاب والحِكمَةَ وَعلمَكَ مَالَم تَكُن تَعلمُ وَكَانَ فَضلُ الله عَلَيكَ عظِيما }**

**صدق الله العلي العظيم سورة النساء الآية (113 )**

**الإهداء**

**الى من**

**تربيت في أحضانه وأكلت و من خيراته وارتويت من**

**أنهاره وطني العزيز**

العراق

**الباحث**

**شكر وتقدير**

**الحمد لله الذي قهر جميع الخلق بقدرته ، الحمد لله جبار السموات ، علام الغيوب منزل البركات كثير الخيرات رحيم ودود ، والصلاة والسلام على خير خلقه محمد وال محمد .**

**وبعد الحمد لله سبحانه وتعالى العلي القدير الذي وفقني لإنجاز هذه الرسالة ، و عرفانا بالجميل اتوجه بفائق الشكر والامتنان الى الاستاذ الدكتور احسان كاظم شريف القرشي رئيس جامعة القادسية ( رحمه الله واسكنه فسيح جناته ) لما قدم لي من جهود علمية ، واتقدم بخالص شكري وعظيم امتناني الى استاذي الفاضل الاستاذ المساعد الدكتور (علي جواد الكناني ) على قبوله الاشراف على هذه الرسالة ، و لما ابداه من ملاحظات قيمة طوال مراحل اعداد الرسالة ، ومالمسته من رعاية علمية وكان له الاثر الكبير في اخراج هذه الرسالة على النحو الذي ظهرت عليه ، اسأل الله ان يوفقه الى كل خير.**

**واتقدم بفائق الشكر والامتنان الى السادة اعضاء لجنة المناقشة لتفضلهم بالموافقة على مناقشة هذه الرسالة وعلى ملاحظاتهم القيمة التي ساهمت في اغناء الرسالة ، واتقدم بالشكر والتقدير الى عمادة كلية علوم الحاسوب وتكنلوجيا المعلومات والى اساتذتي المحترمين ، وشكري وتقديري الى الدكتور (عماد السعبري) في مستشفى الفرات الاوسط ( للأورام السرطانية ) ، الذي قدم المساعدة لي وسهل عملية حصولي وجمعي للبيانات التي تخص المرضى فجزاه الله خير الجزاء ، وختاما اتوجه بالاعتذار الى من قدم لي المساعدة ولم اذكره ومن الله التوفيق .**

**الباحث**

**المستخلص**

يعد موضوع المعّولية وتحليل البيانات باستخدام دالة البقاء من المواضيع المهمة في حياتنا اليومية ، لأنها تدخل في العديد من المجالات الهندسية والطبية والصناعية وسعى الباحث في هذه الرسالة الى تحقيق هدفين هما :

الهدف الاول هو تقدير دالة البقاء والدوال المرتبطة بها للمرضى المصابين بسرطان الثدي، وقد تم جمع البيانات لهذه الرسالة من مركز الفرات الاوسط في النجف الاشرف، وكانت البيانات المدروسة للفترة من (2016-2013) وكان العدد الكلي للمرضى في الدراسة هو(96) مريض وجميعهم توفوا خلال هذه المدة وكان المرضى جميعهم اناث، وقد ركزت هذه الرسالة على تقدير دالة البقاء والدوال المرتبطة بها (دالة الكثافة الاحتمالية f(t) ، ودالة التوزيع التراكمي F(t)، ودالة الخطورة (h(t ).

حيث تم استخدام مجموعة من الطرائق المعلمية و اللامعلمية للتقدير ، اذ استخدمت ثلاث طرائق معلميه لمعلمات توزيع ويبل ذو المعلمتين والدوال المرتبط بها وهذه الطرائق هي (طريقة الامكان الاعظم (Maximum Likelihood Method ( M.L.E، طريقة العزوم (Method Of Moment (M.O.M، وطريقة المربعات الصغرى ((O.L.S Ordinary Least Squares) .

أما الطرائق اللامعلميه فقد استخدمت ثلاث طرائق لامعلميه وهي (طريقة كيرنلMethod Kernel،طريقة نيلسون Method(NL) Nelson والطريقة التجريبية Empirical Method(EM) ) وبعد ذلك عمل الباحث مقارنة بين طرائق تقدير دالة البقاء والمحسوبة من البيانات وبالاعتماد على متوسط مربعات الخطأ ((MSE لجميع الطرائق المدروسة لبيان الافضل بين الطرائق حيث بينت نتائج المقارنة لتقدير دالة البقاء لمرضى سرطان الثدي ، ان طريقة (الامكان الاعظمM.L.E) اعطت افضل النتائج لتقدير دالة البقاء بالنسبة للطرائق المعلميه اما نتائج المقارنة بين الطرائق الامعلميه بينت ان طريقة (كيرنل نوع Cr.V) اعطت نتائج افضل .

أما الهدف الثاني فتمثل باختيار افضل نموذج انحدار يمثل العلاقة بين مجموعة من المتغيرات المدروسة وفترة البقاء على قيد الحياة للمرضى المصابين بسرطان الثدي حيث تم استخدام طريقة Lasso(Least absolute shrinkage and selection operator) وطريقة Adaptive Lasso( Adaptive Least absolute shrinkage and selection operator) لتحقيق هذا الهدف والطريقتان المذكورتان اعلاه تعملان على اختيار اهم المتغيرات المستقلة المؤثرة على فترة البقاء وكذلك تقدير المعالم في آن واحد وليس على مراحل كما هو الحال في طرق اختيار المتغيرات التقليدية .

وقد بينت النتائج الى ان كل من المتغيرات (مستلم الاستروجين – الدلالة السرطانية – الدورة الشهرية – عدد الزيارات العلاجية – العمر– مرحلة المرض – مستلم فيتامين D – الطول – المرض العائلي – نوع العلاج) من اصل المتغيرات المدروسة والتي عددها 16 متغير مستقل هي التي يعتقد ان لها تأثير على فترة بقاء المرضى المصابين بسرطان الثدي باستخدام طريقة ( Lasso )، اما بالنسبة لطريقة (Adaptive Lasso)فقد بينت ان كل من المتغيرات (الدورة الشهرية – عدد الزيارات العلاجية – العمر – مرحلة المرض – مستوى فيتامين Dالطول) هي المتغيرات الاكثر تأثيرا على فترة البقاء .

**المحتويات**

|  |  |  |
| --- | --- | --- |
| **التسلسل** | **العنوان** | **الصفحة** |
| **1** | **الفصل الاول: المقدمة العامة** | **1-13** |
| 1-1 | المقدمة | 1 |
| 2-1 | منهجية البحث | 2 |
| 3-1 | مراحل سرطان الثدي | 3 |
| 4-1 | عوامل الخطورة في سرطان الثدي | 4 |
| 1-5 | العلاج | 6 |
| 1-6 | انواع سرطان الثدي | 7 |
| 1-7 | الاستعراض المرجعي | 8 |
| **2-1** | **الفصل الثاني: الجانب النظري**  **المبحث الاول** | **14-30** |
| 1-2-1 | المقدمة | 14 |
| 2-1-2 | دالة البقاء | 14 |
| 3-1-2 | دالة الكثافة الاحتمالية | 15 |
| 4-1-2 | دالة الخطورة | 15 |
| 2-1-5 | توزيع ويبل | 17 |
| 2-1-6 | طرائق التقدير | 19 |
| 2-1-6-1 | طرائق التقديرالمعلمية | 19 |
| 2-1-6-1-1 | طريقة الامكان الاعظم | 19 |
| 2-1-6-1-2 | طريقة العزوم | 21 |
| 2-1-6-1-3 | طريقة المربعات الصغرى | 25 |
| 2-1-6-2 | طرائق التقدير الامعلميه | 27 |
| 2-1-6-2-1 | الطريقة التجريبية | 27 |
| 2-1-6-2**-**2 | طريقة نيلسون | 28 |
| 2-1-6-2-3 | طريقة كيرنل | 28 |
|  | | |
| **2-2** | **الفصل الثاني الجانب النظري**  **المبحث الثاني** | **32-35** |
| 2-2-1 | المقدمة | 32 |
| 2-2-2 | طرق اختيار افضل نموذج | 32 |
| 2-2-2-1 | الطرائق التقليدية | 33 |
| 2-2-2-1-1 | طريقة اختيار المتغير المتدرج | 33 |
| 2-2-2-1-2 | معيار معلومات اكاكي | 33 |
| 2-2-2-1-3 | معياربيز للمعلومات | 33 |
| 2-2-3 | طرائق حد الجزاء | 34 |
| 2-2-3-1 | طريقة Lasso | 34 |
| 2-2-3-2 | طريقة Adoptive Lasso | 35 |
| **3-1** | **الفصل الثالث الجانب التطبيقي**  **المبحث الاول** | **36-65** |
|  |  |  |
| 1-3-1 | المقدمة | 36 |
| 2-1-3 | وصف البيانات | 36 |
| 3-1-3 | اختبار البيانات | 37 |
| 3-1-4 | اختبار مربع كاي لحسن المطابقة | 37 |
| 5-1-3 | اختبار توزيع ويبل | 37 |
| 3-1-6 | طرائق التقدير المعلمية | 39 |
| 1-6-1-3 | تقدير المعلمات والدوال الاحتمالية لتوزيع ويبل بطريقة الامكان الاعظم | 39 |
| 3-1-6-2 | تقدير المعلمات والدوال الاحتمالية لتوزيع ويبل بطريقة العزوم | 42 |
| 3-1-6-3 | تقدير المعلمات والدوال الاحتمالية لتوزيع ويبل بطريقة المربعات الصغرى | 47 |
| 3-1-7 | رسم الدوال بالطرائق المعلميه | 51 |
| 3-1-8 | الطرائق الامعلمية | 53 |
| -1-38-1 | تقدير دالة البقاء بأستخدام الطريقة التجريبية | 53 |
| 2-8-1-3 | تقدير دالة البقاء بأستخدام طريقة نيلسون | 55 |
| 8-1-3-3 | تقدير دالة البقاء بأستخدام طريقة كيرنل | 58 |
| 3-1-9 | رسم الدوال بالطرائق اللامعلمية | 64 |
| 3-1-10 | المقارنة بين الطرائق المعلمية والطرائق الامعلمية | 64 |
| **3-2** | **الفصل الثالث الجانب التطبيقي**  **المبحث الثاني** | **66-69** |
| 3-2-1 | المقدمة | 66 |
| 3-2-2 | وصف البيانات | 66 |
| 3-2-3 | اختيار افضل نموذج | 67 |
| **4** | **الفصل الرابع** |  |
| 1-4 | الاستنتاجات | 70 |
| 4-2 | التوصيات | 71 |
|  | **المصادر** | 72 |
|  | **الملاحق** |  |
|  | ملحق يمثل البيانات والفئات العمرية حسب المحافظات |  |

**الاشكال**

|  |  |  |
| --- | --- | --- |
| **رقم الشكل** | **الشكل** | رقم الصفحة |
| 3-1-1 | يوضح رسم دالة الكثافة الاحتمالية المقدرة بطريقة الامكان الاعظم | 41 |
| 3-1-2 | يوضح رسم دالة التوزيع التراكمي المقدرة بطريقة الامكان الاعظم | 41 |
| 3-1-3 | يوضح رسم دالة البقاء المقدرة بطريقة الامكان الاعظم | 41 |
| 3-1-4 | يوضح رسم دالة الخطورة المقدرة بطريقة الامكان الاعظم | 42 |
| 3-1-5 | يوضح رسم دالة الكثافة الاحتمالية المقدرة بطريقة العزوم | 46 |
| 3-1-6 | يوضح رسم دالة التوزيع التراكمي المقدرة بطريقة العزوم | 46 |
| 3-1-7 | يوضح رسم دالة البقاء المقدرة بطريقة العزوم | 46 |
| 3-1-8 | يوضح رسم دالة الخطورة المقدرة بطريقة العزوم | 46 |
| 3-1-9 | يوضح رسم دالة الكثافة الاحتمالية المقدرة بطريقة المربعات الصغرى | 49 |
| 3-1-10 | يوضح رسم دالة التوزيع التراكمي المقدرة بطريقة المربعات الصغرى | 49 |
| 3-1-11 | يوضح رسم دالة البقاء المقدرة بطريقة المربعات الصغرى | 49 |
| 3-1-12 | يوضح رسم دالة الخطورة المقدرة بطريقة المربعات الصغرى | 50 |
| 3-1-13 | يوضح رسم دالة الكثافة الاحتمالية لجميع الطرائق المعلميه | 51 |
| 3-1-14 | يوضح رسم دالة التوزيع التراكمي لجميع الطرائق المعلميه | 51 |
| 3-1-15 | يوضح رسم دالة البقاء لجميع الطرائق المعلميه | 52 |
| 3-1-16 | يوضح رسم دالة الخطورة لجميع الطرائق المعلميه | 52 |
| 3-1-17 | يوضح رسم دالة البقاء بالطريقة التجريبية | 55 |
| 3-1-18 | يوضح رسم دالة البقاء بطريقة نيلسون | 57 |
| 3-1-19 | يوضح رسم دالة البقاء بطريقة fixed | 60 |
| 3-1-20 | يوضح رسم دالة البقاء بطريقة Cross Validation | 63 |
| 3-1-21 | يوضح رسم دالة البقاء لجميع الطرائق اللامعلميه | 64 |

**الجداول**

|  |  |  |
| --- | --- | --- |
| الرقم | الجدول | الصفحة |
| 3-1-1 | يمثل تقدير قيم الدوال بطريقة الامكان الاعظم | 39 |
| 3-1-2 | يمثل تقدير قيم الدوال بطريقة العزوم | 42 |
| 3-1-3 | يمثل تقدير قيم الدوال بطريقة المربعات الصغرى | 47 |
| 3-1-4 | يمثل دالة البقاء بالطريقة التجريبية | 53 |
| 3-1-5 | يمثل دالة البقاء بطريقة نيلسون | 55 |
| 3-1-6 | يمثل تقدير دالة البقاء بطريقة كيرنل مع عرض حزمه من نوعfixed | 58 |
| 3-1-7 | يمثل دالة البقاء بطريقة كيرنل مع عرض حزمه من نوع Cr.V | 61 |
| 3-1-8 | يمثل المقارنة بين الطرائق المعلميه واللامعلمية | 65 |

Abstract

Reliability and data analysis using survival function are considered as important topics of our daily life, as they enter in many engineering, medical and industrial fields, the researcher urged through this thesis to achive two goals:-

The first is to estimate survival function and the functions related to for breast cancer patients, he collected the data for this thesis from middle euphrates center located in Holy Najaf, the studied data were for the period (2013-2016), the total number of patients of this study was (96) patients, they all died within this time and they were all females, this study focused on the estimate of survival function and the functions related to, such as probability density function f(t), a cumulative distribution function F(t) and hazard function h(t).

Number of parametric and non - parametric methods were used to estimate survival function and the functions related to, three parametric methods of waybill distribution with two parametric and the functions related to were used to, these methods are represented in (Maximum Likelihood Method (M.L.E), Method Of Moment (M.O.M) and Ordinary Least Squares (O.L.S)).

In this study three non - parametric methods have been used (Kernel Method, Nelson Method (NL) and Empirical Method (EM), after that, the researcher made a comparison between survival function estimate methods accounted from the data depending upon the average of relative squares of error (MSE) of all studied methods to show which is the best among the methods, the results of comparison to estimate the survival function for breast cancer patients showed that Maximum Likelihood Method (M.L.E) has given the best results to estimate survival function in relation to parametric methods,while the results of comparison among non - parametric methods showed that (Kernel Method type Cr. V) has shown the best results.

The second goal is represented in selecting the best descention sample representing the relationship between a group of studied variables and alive survival time for breast cancer patients where a method of Lasso (Least absolute shrinkage and selection operator) and a method of Adaptive Lasso (Adaptive Least absolute shrinkage and selection operator) were used to achive this goal, these two methods work to select the best independent effective variables on survival time, as well as estimate the instructors at the same time, not in stages as in the methods of traditional variables selection.

The results showed that each of variables (Estrogene receiver – cancer indicators – menstrial period – number of treatment visits – age – disease stage – Vitamin D receiver – Length – family disease – type of medication) are original parts of the studied variables,(number 16 independent variable) it’s believed that it has an effect on the survival time for breast cancer patients using (Lasso) method, while (Adaptive Lasso) showed that each of variables (menstrial period – number of treatment visits – age – disease stage – Vitamin D receiver – Length) are the most effective variables on survival time.