Choosing the best method Estimation of Survival Function for breast cancer patients

اختيار افضل طريقة لتقدير دالة البقاء لمرضى سرطان الثدي

ا.م.د علي جواد كاظم الكناني عبدالله احمد حمزة
جامعة القادسية / كلية العلوم والحاسبات /قسم الرياضيات

المستخلص

يعتبر موضوع تحليل البيانات باستخدام دالة البقاء من المواضيع المهمة في حياتنا اليومية وذلك لأنها تدخل في العديد من المجالات الطبية ، الهدف من هذا البحث هو تقدير دالة البقاء للمرضى المصابين بسرطان الثدي بعدة طرائق ، حيث تم جمع بيانات مرض السرطان من مركز الفرات الاوسط في النجف الاشرف ، وكانت البيانات المدروسة للفترة من (2013-2016) وكان العدد الكلي للمرضى في الدراسة هو(96) مريض وجميعهم توفوا خلال هذه المدة ولكن في تواريخ مختلفة .

تم استخدام مجموعة من الطرائق اللامعلمية لتقدير دالة البقاء ، وهذه الطرائق هي (طريقة كيرنل Kernel Method، طريقة نيلسون Nelson Method(NL)، الطريقة التجريبية Empirical Method(EM)) ، وبعد ذلك عمل الباحث مقارنه بين طرائق تقدير دالة البقاء والمحسوبة من البيانات لاختيار افضل هذه الطرائق وبالاعتماد على متوسط مربعات الخطأ (MSE).

Abstract

The analysis of the data using the survival function is one of the most important topics in our daily life because it is included in many medical fields where cancer data were collected from the center of the Middle Euphrates in Najaf. The data studied were for the period (2013-2016), and the total number of patients in the study was (96) and all of them died during this period but on different dates . A number of methods were used to estimate the survival function, The methods used are (the Kernel method, the Nelson Method (NL) method, the Empirical Method (EM) method). The researcher then compared the methods of estimating the survival function and that which calculated from the data to choose the best of these methods depending on the average of the mean squares error (MSE) .

.

1. المقدمة Introduction

يعد مرض السرطان من أخطر الأمراض التي تصيب الانسان في كل الاعمار، ويعرف السرطان بشكل عام على انه نوع من الامراض التي تجعل الخلايا المصابة تنمو وتتكاثر بصورة خارجه عن سيطرة وتحكم الجسم، وتختلف هذه الخلايا عن خلايا الجسم الطبيعية التي تنشأ منها، اذ ينتج عن هذا النمو كتل من الانسجة تسمى بالأورام (tumors) وهذه الأورام اما ان تكون أورام خبيثة وهي أورام سرطانية (Malignant tumors) خطرة تمتاز بسرعة تكاثرها وانتشارها الى أنسجة واعضاء الجسم عن طريق الدم وتدمر الخلايا المجاورة لها ، أو أن تكون أورام حميدة (Benign tumors) وهي أورام ليست سرطانية يمكن ازالتها ولا تنتشر الى مناطق الجسم الاخرى كما انها لا تهدد حياة الانسان [2] [10].

يعطى مرض السرطان اسم الجزء الذي بدا منه وعلى هذا الاساس يعرف سرطان الثدي بانه ورم خبيث يصيب أنسجة الثدي، ويكون واحد من السرطانات الاكثر شيوعا وانتشارا وسبب رئيسي للوفاة بين النساء حول العالم .

ولغاية الان وبالرغم من التطور الذي يشهده العالم وخاصة في مجال الطب الا انه لم يتم الوصول الى السبب الرئيسي للإصابة بسرطان الثدي وهو لا ينتج عن مسبب واحد بل هو ناتج من عدة عوامل محتملة والتي تزيد امكانية الاصابة بالمرض منها (العمر- الحالة الزوجية - الطمث المبكر وعوامل اخرى)، فضلا عن ذلك تشير الاحصائيات الخاصة بأمراض السرطان الى ان نسبة (5-10%) من حالات سرطان الثدي ناتجة عن مسببات وراثية ، حيث ان ضرر وظائف الجينات الطبيعية مثل جيني سرطان الثدي الاول والثاني (BRCA1-BRCA2) يؤدي الى الاصابة بالمرض، وهذه الجينات موجودة عند الرجال والنساء لذلك لا تقتصر الاصابة بمرض سرطان الثدي على النساء كما هو معروف، انما يمكن ان يصيب الرجال ايضا وبنسبة (1%) من مجموع المصابين [2] [9].

ان الاحصائيات الصادرة عن مرض سرطان الثدي في العراق تشير الى ان من بين مجموع النساء المصابات بسرطان الثدي كانت نسبة الاصابة تحت عمر 20 سنة هي 5% ، اما النساء فوق عمر40 سنة فكانت نسبة الاصابة هي 75% [2].

يتضمن هذا البحث استعراضا مرجعياً للمواضيع ذات العلاقة والمتمثلة بطرائق تقدير المعلمات ودالة البقاء والتطبيقات العملية في عام (1994) استخدمت الباحثة (وارتان) [8] بعض الطرائق المعملية وبعض الطرائق اللامعلمية لتقدير دالة البقاء (Survival function) لتوزيع ويبل ذو المعلمتين في حالة البيانات الكاملة وغير الكاملة ، اذ قامت الباحثة بتطبيق عملي لحساب دوال البقاء على مجموعة من مرضى التهاب الكبد الفايروسي في العراق .

وقد قامت الباحثة (الاعظمي) في عام (1999) [1] بتقدير انحدار انموذج كوكس على مجموعة من المريضات المصابات بسرطان الثدي في العراق ، وتوصلت الباحثة الى ان نوع العلاج هو المتغير الوحيد الذي يرشح مدة بقاء المريضات ، بينما لم يكن هناك اي تأثير معنوي بالنسبة للمتغيرات الاخرى .

وفي العام (2005) استعمل الباحث (Rodriguez) [15] عدة طرائق لامعلميه لتحليل بيانات البقاء ومن هذه الطرائق هي طريقة (نيلسون-الين، ومقدر مانتل –هينزن) عند حجوم عينات مختلفة ، وطريقة (كابلن – مير) عند العينة الواحدة ، وتوصل الباحث الى ان طريقة نيلسون –الين اعطت نتائج افضل .

أما في عام (2007) فقد استخدم الباحث (مراد) [7] انموذج كوكس وانموذج الانحدار اللوجستي لدراسة مدة البقاء للمريضات المصابات بسرطان الثدي (Breast cancer) لحين حدوث الوفاة ، والحصول على اهم المتغيرات المؤثرة على مرض سرطان الثدي دون الاعتماد على توزيع اوقات البقاء للمريضات ، وتوصل الباحث الى ان عوامل الخطورة التي حافظت على البقاء في انموذج كوكس هي العمر و هرمون الاستروجين ، بينما عوامل الخطورة التي حافظت على البقاء في انموذج اللوجستيك هي مرحلة الورم .

أما في عام (2013) فقد استعملت الباحثة (الجواد) [3] طريقة الامكان الاعظم وطريقة المربعات الصغرى في تقدير دالة البقاء الذاتي (احتمال البقاء على قيد الحياة) باستعمال توزيع ويبل و توزيع (Gomprtz) ، فضلا عن ذلك استعملت الطريقة اللامعلمية طريقة (Kernel) في تقدير دالة البقاء الذاتي، حيث توصلت الباحثة الى ان طريقة (Kernel) هي الافضل في بناء جداول الحياة الذاتية لبقاء الفرد على قيد الحياة .

وفي عام (2015) قدرت الباحثة (فرحان) [6] دالة البقاء والدوال الاحتمالية (دالة المخاطرة ، دالة الكثافة الاحتمالية والدالة التوزيعية) لبيانات حقيقة كاملة لأوقات الفشل على مرضى سرطان الرئة ، حيث استخدمت الباحثة بعض الطرائق المعلمية وبعض الطرائق اللامعلمية لتقدير الدوال مستخدمة بذلك معياري متوسط الخطأ النسبي المطلق (MAPE) ومتوسط مربعات الخطأ (MSE) للمقارنة بين الطرق، وتوصلت الباحثة الى ان طريقة الاختبار الاولي المقدرة (PR)اعطت افضل تقدير ضمن الطرائق المعلمية و طريقة التجريب (EM) اعطت افضل تقدير ضمن الطرائق اللامعلمية.

أما في عام (2016) فقد قامت الباحثة (العبيدي) [5] بتقدير دالة البقاء باستخدام بعض الطرائق المعلمية وهي طريقة الامكان الاعظم (Maximum likelihood) ومبدأ اعظم انتروبي (Principle of maximum Entropy) وكذلك استخدمت طريقة لا معلمية وهي طريقة (Kernel smoothing) وقامت الباحثة بالمقارنة بين طرائق التقدير باستخدام المتوسط المطلق لنسبة الخطأ (MAPE) وجذر متوسط مربع الخطأ (RMSE) ، وتوصلت الباحثة الى أفضلية طريقة الامكان الاعظم وطريقة Kernel عند اجراء المحاكاة .

2. الجانب النظري

2-1 دالة البقاء Survival function S(t)

تعد هذه الدالة من الدوال الاحصائية المهمة بسبب اهميتها في المجالات الطبية وغيرها من المجالات، وتبرز الحاجة الى دالة البقاء لمعرفة التوزيع الاحتمالي الملائم ليعكس ظاهرة البقاء بشكل واضح [7].

وتعرف دالة البقاء S(t) على انها احتمال بقاء المريض على قيد الحياة هو اكبر من الوقت المحدد ويعبر عنها رياضيا [12] : [13]

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

حيث ان:
T: وقت ظهور الحدث .
t: هو الوقت المحدد .
وان F(t) هي دالة التوزيع التراكمي (cumulative distribution function) وهي دالة مكملة لدالة البقاء .
ان دالة البقاء هي دالة غير متزايدة ومستمرة من جهة اليمين وتتناسب عكسيا مع الزمن ومن خواصها :-

$0 \le S(t) \le 1$

$S(0) = 1 \quad \leftrightarrow \quad F(0) = \infty$

$\lim_{t\to\infty} S(t) = 0 \leftrightarrow \lim_{t\to\infty} F(t) = 1$

اي ان احتمال البقاء على قيد الحياة على الاقل عند الزمن صفر هو 1 ، واحتمال البقاء عند المالانهايه من الزمن هو صفر[7] .

2-2 طرائق التقدير

ان الطرائق اللامعلميه لا تحتاج الى توزيع احتمالي ، اذ تستخدم في حالة ان هناك شك حول التوزيع الدقيق للبيانات وعليه سوف نستخدم في هذه الدراسة ثلاث طرائق لامعلمية هي (الطريقة التجريبية ، طريقة نيلسون ، وطريقة كيرنل) .

2-2-1الطريقة التجريبية Empirical Method (EM)

يمكن كتابة دالة التوزيع التراكمي لتوزيع الفشل بالصيغة الاتية[8] :-

$$F(t) = \begin{cases} 1 & if \quad i = n \\ \frac{i}{n} & if \quad 0 < i < n \\ 0 & o.w \end{cases}$$

اذآ تكتب دالة البقاء بالصيغة الاتية :-

$$\hat{S}(t) = 1 - \frac{i}{n} \qquad (1)$$

كذلك يمكن حساب دالة التوزيع التراكمية بصيغ مختلفة ومنها(35):-

1- دالة التوزيع اعتمادا على رتبة الوسيط :-

$$F(t_i) = \frac{i-0.3}{n+0.4}$$

ودالة البقاء

$$\hat{S}(t_i) = \frac{n- i+0.7}{n+0.4} \qquad (2)$$

2- دالة التوزيع اعتمادا على دالة الكثافة التراكمية المتماثلة :-

$$F(t_i) = \frac{i - 0.5}{n}$$

ودالة البقاء

$$\hat{S}_{EM}(t_i) = \frac{n - i + 0.5}{n} \qquad (3)$$

2-2-2 طريقة نيلسون Nelson Method (NL)

تعتمد طريقة (Nelson) لتقدير دالة البقاء بصوره رئيسية على دالة الخطورة التراكمية H(t) ووقت الحدث عند الزمن t_i حيث ان (i =1,……n) ، ويمكن كتابة هذه الطريقة بالصيغة الاتية [6]:-

$$\hat{H}(t) = \frac{\sum_{t_i \le t} D_i}{n} \qquad t > 0 \qquad (4)$$

حيث ان :-
D_i : تمثل عدد الاشخاص الذين توفوا عند الزمن t_i .
n_i : يمثل عدد الاشخاص المعرضين للخطر عند الزمن t_i

ان احتمال موت شخص خلال فتره t(i) الى t (i+1) هي $\frac{D_i}{n}$ الذي يمثل القيمة التقديرية الى t وعليه فأن دالة الخطورة التراكمية $H(t)$تعرف بالشكل الاتي (15):-

$$H(t) = \log[S(t)]$$
$$= \int_0^\infty h(u)du \qquad (5)$$

اذآ يمكن ايجاد القيمة التقديرية لدالة البقاء بطريقة نيلسون كما بالشكل الاتي :-

$$\hat{S}(t)_{Nl} = \exp\big(H(t)\big) \qquad (6)$$

3-2-2 طريقة كيرنل Kernel Method (K(t))

تهدف هذه الطريقة الى تعديل البيانات بالشكل الذي يمكن ان نحصل على مقدرات تتميز بصفات تقترب مع صفات المعلمات الحقيقية . ان اختيار المعلمة التمهيدية (bandwidth) يكون الاهم من اختيار دوال (kernel)، اذ ان هناك علاقة عكسية بين معلمة التمهيد والتباين ، وتكون العلاقة طردية بين معلمة التمهيد والتحيز [11].
اما الشروط التي يجب ان تحققها دوال (kernel) هي :-

(1) $K(t) = K(-t)$
(2) $\int_{-\infty}^{\infty} K(t)dt = 1$
(3) $\int_{-\infty}^{\infty} t\,K(t)dt = 0$
(4) $\int_{-\infty}^{\infty} t^{2}K(t)\,dt \geq 0$

حيث K(t) هي دالة (Ganssian kernel) وتكتب بالصيغة:-

$$k(t) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{t^2}{2}\right) \qquad s.t. \quad 0 \leq t \leq 1$$

ان تقدير kernel تتأثر بقيمة (bandwidth)، ولذلك تم استخدام هذه الطريقة مع طرق عرض حزمه مختلفة منها [12] :-

طريقة (kernel) مع عرض حزمه من نوع (fixed)

تعني هذه الطريقة هو استخدام المعلمة التمهيدية على طول الخط الحقيقي لتقدير دالة البقاء .
لنفرض ان لدينا مجموعه من المتغيرات المتماثلة والمستقلة التوزيع مع دالة كثافه احتماليه التي تعرف بالصيغة الأتية[4] [12]:-

$$\hat{f}(t) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{t - T_i}{h}\right) \qquad (7)$$

حيث ان :-
t: زمن تقدير دالة البقاء .
n: حجم العينة قيد الدراسة .
$k(t)$: هي دالة (kernel) .
h: المعلمة التمهيدية الثابتة .
ومن تقدير دالة الكثافة الاحتمالية يمكن الحصول تقدير الدالة التجمعية :

$$\hat{F}(t) = \frac{1}{nh}\sum_{i=1}^{n}\int_{-\infty}^{t} K\left(\frac{u - T_i}{h}\right)du \qquad (8)$$

اذ ان تقدير دالة البقاء لدالة (kernel) يكتب بالصيغة [4]:-

$$\hat{S}(t) = 1 - \frac{1}{nh}\sum_{i=1}^{n}\int_{-\infty}^{t} K\left(\frac{u - T_i}{h}\right)du \qquad (9)$$

طريقة kernel مع عرض حزمة نوع Cross Validation

تعتمد هذه الطريقة على تغيير قيمة المعلمة التمهيدية (bandwidth) عند كل نقطة وذلك من خلال استخدام النوافذ الصغيرة عند المناطق ذات الكثافة العالية واستخدام النوافذ الكبيرة عند المناطق ذات الكثافة القليلة ، وان عينة المشاهدات هي متماثلة التوزيع و مستقلة مع دالة الكثافة الاحتمالية وكالآتي [4]:

$$\hat{f}(t) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h(T_i)}K\left(\frac{t-T_i}{h(T_i)}\right) \qquad i = 1,2,\dots,n \qquad (10)$$

حيث أن :

$h(T_i)$: تمثل المعلمة التمهيدية ، وتحسب من الصيغة كالآتي :

$$h(T_i) = h_{(Cr.v)}\ f(T_i)^{-1/2}$$

و أن

$f(T_i)$: تمثل دالة الكثافة المجهولة ويتطلب حسابها استعمال تقدير أولي لهذه الدالة باستعمال مقدر (Kernel) ذو المعلمة التمهيدية الثابتة.

$h_{Cr.v}$: تمثل المعلمة التمهيدية وتقدر بطريقة الامكان الاعظم للعبور الشرعي وحسب الخطوات الآتية[4] [12]:

1. نفرض قيمة أولية للمعلمة h .

2 - حذف المشاهدات (t_i) .

3- حساب مقدر (Leave-one-out) $\hat{f}_{-\mathrm{i}}(t_i)$ من خلال المعادلة الآتية عند نقاط المشاهدات :

$$\hat{f}_{-\mathrm{i}}(t_i) = \frac{1}{(n-1)\,h}\sum_{j\neq i}^{n}K\left(\frac{t_i-t_j}{h}\right) \qquad (11)$$

4- تكوين دالة Cross Validation .

$$\mathrm{Cr.V}(h) = \frac{1}{n}\sum_{i=1}^{n} Log\left[\hat{f}_{-\mathrm{i}}(t_i)\right]$$

5. تكرار الخطوات أعلاه وفي كل مرة حذف مشاهدة واحدة.

6. يتم اختيار المعلمة التمهيدية المقابلة لأعظم { Cr.V(h)}:

$$\hat{h} = argmax\ \mathrm{Cr.V(h)}$$

من تقدير دالة الكثافة $\hat{f}(t)$ يمكن الحصول على تقدير الدالة التوزيع $\hat{F}(t)$ لـ(Kernel) من الصيغة الاتية [4]:

$$\hat{F}(t) = \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{t}\frac{1}{h(T_i)}K\left(\frac{y-T_i}{h(T_i)}\right)dy \qquad (12)$$

اذا تقدير دالة البقاء$\hat{S}(t)$ هو :

$$\hat{S}(t) = 1 - \left[\frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{t}\frac{1}{h(T_i)}K\left(\frac{y-T_i}{h(T_i)}\right)dy\right] \qquad (13)$$

3.الجانب التطبيقي

اعتمد هذا البحث على بيانات كاملة (بيانات المتوفين وعددهم 96) وكان التركيز على فترة البقاء محسوبة بالأيام

243,303,430,435,442,480,553,614,636,660,696,700,719,720,723,731,749,772,790,825,835,848,877,913
،956,960,961,987,988,1006,1032,1044,1051,1069,1070,1071,1072,1080,1083,1085,1107,1110,1119,1123
،1124,1126,1139,1152,1159,1170,1175,1181,1205,1207,1208,1219,1243,1250,1254,1269,1274,1288,1289
،1292,1294,1303,1315,1316,1338,1350,1381,1391,1415,1424,1435,1480,1491,1499,1537,1539,1559,1597
1656,1688,1690,1716,298,931,937

1.3 نتائج تحليل البيانات

لإيجاد القيم التقديرية لدالة البقاء بالطرائق تم استخدام برنامج (MATLAB) للحصول على النتائج وكما موضح في الجدول (1)

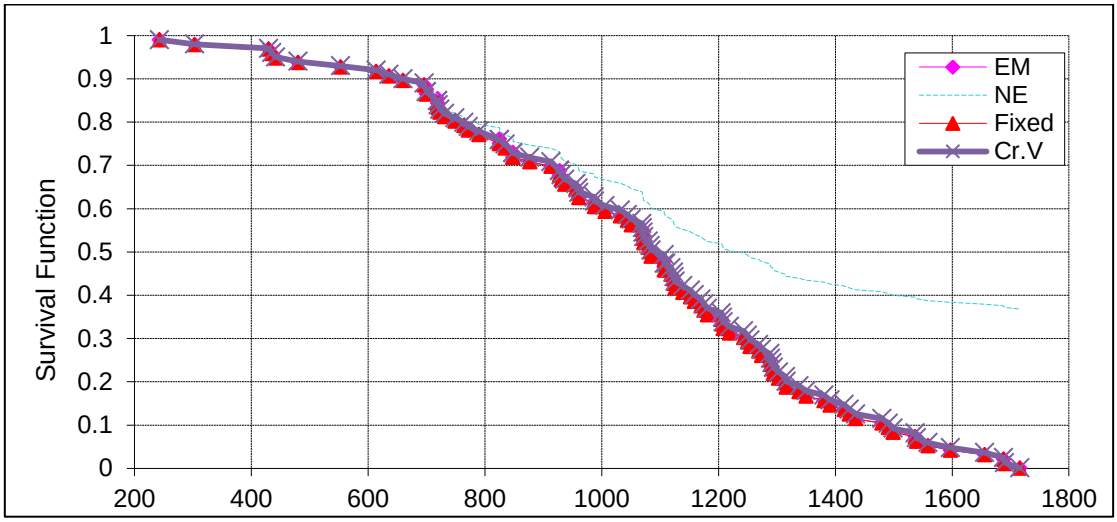
جدول(1): يمثل القيم التقديرية لدالة البقاء.

Time/Days	$\hat{S}_{NL}(t)$	$\hat{S}_{EM}(t)$	$S_f(t)$	$S_{Cr.v}(t)$
243	0.989637	0.989583	0.98958333	0.9900178892
303	0.979382	0.979167	0.97916667	0.9800267387
430	0.969233	0.96875	0.96875000	0.9700255561
435	0.959189	0.958333	0.95833333	0.9600198540
442	0.94925	0.947917	0.94791667	0.9500095362
480	0.939413	0.9375	0.93750000	0.9399938764
553	0.929678	0.927083	0.92708333	0.9299726311
614	0.920044	0.916667	0.91666667	0.9199463887
636	0.91051	0.90625	0.90625000	0.9099154570
660	0.901075	0.895833	0.89583333	0.8998798145
696	0.891738	0.885417	0.88541667	0.8898394522
700	0.882497	0.875	0.86458333	0.8697443405
700	0.873352	0.864583	0.86458333	0.8697443405
719	0.864302	0.854167	0.84375000	0.8496296998
719	0.855345	0.84375	0.84375000	0.8496296998
720	0.846482	0.833333	0.83333333	0.8395648757
723	0.83771	0.822917	0.82291667	0.8294949681
731	0.829029	0.8125	0.81250000	0.8194199244
749	0.820438	0.802083	0.80208333	0.8093397135
764	0.811936	0.791667	0.79166667	0.7992542652
772	0.803523	0.78125	0.78125000	0.7891634816

790	0.795196	0.770833	0.77083333	0.7790673424
825	0.786956	0.760417	0.75000000	0.7588588306
825	0.778801	0.75	0.75000000	0.7588588306
835	0.77073	0.739583	0.73958333	0.7487461530
848	0.762744	0.729167	0.71875000	0.7285035833
848	0.75484	0.71875	0.71875000	0.7285035833
877	0.747018	0.708333	0.70833333	0.7183736475
913	0.739276	0.697917	0.69791667	0.7082379800
928	0.731616	0.6875	0.67708333	0.6879484069
928	0.724034	0.677083	0.67708333	0.6879484069
931	0.716531	0.666667	0.66666667	0.6777941933
937	0.709106	0.65625	0.65625000	0.6676335963
956	0.701758	0.645833	0.64583333	0.6574666371
960	0.694486	0.635417	0.63541667	0.6472930577
961	0.687289	0.625	0.62500000	0.6371127142
987	0.680167	0.614583	0.61458333	0.6269257461
988	0.673119	0.604167	0.60416667	0.6167317710
1006	0.666144	0.59375	0.59375000	0.6065308448
1032	0.659241	0.583333	0.58333333	0.5963227840
1044	0.652409	0.572917	0.57291667	0.5861074537
1051	0.645649	0.5625	0.56250000	0.5758847024
1069	0.638958	0.552083	0.55208333	0.5656543262
1070	0.632337	0.541667	0.54166667	0.5554163025
1071	0.625784	0.53125	0.53125000	0.5451702716
1072	0.619299	0.520833	0.52083333	0.5349160695
1080	0.612882	0.510417	0.51041667	0.5246535258

1083	0.606531	0.5	0.50000000	0.5143825465
1085	0.600245	0.489583	0.48958333	0.5041028854
1107	0.594025	0.479167	0.47916667	0.4938143368
1107	0.58787	0.46875	0.45833333	0.4732102202
1110	0.581778	0.458333	0.45833333	0.4732102202
1119	0.575749	0.447917	0.44791667	0.4628939834
1123	0.569783	0.4375	0.43750000	0.4525681265
1124	0.563878	0.427083	0.42708333	0.4422322536
1126	0.558035	0.416667	0.41666667	0.4318861223
1139	0.552252	0.40625	0.40625000	0.4215296049
1152	0.54653	0.395833	0.39583333	0.4111624180
1159	0.540866	0.385417	0.38541667	0.4007841841
1170	0.535261	0.375	0.37500000	0.3903946410
1175	0.529715	0.364583	0.36458333	0.3799933754
1181	0.524226	0.354167	0.35416667	0.3695800487
1205	0.518793	0.34375	0.34375000	0.3591545305
1207	0.513417	0.333333	0.33333333	0.3487161211
1208	0.508097	0.322917	0.32291667	0.3382643859
1219	0.502832	0.3125	0.31250000	0.3277990138
1243	0.497621	0.302083	0.30208333	0.3173197043
1250	0.492464	0.291667	0.29166667	0.3068256995
1254	0.487361	0.28125	0.28125000	0.2963164046
1269	0.482311	0.270833	0.27083333	0.2857913825
1274	0.477313	0.260417	0.26041667	0.2752498459
1288	0.472367	0.25	0.25000000	0.2646912320
1289	0.467472	0.239583	0.23958333	0.2541145960

1292	0.462627	0.229167	0.22916667	0.2435191504
1294	0.457833	0.21875	0.21875000	0.2329039866
1303	0.453089	0.208333	0.20833333	0.2222682272
1315	0.448394	0.197917	0.19791667	0.2116108326
1316	0.443747	0.1875	0.18750000	0.2009304369
1338	0.439149	0.177083	0.17708333	0.1902260222
1350	0.434598	0.166667	0.16666667	0.1794959322
1381	0.430095	0.15625	0.15625000	0.1687387667
1391	0.425638	0.145833	0.14583333	0.1579522652
1415	0.421227	0.135417	0.13541667	0.1471344442
1424	0.416862	0.125	0.12500000	0.1362825063
1435	0.412542	0.114583	0.11458333	0.1253935008
1480	0.408267	0.104167	0.10416667	0.1144644346
1491	0.404037	0.09375	0.09375000	0.1034904528
1499	0.39985	0.083333	0.08333333	0.0924662723
1537	0.395706	0.072917	0.07291667	0.0813858207
1539	0.391606	0.0625	0.06250000	0.0702399623
1559	0.387548	0.052083	0.05208333	0.0590176345
1597	0.383532	0.041667	0.04166667	0.0477030837
1656	0.379557	0.03125	0.03125000	0.0362723376
1688	0.375624	0.020833	0.02083333	0.0246839600
1690	0.371732	0.010417	0.01041667	0.0128548529
1716	0.367879	0	0.00000	0.0005346279



شكل (1):يوضح رسم دالة البقاء للطرائق الامعلميه .

من خلال الجدوال(1) والشكل (1) نجد ان القيم التقديرية لدالة البقاء انخفضت تدريجيا مع زيادة اوقات البقاء لمرضى سرطان الثدي ، اي ان دالة البقاء عند (t=243) تكون مرتفعة لجميع الطرائق اللامعلمية حيث تبين ان القيم التقديرية لدالة البقاء كانت متقاربة بين الطرق ما عدا طريقة كيرنل نوع Bandwith= Cross Validation التي قيمتها (0.9900178892) ، بينما عند (t=1716) كانت قيمة دالة البقاء منخفضة حيث كانت القيم متساوية بين الطريقة التجريبية وطريقة كيرنل نوع fixed = Bandwith ، بينما القيم التقديرية لدالة البقاء بطريقة نيلسون هي (0.367879) وقيمتها بطريقة كيرنل نوع Cross Validation= Bandwith هي (0.0005346279) .

2.3 المقارنة بين الطرائق اللامعلميه بالاعتماد على مقياس MSE

من اجل المقارنة بين الطرائق اللامعلميه سوف نستخدم معيار متوسط مربعات الخطأ (MSE) بالاعتماد على بيانات حقيقية للمرضى المصابين بسرطان الثدي حيث ان [5]:-

$$MSE[\hat{S}(t_i)] = \frac{\sum_{i=0}^{n}[S(t_i) - \hat{S}(t_i)]^2}{n}$$

حيث ان n تمثل عدد المرضى و i = 1 ... . n

ولإيجاد دالة البقاء الحقيقية $S(t_i)$ نستخدم طريقة (Median rank) على فرض ان [5] :-

$$F(t_i) = \frac{i-0.3}{n+0.4}$$

$$\rightarrow S(t_i) = 1 - \frac{i-0.3}{n+0.4}$$

والجدول (2) يوضح قيم (MSE) للطرائق المختلفة .

جدول(2): يمثل المقارنة بين الطرائق اللامعلمية.

Metheds	$\mathbf{MSE}[\hat{S}(t_i)]$
EM	0.0029
NE	0.0275
fixed	0.0021
Cr.V	0.0001

من النتائج السابقة يمكننا الاستنتاج الى ان طريقة Cross Validation افضل طريقة لتقدير دالة البقاء من بين طرائق التقدير اللامعلميه والتي قيمتها (0.0001).

4. الاستنتاجات Conclusions

من خلال النتائج التي حصلنا عليها للمريضات المصابات بمرض سرطان الثدي توصلنا الى الاستنتاجات الاتية :-

1- لجميع الطرق نلاحظ ان قيم تقدير دالة البقاء تتناقص تدريجيا مع زيادة اوقات الفشل ، اي تتقارب مع دالة البقاء الحقيقية وان هناك علاقه عكسيه بين اوقات البقاء ودالة البقاء .

2- فيما يتعلق بالمقارنة بين طرائق التقدير الامعلميه (التجريبية، نيلسون، كيرنل) باستخدام المؤشر الاحصائي (MSE) تبين ان طريقة كيرنل مع عرض حزمه نوع Cross Validation هي الافضل .

5. التوصيات Recommendations

1- تعد الامراض السرطانية من الامراض الخطرة على حياة الانسان فلابد من اقامة دورات توعيه ومختبرات خاصة للوقاية من المرض ،لان عند اكتشاف المرض في مرحله مبكره تكون نسبة الشفاء 99% .

2- نوصي باستخدام طريقة كيرنل لأنواع اخرى من الامراض السرطانية لتقدير دالة البقاء .

المصادر References

[1] الاعظمي، يسرى طارق اسماعيل (1999) . دراسة وتحليل نموذج (Cox) مع تطبيق عملي. رسالة ماجستير- كلية الإدارة والاقتصاد - جامعة بغداد .

[2] جواد، زينب نزار (2012). العلاقة بين الطفرات في جيني BRCA1،BRCA2 وبعض عوامل الخطورة لدى مرضى سرطان الثدي في محافظة كربلاء . رسالة ماجستير- كلية التربية - جامعة كربلاء .

[3] الجواد، ياسمين عبد الرحمن محمد (2003) . بناء جداول الحياة الذاتية في العراق باستعمال احتمالات البقاء . اطروحة دكتوراه - كلية الادارة والاقتصاد - جامعة بغداد .

[4] الصفار، رواء صالح محمد (13 20). الطرائق اللامعلمية والمعدلة في تقدير دالة المعّولية للبيانات الكاملة مع تطبيق عملي . اطروحة دكتوراه - كلية الادارة والاقتصاد -. رسالة ماجستير - كلية الادارة والاقتصاد - جامعة بغداد.

[5] العبيدي، رغدة زياد طارق (2016). استعمال الانتروبي مع طرائق اخرى في تقديردالة البقاء لتوزيع كاما العام لسكان العراق . رسالة ماجستير - كلية الادارة والاقتصاد -جامعة بغداد.

[6] فرحان، ابتهال حسين (2015). تقدير دالة البقاء لبيانات حقيقيه كاملة لمرضى سرطان الرئة . رسالة ماجستير - كلية التربية ابن الهيثم - جامعة بغداد .

[7] مراد، عباس كول مراد (2007). استخدام نموذجي cox و logistic في تحليل البقاء مع تطبيق عملي . رسالة ماجستير - كلية الادارة والاقتصاد - . رسالة ماجستير - كلية الادارة والاقتصاد - الجامعة المستنصرية .

[8] وارتان، غادة يوسف اسماعيل (1994). تقدير دوال البقاء للمرضى المصابين بالتهاب الكبد الفايروسي في القطر العراقي . رسالة ماجستير- كلية الادارة والاقتصاد - جامعة بغداد .

[9]American cancer society, (2015_2016). Inc, www.breast cancer.org.

[10] - American cancer society , (2017). www.breast cancer.org .

[11] Ferraty,F. ,Veu,P. (2006)." Nonparameteric functional data analysis" Springer.

[12]Hastings,N. , Evans,M. , Peacock,B .(2000). "Statistical distribution". Fourth edition, Wiley.

[13] Lee, E. ,Wang,J. (2003)." Statistical methods for survival data analysis". Third edition, Wiley.

[14]Miladinovic, B. (2008). "Kernel density estimation of reliability with applications to extreme value distribution". university of south Florida, Scholar commons, theses and dissertations.

[10]Rodriguez, G. (2005). "Non-parametric in survival models". Springer.