

مقارنة مقدر Beran باستعمال اوزان (Nadaraya- Waston) واوزان
(Prestley-cho) في تقدير دالة البقاء الشرطية
لمرضى سرطان الثدي

أ.د. علي عبد الحسين الوكيل / كلية الادارة والاقتصاد / جامعة بغداد
د. مها محمد عبد علي / مدينة الطب / مستشفى الاورام التعليمي / وحدة التسجيل السرطاني
الباحث / أمل هادي رشيد الكرخي / كلية الادارة والاقتصاد / جامعة بغداد

تاريخ التقديم: 2018/1/23
تاريخ القبول: 2018/3/4

المستخلص

يتضمن هذا البحث تطبيق لبعض الطرائق اللامعلمية في تقدير دالة البقاء الشرطية والمتمثلة بطريقة (Beran) باستعمال كل من اوزان (Nadaraya- Waston) واوزان (Prestley-cho) في تقدير دالة البقاء الشرطية وباستعمال بيانات البتر (المراقبة) لفترة وبيانات البتر (المراقبة) من اليمين والمتمثلة ببيانات مرض سرطان الثدي ولنوعين من العلاجين الكيميائي والعلاج بالإشعاع واعتبار العمر هو المتغير المشترك ويكون متغيرا مستمر، وتم استعمال عرض الحزمة الامثل وذلك بتطبيق نموذج معكوس دالة البقاء من خلال استعمال برنامج (MATLAB) ومن ثم استعمال معيار متوسط مربعات الخطأ (MSE) للمقارنة بين الاوزان المستعملة وقد اظهرت النتائج افضلية اوزان (Nadaraya- Waston) في تقدير دالة البقاء الشرطية ولكل من العلاج الكيميائي والعلاج بالإشعاع.

المصطلحات الرئيسية في البحث / دالة البقاء، اوزان (Nadaraya- Waston)، واوزان (Prestley-cho)، وزن Kernel، معلمة عرض الحزمة.



مجلة العلوم
الاقتصادية والإدارية
العدد 107 المجلد 24
الصفحات 538-555

*بحث مستل من رسالة ماجستير



1- المقدمة Introduction

ان الدافع الاساسي في بداية الدراسات والبحوث التي تخص تحليل البقاء هو حاجة الانسان الى الاستمرار في الحياة بشكل افضل، اذ ان دراسات تحليل البقاء تهتم بمعرفة مدة البقاء للإنسان الذي يصاب بمرض معين حيث يعرف تحليل البقاء على انه دراسة الوقت الممتد من (بداية الاصابة) اي بداية الدراسة وصولاً الى نقطة النهاية المعينة (نقطة التي يتطور فيها الورم او المرض)، ويتم الحصول على بيانات البقاء التي تخص التجارب الطبية من خلال التجارب السريرية (Clinical trials)، وعادة يستخدم مصطلح دالة البقاء في الدراسات الطبية والحياتية، أما في الدراسات الهندسية أو الميكانيكية فيستخدم مصطلح دالة المعولية (Reliability Function).

اذ يعد تقدير دالة البقاء الاكثر شيوعاً في اغلب الدراسات والبحوث وكذلك يعتبر من المهام المطلوبة في تحليل البيانات التي تخص وقت او زمن البقاء (Survival time) او وقت الفشل (Failure time)، وبذلك تعد العلوم الطبية من اقدم العلوم التي تتلقى اهتمام واسع من قبل الباحثين حيث ان الاهتمام بالبحوث الطبية يأتي من خلال استعمال المعلومات المتاحة لا جراء تقدير يمكن الاعتماد عليه في بيان التأثيرات المختلفة على معدلات وقوع او وفيات المرض وذلك لضرورة تكوين بيئة صحية مناسبة للإنسان من شأنها تحقيق الاستقرار الصحي وايجاد العلاج المناسب للأمراض التي تصيبه ومن ثم يؤدي الى تحسين الوضع الصحي ومكافحة مختلف الامراض.

2- مشكلة البحث problem of research

يتضمن هذه البحث مراقبة عينة من الاناث المصابات بمرض سرطان الثدي (Breast Cancer) لفترة من الزمن والبيانات تكون بيانات بتر (مراقبة) لمدة (Interval-censored) وبتر (مراقبة) من اليمين (Right-censored) وان العينة مقسمة الى مجموعتين تتضمن المجموعة الاولى المرضى الذين تتم معالجتهم بالعلاج الكيميائي والمجموعة الثانية تتضمن المرضى الذين تتم معالجتهم بعلاج بالإشعاع ومن ثم تقدير دالة البقاء الشرطية بالطريقة اللامعلمية التي تتمثل بمقدر (Bearn) التي تفترض بان المجتمعات التي تم سحب العينة منها تبتعد عن التوزيع الطبيعي وتمتلك مشاهداتها توزيعات حرة، ويعد عمر الاناث المصابات بهذا المرض هو المتغير المشترك المعطى (Given).

3- هدف البحث objective of research

ان الهدف الرئيسي من البحث هو دراسة وتقدير دالة البقاء الشرطية باستعمال بعض الطرائق اللامعلمية في حال كون البيانات هي بيانات بتر (مراقبة) لمدة (Interval-censored) وبتر (مراقبة) من اليمين (Right-censored) ومن ثم المقارنة بين هذه الطرائق للوصول الى الطريقة الافضل من بين الطرائق المستعملة في البحث، وذلك من خلال استخدام مؤشر متوسط مربعات الخطأ (MSE) اذ يتم من خلاله الوصول الى افضل طريقة من بين الطرائق المستعملة.

4- الجانب النظري

1-4 تحليل البقاء Survival Analysis

تحليل البقاء يكون له اسماء مختلفة في المجالات الطبية اذ يطلق عليه بتحليل دوال البقاء لان الحدث يمثل الموت اما في الدراسات التي تخص الهندسة والاجهزة والمعدات فيطلق عليها مصطلح المعولية، وتستعمل طرائق تحليل البقاء ايضا لوصف الكمية وفهم السلوك العشوائي من الوقت للأحداث (time to events) حيث يتعامل مع حدوث حالة الموت للمصاب بمرض معين في التجارب الطبية ويتعامل مع حدوث الفشل في الانظمة الميكانيكية، اي اننا في تحليل البقاء نستخدم مصطلح (الفشل) عند حدوث او وقوع الحدث المعني (على الرغم من ان هذا الحدث قد يكون في الواقع هو حدوث النجاح فمثلا الانتعاش او الشفاء جراء العلاج)، عند ذلك فان بيانات البقاء تكون مختطبة (mixture).



أما البيانات الكاملة والتي تشير إلى الوصول إلى حدث النهاية (Terminal event occurred) (Terminal event not occurred) عند انتهاء فترة الملاحظة [1].
الحالة المتابعة، أو بيانات غير كاملة (Censored data occurred) وتعني عدم الوصول إلى حدث

النهاية (Terminal event not occurred) عند انتهاء فترة الملاحظة [1].
ويمكن تطبيق طرائق تحليل البقاء في العديد من المجالات التطبيقية مثل (الطب، الصحة العامة، علم الاحياء، علم الاوبئة، الهندسة، الاقتصاد، المالية، الابحاث الاجتماعية، علم النفس والديموغرافيا)، وتحليل بيانات اوقات الفشل عادة ما يعني معالجة واحدة من المشاكل الثلاثة: تقدير دوال البقاء، مقارنة معالجات دوال البقاء وتقييم العوامل المتغيرة او الاعتماد على وقت الفشل للمتغيرات التفسيرية ويعد الاسلوب الوحيد الذي يتعامل مع البيانات المراقبة وغير المراقبة [10].

2-4 اوقات البقاء Survival Times

وقت البقاء هو المتغير الذي يقيس الوقت من نقطة بداية الدراسة (اي الوقت الذي يبدأ منه العلاج) إلى نقطة النهاية المعينة (اي الوقت حتى تطور الورم او المرض)، وفي معظم الحالات يتم جمع بيانات البقاء على مدى فترة زمنية محددة وذلك لأسباب علمية [10].

3-4 دالة البقاء Survival Function

ان دالة البقاء هي احتمال بقاء الفرد على قيد الحياة حتى الوقت المحدد t ، وان دالة البقاء الشرطية هو احتمال الفرد الذي سيعيش حتى الوقت t المعطى للمتغير المشترك زمنياً (Z)

$$S(t|z_0) = p_r(T > t|Z = z_0) \dots (1)$$

إذا ان Z هو المتغير المشترك و z_0 وهي القيمة الثابتة، وفي بعض الاحيان يكون زمن الحياة والمتغيرات العشوائية غير معروفة لذلك تحتاج دالة البقاء الشرطية إلى التقدير، وان منحنيات البقاء تتبع الخصائص الآتية:
1- تكون دالة رتيبة .

2- $S(t)$ دالة غير متزايدة (non-increasing).

3- عندما يكون الوقت مساوياً إلى الصفر أي ($t=0$) فان دالة البقاء تكون مساوية للواحد ($S(t)=1$) وهذا يعني احتمال بقاء الشخص المصاب على قيد الحياة عند الزمن ($t=0$) يساوي واحد.

$$S(t) \rightarrow 0 \text{ as } t \rightarrow \infty$$

على الرغم من ان منحنيات البقاء على قيد الحياة تتخذ مجموعة متنوعة من الاشكال اعتماداً على بيانات التوزيع التابع وان جميع هذه المنحنيات تتبع الخصائص الأربعة المذكورة [9].

4-4 دالة المخاطرة Hazard Function

اذ تعرف رياضياً على انها دالة الاحتمال الشرطي لحصول الحدث (الموت) خلال المدة ($t, t+\Delta t$) بشرط ان يكون الفرد يبقى حتى الزمن (t) او اكثر ويمكن تسميته بدالة الفشل او معدل الفشل ويرمز له بالرمز $h(t)$ حيث يمكن التعبير عنه رياضياً على النحو الآتي:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr(t < T < t + \Delta t | T > t)}{\Delta t} \dots (2)$$

$$= \lim_{\Delta t \rightarrow 0} \frac{\Pr(t < T < t + \Delta t)}{\Delta t \cdot \Pr(T > t)}$$

$$= \frac{1}{S(t)} \cdot \lim_{\Delta t \rightarrow 0} \frac{F_T(t + \Delta t) - F_T(t)}{\Delta t}$$



$$h(t) = \frac{f(t)}{S(t)} \dots (3)$$

اي ان دالة المخاطرة عبارة عن قسمة دالة الكثافة الاحتمالية على دالة البقاء، ويجب ان تكون دالة المخاطرة غير سالبة اي ان $h(t) \geq 0$ ، اي ان تكاملها يكون ضمن الفترة $[0, \infty)$ اذ يتضح من ذلك بانها دالة تكون اما متزايدة او متناقصة، كما يمكن الحصول على دالة المخاطرة التراكمية (Cumulative Hazard Function) من خلال تكامل دالة المخاطرة ونرمز لها بالرمز (H) وتكون على الشكل الاتي [9]:

$$H(t) = \int_0^t h(u) du \dots (4)$$

4-5- بيانات البتر (المراقبة) Censoring Data

ان ما يميز دراسات البقاء او دوال المعولية عن غيرها من الدراسات الاحصائية هي ظاهرة البتر (المراقبة) (censoring) التي يكون فيها جزء من المعلومات مفقود أي معلومات جزئية عن المتغير العشوائي، وكذلك هي عملية تحديد عدد الوحدات في تجربة الاختبار أو تحديد زمن التجربة وتكون البيانات غير متكاملة عندما تكون هنالك بعض المفردات في العينة تمتلك الحدين الأدنى والأعلى لزمن الحياة [6]. وسوف يتم التطرق الى انواع بيانات البتر (المراقبة) بشكل أكثر توضيحا وكما يلي:-

اولاً:- بيانات البتر (المراقبة) من اليمين (Right-censored data)

يعرف بانه وقت الفشل الواقع (او الذي سيقع) عند نقطة زمنية ويحدث هذا النوع من البيانات عندما يكون للفرد وقت فشل بعد انتهاء وقت المراقبة اي ان اوقات الفشل لبعض الافراد الداخلة في الدراسة تحصل بعد الوقت المحدد، ومن ناحية اخرى فانه من المعروف للأفراد الذين تتم مراقبتهم اي (بتر من اليمين) بان الحدث لم يحدث حتى اخر زيارة ولكن يمكن ان يحدث في اي وقت من هذه اللحظة، لذلك افترضنا في هذه الحالة بان (Ti) يمكن ان يحدث ضمن الفترة $[Li, \infty)$ مع ان (Li) يكون مساويا الى المدة من بداية وقت الدراسة حتى اخر زيارة وان $[10](Ri = \infty)$.

ثانياً:- بيانات البتر (المراقبة) من اليسار (Left-censored)

حيث ان اوقات الفشل تحصل لبعض الافراد او الوحدات الداخلة في الدراسة قبل الوقت المحدد، وانه معروف لبيانات البتر من اليسار بان الحدث المعني يحدث قبل الزيارة الاولى ومن ثم فأننا نفترض بان (Ti) يقع في الفترة $[0, Ri)$ اي ان $(Li=0)$ ويمثل بداية الدراسة وان (Ri) هي وقت المدة من بداية الدراسة حتى اول زيارة، وهذا يؤكد على ان بيانات البتر من اليسار (Left-censored) هي جزء من بيانات البتر لفترة $[6](Interval-censored)$.

ثالثاً:- بيانات البتر (المراقبة) لفترة (Interval-censored)

عندما يكون الحدث المعني معروف فقط اي يحدث ضمن فترة من الزمن حيث ان اوقات الفشل لا تعرف بشكل محدد، يأتي هذا النوع من البيانات كثيراً في الاختبارات الرياضية كسباقات الركض و الجري اذ إن المشاهدات لم تراقب بشكل ثابت وكذلك في مجالات الصحة، اي ان كلا من بيانات البتر (المراقبة) من اليمين وكذلك بيانات البتر (المراقبة) من اليسار بانها جزء من بيانات البتر لفترة [10].

4 - 6- الطرائق اللامعلمية The Non Parameter Methods

تعد الطرائق اللامعلمية من الطرائق المهمة والاكثر اهتماما في الوقت الحاضر اذ تستعمل في التقدير، وذلك لان هذه الطرائق توفر مرونة عالية كما انها لا تتطلب افتراضات كثيرة حول توزيع المجتمع كما هو الحال في الطرائق المعلمية، اذ تعد ميزة كبيرة وذلك لان استعمال هذه الطرائق الاحصائية قد لا يعرف ما اذا كانت الافتراضات التي تقوم عليها متحققة في بياناتهم ام لا حيث تتطلب الطرائق المعلمية افتراضات او معلومات حول خصائص توزيع المجتمع او نوع المتغيرات المدروسة التي تعتمد عليها دالة البقاء، اما الطرائق اللامعلمية لا تفترض أي شروط مسبقة، ويلجأ الى الطرائق اللامعلمية عندما يتعذر ايجاد التوزيع النظري المناسب للبيانات ونقوم بالاستنتاج المباشر لدالة البقاء من خلال ترتيب وتجريب البيانات الخاصة بالبقاء ومن هذه الطرائق طريقة مقدر (Beran):-



طريقة مقدر Beran

تم اقتراح هذا المقدر من قبل الباحث (Beran) حيث يعد واحد من الطرائق اللامعلمية الأكثر استخداما في تقدير انحدار نموذج البقاء ونموذج المعولية وكذلك فان التحقيق في الخصائص الاحصائية لهذا المقدر تكون في الحالة العشوائية عندما تكون قيمة المتغيرات غير ثابتة ويتطلب دراسة هذا المقدر عندما يتم تحديد قيم المتغيرات مقدما [3].

وفي الوقت الحاضر هناك العديد من المنشورات والابحاث وحتى الرسائل والاطاريح التي تناقش مقدرات (kernel) فان الاهتمام الرئيسي يهتم بمشكلة اختيار المعلمة التمهيدية (معلمة عرض الحزمة) الامثل (Optimal Bandwidth)، وفي سياق هذه المشكلة فانه يجب ان نفهم ان الطرائق المرجعية التي تستخدم في الكشف عن الطرائق البديلة وكذلك طريقة العبور الشرعي (C.V) لا تنطبق على مقدر (Beran) كما هو الحال في حالة تحديد دالة (Kernel) لاي وزن من اوزان المشاهدات وفقا لقيم المتغيرات، ومن المعروف ان نوعية مقدر (Beran) يعتمد اساسا على اختيار قيمة معلمة عرض الحزمة [5].

ولذلك نقوم باقتراح فكرة لاختيار معلمة عرض الحزمة الامثل والتي تعتمد التقليل من مسافة مرات الفشل من تقدير (Kernel) لدالة معكوس البقاء (Inverse Survival Function)، ولنفرض ان $T(z)$ عبارة عن زمن البقاء والذي يعتمد على المتغير المشترك وان دالة البقاء تعرف على النحو الاتي [5]:
$$s(t|z_0) = \Pr(T_z \geq t) = 1 - F(t|z_0) \dots (5)$$

اذا ان :

$F(t|z_0)$: تمثل دالة التوزيع الشرطي للمتغير العشوائي T_z .

ان الميزة الرئيسية لزمن البقاء هو وجود مشاهدات خاضعة للبتير من اليمين (Right-censored) وكذلك مشاهدات خاضعة للبتير (المراقبة) لفترة (Interval-censored) وذلك بأخذ متوسط الفترة والتي يمكن ان تكون متمثلة بالشكل الاتي:

$$(Y_1, Z_1, \delta_1), (Y_2, Z_2, \delta_2), \dots, (Y_n, Z_n, \delta_n)$$

اذ ان :

n : تمثل حجم العينة .

Z_i : قيمة المتغير المشارك او المعطى والذي يمثل عمر المرضى لـ (i^{th}) من المتغيرات.

Y_i : وقت البقاء او وقت المراقبة، والتي تساوي

$$Y_i = L_i \quad \text{for Right censored}$$

$$Y_i = \frac{(L_i + R_i)}{2} \quad \text{for Interval censored}$$

δ_i : يمثل مؤشر او عداد للمراقبة والذي يكون على النحو الاتي:

$$\delta_i = \begin{cases} 1 & \text{for Interval censored} \\ 0 & \text{for Right censored} \end{cases}$$

وان مقدر (Beran) لدالة البقاء الشرطية سيكون على النحو الاتي [5]-:

$$\bar{S}_{h_n}(t|z_0) = \prod_{Y(i) \leq t} \left\{ 1 - \frac{W_n^j(z; h_n)}{1 - \sum_{j=1}^{i-1} W_n^j(z; h_n)} \right\}^{\delta_i} \dots (6)$$



مقارنة مقدر Beran باستخدام اوزان [Nadaraya- Waston]
واوزان [Prestley-chao] في تقدير دالة البقاء الشرطية لمرضى سرطان الثدي

اذ ان:

$W_n^j(z; h_n)$, $i=1, \dots, n$ يمثل اوزان (Nadaraya- Waston) والذي يعد من مقدرات الانحدار اللامعلمي الأكثر شيوعا ويكون صيغته بالشكل الآتي [8]:

$$w_n^i [z, h_n] = \frac{k\left(\frac{z - z_i}{h_n}\right)}{\sum_{j=1}^n k\left(\frac{z - z_j}{h_n}\right)} \dots \dots (7)$$

اذ ان: $k\left(\frac{z - z_i}{h_n}\right)$ هي دالة Kernel محققة لشروط الانتظام أي:

$$K(y) = K(-y), 0 \leq K(y) < \infty, \int_{-\infty}^{\infty} K(y) dz = 1$$

ويوجد العديد من دوال kernel الموضحة في الجدول الآتي:

جدول (1) يعرض بعض دوال Kernel [2]

Kernel	K(u)	
Epanchnikov	$(3/4)(1 - u^2)$	$I(u \leq 1)$
Cosine	$(\pi/4) \text{Cos}(\pi u/2)$	$I(u \leq 1)$
Biweight	$(15/16)(1 - u^2)^2$	$I(u \leq 1)$
Triweight	$(35/32)(1 - u^2)^3$	$I(u \leq 1)$
Gaussian	$(1/\sqrt{2\pi}) e^{-\frac{1}{2}u^2}$	$I(u \leq \infty)$
Triangle	$(1 - u)$	$I(u \leq 1)$
Uniform	$(1/2)$	$I(u \leq 1)$
Munaf	$(45/64)(1 - u^4)^2$	$I(u \leq 1)$

وان الدالة المستعملة هي دالة (Gaussian) والتي تكون صيغتها على النحو الآتي:

$$K(y) = \left(\frac{1}{\sqrt{2\pi}}\right) e^{-\frac{1}{2}y^2}, I(|y| \leq \infty) \dots (8)$$



مقارنة مقدر Beran باستخدام اوزان [Nadaraya- Waston]
واوزان [Prestley-chao] في تقدير دالة البقاء الشرطية لمرضى سرطان الثدي

ان معلمة عرض الحزمة (المعلمة التمهيدية) ($h_n > 0$) والتي تحقق الشروط الاتية:

$$\lim_{n \rightarrow \infty} h_n = 0,$$

$$\lim_{n \rightarrow \infty} nh_n = \infty$$

وهناك العديد من طرائق اختيارها وسيقتصر البحث على استعمال الطريقة الاتية :

طريقة معلمة عرض الحزمة الأمثل Optimal Bandwidth Parameter

ان اختيار معلمة عرض الحزمة يحدد قيمة الاوزان اي قيمة $W_n^j(z; h_n)$ والتي بدورها تحدد المتغيرات التي ستشارك في بناء او في تقدير دالة البقاء الشرطية المتمثلة وهكذا تكون معلمة عرض الحزمة متفاوتة بطريقة معينة، فمن الممكن ان تنخفض فيها المشاهدات، وفي هذه الطريقة سوف نقوم باختيار المعلمة الامثل والتي تقوم على التقليل من متوسط انحرافات اوقات البقاء (Y_1, Y_2, \dots, Y_n) وذلك باستخدام التقدير اللامعلمي لمعكوس دالة البقاء اي $s_z^{-1}(p)$ ، حيث يمكن تعريف معكوس دالة البقاء من خلال $g(p|z)$ اي النموذج الذي ذكر في المعادلة (5) سوف يكون على النحو الاتي [4]:

$$T(z) = g(p|z) + \varepsilon \dots (9)$$

اذ ان:

$p \in (0,1)$ ، وان ε هو خطأ المشاهدات.

وان مقدر (kernel) لنموذج المعادلة (9) يمكن كتابته بالشكل الاتي:-

$$\hat{g}(\hat{p}_i|z_i) = \frac{1}{n} \sum_{j=1}^n w_n^j(\hat{p}_i) \cdot y_j \dots \dots (10)$$

اذ ان :

w_n^j يمثل معامل الوزن الذي يحسب باستخدام الاوزان المختلفة للدالة حيث تكون طبيعية اذا كان $\frac{1}{n} \sum_{j=1}^n w_n^j(\hat{p}_i) = 1$ وتكون سالبة اذا كانت دالة الوزن اكبر او تساوي الصفر وبالتالي يقال بانها تكون دالة وزن احتمالية اذا كانت غير سالبة وطبيعية، ومن ثم يتم تطبيق معلمة عرض الحزمة الامثل التي يمكن الحصول على القيمة المثلى لمعلمة عرض الحزمة من خلال مسالة الامثلية الاتية [5] :-

$$h_n^{opt} = arg \min_{h_n} \frac{1}{n} \sum_{i=1}^n \delta_i \cdot |\hat{g}(\hat{p}_i|z_i) - y_i| \dots \dots (11)$$

وسيتم استعمال الاوزان الاتية وهي:

اولاً:- اوزان (Nadaraya- Waston)

ان مقدر اوزان (Nadaraya- Waston) من الدرجة الاولى (First Order) سوف يكون بالصيغة الاتية [6]

$$W_n^j(\hat{p}_i) = \frac{k\left(\frac{\hat{p}_i - \hat{p}_j}{b_n}\right)}{\sum_{k=1}^n k\left(\frac{\hat{p}_i - \hat{p}_k}{b_n}\right)} \dots \dots (12)$$

اذ ان:

b_n : تمثل المعلمة التمهيدية (Smoothing Parameter) ، وان (\hat{p}_i) تمثل الاحتمالات التي يمكن حسابها باستخدام مقدر (Beran) اي وفق الصيغة الاتية [5] :

$$\hat{p}_i = \hat{S}_{b_n}(y_i|z_i) \dots \dots (13)$$



مقارنة مقدر Beran باستخدام اوزان [Nadaraya- Waston]
واوزان [Prestley-chao] في تقدير دالة البقاء الشرطية لمرضى سرطان الثدي

ومن خلال النظر في المشكلة التي تنطوي على استخدام مبهدات (kernel)، فإنه يمكننا استخدام النهج المتقدم في اختيار معلمة عرض الحزمة الامثل لمقدر (kernel)، وكذلك يمكن النظر في طريقة اختيار مؤشر متوسط مربعات الخطأ التكاملية (MISE) وفقا لمعلمة التمهيد والتي تحسب وفق الصيغة الآتية [4]:

$$b_n = 1.059 n^{-\frac{1}{5}} \cdot \hat{\sigma} \quad \dots (14)$$

وان ($\hat{\sigma}$) هو تقدير التباين والذي يقدر وفقا للصيغة الآتية

$$\hat{\sigma}^2 = s_n^2 = \frac{1}{n-1} \sum_{i=1}^n (\hat{p}_i - \bar{\hat{p}})^2 \quad \dots (15)$$

اذ يعاني هذا التباين من اولا: هذا التقدير يكون غير حصين (Not Robust). وثانيا: يمتلك خصائص جيدة فقط اذا كان التوزيع قريب من (Normal)، لذلك فإنه في هذه الطريقة يجب النظر الى قوة التقدير للتباين اي:

$$\hat{\sigma}^2 = s_{rob}^2 = 1.4826 \text{ med}_{i=1, \dots, n} \left| \hat{p}_i - \text{med}_{j=1, \dots, n, k=j+1, \dots, n} \left(\frac{\hat{p}_j + \hat{p}_k}{2} \right) \right| \quad \dots (16)$$

ومن ثم يتم تطبيق معلمة عرض الحزمة الامثل التي تم الاشارة اليها في المعادلة [5](11).

ثانياً :- اوزان (Prestley-chao)

بالرجوع الى المعادلة (10) يتضح ان نوعية التقديرات (عرض الحزمة والتباين) فيها اذ تؤثر بشكل مباشر على مقدر (Beran) ويتم في هذه الطريقة استخدام مقدر اوزان (Prestley-chao) من الدرجة الثانية (Second Order) يكون كالآتي [5]:

$$W_n^j(\hat{p}_i) = \{p_{(i)} - \hat{p}_{(i-1)}\} k \left(\frac{\hat{p}_i - \hat{p}_j}{b_n} \right) \quad \dots (17)$$

وان (\hat{p}_i) تمثل الاحتمالات التي يمكن حسابها باستخدام مقدر (Beran) اي وفق المعادلة (13)، وكما ان معلمة عرض الحزمة المستخدمة تكون على الشكل الآتي [5]:

$$b_n = \left[\frac{8\pi^{\frac{1}{2}} R(k)}{3M_2(k)^2 n} \right]^{\frac{1}{5}} \cdot \hat{\sigma} \quad \dots (18)$$

اذ

$$M_2(k) = \int z^2 k(z) dz, R(k) = \int k^2(z) dz$$

وان التباين يتم تقديره وفق المعادلة (16)، ومن ثم يتم تطبيق معلمة عرض الحزمة الامثل التي تم الاشارة اليها في المعادلة (11) [5].



5- الجانب التطبيقي

5-1- مقدمة عن مرض سرطان الثدي

يعد مرض السرطان من الامراض الخطيرة والاكثر انتشارا في العالم وان هذا المرض الذي يطلق عليه السرطان ذكره القدماء في كتاباتهم الطبية القديمة التي يرجع تاريخها الى الفي سنة قبل الميلاد ، ويعد سرطان الثدي (breast cancer) نوع من أنواع السرطانات الاكثر انتشارا والذي يحتل المرتبة الاولى بين انواع السرطان الاخرى التي تصيب الاثاث ففي الولايات المتحدة الامريكية تسجل حالات التشخيص بسرطان الثدي كل ثلاث دقائق لهذا المرض غير المعروف ولكن هنالك عوامل تزيد من خطورة المرض للمرأة مثل الوراثة ، سن البلوغ المبكر ، الحمل المتأخر ، السمنة وعوامل مختلفة اخرى ، ويظهر في أنسجة الثدي والتي تصيب النساء بنسبة اكثر في العالم من الرجال الذين يصابون به بنسب قليلة ، ويمكن الوقاية من ثلث أنواع السرطان، وذلك بالتشخيص المبكر الصحيح واعطاء العلاج الملائم الذي يمكن شفاء ما يقرب من ثلث المرضى المصابين بالسرطان وكذلك التخلص من الآلام بالنسبة للحالات التي يتعذر الوقاية منها أو شفاؤها.

5-2- وصف البيانات

استطاعت الباحثة الحصول على بيانات مراقبة لعدد من المرضى المصابين بسرطان الثدي في العراق البالغ عددهم (114) مريضا وللمدة من (2015/11/1) الى (2017/1/1) وتشمل البيانات مرض سرطان الثدي اذ تم اخذ عينتين تتكون العينة الاولى من (58 = n_1) مريضة مصابة بالمرض وتتلقى العلاج الكيميائي اي [58,CT] في حين تتكون العينة الثانية من (56 = n_2) مريضة مصابة بالمرض وتتلقى العلاج الاشعاعي اي [56,RT] وقد تم اخذ البيانات من سجلات المرضى وكذلك مراجعتهم للأطباء مستشفى الاورام التعليمي بمدينة الطب وكانت مدة الدراسة (58) اسبوع حيث كانت تتم الزيارات (مراجعة المرضى) كل (4-6) اسبوع كما تختلف اوقات الزيارة الفعلية من مريض لمريض اخر اذ من خلال زيارات المرضى لوحظ عند اخذ العلاج بعض الاعراض التي تصيب المرضى وكان من اهمها اصابة المريض بشلل وخدر في الذراع سواء الذراع اليمين او الايسر ، وتحتوي البيانات على معلومات حول الاصابة بهذا الشلل او الخدر ونلاحظ ان هناك (48) لم يعانون من اي اثار جانبية اثناء الدراسة مما اعطى مشاهدات التي تخضع للبتر من اليمين .

5-3- تقدير دالة البقاء الشرطية باستخدام مقدر Beran

يتم تقدير دالة البقاء الشرطية باستخدام هذا المقدر بعد اخذ المتوسط لبيانات البتر (المراقبة) لفترة (Interval - censored) واعطاؤها رقم (1) واخذ بيانات البتر (المراقبة) من اليمين (Right-censored) واعطاؤه رقم (0) وقد استعملت الاوزان ادناه في ايجاد تقدير مقدر (Beran) في تقدير دالة البقاء الشرطية وذلك باستخدام برنامج (MATLAB) للحصول على النتائج وعلى النحو الاتي :

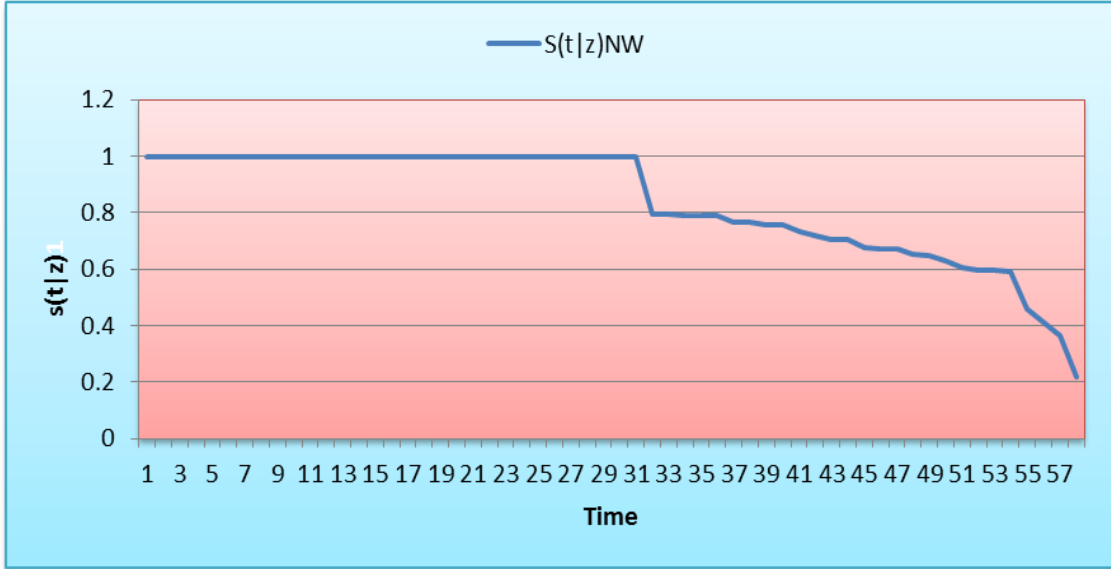
5-3-1- تقدير دالة البقاء الشرطية باستخدام اوزان (Nadaraya- Waston) للعلاج الكيميائي

(CT)

نلاحظ من شكل رقم (1) والجدول (2) بان قيم دالة البقاء الشرطية تبدأ بالتناقص مع ازدياد الزمن اذا ان احتمال البقاء للزمن الاول وحتى الزمن الحادي والثلاثون يكون مستقرا ومساويا الى (1.000) وذلك لان هناك بعض المرضى لم تظهر عليهم الاثار الجانبية وتبدأ قيمة هذا الاحتمال بالتناقص تباعاً ووصولاً الى الزمن الاخير الذي يكون فيه احتمال البقاء مساويا الى (0.2165) وكما هو موضح في الشكل (1) والجدول رقم (2). شكل (1) يوضح تقدير دالة البقاء الشرطية باستخدام اوزان (Nadaraya- Waston) لمقدر (Beran) لبيانات سرطان الثدي باستخدام العلاج الكيميائي (CT)



مقارنة مقدر Beran باستخدام اوزان [Nadaraya- Waston]
واوزان [Prestley-choa] في تقدير دالة البقاء الشرطية لمرضى سرطان الثدي



جدول (2) يوضح دالة البقاء الشرطية باستخدام اوزان (Nadaraya- Waston) لمقدر (Beran) لبيانات سرطان الثدي باستخدام العلاج الكيميائي (CT)

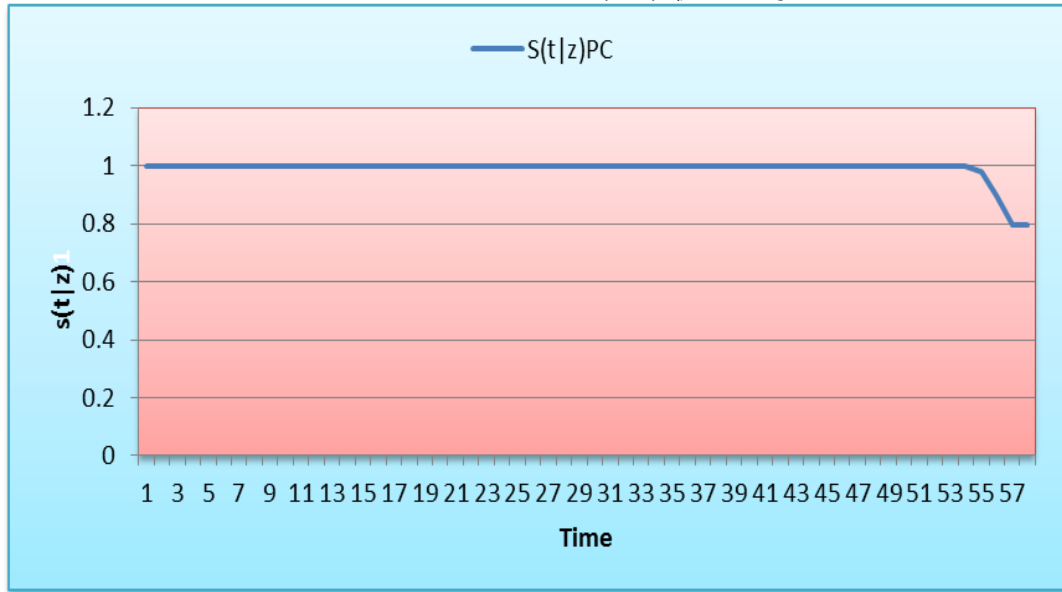
Time	S(t z)NW	Time	S(t z)NW
1	1.000	30	1.000
2	1.000	31	1.000
3	1.000	32	0.7958
4	1.000	33	0.7958
5	1.000	34	0.7911
6	1.000	35	0.7911
7	1.000	36	0.7911
8	1.000	37	0.7682
9	1.000	38	0.7682
10	1.000	39	0.7587
11	1.000	40	0.7587
12	1.000	41	0.7338
13	1.000	42	0.721
14	1.000	43	0.7066
15	1.000	44	0.7066
16	1.000	45	0.6757
17	1.000	46	0.6731
18	1.000	47	0.672
19	1.000	48	0.651
20	1.000	49	0.6496



مقارنة مقدر Beran باستخدام اوزان [Nadaraya- Waston]
واوزان [Prestley-chao] في تقدير دالة البقاء الشرطية لمرضى سرطان الثدي

21	1.000	50	0.6292
22	1.000	51	0.6073
23	1.000	52	0.5967
24	1.000	53	0.5967
25	1.000	54	0.5931
26	1.000	55	0.4572
27	1.000	56	0.4101
28	1.000	57	0.3651
29	1.000	58	0.2165

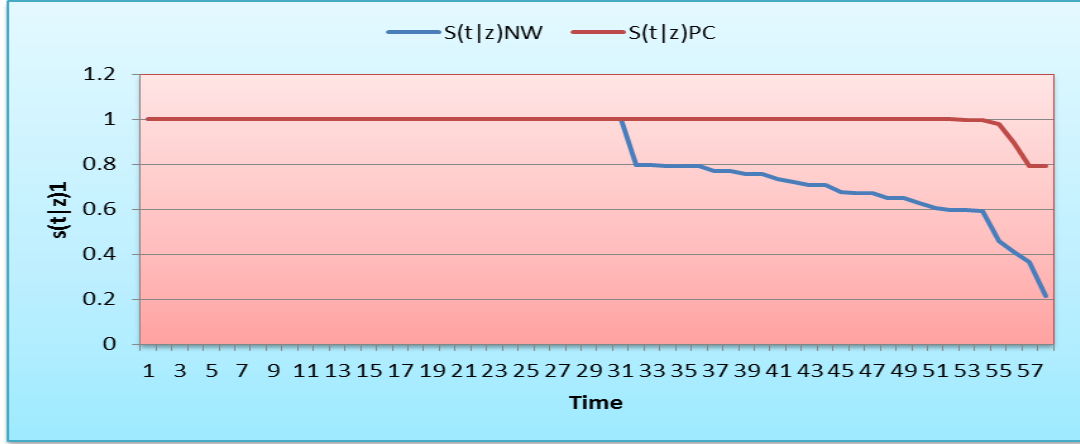
5-3-2- تقدير دالة البقاء الشرطية باستخدام اوزان (Prestley-chao) للعلاج الكيميائي (CT)
نلاحظ من شكل رقم (2) والجدول (3) بان قيم دالة البقاء الشرطية تبدأ بالتناقص مع ازدياد الزمن اذا ان احتمال البقاء للزمن الاول وحتى الزمن الحادي والخمسون يكون مستقرا ومساويا الى (1.000) وذلك لان هناك بعض المرضى لم تظهر عليهم الاثار الجانبية وتبدأ قيمة هذا الاحتمال بالتناقص تباعاً ووصولاً الى الزمن الاخير الذي يكون فيه احتمال البقاء مساويا الى (0.7942) وكما هو موضح في الشكل (2) والجدول رقم (3).
شكل (2) يوضح تقدير دالة البقاء الشرطية باستخدام اوزان (Prestley-chao) لمقدر (Beran) لبيانات سرطان الثدي باستخدام العلاج الكيميائي (CT)



شكل (3) يوضح تقدير دالة البقاء الشرطية باستخدام اوزان (Nadaraya- Waston) واوزان (Prestley-) (chao) لمقدر (Beran) لبيانات سرطان الثدي باستخدام العلاج الكيميائي (CT).



مقارنة مقدر Beran باستخدام اوزان [Nadaraya- Waston]
واوزان [Prestley-choa] في تقدير دالة البقاء الشرطية لمرضى سرطان الثدي



جدول (3) يوضح تقدير دالة البقاء الشرطية باستخدام اوزان (Prestley-choa) لمقدر (Beran) لبيانات سرطان الثدي باستخدام العلاج الكيميائي (CT)

Time	S(t z)PC	Time	S(t z)PC
1	1.000	30	1.000
2	1.000	31	1.000
3	1.000	32	1.000
4	1.000	33	1.000
5	1.000	34	1.000
6	1.000	35	1.000
7	1.000	36	1.000
8	1.000	37	1.000
9	1.000	38	1.000
10	1.000	39	1.000
11	1.000	40	1.000
12	1.000	41	1.000
13	1.000	42	1.000
14	1.000	43	1.000
15	1.000	44	1.000
16	1.000	45	1.000
17	1.000	46	1.000
18	1.000	47	1.000
19	1.000	48	1.000
20	1.000	49	1.000
21	1.000	50	1.000
22	1.000	51	1.000
23	1.000	52	0.9999
24	1.000	53	0.9982
25	1.000	54	0.9963
26	1.000	55	0.9784
27	1.000	56	0.8962
28	1.000	57	0.7942
29	1.000	58	0.7942

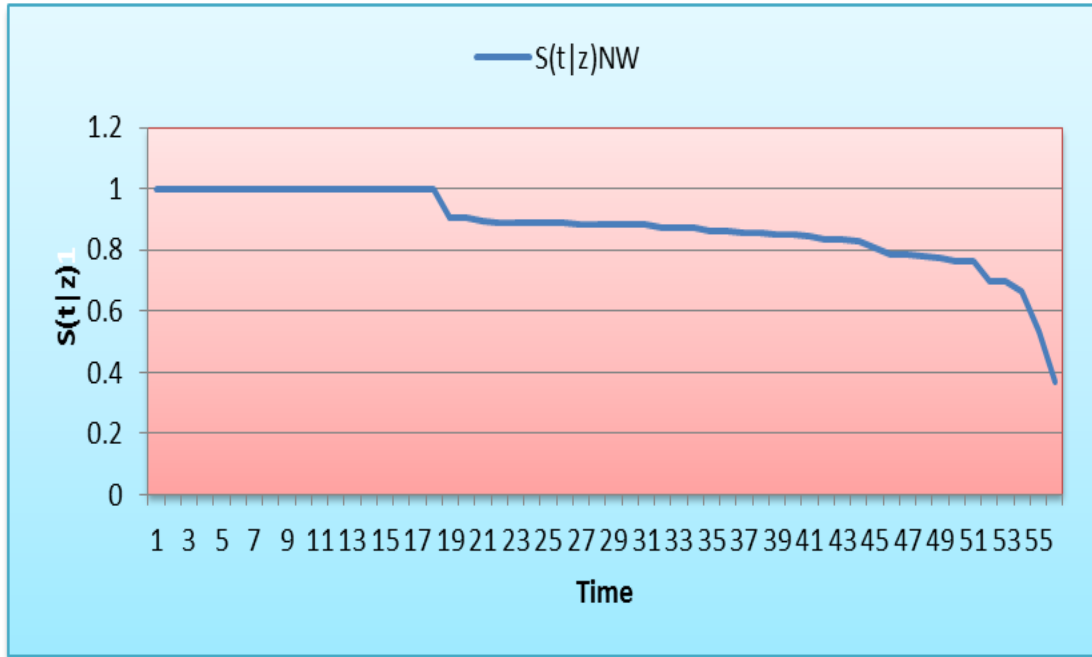


مقارنة مقدر Beran باستخدام اوزان [Nadaraya- Waston]
واوزان [Prestley-chao] في تقدير دالة البقاء الشرطية لمرضى سرطان الثدي

3-3-3- تقدير دالة البقاء الشرطية باستخدام اوزان (Nadaraya- Waston) للعلاج
بالاشعاع (RT)

نلاحظ من شكل رقم (4) والجدول (4) بان قيم دالة البقاء الشرطية تبدأ بالتناقص مع ازدياد الزمن اذا ان احتمال البقاء للزمن الاول وحتى الزمن الثامن عشر يكون مستقرا ومساويا الى (1.000) وتبدأ قيمة هذا الاحتمال بالتناقص تباعاً ووصولاً الى الزمن الاخير الذي يكون فيه احتمال البقاء مساويا الى (0.3682) وكما هو موضح في الشكل (4) والجدول (4).

شكل (4) يوضح تقدير دالة البقاء الشرطية باستخدام اوزان (Nadaraya- Waston) لمقدر (Beran) لبيانات سرطان الثدي باستخدام العلاج بالإشعاع (RT)



جدول (4) يوضح تقدير دالة البقاء الشرطية باستخدام اوزان (Nadaraya- Waston) لمقدر (Beran) لبيانات سرطان الثدي باستخدام العلاج بالإشعاع (RT)

Time	S(t z)NW	Time	S(t z)NW
1	1.000	29	0.886
2	1.000	30	0.886
3	1.000	31	0.886
4	1.000	32	0.8847
5	1.000	33	0.8736
6	1.000	34	0.8736
7	1.000	35	0.8603
8	1.000	36	0.8603
9	1.000	37	0.8573
10	1.000	38	0.8573
11	1.000	39	0.8496



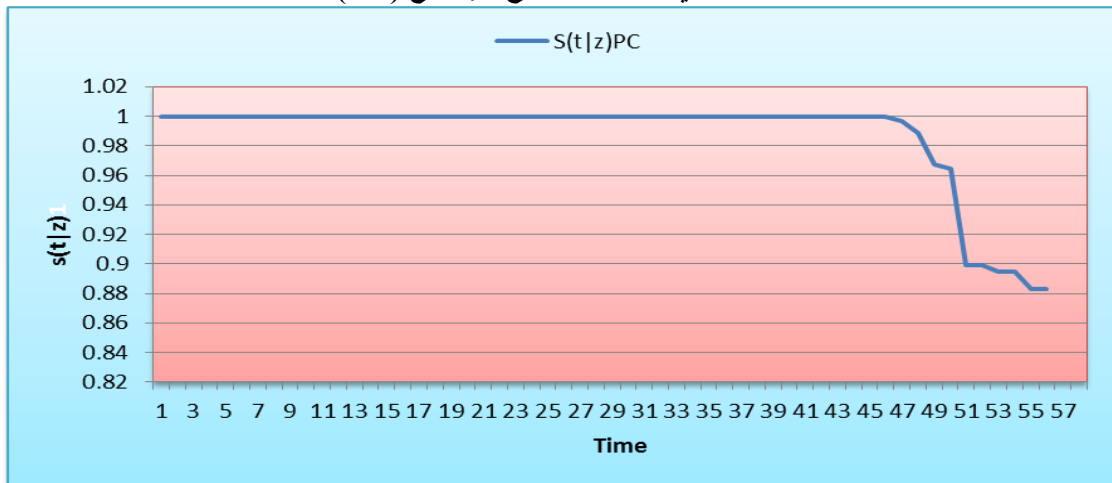
مقارنة مقدر Beran باستخدام اوزان [Nadaraya- Waston]
واوزان [Prestley-choa] في تقدير دالة البقاء الشرطية لمرضى سرطان الثدي

12	1.000	40	0.8496
13	1.000	41	0.8435
14	1.000	42	0.8366
15	1.000	43	0.8323
16	1.000	44	0.8272
17	1.000	45	0.8095
18	1.000	46	0.7866
19	0.9043	47	0.7866
20	0.9043	48	0.7785
21	0.8945	49	0.7752
22	0.8883	50	0.7619
23	0.8883	51	0.7619
24	0.8883	52	0.6993
25	0.8883	53	0.6993
26	0.8883	54	0.6633
27	0.886	55	0.5322
28	0.886	56	0.3682

4-3-5- تقدير دالة البقاء الشرطية باستخدام اوزان (Prestley-choa) للعلاج بالإشعاع (RT)

نلاحظ من شكل رقم (5) والجدول (5) بان قيم دالة البقاء الشرطية تبدأ بالتناقص مع ازدياد الزمن اذا ان احتمال البقاء للزمن الاول وحتى الزمن السابع والثلاثون يكون مستقرا ومساويا الى (1.000) وتبدأ قيمة هذا الاحتمال بالتناقص تباعا ووصولاً الى الزمن الاخير الذي يكون فيه احتمال البقاء مساويا الى (0.8827) وكما هو موضح في الشكل (5) والجدول رقم (5).

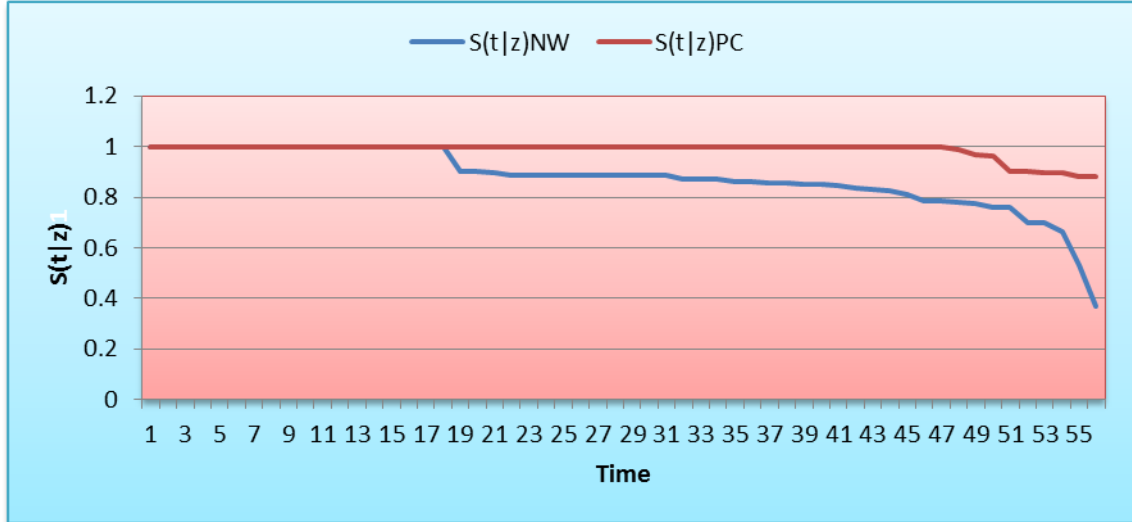
شكل (5) يوضح تقدير دالة البقاء الشرطية باستخدام اوزان (Prestley-choa) لمقدر (Beran) لبيانات سرطان الثدي باستخدام العلاج بالإشعاع (RT)





مقارنة مقدر Beran باستخدام اوزان [Nadaraya- Waston]
واوزان [Prestley-choa] في تقدير دالة البقاء الشرطية لمرضى سرطان الثدي

شكل (6) يوضح تقدير دالة البقاء الشرطية باستخدام اوزان (Nadaraya- Waston) واوزان (Prestley- chao) لمقدر (Beran) لبيانات سرطان الثدي باستخدام العلاج بالإشعاع (RT)



جدول (5) يوضح تقدير دالة البقاء الشرطية باستخدام اوزان (Prestley-choa) لمقدر (Beran) لبيانات سرطان الثدي باستخدام العلاج بالإشعاع (RT)

Time	S(t z)PC	Time	S(t z)PC
1	1.000	29	1.000
2	1.000	30	1.000
3	1.000	31	1.000
4	1.000	32	1.000
5	1.000	33	1.000
6	1.000	34	1.000
7	1.000	35	1.000
8	1.000	36	1.000
9	1.000	37	1.000
10	1.000	38	0.9999
11	1.000	39	0.9999
12	1.000	40	0.9999
13	1.000	41	0.9999
14	1.000	42	0.9999
15	1.000	43	0.9999
16	1.000	44	0.9994
17	1.000	45	0.9994
18	1.000	46	0.9994
19	1.000	47	0.9967
20	1.000	48	0.9883



مقارنة مقدر Beran باستخدام اوزان [Nadaraya- Waston]
واوزان [Prestley-chao] في تقدير دالة البقاء الشرطية لمرضى سرطان الثدي

21	1.000	49	0.9675
22	1.000	50	0.9641
23	1.000	51	0.8994
24	1.000	52	0.8994
25	1.000	53	0.8827
26	1.000	54	0.8827
27	1.000	55	0.8827
28	1.000	56	0.8827

5- 4- مؤشر متوسط مربعات الخطأ (MSE)

يتم استعمال مؤشر متوسط مربعات الخطأ (MSE) للمقارنة بين الاوزان المستعملة في هذا البحث وكما هو موضح في الجدول (6).
جدول (6) يوضح قيم مؤشر متوسط مربعات الخطأ باستخدام الاوزان المذكورة لمقدر (Beran) ولكل من العلاج الكيميائي والعلاج بالإشعاع

الاوزان	العلاج الكيميائي (CT)	العلاج بالإشعاع (RT)
اوزان (Nadaraya- Waston)	0.7435	0.7835
اوزان (Prestley-chao)	0.9829	0.9751

نلاحظ من الجدول اعلاه بان اوزان (Nadaraya- Waston) لمقدر (Beran) هي الطريقة الافضل ولكل من العلاج الكيميائي والعلاج بالإشعاع حيث تمتلك اقل متوسط مربعات خطأ (MSE).

6- الاستنتاجات والتوصيات

1-6- الاستنتاجات

من خلال ما تم طرحه في الجانب التطبيقي فقد تمكنت الباحثة من التوصل الى عدد من الاستنتاجات فيما يأتي:

1- اظهرت نتائج الجانب التطبيقي من خلال الاعتماد على معيار المقارنة متوسط مربعات الخطأ (MSE) والموضحة في الجدول (6) في تقدير دالة البقاء الشرطية فانه تم التوصل الى ان اوزان (Nadaraya- Waston) لمقدر (Beran) هي الافضل اذ تمتلك اقل معيار ولكل من العلاج الكيميائي (CT) والعلاج بالإشعاع (RT).

2- نلاحظ بان منحنيات البقاء المقدرة لا تظهر اختلافات كبيرة جدا ولكن منحنيات المرضى الذين يتعاطون العلاج الكيميائي كانت اكثر تقلباً واقل استقراراً اذ ان الانحلال او الاضمحلال السريع اي التناقص في قيمة الاحتمال لمنحنى المرضى يحدث في ازمة مختلفة ، وكذلك بالنسبة لمنحنيات المرضى الذين يتعاطون العلاج بالإشعاع كانت اقل تقلباً واكثر استقراراً فان الانحلال او الاضمحلال السريع اي التناقص في قيمة الاحتمال لمنحنى المرضى يحدث في ازمة مختلفة ايضاً.

3- نلاحظ بان عرض الحزمة الامثل كان لها تأثير على العلاج الكيميائي باستخدام كلا الوزنين اكثر من العلاج بالإشعاع.



2-6- التوصيات

- من خلال الاستنتاجات واجراءات البحث تم التوصل الى مجموعة من التوصيات يمكن ايجازها بما يأتي:-
- 1- يمكن الاعتماد على اوزان (Nadaraya- Waston) لمقدر (Beran) في تقدير دالة البقاء الشرطية وذلك لان ادائها متفوقا على الاوزان الاخرى ولكل من العلاج الكيميائي والعلاج بالإشعاع وذلك لان منحنيات البقاء للمرضى لا يوجد فيها اختلافات كبيرة على الرغم من انه في العلاج الكيميائي كان اكثر تقلبا واقل استقرارا مقارنة بالعلاج بالإشعاع الذي كان اقل تقلبا واكثر استقرارا.
 - 2- استعمال طرائق لامعتمية اخرى لتقدير دالة البقاء الشرطية وذلك باستعمال بيانات مراقبة لفترة مثلا طريقة مقدر (logspline) وغيرها من الطرائق.
 - 3- استعمال طرائق مختلفة لعرض الحزمة في تقدير دالة البقاء الشرطية مثلا الحزمة لمكيفة (adaptive algorithms) او حزمة (bootstrap) وغيرها من عرض الحزم المعروفة.

7- المصادر

- 1- حسن، رعد فاضل، وصالح، عائدة هادي، (2014)، " استخدام نموذج انحدار كوكس (Cox - Regression) لأوقات بقاء المرضى المصاب ن بمرض سرطان الدماغ ف العراق "، بحث منشور في مجلة كلية الرافيدين لجامعة العلوم، العدد(34)، طبعة(1681-6870).
- 2-حمود ، مناف يوسف، (2000) " مقارنة مقدرات Kernel اللامعلمية لتقدير دوال الانحدار " رسالة ماجستير في الإحصاء ، كلية الإدارة والاقتصاد ، جامعة بغداد.
- 3- Beran, R. (1981), " Nonparametric regression with randomly censored survival data", Technical report, Department of Statistics, University of California, Berkeley.
- 4- Demin, V.A., Chimitova, E.V. ,(2013) ,"Choice of optimal smoothing parameter for nonparametric estimation of regression reliability model ", Tomsk state university. J. Contr. Comput. Sci. 1(22), pp: 22–65 .
- 5-Demin, V.A., Chimitova, E.V, (2014), "'A Method for Selection of the Optimal Bandwidth Parameter for Beran's Nonparametric Estimator", Novosibirsk State Technical University, Novosibirsk, Russia.
- 6-Giolo ,S.R. (2004)," Turnbull's Nonparametric Estimator for Interval-Censored Data", Department of Statistics, Federal University of Parana 81531-990 - Curitiba, Parana, Brazil, Technical Report, August.
- 7-Keilegom, I. V., (1998)," Nonparametric estimation of the conditional distribution in regression with censored data", Dissertation, Limburgs Universitair Centrum te verdedigen door.
- 8- Keilegom, I. V., Akritas,and M.G., Veraverbeke, N., (2001)," Estimation of the conditional distribution in regression with censored data: a comparative study", Comput. Stat. Data Anal. 35, 487–500.
- 9-Klein,J.P. and Moeschberger, M. L.,(2003),"Survival Analysis: Techniques for censored and truncated data", 2rd ed. New York : Springer.
- 10-ZHAO, G. M.A.,(2008) "Nonparametric and Parametric Survival Analysis of Censored Data with Possible Violation of Method Assumptions", A Thesis Submitted to the Faculty of The Graduate School at The University of North Carolina .



Comparison of Beran estimator using Nadaraya-Waston and Prestley-chao weights in estimating the conditional survival function of breast cancer patients

Abstract

This study includes the application of non-parametric methods in estimating the conditional survival function of the Beran method using both the Nadaraya-Waston and the Priestley-chao weights and using data for Interval censored and Right censored of breast cancer and two types of treatment, Chemotherapy and radiation therapy Considering age is continuous variable, through using (MATLAB) use of the (MSE) To compare weights The results showed a superior weight (Nadaraya-Waston) in estimating the survival function and condition of Both for chemotherapy and radiation therapy.

Keywords \ survival function, Nadaraya-Waston ,Prestley-chao weight, Kernel weight, Bandwidth parameter.