مقارنة طريقتي (Turnbull) و (Generalization Turnbull's) اللامعلمية في تقدير دالة البقاء الشرطية (دراسة تطبيقية على مرضى سرطان الثدي)

أ. د. علي عبد الحسين الوكيل / كلية الادارة والاقتصاد / جامعة بغداد
الباحث / أمل هادي رشيد الكرخي

تاريخ التقديم:2017/12/27
تاريخ القبول:2018/1/28

المستخلص

يتضمن هذا البحث تطبيق الطرائق اللامعلمية في تقدير دالة البقاء الشرطية والمتمثلة بطريقة (Turnbull) و (Generalization Turnbull's) باستعمال بيانات البتر (مراقبة) لفترة والمتمثلة ببيانات مرض سرطان الثدي ولنوعين من العلاجين العلاج الكيميائي والعلاج بالإشعاع واعتبار العمر هو المتغير المشترك ويكون متغيرا مستمر، اذ تم تطبيق خوارزمية المقدرين من خلال استعمال برنامج (MATLAB) ومن ثم استعمال معيار متوسط مربعات الخطأ(MSE) للمقارنة بين المقدرين واظهرت النتائج افضلية مقدر (Generalization Turnbull's) في تقدير دالة البقاء الشرطية ولكل من العلاج الكيميائي والعلاج بالإشعاع اذ نلاحظ بان منحنيات البقاء المقدرة لا تظهر اختلافات كبيرة جدا ولكن منحنيات المرضى الذين يتعاطون العلاج الكيميائي كانت اكثر استقرارا ووقت بقاء اطول اذ ان الانحلال او الاضمحلال السريع اي التناقص في قيمة الاحتمال لمنحنى المرضى يحدث في زمن كل من الفترة التاسعة وزمن الفترة الثانية والعشرون وكذلك زمن الفترة السادسة والعشرون ويصل قيمة الاحتمال لدالة البقاء الى الصفر في زمن الفترة الاخيرة .

المصطلحات الرئيسة في البحث/ دالة البقاء، مقدر (Turnbull) ، مقدر (Generalization Turnbull's)، وزن Kernel، معلمة عرض الحزمة.

مجلة العلوم
الاقتصادية والإدارية
العدد 106 المجلد 24
الصفحات 356-374

*بحث مستل من رسالة ماجستير

1- المقدمة Introduction

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 وصولا الى نقطة النهاية المعينة (نقطة التي يتطور فيها الورم او المرض) فمثلا في حالة الاصابة بسرطان الدم (Leukeamia) وحالة المريض بعد زراعة نخاع العظم (Bone Marrow Transplantation) وغيرها من الحالات المرضية اي دراسة تأثير العلاج على حالة المصاب اثناء فترة 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 المعّولية (Reliability Function) ،ويعتبر تقدير دالة البقاء الاكثر شيوعا في اغلب الدراسات والبحوث وكذلك يعتبر من المهام المطلوبة في تحليل البيانات التي تخص وقت او زمن البقاء (Survival time) او وقت الفشل ،و يمكن استعمال دالة البقاء لتقيم صحة افتراض نموذج معين لمتغير البقاء وايضا يحتاج الفرد الى تقدير دالة البقاء لتقدير بعض احتمالات البقاء على قيد الحياة لمقارنة عدة علاجات مختلفة او التنبؤ باحتمالات البقاء للمرضى المصابين على قيد الحياة في المستقبل وكما يُعد موضوع بيانات البتر (المراقبة) (Censored Data) ومختلف مجالاته من المواضيع التي اصبحت تأخذ حيزاً كبيراً في الدراسات والبحوث، فقد كانت بدايات الاهتمام بها تكمن في تحليل وتطبيق بيانات المراقبة في المجال الطبي.

2- مشكلة البحث problem of research

في معظم الدراسات والبحوث التي تتعلق بحالات البقاء فيما يخص بعض الامراض فان حالة المصاب قد يظهر من خلالها الحدث والذي يتمثل بالموت اي الفشل او التحسن وبالتالي يبقى على قيد الحياة اذ يتم مراقبة المرضى لفترة من الزمن وتتضمن هذه الدراسة مراقبة عينة من الاناث المصابات بمرض سرطان الثدي (Breast Cancer) حيث يتم تقسيم هذه العينة الى مجموعتين تتضمن المجموعة الاولى المرضى الذين تتم معالجتهم بالعلاج الكيميائي والمجموعة الثانية تتضمن المرضى الذين تتم معالجتهم بالإشعاع ومن ثم تقدير دالة البقاء الشرطية بالطريقة اللامعلمية التي تتمثل بمقدر (Turnbull) و(Generalization Turnbull's)التي تفترض بان المجتمعات التي تم سحب العينة منها تبتعد عن التوزيع الطبيعي وتمتلك مشاهداتها توزيعات حرة ، ويعتبر عمر الاناث المصابات بهذا المرض هو المتغير المشترك المعطى(Given) .

3- هدف البحث objective of research

ان الهدف الرئيسي من البحث هو تقدير دالة البقاء الشرطية بالطرائق اللامعلمية في حال كون البيانات تتعرض للمراقبة لفترة (Interval-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) حيث يتم من خلاله الوصول الى افضل طريقة من بين الطرائق المستعملة.

4- الجانب النظري

1-4- تحليل البقاء Survival Analysis

تحليل البقاء هو تعبير عن مجموعة من التقنيات الاحصائية المتبعة مع نمذجة بيانات زمن الحياة وتستعمل طرائق تحليل البقاء 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) [10] .

من ناحية اخرى هذا المصطلح يمثل (وقت البقاء على قيد الحياة) الذي يحدد طول الوقت المستغرق لفشل الحدث والذي عادة ما يرمز له بالرمز (Ti) والذي يفترض بان يكون متغير عشوائي ايجابي، وبشكل اوسع هو دراسة الوقت للمصاب منذ معرفة او تشخيص المرض لحين حدوث الحدث (والحدث هنا يمثل الموت في التجارب الطبية) او يمكن ان يكون تحت المراقبة والتي تشمل عدة اسباب منها التعافي او مغادرة المريض من المستشفى الى عنوان مجهول اي (الانسحاب) دون معرفة حالته الصحية او عدم الاستمرار بالعلاج وغيرها من الاسباب، فتحليل البقاء يكون له اسماً مختلفة في المجالات الطبية حيث يطلق عليه بتحليل دوال البقاء لان الحدث يمثل الموت اما في الدراسات التي تخص الهندسة والاجهزة والمعدات فيطلق عليها مصطلح المعولية[11].

2-4- **دالة البقاء** Survival Function

نماذج دالة البقاء هي احتمال بقاء الفرد على قيد الحياة حتى الوقت المحدد t،ويمكن تعريف دالة البقاء رياضيا حسب الصيغة الاتية:

$$s(t) = p(T > t), t \geq 0 \;\ldots\ldots\; (1)$$

حيث ان:

t: يمثل الوقت المحدد، P : يمثل الاحتمال.

T : يمثل متغير زمن البقاء او الزمن المستغرق لحدوث الحدث (time to events) وهو حدث الموت.

وبما ان (T) هو متغير عشوائي مستمر، فان دالة البقاء يمكن تقديرها مرة اخرى بالشكل الاتي:

$$S(t) = p(T > t) = \int_{+t}^{\infty} f(u)du \;\ldots\ldots\quad (2)$$

حيث ان S(t) هو تكامل دالة الكثافة الاحتمالية (p.d.f) لذلك يمكن اخذ سالب المشتقة للمعادلة (2)حيث تشتق بالنسبة للمتغير(t) وبالتالي نحصل على

$$f(t) = -\frac{dS(t)}{d(t)} \;\ldots\ldots\quad (3)$$

ويمكن اعتبار f(t) هو احتمال بان الحدث سوف يحدث عند الوقت (t) من خلال اخذ سالب المشتقة لدالة البقاء بالنسبة للمتغير (t)، فان دالة f(t) المتمثلة في المعادلة (3) ستكون غير سالبة، وان دالة البقاء الشرطية هو احتمال الفرد الذي سيعيش حتى الوقت t المعطى للمتغير المشترك زمنيا(Z_0)

$$S(t|z_0) = p_r(T > t|Z = z_0) \;\ldots\ldots\; (4)$$

اذا ان Z هو المتغير المشترك و Z_0 هي القيمة الثابتة، وفي بعض الاحيان يكون زمن الحياة والمتغيرات العشوائية غير معروفة لذلك تحتاج دالة البقاء الشرطية الى التقدير، وان منحنيات البقاء تتبع الخصائص الاتية:

1-S(t) تكون دالة رتيبة .

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

3- عندما يكون الوقت مساويا الى الصفر اي(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.

$$S(t) \to 0 \;\; as\; t \to \infty$$

على الرغم من ان منحنيات البقاء على قيد الحياة تتخذ مجموعة متنوعة من الاشكال اعتمادا على بيانات التوزيع التابعة وان جميع هذه المنحنيات تتبع الخصائص الاربعة المذكورة [5].

3-4- **دالة الكثافة التراكمية (التجمعية)** Cumulative Density Function

ويطلق عليها ايضا دالة توزيع زمن الحياة(Lifetime distribution function) اذ تعرف على انها دالة مكملة لدالة البقاء ويرمز لها بالرمز F(t) ويمكن التعبير عنها رياضيا كما يلي:

$$F(t) = p(T < t) = \int_{0}^{t} f(u)du = 1 - S(t) \;\ldots\ldots\; (5)$$

حيث تمتاز هذه الدالة بانها تكون محددة في الفترة [0,t] وغير سالبة وكذلك تكون دالة رتيبة ومتزايدة دائما كما ان قيمتها تقع بين الصفر والواحد اي انها $0 \leq F(t) \leq 1$.

وتكون قيمة الغاية لهذه الدالة عندما يقترب (t) من الصفر على الشكل الاتي :

$$\lim_{t \to 0} F(t) = 0$$

وعندما يقترب (t) من ما لانهاية فان قيمة الغاية تكون بالصيغة الاتية:

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

كما ان مشتقة دالة الكثافة التراكمية هي دالة الكثافة الاحتمالية والتي يرمز لها بالرمز f(t) [2]

$$f(t) = \frac{d}{dt} F(t) \quad .... \quad (6)$$

4-4- بيانات البتر (المراقبة) Censoring Data

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، لذلك سوف نعمل مع العينات التي تسمى (Censoring Data) والتي تعني مراقبة وحدات العينة خلال فترة زمنية محددة مسبقا او ان نراقب عدد معين من هذه الوحدات المحددة مسبقا، حيث تشير هذه البيانات الى معلومات جزئية وذلك يعني ان جزء من المعلومات تكون مفقودة وذلك لا سباب مختلفة منها انتهاء زمن الدراسة اي ان التجربة تكون محددة بوقت معين لذلك يتم التضحية ببعض الوحدات التجريبية بعد نهاية التجربة او قد يتم توقف التجربة عند الحصول على عدد معين من حالات الفشل وكما ان الوحدات التجريبية قد تدخل في اوقات مختلفة كان تكون لاستجابة المريض للعلاج او وفاته لسبب اخر غير سبب الدراسة، وهذا ما يميز دراسات دوال البقاء وكذلك دوال المعولية عن غيرها من الدراسات الاحصائية ويمكننا تصنيف مختلف البيانات على النحو الاتي:

1- اذا كان $0 < L_i < R_i < +\infty$، فان (Ti) يمثل بيانات البتر (المراقبة) لفترة (Interval-censored).

2- اذا كان $0 = L_i < R_i < +\infty$، فان (Ti) يمثل بيانات البتر (المراقبة) من اليسار (Left-censored).

3- اذا كان $0 < L_i < R_i = +\infty$ ، فان (Ti) يمثل بيانات البتر (المراقبة) من اليمين (Right-censored).

4- اذا كان $L_i = T_i = R_i$، فان (Ti) يمثل البيانات الكاملة (Complete Data).

5- اذا كان $0 = L_i < R_i = +\infty$، فان (Ti) يمثل بيانات غير مراقبة (Unobserved Data)[4] .

اذا ان(L_i) يمثل بيانات البتر (المراقبة) من اليسار (Left-censored)، وان (R_i) يمثل بيانات البتر (المراقبة) من اليمين (Right-censored).

وسوف يتم التطرق الى انواع بيانات البتر بشكل اكثر توضيحا وكما يلي:-

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

يحدث 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 (مراقبة من اليمين) بان الحدث لم يحدث حتى اخر زيارة ولكن يمكن ان يحدث في اي وقت من هذه اللحظة، لذلك افتراضنا في هذه الحالة بان(Ti) يمكن ان يحدث ضمن الفترة [Li,∞) مع ان (Li) يكون مساويا الى الفترة من بداية وقت الدراسة حتى اخر زيارة وان (Ri=∞)[10] .

ثانيا:- 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 المراقبة من اليسار بان الحدث المعني يحدث قبل الزيارة الاولى وبالتالي فأننا نفترض بان (Ti) يقع في الفترة [0,Ri) اي ان (Li=0) ويمثل بداية الدراسة وان (Ri) هي وقت الفترة من بداية الدراسة حتى اول زيارة ، وهذا يؤكد على ان بيانات المراقبة من اليسار(Left-censored) هي جزء من بيانات المراقبة لفترة (Interval-censored)[4] .

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

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

4-5- الطرائق اللامعلمية The Non Parameteric Methods

تعد الطرائق اللامعلمية في التقدير الإحصائي من الطرائق التي تلاقي اهتمامًا واسعً خصوصا في الوقت الحاضر، وذلك لان هذه الطرائق توفر مرونة عالية كما انها لا تتطلب افتراضات كثيرة حول توزيع المجتمع كما هو الحال في الطرائق المعلمية ، وهذه ميزة كبيرة لان مستخدم 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، اذ يلجا 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:-

1- طريقة مقدر Turnbull Estimator Method

يعتبر التقدير اللامعلمي لدالة البقاء الشرطية $s(t|z_0)$ حيث يكون بناء المقدر اللامعلمي لدالة البقاء الشرطية وفق المعادلة (4)، و يعتبر (T) هو وقت الاصابة بالمرض وان (Z) هو عمر المريض عند التسجيل في الدراسة ، وان(T) تكون غير ملاحظة او مراقبة بالضبط ولكن من المعروف بانها تنتمي الى الفترة (L_i, R_i) حيث يتم تقدير $S(t|z_0)$ بالاعتماد على العينة (n) من المشاهدات المستقلة (L_i, R_i, Z_i) مع الفترة حيث ان $T_i \in (L_i, R_i), i = 1, \dots, n$، ونفترض بان يكون مشروط على المتغيرات لذا فان البيانات البتر لفترة (Interval censored) تكون غير مقيدة وبشكل اكثر تحديدا نفترض بان $(Z = z_0)$ المعطاة لقيمة (T) تكون مستقلة عن عملية توليد الاوقات المحصورة بين(L) و(R)، فمثلا تحت حالة الخليط لبيانات البتر لفترة سوف نلاحظ الفترة (L, R)عندما يكون $0 = U_1 < U_2 < \cdots < U_K < U_{K+1} = \infty$ مع $T \in (U_j, U_{j+1})$، اذ ان $(L = U_j)$ و$(R = U_{j+1})$ ثم نفترض بان $(Z = z_0)$ المعطاة و(T) و $\{K, U_1, U_2, \dots\dots, U_K\}$ تكون مستقلة[3] .

ويطلق على مقدر(Turnbull) ايضا اسم مقدر الامكان الاعظم اللامعلمي(NPMLE) ويستخدم هذا المقدر في تقدير دالة البقاء لبيانات البتر لفترة (Interval censored)، وقد طور (Richard Peto) في عام (1973) طريقة (Newton-Raphson) لتقدير مقدر الامكان الاعظم اللامعلمي(NPMLE) [50pp86]، ثم قدم الباحث (Turnbull) في عام (1974) خوارزمية الاتساق الذاتي(Self-Consistency) والتي هي حالة خاصة من خوارزمية(EM) لحساب مقدر الامكان الاعظم اللامعلمي(NPMLE) لتقدير دالة البقاء لبيانات البتر لفترة ،كما يستوعب هذا المقدر بيانات القطع (Truncated Data)، اذا ان خوارزمية الاتساق الذاتي (Self-consistent) عادة ما تشير الى التقدير الذي يمكن وصفه بمعادلة الاتساق الذاتي والحد من التكرارات التي تم الحصول عليها من تلك المعادلة [9].

وسوف يتم عرض مثال بسيط لمقدر(Turnbull) في حالة عدم وجود اي قطع للبيانات حيث نفترض انه لدينا (5) حالات فشل التي حدث في الدراسة وان اوقات البقاء لـ(5) من المرضى في هذه الدراسة الافتراضية سوف تكون مراقبة لفترة وباتباع مجموعة البيانات الاتية التي تبين ان (n=5) لنقاط بيانات البتر لفترة وتكون على النحو الاتي:

$[L_1,R_1] = [1, 2], [L_2,R_2] = [1, 5], [L_3,R_3] = [4, 7], [L_4,R_4] = [3, 8], [L_5,R_5] = [5, 9]$

وقبل تقدير(NPMLE) باستخدام خوارزمية (Turnbull) يجب ان نحدد فئات التكافؤ(الفترات الداخلية) (equivalence classes) لتحديد في اي وقت نقاط دالة البقاء تاخذ بالقفز(Jumps) ، وفئات التكافؤ تعرف لكل (L) الذي يتبع بشكل مباشر بـ(R) بمجرد ترتيب نقاط النهاية ، ولإيجاد فئات التكافؤ نحتاج للنظر الى جميع الفترات [Li,Ri] حيث ان i=1,…,n ، ثم ترتيب نقاط النهاية من الاصغر الى الاكبر وحسب الجدول الاتي [10]:

جدول(1) بناء فئات التكافؤ(الفترات الداخلية) لمجموعة بيانات المراقبة لفترة [10]

نقاط النهاية الاولية	1	2	1	5	4	7	3	8	5	9
التصنيفات المطابقة	L_1	R_1	L_2	R_2	L_3	R_3	L_4	R_4	L_5	R_5
ترتيب نقاط النهاية	1	1	2	3	4	5	5	7	8	9
التصنيفات المطابقة	L_1	L_2	R_1	L_4	L_3	R_2	L_5	R_3	R_4	R_5
التصنيفات	L	L	R	L	L	R	L	R	R	R

الجدول اعلاه يبن كيف نحصل على فئات التكافؤ(الفترات الداخلية) لمجموعة البيانات الافتراضية اذ ان السطرين الاول والثاني في الجدول هما البيانات الاولية ،والسطرين التاليين يمثلان النقاط النهائية التي تم ترتيبها مع العلامات المقابلة او المطابقة لها ،والسطر الخامس يدل فقط اذا كانت النقطة المطابقة لها هي نقطة النهاية اليسرى (L) والنهاية اليمنى (R)، لذلك فانه لدينا (g=3) من فئات التكافؤ حسب الجدول والتي هي [5,7],[1,2],[4,5] تكون فقط ضمن هذه الفئات المتكافئة(الفترة الداخلية) ، حيث يمكن لدالة البقاء ان يحصل لها قفز(Jumps).

وليكن $D_i = (L_i, R_i)$ تشير الى (ith) من مشاهدات المراقبة لفترة ولتكن $(0 = \tau_0, \tau_1, \ldots, \tau_g, g \le 2n)$ هي القيم المميزة المطلوبة$\{L_i, R_i, i = 1, \ldots, n\}$ ولكن $B_i = (\tau_{j-1}, \tau_j), j = 1, \ldots, g$، ولتكن$\{A_l, l = 1, \ldots, m\}$ عبارة عن مجموعة الفترات الداخلية (Innermost Interval) اي ان:

$$\{A_l, l = 1, \ldots, m\} = \{B_j : \exists i, i^{\cdot} : \tau_{j-1} = L_i \, and \tau_j = R_{i}^{\cdot}, j = 1, \ldots, g\} \ldots .(7)$$

كما ان مقدر (NPMLE) لدالة البقاء يعين كل الكتل الاحتمالية للفترات الداخلية حيث يتم تعريف المتغير الوهمي (α_{ij}) كما هو معروف حسب (Turnbull 1976) [3pp236]:

$$\alpha_{ij} = \begin{cases} 1 & , B_j \subseteq D_i \\ 0 & otherwise \end{cases}, i = 1, \ldots, n, j = 1, \ldots, g \ldots \ldots (8)$$

ولتكن (p_j) تمثل الكتل المخصصة لـ(B_j) في مقدر(Turnbull's NPMLE) لدالة البقاء S(t) حيث ان $p_j = \widehat{P}\,[T \epsilon B_j], j = 1, \ldots, g$، اي ان(p_j) تمثل احتمال الوفاة في B_j ويمكن تعريفها وفق المعادلة الاتية:

$$C_P = \left\{ P \in [0,1]^g : \sum_{j=1}^{g} p_j = 1, p_j \ge 0 \right\} \ldots .(9)$$

باعتبارها مجموعة جزئية ويحدد مقدر الامكان الاعظم اللامعلمي (NPMLE) لـ(P) عن طريق تعظيمها باستخدام دالة الامكان وكما يلي[8]:

$$L_P(S)=\prod_{i=1}^{n}[S(L_i)-S(R_i)]=\prod_{i=1}^{n}\sum_{j=1}^{g}\alpha_{ij}p_j \;.....(10)$$

ومن ثم فان (p_j) يتم ايجاده لحل معادلة الاتساق الذاتي (self-consistency equation) وعلى النحو الاتي:-

$$p_j=\frac{1}{n}\sum_{i=1}^{n}\frac{\alpha_{ij}p_j}{\sum_{k=1}^{g}\alpha_{ik}p_k}\;,j=1,\dots,g\;.....(11)$$

والتي تفسر $(p_j=0)$ اذا كانت (B_j) ليست ضمن الفترة الداخلية وان $(p_j>0)$ في حالات اخرى ، كما انه لا يمكننا مباشرة تطبيق الاستراتيجيات المقترحة من قبل Beran (1981) او Betensky (2000) الذي يجعل المؤشر (α_{ij}) يعتمد على (Z_i) و(z_0)[3] .

وقد بين (Turnbull) بان تقدير دالة البقاء يمكن ان يحدث فقط عندما يكون القفز(Jumps) بين الاحصاءات المرتبة $(\tau_j, i=1,\dots,n)$ لـ (Li) و(Ri) حيث ان $(1\le j\le n)$، واقترح ايضا كيف يمكن التقليل من عدد الاحصاءات المرتبة $\{\tau_j\}$ لتقدير دالة البقاء التي قد تجعلها تقفز او تنتقل سريعا(Jumps) [6] .

خوارزمية (1) خوارزمية مقدر Turnbull

الخطوة الاولى: ضع(r=0).وحدد(ε)التي تساوي عدد حقيقي موجب صغير، وحدد القيم الاولية $p_j^{(0)}(z_0), j=1,\dots,g$ للكتل الاحتمالية على سبيل المثال

$$p_j^{(0)}(z_0)=\begin{cases}\frac{1}{m} & \text{اذا كانت } Bj \text{ فترة داخلية}\\ 0 & \text{حالات اخرى}\end{cases}\;.....(12)$$

اذا لم يتم بعد تحديد الفترات الداخلية(innermost intervals)، ويمكن للفرد تحديد $\forall j$ $p_j^{(0)}(z_0)=\frac{1}{g}$.

الخطوة الثانية: احسب

$$p_j^{(r+1)}(z_0)=\sum_{i=1}^{n}\frac{\alpha_{ij}p_j^{(r)}(z_0)}{\sum_{k=1}^{g}\alpha_{ik}p_k^{(r)}(z_0)}\,j=1,\dots,g\;.....(13)$$

وليكن r=r+1 .

الخطوة الثالثة: اذا كان

$$\frac{\|p^r(z_0)-p^{(r-1)}(z_0)\|}{\|p^r(z_0)\|}>\varepsilon\;....(14)$$

اذ ان $p^{(r)}(z_0)=(p_1^{(r)}(z_0),p_2^{(r)}(z_0)\dots,p_g^{(r)}(z_0))^T$، اعد الخطوة الثانية، غيرها من الحالات توقف.

وان التقدير المقابل لـ$S(t|z_0)$ اذا كان احدهما يعين جميع كتل الاحتمالات من فئات التكافؤ فأنها تكون وفق الشكل الاتي[3]:

$$\hat{S}_h(t|z_0) = \begin{cases} 1 \quad , & t < \tau_1 \\ \sum_{j:\tau_j > t} \hat{p}_j(z_0), & t \geq \tau_1 \end{cases} \quad \dots\dots (15)$$

حيث ان $\tau_1 = \min\{L_i, R_i, i = 1, \dots, n\}$.

2- طريقة مقدر (Turnbull) العام

Generalization of Turnbull's Estimator

لتعميم مقدر (Turnbull) نقوم ببناء مقدر جديد نضيف من خلاله الاوزان في معادلة الاتساق الذاتي لـ(Turnbull) لمقدر الامكان الاعظم اللامعلمي (NPMLE) المستخدم في تقدير دالة البقاء ، وفي الواقع فان اي مرجح من هذا القبيل يجب ان يختلف مع j ،حيث في حالات اخرى يتم الغاءه في معادلة (11) بدلا من ذلك نقوم بوزن او ترجيح (weight) لـ(ith) في مجموع معادلة (11) وعلى النحو الاتي :

$$w_i^h(z_0) = \frac{K_h\{Z_i - z_0\}}{\sum_{q=1}^{n} K_h\{Z_q - z_0\}} \quad , i = 1, .., n \quad \dots\dots(16)$$

حيث ان k(.) هي دالة kernel وان h معلمة عرض الحزمة (المعلمة التمهيدية)، اذا كان dim(Z)=d فان k(.) عادة ما يتم اختياره كحاصل ضرب لدالة الكثافة المتماثلة حول الصفر ، لان اختيار معلمة عرض الحزمة h يكون صعب ولان التقدير اللامعلمي الكامل لـ$S(t|z_0)$ من المرجح ان يكون صعب تحت بيانات المراقبة لفترة (Interval censored) اذا كان (d>1)،وفيما تبقى نحدد الحالة التي يكون فيها المتغير المشترك(Zi) هو ذو بعد واحد(one-dimensional)،وفي واقع الامر نهدف الى اقتراح طريقة لاستخدامها في المراحل الاستطلاعية الاولية للتحليل، على الرغم من ان نتائجنا هي بالنسبة الى h ثابتة من الناحية النظرية اذا كانت(Zi) هي (d-dimensional) [3]،كما ان التطبيق العددي واختيار عرض الحزمة تكون بطبيعة الحال اكثر صعوبة عندما يكون dim(Z)> 1 ، حيث يجب استخدام اساليب التمهيد الاخرى اكثر ملاءمة مثل اساليب (splines, penalized likelihood)، ولتكن $p_j(z_0)$ هي كتلة الاحتمال المعينة او المخصصة لـ(B_j) لدالة البقاء الشرطية $S(t|z_0)$ كقيمة تم الحصول عليها من خلال حل معادلة الاتساق الذاتي:

$$p_j(z_0) = \sum_{i=1}^{n} w_i^h(z_0) \frac{\alpha_{ij} p_j(z_0)}{\sum_{k=1}^{g} \alpha_{ik} p_k(z_0)} \quad , j = 1, \dots, g \quad \dots.(17)$$

اذ تم استعمال دالة kernel الطبيعية القياسية(standard normal kernel) ودالة (Epanechnikov kernel) والتي تكون صيغتها $K(x) = 0.75(1 - x^2)\ 1_{\{x\in[-1,1]\}}$ ، حيث تستخدم(1_A) كمؤشر للحدث A، وغيرها من دوال الكثافة الاخرى المتماثلة حول الصفر مثل دوال (tri-cube or uniform kernels) يمكن استخدامها ايضا ،على الرغم من ان شكل الدالة لـ Kernel ليس لها تأثير كبير علميا على الخصائص الاحصائية للمقدر وان طرق اختيار معلمة عرض الحزمة تعمل جيدا مع دالة Kernel الطبيعية (normal kernel)،كما هو الحال مع خوارزمية (Turnbull) فان المعادلة (17) تقترح خوارزمية الحل التكراري الاتي [3] .

خوارزمية(2)

الخطوة الاولى: ضع(r=0).وحدد(ε)التي تساوي عدد حقيقي موجب صغير، وحدد القيم الاولية $p_j^{(0)}(z_0), j = 1, \dots, g$ للكتل الاحتمالية على سبيل المثال

$$p_j^{(0)}(z_0) = \begin{cases} \frac{1}{m} & \text{اذا كانت } Bj \text{ فترة داخلية} \\ 0 & \text{حالات اخرى} \end{cases} \dots\dots(18)$$

اذا لم يتم بعد تحديد الفترات الداخلية(innermost intervals)، ويمكن للفرد تحديد $p_j^{(0)}(z_0) = \frac{1}{g}\forall j$.

الخطوة الثانية: احسب

$$p_j^{(r+1)}(z_0) = \sum_{i=1}^{n} w_i^h(z_0) \frac{\alpha_{ij} p_j^{(r)}(z_0)}{\sum_{k=1}^{g} \alpha_{ik} p_k^{(r)}(z_0)} j = 1, \dots, g \dots\dots(19)$$

وليكن r=r+1 .

الخطوة الثالثة: اذا كان

$$\frac{\| p^r(z_0) - p^{(r-1)}(z_0) \|}{\| p^r(z_0) \|} > \varepsilon \dots\dots(20)$$

اذ ان $p^{(r)}(z_0) = (p_1^{(r)}(z_0), p_2^{(r)}(z_0) \dots, p_g^{(r)}(z_0))^T$، اعد الخطوة الثانية، غيرها من الحالات توقف.

وان التقدير المقابل لـ$S(t|z_0)$ اذا كان احدهما يعين جميع كتل الاحتمالات لنقاط الجانب الايسر من الفترات الداخلية فأنها تكون وفق الشكل الات[3]:

$$\hat{S}_h(t|z_0) = \begin{cases} 1 \quad , & t < \tau_i \\ \sum_{j:\tau_j > t} \hat{p}_j^h(z_0) , & t \geq \tau_i \end{cases} \dots\dots(21)$$

حيث ان $\tau_1 = \min\{L_i, R_i, i = 1, \dots, n\}$.

اختيار عرض الحزمة Bandwidth selection

ان المعلمة التمهيدية (عرض الحزمة bandwidth) والتي يرمز لها بالرمز(h) والتي تسمى في بعض الاحيان بـ(حجم النافذة وكذلك بمعلمة الانتشار) وغيرها من التسميات ، حيث تؤثر هذه المعلمة بشكل كبير جدا على تمهيد المنحنى المقدر واقترابه من المنحنى الحقيقي، حيث ان الاستعمال العلمي لمقدر Kernel يتطلب تحديد المعلمة التمهيدية على الرغم من ان دوال Kernel مهمة في الحصول على مقدرات تتقارب مع الخواص الاستدلالية الاحصائية، ويعتبر اختيار عرض الحزمة الملائم جزء اساسي من التقريب ،حيث يجب ايجاد الطريقة المثلى لغرض الموازنة بين كلا من التباين والتحيز بحيث يكون الخطأ اقل ما يمكن وان الموازنة لكل من التحيز وكذلك التباين تكن من خلال استعمال المعلمة التمهيدية(عرض الحزمة bandwidth) ، اذ ان هذه المعلمة تؤثر بشكل كبير في التحيز والتباين وان زيادة المعلمة التمهيدية يؤدي الى زيادة التحيز وتقليل التباين والعكس صحيح، وسوف يتم استخدام الطريقة ادناه لمعلمة عرض الحزمة عندما تكون الاوزان تعتمد على (normal kernel) اذا يتم استعمال الطريقة ادناه:-

طريقة قاعدة الابهام Rule of thumb

يتم الاستفادة من حقيقة ان المتغير المشترك Z يكون غير مراقب لفترة وبالتالي نستخدم طريقة قاعدة الابهام او ما تسمى ايضا بطريقة القياس التقريبي لـ(Silverman 1986) وتطبيق هذه الطريقة عند استخدام تقدير كثافة kernel مع دالة (Gaussian kernel) لان التقدير ليس دالة كثافة او متوسط متغير الاستجابة الملاحظ تماما، وان قيمة المعلمة التمهيدية تكون وفق الصيغة الاتية[7] :

$$h_{opt} = 1.06\ \hat{\sigma}_z\ n^{-1/5} \quad \dots\dots (22)$$

حيث ان :

h_{opt} : تمثل تقدير عرض الحزمة.

$\hat{\sigma}_z$: تمثل الانحراف المعياري للعينة Zi,i=1,…,n.

5-الجانب التطبيقي

5-1- مرض سرطان الثدي

يعد مرض السرطان من الامراض الخطيرة والاكثر انتشارا في العالم بمختلف انواعه اذ يصاب به تقريبا ربع السكان ويحتل المركز الثاني بالنسبة الى اكثر اسباب الوفاة شيوعا، ويعتبر سرطان الثدي (breast cancer) نوع من أنواع السرطانات الاكثر انتشارا و يظهر 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، وتسجل في كل عام حوالي (41) الف اصابة جديدة بسرطان الثدي في المملكة المتحدة ،واكثر من مليون اصابة حول العالم وعلى الرغم من ان سرطان الثدي اكثر انتشارا لدى النساء الا انه يصيب الرجال ايضا ،ففي بريطانيا تسجل (300) اصابة للرجال بسرطان الثدي سنويا، وتشير الدراسات الى احتمال اصابة امرأة واحدة من اصل تسع نساء بسرطان الثدي في مرحلة ما من حياتها.

5-2- وصف البيانات

استطاعت الباحثة الحصول على بيانات المراقبة لعدد من المرضى المصابين بسرطان الثدي في العراق البالغ عددهم (114) مريضاً وللفترة من (2015/11/1)الى (2017/1/1) وتشمل البيانات مرض سرطان الثدي حيث تم اخذ عينتين تتكون العينة الاولى من (n_1= 58) مريضة مصابة بالمرض وتتلقى العلاج الكيميائي اي [CT,58] في حين تتكون العينة الثانية من (n_2=56) مريضة مصابة بالمرض وتتلقى العلاج الاشعاعي اي [RT,56] وقد تم اخذ البيانات من سجلات الطبلات المرضى وكذلك مراجعتهم للأطباء مستشفى الاورام التعليمي بمدينة الطب وكانت فترة الدراسة (58) اسبوع حيث كانت تتم الزيارات (مراجعة المرضى) كل (4-6) اسبوع كما تختلف اوقات الزيارة الفعلية من مريض لمريض اخر حيث من خلال زيارات المرضى لوحظ عند اخذ العلاج هناك بعض الاعراض التي تصيب المرضى وكان من اهمها اصابة المريض بشلل و خدر في الذراع سواء الذراع اليمن 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 للمراقبة من اليمين وبالنسبة للمرضى الاخرين فان المشاهدات هي الفترات الزمنية التي حدث فيها الاثار السلبية حيث تم اعطى هذه الفترة بواسطة اخر فحص في اخر زيارة على سبيل المثال الفترة [49,56] اي انه في الاسبوع (49) لم يحدث اي خدر او شلل للمريض ولكن في الاسبوع (56) ظهر للمريض هذه الاعراض وهذا يعني ان ظهور الاعراض تحصر ضهورها في هذه الفترة وهذا يمثل الوقت الذي يستغرقه للإصابة بهذه الاعراض .

5-3- تقدير دالة البقاء الشرطية بالطرائق اللامعلمية

لغرض تقدير دالة البقاء الشرطية نقوم بترتيب البيانات للحصول على فئات التكافؤ(الفترات الداخلية) لمجموعة بيانات المراقبة لفترة ولكلا العلاجين حيث يمكن لدالة البقاء ان يحصل لها قفز(Jumps) او الانتقال السريع، اذ يتم الترتيب حسب جدول رقم (1)، وقد تم استعمال برنامج (MATLAB) للحصول على النتائج وكما يلي:-

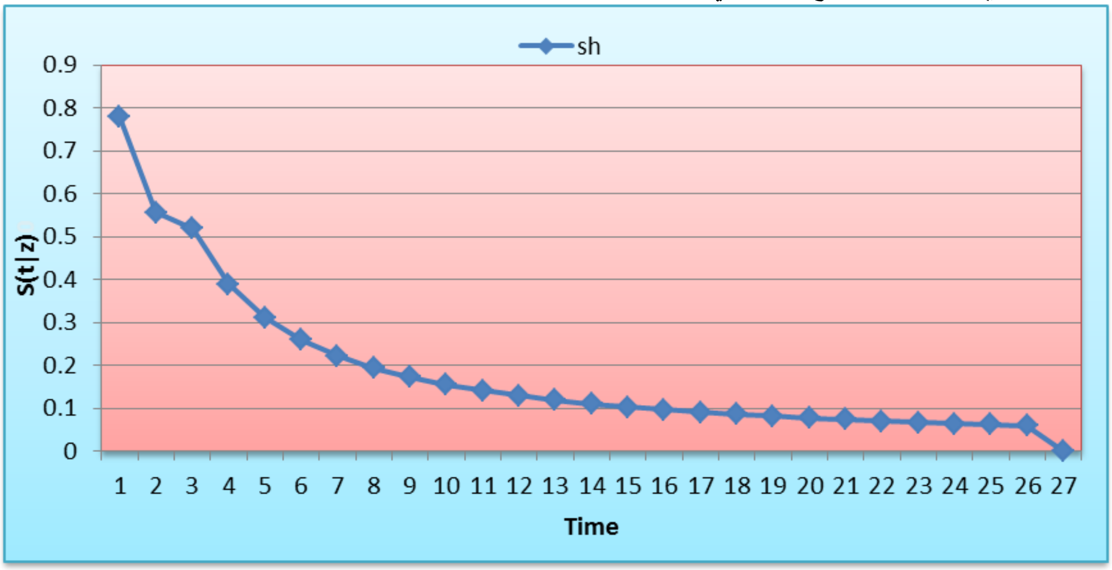
5-3-1 تقدير دالة البقاء الشرطية باستعمال Turnbull Estimator للعلاج الكيميائى(CT)

نلاحظ من جدول رقم (2) والشكل(1) انه تم الحصول على (27) فئة من فئات التكافؤ (الفترات الداخلية) من اصل (58) فئة بعد الترتيب، ونلاحظ ايضا بان قيم دالة البقاء الشرطية تبدا بالتناقص مع ازدياد الزمن اذ ان احتمال البقاء خلال زمن الفترة الاولى يكون مساويا الى (0.778327276445632) وتبدا قيمة هذا الاحتمال بالتناقص تباعا ووصولا الى زمن الفترة الاخيرة التي يكون فيها قيمة الاحتمال مساويا الى (0) وكما يوضح جدول رقم (2)وشكل (1).

جدول (2) يوضح تقدير دالة البقاء الشرطية باستعمال Turnbull Estimator للعلاج الكيميائي (CT)

	L	R	S(t|z)Sh
1	6.000000000000000	7.000000000000000	0.778327276445632
2	7.000000000000000	9.000000000000000	0.556654552891263
3	9.000000000000000	10.000000000000000	0.518884850963754
4	10.000000000000000	13.000000000000000	0.389163638222816
5	13.000000000000000	14.000000000000000	0.311330910578253
6	15.000000000000000	16.000000000000000	0.259442425481877
7	17.000000000000000	18.000000000000000	0.222379221841609
8	18.000000000000000	19.000000000000000	0.194581819111408
9	20.000000000000000	22.000000000000000	0.172961616987918
10	22.000000000000000	25.000000000000000	0.155665455289126
11	25.000000000000000	26.000000000000000	0.141514050262842
12	26.000000000000000	27.000000000000000	0.129721212740939
13	28.000000000000000	29.000000000000000	0.119742657914713
14	30.000000000000000	32.000000000000000	0.111189610920805
15	35.000000000000000	36.000000000000000	0.103776970192751
16	37.000000000000000	38.000000000000000	0.097290909555704
17	38.000000000000000	39.000000000000000	0.091567914875957
18	39.000000000000000	40.000000000000000	0.086480808493959
19	41.000000000000000	42.000000000000000	0.081929186994277
20	44.000000000000000	45.000000000000000	0.077832727644563
21	45.000000000000000	46.000000000000000	0.074126407280536
22	47.000000000000000	48.000000000000000	0.070757025131421
23	49.000000000000000	50.000000000000000	0.067680632734403
24	51.000000000000000	53.000000000000000	0.064860606370469
25	54.000000000000000	55.000000000000000	0.062266182115651
26	56.000000000000000	57.000000000000000	0.059871328957356
27	57.000000000000000	58.000000000000000	0.000000000000000

شكل (1) يوضح تقدير دالة البقاء الشرطية باستعمال طريقة مقدر (Turnbull's Estimator) لبيانات سرطان الثدي باستعمال العلاج الكيميائي



5-3-2 تقدير دالة البقاء الشرطية باستعمال Generalization of Turnbull's Estimator للعلاج الكيميائي (CT)

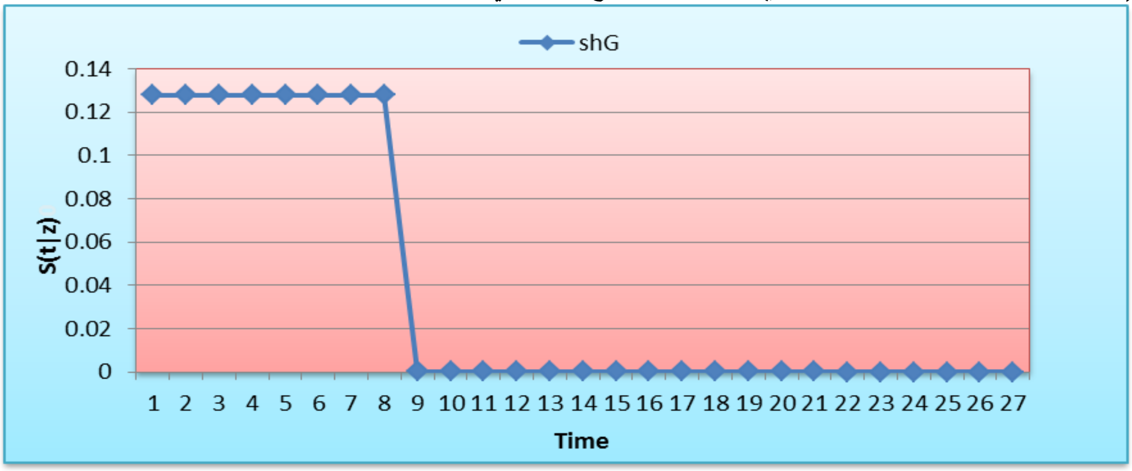
نلاحظ من جدول رقم (3) والشكل(2) انه تم الحصول على (27) فئة من فئات التكافؤ (الفترات الداخلية) من اصل (58) فئة بعد الترتيب، ونلاحظ ايضا بان قيم دالة البقاء الشرطية تبدا بالتناقص اذ ان احتمال البقاء خلال زمن كل من الفترة الاولى الى الفترة التاسعة يكون مساويا الى (0.127934844101704) اي ان التناقص يحدث في زمن الفترة التاسعة ويستقر الاحتمال الذي يكون مساويا الى (0.000019533111465) وصولا الى زمن الفترة الثانية والعشرون ويستقر الاحتمال ايضا الذي يكون مساويا الى (0.000000000008519) وتبدا قيمة هذا الاحتمال بالتناقص تباعا ووصولا الى زمن الفترة الاخيرة التي يكون فيها قيمة الاحتمال مساويا الى (0) وكما يوضح جدول رقم (3)والشكل (2).

جدول (3) طريقة (Generalization of Turnbull's Estimator) باستعمال العلاج الكيميائي(CT)

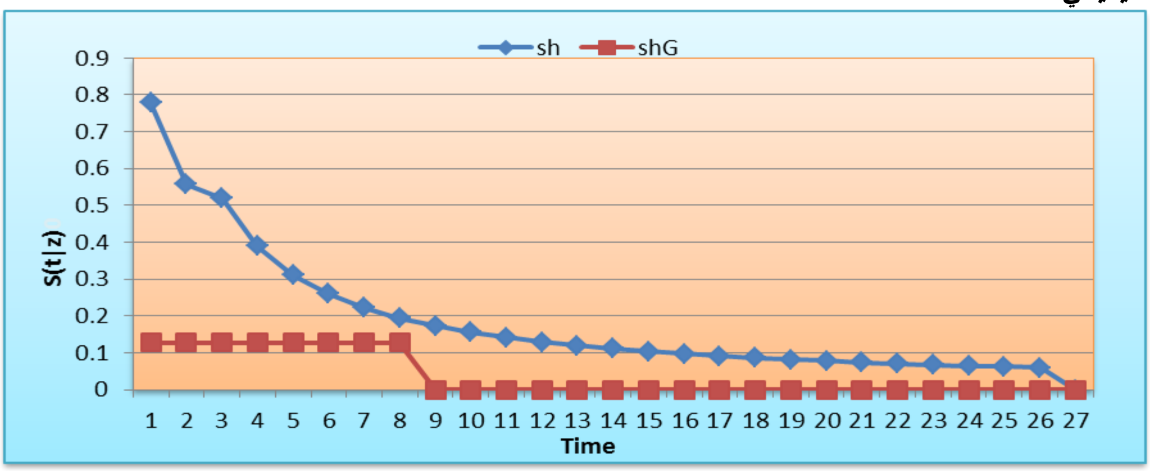
	L	R	S(t|z) shG
1	6.000000000000000	7.000000000000000	0.127934844101704
2	7.000000000000000	9.000000000000000	0.127934844101704
3	9.000000000000000	10.000000000000000	0.127934844101704
4	10.000000000000000	13.000000000000000	0.127934844101704
5	13.000000000000000	14.000000000000000	0.127934844101704
6	15.000000000000000	16.000000000000000	0.127934844101704
7	17.000000000000000	18.000000000000000	0.127934844101704
8	18.000000000000000	19.000000000000000	0.127934844101704
9	20.000000000000000	22.000000000000000	0.000019533111465
10	22.000000000000000	25.000000000000000	0.000019533111465
11	25.000000000000000	26.000000000000000	0.000019533111465
12	26.000000000000000	27.000000000000000	0.000019533111465
13	28.000000000000000	29.000000000000000	0.000019533111465
14	30.000000000000000	32.000000000000000	0.000019533111465
15	35.000000000000000	36.000000000000000	0.000019533111465

16	37.000000000000000	38.000000000000000	0.000019533111465
17	38.000000000000000	39.000000000000000	0.000019533111465
18	39.000000000000000	40.000000000000000	0.000019533111465
19	41.000000000000000	42.000000000000000	0.000019533111465
20	44.000000000000000	45.000000000000000	0.000019533111465
21	45.000000000000000	46.000000000000000	0.000019533111465
22	47.000000000000000	48.000000000000000	0.000000000008519
23	49.000000000000000	50.000000000000000	0.000000000008519
24	51.000000000000000	53.000000000000000	0.000000000008519
25	54.000000000000000	55.000000000000000	0.000000000008519
26	56.000000000000000	57.000000000000000	0.000000000000012
27	57.000000000000000	58.000000000000000	0.000000000000000

شكل(2) يوضح تقدير دالة البقاء الشرطية باستعمال طريقة (Generalization of Turnbull’s Estimator) لبيانات سرطان الثدي باستعمال العلاج الكيميائي



شكل(3) يوضح تقدير دالة البقاء الشرطية باستعمال طريقة مقدر (Generalization of Turnbull’s Estimator) وطريقة مقدر (Turnbull’s Estimator) لبيانات مرض سرطان الثدي باستعمال العلاج الكيميائي



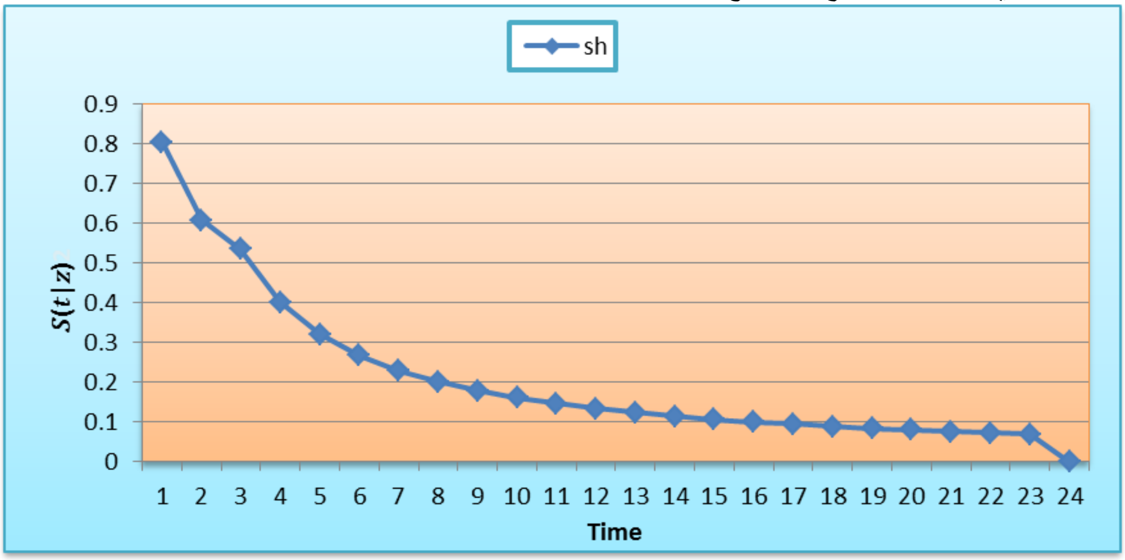
5-3-3-تقدير دالة البقاء الشرطية باستعمال Turnbull Estimator للعلاج الاشعاعي(RT)

نلاحظ من جدول رقم (4) والشكل(4) انه تم الحصول على (24) فئة من فئات التكافؤ (الفترات الداخلية) من اصل (56) فئة بعد الترتيب، ونلاحظ ايضا بان قيم دالة البقاء الشرطية تبدا بالتناقص مع ازدياد الزمن اذ ان احتمال البقاء خلال زمن الفترة الاولى يكون مساويا الى (0.803365241061984) وتبدا قيمة هذا الاحتمال بالتناقص تباعا ووصولا الى زمن الفترة الاخيرة التي يكون فيها قيمة الاحتمال مساويا الى (0) وكما يوضح جدول رقم (4) وشكل (4).

جدول (4) طريقة (Turnbull's Estimator) باستعمال العلاج بالإشعاع (RT)

T	L	R	S(t|z0)Sh
1	6.000000000000000	7.000000000000000	0.803365241061984
2	8.000000000000000	9.000000000000000	0.606730482123968
3	9.000000000000000	10.000000000000000	0.535576827374656
4	10.000000000000000	11.000000000000000	0.401682620530992
5	11.000000000000000	12.000000000000000	0.321346096424794
6	13.000000000000000	14.000000000000000	0.267788413687328
7	17.000000000000000	19.000000000000000	0.229532926017710
8	19.000000000000000	20.000000000000000	0.200841310265496
9	20.000000000000000	21.000000000000000	0.178525609124885
10	22.000000000000000	24.000000000000000	0.160673048212397
11	25.000000000000000	27.000000000000000	0.146066407465815
12	27.000000000000000	28.000000000000000	0.133894206843664
13	28.000000000000000	29.000000000000000	0.123594652471074
14	31.000000000000000	32.000000000000000	0.114766463008855
15	34.000000000000000	35.000000000000000	0.107115365474931
16	35.000000000000000	36.000000000000000	0.100420655132748
17	37.000000000000000	38.000000000000000	0.094513557771998
18	40.000000000000000	41.000000000000000	0.089262804562443
19	44.000000000000000	45.000000000000000	0.084564762217051
20	45.000000000000000	46.000000000000000	0.080336524106198
21	49.000000000000000	50.000000000000000	0.076510975339237
22	53.000000000000000	54.000000000000000	0.073033203732908
23	54.000000000000000	55.000000000000000	0.069857847048868
24	55.000000000000000	56.000000000000000	0.000000000000000

شكل(4) يوضح تقدير دالة البقاء الشرطية باستعمال طريقة مقدر (Turnbull's Estimator) لبيانات سرطان الثدي باستعمال العلاج بالإشعاع



4-3-5 تقدير دالة البقاء الشرطية باستعمال Generalization of Turnbull's Estimator للعلاج بالإشعاع(RT)

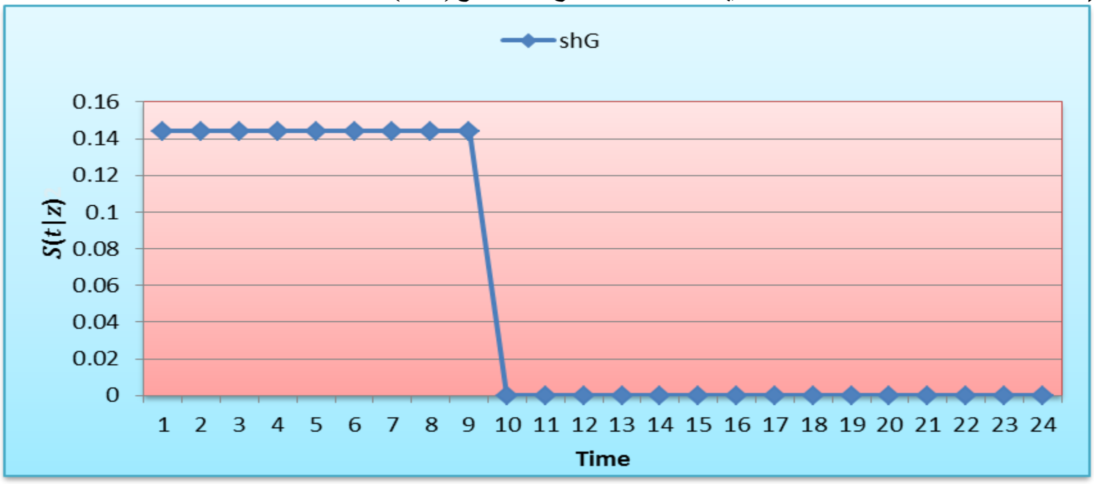
نلاحظ من جدول رقم (5) والشكل(5) انه تم الحصول على (24) فئة من فئات التكافؤ (الفترات الداخلية) من اصل (56) فئة بعد الترتيب، ونلاحظ ايضا بان قيم دالة البقاء الشرطية تبدا بالتناقص اذ ان احتمال البقاء خلال زمن كل من الفترة الاولى الى الفترة العاشرة ويكون مساويا الى (0.143996757239959) اي ان التناقص يحدث في زمن الفترة العاشرة ويستقر الاحتمال الذي يكون مساويا الى (0.000031349231543) وصولا الى زمن الفترة السادسة عشر ويستقر الاحتمال ايضا الذي يكون مساويا الى (0.000000000024699) ويحدث التناقص ايضا في زمن الفترة العشرون والحادية والعشرون الثانية والعشرون وتبدا قيمة هذا الاحتمال بالتناقص تباعا ووصولا الى زمن الفترة الاخيرة التي يكون فيها قيمة الاحتمال مساويا الى (0) وكما يوضح جدول رقم (5) والشكل (5).

جدول (5) طريقة (Generalization of Turnbull's Estimator) باستعمال العلاج بالإشعاع(RT)

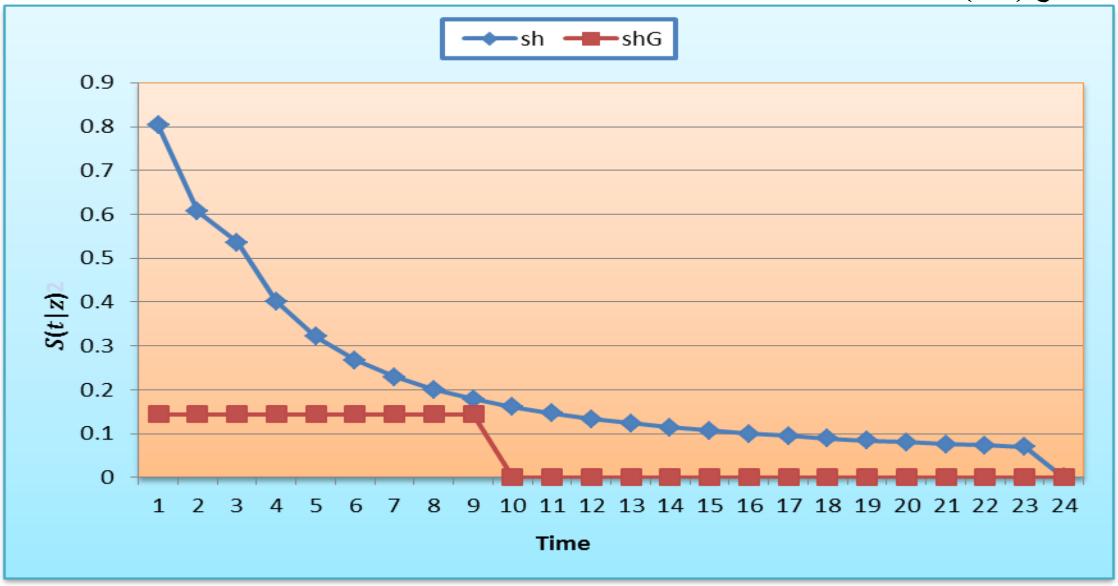
T	L	R	S(t|z0)ShG
1	6.000000000000000	7.000000000000000	0.143996757239959
2	8.000000000000000	9.000000000000000	0.143996757239959
3	9.000000000000000	10.000000000000000	0.143996757239959
4	10.000000000000000	11.000000000000000	0.143996757239959
5	11.000000000000000	12.000000000000000	0.143996757239959
6	13.000000000000000	14.000000000000000	0.143996757239959
7	17.000000000000000	19.000000000000000	0.143996757239959
8	19.000000000000000	20.000000000000000	0.143996757239959
9	20.000000000000000	21.000000000000000	0.143996757239959
10	22.000000000000000	24.000000000000000	0.000031349231543
11	25.000000000000000	27.000000000000000	0.000031349231543
12	27.000000000000000	28.000000000000000	0.000031349231543
13	28.000000000000000	29.000000000000000	0.000031349231543
14	31.000000000000000	32.000000000000000	0.000031349231543
15	34.000000000000000	35.000000000000000	0.000031349231543

16	35.000000000000000	36.000000000000000	0.000000000024699
17	37.000000000000000	38.000000000000000	0.000000000024699
18	40.000000000000000	41.000000000000000	0.000000000024699
19	44.000000000000000	45.000000000000000	0.000000000024699
20	45.000000000000000	46.000000000000000	0.000000000002311
21	49.000000000000000	50.000000000000000	0.000000000001123
22	53.000000000000000	54.000000000000000	0.000000000000101
23	54.000000000000000	55.000000000000000	0.000000000000000
24	55.000000000000000	56.000000000000000	0.000000000000000

شكل(5) يوضح تقدير دالة البقاء الشرطية باستعمال طريقة (Generalization of Turnbull's Estimator) لبيانات سرطان الثدي باستعمال العلاج بالإشعاع(RT)



شكل(6) يوضح تقدير دالة البقاء الشرطية باستعمال طريقة مقدر (Generalization of Turnbull's Estimator) وطريقة مقدر (Turnbull's Estimator) لبيانات مرض سرطان الثدي باستعمال العلاج بالإشعاع (RT)



5-3-5 - مؤشر متوسط مربعات الخطأ MSE

جدول (6) يوضح قيم مؤشر متوسط مربعات الخطأ بالطريقتين ولكل من العلاج الكيميائي والعلاج بالإشعاع

الطرائق	العلاج الكيميائي(CT)	العلاج بالإشعاع(RT)
طريقة(Turnbull's Estimator)	0.144241881874544	0.172362314200145
طريقة (Generalization of Turnbull's Estimator)	0.004849577764543	0.007775650031552

نلاحظ من الجدول اعلاه بان طريقة (Generalization of Turnbull's Estimator) هي الطريقة الافضل ولكل من العلاج الكيميائي والعلاج بالإشعاع حيث تمتلك اقل متوسط مربعات خطا (MSE) .

6- الاستنتاجات والتوصيات

6-1- الاستنتاجات

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

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 طريقة (Generalization of Turnbull's Estimator) هي الافضل حيث 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 استقرارا ووقت بقاء اطول اذ ان الانحلال او الاضمحلال السريع اي التناقص في قيمة الاحتمال لمنحنى المرضى يحدث في زمن كل من الفترة التاسعة وزمن الفترة الثانية والعشرون وكذلك زمن الفترة السادسة والعشرون ويصل قيمة الاحتمال لدالة البقاء الى الصفر في زمن الفترة الاخيرة .

3- بالنسبة لمنحنيات المرضى الذين يتعاطون العلاج بالإشعاع فان الانحلال او الاضمحلال السريع اي التناقص في قيمة الاحتمال لمنحنى المرضى يحدث في زمن كل من الفترة العاشرة وزمن الفترة السادسة عشر وكذلك زمن الفترة العشرون والواحد والعشرون والثانية والعشرون ونرى بان قيمة الاحتمال لدالة البقاء لمنحنى المرضى تصل الى الصفر في زمن الفترة قبل الاخيرة وزمن الفترة الاخيرة .

6-2- التوصيات

من خلال الاستنتاجات واجراءات البحث تم التوصل الى مجموعة من التوصيات يمكن ايجازها بما يلي:-

1- يمكن الاعتماد على طريقة (Generalization of Turnbull's Estimator) في تقدير دالة البقاء الشرطية وذلك لان اداءها متفوقا على الطريقة 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 استقرارا.

2- استعمال طرائق لامعلمية اخرى لتقدير دالة البقاء الشرطية وذلك باستعمال بيانات مراقبة من اليمين مثلا طريقة مقدر (Beran).

3- استعمال طرائق مختلفة لعرض الحزمة في طريقة (Generalization of Turnbull's Estimator) في تقدير دالة البقاء الشرطية.

7- المصادر

1- حسن ، رعد فاضل ، وصالح ،عائدة هادي،(2014)،" استخدام نموذج انحدار كوكس (Cox - Regression)لأوقات بقاء المرضى المصاب نٌ بمرض سرطان الدماغ ف العراق "، بحث منشور في مجلة كلية الرافدين لجامعة العلوم، العدد(34)، طبعة(6870-1681).

2- AL-Nasser, Abdul Majeed ,(2009) "Statistical Reliability" Ithraa Publishing and Distribution ,University of Baghdad .

3- Dehghan , M. H. and Duchesne, T. (2011),'' A generalization of Turnbull's estimator for nonparametric estimation of the conditional survival function with interval-censored data'', Lifetime Data Analysis, 17:234–255.

4-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.

5-Klein,J,P. and Moeschberger, M. L.,(2003),''Survival Analysis: Techniques for censored and truncated data'', 2rd ed. New York : Springer.

6- Pan, W. (1998),'' Smooth estimation of the survival for interval censored data: An Empirical study '', Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN 55455-0378, U.S.A.

7- Silverman, B.W. (1986) " Density estimation for statistics and data analysis " Chapman and Hall, London.

8- Sun, J. (2006),'' The Statistical Analysis of Interval-Censored Failure Time Data'', Springer, New York.

9- Turnbull, B. W. (1976), "The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data" ,Journal fo the Royal Statistical Society, Series B, Vol.38PP.290-295.

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 Two of (Turnbull) and (Generalization Turnbulls)non-parametric methods in estimating conditional survival function (applied study on breast cancer patients)

Abstract

This research includes the application of non-parametric methods in estimating the conditional survival function represented in a method (Turnbull) and (Generalization Turnbull's) using data for Interval censored of breast cancer and two types of treatment, Chemotherapy and radiation therapy and age is continuous variable, The algorithm of estimators was applied through using (MATLAB) and then the use average Mean Square Error (MSE) as amusement to the estimates and the results showed (generalization of Turnbull's) In estimating the conditional survival function and for both treatments ,The estimated survival of the patients does not show very large differences, but the curves of the patients taking the chemotherapy were more stable and the survival time is longer since the rapid decay or decay, the decrease in the probability value of the patient curve occurs at the time of the ninth and twenty-second periods as well as the period of the twenty-sixth The probability of a survival function is zero at the time of the last Interval.

Keywords: **survival function, Turnbull estimator, Turnbull's Generalization, Kernel weight, Bandwidth parameter.**