

استخدام توزيع رايلي المعمم المتقطع في دالة البقاء مع التطبيق

أ.د. جواد كاظم خضير الموسوي
كلية الادارة والاقتصاد الجامعة المستنصرية

الباحثة : نبأ صالح هادي
كلية الادارة والاقتصاد الجامعة المستنصرية

تاريخ استلام البحث: 2022/04/16

تاريخ قبول البحوث: 2022/05/10

نشر البحث في العدد عشرون: ايلول / سبتمبر 2023

رمز التصنيف ديوي / النسخة الالكترونية (Online): 2522-64X/519.5

رمز التصنيف ديوي / النسخة الورقية (Print): 2519-948X/519.5

استخدام توزيع رايلي المعمم المتقطع في دالة البقاء مع التطبيق

أ.د. جواد كاظم خضير الموسوي
كلية الادارة والاقتصاد الجامعة المستنصرية
الباحثة : نبأ صالح هادي
كلية الادارة والاقتصاد الجامعة المستنصرية

المستخلص

إن الاهتمام المتزايد بالبحوث والدراسات الحديثة لنظرية البقاء جاء نتيجة للدور الذي تؤديه في دراسة معدل زمن واحتمال بقاء الكائن الحي بعد مدة محددة من الزمن، حيث تعد دالة البقاء (survival function) من اهم الدوال في علم الإحصاء التي لها دور أساسي في دراسة وتحليل معظم الظواهر اعتماداً على البيانات والمعلومات الإحصائية المتوفرة عن تلك الظاهرة.

ان معظم النماذج التي تناولتها الدراسات الإحصائية يتم بناؤها على المتغيرات المستمرة كالتوزيع الأسي Continuous Exponential Distribution وتوزيع ويبل Continuous Weibull Distribution. إلا أن بعض الدراسات قد عملت على التوزيعات المتقطعة كتوزيع بواسون Discrete Poisson Distribution. ففي السنوات الأخيرة، تم التوجه إلى تكوين نماذج لتوزيعات متقطعة بناءً على توزيعات مستمرة كتوزيع ويبل المتقطع Discrete Weibull Distribution وتوزيع كما المتقطع Discrete Gamma Distribution. وفي هذا البحث تم مناقشة توزيع رايلي المعمم المتقطع Discrete generalized Rayleigh Distribution المقابل لتوزيع رايلي المعمم المستمر continuous generalized Rayleigh Distribution، ومن ثم دراسة الخواص الإحصائية كدالة البقاء (survival function) ودالة المخاطرة (hazard function)، فضلاً عن تقدير معلمات التوزيع بطريقة الإمكان الأعظم Maximum Likelihood Method. وقد تم تطبيق البيانات، والتي تمثل اوقات البقاء في المستشفى لـ 100 شخص اصيبوا بالجلطة الدماغية، وتم التوصل الى مدى ملائمة هذه البيانات لتوزيع رايلي المعمم المتقطع باستخدام اختبار كولموكروف-سميرنوف Kolmogorov-Smirnov (K-S). وقد تم تقدير المعلمات، الخواص الإحصائية ودالة البقاء للتوزيع.

1. المقدمة (1)(2)(3)(4)

يعد تحليل دالة البقاء على قيد الحياة من المواضيع الحديثة حيث تعدى المجالات الطبية الى المجالات الهندسية والاقتصادية والاجتماعية وغيرها من مجالات الحياة حيث يسمح تحليل دالة البقاء باستخدام مجموعة من الاساليب والاختبارات الاحصائية وبناء النماذج التحليلية. ويعتبر من أهم طرائق التحليل الحديثة حيث ان المتغير التابع هو الوقت حتى وقوع الحدث. لكن في الحقيقة أن معظم البحوث المنشورة في مجال البقاء تفترض أن الوقت مستمر، قد يمكن للفرد ملاحظة أن البيانات المتقطعة تظهر في كثير من الحالات العملية على أنها تقريبية من معلومات الحياة الواقعية، خاصة في حالات العينات الصغيرة. عادة ما تعتمد

الدراسات بيانات مستمرة ولم يتم عمل سوى القليل من النماذج المتقطعة. اذ ان مجموعة البيانات المتقطعة هي الاكثر شيوعا في المواقف الحقيقية لكون لديها مقاييس متقطعة للأوقات.

حيث تم اقتراح العديد من التوزيعات المتقطعة في مجال البقاء كتوزيع (Poisson, Geometric, Binomial) لكن التوزيعات المتقطعة المحددة بناءً على الأجزاء المقابلة المستمرة قد تم تقديمها لأول مرة في عام (1975) من قبل الباحثان (Nakagawa & Osaki) وهو توزيع Weibull المستمر. يعد توزيع Weibull المتقطع هو الاكثر انتشاراً من بين التوزيعات الجديدة. اضافة الى انه هناك توزيع آخر تم تطويره حديثاً هو توزيع Gamma المتقطع الذي حظي باهتمام كبير في التطبيقات العملية، اذ تم استخدامه لأول مرة من قبل (Yang, 1994) في مجال البيولوجيا الجزيئية والتطور. تعتمد دراسة تحليل دالة البقاء على بيانات المراقبة حيث ان موضوع بيانات المراقبة ومجال استخدامها من الموضوعات التي تأخذ حيزاً كبيراً ضمن الدراسات والبحوث والتطبيقات العملية الاخرى. لذا ففي هذا البحث تم مناقشة توزيع رايلي المعمم المتقطع Discrete generalized Rayleigh Distribution . وقد تم مناقشة دالة الكتلة الاحتمالية (p.m.f.)، دالة البقاء ودالة المخاطرة للتوزيع . ثم تم تقدير معالم التوزيع باستخدام طريقة الامكان الاعظم .

2. مفاهيم اساسية

2.1. دالة البقاء (4)(2) (Survival Function)

هي احتمال بقاء الكائن حياً بعد مرور الزمن (x) حيث يرمز لها بالرمز S(x) ويمكن التعبير عنها رياضياً:

$$S(x) = p_r(X > x) \quad \dots (1)$$

اذا ان S(x) : تمثل دالة البقاء عند الوقت x
X : متغير عشوائي متقطع

اي ان (x) هو الوقت المحدد و (X) هو وقت ظهور الحدث , ومن الصيغة الرياضية المذكورة اعلاه فان دالة البقاء هي احتمال كون وقت ظهور الحدث X اكبر من الوقت المحدد x اي ان $p_r(X > x)$ وأن صيغة دالة البقاء للتوزيع المتقطع هي :

$$S(x) = p_r(X > x) = \sum_{i=x+1}^{\infty} p_i \quad \dots (2)$$

دالة البقاء هي دالة مكمله للدالة التجميعية حيث زيادة قيمة دالة البقاء يعني صغر للدالة التجميعية وكبر الدالة التجميعية يعني صغر لدالة البقاء فلو كانت دالة البقاء S(x) هي $p_r(X > x)$ والدالة التجميعية F(x) هي $p_r(X \leq x)$ حيث يتم التعبير عنها رياضياً:

$$p_r(X > x) = 1 - \sum_{i=0}^x p_i = 1 - p_r(X < x)$$

وبما ان $p_r(X > x)$ هي دالة البقاء S(x) و $p_r(X < x)$ هي الدالة التجميعية F(x)

$$S(x) = \sum_{i=x+1}^{\infty} p_i = 1 - p_r(X < x)$$

$$S(x) = 1 - F(x) \quad \dots (3)$$

ومن خصائص دالة البقاء انها غير سالبة ودالة غير متزايدة (متناقصة) ترتيبه لجميع قيم المتغير X ، وبما ان دالة البقاء دالة احتمالية لذلك فأن:

$$0 \leq S(x) \leq 1$$

$$S(x = 0) = 1$$

$$S(x \rightarrow \infty) = 0$$

هذا يعني احتمال بقاء المصاب على قيد الحياة في الزمن 0 يساوي 1

2.2. دالة المخاطرة (2)(4) (Hazard function)

تعرف بانها الاحتمال الشرطي لفشل المفردة خلال فترة صغيره جدا من الوقت . بالنظر الى انها لم تفشل حتى الوقت x ويرمز لها بالرمز $h(x)$ ، تكون الدالة الاحتمالية الشرطية للمفردة عند الوقت x هي :

$$p(X = x / X \geq x) = \frac{p(X = x)}{p(X > x - 1)} \quad \dots (4)$$

$$= \frac{S(x - 1) - S(x)}{S(x - 1)}$$

$$h(x) = 1 - \frac{S(x)}{S(x - 1)} \quad \dots (5)$$

ومن خصائص دالة المخاطرة :

1. $0 \leq h(x) \leq 1$
2. $\sum_{x=0}^{\infty} h(x) = \infty$

وان دالة المخاطرة لها الميزات الاتية :

1. تقيس معدل احتمال الفشل الآنية حيث ان دالة البقاء مقياس تراكمي مع الوقت.
2. تستخدم للتعرف على شكل الانموذج الرياضي لبيانات البقاء مثل الانموذج الأسى (exponential)، وويل (Weibull).

2.3. المراقبة (4) (Censoring)

ان ما يميز دراسات دالة البقاء عن الدراسات الاحصائية هي ظاهرة المراقبة (censoring). ومن الصعب تحليل بيانات البقاء لان بعض المفردات تكون مراقبة أي لا يتم مشاهدتها للوقت الكامل وحتى الحدث، فمثلا في التجارب الطبية هناك بعض المرضى الذين بقوا على قيد الحياة في نهاية الدراسة، كذلك قد يكون حالة البقاء في وقت التحليل غير معلومة وذلك لان بعض المفردات قد فقدت المتابعة، على سبيل المثال لنفترض انه بعد التسجيل في دراسة طبية ما انتقل المريض من مدينة الى اخرى ومن ثم فان هذا المريض لا يمكن متابعته ولكن المعلومة المتوفرة لدينا عن بقاء هذا المريض هو اخر تاريخ كان معلوما لدينا بانه مازال على قيد الحياة، وهذا التاريخ هو تاريخ اخر زيارة للمريض لعيادة المتابعة النظامية، هناك سبب

آخر للمراقبة وذلك ان المفردة انسحبت من الدراسة فمثلاً في تجارب الدواء قد تسوء حالة المريض فيعطى دواء اخر).

3. توزيع رايلي المعمم المتقطع⁽¹⁾ Discrete generalized Rayleigh Distribution

ان توزيع generalized Rayleigh المتقطع ذو المعلمتين يعرف بأنه الجزء المقابل من توزيع generalized Rayleigh المستمر، التي تم اقتراحه من قبل Mudholkar and Srivastava (عام (1993) ، يعد أحد التوزيعات المعروفة مدى الحياة باستخدام دالة الكثافة الاحتمالية (p.d.f.) وعضو مميز في فئة توزيعات وييل الأسية ، ففي عام (1995) عرفه (Mudholkar et al.) كتوزيع قيمة عدد صحيح غير سالب. لنفترض المتغير العشوائي X يتبع توزيع $GRD(\alpha, \lambda)$ فان دالة الكثافة الاحتمالية معرفه كالآتي :

$$f(x) = 2\alpha\lambda^2 x e^{-(\lambda x)^2} (1 - e^{-(\lambda x)^2})^{\alpha-1} \quad x > 0 \quad \dots (6)$$

اما دالة البقاء لـ $GRD(\alpha, \lambda)$ تعرف بالشكل الآتي :

$$S(x) = 1 - (1 - e^{-(\lambda x)^2})^\alpha \quad \dots (7)$$

اذ ان α, λ هما معلمتا الشكل والقياس ع التوالي.

فأن لأي توزيع مستمر على الفترة $R^+ = [0, +\infty)$ مع دالة كثافة احتمالية (p.d.f.) يمكننا بناء نظيراً متقطعاً معتمد على مجموعة الاعداد الصحيحة $N = \{0, 1, 2, \dots\}$ سيكون له دالة كتلة احتمالية (p.m.f.) بالاعتماد على الصيغة الآتية :

$$\begin{aligned} p(X = x) &= p(x \leq X < x + 1) \\ p(X = x) &= p(X \geq x) - p(X \geq x + 1) \\ p(X = x) &= S(x) - S(x + 1) \quad x = 0, 1, 2, \dots \quad \dots (8) \end{aligned}$$

$$p(X = x) = 1 - (1 - e^{-(\lambda x)^2})^\alpha - 1 - (1 - e^{-(\lambda(x+1))^2})^\alpha \quad x = 0, 1, \dots$$

$$p(X = x) = (1 - e^{-(\lambda(x+1))^2})^\alpha - (1 - e^{-(\lambda x)^2})^\alpha \quad x = 0, 1, \dots$$

اذ ان : $q = e^{-\lambda^2}$, $0 < q < 1$, $\alpha > 0$,

$$p(X = x) = (1 - q^{(x+1)^2})^\alpha - (1 - q^{(x)^2})^\alpha \quad x = 0, 1, \dots \quad \dots (9)$$

اما الدالة التجميعية لتوزيع $DGR(\alpha, p)$ تعرف بالشكل الآتي :

$$F(x) = (1 - q^{(x+1)^2})^\alpha \quad \dots (10)$$

اما دالة البقاء لتوزيع $DGR(\alpha, p)$ تعرف بالشكل الآتي :

$$S(x) = p(X > x) = 1 - (1 - q^{(x+1)^2})^\alpha \quad \dots (11)$$

اما دالة المخاطرة لتوزيع $DGR(\alpha, p)$ تعرف بالشكل الآتي :

$$h(x) = \frac{(1 - q^{(x+1)^2})^\alpha - (1 - q^{(x)^2})^\alpha}{1 - (1 - q^{(x+1)^2})^\alpha} \quad \dots (12)$$

اما دالة معدل المخاطرة المعكوس لهذا التوزيع غير متناقصة :

$$\bar{h}(x) = \frac{(1 - p^{(x+1)^2})^\alpha - (1 - p^{(x)^2})^\alpha}{(1 - p^{(x+1)^2})^\alpha}$$

$$\bar{h}(x) = 1 - \frac{(1 - p^{(x)^2})^\alpha}{(1 - p^{(x+1)^2})^\alpha} \quad \dots (13)$$

اذ ان: $\frac{(1-p^{(x)^2})^\alpha}{(1-p^{(x+1)^2})^\alpha}$ هي الدالة المتزايدة .

4- طريقة الإمكان الأعظم Maximum Likelihood Method

$$p(X = x) = (1 - q^{(x+1)^2})^\alpha - (1 - q^{x^2})^\alpha \quad x = 0, 1, \dots$$

$$L(x_1, \dots, x_n, \alpha, q) = \prod_{i=1}^n p(X = x)$$

$$L(x_1, \dots, x_n, \alpha, q) = \prod_{i=1}^n [(1 - q^{(x_i+1)^2})^\alpha - (1 - q^{x_i^2})^\alpha] \quad \dots (14)$$

$$L(x_1, \dots, x_n, \alpha, q) = \sum_{i=1}^n [(1 - q^{(x_i+1)^2})^\alpha - (1 - q^{x_i^2})^\alpha] \quad \dots (15)$$

ولغرض تقدير معلمات دالة الإمكان بأخذ اللوغاريتم الطبيعي لطرفي المعادلة السابقة

$$\log L(x_1, \dots, x_n, \alpha, q) = \sum_{i=1}^n \log [(1 - q^{(x_i+1)^2})^\alpha - (1 - q^{x_i^2})^\alpha] \quad \dots (16)$$

وللحصول على دالة الإمكان في نهايته العظمى يتم اشتقاق الدالة بالنسبة الى المعلمات المجهولة وكالاتي :

$$\frac{\partial \log L}{\partial \alpha} = \sum_{i=1}^n \frac{(1 - q^{(x_i+1)^2})^\alpha \log(1 - q^{(x_i+1)^2}) - (1 - q^{x_i^2})^\alpha \log(1 - q^{x_i^2})}{(1 - q^{(x_i+1)^2})^\alpha - (1 - q^{x_i^2})^\alpha} = 0 \quad \dots (17)$$

$$\frac{\partial \log L}{\partial q} = \alpha \sum_{i=1}^n \frac{(1 - q^{x_i^2})^{\alpha-1} q^{x_i^2-1} x_i^2 - (1 - q^{(x_i+1)^2})^{\alpha-1} q^{(x_i+1)^2-1} (x_i+1)^2}{(1 - q^{(x_i+1)^2})^\alpha - (1 - q^{x_i^2})^\alpha} = 0 \quad \dots (18)$$

$$\frac{\partial^2 \log L}{\partial \alpha^2} = \sum_{i=1}^n \left[\frac{z_2}{z_1} - \frac{z_3}{[z_1]^2} \right] \quad \dots (19)$$

اذ ان:

$$\begin{aligned} z_1 &= (1 - q^{(x_i+1)^2})^\alpha - (1 - q^{x_i^2})^\alpha \\ z_2 &= (1 - q^{(x_i+1)^2})^\alpha [\log(1 - q^{(x_i+1)^2})]^2 - (1 - q^{x_i^2})^\alpha [\log(1 - q^{x_i^2})]^2 \\ z_3 &= [(1 - q^{(x_i+1)^2})^\alpha \log(1 - q^{(x_i+1)^2}) - (1 - q^{x_i^2})^\alpha \log(1 - q^{x_i^2})]^2 \\ \frac{\partial^2 \log L}{\partial q^2} &= \alpha \sum_{i=1}^n \frac{z_3 - z_4 + z_5}{z_1} - \frac{z_6}{[z_1]^2} \quad \dots (20) \end{aligned}$$

اذ ان:

$$\begin{aligned} z_3 &= x_i^2 (x_i^2 - 1) (1 - q^{x_i^2})^{\alpha-1} q^{x_i^2-2} - (\alpha - 1) x_i^4 (1 - q^{x_i^2})^{\alpha-2} q^{2x_i^2-2} \\ z_4 &= (x_i + 1)^2 ((x_i + 1)^2 - 1) (1 - q^{(x_i+1)^2})^{\alpha-1} q^{(x_i+1)^2-1} \end{aligned}$$

$$z_5 = (\alpha - 1)(x_i + 1)^4 q^{2(x_i+1)^2 - 2} (1 - q^{(x_i+1)^2})^{\alpha-2}$$

$$z_6 = \left[x_i^2 q^{x_i^2 - 1} (1 - q^{x_i^2})^{\alpha-1} - (x_i + 1)^2 q^{(x_i+1)^2 - 1} (1 - q^{(x_i+1)^2})^{\alpha-1} \right]^2$$

المعادلات اعلاه (17) و (18) هي غير خطية يصعب حلها بالطريقة الاعتيادية ولا بد من استخدام احدى الطرائق العددية لحلها ك (نيوتن رافسن) ليتم الحصول على المعلمات المجهولة.

5- الجانب التطبيقي Application

تم جمع البيانات، والتي تمثل اوقات البقاء في المستشفى لـ 100 شخص اصابوا بالجلطة الدماغية، قد تم الحصول على البيانات من مستشفى الزهراء التعليمي في محافظة واسط، تم احتساب مدة البقاء بالايام، والجدول الآتي يحتوي على هذه البيانات:

جدول (1): بيانات اوقات البقاء للأشخاص المصابين بالجلطة الدماغية

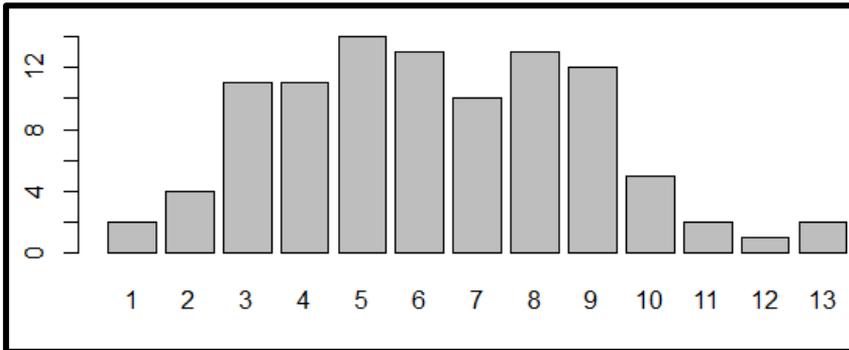
1	1	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	4	4	4
4	4	4	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
5	5	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	7	7	7	7	7	7
7	7	7	7	7	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	9	9
9	9	9	9	9	9	9	9	9	9	9	10	10	10	10	10	10	11	11	12	13	13	13

وقد تم حساب بعض الخواص الإحصائية لهذه البيانات بناءً على توزيع رايلي المعمم المتقطع (DGR)، وكما يأتي:

جدول (2): بعض الإحصاءات الوصفية لبيانات اوقات البقاء للأشخاص المصابين بالجلطة الدماغية

الوسط الحسابي	أقل قيمة	الوسيط	أعلى قيمة	التباين	الانحراف المعياري	الالتواء	التفطح
6.27	1	6	13	7.0476	2.6547	0.2376	2.5461

كما تم رسم الشريط البياني Bar Plot لهذه البيانات وكما يأتي:



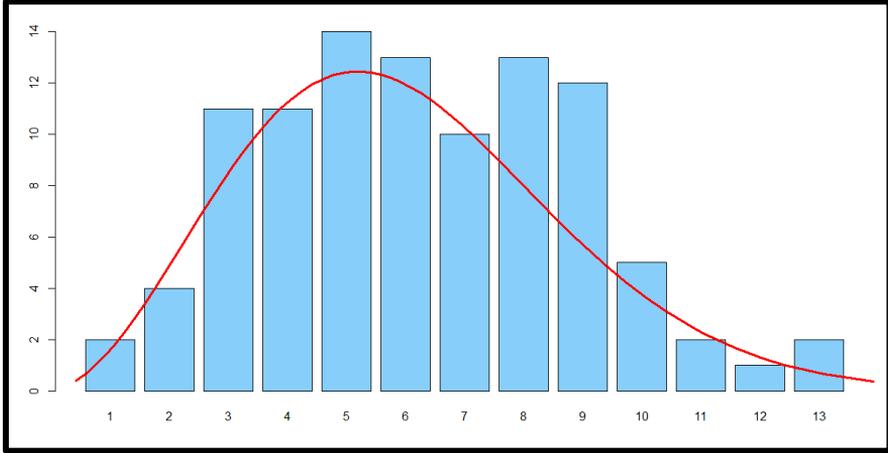
شكل (1): الشريط البياني لبيانات اوقات البقاء للأشخاص المصابين بالجلطة الدماغية

وقد تم تقدير معاملات توزيع رايلي المعمم المتقطع بعد حل المعادلتين (17) و(18)، وكما تم استخدام اختبار كولموكروف-سميرنوف Kolmogorov-Smirnov (K-S) للتأكد من ملائمة هذه البيانات لتوزيع رايلي المعمم المتقطع، وكما يأتي:

جدول (3): نتائج اختبار K-S

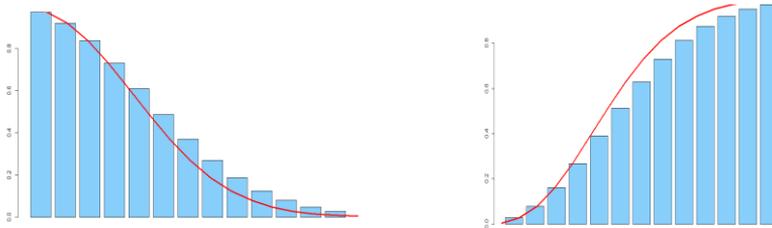
القيمة الحرجة	إحصاءة K-S	تقدير معلمة q	تقدير معلمة α
0.136	0.0876	0.9804	1.4011

حيث نلاحظ أن البيانات تتبع توزيع رايلي المعمم المتقطع لأن إحصاء اختبار كولموكروف-سميرنوف هي أقل من القيمة الحرجة لهذا الاختبار، كما نلاحظ أن القيم التقديرية لمعلمتي التوزيع هما 1.4011 و0.9804 لمعلمة α و q على التوالي، والرسم الآتي يوضح دالة الكتلة الاحتمالية بناءً على القيم التقديرية للمعاملات وقيم البيانات الأصلية وكما يأتي:



شكل (1): رسم دالة الكتلة الاحتمالية المقدر

والشكل الآتي يوضح كل من دالتي التجميع التراكمية والبقاء للبيانات الأصلية والتقديرية.



ب. دالة البقاء

أ. دالة التجميع التراكمية

وقد تم تقدير دالة البقاء بناءً على تقديرات الإمكان الأعظم كما يأتي:

$$\hat{S}(x) = 1 - (1 - \hat{q}^{(x+1)^2})^{\hat{\alpha}} = 1 - (1 - (0.9804)^{(x+1)^2})^{1.4011} \dots (21)$$

وبناءً عليها فقد تم تقدير دالة البقاء عند قيم مختلفة للمتغير x ، وكما يأتي:

جدول (4): القيم التقديرية لدالة البقاء

X	$\hat{S}(x)$	x	$\hat{S}(x)$	X	$\hat{S}(x)$
0	0.9959	6	0.4866	12	0.0488
1	0.9729	7	0.3705	13	0.0287
2	0.921	8	0.2696	14	0.0162
3	0.8389	9	0.1876	15	0.0088
4	0.732	10	0.125	16	0.0046
5	0.6107	11	0.0798	17	0.0023

حيث إن نلاحظ أن قيم دالة البقاء تبدأ بالتناقص مع ازدياد الزمن وهذا ما نسميه الدالة التناقضية.

6- الاستنتاجات Conclusion

في هذا البحث تم دراسة توزيع رايلي المعمم المتقطع، وتم مناقشة دالة البقاء ودالة المخاطرة ومن ثم تقدير معالم التوزيع باستخدام طريقة الامكان الاعظم. قد تم تطبيق بيانات اوقات البقاء في المستشفى لـ 100 شخص اصابوا بالجلطة الدماغية وتم التوصل بأن البيانات التي تم جمعها من مستشفى الزهراء التعليمي في محافظة واسط تتبع توزيع رايلي المعمم المتقطع باستخدام اختبار كولموكروف-سميرنوف (K-S)-Kolmogorov-Smirnov، حيث اظهر الجانب التطبيقي بأن دالة البقاء تتناقص مع ازدياد الزمن وهذا يتطابق مع النظرية الاحصائية لتحليل دوال البقاء. ومن ثم تم تقدير المعالم، الخواص الاحصائية ودالة البقاء للتوزيع.

7. المصادر References

1. Alamatsaz, M.H., Dey,S., Dey,T. & Harandi, S. Shams (2016), "Discrete generalized rayleigh distribution", Pakistan Journal of Statistics, **Vol. 32(1)**, **1-20** Statistics and Management Systems, DOI: 10.1080/09720510.2019.1645400.
2. Alhazzani, N. S. (2012), "Modeling Discrete Life Data in Reliability and its Applications", College of Science King Saud University.
3. A. M. Abouammoh & Najla S. Alhazzani (2015), "On Discrete Gamma Distribution", Communications in Statistics - Theory and Methods, 44:14, 3087-3098, DOI:10.1080/03610926.2013.819924.
4. Collett,D,2003, "Modeling survival data in Medical Research". Champan and Hailondon.

