مقارنة بعض طرائق تقدير دالة المخاطرة لتوزيع (Power Lomax) بإستعمال المحاكاة

الباحث مصطفى عبد الجبار حمدان أ. م. د. عماد حازم عبودي

قسم الإحصاء / كلية الإدارة والاقتصاد / جامعة بغداد

الكلمات المفتاحية: توزيع (POLO) (Power Lomax) , دالة المخاطرة (Hazard Function) , طريقة الإمكان الأعظم (MLE) , طريقة العزوم (MOM) , متوسط مربعات الخطأ التكاملي (IMSE)

المستخلص

بعض الأحيان يواجه الباحثين في مجال الأحصاء مشكلة وجود بيانات لا تتبع التوزيعات الإحتمالية الشائعة لذلك قام الباحثون بإقتراح توزيعات مركبة وأخرى ممتدة من التوزيعات القياسية لتلائم هذه البيانات كما هو الحال مع التوزيع المدروس في هذا البحث وهو توزيع (POLO) (Power Lomax) الذي يُعد أحد إمتدادات توزيع (Lomax) ، وتم تقدير دالة المخاطرة للتوزيع المذكور بطريقة الإمكان الأعظم (MLE) وطريقة العزوم (MOM) والمقارنة بينهما باستعمال معيار المقارنة (IMSE) ، وأظهرت نتائج المحاكاة أفضلية طريقة (MLE) في حجوم العينات المتوسطة والكبيرة بينما كانت الأفضلية لطريقة العزوم في حجوم العينات الصغيرة وإن مجال تطبيق توزيع (POLO) هو في الجانب الصحي والطبي الذي يخص حياة الانسان والامراض التي تصيبه .

Abstract

Sometimes researchers in the field of statistics face the problem of having data that does not follow the standard probability distributions, so the researchers suggested complex and extended distributions from the standard distributions to fit these data, as is the case with the distribution studied in this Paper, which is the(Power Lomax) (POLO) distribution, which is one of the Distribution Extensions (Lomax), The Hazard function of the mentioned distribution was estimated using the maximum Likelihood method (MLE) and the moment method (MOM) and compared them using the comparison criterion (IMSE), The simulation results showed the preference of the (MLE) method in medium and large sample sizes, while The field of application of ,the preference for the moment method was in small sample sizes the distribution of (POLO) is in the health and medical aspect that pertains to human life and the diseases that affect it

.

1- المقدمة Intrudoction

تُعد دراسة و تحليل دوال البقاء والمخاطرة أحد الفروع المهمة في علم الاحصاء لما له أهمية مباشرة بحياة الانسان والأمراض التي تصيبهُ و تمثل دالة البقاء إحتمالية بقاء الانسان على قيد الحياة من بداية المرض حتى حدوث حدث النهاية (الموت أو الشفاء) أما دالة المخاطرة تمثل معدل الفشل الآني للمفردة على قيد الحياة الى وقت البقاء المشاهد (t) للمريض قيد الدراسة خلال المدة $(t+\Delta t,t)$ مع ملاحظة ان المريض كان على قيد الحياة في الزمن (t) ، لذلك اصبح من المهم

تقدير دوال البقاء والمخاطرة لجميع التوزيعات الاحتمالية القياسية والمركبة بمختلف طرائق التقدير، إن المحور الأساس في موضوع دراستنا هو تسليط الضوء على احد امتدادات توزيع (Lomax) وهو توزيع (Power Lomax) (POLO) ذو الثلاث معلمات وهو من التوزيعات الحديثة الممتدة من توزيع (Lomax) والذي اثبت أفضليته على الأمتدادات الأخرى لتوزيع (Lomax) وتقدير دالة المخاطرة (Hazard Function) الخاصة بالتوزيع ومعرفة افضل مقدر لها باستعمال طرائق التقدير التي تم ذكرها سابقاً .

2-هدف البحث The Purpose Of The research

تقدير دالة المخاطرة لتوزيع (POLO) بطريقتين للتقدير والمقارنة بينهما وفق معيار المقارنة الاحصائي متوسط مربعات الخطأ التكاملي (IMSE) للحصول على أفضل مقدر للدالة المذكورة .

3- أهمية البحث The Important Of The research

إن دراسة دوال المخاطرة لتوزيعات مختلفة فضلاً عن توزيع (POLO) يشكل أهمية في مجال توقع احتمالات وفاة الكائنات الحية التي تسلك بياناتها توزيع احصائي معين لذلك فان أهمية البحث تكمن في معرفة افضل مقدر لدالة المخاطرة.

4- الجانب النظري

1-4مفهوم دالة المخاطرة (Concept of Hazard Function)

ويقصد بها إحتمال موت الانسان خلال الفترة (t ، $t+\Delta t$) علماً إن المريض كان حياً خلال الوقت t وهذا يعني ان دالة المخاطرة تمثل الفشل الآني للمفردة على قيد الحياة الى وقت البقاء المشاهد (t)، يرمز لها بالرمز $h(t)$ وتُسمى ايضاً بمعدل الفشل .

و يعبر عنها رياضياً بالشكل الآتي [7:pp.11] :

$$h(t)=\lim_{\Delta t\to 0}\left[\frac{Pr\ (t<T<t+\Delta t\backslash T>t)}{\Delta t}\right] \quad (1)$$

$$h(t)=\lim_{\Delta t\to 0}\left[\frac{\Pr(t<T<t+\Delta t)}{\Delta t.\Pr(T>t)}\right]$$

$$h(t)=\frac{f(t)}{S(t)} \quad (2)$$

2-4توزيع (POLO) (Power Lomax)

تم تقديم توزيع (POLO) (Power Lomax) من قبل راضي الحُسيني وآخرون سنة (2016) وهو أحد إمتدادات توزيع (Lomax) وحقق افضلية على الامتدادات السابقة لتوزيع (Lomax) حسب الدراسات التي أجريت للمقارنة بين توزيع (POLO) والتوزيعات الأخرى ومن خلال النظر في تحويل القوة ادناه لتوزيع (POLO) نلاحظ أن [6:pp.4] :

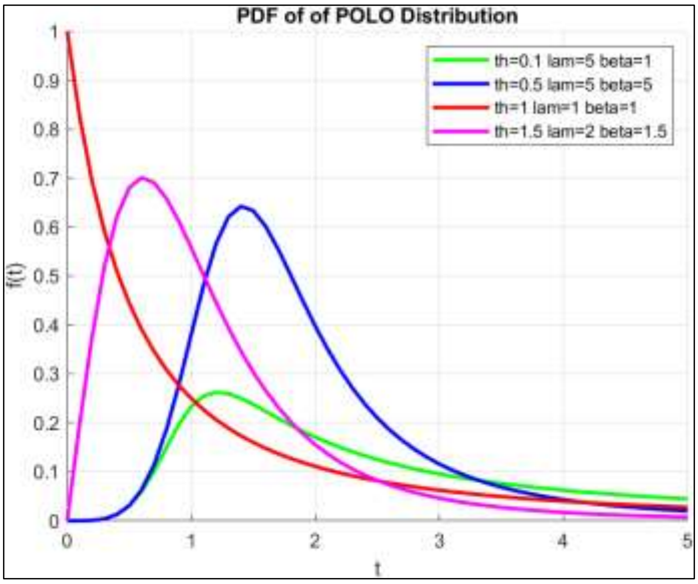
$$T=X^{1/\lambda}$$

إذ إن المتغير العشوائي X يتبع توزيع (Lomax) ذو المعلمتين (θ ، β) ، بينما يتبع المتغير العشوائي (T) توزيع (POLO) (Power Lomax) ذو الثلاث معلمات (θ, λ, β) ، ويشار إلى توزيع (Power Lomax) بالإختصار التالي :

(T ~ POLO (θ, λ, β , وإن مجال تطبيق توزيع (POLO) في الجانب الصحي المرتبط بدوال البقاء.

دالة الكثافة الاحتمالية (PDF) لتوزيع (POLO) تكون كما يلي [6:pp.5]:

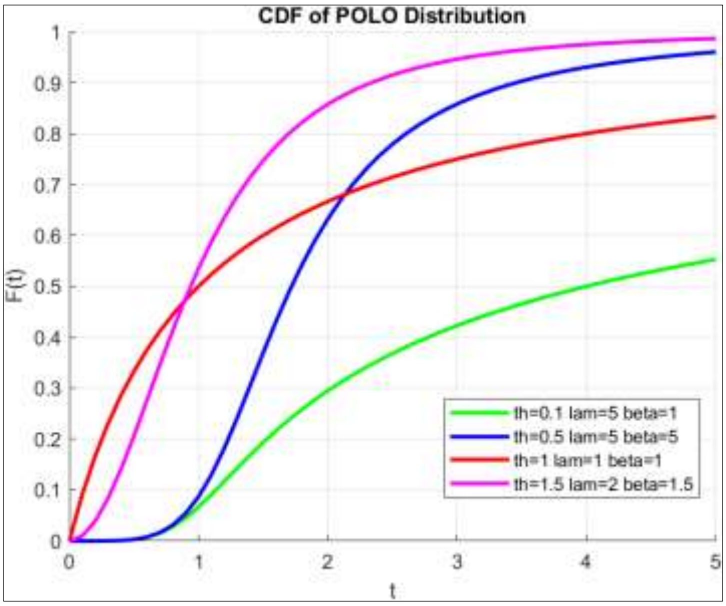
$$f(t)=\theta\lambda\beta^{\theta}t^{\lambda-1}(\beta+t^{\lambda})^{-\theta-1}\ ,\ \ t>0,\qquad \theta,\beta,\lambda>0 \quad (3)$$



شكل (1) دالة الكثافة الإحتمالية لتوزيع (POLO)

وتكون دالة التوزيع التراكمية (CDF) بالشكل التالي [6:pp.5] :

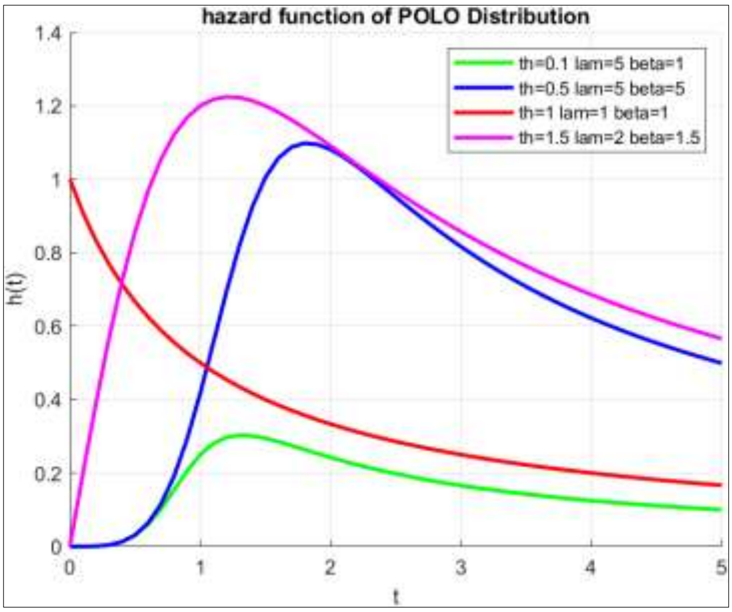
$$F(t) = 1 - \beta^{\theta}(t^{\lambda} + \beta)^{-\theta} \quad , \qquad t > 0 \ , \theta, \beta, \lambda > 0 \qquad (4)$$



شكل (2) الدالة التراكمية لتوزيع (POLO)

وانّ دالة المخاطرة (معدل الفشل) (HRF) لتوزيع (POLO) تكون كما يلي [6:pp.5-6]:

$$h(t) = \frac{t^{\lambda-1}\theta\lambda}{t^{\lambda}+\beta} \quad , \quad t > 0 \ , \theta, \beta, \lambda > 0 \qquad (5)$$



شكل (3) دالة المخاطرة (معدل الفشل) لتوزيع (POLO)

3-4 طرائق التقدير Methods of Estimation

1-3-4 طريقة الإمكان الأعظم (Maximum Likelihood Method) (MLE)

تُعد طريقة الإمكان الأعظم من الطرق المهمة في التقدير وذلك بسبب امتلاكها بعض الخصائص الجيدة مثل الكفاية والاتساق وعدم التحيز خصوصاً عندما يكون حجم العينة كبيراً، وهي اكثر دقة من طرائق التقدير الأخرى لا سيما عند زيادة حجم العينة، ويُعرف التقدير بإستعمال هذه الطريقة بانه قيم المعلمات التي تجعل دالة الإمكان في نهايتها العظمى [5:pp.41] [8:pp.308]:

فإذا كانت t_n، ....، t_2، t_1 عينة عشوائية بحجم (n) من توزيع (POLO) بدالة كثافة إحتمالية :

$$f(t) = \theta\lambda\beta^{\theta}x^{\lambda-1}(\beta+x^{\lambda})^{-\theta-1} \quad , \quad x > 0, \qquad \theta, \beta, \lambda > 0 \qquad (6)$$

فإن دالة الإمكان الأعظم (L) لتوزيع (POLO) تكون كما يلي :

$$L(t;\theta,\lambda,\beta) = \prod_{i=1}^{n} \theta\lambda\beta^{\theta}t^{\lambda-1}(\beta+t^{\lambda})^{-\theta-1} \qquad (7)$$

$$L(t;\theta,\lambda,\beta) = (\theta\lambda)^{n}(\beta^{\theta})^{n}\prod_{i=1}^{n} {t_i}^{\lambda-1}\prod_{i=1}^{n}(\beta+{t_i}^{\lambda})^{-\theta-1} \qquad (8)$$

نأخذ (ln) لطرفي المعادلة ونحصل على :

$$LnL(\theta,\lambda,\beta) = n\ln\theta + n\ln\lambda + n\theta\ln\beta + (\lambda-1)\sum_{i=1}^{n}\ln t_i - (\theta+1)\sum_{i=1}^{n}\ln\left(\beta + t_i^{\lambda}\right) \quad .... \quad (9)$$

نقوم باشتقاق المعادلة (9) جزئياً بالنسبة للمعلمات θ,λ,β على التوالي ونساويها للصفر و نحصل على:

$$\hat{\theta} = \frac{n}{\sum_{i=1}^{n}\ln\left(\hat{\beta} + t_i^{\hat{\lambda}}\right) - n\ln\hat{\beta}} \quad (10)$$

$$\hat{\lambda} = \frac{n}{(\hat{\theta}+1)\sum_{i=1}^{n}\frac{t_i^{\hat{\lambda}}\ln t_i}{\left(\hat{\beta}+t_i^{\hat{\lambda}}\right)} - \sum_{i=1}^{n}\ln t_i} \quad (11)$$

$$\hat{\beta} = \frac{n\hat{\theta}}{(\hat{\theta}+1)\sum_{i=1}^{n}\frac{1}{\left(\hat{\beta}+t_i^{\hat{\lambda}}\right)}} \quad (12)$$

نلاحظ أنَ المعادلات (10) , (11) , (12) هي معادلات لا خطية ولا يمكن حلها بالطرق الرياضية الاعتيادية لذلك نقوم بحلها باستعمال الخوارزمية التكرارية **(Newton Raphson) .**

وإن مقدر دالة المخاطرة لتوزيع **(POLO)** بطريقة الإمكان الأعظم يكون بالشكل التالي :

$$\hat{h}(t) = \frac{t^{\hat{\lambda}-1}\hat{\theta}\hat{\lambda}}{t^{\hat{\lambda}} + \hat{\beta}} \quad (13)$$

2-3-4 طريقة العزوم Moment Method (MOM)

طريقة التقدير بالعزوم من الطرائق واسعة الاستعمال لتقديرمعلمات النماذج الإحصائية وتعتمد على مبدأ مساواة عزم المجتمع بعزم العينة ونحصل على عدد من المعادلات التي تكون مساوية لعدد المعلمات المراد تقديرها وبحل هذه المعادلات نحصل على مقدرات المعلمات **[2:pp.4][4:pp.142]** :

لنفترض أن t_1، t_2 ،.... ، t_n عينة عشوائية بحجم n من توزيع (POLO) بدالة كثافة احتمالية :

$$f(t) = \theta\lambda\beta^{\theta}t^{\lambda-1}(\beta + t^{\lambda})^{-\theta-1} \quad , \quad t > 0, \qquad \theta,\beta,\lambda > 0$$

إذ أن العزم الأول للمجتمع هو [ ]:

$$\mu_1' = E(t^1) \quad (14)$$

$$\mu_1' = \frac{\theta\beta^{\frac{1}{\lambda}}\Gamma\left[\theta - \frac{1}{\lambda}\right]\Gamma\left[\frac{1}{\lambda}\right]}{\lambda\Gamma[1+\theta]} \quad (15)$$

عزم العينة الأول :

$$m_1 = \frac{\sum_{i=1}^{n} t_i}{n} \quad (16)$$

نقوم بمساواة عزم المجتمع مع عزم العينة لنحصل على مقدر المعلمة θ وكما يلي :

$$\mu_1' = m_1 \qquad (17)$$

$$\frac{\hat{\theta}\hat{\beta}^{\frac{1}{\hat{\lambda}}}\Gamma\left[\theta - \frac{1}{\hat{\lambda}}\right]\Gamma\left[\frac{1}{\hat{\lambda}}\right]}{\hat{\lambda}\Gamma[1+\hat{\theta}]} = \frac{\sum_{i=1}^{n} t_i}{n} \qquad (18)$$

$$n\hat{\theta}\hat{\beta}^{\frac{1}{\hat{\lambda}}}\Gamma\left[\theta - \frac{1}{\hat{\lambda}}\right]\Gamma\left[\frac{1}{\hat{\lambda}}\right] = \hat{\lambda}\Gamma[1+\hat{\theta}]\sum_{i=1}^{n} t_i \quad . \; . \; . \; . \; . \quad (19)$$

وبنفس الأسلوب نجد معادلة العزم الثاني والعزم الثالث للمعلمات (λ، β) ونحصل على :

معادلة العزم الثاني بالنسبة للمعلمة λ :

$$n\hat{\theta}\hat{\beta}^{2/\hat{\lambda}}\Gamma\left[\hat{\theta} - \frac{2}{\hat{\lambda}}\right]\Gamma\left[\frac{2+\hat{\lambda}}{\hat{\lambda}}\right] = \Gamma[1+\hat{\theta}]\sum_{i=1}^{n} t_i^2 \quad . \; . \; . \; . \quad (20)$$

معادلة العزم الثالث بالنسبة للمعلمة β :

$$n\hat{\theta}\hat{\beta}^{3/\hat{\lambda}}\,\Gamma\left[\hat{\theta} - \frac{3}{\hat{\lambda}}\right]\Gamma\left[\frac{3+\hat{\lambda}}{\hat{\lambda}}\right] = \Gamma[1+\hat{\theta}]\sum_{i=1}^{n} t_i^3 \quad . \; . \; . \; . \quad (21)$$

نلاحظ أنَ المعادلات (19), (20), (21) هي معادلات لا خطية ولا يمكن حلها بالطرق الرياضية الاعتيادية لذلك نقوم بحلها عن طريق الخوارزمية التكرارية **(Newton Raphson) .**

ونحصل على مقدر تقريبي لدالة المخاطرة وذلك من خلال تعويض مقدرات معلمات توزيع (POLO) بطريقة العزوم في صيغة دالة المخاطرة وكما يلي :

$$\hat{h}(t) = \frac{t^{\hat{\lambda}_{mom}-1}\,\hat{\theta}_{mom}\hat{\lambda}_{mom}}{t^{\hat{\lambda}_{mom}} + \hat{\beta}_{mom}} \qquad (22)$$

5- تجارب المحاكاة [3:pp.29] Simulation experiments

5-1 مراحل تجربة المحاكاة Simulation Stages

المرحلة الأولى:

a. إختيار قيم المعلمات الإفتراضية لتوزيع (POLO) بالاعتماد على البيانات الحقيقية وكما مبين في الجدول ادناه:

التجربة	θ	λ	β
1	0.4	3	65
2	0.6	5	65
3	0.4	3	70
4	0.6	5	70
5	0.4	3	75
6	0.6	5	75

جدول (2) قيم المعلمات الافتراضية لتجربة المحاكاة

.b إختيار حجوم عينات مفترضة وكما يلي :

، وكان تكرار كُل تجربة مساوي الى (1000) . n = (10 , 25 , 50 , 75 , 100)

المرحلة الثانية :
وتتضمن هذه المرحلة توليد متغير عشوائي يتبع التوزيع المنتظم $\mathbf{u \sim U(0,1)}$ وتوليد بيانات تتبع توزيع (POLO) بإستعمال طريقة معكوس الدالة التراكمية (CDF) وحسب الصيغة التالية :

$$F(t) = 1 - \beta^{\theta}(t^{\lambda} + \beta)^{-\theta} \quad (23)$$

$Let \quad F(t) = u$

$$u = 1 - \beta^{\theta}(t^{\lambda} + \beta)^{-\theta} \quad (24)$$

$$t = \beta^{\frac{1}{\lambda}}\left[(1-u)^{-\frac{1}{\theta}} - 1\right]^{\frac{1}{\lambda}} \quad (25)$$

المرحلة الثالثة :
مرحلة التقدير وتعتبر أهم مراحل تجربة المحاكـــاة و تم تقدير دالة المخاطرة بطرائق التقديـــر (MLE, MOM) و وفق الحالات الإفتراضية لقيم المعلمات وحجوم العينات التي تم إفتراضها في المرحلة الأولى من تجربة المحاكاة .

المرحلة الرابعة :
في هذه المرحلة تم إجراء المقارنة بين طرائق التقدير المستعملة وحسب معيار المقارنة الإحصائي متوسط مربعات الخطأ التكاملي (IMSE) الذي يمثل تكامل المساحة الكُلية ل (t_i) ويعبر عنه بالصيغة التالية [1:pp.9] :

$$\text{IMSE}\left[\hat{h}(t)\right] = \frac{1}{r}\sum_{i=1}^{r}\left[\frac{1}{n_t}\sum_{j=1}^{n_t}\left[\hat{h}_i(t_j) - h(t_j)\right]^2\right] \quad (24)$$

المرحلة الخامسة :
تتضمن هذه المرحلة عرض نتائج تجارب المحاكاة لإيجاد أفضل مقدر لدالة المخاطرة لتوزيع (POLO) عند قيم المعلمات الافتراضية ولمختلف حجوم العينات وكما مبين في الجدول ادناه :

$\theta = 0.4, \quad \lambda = 3, \quad \beta = 65$			
N	**Method**		**Best**
	IMSE $(\hat{h})_{MLE}$	**IMSE $(\hat{h})_{MOM}$**	
10	0.005688	0.228531	MLE
25	0.075454	0.158941	MLE
50	0.003379	0.160639	MLE

75	0.000785	0.167149	MLE
100	0.055450	0.190949	MLE
$\theta = 0.6$, $\lambda = 5$, $\beta = 65$			
10	0.025202	0.14599	MLE
25	0.003113	0.160933	MLE
50	0.000940	0.144329	MLE
75	0.001089	0.168955	MLE
100	00.00079	0.149506	MLE
$\theta = 0.4$, $\lambda = 3$, $\beta = 70$			
10	0.016559	0.254168	MLE
25	0.002038	0.241067	MLE
50	0.001821	0.183945	MLE
75	0.002549	0.182132	MLE
100	0.001579	0.220936	MLE
$\theta = 0.6$, $\lambda = 5$, $\beta = 70$			
10	0.025368	0.024615	MOM
25	0.005872	0.206708	MLE
50	0.004381	0.003738	MOM
75	0.002437	0.128652	MLE
100	0.000807	0.145399	MLE
$\theta = 0.4$, $\lambda = 3$, $\beta = 75$			
10	0.002103	0.235757	MLE
25	0.006615	0.236872	MLE
50	0.224795	0.155175	MOM
75	0.000565	0.210246	MLE
100	0.000396	0.203318	MLE
$\theta = 0.6$, $\lambda = 5$, $\beta = 75$			

10	0.204584	0.000662	MOM
25	0.002869	0.140559	MLE
50	0.005559	0.156032	MLE
75	7.69E-05	0.169999	MLE
100	0.000175	0.168102	MLE

جدول (1) قيم متوسط مربعات الخطأ التكــاملي (IMSE) لدالة المخاطرة المقدرة لتوزيع (POLO) وفق طريقتي (MOM ،MLE)

6- الإستنتاجات :

من خلال نتائج تجربة المحاكاة للنماذج الستة تبين ما يلي :

1) عند حجم العينة (10) أظهرت طريقة العزوم (MOM) أفضليتها على طريقة الإمكان الأعظم (MLE) لتسجيلها أقل متوسط لمربعات الخطأ التكاملي (IMSE) بلغ (0.000662) .
2) عند حجوم العينات (25,50,75,100) أظهرت طريقة الإمكان الأعظم (MLE) افضليتها لتسجيلها أقل متوسط لمربعات الخطأ التكاملي (IMSE) .

7- التوصيات :

3) إعتماد طريقة الإمكان الأعظم (MLE) في تقدير دالة المخاطرة لبيانات تتبع توزيع (POLO) عند حجوم العينات المتوسطة والكبيرة .
4) إعتماد طريقة العزوم (MOM) في تقدير دالة المخاطرة لبيانات تتبع توزيع (POLO) عند حجوم العينات الصغيرة .

المصادر :

1. بدر , دريد حسين , الحكيم , معاني أحمد .(2019)" مقارنة بعض طرائق تقدير دالة المعولية لتوزيع وايبل ذي المعلمتين باستخدام المحاكاة", بحث منشور في مجلة كلية الرافدين الجامعة للعلوم العدد 45.
2. محمود , سحر طارق ,(2014), " أستخدام المحاكاة في تقدير معلمة القياس لتوزيع كاما المستمر" , بحث منشور في مجلة الإدارة والاقتصاد ,المجلد 3 , العدد(12) .
3. مناع, احمد سعدون , (2019), " تقدير دالة البقاء باستعمال بيز الحصين في حالة بيانات سابقة متناقضة مع تطبيق عملي ", رسالة ماجستير, كلية الادارة والاقتصاد, جامعة بغداد.
4. الناصر, عبد المجيد حمزة , حسين , ظافر رشيد ,(1988) "الاستدلال الاحصائي " , مطبعة التعليم العالي , بغداد.
5. نعيمة, علي بندر, (2016), " تقدير دوال الفشل للتوزيع الناتج من دمج توزيع بواسون ليندلي مع توزيعات اخرى", أطروحة دكتوراه , كلية الادارة والاقتصاد, جامعة بغداد.
6. El Houssainy. Rady. A, Hassanein W. A., and Elhaddad T. A. (2016) , “The power Lomax distribution with an application to bladder cancer data,” *Springerplus*, vol. 5, no. 1, doi: 10.1186/s40064-016-3464-y.

7. Lee, E. T., and Wang, J. W. , (2003) , " Statistical Methods for Survival Data Analysis", 3rd ed. Wiley, New York .
8. Ronald E. W,Raymond H.M,Sharon L.M & Keying ye(2007)."Probability & Statistics for Engineers & Scientists".