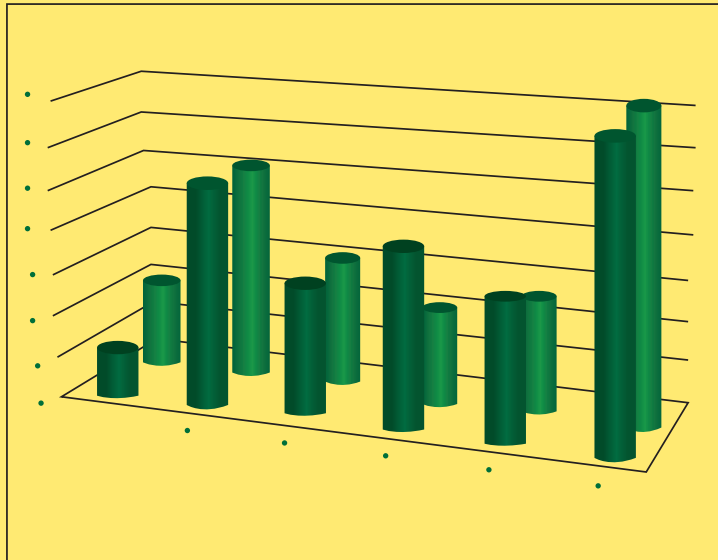


# المعهد العربي للتدريب والبحوث الإحصائية



## مجلة العلوم الإحصائية



العدد رقم 20

مجلة علمية محكمة

يصدرها المعهد العربي للتدريب والبحوث الإحصائية

معتمدة في قائمة المجلات العلمية Ulrich's

[www.ulrichsweb.com](http://www.ulrichsweb.com)

مصنفة في معامل التأثير والاستشادات المرجعية العربي (أرسيف)

[www.emarefa.net/arcif/](http://www.emarefa.net/arcif/)

ISSN 2522-64X (Online), ISSN 2519-948X (Print)

عدد خاص  
أبحاث المحور الإحصائي للمؤتمر الدولي الخامس والسنوي السادس عشر  
لكلية الإدارة والاقتصاد جامعة المستنصرية  
الذي أقيم تحت شعار "نحو عراق رقمي مستدام"

# مجلة العلوم الإحصائية

مجلة علمية محكمة

هيئة التحرير

رئيس هيئة التحرير

الأستاذ الهادي السعيد

أمين التحرير

الدكتور لحسن عبد الله باشيوه

أعضاء هيئة التحرير

- أ. د. فيصل الشرعي  
أ. د. عيسى مصاروه  
أ. م. د. حسان أبو حسان  
أ. د. عبد الخالق التهامي  
أ. م. د. احمد شاكر المتولي  
أ. م. د. حميد بوزيدة  
أ. د. مختار الكوي  
أ. م. د. سلوى محمود عسار

أعضاء الهيئة الاستشارية

- د. قاسم الزعبي  
د. ضياء عواد  
د. لؤي شبانه  
د. نبيل شمس  
د. خليفة البرواني  
د. غازي رحو  
د. علا عوض  
أ. د. عوض حاج علي  
أ. د. ميثم العيبي اسماعيل  
أ. م. د. محمد حسين علي الجنابي

معتمدة في قائمة المجلات العلمية Ulrich's

[www.ulrichsweb.com](http://www.ulrichsweb.com)

مصنفة في معامل التأثير والاستشهادات المرجعية العربي (أرسيف)

[www.emarefa.net/arcif/](http://www.emarefa.net/arcif/)

ISSN 2522-64X (Online), ISSN 2519-948X (Print)

## شروط النشر في مجلة العلوم الإحصائية

- 1 – تنشر-المجلة البحوث والدراسات العلمية في المجالات الإحصائية والمعلوماتية المكتوبة باللغة العربية والانكليزية والفرنسية على أن لا يكون البحث المقدم للنشر. قد نشر. او قدم للنشر في مجلات او دوريات أخرى او قدم ونشر في دوريات لمؤتمرات أو ندوات.
- 2 – ترسل البحوث والدراسات الى أمين التحرير على أن تتضمن اسم الباحث او الباحثين وألقابهم العلمية وأماكن عملهم مع ذكر عنوان المراسلة وأرقام الهواتف والبريد الالكتروني. وان يرسل البحث المراد نشره الكترونياً (على قرص او بالبريد الالكتروني) وفق المواصفات أدناه:
  - أ – أن يكون مطبوعاً على ورق حجم A4 وان يكون على شكل عمود واحد ويستخدم للغة العربية نوع حرف (Simplified Arabic) و(Times New Roman) للإنجليزية والفرنسية وبحجم خط (12). وباستخدام Microsoft Word وعلى وجه واحد للورقة.
  - ب – الهامش مسافة 2.5 سم لجميع جوانب الورقة.
  - ج – يرفق الباحث ملخصاً عن بحثه باللغتين العربية والانجليزية والفرنسية بما لا يزيد عن صفحة واحدة.
  - د - يتم الإشارة الى المصادر العلمية في متن البحث وفي نهايته، مع مراعاة أن لا يتضمن البحث سوى المصادر التي تم الإشارة إليها في المتن ووفق الأصول المعتمدة في ذلك (اسم المؤلف، سنة النشر، عنوان المصدر، دار النشر، البلد).
  - هـ- ترقم الجداول والرسوم التوضيحية وغيرها حسب ورودها في البحث، كما توثق المستعارة منها بالمصادر الأصلية.
  - و- أن لا يزيد عدد صفحات البحث او الدراسة عن (25) صفحة.
- 3 – يتم إشعار الباحث باستلام بحثه خلال مدة لا تتجاوز يومين عمل من تاريخ استلام البحث.
- 4 – تخضع كافة البحوث المرسلة الى المجلة للتقييم العلمي الموضوعي ويبلغ الباحث بالتقييم والتعديلات المقترحة إن وجدت خلال مدة لا تتجاوز اسبوعان من تاريخ استلام البحث.
- 5 – لهيئة تحرير المجلة الحق في قبول او رفض البحث ولها الحق في إجراء أي تعديل او إعادة صياغة جزئية للمواد المقدمة للنشر- بما يتماشى والنسق المعتمد في النشر- لديها بعد موافقة الباحث.
- 6 – يصبح البحث المنشور ملكاً للمجلة ولا يجوز إعادة نشره في أماكن أخرى.
- 7 – تعبر المواد المنشورة بالمجلة عن آراء أصحابها، ولا تعكس وجهة نظر المجلة او المعهد العربي للتدريب والبحوث الإحصائية.
- 8 – ترسل البحوث على العنوان الالكتروني للمجلة:

[journal@aitrs.org](mailto:journal@aitrs.org) / [Info@aitrs.org](mailto:Info@aitrs.org)

## المحتويات

| رقم الصفحة | اسم البحث والباحث  | ت |
|------------|--|---|
| 1          | استخدام توزيع رايلي المعمم المتقطع في دالة البقاء مع التطبيق<br>الباحثة: نبأ صالح هادي / الجامعة المستنصرية / كلية الإدارة والاقتصاد<br>أ.د. جواد كاظم خضير الموسوي / الجامعة المستنصرية / كلية الإدارة والاقتصاد  | 1 |
| 11         | استعمال بعض الدوال اللبية المقترحة في تقدير أنموذج انحدار الخطأ المكاني شبه<br>المعلمي (SPSAR) بطريقة المقدر الخطي الموضوعي<br>أ.م. د. احمد عبد علي عكار كلية الإدارة والاقتصاد / جامعة المستنصرية<br>أ.د. سجي محمد حسين كلية الإدارة والاقتصاد / جامعة بغداد  | 2 |
| 29         | استعمال بعض طرائق التقدير الحصينة في تقدير العلاقة بين الكميات المصدرة<br>والمنتجة من النفط الخام العراقي)<br>الباحث رضا قاسم محمد تميم / الجامعة المستنصرية/ كلية الادارة والاقتصاد<br>أ.م.د. احمد شاكر محمد طاهر/ الجامعة المستنصرية/ كلية الادارة والاقتصاد   | 3 |
| 43         | تصنيف مرضى التهاب الكبد باستعمال الدالة التمييزية<br>مآب نوري حسن جامعة بغداد كلية الإدارة والاقتصاد قسم الأحصاء<br>ابتسام كريم عبد الله جامعة بغداد كلية الإدارة والاقتصاد قسم الأحصاء  | 4 |
| 59         | تقدير انموذج الانحدار الذاتي المكاني ( SAR ) باستعمال طريقة بيز لبيان اثر<br>التجاورات المكانية على مرض السرطان في العراق<br>الباحثة ساره اسامة سعد / الجامعة المستنصرية /كلية الادارة والاقتصاد<br>أ.م.د. هيفاء طه عبد / بالجامعة المستنصرية / كلية الادارة والاقتصاد   | 5 |
| 73         | دراسة تحليلية لمعدلات البطالة في العراق باستخدام تقنية البوتستراب<br>م. فراس منذر جاسم / م.د. زينب فالح حمزة / م. نائرة نجم عبد الله   | 6 |
| 85         | Comparison of Estimating the Survival Function for the Two-<br>Parameter Lindley Distribution by the Bayesian and Maximum-<br>Likelihood Methods<br>Lecturer. Basheer.j.Khaleel/ Ministry of Education, Baghdad. Iraq<br>Prof. Dr. Wadhah S. Ibrahim / Mustansiriyah University / College of<br>Management and Economics Baghdad, Iraq       | 7 |
| 101        | Using gray wave models to predict crude oil production in Iraq<br>Ali Mohammed Ali Assistant lecturer/ Economics department/ College of<br>Administration & Economics/ Mustansiriyah University<br>Dr. Ahmed Shaker Tahir/ Assistant Professor/ Statistics department /<br>College of Administration & Economics<br>Mustansiriyah University | 8 |



## 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 الأسّي 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 الآتية :

1. تقيس معدل احتمال الفشل الآتية حيث ان دالة البقاء مقياس تراكمي مع الوقت.
2. تستخدم للتعرف على شكل الانموذج الرياضي لبيانات البقاء مثل الانموذج الأسّي (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<sup>(1)</sup> 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  $\alpha, \lambda$  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 ل 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

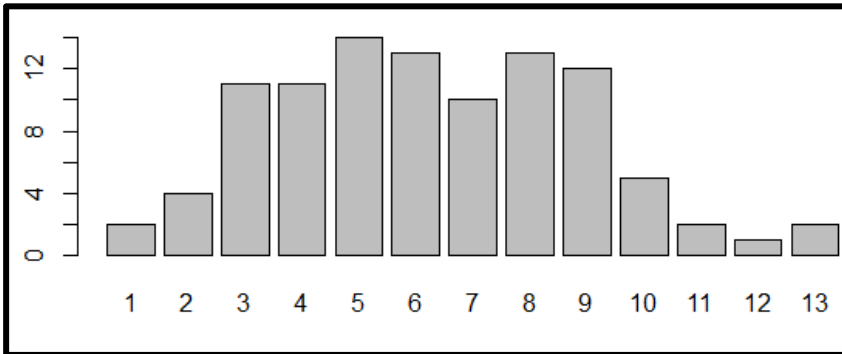
|   |   |   |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |
|---|---|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|
| 1 | 1 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 | 3  | 3  | 3  | 3  | 3  | 3  | 3  | 4  | 4  | 4  |
| 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 5 | 5 | 5 | 5  | 5  | 5  | 5  | 5  | 5  | 5  | 5  | 5  | 5  |
| 5 | 5 | 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  | 9  | 9  |
| 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 10 | 10 | 10 | 10 | 10 | 10 | 11 | 11 | 12 | 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:



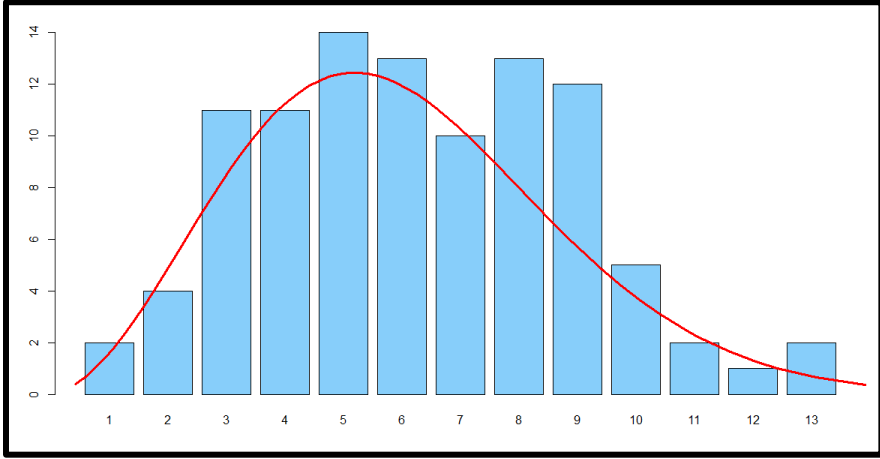
شكل (1): الشريط البياني لبيانات اوقات البقاء للأشخاص المصابين بالجلطة الدماغية

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

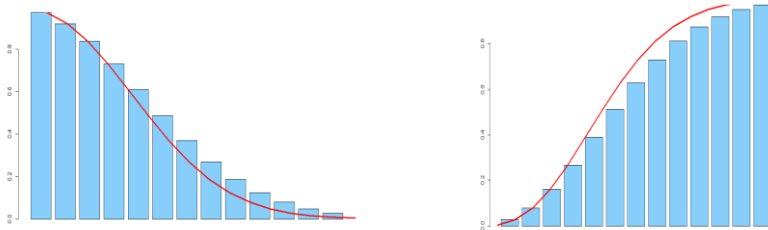
| القيمة الحرجة | إحصاءة K-S | تقدير معلمة $q$ | تقدير معلمة $\alpha$ |
|---------------|------------|-----------------|----------------------|
| 0.136         | 0.0876     | 0.9804          | 1.4011               |

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  $\alpha$  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.



استعمال بعض الدوال اللبية المقترحة في تقدير أنموذج انحدار الخطأ المكاني  
شبه ألمعلمي (SPSAR) بطريقة المقدر الخطي الموضوعي

أ. د. سجي محمد حسين  
كلية الإدارة والاقتصاد / جامعة بغداد

أ. م. د. احمد عبد علي عكار  
كلية الإدارة والاقتصاد / جامعة المستنصرية

تاريخ استلام البحث: 2022/07/22

تاريخ قبول البحوث: 2022/08/10

نشر البحث في العدد عشرون: ايلول / سبتمبر 2023

رمز التصنيف ديوي / النسخة الالكترونية (Online): 2522-64X/512.9

رمز التصنيف ديوي / النسخة الورقية (Print): 2519-948X/512.9



## استعمال بعض الدوال اللبية المقترحة في تقدير أنموذج انحدار الخطأ المكاني شبه المعلمي (SPSAR) بطريقة المقدر الخطي الموضوعي

أ.م. د. احمد عبد علي عكار      أ.د. سجي محمد حسين  
كلية الإدارة والاقتصاد / جامعة المستنصرية      كلية الإدارة والاقتصاد / جامعة بغداد

### المستخلص

في هذا البحث تم استعمال طريقة المقدر الخطي الموضوعي (LLE) اللامعلمية وطريقة الدالة اللبية المقترحة الأولى (SUG1) والثانية (SUG2) اللامعلمية للبيانات المعتمدة مكانيا والتي من ضمنها إيجاد معلمة التمهيد (h) وفقا لمعيار العبور الشرعي (CV) لتقدير أنموذج انحدار التأخر المكاني شبه المعلمي (SPSAR)، كذلك تم استعمال طريقة الإمكان الأعظم لتقدير معلمة الاعتماد المكاني ( $\rho$ ) في انموذج (SPSAR).

ومن خلال إجراء تجربة المحاكاة لثلاث حجوم عينات وثلاث مستويات مختلفة للتباين ولثلاث دوال ولثلاث قيم لمعلمة الاعتماد المكاني ( $\rho$ ) وبتكرار 1000 مرة تم استعمال طرائق التقدير أعلاه لأنموذج (SPSAR) مستعملا مصفوفة التجاورات المكانية المعدلة في ظل معيار تجاور Queen وبمقارنة هذه الطرائق بمعيار متوسط الخطأ النسبي المطلق (MAPE) تبين أن أفضل طريقة في تقدير أنموذج SPSAR فهي طريقة المقدر الخطي الموضوعي (LLE) للدالة المقترحة الأولى (SUG1).

الكلمات الافتتاحية/ أنموذج (SPSAR) - المقدر الخطي الموضوعي - مصفوفة التجاورات المكانية - معيار تجاور Queen.

### 1- المقدمة

ظهر في السنوات الأخيرة اهتمام كبير من قبل الباحثين لموضوع تحليل البيانات المعتمدة مكانيا والخاصة في نماذج القياس الاقتصادي المكاني التي تعد إحدى أنواع النماذج القياسية، إذ إن هذه النماذج تهتم بدراسة التأثيرات المكانية بين مشاهدات الوحدات للظواهر المدروسة، كالظواهر الاقتصادية والزراعية والتجارية والخدمية والصحية وغيرها من الظواهر الأخرى وهذا ما يميز هذه النماذج عن النماذج القياسية التقليدية التي أهملت وبشكل كبير الاعتماد المكاني بين مشاهدات بيانات العينة في النقاط المختلفة من المكان وان هذه النماذج تهتم على الاعتمادية ما بين المشاهدات عند فترة زمنية معينة لذا فإن التأثيرات المكانية هي السبب الأساسي في وجود القياس الاقتصادي المكاني (Spatial Econometrics)، إذ أن الفرق هو الاعتماد المكاني للمشاهدات، وفي حالة عدم الأخذ بنظر الاعتبار الاعتماد المكاني فان ذلك سوف يخالف فرضيات القياس الاقتصادي التقليدي الذي يؤدي إلى مقدرات متحيزة وغير متسقة أي غير كفؤة، لذا فان التأثيرات المكانية لها تأثير واضح على أنموذج القياس الاقتصادي

وهو ما يسمى بالقياس الاقتصادي المكاني الذي يمثل الطريقة العلمية لتحليل السلاسل المكانية الاقتصادية.<sup>[10]</sup> تم في هذا البحث استعمال أنموذج انحدار التأخر المكاني شبه المعلمي (SPSAR) الذي تم تقديره من خلال استعمال بعض طرائق التقدير شبه المعلمية، وان هذا الأنموذج يصف العلاقة بين متغير الاستجابة (المعتمد) والمتغيرات التوضيحية في ظل معيار التجاور المكاني Queen لمصفوفة التجاورات المكانية.<sup>[7]</sup>

## 2- هدف البحث

إن الهدف من هذا البحث هو تقدير أنموذج انحدار التأخر المكاني شبه المعلمي Semi-Parametric Spatial Auto Regressive (SPSAR) والذي يعاني من مشكلة الاعتماد المكاني باستعمال عدة طرائق للتقدير، والمقارنة بين طرائق التقدير بالاعتماد على أسلوب المحاكاة من خلال معيار (MAPE).

## 3- أنموذج الدراسة

ان دراسة أنموذج الانحدار الذاتي المكاني شبه المعلمي هو لتوضيح تأثير المتغيرات التوضيحية على المتغير التابع في ظل وجود التأثيرات المكانية للمواقع المتجاورة لكل موقع من مواقع الدراسة، إذ يمكن دمج الاعتماد المكاني في أنموذج الانحدار الخطي العام من خلال تضمين متغير تابع متأخر مكانياً كأحد المتغيرات التوضيحية (WY) ويسمى هذا الأنموذج بأنموذج التأخر المكاني (Spatial Lag Model)، وهو مشابه بأنموذج السلاسل الزمنية فيما عدا أن ال (Lag) سيكون في مؤثر المكان وليس الزمان.<sup>[1,2]</sup> فإذا كان لدينا أنموذج الانحدار الذاتي المكاني شبه المعلمي (SPSAR) العام والمبين كما في الصيغة الآتية:<sup>[1]</sup>

$$Y = \rho WY + m(X) + u \quad \dots (1)$$

$$u = \lambda Wu + \varepsilon$$

$$\varepsilon \sim N(0, \sigma^2 I_n)$$

حيث أن: Y : يمثل متجه المتغير المعتمد ذات بعد  $n \times 1$ ،  $\rho$ : يمثل معلمة الاعتماد المكاني،  $\lambda$ : تمثل معلمة الخطأ المكاني، W: تمثل مصفوفة التجاور المكاني وهي ثابتة ومحددة مسبقاً ذات بعد  $n \times n$ ، X: تمثل مصفوفة  $n \times k$  من المتغيرات التفسيرية،  $m(X)$ : هي تمثل دالة التمهيد الغير معروفة، u : تمثل الاخطاء المرتبطة مكانياً،  $\varepsilon$ : يمثل حد الخطأ وهو متجه عشوائي ذات بعد  $n \times 1$  والذي يتوزع توزيع طبيعي بمتوسط صفر وتباين  $\sigma^2 I_n$ .

ومن خلال الأنموذج في الصيغة (1) تم اشتقاق أنموذج انحدار التأخر المكاني شبه المعلمي (SPSAR) والذي يعتبر أنموذج خاص بفرض قيد معين.<sup>[2,3]</sup>، أي عندما يكون القيد  $\lambda = 0$  يصبح الأنموذج في (1) أنموذج انحدار التأخر المكاني شبه المعلمي وصيغته كما يلي:<sup>[7]</sup>

$$Y = \rho WY + m(X) + \varepsilon \quad \dots (2)$$

$$\varepsilon \sim N(0, \sigma^2 I_n)$$

وأن حد الخطأ يتوزع توزيع طبيعي بمتوسط صفر وتباين  $\sigma^2 I_n$   
 $I$  : تمثل مصفوفة الوحدة  $(n \times n)$ .

$\rho$  : تمثل معلمة الاعتماد المكاني وقيمتها تكون بين  $-1 < \rho < 1$  وفي حالة إذا كانت قيمة  $\rho$  صفر فإن هذا يدل على انه لا يوجد اعتماد مكاني وهنا يدل على أن  $\rho$  تقيس قوة الاعتماد المكاني في عينة المشاهدات.<sup>[10]</sup>  $W$ : مصفوفة الأوزان المكانية.

#### 4- الاعتماد المكاني Spatial Dependence

عند تحليل البيانات المقطعية غالباً ما يواجه الباحثون مشكلة الاعتماد المكاني عبر وحدات المشاهدات المكانية المنظمة<sup>[7]</sup>، ويمكن إن يعرف الاعتماد المكاني بان مشاهدات بيانات العينة عند الموقع  $(i)$  تعتمد على المشاهدات الأخرى في الموقع  $(j)$  عندما  $i \neq j$  والصيغة تبين ذلك:-

$$Y_i = f(Y_j) \quad , i = 1, \dots, n, i \neq j \quad \dots (3)$$

من الممكن أن يكون الاعتماد على عدة مشاهدات وان يأخذ المؤشر  $(i)$  أي قيمة من  $1, \dots, n$  وان بيانات العينة التي شوهدت في نقطة واحدة في الفضاء هي تعتمد على القيم التي شوهدت في مواقع أخرى<sup>[8]</sup>.

حيث أن قوة الاعتماد المكاني بين مشاهدات الوحدات المكانية تتناقص كلما كانت طول المسافة بعيدة. وبصورة عامة أن إحدى المشاكل الرئيسية في هذه الحالات ينبع من وجود الاعتماد المكاني، والاعتماد المكاني هو وجود علاقة دالية بين ما يحدث عند نقطة واحدة في الفضاء وما يحدث في أماكن أخرى<sup>[3]</sup>، وهناك سبب شائع يؤدي إلى ذلك وهو نتيجة أخطاء القياس للمشاهدات المرتبطة بالوحدات المكانية المتجاورة والذي يؤدي إلى الاعتماد المكاني، وهذا يحدث عندما الحدود الإدارية للمقاطعات والولايات ومساحات التعداد وما إلى ذلك لا تعكس دقة الطبيعة الأساسية التي تولد بيانات العينة. وكمثال نعتبر حالة معدلات البطالة ومقاييس قوة العمل، لان العاملون متنقلون ويمكنهم عبور خطوط المقاطعات أو الولايات للعثور على عمل في المناطق المجاورة أو معدلات البطالة التي تقاس على أساس المكان الذي يعيش فيه الناس ويمكن أن يظهر فيه الاعتماد المكاني.<sup>[9]</sup>

وأن معظم الأمثلة التجريبية في العلوم الإقليمية التطبيقية يتم الحصول على بيانات للمشاهدات لها والتي يتم ترتيبها في الفضاء أو في المكان والزمان. وفي هذه الحالة يمكن وصف المشاهدات بموقعها المطلق باستعمال نظام الإحداثيات أو موقعها النسبي مستندا إلى مقياس المسافة، وبمعنى أكثر عمومية يتم تنظيم البيانات من قبل وحدات مكانية للمشاهدات وبمعنى اعم يمكن النظر في أي حالة تكون فيها البيانات مهياً وفقاً لمقياس الموقع أو المسافة في أي حيز، حيث أن الوحدات المكانية القريبة مع بعضها تكون قوة الارتباط المكاني عالية وهنا تكون قوة التأثير المكاني عالية أيضاً بينما كل ما تبتعد هذه الوحدات المكانية عن بعضها فان قوة الارتباط المكاني تقل وهذا يؤدي إلى تناقص قوة التأثير المكاني وهذا مؤشر إلى أن القياس الاقتصادي يعتمد على التأثير المكاني وتدرج المسافة والموقع، ويمكن تقديره للمناطق الصغيرة

باستعمال بيانات عن عدد من المقاطعات المتجاورة من اجل التخفيف من ندرة السلاسل الزمنية [3].

### 5- مصفوفة التجاور المكاني : Spatial Contiguity Matrix

أن القياسات الطبيعية للاعتمادية المكانية أو الارتباط الذاتي المكاني تعتمد على فكرة التجاور الثنائي بين الوحدات المكانية، فإذا كانت هناك وحدتان مكانيتان لهما حدود مشتركة ذات طول غير صفري فأنها تعتبر متلاصقة وتخصص لها قيمة (1)، أما إذا كانت الوحدتان المكانيتان ليس لهما حدود مشتركة فأنها تعتبر غير متلاصقة وتخصص لها قيمة (0) ومن هذا المنطلق يمكن إيجاد مصفوفة التجاور المكاني والتي تعرف بمصفوفة الوزن المكاني ذات البعد  $n \times n$  وهي تكون مربعة، موجبة، متماثلة وغير عشوائية ويرمز لها بالرمز  $W$  وعناصرها تمثل  $W_{ij}$  في الموقع  $z_i, z_j$  أي يتم تعيين قيمها لكل زوج من المواقع المتجاورة وغير المتجاورة من خلال هذه القواعد المحددة التي تحدد العلاقات المكانية بين المواقع، كذلك نلاحظ أن عناصر القطر الرئيسي للمصفوفة  $W$  تكون أصفار والسبب في ذلك لأن الخلية لا تجاور نفسها، ومن خلال الصيغة (5) يتم تحديد قيم عناصر مصفوفة الأوزان المكانية  $W$  وكالاتي: [3]

$$W_{ij} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are contiguous} \\ 0 & \text{if } i \text{ and } j \text{ are not contiguous} \end{cases} \dots (4)$$

ولبناء مصفوفة التجاور المكانية المعتمدة على الحدود المشتركة تم الاعتماد على معيار تجاور كوين (Queen Contiguity Criterion)، حيث في هذا المعيار يكون التجاور عندما الخليتان المتجاورتان تشتركان في جانب مشترك وكذلك تشتركان في نقطة قمة الرأس المشترك والشكل (1) في أدناه يوضح ذلك :- [13]

|   |   |   |
|---|---|---|
| A | B | C |
| D | E | F |
| G | H | I |

شكل ( 1 ) تجاور معيار Queen

الشكل (1) يبين الحدود المشتركة وقمة الرأس المشتركة أي التجاور بين الخلية E وجميع الخلايا الأخرى A ، B ، C ، D ، F ، G ، H ، I. ومن خلال الشكل والشرح أعلاه يمكن إيجاد مصفوفة أوزان التجاور المكانية (W) وكما يلي :-

$$W = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \end{bmatrix} \quad \dots (5)$$

من خلال مصفوفة التجاور الثنائية المكانية من الممكن إيجاد مصفوفة التجاور المكانية المعدلة والتي يرمز لها بالرمز  $W^{Adj}$  والتي تعتبر امتداد لها بعد إيجاد الإجراءات لها من خلال الصيغة الآتية: [3]

$$W_{ij}^{Adj} = \begin{cases} \frac{W_{ij}}{\sum W_{ij}} & \text{if } i \text{ Contiguity } j \\ 0 & \text{if otherwise} \end{cases} \quad \dots (6)$$

حسب معيار تجاور Queen فيرمز لمصفوفة التجاور المكانية المعدلة بـ  $W_Q^{Adj}$  ويكون مجموع كل صف في هذه المصفوفة المعدلة يساوي واحد أي  $\sum_{j=1}^n W_{ij} = 1$  وان قيمة كل عنصر في المصفوفة المعدلة هو كما في الصيغة (6). [1,3]

## 6- الدوال اللبية المستعملة: Kernel Functions used

إن مفهوم الأسلوب البسيط لتمثيل سلسلة الوزن  $\{W_{hi(x)}\}_{i=1}^n$  هو لوصف شكل دالة الوزن  $W_{hi(x)}$  من خلال دالة الكثافة مع معلمة القياس التي تعدل حجم وشكل الأوزان بالقرب من النقطة  $x$ ، ومن المعروف أن طرائق التقدير هي تعتمد على أساليب التمهيد اللامعلمي وان هذا التمهيد هو توفيق المنحنيات والذي يعمل على إيجاد أفضل منحني ممهد مطابق مع منحني المتغير المعتمد ( $Y$ ) ومن أساليب توفيق المنحنيات هو أسلوب دالة كيرنل والتي تعرف بأنها دالة وزن حقيقية متماثلة تستعمل في تمهيد بيانات المتغير ( $Y$ )، أي تستعمل في تقدير دالة التمهيد المجهولة. [8] ومن بين هذه الدوال الشائعة الاستعمال والتي يرمز لها بالرمز  $K(u)$  هي دالة Gaussian اللبية، أو دالة Tricube اللبية.

وان هذه الدوال يطلق عليها بدوال الكثافة الاحتمالية وذلك كون قيمة تكاملها مساوي إلى الواحد الصحيح والتي تستعمل في المقدار الخطي الموضوعي (LLE) الذي يقدر دالة الانحدار ومن الجدير بالذكر أن الدالة اللبية تحدد كيفية مساهمة المشاهدات لكيرنل الخطي الموضوعي عنده النقطة  $x_0$  المعرفة بأنها أي مشاهدة من المشاهدات  $x_1$ . [9]

وعند استعمال الدالة الكاوسية اللبية يتم تحديد مساهمة النقاط من خلال المسافة  $x_1$  عند النقطة  $x_0$ ، وهذا يدل على أن المسافة الأصغر ( $x_1 - x_0$ ) هي المساهمة الأكبر وهذا بسبب أن الدالة الكاوسية اللبية هي تكون على شكل جرس أي (ناقوس) [13].

أما عند استعمال الدالة (Tricube) اللبية فهي تتبع نفس سلوك الدالة (Gaussian) اللبية وان هذه الدالة سميت بدالة (Tricube) اللبية لأنها ناتجة من اشتقاق متوسط مربعات الخطأ المتكامل (MISE) وهذه الدالة تسمى بالدالة المثلى عندما تقلل (MISE) وكذلك كونها تتمتع ببعض الخصائص المثلى<sup>[8]</sup>.

وهنا فان أفضل دالة في الدوال اللبية هي التي تقلل متوسط مربعات الخطأ المتكامل (MISE). ومن المعروف أن الدالة اللبية تمتلك عدة شروط ومن هذه الشروط هي:-<sup>[7]</sup>

- 1- أن تكون دالة ذات قيم حقيقية وغير سالبة وتمتلك الاستمرارية أي  $0 < K(u) < \infty$
- 2- أن تكون دالة كثافة احتمالية أي قيمة تكامل الدالة مساوي إلى الواحد الصحيح وكالاتي:-

$$\int K(u)du = 1$$

- 3- أن تكون دالة متماثلة أي

$$K(u) = K(-u)$$

- 4- أن تكون الدالة لها متوسط مساوي إلى الصفر أي أن قيمة العزم الأول  $(\mu_1(k))$  يكون كالاتي:

$$E(u) = \mu_1(k) = \int uK(u)du = 0$$

- 5- أن تكون الدالة تمتلك تباين معلوم أي أن قيمة العزم الثاني  $\mu_2(k)$  تكون محددة وكالاتي :

$$E(u^2) = \mu_2(k) = \int u^2K(u)du = \sigma_k^2 < \infty$$

### 7- طريقة المقدر الخطي الموضوعي لتقدير دالة التمهيد $m(x)$ :- Local Linear Estimator (LLE)

تم استعمال طريقة المقدر الخطي الموضوعي (LLE) لتقدير دالة التمهيد  $m(x)$  في أنموذج انحدار التأخر المكاني شبه المعلمي (SPSAR)، والذي يعد إحدى أنواع مميزات اللبية المتعددة الحدود الموضوعية (Local Polynomial Kernel) (LPK) وبذلك فان الطريقة الخطية الموضوعية مبينة على تقليل معادلة المربعات الصغرى الموزونة المعرفة بالصيغة الآتية:-<sup>[11]</sup>

$$\min_{\{a,b\}} (Y - R_X)' K(x)(Y - R_X) \quad \dots (20)$$

وبذلك نحصل على الصيغة النهائية إلى التقدير  $\tilde{m}(x)$  وكالاتي:-<sup>[13]</sup>

$$\tilde{m}(x) = \acute{e} (R'_X K_X R_X)^{-1} R'_X K_{XY} \quad \dots (21)$$

حيث أن:

$$\acute{e} = (1,0)$$

0 : يمثل متجه من الاصفار أي طول d والتي تمثل عدد المتغيرات المستقلة X ،  $1_n$  : يمثل متجه من الواحدات أي طول n والتي تمثل عدد مشاهدات العينة.

$$1'_n = (1, \dots, 1)$$

$$R_x = \begin{bmatrix} 1 & X_1 - x \\ \vdots & \vdots \\ 1 & X_n - x \end{bmatrix}$$

$$K_x = \text{diag}(K_h(X_1 - x)K_h(X_2 - x) \dots K_h(X_n - x))$$

$$K_h(X_i - x) = h^{-1}K(h^{-1}(X_i - x)).$$

حيث أن :  $K_h(X_i - x)$  : تمثل أداة تمهيد اللبي وهي دالة وزن،  $X$  : تمثل نقطة مرشحة تأخذ قيم المشاهدات  $X_1, X_2, \dots, X_n$

$h$  : تمثل عرض الحزمة،  $\hat{m}(x)$  : هو الممهد الخطي الموضوعي (LLE) الذي يمثل المعدل الموزون إلى الاستجابات، حيث تم استعمال دالة Gaussian Kernel لهذا المقدر.<sup>[12]</sup>

### 8- الدوال اللبية المقترحة :- The Proposed Kernel Functions

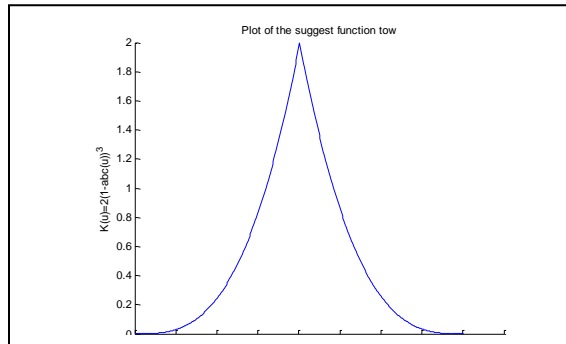
من خلال استعمال طريقة المقدر الخطي الموضوعي لتقدير  $m(x)$  لأنموذج (SPSAR) تم استعمال دالة لبية ل Tricube والتي تعرف بأنها دالة وزن حقيقية متماثلة تستعمل في تمهيد بيانات المتغير  $(Y)$ ، أي تستعمل في تقدير دالة التمهيد المجهولة وتعتبر من أساليب توفيق المنحنيات.<sup>[8]</sup>

وعند استعمال الدالة (Tricube) اللبية فهي تتبع نفس سلوك الدالة (Gaussian) اللبية ومن المعروف أن الدالة اللبية تستعمل لوضع أوزان للمشاهدات القريبة والمجاورة لنقطة التوفيق وبالاعتماد على طريقتي المقدر الخطي الموضوعي ، ومن هذا المنطلق تم اقتراح دالتين وبالشروط والمواصفات المذكورة في أعلاه وهذه الدالتين المقترحة مبينة كما في الصيغتين (22) و(23) وكالاتي :-

1- الدالة المقترحة الأولى (SUG1) وكما يأتي :

$$K(u) = 2(1 - |u|^1)^3 \quad \dots (22)$$

كما تظهر هذه الدالة في الشكل (2) أدناه :

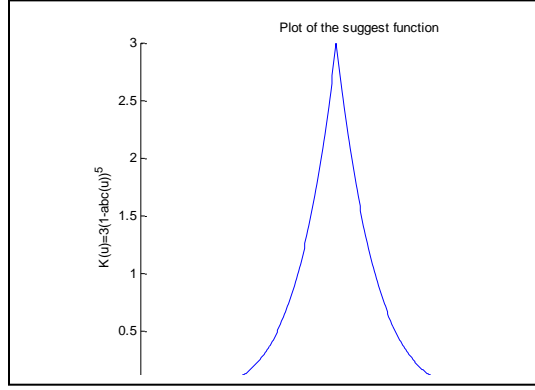


الشكل (2) يوضح الدالة المقترحة الأولى (SUG1)

## 2- الدالة المقترحة الثانية (SUG2) وكما يأتي :

$$K(u) = 3(1 - |u|^1)^5 \quad \dots (23)$$

كما تظهر هذه الدالة في الشكل (3) أدناه :



الشكل (3) يوضح الدالة المقترحة الثانية (SUG2)

## 9- معيار العبور الشرعي لاختيار معلمة التمهيد للطرائق اللامعلمية

تم استعمال معيار العبور الشرعي (CV) لاختيار قيمة معلمة التمهيد (h) وهو من أفضل المعايير وان صيغته تكتب على النحو الآتي :

$$CV(h) = n^{-1} \sum_{i=1}^n [Y_i - \hat{g}(X_i; h)]^2 \quad \dots (22)$$

حيث أن :

$\hat{g}(X_i; h)$  : تمثل مقدر دالة الانحدار اللامعلمي للملاحظات المكانية ولكن بإهمال النقطة  $(X_i, Y_i)$ .

إذ إن القيمة المثلى لمعلم التمهيد  $\hat{h}_{CV}$  هي القيم التي تجعل معيار العبور الشرعي  $CV(h)$  أقل ما يمكن ضمن مدى قيم (h) على التوالي [11]. ( $h > 0$ )

## 10- معيار المقارنة بين طرائق التقدير

تم استعمال معيار متوسط الخطأ النسبي المطلق (Mean Absolute Percentage Error) للوصول للمقدر الأكفأ، والذي يقيس مدى اقتراب أو ابتعاد القيمة المقدرة من القيم الحقيقية ويرمز له (MAPE).

وصيغته الرياضية كالآتي :- [2]

$$MAPE(\hat{Y}) = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{Y}_i - Y_i}{Y_i} \right| \quad \dots (23) \quad i=1, \dots, n$$

حيث أن Y : يمثل القيمة الحقيقية،  $\hat{Y}$  : يمثل القيمة المقدرة، n : يمثل عدد المشاهدات المستعملة لكل تجربة.



**11- تنفيذ تجارب المحاكاة**

من خلال استعمال البرنامج الإحصائي (Matlab 2014) تم تنفيذ تجارب المحاكاة لأنموذج انحدار التأخر المكاني شبه المعلمي (SPSAR) والتي تضمنت عدة خطوات وكالاتي :-  
 أولاً- تم تكرار التجربة 1000 مرة بعد تحديد ثلاث حجوم عينات مختلفة  $n = 25, 75, 150$  مع افتراض ثلاث قيم لمعلمة الاعتماد المكاني ( $\rho$ ) والمتمثلة بالمستويات  $\rho = (0.2, 0.5, 0.8)$   
 ثانياً- توليد المتغيرات المتمثلة بحد الخطأ العشوائي والمتغير التوضيحي ( $X, \varepsilon$ ) والتي تتوزع كالاتي:

$$X \sim U(0,1)$$

$$\varepsilon \sim N(0, \sigma^2) \quad , \quad \sigma = 0.1, 0.2, 0.5$$

ثالثاً- توليد مصفوفة الأوزان المكانية المعدلة حسب معياري Queen والمتمثلة  $W_Q^{Adj}$   
 رابعاً- تم استعمال عدة انواع من الدوال الرياضية والتي تمثل دوال التمهيدي لأنموذج (SPSAR) مع الأخذ بنظر الاعتبار طبيعة البيانات المتعلقة بالبعد المكاني، وهذه الدوال هي كالاتي :-

1- الدالة غير الخطية بوجود حد ثابت وصيغتها هي كالاتي:

$$m(X_1) = 2 + \sin(1.5X)$$

2- الدالة الغير خطية وصيغتها هي كالاتي:

$$m(X_2) = \sin(2X) + 2e^{-2X^2}$$

3- الدالة الغير خطية وصيغتها هي كالاتي:

$$m(X_3) = X + 2e^{-16X^2}$$

خامساً- بعد توليد المتغير التوضيحي  $X$  ومتغير حد الخطأ  $\varepsilon$  ومصفوفة  $W_Q^{Adj}$  وتحديد قيم المعلمة ( $\rho$ ) يتم تعويضها في صيغ النماذج وذلك للحصول على المتغير المعتمد  $Y$ .  
 سادساً- استعمال المتغيرات والمعالم والدوال المولدة في تقدير أنموذج (SPSAR) باستعمال طريقة المقدر الخطي الموضوعي (LLE) ومن ثم المقارنة باستعمال معيار (MAPE).  
 سابعاً- المقارنة لمعيار (MAPE) في الصيغة (23) وكما يلي :  
 1- طريقة المقدر الخطي الموضوعي (LLE) باستعمال الدالة اللبية (Gaussian).  
 2- طريقة المقدر الخطي الموضوعي (LLE) باستعمال الدالة اللبية المقترحة الأولى (SUG1) والمبينة في الصيغة (22).  
 3- طريقة المقدر الخطي الموضوعي (LLE) باستعمال الدالة اللبية المقترحة الثانية (SUG2) والمبينة في الصيغة (23).

**12- نتائج تجارب المحاكاة**

نستعرض نتائج تجارب المحاكاة التي استعملت للمقارنة بين طرائق التقدير شبه المعلمية المستعملة والمقترحة والتي تضم ثلاث دوال لأنموذج انحدار التأخر المكاني شبه المعلمي (SPSAR) في ظل معيار تجاور Queen، ومن هنا تتم عملية المقارنة بين طرائق التقدير وفقا لمعيار متوسط الخطأ النسبي المطلق (MAPE) وعند ثلاث حجوم عينات مختلفة وثلاث مستويات مختلفة للانحراف المعياري وكذلك أيضا ثلاث قيم مختلفة لمعلمة الاعتماد المكاني  $\rho$  وفي ظل مصفوفة الأوزان المكانية المعدلة لمعيار تجاور Queen تتم عملية المقارنة بين

طرائق التقدير شبه المعلمية من خلال الجداول التي تبدأ من الجدول (1) الى الجدول (9) والمبينة وكما يلي :-

جدول (1) يبين المعدل لقيم معيار (MAPE) لجميع حجوم العينات ولكافة طرائق التقدير شبه المعلمية (SPSAR) لدالة التمهيد  $m(X_1)$  ومصفوفة الأوزان المكانية المعدلة باستعمال معيار Queen وعندما  $\sigma_1=0.1$

| حجم العينة | طريقة التقدير<br>LLE | معلمة الاعتماد المكاني |           |           |
|------------|----------------------|------------------------|-----------|-----------|
|            |                      | $\rho_1$               | $\rho_2$  | $\rho_3$  |
| n=25       | GAU.                 | 1.0019604              | 0.9358426 | 0.9441845 |
|            | SUG1                 | 0.9591308              | 0.9742827 | 0.9402497 |
|            | SUG2                 | 0.9818545              | 0.9495375 | 0.9548788 |
| n=75       | GAU.                 | 0.9421484              | 0.9332745 | 0.9448314 |
|            | SUG1                 | 0.9558837              | 0.9322211 | 0.9425780 |
|            | SUG2                 | 0.9684845              | 0.9292995 | 0.9357240 |
| n=150      | GAU.                 | 0.9644682              | 0.9450309 | 0.9599572 |
|            | SUG1                 | 0.9584151              | 0.9503111 | 0.9667003 |
|            | SUG2                 | 0.9646236              | 0.9717755 | 0.9730457 |

جدول (2) يبين المعدل لقيم معيار (MAPE) لجميع حجوم العينات ولكافة طرائق التقدير شبه المعلمية (SPSAR) لدالة التمهيد  $m(X_1)$  ومصفوفة الأوزان المكانية المعدلة باستعمال معيار Queen وعندما  $\sigma_2=0.2$

| حجم العينة | طريقة التقدير<br>LLE | معلمة الاعتماد المكاني |           |           |
|------------|----------------------|------------------------|-----------|-----------|
|            |                      | $\rho_1$               | $\rho_2$  | $\rho_3$  |
| n=25       | GAU.                 | 0.9189044              | 0.9389276 | 0.9440363 |
|            | SUG1                 | 0.9160624              | 0.9832004 | 0.9510599 |
|            | SUG2                 | 0.9300253              | 0.9940329 | 0.9794610 |
| n=75       | GAU.                 | 0.9612840              | 0.9497026 | 0.9440842 |
|            | SUG1                 | 0.9403894              | 0.9467130 | 0.9442580 |
|            | SUG2                 | 0.9513163              | 0.9665651 | 0.9430636 |
| n=150      | GAU.                 | 0.9407617              | 0.9569880 | 0.9416098 |
|            | SUG1                 | 0.9332898              | 0.9424193 | 0.9449306 |
|            | SUG2                 | 0.9682661              | 0.9629684 | 0.9611886 |

جدول (3) يبين المعدل لقيم معيار (MAPE) لجميع حجوم العينات ولكافة طرائق التقدير شبه المعلمية (SPSAR) لدالة التمهيد  $m(X_1)$  ومصفوفة الأوزان المكانية المعدلة باستعمال معيار Queen وعندما  $\sigma_3=0.5$

| حجم العينة | طريقة التقدير<br>LLE | معلمة الاعتماد المكاني |           |           |
|------------|----------------------|------------------------|-----------|-----------|
|            |                      | $\rho_1$               | $\rho_2$  | $\rho_3$  |
| n=25       | GAU.                 | 0.9825199              | 0.9325263 | 0.9447048 |
|            | SUG1                 | 0.9924439              | 0.9346845 | 0.9402968 |
|            | SUG2                 | 0.9954574              | 0.9367775 | 0.9536870 |
| n=75       | GAU.                 | 0.9325625              | 0.9361010 | 0.9443131 |
|            | SUG1                 | 0.9317022              | 0.9411386 | 0.9420627 |
|            | SUG2                 | 0.9997395              | 0.9480260 | 0.9419094 |
| n=150      | GAU.                 | 0.9337419              | 0.9440186 | 0.9437267 |
|            | SUG1                 | 0.9246483              | 0.9338438 | 0.9463552 |
|            | SUG2                 | 0.9258507              | 0.9594230 | 0.9615530 |

جدول (4) يبين المعدل لقيم معيار  $(MAPE)$  لجميع حجوم العينات ولكافة طرائق التقدير شبه المعلمية  $(SPSAR)$  لدالة التمهيد  $m(X_2)$  ومصفوفة الأوزان المكانية المعدلة باستعمال معيار  $Queen$  وعندما  $\sigma_1=0.1$

| حجم العينة | طريقة التقدير<br>LLE | معلمة الاعتماد المكاني |           |           |
|------------|----------------------|------------------------|-----------|-----------|
|            |                      | $\rho_1$               | $\rho_2$  | $\rho_3$  |
| n=25       | GAU.                 | 0.9650911              | 0.9738978 | 0.9482854 |
|            | SUG1                 | 0.9455658              | 0.9856022 | 0.9401338 |
|            | SUG2                 | 0.9570465              | 0.9914668 | 0.9433732 |
| n=75       | GAU.                 | 0.9463638              | 0.9510159 | 0.9406557 |
|            | SUG1                 | 0.9434389              | 0.9589359 | 0.9311371 |
|            | SUG2                 | 0.972822               | 0.9651387 | 0.9487385 |
| n=150      | GAU.                 | 0.9845714              | 0.9527249 | 0.9417404 |
|            | SUG1                 | 0.9482585              | 0.9496051 | 0.9343521 |
|            | SUG2                 | 0.9851369              | 0.9515677 | 0.9326603 |

جدول (5) يبين المعدل لقيم معيار  $(MAPE)$  لجميع حجوم العينات ولكافة طرائق التقدير شبه المعلمية  $(SPSAR)$  لدالة التمهيد  $m(X_2)$  ومصفوفة الأوزان المكانية المعدلة باستعمال معيار  $Queen$  وعندما  $\sigma_2=0.2$

| حجم العينة | طريقة التقدير<br>LLE | معلمة الاعتماد المكاني |           |           |
|------------|----------------------|------------------------|-----------|-----------|
|            |                      | $\rho_1$               | $\rho_2$  | $\rho_3$  |
| n=25       | GAU.                 | 0.9200243              | 0.9518559 | 0.9643507 |
|            | SUG1                 | 0.9491439              | 0.9515009 | 0.9765820 |
|            | SUG2                 | 0.9517712              | 0.9720018 | 0.9473780 |
| n=75       | GAU.                 | 0.9804460              | 0.9532338 | 0.9435295 |
|            | SUG1                 | 0.9016942              | 0.9495943 | 0.9445576 |
|            | SUG2                 | 1.0023952              | 0.9738125 | 0.9521426 |
| n=150      | GAU.                 | 0.9659084              | 0.9463953 | 0.9501485 |
|            | SUG1                 | 0.9632820              | 0.9427624 | 0.9478221 |
|            | SUG2                 | 0.9844753              | 0.9646159 | 0.9409620 |

جدول (6) يبين المعدل لقيم معيار ( $MAPE$ ) لجميع حجوم العينات ولكافة طرائق التقدير شبه المعلمية ( $SPSAR$ ) لدالة التمهيد  $m(X_2)$  ومصفوفة الأوزان المكانية المعدلة باستعمال معيار  $Queen$  وعندما  $\sigma_3=0.5$

| حجم العينة | طريقة التقدير<br>LLE | معلمة الاعتماد المكاني |           |           |
|------------|----------------------|------------------------|-----------|-----------|
|            |                      | $\rho_1$               | $\rho_2$  | $\rho_3$  |
| n=25       | GAU.                 | 0.9578828              | 0.9313945 | 0.9447199 |
|            | SUG1                 | 0.9768569              | 0.9273403 | 0.9464791 |
|            | SUG2                 | 0.9848987              | 0.9478936 | 0.9504187 |
| n=75       | GAU.                 | 0.9382055              | 0.9328980 | 0.9439104 |
|            | SUG1                 | 0.9408636              | 0.9333282 | 0.9410293 |
|            | SUG2                 | 0.9483576              | 0.9419422 | 0.9408271 |
| n=150      | GAU.                 | 0.9504661              | 0.9526187 | 0.9555181 |
|            | SUG1                 | 0.9497944              | 0.9580625 | 0.9568371 |
|            | SUG2                 | 1.0068754              | 0.9622352 | 0.9660613 |

جدول (7) يبين المعدل لقيم معيار ( $MAPE$ ) لجميع حجوم العينات ولكافة طرائق التقدير شبه المعلمية ( $SPSAR$ ) لدالة التمهيد  $m(X_3)$  ومصفوفة الأوزان المكانية المعدلة باستعمال معيار  $Queen$  وعندما  $\sigma_1=0.1$

| حجم العينة | طريقة التقدير<br>LLE | معلمة الاعتماد المكاني |           |           |
|------------|----------------------|------------------------|-----------|-----------|
|            |                      | $\rho_1$               | $\rho_2$  | $\rho_3$  |
| n=25       | GAU.                 | 1.0115852              | 0.9344060 | 0.9467672 |
|            | SUG1                 | 0.9744224              | 0.9342907 | 0.9634862 |
|            | SUG2                 | 0.9960896              | 0.9529372 | 0.9866006 |
| n=75       | GAU.                 | 0.9739702              | 0.9391624 | 0.9398646 |
|            | SUG1                 | 0.9807412              | 0.9340158 | 0.9645601 |
|            | SUG2                 | 0.9781726              | 0.9330691 | 0.9835643 |
| n=150      | GAU.                 | 0.9838235              | 0.9379267 | 0.9507534 |
|            | SUG1                 | 0.9781304              | 0.9349326 | 0.9439154 |
|            | SUG2                 | 0.9515703              | 0.9396968 | 0.9466336 |

جدول (8) يبين المعدل لقيم معيار ( $MAPE$ ) لجميع أحجام العينات ولكافة طرائق التقدير شبه المعلمية ( $SPSAR$ ) لدالة التمهيد  $m(X_3)$  ومصفوفة الأوزان المكانية المعدلة باستعمال معيار  $Queen$  وعندما  $\sigma_2=0.2$

| حجم العينة | طريقة التقدير<br>LLE | معلمة الاعتماد المكاني |           |           |
|------------|----------------------|------------------------|-----------|-----------|
|            |                      | $\rho_1$               | $\rho_2$  | $\rho_3$  |
| n=25       | GAU.                 | 0.9949382              | 0.9440823 | 0.9504683 |
|            | SUG1                 | 0.9894347              | 0.9492315 | 0.9561646 |
|            | SUG2                 | 0.9727047              | 0.9593779 | 0.9575222 |
| n=75       | GAU.                 | 0.9805359              | 0.9633272 | 0.9441745 |
|            | SUG1                 | 0.9565617              | 0.9559245 | 0.9434276 |
|            | SUG2                 | 0.9834909              | 0.9608562 | 0.9444264 |
| n=150      | GAU.                 | 0.9808154              | 0.9473389 | 0.9433870 |
|            | SUG1                 | 0.9419268              | 0.9480019 | 0.9399092 |
|            | SUG2                 | 0.9424613              | 0.9449528 | 0.9384572 |

جدول (9) يبين المعدل لقيم معيار ( $MAPE$ ) لجميع أحجام العينات ولكافة طرائق التقدير شبه المعلمية ( $SPSAR$ ) لدالة التمهيد  $m(X_3)$  ومصفوفة الأوزان المكانية المعدلة باستعمال معيار  $Queen$  وعندما  $\sigma_3=0.5$

| حجم العينة | طريقة التقدير<br>LLE | معلمة الاعتماد المكاني |           |           |
|------------|----------------------|------------------------|-----------|-----------|
|            |                      | $\rho_1$               | $\rho_2$  | $\rho_3$  |
| n=25       | GAU.                 | 0.9340939              | 0.9627669 | 0.9425762 |
|            | SUG1                 | 0.9405995              | 0.9881473 | 0.9438691 |
|            | SUG2                 | 0.9454052              | 0.9758492 | 0.9363122 |
| n=75       | GAU.                 | 0.9868468              | 0.9408479 | 0.9431357 |
|            | SUG1                 | 0.9839688              | 0.9325202 | 0.9425861 |
|            | SUG2                 | 1.0019072              | 0.9309648 | 0.9444639 |
| n=150      | GAU.                 | 0.9798425              | 0.9661562 | 0.9438547 |
|            | SUG1                 | 0.9655303              | 0.9647765 | 0.9428784 |
|            | SUG2                 | 0.9618458              | 0.9601078 | 0.9419561 |

ومن الجدول (10)، الذي يمثل ملخص للجداول (1) إلى (9) والذي يتضمن تكرار ظهور القيمة الأصغر لمعيار  $(MAPE)$  وعدد التكرار والنسب لكل طرائق التقدير يتضح بان طريقة المقدر الخطي الموضوعي باستعمال الدالة اللبية المقترحة الأولى  $(SUG1K)$  هي من أفضل الطرائق ويكون تسلسلها الأول من بين طرائق التقدير لأنموذج التأخر المكاني شبه المعلمي  $(SPSAR)$ ، إذ كان تكرار أفضليتها 38 مرة من مجموع 81 حالة تكرار ونسبة مقدارها 46.91% أي تمتلك أعلى نسبة، في ظل مصفوفة الأوزان المكانية المعدلة لمعيار تجاور  $Queen$ ، تليها طريقة المقدر الخطي الموضوعي باستعمال الدالة اللبية  $(LLEK)$  ونسبة مقدارها 33.33%، أما بعدها فتليها طريقة المقدر الخطي الموضوعي باستعمال الدالة اللبية المقترحة الثانية  $(SUGK2)$  ونسبة مقدارها 19.75%.

جدول (10) يوضح تسلسل نسبة وعدد مرات تكرار أفضلية الطرائق المستعملة والمقترحة لأنموذج  $SPSAR$  في ظل مصفوفة الأوزان المكانية المعدلة لمعيار تجاور  $Queen$

| ت | طريقة التقدير<br>LLE | الأنموذج<br>الأول | الأنموذج<br>الثاني | الأنموذج<br>الثالث | عدد<br>التكرار | مقدار النسبة<br>% |
|---|----------------------|-------------------|--------------------|--------------------|----------------|-------------------|
| 1 | SUG1                 | 13                | 15                 | 10                 | 38             | 46.91             |
| 2 | GAU.                 | 11                | 9                  | 7                  | 27             | 33.33             |
| 3 | SUG2                 | 3                 | 3                  | 10                 | 16             | 19.75             |
|   | $\Sigma$             | 27                | 27                 | 27                 | 81             | 100               |

### 13- الاستنتاجات

بعد إجراء وصف وتنفيذ تجارب المحاكاة على أنموذج انحدار التأخر المكاني شبه المعلمي  $(SPSAR)$  في ظل معيار تجاور  $Queen$  نستنتج وبالاتتماد على معيار متوسط الخطأ النسبي المطلق  $(MAPE)$  نستنتج عند مقارنة طريقة التقدير شبه المعلمية المستعملة في

تقدير أنموذج انحدار التأخر المكاني شبه المعلمي (SPSAR) حسب مصفوفة التجاورات المكانية المعدلة في ظل معيار تجاور *Queen* أن أفضل دالة تم استعمالها لطريقة المقدم الخطي الموضوعي باستعمال الدالة اللبية المقترحة الاولى (*SUG1*) إذ أظهر استعمالها تقدما بشكل كامل كونها حققت اقل قيمة لمعيار (*MAPE*) وبنسبة مقدارها 46.91% في حالة جميع حجوم العينات ومستويات الانحراف المعياري وقيم معلمة الاعتماد المكاني □ المفترضة.

#### 14- التوصيات

بناء على الاستنتاجات التي تم التوصل إليها من خلال تجارب المحاكاة، يمكن أدرج أهم التوصيات وكالاتي :-

- 1- محاولة تطبيق الطرائق التي تعالج مشاكل الاعتماد المكاني على بيانات حقيقية لان هذه المشاكل لها قاعدة عريضة على ارض الواقع.
- 2- استعمال طرائق أخرى لحساب المسافات بين مواقع المشاهدات مثل طريقة قياس *Minkowski*.

#### المصادر

1. Anselin, L. and Bera, A. K. (1998); Spatial Dependence in Linear Regression Model With An Introduction to Spatial Econometrics; In: Ullah A, Giles DEA (eds) Handbook of applied economic statistics. Marcel Dekker, New York, pp (237-289).
2. Anselin, L. (2001). "Spatial econometrics. A companion to theoretical econometrics", 310-330.
3. Basile, R., & Gress, B. (2005). "Semi-parametric spatial auto-covariance models of regional growth behaviour in Europe". *Region et developpement*, 21, 93-118.
4. Gerolimetto, M., & Magrini, S. (2009). "Nonparametric regression with spatially dependent data". Department of Economics WP, 20- 2009.
5. Härdle, W. (1990). *Applied nonparametric regression* (No. 19). Cambridge university press.
6. Ibrahim, W. S., Majeed, G. H., & Hussain, W. J. (2021). Comparison and Estimation of a Spatial Autoregressive (SaR) Model for Cancer in Baghdad Regions. *Int. J. Agricult. Stat. Sci.* Vol, 17(1), 1921-1927.
7. Ibrahim, W. S., & Mousa N. S. (2022). Estimation of the general spatial regression model (SAC) by the maximum likelihood method. *International Journal of Nonlinear Analysis and Applications*, 13(1), 2947-2957.
8. LeSage, J. P. (1999). "The theory and practice of spatial econometrics". University of Toledo. Toledo, Ohio4.

9. Li, Q., & Racine, J. S. (2007). "Nonparametric econometrics: theory and practice". Princeton University Press.
10. Martins-Filho, C., & Yao, F. (2009). "Nonparametric regression estimation with general parametric error covariance. Journal of Multivariate Analysis", 100(3), 309-333.
11. Mauricio S. (2017). "Introduction to Spatial Econometric". Universidad Católica del Norte.
12. Ruppert, D., Wand, M.P., and, Carrol, R.J., 2003. "Semi parametric regression. Cambridge series in statistical and probabilistic mathematics". Cambridge University Press.
13. Wu, H., & Zhang, J. T. (2006). "Nonparametric regression methods for longitudinal data analysis: mixed-effects modeling approaches" (Vol. 515). John Wiley & Sons.





## استعمال بعض طرائق التقدير الحصينة في تقدير العلاقة بين الكميات المصدرة والمنتجة من النفط الخام العراقي

الباحث رضا قاسم محمد تميم  
الجامعة المستنصرية/ كلية الادارة والاقتصاد

أ.م.د. احمد شاكر محمد طاهر  
الجامعة المستنصرية/ كلية الادارة والاقتصاد

تاريخ استلام البحث: 2022/04/12

تاريخ قبول البحوث: 2022/05/09

نشر البحث في العدد عشرون: ايلول / سبتمبر 2023

رمز التصنيف ديوي / النسخة الالكترونية (Online): 2522-64X/310.1

رمز التصنيف ديوي / النسخة الورقية (Print): 2519-948X/310.1

## استعمال بعض طرائق التقدير الحصينة في تقدير العلاقة بين الكميات المصدرة والمنتجة من النفط الخام العراقي

الباحث رضا قاسم محمد تميم  
الجامعة المستنصرية/ كلية الادارة والاقتصاد

أ.م.د. احمد شاكر محمد طاهر  
الجامعة المستنصرية/ كلية الادارة والاقتصاد

### المستخلص:

يبني انموذج الانحدار على عدة افتراضات لكي نحصل على تقديرات المربعات الصغرى التي تتصف بخاصية افضل تقدير خطي غير متحيز. تجانس التباين حدود الخطأ العشوائي يعد احد هذه الافتراضات والذي يتعذر تحقيقه في بعض البيانات الاقتصادية والمالية. ينطرق البحث الى استعمال طرائق التقدير الحصينة MM و TLS و MPV في تقدير معاملات انموذج الانحدار الذي يصف العلاقة بين الكميات المصدرة من النفط الخام العراقي والكميات المنتجة منه الذي يعاني من مشكلة عدم تجانس التباين لحدود الخطأ العشوائي ومن ثم اجراء مقارنة بين نتائج تقدير تلك الطرائق للتوصل الى افضل طريقة تقدير. وافرزت النتائج عن افضلية طريقة TLS إذ حققت اقل قيمة لمعيار المقارنة متوسط مربعات الخطأ (MSE).

الكلمات الافتتاحية: انموذج الانحدار, عدم تجانس التباين, طرائق التقدير المحصنة

### Abstract:

The regression model is built on several assumptions in order to obtain the least squares estimation that have the property of the best linear unbiased estimate. Homogeneity of the random error terms is one of these assumptions which cannot be achieved in some economic and financial data. This research deals with the use of the robust estimation methods MM, TLS, and MPV to estimating the regression model coefficients that describes the relationship between the quantities of exported from Iraqi crude oil and quantities of produced from it, which suffers from the problem of heterogeneity of random error terms, and then a comparison is made between the results of estimating those methods to reach to the best estimation method. The result shows that the TLS is the best method, based on the mean squares error (MSE) criterion.

**1. المقدمة**

يعد تحليل الانحدار أسلوب مهم من الأساليب الإحصائية التطبيقية عند إجراء تحليل العلاقة بين الظواهر الاقتصادية والاجتماعية. ويهتم تحليل الانحدار بدراسة العلاقة بين متغير يسمى بمتغير الاستجابة وعدة متغيرات تدعى بالمتغيرات التوضيحية. كما تعد فرضية تجانس تباين الأخطاء العشوائية إحدى الفرضيات الأساسية المهمة من فرضيات التحليل الخاصة بنموذج الانحدار الخطي، ففي حالة عدم تحقق هذه الفرضية فإن مقدرات المربعات الصغرى الاعتيادية (OLS) ستكون غير كفوءة وان الاعتماد على تلك المقدرات سيؤدي الى نتائج مضللة وتفسيرات خاطئة حول طبيعة العلاقة بين الظواهر قيد البحث. لذا يجب البحث عن طرائق بديلة تكون حصينة لمشكلة عدم تجانس التباين وتكون ذات كفاءة عالية. ويهدف البحث الى استعمال بعض طرائق التقدير الحصينة ( طريقة المربعات الصغرى المشدبة (TLS) وطريقة التقدير (S) وطريقة التقدير (MPV) ) لتقدير معلمات أنموذج الانحدار بوجود مشكلة عدم تجانس تباين الأخطاء العشوائية ومن ثم المقارنة بين تلك الطرائق بالاعتماد على تطبيق عملي على بيانات واقعية للكميات الشهرية المصدرة والمنتجة من النفط الخام العراقي للفترة (2020-2008)، وقد تمت المقارنة باستعمال معيار المقارنة متوسط مربعات الخطأ (MSE)، اذ افرزت نتائج التقدير عن افضلية طريقة التقدير الحصينة TLS.

**2. مشكلة البحث:**

لغرض إجراء تحليل الانحدار الخطي للعلاقة بين الكميات المصدرة والمنتجة من النفط الخام العراقي يجب ان تتحقق الفرضيات الأساسية لأنموذج الانحدار واحدى تلك الفرضيات هي فرضية تجانس تباين الأخطاء العشوائية. إذ ان عدم تحقق هذه الفرضية سيؤدي الى تفسيرات خاطئة حول طبيعة العلاقة بين الكميات المصدرة والمنتجة إذا ما تم استعمال طرائق التقدير التقليدية لتقدير معلمات أنموذج الانحدار، الامر الذي يتطلب للبحث عن طرائق بديلة تتصف بمقدراتها بالكفاءة العالية وتكون حصينة لمشكلة عدم تجانس التباين.

**3. هدف البحث:**

يهدف البحث الى استعمال بعض طرائق التقدير الحصينة لتقدير معلمات أنموذج الانحدار الخطي البسيط الذي يصف العلاقة بين الكميات المصدرة والمنتجة من النفط الخام العراقي في ظل وجود مشكلة عدم تجانس تباين الأخطاء العشوائية ومن ثم ايجاد افضل طريقة حصينة بينها من خلال استعمال معيار المقارنة متوسط مربعات الخطأ.

**4. الجانب النظري:**

يتضمن هذا الجانب عرض نظري لأنموذج الانحدار الخطي البسيط ومشكلة عدم تجانس التباين وتقدير معاملات أنموذج الانحدار بطرائق التقدير الحصينة (TLS) و (MPV) و (MM).

**1.4 أنموذج الانحدار الخطي البسيط:**

يعد تحليل الانحدار من أكثر واوسع الأساليب الإحصائية استخداماً لأنه يحلل العلاقة الخطية والعشوائية بين المتغيرات المطلوب دراستها. وابتسط صيغة لعلاقة عشوائية بين متغيرين يعرف احدهما بمتغير الاستجابة (Y) بينما الآخر هو المتغير التوضيحي (X) فيدعى بالانحدار الخطي البسيط، وان الصيغة الرياضية لهذه العلاقة تعرف بالأنموذج الآتي: [8][2]،

$$Y_i = \alpha + \beta X_i + \epsilon_i \quad (1)$$

إذ أن  $(\epsilon)$  يمثل الخطأ العشوائي لأنموذج الانحدار وان  $(\alpha$  و  $\beta)$  تمثل معلمات الأنموذج والمطلوب تقديرها

ولتقدير هذه المعلمات لا يجب فقط ان يكون شكل أنموذج الانحدار واضحاً ولكن يجب تحديد التوزيع الاحتمالي للأخطاء العشوائية وكذلك قيم المتغير التوضيحي. وهذه المعلومات تؤخذ من خلال ما يسمى الفرضيات الاساسية لأنموذج الانحدار. وهذه الفرضيات هي: [12]، 1- ان الاخطاء العشوائية تتبع التوزيع الطبيعي بوسط حسابي يساوي صفر وتباين ثابت يساوي  $(\sigma^2)$

$$\epsilon_i \sim \text{IN}(0, \sigma^2) \quad (2)$$

ان حدود الخطأ العشوائي غير مرتبطة، اي ان التباين المشترك بينها يساوي صفر  $\text{cov}(\epsilon_i, \epsilon_j) = 0, \forall i \neq j$  (3)

3- ان تباين الخطأ العشوائي ثابت او متساو لجميع مشاهدات العينة

$$E(\epsilon_i^2) = \sigma^2 \quad (4)$$

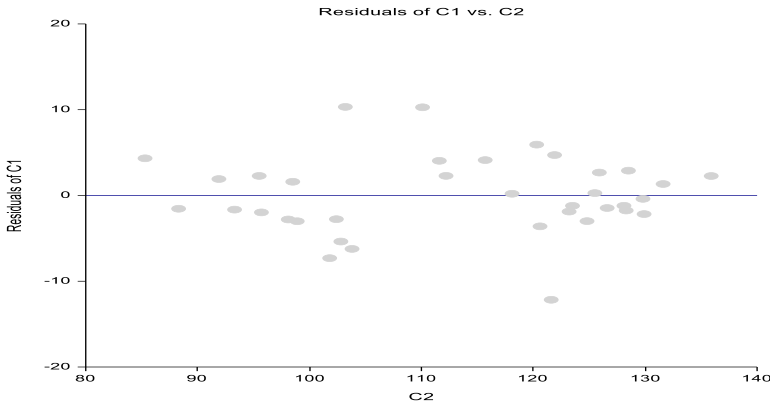
4- متغير الاستجابة  $Y$  هو متغير عشوائي بقيم مستقلة ويتبع التوزيع الطبيعي بمتوسط  $E(Y)$  وتباين يساوي  $(\sigma^2)$ :

$$Y_i \sim \text{IN}(E(Y), \sigma^2) \quad (5)$$

5- ان قيم المتغير التوضيحي  $(X)$  غير عشوائية وقيمها ثابتة في العينات المتكررة، وغير مرتبطة بحدود الخطأ العشوائي.

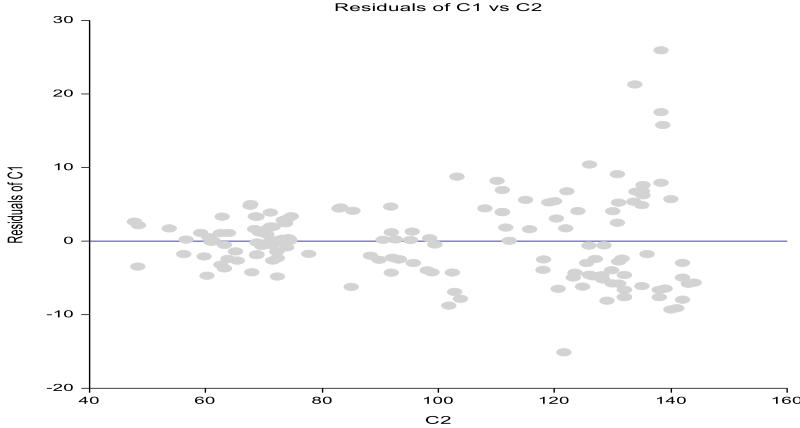
#### 2.4 عدم تجانس تباين الاخطاء العشوائية:

ان احدى الفرضيات الاساسية لأنموذج الانحدار الخطي البسيط ان الاخطاء العشوائية تملك تباين ثابت لكل المشاهدات المستقلة للعينة، كما تم توضيحها في المعادلة (4). بمعنى اخر ان الاخطاء العشوائية تكون متجانسة التباين (Homogeneity)، كما في الشكل (1). ذلك لان طريقة المربعات الصغرى الاعتيادية (OLS) تعتمد على تحقق هذه الفرضية في تقدير معلمات انموذج الانحدار. [12][8][2]،



الشكل (1): يمثل الرسم البياني لفرضية تجانس تباين الاخطاء العشوائية

ان عدم تحقق فرضية تجانس التباين فإنه يصبح استخدام طريقة المربعات الصغرى الاعتيادية امراً غير منطقياً ويؤدي الى نتائج مضللة. ففي معظم الدراسات القياسية فإن تباين مشاهدات المتغير المعتمد قد يختلف اختلافاً كبيراً من مستوى الى اخر من مستويات المتغيرات التوضيحية، وهكذا نجد بأن فرضية تجانس تباين حدود الخطأ العشوائي في هذه الحالة غير متحققة.



الشكل (2): يمثل الرسم البياني لفرضية عدم تجانس تباين الاخطاء العشوائية

وهذا يؤدي الى حدوث مشكلة تعرف بمشكلة عدم تجانس التباين (Heterogeneity). لذا فإن فرضية عدم تجانس التباين تأخذ الصيغة الاتية:

$$E(\epsilon_i^2) = \sigma_i^2, \quad i = 1, 2, \dots, n \quad (6)$$

وفي ظل تحقق فرضية عدم تجانس التباين فإن المعلمات المقدره بطريقة (OLS) لا تمتلك اقل تباين ممكن وبالتالي فإنها لن تكون افضل تقدير خطي غير متحيز [Best Linear Unbiased Estimator (BLUE)]. لذا يجب استعمال طرائق بديلة لطريقة (OLS) وتكون حصينة ضد مشكلة عدم تجانس التباين.

#### 3.4 طرائق التقدير الحصينة لمعاملات أنموذج الانحدار:

استعمال طريقة المربعات الصغرى الاعتيادية في تقدير معاملات أنموذج الانحدار الخطي البسيط تؤدي الى نتائج مضللة عند عدم تحقق فرضية تجانس التباين للأخطاء العشوائية وذلك لان تقدير تباين الاخطاء العشوائية يكون متحيزا مما يؤثر على اختبارات المعنوية للمعاملات المقدره وكذلك على بناء حدود الثقة والقيم التنبؤية. وكبديل لطريقة المربعات الصغرى الاعتيادية يمكن اعتماد طرائق التقدير الحصينة في حالة وجود مشكلة عدم التجانس ومن تلك الطرائق:

1- طريقة المربعات الصغرى المشدبة

2- طريقة تقدير (MM)

3- طريقة (MPV)

### 1.3.4 طريقة المربعات الصغرى المشدبة Trimmed Least Square Method

تم اقتراح هذه الطريقة من قبل الباحث (Rousseeuw) في عام (1984) والتي تعد من طرائق التقدير الحصينة لتقدير معالم أنموذج الانحدار وتتميز هذه الطريقة بكفاءة أفضل من طريقة المربعات الصغرى الاعتيادية (OLS) في حال عدم تجانس تباين الخطأ العشوائي وكذلك عند وجود القيم الشاذة، [4][5]، وتتلخص هذه الطريقة بالخطوات الآتية:

1- ايجاد مربعات البواقي ( $e_i^2$ ) لأنموذج الانحدار الخطي البسيط بطريقة المربعات الصغرى الاعتيادية (OLS).

2- يتم ترتيب تلك البواقي تصاعديا  $e_1^2 \leq e_2^2 \leq \dots \leq e_n^2$ .

3- ايجاد حجم المجموعة الجزئية (k) بالاعتماد على نسبة التشذيب ( $\lambda$ )، [9][16]، وفق الصيغة الآتية:

$$k = n(1 - \lambda) + \lambda(p + 1) \quad (7)$$

حيث n يمثل حجم العينة

p يمثل عدد المتغيرات التوضيحية

وفي حالة كانت نسبة التشذيب مساوية الى (50%) فيتم حساب حجم المجموعة الجزئية (k) من الصيغة التالية حيث تبلغ نقطة الانهيار اقصى حد لها عند هذه النسبة

$$k = \frac{n+p+1}{2} \quad (8)$$

4- يتم حساب مجموع مربعات البواقي بعد استبعاد مربعات البواقي الكبيرة من المجموع للحصول على تقدير لمعاملات أنموذج الانحدار وفق الصيغة الآتية:

$$\hat{\beta}_{LTS} = \min \sum_{i=1}^k e_i^2 \quad (9)$$

ان طريقة عمل (TLS) تشبه طريقة عمل (OLS) لكن الاختلاف يتمثل بان طريقة (TLS) تقوم باستبعاد تأثير القيم ذات البواقي الكبيرة قبل البدء بالتقدير.

ان طريقة (TLS) يمكن ان تكون كفوءة جدا اذا تم تشذيب البيانات بطريقة صحيحة. حيث تكون مقدرات هذه الطريقة مكافئة حسابيا لطريقة (OLS)، وعلى العكس من ذلك اذا كان التشذيب اكبر من المطلوب فقد يتم استبعاد بعض البيانات الجيدة من الحساب وهذا قد يؤثر على كفاءة مقدرات هذه الطريقة.

### 2.3.4 طريقة تقدير (MM) MM-estimation Method

وهي احدى طرائق التقدير الحصينة لتقدير معالم أنموذج الانحدار، والتي تعد تعديلا لطريقة تقدير (M) والمقترحة من قبل الباحث (Yohai) عام (1987). وتتميز مقدرات طريقة تقدير (MM) بامتلاكها نقطة انهيار عالية وكفاءة عالية في حالة اتباع الاخطاء العشوائية التوزيع الطبيعي، [10][15]، إن مقدر (MM) هو حل للمعادلة

$$\sum_{i=1}^n \rho_1 \left( \frac{y_i - \sum_{j=0}^k x_{ij} \hat{\beta}_j}{S_{MM}} \right) x_{ij} = 0 \quad (10)$$

حيث إن ( $S_{MM}$ ) هو الانحراف المعياري لبواقي طريقة تقدير (S)، وان ( $\rho_1$ ) هي دالة الوزن ل (Tukey's Bisquare) وكما يمكن اقتراح استعمال دالة الوزن ل (Huber). ويتم حساب مقدرات (MM) وفق الخطوات الآتية:

1- تقدير معالم أنموذج الانحدار بطريقة (OLS)

2- حساب قيم البواقي ( $e_i = Y_i - \hat{Y}_i$ ) باستعمال طريقة تقدير (S)

3- حساب قيمة  $(\hat{\sigma}_i = \hat{\sigma}_{S_n})$

4- حساب قيمة الاوزان  $(w_i)$  لكل دالة من دوال الاوزان المذكورة سابقا وكالآتي:  
أ- دالة الوزن (Tukey's Bisquare) وتحسب وفق الصيغة الآتية:

$$w_i = \begin{cases} \left(1 - \left(\frac{e_i/\hat{\sigma}_i}{4.685}\right)^2\right)^2, & |e_i| \leq 4.685 \\ 0, & |e_i| > 4.685 \end{cases} \quad (11)$$

ب- دالة الوزن (Huber) وتحسب وفق الصيغة الآتية:

$$w_i = \begin{cases} 1, & |e_i| < 1.345 \\ \frac{1.345}{|e_i|}, & |e_i| \geq 1.345 \end{cases} \quad (12)$$

5- حساب  $(\hat{\beta}_{MM})$  باستعمال طريقة المربعات الصغرى الموزونة (WLS)

6- تكرار الخطوات (2-5) حتى الحصول على قيم متقاربة لـ  $(\hat{\beta}_{MM})$

### 3.3.4 طريقة التقدير (MPV) MPV Estimation Method

اقترح الباحثون (Montgomery, Peck and Vining) في عام (2001)، [14][13][3]، تعديلاً على طريقة المربعات الصغرى الموزونة من خلال حساب قيمة الاوزان المعتمدة لمعالجة مشكلة عدم تجانس تباين الاخطاء العشوائية وذلك من خلال تقسيم قيم متغير الاستجابة الى مجاميع بالاعتماد على القيم المتقاربة للمتغير التوضيحي. وقد سميت هذه الطريقة بالمختصر (MPV) نسبة الى اسماء الباحثين. وتتلخص هذه الطريقة بالخطوات الآتية:

- 1- يتم ترتيب قيم المتغير التوضيحي  $(X_i)$  تصاعدياً، ثم تفرز القيم المتقاربة لهذا المتغير على شكل مجاميع
- 2- تفرز قيم متغير الاستجابة  $(Y_i)$  المناظرة لقيم المتغير التوضيحي على شكل مجاميع
- 3- يتم حساب الاوساط الحسابية لمجاميع المتغير التوضيحي  $(\bar{X}_j)$ ،  $(j = 1, 2, \dots, k)$  إذ ان  $(k)$  تمثل عدد المجاميع
- 4- يتم حساب تباين متغير الاستجابة ولكل مجموعة  $(S_j^2)$ ،  $(j = 1, 2, \dots, k)$
- 5- بناء أنموذج انحدار  $(S_j^2)$  باعتباره متغير استجابة على المتغير التوضيحي  $(\bar{X}_j)$  وفق الصيغة الآتية:

$$S_j^2 = \gamma_0 + \gamma_1 \bar{X}_j + e_j, \quad j = 1, 2, \dots, k \quad (13)$$

- 6- يتم تقدير معلمات الأنموذج (2-46)، إذ يمكن الحصول على تقدير لكل مجموعة
- 7- باستعمال الصيغة التقديرية لأنموذج الانحدار المبين بالصيغة (12) يمكن تقدير تباين الاخطاء  $(\hat{S}_i^2)$  لكل قيمة من قيم متغير الاستجابة  $(Y_i)$  وذلك بتعويض قيم المتغير التوضيحي  $(X_i)$  وكالآتي:

$$\hat{S}_i^2 = \hat{\gamma}_0 + \hat{\gamma}_1 X_i, \quad i = 1, 2, \dots, n \quad (14)$$

8- يتم حساب تقدير الاوزان  $(\hat{w}_i)$  وفق الصيغة الآتية:

$$\hat{w}_i = \frac{1}{\hat{S}_i^2} \quad (15)$$

- 9- يتم تقدير معلمات أنموذج الانحدار بموجب طريقة المربعات الصغرى الموزونة وفق الاوزان المقدره بالصيغة (15) وكالآتي:

$$\hat{\beta}_{MPV} = (X' \hat{W} X)^{-1} X' \hat{W} Y \quad (16)$$



وتعتبر هذه الطريقة كفاءة لتقدير معلمات أنموذج الانحدار في حالة وجود مشكلة عدم تجانس التباين، ولكنها تفقد خصائصها عندما تحتوي البيانات على قيم شاذة.

#### 4.4 اختبارات الكشف عن مشكلة عدم تجانس التباين

هناك عدد كثير من الاختبارات التي تعتمد في الكشف عن مشكلة عدم تجانس تباين الاخطاء العشوائية نذكر منها الآتي:

- 1- اختبار بارك (Park test)
- 2- اختبار غولدفيلد-كوانت (Goldfeld-Quandt test)
- 3- اختبار كوينكر-باسيت (Koenker-Bassett test)
- 4- اختبار هارفي-غودفري (Harvey-Godfrey test)
- 5- اختبار بروش-باغان (Breusch-Pagan-Godfrey test)
- 6- اختبار وايت العام (White's general test)
- 7- اختبار غليجر (Glejser test)

وسنتطرق بشيء من التفصيل لاختبارين من تلك الاختبارات.

#### 1.4.4 اختبار هارفي-غودفري

اقترح هذا الاختبار من قبل الباحث (Harvey) عام (1976) وطوره الباحث (Godfrey) عام (1978) وخطوات هذا الاختبار كالتالي: [1][11]

- 1- يتم تقدير معلمات أنموذج الانحدار في المعادلة (1) باستعمال طريقة المربعات الصغرى الاعتيادية ثم حساب قيم البواقي المقدرة ( $e_i$ ).
- 2- صياغة أنموذج الانحدار مساعد بمتغير استجابة يمثل اللوغاريتم الطبيعي لمربعات البواقي والمتغيرات التوضيحية لهذا الأنموذج عبارة عن المتغيرات التوضيحية لأنموذج الانحدار الاساسي كالتالي:

$$\ln(e_i^2) = \delta_0 + \delta_1 Z_{1i} + \delta_2 Z_{2i} + \dots + \delta_m Z_{mi} + v_i \quad (17)$$

3- صياغة الفرضية المطلوب اختبارها ولتكن:

$$H_0 : \delta_0 = \delta_1 = \dots = \delta_m = 0$$

على الاقل توجد معلمة واحدة لا تساوي صفر :  $H_1$

4- حساب قيمة إحصاء الاختبار من الصيغة الآتية:

$$HG = nR^2 \quad (18)$$

إذ ان ( $n$ ) تمثل عدد المشاهدات

( $R^2$ ) هو معامل التحديد لأنموذج الانحدار المساعد

إن إحصاء الاختبار تتبع توزيع مربع كاي ( $\chi^2$ ) بدرجة حرية ( $m - 1$ ). يتم رفض فرضية العدم اذا كانت قيمة إحصاء الاختبار اكبر من قيمة ( $\chi^2$ ) الجدولية عند درجة حرية ( $m - 1$ ) ومستوى معنوية (0.05)، هذا يعني بأن حدود الخطأ العشوائي غير متجانسة التباين.

#### 2.4.4 اختبار غولدفيلد-كوانت:

اقترح الباحثان (Goldfeld و Quandt) عام (1965) طريقة اختبار جديدة تعتمد على فكرة اذا كانت تباينات البواقي هي نفسها لجميع المشاهدات ( اي انها متجانسة التباين) فإن تباين

الجزء الاول من العينة يجب ان يكون متساوٍ مع تباين الجزء الاخر منها. ويمكن توضيح خطوات الاختبار كما يلي: [11][6][1]

1- يتم ترتيب مشاهدات المتغير التوضيحي لأنموذج الانحدار في المعادلة (1) تنازليا. اما اذا كان أنموذج الانحدار المستخدم هو أنموذج انحدار متعدد فسوف يتم تحديد المتغير التوضيحي الذي يعتقد انه يسبب مشكلة عدم تجانس التباين ومن ثم ترتيب مشاهدات ذلك المتغير تنازليا.

2- تقسيم مشاهدات العينة الى جزئين متساويين في الحجم كل عينة جزئية لها نفس عدد المشاهدات مع اهمال قيم المشاهدات المركزية (c)، هذا يعني ان كل عينة جزئية سوف تحتوي على  $\left[ \frac{1}{2}(n - c) \right]$  من المشاهدات. حيث تضم العينة الجزئية الاولى المشاهدات ذات القيم الكبيرة بينما ستضم العينة الثانية المشاهدات ذات القيم الصغيرة. ويجب الاشارة الى ان قيمة المشاهدات المركزية (c) التي يتم اهمالها يكون اختيارها بشكل عشوائي على ان يكون عدد المشاهدات المهملة بين  $\left(\frac{1}{3}\right)$  و  $\left(\frac{1}{6}\right)$  من عدد المشاهدات الكلية.

3- يتم صياغة أنموذج انحدار لكل عينة جزئية وكالآتي:

$$Y_{i1} = \alpha_1 + \beta_1 X_{i1} + \epsilon_{i1} \quad (19)$$

$$Y_{i2} = \alpha_2 + \beta_2 X_{i2} + \epsilon_{i2} \quad (20)$$

إذ ان المعادلة (19) تمثل أنموذج الانحدار لبيانات العينة الجزئية الاولى. والمعادلة (20) تمثل أنموذج الانحدار لبيانات العينة الجزئية الثانية.

4- استعمال طريقة المربعات الصغرى الاعتيادية لتقدير معلمات أنموذج الانحدار في المعادلتين (19) و (20).

5- حساب مجموع مربعات الانحدار (RSS) لكل عينة جزئية.

6- يتم صياغة فرضية الاختبار وكالآتي:

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

7- حساب إحصاءة الاختبار (F) من الصيغة الآتية:

$$F = \frac{RSS_1}{RSS_2} \quad (21)$$

إذ ان (RSS<sub>1</sub>) يمثل مجموع مربعات الانحدار للعينة الجزئية الاولى. وان (RSS<sub>2</sub>) هو مجموع مربعات الانحدار للعينة الجزئية الثانية.

يتم رفض فرضية العدم اذا كانت قيمة إحصاءة الاختبار اكبر من قيمة (F) الجدولية بدرجة حرية تساوي  $\left[ \frac{1}{2}(n - c) - k, \frac{1}{2}(n - c) - k \right]$ ، وان (k) تمثل عدد المتغيرات التوضيحية لأنموذج الانحدار.

ان المشكلة في اختبار غولدفيلد-كوانت انه لا يأخذ في الاعتبار الحالات التي تحدث فيها مشكلة عدم تجانس التباين بسبب وجود أكثر من متغير توضيحي، كما انه ليس مناسب دائما لبيانات السلاسل الزمنية، إلا انه اختبار شائع جدا لنماذج الانحدار الخطية البسيطة.

## 5. الجانب العملي

يتضمن الجانب العملي نتائج اختبارات الكشف عن مشكلة عدم تجانس تباين الاخطاء العشوائية وطرائق التقدير الحصينة لمعاملات أنموذج الانحدار الخطي البسيط باستعمال (TLS)

و (MM و MPV). إذ تم تطبيق طرائق التقدير الحصينة تلك على بيانات حقيقية تمثل الكميات المنتجة من النفط الخام العراقي والكميات المصدرة منه للفترة (2008-2020). وتمت المقارنة بين طرائق التقدير بالاعتماد على معيار المفاضلة متوسط مربعات الخطأ (MSE).

### 1.5 وصف عينة البحث

تمثل بيانات عينة البحث المشاهدات الشهرية للكميات المنتجة من النفط الخام العراقي للفترة (2008-2020) وقد تم تمثيلها بالمتغير (P)، بينما تمثل المشاهدات الشهرية للكميات المصدرة من النفط الخام العراقي لنفس الفترة بالمتغير (E). إذ تكونت لدينا عينة عشوائية مؤلفة من (156) زوج من المشاهدات وقد تم الحصول على هذه البيانات من وزارة التخطيط / الجهاز المركزي للإحصاء / المجموعة الإحصائية السنوية. ولتحليل العلاقة بين الكميات المصدرة، المنتجة من النفط الخام العراقي تم صياغتها بأنموذج الانحدار الخطي البسيط وفق العلاقة الآتية:

$$(22)E_i = \alpha + \beta P_i + \epsilon_i$$

إذ ان ( $E_i$ ) يمثل الكميات المصدرة من النفط الخام وهو متغير الاستجابة ( $P_i$ ) يمثل الكميات المنتجة من النفط الخام وهو المتغير التوضيحي ( $\beta$ ) و ( $\alpha$ ) تمثل معاملات أنموذج الانحدار المطلوب تقديرها ( $\epsilon_i$ ) يمثل حد الخطأ العشوائي والذي تكون تبايناته غير متجانسة

### 2.5 الاختبارات المعتمدة في الكشف عن مشكلة عدم تجانس التباين

قبل اجراء عملية تقدير معاملات أنموذج الانحدار تم اجراء اختبارات الكشف عن مشكلة عدم تجانس التباين، إذ تم استعمال الاختبارين (هارفي وغولدفيلد-كوانت) الذي تم عرضها في المبحث الثاني إذ تم اختبار فرضية عدم تجانس التباين للاختبارين المذكورين سابقا، وأشارت النتائج الى معنوية إحصاءة الاختبار لكلا الاختبارين وكانت قيمة إحصاءة اختبار هارفي تساوي (44.26) بينما كانت قيمتها لاختبار غولدفيلد-كوانت تساوي (4.73) وكما موضح في الجدول رقم (1)

الجدول(1): نتائج اختبارات الكشف عن مشكلة عدم تجانس التباين

| Harvey-Godfrey |         | Goldfeld-Quandt |         |
|----------------|---------|-----------------|---------|
| HG             | P-value | F               | P-value |
| 44.26          | 0.0000  | 4.73            | 0.0000  |

من نتائج الجدول اعلاه نستنتج ان تباينات الاخطاء العشوائية لأنموذج الانحدار وفق المعادلة (22) غير متجانسة. وبالنظر لعدم تحقق فرضية تجانس تباين الاخطاء العشوائية، سوف يتم استعمال طرائق التقدير الحصينة التي تم ذكرها في الجانب النظري لتقدير معاملات أنموذج الانحدار موضوع البحث.

### 3.5 تقدير معاملات أنموذج الانحدار بطرائق التقدير الحصينة

بالاعتماد على الطرائق الحصينة التي تم توضيحها في الجانب النظري وبوجود مشكلة عدم تجانس تباين الاخطاء العشوائية سيتم تقدير معاملات أنموذج الانحدار الخطي البسيط الذي يصف العلاقة بين الكميات المصدرة من النفط الخام العراقي والكميات المنتجة منه.

### 1.3.5 طريقة تقدير المربعات الصغرى المشدبة

تم اجراء عملية تقدير معلمات أنموذج الانحدار باستعمال طريقة المربعات الصغرى المشدبة (TLS) بالاعتماد على اربعة نسب للتشذيب ( $\lambda_1 = 10\%$ ,  $\lambda_2 = 25\%$ ,  $\lambda_3 = 40\%$ ,  $\lambda_4 = 50\%$ ). والجدول (2) يعرض نتائج تقدير التقدير وفق طريقة التقدير هذه، إذ اشارت النتائج الى افضلية نسبة التشذيب (50%) إذ حققت اقل قيمة لمعيار المفاضلة متوسط مربعات الخطأ وكانت قيمته تساوي (1.62)، كما اشارت النتائج الى معنوية تأثير الكميات المنتجة من النفط الخام وكانت قيمة معامل الميل تساوي (0.82) وان قيمة معامل التحديد لهذه النسبة بلغت (99.17).

الجدول(2): نتائج تقدير المعلمات بطريقة المربعات الصغرى المشدبة

| نسبة التشذيب       | المعلمات | تقدير المعلمات | S.E. | t-value | P-value | R <sup>2</sup> | MSE   |
|--------------------|----------|----------------|------|---------|---------|----------------|-------|
| $\lambda_1 = 10\%$ | $\alpha$ | -1.04          | 1.10 | -0.94   | 0.3135  | 97.54          | 12.59 |
|                    | $\beta$  | 0.81           | 0.01 | 73.96   | 0.0000  |                |       |
| $\lambda_2 = 25\%$ | $\alpha$ | -1.57          | 0.97 | -1.63   | 0.1067  | 98.21          | 6.60  |
|                    | $\beta$  | 0.82           | 0.01 | 79.37   | 0.0000  |                |       |
| $\lambda_3 = 40\%$ | $\alpha$ | -1.97          | 0.86 | -2.29   | 0.0245  | 98.73          | 3.20  |
|                    | $\beta$  | 0.83           | 0.01 | 84.52   | 0.0000  |                |       |
| $\lambda_4 = 50\%$ | $\alpha$ | -1.25          | 0.73 | -1.71   | 0.0921  | 99.17          | 1.62  |
|                    | $\beta$  | 0.82           | 0.01 | 95.66   | 0.0000  |                |       |

### 2.3.5 طريقة تقدير (MM)

تم تقدير معلمات أنموذج الانحدار الخطي باستعمال طريقة تقدير (MM) بالاعتماد على دالتي الوزن (Huber و Tukey's Bisquare)، والجدول (3) يوضح نتائج التقدير لهذه الطريقة، إذ اشارت تلك النتائج الى معنوية تأثير الكميات المنتجة من النفط الخام لكلا الدالتين وكانت قيمة معلمة الميل لدالة (Tukey's Bisquare) تساوي (0.80) بينما كانت قيمتها لدالة (Huber) تساوي (0.75). وللمقارنة بين الدالتين تم الاعتماد على معيار المفاضلة (MSE) وكانت قيمته لدالة (Tukey's Bisquare) تساوي (32.90) مما يشير الى افضلية هذه الدالة، وكانت قيمة معامل التحديد لها تساوي (74.09).

الجدول(3): نتائج تقدير طريقة (MM)

| دالة الوزن | المعلمات | تقدير المعلمات | S.E. | t-value | P-value | R <sup>2</sup> | MSE   |
|------------|----------|----------------|------|---------|---------|----------------|-------|
| Tukey      | $\alpha$ | 0.04           | 1.31 | 0.03    | 0.9767  | 74.09          | 32.90 |
|            | $\beta$  | 0.80           | 0.01 | 62.85   | 0.0000  |                |       |
| Huber      | $\alpha$ | 3.45           | 1.07 | 3.22    | 0.0013  | 42.67          | 44.09 |
|            | $\beta$  | 0.75           | 0.01 | 72.12   | 0.0000  |                |       |

### 3.3.5 طريقة تقدير (MPV)

تم استعمال طريقة (MPV) لتقدير معلمات أنموذج الانحدار الخطي البسيط بوجود مشكلة عدم تجانس التباين. وقد اشارت النتائج الى معنوية تأثير الكميات المنتجة من النفط الخام وكانت قيمة معلمة الميل تساوي (0.83)، وكانت قيمة معامل التحديد لهذه الطريقة تساوي (88.95) وان قيمة متوسط مربعات الخطأ تساوي (7.40).

الجدول(4): نتائج تقدير طريقة (MPV)

| المعاملات | تقدير المعاملات | S.E. | t-value | P-value | R <sup>2</sup> | MSE  |
|-----------|-----------------|------|---------|---------|----------------|------|
| $\alpha$  | -2.44           | 0.91 | -2.68   | 0.0082  | 88.95          | 7.40 |
| $\beta$   | 0.83            | 0.01 | 70.93   | 0.0000  |                |      |

#### 4.5 اختيار افضل طريقة تقدير حصينة

الجدول (5) يبين نتائج التقدير المثلى لكل طريقة تقدير، و اشارت النتائج الى افضلية طريقة المربعات الصغرى المشدبة (TLS) بنسبة تشذيب (50%) إذ حققت اقل قيمة لمعيار المفاضلة متوسط مربعات الخطأ وكانت قيمته تساوي (1.62)، تلتها طريقة (MPV) وكانت قيمة متوسط مربعات الخطأ لها تساوي (7.40)، في حين حققت طريقة (MM) اكبر قيمة لمعيار المفاضلة والبالغة (32.90).

الجدول(5): نتائج التقدير الحصينة المثلى

| طريقة التقدير              | R <sup>2</sup> | MSE   |
|----------------------------|----------------|-------|
| TLS ( $\lambda_4 = 50\%$ ) | 99.17          | 1.62  |
| MM (Tukey)                 | 74.09          | 32.09 |
| MPV                        | 88.95          | 7.40  |

مما ورد اعلاه نستنتج ان طريقة (TLS) هي افضل طريقة حصينة لتقدير معاملات أنموذج الانحدار. وان الأنموذج المقدر له كالتالي إذ بلغت قدرته التفسيرية (99.17):

$$\hat{Y}_i = -1.25 + 0.82 X_i \quad (22)$$

#### 6. الاستنتاجات والتوصيات

##### 1.6 الاستنتاجات

من خلال ما تم التوصل اليه من نتائج تقدير معاملات انموذج الانحدار موضوع البحث باستعمال طرائق التقدير المستعملة، فقد تم التوصل الى الاستنتاجات الآتية:

- 1- ان نسبة التشذيب (50%) لطريقة المربعات الصغرى المشدبة كانت افضل نسبة تشذيب معتمدة لتحقيقها اقل قيمة لمتوسط مربعات الخطأ.
- 2- ان دالة الوزن (Tukey's Bisquare) كانت افضل من دالة الوزن (Huber) لطريقة التقدير (MM).

- 3- طريقة المربعات الصغرى المشدبة كانت افضل طريقة حصينة لتقدير معاملات أنموذج الانحدار.

##### 2.4 التوصيات

- 1- استعمال طرائق تقدير حصينة اخرى لتقدير معاملات انموذج الانحدار بوجود مشكلة عدم تجانس تباين الاخطاء العشوائية.
- 2- استعمال طرائق التقدير موضوع البحث لتقدير معاملات أنموذج الانحدار لبيانات اقتصادية اخرى.
- 3- استعمال طرائق التقدير اللامعلمية ومقارنتها مع طرائق التقدير الحصينة موضوع البحث.

## المصادر

- [1]. السواعي، خالد محمد. (2018). مبادئ الاقتصاد القياسي. عمان، الاردن: دار الكتاب الثقافي.
- [2]. بخيت، حسين علي ، و فتح الله، سحر. (2008). الاقتصاد القياسي. عمان، الاردن: دار اليازوري العلمية للنشر والتوزيع.
- [3]. بندر، علي قاسم محمد. (2018). تقدير انموذج الانحدار باستخدام طريقة المربعات الصغرى الموزونة الحصينة في حالة وجود مشكلة عدم تجانس التباين والقيم الشاذة مع تطبيق عملي. مجلة الادارة والاقتصاد، العدد 117، الصفحات 285-299.
- [4]. حسن، تارا احمد ، ورضا، مهدي صابر. (2011). استخدام الانحدار الحصين لإيجاد أنسب نموذج لتمثل بيانات الأنواء الجوية في مدينة اربيل خلال الفترة (1998-2010). مجلة الادارة والاقتصاد، العدد التاسع والثمانون، الصفحات 195-213.
- [5]. حميد، زينب حامد. (2018). مقارنة بين الطرائق الاعتيادية (LS,4SIV) والطرائق الحصينة (LTS,2SLWLS,RA) لتقدير معلمات أنموذج (ARX(1,1,1)) للاحمال الكهربائية. مجلة العلوم الاقتصادية والادارية، العدد 109، الصفحات 496-514.
- [6]. عبدالله، رائد اسمر. (2018). إختيار أفضل طريقة اختبار لمشكلة عدم تجانس التباين في نموذج الانحدار المتعدد (مع تطبيق عملي). رسالة ماجستير، كلية الادارة والاقتصاد جامعة كربلاء، قسم الاحصاء
- [7]. علي، سيف حسين. (2019). اختيار افضل تقدير معالم إنموذج الانحدار الحصين مع تطبيق عملي. رسالة ماجستير، كلية الادارة والاقتصاد جامعة كربلاء، قسم الاحصاء.
- [8]. كاظم، اموري هادي. (2009). مقدمة في القياس الاقتصادي. عمان، الاردن: دار زهران للنشر والتوزيع.
- [9]. كاظم، خالد جمال. (2018). المربعات الصغرى المشدبة الموزونة لتقدير تأثير مياه الصرف الصحي في تلوث مياه نهر دجلة/محافظة واسط. مجلة العلوم الاقتصادية والادارية، العدد 109، الصفحات 486-495.
- [10]. Almetwally, E. M., & Almongy, H. M. (2018). Comparison Between M-Estimation, S-Estimation, and MM-Estimation Methods of Robust Estimation with Application and Simulation. International Journal of Mathematical Archive, 9(11), pp. 55-63.
- [11]. Asteriou, Dimitrios & Hall, Stephen G. (2021). Applied Econometrics. (Forth Edition). London, England, Uk: Macmillan Education Limited. ISBN 978-1-353-01202-6 paperback.
- [12]. kmenta, J. (1986). Elements of Econometrics (Vol. Second Edition). New York: Macmillan Inc.
- [13]. MIDI, H., RANA, S., & IMON, A. R. (2009). The Performance of Robust Weighted Least Squares in the Presence of Outliers and Heteroscedastic Errors. WSEAS Transactions on Mathematics, vol. 8(issue 7), pp. 351-361.
- [14]. Montgomery, D. C., Peck, E. A., & Vining, G. G. (2006). Introduction to Linear Regression Analysis (4th ed.). Hoboken, New Jersey: John Wiley & Sons, Inc.
- [15]. Susanti, Y., Pratiwi, H., Sulistijowati, S., & Liana, T. (2014). M Estimation, S Estimation, and MM Estimation in Robust Regression. International Journal of Pure and Applied Mathematics, Vol. 91(No. 3), pp. 349-360. doi: <http://dx.doi.org/10.12732/ijpam.v91i3.7>.
- [16]. Zaman, A., Rousseeuw, P. J., & Orhan, M. (2001). Econometric applications of high-breakdown robust regression techniques. Economic Letters, 71, pp. 1-8. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0165176500004043>.



## تصنيف مرضى التهاب الكبد باستعمال الدالة التمييزية

مآب نوري حسن  
جامعة بغداد / كلية الادارة والاقتصاد / قسم  
الأحصاء

ابتسام كريم عبد الله  
جامعة بغداد / كلية الادارة والاقتصاد / قسم  
الأحصاء

تاريخ استلام البحث: 2022/07/13

تاريخ قبول البحث: 2022/08/15

نشر البحث في العدد عشرون: سبتمبر / ايلول 2023

رمز التصنيف ديوي / النسخة الالكترونية (Online): 2522-64X/512:616.3

رمز التصنيف ديوي / النسخة الورقية (Print): 2519-948X/512:616.3



## تصنيف مرضى التهاب الكبد باستعمال الدالة التمييزية

مآب نوري حسن  
جامعة بغداد / كلية الادارة والاقتصاد / قسم  
الأحصاء

ابتسام كريم عبد الله  
جامعة بغداد / كلية الادارة والاقتصاد / قسم  
الأحصاء

### مستخلص البحث

يهدف البحث الى إنشاء دالة تمييزية تقوم بتحديد العوامل المؤثرة في الأصابة بأختلال انزيمات الكبد، من خلال التمييز والفصل بين المصابين إلى مجموعتين (الذكور، الأناث) ودراسة عدد من العوامل المؤثرة (Age ,weight , ALP, AST ,ALT ,GGT) عن طريق أخذ عينة من 100 شخص وبعد التأكد من ملائمة التحليل التمييزي للبيانات وتكوين دالة تمييزية للفصل والتمييز

### Abstract

The research aims to create a discriminatory function that determines the factors affecting the incidence of liver enzyme abnormalities, by distinguishing and separating the sufferers into two groups (males, females) and studying a number of influencing factors (Age, weight, ALP, AST, ALT, GGT) by taking A sample of 100 people and after ensuring the appropriateness of the discriminatory analysis of the data and the formation of a discriminatory function for segregation and discrimination.

### مقدمة (Introduction):

أدى تطور الأبحاث ووجود العلوم الحديثة المختلفة الى وجود تعدد في المتغيرات سواء أكانت نوعية أم كمية والتي ادت الى ظهور بيانات متعددة, وان وجود هذه المتغيرات يتطلب من الباحثين استعمال التحليل الأحصائي المتعدد المتغيرات والذي يقدم للباحثين طرائق احصائية لتحليل واختبار تلك البيانات والتنبؤ بها ويعد التحليل التمييزي نوع من انواع التحليل الأحصائي متعدد المتغيرات [12]. التي تستعمل في معالجة البيانات الوصفية ويعتمد على بناء دالة تسمى دالة التمييز الخطية وهي عبارة عن توليفة خطية لمجموعة من المتغيرات المستقلة وهذه الدالة تعمل على تقليل التشابه في أخطاء التصنيف ويهدف التحليل التمييزي إلى تصنيف المشاهدات إلى مجموعاتها الصحيحة بأقل خطأ تصنيف ممكن، و يختلف أسلوب التمييزي مع تحليل التباين وتحليل الانحدار و المتغير التابع نوعي، بينما في الاسلوبين الاخيرين يكون تغير كمي، كما يتشابه التحليل التمييزي مع الانحدار اللوجستي إذ يفسر كل منهما متغير وصفي الا ان الانحدار اللوجستي لا يطلب ان تكون المتغيرات المستقلة تتبع التوزيع الطبيعي. [2] ويستعمل التحليل التمييزي في مجالات متعددة فعلى سبيل المثال يوضف التحليل التمييزي في مجالات الطب والزراعة والتعليم وعلم الاجتماع وغيرها من المجالات التطبيقية. ومن الناحية الطبية فإن التطرق الى أنزيمات الكبد هو من الدراسة المهمة وأن الكبد من الاعضاء الاساسية في جسم الانسان وان اي زيادة او نقصان في انزيمات الكبد تشير إلى وجود التهاب أو

تلف في خلايا الكبد، وتسبب خلاياه الملتهبة أو المصابة أعلى من الكميات الطبيعية من بعض المواد الكيميائية إلى مجرى الدم، مما قد يؤدي إلى ارتفاع إنزيمات الكبد في اختبارات الدم و يؤدي ارتفاع احد الأنزيمات الى اختلال في عمل وظائف الكبد ويتم في هذا البحث دراسة هذه الانزيمات ومدى تأثيرها من خلال اصابة الكبد بالالتهاب أو تلف خلاياه. الدراسات السابقة

- في العام 2015م قامت الباحثة الاء عبد الستار حمودات<sup>[6]</sup> بدراسة الدالة التمييزية وطرق تحديد متغيراتها ولغرض التوصل إلى معلمات النموذج المستعمل لوصف المجتمعين هي ( $Hotelling T^2$ ) وجدول تحليل التباين وهذا يدل على أن الدالة التمييزية ذات كفاءة في التمييز وكذلك في معرفة معنوية المتغيرات الداخلة واستعمال طريقة روي بوس لحدود الثقة وأختبار t.
- في عام 2008م قامت الباحثة عائدة<sup>[13]</sup> بكتابة مقال عن استعمال التحليل المميز لتشخيص بعض أمراض العيون و تم استعمال التحليل المميز لتصنيف بعض أمراض العيون إلى ثلاث مجموعات إعتما داً على متغيرات ذات صفات تمييزية . و تم استعمال المقاييس التالية مقياس مربع كاي ومقياس F ومقياس ولكس، وباستعمال الدالة المميزة الخطية تم تصنيف بعض أمراض العيون على أساس مجموعة من المتغيرات المصاحبة للمرض.
- في عام 2009م قام الباحث بدر<sup>[14]</sup> بدراسة عن استعمال بعض طرائق التمييز الحصينة لتشخيص مرضى سرطان الدم اللوكيميا عن طريق ايجاد مقدرات حصينة للمتوسطات ودالة التباين والتباين المشترك للدوال الخطية والتربيعية ، وكذلك ايجاد دوال خطية وتربيعية حصينة في حالة وجود قيم شاذة باستعمال طرق تحصين ضد وجود الشواذ في البيانات . ومن هذه الطرق طريقة اصغر محدد تباين مشترك وطريقة اصغر قطع بيضوي
- في عام 2018م قامت الباحثة شهد نجم عبد الله<sup>[15]</sup> بدراسة حول أستعمال التحليل التمييزي في تصنيف درجة التشوه للولادات ، وأن الدراسة شملت ثلاث مجموعات هي التشوه البسيط والتشوه المتوسط والخطير . وقد اظهرت المتغيرات ان هنالك تأثير معنوي كبير لمتغير التدخين والمتغير R و عدد الاجهازات .
- في عام 2014م قامت الباحثة بسمة<sup>[2]</sup> بدراسة بعنوان تمييز الكادر الطبي حسب معرفتهم للتصنيف الدولي (ICD-10) باستعمال الدالة التمييزية ، وطريقة الرتب لأستخراج الدالة التمييزية لمعرفة الاطباء الذين لديهم المام بالتصنيف الدولي للأمراض وتمت المقارنة بين الدالتين بالأعتماد على خطأ التصنيف.

#### مراحل اختيار الدالة التمييزية:

#### أ. مرحلة الاشتقاق Derivation

تتضمن هذه المرحلة عدة خطوات أبتداء من اختيار المتغيرات وتقسيم العينة ثم اجراء الاختبارات الاحصائية. بالنسبة إلى خطوة اختيار المتغيرات، تتضمن تحديد ما هي المتغيرات التفسيرية و من هو المتغير التابع، الخطوة الثانية هي تقسيم العينة. في حين تتضمن الخطوة الثالثة أسلوبين إجراء التحليل التمييزي وهو أسلوب المباشر وأسلوب الثاني هو أسلوب الخطوات (stepwise)، وسيتم الأعتما د في هذا البحث على هذا الأسلوب و يكون لدينا عدد كبير من المتغيرات التفسيرية للوصول الى أكثر المتغيرات معنوية للتمييز بين المشاهدات<sup>[6]</sup>

**ب. مرحلة التحقق Validation**

في هذه المرحلة يتم الاعتماد على أكثر من مؤشر لتقييم مدى دقة المعادلة التمييزية، فإذا كانت معنوية ينصب الهدف على تحديد نقطة القطع (cut off value)، ويتم اعتماد هذه الدرجة كمقياس لمعرفة ما إذا كانت كل مفردة تعود إلى إحدى المجموعات.

**ت. مرحلة التفسير**

إذا كانت الدالة التمييزية معنوية إحصائياً والتصنيف مقبولاً يمكن الاستمرار في هذه المرحلة، وأنها تتضمن خطوتين مهمتين، الخطوة الأولى تكون تتضمن الدوال المميزة لتحديد الأهمية النسبية لكل متغير مستقل في التمييز بين المجاميع. أما الخطوة الثانية تكون اختبار متوسطات المشاهدات في تحليل التباين الأحادي لكل متغير توضيحي ذي معنوية إحصائية في الدالة المميزة.

**الانحدار التدريجي (Stepwise-regression)**

وسيتيم استعمال الأنحدار التدريجي في هذا البحث إذ تكون أهمية الأنحدار التدريجي في فحص الأهمية الإحصائية بصورة متكرره لكل المتغيرات وبشكل مستقل

**خطوات الانحدار التدريجي**

يتم استعمال المتغيرات التفسيرية أثناء إنشاء دالة التحليل التمييزي بالترتيب و أن ترتيب المتغيرات فيه يكون وفق معيار احصائي و يتم البدء بأكثر المتغيرات تميزاً وهو من أشهر الاساليب الاحصائية<sup>[7]</sup>

أ- نأخذ المتغير الأكثر معنوية إلى الانموذج، وذلك بعد أن نختبر فيما إذا كان غير معنوياً، فإذا كان غير معنوي يتم التوقف.

ب- تختبر جميع المتغيرات في الانموذج فإذا كان منها متغيرات غير معنوية تحذف من الانموذج.

ج- إعادة الخطوات أ و ب إلى أن يتم التوقف ويلاحظ في الخطوة (أ) إن اختيار المتغير الأكثر معنوية يكون أنياً وليس متفرداً<sup>[9]</sup>.

**تقدير حدود الثقة لروي بوس (Roy-Bose Confidence Interval Estimation):**

يتم استعمال حدود الثقة لروي- بوس لتعيين المتغيرات المهمة التي ستدخل في الدالة، ويمكن استعمال هذه الطريقة بالخطوات التالية:

1- نجد قيمة F الجدولية بالرجوع إلى جداول خاصة وبدرجة حرية:

$$Tab'F = F(\alpha/2, m, n1 + n2 - 1)$$

2- ايجاد قيمة T الجدولية:

$$TabT = \left[ \frac{(n_1+n_2-2)m}{n_1+n_2-m-1} tab \cdot F \right]^2$$

3- إيجاد المتجه (Vector Selection) ويرمز له بالرمز a وهو متجه غير صفري ويكون عدد هذه المتجهات بعدد المتغيرات التوضيحية

$$\begin{aligned} a_1 &= [1 \ 0 \ 0 \ \dots] \\ a_2 &= [0 \ 1 \ 0 \ 0 \ \dots] \\ &\vdots \\ \dots \dots a_m &= [0 \ 0 \ \dots \dots 1] \end{aligned}$$

4- ثم بعد ذلك نجد حدود الثقة ( Roy-Bose ) وتكتب بالصيغة التالية:

$$\underline{ak}(\bar{X}_1 - \bar{X}_2) - \sqrt{\underline{ak} \sum a_k \left[ \frac{n_1 + n_2}{n_1 \cdot n_2} \right]} T < \underline{ak} < \underline{ak}(\bar{X}_1 - \bar{X}_2) + \sqrt{\underline{ak} \sum a_k \left[ \frac{n_1 + n_2}{n_1 \cdot n_2} \right]}$$

5- ثم نكتب المتباينة

$$C.I(L) < a'm < C.I$$

والقرار هو إذا احتوت المتباينة على الصفر يعني عدم وجود أهمية لذلك المتغير أي أن المتغير يختلف وسطه الحسابي في كلا المجموعتين ، إما إذا لم تحتوي المتباينة على الصفر يعني وجود أهمية لذلك المتغير.

### نقطة القطع (Cut Off Value)

بعد تكوين تصنيف الدالة التمييزية واختبار قدرتها على التمييز والفصل بين المجموعتين يبدأ الاستعمال الثاني لها وهو كيفية تصنيف المشاهدة الجديدة إلى أي المجموعتين تنتمي. أي ان لتصنيف الأفراد فإننا نحتاج إلى نقطة تفصل بين المجموعتين وإذا قلت ( مجموعة نقاط الأفراد ) عن تلك النقطة فإن الفرد عندئذ يتبع إلى مجموعة معينة وإذا زادت مجموع نقاطه عن تلك النقطة الفاصلة أصبح تابعاً للمجموعة الأخرى. ويتم ذلك من خلال الخطوات الآتية [11]:

أ. تحديد نقطة التصنيف وهي تمثل متوسط المتوسطين:

$$\bar{y} = \frac{\bar{y}_1 + \bar{y}_2}{2}$$

و ان :

$\bar{y}$ : نقطة التصنيف.

$\bar{y}_1$ : متوسط القيم التمييزية للمجموعة الأولى.

$\bar{y}_2$ : متوسط القيم التمييزية للمجموعة الثانية.

### • اختبار كفاءة دالة التمييز الخطي Testing of Linear Discriminate Function

عندما يراد التمييز بين مجموعتين فإنه يمكننا إختبار الفرضية التي تنص على تساوي متوسط المجموعات.

$$H_0: b_1 = b_2 = \dots = b_k$$

$$H_1: b_1 \neq b_2 \neq \dots \neq b_k$$

وبذلك فإن اختبار الدالة المميزة تعتمد على رفض أو قبول فرضية العدم  $H_0$ .

وان إحصاءة الاختبار المستعملة في حالة التمييز بين مجموعتين هي إحصاءة الاختبار (F) والتي تعتمد على إحصاءة الاختبار هوتلنك (Hotelling) وصيغته الرياضية كالآتي [7]:

$$F = \frac{n_1 + n_2 - p - 1}{(n_1 + n_2 - 2)p} T^2$$

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} D_i^2$$

و ان :-

$$D_i^2 = (\bar{X}_{i1} - \bar{X}_{i2})' S^{-1} (\bar{X}_{i1} - \bar{X}_{i2})$$

$D_i^2$  :- يمثل إحصاءة (Mahalanobis) (مهالنوبيس)

p: عدد المتغيرات

وبهذا نرفض فرضية العدم  $H_0$ ، عندما تكون  $F$  (المحسوبة) اكبر من  $F$  (الجدولية) ونقبل الفرضية البديلة  $H_1$ ، بمستوى معنوية  $\alpha$  وبدرجة حرية  $(n_1 + n_2 - p - 1)$ .

$$F > F_{\alpha(p, n_1+n_2-p-1)}$$

وهذا يدل على إن متوسطات المجموعتين غير متساوية إي يوجد فروق معنوية بين المجموعتين بمعنى إن الدالة المميزة الخطية قابلة للتمييز بدرجة عالية.

### اختبار تساوي مصفوفة التباين والتباين المشترك لجميع المجاميع:

يستعمل هذا الاختبار لمعرفة النوع الملائم من النماذج لتمثيل دالة التميز بين المجاميع. وتكون فرضيتي العدم والبديلة بالشكل التالي [13]:

$$H_0: \sum_1 = \sum_2 = \dots \sum_k \dots$$

يوجد على الاقل اثنان منهم غير متساويان:  $H_1$   
فعندما تكون فرضية العدم  $H_0$  صحيحة فأن [20]:

$$S = \frac{\sum_{i=1}^k n_i s_i}{\sum_{i=1}^k n_i}$$

وأن:

$S$  هو تقدير غير متحيز ل  $\sum$  ، وبذلك فأن  $S$  تمثل تقدير مصفوفة التباين والتباين المشترك وتكون إحصاء الاختبار المستعملة في هذا كما يلي:

$$M = (\sum_{i=1}^k n_i) Ln |S| - \sum_{i=1}^k (n_i Ln |S|)$$

وقد اثبت Box عام (1949) م انه إذا ضرب  $M$  في ثابت  $C-1$  والذي يساوي [8]:

$$C^{-1} = 1 - \frac{2p^2+3(p-1)}{6(p+1)(k-1)} \left[ \sum_{i=1}^k \frac{1}{n_i} - \frac{1}{\sum_{i=1}^k n_i} \right]$$

نتوصل بذلك إلى مقياس يتوزع بالتقريب توزيع  $(\chi^2)$  وخاصة عندما يكون حجم العينة كبير وبدرجة حرية  $(k-1)(p-1)$

$$MC^{-1} \sim \chi^2 \text{ with } (k-1)(p-1)$$

وإن:

$p$ : عدد المتغيرات المدروسة في المصفوفة .

$k$ : عدد المجاميع.

### أخطاء التصنيف (Classification Error)

تعرف أخطاء التصنيف بأنه احتمال تصنيف مفردة معينة إلى المجموعة الأولى وهي في الحقيقة تعود للمجموعة الثانية وبالعكس وعند حساب خطأ التصنيف نأخذ بعين الاعتبار أنه كلما كان حجم العينة كبيراً فأن توزيع المفردات يقترب من التوزيع الطبيعي وهذا الأمر مهم في حساب احتمال اخطاء التصنيف. [5]

ويقسم خطأ التصنيف الى نوعين:

أ- خطأ التصنيف الظاهري:

ويتم ايجاد الخطأ الظاهري كما يلي:

$$p_{12} = \frac{n_{12}}{n_1}$$

$p_{12}$  : نسبة المفردات التي تنتمي للمجموعة الأولى وصنفت خطأً للثانية.

$$p_{21} = \frac{n_{21}}{n_2}$$

$p_{21}$  : نسبة المفردات التي تنتمي للمجموعة الثانية وصنفت خطأً في الأولى،  
و يتم الحساب عن طريق الجدول التالي :

جدول (1) خطأ التصنيف الظاهري

| المجموعة | مشاهدات المجموعة الأولى | مشاهدات المجموعة الثانية | المجموع |
|----------|-------------------------|--------------------------|---------|
| الأولى   | $n_{11}$                | $n_{12}$                 | $n_1$   |
| الثانية  | $n_{21}$                | $n_{22}$                 | $n_2$   |

$n_{11}$  : عدد المفردات من المجموعة الأولى والتي تم تصنيفها في نفس المجموعة وبالتالي هي صنفت بطريقة صحيحة.

$n_{12}$  : عدد المفردات من المجموعة الأولى والتي تم تصنيفها خطأً في المجموعة الثانية .  
 $n_{21}$  : عدد المفردات التي تنتمي بالأصل إلى المجموعة الثانية وتم تصنيفها خطأً في المجموعة الأولى .  
 $n_{22}$  : عدد المفردات في المجموعة الثانية التي تم تصنيفها في نفس المجموعة وهي صنفت بطريقة صحيحة.

ويمكن حساب معدل الخطأ الظاهري باستعمال المعادلة ادناه :

$$\frac{n_{12} + n_{21}}{n_1 + n_2}$$

### الخطأ الحقيقي

يمثل نسبة التصنيف الخاطئ في المجتمع [4]

$$P_{12} = P_{21} = L \left[ \frac{-\sqrt{D^2}}{2} \right]$$

و ( $L$ ) دالة التوزيع الطبيعي المعياري،  $D$  أحصائية (Mahalanobis) تحسب القيمة بين القوسين ويحسب الأختمال المقابل لها من جدول التوزيع الطبيعي المعياري وكلما اقترب الأختمال من الصفر دل على نوع وانخفاض في خطأ التصنيف وبالتالي قدرة الدالة على التصنيف والتمييز اما إذا كان الأختمال قريب من الواحد يدل على ارتفاع خطأ التصنيف وانخفاض قدرة الدالة على التصنيف والتمييز ويعتبر خطأ التصنيف عامل مهم عند الحكم على كفاءة الدالة التمييزية. [2]

الدالة التمييزية بمعاملات غير معيارية [3] - :

تأخذ الشكل التالي:

$$y = b_0 + b_1x_1 + b_2x_2 \dots \dots \dots b_kx_k$$

( $Y$ ) الدالة التمييزية غير المعيارية.

( $b_0$ ) ثابت التمييز.

( $b_n$ 's) معاملات التمييز غير المعيارية.

( $x_n's$ ) قيم المتغيرات غير المعيارية. وللحكم على جودة الانموذج التمييزي من خلال معامل الارتباط القانوني Canonical correlation وان القيم المرتفعة لمعامل الارتباط القانوني تكون مؤشر للجودة العالية للأنموذج التمييزي وبتربيع قيمة عامل الارتباط القانوني تحصل على قيمة معامل التحديد  $R^2$  الذي يحدد نسبة مساهمة المتغيرات المستقلة في التصنيف والتمييز [18].

### أختبار قدرة الدالة على التمييز والفصل بين المجموعات:

تستعمل الأختبارات الآتية :

اختبار  $F$  ( $F - TEST$ )

وذلك لأختبار قدرة الدالة على التمييز وعن طريق الفرضية التي تنص على ان الدالة ليس لديها القدرة على التمييز  $H_0$  ضد دالة لديها القدرة على التمييز ( $H_1$ ) ويعتمد هذا الأختبار على قياس الاختلافات بين المجموعات وداخل المجموعات بين المفردات ويتم ذلك من خلال جدول تحليل التباين (ANOVA Table) الآتي [7]:

جدول(2) ANOVA Table

| Source of Variation | $SS$  | $df$    | $MSS$    | $F$                     |
|---------------------|-------|---------|----------|-------------------------|
| بين المجموعات $x's$ | $SSB$ | $k - 1$ | $M_{SB}$ | $\frac{M_{SB}}{M_{SE}}$ |
| الخطأ $x's$ Within  | $SSE$ | $n - k$ | $M_{SE}$ |                         |
| الكل Total          | $SST$ | $n - 1$ |          |                         |

وان :-

1. مجموع مربعات الاخطاء يحسب كالتالي<sup>[10]</sup>  

$$SSE = D^2 = \hat{\alpha}_1 d_1 + \hat{\alpha}_2 d_2 + \dots + \hat{\alpha}_k d_k$$
2. مجموع مربعات بين المتغيرات يحسب كالتالي:  

$$SSB = \frac{n_1 n_2}{(n_1 + n_2)(n_1 + n_2 - 2)} \times (D^2)^2$$
3. مجموع المربعات الكلي يحسب كالتالي:  

$$SST = SSB + SSE$$

ويتم الأختبار كالتالي:-

1- صياغة الفروض:

الدالة ليس لها القدرة على التمييز:  $H_0$

الدالة لها القدرة على التمييز:  $H_1$

2- القيمة المحسوبة:

$$F = \frac{M_{SB}}{M_{SE}}$$

3- القيمة الجدولية:

$$F_{\alpha(k-1, n-k)}$$

### 4- القرار:

إذا كانت  $F$  المحسوبة أكبر من  $F$  الجدولية نرفض الفرض العدمي ونقبل بالفرض البديل ويكون للدالة قدرة عالية على التمييز والعكس صحيح.

## جمع البيانات:

جمعت البيانات الخاصة بموضوع البحث من مختبرات التعليمية لمدينة الأمامين الكاظمين الطبية واحدى مختبرات المستشفيات الاهلية، وقد تم الاستعانة بمجموعة من المختصين بامراض الكبد للوقوف على اهم العوامل المؤثرة اذا تم جمع البيانات الخاصة لسنة 2021م وبعينة حجمها 100 مراجع، ولغرض تحليل البيانات فقد تم اعتبار متغير الاستجابة (y) هو (y=1) ذكور (y=2) أناث وكان عدد الاهلي 50 والحكومي 50 بالاعتماد على المتغيرات التفسيرية.

- (Age) متغير يمثل العمر
- (Weight) متغير يمثل الوزن
- (ALT) متغير يمثل انزيم ناقل امين ألانين (Alanine transaminase) يتركز هذا الانزيم بشكل رئيسي في الكبد وهو يحول البروتينات الى طاقة لذلك فإن اي خلل في النسب الطبيعية لهذا الأنزيم يدل على اصابة الكبد.
- (AST) متغير يمثل أنزيم ناقل أمين السباتات (Aspartate transaminase)
- وهو انزيم يوجد في العديد من اعضاء الجسم ويساعد في استقلاب الاحماض الأمينية.
- (ALP) متغير يمثل أنزيم الفوسفاتيز القلوي (Alkaline phosphatase) يوجد هذا الأنزيم بصورة اساسية في الكبد والذي يعمل على تحلل البروتين .
- (GGT) متغير يمثل أنزيم ناقلة الببتيد (Gamma-glutamyl transpeptidase) وهو احد الأنزيمات التي يتم تصنيعها في الكبد وتتركز بشكل رئيسي في الدم. وتدل الزيادة او النقصان في هذه الأنزيمات الى العديد من المشكلات مثل التهاب الكبد الحاد او المزمن او تليف الكبد او أمراض الغدة الدرقية.

## تحليل البيانات:

تتم عملية تحليل البيانات باستعمال البرنامج الاحصائي (SPSS) وتتم مرحلة التحليل بثلاث

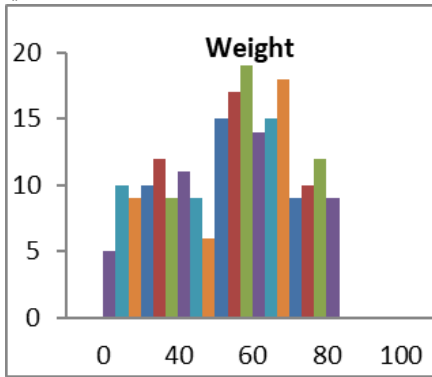
| جدول (3): اختبار Kolmogorov-Smirnov      |                         |                    |                    |                    |        |                    |        |   |
|--|-------------------------|--------------------|--------------------|--------------------|--------|--------------------|--------|---|
| One Kolmogorov-Smirnov Test              |                         |                    |                    |                    |        |                    |        |   |
|  |                         | Age                | Weight             | ALP                | AST    | ALT                | GGT    |   |
| Most Extreme Differences                 | Absolute                | 0.060              | 0.067              | 0.062              | 0.089  | 0.0                | 0.203  |   |
|  | Positive                | 0.053              | 0.067              | 0.062              | 0.040  | 0.043              | 0.192  |   |
|  | Negative                | -0.060             | -0.064             | -0.054             | -0.089 | -0.045             | -0.203 |   |
| Test Statistic                           |                         | 0.060              | 0.067              | 0.062              | 0.089  | 0.045              | 0.203  |   |
| Asymp. Sig. (2-tailed) <sup>c</sup>      |                         | 0.200 <sup>d</sup> | 0.200 <sup>d</sup> | 0.200 <sup>d</sup> | 0.048  | 0.200 <sup>d</sup> | 0      |   |
| Monte Carlo Sig. (2-tailed) <sup>e</sup> | Sig.                    | 0.501              | 0.330              | 0.437              | 0.052  | 0.889              | 0      |   |
|  | 99% Confidence Interval | Lower Bound        | 0.488              | 0.318              | 0.424  | 0.046              | 0.881  | 0 |
|  |                         | Upper Bound        | 0.514              | 0.342              | 0.450  | 0.058              | 0.897  | 0 |



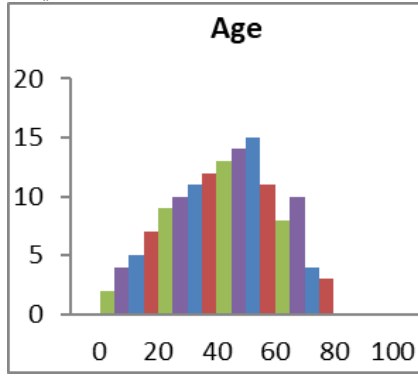
التأكد من توفر شروط التحليل التمييزي عن طريق:  
ايجاد التوزيع الطبيعي يتم ايجاد هل أن البيانات تتبع التوزيع الطبيعي عن طريق اجراء اختبار  
Kolmogorov-Smirnov

| One Kolmogorov-Smirnov Test              |                         |                    |                    |                    |        |                    |        |   |
|--|-------------------------|--------------------|--------------------|--------------------|--------|--------------------|--------|---|
|  |                         | Age                | Weight             | ALP                | AST    | ALT                | GGT    |   |
| Most Extreme Differences                 | Absolute                | 0.060              | 0.067              | 0.062              | 0.089  | 0.0                | 0.203  |   |
|  | Positive                | 0.053              | 0.067              | 0.062              | 0.040  | 0.043              | 0.192  |   |
|  | Negative                | -0.060             | -0.064             | -0.054             | -0.089 | -0.045             | -0.203 |   |
| Test Statistic                           |                         | 0.060              | 0.067              | 0.062              | 0.089  | 0.045              | 0.203  |   |
| Asymp. Sig. (2-tailed) <sup>c</sup>      |                         | 0.200 <sup>d</sup> | 0.200 <sup>d</sup> | 0.200 <sup>d</sup> | 0.048  | 0.200 <sup>d</sup> | 0      |   |
| Monte Carlo Sig. (2-tailed) <sup>e</sup> | Sig.                    | 0.501              | 0.330              | 0.437              | 0.052  | 0.889              | 0      |   |
|  | 99% Confidence Interval | Lower Bound        | 0.488              | 0.318              | 0.424  | 0.046              | 0.881  | 0 |
|  |                         | Upper Bound        | 0.514              | 0.342              | 0.450  | 0.058              | 0.897  | 0 |

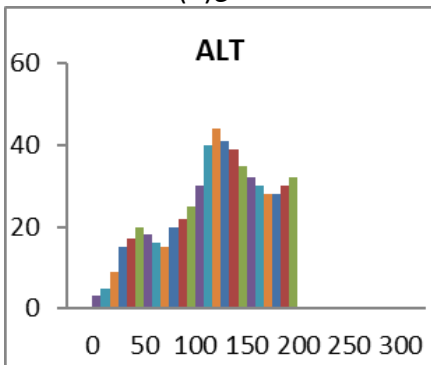
وتوضح نتائج الجدول (3) من المخرجات اعلاه ان البيانات المجمعة في كل من المتغيرات الستة تتبع التوزيع الطبيعي وان مستوى الدلالة لكل فيها أكبر من 0.05 المستوى المعتمدة لهذه الدراسة ويمكن ايجاد التوزيع الطبيعي للبيانات عن طريق المدرجات التكرارية وكما يلي:



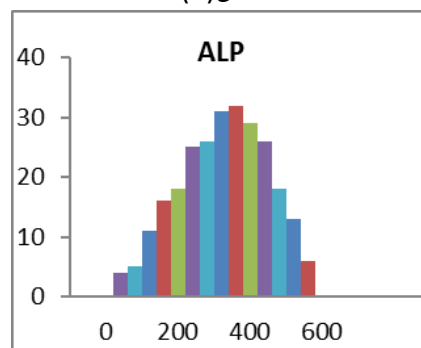
شكل (2)



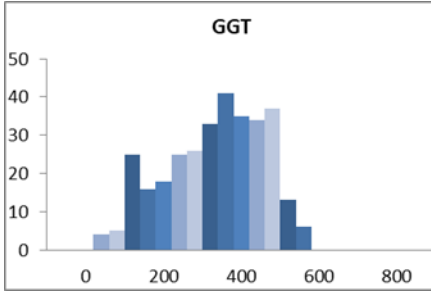
شكل (3)



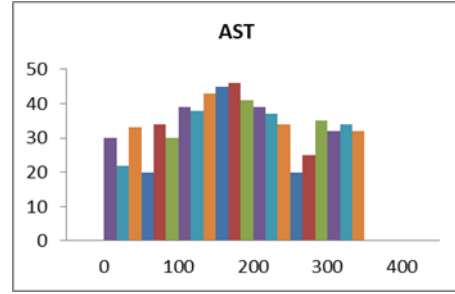
شكل (4)



شكل (5)



شكل (6)



شكل (7)

وللتأكد من ان البيانات لا تحتوي على قيم شاذة في كافة المتغيرات المستقلة تجري اختبار (Mahalanbis) ونلاحظ بوجود عمود جديد يحتوي على متغير جديد بأسم (MAH.1)، وبمراجعة هذه القيم الموجودة تحت عمود (MAH.1) نجد ان هذه القيم أقل من القيم الجدولة لمربع كاي والتي كانت تساوي 98 وبدرجة حرية 5 وذلك بعدم وجود قيم شاذة بكافة المتغيرات المستقلة .

التأكد من عدم وجود ارتباط ذاتي بين المتغيرات المستقلة يتضح من جدول رقم (4) أن قيم (VIF) لكل من المتغيرات التفسيرية الستة كانت 1.335 , 1.241 , 1.190 , 1.473 , 1.477 , 1.255, على التوالي وبما أن النتائج كانت أقل من 5 , يمكن الاستنتاج بأنة لا يوجد ارتباط , وايضا يمكن الاستنتاج من خلال قيم (Tolerance) وبما أن القيم أكبر من 0.57 (  $1 - R^2$  ) فإن ذلك يعني عدم وجود ارتباط بين المتغيرات التفسيرية .

جدول (4): يوضح الارتباط بين المتغيرات

| Classification Function Coefficients |         |         |
|--------------------------------------|---------|---------|
|                                      | genval  |         |
|                                      | 1.00    | 2.00    |
| Age                                  | -.034   | .005    |
| Weight                               | .408    | .322    |
| ALP                                  | .007    | .013    |
| AST                                  | .024    | .017    |
| (Constant)                           | -17.207 | -13.282 |

التأكد من تجانس المجتمع Box's M

الجدول (5) يوضح نتائج اختبار Box's M

| Test Results |         |             |
|--------------|---------|-------------|
| Box's M      |         | 36.823      |
| F            | Approx. | 12.003      |
|              | df1     | 3           |
|              | df2     | 1728720.000 |
|              | Sig.    | <0.23       |

ونلاحظ من الجدول (5) أن قيمة مستوى الدلالة المستخرج تساوي 0.23 وهي أكبر من مستوى المعنوية المستعملة في الدراسة هذا يدل على التجانس بين أفراد المجموعتين مما يستوجب بقبول الفرضية الصفرية .

جدول (6): نتائج القيمة النسبية

| Log Determinants     |      |                  |
|----------------------|------|------------------|
| Type                 | Rank | Log Determinants |
| 1                    | 2    | 10.602           |
| 2                    | 2    | 10.786           |
| Pooled Within-groups | 2    | 10.070           |

كما يلاحظ من الجدول رقم (6) أن قيم (Log Determinants) تقريبا متساوية بين المجموعتين وهذا يعني ان مصفوفات التباين المشتركة متساوية.

أجراء التحليل التمييزي:

الجدول (7): الأوساط الحسابية والانحرافات المعيارية لكل مجموعة من المجموعات.

| Group statistics |        |         |                |
|------------------|--------|---------|----------------|
|                  | Type   | Mean    | Std. Deviation |
| الذكور           | Age    | 64.040  | 7.5376         |
|                  | Weight | 76.580  | 9.7940         |
|                  | ALP    | 152.920 | 117.3491       |
|                  | AST    | 42.370  | 52.0797        |
|                  | ALT    | 24.772  | 27.9268        |
|                  | GGT    | 130.100 | 151.3071       |
| الاناث           | Age    | 34.560  | 14.9805        |
|                  | Weight | 68.460  | 18.8487        |
|                  | ALP    | 139.140 | 83.2422        |
|                  | AST    | 63.974  | 87.2415        |
|                  | ALT    | 38.078  | 46.2523        |
|                  | GGT    | 88.840  | 69.3649        |
| Total            | Age    | 49.300  | 18.9382        |
|                  | Weight | 72.520  | 15.4909        |
|                  | ALP    | 146.030 | 101.4566       |
|                  | AST    | 53.172  | 72.3009        |
|                  | ALT    | 31.425  | 38.5948        |

## جدول (8): جدول مصفوفة الارتباط داخل المتغيرات التفسيرية

Pooled Within-Groups Matrices

|             |        | Age    | Weight | ALP    | AST    | ALT    | GGT    |
|-------------|--------|--------|--------|--------|--------|--------|--------|
| Correlation | Age    | 1      | 0.505  | -0.047 | 0.014  | -0.055 | -0.040 |
|             | Weight | 0.505  | 1      | -0.042 | -0.054 | 0.065  | -0.216 |
|             | ALP    | -0.047 | -0.042 | 1      | 0.050  | 0.165  | 0.341  |
|             | AST    | 0.014  | -0.054 | 0.050  | 1      | 0.413  | -0.001 |
|             | ALT    | -0.055 | 0.065  | 0.165  | 0.413  | 1      | -0.088 |
|             | GGT    | -0.040 | -0.216 | 0.341  | -0.001 | -0.088 | 1      |

يشير جدول رقم (8) الى معاملات الارتباط الثنائي بين المتغيرات التفسيرية الستة، فمعامل الارتباط مثلاً بين (Age, Weight) كانت قيمته 0,505 وبين (AST,ALP) كان يساوي 0.050 وهكذا بين باقي المتغيرات

## جدول رقم(9) جدول التحليل التمييزي المتدرج

| Variables Not in the Analysis |           |                |            |         |                |                 |
|-------------------------------|-----------|----------------|------------|---------|----------------|-----------------|
| Step                          | Tolerance | Min. Tolerance | F to Enter | Min. F  | Between Groups |                 |
| 0                             | Age       | 1              | 1          | 154.513 | 154.513        | الحكومي والاهلي |
|                               | Weight    | 1              | 1          | 7.307   | 7.307          | الحكومي والاهلي |
|                               | ALP       | 1              | 1          | 0.459   | 0.459          | الحكومي والاهلي |
|                               | AST       | 1              | 1          | 2.261   | 2.261          | الحكومي والاهلي |
|                               | ALT       | 1              | 1          | 3.033   | 3.033          | الحكومي والاهلي |
|                               | GGT       | 1              | 1          | 3.072   | 3.072          | الحكومي والاهلي |
| 1                             | Weight    | 0.745          | 0.745      | 6.607   | 84.981         | الحكومي والاهلي |
|                               | ALP       | .998           | 0.998      | 0.612   | 77.256         | الحكومي والاهلي |
|                               | AST       | 1              | 1          | 1.089   | 77.871         | الحكومي والاهلي |
|                               | ALT       | 0.997          | 0.997      | 0.428   | 77.020         | الحكومي والاهلي |
|                               | GGT       | 0.998          | 0.998      | 1.940   | 78.967         | الحكومي والاهلي |
| 2                             | ALP       | 0.997          | 0.744      | 0.493   | 56.522         | الحكومي والاهلي |
|                               | AST       | 0.995          | 0.741      | 1.397   | 57.351         | الحكومي والاهلي |
|                               | ALT       | 0.985          | 0.736      | 0.133   | 56.191         | الحكومي والاهلي |
|                               | GGT       | 0.947          | 0.706      | 0.641   | 56.658         | الحكومي والاهلي |

يوضح الجدول (9) خطوات التحليل التمييزي المتدرج (Stepwise) و توضح الخطوة رقم(0) عدم استبعاد اي متغير وتوضح الخطوة رقم(1) تم تحديد متغير العمر Age كأول متغير يدخل في الدالة التمييزية لأنه يتمتع باكبر قيمة لدالة الاختبار F و ان قيمتها(154.513) وتوضح الخطوة رقم(2) استبعاد متغير الوزن (Weight) لأن قيمتها اكبر من قيمة F ، في الخطوة (2) لم يتم ادخال اي من المتغيرات الاربعة المتبقية (GGT,ALT,ALP, AST) الى الدالة اطلاقاً وذلك لأن قيم الأختبار لهذه المتغيرات اصغر من الحد الأدنى الموضوع مسبقاً والذي يساوي (3.84).

## جدول (10): يشير الى نتائج التحليل التمييزي التدريجي

| Walk's' Lambda |                     |        |     |     |     |           |     |        |       |
|----------------|---------------------|--------|-----|-----|-----|-----------|-----|--------|-------|
| Step           | Number of Variables | Lambda | df1 | df2 | df3 | Exact F   |     |        |       |
|                |                     |        |     |     |     | Statistic | df1 | df2    | Sig.  |
| 1              | 1                   | 0.388  | 1   | 1   | 98  | 154.513   | 1   | 98.000 | <.001 |
| 2              | 2                   | 0.363  | 2   | 1   | 98  | 84.981    | 2   | 97.000 | <.001 |

بين جدول (10) النتائج المعتمدة للتحليل التمييزي المتدرج يوضح ان المتغيرين (1,2) هي التي ستعتمد عليها دالة التمييز للمصابين وان اول هذه المتغيرات هو العمر (Age) لان له اكبر قيمة لنتيجة اختبار F واكبر قيمة ل(Walk's' Lambda) ويليها متغير التفسير الثاني الوزن (Weight).

جدول(11): يشير الى معاملات التمييز المعيارية وغير المعيارية

## Standardized Canonical Discriminant Function Coefficients

|        | Function |
|--------|----------|
|        | 1        |
| Age    | -.623    |
| Weight | 1.051    |
| ALP    | -.492    |
| AST    | .404     |

يوضح الجدول (11) معاملات التمييز المعيارية وغير المعيارية ويمكن كتابة نموذج الدالة التمييزية كما في المعادلة التالية

$$y = -0.623x_1 + 1.051x_2 - 0.493x_3 + 0.404x_4$$

التصنيف (Classification):

جدول(12): نتائج التصنيف

| Classification Results                                  |       |         |                            |        |       |
|---|-------|---------|----------------------------|--------|-------|
|   |       | Type    | Predicted Group Membership |        | Total |
|   |       |         | الحكومي                    | الاھلي |       |
| Original  | Count | الحكومي | 50                         | 0      | 50    |
|   |       | الاھلي  | 5                          | 45     | 50    |
|   | %     | الحكومي | 100.0                      | .0     | 100.0 |
|   |       | الاھلي  | 10.0                       | 90.0   | 100.0 |
| A 95.0% of original grouped cases correctly classified. |       |         |                            |        |       |

يشير جدول رقم (12) الى التصنيف الصحيح لمفردات العينة بالأعتماد على الدالة التمييزية وبيّن الجدول ان معدل التصنيف الناجح هو %95 اي ان تم تصنيف مرضى الحكومي والأهلي تصنيفاً صحيحاً بدقة %95 وهذا يعني ان نسبة خطأ التصنيف قليلة وهي %5 .

### الأستنتاجات والتوصيات

1. ملائمة اسلوب التحليل التمييزي لأية بيانات يحتاجها الباحث في تمييز وتصنيف المفردات الجديدة.

2. تم التوصل الى نموذج يوضح اهمية المتغيرات الداخلة في التمييز

3. تطبيق الدوال التمييزية في تحليل نتائج البيانات ولاسيما الطبية منها

### المصادر:

- 1) بسيوني، عبدالرحيم عوض عبدالخالق (2021)، "استخدام التحليل التمييزي في التنبؤ". المجلة العلمية / التجارة والتمويل، المجلد 41، العدد 3.
- 2) الحمداني، بسمة رشيد، (2014)، "تمييز الكادر الطبي حسب معرفتهم للتصنيف الدولي بأستخدام الدالة المميزة، بحث دبلوم عالي جامعة بغداد، كلية الادارة والاقتصاد
- 3) بغرش، سعيدة، (2020) "استخدام التحليل التمييزي في تقدير خطر عدم تسديد القرض من طرق الوكالة الوطنية لتسيير القرض المصغر" مجلة نماء للاقتصاد والتجارة، المجلد 4 العدد1، الجزائر.
- 4) الجاعوني، فريد عدنان غانم (2007)، "التحليل الإحصائي متعدد المتغيرات (التحليل التمييزي) في توصيف وتوزيع الأسر داخل الهيكل الاقتصادي الاجتماعي في المجتمع"، مجلة جامعة دمشق للعلوم الاقتصادية والقانونية- المجلد 23 -العدد الثاني.
- 5) الحمداني، بسمة رشيد، (2014)، "تمييز الكادر الطبي حسب معرفتهم للتصنيف الدولي بأستخدام الدالة المميزة، بحث دبلوم عالي جامعة بغداد، كلية الادارة والاقتصاد
- 6) حمودات، الاء عبد الستار، (2015)، "الدالة التمييزية وطرق تحديد متغيراتها"، بحث ماجستير، جامعة الموصل، كلية علوم الحاسبات والرياضيات قسم الرياضيات.
- 7) خلف عماد محمد وعقيل محمد رشيد (2021) "مديونية القطاع المالي وأثرها على الاقتصاد المالي في العراق" مجلة الدراسات النقدية والمالية العدد التاسع.
- 8) خياط، شذى بنت شاكر وجواهر بنت محمد المزيد، (2018)، "القوة التمييزية للذكاءات المتعددة بين طالبات الكليات الأنسانية والعلمية بجامعة الملك سعود"، بحث بكالوريوس، كلية التربية- جامعة جدة، كلية التربية جامعة الملك سعود.
- 9) الراوي، خاشع، (1987)، "المدخل الى تحليل الانحدار"، جامعة الموصل، كلية الزراعة والغابات.
- 10) رجب، وليد خالد وحسين جميل محمود، (2012)، "التحليل التمييزي لبعض المتغيرات المهارة لدى حراس المرمى لكرة القدم"، مجلة الرافدين للعلوم الرياضية (نصف سنوية)، المجلد (18)، العدد (58)
- 11) سكيينة، شامل جاسم (2020)، "دراسة مقارنة بين اسلوب التحليل التمييزي الخطي واسلوب التحليل التمييزي اللبي"، المجلة العراقية للعلوم الادارية.
- 12) سليمان، علي أبشر فضل المولى، (2015)، "المقارنة بين التحليل التمييزي والنموذج اللوجستي الثنائي ونماذج الشبكات العصبية في تصنيف المشاهدات.
- 13) هادي، عائدة صالح، (2008)، "إستخدام التحليل المميز لتشخيص بعض أمراض العيون"، مجلة الإدارة والاقتصاد، العدد السابع والستون.
- 14) الياسين، دريد حسين بدر، (2009)، "أستخدام بعض طرائق التمييز الحصينة لتشخيص أمراض سرطان الدم"، بحث ماجستير، الجامعة المستنصرية، كلية الإدارة والاقتصاد.
- 15) عبدالله، شهد نجم، (2018)، "استعمال التحليل التمييزي في تصنيف درجة التشوه للولادات"، بحث دبلوم عالي، الجامعة المستنصرية، قسم الاحصاء



## تقدير انموذج الانحدار الذاتي المكاني ( SAR ) باستعمال طريقة ييز لبيان اثر التجاورات المكانية على مرض السرطان في العراق

الباحثة ساره اسامة سعد  
أ.م.د هيفاء طه عبد  
الجامعة المستنصرية / كلية الادارة والاقتصاد      الجامعة المستنصرية / كلية الادارة والاقتصاد

تاريخ استلام البحث: 2022/04/12

تاريخ قبول البحوث: 2022/05/02

نشر البحث في العدد عشرون: ايلول / سبتمبر 2023

رمز التصنيف ديوي / النسخة الالكترونية (Online): 2522-64X/515.7

رمز التصنيف ديوي / النسخة الورقية (Print): 2519-948X/515.7



## تقدير انموذج الانحدار الذاتي المكاني ( SAR ) باستعمال طريقة بيز لبیان اثر التجاورات المكانية على مرض السرطان في العراق

الباحثة ساره اسامة سعد  
أ.م.د هيفاء طه عبد  
الجامعة المستنصرية / كلية الادارة والاقتصاد      الجامعة المستنصرية / كلية الادارة والاقتصاد

### المستخلص:

لقد جذب تحليل البيانات المكانية في الآونة الأخيرة اهتمام الباحثين الإحصائيين وخاصة لان إهمال البعد المكاني في التحليل يؤثر على النتائج ويؤدي إلى ضياع معلومات مهمة، ولذلك تم اللجوء إلى نماذج الانحدار المكانية التي يمكن من خلالها دراسة مدى تأثير المتغير المعتمد بالمتغيرات التوضيحية في ضل وجود الاعتماد المكاني لمفردات الظاهرة المدروسة، ولقد تم استعمال طريقة بيز المعلمية في هذا البحث في تقدير انموذج الانحدار الذاتي المكاني (SAR) في ضل وجود مصفوفة الأوزان المكانية المعدلة التي تم بناءها بلاء اعتماد على معيار روك Rock للتجاور ومصفوفة الأوزان المكانية المقترحة التي تجمع ما بين معيار روك للتجاور وعلى المسافة الإقليدية في بناءها، وبلاء اعتماد على بيانات مرضى السرطان المتمثلة بعدد مرضى السرطان في كل محافظة من محافظات العراق كمتغير معتمد اما المتغيرات التوضيحية فقد تمثلت بمعدل العمر ومعدل حجم الورم لمرضى السرطان وعدد المناطق الملوثة باليورانيوم في المحافظة، وقد تم من خلال استعمال معيار موران الكشف عن وجود الاعتماد المكاني لبيانات مرضى السرطان مما يدل على ان الموقع الجغرافي له تأثير على الإصابة بمرض السرطان، ومن خلال نتائج التقدير نلاحظ ان قيم المتغير المعتمد التقديرية في ضل مصفوفة الأوزان المكانية المعدلة تكون اقرب للقيم الحقيقية من قيم المتغير المعتمد التقديرية في ضل مصفوفة الأوزان المكانية المقترحة، وهذا ما اكده قيمة معيار MAPE الذي تم حسابه في ضل مصفوفة الأوزان المكانية المعدلة اصغر من القيمة التي تم حسابها في ضل مصفوفة الأوزان المكانية المقترحة.

**الكلمات المفتاحية:** انموذج الانحدار الذاتي المكاني، الاعتماد المكاني، معامل موران، طريقة بيز Bayes.

### Estimate the Spatial Auto Regressive Model by using Bayesian Methods to show the impact of spatial juxtapositions on cancer in Iraq

#### Abstract:

Recently, spatial data analysis has attracted the attention of statistical researchers, especially because neglecting the spatial dimension in the analysis affects the results and leads to the loss of important information. Therefore, spatial regression models have been resorted to, through which it is possible to study the extent to which the dependent variable is affected by the explanatory variables in light of the presence of spatial dependence for the studied

vocabulary, The Bayes method was used in this research to estimate the SAR model in the presence of the modified spatial weight matrix built on the Rock juxtaposition standard and Suggested weights matrix by the researcher that was built Depending on the Rock juxtaposition criterion and the Euclidean distance, and based on cancer patient data on the number of cancer patients in each governorate of Iraq as an approved variable, but the illustrative variables were the rate of age, the rate of tumor size of cancer patients and the number of areas contaminated with uranium in the governorate, Through the use of Moran's criterion, the presence of spatial dependence of cancer patients' data was revealed, which indicates that the geographical location has an impact on the incidence of cancer, Through the estimation results, we note that the estimated values of the dependent variable in light of the modified spatial weights matrix are closer to the real values than the estimated values of the dependent variable in light of the proposed spatial weights matrix, This was confirmed by the value of the MAPE standard, which was calculated within the modified spatial weights matrix is smaller than the value that was calculated under the proposed spatial weights matrix.

**Key words:** Spatial Auto Regressive Model, Spatial Depends, Moran Coefficient, Bayes method.

## 1- المبحث الاول : منهجية البحث

### 1-1 المقدمة

يمثل الاقتصاد القياسي المكاني أحد أهم الفروع في الاقتصاد القياسي، وتأتي أهميته من كونه يتعامل مع البيانات المكانية التي تتميز بأحتواءها على صفة التبعية المكانية (الارتباط المكاني) وصفة عدم التجانس المكاني، وقد قام الباحث Anselin بمناقشة هذا الأمر في ابحاثه عام (1988-2001)، وإن هاتان الصفتين تجعل من غير المناسب التعامل مع البيانات المكانية بتقنيات الاقتصاد القياسي التقليدي التي تهتم النماذج الخاصة به بالاعتمادية بين المشاهدات خلال فترة زمنية معينة دون الاخذ بنظر الاعتبار الاعتمادية المكانية مما يؤدي الى الحصول على تقديرات غير كفؤه نتيجة لعدم تحقق فريات التحليل بالاضافة الى إهمال الكثير من المعلومات المكانية الخاصة بالبيانات نتيجة عدم القدرة على توضيفها والاستفادة منها عند استعمال نماذج القياس الاقتصادي الغير مكانية، وقد قام الباحثين في الاقتصاد القياسي بوضع عدد من النماذج المكانية التي اصبحت تنتمي لاحد فروع الاقتصاد القياسي الذي يقوم بعملية التحليل الاحصائي المكاني والتعامل مع الاعتمادية المكانية.<sup>(1)</sup>

### 1-2 مشكلة البحث

ان مشكلة البحث تتلخص في وجود الاعتماد المكاني في البيانات للظاهرة المدروسة والتي لا تؤخذ بنظر الاعتبار عند دراسة هذا النوع من الظواهر مما يؤدي الى ضياع معلومات مهمة تنعكس على دقة تقدير معلمات الانموذج وعلى النتائج التي تم الحصول عليها، الامر الذي يستدعي الى البحث عن انموذج قياسي يأخذ بنظر الاعتبار الارتباط المكاني لبيانات الظاهرة.

**3-1 هدف البحث**

يهدف هذا البحث الى تقدير معلمات إنموذج الانحدار الذاتي المكاني (SAR) باستعمال طريقة بيز Bayes لبيانات مرض السرطان عن طريق أخذ عينة من كل محافظة من محافظات العراق لبيان هل هناك اثر للتجاور المكاني للمحافظات في انتشار المرض.

**2- المبحث الثاني: الجانب النظري**

يتضمن هذا المبحث عرض نظري لانموذج الانحدار الذاتي المكاني وطريقة بيز في التقدير مع توضيح بعض المفاهيم الاساسية الخاصة بالانموذج.

**1-2 انموذج الانحدار الذاتي المكاني (SAR)**

ويسمى ايضا انموذج الانحدار الذاتي المكاني المختلط Mixed Spatial Autoregressive Model ويمثل هذا الانموذج حالة خاصة من انموذج الانحدار الذاتي المكاني العام (SAC) الذي تم اقتراحه من قبل (Anselin)، ويمكن التعبير عن انموذج الانحدار الذاتي المكاني رياضيا من خلال الصيغة التالية: (5)

$$\underline{y} = \lambda W \underline{y} + X \underline{B} + \underline{u} \dots \dots \dots (1)$$

$$\underline{u} \sim N(0, \sigma^2 I_n)$$

حيث ان :

$\underline{y}$ : متجة ابعاده (nx1) ويمثل أمتغير المعتمد.

$\lambda$ : معلمة الاعتماد المكاني.

X: مصفوفة ابعاده (n x k) للمتغيرات التوضيحية .

B: متجة ابعاده (k x 1) للمعلمات المتعلقة بالمتغيرات التوضيحية x.

$\underline{u}$ : متجة الاخطاء ابعاده (nx1) والذي يتوزع توزيع طبيعي بمتوسط صفر وتباين  $\sigma^2 I_n$ .

W: مصفوفة الاوزان المكانية ابعاده (nxn) وهي ثابتة وتكون محددة مسبقاً.

**2-2 مصفوفة التجاور المكانية**

وهي احدى مصفوفات الاوزان المكانية ويتم بناءها بالاعتماد على معيار التجاور بين الوحدات المكانية ، فاذا كانت الوجدتين المكانيةين متجاورة فيعطى لها القيمة (1) اما اذا كانت الوجدتين المكانيةين غير متجاورة فيعطى لها القيمة (0) وكذلك فإن الوحدة المكانية لا تجاور نفسها فيعطى لها القيمة (0) ايضا في مصفوفة الاوزان المكانية عند بناءها ، وهناك عدة معايير لبناء لمصفوفة التجاور المكانية وقد تم الاعتماد على معيار روك للتجاور في هذا البحث، حيث يكون التجاور بالاعتماد على اذا كان الوجدتين المكانية مشتركتين بحد طولي غير صفري.(3)

**3-2 مصفوفة الاوزان المكانية المعدلة**

لقد سميت هذه المصفوفة بالمصفوفة المعدلة لانها تتكون من مصفوفة التجاور المكانية ولكن بعد اجراء بعض العمليات عليها من أجل ان يصبح مجموع كل صف فيها يساوي الواحد الصحيح اي ان: (3)

$$\sum_{j=1}^n W_{ij}^{Adj} = 1$$

ويتم بناء مصفوفة الاوزان المعدلة من خلال تطبيق الصيغة التالية على قيم اوزان مصفوفة التجاور:

$$W_{ij}^{Adj} = \begin{cases} \frac{W_{ij}}{\sum W_{ij}} & \text{if } W_{ij} = 1 \\ 0 & \text{if } W_{ij} = 0 \end{cases} \quad \dots (2)$$

#### 4-2 مصفوفة الاوزان المكانية المقترحة

ان المصفوفة المقترحة من قبل الباحث تم بناءها بالاعتماد على عامل الجوار و المسافة معا بين الوحدات المكانية حيث ان الاوزان المكانية ستتألف من المسافة الاقليدية بين مراكز الوحدات المكانية (المحافظات ) المتجاورة وفق معيار روك للتجاور، ولقد تم بناء الاوزان المكانية للمصفوفة المقترحة وفق الصيغة التالية : (9)

$$W_{ij}^{Rd} = \begin{cases} \frac{1}{1 + d_{ij}} & \text{if } i \neq j \text{ and } i \text{ is contiguous with } j \\ \frac{1}{\sum_{i=1}^n (1 + d_{ij})} & \dots (3) \\ 0 & \text{if } i = j \text{ and } i \text{ isn't contiguous with } j \end{cases}$$

#### 5-2 طريقة بيز في التقدير

إن طريقة بيز في التقدير قد نشأت من قبل العالم البريطاني توماس بيز (1702-1761) وهو صاحب مبرهنة بيز التي تستند على الاحتمالات الشرطية، وتختلف طريقة بيز عن الطرق الاخرى في التقدير بكونها تتعامل مع المعلمات المراد تقديرها على إنها متغيرات عشوائية، بمعنى ان المعلمات تمتلك توزيع احتمالي بفترة معينة بعكس الطرائق التقليدية التي تتعامل مع معلمات الانموذج كثوابت غير معلومة القيمة، ويتألف الانموذج البيزي من ثلاثة اجزاء ويمكن توضيحها من خلال الصيغة الاتية : (4)

$$p(Z|D) = \frac{P(D|Z) P(Z)}{P(D)} \quad \dots (4)$$

حيث ان :

$P(Z)$ : التوزيع المسبق للمعلمة  $Z$ ، ويمثل أعتقاد الباحث السابق لتوزيع المعلمة ( $Z$ ) قبل ملاحظة البيانات وهذه الاعتقادات تكون إما معلوماتية او غير معلوماتية .

$P(D|Z)$ : دالة الامكان الاعظم للإنموذج، وهي التي توضح العلاقة بين البيانات ومعلمة الإنموذج  $Z$ .  
 $p(Z|D)$ : التوزيع اللاحق للمعلمة  $Z$ ، ويحتوي على البيانات المحدثة الخاصة بتوزيع المعلمة  $Z$  بعد أخذ بيانات العينة الحالية ( $D$ ) بنظر الاعتبار.

ويمثل التوزيع اللاحق قاعدة أرتكاز الاستدلال البيزي حيث تم إشتقاقه من قاعدة بيز، وإذا تم تجاهل  $P(D)$  في الصيغة (4) بسبب كونها لا تتضمن المعلمة ( $Z$ ) فيمكن كتابة التوزيع اللاحق بالشكل الأتي : (6)

$$p(Z|D) \propto P(D|Z) P(Z) \quad \dots (5)$$

وهذا يعني إن التوزيع اللاحق ناتج من حاصل ضرب دالة الامكان الأعظم  $P(D|Z)$  بالمقدار  $P(Z)$  والذي يمثل التوزيع المسبق للمعلمة ( $Z$ )، وفي حالة إحتواء الإنموذج على أكثر من معلمة كما هو الحال في إنموذج SAR المستعمل في البحث فإن التوزيع اللاحق سيمثل التوزيع المشترك لكل المعالم مشروطاً على بيانات المشاهدة  $y$ ، حيث يتم الحصول عليه عن طريق ضرب دالة الامكان الاعظم  $L(\underline{B}, \sigma^2, \lambda)$  بالتوزيعات المسبقة

للمعلمات  $p(\underline{B})$ ,  $p(\lambda)$ ,  $p(\sigma^2)$ ، ومن خلال قاعدة بيز يمكن كتابة التوزيع اللاحق لانموذج الانحدار الذاتي المكاني SAR بالشكل الآتي: (6)

$$p(\underline{B}, \sigma^2, \lambda | D) = \frac{p(D | \underline{B}, \sigma^2, \lambda)}{p(D)} \pi(\underline{B}, \sigma^2) \pi(\lambda) \dots (6)$$

حيث ان  $p(D)$  يتضمن اي جزء من التوزيع اللاحق الذي لا يحتوي على معلمات الانموذج ويمكن كتابة التوزيع اللاحق في الصيغة (6) بالشكل الآتي :

$$p(\underline{B}, \sigma^2, \lambda | D) \propto p(D | \underline{B}, \sigma^2, \lambda) \pi(\underline{B}, \sigma^2) \pi(\lambda) \dots (7)$$

$$\pi(\underline{B}, \sigma^2) \sim \text{NIG}(c, T, a, b)$$

$$\pi(\underline{B}, \sigma^2) = \pi(\underline{B} | \sigma^2) \pi(\sigma^2)$$

$$= N(c, \sigma^2 T) \text{IG}(a, b)$$

$$\pi(\underline{B}, \sigma^2) = \frac{b^n}{(2\pi)^{k/n} |\Gamma|^{1/2} \Gamma(a)} (\sigma^2)^{-(a+(k/2)+1)} \times \exp\left[-\{(B - c)'T^{-1}(B - c) + 2b\}/(2\sigma^2)\right] \dots (8)$$

$$\pi(\sigma^2) = \frac{b^a}{\Gamma(a)} (\sigma^2)^{-(a+1)} \exp(-b/\sigma^2) \dots (9)$$

$$\Gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt$$

$$\sigma^2 > 0, \quad a, b > 0$$

$$\pi(\lambda) \sim U(\omega_{\min}^{-1}, \omega_{\max}^{-1}) \dots (10)$$

حيث ان :

$$\pi(\underline{B}, \sigma^2) : \text{التوزيع الحدي (الهامشي) السابق للمعلمتين } \underline{B}, \sigma^2$$

$$\pi(\sigma^2) : \text{التوزيع الحدي السابق (الاولي) للمعلمة } \sigma^2$$

$$\pi(\lambda) : \text{التوزيع الحدي السابق للمعلمة } \lambda$$

ومن خلال الصيغة (8)،(9)،(10) يمكن اعادة كتابة التوزيع اللاحق لانموذج الانحدار الذاتي SAR في الصيغة (5) بالشكل الآتي :

$$p(\underline{B}, \sigma^2, \lambda | D) \propto (\sigma^2)^{a^*+(k/2)+1} |A| \times \exp\left\{-\frac{1}{2\sigma^2} [2b^* + (B - c^*)'(T^*)^{-1} (B - c^*)]\right\} \dots (11)$$

حيث ان :

$$(B - c^*)'(T^*)^{-1} (B - c^*) \equiv (Ay - XB)'(Ay - XB) + (B - c)'T^{-1}(B - c) + 2b^*$$

$$c^* = (X'X + T^{-1})^{-1}(X'Ay + T^{-1}c)$$

$$c = \text{zero matrix}$$

$$T^* = (X'X + T^{-1})^{-1}$$

$$T = I \cdot 10^3$$

$$a^* = a + n/2$$

$$b^* = b + (c'T^{-1}c + y'A'Ay - (c^*)'(T^*)^{-1}c^*)/2$$

$$A = I - \lambda W$$

ويعد التوزيع اللاحق في الصيغة (11) العنصر الرئيسي للتقدير بطريقة بيز حيث يتم من خلاله اشتقاق التوزيعات الحدية التي تحتوي على كل المعلومات الخاصة بمعلمات الانموذج  $(\underline{B}, \sigma^2, \lambda)$  مشروطة على بيانات المشاهدة  $y$ ، وان عملية الحصول على التوزيعات الحدية الشرطية  $p(\sigma^2 | y)$ ,  $p(\underline{B} | y)$ ,  $p(\lambda | y)$  من التوزيع اللاحق المشترك  $p(\underline{B}, \sigma^2, \lambda | D)$  عملية

صعبة وذلك لكون عملية الاشتقاق معقدة وشبه مستحيلة باستعمال الورقة والقلم، وهذه المشكلة سببت صعوبة للطريقة البيزية في ايجاد التوزيع الخاص بكل معلمة من معلمات الانموذج على حدة ومن اجل حل هذه المشكلة تم اللجوء الى استعمال سلاسل ماركوف مونت كارلو (MCMC)، وهي عبارة عن طرائق توليد مرنة ومن اشهرها طريقة كبز Gibbs وطريقة Metropolis-Hastings، ويتلخص مبدأ عملها بتحليل التوزيع اللاحق لمجموعة من التوزيعات الشرطية الكاملة لكل معلمة من معلمات انموذج SAR وهي :

$$p(\sigma^2 | \underline{B}, \lambda), p(\underline{B} | \sigma^2, \lambda), (\lambda | \underline{B}, \sigma^2) \dots (12)$$

ثم القيام بعملية سحب عينات عشوائية من هذه التوزيعات الشرطية الكاملة مما يمكن الباحث من استعمال هذه العينة من السحبات في الحصول على التقديرات البيزية الصحيحة لمعلمات الانموذج  $(\underline{B}, \sigma^2, \lambda)$ ، حيث تستعمل طريقة كبز Gibbs مع التوزيعات الشرطية التي تمتلك صيغة معروفة فقط بينما طريقة Metropolis-Hastings يمكن استعمالها مع التوزيعات الشرطية الكاملة التي لا تمتلك صيغة معروفة. (6)

## 2-5-1 التوزيعات الشرطية الكاملة

ان عملية الحصول على التوزيعات الشرطية الكاملة هي اولى متطلبات معاينة كبز حيث تم اشتقاق التوزيعات الشرطية الكاملة لانموذج SAR من التوزيع اللاحق في الصيغة (11)، ومن الجدير بالذكر ان هنالك ثلاثة افكار رياضية تم تطبيقها والالتزام بها عند الاشتقاق وهي : (4) اولاً: اهمال كل الثوابت الغير ضرورية.

ثانياً: جمع الحدود الخاصة فقط بالتوزيع الشرطي عند القيام بعملية اشتقاق التوزيع الشرطي المطلوب .

### ثالثاً: إكمال المربع.

ان التوزيعات الشرطية الكاملة لانموذج SAR التي تم اشتقاقها كما يلي : (6)

■ التوزيع الشرطي الكامل للمعلمة  $\sigma^2$ ،  $p(\sigma^2 | \underline{B}, \lambda)$ : يملك توزيع كما المعكوس (IG) المبين كما يلي:

$$p(\sigma^2 | \underline{B}, \lambda) \sim \text{IG}(a^*, b^*)$$

$$p(\sigma^2 | \underline{B}, \lambda) = \frac{b^{*a^*}}{\Gamma(a^*)} \sigma^{2-(a^*+1)} \cdot e^{-b^*/\sigma^2} \dots (13)$$

$$a^* = a + n/2$$

$$b^* = b + (\underline{A}\underline{y} - \underline{X}\underline{B})'(\underline{A}\underline{y} - \underline{X}\underline{B})/2$$

$$\pi\sigma^2 \sim \text{IG}(a, b)$$

$$a = \frac{n(\bar{x}^2 - 2 \sum x_i^2)}{n\bar{x}^2 - \sum x_i^2}$$

$$b = (a - 1)\bar{x}$$

■ التوزيع الشرطي الكامل للمعلمة  $\underline{B}$ ،  $p(\underline{B} | \sigma^2, \lambda)$ : يملك التوزيع الطبيعي المتعدد المتغيرات (N) كما يلي:

$$p(\underline{B} | \sigma^2, \lambda) \sim N(C^*, \sigma^2 T^*)$$

$$p(\underline{B} | \sigma^2, \lambda) = \frac{1}{(2\pi)^{-p/2} |\sigma^2 T^*|^{1/2}} \exp \left\{ -\frac{1}{2} (\underline{B} - C^*)' (\sigma^2 T^*)^{-1} (\underline{B} - C^*) \right\} \dots (14)$$

$$c^* = (X'X + T^{-1})^{-1} (X' Ay + T^{-1} c)$$

$$c = \text{zero matrix}$$

$$T^* = (X'X + T^{-1})^{-1}$$

$$T = I \cdot 10^3$$

■ التوزيع الشرطي الكامل للمعلمة  $p(\lambda | \underline{B}, \sigma^2)$  يملك توزيع بصيغة غير معروفة يتم الحصول عليه من الصيغة التالية :

$$p(\lambda | \underline{B}, \sigma^2) \propto \frac{p(\lambda, \underline{B}, \sigma^2 | D)}{p(\underline{B}, \sigma^2 | D)} \quad \dots (15)$$

$$p(\lambda | \underline{B}, \sigma^2) \propto |I - \lambda W| \exp\left(-\frac{1}{2\sigma^2} (A\mathbf{y} - X\mathbf{B})' (A\mathbf{y} - X\mathbf{B})\right) \quad \dots (16)$$

$$p(\underline{B}, \sigma^2 | D) \sim \text{NIG}(c, T, a, b)$$

## 2-5-2 معاينة (M-H) Metropolis-Hastings

ان معاينة (M-H) عبارة عن خوارزمية رفض وقبول حيث يتم فيها اقتراح توزيع معين نولد منه قيمة مرشحة للمعلمة  $\lambda$  والتي سنسميها  $\lambda^*$  ثم يتم المقارنة او تقييم القيمة المرشحة مع القيمة الحالية للمعلمة التي نحصل عليها من الصيغة (16) والتي سنسميها  $\lambda^Q$ ، ومن اجل حساب احتمال القبول للقيم المرشحة بعد المقارنة نستعمل الصيغة التالية : (6)

$$\psi H(\lambda^Q, \lambda^*) = \min \left[ 1, \frac{p(\lambda^* | B, \sigma^2)}{p(\lambda^Q | B, \sigma^2)} \right] \quad \dots (17)$$

وغالبا ما يتم ضبط احتمالية القبول ضمن المدى (40-60) وذلك لكي يتم سحب نصف العينات المرشحة ، ولقد تم اختيار التوزيع الطبيعي كتوزيع مقترح جنبا الى جنب مع اجراءات السير العشوائي التي اقترحها كل من Holloway, Shankara و Rahman عام 2002 من اجل الحصول على القيمة المرشحة للمعلمة  $\lambda$ ، وان عملية المعاينة ومعلمة الضبط Q يمكن توضيحها من خلال الصيغة التالية :

$$\lambda^* = \lambda^Q + Q \cdot N(0,1) \quad \dots (18)$$

حيث ان الصيغة (18) توضح سبب تسمية اجراء توليد البيانات بالسير العشوائي، وان الهدف من ضبط وتقييم العينات التي تسحب من التوزيع الطبيعي المقترح هو للتأكد من ان عملية سحب العينات العشوائية تتم خلال التوزيع الشرطي بأكمله.

## 2-5-3 معاينة كايبز Gibbs

تمثل معاينة كايبز Gibbs احدى طرائق توليد البيانات وهي حالة خاصة من معاينة (M\_H) وذلك لكونها تتعامل مع التوزيعات الشرطية الكاملة ذات الصيغ المعروفة فقط ، حيث يتم سحب عينات عشوائية من التوزيعات الشرطية الكاملة الخاصة بمعالم نموذج SAR ويتم اولا اختيار قيم افتراضية لمعلمات الانموذج  $\lambda_{(0)}, B_{(0)}, \sigma^2_{(0)}$ ، وبعدها يبدأ اخذ عينات بشكل متسلسل من التوزيعات الشرطية الكاملة الثلاث وكالاتي : (6)

1- تسحب عينة من التوزيع الشرطي الكامل للمعلمة B، حيث تمتلك  $p(\underline{B} | \sigma^2_{(0)}, \lambda_{(0)})$  التوزيع الطبيعي المتعدد  $N(C^*, \sigma^2 T^*)$  المبين في الصيغة (14) ونرمز للمعلمة المسحوبة بالرمز  $\underline{B}_{(1)}$ ، ويتم استعمال هذه المعلمة بدل من المعلمة  $\underline{B}_{(0)}$  بالخطوة التالية.

2- تسحب عينة من التوزيع الشرطي الكامل للمعلمة  $\sigma^2$ ، حيث تمتلك  $p(\sigma^2 | \underline{B}_{(1)}, \lambda_{(0)})$  توزيع كما المعكوس  $IG(a^*, b^*)$  المبين في الصيغة (15) ونرمز للمعلمة المسحوبة بالرمز  $\sigma^2_{(1)}$  ويتم استعمال هذه المعلمة بدل من المعلمة  $\sigma^2_{(0)}$  بالخطوة التالية.

3- تسحب عينة من التوزيع الشرطي الكامل للمعلمة  $\lambda$ ، حيث تمتلك  $p(\lambda|B_{(1)}, \sigma^2_{(1)})$  توزيع غير معروفة الصيغة و المبين في الصيغة (16) ولهذا تم سحب العينة باستعمال طريقة معاينة (M-H) ونرمز للعينة الجديدة بالرمز  $\lambda_{(1)}$  ونعوضها بالخطوه رقم (1) حيث ان الخطوات المتسلسلة من 1 الى 3 تمثل مسار واحد او دورة واحدة وفي كل دوره نقوم بسحب عينات جديدة للمعلمات، وكمثال توضيحي من الممكن ان نسحب 7500 مرة ونتجاهل اول 2500 سحبة من العينات ونستعمل الباقي للتوصل للتوزيع الاحق والاستدلالات الخاصة به.

## 2-6 اختبار معامل موران :

ان معامل موران عبارة عن اداة لقياس الاعتماد المكاني في بيانات الظاهرة المراد دراستها ومدى التشابه فيما بينها ويمكن ان نرمز له بالرمز  $I_{MC}$  ويكون مناضر لاختبار ديرين واتسون ( Durbin Watson) في بيانات السلاسل الزمنية، وتتراوح قيمة معامل موران بين (+1, -1) حيث نلاحظ انه اذا اقتربت قيمة معامل موران من (+1) يكون نمط الانتشار للبيانات متقارب اما اذا كانت قيمة معامل موران قريبة من (-1) يكون انتشار البيانات متباعد فيما بينها اما اذا كانت قيمة معامل موران تقترب من (0) نلاحظ ان البيانات عشوائية الانتشار المكاني،<sup>(7)</sup> وان صيغة معامل موران كما يلي: (2)

$$I_{MC} = \frac{n(u'w u)}{S_{\theta}(u'u)} \dots \dots \dots (19)$$

$$S_{\theta} = \sum_{i=1}^n \sum_{j=1}^n W_{ij}$$

حيث ان :

$S_{\theta}$  : مجموع العناصر في المصفوفة  $W$ .

$n$  : حجم العينة .

$W$  : مصفوفة الازان (التجاورات) ذو الأبعاد  $n \times n$ .

$u$  : متجة الاخطاء (البواقي) ذو البعد  $n \times 1$ .

اما اذا كانت مصفوفة الازان المستعملة من النوع التي يكون فيها مجموع الصف يساوي الواحد الصحيح ( Row- Standardized ) في هذه الحال سيكون  $(n=S_{\theta})$  مما يمكن الباحث من كتابة الصيغة (19) بالشكل الآتي:

$$I_{MC} = \frac{(u'w u)}{(u'u)} \dots \dots \dots (20)$$

ولقد تم التوصل للتوزيع التقاربي الخاص باحصاءة موران من قبل الباحثان Ord و Cliff في عام (1972) حيث انه يتوافق مع التوزيع الطبيعي القياسي ويمكن اجراء اختبار موران (Z) من خلال الصيغة التالية : (2)

$$Z_{I_{MC}} = \frac{I_{MC} - E(I_{MC})}{\sqrt{V(I)}} \dots \dots \dots (21)$$

$$E(I_{MC}) = E(I_{MC}) = \frac{\text{tr}(M W)}{n - k}$$

$$V(I_{MC}) = \frac{\text{tr}(M W M W') + \text{tr}(M W)^2 + (\text{tr}(M W))^2}{(n - k)(n - k + 1)} - (E(I_{MC}))^2$$



$$M = (I_n - X(X'X)^{-1}X')$$

حيث ان :

tr: مجموع العناصر للقطر الرئيسي.

K: يمثل عدد المتغيرات التفسيرية.

M: تمثل مصفوفة صماء وتكون مربعة و متماثلة.

ومن اجل اختبار وجود الاعتماد المكاني من عدمه نستخدم الفرضية التالية :

$$H_0: \lambda = 0 \quad \text{لا يوجد اعتماد مكاني}$$

$$H_1: \text{At least one of } \lambda \neq 0 \quad \text{يوجد اعتماد مكاني}$$

ومن خلال الصيغة (20) اذا كانت قيمة  $Z_{IMC}$  المحسوبة اكبر من قيمة  $Z$  الجدولية عند مستوى دلالة معينة نقبل الفرضية البديلة وهذا يعني وجود اعتماد مكاني بين مفردات الظاهرة المدروسة، اما اذا كانت  $Z_{IMC}$  المحسوبة اصغر من قيمة  $Z$  الجدولية فإن ذلك يؤدي الى قبول الفرضية العدم بمعنى انه لا يوجد اعتماد مكاني بين مفردات الظاهرة المدروسة .

### 7-2 معيار متوسط الخطأ المطلق النسبي

يمثل معيار (MAPE) الوسيلة التي سيتم بها المقارنة بين مصفوفات الاوزان المكانية المستعملة في البحث ويتم حسابة من خلال الصيغة الرياضية التالية :<sup>(10)</sup>

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad \dots (22)$$

حيث ان مصفوفة الاوزان المكانية التي تحصل على القيمة الاصغر ستكون هي المصفوفة الافضل في تمثيل ظاهرة الاعتماد المكاني لمفردات الظاهرة المدروسة.

### 3- المبحث الثالث: الجانب التطبيقي

تتضمن هذه الفقرة التطبيق العملي لاختبار الكشف عن الاعتماد المكاني للبيانات الحقيقية المستعملة في البحث وبعدها تطبيق تقدير معلمات نموذج الانحدار الذاتي المكاني باستعمال طريقة بيز في ضل مصفوفة الاوزان المكانية المعدلة المبنية وفق معيار روك للتجاور ومصفوفة الاوزان المكانية المقترحة من قبل الباحث ومن ثم المقارنة بين النتائج باستعمال معيار متوسط الخطأ المطلق النسبي (MAPE) كما يلي:

#### 3-1 وصف عينة البحث

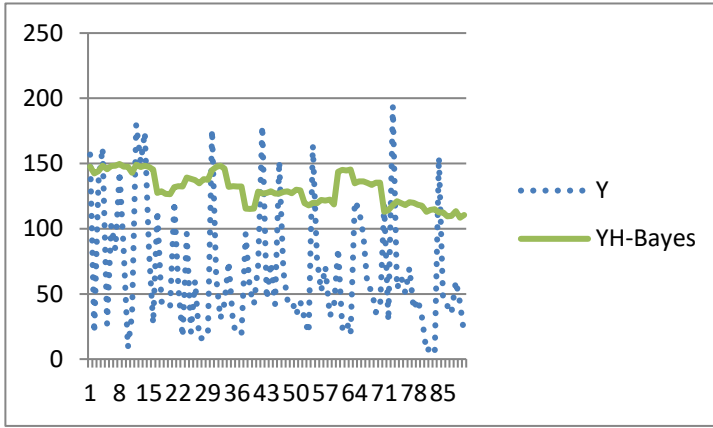
ان البيانات الحقيقية التي سيتم استعمالها في الجانب التطبيقي تعد بيانات مكانية لمرض السرطان موزعة على محافظات العراق الثمانية عشر والمتمثلة بقيم المتغيرات التوضيحية  $x$  (معدل العمر، معدل حجم الورم، عدد المناطق الملوثة باليورانيوم في المحافظة<sup>(8)</sup>) و بقيم المتغير المعتمد  $y$  والذي يمثل عدد المرضى في كل محافظة و يبلغ عدده 90 مشاهدة تم اختيارها بشكل عشوائي من محافظات العراق المختلفة ولعدة انواع من السرطان التي تمثلت باكثر انواع السرطان انتشارا في المحافظة لسنة 2018 ولقد تم الحصول على البيانات من وزارة الصحة /مركز السرطان، ومن خلال اختبار الكشف عن الاعتماد المكاني كانت نتيجة اختبار موران (4.0731) وعند المقارنة بين القيمة المحسوبة لاختبار موران والقيمة الجدولية التي تساوي (1.96) عند مستوى معنوية (0.05) نلاحظ ان القيمة الجدولية اصغر من القيمة المحسوبة وهذا دليل على وجود الاعتماد المكاني في البيانات.

### 2-3 التقدير بطريقة بيز في ضل مصفوفة الاوزان المكانية المعدلة

باستعمال البرنامج (Matlab) تم تقدير النموذج الانحدار الذاتي المكاني باستعمال طريقة بيز Bayes في ضل وجود مصفوفة الاوزان المكانية المعدلة والحصول على قيم المتغير المعتمد التقديرية  $\hat{y}$  المبينة في الجدول (1) الذي يضم القيم الحقيقية والتقديرية للمتغير المعتمد نفسة  $y$  الذي يمثل عدد مرضى السرطان، وبعد ان تم حساب القيم التقديرية تم ايجاد معيار متوسط الخطأ المطلق النسبي (MAPE) حيث كانت تبلغ قيمة (2.6724)، وكذلك تم تمثيل القيم الحقيقية والتقديرية للمتغير  $y$  في الجدول (1) بالرسم لبياني في الشكل (1). وبعد ان تم حساب القيم التقديرية تم ايجاد معيار متوسط الخطأ المطلق النسبي (MAPE) حيث كانت تبلغ قيمة (2.6724).

جدول (1) القيم الحقيقية والتقديرية للمتغير المعتمد  $y$  باستعمال طريقة بيز Bayes في ضل مصفوفة الاوزان المكانية المعدلة

| T  | y   | $\hat{y}$ | T  | y   | $\hat{y}$ | T  | y   | $\hat{y}$ | T  | y   | $\hat{y}$ |
|----|-----|-----------|----|-----|-----------|----|-----|-----------|----|-----|-----------|
| 1  | 157 | 148       | 24 | 96  | 139       | 47 | 67  | 128       | 70 | 39  | 135       |
| 2  | 20  | 143       | 25 | 21  | 138       | 48 | 45  | 129       | 71 | 118 | 113       |
| 3  | 136 | 144       | 26 | 60  | 137       | 49 | 46  | 127       | 72 | 29  | 114       |
| 4  | 162 | 148       | 27 | 16  | 135       | 50 | 34  | 130       | 73 | 194 | 118       |
| 5  | 27  | 146       | 28 | 16  | 138       | 51 | 44  | 129       | 74 | 54  | 121       |
| 6  | 103 | 148       | 29 | 18  | 137       | 52 | 31  | 120       | 75 | 63  | 120       |
| 7  | 84  | 148       | 30 | 176 | 145       | 53 | 20  | 118       | 76 | 51  | 118       |
| 8  | 144 | 150       | 31 | 63  | 147       | 54 | 164 | 120       | 77 | 70  | 120       |
| 9  | 83  | 148       | 32 | 31  | 148       | 55 | 77  | 120       | 78 | 38  | 120       |
| 10 | 9   | 148       | 33 | 43  | 146       | 56 | 52  | 122       | 79 | 46  | 118       |
| 11 | 34  | 143       | 34 | 75  | 132       | 57 | 71  | 122       | 80 | 32  | 118       |
| 12 | 180 | 149       | 35 | 26  | 133       | 58 | 33  | 122       | 81 | 6   | 113       |
| 13 | 151 | 147       | 36 | 20  | 132       | 59 | 40  | 119       | 82 | 9   | 114       |
| 14 | 174 | 148       | 37 | 19  | 132       | 60 | 85  | 144       | 83 | 5   | 115       |
| 15 | 83  | 147       | 38 | 96  | 115       | 61 | 24  | 145       | 84 | 155 | 113       |
| 16 | 27  | 145       | 39 | 54  | 115       | 62 | 29  | 145       | 85 | 45  | 112       |
| 17 | 114 | 127       | 40 | 42  | 116       | 63 | 19  | 145       | 86 | 41  | 110       |
| 18 | 44  | 129       | 41 | 72  | 128       | 64 | 121 | 135       | 87 | 35  | 110       |
| 19 | 44  | 127       | 42 | 179 | 127       | 65 | 115 | 136       | 88 | 60  | 114       |
| 20 | 40  | 127       | 43 | 46  | 127       | 66 | 95  | 136       | 89 | 43  | 108       |
| 21 | 121 | 132       | 44 | 74  | 129       | 67 | 55  | 135       | 90 | 19  | 110       |
| 22 | 54  | 133       | 45 | 42  | 127       | 68 | 53  | 134       |    |     |           |
| 23 | 17  | 132       | 46 | 152 | 127       | 69 | 36  | 135       |    |     |           |



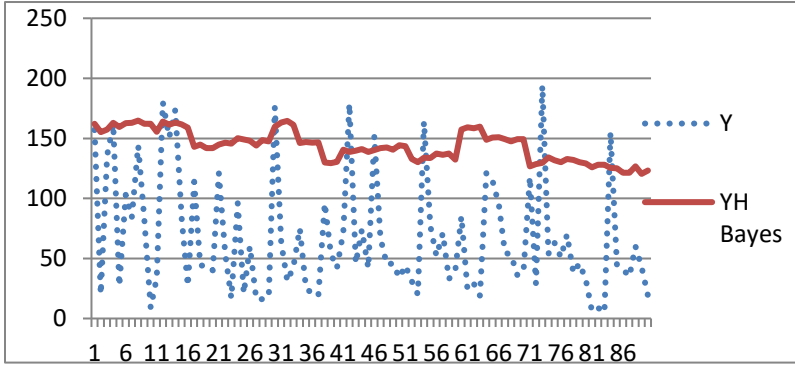
شكل (1) الرسم البياني للقيم الحقيقية للمتغير المعتمد  $y$  والقيم التقديرية  $\hat{y}$  باستعمال طريقة Bayes في ضل مصفوفة الاوزان المكانية المعدلة.

### 3-3 التقدير بطريقة بيز في ضل مصفوفة الاوزان المكانية المقترحة

باستعمال البرنامج (Matlab) تم تقدير انموذج الانحدار الذاتي المكاني باستعمال طريقة بيز Bayes في ضل مصفوفة الاوزان المكانية المقترحة والحصول على قيم المتغير المعتمد التقديرية  $\hat{y}$  المبينة في الجدول (2) الذي يضم القيم الحقيقية والتقديرية للمتغير المعتمد  $y$  الذي يمثل عدد مرضى السرطان، وكذلك تم تمثيل القيم الحقيقية والتقديرية للمتغير  $y$  في الجدول (2) بالرسم لبياني في الشكل (2). وبعد ان تم حساب القيم التقديرية تم ايجاد معيار متوسط الخطأ المطلق النسبي (MAPE) حيث كانت تبلغ قيمته (3.0239).

جدول (2) القيم الحقيقية والتقديرية للمتغير المعتمد  $y$  باستعمال طريقة بيز Bayes في ضل مصفوفة الاوزان المكانية المقترحة.

| T  | y   | $\hat{y}$ | T  | y   | $\hat{y}$ | T  | y   | $\hat{y}$ | T  | y   | $\hat{y}$ |
|----|-----|-----------|----|-----|-----------|----|-----|-----------|----|-----|-----------|
| 1  | 157 | 162       | 24 | 96  | 150       | 47 | 67  | 142       | 70 | 39  | 149       |
| 2  | 20  | 155       | 25 | 21  | 149       | 48 | 45  | 143       | 71 | 118 | 127       |
| 3  | 136 | 157       | 26 | 60  | 148       | 49 | 46  | 141       | 72 | 29  | 129       |
| 4  | 162 | 163       | 27 | 16  | 144       | 50 | 34  | 144       | 73 | 194 | 130       |
| 5  | 27  | 159       | 28 | 16  | 148       | 51 | 44  | 143       | 74 | 54  | 134       |
| 6  | 103 | 163       | 29 | 18  | 148       | 52 | 31  | 133       | 75 | 63  | 132       |
| 7  | 84  | 163       | 30 | 176 | 160       | 53 | 20  | 130       | 76 | 51  | 130       |
| 8  | 144 | 165       | 31 | 63  | 163       | 54 | 164 | 134       | 77 | 70  | 133       |
| 9  | 83  | 162       | 32 | 31  | 165       | 55 | 77  | 134       | 78 | 38  | 132       |
| 10 | 9   | 162       | 33 | 43  | 161       | 56 | 52  | 137       | 79 | 46  | 130       |
| 11 | 34  | 156       | 34 | 75  | 146       | 57 | 71  | 136       | 80 | 32  | 129       |
| 12 | 180 | 164       | 35 | 26  | 147       | 58 | 33  | 137       | 81 | 6   | 126       |
| 13 | 151 | 161       | 36 | 20  | 147       | 59 | 40  | 132       | 82 | 9   | 128       |
| 14 | 174 | 163       | 37 | 19  | 147       | 60 | 85  | 157       | 83 | 5   | 128       |
| 15 | 83  | 162       | 38 | 96  | 130       | 61 | 24  | 159       | 84 | 155 | 125       |
| 16 | 27  | 159       | 39 | 54  | 129       | 62 | 29  | 158       | 85 | 45  | 125       |
| 17 | 114 | 143       | 40 | 42  | 130       | 63 | 19  | 160       | 86 | 41  | 121       |
| 18 | 44  | 145       | 41 | 72  | 141       | 64 | 121 | 149       | 87 | 35  | 121       |
| 19 | 44  | 142       | 42 | 179 | 139       | 65 | 115 | 151       | 88 | 60  | 127       |
| 20 | 40  | 142       | 43 | 46  | 140       | 66 | 95  | 151       | 89 | 43  | 120       |
| 21 | 121 | 145       | 44 | 74  | 141       | 67 | 55  | 149       | 90 | 19  | 123       |
| 22 | 54  | 146       | 45 | 42  | 139       | 68 | 53  | 147       |    |     |           |
| 23 | 17  | 146       | 46 | 152 | 140       | 69 | 36  | 149       |    |     |           |



شكل (2) الرسم البياني للقيم الحقيقية للمتغير المعتمد  $y$  والقيم التقديرية  $\hat{y}$  باستعمال طريقة بيز Bayes في ضل مصفوفة الازان المكانية المقترحة .

### 4-3 تحليل نتائج التطبيقي

نلاحظ من خلال جدول رقم (1) و جدول رقم (2) ان قيم المتغير المعتمد التقديرية في ضل مصفوفة الازان المكانية المعدلة تكون اقرب للقيم الحقيقية من قيم المتغير المعتمد التقديرية في ضل مصفوفة الازان المكانية المقترحة ويمكن ملاحظة ذلك من خلال الرسم رقم (3) ، وهذا ما جعل قيمة معيار MAPE الذي تم حسابة في ضل مصفوفة الازان المكانية المعدلة اصغر من القيمة التي تم حسابها في ضل مصفوفة الازان المكانية المقترحة .

### 4- المبحث الرابع: الاستنتاجات والتوصيات

بعد تقدير معاملات انموذج الانحدار الذاتي المكاني وباستعمال طريقة بيز في التقدير في ضل وجود مصفوفة التجاور المكانية باستعمال معيار روك للتجاور ندرج ادناه اهم ما توصل اليه :

#### 1-4 الاستنتاجات

- 1- وجود فروقات كبيرة بين القيم الحقيقية والقيم المقدرة بطريقة بيز.
- 2- ان المصفوفة المكانية المعدلة التي تم بناءها بالاعتماد على معيار روك للتجاور افضل من مصفوفة الازان المكانية المقترحة من قبل الباحث عند الاعتماد على نتائج معيار ( MAPE)
- 3- وجود اعتماد مكاني لبيانات مرضى السرطان مما يدل على ان الموقع الجغرافي ذو تاثير على المرض.

#### 2-4 التوصيات

- 1- استعمال طرائق معلمية اخرى لتقدير انموذج الانحدار الذاتي المكاني .
- 2- استعمال مصفوفة الازان المكانية المعدلة المبنية وفق معيار روك للجوار كونها اظهرت كفاءتها .
- 3- كدراسات مستقبلية نوصي بالمقارنة بين مصفوفة التجاور المكانية المعدلة ومصفوفة المسافة.
- 4- كدراسات مستقبلية نوصي بدراسة مكانية للعوامل التي تؤدي الى الاصابة بمرض السرطان وتأثيرها على عدد الاصابات في كل منطقة .

## المصادر

- 1-Anselin, L. (1992), "Spatial Data Analysis with GIS: An Introduction to Appalachian the Social Sciences", National Center for Geographic Information and Analysis, University of California, Santa Barbara.
- 2- Anselin, L. and Bera, A. (1998). "Spatial dependence in linear regression models with an introduction to spatial econometrics ", Handbook of Applied Economic Statistics. Marcel Dekker, New York.
- 3- Gumprecht, D.,(2005), "Spatial Methods in Econometrics: An Application to R and D Spillovers", Department of statistics and Mathematics.
- 4- Lacombe, Donald J. (2008). " An Introduction to Bayesian Inference in Spatial Econometrics". Department of Economics Ohio University.
- 5- Lesage, J. P., ( 1999 ), "The Theory and Practice of Spatial Econometrics", Department of Economics University of Toledo.
- 6- Lesage, J. P. (2009). "Introduction to Spatial Econometrics". Statistics, a series of textbooks and monographs.
- 7- Michiel, D., Pooter, D., Segers, R. Herman, K. Dijk, v., (2006). " Gibbs Sampling in Econometric Practice", Econometric Institute and Tinbergen Institute Erasmus University Rotterdam, The Netherlands.
- 8- Oakford, S., (2016 )," Iraq War records reignite debate over US use of depleted uranium" .  
<https://www.thenewhumanitarian.org/investigations/2016/10/06/exclusive-iraq-war-records-reignite-debate-over-us-use-depleted-uranium>.
- 9- Permai1, S D., Mukhaiyar, U., Satyaning, N. L., Soleh, M., Aini, Q., (2018), "Spatial weighting approach in numerical method for dis Statistics Department, , Bina Nusantara University,Jakarta, Indonesia
- 10- Rahim, Sh. A., Hussein, M. F.,(2020), "A study on atmospheric pressure in krg using spatial regression (SAR and SEM)" A Scientific Quarterly Refereed Journal Issued by Lebanese French University.

## دراسة تحليلية لمعدلات البطالة في العراق باستخدام تقنية البوتستراپ

م. فراس منذر جاسم      م.د. زينب فالح حمزة      م. ثائرة نجم عبد الله

تاريخ استلام البحث: 2022/04/24

تاريخ قبول البحوث: 2022/05/11

نشر البحث في العدد عشرون: ايلول / سبتمبر 2023

رمز التصنيف ديوي / النسخة الالكترونية (Online): 2522-64X/313.13

رمز التصنيف ديوي / النسخة الورقية (Print): 2519-948X/313.13

## دراسة تحليلية لمعدلات البطالة في العراق باستخدام تقنية البوتستراب

م. فراس منذر جاسم      م.د. زينب فالح حمزة      م. ثائرة نجم عبد الله

### الملخص:

يقدم البحث دراسة إحصائية تحليلية لمعدلات البطالة في العراق للسنوات (2008-2003)، إذ تم توظيف تجارب المحاكاة بمعلمات إفتراضية وأحجام عينات مختلفة في عملية التقدير من خلال استخدام طريقة الإمكان الأعظم، وتقنية البوتستراب بأسلوب الوسيط والمتوسط، تم إجراء المقارنة لتلك التقديرات بالاعتماد على متوسط مربعات الخطأ، وقد رجحت تجارب المحاكاة تقنية البوتستراب وفقاً لإسلوب الوسيط، لذلك تم استخدامها في تقدير وتحليل معلمة ومتوسط زمن البقاء ودالة البقاء لبيانات معدلات البطالة في العراق التي تتوفر بستة مفردات تتوزع التوزيع الأسي. أثبتت النتائج إن معدل البطالة في العراق يتراوح عند (20%) وإن متوسط زمن بقاء هذا المعدل هو (20) سنة، وإن احتمال بقاء هذا المعدل على حاله هو (63%) في العام الحالي، وسيكون (49.5%) في عام (2023)، و (38.5%) في العام (2028).

**الكلمات المفتاحية:** التوزيع الأسي، دالة البقاء، تقنية البوتستراب، معدلات البطالة.

### An analytical study of unemployment rates in Iraq using the Bootstrap technique

#### Abstract

The research presents a statistical analysis of the unemployment rates in Iraq since 2003-2008. A simulation experiments were use procedure with different parameters and different sizes of samples in the estimation process which include the Maximum Likelihood Method and the Bootstrap technique with Mean and Median methods. A comparison was done dependent on the Mean Square Error. The simulations outbalanced the second Bootstrap technique, so it was use to estimate and analyze the parameter, average survival time, and survival function of the unemployment data in Iraq, which is available in six observations distributed exponentially. The results show that the unemployment rate in Iraq ranges in (20%) and the average survival rate is (20) years, and the probability of survival of this rate is the same Circumstances was (63%) this year and (49.5%) in (2023) and (38.5%) in the (2028).

#### 1- المقدمة والمشكلة والهدف

تعتبر البطالة من المشاكل الخطيرة التي تواجه الاقتصاد العالمي عامة والاقتصاد العراقي بصورة خاصة لما لها من آثار اجتماعية واقتصادية وسياسية سلبية. إن الدراسات الاقتصادية في مجال تحليل معدلات البطالة في العراق يحتاج الى تحليل دقيق لا يقتصر على سوق العمل لكونها تتأثر بالطلب والعرض فقط، بل هناك ظروف أخرى تجعلها تتأثر بعوامل خارجية وداخلية

كثيرة قد يصعب السيطرة على البعض منها، مما أدى الى وجود تفاوت في تقديرات معدلات البطالة إذ بلغ معدل البطالة (28.1%) في المسح الذي تم إجراؤه عام (2003) حسب مفهوم منظمة العمل الدولية في حين قدرت الأمم المتحدة معدل البطالة في العراق بحوالي (50%). إن هذا الاختلاف في تقدير معدل البطالة قد يعود الى تداخل وتعدد العوامل الاقتصادية المؤثرة عليه كمعدل النمو والناتج المحلي والاستثمار وغيرها. أما من الجانب الإحصائي فإن التعامل مع عينة صغيرة جدا قد يكون أمر صعب وغير مقنع، كما وإن تساؤلا مهما قد يطرح هنا (كم ستستمر مشكلة المعدلات العالية للبطالة في العراق إذا ما بقيت الظروف كما هي؟). لذلك تم في هذا البحث توظيف بعض التقنيات والأساليب الإحصائية التي تتعامل بدقة مع أحجام العينات الصغيرة وهي تقنية البوتستراتب وفق اسلويين هما المتوسط والوسيط، فضلا عن تطبيق طريقة الإمكان الأعظم، ومن خلال تجارب المحاكاة المختلفة تم تقدير معلمة وزمن البقاء ودالة البقاء بهدف إجراء المقارنة بين تلك التقديرات وبالاعتماد على متوسط مربعات الخطأ كمعيار مقارنة، ومن ثم استخدام الطريقة أو الاسلوب الأفضل في دراسة وتحليل معدلات البطالة لإيجاد المعلمة ومتوسط زمن البقاء ودالة البقاء على البيانات الحقيقية التي تبين إنها تتبع التوزيع الأسي .

## 2- الجانب النظري

### 1-2 التوزيع الأسي : Exponential Distribution

الصيغة التالية تمثل دالة الكثافة الاحتمالية (Probability Density Function) للمتغير العشوائي ( $v$ ) الذي يتبع التوزيع الأسي بمعلمة القياس ( $\tau$ ):

$$f(v; \tau) = \frac{1}{\tau} e^{-\frac{v}{\tau}} \quad \tau, v > 0 \quad \dots (1)$$

وإن دالة الكثافة التجميعية (Probability Cumulative Function) تكون وفق الصيغة التالية:

$$F(v; \tau) = Pr(V \leq v) = \int_0^v f(v) dv = 1 - e^{-\frac{v}{\tau}} \quad \dots (2)$$

كذلك فإن دالة البقاء (Survival Function) أو دالة المعولية (Reliability Function) هي:

$$S(v; \tau) = 1 - F(v; \tau) = e^{-\frac{v}{\tau}} \quad \dots (3)$$

كما تعرف دالة المخاطرة (Hazard Function) كما يأتي:

$$h(v; \tau) = \frac{f(v; \tau)}{S(v; \tau)} = \frac{\frac{1}{\tau} e^{-\frac{v}{\tau}}}{e^{-\frac{v}{\tau}}} = \frac{1}{\tau} \quad \dots (4)$$

وإن متوسط الزمن المستغرق لحين حدوث الوفاة (الفشل) (MTTD) هو:

$$MTTD = \int_0^{\infty} S(v; \tau) dv = \int_0^{\infty} e^{-\frac{v}{\tau}} dv = \tau \quad \dots (5)$$



## 2-2 طرائق تقدير معلمة ودالة معولية التوزيع الأسي:

## 1-2-2 طريقة الإمكان الأعظم : Maximum Likelihood method

إذا كانت  $(v_1, v_2, \dots, v_n)$  تمثل عينة عشوائية مستقلة لمشاهدات متغير يتبع التوزيع الأسي، فإن دالة الإمكان لمشاهدات المتغير العشوائي  $(v)$  تكون:

$$L(v_i|\tau) = \prod_{i=1}^n f(v_i|\tau) = \prod_{i=1}^n \frac{1}{\tau} e^{-\frac{v_i}{\tau}} = \tau^{-n} e^{-\frac{1}{\tau} \sum_{i=1}^n v_i} \quad \dots (6)$$

نحصل على صيغة مقدر طريقة الإمكان الأعظم للمعلمة  $(\tau)$  وذلك بأخذ اللوغاريتم الطبيعي للمعادلة (6)، ومن ثم الإشتقاق الجزئي بالنسبة للمعلمة  $(\tau)$ ، ثم مساواة المشتقة بالصفر:

$$\begin{aligned} \ln L(v_i|\tau) &= -n \ln(\tau) - \frac{\sum_{i=1}^n v_i}{\tau} \\ \frac{\partial L(v_i|\tau)}{\partial \tau} &= -\frac{n}{\tau} + \frac{\sum_{i=1}^n v_i}{\tau^2} = \frac{-n\tau + \sum_{i=1}^n v_i}{\tau^2} \\ \frac{\partial L(v_i|\tau)}{\partial \tau} &= 0 \end{aligned}$$

$$-n\hat{\tau} + \sum_{i=1}^n v_i = 0$$

$$\hat{\tau}_{ML} = \frac{\sum_{i=1}^n v_i}{n} = \bar{v} \quad \dots (7)$$

أما فيما يخص تقدير دالة البقاء للتوزيع الأسي، فيمكن تقديرها إستناداً الى خاصية الثبات (Invariance Property) التي تمتع بها طريقة الإمكان الأعظم، وذلك بتعويض المعادلة (7) في المعادلة (3) وفق الصيغة التالية:

$$\hat{S}_{ML} = e^{-\frac{v}{\hat{\tau}_{ML}}} = e^{-\frac{v}{\bar{v}}} \quad \dots (8)$$

## 2-2-2 تقنية البوتستراپ Bootstrap Technique

تقنية البوتستراپ من أساليب إعادة المعاينة التي تهدف الى الحصول على تقديرات وتنبؤات تتصف بخواص وصفات معينة، وتعتمد هذه التقنية على استخدام بيانات العينة نفسها في تقدير احصاءة معينة بشروط وفروض بسيطة يخضع لها النموذج المعني بالدراسة، وبذلك فإنه يمكن من خلال توظيف هذه التقنية تجاوز العديد من المشاكل التي قد تواجه عملية التقدير أو التنبؤ مثل وجود القيم الشاذة في البيانات المدروسة، التعقيد الرياضي في صيغ التقدير، الضبابية في معرفة التوزيع الاحتمالي الذي تتبعه البيانات، صغر حجم العينة المتوفر لتوضيح الفكرة الأساسية التي تعتمد عليها تقنية البوتستراپ نفترض إن عينة عشوائية بحجم  $(n)$  لها دالة التوزيع الاحتمالية  $(F)$ ، عندها يمكن سحب عينة بحجم  $(m)$  تعرف بعينة البوتستراپ، ويتم تطبيق طريقة التقدير المطلوبة من تلك العينة، وإرجاع العينة الأولى يمكن سحب عينة ثانية بنفس الحجم وإجراء التقدير مرة أخرى وهكذا الى أن يتم الحصول على  $(D)$  من عينات البوتستراپ والمقدرات، عندها فإن المقدر المطلوب إيجاده وفق تقنية البوتستراپ سيكون عبارة عن متوسط أو وسيط مقدرات عينات البوتستراپ. تجدر الإشارة الى إنه ليس هنالك عدد محدد لعينات البوتستراپ التي ينبغي سحبها إذ يختلف عدد تلك العينات باختلاف الغرض الذي من اجله تم استخدام هذه التقنية. وقد تمت مناقشة مسألة عدد عينات البوتستراپ بتعمق واستفاضة من قبل الكثير من الباحثين وتم تحديدها بين (100-1000)

مفردة حسب عدد المشاهدات المتوفرة للمتغيرات العشوائية المدروسة وحسب درجة تعقيد الصيغ الرياضية التي تعتمد عليها عملية التقدير.

### 3- الجانب التجريبي

تضمنت تجارب المحاكاة في هذا البحث والتي تم تنفيذها باستخدام البرنامج الإحصائي  $(R_{3.5.1})$  المراحل والخطوات التالية:

**المرحلة الأولى:** اختيار قيم افتراضية لمعلمة القياس  $(\tau)$  وإحجام عينات مختلفة وقد تم اختيار القيم التالية وعلى التوالي  $(n = 5, 10, 30, 100)$  ،  $(\tau = 0.05, 0.1, 1)$  ، بينما تم سحب (500) عينة بوتستراب بحجم  $(m = 25)$  .

**المرحلة الثانية:** توليد المتغيرات العشوائية  $(V_1, V_2, \dots, V_j)$  التي تتبع التوزيع الأسّي بالمعلمة  $(\tau)$  واستنادا الى القيم الافتراضية وأحجام العينات التي تم تحديدها في المرحلة السابقة ، وذلك باستخدام طريقة التحويل العكسي (Inverse Transform Method) التي تتضمن الخطوات التالية:

$$1- \text{توليد المتغير العشوائي } U \sim \text{Uniform}(0,1)$$

2- استخدام التحويل

$$U = F(v) = 1 - e^{-\frac{v}{\tau}}$$

$$V = F^{-1}(U) = -\frac{1}{\tau} \text{Log}(1 - u)$$

**المرحلة الثالثة:** إيجاد تقديرات معلمة القياس  $(\tau)$  ودالة البقاء  $S(v; \tau)$  حسب الطرائق التي تم تناولها في الجانب النظري وهي :

1- طريقة الإمكان الأعظم (MLM) : من خلال المعادلتين (7) و (8) يمكن إيجاد  $\hat{\tau}_{(ML)j}$  و  $\hat{S}_{(ML)j}$  ، وبإعادة الخطوات 1 ، 2 من المرات ، إذ إن  $(j)$  يمثل عدد تكرارات عملية التوليد  $(j = 1000)$  ، عندها فإن مقدري طريقة الإمكان الأعظم للمعلمة ودالة البقاء سيكونان :

$$\hat{\tau}_{(ML)} = \frac{\sum_{j=1}^{1000} \hat{\tau}_{(ML)j}}{1000} , \quad \hat{S}_{(ML)} = \frac{\sum_{j=1}^{1000} \hat{S}_{(ML)j}}{1000}$$

2- تقنية البوتستراب : يتم إيجاد المقدرات وفقا للخطوات التالية :

- سحب عينة بوتستراب ولتكن  $V_1 \sim (v_1, v_2, \dots, v_m)$  بحجم  $(m)$  من بيانات

العينة الأصلية التي تم توليدها  $V = (v_1, v_2, \dots, v_n)$  .

- إيجاد  $\hat{\tau}_{B_1(ML)j}$  و  $\hat{S}_{B_1(ML)j}$  من المعادلتين (7) و (8) لعينة البوتستراب الأولى ، ثم إرجاع تلك العينة البيانات .

- تكرار عملية سحب عينات البوتستراب  $(D = 500)$  وإيجاد التقدير للمعلمة ودالة البقاء لكل العينات

$V_{11} \sim (v_1, v_2, \dots, v_m), V_{21} \sim (v_1, v_2, \dots, v_m), \dots, V_{D1} \sim (v_1, v_2, \dots, v_m)$

عندها نحصل على  $(D)$  من مقدرات المعلمة  $(\tau)$  للعينة الأولى التي تم توليدها وهي  $(\hat{\tau}_{B_1(ML)1}, \hat{\tau}_{B_2(ML)1}, \dots, \hat{\tau}_{B_{500}(ML)1})$  ، وكذلك سنحصل  $(D)$  من

مقدرات دالة البقاء وهي  $(\hat{S}_{B_1(ML)1}, \hat{S}_{B_2(ML)1}, \dots, \hat{S}_{B_{500}(ML)1})$  .

- يمكن إيجاد مقدري تقنية البوتستراب بأسلوب المتوسط للعينة الأولى التي تم توليدها كما يأتي:

$$\hat{\tau}^{\sim}_{(BM)1} = \frac{\sum_{D=1}^{500} \hat{\tau}^{\sim}_{(B)1}}{500}, \quad \hat{S}^{\sim}_{(BM)1} = \frac{\sum_{D=1}^{500} \hat{S}^{\sim}_{(B)1}}{500}$$

أما بالنسبة لاسلوب الوسيط فيمكن إيجادهما بعد ترتيب المقدرات تصاعديا كما يأتي :

$$\hat{\tau}^{\sim}_{(Bm)1} = \text{median}\{\hat{\tau}^{\sim}_{(B)1}\}, \quad \hat{S}^{\sim}_{(Bm)1} = \text{median}\{\hat{S}^{\sim}_{(B)1}\}$$

- يتم إعادة خطوات البوتستراب أعلاه للحصول على مقدري اسلوب البوتستراب لمعلمة القياس ( $\tau$ ) ودالة البقاء  $S(v; \tau)$  ووفق إسلوب المتوسط والوسيط ، كما يأتي :

$$\hat{\tau}^{\sim}_{(B)} = \frac{\sum_{j=1}^{1000} \hat{\tau}^{\sim}_{(B)j}}{1000}, \quad \hat{R}^{\sim}_{(B)} = \frac{\sum_{j=1}^{1000} \hat{R}^{\sim}_{(B)j}}{1000}$$

**المرحلة الرابعة:** حساب متوسط مربعات الخطأ (MSE) لمقدرات طريقة الإمكان الأعظم التي تم إيجادها في الخطوات السابقة وفق الصيغ التالية :

$$MSE(\hat{\tau}_{ML}) = \frac{\sum_{j=1}^{1000} \{\tau - \hat{\tau}_{(ML)j}\}^2}{1000}, \quad MSE(\hat{R}_{(ML)}) = \frac{\sum_{j=1}^{1000} \{S - \hat{S}_{(ML)j}\}^2}{1000}$$

وحساب متوسط مربعات الخطأ (MSE) لمقدرات تقنية البوتستراب بأسلوب المتوسط سيكون :

$$MSE(\hat{\tau}^{\sim}_{BMj}) = \frac{\sum_{D=1}^{500} \{\tau - \hat{\tau}^{\sim}_{(BM)D}\}^2}{500}, \quad MSE(\hat{R}^{\sim}_{BMj}) = \frac{\sum_{D=1}^{500} \{S - \hat{S}^{\sim}_{(BM)D}\}^2}{500}$$

$$MSE(\hat{\tau}^{\sim}_{BM}) = \frac{\sum_{j=1}^{1000} MSE(\hat{\tau}^{\sim}_{BMj})}{1000}, \quad MSE(\hat{S}^{\sim}_{BM}) = \frac{\sum_{j=1}^{1000} MSE(\hat{S}^{\sim}_{BMj})}{1000}$$

وبأسلوب الوسيط هو:

$$MSE(\hat{\tau}^{\sim}_{Bmj}) = \text{median}[MSE\{\hat{\tau}^{\sim}_{(Bm)D}\}], \quad MSE(\hat{S}^{\sim}_{Bmj}) = \text{median}[MSE\{\hat{S}^{\sim}_{(Bm)D}\}]$$

$$MSE(\hat{\tau}^{\sim}_{Bm}) = \frac{\sum_{j=1}^{1000} \{\tau - \hat{\tau}^{\sim}_{(Bm)j}\}^2}{1000}, \quad MSE(\hat{S}^{\sim}_{Bm}) = \frac{\sum_{j=1}^{1000} \{S - \hat{S}^{\sim}_{(Bm)j}\}^2}{1000}$$

### 1-3 عرض وتحليل نتائج المحاكاة:

الجداول من (1) الى (6) توضح نتائج تجارب المحاكاة وكما يأتي:

جدول (1) تقديرات المعلمة ( $\tau$ ) وقيم متوسطات مربعات الخطأ ( $MSE$ )

| $\tau_{Real}$ | N   | $\hat{\tau}_{ML}$ | MSE       | $\hat{\tau}_{BM}$ | MSE      | $\hat{\tau}_{Bm}$ | MSE      |
|---------------|-----|-------------------|-----------|-------------------|----------|-------------------|----------|
| 20            | 5   | 16.130671         | 14.971960 | 17.835788         | 4.684025 | 18.760730         | 1.535980 |
|               | 10  | 18.402995         | 2.550522  | 18.979539         | 1.041434 | 19.371353         | 0.395286 |
|               | 30  | 19.279959         | 0.518504  | 19.560759         | 0.192976 | 19.573176         | 0.182223 |
|               | 100 | 19.851913         | 0.021951  | 19.710694         | 0.083719 | 19.889806         | 0.012164 |
| 10            | 5   | 7.289427          | 7.348462  | 9.044323          | 0.914134 | 9.400378          | 0.360301 |
|               | 15  | 9.106639          | 0.798496  | 9.556867          | 0.196732 | 9.631443          | 0.136193 |
|               | 30  | 9.750533          | 0.062409  | 9.832378          | 0.028270 | 9.836043          | 0.027054 |
|               | 100 | 9.914055          | 0.007471  | 9.884480          | 0.013430 | 9.963444          | 0.001420 |
| 5             | 5   | 4.761186          | 0.059973  | 5.041474          | 0.004343 | 5.021449          | 0.003104 |
|               | 10  | 4.826110          | 0.031669  | 5.020572          | 0.001746 | 5.008853          | 0.001407 |
|               | 30  | 4.917122          | 0.007558  | 4.983249          | 0.000952 | 5.000325          | 0.000667 |
|               | 100 | 4.992272          | 0.000394  | 4.982576          | 0.000639 | 5.001203          | 0.000335 |
| 2             | 5   | 1.549538          | 0.230682  | 1.822017          | 0.051760 | 1.862478          | 0.038131 |
|               | 10  | 1.787957          | 0.055389  | 1.908602          | 0.017504 | 1.936729          | 0.012890 |
|               | 30  | 1.895745          | 0.015507  | 1.955225          | 0.006364 | 1.968793          | 0.005274 |
|               | 100 | 1.974998          | 0.002762  | 1.920360          | 0.008602 | 1.983211          | 0.002401 |
| 1             | 5   | 0.830858          | 0.125182  | 0.903320          | 0.091048 | 0.941602          | 0.078603 |
|               | 10  | 0.882389          | 0.056644  | 0.963194          | 0.037284 | 0.969422          | 0.036404 |
|               | 30  | 0.982698          | 0.017558  | 0.977281          | 0.017967 | 0.984369          | 0.017444 |
|               | 100 | 0.985662          | 0.008783  | 0.984502          | 0.008838 | 0.991594          | 0.008546 |

جدول (2) تقديرات دالة البقاء التوزيع الأسّي وقيم متوسطات مربعات الخطأ عندما  $\tau = 20$

| n   | t | $S_{Rael}$ | $\hat{S}_{ML}$ | MSE     | $\hat{S}_{BM}$ | MSE     | $\hat{S}_{Bm}$ | MSE     |
|-----|---|------------|----------------|---------|----------------|---------|----------------|---------|
| 5   | 1 | 0.95122    | 0.93988        | 0.01736 | 0.94547        | 0.00460 | 0.94809        | 0.00118 |
|     | 5 | 0.77880    | 0.73347        | 0.27740 | 0.75553        | 0.07529 | 0.76604        | 0.01952 |
|     | 1 | 0.60653    | 0.53797        | 0.63442 | 0.57082        | 0.17726 | 0.58682        | 0.04660 |
|     | 2 | 0.36787    | 0.28942        | 0.83102 | 0.32584        | 0.24571 | 0.34436        | 0.06637 |
| 10  | 1 | 0.95122    | 0.94867        | 0.00088 | 0.94711        | 0.00235 | 0.94968        | 0.00028 |
|     | 5 | 0.77880    | 0.76208        | 0.03771 | 0.76840        | 0.01503 | 0.77250        | 0.00475 |
|     | 1 | 0.60653    | 0.58077        | 0.08954 | 0.59044        | 0.03598 | 0.59676        | 0.01143 |
|     | 2 | 0.36787    | 0.33730        | 0.12623 | 0.34862        | 0.05156 | 0.35613        | 0.01656 |
| 30  | 1 | 0.95122    | 0.94945        | 0.00042 | 0.95016        | 0.00015 | 0.95019        | 0.00012 |
|     | 5 | 0.77880    | 0.77156        | 0.00707 | 0.77444        | 0.00264 | 0.77456        | 0.00215 |
|     | 1 | 0.60653    | 0.59531        | 0.01699 | 0.59975        | 0.00637 | 0.59995        | 0.00519 |
|     | 2 | 0.36787    | 0.35439        | 0.02454 | 0.35971        | 0.00928 | 0.35994        | 0.00755 |
| 100 | 1 | 0.95122    | 0.95087        | 0.00002 | 0.95053        | 0.00007 | 0.95096        | 0.00001 |
|     | 5 | 0.77880    | 0.77735        | 0.00028 | 0.77594        | 0.00113 | 0.77772        | 0.00013 |
|     | 1 | 0.60653    | 0.60427        | 0.00069 | 0.60209        | 0.00274 | 0.60485        | 0.00034 |
|     | 2 | 0.36787    | 0.36514        | 0.00100 | 0.36251        | 0.00399 | 0.36584        | 0.00049 |

جدول (3) تقديرات دالة البقاء للتوزيع الأسّي وقيم متوسطات مربعات الخطأ عندما  $\tau = 10$ 

| n   | t | $S_{Rael}$ | $\hat{S}_{ML}$ | MSE     | $\hat{S}_{BM}$ | MSE     | $\hat{S}_{Bm}$ | MSE     |
|-----|---|------------|----------------|---------|----------------|---------|----------------|---------|
| 5   | 1 | 0.90483    | 0.87180        | 0.01472 | 0.89532        | 0.00116 | 0.89908        | 0.00040 |
|     | 5 | 0.60653    | 0.50362        | 0.14296 | 0.57531        | 0.01254 | 0.58749        | 0.00439 |
|     | 1 | 0.36787    | 0.25363        | 0.17619 | 0.33099        | 0.01752 | 0.34514        | 0.00626 |
|     | 2 | 0.13533    | 0.06433        | 0.06806 | 0.10955        | 0.00855 | 0.11912        | 0.00318 |
| 10  | 1 | 0.90483    | 0.89600        | 0.00105 | 0.90065        | 0.00022 | 0.90138        | 0.00014 |
|     | 5 | 0.60653    | 0.57749        | 0.01137 | 0.59263        | 0.00248 | 0.59503        | 0.00160 |
|     | 1 | 0.36787    | 0.33350        | 0.01595 | 0.35121        | 0.00357 | 0.35406        | 0.00231 |
|     | 2 | 0.13533    | 0.11122        | 0.00784 | 0.12334        | 0.00185 | 0.12536        | 0.00120 |
| 30  | 1 | 0.90483    | 0.90252        | 0.00007 | 0.90329        | 0.00003 | 0.90333        | 0.00002 |
|     | 5 | 0.60653    | 0.59882        | 0.00080 | 0.60138        | 0.00034 | 0.60149        | 0.00030 |
|     | 1 | 0.36787    | 0.35858        | 0.00116 | 0.36166        | 0.00049 | 0.36179        | 0.00044 |
|     | 2 | 0.13533    | 0.12858        | 0.00061 | 0.13079        | 0.00026 | 0.13089        | 0.00023 |
| 100 | 1 | 0.90483    | 0.90405        | 0.00000 | 0.90378        | 0.00001 | 0.90450        | 0.00000 |
|     | 5 | 0.60653    | 0.60390        | 0.00009 | 0.60299        | 0.00016 | 0.60541        | 0.00001 |
|     | 1 | 0.36787    | 0.36470        | 0.00013 | 0.36360        | 0.00023 | 0.36653        | 0.00002 |
|     | 2 | 0.13533    | 0.13300        | 0.00007 | 0.13220        | 0.00012 | 0.13434        | 0.00001 |

جدول (4) تقديرات دالة البقاء للتوزيع الأسّي وقيم متوسطات مربعات الخطأ عندما  $\tau = 5$ 

| n   | t | $S_{Rael}$ | $\hat{S}_{ML}$ | MSE      | $\hat{S}_{BM}$ | MSE      | $\hat{S}_{Bm}$ | MSE      |
|-----|---|------------|----------------|----------|----------------|----------|----------------|----------|
| 5   | 1 | 0.81873    | 0.00001        | 0.00090  | 0.82007        | 0.00002  | 0.81943        | 0.00000  |
|     | 5 | 0.36787    | 0.00008        | 0.00437  | 0.37091        | 0.00011  | 0.36945        | 0.00003  |
|     | 1 | 0.13533    | 0.00004        | 0.00225  | 0.13758        | 0.00006  | 0.13649        | 0.00001  |
|     | 2 | 0.01831    | 0.00000        | 0.00015  | 0.01892        | 0.00000  | 0.01863        | 0.00000  |
| 10  | 1 | 0.95122    | 0.00478        | 0.25850  | 0.81902        | 0.22504  | 0.81940        | 0.21071  |
|     | 5 | 0.77880    | 0.04493        | 2.42629  | 0.36853        | 2.16714  | 0.36939        | 2.03235  |
|     | 1 | 0.60653    | 0.05774        | 3.11824  | 0.13581        | 2.85276  | 0.13644        | 2.67934  |
|     | 2 | 0.01831    | 0.00000        | 0.00008  | 0.01861        | 0.00000  | 0.01844        | 2.78E-10 |
| 30  | 1 | 0.95122    | 0.00457        | 0.24696  | 0.81818        | 0.22791  | 0.81874        | 0.21283  |
|     | 5 | 0.77880    | 0.04348        | 2.34829  | 0.36664        | 2.18710  | 0.36790        | 2.04714  |
|     | 1 | 0.60653    | 0.05656        | 3.05468  | 0.13442        | 2.86958  | 0.13535        | 2.69185  |
|     | 2 | 0.01831    | 0.00000        | 0.00001  | 0.01807        | 0.00000  | 0.01832        | 2.07E-07 |
| 100 | 1 | 0.95122    | 0.81847        | 0.23791  | 0.00442        | 0.22798  | 0.81877        | 0.21273  |
|     | 5 | 0.77880    | 0.36731        | 2.28587  | 0.04247        | 2.18763  | 0.36796        | 2.04650  |
|     | 1 | 0.60653    | 0.13491        | 3.00266  | 0.05572        | 2.87002  | 0.13540        | 2.69131  |
|     | 2 | 0.01831    | 0.01820        | 1.70E-07 | 1.62E-08       | 8.30E-07 | 0.01833        | 3.81E-09 |

جدول (5) تقديرات دالة البقاء للتوزيع الأسّي وقيم متوسطات مربعات الخطأ عندما  $\tau = 2$ 

| n   | t  | $S_{Rael}$ | $\hat{S}_{ML}$ | MSE    | $\hat{S}_{BM}$ | MSE    | $\hat{S}_{Bm}$ | MSE    |
|-----|----|------------|----------------|--------|----------------|--------|----------------|--------|
| 5   | 1  | 0.606531   | 0.524477       | 9.09E- | 0.577618       | 1.08E- | 0.584546       | 5.86E- |
|     | 5  | 0.082085   | 0.039686       | 2.43E- | 0.064299       | 9.63E- | 0.068249       | 2.32E- |
|     | 10 | 0.006738   | 0.001575       | 3.60E- | 0.004134       | 8.73E- | 0.004658       | 5.25E- |
|     | 20 | 0.000045   | 0.000002       | 2.00E- | 0.000017       | 1.00E- | 0.000022       | 6.88E- |
| 10  | 1  | 0.606531   | 0.592181       | 2.78E- | 0.571610       | 1.57E- | 0.596704       | 1.17E- |
|     | 5  | 0.082085   | 0.072823       | 1.16E- | 0.061024       | 5.71E- | 0.075647       | 5.03E- |
|     | 10 | 0.006738   | 0.005303       | 2.78E- | 0.003724       | 1.17E- | 0.005723       | 1.25E- |
|     | 20 | 0.000045   | 0.000028       | 3.66E- | 0.000014       | 1.00E- | 0.000033       | 1.96E- |
| 30  | 1  | 0.606531   | 0.590080       | 3.65E- | 0.599625       | 6.14E- | 0.601743       | 2.78E- |
|     | 5  | 0.082085   | 0.071541       | 1.50E- | 0.077518       | 2.69E- | 0.078896       | 1.23E- |
|     | 10 | 0.006738   | 0.005118       | 3.54E- | 0.006009       | 6.84E- | 0.006225       | 3.20E- |
|     | 20 | 0.000045   | 0.000026       | 4.52E- | 0.000036       | 1.06E- | 0.000039       | 6.88E- |
| 100 | 1  | 0.606531   | 0.603969       | 8.86E- | 0.594083       | 1.99E- | 0.602704       | 1.78E- |
|     | 5  | 0.082085   | 0.080366       | 3.99E- | 0.074001       | 8.41E- | 0.079528       | 7.93E- |
|     | 10 | 0.006738   | 0.006459       | 1.05E- | 0.005476       | 2.05E- | 0.006325       | 2.07E- |
|     | 20 | 0.000045   | 0.000042       | 1.66E- | 0.000030       | 2.91E- | 0.000040       | 3.57E- |

جدول (6) تقديرات دالة البقاء للتوزيع الأسّي وقيم متوسطات مربعات الخطأ عندما  $\tau = 1$ 

| $\tau = 1$ |    |            |                |          |                |          |                |          |
|------------|----|------------|----------------|----------|----------------|----------|----------------|----------|
| n          | t  | $S_{Rael}$ | $\hat{S}_{ML}$ | MSE      | $\hat{S}_{BM}$ | MSE      | $\hat{S}_{Bm}$ | MSE      |
| 5          | 1  | 3.68E-01   | 3.00E-01       | 6.20E-02 | 3.31E-01       | 1.80E-02 | 3.46E-01       | 5.93E-03 |
|            | 5  | 6.74E-03   | 2.43E-03       | 2.50E-04 | 3.95E-03       | 1.00E-04 | 4.94E-03       | 3.91E-05 |
|            | 10 | 4.54E-05   | 5.93E-06       | 2.00E-08 | 1.56E-05       | 1.00E-08 | 2.44E-05       | 1.00E-08 |
|            | 20 | 2.06E-09   | 3.51E-11       | 5.23E-17 | 2.42E-10       | 4.09E-17 | 5.96E-10       | 2.60E-17 |
| 10         | 1  | 3.68E-01   | 3.22E-01       | 2.84E-02 | 3.54E-01       | 2.45E-03 | 3.56E-01       | 1.58E-03 |
|            | 5  | 6.74E-03   | 3.46E-03       | 1.45E-04 | 5.57E-03       | 1.77E-05 | 5.75E-03       | 1.17E-05 |
|            | 10 | 4.54E-05   | 1.20E-05       | 2.00E-08 | 3.10E-05       | 2.57E-09 | 3.31E-05       | 1.83E-09 |
|            | 20 | 2.06E-09   | 1.43E-10       | 4.69E-17 | 9.60E-10       | 1.50E-17 | 1.10E-09       | 1.13E-17 |
| 30         | 1  | 3.68E-01   | 3.59E-01       | 9.65E-04 | 3.61E-01       | 5.31E-04 | 3.62E-01       | 4.07E-04 |
|            | 5  | 6.74E-03   | 6.00E-03       | 7.38E-06 | 6.17E-03       | 4.15E-06 | 6.22E-03       | 3.21E-06 |
|            | 10 | 4.54E-05   | 3.60E-05       | 1.13E-09 | 3.81E-05       | 6.65E-10 | 3.87E-05       | 5.39E-10 |
|            | 20 | 2.06E-09   | 1.29E-09       | 7.49E-18 | 1.45E-09       | 4.63E-18 | 1.50E-09       | 3.81E-18 |
| 100        | 1  | 3.68E-01   | 3.63E-01       | 3.81E-04 | 3.62E-01       | 4.25E-04 | 3.65E-01       | 1.17E-04 |
|            | 5  | 6.74E-03   | 6.27E-03       | 3.02E-06 | 6.23E-03       | 3.35E-06 | 6.46E-03       | 9.50E-07 |
|            | 10 | 4.54E-05   | 3.93E-05       | 4.72E-11 | 3.88E-05       | 5.41E-10 | 4.17E-05       | 1.65E-10 |
|            | 20 | 2.06E-09   | 1.54E-09       | 3.55E-18 | 1.50E-09       | 3.84E-18 | 1.74E-09       | 1.25E-18 |

نتائج تجارب المحاكاة الجدول (1) تخص القيم التقديرية لكل من المعلمة ( $\tau$ ) ومتوسطات مربعات الخطأ ، بينما تتضمن الجداول (2) الى (6) القيم التقديرية لدالة البقاء ( $S$ ) المقابلة لقيم المعلمة المقدر وقيم متوسطات مربعات الخطأ لتلك القيم التقديرية ، والتي تم الحصول عليها من خلال طريقة الإمكان الأعظم وتقنية البوتستراب بإسلوبها المتوسط والوسيط . وقد تبين ما يلي:

- 1- التقديرات التي قدمتها تقنية البوتستراب وفق الاسلوبين المتوسط والوسيط كانت قريبة من القيم الحقيقية (الافتراضية)، وتقترب تلك التقديرات من القيم الحقيقية كلما ازداد حجم العينة، وكذلك الحال بالنسبة لطريقة الإمكان الأعظم.
- 2- إن القيم التقديرية للمعلمة ( $\tau$ ) ودالة البقاء ( $S$ ) باستخدام تقنية البوتستراب وفق اسلوب الوسيط كانت الأقرب الى القيم الحقيقية في كل التجارب، تلتها القيم التقديرية لتقنية البوتستراب بإسلوب المتوسط عندما ( $n = 5, 10, 30$ ) ، أما في حالة ( $n = 100$ ) فإن طريقة الإمكان الأعظم قد حلت ثانياً.
- 3- فيما يخص قيم لمتوسطات مربعات الخطأ للمعلمة ( $\tau$ ) فإن المقدر ( $\hat{\tau}_{BM}$ ) قد أعطى أقل متوسطات مربعات خطأ يليه المقدر ( $\hat{\tau}_{ML}$ ) ثم المقدر ( $\hat{\tau}_{ML}$ ) عندما ( $n = 5, 10, 30$ ) ، كما وإن قيم متوسطات مربعات الخطأ لدالة البقاء ( $S$ ) كانت بنفس الترتيب، أي  $MSE(\hat{S}_{BM}) < MSE(\hat{S}_{ML}) < MSE(\hat{S}_M)$  ، أما عندما ( $n = 100$ ) فإن الترتيب كان  $MSE(\hat{S}_{BM}) < MSE(\hat{S}_{ML}) < MSE(\hat{S}_M)$  مع ملاحظة إن القيم التقديرية لمتوسطات مربعات الخطأ تتناسب بشكل عكسي مع حجم العينة.

#### 4- الجانب التطبيقي

من خلال النشرات المتوفرة لدى الجهاز المركزي للإحصاء تم الحصول على بيانات معدلات البطالة في العراق للسنوات (2003-2008) وكانت بالترتيب

$$V = (0.281, 0.268, 0.1797, 0.175, 0.117, 0.1534)$$

ولمعرفة فيما اذا كانت البيانات تتبع التوزيع الأسّي أم لا، تم إجراء اختبار حسن المطابقة (Goodness of Fit) لاختبار فرضية العدم التي تنص على إن بيانات معدلات البطالة في العراق للسنوات من (2003-2008) تتبع التوزيع الأسّي وذلك باستخدام اختبار (Kolmogorov Smirnov) بمستوى المعنوية (0.05).

وقد أشارت نتائج الاختبار الى إن قيمة إحصاء الاختبار (0.45004) وقيمة قوة الاختبار ( $p$ -value) (0.1253) ، وهذا يعني قبول فرضية العدم أي إن بيانات معدلات البطالة في العراق للسنوات من (2003-2008) تتبع التوزيع الأسّي.

تم حساب معلمة القياس باستخدام تقنية البوتستراب بإسلوب الوسيط والتي قدمت افضل النتائج في الجانب التجريبي لبيانات المتغير العشوائي ( $V$ ) الذي يتبع التوزيع الأسّي ويمثل معدلات البطالة، وذلك بسحب (500) عينة بوتستراب من بيانات معدلات البطالة السنوية في العراق مع الإرجاع وكل عينة بحجم (25) مشاهدة وبالإستعانة بالبرنامج الإحصائي ( $R_{3.5.1}$ )، وقد كانت قيمة المعلمة المحسوبة ( $\tau = 19.90398$ ) ، وهي أيضا تساوي معدل زمن الوفاة، أي ( $MTTD = 19.90398$ ). كما تم حساب دالة البقاء ( $S$ ) لفترات زمنية مختلفة (Time) بالاعتماد على نفس التقنية والاسلوب، وتفاصيل النتائج مبينة في الجدول (7).

جدول (7) قيم دالة البقاء لبيانات معدلات البطالة في العراق

| Time (years) | S        | Time (years) | S        |
|--------------|----------|--------------|----------|
| 5            | 0.817941 | 30           | 0.232935 |
| 10           | 0.636245 | 35           | 0.181191 |
| 15           | 0.494911 | 40           | 0.140942 |
| 20           | 0.384973 | 45           | 0.109633 |
| 25           | 0.299456 | 50           | 0.08528  |

## 5- الإستنتاجات

- 1- جميع النتائج كانت مطابقة للنظرية الإحصائية، فكلما إزداد حجم العينة إقتربت التقديرات من القيم الحقيقية ونقص متوسط مربعات الخطأ ، كما وإن دالة البقاء تتناقصت بمرور الزمن.
- 2- إن عملية التقدير بإستخدام تقنية البوتستراب سواءا تم تطبيقها بإسلوب المتوسط أو بإسلوب الوسيط كانت تقنية فعالة حتى لو كان حجم العينة صغير جدا، وإن إسلوب الوسيط ضمن تقنية البوتستراب كان أفضل من إسلوب المتوسط وأفضل من طريقة الإمكان الأعظم، رغم إن الأخيرة قد سجلت تقديرات متقاربة مع إسلوب الوسيط في حالة أحجام العينات الكبيرة .
- 3- تبين من خلال تطبيق تقنية البوتستراب بإسلوب الوسيط إن بيانات معدلات البطالة في العراق والمتوفرة بشكل رسمي للسنوات (2003-2008) تتوزع توزيع أسّي بمعلمة قياس  $\left(\frac{1}{\tau} = 19.90398\right)$  ، وهذا يدل على إن معدل البطالة في العراق لتلك السنوات يتراوح عند (20%) .
- 4- بلغ متوسط زمن البقاء (MTTD) لمعدلات البطالة في العراق (19.90398) ، هذا يعني إن هذه البطالة يمكن أن تنتهي بعد (20) سنة تقريبا من عام (2008)، أي في عام (2028) .
- 5- من خلال قيم دالة البقاء التي تم حسابها بإستخدام تقنية البوتستراب بإسلوب الوسيط في الجدول (7)، فإنه يمكن القول وفي حالة إستمرار نفس الظروف الحالية: إن إحتمال أن يبقى معدل البطالة في العراق على حاله في عام (2013) هو (81.8%)، وفي العام الحالي (2018) فإن إحتمال بقاء معدل البطالة في العراق بنفس المعدل سيكون (63.6%) ، أما بعد مرور (15) سنة أي في العام (2023) فإن إحتمال أن يبقى معدل البطالة عند (20%) هو (38.5%) ، وإن إحتمال ان يبقى معدل البطالة في العراق على حاله بعد مرور (40) سنة من الان أي في العام (2048) سيكون (8.5%) إذا ما بقيت الظروف عند هذا المنوال.



**6- المصادر:**

- 1- كامل علاوي كاظم، (2005)، "دراسة تحليلية لواقع الاقتصاد العراقي" مجلة الغري للعلوم الاقتصادية والادارية، العدد (2).
- 2- وزارة التخطيط والتعاون الإنمائي " مسح التشغيل والبطالة لسنة 2003- 2008 "
- 3- Al-Kutubi, H., (2005). "On Comparison estimation, procedures for parameter and survival function exponential distribution using simulation". Ph.D.Thesis, College of Ibn Al-Hatham, Baghdad University, Iraq.
- 4- Casella, G. and Berger, R., (2002). Statistical Inference, volume (2). Duxbury Pacific Grove, CA.
- 5- Efron, B, and Tibshirani, R., (1993). "An introduction to the bootstrap". Chapman and Hall, London.
- 6- Hall, P., (1989). "On efficient bootstrap simulation". Boimetrika. 76, 613-617.
- 7- Lawless J., "Statistical Models and Method for Lifetime Data", John Wiley & Sons, 1982.
- 8- Politis, D. and Romano, J., (1994). "The stationary bootstrap". Jour. Amer. Stat. Assoc. 89, 1303 – 1313.
- 9- Shao, J., (1994). "Bootstrap sample size in non-regular cases". Proc. of the Amer. Math. Soc. 1221251-1262.

## Comparison of Estimating the Survival Function for the Two-Parameter Lindley Distribution by the Bayesian and Maximum-Likelihood Methods

Lecturer. Basheer.j.Khaleel  
Ministry of Education  
Baghdad. Iraq

Prof. Dr. Wadhah S. Ibrahim  
Mustansiriyah University / College of  
Management and Economics  
Baghdad, Iraq

تاريخ استلام البحث: 2022/04/14

تاريخ قبول البحوث: 2022/04/25

نشر البحث في العدد العشرون: سبتمبر / ايلول 2023

64-2522X/519.2

رمز التصنيف ديوي / النسخة الالكترونية (Online):

948-2519X/519.2

رمز التصنيف ديوي / النسخة الورقية (Print):

# Comparison of Estimating the Survival Function for the Two-Parameter Lindley Distribution by the Bayesian and Maximum-Likelihood Methods

Lecturer. Basheer.j.Khaleel  
Ministry of Education  
Baghdad. Iraq

Prof. Dr. Wadhah S. Ibrahim  
Mustansiriyah University / College of  
Management and Economics  
Baghdad, Iraq

## Abstract

In this paper the survival function was estimated by estimating two parameters of the two-parameter Lindley distribution using the maximum likelihood method as well as the Bayesian method, and the difficulty of obtaining the normalizing constant for the posterior distribution function, therefore the difficulty of obtaining a Bayesian estimator for this function and an approximate method was resorted to represented by the Laplace approximation method. Then the behavior of these estimates was studied in terms of having the properties of good estimator, where a simulation study was conducted at different sample sizes (50, 100, 150) and for different cases that differ by changing the values of two parameters of the Lindley distribution. It was found that the estimates of the two methods have the property of consistency because the value of the associated MSE decreases with the large sample size and the simulation showed the superiority of the Bayesian method over the method of maximum likelihood in estimating the survival function at all sample sizes and for all cases.

**Keywords:** Lindley Distribution, Survival Function, Laplace Approximation.

## 1. Introduction

Reliability theory and Survival Analysis have become very important statistical methods since the beginning of the twentieth century. These two theories are based on non-negative probability distributions that are often right-skewed. The most common distributions used in this field are the Exponential Distribution, Gamma Distribution, Weibull Distribution, Rayleigh Distribution, Log-Normal Distribution and many other probability distributions [8]. In addition to the mixed distributions, which we mean that a part of the sample members follows a certain distribution and the rest own another distribution, the most common mixed distributions are the t-Student Distribution, the Laplace Distribution, the Lomax Distribution and the Lindley Distribution [14]. The Lindley distribution is a mixture of the Exponential distribution and the Gamma distribution, with different ratios for each. Although this distribution was presented by Lindley in 1958 [13], interest in it appeared recently for nearly a decade and a half, so its statistical properties were studied and its parameters

were estimated by many researchers such as Ghitany et al. [9], and Jodrá [12]. The idea of the Lindley distribution is summarized as a distribution consisting of mixing two continuous distributions (the first is the Exponential distribution and the second is Gamma) with different weights. where its general form will be:

$$f(x; \theta) = wf_1(x; \theta) + (1 - w)f_2(x; \theta) \quad \dots(1)$$

Where:

$$f_1(x; \theta) = \theta e^{-\theta x} \quad ; \quad f_2(x; \theta) = \theta^2 x e^{-\theta x} \quad \dots(2)$$

So, the resulting function is:

$$f(x; \theta) = w\theta e^{-\theta x} + (1 - w)\theta^2 x e^{-\theta x} \quad \dots(3)$$

Lindley [13] used the weight  $w = \theta/(\theta + 1)$  so the probability density function will be:

$$f(x; \theta) = \frac{\theta^2}{\theta + 1} (1 + x)e^{-\theta x} \quad , x > 0, \theta \geq 0 \quad \dots(4)$$

The Lindley distribution consists of one parameter, and this makes it less flexible compared to distributions that contain a larger number of parameters, so many generalizations have appeared for this distribution, and these generalizations have a different number of parameters, some with two parameters, the other with three, and some reaching five parameters [16]. A two-parameter Lindley distribution was proposed by Shanker et al. [15], in addition to studying its properties and estimating its parameters using the maximum likelihood method, and its parameters were estimated by the Bayes method for the fuzzy data by Al-Bayati [2] using Lindley approximation method. In this paper, the Bayesian estimation will be done using the Laplace approximation method, and then the Bayesian estimates will be compared with the maximum likelihood estimates using the MSE criterion.

The paper was divided into several sections. The first section included the introduction and the second section a study of the two-parameter Lindley distribution in terms of clarifying the mechanism of obtaining it in addition to important functions such as the probability density function, the cumulative distribution function and the survival function, while the third section was devoted to the maximum likelihood estimators, and the fourth section to the Bayesian estimators using Laplace approximation, while the fifth section included the simulation and comparison between the two methods, and the sixth section included the conclusions.

## 2. Two-Parameter Lindley Distribution

The two-parameter Lindley distribution was presented by (Shanker, Sharma & Shanker) in 2013 [15], The researchers used the weight  $w = \theta/(\theta + \alpha)$  with the same distributions used in the one-parameter Lindley distribution, as follows:

$$f(x; \alpha, \theta) = \frac{\theta}{\theta + \alpha} \theta e^{-\theta x} + \frac{\alpha}{\theta + \alpha} \theta^2 x e^{-\theta x} \quad \dots (5)$$

Thus, the probability density function (p.d.f) for a random variable that follows a two-parameter Lindley distribution is:

$$f(x; \alpha, \theta) = \frac{\theta^2}{\theta + \alpha} (1 + \alpha x) e^{-\theta x}, x > 0, \theta \geq 0, \alpha > -\theta \quad \dots (6)$$

This function will be plotted at different values for  $\alpha$  parameter and with the same values for  $\theta$  parameter shown in Figure (2-1) and as in the following figure:

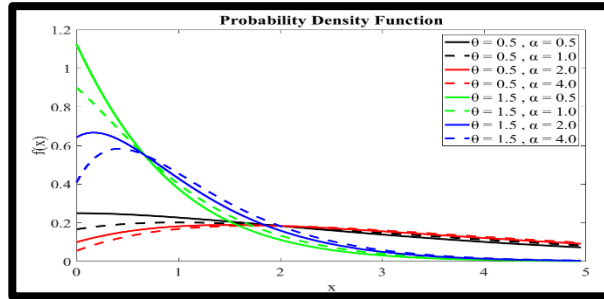


Figure 1. The p.d.f of a two-parameter Lindley distribution at different values of the parameters.

The cumulative distribution function (c.d.f) for a two-parameter Lindley distribution can be obtained as follows:

$$F(x; \theta, \alpha) = \int_0^x f(t) dt = \int_{-\infty}^x \frac{\theta^2}{\theta + \alpha} (1 + \alpha t) e^{-\theta t} dt = \frac{\theta^2}{\theta + \alpha} \int_{-\infty}^x (1 + \alpha t) e^{-\theta t} dt$$

Solving the integral, the c.d.f will be:

$$F(x; \theta, \alpha) = 1 - \left[ \frac{\theta + \alpha + \alpha \theta x}{\theta + \alpha} \right] e^{-\theta x} \quad x > 0, \theta \geq 0, \alpha > -\theta \quad \dots (7)$$

The following figure shows the cumulative distribution function at different values of the parameters.

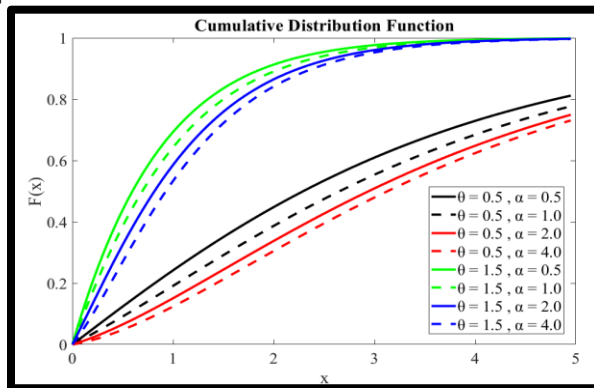


Figure 2. The c.d.f of a two-parameter Lindley distribution at different values of the parameters.

The Survival Function is defined as the probability of an organism surviving until time t, and it can be obtained through the cumulative distribution function as follows:

$$S(x; \theta, \alpha) = 1 - F(x; \theta, \alpha) = 1 - \left\{ 1 - \left[ \frac{\theta + \alpha + \alpha\theta x}{\theta + \alpha} \right] e^{-\theta x} \right\}$$

$$\therefore S(x; \theta, \alpha) = \left[ \frac{\theta + \alpha + \alpha\theta x}{\theta + \alpha} \right] e^{-\theta x}, x > 0, \theta \geq 0, \alpha > -\theta \quad \dots (8)$$

The following figure shows the survival function at different values of the parameters.

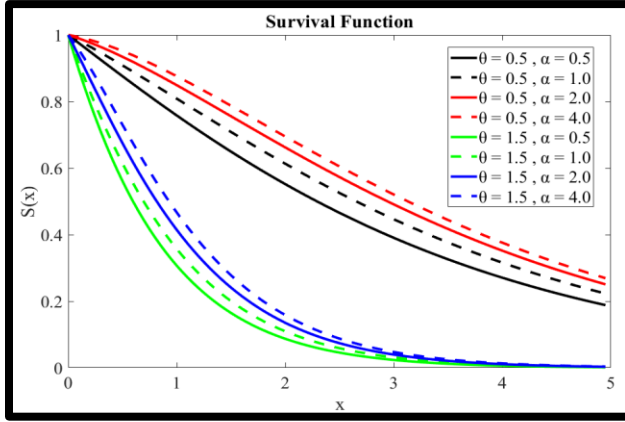


Figure 3. The survival function of a two-parameter Lindley distribution at different values of the parameters.

### 3. Maximum Likelihood Estimators

The idea of the maximum likelihood method is summarized in finding the value that makes the logarithm of the possibility function as large as possible and the likelihood function for a two-parameter Lindley distribution that is as follows [15]:

$$L(\theta, \alpha) = \prod_{i=1}^n \left[ \frac{\theta^2}{\theta + \alpha} (1 + \alpha x_i) e^{-\theta x_i} \right] = \left( \frac{\theta^2}{\theta + \alpha} \right)^n \prod_{i=1}^n (1 + \alpha x_i) e^{-n\theta \bar{x}} \quad \dots (9)$$

Taking the natural logarithm, we obtained:

$$\ell(\theta, \alpha) = \ln[L(\theta, \alpha)] = 2n \ln \theta - n \ln(\theta + \alpha) + \sum_{i=1}^n \ln(1 + \alpha x_i) - n\theta \bar{x} \quad \dots (10)$$

Now, the logarithm of the likelihood function is derived for the two parameters of the distribution, as follows:

$$\frac{\partial \ell(\theta, \alpha)}{\partial \theta} = \frac{2n}{\hat{\theta}} - \frac{n}{\hat{\theta} + \hat{\alpha}} - n\bar{x} = 0 \quad \dots (11)$$

$$\frac{\partial \ell(\theta, \alpha)}{\partial \alpha} = \frac{-n}{\hat{\theta} + \hat{\alpha}} + \sum_{i=1}^n \frac{x_i}{1 + \hat{\alpha} x_i} = 0 \quad \dots (12)$$

Then, the second derivative is found, as follows:

$$\frac{\partial^2 \ell(\theta, \alpha)}{\partial \theta^2} = \frac{-2n}{\hat{\theta}^2} + \frac{n}{(\hat{\theta} + \hat{\alpha})^2} \quad \dots (13)$$

$$\frac{\partial^2 \ell(\theta, \alpha)}{\partial \alpha^2} = \frac{n}{(\hat{\theta} + \hat{\alpha})^2} - \sum_{i=1}^n \frac{x_i^2}{(1 + \hat{\alpha}x_i)^2} \quad \dots (14)$$

$$\frac{\partial^2 \ell(\theta, \alpha)}{\partial \theta \partial \alpha} = \frac{\partial^2 \ell(\theta, \alpha)}{\partial \alpha \partial \theta} = \frac{n}{(\hat{\theta} + \hat{\alpha})^2} \quad \dots (15)$$

Thus, the matrix of the second derivatives is as follows:

$$H_{ML} = \begin{bmatrix} \frac{\partial^2 \ell}{\partial \theta^2} & \frac{\partial^2 \ell}{\partial \theta \partial \alpha} \\ \frac{\partial^2 \ell}{\partial \alpha \partial \theta} & \frac{\partial^2 \ell}{\partial \alpha^2} \end{bmatrix} \quad \dots (16)$$

After obtaining the estimates of  $\alpha$  and  $\theta$  by solving equations (11) and (12) using the Newton-Raphson method, they are substituted into the  $H_{ML}$  matrix. If this matrix is Negative Definite, then the estimates are maximum likelihood estimates.

The estimated survival function can be obtained by the method of maximum likelihood by substituting the estimates obtained from solving equations (11) and (12) into the survival function for a two-parameter Lindley distribution, which was previously defined in equation (8) as follows:

$$\hat{S}_{ML} = \left[ \frac{\hat{\theta}_{ML} + \hat{\alpha}_{ML} + \hat{\alpha}_{ML} \hat{\theta}_{ML} x}{\hat{\theta}_{ML} + \hat{\alpha}_{ML}} \right] e^{-\hat{\theta}_{ML} x} \quad \dots (17)$$

#### 4. Bayesian Estimators

In the classical estimation method, the parameter is considered fixed although it is unknown, but in the Bayesian method it is considered random and has a probability distribution. This process takes place by adopting a probability distribution for the parameter, let it be  $\lambda$ , that contains all the information about this parameter from previous experience, that is, that has nothing to do with the data called the prior distribution and a probability distribution that contains the previous information as well as the data, which is called the posterior distribution, as follows [10]:

$$p(\lambda | \underline{x}) = \frac{p(\underline{x}, \lambda)}{p(\underline{x})} = \frac{p(\underline{x}, \lambda)}{\int p(\underline{x}, \lambda) d\lambda} \quad \dots (18)$$

Where:

$\underline{x}$ : Observations vector.

$p(\lambda | \underline{x})$ : The posterior distribution for  $\lambda$ .

$p(\underline{x})$ : The probability function related to the data.

$p(\underline{x}, \lambda)$ : The joint distribution between  $\lambda$  and observations.

Now, by substituting  $p(\underline{x}, \lambda)$  with its equal to the conditional probability, we get:

$$p(\lambda|\underline{x}) = \frac{p(\underline{x}|\lambda)p(\lambda)}{\int p(\underline{x}|\lambda)p(\lambda)d\lambda} \quad \dots (19)$$

Where:

$p(\underline{x}|\lambda)$ : The possible function, which we usually symbolize  $L(\lambda)$ .

$p(\lambda)$ : The prior distribution for  $\lambda$ .

The previous equation called the Bayes Theorem, and it is clear from it that the posterior probability is the outcome of merging the previous information with the sample information [3].

The Bayes estimator is obtained using loss functions, which is usually symbolized by the symbol  $L(\hat{\lambda}, \lambda)$ , where the expectation is found for these functions, which is called the risk function, and as follows [1]:

$$R(\hat{\lambda}, \lambda) = E(L(\hat{\lambda}, \lambda)) = \int L(\hat{\lambda}, \lambda)p(\lambda|\underline{x})d\lambda \quad \dots (20)$$

The squared error loss function will be used in this research, which can be known as follows [6]:

$$L(\hat{\lambda}, \lambda) = (\lambda - \hat{\lambda})^2 \quad \dots (21)$$

For a two-parameter Lindley distribution, the number of parameters is two  $\alpha$  and  $\theta$  instead of one  $\lambda$ , and the likelihood function for this distribution has been defined in equation (9), and the prior joint distribution of these parameters was obtained after assuming the following:

$\theta \sim \text{Gamma}(a, b), \alpha | \theta \sim \text{TwoPExp}(-\theta, c)$

That is,  $\theta$  follows the Gamma distribution and  $\alpha$  conditioned by the variable  $\theta$  follows a two-parameter exponential distribution, and the reason for this is that the space of  $\theta$  is devoid of any dependence on  $\alpha$  while the space of  $\alpha$  depends on  $\theta$  ( $\alpha > -\theta$ ), that is, the probability density function for each of them can be written as follows:

$$p(\theta) = \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta} \quad \dots (22)$$

$$p(\alpha|\theta) = ce^{-c(\alpha+\theta)} \quad \dots (23)$$

Thus, the prior joint distribution is as follows:

$$p(\theta, \alpha) = p(\alpha|\theta)p(\theta) = \frac{cb^a}{\Gamma(a)} \theta^{a-1} e^{-[(b+c)\theta+c\alpha]} \quad \dots (24)$$

By applying this equation as well as the likelihood function defined in equation (9) in equation (19), we get the joint posterior probability as follows:

$$p(\theta, \alpha|\underline{x}) = \frac{\left(\frac{\theta^2}{\theta + \alpha}\right)^n \prod_{i=1}^n (1 + \alpha x_i) e^{-n\theta\bar{x}} \theta^{a-1} e^{-((b+c)\theta+c\alpha)}}{\iint \left(\frac{\theta^2}{\theta + \alpha}\right)^n \prod_{i=1}^n (1 + \alpha x_i) e^{-n\theta\bar{x}} \theta^{a-1} e^{-((b+c)\theta+c\alpha)} d\theta d\alpha} \quad \dots (25)$$

The estimator of any function can be obtained in terms of the two parameters, let it be,  $u$ , using the squared loss function as follows:

$$\hat{u} = E[u(\theta, \alpha)] = \frac{\iint u(\theta, \alpha) \left(\frac{\theta^2}{\theta + \alpha}\right)^n \prod_{i=1}^n (1 + \alpha x_i) e^{-n\theta\bar{x}} \theta^{a-1} e^{-((b+c)\theta+c\alpha)} d\theta d\alpha}{\iint \left(\frac{\theta^2}{\theta + \alpha}\right)^n \prod_{i=1}^n (1 + \alpha x_i) e^{-n\theta\bar{x}} \theta^{a-1} e^{-((b+c)\theta+c\alpha)} d\theta d\alpha} \quad \dots (26)$$



Because of the inability to obtain an algebraic formula (Closed Form), approximate methods are resorted to. There are many of these methods useful for this purpose, the most important of which are Lindley's approximation and Laplace's approximation.

The researchers (Tierney & Kadane) introduced an approximation in 1986 to obtain the product of the integration in equation (26) based on the work of Laplace in 1774 in a research paper entitled (Accurate approximations for posterior moments and marginal densities). Its purpose is to find the expectation in cases that are difficult to find [17]. This approximation is called the Laplace Approximation in relation to the method based on the Laplace method for finding the integral [5]. Or the Tierney-Kadane Approximation related to the aforementioned researchers who introduced this method [4].

According to the Laplace approximation method, the expectation for the posterior distribution is found when this distribution includes an integral in the denominator that cannot be reduced to a simplified form. If the posterior expectation is as mentioned in equation (26), then we can formulate it as follows [11]:

$$I(\underline{x}) = E[u(\underline{\lambda}|\underline{x})] = \frac{\int e^{(\ln[u(\underline{\lambda})]+\ell+\rho)} d\underline{\lambda}}{\int e^{(\ell+\rho)} d\underline{\lambda}} \quad \dots (27)$$

Where:

$\rho$ : The natural logarithm of the prior joint distribution of the parameters.

$\ell$ : The natural logarithm of the likelihood function.

Assuming that:

$$h(\underline{\lambda}) = \frac{1}{n}(\ell + \rho) \quad \dots (28)$$

$$h^*(\underline{\lambda}) = \frac{1}{n} \ln(u(\underline{\lambda})) + h(\underline{\lambda}) \quad \dots (29)$$

Then equation (27) becomes:

$$I(\underline{x}) = E[u(\underline{\lambda}|\underline{x})] = \frac{\int e^{nh^*(\underline{\lambda})} d\underline{\lambda}}{\int e^{nh(\underline{\lambda})} d\underline{\lambda}} \quad \dots (30)$$

Thus, Laplace's estimation for this equation is as follows [17]:

$$I(\underline{x}) = \hat{E}[u(\underline{\lambda}|\underline{x})] = \left[ \frac{|\Sigma^*|}{|\Sigma|} \right]^{1/2} \exp \left\{ n \left( h^*(\hat{\underline{\lambda}}^*) - h(\hat{\underline{\lambda}}) \right) \right\} \quad \dots (31)$$

Where  $|\cdot|$  denotes the determinant of the matrix,  $\hat{\underline{\lambda}}^*$  and  $\hat{\underline{\lambda}}$  are estimated vectors that maximize functions  $h^*(\underline{\lambda})$  and  $h(\underline{\lambda})$  respectively, and  $\Sigma^*$  and  $\Sigma$  are the inverse of the negative Hessian matrix for  $h^*(\underline{\lambda})$  and  $h(\underline{\lambda})$  at  $\hat{\underline{\lambda}}^*$  and  $\hat{\underline{\lambda}}$  respectively.

In the case of two parameters, as is the case in this research, the equations (28), (29) and (30) are written as follows [7]:

$$h(\theta, \alpha) = \frac{1}{n}(\ell + \rho) \quad \dots (32)$$

$$h^*(\theta, \alpha) = \frac{1}{n} \ln(u(\theta, \alpha)) + h(\theta, \alpha) \quad \dots (33)$$

$$l(\underline{x}) = \hat{E}[u(\theta, \alpha | \underline{x})] = \left( \frac{|\Sigma^*|}{|\Sigma|} \right)^{1/2} \exp \left\{ n \left( h^*(\hat{\theta}^*, \hat{\alpha}^*) - h(\hat{\theta}, \hat{\alpha}) \right) \right\} \dots (34)$$

Where  $(\hat{\theta}^*, \hat{\alpha}^*)$  and  $(\hat{\theta}, \hat{\alpha})$  maximize functions  $h^*(\theta, \alpha)$  and  $h(\theta, \alpha)$  respectively, and  $\Sigma^*$  and  $\Sigma$  are the inverse of the negative Hessian matrix for  $h^*(\theta, \alpha)$  and  $h(\theta, \alpha)$  at  $(\hat{\theta}^*, \hat{\alpha}^*)$  and  $(\hat{\theta}, \hat{\alpha})$  respectively, so that both  $\Sigma^*$  and  $\Sigma$  can be defined as follows:

$$\Sigma = [-H_h]^{-1} \Big|_{\substack{\theta=\hat{\theta} \\ \alpha=\hat{\alpha}}} = \begin{bmatrix} -\frac{\partial^2 h}{\partial \theta^2} & -\frac{\partial^2 h}{\partial \theta \partial \alpha} \\ -\frac{\partial^2 h}{\partial \theta \partial \alpha} & -\frac{\partial^2 h}{\partial \alpha^2} \end{bmatrix}^{-1} \Big|_{\substack{\theta=\hat{\theta} \\ \alpha=\hat{\alpha}}} \dots (35)$$

$$\Sigma^* = [-H_{h^*}]^{-1} \Big|_{\substack{\theta=\hat{\theta}^* \\ \alpha=\hat{\alpha}^*}} = \begin{bmatrix} -\frac{\partial^2 h^*}{\partial \theta^2} & -\frac{\partial^2 h^*}{\partial \theta \partial \alpha} \\ -\frac{\partial^2 h^*}{\partial \theta \partial \alpha} & -\frac{\partial^2 h^*}{\partial \alpha^2} \end{bmatrix}^{-1} \Big|_{\substack{\theta=\hat{\theta}^* \\ \alpha=\hat{\alpha}^*}} \dots (36)$$

As for the case of the two-parameter Lindley distribution, the researchers derived in this research the values that maximize the two previously mentioned functions in equations (32) and (33) in addition to the Hessian matrices for both functions, taking into account that equation (32) does not change with the change of the u function and thus it will be the same for all cases, and the maximum values for it will be found once, unlike equation (33) which will be maximized by three different cases, and the following is the finding of the maximum points for the function  $h(\theta, \alpha)$ :

$$h(\theta, \alpha) = \frac{1}{n} \left\{ 2n \ln \theta - n \ln(\theta + \alpha) + \sum_{i=1}^n \ln(1 + \alpha x_i) - n \bar{x} + (a - 1) \ln \theta - [(b + c)\theta + c\alpha] \right\} \dots (37)$$

$$\frac{\partial h(\theta, \alpha)}{\partial \theta} = \frac{1}{n} \left\{ \frac{2n + a - 1}{\theta} - \frac{n}{\theta + \hat{\alpha}} - n \bar{x} - (b + c) \right\} = 0 \dots (38)$$

$$\frac{\partial h(\theta, \alpha)}{\partial \alpha} = \frac{1}{n} \left\{ -\frac{n}{\theta + \hat{\alpha}} + \sum_{i=1}^n \frac{x_i}{1 + \hat{\alpha} x_i} - c \right\} = 0 \dots (39)$$

The values  $\hat{\theta}$  and  $\hat{\alpha}$  are obtained after solving equations (38) and (39). As for the Hessian matrix, we obtain it by differentiating equations (38) and (39) again, as follows:

$$\frac{\partial^2 h(\theta, \alpha)}{\partial \theta^2} = \frac{1}{n} \left\{ -\frac{2n + a - 1}{\theta^2} + \frac{n}{(\hat{\theta} + \hat{\alpha})^2} \right\} \dots (40)$$

$$\frac{\partial^2 h(\theta, \alpha)}{\partial \alpha^2} = \frac{1}{n} \left\{ \frac{n}{(\hat{\theta} + \hat{\alpha})^2} - \sum_{i=1}^n \frac{x_i^2}{(1 + \hat{\alpha} x_i)^2} \right\} \dots (41)$$

$$\frac{\partial^2 h(\theta, \alpha)}{\partial \theta \partial \alpha} = \frac{\partial^2 h(\theta, \alpha)}{\partial \alpha \partial \theta} = \frac{1}{(\hat{\theta} + \hat{\alpha})^2} \dots (42)$$

Thus, the matrix of the second derivatives is as follows:

$$H_h = \begin{bmatrix} \frac{\partial^2 h}{\partial \theta^2} & \frac{\partial^2 h}{\partial \theta \partial \alpha} \\ \frac{\partial^2 h}{\partial \alpha \partial \theta} & \frac{\partial^2 h}{\partial \alpha^2} \end{bmatrix} \quad \dots (43)$$

As for equation (33), its maximization varies according to the u function, but here we will find general equations and then substitute for special cases, as follows:

$$h^*(\theta, \alpha) = \frac{1}{n} \ln(u(\theta, \alpha)) + h(\theta, \alpha) \quad \dots (44)$$

$$\frac{\partial h^*(\theta, \alpha)}{\partial \theta} = \frac{1}{n} \left\{ \frac{u_\theta}{u(\theta, \alpha)} \Big|_{\substack{\theta=\hat{\theta}^* \\ \alpha=\hat{\alpha}^*}} + \frac{2n+a-1}{\hat{\theta}} - \frac{n}{\hat{\theta} + \hat{\alpha}} - n\bar{x} - (b+c) \right\} = 0 \quad \dots (45)$$

$$\frac{\partial h^*(\theta, \alpha)}{\partial \alpha} = \frac{1}{n} \left\{ \frac{u_\alpha}{u(\theta, \alpha)} \Big|_{\substack{\theta=\hat{\theta}^* \\ \alpha=\hat{\alpha}^*}} - \frac{n}{\hat{\theta} + \hat{\alpha}} + \sum_{i=1}^n \frac{x_i}{1 + \hat{\alpha}x_i} - c \right\} = 0 \quad \dots (46)$$

The values of  $\hat{\theta}^*$  and  $\hat{\alpha}^*$  are obtained after solving equations (45) and (46). As for the Hessian matrix, we obtain it by differentiating equations (45) and (46) again, as follows:

$$\frac{\partial^2 h^*(\theta, \alpha)}{\partial \theta^2} = \frac{1}{n} \left\{ \frac{u_{\theta\theta}}{u(\theta, \alpha)} \Big|_{\substack{\theta=\hat{\theta}^* \\ \alpha=\hat{\alpha}^*}} - \frac{[u_\theta]^2}{[u(\hat{\theta}^*, \hat{\alpha}^*)]^2} \Big|_{\substack{\theta=\hat{\theta}^* \\ \alpha=\hat{\alpha}^*}} - \frac{2n+a-1}{\hat{\theta}^{*2}} + \frac{n}{(\hat{\theta}^* + \hat{\alpha}^*)^2} \right\} \quad \dots (47)$$

$$\frac{\partial^2 h^*(\theta, \alpha)}{\partial \alpha^2} = \frac{1}{n} \left\{ \frac{u_{\alpha\alpha}}{u(\theta, \alpha)} \Big|_{\substack{\theta=\hat{\theta}^* \\ \alpha=\hat{\alpha}^*}} - \frac{[u_\alpha]^2}{[u(\hat{\theta}^*, \hat{\alpha}^*)]^2} \Big|_{\substack{\theta=\hat{\theta}^* \\ \alpha=\hat{\alpha}^*}} + \frac{n}{(\hat{\theta}^* + \hat{\alpha}^*)^2} - \sum_{i=1}^n \frac{x_i^2}{(1 + \hat{\alpha}^*x_i)^2} \right\} \quad \dots (48)$$

$$\frac{\partial^2 h^*(\theta, \alpha)}{\partial \theta \partial \alpha} = \frac{\partial^2 h^*(\theta, \alpha)}{\partial \alpha \partial \theta} = \frac{1}{n} \left\{ \frac{u_{\theta\alpha}}{u(\theta, \alpha)} \Big|_{\substack{\theta=\hat{\theta}^* \\ \alpha=\hat{\alpha}^*}} - \frac{u_\theta u_\alpha}{[u(\theta, \alpha)]^2} \Big|_{\substack{\theta=\hat{\theta}^* \\ \alpha=\hat{\alpha}^*}} + \frac{n}{(\hat{\theta}^* + \hat{\alpha}^*)^2} \right\} \quad \dots (49)$$

Thus, the matrix of the second derivatives is as follows:

$$H_{h^*} = \begin{bmatrix} \frac{\partial^2 h^*}{\partial \theta^2} & \frac{\partial^2 h^*}{\partial \theta \partial \alpha} \\ \frac{\partial^2 h^*}{\partial \alpha \partial \theta} & \frac{\partial^2 h^*}{\partial \alpha^2} \end{bmatrix} \quad \dots (50)$$

As special cases of the function u, we take the following three cases and substitute  $h(\theta, \alpha)$  and  $h^*(\theta, \alpha)$  into equation (34) after taking into account u and its accompanying derivatives for each case:

**Bayesian Estimator of  $\theta$ :** in this case  $u(\theta, \alpha) = \theta$ , then:

$$u_\theta = 1, u_\alpha = u_{\theta\alpha} = u_{\alpha\theta} = u_{\theta\theta} = u_{\alpha\alpha} = 0$$

**Bayesian Estimator of  $\alpha$ :** in this case  $u(\theta, \alpha) = \alpha$ , then:

$$u_\alpha = 1, u_\theta = u_{\theta\alpha} = u_{\alpha\theta} = u_{\theta\theta} = u_{\alpha\alpha} = 0$$

**Bayesian Estimator of the survival function:** in this case

$$u(\theta, \alpha) = \left[ \frac{\theta + \alpha + \alpha\theta t}{\theta + \alpha} \right] e^{-\theta t}, \text{ then:}$$

$$u_\theta = -\theta t \left[ \frac{\theta + 2\alpha + \alpha\theta t + \alpha^2 t}{(\theta + \alpha)^2} \right] e^{-\theta t} \quad \dots (51)$$

$$u_\alpha = \frac{\theta^2 t}{(\theta + \alpha)^2} e^{-\theta t} \quad \dots (52)$$

$$u_{\theta\alpha} = u_{\alpha\theta} = -\theta x \left[ \frac{\theta^2 t + \alpha\theta t - 2\alpha}{(\theta + \alpha)^3} \right] e^{-\theta x} \quad \dots (53)$$

$$u_{\theta\theta} = \left[ \frac{(2\alpha^2\theta^2 + \alpha\theta^3 + \alpha^3\theta)t^3 + (\theta^2 + 3\alpha\theta^2 + \alpha^2\theta - \alpha^3)t^2 - 2\alpha^2 t}{(\theta + \alpha)^3} \right] e^{-\theta t} \quad \dots (54)$$

$$u_{\alpha\alpha} = -\frac{2\theta^2 x}{(\theta + \alpha)^3} e^{-\theta t} \quad \dots (55)$$

Where the Bayesian estimator of the survival function is obtained after substituting equations (51) to (55) in equation (31).

## 5. Simulation

A comprehensive simulation was carried out to generate data distributives as two-parameter Lindley distribution, and based on these data, the parameters and survival function were estimated by the maximum likelihood and Bayesian methods by Laplace' approximation, and then the two methods were compared using the MSE criterion. The data was generated by the quantile function of the two-parameter Lindley distribution, which is obtained as follows:

$$\begin{aligned} F(x) &= 1 - \left[ \frac{\theta + \alpha + \alpha\theta x}{\theta + \alpha} \right] e^{-\theta x} = u \\ \Rightarrow 1 - u &= \left[ \frac{\theta + \alpha + \alpha\theta x}{\theta + \alpha} \right] e^{-\theta x} \\ \Rightarrow (\theta + \alpha)(1 - u) &= [\theta + \alpha + \alpha\theta x] e^{-\theta x} \\ \Rightarrow \frac{(\theta + \alpha)(1 - u)}{\alpha} &= \left[ \frac{\theta + \alpha}{\alpha} + \theta x \right] e^{-\theta x} \\ \Rightarrow \frac{(\theta + \alpha)(1 - u)e^{-\frac{(\theta + \alpha)}{\alpha}}}{\alpha} &= \left[ \frac{\theta + \alpha}{\alpha} + \theta x \right] e^{-\left(\theta x + \frac{(\theta + \alpha)}{\alpha}\right)} \\ \Rightarrow -\frac{(\theta + \alpha)(1 - u)e^{-\frac{(\theta + \alpha)}{\alpha}}}{\alpha} &= -\left[ \frac{\theta + \alpha}{\alpha} + \theta x \right] e^{-\left(\theta x + \frac{(\theta + \alpha)}{\alpha}\right)} \\ \Rightarrow W_{-1} \left[ -\frac{(\theta + \alpha)(1 - u)e^{-\frac{(\theta + \alpha)}{\alpha}}}{\alpha} \right] &= -\left[ \frac{\theta + \alpha}{\alpha} + \theta x \right] \\ \Rightarrow -W_{-1} \left[ -\frac{(\theta + \alpha)(1 - u)e^{-\frac{(\theta + \alpha)}{\alpha}}}{\alpha} \right] &= \left[ \frac{\theta + \alpha}{\alpha} + \theta x \right] \\ \Rightarrow -W_{-1} \left[ -\frac{(\theta + \alpha)(1 - u)e^{-\frac{(\theta + \alpha)}{\alpha}}}{\alpha} \right] - \left[ \frac{\theta + \alpha}{\alpha} \right] &= \theta x \\ \Rightarrow x &= -\frac{1}{\theta} W_{-1} \left[ -\frac{(\theta + \alpha)(1 - u)e^{-\frac{(\theta + \alpha)}{\alpha}}}{\alpha} \right] - \left[ \frac{\theta + \alpha}{\theta\alpha} \right] \quad \dots (55) \end{aligned}$$

Where  $W_{-1}(\cdot)$  is the Lambert W Function with a negative branch.

The generation process was performed for four different cases as shown in Table 2, and for each case different sample sizes (50, 100, 150) were used. The experiment was repeated 1000 for each of the four cases, after which the estimation process was carried out using the maximum likelihood and Bayesian methods, and their comparison was done using the MSE criteria. All of these operations took place in the R environment.

Table 1. Different cases of simulation.

| Cases  | $\theta$ | $\alpha$ |
|--------|----------|----------|
| Case 1 | 0.5      | 0.5      |
| Case 2 | 0.5      | 2        |
| Case 3 | 1.5      | 0.5      |
| Case 4 | 1.5      | 2        |

The following tables show the simulation results for the parameters:

Table 2. The simulated MSEs values for  $\theta = 0.5; \alpha = 0.5$ .

| Sample Size | Parameters     | MLE     | Bayes   | Best  |
|-------------|----------------|---------|---------|-------|
| n=50        | $\theta = 0.5$ | 0.50786 | 0.51654 | Bayes |
|             | $\alpha = 0.5$ | 0.52712 | 0.51706 |       |
|             | MSE            | 0.00671 | 0.00653 |       |
| n=100       | $\theta = 0.5$ | 0.50475 | 0.50916 | Bayes |
|             | $\alpha = 0.5$ | 0.52593 | 0.52129 |       |
|             | MSE            | 0.00392 | 0.00372 |       |
| n=150       | $\theta = 0.5$ | 0.50476 | 0.50779 | Bayes |
|             | $\alpha = 0.5$ | 0.52534 | 0.52265 |       |
|             | MSE            | 0.00290 | 0.00276 |       |

Table 3. The simulated MSEs values for  $\theta = 0.5; \alpha = 2$ .

| Sample Size | Parameters     | MLE     | Bayes   | Best  |
|-------------|----------------|---------|---------|-------|
| n=50        | $\theta = 0.5$ | 0.5099  | 0.57521 | Bayes |
|             | $\alpha = 2$   | 2.19657 | 2.08677 |       |
|             | MSE            | 0.06138 | 0.03087 |       |
| n=100       | $\theta = 0.5$ | 0.50538 | 0.53706 | Bayes |
|             | $\alpha = 2$   | 2.16786 | 2.0951  |       |
|             | MSE            | 0.04702 | 0.02408 |       |
| n=150       | $\theta = 0.5$ | 0.5075  | 0.50918 | Bayes |
|             | $\alpha = 2$   | 2.14852 | 2.11255 |       |
|             | MSE            | 0.03768 | 0.02099 |       |

Table 4. The simulated MSEs values for  $\theta = 1.5; \alpha = 0.5$ .

| Sample Size | Parameters     | MLE     | Bayes   | Best  |
|-------------|----------------|---------|---------|-------|
| n=50        | $\theta = 1.5$ | 1.5374  | 1.53958 | Bayes |
|             | $\alpha = 0.5$ | 0.57628 | 0.54821 |       |
|             | MSE            | 0.04759 | 0.04757 |       |
| n=100       | $\theta = 1.5$ | 1.53795 | 1.55446 | Bayes |
|             | $\alpha = 0.5$ | 0.57571 | 0.5622  |       |
|             | MSE            | 0.02855 | 0.02811 |       |
| n=150       | $\theta = 1.5$ | 1.53313 | 1.54472 | Bayes |
|             | $\alpha = 0.5$ | 0.58082 | 0.57253 |       |
|             | MSE            | 0.02258 | 0.0221  |       |

Table 5. The simulated MSEs values for  $\theta = 1.5; \alpha = 2$ .

| Sample Size | Parameters     | MLE     | Bayes   | Best  |
|-------------|----------------|---------|---------|-------|
| n=50        | $\theta = 1.5$ | 1.52705 | 1.53752 | Bayes |
|             | $\alpha = 2$   | 2.11779 | 2.00048 |       |
|             | MSE            | 0.07729 | 0.07394 |       |
| n=100       | $\theta = 1.5$ | 1.51477 | 1.57137 | Bayes |
|             | $\alpha = 2$   | 2.10936 | 2.05143 |       |
|             | MSE            | 0.04811 | 0.04132 |       |
| n=150       | $\theta = 1.5$ | 1.51474 | 1.55194 | Bayes |
|             | $\alpha = 2$   | 2.10363 | 2.06154 |       |
|             | MSE            | 0.03746 | 0.03233 |       |

The following tables show the simulation results for the survival function:

Table 6. The simulated MSEs values of survival function for  $\theta = 0.5; \alpha = 0.5$ .

| Sample Size | t    | Real    | MLE     | Bayes   | Best  |
|-------------|------|---------|---------|---------|-------|
| n=50        | 0.05 | 0.9875  | 0.98742 | 0.98698 | Bayes |
|             | 0.5  | 0.87615 | 0.87539 | 0.87152 |       |
|             | 1.5  | 0.6495  | 0.64821 | 0.6397  |       |
|             | 3.5  | 0.32583 | 0.32619 | 0.31637 |       |
|             | 5    | 0.18469 | 0.18645 | 0.1785  |       |
|             | MSE  |         |         | 0.00899 |       |
| n=100       | 0.05 | 0.9875  | 0.98757 | 0.98735 | Bayes |
|             | 0.5  | 0.87615 | 0.87659 | 0.87467 |       |
|             | 1.5  | 0.6495  | 0.65017 | 0.64592 |       |
|             | 3.5  | 0.32583 | 0.32679 | 0.32181 |       |
|             | 5    | 0.18469 | 0.18589 | 0.18185 |       |
|             | MSE  |         |         | 0.00442 |       |
| n=150       | 0.05 | 0.9875  | 0.98758 | 0.98744 | Bayes |
|             | 0.5  | 0.87615 | 0.87666 | 0.87539 |       |
|             | 1.5  | 0.6495  | 0.65004 | 0.6472  |       |
|             | 3.5  | 0.32583 | 0.32601 | 0.32267 |       |
|             | 5    | 0.18469 | 0.18491 | 0.18219 |       |
|             | MSE  |         |         | 0.00286 |       |

Table 7. The simulated MSEs values of survival function for  $\theta = 0.5; \alpha = 2$ .

| Sample Size | t    | Real    | MLE     | Bayes   | Best  |
|-------------|------|---------|---------|---------|-------|
| n=50        | 0.05 | 0.99482 | 0.99491 | 0.9948  | Bayes |
|             | 0.5  | 0.93456 | 0.93456 | 0.9186  |       |
|             | 1.5  | 0.75579 | 0.75397 | 0.70765 |       |
|             | 3.5  | 0.41706 | 0.41608 | 0.34978 |       |
|             | 5    | 0.24625 | 0.24795 | 0.19087 |       |
|             | MSE  |         |         | 0.0229  |       |
| n=100       | 0.05 | 0.99482 | 0.99497 | 0.99427 | Bayes |
|             | 0.5  | 0.93456 | 0.93544 | 0.92757 |       |
|             | 1.5  | 0.75579 | 0.75635 | 0.73328 |       |
|             | 3.5  | 0.41706 | 0.41754 | 0.38356 |       |
|             | 5    | 0.24625 | 0.24758 | 0.21779 |       |
|             | MSE  |         |         | 0.00632 |       |
| n=150       | 0.05 | 0.99482 | 0.99492 | 0.99486 | Bayes |
|             | 0.5  | 0.93456 | 0.93483 | 0.92963 |       |
|             | 1.5  | 0.75579 | 0.75449 | 0.73901 |       |
|             | 3.5  | 0.41706 | 0.41422 | 0.39105 |       |
|             | 5    | 0.24625 | 0.24421 | 0.22377 |       |
|             | MSE  |         |         | 0.00463 |       |

Table 8. The simulated MSEs values of survival function for  $\theta = 1.5; \alpha = 0.5$ .

| Sample Size | t    | Real    | MLE     | Bayes   | Best  |
|-------------|------|---------|---------|---------|-------|
| n=50        | 0.05 | 0.94514 | 0.94508 | 0.94297 | Bayes |
|             | 0.5  | 0.56094 | 0.56117 | 0.54933 |       |
|             | 1.5  | 0.16469 | 0.16649 | 0.15681 |       |
|             | 3.5  | 0.01213 | 0.01313 | 0.01153 |       |
|             | 5    | 0.00159 | 0.00189 | 0.00158 |       |
|             | MSE  |         |         | 0.00477 |       |
| n=100       | 0.05 | 0.94514 | 0.94519 | 0.94413 | Bayes |
|             | 0.5  | 0.56094 | 0.56067 | 0.55468 |       |
|             | 1.5  | 0.16469 | 0.16418 | 0.15927 |       |
|             | 3.5  | 0.01213 | 0.0122  | 0.01142 |       |
|             | 5    | 0.00159 | 0.00164 | 0.0015  |       |
|             | MSE  |         |         | 0.00243 |       |
| n=150       | 0.05 | 0.94514 | 0.94558 | 0.94488 | Bayes |
|             | 0.5  | 0.56094 | 0.56249 | 0.55846 |       |
|             | 1.5  | 0.16469 | 0.165   | 0.16167 |       |
|             | 3.5  | 0.01213 | 0.01215 | 0.01163 |       |
|             | 5    | 0.00159 | 0.00161 | 0.00151 |       |
|             | MSE  |         |         | 0.00164 |       |

Table 9. The simulated MSEs values of survival function for  $\theta = 1.5; \alpha = 2$ .

| Sample Size | t    | Real    | MLE     | Bayes   | Best  |
|-------------|------|---------|---------|---------|-------|
| n=50        | 0.05 | 0.9675  | 0.96744 | 0.9673  | Bayes |
|             | 0.5  | 0.67481 | 0.6742  | 0.66101 |       |
|             | 1.5  | 0.24091 | 0.2429  | 0.24077 |       |
|             | 3.5  | 0.02099 | 0.0231  | 0.01587 |       |
|             | 5    | 0.00292 | 0.00366 | 0.00212 |       |
|             | MSE  |         |         | 0.00592 |       |
| n=100       | 0.05 | 0.9675  | 0.96774 | 0.96539 | Bayes |
|             | 0.5  | 0.67481 | 0.67551 | 0.65562 |       |
|             | 1.5  | 0.24091 | 0.24203 | 0.2236  |       |
|             | 3.5  | 0.02099 | 0.02191 | 0.01812 |       |
|             | 5    | 0.00292 | 0.00324 | 0.00246 |       |
|             | MSE  |         |         | 0.00230 |       |
| n=150       | 0.05 | 0.9675  | 0.96774 | 0.96717 | Bayes |
|             | 0.5  | 0.67481 | 0.67527 | 0.67397 |       |
|             | 1.5  | 0.24091 | 0.24105 | 0.22871 |       |
|             | 3.5  | 0.02099 | 0.02141 | 0.01888 |       |
|             | 5    | 0.00292 | 0.0031  | 0.00258 |       |
|             | MSE  |         |         | 0.00206 |       |

We note from all the previous tables that the Bayesian method is better than the maximum likelihood method in the four cases and for the three sample sizes. The sample. As for the consistency property, the estimates of the two methods have this property because the MSE values of these estimates decrease with the increase in the sample size.

### 6. Conclusions

We conclude from the above the following:

1. All estimators of the two methods have a property of consistency.
2. The priority of the Bayesian method over the maximum likelihood method.
3. The performance of the two methods converges with increasing the sample size.

**References**

1. Al-Badran, F. M. J. (2014). Estimation of Reliability of Stress & Strength Model for Lindley Distribution (A Comparison Study with Application). A Master Thesis. Mustansiriyah University. Iraq.
2. Al-Bayati, R. S. S. (2018). Some Estimation Methods for Lindley Distribution. A Master Thesis. Mustansiriyah University. Iraq.
3. Carlin, B. P., & Louis, T. A. (2009). Bayesian methods for data analysis. CRC Press.
4. Danish, M. Y., & Aslam, M. (2013). Bayesian analysis of randomly censored generalized exponential distribution. *Austrian Journal of Statistics*, 42(1), 47-62.
5. dos SANTOS, M. A., MOALA, F. A., & TACHIBANA, V. M. (2009). Approximate Bayesian methods for logistic regression model. *Rev. Bras. Biom.*, 27(2), 288-300.
6. Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2014). Bayesian data analysis. 3rd Edition, Chapman and Hall/CRC.
7. Gencer, G., & Gencer, K. (2019). Comparison of Maximum Likelihood and Bayes Estimators Under Symmetric and Asymmetric Loss Functions by means of Tierney Kadane's Approximation for Weibull Distribution. *Turkish Journal of Science and Technology*, 14(2), 69-78.
8. Georges, P. E. (2006). Building a Simulation Model to Find the Reliability of Electrical Power System. A Master Thesis. Baghdad University, Iraq.
9. Ghitany, M. E., Atieh, B., & Nadarajah, S. (2008). Lindley distribution and its application. *Mathematics and computers in simulation*, 78(4), 493-506.
10. Heard, N. (2021). *An Introduction to Bayesian Inference, Methods and Computation*. Springer.
11. Jan, U., & Ahmad, S. P. (2019). Estimating the Parameter of Weighted Ailamujia Distribution Using Bayesian Approximation Techniques. *Bayesian Analysis and Reliability Estimation of Generalized Probability Distributions*, edited by Afaq Ahmad, AIJR Publisher, 2019, pp. 50–58., ISBN: 978-81-936820-7-4, DOI: 10.21467/books.44.5.
12. Jodrá, P. (2010). Computer generation of random variables with Lindley or Poisson–Lindley distribution via the Lambert W function. *Mathematics and Computers in Simulation*, 81(4), 851-859.
13. Lindley, D.V. (1958). Fiducial Distribution and Bayes Theorem. *Journal of the Royal Statistical Society, Series B*, 20(1),102-107.
14. Salh, A. A. (2016). Comparison of some methods for estimation of Hazard function of distribution Quasi Lindley with application, A Master Thesis, Baghdad University, Iraq.
15. Shanker, R., Sharma, S., & Shanker, R. (2013). A two-parameter Lindley distribution for modeling waiting and survival times data. *Applied Mathematics* 4: 363-368.
16. Tharshan, R., & Wijekoon, P. (2020). A comparison study on a new five-parameter generalized Lindley distribution with its sub-models. *STATISTICS*, 89.
17. Tierney, L., & Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, 81(393), 82-86.





## Using gray wave models to predict crude oil production in Iraq

Ali Mohammed Ali

Assistant lecturer

Economics department

college of Administration & Economics

Misan University

Dr. Ahmed Shaker Mohammed Tahir

Assistant Professor

Statistics department

college of Administration & Economics

Mustansiriyah University

تاريخ استلام البحث: 2022/04/14

تاريخ قبول البحث: 2022/05/15

نشر البحث في العدد العشرون: ايلول / سبتمبر 2023

64-2522X/519.5

رمز التصنيف ديوي / النسخة الالكترونية (Online):

2519-948X/519.5

رمز التصنيف ديوي / النسخة الورقية (Print):

## Using gray wave models to predict crude oil production in Iraq

Ali Mohammed Ali

Assistant lecturer

Economics department

college of Administration & Economics

Misan University

Dr. Ahmed Shaker Mohammed Tahir

Assistant Professor

Statistics department

college of Administration & Economics

Mustansiriyah University

### ABSTRACT:

The gray system theory which proposed by the Chinese researcher Julong Deng is the theory that deals with the uncertain information of phenomena, according to which few data can be used to create models without taking into account the distribution of those data. Therefore, this theory has been widely used in the field of image processing, time-series prediction, system optimization, control and decision, etc. Gray system prediction is one of the most important components of gray system theory that works to find rules for system development through certain data processing procedures in order to predict the future of the system scientifically and quantitatively.

When the time series data have large fluctuations and a greater oscillation range, it is difficult to build an appropriate model to describe the behavior of that time series, in this case we can think of predicting the wave of behavioral data development in the future based on the original wave of data, this kind of prediction is called wave prediction, and its modeling process is obvious and simple, so wave prediction is gaining high application value in many fields.

In this paper, gray wave prediction models will be used to predict crude oil production in Iraq based on the time series data of the monthly production rate, whose observations suffer from large fluctuations.

**Keywords:** Time series, Grey Model, Grey Wave model, Contour lines.

### استعمال النماذج الموجية الرمادية للتنبؤ بإنتاج النفط الخام في العراق المستخلص

نظرية النظام الرمادي المقترحة من قبل الباحث الصيني Julong Deng هي النظرية التي تتعامل مع المعلومات غير المؤكدة للظواهر إذ يمكن بموجبها استعمال بعض البيانات لإنشاء نماذج دون مراعاة توزيع تلك البيانات. لذا تم استعمال هذه النظرية على نطاق واسع في مجال معالجة الصور والتنبؤ في مجال السلاسل الزمنية وتحسين النظام والتحكم والقرار وما إلى ذلك. يعد التنبؤ بالنظام الرمادي أحد أهم المكونات في نظرية النظام الرمادي الذي يعمل على إيجاد قواعد تطوير النظام من خلال إجراءات معينة لمعالجة البيانات من أجل التنبؤ بمستقبل النظام علمياً وكمياً.

عندما تكون بيانات السلسلة الزمنية ذات تقلبات كبيرة ونطاق تأرجحها أكبر، يكون من الصعب بناء نموذج مناسب لوصف سلوك تلك السلسلة الزمنية، في هذه الحالة يمكننا التفكير في التنبؤ بموجة تطوير البيانات السلوكية في المستقبل بناءً على موجة البيانات الأصلية، ويسمى هذا النوع من التنبؤ بالتنبؤ بالموجات. يعد التنبؤ بالموجات أحد المحتويات المهمة لنظام التنبؤ الرمادي، وعملية النمذجة الخاصة به واضحة وبسيطة، لذا فإن التنبؤ بالموجات يكتسب قيمة عالية للتطبيق في العديد من المجالات.

في هذا البحث سيتم استعمال نماذج التنبؤ الموجية الرمادية للتنبؤ بإنتاج النفط الخام في العراق بالاعتماد على بيانات السلسلة الزمنية لمعدل الانتاج الشهري والتي تعاني مشاهداتها من التقلبات الكبيرة.

## 1. Introduction

In systems theory, colors are used to express the availability of information or not for the studied system, as the white color is used to express the availability of complete information for the system, so it is called the white system, while the black color used to express the complete lack of information and is called the black system, while the system is called the gray system if the information about it is tainted by uncertainty or accuracy, and for this system, we find that some of the information is known and the other part is unknown.

Gray system theory can be defined as a research in uncertainties due to its use of little data and information. This theory is used with little and uncertain data for small samples of data and partial and weak information as the main research inputs.

The gray system was proposed by the Chinese scientist Deng Ju-Long in 1982 [8]. Which is adopted as a method for forecasting non-linear time series, and this system has been developed by researchers and adopted in various fields, economic, industrial and natural phenomena.

Gray models are useful for predicting the behavior of a monotonous time series, meaning that it is characterized as ascending or descending pattern, as it gives predictions with high accuracy, but it is inefficient in the case of time series with periodic and transverse fluctuations, which prompted researchers to think about predicting the wave of data behavior in the future based on the wave of the original data and this type of prediction is called the wave prediction, which is one of the most important gray forecasting and the adopted model is called the gray wave model and it is of interest in this research.

This research has been divided into five main sections, the first includes the introduction, the second includes the gray prediction model and clarification of its mechanism of action and steps for its construction, the third includes the gray wave model and clarification of the justifications for its existence and steps for its construction, and the fourth section includes the applied aspect of the research where the application is applied to two series of data, and finally the section fifth for conclusions.

## 2. Gray Forecasting Model

The gray forecasting model is one of the quantitative models used in prediction, which depends on the gray generation function, where the cumulative generation process is used to form differential equations that have properties that require less data, so this model appears as an effective tool for obtaining valuable information from small data sets [9]. This model is referred to as  $GM(p,q)$ , as the symbol  $p$  indicates the order of the differential equation and the symbol  $q$  indicates the number of variables included in the model [4].

As a special case of the gray forecasting model  $GM(p,q)$ , the  $GM(1,1)$  model, which is characterized by its simplicity and accuracy, as well as being the most widely used despite the presence of several other types, but most researchers focus on using this model for its computational efficiency, and this model is

described as A first-order linear gray model that consists of a first-order differential equation with only one variable [18].

The idea of GM(1,1) can be summarized in transforming the original data series of random nature into a regular data series to achieve an arrangement that meets the requirements of the forecasting model, where the raw data that suffers from randomness undergoes a cumulative assembly process called accumulative generation operator (AGO), then the differential equation of the model is solved depending on the data series resulting from the accumulation process to get the expected value, then the inverse of the accumulation process is applied to get the predictive value [14]. Thus, we can say that prediction in the gray models includes three main processes: the cumulative generation process, the gray modelling, and the inverse accumulated generating operation [5], and these steps will be explained as follows [1], [10], [16]:

I. Preparation the original time series data assuming that it is represented as follows [1]:

$$x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)) \quad , n \geq 4 \quad \dots (1)$$

Where n represent the time series length.

And then using the accumulative generation operator to create a new sequence for the original data series, as follows:

AGO:  $x^{(0)} \rightarrow x^{(1)}$

$$x^{(1)} = [x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)] \quad \dots (2)$$

Where  $x^{(1)}$  represents the new series after cumulative additions, which is obtained as follows:

$$x^{(1)}(m) = \sum_{k=1}^m x^{(0)}(k) \quad , m = 1, 2, \dots, n \quad \dots (3)$$

The effect of randomness and irregularity in the original data series will be weak and decreases after the accumulative generation operator takes place [6].

II. Generate a new time series denoted by  $z(1)$  , whose values are the averages of the adjacent values of the time series  $x^{(1)}$  which generated using the accumulative generation operator eq.(2), as following:

$$z^{(1)}(m) = 0.5 (x^{(1)}(m) + x^{(1)}(m - 1)) \quad , m = 2, 3, \dots \quad \dots (4)$$

III. Create a grey differential equation as shown in the following equation which represents the first-order differential equation through which we can model  $x^{(1)}$ :

$$x^{(0)}(m) + \alpha z^{(1)}(m) = \beta \quad , m = 2, 3, \dots, n \quad \dots (5)$$

Where:  $\alpha$ : the developed coefficient.  $\beta$ : the grey controlled variable.

Equation (3) can be reformulated as follow:

$$x^{(0)}(m) = -\alpha z^{(1)}(m) + \beta \quad , m = 2, 3, \dots, n \quad \dots (6)$$

We can formulate equation (6) in the form of matrices, as follows:

$$\begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix} = \begin{bmatrix} -z^{(0)}(2) & 1 \\ -z^{(0)}(3) & 1 \\ \vdots & \vdots \\ -z^{(0)}(n) & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \quad \dots (7)$$

The parameters of the gray model are estimated in equation (7) using the least square method, as follows:

$$\hat{p} = \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = (Z'Z)^{-1}Z'y$$

Where:

$$Z = \begin{bmatrix} -z^{(0)}(2) & 1 \\ -z^{(0)}(3) & 1 \\ \vdots & \vdots \\ -z^{(0)}(n) & 1 \end{bmatrix}, \quad y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}$$

After getting the values  $\hat{\alpha}$  and  $\hat{\beta}$  the response time values  $\hat{x}^{(1)}(m)$  are calculated as follows:

$$\hat{x}^{(1)}(m) = \left( x^{(0)}(1) - \frac{\hat{\beta}}{\hat{\alpha}} \right) e^{-\hat{\alpha}(m-1)} + \frac{\hat{\beta}}{\hat{\alpha}} \quad ; m = 1, 2, \dots, n, n + 1, \dots \quad \dots (8)$$

IV. After all the values of  $\hat{x}^{(1)}$  are obtained, the inverse accumulated generating operation is performed using the following equation [7]:

$$\hat{x}^{(0)}(m + 1) = \hat{x}^{(1)}(m + 1) - \hat{x}^{(1)}(m) \quad \dots (9)$$

### 3. Gray Wave Forecasting Model

Gray models of one or more variables are useful for predicting the behavior of a monotonous time series, as they give high accuracy predictions, but they are inefficient in the case of time series with cyclical and episodic fluctuations, which prompted researchers to think about predicting the wave of data behavior in the future based on the original data wave. The type of prediction of the wave prediction, which is one of the most important gray predictions and the adopted model is called the gray-wave model [12], the idea of the gray wave forecasting model is summarized in three main steps: selecting the contour lines, determining the time sequences, and applying the gray model GM(1, 1) on the time sequences and these steps will be explained as follows [13]:

I. Choosing contour lines: In the gray wave, equal contour lines are originally used based on the maximum and minimum values in the data. This method is suitable when fluctuations are regular, however, it may not be, so Chen et al. [3] suggested unequal contour lines. Now, let's symbolize the original series as follows:

$$X = (x(1), x(2), \dots, x(n)) \quad \dots (10)$$

then we arrange this series in ascending order to get the following series:

$$X^a = (x^a(1), x^a(2), \dots, x^a(n)) \quad \dots (11)$$

Let  $\xi_0 = x^a(1)$ ,  $\xi_s = x^a(n)$ , and  $\xi_1, \xi_2, \dots, \xi_{s-1}$  be s-quantile data, where:

$$\xi_i = \begin{cases} x^a(\lceil w \rceil) & w \text{ is not integer} \\ \frac{[x^a(w) + x^a(\lceil w + 1 \rceil)]}{2} & w \text{ is integer} \end{cases} \quad \dots (12)$$

Where:

$$w = n(i/s) \quad \dots (13)$$

$\xi_i$  represents the unequal-interval contour lines of the original series.

II. Identifying contour time sequences: which are determined based on the intersects between the unequal interval contour lines  $\xi_i$  obtained from equation (12), and the original time series X, let  $X_{\xi_i} = (P_1, P_2, \dots, P_{m_i})$  be a set of intersection points,  $P_j$  lies on the  $t_j$  broken line and the coordinates of  $P_j$  are  $(q(j), \xi_j)$ , where [15]:

$$q(j) = t_j + \frac{\xi_j - x(t_j)}{x(t_j + 1) - x(t_j)}, \quad j = 1, 2, \dots, m \quad \dots (14)$$

Thus  $Q_i^{(0)} = (q_i(1), q_i(2), \dots, q_i(m_i)), i = 0, 1, 2, \dots, s$  represent the time sequences of the contour line  $\xi_i$ .

III. Filtering contour time sequences and forecasting with GM(1,1): For gray wave forecasting, originally used time sequences that contain 4 or more elements [15], but when the time series fluctuates, this is not convenient to predict but it is necessary to filter these sequences when used with GM(1,1) [2]. The idea of filtering depends on the autocorrelation of the time series, as the autocorrelation measures the similarity between the observations as a function of time, based on it we can choose the time sequences whose ends are close to the sequence of the first future value, then we can use the GM(1,1) model based on the sequences that are selected and as follows [17]:

Let  $t_1^f$  represent the serial number of the first forecasted value and  $Q_i^{(0)} = (q_i(1), q_i(2), \dots, q_i(m_i))$  represent the time sequences chosen for the contour line  $\xi_i$ , for which  $(t_1^f - q_i(m_i))$  is very small. Establish the gray model GM(1,1) based on the chosen time sequences  $Q_i^{(0)}$  to get the forecasting values of

$q_i(m_i + 1), q_i(m_i + 2), \dots, q_i(m_i + k_i)$ . Arranged all the elements in contour time sequences  $Q_0^{(0)}, Q_1^{(0)}, \dots, Q_s^{(0)}$  in ascending order and deleting the invalid values, let  $(q_i(1) < q_i(2) < \dots < q_i(n_s))$  are the forecasting series, such that  $n_s \leq \sum_{i=1}^s (m_i + k_i)$ . If  $q(k)$  is a point located on the contour line  $\xi_{q(k)}$ , the in sample estimated and out sample forecasting of the generated wave is as follows:

$$X = X^{(0)} = \left\{ \xi_{q(k)} + \frac{t - q(k)}{q(k+1) - q(k)} [\xi_{q(k+1)} - \xi_{q(k)}] \right\}, \quad k = 1, 2, \dots, n_s \quad \dots (15)$$

**4. Application**

The data under research, represented the monthly average of crude oil production, in millions of barrels in Iraq, for the period from January 2016 to December 2021, that is we have a monthly time series of 72 observations. These data are presented in Table (1), and Figure (1) illustrates the plot of this series.

Table (1): Monthly crude oil production from Jan. 2016 to Dec. 2021.

|           | 2016  | 2017  | 2018  | 2019 | 2020  | 2021 |
|-----------|-------|-------|-------|------|-------|------|
| January   | 88.3  | 102.4 | 129.8 | 132  | 135.2 | 142  |
| February  | 95.5  | 85.3  | 118.1 | 118  | 122.1 | 127  |
| March     | 95.7  | 111.6 | 125.9 | 132  | 135.1 | 140  |
| April     | 91.9  | 103.2 | 121.9 | 126  | 130.8 | 135  |
| May       | 98.5  | 110.1 | 125.5 | 131  | 135.2 | 142  |
| June      | 93.3  | 115.7 | 123.2 | 126  | 130.8 | 138  |
| July      | 98.9  | 123.5 | 128.3 | 132  | 138.3 | 143  |
| August    | 101.8 | 124.8 | 129.9 | 132  | 138.3 | 144  |
| September | 103.8 | 120.6 | 126.6 | 129  | 133.8 | 139  |
| October   | 102.8 | 121.6 | 131.6 | 131  | 138.3 | 142  |
| November  | 98.1  | 120.3 | 128.5 | 124  | 133.6 | 138  |
| December  | 112.2 | 128.1 | 135.9 | 130  | 138.6 | 141  |

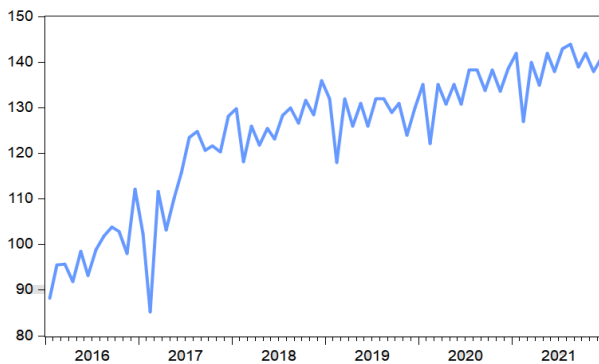


Figure (1): Monthly crude oil production from Jan. 2016 to Dec. 2021.



We note from Figure 1 that the series is fluctuant, and this is supported by the descriptive statistics related to it, as the difference between the maximum value and the minimum value is relatively large, and the results of the Jarque-Bera test indicate that these data do not follow the Normal distribution and the following table shows these measures:

Table (2): Description statistics of monthly crude oil production.

| Mean     | Max | Min  | Std. Dev. | Jarque-Bera Test       |
|----------|-----|------|-----------|------------------------|
| 123.5736 | 144 | 85.3 | 15.23413  | 9.494792<br>(0.008674) |

Table 3 shows the autocorrelation coefficients from the first lag to the tenth lag, which indicate that this series is highly auto-correlative, so we can choose the time sequences from it.

Table (3): Autocorrelation analysis of crude oil production.

| Lag   | 1     | 2     | 3     | 4     | 5     |
|-------|-------|-------|-------|-------|-------|
| ACF   | 0.834 | 0.820 | 0.783 | 0.710 | 0.675 |
| prob. | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Lag   | 6     | 7     | 8     | 9     | 10    |
| ACF   | 0.587 | 0.583 | 0.523 | 0.495 | 0.465 |
| prob. | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |

For prediction purposes, this series consisting of 72 observations has been divided into two parts, the first part for the period from January 2016 to December 2020, meaning that it is 60 observations, and the second part is 12 observations for the period between January and December of 2021. The first part was used to select the contour lines and determine the time sequences, and the second part was used to evaluate the prediction, and the prediction in this research is a prediction of 12 steps.

Now we start with the first step of the gray wave prediction, which is represented by choosing the contour lines, where  $s=9$  was used, meaning there are ten contour lines as shown in figure (2), Table (4) also shows the number of broken lines that intersect each of the contour lines. while table (5) presents the time sequences which represent the second step.

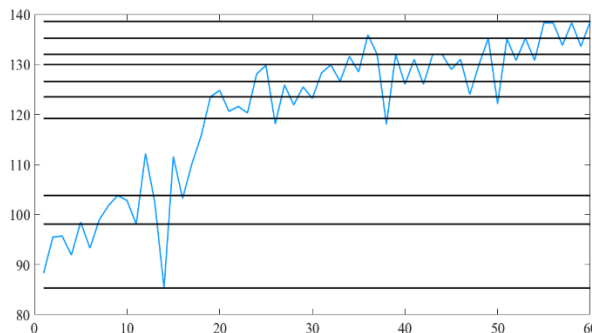


Figure (2): Ten contour lines for crude oil production.

Table (4): Contour lines and their intersection points.

| $i$     | 0    | 1    | 2     | 3     | 4     | 5     | 6      | 7   | 8     | 9     |
|---------|------|------|-------|-------|-------|-------|--------|-----|-------|-------|
| $\xi_i$ | 85.3 | 98.1 | 103.8 | 119.2 | 123.5 | 126.6 | 129.95 | 132 | 135.2 | 138.6 |
| $m_i$   | 1    | 6    | 6     | 5     | 13    | 14    | 17     | 12  | 10    | 1     |
| $t_i$   | 14   | 4    | 9     | 18    | 19    | 23    | 25     | 35  | 35    | 60    |
|         |      | 5    | 11    | 25    | 20    | 25    | 32     | 37  | 36    |       |
|         |      | 6    | 12    | 26    | 23    | 30    | 33     | 39  | 49    |       |
|         |      | 11   | 14    | 37    | 25    | 33    | 34     | 43  | 51    |       |
|         |      | 13   | 15    | 38    | 26    | 37    | 35     | 44  | 53    |       |
|         |      | 14   | 16    |       | 27    | 38    | 37     | 48  | 54    |       |
|         |      |      |       |       | 28    | 39    | 38     | 49  | 56    |       |
|         |      |      |       |       | 29    | 40    | 39     | 50  | 57    |       |
|         |      |      |       |       | 30    | 41    | 40     | 51  | 58    |       |
|         |      |      |       |       | 37    | 42    | 41     | 52  | 59    |       |
|         |      |      |       |       | 38    | 46    | 42     | 53  |       |       |
|         |      |      |       |       | 49    | 47    | 44     | 54  |       |       |
|         |      |      |       |       | 50    | 49    | 45     |     |       |       |
|         |      |      |       |       |       | 50    | 46     |     |       |       |
|         |      |      |       |       |       |       | 48     |     |       |       |
|         |      |      |       |       |       | 49    |        |     |       |       |
|         |      |      |       |       |       | 50    |        |     |       |       |

Table (5): Contour time sequences.

| $i$         | 0     | 1     | 2     | 3     | 4     | 5     | 6     | 7     | 8     | 9     |
|-------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| $Q_i^{(0)}$ | 14.00 | 4.94  | 9.00  | 18.45 | 19.00 | 23.81 | 24.99 | 35.47 | 35.91 | 60.00 |
|             |       | 5.08  | 11.40 | 25.91 | 20.31 | 25.27 | 31.98 | 37.00 | 36.18 |       |
|             |       | 6.86  | 12.86 | 26.14 | 23.41 | 30.67 | 33.67 | 39.00 | 49.00 |       |
|             |       | 11.00 | 14.70 | 37.91 | 25.54 | 33.00 | 34.53 | 43.00 | 50.98 |       |
|             |       | 13.25 | 15.93 | 38.09 | 26.69 | 37.39 | 35.20 | 44.00 | 53.00 |       |
|             |       | 14.49 | 16.09 |       | 27.60 | 38.61 | 37.15 | 48.38 | 54.59 |       |
|             |       |       |       |       | 28.44 | 39.90 | 38.85 | 49.24 | 56.69 |       |
|             |       |       |       |       | 29.87 | 40.12 | 39.34 | 50.76 | 57.31 |       |
|             |       |       |       |       | 30.06 | 41.88 | 40.79 | 51.72 | 58.66 |       |
|             |       |       |       |       | 37.61 | 42.10 | 41.21 | 52.27 | 59.32 |       |
|             |       |       |       |       | 38.39 | 46.63 | 42.66 | 53.73 |       |       |
|             |       |       |       |       | 49.89 | 47.43 | 44.68 | 54.16 |       |       |
|             |       |       |       |       | 50.11 | 49.66 | 45.47 |       |       |       |
|             |       |       |       |       |       | 50.35 | 46.15 |       |       |       |
|             |       |       |       |       |       |       | 47.99 |       |       |       |
|             |       |       |       |       |       | 49.40 |       |       |       |       |
|             |       |       |       |       |       | 50.60 |       |       |       |       |

Now, we start with the third step, which represents the filtering of the time sequences and the formation of the GM(1,1) model, we know that the serial number of the first future value is 61, and the series showed strong autocorrelation as shown in Table (3), and  $Q_6^{(0)}, Q_7^{(0)}$  and  $Q_8^{(0)}$  in GM(1,1) modeling because it is the closest to the future value, and the accumulative generation operator will be found according to equation (2) for all of these sequences, as shown in table (6).

Table (6): Accumulating generator operators.

| <i>i</i>   | 6      | 7      | 8      |
|------------|--------|--------|--------|
| <b>AGO</b> | 24.99  | 35.47  | 35.91  |
|            | 56.97  | 72.47  | 72.08  |
|            | 90.64  | 111.47 | 121.08 |
|            | 125.17 | 154.47 | 172.06 |
|            | 160.37 | 198.47 | 225.06 |
|            | 197.52 | 246.86 | 279.65 |
|            | 236.37 | 296.10 | 336.34 |
|            | 275.71 | 346.86 | 393.65 |
|            | 316.50 | 398.58 | 452.31 |
|            | 357.71 | 450.86 | 511.63 |
|            | 400.37 | 504.58 |        |
|            | 445.05 | 558.74 |        |
|            | 490.53 |        |        |
|            | 536.68 |        |        |
|            | 584.67 |        |        |
|            | 634.07 |        |        |
| 684.67     |        |        |        |

Then we get the values of  $z^{(1)}(m)$  according to equation (4) and as in the following table:

Table (7). The mean generation of consecutive neighbors' sequences.

| <i>i</i>    | 6      | 7      | 8      |
|-------------|--------|--------|--------|
| $z_i^{(1)}$ | 40.98  | 53.97  | 54.00  |
|             | 73.81  | 91.97  | 96.58  |
|             | 107.91 | 132.97 | 146.57 |
|             | 142.77 | 176.47 | 198.56 |
|             | 178.94 | 222.67 | 252.35 |
|             | 216.94 | 271.48 | 307.99 |
|             | 256.04 | 321.48 | 364.99 |
|             | 296.11 | 372.72 | 422.98 |
|             | 337.11 | 424.72 | 481.97 |
|             | 379.04 | 477.72 |        |
|             | 422.71 | 531.66 |        |
|             | 467.79 |        |        |
|             | 513.60 |        |        |
|             | 560.67 |        |        |
|             | 609.37 |        |        |
|             | 659.37 |        |        |

Using least squares method, the following estimates were obtained:

$$\hat{p}_6 = \begin{bmatrix} \hat{\alpha}_6 \\ \hat{\beta}_6 \end{bmatrix} = \begin{bmatrix} -0.0295 \\ 31.534 \end{bmatrix}$$

$$\hat{p}_7 = \begin{bmatrix} \hat{\alpha}_7 \\ \hat{\beta}_7 \end{bmatrix} = \begin{bmatrix} -0.0354 \\ 37.67 \end{bmatrix}$$

$$\hat{p}_8 = \begin{bmatrix} \hat{\alpha}_8 \\ \hat{\beta}_8 \end{bmatrix} = \begin{bmatrix} -0.0417 \\ 42.09 \end{bmatrix}$$

After getting the values of  $\hat{\alpha}$  and  $\hat{\beta}$  the response time values are calculated as follows:

$$q_6^{(1)}(m) = \left( x^{(0)}(1) - \frac{\hat{\beta}_6}{\hat{\alpha}_6} \right) e^{-\hat{\alpha}_6(m-1)} + \frac{\hat{\beta}_6}{\hat{\alpha}_6} = 1093.939e^{0.0295(m-1)} - 1068.949$$

$$q_7^{(1)}(m) = \left( x^{(0)}(1) - \frac{\hat{\beta}_7}{\hat{\alpha}_7} \right) e^{-\hat{\alpha}_7(m-1)} + \frac{\hat{\beta}_7}{\hat{\alpha}_7} = 1099.594e^{0.0354(m-1)} - 1064.124$$

$$q_8^{(1)}(m) = \left( x^{(0)}(1) - \frac{\hat{\beta}_8}{\hat{\alpha}_8} \right) e^{-\hat{\alpha}_8(m-1)} + \frac{\hat{\beta}_8}{\hat{\alpha}_8} = 1045.263e^{0.0417(m-1)} - 1009.353$$

then:

$$q_i(m) = q_i^{(1)}(m) - q_i^{(1)}(m-1)$$

After the estimated values are obtained, we can predict the time series, and the value of the root mean squares error criterion was calculated and its value was equal to 3.0723, the original series and the predictive series was drawn together in the following figure, which shows the extent of convergence between them.

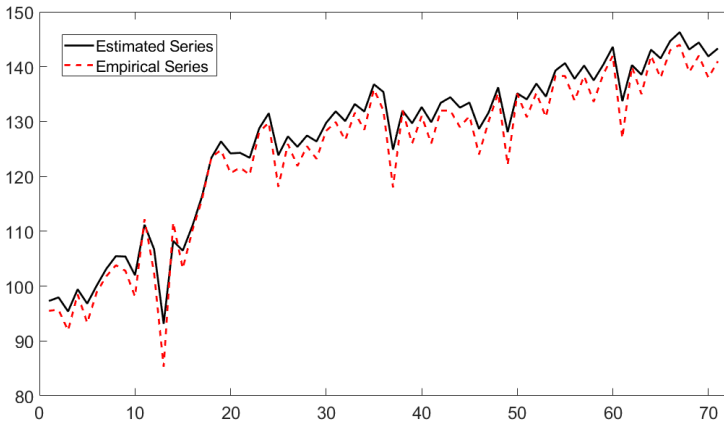


Figure (3): The estimated and the empirical series.

## 5. Conclusion

In this research, the monthly rate of crude oil production in Iraq was studied, and the results showed that this series is not a monotonous series, and had irregular fluctuation ranges, the results also showed that these data do not follow the Normal distribution and that the autocorrelation is strong. The Contour time sequences (6, 7, 8) were chosen and used to predict the series because they are the closest to the first future value. A graph showed how close the original empirical series was to what was estimated by the gray wave method.

**References**

1. Chang, C. J., Der-Chiang, L. I., Chien-Chih, C. H. E. N., & Wen-Chih, C. H. E. N. (2019). EXTRAPOLATION-BASED GREY MODEL FOR SMALL-DATASET FORECASTING. *Economic Computation & Economic Cybernetics Studies & Research*, 53(1).
2. Chen, Y., Liu, B., & Wang, T. (2020). Analyzing and forecasting China containerized freight index with a hybrid decomposition–ensemble method based on EMD, grey wave and ARMA. *Grey Systems: Theory and Application*.
3. Chen, Y., Zou, Y., & Wang, S. (2015). Grey wave forecasting with qualified unequal-interval contour lines and its application in metal prices prediction. In 2015 IEEE International Conference on Grey Systems and Intelligent Services (GSIS) (pp. 261-265). IEEE.
4. Duman, G. M., Kongar, E., & Gupta, S. M. (2019). Estimation of electronic waste using optimized multivariate grey models. *Waste Management*, 95, 241-249.
5. Geng, N., Zhang, Y., Sun, Y., Jiang, Y., & Chen, D. (2015). Forecasting China's annual biofuel production using an improved grey model. *Energies*, 8(10), 12080-12099.
6. Guo, Y. L., Zhao, M. H., & Deng, Z. W. (2014). Tunnel surrounding rock deformation forecast analysis based on GM and FEM. *Electronic Journal of Geotechnical Engineering*, 19, 1379-1394.
7. Hu, Y. C. (2020). Energy demand forecasting using a novel remnant GM (1, 1) model. *Soft Computing*, 24(18), 13903-13912.
8. Julong, D. (1982). Grey control system. *Journal of Huazhong University of Science and Technology*, 3(9), 18.
9. Kahraman, C., & Yavuz, M. (2010). *Production engineering and management under fuzziness* (Vol. 252). Springer.
10. Lin, L., & Lind, H. (2010). AN APPLICATION OF GREY SYSTEM THEORY INTO REAL ESTATE INVESTMENT DECISION-MAKING (Doctoral dissertation, MS thesis, Department of Real Estate and Construction Management, KTH, Stockholm, Sweden).
11. Liu, S., Dang, Y., Fang, Z. (2010). *Grey System Theory and Its Application*. Fifth edition Beijing China: Science Press.
12. Liu, S., & Forrest, J. Y. L. (2010). *Grey systems: theory and applications*. Springer Science & Business Media.
13. Liu, S. & Lin, Y. (2010). *Grey Systems: Theory and Applications*. Springer.
14. Liu, S., Yang, Y., Xie, N., & Forrest, J. (2016). New progress of grey system theory in the new millennium. *Grey Systems: Theory and Application*.
15. Wan, Q., Wei, Y., & Yang, X. (2010). Research on grey wave forecasting model. In *Advances in Grey Systems Research* (pp. 349-359). Springer, Berlin, Heidelberg.

16. Wu, W. Z., & Zhang, T. (2020). An improved gray interval forecast method and its application. *Communications in Statistics-Theory and Methods*, 49(5), 1120-1131.
17. Yin, K., Lu, D., & Li, X. (2017). A novel grey wave method for predicting total Chinese trade volume. *Sustainability*, 9(12), 2367.
18. Zhang, L., Wang, L., Zheng, Y., Wang, K., Zhang, X., & Zheng, Y. (2017). Time prediction models for echinococcosis based on gray system theory and epidemic dynamics. *International journal of environmental research and public health*, 14(3), 262.

**Publication Prerequisites and terms**

- 1- The journal publishes scientific research and studies in statistics and informatics written in Arabic, English and French, to make it clear that research submitted for publication has been published or submitted for publication in magazines or other periodicals or presented and published in periodicals for conferences or seminars.
- 2- Send electronic copies (word & PDF) of the research and studies to the editor should include the name of the researcher or researchers and their scientific titles and places of work with the address of the correspondence, the numbers of telephones and e-mail. The research to be published should be sent electronically in accordance with the specifications below:
  - a. To be printed on A4 paper and be in the form of a single column and use the Type simplified Arabic and Times New Roman for English and French and with a font size (12). Using Microsoft Word and on one face of the paper.
  - b. The margin is 2.5 cm for all sides of the paper.
  - c. The researcher will attach a summary of his research in Arabic, English, or French in no more than one page.
  - d. Place references at the end of the paper and separated page. It is recommended to use the Harvard system of referencing, which (author's name, year of publication, source address, publishing house, country).
  - e. Numbered tables, illustrations, and others as they are received in the research, documents as aliases of the original sources.
  - f. The number of search or study pages should not exceed (25) pages.
- 3- Authors will notified of receiving their research within two working days from the date of receipt of the research.
- 4- Referees will evaluate all submitted research, the Authors will informed of the proposed evaluation and modifications if any within two weeks of receipt of the research.
- 5- The editorial board of the Journal has the right to accept or reject the research and has the right to make any modification or partial redrafting of the material submitted for publication in accordance with the format adopted in its publication after the approval of the researcher.
- 6- Published research becomes the property of the Journal and may not republished elsewhere.
- 7- The articles published in the magazine reflect the opinions of the authors, and do not necessarily reflect the view of the Journal or the Arab Institute for Training and Research in Statistics.
- 8- The research is sent to the magazine's e-mail address:  
[journal@aitrs.org](mailto:journal@aitrs.org) or [Info@aitrs.org](mailto:Info@aitrs.org)

# Journal of Statistical Sciences

Scientific Referred Journal

## Editorial Board

### Editor-in Chief

Mr. Hedi Saidi

### Editorial Secretary

Dr. Bachioua Lahcene

## Editorial Board Members

Prof. Dr. Faisal Al-Sharabi

Dr. Salwa Mahmoud Assar

Dr. Hassan Abuhassan

Prof. Dr. Abed Khaliq Tohami

Dr. Ahmed Shaker Almutwali

Dr. Hamid Bouzida

Prof. Dr. Mukhtar Al-Kouki

Prof. Dr. Issa Masarweh

## Scientific Consulting Committee

Dr. Qassim Al-Zoubi

Dr. Diaa Awad

Dr. luay shabaneh

Dr. Nabeel M. Shams

Dr. Khalifa Al-Barwani

Prof. Dr. Ghazi Raho

Dr. Ola Awad

Prof. Dr. Awad Haje Ali

Prof. Dr. Maytham Elaibi Ismael

Dr. Mohammed Husain Ali Al-Janabi

Listed in Ulrich's website

[www.ulrichsweb.com](http://www.ulrichsweb.com)

Classified in The Arab Citation & Impact Factor (Arcif)

[www.emarefa.net/arcif/](http://www.emarefa.net/arcif/)

ISSN 2522-64X (Online), ISSN 2519-948X (Print)

P.O. Box: 851104- Amman 11185 Jordan ص.ب: 851104 عمان 11185 الأردن

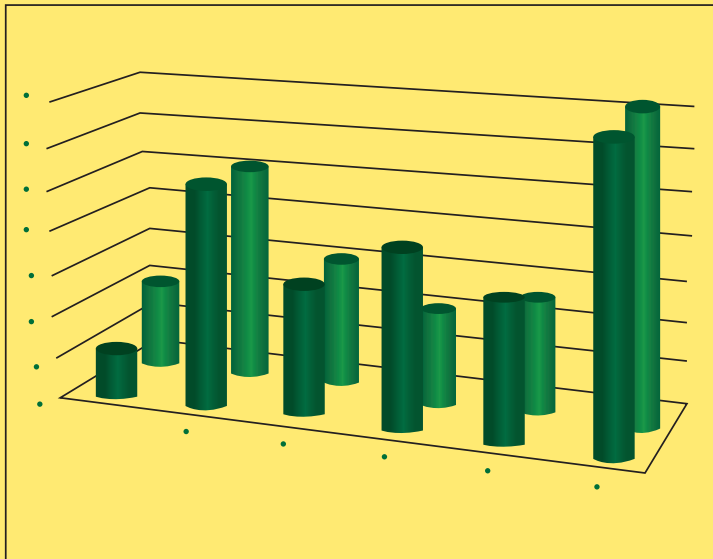
Tel: +96265549805 – Fax: +96265549804- [www.aitrs.org](http://www.aitrs.org) , [info@aitrs.org](mailto:info@aitrs.org)





**Arab Institute for Training and Research in Statistics**

# Journal of Statistical Sciences



**Issue No. 20**

**Scientific Peer-reviewed Journal issued by  
Arab Institute for Training and Research in Statistics**

Listed in Ulrich's website

[www.ulrichsweb.com](http://www.ulrichsweb.com)

Classified in The Arab Citation & Impact Factor (Arcif)

[www.emarefa.net/arcif/](http://www.emarefa.net/arcif/)

ISSN 2522-64X (Online), ISSN 2519-948X (Print)