

## تحسين تقدير نموذج انحدار بواسون باستعمال خوارزمية مستعمرة النمل: دراسة تطبيقية على جرائم الخطف في العراق

أمنة نافع محسن<sup>1</sup> أسيل عبد الرزاق رشيد<sup>2</sup>

الجامعة المستنصرية/ كلية الادارة والاقتصاد/ قسم الاحصاء  
الجامعة المستنصرية/ كلية الادارة والاقتصاد/ قسم الاحصاء

[aseelstat@uomustansiriyah.edu.iq](mailto:aseelstat@uomustansiriyah.edu.iq) [Amna.nafi@uomustansiriyah.edu.iq](mailto:Amna.nafi@uomustansiriyah.edu.iq)

### مستخلص البحث:

يعد انموذج انحدار بواسون من النماذج الإحصائية الأساسية المستخدمة لتحليل بيانات العد إذ يتمثل الهدف في تقدير عدد مرات حدوث حدث معين ضمن فترة زمنية أو مساحة محددة. يعتمد هذا النموذج على فرضية أن المتغير التابع يتبع توزيع بواسون، والذي يفترض أن يكون المتوسط مساوياً للتباين، يُستخدم هذا النموذج في العديد من التطبيقات مثل تقدير عدد العملاء أو الحوادث أو الحالات المرضية. تُعدّ خوارزميات الذكاء الاصطناعي من الأدوات الحديثة ذات القدرة العالية على تحسين دقة النماذج الإحصائية والتغلب على القيود التقليدية لطرائق التقدير الكلاسيكية. ومن بين هذه الخوارزميات، تبرز خوارزمية مستعمرة النمل (Ant Colony Optimization - ACO) كإحدى الخوارزميات المبتكرة الفعالة في إيجاد الحلول المثلى ضمن فضاءات معقدة وعالية الأبعاد. يهدف هذا البحث إلى تحسين تقدير أنموذج انحدار بواسون باستعمال خوارزمية مستعمرة النمل وتطبيقه في تحليل البيانات الجنائية، ولا سيما بيانات جرائم الخطف في العراق خلال الفترة (2024-2021)، والتي تشمل المناطق الشمالية والوسطى والجنوبية من البلاد.

تم مقارنة أداء النموذج المحسّن باستعمال الخوارزمية مع الطريقة التقليدية، والتحقق من ملاءمة النموذج من خلال اختباري Scaled Deviance و Pearson Chi-square. وقد أظهرت النتائج وجود تشتت زائد (Over-Dispersion) وتشتت ناقص (under - dispersion) في البيانات، إلا أنه تم معالجته بنجاح، مما حسن من أداء النموذج واستقراره. كما تم اختبار مشكلة التعدد الخطي (Multi-Collinearity) بين المتغيرات المستقلة، وأثبتت النتائج خلو النموذج منها.

تؤكد نتائج الدراسة أن دمج خوارزميات الذكاء الاصطناعي مع النماذج الإحصائية الكلاسيكية يسهم في رفع كفاءة ودقة التقدير وتحسين تفسير الظواهر الجنائية، مما يعزز من موثوقية النتائج الإحصائية ويدعم صنّاع القرار الأمني في وضع استراتيجيات أكثر فاعلية لمكافحة الجريمة وتوزيع الموارد الأمنية.

**الكلمات المفتاحية:** انموذج انحدار بواسون، خوارزمية مستعمرة النمل، جرائم الخطف.

### 1. المقدمة: [1]

يعد انموذج انحدار بواسون من النماذج الإحصائية الأساسية المستخدمة لتحليل بيانات العد، حيث يتمثل الهدف في تقدير عدد مرات حدوث حدث معين ضمن فترة زمنية أو مساحة محددة. يعتمد هذا النموذج على فرضية أن المتغير التابع يتبع توزيع بواسون، والذي يفترض أن يكون المتوسط مساوياً للتباين، يُستخدم هذا النموذج في العديد من التطبيقات مثل تقدير عدد العملاء أو الحوادث أو الحالات المرضية ضمن سياق معين. رغم أهمية انموذج انحدار بواسون، فإنه يواجه عدة تحديات في التطبيق العملي، من أبرزها افتراض تماثل المتوسط والتباين. في العديد من الحالات الواقعية، قد يكون التباين أعلى من المتوسط، وهو ما يُعرف بالتشتت الزائد، مما يؤدي إلى تقليل كفاءة النموذج في تقديم توقعات دقيقة، إضافة إلى ذلك، يعتمد النموذج على فرضية العلاقات الخطية بين المتغيرات المستقلة

والمتغير التابع، مما قد لا يعكس بشكل دقيق العلاقات المعقدة في البيانات، لتجاوز التحديات المرتبطة بدقة النماذج الإحصائية مثل انحدار بواسون، يتم الاعتماد على خوارزميات الذكاء الاصطناعي والتعلم الآلي، من بين هذه الخوارزميات، خوارزمية مستعمرات النمل (ACO) نموذجًا مستوحى من السلوك الطبيعي للنمل في البحث عن الغذاء، حيث تُحاكي آلية التواصل الجماعي بين الأفراد من خلال إفراز الفيرومونات (Pheromones) التي تُستخدم كوسيلة لتوجيه النمل نحو المسارات المثلى. تعتمد هذه الخوارزمية على مبدأ التعاون والتعلم التراكمي بين الأفراد في المستعمرة، إذ يقوم كل نمل بتجربة مسار محتمل، وتزداد قوة المسارات التي تؤدي إلى حلول أفضل من خلال تحديث تراكيز الفيرومونات عليها. يُسهّم هذا السلوك الجماعي المنظم في تحقيق توازن بين الاستكشاف (Exploration) لفضاء الحلول والبحث المكثف حول المناطق الواعدة (Exploitation)، مما يجعل الخوارزمية فعالة في تحسين عملية التقدير والوصول إلى الحلول المثلى ضمن المسائل الإحصائية المعقدة.

#### انحدار بواسون (Poisson regression) [2,4]

يُعرف انحدار بواسون (Poisson Regression) كمنهجية نمذجة يتم فيها تقدير المتغير التابع (Response Variable) الذي يأخذ قيمًا على هيئة بيانات معدودة (Count Data) أو معدلات (Rate Data). يعتمد توزيع المتغير التابع على توزيع بواسون (Poisson Distribution) بمعلمة  $(\lambda)$ ، حيث يكون التوقع والتباين متساويين  $(E(Z) = V(Z) = \lambda)$  تتأثر هذه القيم بالمتغيرات التوضيحية (Explanatory Variables)، ويتم استخدام الانحدار الخطي المعمم (Generalized Linear Model) مع الدالة اللوغاريتمية (Log Function) لضمان أن  $\lambda$  تأخذ قيمًا موجبة، وبناء علاقة خطية بين  $(\log \lambda)$  والمتغيرات التوضيحية. وهو أحد أنواع النماذج الإحصائية التي تُستخدم في تحليل البيانات التي تعتمد على العد (count data)، حيث يكون المتغير التابع (المتغير المستجيب) عبارة عن عدد صحيح يمثل عدد مرات حدوث حدث معين خلال فترة معينة أو مكان معين، ويمكن أن يصف عدد الحوادث الجنائية التي تحدث في منطقة معينة خلال شهر معين. الصيغة العامة لنموذج انحدار بواسون

$$\lambda = e^{A\beta + U} \dots \dots \dots (1)$$

#### 2. طريقة الامكان الاعظم (Maximum Likelihood Estimation – MLE) لتقدير

##### معلومات نموذج انحدار بواسون [6,5]

تُعدّ طريقة الإمكان الأعظم (MLE) واحدة من أكثر طرائق التقدير استخدامًا في الإحصاء الرياضي والاستدلال الإحصائي، وتُستخدم لتقدير معلمات النماذج الإحصائية من خلال تعظيم دالة الإمكان (Likelihood Function) بناءً على القيم المرصودة.

الفكرة الأساسية لطريقة الإمكان الأعظم هي البحث عن مجموعة القيم المثلى للمعلمات (Parameters) التي تجعل البيانات المشاهدة أكثر احتمالاً ضمن إطار النموذج المفترض. وتُطبق هذه الطريقة على نطاق واسع في نماذج العد، مثل نموذج انحدار بواسون، وخاصة في تحليل البيانات الجنائية التي تشمل أعداد المجني عليهم أو تكرار الجرائم.

نفترض  $Z_i = Z_1, Z_2, \dots, Z_k$  تتوزع بواسون بمعلمة  $(\lambda)$  حيث  $\text{pr}(Z_i | \lambda)$  هي دالة الكثافة الاحتمالية أو دالة كتلة الاحتمال لمتغير عشوائي  $Z_i$  تحت المعلمة  $\lambda$ .

$$\text{Pr}(Z_i | \lambda) = \frac{\lambda^Z e^{-\lambda}}{Z!} \quad Z = 1, 2, 3, \dots$$

دالة الامكان الاعظم تكتب بالشكل التالي

$$L(\lambda) = \prod_{i=1}^n \Pr(Z_i | \lambda) \dots \dots \dots (2)$$

يتم التعويض عن دالة الاحتمالية لتوزيع بواسون بدالة الامكان الاعظم

$$L(\lambda) = \prod_{i=1}^n \frac{\lambda^{Z_i} e^{-\lambda}}{Z_i!} \dots \dots \dots (3)$$

يتم فك الحد السابق

$$L(\lambda) = \frac{\prod_{i=1}^k \lambda^{Z_i} \prod_{i=1}^k e^{-\lambda}}{\prod_{i=1}^n Z_i!} \dots \dots \dots (4)$$

بأخذ Ln لدالة الامكان الاعظم للملاحظات

$$\ln L(\lambda) = \ln \left[ \frac{\prod_{i=1}^k \lambda^{Z_i} \prod_{i=1}^k e^{-\lambda}}{\prod_{i=1}^n Z_i!} \right] \dots \dots \dots (5)$$

$$= \ln(e^{-k\lambda}) + \ln(\lambda^{\sum_{i=1}^k Z_i}) + \ln \left( \prod_{i=1}^k Z_i! \right) \dots \dots \dots (6)$$

$$= -n\lambda + \sum_{i=1}^k Z_i \ln(\lambda) - \ln \left( \prod_{i=1}^n Z_i! \right) \dots \dots \dots (7)$$

يتم اشتقاق مشتقة اولى و مشتقة ثانية

$$\frac{\partial \ln L(\lambda)}{\partial \ln(\lambda)} = -n\lambda + \sum_{i=1}^k Z_i - ZERO \dots \dots \dots (8)$$

$$\frac{\partial^2 \ln L(\lambda)}{\partial \lambda^2} = -n \dots \dots \dots (9)$$

ومن خلال النظر الى المعادلات (8,9) نلاحظ عدم امكانية حل هذه المعادلات بالنسبة الى المعلمة ( $\lambda$ ) بشكل مباشر، ولهذا السبب نلجأ الى استخدام خوارزمية نيوتن رافسن والتي نسعى من خلالها الى تعظيم دالة الامكان الاعظم لتقدير معالمات نموذج انحدار بواسون.

### 3. خوارزمية مستعمرة النمل (Ant Colony Optimization - ACO) [7]

تعد خوارزمية مستعمرة النمل (ACO) إحدى خوارزميات الذكاء السربي (Swarm Intelligence)، استلهمت من السلوك الجماعي للنمل في الطبيعة، وخاصة قدرته على إيجاد أقصر طريق بين المستعمرة ومصدر الغذاء عبر تبادل المعلومات باستخدام مادة الفيرومون (Pheromone). تُحاكي الخوارزمية هذا السلوك من خلال مجموعة من "النملات الاصطناعية" التي تبحث في فضاء الحلول، حيث تُحدّث خريطة الفيرومون لتوجيه عملية البحث نحو الحلول الأفضل، بينما تُخفّض أهمية المسارات الضعيفة عبر تبخر الفيرومون، مما يحقق توازناً بين الاستكشاف والاستغلال. تُستخدم خوارزمية ACO في هذا البحث لتحسين تقدير معاملات نموذج انحدار بواسون لجريمة الخطف في العراق، إذ تعمل على إيجاد القيم المثلى للمعاملات التي تُقلل الخطأ الإحصائي وتُحسن من دقة النموذج التنبؤية. وتحسين كفاءة النموذج في تفسير العوامل المؤثرة في الظاهرة الإجرامية. تعتمد الخوارزمية على نظام ديناميكي لتحديث الفيرومون على المسارات، وهو يُعبّر عنه رياضياً كالتالي:

$$N_{ij}^{(k+1)} = N_{ij}^{(k)} \cdot (N - 1) + \Delta_{ij} N$$

$N_{ij}^{(k)}$ : تمثل كمية الفيرومونات على المسار الواقع بين الموقعين  $i$  و  $j$  في الدورة  $k$ .

$N$ : معامل التبخر، ويحدد نسبة التناقص في الفيرومونات من دورة إلى أخرى،  $1 > N > 0$ .

$\Delta_{ij} N$ : كمية الفيرومونات الجديدة المضافة على المسار  $ij$  بناءً على جودة الحل المكتشف في الدورة

الأخيرة.

$N_{ij}^{(k+1)}$ : كمية الفيرومونات بعد التحديث في الدورة التالية:

الربط بتطبيق تحليل الجريمة

- المسار  $ij$  قد يعبر عن علاقة بين متغيرين في الجريمة، كالعلاقة بين الكثافة السكانية وعدد حالات الانتحار أو بين المناطق الجغرافية وتكرار الجريمة.
- الفيرومونات  $N_{ij}^{(k)}$  تمثل "درجة الأهمية الإحصائية" أو "قوة العلاقة" بين متغيرين من خلال ما اكتشفته

الخوارزمية.

• معامل التبخر  $N$  يُساعد على إزالة الآثار القديمة تدريجياً، حتى لا تُكرَّر التفسيرات الضعيفة أو المؤقتة

#### - الربط خوارزمية ACO بانحدار بواسون

يُعد نموذج انحدار بواسون أحد أهم النماذج الإحصائية المستخدمة في تحليل البيانات التي تتبع توزيعات عددية (count data)، كما في حالات تحليل عدد الجرائم شهرياً أو سنوياً. إلا أن هذا النموذج يعتمد في دقته على اختيار مجموعة مناسبة من المتغيرات التفسيرية التي تؤثر بشكل فعلي على متغير الاستجابة. وهنا تبرز أهمية خوارزمية أمثلية مستعمرة النمل (ACO)، حيث يمكن توظيفها كأداة ذكية لأغراض اختيار المتغيرات (Feature Selection)، والتي تُعد خطوة حاسمة في بناء نماذج بواسون دقيقة وفعالة. تعمل خوارزمية النمل على البحث ضمن جميع التركيبات الممكنة من المتغيرات التفسيرية، لاختيار المجموعات التي تُحقق أفضل أداء إحصائي لنموذج بواسون، وذلك من خلال تقييم الحلول باستخدام معايير دقة التقدير مثل معيار المعلومات أكيكي (AIC) أو معيار معلومات بيزي (BIC) أو معامل التحديد الزائف ( $R^2$ ) في كل دورة من الخوارزمية، تمثل

"النملة الاصطناعية" اختياراً لمجموعة من المتغيرات، ويتم تقدير نموذج بواسون بناءً على هذه المجموعة، ومن ثم يُحسب أداء النموذج. وتُستخدم نتيجة هذا التقدير لتحديث خريطة الفيرومونات، بحيث تُعزز التركيبات الناجحة وتُضعف غير الفعالة. وبهذه الطريقة، يتم دمج ACO مع انحدار بواسون ضمن إطار هجين يجمع بين الذكاء الاصطناعي والنماذج الإحصائية الكلاسيكية، مما يتيح الحصول على نماذج أكثر كفاءة وقدرة على تفسير سلوك الجريمة بناءً على المتغيرات الأكثر تأثيراً. حيث كل نملة ( $Ant$ ) تمثل مجموعة فرعية من المتغيرات المستقلة ( $A_1, A_2, \dots, A_n$ ) كل مسار تبنيه النملة هو اختيار معين للمتغيرات لتضمينها في النموذج

#### - مراحل عمل خوارزمية أمثلية مستعمرة النمل ACO

تمر خوارزمية أمثلية مستعمرة النمل (ACO) بعدة مراحل أساسية تهدف إلى بناء وتحسين الحلول لمسألة الأمثلية، ويُمكن تكيف هذه المراحل لتلائم اختيار أفضل مجموعة من المتغيرات

التفسيرية التي تُستخدم ضمن نموذج انحدار بواسون Poisson Regression لتحليل بيانات الجرائم. وتتمثل هذه المراحل بما يلي:

- 1- تهيئة المعلمات (Initialization): في هذه المرحلة تُحدد المعلمات الأولية للخوارزمية، مثل:
  - عدد النملات (عدد الحلول المحتملة التي تُبنى في كل دورة).
  - عدد الدورات. (Iterations)
  - معاملات الفيرومون، ودالة الجاذبية .
  - معدل تبخر الفيرومون. ( $\alpha$ )

• قيم الفيرومون الابتدائية على كل مسار (بين المتغيرات المحتملة).

- 2- تهيئة الفيرومون (Pheromone Initialization): تُوزع كميات متساوية من الفيرومون على جميع المتغيرات أو المسارات المحتملة، وذلك لعدم وجود تفضيل مبدئي لأي متغير عند بداية الخوارزمية.

- 3- بناء الحلول (Solution Construction): يقوم كل "نمل اصطناعي" ببناء حل تدريجي من خلال اختيار مجموعة من المتغيرات التفسيرية المناسبة، حيث يتم:
  - التنقل بين المتغيرات المحتملة اعتمادًا على دالة احتمالية تعتمد على كمية الفيرومون ودالة الجاذبية.

• بناء مجموعة فرعية من المتغيرات (Subset) تمثل حلاً مقترحاً لنموذج بواسون.

- 4- تقييم الحلول (Solution Evaluation): بعد بناء كل حل، يُستخدم لتكوين نموذج انحدار بواسون. ويُقيم النموذج باستخدام معايير إحصائية مثل:

• Akaike Information Criterion (AIC).

• Bayesian Information Criterion (BIC).

• Residual Deviance.

• Mean Squared Error (MSE) في حال التحقق من جودة التنبؤ.

- 5 تحديث الفيرومون (Pheromone Update): يتم تحديث قيم الفيرومون على المتغيرات المختارة في كل حل، بحيث:

• تزداد كمية الفيرومون على المتغيرات التي ساهمت في بناء نماذج بواسون عالية الجودة.

• يتم تقليل كمية الفيرومون تدريجيًا على المسارات الأقل فاعلية، وفقًا لمعدل التبخر

- 6 تكرار العملية (Iteration) تُعاد خطوات البناء والتقييم والتحديث لعدد من الدورات (generations)، ما يسمح للنظام بالتعلم التدريجي والتوجه نحو الحل الأمثل.

- 7- اختيار أفضل الحلول: (Best Solution Selection) بعد انتهاء جميع الدورات، يُختار الحل الذي أدى إلى أقل قيمة AIC أو أعلى دقة تنبؤية كمجموعة المتغيرات المثلى التي سيتم اعتمادها في نموذج بواسون النهائي.

اختبار والد (Wald Test) للكشف عن مشكلة التشنت

ان مشكلة التشنت في نموذج انحدار بواسون، تعد مسألة شائعة في تحليل البيانات العددية. في حالة توزيع بواسون، يُفترض أن يكون التباين مساويًا للمتوسط، وهذه الخاصية تُعرف باسم تساوي التشنت (Equidispersion). ومع ذلك، في العديد من التطبيقات العملية، قد يكون التباين أكبر من المتوسط، وهذا يُعرف باسم فوق التشنت (Over dispersion) أما إذا كان التباين أقل من المتوسط، فهذه الحالة تسمى تحت التشنت (Under dispersion).

ويعتبر Wald Test هو أحد الاختبارات الإحصائية المستخدمة للتحقق من وجود تشتت عالي (Over dispersion) في نماذج العد، حيث يُستخدم اختبار والد للتحقق من الفرضية القائلة بأن التباين يساوي المتوسط (أي لا يوجد زيادة في التشتت).  
الفرضيات

$H_0: \theta = 1$  الفرضية الصفرية أي لا يوجد فوق التشتت

$H_1: \theta \neq 1$  البديلة الفرضية يوجد فوق أو تحت التشتت

$$W = \frac{(1 - \bar{\theta})^2}{\text{var}(\bar{\theta})} \dots \dots \dots (11)$$

حيث ان :-

$\bar{\theta}$  التقدير المقدر لمعامل التشتت

$\text{var}\bar{\theta}$  تباين التقدير

إذا كانت  $W > X^2_{1,\alpha^2}$  أي يوجد فوق التشتت.

إذا كانت  $W \leq X^2_{1,\alpha^2}$  أي لا يوجد فوق التشتت.

### 5 متوسط مربع الخطأ لدقة النموذج (Mean Squared Error - MSE)

يُستخدم متوسط مربع الخطأ (MSE) كمقياس لمدى دقة النموذج في التنبؤ بالقيم الفعلية. يعبر هذا المقياس عن متوسط مربعات الفروقات بين القيم الفعلية (Actual Values) والقيم المتوقعة (Predicted Values) بواسطة النموذج. يتم حساب MSE باستخدام المعادلة التالية:

$$MSE = \frac{1}{n} \sum_{i=1}^n (Z_i - \bar{Z}_i)^2 \dots \dots \dots (12)$$

$Z_i$  هي القيم الفعلية (Actual Values).

$\bar{Z}_i$  هي القيم المتوقعة (Predicted Values) من النموذج.

$n$  هو عدد النقاط أو البيانات في العينة (Sample Size).

كلما كانت قيمة MSE أصغر، كان النموذج أكثر دقة في التنبؤ. ومع ذلك، فإن MSE حساس للقيم المتطرفة (Outliers)، لأن التربيع يزيد من تأثير الأخطاء الكبيرة.

### 6 معامل التحديد الزائف لاختبار كفاءة النموذج (Pseudo-R<sup>2</sup>)

يُعد معامل التحديد  $R^2$  أداة أساسية في الانحدار الخطي، حيث يقيس نسبة التباين المفسر بواسطة المتغيرات المستقلة. غير أنّ هذا المفهوم لا يمكن نقله مباشرة إلى نماذج الانحدار العام (Generalized Linear Models – GLMs) مثل انحدار بواسون، نظراً لاختلاف طبيعة دالة الارتباط (link function) وتوزيع الاستجابة. ولهذا طُوّرت عدة مؤشرات بديلة تُعرف باسم معاملات التحديد الزائفة (Pseudo-R<sup>2</sup>) ، تهدف إلى تقييم جودة النموذج بشكل تقريبي ومقارن. باستخدام المعادلة التالية:

$$R^2 = 1 - \frac{D_{model}}{D_{null}} \dots \dots \dots (13)$$

حيث ان :-

$D_{model}$  الـ Residual Deviance للنموذج الكامل

$D_{null}$  الـ Null Deviance للنموذج الصفري

### 7 معالجة مشكلة التشتت الزائد و الناقص في أنحدار بواسون [3]

بعد توفيق نموذج انحدار بواسون للبيانات، من الضروري تقييم مدى ملاءمته وجودته. يُعد توزيع بواسون من التوزيعات الخاصة، إذ يتميز بتساوي المتوسط والتباين. وبناءً على هذه الخاصية، يمكن استخدام نسبة التباين إلى المتوسط كمؤشر يُعرف ب معلمة التشتت. إذا كانت قيمة معلمة التشتت تساوي 1 فإن ذلك يدل على أن البيانات تتبع توزيع بواسون القياسي، وبالتالي يُعد النموذج ملائمًا، أما إذا كانت القيمة أكبر من 1 ، فإن هذا يشير إلى وجود فوق التشتت (Over dispersion) أي أن التباين في البيانات أكبر من المتوسط، مما يعني أن نموذج بواسون القياسي غير ملائم، وفي حال كانت معلمة التشتت أقل من 1 ، فإننا نواجه حالة تحت التشتت (Under dispersion) والتي تُعد أيضًا دليلًا على عدم ملاءمة النموذج .  
في حال اكتشاف وجود فوق التشتت أو تحت التشتت، فمن الضروري إعادة توفيق النموذج ويمكن التعبير بالشكل الاتي :-

$$\text{Var}(z) = \zeta \lambda \quad \dots(14)$$

بدلا من

$$\text{Var}(z) = \lambda$$

### 8 الجانب العملي (التطبيقي): وصف البيانات

#### 1-8 العينة ومجتمع الدراسة

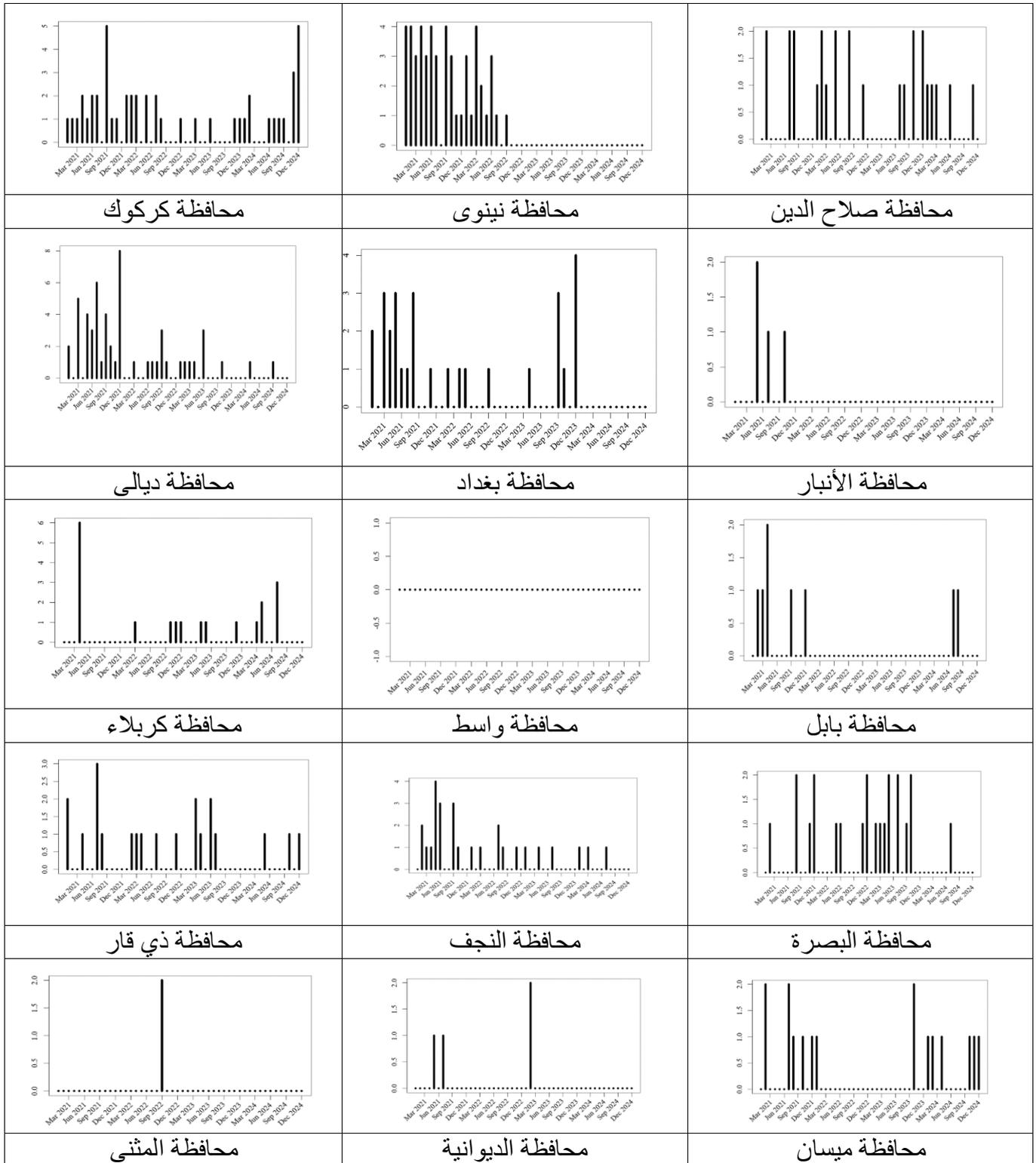
يمثل مجتمع الدراسة جميع حالات جرائم الخطف المسجلة في العراق خلال الفترة موضوع البحث، موزعة حسب المناطق الجغرافية: الشمالية (صلاح الدين، نينوى، كركوك)، الوسطى (بغداد، ديالى، بابل، واسط، الأنبار)، والجنوبية (البصرة، كربلاء المقدسة، النجف الأشرف، ذي قار، ميسان، الديوانية، المثنى). ولأغراض الدراسة، تم الاعتماد على عينة مكونة من 144 في المنطقة الشمالية، و240 في المنطقة الوسطى، و336 في المنطقة الجنوبية، مأخوذة من السجلات الرسمية لوزارة الداخلية العراقية، مما يضمن موثوقية ودقة البيانات. يمثل المتغير التابع عدد مجني عليهم، بينما تشمل المتغيرات المستقلة 9 متغيرًا (A1-A9) تمثل العوامل المباشرة وغير المباشرة المؤثرة في وقوع الجريمة، وقد تم ترميزها لتسهيل إدخالها في النماذج الإحصائية. وسيتم تحليل هذه البيانات باستخدام نموذج انحدار بواسون (Poisson Regression) لدراسة تأثير هذه العوامل على عدد حالات الخطف، مع تحسين تقدير المعلمات باستخدام خوارزمية مستعمرة النمل (ACO) لتعزيز دقة النموذج وكفاءته التفسيرية والتنبؤية لتسهيل إدخالها في النماذج الإحصائية وكما موضحة في الجدول رقم (1)

#### جدول رقم (1) ترميز اسباب الجريمة

الرمز	سبب الجريمة	الرمز	سبب الجريمة	الرمز	سبب الجريمة
A1	الانتقام والعداوة	A4	الإفلاس والحاجة	A7	المشاكل العائلية
A2	تحقيق النفع الذاتي	A5	الشذوذ الجنسي	A8	مجهولة
A3	المطالبة بالدين	A6	العلاقات الغرامية	A9	أخرى

### 2-8 التحليل الإحصائي للخطف في المحافظات العراقية

يتضمن الشكل (1) بيانات الحالات الشهرية للخطف في المحافظات العراقية للسنوات من 2021 إلى 2024، وكما يأتي:



شكل رقم (1) يمثل الشهرية للخطف في المحافظات العراقية

### 3-8 حسن المطابقة للبيانات

للتحقق من ملائمة نماذج انحدار بواسون في تمثيل بيانات عدد المجنى عليهم في المناطق الشمالية والوسطى والجنوبية، وتم تقدير معلمة التشتت بطريقتين: الأولى باستخدام إحصائية مربع كاي لبيرسون (Pearson Chi-square)، والثانية من خلال الانحراف المقاس (Scaled Deviance). كما استُخدم اختبار Wald للتحقق من معنوية معاملات النماذج، وقد بينت النتائج أن المؤشرات المتحصلة توفر دليلاً على ملائمة النماذج وقدرتها التفسيرية.

#### جدول (2) نتائج معاملات التشتت واختبار Wald لحالات الخطف في المحافظات العراقية

المناطق	Methods	$\hat{\sigma}$	W	p-value
المنطقة الشمالية	Pearson	0.39914	24.18908	<0.00001
	Deviance	0.46878	18.90727	0.00001
المنطقة الوسطى	Pearson	0.62816	15.90007	0.00007
	Deviance	0.48292	30.74826	<0.00001
المنطقة الجنوبية	Pearson	1.39122	24.94817	<0.00001
	Deviance	1.44053	31.02075	<0.00001

أظهرت نتائج تحليل معاملات التشتت لحالات الخطف في المحافظات العراقية تفاوتاً ملحوظاً بين المناطق الثلاث. ففي المنطقة الشمالية، بلغت معاملات التشتت وفق طريقة بيرسون (0.39914) والانحراف المقاس (0.46878)، بينما سجلت المنطقة الوسطى 0.62816 و0.48292 على التوالي، وهذه القيم جميعها أقل من القيمة المثالية (1)، مما يشير إلى وجود تشتت ناقص في البيانات، أي أن التباين الفعلي للبيانات أقل مما يتوقعه نموذج بواسون. وعلى الجانب الآخر، سجلت المنطقة الجنوبية معاملات تشتت أكبر من 1، حيث بلغت 1.39122 وفق بيرسون و1.44053 وفق الانحراف المقاس، ما يعكس ميلاً خفيفاً نحو تشتت زائد، مع بقائها قريبة نسبياً من التباين المتوقع، مما يشير إلى تحقق افتراضات نموذج بواسون بشكل أفضل في هذه المنطقة مقارنة بالشمال والوسط. وقد دعمت نتائج اختبار Wald هذا التفسير في جميع المناطق، إذ جاءت قيم الاحتمال  $p < 0.05$  ما أدى إلى رفض فرضية العدم القائلة بأن التباين مطابق لمتوقع نموذج بواسون. وبناءً على ذلك، يظهر أن البيانات في المناطق الشمالية والوسطى تتطلب تصحيح التباين لضمان دقة التقديرات الإحصائية وتحسين ملائمة النموذج، بينما يمكن اعتبار النموذج صالحاً نسبياً للمنطقة الجنوبية، مع الاستفادة من معادلة تصحيح التشتت لضمان استدلال أكثر دقة وموثوقية في جميع المناطق. باستعمال معادلة التصحيح لضمان دقة الاستدلال الإحصائي كما في معادلة (14).

لاختبار وجود مشكلة التعدد الخطي بين المتغيرات التوضيحية، وللتأكد من خلو متغيرات الدراسة من ظاهرة التعدد الخطي (Multi-collinearity)، تم حساب معاملات تضخم التباين (Variance Inflation Factor - VIF) باستخدام برنامج R، وتشير نتائج ان جميع القيم كانت أقل من الحدّ التشخيصي الشائع (5)، وهو ما يُعد مؤشراً قوياً على عدم وجود مشكلة التعدد الخطي بين المتغيرات المستقلة. وبذلك يمكن القول إن المتغيرات التوضيحية مستقلة عن بعضها ولا يوجد ترابط خطي مرتفع قد يؤثر في استقرار تقديرات معاملات النموذج أو يؤدي إلى تضخيم الأخطاء المعيارية للمعلمات. وهذا يعزز من موثوقية النتائج اللاحقة، ويؤكد أن النموذج يمكنه تقدير أثر كل متغير بشكل مستقل دون تداخل إحصائي جوهري.

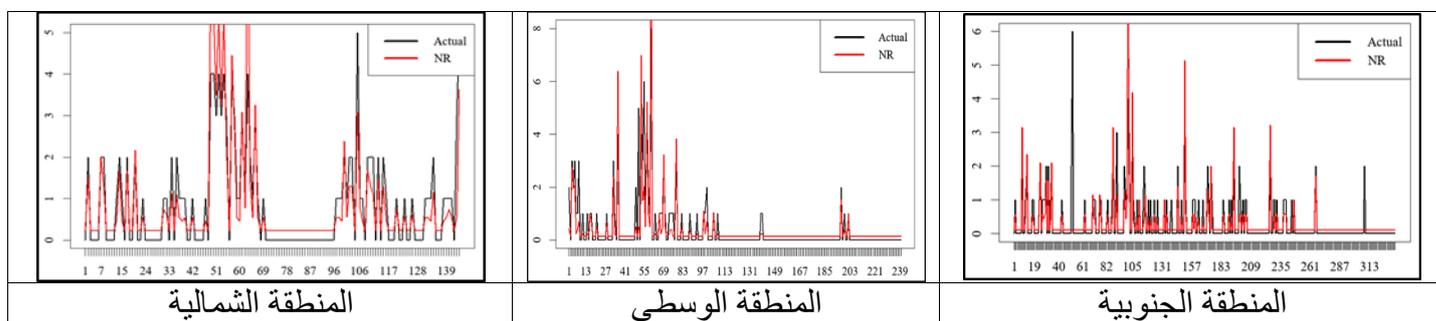
جدول رقم (3) معاملات تضخم التباين لبيانات الخطف في المحافظات العراقية

المحافظات الشمالية	المحافظات الجنوبية	المحافظات الوسطى
VIF	VIF	VIF
A1	1.02660	1.05696
A2	1.23344	1.02808
A3	1.00076	1.06438
A4	1.01311	1.00071
A5	1.08466	1.07424
A6	1.03153	1.13780
A7	1.10271	1.00270
A8	1.09229	1.00414
A9	1.00148	1.01226

4-8 تقدير نماذج انحدار بواسون

تم تقدير نموذج انحدار بواسون باستخدام خوارزمية NR وكانت النتائج كما في الجدول الآتي:  
جدول (4) نتائج اختبار المعلمات عبر خوارزمية NR لحالة الخطف في العراق

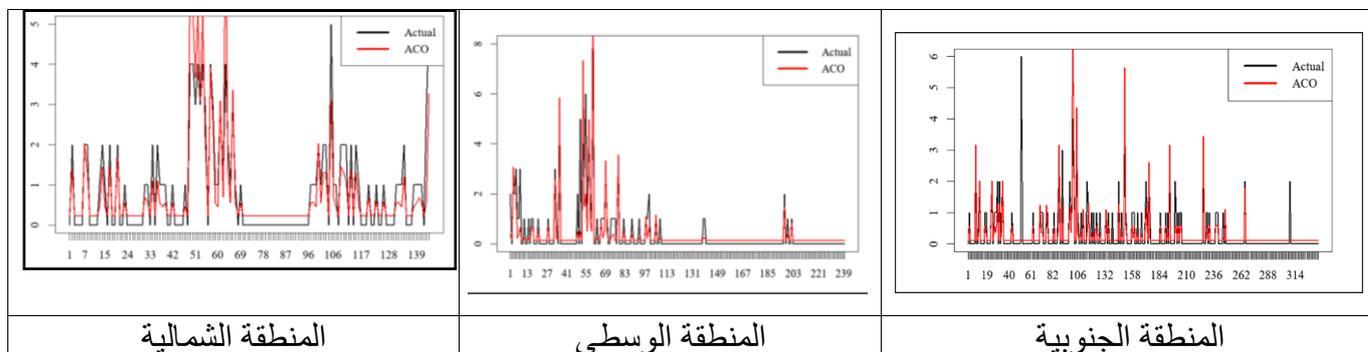
Methods	المنطقة الشمالية				المنطقة الوسطى				المنطقة الجنوبية			
	Estimate	SE	t-value	p-value	Estimate	SE	t-value	p-value	Estimate	SE	t-value	p-value
$\beta_0$	-1.46834	0.11494	-12.77453	0.00000	-2.07788	0.14333	-14.49743	0.00001	-2.22791	0.19799	-11.25290	0.00001
$\beta_1$	0.99170	0.13679	7.24984	0.00000	1.71362	0.28357	6.04312	0.00001	1.28551	0.29512	4.35587	0.00001
$\beta_2$	0.88247	0.08052	10.95948	0.00000	0.62327	0.09559	6.52042	0.00001	1.68769	0.19687	8.57269	0.00001
$\beta_3$	0.91938	0.15513	5.92644	0.00000	2.07788	0.80543	2.57986	0.00988	1.08862	0.24600	4.42525	0.00001
$\beta_4$	1.16979	0.46963	2.49090	0.01274	2.12392	0.47992	4.42558	0.00001	1.46053	0.42860	3.40763	0.00066
$\beta_5$	1.11972	0.20604	5.43449	0.00000	2.17639	0.34715	6.26922	0.00001	0.30206	0.62759	0.48130	0.63030
$\beta_6$	0.70146	0.08849	7.92718	0.00000	1.10333	0.11488	9.60425	0.00001	1.39965	0.32856	4.25999	0.00002
$\beta_7$	1.21379	0.27439	4.42357	0.00001	1.23130	0.17915	6.87307	0.00001	2.22790	0.62210	3.58128	0.00034
$\beta_8$	0.86487	0.05875	14.72115	0.00000	1.39555	0.12707	10.98228	0.00001	1.70021	0.31462	5.40395	0.00001
$\beta_9$	1.16933	0.19638	5.95443	0.00000	2.07789	0.57847	3.59206	0.00033	2.37681	0.42638	5.57439	0.00001
Pseudo R <sup>2</sup>	72.9%				66.2%				59.2%			
MSE	0.45616				0.51013				0.39767			



شكل رقم (2) القيم الحقيقية و القيم المقدرة لمحافظة العراقية باستخدام NR أظهرت نتائج تقدير نموذج انحدار بواسون لجريمة الخطف في العراق أن كافة المتغيرات التوضيحية كانت ذات دلالة إحصائية في المنطقة الشمالية و الوسطى، بينما كانت احدى المتغيرات غير معنوية في المناطق الجنوبية. سجل النموذج  $Pseudo R^2$  بقيم 72.9% و 66.2% و 59.2% على التوالي، فيما كانت MSE ضمن الحدود المقبولة، مما يعكس دقة التقديرات وكفاءة النموذج في تفسير عدد حالات الخطف. تم تقدير نموذج انحدار بواسون باستخدام خوارزمية ACO وكانت النتائج كما في الجدول الآتي:

جدول رقم (5) تقدير واختبار المعلمات باستخدام ACO

Method s	المنطقة الشمالية				المنطقة الوسطى				المنطقة الجنوبية			
	Estimate	SE	t-value	p-value	Estimate	SE	t-value	p-value	Estimate	SE	t-value	p-value
$\beta_0$	-1.45807	0.11503	12.67553	0.00000	2.07786	0.14375	14.45502	0.00001	-2.25024	0.19934	-11.28874	0.00001
$\beta_1$	0.91725	0.14313	6.40850	0.00000	1.71095	0.28503	6.00264	0.00001	1.25212	0.30263	4.13741	0.00004
$\beta_2$	0.86334	0.08238	10.47959	0.00000	0.58768	0.09703	6.05669	0.00001	1.70266	0.19793	8.60225	0.00001
$\beta_3$	0.94635	0.15242	6.20876	0.00000	1.61484	1.00932	1.59992	0.10962	1.13911	0.23855	4.77508	0.00001
$\beta_4$	1.24719	0.46886	2.66003	0.00781	2.22802	0.46270	4.81530	0.00001	1.60372	0.37879	4.23385	0.00002
$\beta_5$	0.97987	0.23086	4.24449	0.00002	2.17421	0.35392	6.14323	0.00001	0.60193	0.62598	0.96158	0.33626
$\beta_6$	0.69294	0.08961	7.73325	0.00000	1.08520	0.11896	9.12237	0.00001	1.25107	0.33860	3.69483	0.00022
$\beta_7$	1.10223	0.28935	3.80936	0.00014	1.30854	0.17369	7.53378	0.00001	2.35444	0.59424	3.96207	0.00007
$\beta_8$	0.86214	0.05931	14.53644	0.00000	1.38156	0.12598	10.96675	0.00001	1.74279	0.30755	5.66667	0.00001
$\beta_9$	1.08437	0.20950	5.17603	0.00000	1.89090	0.63191	2.99235	0.00277	2.47101	0.41856	5.90366	0.00001
Pseudo $R^2$	73.0%				67.4%				59.7%			
MSE	0.38936				0.40396				0.30021			



المنطقة الشمالية

المنطقة الوسطى

المنطقة الجنوبية

### شكل رقم (3) القيم الحقيقية والتقديرية لمحافظة العراقية باستخدام خوارزمية ACO

أظهرت النتائج أن خوارزمية ACO قادرة على تقدير المتغيرات بكفاءة عالية، كما هو موضح في الشكل رقم (3) والجدول رقم (5)، إذ تمكنت من تمييز المتغيرات المؤثرة عن غير المؤثرة بدقة. وتشير قيم **p-value** المنخفضة ( $0.05 <$ ) إلى دلالة معلمات المتغيرات المهمة، مما يعكس قوة النموذج في تفسير التباين في البيانات.

أما القيم غير الدالة ( $0.05 \geq \text{p-value}$ ) فتشير إلى أن بعض المتغيرات لا تسهم بشكل معنوي في النموذج، ويمكن الاستفادة من هذه النتيجة في تبسيط النموذج دون التأثير على مستوى الدقة.

### جدول (6) مقارنة بين خوارزميات الذكاء الاصطناعي في تحليل جريمة الخطف في العراق

Methods	المنطقة الشمالية		المنطقة الوسطى		المنطقة الجنوبية	
	Pseudo R <sup>2</sup>	MSE	Pseudo R <sup>2</sup>	MSE	Pseudo R <sup>2</sup>	MSE
NR	72.9%	0.45616	66.2%	0.51013	59.2%	0.39767
ACO	73.0%	0.38936	67.4%	0.40396	59.7%	0.30021

### 8. الاستنتاجات والتوصيات

1. في المناطق الشمالية والوسطى والجنوبية تفوقت خوارزمية ACO على خوارزمية NR مما يعكس تحسين دقة نموذج انحدار بواسون.
2. النموذج يحقق أفضل أداء في المنطقة الشمالية، حيث التقديرات دقيقة جداً والقدرة على تفسير التباين قوية، بينما تراجع الأداء في المنطقة (الوسطى والجنوبية)، وقد يعود هذا إلى وجود تباين أعلى في البيانات أو عوامل خارجية لم يتم تضمينها في النموذج أو ضعف العلاقة بين المتغيرات التوضيحية في هذه المناطق مقارنة بالشمالية حيث ان هذه النتائج تؤكد أن فعالية النموذج ليست متساوية عبر جميع المناطق، وأن المنطقة الشمالية تمثل حالة أكثر استقراراً وصلاحيّة لتطبيق النموذج.
3. توضح نتائج تحليل جريمة الخطف في العراق باستعمال خوارزمية ACO وجود تأثير لكافة الاسباب المؤثرة على جرائم الخطف في العراق، بينما في المنطقة الوسطى هناك تأثير لكافة المتغيرات عدا المتغير A3 والذي يمثل (المطالبة بالدين)، بينما اظهرت النتائج في المنطقة الجنوبية وجود تأثير لكافة المتغيرات المسببة لجريمة الخطف في العراق عدا المتغير A5 والذي يمثل (الشذوذ الجنسي).
4. اظهر نموذج انحدار بواسون باستعمال خوارزمية مستعمرة النمل بكفائته العالية على تمييز العوامل المؤثرة وحسب المنطقة المدروسة.
5. دراسة ظواهر اخرى باستعمال خوارزمية مستعمرة النمل في تقدير الانموذج.

6. استعمال طرائق تقدير اخرى ومقارنتها مع طريقة MLE بالاعتماد على خوارزمية مستعمرة النمل.

#### 9- المصادر

- 1- العوادي، إيثار حسين جواد. (2017). مقارنة بعض طرائق تقدير معلمات نموذج انحدار بواسون الهرمي الجزئي مع تطبيق عملي (رسالة ماجستير غير منشورة). قسم الإحصاء، كلية الإدارة والاقتصاد، جامعة بغداد.
- 2- ذاكر، سلمى ثابت، & جاسم، انتصار مجيد. (2017). تقدير معدل تكرار الإصابة باحتشاء العضلة القلبية باستخدام نموذج بواسون. مجلة الإدارة والاقتصاد & Economics Administration Journal، 40(111).
- 3- عبد الله، غادة، & يوسف، إسماعيل. (2019). اختيار المتغيرات في نموذج انحدار بواسون باستخدام خوارزمية الأعداب الضارة. المجلة العراقية للعلوم الإحصائية، 30(1).
- 4- محمد، غازي جواد. (2023). استخدام طريقة M الحصينة في تقدير أنموذج انحدار بواسون لأعداد العاطلين عن العمل في العراق. مجلة الإدارة والاقتصاد، 48(139)، 213-220.
- 5- Abonazel, M. R., & Saber, O. M. (2020). A comparative study of robust estimators for Poisson regression model with outliers. *Journal of Statistics Applications & Probability*, 9(2), 279–286.
- 6- Algamal, Z. Y. (2019). Variable selection in count data regression model based on Firefly Algorithm. *Statistics, Optimization & Information Computing*, 7, 520–529.
- 7- Alangood, N. A., Algamal, Z. Y., & Khaleel, M. A. (2024). Variable selection in Poisson regression model based on chaotic meta-heuristic search algorithm. *BIO Web of Conferences*, 97, 00161.  
<https://doi.org/10.1051/bioconf/2024970016>
- 8- Agresti, A. (2013). *Categorical data analysis* (3rd ed.). John Wiley & Sons.
- 9- Al-Zubaidi, M. S., & Ebtisam, K. (2024). Comparison of estimation methods for zero truncated Poisson regression model. *Journal of Economics and Administrative Sciences (JEAS)*, 30(142), 492–508.
- 10- Alamgir, A. A., Khan, S. A., Khan, D. M., & Khalil, U. (2013). A new efficient redescending M-estimator: Alamgir redescending M-estimator. *Research Journal of Recent Sciences*, 2(8), 79–91.
- 11- Dorigo, M., Birattari, M., & Stützle, T. (2006). Ant colony optimization: Artificial ants as a computational intelligence technique. *IEEE Computational Intelligence Magazine*, 1(4), 28–39.
- 12- Gabriella, A., Abdullah, S., & Soemartojo, S. (2019). Comparison of Poisson and quasi-Poisson regression: A simulation study. In *Proceedings of the IEEE International Conference on Software Architecture* (pp. 1–8).  
<https://doi.org/10.29244/icsa.2019.pp1-8>

- 13- Jassim, S. A. (2024). *Comparison of some methods for estimating (COM-Poisson) regression model* (Unpublished master's thesis). Al-Mustansiriyah University, College of Administration and Economics.
- 14- Lokonon, E. B. (2014). *Generalized linear models with Poisson family: Applications in ecology* (Master's thesis). Faculty of Agronomic Sciences, University of Abomey-Calavi.
- 15- McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models*. Chapman & Hall.
- 16- Ortíz-Barrios, M. A., Cleland, I., Nugent, C., Pancardo, P., Järpe, E., & Synnott, J. (2020). Simulated data to estimate real sensor events—A Poisson-regression-based modelling. *Sensors*, 20(5), 1400.  
<https://doi.org/10.3390/s20051400>
- 17- Rashedi, E., Nezamabadi-Pour, H., & Saryazdi, S. (2009). GSA: A gravitational search algorithm. *Information Sciences*, 179(13), 2232–2248.  
<https://doi.org/10.1016/j.ins.2009.03.004>
- 18- Rashedi, E., Rashedi, E., & Nezamabadi-Pour, H. (2018). A comprehensive survey on gravitational search algorithm. *Swarm and Evolutionary Computation*, 41, 141–158.  
<https://doi.org/10.1016/j.swevo.2017.11.006>
- 19- Samiman, A. S. D. (2019). Using Ant Colony Optimization Algorithm to achieve optimal allocation. *Journal of University of Anbar for Economic and Administrative Sciences*, 11(25), 497–513.
- 20- Spinelli, J. J., Lockhart, R. A., & Stephens, M. A. (2002). Tests for the response distribution in a Poisson regression model. *Journal of Statistical Planning and Inference*, 108, 137–154.
- 21- Winkelmann, R. (2008). *Econometric analysis of count data* (5th ed.). Springer-Verlag.
- 22- Zakariya, Y. (2012). Diagnostic in Poisson regression models. *Electronic Journal of Applied Statistical Analysis*, 5(2), 178–186.

**Improving the estimation of the Poisson regression model using the ant colony Optimization: An applied study on kidnapping crimes in Iraq**

Anna Nafi Mohsen<sup>1</sup>

Asst. Prof. Aseel Abdulrazzak  
Rasheed<sup>2</sup>

Department of Statistics, College of Administration and Economics,  
Mustansiriyah University, Iraq

[anna.nafi@uomustansiriyah.edu.iq](mailto:anna.nafi@uomustansiriyah.edu.iq) [aseelstat@uomustansiriyah.edu.iq](mailto:aseelstat@uomustansiriyah.edu.iq)

**Abstract**

Poisson regression is a fundamental statistical model used to analyze count data. The research objective is to estimate the frequency of a specific event within a defined time period or area. This model is based on the assumption that the dependent variable follows a Poisson distribution, which assumes that the mean equals the variance. This model is used in many applications, such as estimating the number of customers, accidents, or illnesses. Artificial intelligence algorithms are modern tools with a high capacity for improving the accuracy of statistical models and overcoming the traditional limitations of classical estimation methods. Among these algorithms, Ant Colony Optimization (ACO) stands out as a meta-innovative algorithm effective in finding optimal solutions within complex, high-dimensional spaces. This research aims to improve the estimation of the Poisson regression model using the ant colony Optimization and its application in the analysis of criminal data, particularly kidnapping crime data in Iraq during the period (2021–2024), which includes the northern, central and southern regions. The performance of the improved model using the algorithm was compared with that of the traditional method, and the model's suitability was verified using the Scaled Deviance and Pearson Chi-square tests. The results showed both over-dispersion and under-dispersion in the data, but these were successfully addressed, thus improving the model's performance and stability. The multi-collinearity problem between the independent variables was also tested, and the results confirmed that the model was free of this problem. The study results confirm that integrating artificial intelligence algorithms with classical statistical models contributes to increasing the efficiency and accuracy of estimation and improving the interpretation of criminal phenomena, which enhances the reliability of statistical results and supports security decision-makers in developing more effective strategies to combat crime and allocate security resources.

**Key Words:** Poisson regression model, Ant Colony Optimization (ACO), kidnapping crimes.