Journal of Al-Turath University College (JATUC)
Open Access Journal
ISSN : 2074-5621
https://jatuc.uoturath.edu.iq/

مقارنة تقدير أنموذج انحدار بواسون للبيانات الطولية باستعمال طريقة الإمكان الأعظم (MLE)وطريقة المعادلات المُعمَّمة (GEE)

مروان عبدالله قحطان*، نازك جعفر صادق

قسم الإحصاء – كلية الإدارة والاقتصاد – جامعة بغداد ، بغداد، العراق.

*Corresponding Author: marwan.qahtan2201@coadec.uobaghdad.edu.iq

الملخص

يهدف هذا البحث إلى تقدير أنموذج انحدار بواسون للبيانات الطولية ومقارنة كفاءة طريقتين من طرائق التقدير هما: طريقة الإمكان الأعظم (Maximum Likelihood Estimation - MLE) وطريقة المعادلات المعممة (Generalized Estimating Equations - GEE).

تمثل البيانات الطولية نوعاً خاصاً من البيانات الإحصائية التي تُجمع عن نفس الأفراد عبر فترات زمنية متعددة، مما يؤدي إلى وجود ترابط داخلي بين المشاهدات، وهو ما يجعل تحليلها أكثر تعقيداً من البيانات المقطعية.

تم اجراء دراسة محاكاة اعتمدت على توليد بيانات طولية خاضعة لتوزيع بواسون، حيث تم تقدير معلمات الأنموذج باستخدام الطريقتين المذكورتين ومقارنة نتائجهما باستخدام مقاييس الدقة مثل متوسط مربع الخطأ (Mean Squared Error - MSE)ومتوسط الخطأ المطلق (Mean Absolute Error - MAE).

أظهرت النتائج أن طريقة المعادلات المعممة توفر تقديرات أكثر ثباتاً في وجود ارتباط زمني بين المشاهدات، بينما تميزت طريقة الإمكان الأعظم بدقة أعلى عند ضعف الترابط الداخلي.

تُسهم هذه المقارنة في توضيح ملاءمة كل طريقة بحسب طبيعة البيانات والهدف من التحليل، مما يعزز من دقة النماذج الإحصائية المستخدمة في تحليل البيانات الطولية .

الكلمات المفتاحية

البيانات الطولية، انحدار بواسون، طريقة الإمكان الأعظم، طريقة المعادلات المعممة، المقارنة الإحصائية.

Abstract

This study aims to estimate the Poisson regression model for longitudinal data and to compare the efficiency of two estimation methods: Maximum Likelihood Estimation (MLE) and Generalized Estimating Equations (GEE).

Longitudinal data represent a special type of statistical data collected from the same subjects over multiple time periods, leading to within-subject correlation that complicates analysis compared to cross-sectional data.

The study utilized Poisson-distributed longitudinal data, where model parameters were estimated using both methods, and their performance was compared based on accuracy measures such as the Mean Squared Error (Mean Squared Error - MSE)and the Mean Absolute Error (Mean Absolute Error - MAE).

Results indicated that the GEE method provides more stable estimates in the presence of temporal correlation, whereas the MLE method yielded higher precision when within-subject correlation was weak.

This comparison highlights the suitability of each estimation approach depending on the data structure and analytical goals, thereby enhancing the reliability of statistical modeling for longitudinal data.

Keywords

Longitudinal data, Poisson regression, Maximum Likelihood Estimation (MLE), Generalized Estimating Equations (GEE), statistical comparison.

1. المقدمة

تُعدُّ البيانات الطولية (Longitudinal Data) من الأنواع المهمة في التحليل الإحصائي، إذ تُجمع عن نفس الوحدات أو الأفراد عبر فترات زمنية متعددة، مما يتيح دراسة التغيرات التي تطرأ على الظاهرة محل الاهتمام بمرور الزمن. وتمتاز هذه البيانات بوجود ترابط داخلي بين المشاهدات المتكررة للفرد الواحد، وهو ما يجعل تحليلها أكثر تعقيدًا مقارنةً بالبيانات المقطعية التي تفترض استقلالية الملاحظات.

وقد استُخدمت نماذج الانحدار العدّية، وبخاصة أنموذج انحدار بواسون(Poisson Regression) ، على نطاق واسع في تحليل البيانات التي تمثل عددًا من الأحداث أو التكرارات، مثل عدد الزيارات أو الإصابات أو مرات العلاج. إلا أن تطبيق هذا الأنموذج التقليدي على البيانات الطولية قد يكون غير كافٍ في كثير من الحالات، بسبب افتراض استقلالية المشاهدات، وهو افتراض غالبًا ما لا يتحقق في البيانات المتكررة عبر الزمن.

أشارت العديد من الدراسات السابقة إلى هذه الإشكالية، حيث قدم Liang and Zeger (1986) طريقة المعادلات المعممة (Generalized Estimating Equations – GEE)كبديل ملائم لتحليل البيانات الطولية، لكونها تأخذ بنظر الاعتبار الترابط الداخلي بين المشاهدات دون الحاجة إلى تحديد التوزيع الكامل للبيانات. كما ناقش Hardin and Hilbe (2012) كفاءة هذه الطريقة مقارنةً بطريقة الإمكان الأعظم (Maximum Likelihood Estimation – MLE)، موضحين أن أداء كل طريقة يعتمد على حجم العينة وطبيعة الترابط بين المشاهدات.

وفي المقابل، ما تزال طريقة الإمكان الأعظم تُعد من أكثر طرائق التقدير شيوعًا في النماذج الإحصائية، لما تتمتع به من دقة وكفاءة عالية عند تحقق افتراضات النموذج، ولا سيما في حالة ضعف الترابط الداخلي بين المشاهدات. ومن هنا برزت الحاجة إلى إجراء مقارنات منهجية بين هاتين الطريقتين عند تطبيقهما على أنموذج انحدار بواسون للبيانات الطولية، بهدف بيان مدى ملاءمة كل منهما في مواقف تحليلية مختلفة.

في المقابل، ظهرت طريقة المعادلات المعممة (Generalized Estimating Equations – GEE) التي قدّمها Liang و Zegerعام 1986، كطريقة بديلة لتقدير نماذج البيانات الطولية دون الحاجة إلى تحديد التوزيع الكامل للمشاهدات، إذ تعتمد على وصف الارتباط داخل الفرد من خلال مصفوفة ارتباط عاملة .(Working Correlation Matrix) وتُعد هذه الطريقة أكثر مرونة في التعامل مع البيانات المرتبطة، كما توفر تقديرات متسقة حتى في حال عدم صحة افتراضات التوزيع.

بناءً على ما تقدّم، يهدف هذا البحث إلى تقدير أنموذج انحدار بواسون للبيانات الطولية ومقارنة نتائج طريقتي الإمكان الأعظم والمعادلات المعممة من حيث دقتهما وثبات التقدير تحت ظروف مختلفة من الارتباط بين المشاهدات ¸ إن إجراء هذه المقارنة يُسهم في تحديد الطريقة الأنسب لتحليل هذا النوع من البيانات، مما ينعكس إيجاباً على جودة النماذج التفسيرية المستخدمة في الدراسات التطبيقية.

1.1 مشكلة البحث:

تُعدُّ البيانات الطولية من أكثر أنواع البيانات تعقيدًا في التحليل الإحصائي، لما تتضمنه من ترابط داخلي بين المشاهدات المتكررة للفرد الواحد عبر الزمن. إن تجاهل هذا الترابط عند استخدام النماذج التقليدية، مثل أنموذج انحدار بواسون العادي، قد يؤدي إلى تقديرات غير دقيقة أو متحيزة لمعلمات النموذج، مما ينعكس سلبًا على صحة الاستنتاجات الإحصائية.

وعلى الرغم من شيوع استخدام طريقة الإمكان الأعظم في تقدير معلمات أنموذج انحدار بواسون، إلا أن كفاءتها قد تتأثر بوجود الترابط بين المشاهدات الطولية. في المقابل، تُعد طريقة المعادلات المعممة من الأساليب الحديثة التي صُممت خصيصًا لمعالجة هذا النوع من البيانات، لما توفره من مرونة في التعامل مع الارتباط الداخلي دون الحاجة إلى افتراض توزيع كامل للمشاهدات.

من هنا تنبع مشكلة البحث الحالية في تحديد مدى كفاءة كل من طريقتي الإمكان الأعظم والمعادلات المعممة في تقدير معلمات أنموذج انحدار بواسون للبيانات الطولية، ولا سيما تحت ظروف مختلفة من حجم العينة ودرجة الترابط بين المشاهدات، الأمر الذي يستدعي إجراء مقارنة إحصائية منهجية بين الطريقتين.

1.2 أهداف البحث :

1.تقديم إطار نظري وتحليلي لنموذج انحدار بواسون للبيانات الطولية وبيان خصائصه الإحصائية وأهم استخداماته التطبيقية.

2. تقدير معلمات النموذج باستخدام طريقة الإمكان الأعظم(Maximum Likelihood Estimation - MLE) ، وتحليل دقتها وكفاءتها في ظل افتراضات الترابط المختلفة بين المشاهدات.

3. تطبيق طريقة المعادلات المعممة (Generalized Estimating Equations - GEE) لتقدير نفس النموذج، ودراسة قدرتها على التعامل مع الارتباط داخل الوحدات الطولية.

4. إجراء مقارنة كمية وإحصائية بين الطريقتين (MLEوGEE) بالاعتماد على مقاييس دقة مثل متوسط مربع الخطأ (MSE) ومتوسط الخطأ المطلق(MAE) ، لبيان أفضلية كل منهما في مواقف مختلفة.

2.أنموذج انحدار بواسون Poisson Regression Model:

ان أنموذج انحدار بواسون من اهم النماذج الخطية اللوغاريتمية، حيث انه يحصل على الاسم من اخذ اللوغاريتم الطبيعي للصيغة الاصلية للتوزيع ويتحول الى نموذج خطي وهي احد مميزات هذا التوزيع[15],[16]، وهو الاداة التي تستخدم في نمذجة المتغير المعتمد عندما تكون قيمه على شكل قيم قابلة للعد. وكما هو الحال في بقية نماذج الانحدار، قد يحتوي الأنموذج على الكثير من المتغيرات المستقلة مما يؤثر سلباً على دقة الأنموذج وسهولة تفسير نتائجه. كما يفترض انموذج الانحدار بان المتغير المعتمد y_i هو متغير استجابة يتبع توزيع بواسون بمعلمة قدرها (μ) ،وتتبع الاخطاء العشوائية في الانموذج توزيع بواسون وبمعلمة قدرها ايضاً (μ)[4] ،كما يعرف أنموذج انحدار بواسون بانه الطريقة التي تستخدم لغرض نمذجة المتغير المعتمد في حال توفر قيمه بشكل قيم معدودة او بشكل معدلات لأنه اداة جيدة في تحليل بيانات الاحداث النادرة تحت شرط الا تكون هذه البيانات سالبة[3] ، ويعتمد هذا الانموذج بالأساس على فقرتين:

الاولى تخص الافتراضات المتعلقة بالتوزيع ، حيث انه يختلف عن توزيع الاخطاء العشوائية في أنموذج الانحدار الخطي التقليدي .

الثانية تخص المعلمة وتميزها في توزيع بواسون ، والتي تمثل التوقع والتباين والتي تعتبر كدالة للمتغيرات التفسيرية.

ويعرف وفق الدالة الاحتمالية المعرفة بالصيغة ادناه[5]:

$$y = Exp(X\beta + U) \quad .....(1)$$

اذ ان :

y : موجه المتغير التابع ذي درجة (n×1)

X: مصفوفة المتغيرات المستقلة ذات الدرجة ((n×(p+1)

β: موجه المعلمات ذو الدرجة (1× (p +1))

U : موجه الأخطاء العشوائية ذي الدرجة (n×1)

n: حجم العينة

P : عدد المتغيرات المستقلة

يفترض الانموذج ان المتغير التابع يتبع توزيع بواسون ، اذ ان هذا التوزيع هو واحد من التوزيعات الاحتمالية المتقطع , في عام 1873 قام العالم الفرنسي (Simeon Poisson) باشتقاق الدالة الاحتمالية للتوزيع وسمي على اسمه .يصف هذا التوزيع احتمال حدوث المتغير العشوائي وان هذا المتغير يبين عدد مرات وقوع حدث معين خلال فترة زمنية على فرض ان Y_i متغير عشوائي يتوزع توزيع بواسون [4].

3.البيانات الطولية Longitudinal data:

المعلومات والبيانات التي يتم جمعها خلال فترات زمنية (حالات معينة متكررة) تسمى البيانات الطولية والتي تعرف بانها القياسات المأخوذة بشكل متكرر في كل قطاع خلال مدة من الزمن كان تكون اشهر او سنوات , وكذلك تعرف بانها قياسات متكررة لكل قطاع من القطاعات المشاهدة , اذ ان المشاهدات في القطاع الواحد تتميز بكونها مشاهدات مرتبطة , ويسمى احياناً بالبيانات المجدولة[13],[14] , تتعقب البيانات الطولية عينة معينة في نقاط مختلفة ، يمكن ان تتألف العينة من الخصائص الطبية و الفردية والاقتصادية والمعيشية ، تسمى البيانات الطولية المستعملة في الدراسات السريرية والبيئية *Environmental &Clinical* اما في البيانات الاقتصادية فتسمى البيانات اللوحية *Panel data*[4],[1] وقد تناولت بعض الدراسات العربية مشكلات تحليل البيانات المزدوجة والمتكررة في التطبيقات الاقتصادية، حيث قدّم عبد الحافظ (2012) تقديرًا لامعلميًا لهذا النوع من البيانات، مبيّنًا الحاجة إلى استخدام أساليب إحصائية ملائمة تراعي طبيعة الترابط بين المشاهدات[2].

4. الطرائق المستخدمة لتقدير الانموذج:

4.1 طريقة الإمكان الأعظم (Maximum Likelihood Estimation - MLE)

تُعد طريقة الامكان الاعظم (MLE) من ابرز الطرائق الاحصائية لتقدير معلمات النماذج ، وتستند الى مبدأ تعظيم دالة الامكان للحصول على قيم المعلمات التي تجعل البيانات المرصودة اكثر احتمالاً , ففي حالة أنموذج انحدار بواسون تُفرض البيانات على شكل قيم عددية تمثل عدد مرات وقوع حدث معين ، وتعد هذه الطريقة فعالة عند تحقيق افتراضات النموذج بدقة [6],[7] .

دالة الامكان لنموذج انحدار بواسون :

لنفترض ان $y1, y2, \ldots, yn$ تمثل المتغير التابع الذي يتبع توزيع بواسون بوسيط $\lambda_i = \exp(x_i^T\beta)$, فإن دالة الامكان تكون:

$$.....(2) L(\beta; y1, y2, \ldots, yn) = \prod_{i=1}^{n} \frac{e^{-\lambda_i}\lambda_i^{yi}}{yi!} \quad , \quad \lambda_i = \exp(x_i^T\beta)$$

تحويل دالة الامكان الى اللوغاريتم (دالة اللوغاريتم الإمكاني) :

للتبسيط نأخذ اللوغاريتم الطبيعي لدالة الامكان :

$$.....(3) \ell(\beta) = \sum_{i=1}^{n}[yi * (x_i^T\beta) - \exp(x_i^T\beta) - log(yi!)]$$

معادلة التقدير :

نشتق دالة اللوغاريتم الإمكاني بالنسبة الى β ، ونساوي المشتقة بالصفر للحصول على معادلة تقدير :

$$.....(4) \frac{\partial \ell(\beta)}{\partial \beta} = \sum_{i=1}^{n} x_i * (yi - \lambda_i) = 0$$

حيث :

$$.....(5) \lambda_i = \exp(x_i^T\beta)$$

استخدام خوارزمية نيوتن-رافسون:

لان المعادلة اعلاه غير خطية يُلجأ الى طرائق عددية لحلها ، مثل خوارزمية نيوتن-رافسون :

$$\beta^{(t+1)} = \beta^{(t)} - \left[H\left(\beta^{(t)}\right)\right]^{-1} S\left(\beta^{(t)}\right) \quad .....(6)$$

حيث :

$S(\beta)$: هو متجه المشتقة الاولى (Score Function).

$H(\beta)$: هو مصفوفة المشتقة الثانية (مصفوفة هيسيان Hessian) [8] , [1] ,[20].

$$H(\beta) = \sum_{i=1}^{n} x_i x_i^T * \lambda_i \quad .....(7)$$

4.2 طريقة المعادلات المعممة (Generalized Estimating Equations)

تعد طريقة المعادلات المعممة (GEE) من اهم الطرائق الاحصائية المستخدمة في تحليل البيانات الطولية او البيانات المترابطة ، وهي مناسبة عندما تتكرر القياسات على نفس المفردة بمرور الوقت ، كما هو في الدراسات الطبية الطولية.

قدمها Zeger و Liang في سنة 1986 لتكون امتداداً للأنموذج الخطي المعمم (GLM) ، حيث تسمح بتقدير معلمات النموذج مع الاخذ بعين الاعتبار الترابط الداخلي بين المشاهدات دون الحاجة الى نمذجة التوزيع الكامل للبيانات[10].

تعتمد GEE بناء معادلة تقديرية تربط بين المتوسط الشرطي للاستجابة والمتغيرات التفسيرية ، دون الحاجة الى تحديد الشكل الكامل لتوزيع البيانات , وتُستخدم بشكل خاص في الدراسات التي على قياسات متكررة لنفس الافراد مثل التجارب السريرية او المسوح الطبية [11].

النموذج العام لـGEE:

يفترض ان متغير الاستجابة y_i للوحدة i ذات البُعد T_i، وان:

- $\mu_i = E(y_i)$:المتوسط الشرطي .
- X_i: مصفوفة التصميم (تصميم المتغيرات التفسيرية).
- β: متجه المعلمات المطلوب تفسيره .
- $g(0)$: دالة الربط المناسبة (لبيانات بواسون نستخدم اللوغاريتم).

صيغة الربط :

$$g(\mu_i) = x_i\beta \Rightarrow \log(\mu_i) = x_i\beta \quad .....(11)$$

معادلة التقدير:

يتم تقدير β بحل المعادلة التالية :

$$U(\beta) = \sum_{i=1}^{n} D_i^T V_i^{-1}(y_i - \mu_i) = 0 \quad .....(12)$$

حيث :

$D_i = \frac{\partial \mu_i}{\partial \beta^T}$: هو مشتق المتوسط بالنسبة لـ(β).

V_i: هو مصفوفة التغاير للمتجه y_i، ويكتب: $V_i = A_i^{\frac{1}{2}} R(\alpha) A_i^{\frac{1}{2}}$ ، وان :

A_i: مصفوفة قطرية تحتوي على تباينات العناصر في y_i.

$R(\alpha)$: مصفوفة الارتباط (Working Correlation Matrix)، وتفرض بأحد الاشكال:

- مستقلة (Independence): لا يوجد ارتباط بين المشاهدات.
- تبادلية (Exchangeable): جميع الازواج لها نفس معامل الارتباط.

- انحدار ذاتي من الرتبة الاولى AR(1) : الارتباط يتناقص مع الزمن.

التوزيع التقاربي للمقدر :

وفقاً لنظرية المقدرات التكرارية (M-estimators) فإن المقدر الناتج من GEE يتمتع بالتقارب التالي :

$$\hat{\beta}_{GEE} \xrightarrow{d} N(\beta_0, B^{-1}MB^{-1}) \quad .....(13)$$

حيث:

$B = \sum_{i=1}^{n} D_i^T V_i^{-1} D_i$: مصفوفة المعلومات (يُعرف احياناً بمصفوفة شبه Fisher).

$M = \sum_{i=1}^{n} D_i^T V_i^{-1}(y_i - \mu_i)(y_i - \mu_i)^T V_i^{-1} D_i$: تمثل مصفوفة التغاير المحسوبة من البواقي .

هذا التقدير معروف باسم Sandwich Estimator او (مقدر الشطيرة) لأنه يتكوّن من ثلاث اجزاء ، يشبه في بنيته شكل الشطيرة:

$$\text{Sandwich Estimator} = B^{-1}MB^{-1} \quad .....(14)$$

ويُعبَّر عن هذا التوزيع بشكل موسع كما في الصيغة التالية [10],[11]:

$$\hat{\beta} \sim N\left(\beta\left(\sum_{i=1}^{N} D_i^T V_i^{-1} D_i\right)^{-1}\left(\sum_{i=1}^{N} D_i^T V_i^{-1}(Y_i - \mu_i)\,(Y_i - \mu_i)^T\right)\left(\sum_{i=1}^{N} D_i^T V_i^{-1} D_i\right)^{-1}\right) \quad .....(15)$$

وتعرف صيغة التغاير هذه باسم (Estimator Sandwich Variance) وتستخدم لتقدير التابين بطريقة مستقلة عن الشكل المفترض لمصفوفة الارتباط العامة $R(\alpha)$.

4.2.1 خصائص ومميزات طريقة المعادلات المعممة (GEE):

1. الاستقرار (*Robustness*): حتى في حالة الخطأ في تحديد شكل المصفوفة فهي تعطي مقدرات صحيحة للمعلمات .
2. عدم الحاجة لنمذجة التوزيع الكامل : لان طريقة *GEE* لا تفترض معرفة التوزيع المشترك للاستجابة .
3. التركيز على التقدير السكاني (*Averaged Population*): اي ان النتائج تعكس السلوك العام للسكان وليس لمفردة محددة وهذا ما يميزها عن النماذج ذات التاثيرات العشوائية التي تركز على الفرد[11],[12].

5. الجانب التطبيقي لطريقتي MLE و GEEلتقدير نموذج بواسون:

تم تطبيق طريقتي التقدير الإمكان الأعظم (MLE) والمعادلات التقديرية المعممة (GEE) على بيانات محاكاة لتقدير معلمات نموذج انحدار بواسون للبيانات الطولية، ويهدف هذا التطبيق إلى مقارنة أداء الطريقتين من حيث جودة التقدير.

5.1 طرائق المقارنة الاحصائية المستخدمة في تقييم كفاءة التقدير [9]:

لغرض مقارنة كفاءة طرائق التقدير المختلفة لأنموذج انحدار بواسون للبيانات الطولية، تم الاعتماد على مقاييس الخطأ الشائعة في الأدبيات الإحصائية، ولاسيما متوسط مربع الخطأ (MSE)ومتوسط الخطأ المطلق(MAE) ، لما لهما من قدرة على قياس دقة التقدير ومقارنة أداء النماذج الإحصائية بصورة موضوعية:

5.1.1 متوسط مربع الخطأ(Mean Squared Error - MSE) :

يُعد متوسط مربع الخطأ من أكثر المقاييس استخداماً في تقييم جودة التقدير، ويقيس متوسط مربعات الفروق بين القيم الحقيقية والقيم المقدّرة، ويُعطى بالعلاقة:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad .....(16)$$

حيث:

y_i: القيمة الفعلية للمشاهدة i.

$\hat{y}_i$: القيمة المقدرة للمشاهدة i.

n: عدد المشاهدات.

ويمتاز هذا المقياس بحساسيته العالية للأخطاء الكبيرة نتيجة تربيع الفروق، مما يجعله مناسباً في الحالات التي يكون فيها الانحراف الكبير عن القيم الحقيقية غير مقبول إحصائياً.

5.1.2 متوسط الخطأ المطلق(Mean Absolute Error - MAE) :

يمثل متوسط الخطأ المطلق مقياساً بديلاً لقياس دقة التقدير، ويعتمد على متوسط القيم المطلقة للفروق بين القيم الحقيقية والمقدّرة، ويُعطى بالعلاقة:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad .....(17)$$

ويتميّز MAE بكونه أقل تأثراً بالقيم المتطرفة مقارنةً بـMSE ، مما يجعله أكثر استقراراً
في حالة وجود مشاهدات شاذة.

5.2 اهمية MSE و MAE في تحليل البيانات الطولية وعدّ الاحداث:

يُستخدم كل من MSE و MAE بصورة تكاملية عند مقارنة أداء طرائق التقدير المختلفة مثل طريقة الإمكان الأعظم (MLE) وطريقة المعادلات المعممة(GEE) ، حيث يوفر MSE مؤشراً على حساسية الطريقة للأخطاء الكبيرة، بينما يعطي MAE تصوراً أوضح لمتوسط حجم الخطأ الفعلي.

5.3 مقارنة المقاييس:

- MSE اكثر حساسية للأخطاء الكبيرة ، ويوضح الانحراف الكلي للنموذج .
- MAE يعد اوضح تمثيلاً للخطأ المتوسط الفعلي دون المبالغة في الاخطاء الشاذة .284

6. النتائج المستخلصة من المحاكاة:

طريقة التقدير	MSE	MAE
MLE	0.024	0.106
GEE	0.037	0.127

6.1 النتائج والمناقشة

ولغرض تفسير النتائج المبينة في الجدول اعلاه ،أجريت تجربة المحاكاة باستخدام أسلوب مونت كارلو (Monte Carlo Simulation) لتقدير معلمات أنموذج انحدار بواسون للبيانات الطولية باستخدام طريقتين مختلفتين للتقدير، هما: طريقة الإمكان الأعظم (MLE) و طريقة المعادلات المعممة (GEE).

هدفت التجربة إلى مقارنة أداء الطريقتين من حيث دقة التقدير تحت ظروف مختلفة، شملت تغيّر حجم العينة (N=25, 50, 100, 150) وعدد التكرارات(r=3, 5) ، فضلاً عن اختلاف عدد المتغيرات التوضيحية بين مجموعتين:

• المجموعة الأولى: تضمنت p=5 متغيرات مستقلة.

• المجموعة الثانية: تضمنت p=11 متغيرًا مستقلاً.

تم اعتماد معياري متوسط مربع الخطأ (MSE) و متوسط الخطأ المطلق (MAE) للحكم على كفاءة الطريقتين، إذ تُعد الطريقة التي تعطي أقل قيم لهذين المعيارين هي الأكثر دقة في التقدير.

أظهرت النتائج أنّ قيمي MSE و MAEتقلّ تدريجياً مع ازدياد حجم العينة وعدد التكرارات في كلتا الطريقتين، مما يؤكد أن زيادة حجم البيانات تؤدي إلى تحسين دقة التقدير وتقليل التشتت في المعلمات المقدَّرة, كما تبيّن أن الطريقتين أعطتا تقديرات متقاربة جداً لمعلمات النموذج، مما يشير إلى توافقٍ عام في الأداء تحت الظروف التجريبية المختلفة.

ورغم هذا التقارب، أظهرت طريقة الإمكان الأعظم (MLE) تفوقاً طفيفاً في معظم الحالات، إذ كانت تمتلك قيماً أقل لـ MSE و MAE مقارنة بطريقة GEE، خصوصاً عند أحجام العينات الصغيرة (N=25, 50) وعدد التكرارات المنخفض (r=3) يُعزى هذا التفوق إلى أن طريقة الإمكان الأعظم تعتمد على دالة الاحتمال الكاملة، مما يجعلها أكثر كفاءة عندما تكون البيانات محدودة والعلاقات الترابطية بين المشاهدات ضعيفة نسبياً.

أما عند الأحجام الكبيرة من العينات وعدد التكرارات الأعلى(r=5) ، فقد اقترب أداء الطريقتين بشكل واضح، حيث أعطت طريقة GEE نتائج مُماثلة تقريباً لـ MLE من حيث دقة التقدير, ويُعزى ذلك إلى أن طريقة GEE تُعد أكثر ملاءمة للبيانات الطولية التي تتضمن ترابطاً داخلياً بين المشاهدات داخل الفرد الواحد، إذ تأخذ هذا الترابط بنظر الاعتبار في مصفوفة التغاير المفترضة(Working Correlation Matrix) ، مما يجعلها أكثر استقراراً في حالة العينات الكبيرة.

بشكل عام، تُشير نتائج المحاكاة إلى أن كلا الطريقتين تؤديان إلى تقديرات مستقرة ومتقاربة لمعاملات أنموذج بواسون الطولي، إلا أن:

• طريقة MLE تتفوّق من حيث الدقة الإحصائية في العينات الصغيرة أو في حال ضعف الترابط بين الملاحظات.

• بينما طريقة GEE تُظهر كفاءة أفضل عند التعامل مع العينات الأكبر ومع وجود ترابط واضح بين الملاحظات الطولية.

تؤكد هذه النتائج ما ورد في الأدبيات الإحصائية مثل دراسات Liang and Zeger(1986) و Hardin and Hilbe(2012) التي بيّنت أن أداء طريقة GEE يقترب من طريقة الإمكان الأعظم عند الأحجام الكبيرة، مع ميزة إضافية هي قدرتها على التعامل مع الارتباط الداخلي للبيانات المتكررة دون الحاجة إلى افتراض توزيع دقيق للمشاهدات.

7. الاستنتاجات والتوصيات :

في ضوء تجربة المحاكاة التي أُجريت لتقدير معلمات أنموذج انحدار بواسون للبيانات الطولية، وبالاعتماد على طريقتي التقدير الإمكان الأعظم (MLE) و المعادلات المعممة(GEE) ، يمكن استخلاص النقاط الآتية:

1. كفاءة الطريقتين:

أظهرت كل من طريقتي MLE و GEEقدرة عالية على تقدير معاملات أنموذج بواسون الطولي بدقة، إذ كانت التقديرات متقاربة جداً، مما يؤكد صلاحية الطريقتين للتطبيق على البيانات الطولية ذات الطبيعة العدّية.

2. تأثير حجم العينة وعدد التكرارات:

بيّنت النتائج أن دقة التقدير تتحسن بشكل ملحوظ مع ازدياد حجم العينة وعدد التكرارات، حيث تنخفض قيم MSE و MAEتدريجياً كلما ازدادت البيانات، وهو ما يتوافق مع القاعدة الإحصائية التي تنص على أن حجم العينة الأكبر يؤدي إلى مقدرات أكثر استقراراً وأقرب إلى القيم الحقيقية.

3. تفوق طريقة الإمكان الأعظم:(MLE)

تفوقت طريقة MLE تفوقاً طفيفاً في معظم التجارب، ولا سيّما في العينات الصغيرة وعدد التكرارات المحدود، ويُفسر ذلك باعتمادها الكامل على دالة الاحتمال، مما يمنحها كفاءة عالية في ظل البيانات المحدودة أو ضعف الترابط بين الملاحظات.

4. استقرار طريقة المعادلات المعممة:(GEE)

أظهرت طريقة GEE أداءً متقارباً جداً من MLE في العينات الكبيرة، مع ميزة إضافية تمثلت في استقرار التقديرات عند وجود ارتباط داخلي بين الملاحظات داخل كل وحدة طولية، بفضل اعتمادها على مصفوفة الارتباط العاملة.(Working Correlation Matrix)

8. الاستنتاج العام:

وبناءً على نتائج المحاكاة , تبين ان اختيار طريقة التقدير الملائمة يعتمد بشكل اساسي على حجم العينات وطبيعة الترابط الداخلي في البيانات الطولية

استناداً إلى النتائج السابقة وما أظهرته المقارنات الإحصائية بين الطريقتين، يمكن تقديم التوصيات الآتية:

1.اعتماد طريقة المعادلات المعممة (GEE) في الدراسات التي تتضمن بيانات طولية ذات ترابط زمني أو فردي واضح بين المشاهدات، لكونها تتعامل بمرونة مع الارتباط الداخلي دون الحاجة إلى افتراض توزيع معيّن للمشاهدات.

2. استخدام طريقة الإمكان الأعظم (MLE) عند التعامل مع عينات صغيرة أو بيانات قليلة التكرار، إذ توفر تقديرات أكثر دقة واستقراراً في مثل هذه الحالات.

3. التوسع في دراسات المحاكاة المستقبلية لتشمل نماذج عدّية أخرى مثل أنموذج بواسون ذي التشتت الزائد (Overdispersed Poisson) أو أنموذج الانحدار ذي الحدين السالب (Negative Binomial Regression) لمقارنة أداء طرائق التقدير المختلفة.

4. تطبيق الطرق على بيانات حقيقية (مثل بيانات المرضى أو التجارب الطبية الطولية) للتحقق من كفاءة الطريقتين في البيئات التطبيقية الواقعية ودراسة سلوك المقدّرات تحت ظروف عملية.

5. الاهتمام بتحليل مصفوفة الارتباط في GEE لاختيار البنية الأنسب (مثل تبادلية، مستقلة، أو تبعية زمنية)، لما لها من أثر مباشر في تحسين دقة التقدير وتقليل الخطأ.

تضارب المصالح
لا يوجد تضارب مصالح

الموافقة على النشر
"لقد قرأنا ووافقنا على النسخة النهائية من المخطوطة للنشر".

توافر البيانات والمواد
جميع البيانات قد تم تضمينها ضمن المخطوطة.

مساهمات المؤلفين
م. ق. تنفيذ الدراسة، وتحليل النتائج، وكتابة المقال

ن.ص. تصميم الفكرة

التمويل
لا يوجد تمويل

الشكر والتقدير
شكر لكل الأشخاص أو المؤسسات التي قدّمت لنا مساعدة علمية أو مادية في إنجاز الدراسة وكتابة المقال.

المراجع

[1] فدعم، انتصار عريبي، و الخفاجي، يوسف خليل. (2017). مقارنة نماذج الانحدار اللوجستي الشرطية مع التأثيرات الثابتة والمختلطة في حالة البيانات الطولية. كلية الإدارة والاقتصاد، جامعة بغداد.

[2] عبد الحافظ، علي سيف الدين. (2012). تقدير أنموذج لامعلمي للبيانات المزدوجة في بعض الأنشطة الاقتصادية في العراق (رسالة دكتوراه غير منشورة). كلية الإدارة والاقتصاد، جامعة بغداد.

[3]Cameron, A. C., & Trivedi, P. K. (2013). Regression analysis of count data (Vol. 53). Cambridge University Press.

[4]Rodriguez, G. (2007). Generalized linear models. Princeton University, Revised September 2007.

[5]McCullagh, P., & Nelder, J. A. (1989). Generalized linear models. Chapman and Hall.

[6]Tajuddin, R. M., Lamail, N., & Ibrahim, K. (2021). Comparison of estimation methods for one-inflated positive Poisson distribution. Journal of Taibah University for Science, 15(1), 1–12.

[7]Al Rahamneh, A. A., Al Zubeidi, M. A., Arabyat, Y. A., & Al Hawary, S. I. (2023). Forecasting the number of traffic accidents in Jordan using the Poisson regression model. Information Sciences Letters, 12(4), 1–10.

[8]Pardo-Fernández, J. C., & de Uña-Álvarez, J. (Eds.). (2008). ISNI 2008: Scientific programme and proceedings, Vigo, 5–7 November 2008.

[9]Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). Introduction to linear regression analysis (5th ed.). Wiley.

[10]Liang, K.-Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. Biometrika, 73(1), 13–22.

[11]Hardin, J. W., & Hilbe, J. M. (2012). Generalized estimating equations (2nd ed.). Chapman & Hall/CRC.

[12]Zorn, C. J. W. (2001). Generalized estimating equation models for correlated data. American Journal of Political Science, 45(2), 470–490.

[13]Winkelmann, R. (2008). Econometric analysis of count data. Springer.

[14]Rodriguez, G. (2007). Generalized linear models. Princeton University.

[15]Winkelmann, R. (2008). Econometric analysis of count data (5th ed.). Springer-Verlag Berlin Heidelberg.

[16]Long, J. S. (1997). Regression models for categorical and limited dependent variables (Vol. 7). Advanced Quantitative Techniques in the Social Sciences, pp. 219. Sage Publications.