مقارنة بعض الطرائق الحصينة لتقدير انموذج انحدار بواسون بأستخدام المحاكاة

الباحث محمد غازي جواد
مديرية تربية محافظة بغداد / الكرخ الثانية

أ.م اسيل عبد الرزاق رشيد
الجامعة المستنصرية / كلية الادارة والاقتصاد

تاريخ استلام البحث: 2023/03/01

تاريخ قبول البحوث: 2023/03/17

نشر البحث في العدد الواحد والعشرون: ديسمبر / كانون اول 2023

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

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

مقارنة بعض الطرائق الحصينة لتقدير انموذج انحدار بواسون بأستخدام المحاكاة

الباحث محمد غازي جواد **أ.م اسيل عبد الرزاق رشيد**
مديرية تربية محافظة بغداد / الكرخ الثانية **الجامعة المستنصرية / كلية الادارة والاقتصاد**

المستلخص
يعد انموذج انحدار بواسون احد نماذج العد المهمة حيث استخدم لنمذجة العديد من الظواهر الأقتصادية والصحية وحوادث المرور وغيرها. يتم في هذا البحث تقدير انموذج انحدار بواسون في حالة وجود القيم الشاذة, حيث تم اجراء دراسة محاكاة مونت كارلو, لمقارنة المقدرات الحصينة (M, S, MM) مع المقدر الاعتيادي، وهي طريقة الامكان الاعظم (MLE)، وقد اظهرت نتائج المحاكاة اعتمادا على معايير المفاضلة(MAE,MSE) ان طريقة (M) الحصينة اكثر كفاءة من باقي الطرائق الحصينة، وكذلك من الطريقة الاعتيادية للتقدير (MLE) .
الكلمات المفتاحية : انحدار بواسون، طريقة مقدرات M، طريقة مقدرات S، طريقة مقدرات MM

Comparison of some robust methods for estimating the Poisson regression model using simulation

Abstract
The Poisson regression model is one of the important counting models, as it was used to model many economic and health phenomena, traffic accidents, and others. In this research, the Poisson regression model is estimated in the case of the presence of outliers, where a Monte Carlo simulation study was conducted, to compare the robust estimators (M, S, MM) with the usual estimator, which is the method of Maximum Likelihood Estimator (MLE), and the results of the simulation were shown depending on comparison criteria (MSE, MAE) The robust (M) method is more efficient than the rest of the robust methods, as well as the conventional method of estimation (MLE).
Keywords: Poisson regression, M estimators method, S estimators method, MM estimators method.

المقدمة

يعتبر انموذج انحدار بواسون احد نماذج الانحدار غير الخطية، وكذلك احد اهم نماذج الانحدار اللوغارتمية الخطية وهو الاداة التي يتم من خلالها نمذجة المتغير المعتمد عندما تكون قيم ذلك المتغير على شكل قيم قابلة للعد . اذ يفترض هذا الانموذج ان متغير الاستجابة يتبع توزيع بواسون بمتوسط وتباين متساويان .وفي حالة وجود القيم الشاذة في البيانات يصبح من الصعوبة الحصول على نتائج دقيقة لتقدير معالم الانموذج بالطرائق الاعتيادية لذلك سوف نلجأ الى الطرائق الحصينة للتقدير وهي (طريقة مقدرات M ، طريقة مقدرات MM ، طريقة مقدرات S) ونجري مقارنة بينها اعتمادا على معايير المقارنة لنعرف ماهي الطريقة الافضل لتقدير معلمات انموذج انحدار بواسون.

انموذج انحدار بواسون [9][10]

يعرف انموذج انحدار بواسون بأنه احد انواع نماذج الانحداراللاخطية والمستخدم في نمذجة البيانات المعدودة ويمكن كتابة انموذج انحدار بواسون على النحو التالي:

$$\begin{aligned} Y_i &= \mathrm{E}(Y_i) + u_i \\ &= \lambda_i + u_i \end{aligned} \quad .... (4)$$

حيث ان $i = 1,2, \dots n$

وان متغير الاستجابة Yi يتبع توزيع بواسون بمتوسط وتباين يساوي (λi)

$$\lambda_i = E(Y_i|X_i) = e^{\underline{X}_i \underline{B}} \quad .... (5)$$

وعلى وفق هذه الصيغة فان المتغيرات التوضيحية تقترح اي قيمة حقيقية لمتوسط متغير الاستجابة (بمعنى ان المتوسط اما يكون موجب او سالب) ، وهذا يتعارض مع خاصية معلمة توزيع بواسون وهي ان قيمة λ موجبة فقط ، لذلك تم اقتراح نمذجة علاقة خطية لوغارتم المتوسط (Log λ) والمتغيرات التوضيحية (Xi) اي اعتماد النموذج الخطي اللوغارتمي للتخلص من القيم السالبة لمتوسط متغير الاستجابة (Y)

$$\log (E(Y_i)) = \mathrm{Log}\ (\lambda_i) = \underline{X_i}'\underline{B} \quad ....(6)$$

حيث ان :

$\underline{B}$: يمثل متجه المعلمات ذو الدرجة 1*(K+1) المطلوب تقديرها

$\underline{X_i}'$: يمثل متجه المتغيرات التوضيحية ذو الدرجة 1*(k+1)

اذ تقوم الفكرة الاساسية على ربط الدالة الاحتمالية لمتغير الاستجابة (Yi) حيث (i=1,2,....,n) و n هي عدد مشاهدات العينة بمتجه $\underline{X}'$ الذي هو عبارة عن مجموعة من المتغيرات التوضيحية (Xj) حيث (j=0,1,2,...,K) وبالاعتماد على هذا الانموذج فان معاملات الانحدار Bj تمثل التغير المتوقع في لوغارتم المتوسط لمتغير الاستجابة Y لكل وحدة تغير حدثت في المتغير التوضيحي Xj .

وبأخذ exp لطرفي العلاقة رقم (6) نحصل

$$\lambda_i = \exp\ (\underline{X_i}'\underline{B}) \quad ....(7)$$

وبالتالي فأن معامل الانحدار الاسي exponentionl coefficient regression يمثل التاثير المضاعف للمتغير التوضيحي Xj على متوسط الاستجابة Y حيث ان انموذج انحدار بواسون يستند على ثلاث فرضيات :

الافتراض الاول : متغير الاستجابة Y يتوزع وفق توزيع بواسون بمعلمة λ وان (λ>0) وحسب العلاقة (1) .

الافتراض الثاني : المعلمة λ والتي تمثل متوسط قيم متغير الاستجابة Y وتمثل التباين ايضا يمكن ان يعبر عنها استنادا الى متجه المتغيرات التوضيحة وكما مبين بالصيغة (7) .

الافتراض الثالث : ان الازواج المرتبة للمشاهدات (Xi,Yi) لها توزيع مستقل .

اجمالا يمكن دمج الافتراضين الاول والثاني الحصول على الدالة الاحتمالية الشرطية التالية :

$$f(yi,xi) = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!} = \frac{e^{-e^{(\underline{x}_i{}'\underline{B})}}e^{y_i\underline{x}_i{}'\underline{B}}}{y_i!} \qquad y=0,1,2,...... \qquad ....(8)$$

ان انموذج انحدار بواسون والذي يمتلك الخصائص التي ذكرناها سابقا يحقق دالة الوسط الحسابي الشرطية الاسية (الخطية اللوغارتمية) لان الانموذج يتحول بأخذ اللوغارتم الطبيعي الى انموذج خطي .

$$E(Y/X) = \lambda = \exp(\underline{x}_i{}'\underline{B}) \qquad ......(9)$$

وايضا دالة التباين الشرطية الاسية (الخطية اللوغارتمية)

$$Var(Y/X) = \lambda = \exp(\underline{x}_i{}'\underline{B}) \qquad ....(10)$$

وذكرنا سابقا عند مناقشة توزيع بواسون خاصية تساوي الوسط الحسابي والتباين، وهنا ايضا تنطبق هذه الخاصية في انموذج انحدار بواسون (التوزيع المتساوي) .

وعليه فأن الصيغة الخطية ل λ_i تصبح $\text{Log } \lambda_i = {}'\underline{x}_i\ \underline{B}$

طرائق التقدير الاعتيادية

يوجد الكثير من الطرق الاعتيادية (الكلاسيكية) للتقدير وابرزها والاكثر استخداما طريقة المربعات الصغرى الاعتيادية (OLS) وطريقة الامكان الاعظم (MLE) .

وسوف يتم التطرق الى طريقة الامكان الاعظم (MLE) بالتفصيل لتقدير انموذج انحدار بواسون والتي تهدف الى تعظيم الدالة الاحتمالية وجعل دالة الامكان في نهايتها العظمى ، حيث تتصف مقدرات هذه الطريقة بكونها لاتمتلك خصائص مرغوبة في العينات الصغيرة ، بينما تمتلك هذه المقدرات في العينات الكبيرة خصائص مرغوبة اضافة الى الاتساق [1] .

طريقة الامكان الاعظم [9]

نأخذ عينة بحجم n من ازواج المشاهدات (yi,xi) المستقلة فان دالة الامكان سوف تمثل حاصل ضرب دوال الكتلة الاحتمالية الشرطية لكل مشاهدة (yi) فتكون :

$$L(B; y_1, y_2, ..., y_n, x_1, x_2, ..., x_n) = \prod_{i=1}^{n} f(yi/xi\ ; B) \qquad ....(11)$$

وان مقدر الامكان الاعظم هو المقدر الذي يجعل دالة الامكان اكبر مايمكن

وهنا سوف نعوض باللوغارتم في دالة الامكان لتسهيل العمليات الحسابية

$$L(B;Y,X) = \text{Log} \prod_{i=1}^{n} f(y_i/x_i\ ; B) \qquad ....(12)$$

$$L(B;Y,X) = \sum_{i=1}^{n} Log\ f(y_i/x_i; B) \qquad ....(13)$$

$$L(B;Y,X) = \sum_{i=1}^{n}\left[-\exp\left(\underline{x}_i{}'\underline{B}\right) + y_i\underline{x}_i{}'\underline{B} - Log(y_i!)\right] \qquad ....(14)$$

الان نشتق لوغارتم دالة الامكان من خلال اشتقاق متجه المعلمات $\underline{B}$ وبالتالي سوف نحصل على ($k+1$) من المشتقات :

$$A_n = \frac{dL(\underline{B};\underline{Y},\underline{X})}{d\underline{B}} = \sum_{i=1}^{n}[-\exp(\underline{x_i}'\underline{B})\,\underline{x_i} + y_i\underline{x_i}'] \quad ..... (15)$$

$$A_n = \frac{dL(\underline{B};\underline{Y},\underline{X})}{d\underline{B}} = \sum_{i=1}^{n}[y_i - \exp(\underline{x_i}'\underline{B})]\underline{x_i} \quad ...... (16)$$

ثم نساوي المشتقة الاولى بالصفر

$$A_n = \frac{dL(\underline{B};\underline{Y},\underline{X})}{d\underline{B}} = 0$$

$$\sum_{i=1}^{n}[y_i - \exp(\underline{x_i}'\underline{B})]\underline{x_i} = 0 \quad .... (17)$$

واذا اردنا التاكد من صحة الاشتقاق نشتق المعادلة (15) مرة ثانية فنحصل :

$$T_n = \frac{d^2L(\underline{B};\underline{Y},\underline{X})}{d\underline{B}\,d\underline{B}'} = -\sum_{i=1}^{n}\exp(\underline{x_i}'\underline{B})\underline{x_i}\underline{x_i}' \quad .... (18)$$

نلاحظ ان المشتقة الثانية تكون سالبة معناها اشتقاقنا صحيح . ان معلمات المعادلة (17) غير خطية وتحل باستخدام احدى الطرق التكرارية ، وسوف نستخدم طريقة نيوتن رافسون حيث نقوم بأعطاء قيم تقديرية اولية لمعالم الانموذج $\hat{B}^0$ بحيث نحصل على تقريب من الدرجة الثانية في لوغارتم دالة الامكان L(B) حول القيم الاولية $\hat{B}^0$ وكالتالي :-

$$L^*(B) = L(\hat{B}^0) + A_n(\hat{B}^0)'(B-\hat{B}^0) + \frac{1}{2}(B-\hat{B}^0)'T_n(\hat{B}^0)(B-\hat{B}^0) \approx L(B) \quad ...(19)$$

وبتعظيم L^*(B) بدلا من L(B) بالنسبة للمعالم فنحصل على تقدير جديد لها وليكن $\hat{B}^1$ والقيد الضروري لتعظيم المعادلة يكون كالتالي :-

$$A_n(\hat{B}^0) + T_n(\hat{B}^0)(\hat{B}^1 - \hat{B}^0) = 0 \quad ......(20)$$

ثم نحصل على $\hat{B}^1$

$$\hat{B}^1 = \hat{B}^0 - [T_n(\hat{B}^0)]^{-1}A_n(\hat{B}^0) \quad ......(21)$$

اذن صيغة نيوتن رافسون بأعتبار $\hat{B}_0$ قيمة تقديرية اولية يكون كالتالي :-

$$\hat{B}^{k+1} = \hat{B}^k - [T_n(\hat{B}^k)]^{-1}A_n(\hat{B}^k) \quad .... (22)$$

حيث k=0,1,2,.....

وعندما يكون الفرق بين $[\hat{B}^{k+1} - \hat{B}^k]$ قريبة جدا من الصفر (مثلا عند النقطة 0.00001) نتوقف عن تكرار التقدير .

مشكلة وجود القيم الشاذة في البيانات [2][3][6]

ان مشكلة وجود القيم الشاذة في البيانات قد افقد طرائق التقدير الأعتيادية اهميتها وبالتالي اصبحت غير دقيقة في تقدير المعلمات اي اننا نحصل من خلالها على معلمات ذات خصائص غير جيدة بسبب عدم توفر اهم شرط لتطبيق هذه الطرائق وهو(عدم وجود القيم الشاذة في البيانات) .

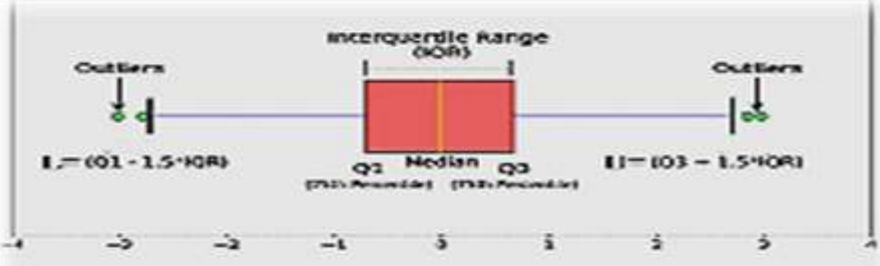
ويمكننا ان نعرف القيم الشاذة outliers بأنها القيم التي تشذ عن بقية المشاهدات ولاتنسجم معها فعند ترتيب البيانات بشكل تصاعدي تكون موقع القيم الشاذة اسفل او اعلى الترتيب وبعيدة عنه ايضا , حيث تتبع القيم الشاذة توزيع غير التوزيع الاصلي للبيانات , وتنتج القيم الشاذة من اخطاء القياس او اخطاء التسجيل او اخطاء المعاينة او تكون نتيجة حدث غير عادي مثل الحروب او الكوارث الطبيعية وغيرها .

يكون وجود القيم الشاذة في البيانات بثلاثة انواع
النوع الأول : وجود القيم الشاذة في مشاهدات متغير الأستجابة y
النوع الثاني : وجود القيم الشاذة في مشاهدات المتغيرات التوضيحية x_j
النوع الثالث : وجود القيم الشاذة في مشاهدات متغير الأستجابة y والمتغيرات التوضيحية x_j
هناك عدة طرائق للكشف عن وجود القيم الشاذة في البيانات مثل طريقة الرسم الصندوقي والقطع المخططة ، وطريقة مصفوفة H ، وطريقة مسافات مهلنوبس التربيعية العادية ، وطريقة مسافات مهلنوبس التربيعية الحصينة وغيرها .
وسوف نكتفي هنا بالتطرق الى الطريقة الأولى للكشف عن وجود القيم الشاذة في البيانات

طريقة الرسم الصندوقي والقطع المخططة [3]
هي اكثر الطرائق شيوعا وسهولة في الاستخدام وتعتمد على الربيعات حيث ان الربيع الثاني او نصف المدى الربيعي الذي يكشف 50% من انحرافات المشاهدات ونحدد الربيع الأدنى (Q_1) ونحدد الربيع الأعلى (Q_3) ونحدد اقل قيمة غير شاذة (L) ونحدد اكبر قيمة غير شاذة (U) وبالتالي اي قيمة تزيد عن U او تقل عن L نعتبرها قيمة شاذة .
نحدد قيمة L ، U من خلال المعادلات الأتية :

$$\mathrm{L} = Q_1 - 1.5(Q_3 - Q_1) = Q_1 - 1.5 * IQR$$
$$\mathrm{U} = Q_3 + 1.5(Q_3 - Q_1) = Q_3 + 1.5 * IQR$$

الخطوط العمودية الثلاثة للمربع تدل على الربيعات الثلاثة والخطوط المتباعدة تجاه اليمين واليسار تسمى بالشوارب معنى ذلك ان اي قيمة خارج هذه الشوارب تعتبر قيمة شاذة كما موضح بالشكل (2-1) .



شكل (2– 1) القيم الشاذة والرسم الصندوقي

بالتالي لتلافي مشكلة القيم الشاذة في البيانات وحتى لاتؤثر على عملية تقدير المعلمات سوف نستخدم التقديرات الحصينة للحصول على مقدرات ذات خصائص جيدة .

طرائق التقدير الحصينة
اول من اطلق مصطلح المقدرات الحصينة هو الباحث Box في عام 1953 ثم توالت الدراسات بخصوصها من قبل عدة باحثين مثلا الباحث Johan Tukey في الفترة (1960-1962) تلاه الباحث Peter Huber في الفترة (1964-1967) ثم الباحث Frank Hampel في الفترة

(1971-1974) حيث اقترحوا عدة اساليب للأحصاءات الحصينة من خلال اصدار عدد من البحوث والدراسات في هذا المجال .[8]
ان كلمة الحصانة تطلق على المقدرات التي لاتتأثر او تتحسن بسهولة لوجود مخالفة في احدى فرضيات التوزيع الطبيعي ، وان الهدف الاساسي من ايجاد طرائق التقدير الحصين هو لتقليل تأثير القيم الشاذة على المقدر.[2]
ان اتباع الطرق الكلاسيكية في تقدير المعالم سواء كان الانموذج خطيا او غير خطي ليس اكثر امانا من تطبيق الطرق الحصينة وذلك بسبب ان الفروض الواجب توفرها لتطبيق الطرق الكلاسيكية ليست بالسهلة مثل عدم وجود القيم الشاذة او اتباع الخطا العشوائي توزيعا غير التوزيع الذي يناسب الطريقة المعتمدة في التقدير .[5]
ان الهدف من الاحصاءات الحصينة هو تحسين الاجراءات التي يمكن الوثوق بها والتي تكون كفوءة في ظل وجود القيم الشاذة في البيانات، ان الاحصاءات الحصينة هي امتداد للاحصاءات المعلمية مع الاخذ بنظر الاعتبار ان النماذج المعلمية هي افضل وصف للواقع.[8]
يمكن استخدام نقطة الانهيار[6] كمقياس لحصانة المقدرات والتي تمثل الحد الذي يصف مقدار مقاومة المقدر للبيانات الشاذة (الملوثة) قبل ان يصبح بدون فائدة ، ويعد المقدر مقاوم اذا كانت نقطة انهياره اكبر من الصفر ، ويتصف المقدر بأنه الاكثر حصانة اذا كان يمتلك نقطة انهيار عالية، إذ ان اعلى نقطة انهيار لاتتجاوز 0.50 ، واذا تجاوزت يصبح من المستحيل التمييز بين الجزء الجيد وغير الجيد (الملوث) من العينة، وبالتالي من غيرالممكن التمييز بين التوزيع الاساس والتوزيع الملوث . ان ابسط مثال على المقدر غير الحصين (ذو نقطة الانهيار الواطئة) هو الوسط الحسابي حيث ان مشاهدة واحدة شاذة تجعله ينهار. بينما ابسط مثال على المقدر الحصين (ذو نقطة الانهيار العالية) هو الوسيط لانه لايعتمد على البيانات الشاذة. ويمكن ان نُعرَف المقدر الحصين هو المقدر الذي يحافظ على الخصائص المرغوبة في انموذج الانحدارعند انتهاك بعض فرضياته . وفي هذا البحث سوف نستخدم ثلاث طرق حصينة لتقدير انموذج انحدار بواسون وكالآتي :

طريقة مقدرات M [3]

هي مقدر انحدار حصين امتداد لطريقة (MLE) وتستعمل على نطاق واسع وتسمى احيانا تقديرات (Huber) وتعتبر تقديرات قوية ضد القيم الشاذة في المتجه المعتمد (Y) ولكنها ليست قوية ضد القيم الشاذة في المتجه (X) وتكون مقدرات هذه الطريقة غير متحيزة وتمتلك اقل تباين ويكون هدف تقدير M هو تقليل البواقي للاخطاء .

$\tilde{B}_{M\,=}\ \min p(y_i - \exp(\acute{X}B)$ .... (23)

وبتقليل دالة البواقي الى اقل ما يمكن

$\hat{B}_M = \min \sum_{i=1}^{n} P(u_i) = \min \sum_{i=1}^{n} P\left(\frac{e_i}{\hat{\sigma}}\right) = min \sum_{i=1}^{n} P\left[\frac{y_i - \exp(\acute{X}B)}{\hat{\sigma}}\right]$ ....(24)

إذ ان e_i : تمثل البواقي

حيث يقدر التباين وفق الصيغة الاتية :

$\hat{\sigma} = \frac{MAD}{0.6745}$ .... (25)

MAD = median $|e_i - median(e_i)|$ .... (26)

إذ ان MAD : تمثل الوسيط لمطلق الفروقات بين البواقي ووسيطها

عند اشتقاق المعادلة (2-24) بالنسبة الى B نحصل :

$$\sum_{i=1}^{n} x_i \varphi[\frac{y_i-\exp(XB)}{\hat{\sigma}_M}] = 0 \qquad .... (27)$$

إذ ان $\varphi(.)$: تمثل مشتقة دالة P(.) وتسمى ايضا بدالة التأثير Influence function ويتم حساب الاوزان للقيم

$$W_i = \frac{\varphi[\frac{y_i-\exp(XB)}{\hat{\sigma}_M}]}{\frac{y_i-\exp(XB)}{\hat{\sigma}_M}} \qquad .... (28)$$

وبالتالي يمكن صياغة المعادلة المشار إليها بالاعتماد على دالة توكي وكما يلي :

$$W_i = \begin{cases} \left[1-\left(\frac{u_i}{c}\right)^2\right]^2 & , |u_i| \le c \\ 0 & , |u_i| > c \end{cases} \qquad .... (29)$$

حيث ان قيمة (c=4.685) حسب دالة Tukey

Where $u_i = \frac{e_i}{\hat{\sigma}}$

$$\sum_{i=1}^{n} x_i W_i (y_i - \exp(XB)) = 0 \qquad .... (30)$$

2-5 طريقة مقدرات S [3]

ان هذه الطريقة تعتمد على تقدير مقياس الاخطاء حيث نقوم بتقليل مجموع الاخطاء الى ادنى حد ممكن وهذه الطريقة قوية للغاية ضد القيم الشاذة الموجودة في بيانات المتجه X ، وضعيفة ضد القيم الشاذة في بيانات المتجه Y ، ويتم التقدير بالأعتماد على بواقي طريقة M وكالأتي :

$$\hat{B} = \min \hat{\sigma}_s \left(e_1(B), e_2(B), \dots\dots , e_n(B)\right) \qquad .... (31)$$

وبأستخدام الحد الادنى من مقدار التباين الحصين $\hat{\sigma}_s$

$$min_{B_I} \sum_{i=1}^{n} p[\frac{y_i-\exp(\acute{X}B)}{\hat{\sigma}_s}] \qquad .... (32)$$

$$\hat{\sigma}_s = \sqrt{\frac{\sum_{i=1}^{n} W_i e_i^2}{nk}} \qquad .... (33)$$

عندما

K = 0.199 , $W_i = \frac{p(u_i)}{u_i^2}$

إذ ان : e_i تمثل البواقي

ويكون التقدير الاولي لمقدر التباين كالاتي :

$$\hat{\sigma}_s = \frac{median\ |e_i - median(e_i)|}{0.6745} = \frac{MAD}{0.6745} \qquad .... (34)$$

إذ ان MAD : تمثل الوسيط لمطلق الفروقات بين البواقي ووسيطها ثم نشتق المعادلة (32) بالنسبة الى B فيكون الناتج كالاتي :

$$\sum_{i=1}^{n} x_i\ \varphi(\frac{y_i-\exp(XB)}{\hat{\sigma}_s}) = 0 \qquad .... (35)$$

$$\varphi(u_i) = P(u_i) = \begin{cases} u_i\left[1-\left(\frac{u_i}{c}\right)^2\right]^2 & , |u_i| \le c \\ 0 & , |u_i| > c \end{cases} \qquad .... (36)$$

وتكون دالة الاوزان حسب صيغة Tukey كالتالي : where c = 1.547

$$W_i = \begin{cases} [1-\left(\frac{u_i}{c}\right)^2]^2 & , |u_i| \le c \\ 0 & , |u_i| > c \end{cases} \qquad .... (37)$$

طريقة مقدرات MM [4]

وهي احدى الطرق الحصينة لتقدير معلمات انموذج انحدار بواسون والتي تعد تعديلا لطريقة M . والمقترحة من قبل الباحث (Yohai) عام 1987 .

$$\sum_{i=1}^{n} p(u_i) x_i = 0 \qquad .... (38)$$

$$\sum_{i=1}^{n} p\left(\frac{y_i - \exp(XB)}{S_{MM}}\right) x_i = 0 \qquad .... (39)$$

حيث S_{MM} هو الانحراف المعياري لبواقي طريقة (S) الذي يتم الحصول عليه من بقايا تقدير S و P هي دالة Tukey الموزونة .

$$\hat{\sigma}_{\mathrm{MM}} = \begin{cases} MAD & iteration = 1 \\ \frac{\sum_{i=1}^{n} W_i e_i^2}{nk} & iteration > 1 \end{cases} \qquad .... (40)$$

إذ ان e_i تمثل البواقي

Where MAD = median $|e_i - median(e_i)|$ .... (41)

إذ ان MAD : تمثل الوسيط لمطلق الفروقات بين البواقي ووسيطها

$$u_i = \frac{e_i}{\hat{\sigma}_{\mathrm{MM}}} \qquad .... (42)$$

$$W_i = \begin{cases} (1 - \left(\frac{u_i}{c}\right)^2)^2 & |u_i| \le \mathrm{c} \\ 0 & |u_i| > \mathrm{c} \end{cases} \qquad .... (43)$$

حيث c = 4.685

$$\mathrm{p}(u_i) = \begin{cases} \frac{u_i^2}{2} - \frac{u_i^4}{2c^2} + \frac{u_i^6}{6c^2} & u_i \le |c| \\ \frac{c^2}{6} & u_i > |c| \end{cases} \qquad .... (44)$$

معايير المقارنة

حتى يتم اختيار افضل طريقة لتقدير معالم انموذج بواسون وبالتالي نحصل على انموذج ذو معنوية عالية ، سوف نعتمد على المعايير الاتية :

متوسط مربعات الخطأ (MSE) [11]

يعد هذا المعيار من معايير المفاضلة المهمة في نماذج الانحدار ويحسب كالاتي :

$$MSE = \frac{1}{p} \sum_{i=1}^{p} (\hat{B}_i - B)^2 \qquad .... (45)$$

حيث ان :

p : تمثل عدد التجارب

$\hat{B}_i$: تمثل قيمة المعلمة i التقديرية

B : تمثل قيمة المعلمة الحقيقية

وهناك علاقة عكسية بين MSE ومعنوية الانموذج حيث كلما قل MSE زادت معنوية الانموذج والعكس بالعكس ايضا .

متوسط مطلق الخطأ (MAE) [11]

يشير هذا المعيار الى متوسط مدى قرب القيمة التقديرية الى القيمة الحقيقية ويستخدم للأشارة الى مدى دقة التقدير ايضا ويحسب كالأتي :

$$MAE = \frac{1}{p} \sum_{i=1}^{p} |\hat{B}_i - B| \qquad .... (46)$$

حيث p : تمثل عدد التجارب

$\hat{B}_i$: تمثل قيمة المعلمة i التقديرية ، B : تمثل قيمة المعلمة الحقيقية

وهناك علاقة عكسية بين MAE ومعنوية الانموذج ، حيث كلما قل MAE زادت معنوية الانموذج والعكس بالعكس ايضا .

مفهوم المحاكاة [6][7]

تعد المحاكاة من الاساليب المهمة التي من خلالها يمكن تقليد الواقع العملي ذلك عند تكوين مجتمع نظري يشابه المجتمع الحقيقي حيث نضع برنامج على الحاسبة الالكترونية يكوَن عدد من النماذج الاحصائية لتوليد البيانات ، ثم تكرر التجربة التي تسمح للباحث في كل مرة بتغيير المدخلات مثل حجم العينة وقيمة المعلمة المولدة ، وبالتالي يمكن استعمال اسلوب المحاكاة بديلا عن عدم توفر البيانات المطلوبة او صعوبة الحصول عليها او صعوبة تحليل بعض النظريات واشتقاقها .

هناك عدة طرائق للمحاكاة وهي الطريقة التناظرية والطريقة المختلطة وطريقة مونتي كارلو التي تعتبر من اهم واكثر الطرائق شيوعا كونها تمتاز بالمرونة من حيث تكرار العملية مرات عديدة للوصول الى النتائج المرجوة وتمتاز ايضا بالعشوائية حيث ان مجموعة البيانات العشوائية في التجربة الاولى مستقلة عن مجموعة البيانات العشوائية في التجربة الثانية وهكذا.

مراحل تجربة المحاكاة

تم استعمال لغة البرمجة ماتلاب 2020 لكتابة برنامج المحاكاة والذي تم إرفاقه في الملحق (A)، ويتضمن البرنامج المكتوب أربعة مراحل أساسية لتقدير أنموذج الانحدار، وكما يأتي:

المرحلة الأولى: مرحلة تحديد القيم الافتراضية

إذ يتم في هذه المرحلة اختيار القيم الافتراضية للمعلمات باعتبار وجود متغيرين توضيحيين، وقد تم اختيار القيم كما يأتي:

اختيرت القيم الافتراضية المختلفة للمعلمات وهذه القيم موضحة في الجدول (1).

جدول (1): القيم الافتراضية للمعلمات

Model	β_0	β_1	β_2
I	2	0.5	-1
II	0	1	1
III	0.5	-1	1.5
IV	2	0.9	0.8
V	3	2.9	2.5
VI	10	5	6

1. تم اختيار ثلاثة أحجام مختلفة للعينات (25، 50، 100).
2. اختيرت نسب مختلفة للتلوث (10%، 15%، 20%)
3. تم تكرار كل تجربة 1000 مرة.

المرحلة الثانية: توليد البيانات

وهي مرحلة مهمة جداً لاعتماد الخطوات التي تليها عليها، إذ يتم فيها توليد المتغيرات التوضيحية باعتبارها مولدة من التوزيع المنتظم ومن ثم توليد المتغير التابع من توزيع بواسون.

المرحلة الثالثة: التقدير

يتم في هذه المرحلة إجراء عملية التقدير لمعلمات الانحدار ومن ثم تقدير أنموذج الانحدار وذلك باستعمال طرائق التقدير الواردة في الفصل الثاني، وكما يأتي:

1. طريقة الإمكان الأعظم MLE.
2. طريقة مقدرات M.
3. طريقة مقدرات S.
4. طريقة مقدرات MM.

المرحلة الرابعة: مرحلة المقارنة بين الطرائق

لغرض المقارنة بين طرائق التقدير المختلفة للمعلمات وإيجاد أفضل المقدرات تم استعمال معيار MSE وMAE حيث إن الطريقة التي تمتلك أقل قيمة MSE وأقل قيمة MAE تعتبر أفضل.

نتائج عمليات المحاكاة

لغرض تطبيق طرائق التقدير لأنموذج الانحدار وتحديد الطريقة الأفضل، والتي ستستعمل في تقدير الأنموذج للبيانات الحقيقية في الجانب التطبيقي مستعملين لذلك برنامج تمت كتابته بلغة ماتلاب ، وسيتم عرض النتائج التي تمثل قيم MSE وMAE لكل طريقة وفقاً لأحجام العينات والقيم المختلفة للمعلمات ونسبة التلوث وكما يأتي:

جدول (2): قيم MSE وMAE للطرائق المختلفة وللحالة الأولى (I)

PR [1]	Sample Size	Criteria	MLE	M	S	MM
10%	25	MSE	1.39608	0.93375	1.39256	1.14597
		MAE	1.68495	1.41146	1.6794	1.53438
	50	MSE	1.38106	0.64816	1.38217	1.03164
		MAE	1.66917	1.23404	1.67826	1.45804
	100	MSE	1.26681	0.49818	1.26530	0.98240
		MAE	1.52357	1.08589	1.52906	1.34659
15%	25	MSE	1.50028	1.04742	1.49599	1.38257
		MAE	1.74171	1.48303	1.74031	1.68178
	50	MSE	1.47195	0.85684	1.47007	1.33238
		MAE	1.7359	1.37523	1.73488	1.64577
	100	MSE	1.45851	0.65442	1.45963	1.29765
		MAE	1.71882	1.2758	1.7277	1.5233
20%	25	MSE	1.52413	1.15095	1.52627	1.46185
		MAE	1.74557	1.54687	1.7564	1.72727
	50	MSE	1.50953	1.0033	1.50619	1.36514
		MAE	1.75717	1.46147	1.74971	1.6609
	100	MSE	1.50169	0.81627	1.50134	1.28341
		MAE	1.74983	1.36588	1.73999	1.64722

[1]. PR هو اختصار للكلمتين Pollution Rate والتي تعني نسبة التلوث والتي بدورها تعني نسبة وجود الشواذ في متغير الاستجابة وكذلك في المتغيرين التفسيريين.

جدول (3): قيم MSE وMAE للطرائق المختلفة وللحالة الثانية (II)

PR	Sample Size	Criteria	MLE	M	S	MM
10%	25	MSE	2.3441	2.14388	2.2826	2.26486
		MAE	2.53997	2.39721	2.50445	2.48419
	50	MSE	2.27399	2.06369	2.20369	2.20181
		MAE	2.56184	2.40813	2.52054	2.50908
	100	MSE	2.23046	2.01891	2.15567	2.17831
		MAE	2.56769	2.40434	2.52533	2.53172
15%	25	MSE	2.40046	2.18482	2.32485	2.34095
		MAE	2.71563	2.56773	2.67315	2.67453
	50	MSE	2.28615	2.08078	2.21414	2.22948
		MAE	2.58631	2.44546	2.54586	2.54717
	100	MSE	2.27052	2.05638	2.20108	2.2299
		MAE	2.58274	2.42552	2.55423	2.5692
20%	25	MSE	2.43498	2.19929	2.36014	2.38575
		MAE	2.74042	2.58889	2.69938	2.7088
	50	MSE	2.31903	2.09456	2.24775	2.27214
		MAE	2.60992	2.46561	2.57084	2.57981
	100	MSE	2.30818	2.06587	2.23684	2.26741
		MAE	2.61503	2.46647	2.57619	2.59311

جدول (4): قيم MSE وMAE للطرائق المختلفة وللحالة الثالثة (III)

PR	Sample Size	Criteria	MLE	M	S	MM
10%	25	MSE	2.93018	0.21094	2.86704	1.69096
		MAE	2.73804	0.61685	2.66965	1.90724
	50	MSE	2.79065	0.20090	2.73051	1.61044
		MAE	2.60765	0.58748	2.54252	1.81642
	100	MSE	2.65776	0.19133	2.60049	1.53375
		MAE	2.48348	0.55950	2.42145	1.72992
15%	25	MSE	3.18131	0.42035	3.11315	2.74179
		MAE	2.87539	0.80812	2.80877	2.61763
	50	MSE	3.02982	0.40033	2.96490	2.61123
		MAE	2.73847	0.76964	2.67502	2.49298
	100	MSE	2.88554	0.38127	2.82371	2.48689
		MAE	2.60807	0.73299	2.54764	2.37427
20%	25	MSE	3.34106	0.82904	3.27639	3.20643
		MAE	2.96280	1.17237	2.90029	2.88347
	50	MSE	3.18196	0.78956	3.12037	3.05374
		MAE	2.82171	1.11654	2.76218	2.74616
	100	MSE	3.03044	0.75196	2.97178	2.90832
		MAE	2.68734	1.06337	2.63065	2.61539

جدول (5): قيم MSE وMAE للطرائق المختلفة وللحالة الرابعة (IV)

PR	Sample Size	Criteria	MLE	M	S	MM
10%	25	MSE	1.65152	1.34366	1.62324	1.54831
		MAE	2.31638	2.01915	2.29742	2.23542
	50	MSE	1.57288	1.27968	1.54594	1.47458
		MAE	2.20608	1.92300	2.18802	2.12897
	100	MSE	1.49798	1.21874	1.47232	1.40436
		MAE	2.10103	1.83143	2.08383	2.02759
15%	25	MSE	1.68235	1.48050	1.65199	1.60863
		MAE	2.34142	2.19273	2.32125	2.28776
	50	MSE	1.60224	1.41000	1.57332	1.53203
		MAE	2.22992	2.08831	2.21071	2.17882
	100	MSE	1.52594	1.34286	1.49840	1.45908
		MAE	2.12373	1.98887	2.10544	2.07507
20%	25	MSE	1.72722	1.51155	1.69010	1.67346
		MAE	2.37198	2.22116	2.34767	2.33575
	50	MSE	1.64497	1.43957	1.60962	1.59377
		MAE	2.25903	2.11539	2.23588	2.22452
	100	MSE	1.56664	1.37102	1.53297	1.51788
		MAE	2.15146	2.01466	2.12941	2.11859

جدول (6): قيم MSE وMAE للطرائق المختلفة وللحالة الخامسة (V)

PR	Sample Size	Criteria	MLE	M	S	MM
10%	25	MSE	24.80634	12.83330	24.12417	21.56363
		MAE	8.97836	7.33459	8.86391	7.88489
	50	MSE	23.62509	12.22219	22.97540	20.53679
		MAE	8.55082	6.98532	8.44182	7.50942
	100	MSE	22.50009	11.64018	21.88133	19.55885
		MAE	8.14364	6.65269	8.03983	7.15183
15%	25	MSE	24.96664	17.33314	24.32649	23.90185
		MAE	9.00648	6.90739	8.89893	8.65880
	50	MSE	23.77775	16.50775	23.16809	22.76367
		MAE	8.57760	6.57847	8.47517	8.24648
	100	MSE	22.64548	15.72167	22.06485	21.67969
		MAE	8.16914	6.26521	8.07159	7.85379
20%	25	MSE	25.09571	19.30146	24.50446	24.50035
		MAE	9.03097	7.65601	8.93155	8.83676
	50	MSE	23.90068	18.38234	23.33758	23.33367
		MAE	8.60092	7.29144	8.50624	8.41596
	100	MSE	22.76255	17.50699	22.22627	22.22254
		MAE	8.19135	6.94423	8.10118	8.01520

جدول (7): قيم MSE وMAE للطرائق المختلفة وللحالة السادسة (VI)

PR	Sample Size	Criteria	MLE	M	S	MM
10%	25	MSE	124.12461	42.31230	121.09400	71.23852
		MAE	19.92541	8.09803	19.70914	11.81062
	50	MSE	118.21391	40.29743	115.32762	67.84621
		MAE	18.97658	7.71241	18.77061	11.24821
	100	MSE	112.58468	38.37850	109.83583	64.61544
		MAE	18.07293	7.34515	17.87677	10.71258
15%	25	MSE	124.56729	49.03915	120.92441	77.63511
		MAE	19.93840	9.34597	19.70226	12.85716
	50	MSE	118.63551	46.70395	115.16610	73.93820
		MAE	18.98895	8.90092	18.76406	12.24491
	100	MSE	112.98620	44.47995	109.68200	70.41733
		MAE	18.08471	8.47707	17.87053	11.66182
20%	25	MSE	125.34374	63.73220	121.31534	91.12018
		MAE	19.94069	12.12219	19.84176	15.04016
	50	MSE	119.37499	60.69733	115.53842	86.78112
		MAE	18.99113	11.54494	18.89691	14.32396
	100	MSE	113.69047	57.80698	110.03659	82.64869
		MAE	18.08679	10.99518	17.99706	13.64187

الاستنتاجات

من خلال تحليل نتائج المحاكاة وجدنا ان قيم متوسط مربعات الخطا (MSE) وقيم متوسط مطلق الخطأ (MAE) لتقدير انموذج انحدار بواسون بالطريقة الأعتيادية (MLE) والطرائق الحصينة الموضحة في الجداول (2) الى (7) تشير الى :

1- افضلية طريقة M على الطرائق الحصينة الاخرى والتي بدورها تتفوق اغلب الاحيان على طريقة MLE لتقدير معلمات انموذج انحدار بواسون في حالة وجود القيم الشاذة .
2- ان قيم متوسط مربعات الخطأ MSE وقيم متوسط مطلق الخطأ MAE تقل كلما ازدادت حجوم العينات في كل الطرائق ولجميع النماذج المدروسة .
3- ان قيم متوسط مربعات الخطأ MSE وقيم متوسط مطلق الخطأ MAE تزداد كلما ازدادت نسبة التلوث في كل الطرائق ولجميع النماذج المدروسة .

المصادر

1- سلطان، مها حسن، 2017 "طريقة الامكان الاعظم وبعض الطرائق اللامعلمية في تقدير انموذج انحدار بواسون" رسالة ماجستير، كلية الادارة والاقتصاد، الجامعة المستنصرية.
2- هادي، خديجة عبد الكريم، 2021 "مقارنة بعض الطرائق الحصينة لتقدير انموذج الانحدار للبيانات ذات التوزيع الطبيعي الملتوي "رسالة ماجستير ، كلية الادارة والاقتصاد، الجامعة المستنصرية.
3- باهض، زهراء خالد، 2022 "التقدير الحصين لأنموذج الانحدار شبه المعلمي مع تطبيق عملي "رسالة ماجستير، كلية الادارة والاقتصاد، الجامعة المستنصرية.
4- محمد، رضا قاسم، 2022 "المقدرات الحصينة لمعلمات انموذج الانحدار بأخطاء عشوائية غير متجانسة التباين" رسالة ماجستير، كلية الادارة والاقتصاد، الجامعة المستنصرية.

5- الجشعمي، حسين علي، 2007 "مقارنة لبعض المقدرات الحصينة لمعالم النماذج اللاخطية"، اطروحه دكتوراه، كلية الادارة والاقتصاد، الجامعة المستنصرية.
6- صالح، طارق عزيز، 2009 "مقارنة بعض الطرائق الحصينة في تحليل الارتباط القويم الخطي بأستخدام المحاكاة مع تطبيق عملي" رسالة ماجستير، كلية الادارة والاقتصاد، جامعة بغداد .
7- لعيبي، اماني عماد، 2018 "مقارنة بعض الطرائق الحصينة لتقدير معلمتي توزيع دالة القوى مع تطبيق عملي" رسالة ماجستير ، كلية الأدارة والأقتصاد، الجامعة المستنصرية.
8- خليل، منال اسماعيل، 2011 "استخدام المقدرات الحصينة في التحليل العنقودي مع تطبيق عملي في مجال الفساد الاداري والمالي" رسالة ماجستير، كلية الأدارة والأقتصاد، جامعة بغداد.

9- Winklmann Rainer, 2008, " Econometric Analysis of Count Data" Fifth edition , springer-verlag Berlin Heidelberg .
10- Philip, N. and Sebastian, N., 2015, "Application of Poisson Regression on Traffic Safety"
11- Mohamed, R. A. and Omnia, M. S., 2020, "A comparative study of Robust Estimation for Poisson Regression Model with outliers" , Journal of statistics Application and probability, No. 2, PP. (279-286).