



مجلة الإدارة والاقتصاد Journal of Administration & Economics

Mustansiriyah
University

College of
Administration &
Economics

P-ISSN: 1813 - 6729

E-ISSN: 2707-1359

المقارنة بين مقدرات طريقة انحدار ليو وطريقة انحدار الحرف لنموذج انحدار بواسون في ظل وجود مشكلة التعدد الخطي

رائد فاضل محمد الحسني

قسم المحاسبة، كلية الإدارة والاقتصاد، جامعة المستنصرية، بغداد، العراق

Email: raad@uomustansiriyah.edu.iq, ORCID ID: <https://orcid.org/0000-0002-7309-8702>

المستخلص

عتبر أنموذج انحدار بواسون أحد أهم نماذج الانحدار اللوغاريتمية الخطية، وهو الأداة التي يتم من خلالها نمذجة المتغير المعتمد عندما تكون قيمه موجبة وبهيئة معدلات أو بيانات معدودة، فضلاً عن كونه الأنموذج الملائم لتحليل الأحداث النادرة. وكغيره من سائر نماذج الانحدار، قد تتعرض المتغيرات التوضيحية الداخلة في بناءه إلى ارتباط عالي بين متغيرين أو أكثر ما يؤثر سلباً على عملية تقدير معاملات الأنموذج. في هذا البحث سنستعرض أبرز طرائق تقدير معاملات أنموذج انحدار بواسون عندما تعاني البيانات من مشكلة التعدد الخطي شبه التام كطريقة انحدار الحرف وطريقة مقدرات ليو، كما يهدف البحث الى المقارنة بين الطريقتين من خلال اعتماد معيار معلومات أكيائي كمعيار للمقارنة. تم تطبيق طرائق التقدير على بيانات حقيقية تم الحصول عليها من مستشفى الطفل المركزي في بغداد تمثل عدد حالات العيوب الخلقية للأطفال في القلب وجهاز الدوران للفترة من عام (2012) وحتى نهاية عام (2019)، وأظهرت النتائج تفوق طريقة انحدار ليو على طريقة انحدار الحرف.

معلومات البحث

تواريخ البحث:

تاريخ تقديم البحث: 1 / 2 / 2024

تاريخ قبول البحث: 10 / 6 / 2024

عدد صفحات البحث 36 - 45

الكلمات المفتاحية:

انحدار بواسون ، مقدرات ليو ، مقدر

انحدار الحرف ، مشكلة التعدد الخطي .

المراسلة:

أسم الباحث: رائد فاضل الحسني

Email:

raad@uomustansiriyah.edu.iq

1. المقدمة

تناقش العديد من البحوث والدراسات مشكلة التعدد الخطي من حيث نشأتها وطرق الكشف عنها وأثارها في عملية تقدير معاملات نموذج الانحدار الخطي عندما تتبع الأخطاء العشوائية توزيعاً طبيعياً فضلاً عن وضع الحلول للحل من آثار تلك المشكلة، كما ناقشت بحوث أخرى ذات المشكلة في النماذج التي تتبع الأخطاء العشوائية فيها توزيعات احتمالية أخرى. وسرعان ما تنوعت الأفكار في تلك الحلول، كما أملت ضرورات التطبيق كندرة الحوادث المدروسة أو تنوع المتغيرات الداخلة في بناء نموذج قيد الدراسة وعدم اتباع بيانات الدراسة للتوزيع الطبيعي ظهور أنواع أخرى من نماذج الانحدار المختلفة خصوصاً عندما لا تصلح نماذج الانحدار الخطية كأنموذج انحدار كوكس (Cox Regression) وأنموذج انحدار بواسون (Poisson Regression Model) وغيرها، فقدمت تلك النماذج الأخرى إضافة كبيرة ودعمًا قوياً للنظريات الإحصائية لتطبيقات عدة في علوم الحياة الأخرى.

وعلى الرغم من عراقة نماذج الانحدار في أدبيات البحث العلمي في المجال الإحصائي، إلا أن الاتجاه العام في دراسة أنموذج انحدار بواسون الذي تتبع الأخطاء العشوائية فيه توزيع بواسون والمستعمل في تحليل الأحداث نادرة الوقوع من حيث عملية تقدير المعلمات عندما يعاني من مشكلات مختلفة، وقد بدأ في الآونة الأخيرة يأخذ حيزاً واسعاً في الدراسات الإحصائية الحديثة. لذلك، سيتم التطرق في هذا البحث الى أنموذج انحدار بواسون عند وجود مشكلة التعدد الخطي شبه التام.

2. هدف البحث

يهدف البحث الى استعراض أبرز طرائق تقدير معاملات أنموذج انحدار بواسون عندما تعاني البيانات من مشكلة التعدد الخطي شبه التام كطريقة انحدار الحرف وطريقة مقدرات ليو. وكذلك المقارنة بين هذه الطرائق من خلال اعتماد معيار معلومات أكيائي كمعيار للمقارنة.

3. توزيع بواسون:

يعد توزيع بواسون من التوزيعات المتقطعة المهمة جداً في الكثير من التطبيقات الإحصائية، ويسمى في بعض الأحيان توزيع الحوادث النادرة الوقوع كحوادث سقوط الطائرات وعدد الوحدات المعيبة في إنتاج مصنع معين، وأول من اشتق هذا التوزيع هو العالم الرياضي الفيزيائي الفرنسي (Simeon Poisson) وأسس ذلك الاشتقاق على أن التوزيع يمثل حالة تقاربيه لتوزيع ثنائي الحدين، ونشر اشتقاقه عام (1873) م كما أطلق اسمه على ذلك التوزيع. [2]

ولنفرض أن (Y_i) متغير عشوائي يشير إلى عدد الأوقات لحصول حدث معين خلال فترة من الزمن، فإن (Y_i) يتوزع تبعاً لتوزيع بواسون بمعلمة قدرها (μ) وإن دالة الكثافة الاحتمالية للتوزيع يمكن التعبير عنها في المعادلة (1) التالية:

$$P_r(Y) = \frac{e^{-\mu} \mu^Y}{Y!}, Y_i = 0,1,2,\dots \quad (1)$$

حيث تمثل (μ) معلمة التوزيع وهي ذات قيمة أكبر من الصفر $(\mu > 0)$. ومن أهم خصائص توزيع بواسون، إن معدل عدد أوقات حصول حدث معين خلال فترة من الزمن تمثل معلمة التوزيع وهي بذات الوقت وسطه الحسابي، أي أن:

$$E(Y) = \mu \quad (2)$$

وإن قيمة التباين لتوزيع بواسون تكون مساوية إلى قيمة الوسط الحسابي، أي أن:

$$Var(Y) = \mu \quad (3)$$

إذ تعرف هذه الخاصية بخاصية التكافؤ (Equidispersion)، وفي التطبيق غالباً ما يكون التباين للمتغيرات المعدودة أكبر من الوسط الحسابي، وفي هذه الحالة تدعى الخاصية أعلاه بخاصية فوق التشنت (Over Dispersion). كما يتصف توزيع بواسون بأنه من التوزيعات المتلوية باتجاه اليمين، وكلما ازدادت قيمة معلمة التوزيع، تبدأ قيمة احتمال الصفر بالانخفاض، وحالة خاصة عندما تكون قيمة المعلمة $(\mu=10.5)$ ، فإن توزيع بواسون سيقترب من التوزيع الطبيعي. [1]

4. أنموذج انحدار بواسون:

أحد أنواع نماذج الانحدار التي تنضوي تحت مظلة نماذج الانحدار الخطية – اللوغاريمية، إذ من خلال اخذ اللوغاريم الطبيعي لصيغة التوزيع فإنها تتحول إلى صيغة خطية، كما تتبع الأخطاء العشوائية في الأنموذج توزيع بواسون بمعلمة قدرها (μ) . ويرتكز الأنموذج على أمرين مهمين، يتمثل الأول بالافتراضات الخاصة بالتوزيع كونه يختلف عن توزيع الأخطاء العشوائية كما في أنموذج الانحدار الخطي، والثاني مميزات معلمة توزيع بواسون (μ) باعتبارها دالة للمتغيرات التوضيحية. [3]

$$\underline{Y} = e^{X\underline{\beta} + \underline{U}} \quad (4)$$

إذ يمثل الرمز \underline{Y} موجه متغير الاستجابة ذي درجة $(n \times 1)$ ، ويمثل الرمز X مصفوفة المتغيرات التوضيحية ذات الدرجة $(n \times (p+1))$ ، وان $\underline{\beta}$ هو متجه المعلمات ذو الدرجة $((p+1) \times 1)$ ، وان \underline{U} يمثل متجه الأخطاء العشوائية ذي الدرجة $(n \times 1)$ ، أما الرمز n فيمثل حجم العينة، والرمز p يمثل عدد المتغيرات التوضيحية في النموذج.

ويستعمل أنموذج انحدار بواسون عندما تكون قيم المتغير (Y) بهيئة قيم معدودة (Count Data)، كعدد حوادث تصادم السفن، وعدد حوادث التأمين المسجلة في إحدى الشركات، وعدد براءات الاختراع التي تحصل عليها جهة طبية في إنتاج دواء معين، وعدد حوادث سقوط الطائرات، وعدد الأطفال المصابين بالعيوب الخلقية في القلب وجهاز الدوران، وعدد الرجال المصابين بسرطان الثدي وغيرها. [9]

كما يستعمل أنموذج انحدار بواسون في حالة كون قيم متغير الاستجابة بهيئة معدلات (Rate Data)، كمعدل حوادث السيارات في إحدى المدن، ومعدل النداءات الهاتفية المستلمة من بدالة هاتف خلال فترة زمنية معينة، ومعدل الأخطاء المطبعية في كتب إحدى المكتبات، ومعدل حوادث خروج القطارات عن مسارها ضمن السكة الحديدية. بالتالي، فإن أنموذج انحدار بواسون يكون الحل الأمثل للحالات التي تملحها طبيعة البيانات من حيث كونها معدودة أو معدلات، فضلاً عن درجة حدوث تلك الحوادث التي تعبر عن البيانات من حيث ندرتها. [11]

5. مشكلة التعدد الخطي:

تحصل مشكلة التعدد الخطي عندما يرتبط متغيران توضيحيان أو أكثر بعلاقة خطية قوية، بحيث يصعب فصل تأثير كل متغير توضيحي عن المتغير المعتمد في الواقع التطبيقي، كما تحدث مشكلة التعدد الخطي حينما تكون قيم أحد المتغيرات متساوية لكافة المشاهدات، أو عندما تعتمد قيمة أحد المتغيرات التوضيحية على واحد أو أكثر من المتغيرات التوضيحية الأخرى في النموذج قيد الدراسة، وكذلك في حال أخذت البيانات شكل السلسلة الزمنية (Time Series) أو البيانات المقطعية (Cross – Section Data). كما يمكن تعريف المشكلة من خلال مفهوم التعامد عندما تكون المتغيرات التوضيحية متعامدة (غير مرتبطة ببعضها البعض) فإن جميع الجذور الكامنة أو القيم المميزة (Eigen values) لمصفوفة التصميم (Design Matrix) ذات الرتبة الكاملة

(Full rank) تكون مساوية إلى الواحد، أما إذا كانت على الأقل إحدى تلك القيم المميزة أقل من الواحد خصوصاً عندما تكون مساوية أو قريبة من الصفر، فإن ذلك يدل على عدم التعامد ووجود تعدد خطي. وبشكل عام يمكن تصنيف مشكلة التعدد الخطي على نوعين: [5]

1.5.1. التعدد الخطي التام:

تكون قيمة محدد مصفوفة المعلومات مساوية للصفر $|X'X| = 0$. ويترتب على ذلك استحالة إيجاد تقدير لمعاملات نموذج الانحدار نتيجة لعدم إمكانية حساب معكوس المصفوفة $(X'X)$ وأبرز طرائق معالجة هذه الحالة استعمال طريقة المكونات الرئيسية (Principle Component Analysis).

2.5.2. التعدد الخطي شبه التام:

في هذه الحالة تكون قيمة محدد مصفوفة المعلومات صغيرة جداً (قريبة جداً من الصفر) عندها تكون المعلمات المقدرة ذات تباين كبير جداً، وأبرز طرائق معالجة هذه الحالة هو استعمال طريقة انحدار الحرف (Ridge Regression Method). [8] ومن أهم تأثيرات مشكلة التعدد الخطي هو أن مقدرات معاملات نموذج الانحدار تكون غير مستقرة (Unstable) في حالة التعدد الخطي شبه التام نتيجة لصغر قيمة محدد مصفوفة المعلومات مما يجعل قيمة تباين المعلمة المقدرة كبيرة جداً (Inflated) وهذا بدوره سيؤدي إلى اختلال دقة المقدرات. ويمكن التعبير عن مصفوفة التباين والتباين المشترك للمعاملات المقدرة بالصيغة (5) التالية:

$$Var - Cov(\hat{\beta}_i) = S_e^2 (X'X)^{-1} = S_e^2 \frac{adj(X'X)}{|X'X|} \quad (5)$$

حيث يمثل الرمز (S_e^2) تباين الخطأ العشوائي، ويمثل الرمز $adj(X'X)$ المرافقة الطبيعية لمصفوفة المعلومات $(X'X)$ ، كما تتسبب مشكلة التعدد الخطي في إظهار معاملات المعلمات بإشارة خاطئة، وإن اختبار معنوية الأنموذج قد يقصي متغيرات مهمة من الأنموذج المدروس. [7]

6. الكشف عن مشكلة التعدد الخطي:

في التطبيقات الاقتصادية، غالباً ما يكون هناك قدر من الارتباط بين المتغيرات المستقلة بسبب أو آخر، وذلك نتيجة لميل تلك المتغيرات نحو التحرك سوية عبر الزمن، فمثلاً عند فترات الرواج الاقتصادي تميل متغيرات الدخل، الأسعار، الإنتاج، الاستهلاك وغيرها نحو الارتفاع، بينما يحدث العكس في فترات الكساد، بما يدل على وجود علاقات خطية بين تلك المتغيرات وبالتالي حدوث التعدد الخطي، إلا أن التعدد الخطي بين تلك المتغيرات لا يعتبر مشكلة إلا إذا تجاوزت شدته حد معين يتعدى عنده تحديد أثر أي من تلك المتغيرات التوضيحية في المتغير المعتمد بشكل منفصل. [4]

لعل أفضل الوسائل التي تقيس شدة التعدد الخطي، هو مقياس معامل تضخم التباين (Variance Inflation Factor) والذي يرمز له اختصاراً (VIF)، وتكون صيغة هذا المقياس كما في المعادلة (6) الآتية:

$$VIF = \frac{1}{1 - R_j^2} \quad , j = 1, 2, \dots, p \quad (6)$$

حيث يمثل (R_j^2) معامل التحديد الخاص بالمتغير (x) على بقية المتغيرات التوضيحية الأخرى في الأنموذج، وقد أشار الباحثان (Eyup & Sana) بأن قيمة معامل تضخم التباين في حال تجاوزها (10) دل ذلك على وجود تأثير كبير لمشكلة التعدد الخطي. كما وضع (Klein) قاعدة لاختبار وجود مشكلة التعدد الخطي مفادها في حالة كون قيمة معامل الارتباط البسيط لأي متغيرين أكبر من قيمة معامل التحديد دل ذلك على وجود مشكلة التعدد الخطي. [6]

7. تقدير معاملات أنموذج انحدار بواسون:

تستخدم طريقة مقدرات الإمكان الأعظم لتقدير معاملات أنموذج انحدار بواسون، فإذا كان المتغير المعتمد (Y_i) يتبع توزيع بواسون بمعلمة قدرها (μ) فإن دالة التوزيع تكون بالصيغة (1) والمبينة سلفاً، ومن خلال تعظيم المشاهدات لتوزيع المتغير المعتمد (Y_i) الوارد في الصيغة (1) وأخذ اللوغاريتم الطبيعي لدالة الإمكان للمشاهدات، واشتقاق دالة الإمكان وحل المعادلة باستعمال الطرائق التكرارية، فإن مقدرات المعلمات لأنموذج انحدار بواسون تكون بالصيغة (7) التالية:

$$\hat{\beta}_{MLE} = (X'WX)^{-1}(X'WZ) \quad (7)$$

حيث يمثل $\hat{\beta}_{MLE}$ متجه معاملات أنموذج انحدار بواسون المقدرة وفق طريقة الإمكان الأعظم. ويمثل الرمز W مصفوفة قطرية عناصر قطرها الرئيسي هي القيم المقدرة لمعلمة توزيع بواسون $(\hat{\mu}_i)$ ، أي أن:

$$W = \begin{pmatrix} e^{\hat{\mu}} & & & \\ & e^{\hat{\mu}} & & \\ & & \ddots & \\ & & & e^{\hat{\mu}} \end{pmatrix}$$

ويمثل \underline{Z} متجه، حيث أن العنصر (i) في المتجه \underline{Z} يكون مساوياً إلى:

$$Z_i = \text{Log}(\hat{\mu}_i) + \frac{Y_i - \hat{\mu}_i}{\hat{\mu}_i} \quad (8)$$

وإن مصفوفة التباين المشترك لمقدرات الإمكان الأعظم لأنموذج انحدار بواسون تكون بالصيغة (9) التالية:

$$\text{Cov}(\hat{\underline{\beta}}_{ML}) = \left[E \left\{ \frac{\partial^2 L(X; \underline{\beta})}{\partial \underline{\beta} \partial \underline{\beta}'} \right\} \right]^{-1} = \sigma_u^2 (X' \widehat{W} X)^{-1} \quad (9)$$

حيث يمثل الرمز σ_u^2 تباين الخطأ العشوائي للمجتمع. وبذلك يكون متوسط مربعات الخطأ (Mean Squared Error) لمعاملات أنموذج انحدار بواسون المقدر وفق طريقة الإمكان الأعظم كما في المعادلة (10) الآتية:

$$\text{MSE}(\hat{\underline{\beta}}_{ML}) = E(\hat{\underline{\beta}}_{ML} - \underline{\beta})'(\hat{\underline{\beta}}_{ML} - \underline{\beta}) = \text{tr}(X' \widehat{W} X)^{-1} = \sum_{j=1}^p \frac{1}{\lambda_j} \quad (10)$$

إذ تمثل (λ_j) القيمة المميزة للعنصر (j) في المصفوفة ($X' \widehat{W} X$). وعند وجود علاقة قوية بين المتغيرات التوضيحية، سيكون أداء المصفوفة الموزونة ($X' \widehat{W} X$) ضعيفاً في حال احتساب تقدير لمعاملات الأنموذج، مما يؤدي إلى عدم استقرار المقدرات، وتضخم تباين المعلمات المقدر وفق هذه الطريقة، ما يدعو إلى عدم اعتماد هذه الطريقة على الرغم من كون المقدرات المتحصل عليها غير متحيزة. [9]

8. تقدير معاملات أنموذج انحدار بواسون في حالة وجود مشكلة التعدد الخطي:

1.8. طريقة مقدرات انحدار الحرف:

إحدى البدائل المهمة لتقدير معاملات نماذج الانحدار عندما يكون هناك تعدد خطي أو ازدواج خطي بين المتغيرات التوضيحية، وتنشأ هذه الطريقة وفقاً لمبدأ الباحثان (Hoerl & Kennard) وهو بإضافة كمية صغيرة موجبة إلى عناصر القطر الرئيسي لمصفوفة المعلومات مع اختلاف نوع الأنموذج باعتباره أنموذج انحدار بواسون لوغاريتمي خطي. وتبدأ هذه الطريقة بملاحظة أن طريقة المكان الأعظم لتقدير معاملات الأنموذج تعمل على تصغير مجموع مربعات الأخطاء الموزون، وإذا تم اختبار مقدر آخر وليكن $\underline{\beta}_{RR}$ ، فيمكن كتابة مجموع مربعات الأخطاء الموزون كما في الصيغة (11) التالية. [4]

$$\underline{u}'\underline{u} = (\underline{Y} - X\underline{\beta}_{ML})'(\underline{Y} - X\underline{\beta}_{ML}) + (\underline{\beta}_{RR} - \underline{\beta}_{ML})'(X' \widehat{W} X)(\underline{\beta}_{RR} - \underline{\beta}_{ML}) = \theta_{min} + \theta(\underline{\beta}_{RR}) \quad (11)$$

حيث تمثل θ_{min} الزيادة في متوسط مربعات الأخطاء الموزون في حال استبدال المعلمات المقدر بطريقة الإمكان الأعظم $\underline{\beta}_{ML}$ بالمعاملات المزعم إيجادها $\underline{\beta}_{RR}$. ومن خلال العلاقة العكسية بين القيم المميزة وتباين المعلمات المقدر وفق طريقة الإمكان الأعظم والتي تمت الإشارة إليها في الصيغة (10)، فإن متجه المعلمات المزعم احتسابه $\underline{\beta}_{RR}$ سيصغر مجموع مربعات الأخطاء الموزون وفقاً للقيود التالي:

$$\text{Minimize } F = \underline{\beta}_{RR}'\underline{\beta}_{RR} + \left(\frac{1}{K}\right)(\underline{\beta}_{RR} - \underline{\beta}_{ML})'(X' \widehat{W} X)(\underline{\beta}_{RR} - \underline{\beta}_{ML} - \theta_0) \quad (12)$$

حيث أن $\left(\frac{1}{K}\right)$ مضاعف لاكرانج، وأن $\theta_0 = \theta(\underline{\beta}_{RR})$ وبذلك سيكون مجموع مربعات البواقي الموزون كما في الصيغة (13):

$$\underline{u}'\underline{u} = (\underline{Y} - X\underline{\beta}_{ML})'(\underline{Y} - X\underline{\beta}_{ML}) + \underline{\beta}_{RR}'\underline{\beta}_{RR} + \left(\frac{1}{K}\right)(\underline{\beta}_{RR} - \underline{\beta}_{ML})'(X' \widehat{W} X)(\underline{\beta}_{RR} - \underline{\beta}_{ML} - \theta_0) \quad (13)$$

ومن خلال اشتقاق الصيغة (13) بالنسبة لمتجه المعلمات $\underline{\beta}_{RR}$ وبمساواة قيمة المشتقة إلى الصفر نحصل على مقدرات انحدار الحرف لأنموذج انحدار بواسون وكما يلي: [6]

$$\hat{\underline{\beta}}_{RR} = (X' \widehat{W} X + KI)^{-1} (X' \widehat{W} X) \hat{\underline{\beta}}_{ML} = Z \hat{\underline{\beta}}_{ML} \quad (14)$$

وتكون مقدرات انحدار الحرف متحيزة عندما ($K > 0$)، وان مقدار التحيز يكون بالصيغة (15) التالية:

$$\text{Bias}(\hat{\underline{\beta}}_{RR}) = E(\hat{\underline{\beta}}_{RR}) - \underline{\beta} = Z\underline{\beta} - \underline{\beta} = (Z - I)\underline{\beta} \quad (15)$$

كما أن مصفوفة تباين مقدرات انحدار الحرف تكون كما في الصيغة (16) الآتية:

$$Var - Cov(\hat{\beta}_{RR}) = Z Var - Cov(\hat{\beta}_{ML}) Z' = Z \sigma_u^2 (X' \hat{W} X)^{-1} Z' \quad (16)$$

أما متوسط مربعات الخطأ $MSE(\hat{\beta}_{RR})$ لمعاملات نموذج انحدار بواسون المقدره وفق طريقة انحدار الحرف فيكون بالصيغة (17) الآتية:

$$MSE(\hat{\beta}_{RR}) = E(\hat{\beta}_{RR} - \beta)'(\hat{\beta}_{RR} - \beta) = \sum_{j=1}^p \frac{\lambda_j}{(\lambda_j + K)^2} + K^2 \sum_{j=1}^p \frac{\alpha_j^2}{(\lambda_j + K)^2} \quad (17)$$

حيث أن:

$$\alpha_j = \gamma \hat{\beta}_{ML} \quad (18)$$

وأن α_j تمثل المعلمة (j) من معاملات نموذج انحدار بواسون عند اخذ اللوغاريتم الطبيعي للنموذج، وان (γ) تمثل المتجه الكامن للمصفوفة $(X' \hat{W} X)$ ، وأن (K) تمثل معلمة الحرف أو معلمة التحيز، ويمكن تقديرها وفقاً لصيغة (hoerl & kennard) كما في المعادلة (19) الآتية:

$$K_1 = \frac{\hat{\sigma}_u^2}{\hat{\alpha}_{max}^2} \quad (19)$$

حيث تمثل $\hat{\sigma}_u^2$ متوسط مربعات الأخطاء المقدر والذي يحسب في نموذج انحدار بواسون وفق الصيغة (20) التالية:

$$S_e^2 = \frac{\sum_{i=1}^n (Y_i - \hat{\mu}_i)^2}{n - p - 1} \quad (20)$$

وتمثل $\hat{\alpha}_{max}^2$ أكبر عنصر في الصيغة (19)، وتقع قيمة معلمة التحيز ضمن المجال أدناه: [9]

$$K \in \left\{ 0, \frac{\partial \sigma_u^2}{\alpha_{max}} \right\}$$

2.8. طريقة مقدرات ليو:

وضع أسس هذه الطريقة الباحث (Liu) في عام (1993) م لمعالجة مسألة تضخم تباين معاملات النموذج المقدره، في ظل وجود مشكلة التعدد الخطي، وتنشأ هذه الطريقة بذات الأسلوب الذي اتبعه (Liu) مع اختلاف طبيعة نوع نموذج انحدار بواسون كونه نموذج غير خطي.

$$Y = e^{X\beta + U} \quad (21)$$

بأخذ اللوغاريتم الطبيعي لها نحصل على

$$\text{Log}(Y) = \text{Log}\{e^{X\beta + U}\}$$

$$Y^* = X\beta + U$$

حيث أن:

$$Y^* = \text{Log}(Y)$$

وباعتماد القيد الذي وضعه (Liu) فإن مقدار زيادة متوسط مربعات الخطأ الموزون يعبر عنه بالمعادلة (22) التالية:

$$\epsilon' \epsilon = (d\hat{\beta} - \beta^*)'(d\hat{\beta} - \beta^*) \quad (22)$$

حيث يمثل $\epsilon' \epsilon$ مقدار الزيادة في متوسط مربعات الخطأ الموزون في حال تم استبدال متجه المعاملات المقدره بطريقة الإمكان الأعظم $\hat{\beta}$ بمتجه المعاملات المقدره وفقاً لطريقة ليو β^* ، ويمثل d معلمة الإضافة (معلمة التحيز). وبأخذ مربع الأخطاء الموزون

للنموذج المبحوث مع إضافة القيد ومن ثم اشتقاق الناتج بالنسبة إلى $\hat{\beta}$ نحصل على: [10]

$$\begin{aligned} \epsilon' W \epsilon &= (Y^* - X\beta^*)' \hat{W} (Y^* - X\beta^*) + (d\hat{\beta} - \beta^*)'(d\hat{\beta} - \beta^*) \\ &= Y^{*'} \hat{W} Y^* - 2\beta^{*'} X' \hat{W} Y^* + \beta^{*'} X' \hat{W} X \beta^* + d' \hat{\beta}' \hat{\beta} d - 2d\hat{\beta}' \beta^* + \beta^{*'} \beta^* \end{aligned} \quad (23)$$

باشتقاق المعادلة (23) بالنسبة لمتجه المعلمات β^* ، ومساواة ناتج الاشتقاق بالصفر يتم الحصول على مقدرات (Liu) لمعاملات أنموذج انحدار بواسون، وكما في الصيغة (24) التالية:

$$\hat{\beta}_{Liu} = (X'WX + I)^{-1}(X'WY^* + d\hat{\beta}) \quad (24)$$

إذ إن مقدرات (Liu) تعد متحيزة عندما تكون قيمة $(d > 0)$ ومقدار التحيز يكون كما في المعادلة (25) التالية:

$$Bias(\hat{\beta}_{Liu}) = E(\hat{\beta}_{Liu}) - \beta = Z\hat{\beta} - \beta = (Z - I)\beta \quad (25)$$

حيث أن:

$$Z = (X'WX + I)^{-1}(X'WX + dI) \quad (26)$$

وأن مصفوفة تباين مقدرات (Liu) تكون بالصيغة (27) التالية:

$$Var - Cov(\hat{\beta}_{Liu}) = ZVar - Cov(\hat{\beta}_{ML})Z' = Z\sigma_u^2(X'WX)^{-1}Z' \quad (27)$$

أما متوسط مربعات الخطأ لمعاملات أنموذج انحدار بواسون المقدره وفق طريقة مقدرات (Liu) فيكون بالصيغة (28) التالية:

$$MSE(\hat{\beta}_{Liu}) = E(\hat{\beta}_{Liu} - \beta)'(\hat{\beta}_{Liu} - \beta) = \sum_{j=1}^p \frac{(\lambda_j + d)^2}{\lambda_j(\lambda_j + 1)^2} + (d - 1)^2 \sum_{j=1}^p \frac{\alpha_j^2}{(\lambda_j + 1)^2} \quad (28)$$

تعد مقدرات (Liu) متحيزة وان سبب التحيز هو وجود قيمة مضافة (d) ، والتي تتراوح بين $(1, 0)$ إذ يمكن ملاحظة ذلك بسهولة في الصيغتين (24) و (26)، كما أن متوسط مربعات الخطأ المحتسب وفق طريقة (Liu) أقل من متوسط مربعات الخطأ لذات المعلمات إن قدرت وفق طريقة الإمكان الأعظم، على الرغم من كونها متحيزة. [8]

ولإيجاد المقدرات المقترحة لمعلمة التحيز (d) ينبغي تحديد القيمة المثلى لها، من خلال اشتقاق المعادلة (28) بالنسبة الى (d) ومساواة قيمة المشتقة بالصفر فنحصل على:

$$\hat{d} = \frac{\alpha_j^2 - 1}{\frac{1}{\lambda_j} + \alpha_j^2} = \frac{\alpha_j^2 - 1}{\frac{1}{\lambda_j} + \alpha_j^2} \quad (29)$$

وهي القيمة المثلى للمعلمة (d) ، حيث يمكن إيجاد القيمة المفردة للمعلمة كما في الصيغة (30) الآتية:

$$D_1 = \text{Max} \left[0, \frac{\hat{\alpha}_{max}^2 - 1}{\frac{1}{\lambda_{max}} + \hat{\alpha}_{max}^2} \right] \quad (30)$$

حيث تمثل (λ_{max}) أكبر قيمة مميزة في المصفوفة $(X'WX)$ ، إذ تم الاستناد بهذا المقدر على الفكرة الأساسية التي تبناها (Horel & Kennard) في إضافة قيمة موجبة لعناصر قطر المصفوفة $(X'WX)$ لأنموذج الانحدار الخطي، فضلاً عن إيجاد أكبر عنصر في الصيغة (18). [2]

9. الجانب التطبيقي:

لغرض تطبيق الطرائق التي تم استعراضها أعلاه، استحصل الباحث على بيانات حقيقية تخص العيوب الخلقية في القلب وجهاز الدوران لدى الأطفال حديثي الولادة من مستشفى الطفل المركزي التعليمي في حي الإسكان غربي محافظة بغداد - العراق، حيث تم دراسة توزيع متغير المعتمد (y_i) والذي يمثل تشوهات القلب وجهاز الدوران لدى الأطفال والكشف عن وجود مشكلة التعدد الخطي بين المتغيرات التوضيحية (x_i) قيد الدراسة.

إذ يعاني الأطفال حديثي الولادة الوافدين إلى مستشفى الطفل المركزي التعليمي باعتباره المستشفى الرئيسي المتخصص بالأطفال في بغداد، من حالات العيوب الخلقية المتنوعة تأتي في مقدمتها العيوب الخلقية التي تصيب القلب وجهاز الدوران فضلاً عن أنواع أخرى من العيوب الخلقية وعند وصول تلك الحالات من مستشفيات الولادة والمستشفيات الأخرى في بغداد، تدخل تلك الحالات إلى جناح الخُدج ويتابعها الأطباء والممرضون ضمن العناية المركزة بغية تشخيصها ووضع الحلول الممكنة لها، ونتيجة للتأثيرات المباشرة لتلك التشوهات يتعرض بعض أولئك الأطفال إلى الوفاة.

يتم تثبيت حالات العيوب الخلقية الوافدة إلى مستشفى الطفل المركزي التعليمي في استمارة يتم إعدادها من قبل شعبة الإحصاء في المستشفى، حيث تكون بهيئة بيانات معدودة (count data)، وعلى شكل مجاميع ضمن فترات نصف شهرية، وتم اخذ عينة

للفترة من عام (2012) وحتى نهاية عام (2019)، كما تم بناء أنموذج انحدار بواسون بوصفه أحد النماذج الملائمة لوصف تلك الظاهرة، إذ كانت صيغته كما في المعادلة (31) التالية:

$$Y_i = e^{(\beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} + \beta_6 X_{i6} + \beta_7 X_{i7} + u_i)} \quad (31)$$

اذ أن:

Y_i : مجموع الأطفال المصابين بعيوب خلقية في القلب وجهاز الدوران ضمن كل فترة زمنية.

$\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7$: معاملات الميل الحدي في الأنموذج، وان β_0 تمثل الحد الثابت

X_{i1} : يمثل مجاميع أوزان الأطفال المصابين ضمن كل فترة زمنية.

X_{i2} : يمثل مجاميع أعمار آباء الأطفال المصابين ضمن كل فترة زمنية.

X_{i3} : يمثل مجاميع أعمار أمهات الأطفال المصابين ضمن كل فترة زمنية.

X_{i4} : يمثل أعداد الأطفال الذكور المصابين ضمن كل فترة زمنية.

X_{i5} : يمثل أعداد الأطفال الإناث المصابين ضمن كل فترة زمنية.

X_{i6} : يمثل أعداد الأطفال المصابين والمولودين من زواج الأقارب ضمن كل فترة زمنية.

X_{i7} : يمثل أعداد الأطفال المصابين والذين تعرضت أمهاتهم إلى إشعاع أو مؤثرات حياتية كتناول أدوية وعقاقير معينة في

فترة الحمل ضمن كل فترة زمنية.

u_i : يمثل حد الخطأ العشوائي في الأنموذج.

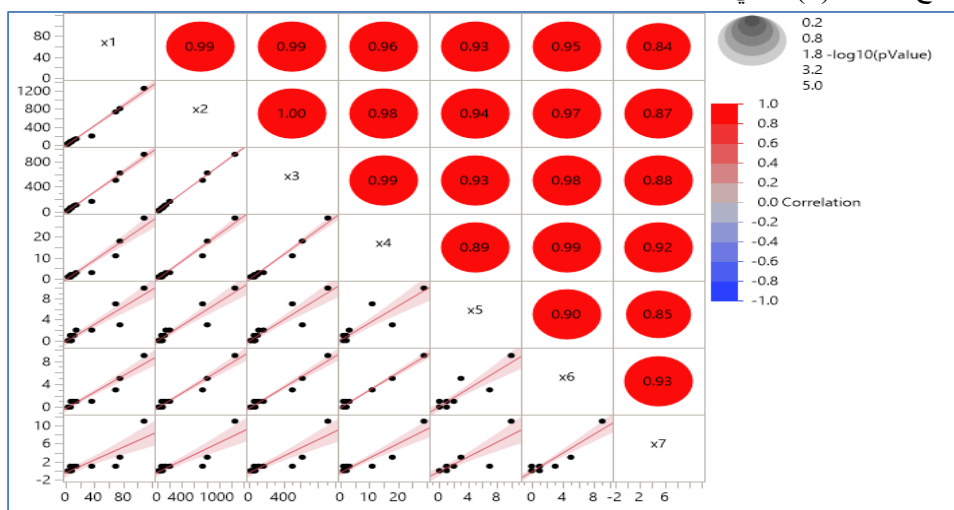
1.9 اختبار البيانات وتشخيص مشكلة التعدد الخطي:

لمعرفة طبيعة التوزيعات الاحتمالية التي من الممكن أن يتوزع متغير الاستجابة تبعاً لها، تم استعمال برنامج (JMPPro17)، وقد تبين أن (Y_i) يتبع توزيع بواسون بمعلمة توزيع قدرها $(\mu = 6.5)$. ولأجل التحقق من موائمة توزيع بواسون لمتغير الاستجابة (Y_i) في الجانب التطبيقي، تم إجراء اختبار حسن المطابقة (Goodness of Fit) لمتغير عدد الأطفال المصابين بالعيوب الخلقية في القلب وجهاز الدوران والذي من خلاله يتم التأكيد على أن توزيع بواسون هو أبرز التوزيعات التي من الممكن أن يتبع لها المتغير المعتمد، حيث نلاحظ أن قيمة اختبار (Anderson Darling) قد بلغت (495.845) بمستوى معنوية اقل من (0.001) وكما موضح بالجدول (1) التالي:

الجدول (1): اختبار حسن المطابقة لمتغير الاستجابة.

المعلمة	التقدير	الخطأ المعياري
λ	6.5	0.6374
Goodness of Fit test	Pearson X ²	p-value
	495.845	0.0001*

وللكشف عما إذا كان هناك تعدد خطي بين المتغيرات التوضيحية قيد الدراسة، تم احتساب مصفوفة الارتباط بين المتغيرات التوضيحية كما موضح بالشكل (1) التالي:



شكل (1): يبين مصفوفة ارتباط المتغيرات التوضيحية

حيث نلاحظ من الشكل (1) أن قيم معاملات ارتباط كبيرة وطردية الاتجاه لجميع المتغيرات التوضيحية، إذ يرتبط كل واحد منهم مع كافة المتغيرات التوضيحية الأخرى بعلاقات خطية طردية قوية. ويوضح الجدول (2) التالي قيم عامل تضخم التباين (VIF):

جدول (2): قيم معامل تضخم التباين (VIF)

Term	x1	x2	x3	x4	x5	x6	x7
VIF	124.569	3491.011	2728.371	643.137	114.146	80.196	60.149

إذ تعكس نتائج الجدول (2) قيم كبيرة جداً لمقياس عامل تضخم التباين (VIF) لكافة متغيرات النموذج، إذ كان أكبرها تلك الخاصة بالمتغيرات التوضيحية (x_4, x_3, x_2) كما فاق عامل تضخم التباين للمتغيرات التوضيحية الأخرى المتبقية حاجز العدد (15) الذي يشير إلى وجود مشكلة التعدد الخطي، من ذلك نستنتج بأن هناك تعدد خطي بين المتغيرات التوضيحية في النموذج. تم تطبيق طريقة انحدار ليو (Liu Regression) لتقدير معاملات نموذج انحدار بواسون عند وجود مشكلة التعدد الخطي شبه التام، وكما مبين في الجدول (3) التالي:

جدول (3): تقديرات معاملات نموذج انحدار بواسون بطريقة انحدار Liu

term	$\hat{\beta}_i$	$S(\hat{\beta}_i)$	t	p-value
intercept	0.3197	3.8554	0.083	0.9339
X_{i1}	0.0380	0.0244	1.554	0.1203
X_{i2}	0.0500	0.0071	7.056	1.71e-12 ***
X_{i3}	0.0493	0.0115	4.269	1.97e-05 ***
X_{i4}	0.2418	0.1145	2.112	0.0347 *
X_{i5}	0.1970	0.1248	2.579	0.0144 *
X_{i6}	0.3816	0.1544	2.472	0.0134 *
X_{i7}	0.7310	0.1041	7.023	2.18e-12 ***
AIC = 35.29				
d = 0.1				

من الجدول (3) أعلاه، نلاحظ أن إجمالي عدد الأطفال المصابين بعيوب خلقية في القلب وجهاز الدوران مرهون بمدى الزيادة الحاصلة في كافة متغيرات النموذج على الرغم من عدم معنوية المتغير (X_{i1}) والذي يمثل مجاميع أوزان الأطفال المصابين ضمن كل فترة زمنية. وذلك لأن جميع المعلمات مؤثرة في زيادة أعداد الأطفال المصابين بتلك العيوب الخلقية ولكن بنسب متفاوتة، كما تشير النتائج إلى أن قيمة معيار أكيائي قد بلغت (AIC = 35.92) عند معلمة تحيز (d=0.1). وعند تطبيق طريقة انحدار الحرف (Ridge Regression) لتقدير معاملات نموذج انحدار بواسون عند وجود مشكلة التعدد الخطي شبه التام، كانت النتائج كما في الجدول (4) التالي:

جدول (4): تقديرات معاملات نموذج انحدار بواسون بطريقة انحدار الحرف

term	$\hat{\beta}_i$	$S(\hat{\beta}_i)$	t	p-value
intercept	0.4759	0.1994	2.39	0.0441*
X_{i1}	0.0454	0.0328	1.38	0.2035
X_{i2}	0.0559	0.0166	3.38	0.0097 ***
X_{i3}	-0.0550	0.0206	2.67	0.0285 *
X_{i4}	0.0781	0.2751	0.28	0.7838
X_{i5}	-0.3780	0.3390	1.11	0.2972
X_{i6}	0.5837	0.2945	1.98	0.0828
X_{i7}	0.8351	0.2932	2.85	0.0215 *
AIC = 86.496				
k = 0.12				

من الجدول (4)، نلاحظ أن إجمالي عدد الأطفال المصابين بعيوب خلقية في القلب والدورة الدموية في كل فترة يعتمد على الزيادة الحاصلة في جميع معاملات النموذج. مع ذلك فإن معظم المتغيرات غير معنوية (x_6, x_5, x_4, x_1) بسبب تأثير التعدد الخطي شبه التام، كما تشير النتائج إلى أن قيمة معيار أكيائي قد بلغت (AIC = 86.496) وان قيمة معامل التحيز قد بلغت (k=0.12).

وعند المقارنة بين الطريقتين نلاحظ أن طريقة انحدار ليو (Liu Regression) تعطي قيمة أقل لمعيار معلومات أكيائي، كما تمتلك معاملات أكثر أهمية إذا ما قورنت بطريقة انحدار الحرف (Ridge Regression)، وبهذا تكون معادلة الانحدار المقدر لإجمالي عدد الأطفال المصابون بعيوب خلقية في القلب والدورة الدموية حسب طريقة مقدرات (Liu) كما يلي:

$$\hat{Y}_i = e \left\{ \begin{array}{l} 0.3197 + 0.0380X_{i1} + 0.0500X_{i2} - 0.0493X_{i3} + 0.2418X_{i4} \\ + 0.1970X_{i5} + 0.3816X_{i6} + 0.7310X_{i7} \end{array} \right\} \quad (31)$$

10. الاستنتاجات:

1. من خلال تطبيق طرائق تحليل الانحدار في ظل وجود مشكلة تعدد خطي شبه تام على بيانات حقيقية تم الحصول عليها من مستشفى الطفل المركزي في بغداد تمثل عدد حالات العيوب الخلقية للأطفال في القلب وجهاز الدوران للفترة من (2012 - 2019)، تبين أن طريقة مقدرات ليو هي أكثر كفاءة من طريقة انحدار الحرف كونها تمتلك معيار معلومات أكيائي أقل، كما أنه يعطي نتائج أكثر موثوقية وقيم P أكثر دقة.
2. من خلال طريقة مقدرات ليو، يتضح أن جميع المتغيرات التنبؤية قيد الدراسة مؤثرة في نموذج الانحدار، حتى لو لم تكن ذات دلالة إحصائية، حيث أن جميع المتغيرات مؤثرة في زيادة عدد الأطفال ذوي الإعاقة الخلقية ولكن بنسب متفاوتة.
3. أن تعرض الأمهات إلى إشعاع أو مؤثرات حياتية كتناول الأدوية خلال فترة الحمل، وزيادة أعمار آباء وأمهات الأطفال، تعد من أبرز واهم العوامل مؤثرة في زيادة إصابة الأطفال بالعيوب الخلقية.

المصادر:

- [1] Akdeniz, F., & Erol, H. (2003), "Mean squared error matrix comparisons of some biased estimators in linear regression", *Communications in Statistics-Theory and Methods*, 32(12), 2389-2413.
- [2] Akdeniz, F., Wan, A. T., & Akdeniz, E. (2005), "Generalized Liu type estimators under Zellner's balanced loss function", *Communications in Statistics—Theory and Methods*, 34(8), 1725-1736.
- [3] Arslan, O., & Billor, N. (2000), "Robust Liu estimator for regression based on an M-estimator", *Journal of applied statistics*, 27(1), 39-47.
- [4] Hoerl, A. E., & Kennard, R. W. (1970), "Ridge regression: Biased estimation for nonorthogonal problems", *Technometrics*, 12(1), 55-67.
- [5] Hubert, M. H., & Wijekoon, P. (2006), "Improvement of the Liu estimator in linear regression model", *Statistical Papers*, 47(3), 471-479.
- [6] James, W., & Stein, C. (1992), "Estimation with quadratic loss". In *Breakthroughs in statistics* (pp. 443-460), Springer, New York, NY.
- [7] Kejian, L. (1993), "A new class of biased estimate in linear regression", *Communications in Statistics-Theory and Methods*, 22(2), 393-402.
- [8] Mardikyan, S., & Çetin, E. (2008), "Efficient choice of biasing constant for ridge regression", *Int. J. Contemp. Math. Sciences*, 3(11), 527-536.
- [9] Maronna, R. A., Martin, R. D., Yohai, V. J., & Salibián-Barrera, M. (2019), "Robust statistics: theory and methods (with R)", John Wiley & Sons.
- [10] Mansson, K & Shukur, G (2011), "A Poisson Ridge Regression Estimator", *Economic Modeling*, Vol. 28, Issue. 4, pp. 1475-1491.
- [11] Yang, H., & Xu, J. (2009), "An alternative stochastic restricted Liu estimator in linear regression", *Statistical Papers*, 50(3), 639-647.



Comparison Between Estimators Leu Regression Method and Ridge Regression Method of the Poisson Regression Model in The Presence of Multicollinearity Problem

Raaed F. Mohammed Al-Hasani

Dep. of Accounting, College of Administration and Economics, Mustansiriya University, Baghdad, Iraq.

Email: raad@uomustansiriyah.edu.iq, ORCID ID: <https://orcid.org/0000-0002-7309-8702>

Article Information

Article History:

Received: 1 / 2 / 2024

Accepted : 10 / 6 / 2024

Available Online: 25 / 7 / 2024

Page no : 36 – 45

Keywords:

Poisson regression , Liu estimators , Ridge regression , Multicollinearity problem .

Correspondence:

Researcher name Raaed F.

Mohammed Al-Hasani:

Email:

raad@uomustansiriyah.edu.iq

Abstract

The Poisson regression model is one of the most essential linear logarithmic regression models, and it is the tool through which the dependent variable is ratios when its values are positive and in the form of percentages or countable data, as well as the fitted model for analyzing the rare events. Like many other regression models, the predictor variables included in its construction may be exposed to a high correlation between two or more variables, negatively affecting the estimation of the model parameters. In this paper, we will review the most prominent methods for estimating parameters of the Poisson regression model when data suffers from a semi-multicollinearity problem, such as Ridge regression and Liu estimator's method. Estimation methods were applied to real data obtained from Central Child Hospital in Baghdad, representing the number of cases of congenital disabilities of children in the heart and circulatory system from 2012 to 2019. The results showed the superiority of the Liu estimators' method over the ridge regression method based on (AIC) as a criterion for comparison.