Warith Scientific Journal

توظيف اسلوب (Kernel) في تقدير انموذج انحدار بواسون مع تطبيق عملي

Employing the Kernel method in estimating the Poisson regression model with application

م.م مها حسن سلطان
Maha Hassan Sultan
جامعة البيان- كلية ادارة الاعمال ـ قسم ادارة الاعمال
Al-Bayan University- College of Business Administration
Business Administration Department
Maha.h@albayan.edu.iq

المستخلص

شكل علم الاحصاء يداً ممتدة تقدم خدماتها عبر الزمن الى العلوم الأخرى ورافداً يساهم في تحليل نتائج ما تتوصل اليه الدراسات في مختلف المجالات وحقول المعرفة, و يعد انموذج انحدار بواسون من النماذج الاحصائية الحيوية لتمثيل الظواهر الطبيعية , ويوصف اسلوب (kernel) من اهم الاساليب الامعلمية التي تستعمل لتقدير معالمه . لذا تضمنت الدراسة استعمال أنموذج الانحدار بواسون في صياغة العلاقات التي يكون فيها متغير الاستجابة ذو قيم معدودة بخطأ عشوائي يتبع توزيع ذاته هو وتقدير المعلمات بطريقة اسلوب (kernel) واختبار جودة التوفيق الانموذج بشكل عام من خلال اقتراح اختبار يعتمد بشكل اساسي على توزيع فيشر . وتم تطبيق ذلك على البيانات الطبية كمجال للتطبيق العملي, والمتعلقة بالأفراد الذين تعرضوا للجلطة القلبية بمتغيرين احدهما يمثل متغير الاستجابة (Y) والذي يعبر عن عدد مرات تعرض المريض للجلطة القلبية والمتغير الاخر (X) الذي يعبر عن عمر المريض كمتغير مستقل . و بينت النتائج ان الأنموذج بواسون مثل البيانات بصوره صحيح من خلال اجتياز الأنموذج المقدر لاختبار جودة التوفيق فضلاً عن نتائج المعلمات المقدرة ظهرت ذات دلالة احصائية معنوية , على ضوء قيم معايير المفاضلة التي افرزتها نتائج التقدير لنماذج انحدار بواسون الخاصة بعينات التطبيق مما يشير الى وجود تأثير معنوي موجب لعمر المريض على العدد المتوقع لمرات تعرضه للجلطة القلبية مما يظهر ان لعمر المريض اثراً كبيراً في عدد مرات اصابته بالجلطة الدماغية للمرضى الراقدين في مستشفى ابن النفيس في محافظة بغداد لعام 2024.

الكلمات المفتاحية : انموذج انحدار بواسون , طريقة انحدار النواة , مقاييس جودة التوفيق

Abstract

Statistics has been an extended hand that provides its services over time to other sciences and a tributary that contributes to analyzing the results of studies in various fields and areas of knowledge. The Poisson regression model is one of the vital statistical models to represent natural phenomena, and the (kernel) method is described as one of the most important parametric methods used to estimate its parameters. Therefore, the study included the use of the Poisson regression model in formulating relationships in which the response variable has numerical values with a random error that follows its own distribution and estimating the parameters using the (kernel) method and testing the goodness of fit of the model in general by proposing a test that relies primarily on the Fisher distribution. This was applied to medical data as a field of practical application, related to individuals who were exposed to a heart attack with two variables, one of which represents the response variable (Y), which expresses the number of times the patient was exposed to a heart attack, and the other variable (X), which expresses the patient's age as an independent variable. The results showed that the Poisson model represented the data correctly by passing the estimated model to the goodness of fit test, in addition to the results of the estimated parameters that appeared to be statistically significant, in light of the values of the comparison criteria that resulted from the estimation results of the Poisson regression models for the application samples, which indicates the presence of a positive significant effect of the patient's age on the expected number of times he is exposed to a heart attack, which shows that the patient's age has a significant effect on the number of times he is exposed to a stroke for patients hospitalized in Ibn Al-Nafis Hospital in Baghdad Governorate for the year 2024.

Keywords: Poisson Regression Model , Kernel regression Method Goodness of fit Measures,

1-منهجية البحث

1-1 المقدمة

يهتم الكثير من الباحثين عند تحليل مجموعة من المتغيرات الى الوصول الى طبيعة العلاقة التي تربط بين هذه المتغيرات , ويعتبر تحليل الانحدار واحد من اهم المواضيع التي تدرس شكل العلاقة بين متغير واحد يسمى المتغير التابع و واحد او اكثر من المتغيرات التي تسمى المتغيرات المستقلة . ويمكن وصف أنموذج انحدار بواسون بانه احد صيغ نماذج الانحدار غير الخطية والذي يستعمل في صياغة طبيعة العلاقة بين متغير الاستجابة ذو القيم المعدودة countable data وعدة متغيرات توضيحية , اذ يفترض هذا الانموذج ان متغير الاستجابة يتبع توزيع بواسون بمتوسط وتباين متساويان كما ويفترض ان لوغاريتم القيمة المتوقعة لمتغير الاستجابة يمكن ان صياغتها على انها تركيب خطي بعدد من المعلمات غير المعلومة , لذا يسمى انموذج انحدار بواسون احيانا بالأنموذج اللوغاريتمي الخطي Log-Linear model . لنموذج انحدار بواسون تطبيقات واسعة في الاحصاء و الاقتصاد القياسي, فهناك الكثير من الامثلة على تطبيقاته, كنمذجة اعداد (او معدلات) الاصابات بالأمراض, و اصابات الحوادث المرورية, و التغيب عن العمل, و الرحلات الترفيهية او التسويقية, و التأمين على السيارات, و انتقال العمالة, و غيرها من المتغيرات ذات القيم القابلة للعد.

1- 2 هدف البحث

يهدف البحث إلى اعتماد طريقة تقدير (Kernel Regression Estimated method) الامعلمية بغية تقدير معلمات انموذج انحدار بواسون بالاعتماد على معايير المفاضلة لدقة التقدير ومن ثم تطبيقها على بيانات حقيقية والمتمثلة بالأفراد الذين تعـــرضوا للجلطة القلبية.

1- 3 مشكلة البحث

تكمن مشـــكلة البحث حول عدم امكانية الاعتماد على طريقه المربعات الصـــغرى الاعتيادية (ols) لتقدير معالم انموذج انحدار بواسـون المسـتخدم في مزج العلاقة بين متغير الاسـتجابة ذات القيم المعدودة وبين متغيرا او اكثر من المتغيرات المسـتقلة, وذلك نتيجة لعدم تحقق الفروض الخاصـــة بتلك الطريقة مما يتطلب البحث حول ايجاد طرائق تقدير بديله وسـيتم التركيز على الطرق الا معلمية والتي يمكن اعتمادها في تقدير معلمات ذلك الانموذج , ومن ثم تطبيقها على بيانات حقيقية والمتمثلة بالأفراد الذين تعرضوا للجلطة القلبية.

1-4:فرضية البحث:

تتلخص فرضية البحث بالآتي:

تعد الطريقة الامعلمية (انحدار النواة) بديل جيد لطريقة الامكان الاعظم ,التي تعد من الطرائق المعلمية, وذات مرونة عالية في التطبيق .

2- الجانب النظري

1-2 إنموذج إنحدار بواسون:Poisson Regression Model

إنحدار بواسون هو أحد أشكال تحليل الانحدار والمستعمل في نمذجة البيانات المصنفة (Countable Data)، إذ يعد الانموذج أداة مناسبة لتحليل الأحداث النادرة الوقوع، يفترض هذا الانموذج أن متغير الاستجابة(Y_i) يتبع توزيع بواسون بمتوسط وتباين يساوي (μ_i) كما ويفترض أن لوغاريتم القيمة المتوقعة لمتغير الاستجابة يمكن نمذجتها على أنها تركيبة خطية بعدد من المعلمات غير المعلومة، لذا يسمى أنموذج انحدار بواسون أحياناً بالأنموذج اللوغارتمي – الخطي Log – Linear Model، [1].

لنفرض عينة بحجم n من المشاهدات ($y_1, y_2, \dots, y_n$)التي تمثل متغير الاستجابة والذي يتوزع وفق توزيع بواسون بمعلمة التوزيع (μ) وعلى افتراض أن معلمة التوزيع التي تمثل الوسط الحسابي لقيم متغير الاستجابة (وكذلك تباينها) بأنها تعتمد على متجه من المتغيرات المستقلة X فإذا تم الاعتماد على أنموذج الإنحدار الخطي في وصف تلك العلاقة فإن الأنموذج يكون كالآتي:

$$E(Y_i) = \mu_i = \underline{X}_i'\beta \qquad \dots\dots\dots\dots (1)$$

وعلى وفق هذه الصيغة الخطية فإن المتغيرات التوضيحية تقترح أية قيمة حقيقية لمتوسط متغير الاستجابة Y غير أن ذلك يتناقض مع ما تمتاز به معلمة توزيع بواسون (μ) إذ أنها تأخذ قيماً غير سالبة. ولمعالجة هذا الإخفاق يمكن نمذجة علاقة خطية بين لوغاريتم المتوسط (Logμ) والمتغيرات التوضيحية (X) ويتم ذلك من خلال تطبيق الأنموذج الخطي المعمم. باستخدام دالة اللوغاريتم كدالة ربط بمعنى آخر اعتماد الأنموذج الخطي – اللوغارتمي على وفق الصيغة الآتية: [9]

$$\mathrm{Log}\big(E(Y_i)\big) = \mathrm{Log}(\mu_i) = \underline{X}_i'\beta \qquad \dots\dots\dots\dots (2)$$

إذ أن:

β : هو متجه المعلمات ذو السعة (1*(k+1)).المطلوب تقديرها

$\underline{X}_i$: متجه المتغيرات التوضيحية ذو السعة ((k+1)**1) بما في ذلك الحد الثابت.

الصيغة (2) تمثل الصيغة العامة لأنموذج إنحدار بواسون والفكرة الأساس لأنموذج إنحدار بواسون تقوم على ربط الدالة الاحتمالية لمتغير الاستجابة y_i، $i = 1, 2, \dots, n$(إذ أن n تمثل عدد مشاهدات العينة) بمجموعة من المتغيرات التوضيحية (X_j)، $j = 0,1, 2, \dots, k$.

وبموجب هذا الأنموذج فإن معاملات الإنحدار β_jتمثل التغير المتوقع في لوغاريتم المتوسط لمتغير الاستجابة Y لكل وحدة تغير حاصلة في المتغير التوضيحي X_jوبأخذ الدالة الأسية لطرفي الإنموذج (2 – 2) نحصل على الأنموذج المضاعف لمتوسط متغير الاستجابة على وفق الصيغة الآتية:

$$\mu_i = \exp(\underline{X}_i'\beta) \quad \dots\dots\dots\dots (3)$$

وبتالي فإن معامل الإنحدار الأسي exponential coefficient regression يمثل التأثير المضاعف للمتغير التوضيحي X_j على متوسط متغير الاستجابة Y فزيادة المتغير التوضيحي X_j بوحدة واحدة يضاعف متوسط متغير الاستجابة Y بمقدار $\exp(\beta_j)$.

2-2 : تقدير معلمات الأنموذج[7][11][5] Estimating model parameters

لتقدير معالم هذا الأنموذج نستخدم طريقة انحدار النواه (Kernel Regression Estimated method), على فرض أن الخطأ هو يتبع توزيع بواسون وبهذا يمكن القول أن مشاهدات المتغير التابع هي الأخرى تتبع توزيع بواسون , وبشكل عام ولعينة حجمها (n) يأخذ المتغير التابع (y) المشاهدات التالية (y_1, y_2, …… y_n) وأن دالة الكثافة الاحتمالية (p.d.f) لكل مشاهدة من مشاهدات المتغير المعتمد لهذه العينة يمكن أن توضع بدلالة المعالم المقدرة وكما يأتي :

2- 2-1 : طريقة إنحدار النواة: Kernel regression Method

يتم اخذ لوغاريتم دالة الإمكان لأنموذج إنحدار (Poisson) لنحصل على دالة الامكان السالبة وتكون كالآتي :

$$\ell(\mu) = \frac{1}{n}\{e^{\eta(x_i)} - y_i\eta(x_i)\} \quad \dots\dots\dots\dots (4)$$

إذ أن $\eta(x_i) = \log(\mu(x_i))$. تقدير النواة غير الخطي لا نموذج انحدار بواسون يتلخص بتقدير متجه معلمات الانموذج $\eta(x_i)$ وباستخدام الانموذج الخطي الآتي[11]، [8] :

$$\eta(x_i) = b_0 + w^T\phi(x_i) \quad \dots\dots\dots\dots (5)$$

إذ أن$\phi^T(x_i)\phi(x_l) = K(x_i, x_l)$, وأن $K(x_i, x_l)$تمثل احدى دوال النواة.

تقدير متجه معلمات الانموذج$\eta(x_i)$نحصل عليه بتصغير لوغاريتم دالة الإمكان الجزائية السالبة (Penalized negative log-Likelihood)المبينة أدناه:

$$\ell(w, b) = \sum_{i=1}^{n}\left[e^{b_0+w^T\phi(x_i)} - y_i[b_0 + w^T\phi(x_i)]\right] + \frac{\lambda}{2}\|w\|^2$$

$$\dots\dots\dots\dots (6)$$

إذ أن λ تمثل معلمة الجزاء (Penalty parameter) والتي تسيطر على الموازنة بين جوده حسن المطابقه (goodness of fit) بالاعتماد على البيانات ومقدار التمهيد (Smoothness)$\|w\|^2$ للدالة $\eta(x_i)$.الذي يجعل لوغاريتم دالة الإمكان الجزائية السالبة في نهايتها الصغرى يكون متجه معلمات الأنموذج $\eta(x_i)$ وفق الصيغة الآتية:

$$\eta(x_i) = b_0 + k_i^T\alpha \quad \dots\dots\dots\dots (7)$$

إذ أن:

k_i هو العمود i لمصفوفة النواة K ذات البعد (n*n) وبعناصر $k(x_i, x_l)$ ،$i, l = 1,2, \dots, n$

α : هو متجه معلمات التمهيد ذو السعة(n*1).

b_0 : الحد الثابت.

وبهذا فإن صيغة لوغاريتم دالة الإمكان الجزائية السالبة تكون كالآتي: [6]

$$\ell(b_0, \alpha) = \sum_{i=1}^{n} \left[e^{b_0 + k_i^T \alpha} - y_i [b_0 + k_i^T \alpha] \right] + \frac{\lambda}{2} \alpha^T K \alpha \ldots (8)$$

وللحصـــول على مقدرات الإمكان الأعظم لنماذج الانحدار غير الخطية والتي يكون فيها متغير الاســتجابة ذو توزيع بواســون يمكن اعتماد طرائق التقدير العددية كطريقة نيوتن – رافسون أو طريقة المربعات الصـغرى للأوزان المعادة ذات المراحل المتعددة Iterative reweighted least squares (IRWLS) والتي تعـد أســهـل من الطريقـة الأولى في اشــتقـاق معيـار العبور الشـرعي(Cross – validation criterion (C.V وأن لها معدل تقارب أقل من معدل التقارب لطريقة نيوتن – رافسـون، [6]، [14].

ولإيجاد مقدرات الإمكان الأعظم للمعلمات b_0 , $\boldsymbol{\alpha}$ نجد أولاً المشتقات الجزئية للوغاريتم دالة الإمكان الجزائية السالبة بالنسبة لـ b_0 , $\boldsymbol{\alpha}$ ومساواتها بالصفر، أي إيجاد:

$$\left.\begin{array}{l} \frac{\partial \ell(b_0,\alpha)}{\partial b_0} = 0 \\ \frac{\partial \ell(b_0,\alpha)}{\partial \alpha} = O_n \end{array}\right\} \quad \ldots\ldots\ldots\ldots (9)$$

ثم إيجاد الحل لمجموعة المعادلات الآتية باستخدام طريقة (IRWLS):

$$(V^T W V + U)\beta = V^T W y^* \quad \ldots\ldots\ldots\ldots (10)$$

إذ أن:

V : مصفوفة ذات بعد n*(n+1)والمعرفة كالآتي:

$$V = (1_n, K) \quad \ldots\ldots\ldots\ldots (11)$$

1_n : تمثل متجه الواحد الصحيح ذو البعد (n*1) و K هي مصفوفة النواة كما تم تعريفها سابقاً.

W : مصفوفة قطرية ذات بعد (n*n) والمعرفة كالآتي:

$$W = \operatorname{diag}\left(\exp(b_0 + k_1^T \alpha), \ldots., \exp(b_0 + k_n^T \alpha)\right) \ldots\ldots\ldots\ldots (12)$$

β: تمثل متجه المعلمات التي نريد تقديرها ذو البعد 1*(n+1) والمعرف كالآتي:

$$\beta = \left(b_0, \alpha^T\right)^T \quad \ldots\ldots\ldots\ldots (13)$$

U: هي مصفوفة ذات بعد (n+1)*(n+1)والمعرفة وفق الآتي:

$$U = \begin{bmatrix} O & O_n^T \\ O_n & \lambda K \end{bmatrix} \quad \ldots\ldots\ldots\ldots (14)$$

إذ أن O_n هو متجه صفري ذوالسعة (n*1).

$y^* = (y_1^*, \ldots., y_n^*)$: يمثل متجه الاســتجابة الخاص بطريقة النواة لتقدير انموذج انحدار بواســون ذو الســعة(n*1)، و أن y_i^* تكون كالآتي:

$$y_i^* = \frac{y_i - \exp(b_0 + k_i^T \alpha)}{\exp(b_0 + k_i^T \alpha)} + (b_0 + k_i^T \alpha) \quad \ldots\ldots\ldots\ldots (15)$$

في حالة اســـتخدام طريقة نيوتن – رافســـون فإن تقدير متجه المعلمات αوالمعلمة b_0، والتي تحدث قيمها التقديرية بالمراحل المتعددة للطريقة، تكون كالآتي:

$$\begin{pmatrix}\alpha^{(t+1)}\\ b_0^{(t+1)}\end{pmatrix}=\begin{pmatrix}\alpha^{(t)}\\ b_0^{(t)}\end{pmatrix}-\begin{pmatrix}Kw^{(t)}K+\lambda K & K\mu^{(t)}\\ \mu^{(t)\prime}K & 1'\mu^{(t)}\end{pmatrix}^{-1}\begin{pmatrix}K\mu^{(t)}-y+\lambda\alpha^{(t)}\\ 1'(\mu^{(t)}-y)\end{pmatrix}$$

..... (16)

مقدر دالة المتوسط عند متجه المدخلات x_0 وللقيم المثلى التقديرية لمتجه المعلماتα والمعلمةb_0 تكون كالآتي:

$$\hat{\mu}(x_0)=\exp\left(\hat{b}_0+k_0^T\hat{\alpha}\right) \quad \text{..... (17)}$$

إذ أن:

$$k_0=\left(k(x_1,x_0),\dots\dots,k(x_n,x_0)\right)^T \quad \text{..... (18)}$$

لاختيار معلمة الجزاء (λ) ومعلمة الدالة اللبيه σ^2 التي تؤدي إلى أفضل تمهيد يمكن استخدام معيار العبور الشرعي الاعتيادي ordinary cross – validation (OCV) وفق الصيغة الآتية: [3]

$$OCV(\theta)=\frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i-e^{\hat{\eta}_\theta(x_i)}}{(1-S_{ii})}\right)^2 \quad \text{..... (19)}$$

إذ أن S_{ii} هو العنصر القطري(i) للمصفوفة S والمعرفة كالآتي:

$$S=V(V^TWV+U)^{-1}V^TW \quad \text{..... (20)}$$

والتي تمثل مصفوفة التمهيد.

وبالتعويض عن S_{ii} بمتوسطها الذي يساوي $(\frac{1}{n}tr(S))$ في صيغة $OCV(\theta)$ فإننا نحصل على معيار العبور الشرعي المعمم Generalized cross – validation GCV)والذي يحسب لـ θ من المعلمات وفق الصيغة الآتية:

$$GCV(\theta)=\frac{n\sum_{i=1}^{n}\left(y_i-e^{\hat{\eta}_\theta(x_i)}\right)^2}{\left(n-tr(S)\right)^2} \quad \text{..... (21)}$$

إذ أن:

$\hat{\eta}_\theta(x_i)$: هي تقدير لـ $\left(\eta_\theta(x_i)=\log\mu_\theta(x_i)\right)$ لكل معلمة بعد استبعاد المشاهدة (i) وفق طريقة (Leave – one – out) ،

معلمة التمهيد تعمل على السيطرة بين دقة التقدير و مقدار التمهيد الذي يساوي $(\theta^TP\theta)$ إذ ان:

$$\theta^TP\theta=\sum_{k=1}^{K}(\Delta^2\theta_k)^2 \quad \text{..... (22)}$$

إذ ان :

$$\Delta^2\theta_k=\theta_k-2\theta_{k-1}+\theta_{k-2} \quad \text{..... (23)}$$

وأن:

$$P=D_2^TD_2 \quad \text{.....(24)}$$

و ان D_2 تمثل مصفوفة عامل الفروق Δ^2ذات السعة (K-2)*K [3]، [4].

3-2: اختبار الأنموذج بشكل عام [3][11][7] General model test

بعد تحديد الأنموذج وتقدير معالمه تأتي مرحلة تقييم الأنموذج المقدر واختباره لبيان مدى مطابقته و ملائمته للبيانات ويتم ذلك من خلال تحليل البواقي التي يفرزها الأنموذج المقدر واعتماد اختبارات حسن المطابقة.

2-3-1 **مقاييس حسن المطابقة:** Goodness of fit Measures

تعد مقاييس حسن المطابقة الشائعة الاستخدام بالنسبة لنماذج الخطية (المعلمية ولامعلمية) هما الإحصائيتان بيرسون **Pearson** والإنحراف **deviance** والتي هي عبارة عن مجموع موزون للبواقي.

1. إحصاءة بيرسون: Pearson Statistic

تعد إحصاءه بيرسون احد مقاييس حسن المطابقة لأي إنموذج لمتغير الاستجابة (y_i) بمتوسط μ_i وتباين w_i، [9],[7],[13]:

$$\mathbf{P} = \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{\hat{w}_i} \qquad \ldots\ldots\ldots\ldots (25)$$

إذ أن $\hat{\mu}_i$ و $\hat{w}_i$ هما تقدير كل منμ_i و w_i على التتابع. في حالة كون المتوسط والتباين محددين بشكل صحيح فإن:

$$E\left[\sum_{i=1}^{n} \frac{(y_i - \mu_i)^2}{w_i}\right] = n \qquad \ldots\ldots\ldots\ldots (26)$$

وذلك لأن $E((y_i - \mu_i)^2/w_i) = 1$, ولكن عملياً إحصاءه P تقارن مع (n-k) وذلك لتصحيح الفقدان في درجات الحرية نتيجة لتقدير μ_i.

بالنسبة للأنموذج انحدار بواسون فإن $(w_i = \mu_i)$ لذا فإن إحصاءة بيرسون تكون على وفق الصيغة الآتية:

$$P_p = \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i} \qquad \ldots\ldots\ldots\ldots (27)$$

2. إحصاءه الانحراف Deviance statistic

تعد إحصاءه الانحراف أحد مقاييس حسن المطابقة الخاصة بالنماذج الخطية العامة والتي تحسب وفق الصيغة الآتية:

$$D(y, \hat{\mu}) = 2[\ell(y) - \ell(\hat{\mu})] \qquad \ldots\ldots\ldots\ldots (28)$$

إذ أن $D(y, \hat{\mu})$ تمثل إحصاءه الانحراف و $\ell(\hat{\mu})$ تمثل لوغاريتم دالة الإمكان للأنموذج المقدر $\left(\ell(\hat{\mu}) = \text{LogL}(\hat{\mu})\right)$ و $\ell(y)$ تمثل لوغاريتم دالة الإمكان الأعظم المتحققة Maximum Log –Likelihood achievable، y و $\hat{\mu}$ هما متجهين كل منهما ذو سعة $(n \times 1)$ بمدخلات y_i و $\hat{\mu}_i$ على التتابع, [13]، [7].

اما بالنسبة للأنموذج انحدار بواسون فإن صيغة إحصاءه الانحراف تكون على وفق الصيغة الآتية:

$$Dev_p = \sum_{i=1}^{n} \left[y_i \text{Log}\left(\frac{y_i}{\hat{\mu}_i}\right) - (y_i - \hat{\mu}_i)\right] \qquad \ldots\ldots\ldots\ldots (29)$$

3. مقياس معامل التفسيرالزائف: Pseudo R- square Measures

في نماذج الإنحدار الخطية يستخدم مقياس معامل التفسير R- square في تحديد مدى قدرة الإنموذج في تفسير التغيرات الحاصلة في متغير الاستجابة Y غير أن هذه الميزة لمعامل التفسير تكون غير متوفرة بالنسبة لنماذج الإنحدار غير الخطية.

في نماذج الإنحدار الخطية يتم الحصول على معامل التفسير من خلال تحليل مربعات الإنحرافات الكلية إذ أن: [13]

$$\sum_{i=1}^{n} (y_i - \overline{Y})^2 = \sum_{i=1}^{n} (y_i - \hat{\mu}_i)^2 + \sum_{i=1}^{n} (\hat{\mu}_i - \overline{Y})^2 + 2\sum_{i=1}^{n} (y_i - \hat{\mu}_i)(\hat{\mu}_i - \overline{Y})$$

$$\ldots\ldots\ldots\ldots (30)$$

إذ أن:

$\sum_{i=1}^{n}(y_i - \overline{Y})^2$: يمثل مجموع مربعات الإنحرافات الكليه (TSS).

$\sum_{i=1}^{n}(y_i - \hat{\mu}_i)^2$: يمثل مجموع مربعات البواقي (RSS).

$\sum_{i=1}^{n}(\hat{\mu}_i - \overline{Y})^2$: يمثل مجموع مربعات الإنحرافات المفسرة (ESS).

غير أن العلاقة (27) تكون غير متحققة (أي لا تساوي صفراً) في حالة استخدام طرائق التقدير الأخرى وكذلك بالنسبة لنماذج الإنحدار غير الخطية، بما في ذلك أنموذج إنحدار بواسون لذا فإن الميزة المرغوبة في مقياس معامل التفسير تكون غير متوفرة في نماذج الإنحدار غير الخطية، الأمر الذي يدعو إلى استخدام مقاييس أخرى في حساب R^2 , بتالي فإن معامل التفسير الزائف يكون وفق الصيغة الآتية: [4]:

$$R^2_{DEV,p} = \frac{\sum_{i=1}^{n}\left[y_i Log\left(\frac{\hat{\mu}_i}{\overline{y}}\right) - (y_i - \hat{\mu}_i)\right]}{\sum_{i=1}^{n} y_i Log\left(\frac{y_i}{\overline{y}}\right)} \quad \ldots\ldots\ldots\ldots (31)$$

3- الجانب العملي

تم الاعتماد على بيانات حقيقية لتطبيق انموذج انحدار المستخدم بالاعتماد على طريقة تقدير انحدار النواة والوصول الى النتائج الدقيقة المستندة على اسس علمية ومن ثم تطبيقها على البرامج الحاسوبية .

3-1 وصف البيانات Data description

تم الاعتماد على البيانات الطبية كمجال للتطبيق العملي, والمتعلقة بالأفراد الذين تعرضوا للجلطة القلبية. اذ تم اختيار عينة عشوائية بسيطة بلغت (n=100) جمعت من المرضى الراقدين في مستشفى الكندي في محافظة بغداد من شهر (اب)الى شهر(ايلول) لعام 2024 اذ تم تسجيل عدد مرات تعرض المريض للجلطة القلبية وعمره, اذ مثل عدد مرات تعرض المريض للجلطة القلبية بالمتغير (Y) وعمر المريض بالمتغير (X) .تم تسجيل البيانات من الطبلات الخاصة بالمرضى الراقدين في المستشفى مجال البحث [3] [13] .

3-2:صياغة العلاقة بموجب انموذج انحدار بواسون:

بالاعتماد على انموذج انحدار بواسون وباستخدام الانموذج الامعلمي باستخدام دالة اللوغاريتم كدالة ربط ، وعلى افتراض ان متغير الاستجابة (عدد مرات تعرض المريض للجلطات القلبية) يتبع توزيع بواسون وذلك لكونه متغيرا ذو قيما معدودة , يمكن صياغة علاقة خطية بين لوغاريتم القيمة المتوقعة لعدد مرات تعرض المريض للجلطة القلبية E(Y) وعمر المريض X كمتغير توضيحي وكالاتي:

$$Log(E(y_i)) = Log(\mu_i) = \beta_0 + \beta_1 x_i \quad \ldots\ldots\ldots\ldots (1)$$

3-3: تقدير انموذج انحدار بواسون:

بالاعتماد على طريقة تقدير انحدار النواة والخاصة بموضوع البحث سيتم تقدير انموذج انحدار(Poisson)لعينة تطبيق الدراسة

3-3-1:التقدير بموجب Kernel regression Method

بالاعتماد على طريقة انحدار النواة التي تم توضيحها في الجانب النظري تم تقدير معلمات انموذج انحدار بواسون المبين بالصيغة (21) لعينة التطبيق, تم تقدير النماذج باستخدام حزمة البرامج الاحصائية SAS والبرنامج الاحصائيSPSS version 20, نتائج التقديرات مبينة في الجدول (3-1), وبملاحظة تلك النتائج نجد ان النماذج المقدرة معنوية اذ ان قيمة المؤشر الاحصائي Chi-square ذو قيمة معنوية ولمستوى معنوية (0.05) لعينة البحث, فقد كان معنويا عند مستوى معنوية (0.10), مما يدل على وجود تأثير معنوي للمتغير التوضيحي على متغير الاستجابة . ان معلمة التمهيد المثلى تم تحديدها بالاعتماد معيار العبور الشرعي المعمم GCV, قيمة إحصاءه الانحراف المحسوبة للنموذج المقدرة تشير الى ان بيانات العينة تعاني من مشكلة تحت التشتت اذ كانت اقل من عدد درجات الحرية, القدرة التفسيرية للنموذج المقدر والتي تعكسها قيمة معامل التفسير الزائف (Pseudo R- square) كانت تشير الى قدرة تفسيرية جيدة اذ كانت تتراوح بين (0.55) و (0.59), القيم التقديرية للعدد المتوقع لمرات تعرض المريض للجلطة القلبية لعينة التطبيق البحث .

جدول (3-1): نتائج تقديرات طريقة انحدار النواة لانموذج انحدار بواسون لعينة التطبيق .

التطبيق	العينة	Chi-Square	Sig	معلمة التمهيد	GCV	Deviance	MSE	R^2_{Dev}
n=100	1	8.0701	0.005	0.3600	0.00282	9.3200	0.3204	0.5959
	2	9.4509	0.02	0.5100	0.00173	8.9750	0.1504	0.6265
	3	12.2415	0.01	0.3100	0.00274	8.0442	0.1579	0.8449
	4	7.4849	0.07	0.3900	0.00369	9.7567	0.2124	0.5530

2-3-3 النتائج Results

تم استعمال لغة البرمجة (R) و كذلك البرنامج المعروف مايكروسوفت – اكسل (Microsoft Excel) في استخراج النتائج التي تم الحصول عليها من تقدير معلمات وكذلك الاختبارات التي تخص أنموذج انحدار بواسون بخطأ يتبع توزيع ذاته وسيتم مناقشتها مع تبويبها في جداول ليسهل تحليلها وفهمها وهي كما يلي .

بالاعتماد على طريقة انحدار النواة التي تم توضيحها في الجانب النظري تم تقدير معلمات انموذج انحدار بواسون المبين بالصيغة (1-3) الخاصة بالعينة المستعملة, تم تقدير النموذج باستخدام حزمة لغة البرمجة (R)والبرنامج الاحصائيSPSS version 20, تم الحصول على نتائج التقديرات مبينة في الجدول (1-3), وبملاحظة تلك النتائج نجد ان النموذج المقدر ذو دلالة احصائية اذ ان قيمة المؤشر الاحصائي Chi-square ذو قيمة معنوية ولمستوى معنوية (0.05) لكل عينة من عينات التطبيق, عدا الانموذج المقدر للعينة الرابعة للتطبيق الاول فقد كان معنويا عند مستوى معنوية (0.10), مما يدل على وجود تأثير معنوي للمتغير التوضيحي على متغير الاستجابة . ان معلمة التمهيد المثلى تم تحديدها بالاعتماد معيار العبور الشرعي المعمم GCV, قيمة إحصاءه الانحراف المحسوبة للنماذج المقدرة تشير الى ان بيانات العينات جميعها تعاني من مشكلة تحت التشتت اذ كانت اقل من عدد درجات الحرية, القدرة التفسيرية للنماذج المقدرة والتي تعكسها قيمة معامل التفسير الزائف (Pseudo R- square) للنموذج المقدر كانت تشير الى قدرة تفسيرية جيدة اذ كانت تتراوح بين (0.55) و (0.84), القيم التقديرية للعدد المتوقع لمرات تعرض المريض للجلطة القلبية لعينات التطبيق .

3-3-3 : اختبار جودة التوفيق للأنموذج المقدرة Quality-of-fit test for estimated model

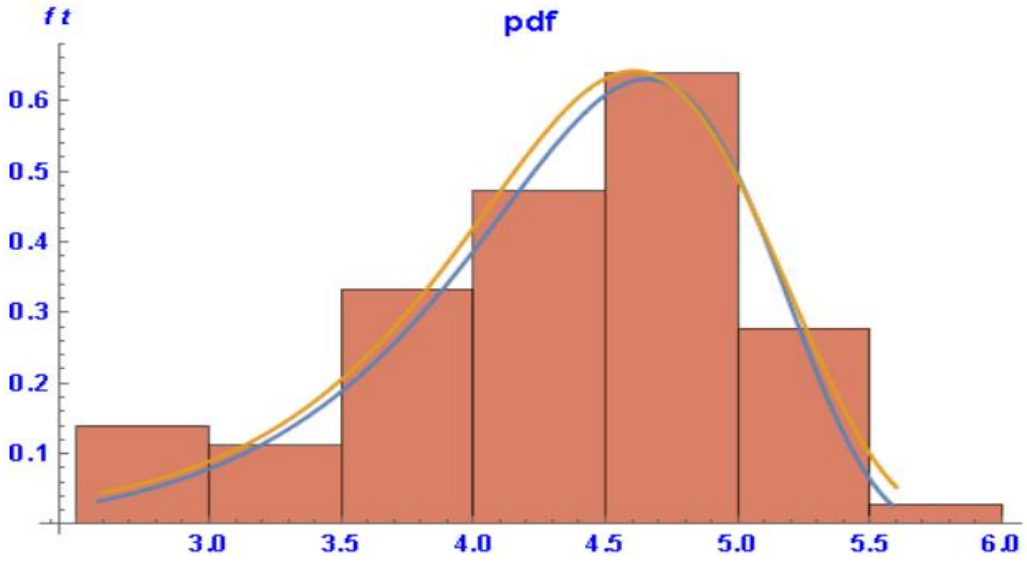
لاختبار العلاقات المقدرة للأنموذج الذي تم تقديره سنختبر الفرضية والتي ينص فرض العدم منها على أن العلاقة المقدرة غير معنوية أما الفرض البديل فينص على أن العلاقة المقدرة معنوية و تم حساب المقادير الضرورية لأجراء الاختبار وهي كما موضحة في الجدول التالي :

الجدول (2-3)

اختبار جودة التوفيق لأنموذج انحدار بواسون المقدر

SS_E	MS_E	SS_R	MS_R	Fالمحسوبة	Fالجدولية
42.17	0.346	44.3	44.3	128.03	3.92

نلاحظ من الجدول اعلاه أن قيمة F المحسوبة اكبر من قيمة F الجدولية بمستوى ثقة مقداره (0.95) ودرجة حرية للبسط مقدارها واحد وللمقام مقدارها (122) فيكون القرار هو رفض فرضية العدم أي أن العلاقة المقدرة من أنموذج انحدار المقدر اعتماداً على طريقة انحدار النواه هي علاقة معنوية وممثلة للبيانات بصوره صحيحة , ويمكن تمثيلها على وفق الشكل البياني الاتي:



المصدر من اعداد الباحثة بالاعتماد على مخرجات الحاسبة

4- الاستنتاجات والتوصيات

من خلال تطبيق طريقة التقدير للمعلمات و اختبار جودة التوفيق للأنموذج يمكن الإشارة الى بعض الاستنتاجات التي توصلت اليها الباحثة والتوصيات التي يعتقد انها ضرورية.

4-1 الاستنتاجات

1- يعد انموذج انحدار بواسون المقدرة باعتماد طريقة انحدار النواة والمستخدمة في تمثيل العلاقة بين عدد مرات تعرض المريض للجلطة القلبية وعمر المريض كانت معنوية لجميع عينات الدراسة وعند مستوى معنويه 0.05 , اذ كانت المعلمات المقدرة ذات قيما معنوية مما يشير الى وجود تأثير معنوي موجب لعمر المريض على العدد المتوقع لمرات تعرضه للجلطة القلبية.

2- تعد طريقة انحدار النواة طريقة كفؤة على ضوء قيم معايير المفاضلة التي افرزتها نتائج التقدير لنماذج انحدار بواسون لعينات التطبيقات المدروسة .

3- تأثير التغير في حجم العينة على نتائج تقدير انموذج انحدار بواسون ظهر فقط في قيمة الانحراف (Deviance) والمتعلقة بطريقة تقدير انحدار النواة اذ نجد ان قيمة هذا المؤشر تزداد بزيادة حجم العينة.

4- اظهرت قيمتي إحصاءه بيرسون Pearson Statistic و إحصاءه الانحراف Deviance Statistic لنماذج انحدار بواسون المقدرة لعينة البحث ان متغير الاستجابة يعاني من ظاهرة تحت التشتت under dispersion, اذ كانت قيمتهما اقل من عدد درجات الحرية.

4-2 التوصيات

1- نوصي باستعمال طريقة تقدير انحدار النواة اللامعلمية والتي تفترض فروضا قوية على عملية التقدير, في تقدير انموذج انحدار(Poisson) المستخدم في صياغة طبيعة العلاقة بين متغير استجابة ذو قيما معدودة ومتغيراً او اكثر من المتغيرات التوضيحية.

2- اجراء بحوث مستقبلية تأخذ بنظر الاعتبار تأثير العوامل الاخرى على عدد مرات تعرض المريض للجلطة القلبية بالإضافة لعمر المريض.

3- نوصي بتوسيع طريقة اختبار الأنموذج من خلال اختبار مدى معنوية المعالم المقدرة كل معلمة على حده.

4- نوصي بأجراء بحوث مستقبلية تتناول طرائق تقدير اخرى (معلمية او لامعلمية) كطريقة العزوم وطريقة تحليل الموجة الصغيرة Wavelet لتقدير انموذج انحدار بواسون واجراء مقارنة بين تلك الطرائق لفرز الافضل منها.

Sources : المصادر

1.Algamal, Z. Y. and Rasheed, K. B. , 2021 , " Re- sampling Techniques in count data regression models " , Iraqi journal of statistical Science , Vol. 22 , pp.(15-25)

2.Allahham, N. (2015). ON THE SIMPLE ANGULAR REGRESSION MODEL WITH WRAPPED CAUCHY ERROR (Doctoral dissertation, Al Azhar University-Gaza).

3. Aboudi, Imad Hazem and Sabry, Hussam Muwaffaq, 2013, "A comparison between methods of estimating the parameters of the Poisson regression model in the presence of the multicollinearity problem", PhD thesis, College of Administration and Economics, University of Baghdad.

4.Camarda, C. G., 2018, "Smoothing methods for the analysis of mortality data", PH.D thesis, Carlos III University, Madrid, Spain.

5.Cameron,A. C. and Thrived, P.R, 1998, "Regression Analysis of count Data", Cambridge University press, pp.(1-434).

6. Davidson, Chris; translated by Mark, Aboud, 2013, "Angina and Heart Attacks," Family Physician Books, Arab Journal, King Abdulaziz City for Science and Technology KACST.

7.Hwang, G. and Shim., 2020, "Semi parametric Kernel Poisson Regression for longitudinal Count Data", Communication of the Korean Statistical Society. vol. 15, No. 6, pp. (1003-1011).

8.Mardia, K. V. (1972). Statistics of directional data Academic Press. New York, 357.

9.Rodriguez, G., 2007, "generalized linear Models", Princeton University, Revised September 2007 .

10 .Rodriguez, G., 2017, "generalized linear Models", Princeton University, Revised September 2007.

11.Shim, J.,2012, "Kernel Poisson regression for Mixed input variables", Journal of The Korean Data and Information Science Society, 23(6), pp. (1231-1239).

12.Winklmann Rainer ,2022, "Econometric Analysis of Count Data", Fifth edition, Springer – verlag Berlin Heidelberg.

13. Jassim, Wafaa Mahmoud and Ahmed, Lamia Mohammed, 2011 "Factors affecting the increase in the prevalence of heart attack and stroke in Kirkuk city" Technical Journal, Technical Education Authority, Volume 24, Issue 4.