Comparison Between Two Methods Maximum Likelihood and Weighted Likelihood for Estimate of Negative Binomial Model

مقارنة بين طريقتي الامكان الاعظم والامكان الموزونة لتقدير إنموذج أنحدار ثنائي الحدين السالب

أ.م.د.شروق عبد الرضا سعيد عدنان فاضل طعمة
جامعة كربلاء / كلية الادارة والاقتصاد قسم الاحصاء
(بحث مستل من رسالة ماجستير)

المستخلص

يعد إنموذج أنحدار ذي الحدين السالب أحد النماذج العددية التي تستعمل لتمثيل بعض الظواهر والحالات التي لايمكن التعبير عنها بالنماذج الاعتيادية وفي هذا تم تقدير معلمات إنموذج أنحدار ثنائي الحدين السالب بطريقتي (طريقة الأمكان الأعظم maximum likelihood estimation (MLE وطريقة الأمكان الموزونة weighted likelihood estimation (WLE) بهدف الوصول الى أفضل طريقة ، أذ سحبت عينة عشوائية بسيطة حجمها 257 حالة من حديثي الولادة يعانون من تشوهات خلقية مسجلين في دائرة صحة بابل ، وقد بينت النتائج أن طريقة الأمكان الأعظم هي الأفضل كونها أمتلكت أقل متوسط مربعات للخطأ MSE وأعلى معامل تحديد R^2 .

Abstract:

The negative binomial regression model is considered one of the count models , It is used to represent many phenomena and cases which cannot be expressed by ordinary models The negative they use a count models generally and negative binomial regression spetially to represent these phenomena and cases The parameters of negative binomial regression model by two different methods : [maximum likelihood estimation (MLE) , weighted likelihood estimation (WLE). The goal of the study to reach the best method for estimation , where taked a random sample (257) patients Who Where fined newborn suffering from congenital anomalies , they registered in the department of health Babylon. The show results that the method of (MLE)is the best because it has the lowest mean square error MSE and the highest R^2

1-المقدمة ومنهجية البحث

(1-1) المقدمة

هنالك بعض الظواهر الطبيعية كالظواهر الطبية ، الهندسية ، المالية ، الجيوفيزيائية والطبيعية (الامطار والاعاصير والزلازل) وغيرها، لايمكن تمثيلها بتوزيع منفرد بل تحتاج الى دمج توزيعين مثل توزيع بواسون وتوزيع كاما للحصول على توزيع اكثر مرونه لتمثيل الظواهر المعقده والمجتمعات غير المتجانسة .

يعد توزيع ذي الحدين السالب احد التوزيعات الاحصائية المتقطعة المهمة في الحقول العلمية المتعددة مثل الدراسات الحياتية والبايولوجية ، البيئية ، العلوم الزراعية ، الهندسية ، علوم البكتريا . فهو اساس لنموذج احصائي للبيانات العددية (count data) .

فمن المعلوم ان الوسط الحسابي والتباين لتوزيع بواسون متساوي ، وكلما تزداد قيمة المتوسط تزداد قيمة التباين ، وهذه الخاصية للبيانات يطلق عليها متعادلة التشتت (equidispersion) في حالة البيانات تمتلك توزيع بواسون ، لكن اذا لم يتحقق هذا الافتراض اي ان التباين اكبر من المتوسط للبيانات حيث البيانات تمتلك خاصية فرط التشتت (overdispersion) ، سوف نلجأ الى نموذج ثنائي الحدين السالب ، اي نموذج بواسون – كاما المركب (poisson – gamma mixture model) اكثر ملائمه في حالة فرط التشتت . لهذا تناولنا هذا النموذج المهم والتطبيقات العملية وفي كافة المجالات الطبية والهندسية والحياتية وغيرها .

(2-1) مشكلة البحث

تقسم مشكلة البحث الى :-

1. مشكلة البحث النظرية : كافتيقارنا الى انموذج عددي يعالج الظواهر والحالات العددية التي يمكن تقدير معلماته.
2. مشكلة البحث العملية : زيادة استخدام هذا الانموذج العددي لتسهيل زيادة حدوث التشوهات الخلقية في الفترة هذه بسبب أوضاع العراق والحروب والاوضاع التي عاشها البلد الامر الذي يدفعنا لدراسة العوامل والاسباب التي ساعدت في حصولها والظروف .

(3-1) أهمية البحث

تكمن أهمية البحث في تقدير معلماته بطريقتي : طريقة الأمكان الأعظم وطريقة الأمكان الموزونة ومعرفة افضل هذه الطرق لتقدير المعلمات .فضلا عن اهمية مجال البحث وتطبيقة على الخدج المصابين بالتشوهات الخلقية والمنتشرة بصورة واسعة في الفترة الاخيرة .

(4-1)منهجية البحث

استخدام المنهج التطبيقي بعد المنهج الاستقرائي في تقدير معلمات انموذج ثنائي الحدين السالب.

(5-1) هدف البحث

يهدف البحث الى:

1. دراسة توزيع ثنائي الحدين السالب وخصائصة ونموذج الانحدار الخاص به .

2. تقدير معلمات نموذج انحدار ثنائي الحدين السالب باستعمال بعض طرائق التقدير(طريقة الأمكان الأعظم ، طريقة الأمكان الأعظم الموزونة) والمقارنة بين نتائج التقدير للطريقتين وتحديد افضل طريقة لتقدير المعلمات .

2- الجانب النظري :

(1-2) التشوهات الخلقية[1]

التشوهات الخلقية هي تكوين غير طبيعي في احد أجزاء الجسم في مرحلة تكوين الجنين ، او هي ولادة الطفل بنقص او زيادة ليست طبيعية في عضو بجسم المولود او جزء منه من خلال وجود خلل جيني اثناء المراحل المبكرة لتطوير الخلية التكوينية بسبب تعرضها لبعض التأثيرات والعوامل الشاذة عن طبيعة تطور الجنين .وتعتبر التشوهات الخلقية اما ظاهرية او داخلية ويمكن التاكد منة بالفحوصات المختبرية والفحص بالأشعة والسونار .[7][4]

يمكن ان نصف دالة الكتلة الاحتمالية لثنائي الحدين السالب هو احتمال فشل المشاهده y قبل r-th من النجاحات لسلسلة محاولات برنولي ، علما" ان r هي عدد صحيح موجب . هذا هو الاختلاف في فلسفة حالات الفشل والنجاح بين نماذج ثنائي الحدين والهندسي وثنائي الحدين السالب ، في ثنائي الحدين يصف عدد النجاحات في n من محاولات برنولي ، في الهندسي يصف عدد حالات الفشل قبل النجاح الاول ، بينما ثنائي الحدين السالب يصف عدد حالات الفشل قبل r-th من النجاحات .اي ان توزيع ثنائي الحدين يصف عدد نجاحات بينما ثنائي الحدين السالب يصف حلات الفشل قبل حدوث r من النجاحات ومن هنا جائت تسميته بالسالب.

نفرض y يتوزع ثنائي الحدين السالب بمعلمات α وθ اذ ان : $y \sim NB(\alpha, \theta)$

$$p(Y = y) = \frac{\Gamma(\alpha + y)}{\Gamma(\alpha)\Gamma(y+1)}\left(\frac{1}{1+\theta}\right)^{\alpha}\left(\frac{\theta}{1+\theta}\right)^{y} \quad ; \alpha, \theta \in R^{+} ; y = 0,1,2, \dots \qquad (1-2)$$

$\Gamma(.)$ ترمز لدالة كاما في المناهج الاحصائيه توجد معلمات بديلة $p = {}^{1}/_{1+\theta}$ كذلك $\varphi = 1/\theta$ واذا كان α عدد صحيح سمي التوزيع بتوزيع باسكال .

(2-2) أنحدار ثنائي الحدين السالب[5][4]

انحدار ثنائي الحدين السالب مشابه لحد ما للانحدار المتعدد المنتظم وهنا المتغير المعتمد y مشاهداته عددية تتوزع ثنائي الحدين السالب ، وقيم y الممكنه هي اعداد صحيحه غير سالبه .... , 3, 2, 1 , 0.

نموذج ثنائي الحدين السالب (negative binomial model) هو نوع من النماذج الخطية العامة (generalized linear models) يحوي المتغير المعتمد (Y) ياخذ ارقام قابلة للعد لاي ظاهرة او حدث ما ، المعلمات الملائمة لتوزيع ثنائي الحدين السالب هي:

$$\theta = \alpha\mu$$

$$p(y) = p(Y = y) = \frac{\Gamma(y + 1/\alpha)}{\Gamma(y+1)\Gamma(1/\alpha)}\left(\frac{1}{1+\alpha\mu}\right)^{1/\alpha}\left(\frac{\alpha\mu}{1+\alpha\mu}\right)^{y} \qquad \dots (2-2)$$

اذ ان : $\mu > 0$ وهو الوسط الحسابي لل(Y) ، $\alpha > 0$ معلمة التشتت (heterogeneity parameter) يمكن اشتقاقها من توزيع بواسون كاما المركب (poisson-gamma mixture) او من خلال عدد حالات الفشل قبل $(1/\alpha)^{th}$ من حالات النجاح ، علما ان $1/\alpha$ ليس بالضروري ان تكون عددا صحيحا" .نموذج ثنائي الحدين السالب التقليدي (NB2) هو :

$$\ln\mu = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p \qquad \dots\dots(3-2)$$

اذ ان :$X_1, \dots\dots, X_p$ المتغيرات المستقلة ، $\beta_1, \dots\dots, \beta_p$ تمثل معاملات الأنحدار .

في حالة العينة العشوائية (n) لأي موضوع قيد الدراسة سوف نرى ان الموضوع (i) يأخذ متغير معتمد Y_i ومتغيرات مستقلة X_i وتأخذ قيم $(X_{1i}, X_{2i}, \dots, X_{pi})$ ويأخذ متجه للمعلمات $\beta = (\beta_0 \beta_1 \dots\dots \beta_p)^T$ ، ومصفوفة المتغيرات المستقلة :

$$X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix}$$

ويمكن اعادة ترتيب صف i^{th} للx ليكون x_i :

$$p(y_i) = \frac{\Gamma(y_i + 1/\alpha)}{\Gamma(y_i+1)\Gamma(1/\alpha)} \left(\frac{1}{1+\alpha e^{x_i\beta}}\right)^{\frac{1}{\alpha}} \left(\frac{\alpha e^{x_i\beta}}{1+\alpha e^{x_i\beta}}\right)^{y_i}, i = 1,2,\dots,n \qquad \dots(4-2)$$

(3-2) طرق التقدير

(1-3-2) طريقة الأمكان الأعظم [2][4]:

هي الطريقه الاكثر شيوعا والمالوفه في التقدير ، تعد مقدرات الامكان الاعظم ثابته وتعطي مقدرات الامكان الاعظم معيارا عالي للكفاءه وكذلك بخاصية الاتساق وبالتالي فهي العملية التي تعطي دالة الامكان للمتغيرات العشوائيه اعظم مايمكن .

$$L(\mu; y, \alpha) = \prod_{i=1}^{n} \exp\left\{ y_i \ln\left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right) - \frac{1}{\alpha}\ln(1+\alpha\mu_i) + \ln\Gamma\left(y_i + \frac{1}{\alpha}\right) - \ln\Gamma(y_i+1) - \ln\Gamma\left(\frac{1}{\alpha}\right)\right\} \qquad \dots\dots(5-2)$$

دالة الامكان اللوغارتميه يمكن ان نحصل عليها وذلك باخذ اللوغارتم الطبيعي لكلا الجانبين للمعادلة، فتصبح الداله جمعيه اكثر مما هي ضربيه :

$$L(\mu; y, \alpha) = \sum_{i=1}^{n} y_i \ln\left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right) - \frac{1}{\alpha}\ln(1+\alpha\mu_i) + \ln\Gamma\left(y_i + \frac{1}{\alpha}\right) - \ln\Gamma(y_i+1) - \ln\Gamma\left(\frac{1}{\alpha}\right) \qquad \dots\dots(6-2)$$

ان الامكان اللوغارتمي لثنائي الحدين السالب له معلمةβ ومعاملات النموذج كما يعبر عنها بالمعادله :

$$L(\beta_j; y, \alpha) = \sum_{i=1}^{n} y_i \ln\left(\frac{\alpha\exp(x_i'\beta)}{1+\alpha\exp(x_i'\beta)}\right) - \frac{1}{\alpha}\ln[1+\alpha\exp(x_i'\beta)] + \ln\Gamma\left(y_i + \frac{1}{\alpha}\right) - \ln\Gamma(y_i+1) - \ln\boldsymbol{\Gamma}\left(\frac{1}{\alpha}\right) \qquad \dots(6-2)$$

ان المشتقه الاولى للوغارتميه دالة الامكان تعرف باسم (gradient) ، اما المشتقه الثانيه فتسمى مشتقة حسن (Hessian) ، ان مقدرات الامكان الاعظم لمعلمات النموذج يتم تحديدها بايجاد المشتقه الاولى (gradient) لوغاريتمية الامكان بالنسبه الى β ومساواتها بالصفر وحلها ، بعباره اخرى توجد قيمه للمشتقات الاولى الجزئيه لوغاريتمية الامكان لتقدير عدة معالم . ويمكن تقدير معلمة عدم التجانس (heterogeneity parameter) لثنائي الحدين السالب بايجاد المشتقه الاولى لوغاريتمية الامكان بالنسبه الى α .

مصفوفة المشتقات الجزئيه الثانيه للامكان اللوغارتمي غالبا ماتعرف باسم حسن ومصفوفة المعلومات المشاهدة (observed information matrix) تعرف بسالب حسن (Hessian negative matrix) تقيم من خلال مقدر الامكان الاعظم .الاخطاء القياسيه للنموذج يمكن الحصول عليها من الجذر التربيعي للعناصر القطريه لمصفوفة التباين

والتباين المشترك ، والتي تعتبر معكوس المعلومات لذلك الاخطاء القياسيه للنموذج هي الجذر التربيعي للحدود القطريه لمصفوفه معكوس سالب حسن (negative inverse Hessian matrix) .
نيوتن- رافسون تعتمد على تقدير الامكان الاعظم ويستلزم مصفوفة معلومات المشاهده ، المشتقتان الاولى والثانيه لدالة الامكان الاعظم اللوغارتميه بالنسبه الى β , α كما يلي :

$$\mu = \exp(x'\beta)$$

NB GRADIENT – β

$$\frac{\partial L}{\partial \beta} = \sum_{i=1}^{n} \frac{x_i(y_i - \mu_i)}{1 + \alpha\mu_i} \quad \dots\dots\dots (7-2)$$

المشتقه الاولى بالنسبه الى μ (فقط في حدود الـμ)

$$\frac{\partial L}{\partial \mu} = \sum_{i=1}^{n} \frac{y_i - \mu_i}{\mu_i(1 + \alpha\mu_i)} \quad \dots\dots (8-2)$$

NB GRADIENT – α

$$\frac{\partial L}{\partial \alpha} = \sum_{i=1}^{n} \left[\frac{1}{\alpha^2} \left(\ln(1 + \alpha\mu_i) + \frac{\alpha(y_i - \mu_i)}{1 + \alpha\mu_i} \right) + \Psi\left(y_i + \frac{1}{\alpha}\right) - \Psi\left(\frac{1}{\alpha}\right) \right] \quad \dots (9-2)$$

NB –HESSIAN –β

$$\frac{-\partial^2 L}{\partial \beta \, \partial \beta'} = \sum_{i=1}^{n} \frac{\mu_i(1 + \alpha y_i)}{(1 + \alpha\mu_i)^2} x_i x_i' \quad \dots\dots (10-2)$$

او

$$\frac{\partial^2 L}{\partial \beta_j' \, \partial \beta_j} = \sum_{i=1}^{n} \left[-x_{ij} x_{ij}' \frac{\mu_i(1 + \alpha\mu_i)}{(1 + \alpha\mu_i)^2} \right] \quad \dots\dots (11-2)$$

NB –HESSIAN –β ; α

$$\frac{\partial^2 L}{\partial \beta \, \partial \alpha} = E\left[-\sum_{i=1}^{n} \frac{\mu_i(y_i - \mu_i)x_{ij}}{(1 + \alpha\mu_i)^2} \right] \quad \dots\dots (12-2)$$

NB –HESSIAN –α

$$\frac{\partial^2 L}{\partial \alpha^2} = \sum_{i=1}^{n} \left[-\frac{1}{\alpha^3} \left(\frac{\alpha(1 + 2\alpha\mu_i)(y_i - \mu_i) - \alpha\mu_i(1 + \alpha\mu_i)}{(1 + \alpha\mu_i)^2} \right) + 2\ln(1 + \alpha\mu_i) + \Psi'\left(y_i + \frac{1}{\alpha}\right) - \Psi'\left(\frac{1}{\alpha}\right) \right] \quad \dots\dots (13-2)$$

مشتقة دالة كاما اللوغارتميه هي دالة كاما الثنائيه (digamma) كذلك المشتقه الثانيه لدالة كاما اللوغارتميه دالة كاما الثلاثيه (trigamma)

$$\text{DIGAMMA} = \Psi_1 = \Psi(x) = \frac{\ln\Gamma(x + 0.0001) - \ln\Gamma(x - 0.0001)}{0.0002} \quad \dots\dots (14-2)$$

وكذلك

$$\text{TRIGAMMA} = \Psi_2 = \Psi'(x) = \frac{-\ln\Gamma(x + 0.002) + 16\ln\Gamma(x + 0.001) - 30\ln\Gamma(x) + 16\ln\Gamma(x - 0.001) - \ln\Gamma(x - 0.002)}{0.000012}$$

او

$$= \frac{\left\{\ln\Gamma\left(\left(1/\alpha\right) + 0.0001\right) - \ln\Gamma\left(\left(1/\alpha\right) - 0.0001\right)\right\}}{0.0002} \quad \dots\dots (15-2)$$

(2-3-2) طريقة الأمكان الموزونة[3]

طريقة الامكان الموزونة تعطينا مقدرات ذات كفائة عالية في حالات ومواقف تكون فيها الاخطاء متغيرات مستقلة توضيحية. الاوزان تتكون من مقارنة التوزيع التجريبي للبواقي والتوزيع النظري .

هذه الطريقة يمكن تطبيقها في حالات عندما توزيع الاخطاء يكون معتمد على المتغيرات المستقلة (covariates) مثل انحدار بواسون او انحدار ثنائي الحدين السالب . اما انحدار ثنائي الحدين السالب ، فنفترض $NB_{\alpha,\mu}$ عائلة توزيعات ثنائي الحدين السالب كذلك $Y_{\alpha,\mu} \sim NB_{\alpha,\mu}$

اذ ان :

$$E(Y_{\alpha,\mu}) = \mu \qquad \text{and} \qquad var(Y_{\alpha,\mu}) = \mu + \alpha\mu^2$$

متغير الاستجابة (Y)

$$Y_{\alpha_0,\mu_0(x)} \sim NB_{\alpha_0,\mu_0(x)}$$

اذ ان :

X : متجة المتغيرات المستقلة ، كذلك $\mu_0(x) = \theta^{-1}(\beta_0^T x)$

θ : دالة الربط (link function)

β_0 : متجة المعلمات المجهولة unknown parameters

الاخطاء $Y_{\alpha_0,\mu_0(x)} - \mu_0(x)$ هي ليست (iid) ، بل هي تعتمد على المتغيرات المستقلة ، ولايمكن تحويلها للشكل المعياري ضمن النموذج الطبيعي .

الغرض الرئيسي من هذة الطريقة لتقدير α_0, β_0 . نفرض لدينا عينة عشوائية $(x_1, y_1), \dots, (x_n, y_n)$ ولدينا $\theta = (\alpha, \beta)$ كذلك $z_i = (x_i, y_i)$ سوف تكون الاوزان $w(z_i, h)$ ، ويمكن توضيح مقدر (h) :

$$\sum_{i=1}^{n} w(z_i, h)s(h, z_i) = 0 \qquad \dots\dots\dots (16-2)$$

اذ ان :

S(h,z) قيمة دوال التسجيل (usual score function)

خلال عملية تكوين الاوزان : يمكن تعريف الاحتمالات :

$$P_h(z_i) = P(Y_{\alpha,\mu(x_i)} \le y_i) - u_i P(Y_{\alpha,\mu(x_i)} = y_i) \qquad \dots (17-2)$$

اذ ان : $u_1, \dots\dots, u_n$ ارقام عشوائية تولدت من التوزيع المنتظم على $[0,1]$ ، لان $P_\theta(z_i), i = 1, \dots\dots, n$ ، كذلك $\theta = (\alpha_0, \beta_0)$ هي عينة من التوزيع المنتظم على $[0,1]$

(4-2) معايير المقارنة [6](comparative criterions)

❖ **متوسط مربعات الخطأ (Mean Square Error)(MSE)**

$$MSE = \frac{SSE}{n-p} \qquad \dots\dots\dots (18-2)$$

اذ ان : n حجم العينة ، p عدد معلمات النموذج .

$$SSE = \sum_{i=1}^{n} u_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \qquad \dots\dots\dots (19-2)$$

❖ **معامل التحديد (coefficient of determination)**

يستعمل هذا المعيار في تحديد قدرة النموذج في تفسير التغيرات التي تحصل على متغير الأستجابه (y) .

$$\sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} (y_i - \hat{\theta}_i)^2 + \sum_{i=1}^{n} (\hat{\theta}_i - \bar{y})^2 + 2\sum_{i=1}^{n} (y_i - \hat{\theta}_i)(\hat{\theta}_i - \bar{y}) \quad \dots (20-2)$$

أذ أن :

$\sum_{i=1}^{n} (y_i - \bar{y})^2$: مجموع مربعات الأنحرافات الكلية .

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

$\sum_{i=1}^{n} (\hat{\theta}_i - \bar{y})^2$: مجموع مربعات الأنحرافات المفسره .

- قيمة معامل التحديد تتراوح بين الصفر والواحد $0 \le R^2 \le 1$

3- الجانب التطبيقي

(3-1) عينة الدراسة (The Sample) :

تم سحب عينة عشوائية بسيطة من حالات التشوهات الخلقية (congenital anomalies) للخدج من دائرة صحة بابل حجمها(257) ولغرض حصر التشوهات الخلقية تم أعتماد تصنيف وترميز (ICD 10) وكما في الجدول (3-1) .

الجدول (3-1) يبين العيوب الخلقية وترميزها حسب التصنيف (ICD10)

	قائمة العيوب الخلقية باللغة العربية	الرمز وفق التصنيف (ICD 10)	قائمة العيوب الخلقية باللغة الانكليزية	الرمز
1	انعدام الدماغ	Q00	ANENCEPHALY	0
2	صغر الرأس	Q02	MIROCEPHALUS	1
3	أستسقاء الرأس الخلفي (موه الرأس الخلفي)	Q03	CONGENITAL HYDROCEPHALUS	2
4	العيوب الخلقية في القلب وجهاز الدوران	Q28	CONGENITAL ANOMALIES OF HEART AND CIRCULATORY SYSTEM	3
5	المنغولية	Q90	MONGOLISM	4
6	العيوب الخلقية الكروموزومية الاخرى	Q91-99	OTHER CHROSOMAL ANOMALIES	5
7	شق خلقي بالشفة (شق الارنب)	Q36	CLEFTLIP	6
8	شق خلقي بالحنك	Q35	CLEFT PALATE	7
9	شق خلقي بالشفة والحنك	Q37	CLEFT LIP AND PALATE	8
10	الصلب الاشرم	Q05	SPINABIFIDE	9
11	العيوب الخلقية الاخرى في الدماغ والحبل الشوكي	Q06	OTHER ANOMALIES OF BRAIN AND SPINAL CORD	10
12	الخنوثة	Q56	AMBIGOUS GENITALIA	11
13	قيلة مائية خلقية	Q83	HYDROCELE CONGENITAL	12
14	خصية غير نازلة	Q53	UNDEACENDED TESTIS	13
15	فتحة احليل فوقانية او تحتانية	Q64	HYPOSPADIAS AND EPISPADIAS	14
16	العيوب الخلقية الاخرى في الاعضاء التناسلية	Q52,Q54 ,Q55	OTHER ANOMALIES OF GENITO - URINARY ORGANS	15
17	العيوب الخلقية في الجلد	Q82	CONGENITAL ANOMALIES OF THE SKIN	16
18	تضيق فتحة الشرج	Q42	ANAL STENOSIS	17
19	تشوهات خلقية بالجهاز الهضمي	Q38-Q41,Q43 -Q45	Other congenital malformation of the digestive system	18
20	العيوب الخلقية في العين	Q15	CONGENITAL ANOMALIES OF THE EYE	19
21	صيوان الاذن الاضافي	Q17	ACCESSORY AURICLE	20
22	العيوب الخلقية في الطرف العلوي	Q71	CONGENITAL ANOMALIES OF UPPER LIMB	21
23	العيوب الخلقية في الطرف السفلي	Q72	CONGENITAL ANOMALIES OF LOWAR LIMB	22
24	تشوهات خلقية اخرى بالوجة والرقبة	Q18	Other congenital malformations of face and neck	23
25	تشوهات خلقية اخرى بالجهاز التنفسي	Q34	Other Congenital malformations of respiratory system	24
26	تشوهات خلقية بالجهاز العضلي	Q79	Congenital malformations of	25

	musculoskeletal system , not elsewhere classified		الهيكلي غير مصنفة في مكان اخر	
26	Other specified congenital malformation syndromes affecting multiple systems	Q87	متلازمات معينة اخرى للتشوه الخلقي المؤثرة باجهزه متعددة	27
27	Other congenital malformations, not elsewhere classified	Q89	تشوهات خلقية اخرى غير مصنفة في مكان اخر	28

(2-3) متغيرات الدراسة

وشملت الدراسة (14) متغير ورمزنا للمتغير المعتمد (y) وهو متغير الاستجابة العددية (count response variable) ويأخذ رقم الرمز من (27- 0) اي نوع العوق التشوه الخلقي ، اي ان المتغير المعتمد يأخذ اعداد رقمية ومن هنا جاءت تسمية نماذج الاستجابة العددية ، المتغيرات المستقلة $\mathbf{x_i}$ ويتألف الجدول(3-2) من :

الجدول (3 – 2) يبين المتغيرات وتفاصيلها

الرمز	أسم المتغير	التفاصيل
X_1	عمر الام	
X_2	مهنة الام	1-ربة بيت 2-موظفة حكومية 3-موظفة اهلية 4-اعمال حرة
X_3	عمر الاب	
X_4	مهنة الاب	1-لايعمل 2-موظف حكومي 3-موظف اهلي 4-اعمال حرة
X_5	درجة القرابة بين الابوين	1-يوجد 2-لايوجد
X_6	الولادات السابقة	هل يوجد عوق ولادي للولادات السابقة : 1-نعم 2- كلا
X_7	نوع الولادة	1- مفردة 2- متعددة
X_8	الولادات الحالية	1- حية 2-ميتة
X_9	تعرض الام	1- الحمى 2- اشعاع 3- تناول ادوية 4-لايوجد مما ذكر
X_{10}	عدد الاسقاط السابق	
X_{11}	نوع السكن	1- حضر 2- ريف
X_{12}	جنس الطفل	1- ذكر 2- انثى 3- خنثى
X_{13}	وزن الطفل	
Y	نوع العوق ورمزة	يكتب حسب التصنيف الدولي العاشر للمراضة

(3-3) معرفة التوزيع الأحصائي الملائم للبيانات

لمعرفة التوزيع الأحصائي لبيانات التشوهات الخلقية لحديثي الولادة تم تطبيق اختبار حسن المطابقة **(Goodness of Fit)** للبيانات من خلال البرنامج الأحصائي **(Easy Fit)** وكانت النتائج ان البيانات تتبع لتوزيع ثنائي الحدين السالب بمعلمات **(n=2 , p=0.15498)** وكانت قيمة **(p-value =0.000)** . علما ان البيانات تمتلك خاصية فرط التشتت (Overdispertion) اي ان قيمة التباين للبيانات 74.28 اعلى من قيمة الوسط الحسابي وهي 11.47 وهذا يعطي دليل قوي بأن البيانات تتبع لتوزيع ثنائي الحدين السالب .

(4-3) تقدير معلمات نموذج انحدار ثنائي الحدين السالب :-

توجد العديد من الطرائق الأحصائية لتقدير معلمات نموذج انحدار ثنائي الحدين السالب ومنها طرائق الأمكان الأعظم (MLE) وطريقة الأمكان الموزونة (WLE) .

(1-4-3) طريقة الامكان الاعظم (Maximum Likelihood Estimation)

تعد طريقة الامكان الاعظم من الطرائق المهمة في تقدير معلمات نموذج انحدار ثنائي الحدين السالب، ويهدف أختبار فرضية العدم :

$$H_0 : \beta_0 = \beta_1 = \cdots = \beta_{13} = 0$$
$$H_1 : \beta_0 \neq \beta_1 \neq \cdots \neq \beta_{13} \neq 0$$

وبأستعمال البرنامج الأحصائي (Minitab 17) تم تقدير نموذج الانحدار اي تقدير وأختبار معلمات انحدار ثنائي الحدين السالب حيث كان قيمة معامل التحديد (15.25%) ومتوسط مربعات الخطأ(66.32) وكالاتي:

جدول رقم(3-3) يبين قيم معاملات الأنحداروفق طريقة MLE

المتغيرات المستقلة	معاملات الأنحدار (b)	الخطأ المعياري	P- value
(عمر الأم) X1	-0.0076193	0.0095585	0.425
(مهنة الأم)X2	-0.2640292	0.3827613	0.490
(عمرالأب)X3	0.0024065	0.00763	0.752
(مهنة الاب) X4	-0.0790277	0.0426701	0.064
(درجة القرابة) X5	-0.0517542	0.1166073	0.657
(الولادات السابقة) X6	0.2388692	0.118685	0.044
(نوع الولادة) X7	-0.5673884	0.2730285	0.038
(الولادة الحاليه) X8	-0.3568640	0.127135	0.005
(تعرض الأم) X9	-0.1255543	0.0389791	0.001
(عدد الأسقاط) X10	-0.0488763	0.094799	0.606
(نوع السكن) X11	0.0274382	0.1042203	0.792
(جنس الطفل) X12	0.0646894	0.0958825	0.500
(وزن الطفل) X13	0.0001033	0.0000707	0.144
الحد الثابت	3.759862	0.6895661	0.000

يتضح من الجدول (3-3) ان القيم المعنوية للمتغيرات (x_6, x_7, x_8, x_9) اي ان قيم $\mathrm{p-value}$ اقل من 0.05 اي ان العوامل المؤثره في نوع العوق هي وجود العوق في الولادات السابقة ونوع الولاده مفرده ام متعدده كذلك الولادة الحالية حية ام ميتة واخيرا" تعرض الأم الى حمى او اشعاع او تناول ادويه تساعد في حصول التشوة الخلقي خصوصا" في المرحلة الأولى من الحمل. اما بقية العوامل فليس لها تأثير على نوع العوق او تأثيرها قليل جدا .

(3-4-2) طريقة الأمكان الموزونة

في هذة الطريقة تم استعمال البرنامج الأحصائي STATA لتقدير وأختبار معلمات نموذج انحدار ثنائي الحدين السالب ، علما ان فرضية العدم :

$$H_0 : \beta_0 = \beta_1 = \cdots = \beta_{13} = 0$$

$$H_1 : \beta_0 \neq \beta_1 \neq \cdots \neq \beta_{13} \neq 0$$

وباستعمال البرنامج الأحصائي (Mini tab 17) حيث كان معامل التحديد (14.30%) ومتوسط مربعات الخطأ (69.82) وكانت النتائج مبينة في الجدول التالي :

جدول رقم(3-4) يبين قيم معاملات الأنحدار بطريقة WLE

المتغيرات المستقلة	معاملات الأنحدار	الخطا المعياري	p-value
(عمر الأم) X1	-0.0076094	0.604944	0.900
(مهنة الأم)X2	-0.1626012	1.267844	0.898
(عمرالأب)X3	0.0019222	0.0514333	0.970
(مهنة الاب) X4	-0.066428	0.2846189	0.815
(درجة القرابة) X5	-0.0388292	0.7517452	0.959
(الولادات السابقة) X6	0.2415253	0.8126605	0.766
(نوع الولادة) X7	-0.6431148	1.403072	0.647
(الولادة الحاليه) X8	-0.337303	0.7623365	0.658
(تعرض الأم) X9	-0.1550309	0.25427	0.542
(عدد الأسقاط) X10	-0.090717	0.6770126	0.893
(نوع السكن) X11	0.020613	0.6622849	0.975
(جنس الطفل) X12	0.0696726	0.5882673	0.906
(وزن الطفل) X13	0.0001332	0.0004594	0.772
الحد الثابت	3.666405	3.703025	0.322

لم يظهر اي عامل من العوامل (المتغيرات المستقلة) تأثيرا" معنويا واضحا" على المتغير التابع (نوع التشوة) اي ان جميع المتغيرات المستقلة كانت قيم p-value اكبر من 0.05 ونستنتج عدم رفض فرضية العدم ولجميع المتغيرات المستقله .

المقارنة بين الطريقتين :

ومن خلال الجدول التالي لمعامل التحديد ومتوسط مربعات الخطأ لكل طريقه من طرق التقدير المذكوره :

الجدول (3-5) يبين معامل التحديد ومتوسط مربعات الخطأ لطرق التقدير MLE,WLE

المعيار	MLE	WLE
MSE	66.32	69.82
R^2	%15.25	%14.3

من خلال الجدول اعلاه يظهر ان طريقة الأمكان الأعظم MLE هي أفضل طريقة للتقدير ومن خلال النتائج التي تطرقنا لها حيث تمتلك أقل متوسط مربعات للخطأ (66.32) وأعلى معامل تحديد أذ بلغ (15.25%) ثم طريقة الأمكان الأعظم الموزونه بمتوسط مربعات الخطأ (69.82) ومعامل تحديد (14.3%) .

4- الاستنتاجات والتوصيات

(4-1) الاستنتاجات :

1- توصلت الدراسة الى ان افضل توزيع لبيانات حديثي الولادة الذين يعانون من التشوهات الخلقيه هو توزيع ثنائي الحدين السالب .

2- وان طريقة الامكان الاعظم (MLE) تعتبر افضل من طريقة الامكان الموزونة WLE .

(4-2) التوصيات

1. ان استعمال نموذج انحدار ثنائي الحدين السالب تمثيل البيانات التي تحتوي خاصية فرط التشتت overdispertion اي يكون فيها التباين اعلى من الوسط الحسابي
2. تطبيق طريقة الامكان الاعظم في انموذج انحدار ثنائي في الحدين السالب فهي افضل من مقدرات طريقة الامكان العنقودية.
3. التوسع في البيانات العددية count data وبناء النماذج الخاصة بها وخصوصا البيانات الطبية منها لافتقار البحوث الحديثة لهذه المواضيع والظواهر المتعلقة بها .
4. زيادة الوعي الصحي من خلال الأعلانات والنشرات الصحية ، كذلك تحديث بيانات دوائر الصحة وعدم الأعتماد على السجلات القديمة التي أصبح دورها ضعيف جدا في حدوث التشوهات الخلقية ، وتفعيل دور فرق البيئة ومكافحة كل أشكال التلوث البيئي والحياتي .

المصادر

أولا : المصادر العربية

1. الثعلبي ، ساهره حسين زين "تحليل البيانات الثنائية لدراسة العوامل المؤثرة في حدوث التشوهات الولادية في مستشفى البصرة للنسائية والأطفال "رسالة ماجستير ، مقدمة الى مجلس كلية الأدارة والأقتصاد جامعة البصره (2008) .
2. محمد ، نور أياد " تقدير معلمات توزيع بواسون المركب مع تطبيق عملي " ، رسالة ماجستير ، مقدمة الى مجلس كلية الأدارة والأقتصاد جامعة بغداد (2017)

ثانيا : المصادر الاجنبية

3-Amiguet , M ."Weighted likelihood Negative Binomial Regression , Institute for social and preventive medicine , University of Lausanne , Saint Petersburg (2013)

4-Hilbe , J . M ."Negative Binomial Regression , second edition , Jet Propulsion Laboratory , California Institute of Technology and Arizona State University , Cambridge University Press(2011)

5-Hilbe , J .M ., Log negative binomial regression as a generalized linear model , technical report COS 93/94-5-26 ,Department of Sociology , Arizona state university (1993).

6-Nelder , J . A ., and Y. Lee , likelihood , quasi-likelihood and pseudo-likelihood : some comparisons , Journal of the Royal Statistical society , B 54 :pp(273-284) .(1992)

7-Winkelmann , R ."Count Data Models , Econometric theory and an application to labor mobility , Lecture Notes in Economics and Mathematical System No.(410)