Comparison Between Two Methods Iteratively Re-weighted Least Squares and Weighted Least Squares for Estimate of Negative Binomial Model

مقارنة بين طريقتي المربعات الصغرى المعادة الوزن التكرارية والمربعات الصغرى الموزونة لتقدير إنموذج أنحدار ثنائي الحدين السالب

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

المستخلص

تم تقدير معلمات إنموذج أنحدار ثنائي الحدين السالب بطريقتي (طريقة المربعات الصغرى المعادة الوزن التكرارية iteratively re-weighted least square (IRLS) وطريقة المربعات الصغرى الموزونة weighted least square (WLS)) 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 المربعات الصغرى المعادة الوزن التكرارية هي الأفضل كونها أمتلكت أقل متوسط مربعات للخطأ MSE وأعلى معامل تحديد R^2 .

Abstract

The parameters of negative binomial regression nodel :[iteratively re-weighted least square (IRLS) and weighted least square (WLS)]. The goal of the study to reach the best method for estimation , where random sample taken of (257) patients consider of newborn suffering from congenital anomalies, who registered in the department of health Babylon . The result show us that the method of (IRLS)is the best because it has the least 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 : أفتقار البحوث الحديثة التي تعالج الحالات والظواهر التي لايمكن تمثيلها ألا من خلال النماذج العددية (count models) وخصوصا" إنموذج ثنائي الحدين السالب .
2- مشكلة البحث العملية : زيادة حدوث التشوهات الخلقية في الفترة الأخيرة 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 ان نصل ألى فهم أعمق للنماذج العدديه بصوره عامة ونموذج أنحدار ثنائي الحدين السالب على وجه الخصوص وتقدير معلماته بطريقتي : طريقة المربعات الصغرى المعادة الوزن التكرارية وطريقة المربعات الصغرى الموزونة ومعرفة الأفضل بينهما الطريقتين لتقدير المعلمات .فضلا عن اهمية مجال البحث وتطبيقة على الخدج المصابين بالتشوهات الخلقية والمنتشرة بصورة واسعة في العراق في الفترة الاخيرة .

(4-1) 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 المربعات الصغرى المعادة الوزن التكرارية ، طريقة المربعات الصغرى الموزونة) والمقارنة بين نتائج التقدير الطريقتين وتحديد الافضل 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 بجسمه 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.

(2-2) توزيع ثنائي الحدين السالب[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 ، بينما الهندسي يصف عدد حالات الفشل قبل النجاح الاول ، اما ثنائي الحدين السالب فهو 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 جاءت 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 . α: معلمة الشكل ،θ :معلمة الموقع.

(3-2) 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 انحدار ثنائي الحدين السالب هو تعميم لانحدار بواسون ، لكن في نموذج بواسون يتساوى المتوسط والتباين ، اما انحدار ثنائي الحدين السالب يكون التباين اكبر من المتوسط ، (NB1 العلاقة الخطية بين المتوسط والتباين) أما(NB2 العلاقة التربيعية بين المتوسط والتباين) يعتمد على توزيع بواسون – كاما المركب ، وهذه الصيغه شائعه لانها تسمح لاستعمال توزيع كاما لعدم تجانس بواسون poisson heterogeneity .

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 : $0 < \mu$ 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 :

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

اذ ان : $x_1, \ldots \ldots, x_p$ المتغيرات المستقلة ، $\beta_1, \ldots \ldots, \beta_p$ 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 $(x_{1i}, x_{2i}, \ldots, x_{pi})$ ويأخذ متجه للمعلمات $\beta = (\beta_0 \beta_1 \ldots \ldots \beta_p)^T$ ، 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 :

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

ويمكن تقدير المعلمات α وβ باستعمال تقدير الامكان الاعظم ، فتصبح دالة الامكان الاعظم :

$$L(\alpha, \beta) = \prod_{i=1}^{n} P(y_i) = \prod_{i=1}^{n} \frac{\Gamma(y_i + 1/\alpha)}{\Gamma(y_i + 1)\Gamma(1/\alpha)} \left(\frac{1}{1 + \alpha e^{x_i\beta}}\right)^{\frac{1}{\alpha}} \left(\frac{\alpha e^{x_i\beta}}{1 + \alpha e^{x_i\beta}}\right)^{y_i} \qquad \ldots(5-2)$$

وبذلك تصبح دالة الامكان بعد اخذ اللوغارتمية :

$$\ln L(\alpha, B) = \sum_{i=1}^{n} \left(y_i \ln\alpha + y_i(x_i.\beta) - \left(y_i + \frac{1}{\alpha}\right)\ln\left(1 + \alpha e^{x_i\beta}\right) + \ln\Gamma\left(y_i + \frac{1}{\alpha}\right) - \ln\Gamma(y_i + 1) - \ln\Gamma\left(\frac{1}{\alpha}\right)\right) \qquad \ldots(6-2)$$

قيم α وβ التي تعظم $\ln L(\alpha, B)$ سوف تكون مقدرات الامكان الاعظم **بعد مساواة** $\ln L$ **بالصفر**. ومصفوفة التباين والتباين المشترك للمقدرات هي $\Sigma = -H^{-1}$ اذ ان : H هي مصفوفة هيسن (Hessian matrix) للمشتقه الثانية لدالة الأمكان اللوغارتمية .

(4-2) طرق التقدير

(1-4-2) طريقة المربعات الصغرى المعادة الوزن التكرارية [4]:-

Iteratively re – weighted least squares (IRLS)

تعتمد هذه الطريقه على ترميز فشر (fisher scoring) وتعتبر هذه الطريقه كجزء من طريقة الأمكان الاعظم والتي يمكن استخدامها لتقدير النموذج الخطي بالنسبه لمعلمات النموذج تسمى مصفوفة حسن Hessian matrix .

ومن المعلوم ان دالة الكثافه الاحتماليه للعائله الاسيه (exponential family) :-

اذ ان : $f(y;\theta,\Phi) = \exp\left[\frac{y_i\theta_i - b(\theta_i)}{\alpha_i(\Phi)} + c(y_i;\Phi)\right]$ ...... (7 – 2)

θ_i هي المعلمه القانونيه او دالة الربط (link function)

$b(\theta_i)$ هي معلمة الموقع (Location)

$\alpha(\Phi)$ معلمة القياس (scale parameter)

$c(y_i;\Phi)$ حد التطبيع (normalization)

شكل العائلة الأسية يعتبر وحيد (unique) في المشتقتان الاولى والثانية لمعلمة الموقع (Location) بالنسبة الθ فينتج دوال المتوسط والتباين على التوالي ،

$$b'(\theta_i) = \text{mean}$$
$$b''(\theta_i) = \text{variance}$$

ان الداله الاحتماليه للنموذج الخطي العام GLM :-

$$f(y;\theta,\Phi) \qquad \ldots\ldots\ldots .(8-2)$$

حيث ان :

Y متغير الاستجابه (response variable)

θ معلمة الموقع (location parameter)

Φ معلمة القياس (scale parameter)

في النماذج العدديه count models لدينا قيم متغير الاستجابه y لها خواص توزيعيه تتلائم مع التوزيع المستخدم في التقدير ، الدوال الاحتماليه تحدد خواص الاستجابه ، طريقة التقدير بالامكان الاعظم تعتمد على تعظيم الاحتمال ، دالة الامكان هي متممه للداله الاحتماليه . وبالرغم من ان البيانات هي التي تحدد قيم المتوسط وقيم القياس لكن يبقى الهدف المهم للامكان هو تحديد قيم المعالم المجهوله.

وباخذ اللوغارتم الطبيعي لدالة الامكان لتسهيل التقدير وناخذ الشكل الضربي multiplicative manner لتقدير المعلمات .

ان دالة الامكان اللوغارتميه يمكن كتابتها :

$$L(\theta,\Phi;y)$$

وبالاعتماد على تعديل تايلر الموسعة لدالة الامكان بعد اخذ اللوغاريتم :

$$0 = f(y_0) + (y_1 - y_0)f'(y_0) + \frac{(y_1-y_0)^2}{2!}f''(y_0) + \frac{(y_1-y_0)^3}{3!}f'''(y_0) + \cdots$$

و تختزل دالة او حدين فقط :

$$0 = f(y_0) + (y_1 - y_0)f'(y_0)$$

ويمكن كتابتها

$$y_1 = y_0 - \frac{f(y_0)}{f'(y_0)}$$

المشتقه الاولى لدالة الامكان اللوغارتميه تدعى ايضا بدالة فشر fisher score ، اذا كان لوغاريتم داله الامكان مقعره ، سوف نجعل المشتقه مساويه للصفر ونحل بالنسبه الى β لتقدير معلمات الامكان الاعظم .

المشتقه الثانيه لدالة الامكان اللوغارتميه التي تسمى مصفوفة هيسن (Hessian matrix)، عندما يكون شكل دالة الامكان اللوغارتميه محدب اكثر مما هو مفلطح ، معكوس حسن السالب(negative inverse Hessian matrix) يعطي مصفوفة التباين والتباين المشترك .

ان معلمة الاخطاء المعياريه تعتمد على عناصر القطر لمصفوفة التباين . وتسمى كذلك بمصفوفة المعلومات information (matrix) . وعليه نحصل على :

$$U = \partial L \qquad \text{and} \qquad H = \partial^2 L \qquad \ldots\ldots .(9-2)$$

اذ ان:

U : هي المشتقة الاولى لدالة لوغاريتم الامكان ، H : المشتقة الثانية لدالة لوغاريتم الامكان .

وبعدها نجد مقدرات المعلمه β_r وذلك بتوظيف نيوتن رافسون نحصل على :

$$\beta_r = \beta_{r-1} - H^{-1}U \qquad \ldots\ldots\ldots\ldots(10-2)$$

عندما

$$H = H_{r-1} \qquad \text{and} \qquad U = U_{r-1}$$

ومن خلال خوارزمية نيوتن – رافسون نحصل على β_r وهي مقدرات معلمة النموذج ، و بتكرار ايجاد الحلول للH يصبح لدينا مصفوفة المعلومات المشاهدة

(Observed information matrix)، كذلك مصفوفة Hessian المستخدمه في IRLS تسمى مصفوفة المعلومات المتوقعه(EIM) (expected information matrix)

ولايجاد المشتقه الاولى the gradient(U)

في شكل العائله الاسيه دالة الامكان اللوغارتميه تكون :

$$L(\theta; y, \Phi) = \sum_{i=1}^{n} \frac{y_i\theta_i - b(\theta_i)}{\alpha_i(\Phi)} + c(y_i; \Phi) \qquad \ldots\ldots(11-2)$$

نشتق الداله بالنسبه الى β باستعمال قانون السلسله :

$$\frac{\partial L}{\partial \beta_j} = \sum_{i=1}^{n} \left(\frac{\partial l_i}{\partial \theta_i}\right)\left(\frac{\partial \theta_i}{\partial \mu_i}\right)\left(\frac{\partial \mu_i}{\partial \eta_i}\right)\left(\frac{\partial \eta_i}{\partial \beta_i}\right) \qquad \ldots\ldots(12-2)$$

نحل كل حد

$$\frac{\partial L}{\partial \beta_j} = \sum_{i=1}^{n} \frac{y_i\theta_i - b'(\theta_i)}{\alpha_i(\Phi)} + \sum_{i=1}^{n} \frac{y_i - \mu_i}{\alpha_i(\Phi)} \qquad \ldots\ldots(13-2)$$

حصلنا على الصيغه اعلاه من اشتقاق كل حد للسلسله ولدينا $b'(\theta_i) = \mu_i$

$$\frac{\partial \mu_i}{\partial \theta_i} = \frac{\partial b'(\theta_i)}{\partial \theta_i} = b''(\theta_i) = V(\mu_i)\,, \frac{\partial \theta_i}{\partial \mu_i} = \frac{1}{V(\mu_i)} \qquad \ldots\ldots(14-2)$$

كذلك

$$\frac{\partial \eta_i}{\partial \beta_j} = \frac{\partial (x_i \beta_j)}{\partial \beta_j} = x_{ij} \quad , \text{since} \quad \eta_i = x_i \beta_j \qquad \ldots\ldots\ldots(15-2)$$

كذلك

$$\frac{\partial \mu_i}{\partial \eta_i} = [g^{-1}(\eta_i)]' = \frac{1}{\partial \eta_i / \partial \mu_i} = \frac{1}{g'(\mu_i)} \qquad \ldots\ldots\ldots(16-2)$$

هذه مشتقة داله الربط بالنسبه الى μ ، η معكوس دالة الربط . استبدال التعابير لمقدر الامكان الاعظم لل β هو الحل للمعادله التقديريه .

$$\sum_{i=1}^{n} \frac{(y_i - \mu_i) x_i}{\alpha_i(\Phi) V(\mu_i) g'(\mu_i)} = \sum_{i=1}^{n} \frac{(y_i - \mu_i) x_i}{\alpha_i(\Phi) V(\mu_i)} \left(\frac{\partial \mu_i}{\partial \eta_i}\right) = 0 \qquad \ldots\ldots\ldots(17-2)$$

عندما Y تمثل الاستجابه ، μ تمثل المتغيرات المجهزه (fitted variables) ، X متجه صفي ، المجموع الناتج متجه عمودي .

ولايجاد المشتقه الثانيه نستبدل H بالI

$$I = -E\left[\frac{\partial^2 L}{\partial \beta_j \, \partial \beta_k}\right] = E\left[\frac{\partial L}{\partial \beta_j} \frac{\partial L}{\partial \beta_k}\right] \qquad \ldots\ldots\ldots(18-2)$$

$$I = \frac{\partial}{\partial \beta_j}\left[\frac{(y_i - \mu_i) x_j}{\alpha_i(\Phi) V(\mu_i)} \left(\frac{\partial}{\partial \eta}\right)_i\right] * \frac{\partial}{\partial \beta_k}\left[\frac{(y_i - \mu_i) x_k}{\alpha_i(\Phi) V(\mu_i)} \left(\frac{\partial}{\partial \eta}\right)_i\right] \qquad \ldots(19-2)$$

$$I = \frac{(y_i - \mu_i)^2 x_j x_k}{\{\alpha_i(\Phi) V(\mu_i)\}^2} \left(\frac{\partial \mu}{\partial \eta}\right)_i^2 \qquad \ldots\ldots\ldots(20-2)$$

اذ ان :

$$(y_i - \mu_i)^2 = \alpha_i(\Phi) V(\mu_i)$$

نفرض ان :

$$V(y_i) = \alpha_i(\Phi) V(\mu_i) = (y_i - \mu_i)^2$$

$$I = \frac{x_j x_k}{V(y_i)} \left(\frac{\partial \mu}{\partial \eta}\right)_i^2 = \frac{x_j x_k}{V(y_i) g'^2}$$

نضع المعادلات مع بعضها سوف نحصل :

$$\beta_r = \beta_{r-1} - \left[\frac{x_j x_k}{V(y_i)} \left(\frac{\partial \mu}{\partial \eta}\right)_i^2\right]^{-1} \left[\frac{(y_i - \mu_i) x_k}{V(y_i)} \left(\frac{\partial \mu}{\partial \eta}\right)_i\right] \qquad \ldots\ldots(21-2)$$

بضرب طرفي المعادله في I

$$\left[\frac{x_j x_k}{V(y_i)}\left(\frac{\partial\mu}{\partial\eta}\right)_i^2\right]\beta_r = \left[\frac{x_j x_k}{V(y_i)}\left(\frac{\partial\mu}{\partial\eta}\right)_i^2\right]\beta_{r-1} + \left[\frac{(y_i-\mu_i)x_k}{V(y_i)}\left(\frac{\partial\mu}{\partial\eta}\right)_i\right] \qquad \dots (22-2)$$

نفرض ان w يساوي

$$W = \frac{1}{V(y_i)}\left(\frac{\partial\mu}{\partial\eta}\right)_i^2$$

ومع التركيبه الخطيه η يعطي

$$\eta_i = x_{ik}\beta_{r-1}$$

حيث ال η_i يبلغ قيمته في التكرار (n−1)

$$\left[\frac{x_j x_k}{V(y_i)}\left(\frac{\partial\mu}{\partial\eta}\right)_i^2\right]\beta_r = [X'WX]\beta_r \qquad \dots\dots\dots (23-2)$$

نعرف w, V(y)

$$\frac{(y_i-\mu_i)x_k}{V(y_i)}\left(\frac{\partial\mu}{\partial\eta}\right)_i = \frac{(y_i-\mu_i)x_k}{\frac{1}{w}\left(\frac{\partial\mu}{\partial\eta}\right)_i^2}\left(\frac{\partial\mu}{\partial\eta}\right)_i \qquad \dots\dots\dots (24-2)$$

وهنا

$$\left[\frac{x_j x_k}{V(y_i)}\left(\frac{\partial\mu}{\partial\eta}\right)_i^2\right]\beta_{r-1} = x'w\eta_i \qquad \dots\dots\dots (25-2)$$

بدمج التركيبين

$$[x'wx]\beta_r = x'w\eta_i + \left[\frac{(y_i-\mu_i)x_k}{\frac{1}{w}\left(\frac{\partial\mu}{\partial\eta}\right)_i^2}\left(\frac{\partial\mu}{\partial\eta}\right)_i\right] \qquad \dots\dots\dots (26-2)$$

$$[x'wx]\beta_r = x'w\eta_i + \left[x_k w(y_i-\mu_i)\left(\frac{\partial\eta}{\partial\mu}\right)_i\right] \qquad \dots\dots\dots (27-2)$$

نفرض z :

$$z_i = \eta_i + (y_i-\mu_i)\left(\frac{\partial\eta}{\partial\mu}\right)_i$$

لدينا

$$[x'wx]\beta_r = x'wz$$

وبالتعويض

$$\beta_r = [x'wx]^{-1}x'wz \qquad \dots\dots\dots(28-2)$$

(2-5-2) طريقة المربعات الصغرى الموزونة [2](weighted least squares)

تستعمل هذه الطريقة لزيادة كفاءة مقدرات المربعات الصغرى عند وجود مشكلة عدم تجانس التباين ، ويمكن ان نحصل على تقدير لمعلمات الانحدار من خلال أوزان تعطى لكل مشاهدة وذلك حسب تباين البواقي اذ ان الفرق بين قيم الحد الثابت والميل الحدي المقدره للنموذج هو المقدار ($\sqrt{w_i}$) .

$$\begin{bmatrix} \sqrt{w_1}Y_1 \\ \sqrt{w_2}Y_2 \\ \vdots \\ \sqrt{w_n}Y_n \end{bmatrix} = \begin{bmatrix} \sqrt{w_1} & \sqrt{w_1}X_{12} & \cdots & \sqrt{w_1}X_{1k} \\ \sqrt{w_2} & \sqrt{w_2}X_{22} & \cdots & \sqrt{w_2}X_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ \sqrt{w_n} & \sqrt{w_n}X_{n2} & \cdots & \sqrt{w_n}X_{nk} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \sqrt{w_1}U_1 \\ \sqrt{w_2}U_2 \\ \vdots \\ \sqrt{w_n}U_n \end{bmatrix}$$

ويمكن ان نعبر عن اعلاه بشكل مختصر وكالاتي :

$$p^{-1}Y = p^{-1}X\beta + p^{-1}U \qquad \dots\dots\dots(29-2)$$

$$p^{-1} = \begin{bmatrix} \sqrt{w_1} & 0 & 0 & 0 \\ 0 & \sqrt{w_2} & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \sqrt{w_n} \end{bmatrix}$$

$$p = \begin{bmatrix} 1/\sqrt{w_1} & 0 & 0 & 0 \\ 0 & 1/\sqrt{w_2} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 1/\sqrt{w_n} \end{bmatrix}$$

ويمكن ان نحصل على :

$$pp' = \begin{bmatrix} 1/w_1 & 0 & 0 & 0 \\ 0 & 1/w_2 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 1/w_n \end{bmatrix}$$

$$W^{-1} = (pp')^{-1} = \begin{bmatrix} w_1 & 0 & 0 & 0 \\ 0 & w_2 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & w_n \end{bmatrix}$$

في حالة تحقق الفروض الخاصة بتطبيق المربعات الصغرى الاعتيادية وذلك لان

$$E(U.U') = E[(p^{-1}U)(p^{-1}U)'] = E(p^{-1}UU'p'^{-1}) = p^{-1}E(UU')p'^{-1}$$

$$= p^{-1}\sigma_u^2 wp'^{-1} = \sigma_u^2 p^{-1}pp'p'^{-1}$$

$$E(U.U') = \sigma_u^2 I_n I_n = \sigma_u^2 I_n \qquad \ldots\ldots\ldots(30-2)$$

اي تم تحقق الفروض الخاصة بتجانس تباين الخطا وعدم وجود الارتباط الذاتي ، ان الفرضيات الخاصة بنموذج الانحدار قد تحققت وبالتالي يمكن تطبيق طريقة المربعات الصغرى الاعتيادية في تقدير موجه معالم النموذج وكالاتي :

$$p^{-1}y = p^{-1}X\beta + p^{-1}U$$

$$p^{-1}U = p^{-1}Y - p^{-1}X\beta$$

$$(p^{-1}U)'(p^{-1}U) = (p^{-1}Y - p^{-1}X\beta)'(p^{-1}Y - p^{-1}X\beta)$$

$$= (Y'p'^{-1} - \beta'X'p'^{-1})(p^{-1}Y - p^{-1}X\beta)$$

$$= Y'p'^{-1}p^{-1}Y - Y'p'^{-1}p^{-1}X\beta - \beta'X'p'^{-1}p^{-1}Y + \beta'X'p'^{-1}p^{-1}X\beta$$

$$(U'W^{-1}U) = Y'W^{-1}Y - 2Y'W^{-1}X\beta - \beta'X'W^{-1}X\beta$$

$$= Y'W^{-1}Y - 2\beta'X'W^{-1}Y + \beta'X'W^{-1}X\beta \qquad \ldots\ldots\ldots(31-2)$$

وباخذ المشتقة بالنسبة الى β' نحصل على :

$$\frac{\partial(U'W^{-1}U)}{\partial\beta'} = -2X'W^{-1}Y + 2X'W^{-1}Xb_{WLS} = 0$$

$$X'W^{-1}Y = X'W^{-1}Xb_{WLS}$$

$$b_{WLS} = (X'W^{-1}X)^{-1}X'W^{-1}Y \qquad \ldots\ldots\ldots(32-2)$$

(5-2) معايير المقارنة [6][3](Comparative Criterions)

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

$$MSE = \frac{SSE}{n-p} \qquad \ldots\ldots\ldots(33-2)$$

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

$$SSE = \sum_{i=1}^{n} u_i^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \qquad \ldots\ldots\ldots(34-2)$$

❖ معامل التحديد (Coefficient of Determination)

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

$$\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}(y_i-\hat{\theta}_i)^2+\sum_{i=1}^{n}(\hat{\theta}_i-\bar{y})^2+2\sum_{i=1}^{n}(y_i-\hat{\theta}_i)(\hat{\theta}_i-\bar{y}) \quad \dots(35-2)$$

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 .

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

3- الجانب التطبيقي

(1-3) 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 (1-3) .

الجدول (1-3) 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 الخلقي (موه الرأس الخلقي)	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

13	UNDEACENDED TESTIS	Q53	خصية غير نازلة	14
14	HYPOSPADIAS AND EPISPADIAS	Q64	فتحة احليل فوقانية او تحتانية	15
15	OTHER ANOMALIES OF GENITO – URINARY ORGANS	Q52,Q54, Q55	العيوب الخلقية الاخرى في الاعضاء التناسلية	16
16	CONGENITAL ANOMALIES OF THE SKIN	Q82	العيوب الخلقية في الجلد	17
17	ANAL STENOSIS	Q42	تضيق فتحة الشرج	18
18	Other congenital malformation of the digestive system	Q38–Q41,Q43 –Q45	تشوهات خلقية بالجهاز الهضمي	19
19	CONGENITAL ANOMALIES OF THE EYE	Q15	العيوب الخلقية في العين	20
20	ACCESSORY AURICLE	Q17	صيوان الأذن الاضافي	21
21	CONGENITAL ANOMALIES OF UPPER LIMB	Q71	العيوب الخلقية في الطرف العلوي	22
22	CONGENITAL ANOMALIES OF LOWAR LIMB	Q72	العيوب الخلقية في الطرف السفلي	23
23	Other congenital malformations of face and neck	Q18	تشوهات خلقية اخرى بالوجة والرقبة	24
24	Other Congenital malformations of respiratory system	Q34	تشوهات خلقية اخرى بالجهاز التنفسي	25
25	Congenital malformations of musculoskeletal system , not elsewhere classified	Q79	تشوهات خلقية بالجهاز العضلي الهيكلي غير مصنفة في مكان اخر	26
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 X_i ويتألف الجدول(2-3) 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 ورمزه	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 -:

لتقدير معلمات نموذج انحدار ثنائي الحدين السالب تم تطبيق طرائق المربعات الصغرى الموزونة المكررة IRLS وطريقة المربعات الصغرى الموزونة WLS .

(1-4-3) طريقة المربعات الصغرى المعادة الوزن التكرارية (IRLS)

يستعمل فيها البرنامج الأحصائي (STATA) لتقدير وأختبار معلمات إنموذج انحدار ثنائي الحدين السالب علما ان فرضية العدم هي :

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

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

أذ كانت قيمة معامل التحديد (%19) ومتوسط مربعات الخطأ (54.73) وكانت النتائج كما يلي :

جدول رقم (3-3) يبين مقدرات قيم معاملات الأنحدار لطريقة IRLS

المتغيرات المستقلة	معاملات الأنحدار	الخطأ المعياري	P-value
X1 (عمر الأم)	-0.0829436	0.1041408	0.427
X2 (مهنة الأم)	-2.869986	4.070355	0.481
X3 (عمرالأب)	0.0552933	0.0870725	0.526
X4 (مهنة الاب)	-0.9558305	0.4748609	0.045
X5 (درجة القرابة)	-0.6986716	1.236127	0.527
X6 (الولادات السابقة)	3.061872	1.279343	0.017
X7 (نوع الولادة)	-5.0951	2.809135	0.071
X8 (الولادة الحاليه)	-3.5937	1.361783	0.009
X9 (تعرض الأم)	-1.433926	0.4265287	0.001
X10 (عدد الأسقاط)	-0.4209224	1.074158	0.969
X11 (نوع السكن)	0.3859155	1.142004	0.736
X12 (جنس الطفل)	0.7500499	1.014451	0.460
X13 (وزن الطفل)	0.0011191	0.0007725	0.149
الحد الثابت	23.76104	7.469959	0.002

من الجدول (3-3) تبين وجود خمسة عوامل أظهرت تأثيرأ معنويا على المتغير المعتمد(نوع العوق) اي ان قيم p – value هي أقل من 0.05 وهذه العوامل هي : مهنة الأب x_4 ، وجود العوق في الولادات السابقة x_6 ، والولادة الحالية حية او ميتة x_8 وتعرض الأم أثناء فترة الحمل x_9 ، أما بقية العوامل فكانت غير معنوية أي ليس لها تأثير على المتغير المعتمد (y) . **(2-4-3)**

طريقة المربعات الصغرى الموزونة (WLS)

بهذه الطريقة لتقدير وأختبار معلمات نموذج أنحدار ثنائي الحدين السالب أستعملنا البرنامج الأحصائي SPSS ، علما ان فرضية العدم :

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

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

وبأستعمال البرنامج الأحصائي (mini tab 17) حيث كان معامل التحديد (%17.5) ومتوسط مربعات الخطأ (62.76) وكانت النتائج كما يلي :

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

المتغيرات المستقلة	معاملات الأنحدار (b)	الخطأ المعياري	P-value
X1 (عمر الأم)	-0.066	0.091	0.466
X2 (مهنة الأم)	-2.087	1.902	0.274
X3 (عمرالأب)	0.037	0.080	0.650
X4 (مهنة الاب)	-0.672	0.440	0.128
X5 (درجة القرابة)	-0.357	1.104	0.747
X6 (الولادات السابقة)	2.724	1.221	0.027
X7 (نوع الولادة)	-4.752	1.975	0.017
X8 (الولادة الحاليه)	-2.903	1.141	0.012
X9 (تعرض الأم)	-1.545	0.385	0.000
X10 (عدد الأسقاط)	-0.869	1.075	0.419
X11 (نوع السكن)	0.206	1.010	0.383
X12 (جنس الطفل)	0.762	0.866	0.379
X13 (وزن الطفل)	0.001	0.001	0.057
الحد الثابت	21.011	5.593	0.000

تم تقدير معلمات النموذج ، ويتضح من الجدول (3-4)ان خمسة من المتغيرات المستقلة لها تأثير واضح في نوع العوق وهي : وجود العوق في الولادات السابقة ، نوع الولادة مفرده او متعددة ، والولادة الحالية حية ام ميتة ، تعرض الأم الى حمى اشعاع تناول ادوية ، وزن الطفل . اما بقية العوامل فليس لها تاثير مباشر .
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 التقديرIRLS,WLS

المعيار	IRLS	WLS
MSE	54.73	62.76
R^2	%19	%17.5

\ يظهر ان طريقة المربعات الصغرى المعادة الوزن التكرارية **IRLS** هي أفضل طريقة للتقدير ومن خلال النتائج التي تطرقنا لها حيث تمتلك أقل متوسط مربعات للخطأ **(54.73)** وأعلى معامل تحديد أذ بلغ **(%19)** أما طريقة المربعات الصغرى الموزونه WLS بمتوسط مربعات الخطأ (62.76) ومعامل تحديد (%17.5)

4- الاستنتاجات والتوصيات

(4-1) الاستنتاجات :

1- تم التوصل الى ان توزيع ثنائي الحدين السالب افضل توزيع لبيانات حديثي الولادة الذين يعانون من التشوهات الخلقيه هو.
2- طريقة تقدير المربعات الصغرى التكرارية (IRLS) هي افضل طريقة معلمات الانموذج.

(2-4) التوصيات

1. ضرورة استعمال نموذج انحدار ثنائي الحدين السالب عند دراسة البيانات التي تحتوي خاصية فرط التشتت overdispertion اي يكون فيها التباين اعلى من الوسط الحسابي
2. ضرورة تطبيق طريقة المربعات الصغرى المعادة الوزن التكرارية في حالة انموذج ثنائي الحدين السالب فهي افضل من طريقة المربعات الصغرى الموزونة عند تطبيقها على نموذج ثنائي في الحدين السالب .
3. ضرورة البحث في البيانات العددية count data وبناء النماذج الخاصة بها وخصوصا البيانات الطبية منها لافتقار البحوث الحديثة لهذه المواضيع والظواهر المتعلقة بها .
4. ا- زيادة الوعي الصحي من خلال الأعلانات والنشرات الصحية.

ب- كذلك تحديث بيانات دوائر الصحة وعدم الأعتماد على السجلات القديمة التي أصبح دورها ضعيف جدا في حدوث التشوهات الخلقية.

ج- وتفعيل دور فرق البيئة ومكافحة كل أشكال التلوث البيئي.

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. كاظم ، أموري هادي " القياس الأقتصادي المتقدم (النظرية والتطبيق) ، بغداد (2002) مطبعة الطيف .
3. 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

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)