جمهورية العراق

وزارة التعليم العالي والبحث العلمي

جامعة كربلاء

كلية الادارة والاقتصاد

أختيار أفضل تقدير نموذج إنحدار لتوزيع ذي الحدين السالب مع تطبيق عملي

رسالة مقدمة الى

مجلس كلية الادارة والاقتصاد / جامعة كربلاء

وهي جزء من متطلبات نيل درجة ماجستير

علوم في الاحصاء

من قبل

عدنان فاضل طعمه

باشراف

الاستاذ المساعد الدكتور

شروق عبدالرضا سعيد

1439 هجري **2018 م**

Repub

بسم الله الرحمن الرحيم

قَالُوا

سُبْحَانَكَ لاَ عِلْمَ لَنَا إلاَّ مَا عَلَّمْتَنَا

إنَّكَ أنْتَ

العَلِيْمُ الحَكِيْمُ

صَدَقَ اللهُ العَلِيُّ العَظيم

(البقرة 32)

الأهداء

الى من بلغ الرسالة وادى الامانة ونصح الأمة نبي الرحمة ونور العالمين محمد(ص)...

الى ملاذي واماني وملجئي عند شدتي من تشرفت وآنست بالسكن بجوارهم ابا عبد الله الحسين وابا الفضل العباس (عليهما السلام) ...

الى كل دم طاهر سال دفاعا عن ارض بلاد الرافدين ...

الى امي الغالية والدي الحبيب لكما كل التجلي والاحترام ...

الى سندي في الحياة زوجتي وبناتي الحبيبات ...

الى اخوتي واخواتي واصدقائي وكل من وقف بجانبي ودعمني ...

اهدي وبأنحناء جهدي المتواضع هذا ...

ع . ف .ط

الشكر والامتنان

الحمد لله الذي انعم فزاد في النعم , وأكرم ففاض في الكرم , وعلم الأنسان مالم يعلم والصلاة والسلام على رسولة الكريم محمد "صلى اللة عليه وعلى اله وصحبة وسلم " وبعد ...

يدفعني واجب الوفاء والعرفان الى ان أقدم من الشكر أجزله والتقدير ارفعه والامتنان اعظمه للأستاذة المشرفة الدكتورة " شروق عبد الرضا " التي قدمت لي كل ما يمكن ان تقدمة من دعم علمي وعون فكري نابعين من ضمير وأصالة لما بذلتة من متابعة مستمرة في توجيهي , لها مني جزيل الشكر والعرفان وجزاها اللّٰه عني خير الجزاء .

واتقدم بالشكر الجزيل الى السيدات والسادة اعضاء لجنة المناقشة على تفضلهم بقبول مناقشة رسالتي وعلى ماسيبذلوه من جميل ملاحظاتهم القيمة والتي لاغنى لي عنها .

وعرفانا مني بالجهود التي قدمت في سبيل أخراج هذة الرسالة بشكلها النهائي الى كل من مد يد العون لي وآزرني في أعدادها من أساتذة وموظفي قسم الأحصاء المحترمين وزملاء وزميلات الدراسة فجزى اللّٰه الجميع عني أوفى الجزاء وأحسنة أنه سميع مجيب الدعاء .

وختاما أتقدم بأسمى آيات الشكر والمحبة والامتنان لكل القلوب التي دعت لي بالتوفيق .

ع . ف . ط

قائمة الرموز

الرمز	التفاصيل
Y	متغير الاستجابة
X	متجة المتغيرات المستقلة
β	متجة المعلمات المجهولة
μ_{ij}	متوسط ال y_{ij}
α	معلمة التشتت
K	هي المعلمة اللاخطية تسمح بأخذ اي قيمة مستمره
μ=k	
$V=\frac{1}{\alpha}$	
Hessian	المشتقات الجزئية الثانية لدالة الأمكان اللوغارتمية
مصفوفة سالب حسن Hessian negative matrix	observed information matrix مصفوفة المعلومات المشاهدة
Gradient	المشتقات الجزئية الاولى لدالة الامكان اللوغارتمية
$\psi_1 = \psi'_{(x)}$	مشتقة دالة كاما اللوغارتمية (digamma) دالة كاما الثنائية
$\psi_2 = \psi''_{(x)}$	المشتقة الثانية لدالة كاما اللوغارتمية (trigamma) دالة كاما الثلاثية
θ_i	المعلمة القانونية او دالة الربط (link function)
$b(\theta_i)$	معلمة الموقع (cumulant)
$\alpha(\phi)$	معلمة القياس (scale parameter)
$c(y_i, \phi)$	حد التطبيع (normalization)
θ	معلمة الموقع (location parameter)
ϕ	معلمة القياس (scale parameter)
η	معكوس دالة الربط (link function inverse)
$s(h, z)$	دوال التسجيل (usual score functions)

المستخلص

يعد إنموذج أنحدار ذي الحدين السالب أحد النماذج العددية التي تستعمل لتمثيل بعض الظواهر والحالات التي لايمكن التعبير عنها بالنماذج الأعتيادية والذي يأتي من أنموذج(بواسون – كاما) المركب ويتحقق عندما يكون التباين أكبر من المتوسط للبيانات .

إذ تم تقدير معلمات إنموذج أنحدار ثنائي الحدين السالب بأربع طرائق مختلفة هي [طريقة الأمكان الأعظم (MLE) Maximum Likelihood Estimation وطريقة المربعات الصغرى المعادة الوزن التكرارية (الحصينة) Iteratively Re-weighted Least Square (IRLS) وطريقة الأمكان الموزونة (WLE) Weighted Likelihood Estimation وطريقة المربعات الصغرى الموزونة (WLS) Weighted Least Square] 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 . وتم أستعمال البرامج الأحصائية (stata , spss ,minitab) لتقدير معلمات إنموذج أنحدار ثنائي الحدين السالب وتحديد أفضل طريقة , وقد بينت النتائج أن طريقة المربعات الصغرى المعادة الوزن التكرارية (الحصينة) هي الأفضل كونها أمتلكت أقل متوسط مربعات للخطأ MSE وأعلى معامل تحديد R^2

الفصل الاول

المقدمة ومنهجية البحث

1 -1	المقدمة
2-1	مشكلة الدراسة
3-1	أهمية البحث
4-1	منهجية البحث
5-1	هدف البحث
6-1	الأستعراض المرجعي

(1-1) المقدمة (introduction)

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 لإنموذج 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 إذ البيانات تمتلك خاصية فرط التشتت (overdispersion) , سوف نلجأ الى إنموذج ثنائي الحدين السالب , اي إنموذج بواسون – كاما المركب (poisson – gamma mixture model) اكثر ملائمة في حالة فرط التشتت . لهذا تناولنا هذا الإنموذج المهم والتطبيقات العملية وفي كافة المجالات الطبية والهندسية والحياتية .

أن إنموذج ثنائي الحدين السالب ياتي من إنموذج (بواسون – كاما) المركب بصورة تقليدية , فضلا عن ذلك قد يأتي إنموذج ثنائي الحدين السالب كعضو من توزيعات العائلة الاسية ذات المعلمة المفرده , هذه العائلة من التوزيعات التي تختص بما يسمى النماذج الخطية العامة (Generalized Linear Models)(GLM). ولغرض عرض وتوضيح مضمون وتفاصيل الرسالة اذ تم تقسيم الرسالة الى اربعة فصول :

الفصل الاول: تضمن مقدمة عن النماذج العددية بصورة عامة وإنموذج انحدار ثنائي الحدين السالب بصوره خاصة ,مشكلة البحث , هدف البحث , فضلا عن عرض البحوث والدراسات السابقة ضمن الاستعراض المرجعي المتناولة لإنموذج انحدار ثنائي الحدين السالب .

الفصل الثاني : تضمن الجانب النظري والذي شمل عرض الطرائق الأربعة لتقدير معلمات إنموذج انحدار ثنائي الحدين السالب وهي:

1. طريقة الامكان الاعظم (maximum likelihood estimation)
2. طريقة المربعات الصغرى المعادة الوزن التكرارية (الحصينة) (iteratively re-weighted least squares (IRLS))
3. طريقة الامكان الموزونة (weighted likelihood estimation)
4. طريقة المربعات الصغرى الموزونة (weighted least squares)

وتضمن بعض المفاهيم الاساسية المتعلقة بموضوع الرسالة إنموذج انحدار ثنائي الحدين السالب

الفصل الثالث : تطرق ألى عينة الدراسة ومعرفة التوزيع المناسب للبيانات , تقدير معلمات إنموذج انحدار ثنائي الحدين السالب بأربع طرائق (طريقة الأمكان الأعظم , طريقة المربعات الصغرى المعادة الوزن التكرارية , طريقة الأمكان الأعظم الموزونة , طريقة المربعات الصغرى الموزونة) وتم المقارنة بين الطرائق الأربعة .

الفصل الرابع : تضمن الاستنتاجات والتوصيات وبالاعتماد على الجانب التطبيقي يمكن النظر في الدراسات المستقبلية .

(2-1) مشكلة البحث (the problem of the research)

تقسم مشكلة البحث الى :-

1. مشكلة البحث النظرية : أفتقار البحوث الحديثة التي تعالج الحالات والظواهر التي لايمكن تمثيلها ألا عن طريق النماذج العددية (count models) ولاسيما إنموذج ثنائي الحدين السالب .
2. مشكلة البحث العملية : زيادة حدوث التشوهات الخلقية في المدة الأخيرة بسبب أوضاع العراق والحروب والظروف التي عاشها البلد الامر الذي يدفعنا لدراسة العوامل والاسباب التي ساعدت في حصول تلك التشوهات الخلقية , فضلا عن الظروف والحروب والتلوث

البيئي بكل انواعة والاشعاع وظهور التكنلوجيا ووسائل الاتصال والبث الكهرومغناطيسي وغيرها فضلا عن متغيرات البحث المستقلة , كانت عوامل مهمة يجب الوقوف عندها .

(1-3) أهمية البحث (the importance of the research)

تكمن أهمية البحث في ان نصل إلى فهم أعمق للنماذج العددية بصورة عامة وإنموذج أنحدار ثنائي الحدين السالب على وجه الخصوص وتقدير معلماتة بالطرائق الأربع : طريقة الأمكان الأعظم وطريقة المربعات الصغرى المعادة الوزن التكرارية وطريقة الأمكان الموزونة وطريقة المربعات الصغرى الموزونة ومعرفة افضل هذه الطرائق لتقدير المعلمات .فضلا عن اهمية مجال البحث وتطبيقة على الخدج المصابين بالتشوهات الخلقية والمنتشرة بصورة واسعة في المدة الاخيرة .

(1-4) منهجية البحث :

أعدت البحث ضمن المنهج الأستقرائي إذ تم سحب العينة ودراستها وتحليل النتائج وإعمامها على مجتمع الدراسة .

(1-5) فرضية البحث

يفترض البحث وجود تأثير معنوي لمتغيرات البحث (عمر الأم , مهنة الأم , عمر الأب , مهنة الأب , درجة القرابة بين الأبوين , الولادات السابقة , نوع الولادة , الولادات الحالية , تعرض الأم لبعض المؤثرات , عدد الاسقاط السابق , نوع السكن , جنس الطفل , وزن الطفل) على (نوع العوق) .

(1-6) هدف البحث (the objective of the research)

يرمي البحث الى:

1. دراسة توزيع ثنائي الحدين السالب وخصائصه وإنموذج الانحدار الخاص به .

2. تقدير معلمات إنموذج انحدار ثنائي الحدين السالب باستعمال بعض طرائق التقدير(طريقة الأمكان الأعظم , طريقة المربعات الصغرى المعادة الوزن التكرارية , طريقة الأمكان الأعظم الموزونة , طريقة المربعات الصغرى الموزونة) والمقارنة بين نتائج التقدير المختلفة وتحديد افضل طريقة لتقدير المعلمات .

(7-1) الاستعراض المرجعي لبعض الدراسات السابقة (review of literature)

- في عام 1907 كان الباحث (Gosset) [24]اول من كتب في اشتقاق ثنائي الحدين السالب في المختبر للقياس الحياتي في لندن والبحث عن خطأ المعاينة بالاعتماد على حساب عدد خلايا الخميرة .
- خلال عام 1920(Later Green wood and Yule [26] قدموا ثنائي الحدين السالب وهو احتمال حدوث y عدد حالات الفشل قبل r–th من النجاحات في سلسله من محاولات برنولي .
- في عام 1923(Eggen berger and polya) [17] قدموا فكرة خلط التوزيعات ومنها توزيع بواسون المركب للحصول على ثنائي الحدين السالب , بوضع معلمة بواسون λ كمتغير عشوائي لتوزيع كاما , وهو الاشتقاق الاول لثنائي الحدين السالب هو توزيع بواسون – كاما المركب , وقدموا معلمة بواسون لمربع كاي بدرجة حرية (2) .
- في عام 1943 (Fisher et al) [21] تناول توزيع (بواسون – كاما المركب)اذ ان معلمة بواسون تتبع توزيع كاما وعن طريق ضرب الدالتين وتكامل بالنسبة لحدود معلمة بواسون وهي متغير عشوائي يتبع توزيع كاما ينتج عنه توزيع ثنائي الحدين السالب.
- خلال الاعوام1942(George Beall) [13] و1947(Maurice Bartlett) [12] و1949(F.J.Anscombe) [9] خلال مدة الاربعينات من القرن الماضي عمل متواصل من لدن الباحثين ولاسيما على النماذج العددية , قاموا بتطوير مقاييس التحويل للبيانات التي لاتخضع للتوزيع الطبيعي لجعل هذه البيانات تتوزع توزيعا" طبيعيا" إذ قدم (Bartlett), تحليل لتحويل الجذر التربيعي على بيانات بواسون وذلك عن طريق اختبار التحويلات ثبات التباين للبيانات المشتتة .

- في عام 1950((Anscombe)[10]كان اول من قدم أنموذج انحدار ثنائي الحدين السالب كسلسله للتوزيعات اللوغارتميه وتناول اشتقاقات بديله مثل معاينة معكوس ثنائي الحدين , معاينة بواسون غير المتجانسة عندما λ تعد جزءا" من مربع كاي , ونموذج ثنائي الحدين السالب كأنموذج نمو المجتمع , ثنائي الحدين السالب مشتق من سلسلة هندسية .
- عام 1953(Evans)[19] طور مايعرف ثنائي الحدين السالب الخطي (negative binomial 1(NB1)) وجود علاقة خطية بين الوسط الحسابي والتباين) ومعلمات ثنائي الحدين السالب .
- في عام 1957 (Gurland)[27] تناول العديد من العلاقات بين التوزيعات العامة (generalized distributions) والتوزيعات المركبة (mixture distributions) والتوزيعات المكافئة (equivalent distributions) وأخذ بعض الأمثلة التي تناولت العلاقات بين التوزيعات المختلفة الثلاثة (العامة , المركبة , المكافئة) , من التوزيعات التي تطرق إليها البحث هو توزيع ثنائي الحدين السالب باعتباره توزيع بواسون المركب مع توزيع كاما , وسمي توزيع ثنائي الحدين السالب هذا بتوزيع باسكال (Pascal distribution) .
- في عام 1959(Jain) [33]قام بتشكيل توزيع ثنائي الحدين السالب الاولي إذ تم بحث عدد الوفيات لعدد المتعرضين للاصابة بالامراض , هذه الصياغه منبثقه من مبدأ معاينة ثنائي الحدين السالب .
- في عام 1961(Leroy Simon)[48] نشر خوارزمية تعظيم الاحتمال لثنائي الحدين السالب وبواسون فضلا عن مشاركته في دراسة جودة توافق ومدى مطابقة بيانات أنموذج ثنائي الحدين السالب .
- في عام 1970 (Sankaran)[46]قدر معلمة التوزيع المتقطع (بواسون – ليندلي المركب) بطريقة الامكان الاعظم وكانت دراستة على مجموعتين لها توزيع مركب (ليندلي – بواسون) تمثل المجموعة الأولى للبيانات عدد النساء العاملات في القذائف ذات الأنفجار الشديد أما المجموعة الثانية فهي عدد الأخطاء لأرقام عشوائية لمجموعات

مطبوعة وتمت دراسة مطابقة البيانات وملائمتها لتوزيع بواسون ليندلي المركب ومقارنتها مع توزيع ثنائي الحدين السالب والتوزيع الهرمتي وتوزيع بواسون لبيانات المجموعتين نفسها وتبين ان توزيع بواسون ليندلي المركب كان هو الأفضل في مدى ملائمة البيانات للتوزيع .

- خلال عام 1972 (John Nelder and R.W.M.Wedder and Burn)[42] شددوا على خواص التوزيعات والعلاقات بين التوزيعات فبدئوا بتطوير النماذج الخطية العامة GLMs و التركيز على النماذج غير الخطية .
- في عام 1974 (Bulmer)[14] قام العالم بدراسة على مجموعتين من البيانات الحقيقية , المجموعة الأولى هي عدد الحيوانات حرشفية الأجنحة والتي تم صيدها عن طريق فخ الضوء , اما المجموعة الأخرى تمثل عدد الفراشات من نوع ملايو المجمعة ومقارنة مجموعتي البيانات عن طريق مدى ملائمتها للتوزيعات (ثنائي الحدين السالب وتوزيع بواسون وتوزيع بواسون – اللوغارتمي الطبيعي المركب) وظهر ان البيانات كانت تلائم توزيع ثنائي الحدين السالب أكثر من بقية التوزيعات وتم تقدير معلمات التوزيع بطريقة الأمكان الأعظم .
- في عام 1981 (Shaban)[47] تناول (توزيع بواسون – كاوس العكسي) المركب والتطرق لبعض خصائص التوزيع المركب الناتج كالدالة المولده الأحتمالية والعزمين الأول والثاني وغيرها ثم تقدير معلمات التوزيع باستعمال طريقة الأمكان الأعظم , كذلك تم أستعمال مجموعتين من البيانات الحقيقية , المجموعة الأولى هي عدد الحوادث بين الرجال وتكرارها في مصنع ما , أما الأخرى هي عدد الحوادث وتكرارها الى 646 امرأه عاملة على مواد شديدة الأنفجار وتم قياس ملائمة المجموعتين للبيانات وللتوزيعات الآتية (توزيع ثنائي الحدين السالب , توزيع بواسون , توزيع بواسون كاوس العكسي المركب , توزيع بواسون ليندلي المركب) بأستعمال مربع كاي وعن طريق النتائج تبين ان توزيع ثنائي الحدين السالب كان افضل من بقية التوزيعات في ملائمة بيانات المجموعة الأولى وتوزيع ثنائي الحدين السالب وتوزيع بواسون كاوس العكسي المركب هما الأفضل ملائمة لبيانات المجموعة الثانية.

- خلال عام 1982(Nelder and McCullaph)[39] كان الباحث أول من تناول النماذج الخطية العامة (GLM) (generalized linear models) وكيفية أيجاد أنموذج انحدار ثنائي الحدين السالب واشتقاقه من توزيع ثنائي الحدين السالب, وتقدير معالمه ودراسة خصائصه .
- في عام 1987(Nelder)[43] استعمل نموذج ثنائي الحدين السالب لتحليل مصائد الحشرات في تصميم القطاعات المتداخل . ودراسة الخصائص الأحصائية لدالة شبة الأمكان الموسعة وفقا لهذا التصميم .
- عام 1988 (Kocherlakota)[37] أستعمل معالجة متطورة للتوزيعات المركبة وقدم توزيع بواسون ثنائي المتغير المركب (compound bivariate poisson distribution) وأفترض هناك متغيرين لهما دالة مولدة احتمالية مشتركة يمثل توزيع بواسون ومعلماتة تمثل الثوابت وتحديد المعلمة العشوائية التي تخص الأفراد واختيار(4) توزيعات لها التوزيع الأول توزيع كاما وينتج عنه توزيع ثنائي الحدين السالب الثنائي, والتوزيع الثاني هو التوزيع الطبيعي والذي نتج عنه التوزيع الهرمتي ثنائي الحدود , والثالث توزيع بواسون والذي نتج عنه توزيع نيمان من نوع (A) , والرابع فهو توزيع كاوس العكسي وينتج عنة توزيع بواسون كاوس العكسي المركب ثنائي الحدود , اذ تم التطرق لبعض الخصائص الأحصائية والدوال الأحصائية لكل توزيع ناتج ومنها توزيع ثنائي الحدين السالب ثنائي الحدود.
- في عام 1989(McCullaph and Nelder)[38] درسوا بالتفصيل النماذج الخطية العامة وخصوصا" أنموذج انحدار ثنائي الحدين السالب , وخصائصه الأحصائية والدالة الأحصائية لأنموذج انحدار ثنائي الحدين السالب .وتقدير معلمات الأنموذج .
- خلال عام 1992(Nelder)[40] طور هذا العالم النماذج الخطية التفاعلية العامة الى مايسمى بال(kk system) لتقدير معلمات ثنائي الحدين السالب بهيأة (GenStat Macro) .
- في العام نفسه 1992 (Dey & Chung)[16] أستعملوا توزيع ثنائي الحدين السالب متعدد المتغيرات , ودراسة الدالة الأحتمالية للتوزيع الذي تم الحصول عليها من دمج n

من توزيعات بواسون مع توزيع كاما , وتم التطرق لبعض الخصائص الأحصائيه لتوزيع ثنائي الحدين السالب المتعدد كالدوال الشرطيه في حالتي متعدد المتغيرات واحادي المتغيرات كذلك الداله المولده للعزوم العامليه وتقدير معلمات التوزيع بطريقة بيز.

- في عام 1993(Hilbe)[32] عمل اول تقدير أنموذج لثنائي الحدين السالب NB_2 (عند وجود علاقة تربيعية بين المتوسط والتباين) كجزء من خوارزمية الأنموذج الخطي العام , مع تطبيقة في البرامجيات stata and xplore.
- في عام 1993 (Zaky et al)[51] قاموا بدراسة عن التوزيعات المركبة وكان ثنائي الحدين السالب احد هذه التوزيعات إذ وجدوا (3) أساليب لربط التوزيعات وخلط التوزيعات , أذ تم خلط توزيع بواسون مع توزيع ثنائي الحدين السالب وأفترضوا ان معلمة بواسون تتبع توزيع ثنائي الحدين السالب ولمعرفة خصائص التوزيع المركب , أعتمدوا برنامج (pearson) للمنحنيات الذي يعتمد على ثابت يمكن حسابة من قيمة معاملي التفلطح والألتواء للتوزيع الناتج المركب .
- ضمن عام 1994(Nelder)[41] ترأس فريق من الاحصائيين وقاموا بالعمل على تطوير (GLIM)هوالنماذج الخطية التفاعلية العامة (generalized linear interactive models) والتطبيقات الجاهزه لتنفيذ نظرية النماذج الخطية العامة , تطبيقات النماذج الخطية التفاعلية العامة تسمح للمستخدمين بتقدير نماذج الخطية العامة لمجموعة محدده من اعضاء العائلة الاسية مثل ثنائي الحدين السالب , بواسون , ثنائي الحدين .
- في عام 2005c(Hilbe)[30] استعملوا تحليل ثنائي الحدين السالب التتابعي (squantial negative binomial analysis) لآلية إدارة الافات الحشرية والحد من خطورتها.
- في عام 2005 (Karlis)[36] أستعمل خوارزمية تعظيم التوقع (expectation maximization) لتوزيع ثنائي الحداين السالب وتوزيعات بواسون المركبة الاخرى بأفتراض ان المتغير العشوائي يتبع توزيع بواسون والمعلمة تسلك سلوك المتغير العشوائي وسميت التوزيعات الناتجة بعائلة بواسون للتوزيعات المختلطة (mixing distributions) فضلا" عن تناول الباحث الكثير من التوزيعات المركبة مع توزيع

بواسون (كالتوزيع الأسي والتوزيع الناتج هو التوزيع الهندسي وهو حالة خاصه من توزيع ثنائي الحدين السالب , توزيع كاما والتوزيع الناتج هو ثنائي الحدين السالب , توزيع بواسون ليندلي المركب مع التوزيع الطبيعي والتوزيع الناتج هو التوزيع الهرمتي , توزيع بواسون مع التوزيع اللوغارتمي الطبيعي وينتج عنه توزع بواسون – اللوغارتمي الطبيعي المدمج , تم أستعمال خوارزمية تعظيم التوقع في كل التوزيعات المذكورة لتقدير معلماتها بطريقة الأمكان الأعظم .

- في عام 2006(Faraway)[20] كتب العالم في موضوع ثنائي الحدين السالب وبين انه انموذج خطي عام حقيقي . وتناول خصائص الأنموذج الأحصائية وتقدير معلمات أنموذج ثنائي الحدين السالب .
- في عام2009(Jones et al)[34] تم تطبيق أنموذج ثنائي الحدين السالب على اعداد الطفيليات حسب نوع المضيف (host species) في مقارنة بين المضيف الوسطي والمضيف النهائي .
- عام 2009 (Ghitany & AL-Mutairi)[22] تم تقدير معلمات توزيع ثنائي الحدين السالب بطريقتي الأمكان الأعظم والعزوم وكانت مقدرات الطريقتين متسقة ومتقاربة طبيعيا".
- في عام 2010_b(Hilbe)[29] حللوا الاستجابة السمية المزمنة (chronic toxicity response) لبعض السموم باستعمال نموذج ثنائي الحدين السالب .
- في عام 2010 (Ozel & Inal)[45] أستعملوا خوارزمية لصياغة دالة احتمالية لتوزيع ثنائي الحدين السالب (توزيع بواسون كاما المركب) وطبقت الخوارزميه على بيانات حقيقية لحوادث السير .
- عام 2011 (Hilbe)[31] تناول أنموذج إنحدار ثنائي الحدين السالب وأشتقاقة كتوزيع من بواسون – كاما المركب وكعضو من العائلة الأسية كذلك وبأستعمال اللوغارتم الطبيعي لدالة الأمكان الأعظم اوجد المشتقة الأولى التي تسمى gradient والمشتقة الثانية التي تدعى مشتقة حسن , إذ تم تقدير معالم أنموذج إنحدار ثنائي الحدين السالب بطريقتين (طريقة الأمكان الأعظم , وطريقة المربعات الصغرى الموزونة المكررة)

وتطرق لكل أنواع نموذج ثنائي الحدين السالب ومنها (NB1 , NB2) إذ تعد (NB1 وجود علاقة خطية بين المتوسط والتباين , NB2 وجود علاقة تربيعية بين المتوسط والتباين), كذلك عمل على معيار أكاكي للمعلومات AIC وعده أحد أكثر المعايير أستعمالا" لمقارنات الاحصاءات , كذلك عمل على معيار بيز للمعلومات BIC كمعيار مهم اخر للمقارنات

- عام 2013 (Cameron)[15] عمل على تقدير معلمات أنموذج انحدار ثنائي الحدين السالب بطريقة الأمكان الأعظم فضلا" عن أيجاد المشتقة الأولى لدالة الأمكان اللوغارتمية كذلك أعطى (asymptotic distribution) لمقدرات الأمكان الاعظم كمتعدد متغيرات طبيعي (multivariate normal) وطبقها على بيانات حقيقية على أمراض السرطان (melanoma) بمنطقتين مختلفتين وبأعمار مختلفة 12 فئة عمرية .
- عام 2015 (Coffie Emmanual)[18] عمل مقارنة بين أنموذج إنحدار ثنائي الحدين السالب أو أنموذج إنحدار بواسون ونماذج لي كارتر (Lee Carter models) للتنبؤ بمعدلات وفيات النرويجيين الذكور , اي مقارنة عملية الدقة في التنبؤ بالأعتماد على بيانات سابقة للوفيات لأيجاد الأنموذج الأكثر دقة في التنبؤ لوفيات الذكور النرويجيين .
- عام 2017 (محمد)[3] تم تقدير معلمات توزيع بواسون المركب بأربع طرائق وهي طريقة الأمكان الأعظم , طريقة العزوم , طريقة تعظيم التوقع EM , خوارزمية داونهيل وقد تم مزج توزيع بواسون مع توزيع كاما ونتج عنه توزيع ثنائي الحدين السالب وتم تقدير معالم التوزيع الناتج بطرائق التقدير .

الفصل الثاني

الجانب النظري

تمهيد	1-2
مفاهيم اساسية	2-2
توزيع ثنائي الحدين السالب	1-2-2
نماذج الانحدار	2-2-2
النماذج الخطية العامة	3-2-2
نماذج الاستجابة العددية	4-2-2
فرط التشتت	5-2-2
تعريف أنموذج انحدار ثنائي الحدين السالب	6-2-2
انحدار ثنائي الحدين السالب	7-2-2
اشتقاق توزيع ثنائي الحدين السالب	3-2
طرائق التقدير	4-2
طريقة الامكان الاعظم	1-4-2
طريقة المربعات الصغرى المعادة الوزن التكرارية	2-4-2
طريقة الامكان الموزونة	3-4-2
طريقة المربعات الصغرى الموزونة	4-4-2
معايير المقارنة	5-2

(1−2) تمهيد (preface)

في هذا الفصل نتناول المفاهيم الاساسية عن نماذج الاستجابة العددية بصورة عامة وإنموذج ثنائي الحدين بصورة خاصة إذ تم التعرف على توزيع ثنائي الحدين السالب وكيفية تحويله الى إنموذج انحدار ثنائي الحدين السالب فضلا عن تقدير معلمات الإنموذج بثلاث طرائق مختلفة.

(2−2) مفاهيم اساسيه (basic concepts)

(1−2−2) توزيع ثنائي الحدين السالب [31] [50] (negative binomial distribution)

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 , 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 حالات الفشل قبل حدوث r من النجاحات ومن هنا جاءت تسميته بالسالب.

نفرض y يتوزع ثنائي الحدين السالب بمعلمات α و θ

α : معلمة التشتت

θ : معلمة الموقع

Γ(.) : ترمز لدالة كاما

اذ ان : $y \sim NB(\alpha, \theta)$

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

وقد يرمز الى المقدار $^{1}/_{1+\theta}$ بالرمز p اي ان $p = {}^{1}/_{1+\theta}$ كذلك $\varphi = 1/\theta$.

واذا كان α عدد صحيح سمي التوزيع بتوزيع باسكال , أما التوزيع الهندسي (geometric distribution) فهو حالة خاصة من توزيع ثنائي الحدين السالب عندما α تساوي (1) يستعمل توزيع ثنائي الحدين السالب في التطبيقات الاقتصادية بصياغة بديلة تجنبا لمقادير الكاما ولنجعلها اكثر ملائمة للاغراض الحسابية , باستعمال :

$$p(Y = y) = (1+\theta)^{-\alpha} \times \begin{cases} \prod_{j=1}^{k} \dfrac{\theta(\alpha + j - 1)}{(1+\theta)j} & \text{for } y = 1,2,\ldots \\ 1 & \text{for } y = 0 \end{cases} \qquad (2-2)$$

نفرض X متغير عشوائي معرف على اعداد صحيحة غير صفرية , الدالة المولده الاحتمالية p تعطى بمتعددة الحدود هذه :

$$p^{(Y)}(s) = p_0 + p_1 s + p_2 s^2 + \cdots = \sum_{j=0}^{\infty} p_j s^j = E(s^Y) \qquad \ldots.(3-2)$$

دالة $p(s)$ تعرف ب$p_j's$ مفكوك متعددة الحدود (polynomial expansion) وتعد وحيدة (unique) .

معلمتي التوزيع لها دالة مولدة احتمالية (probability generating function) :

$$p(s) = E(s^Y) = [1 + \theta(1-s)]^{-\alpha} \qquad \ldots\ldots(4-2)$$

المتوسط والتباين :

$$E(Y) = p'(1) = \alpha\theta \quad ; \quad V(Y) = \alpha\theta(1+\theta) \qquad \ldots\ldots(5-2)$$

اذ ان : $\theta > 0$, تباين توزيع ثنائي الحدين السالب دائما يفوق الوسط الحسابي اي توجد حالة فرط التشتت (over dispersion) لكن فرط التشتت يتلاشى في غاية (limit) $\theta \to 0$.

(2-2-2) نماذج الأنحدار[4] (regression models)

في العديد من الدراسات العلمية , يقع الاهتمام على العلاقة بين اثنين او اكثر من مجاميع المشاهدات , تحليل الانحدار يسمح بتقدير التغير في كمية او نوع من المشاهدات بمتغير معتمد (y) كدالة لمتغير (x) ويسمى المتغير المعتمد بمتغير الاستجابة (response variable) .

في الإنموذج الاحصائي قيمة (y) هي عشوائية بقيم المشاهدات نتيجة المعاينة او الاختلاف الطبيعي للمجتمع (x) تعرف بالمتغيرات , التوزيع الشرطي لل(y) بالنسبة الى (x) يدرس مجموعة من النقاط (i=1,…,n) بالنسبه للyi و xi , المجموعة p وهي المتغيرات المستقلة المستعملة في الأنموذج والمعادلة الخطية يمكن التعبير عنها :

$$\beta_1 x_{i1} + \cdots + \beta_p x_{ip} \qquad (6-2)$$

اذ ان β_p . … . β_1 مقدرات معلمات لتاثير x_{ip} … … … x_{i1} على (y) .

في إنموذج الانحدار الخطي التقليدي متغير الاستجابه يتم وصفه عن طريق الكثافه الاحتمالية الشرطية لل(y) بالنسبه للتركيبة الخطية :

$$\underline{y} = x\underline{\beta} + \underline{u} \qquad \dots\dots (7-2)$$

اذ ان $\underline{y} = (y_1 \dots\dots\dots y_n)^T$ هي استجابات مستمره والبواقي $\underline{u} = (u_1 \dots\dots\dots u_n)^T$ تتوزع توزيعا طبيعيا بصورة مستقلة بمتوسط (صفر) وتباين ثابت(σ^2) لكل الوحدات $u_i \sim N(0, \sigma^2)$

وإنموذج الانحدار الخطي التقليدي نستطيع ان نعرفه بانه التوقع الشرطي للاستجابات μ_i بالنسبه للتركيبة الخطية :

$$\mu_i = E\left(y_i / \beta, \sigma, x_i^T\right) = x_i^T \beta \qquad \dots\dots (8-2)$$

اذ ان y_i تتوزع طبيعيا بمتوسط μ_i وتباين (σ^2).

(3-2-2) النماذج الخطية العامة [4] (generalized linear models)

نماذج الانحدار الخطية هي جزء من النماذج الاحصائية وتعرف باسم النماذج الخطية العامة (GLM) هذه النماذج ترتبط معا لمتغيرات الاستجابه العشوائية كتركيبة خطية نظامية وهي تتضمن نماذج بافتراض العلاقة الخطية او اختلافات طبيعية لمتغير الاستجابة ربما تكون غير ملائمة مثل العلاقة الخطية اللوغارتمية , او الاستجابة العددية لبواسون , وهنا يمكن القول لدينا تركيبتين أساسيتين لتشكيل النماذج الخطية العامة :

1- تركيبة عشوائية تحدد المعلمة الطبيعية θ_i .

عندما توزيع متغير الاستجابة هو عضو لتوزيعات العائلة الاسية الطبيعي , اي يمتلك دالة كثافة احتمالية كما يأتي :

$$f(y;\theta,\Phi)=\exp\left[\frac{y_i\theta_i-b(\theta_i)}{\alpha_i(\Phi)}+c(y_i;\Phi)\right] \quad \ldots\ldots(9-2)$$

إذ ان :

θ_i هي المعلمه القانونيه او دالة الربط (link function)

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

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

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

2- تركيبه نظامية متعلقه بالمتجه $\eta=(\eta_1\ldots\ldots\ldots\eta_p)^T$

بقيم توضيحية ضمن الأنموذج الخطي , باستعمال التركيبة الخطية $\eta_i=x_i^T\beta$

دالة الربط $g(.)$ للربط العشوائي بالمركبات النظامية عندما μ_i بالη_i بواسطة $\eta_i=g(\mu_i)$ ويمكن ان نعبر عن المعادلة للمتغيرات التوضيحية $g(\mu_i)=x_i^T\beta$.

(4-2-2) نماذج الاستجابة العددية [50] (count response models)

تعد نماذج الاستجابة العدديه جزءا" من نماذج انحدار الاستجابة المتقطعة (discrete response regression models), وهذه النماذج تعرف بالاستجابات الصحيحه غير السالبة (non−negative integer responses) ولدينا امثلة عن النماذج المتقطعة هي :

الثنائي binary : الانحدار اللوجستي الثنائي (binary logistic regression) والانحدار الاحتمالي (probit regression) .

النسبي (proportional) : الانحدار اللوجستي التجميعي (grouped logistic) .

الرتبي (ordered) : الانحدار اللوجستي الرتبي (ordinal logistic regression) والانحدار الاحتمالي الرتبي. (ordered probit regression) .

متعدد الحدود (multinomial) : الانحدار اللوجستي الاختياري المتقطع (discrete choice logistic regression)

العددية (count) : انحدار بواسون , انحدار ثنائي الحدين السالب .

الاستجابة العددية تتكون من استجابات عددية متقطعة مثل : ايام الرقود في المستشفى للمرضى كذلك الاهداف المسجلة في المسابقات , جميع النماذج العددية تهدف لتوضيح عدد الحدوث اي انها تتعامل مع أعداد وهذه الأعداد بطبيعتها لاتملك تجانسا" أو تعاني من التواء أيمن أو تمتلك تباينا" أكبر من المتوسط . اذ ان المتغير العشوائي (y) هو متغير الاستجابة العددية , كذلك المعلمة (λ) تعد المتوسط بالنسبة لإنموذج بواسون , إنموذج بواسون يعد الإنموذج الاساس بالنسبة للنماذج العددية وبقية النماذج العددية تعتمد عليه اعتمادا" كبيرا" , إنموذج بواسون هو :

$$f(y) = \frac{e^{-\lambda_i}(\lambda_i)^{y_i}}{y_i!} \qquad y_i = 0, 1, \dots, n_i \quad ; \ \lambda > 0 \qquad \dots (10-2)$$

إنموذج بواسون لايمتلك معلمة قياس محددة فضلا عن ان القياس يفرض يساوي واحدا" . ويعبر عن (λ) بما يعادلها وهو (μ) إذ تشير لإنماذج بواسون وثنائي الحدين السالب , فضلا عن ان (μ) هي الاداة القياسية للتعبير عن المتوسط (mean parameter) في النماذج الخطية

العامة . كذلك في نماذج بواسون وثنائي الحدين السالب تتضمن متغير التعرض (exposure variable) مرتبط بال (μ) , المتغير (t) يعد هو المسافة او طول الوقت المستغرق لحصول الحدث المعني او حدوث الاعداد المحددة , هذا مايعرف بالتعرض او المواجهه (exposure) , اذا $t=1$ إذ الدالة الاحتمالية لبواسون تختزل الى الشكل القياسي وهو :

$$f(y;\mu)=\frac{e^{-t_i\mu_i}(t_i\mu_i)^{y_i}}{y_i!} \qquad \ldots\ldots(11-2)$$

الصورة الوحيدة في توزيع بواسون هي العلاقة بين المتوسط والتباين إذ يتساوى المتوسط والتباين وتسمى هذه العلاقة متعادلة التشتت (equidispersion) , الحقيقة المتغير المعتمد (y) في بواسون او يسمى باعداد بواسون هو مثال لعملية بواسون كل حدث ياخذ وقتا" مطلوبا" او مساحة لعملية العد ويعد مستقلا" كل واحد عن الاخر , اي عملية استغراق الوقت المطلوب لكل حدث مستقلة واحدة عن الاخرى مثل التوزيع المنتظم اي ان الحدث (y) في المدة (A) ليس له اي علاقة كم عدد الاحداث في المدة (B) . معدل الاحداث الذي يحصل في المدة او المساحة يسمى (μ) باحتمالية حدوثه احيانا طول المدة او حجم المساحة .

إنموذج انحدار بواسون يشتق من توزيع بواسون , العلاقة بين x, β, μ سوف تنتج متوسط الإنموذج , معالم الإنموذج , والمتغيرات المستقلة على الترتيب .

المعالم مثل $\mu=\exp(x\beta)$ هنا $x\beta$هي التركيبة الخطية والتي نرمز لها بالرمز η ضمن النماذج الخطية العامة $x\beta$ تضمن ان μ هو موجب لجميع قيم الη ولكل مقدرات المعالم . المتغير المعتمد (y) هو متغير الاستجابة لإنموذج بواسون وإنموذج ثنائي الحدين السالب ويعرف بارقام لاحداث محددة , المتغيرات المستقلة x او المتغيرات التوضيحية تعطى بمجاميع غير عشوائية للمشاهدات , اما المشاهدات ضمن المتغيرات المستقلة يفترض ان تكون مستقلة واحدة عن الاخرى .

يتم تقدير معلمات النماذج العددية بطريقتين رئيستين هما :

اولا: طريقة الامكان الاعظم maximum likelihood estimation

ثانيا: طريقة المربعات الصغرى المعادة الوزن التكرارية (الحصينة) iteratively re-weighted least squares (IRLS) والتي تعمل على تبسيط طريقة الامكان الاعظم (MLE) وتتميز بتقديرها للنماذج الخطية العامة (generalized linear models), وبالتقدير بتلك الطريقتين يصبح لدينا سهولة حساب التركيبة الخطية :

$$x_i'\beta = \eta_i = \alpha_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_n x_{in} \qquad \ldots\ldots(12-2)$$

كل مشاهدة ضمن الإنموذج لها قيمة بالتركيبة الخطية , بواسون القانوني يمتلك ربط لوغارتمي طبيعي إذ ان:

$$\eta = \ln(\mu) = \ln\big(\exp(x'\beta)\big) = x'\beta$$

إن الشكل التقليدي لثنائي الحدين السالب (NB2) (العلاقة التربيعية بين المتوسط والتباين) يحتوي ايضا على الربط اللوغارتمي , أما البواقي الاساسية تعرف على انها الاختلاف بين استجابة المشاهدة واستجابة التنبؤ , إذ (y) الاستجابة $\hat{y}$ او μ تستعمل للتنبؤ predicted .

$$\text{residual} = (y_i - \hat{y}_i) \quad \text{or} \quad (y_i - \mu_i) \quad \text{or} \quad \big(y_i - E(y_i)\big)$$

(5-2-2) فرط التشتت (overdispersion)[29][31]

فرط التشتت في نماذج بواسون يحدث عندما يكون تباين الاستجابة أكبر من المتوسط ويعود السبب للارتباط الموجب بين الاستجابات او نتيجة التغاير الكبير بين اعداد الاستجابات او احتمالات الاستجابات لذلك يحدث فرط التشتت ويزداد عندما لاتتحقق فرضيات ترتيب البيانات كأن تكون البيانات عنقودية فيعد فرط التشتت هو مشكلة يسبب اخطاء معيارية (standard errors) للمقدرات . اي ان المتغير ربما يظهر ليكون متغيرا" مستقلا" معنويا" وهو بالحقيقة غير معنوي وكيف نميز فرط التشتت اي عندما يحوي الإنموذج قيمة احصاءة مربع كاي بيرسون X^2 مقسومة على درجات الحرية اكبر من1.0 فهذا يعني وجود حالة فرط التشتت (overdispertion)وبالعكس اذا كانت قيمة احصاءة التشتت أقل من 1.0 كان الإنموذج معتدل والحالة تصبح (equidispertion) , نلحظ ان فرط التشتت يكون واضحا عن طريق

قيمة احصاءة مربع كاي في نماذج انحدار بواسون وثنائي الحدين السالب والتي تعكس لنا اختلافات البيانات في الإنموذج وهي كما في الصيغة التالية :

$$\chi^2 = \sum_i (y_i - \mu_i)^2 / V(\mu_i) \qquad \ldots\ldots(13-2)$$

(6-2-2) تعاريف إنموذج ثنائي الحدين السالب

تعريف 1 [28]:

نفرض ان X متغير عشوائي تابع لثنائي الحدين السالب بمعلمات (p,α) له الدالة الاحتمالية :

$$p(x=y) = \binom{x+\alpha-1}{y} p^y (1-p)^{\alpha} \qquad \ldots\ldots.(14-2)$$

يمكن الحصول على ثنائي الحدين السالب من توزيع بواسون المركب وذلك بمزجه بتوزيع كاما بعملية خلط التوزيعات . المتغير العشوائي y يتبع لتوزيع كاما بمعلمات (α,θ) . $y \sim Gam(\alpha,\theta)$ له الدالة احتمالية التالية :

$$f(y) = \frac{y^{\alpha-1}\theta^{\alpha}e^{-\theta y}}{\Gamma(\alpha)} \qquad \ldots\ldots.(15-2)$$

اذ ان $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1}e^{-y}\,dy$ وكذلك اذا $\alpha \in N$ فأن $\Gamma(\alpha) = (\alpha-1)!$

عند استعمال توزيع كاما مع بواسون نحصل على :

$$p(x=y) = \int_0^\infty \frac{e^{-\lambda}\lambda^y}{y!}\frac{\lambda^{\alpha-1}\theta^{\alpha}e^{-\theta\lambda}}{\Gamma(\alpha)}d\lambda$$

$$= \frac{\theta^{\alpha}}{y!\,\Gamma(\alpha)}\int_0^\infty e^{-\lambda(1+\theta)}\lambda^{y+\alpha-1}d\lambda$$

$$= \frac{\theta^{\alpha}}{y!\,(\alpha-1)!}(y+\alpha-1)!\frac{1}{(1+\theta)^{y+\alpha}}$$

$$= \binom{y+\alpha-1}{y}\left(\frac{1}{1+\theta}\right)^{y}\left(\frac{\theta}{1+\theta}\right)^{\alpha}$$

$$= \binom{y+\alpha-1}{y} p^{y}(1-p)^{\alpha} \qquad \ldots\ldots(16-2)$$

اذ ان α هي عدد صحيح . نحصل على توزيع ثنائي الحدين السالب بمعلمات $\left(\alpha, \frac{1}{1+\theta}\right)$. طالما توزيع ثنائي الحدين السالب له معلمتان , فهذا يعطينا سيطرة اكثر مقارنة بتوزيع بواسون .

$$E(y) = \frac{\alpha}{\theta} \qquad\qquad Var(y) = \frac{(1+\theta)\alpha}{\theta^2}$$

تعريف 2 [18]:

انحدار ثنائي الحدين السالب هو نوع من انحدار بواسون وفيه المتغير المعتمد y_{ij} في انحدار ثنائي الحدين السالب توجد معالم ملائمه لثنائي الحدين السالب هي :

$$f(y_{ij}) = \frac{\Gamma\left(y_{ij}+\frac{1}{\alpha}\right)}{y_{ij}!\,\Gamma\left(\frac{1}{\alpha}\right)}\left(\frac{1}{1+\alpha\mu_{ij}}\right)^{\frac{1}{\alpha}}\left(\frac{\alpha\mu_{ij}}{1+\alpha\mu_{ij}}\right)^{y_{ij}} \qquad \ldots\ldots(17-2)$$

اذ ان : $\mu_{ij} > 0$ متو سط ال y_{ij} وكذلك $\alpha = \frac{1}{\theta} > 0$ هي معلمة التشتت .

المتوسط والتباين لانحدار ثنائي الحدين السالب هو :

$$E(y_{ij}) = \mu_{ij} \qquad\qquad Var(y_{ij}) = u_{ij}(1+\alpha u_{ij})$$

$$y_{ij} \sim NegBin(u_{ij})$$

$$\log(u_{ij}) = \log(n_{ij}) + (x_{ij})^{T}\beta$$

$$u_{ij} = \exp\left\{\log(n_{ij}) + (x_{ij})^{T}\beta\right\}$$

دالة الكتلة الاحتمالية لانحدار ثنائي الحدين السالب تصبح :

$$f(y_{ij}) = \frac{\Gamma\left(y_{ij}+\frac{1}{\alpha}\right)}{y_{ij}!\,\Gamma\left(\frac{1}{\alpha}\right)}\left(\frac{1}{1+\alpha\left(\exp\left\{\log(n_{ij})+(x_{ij})^{T}\beta\right\}\right)}\right)^{\frac{1}{\alpha}}\left(\frac{\alpha\left(\exp\left\{\log(n_{ij})+(x_{ij})^{T}\beta\right\}\right)}{1+\alpha\left(\exp\left\{\log(n_{ij})+(x_{ij})^{T}\beta\right\}\right)}\right)^{y_{ij}} \quad \ldots(18-2)$$

اذ ان : $E(y_{ij}) = \mu_{ij}$ $\quad$ $Var(y_{ij}) = \mu_{ij}(1+\alpha\mu_{ij})$

α : مقياس للتشتت

تعريف 3 [4]:

توزيع ثنائي الحدين السالب يمتلك معلمتين والذي يسمح للمتوسط والتباين للتطابق بصوره منفصله , بعكس المعلمه المفرده لبواسون , وهذه تتم بطريقتين : كتوزيع حدي لمتغير عشوائي بواسون عندما معلمة المتوسط لها توزيع كاما او كدالة احتمالية للمشاهده y للفشل او لحالات الفشل قبل r−th من حالات النجاح في سلسلة محاولات برنولي المستقلة , دالة الكثافة الاحتمالة هي :

$$f(y_i/\mu_i\,,\alpha) = \frac{\Gamma(y_i+\alpha^{-1})}{y_i!\,\Gamma(\alpha^{-1})}\left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right)^{y_i}\left(\frac{1}{1+\alpha\mu_i}\right)^{\alpha^{-1}} \quad \ldots\ldots(19-2)$$

$y_i = 0,1,2,\ldots\ldots\ldots,\alpha \geq 0$

$\mu_i = \exp(x_i^T\beta)$

اذ ان : μ_i هو المتوسط لتوزيع بواسون و α معلمة التشتت , إنموذج انحدار ثنائي الحدين السالب ياخذ تنسيق إنموذج انحدار بواسون نفسه , الإنموذج يمتلك متوسط μ_i ودالة التباين $\mu_i + \alpha\mu_i^2$, عندما التشتت يساوي صفر نحصل على إنموذج بواسون .

تعريف 4 [50]:

إنموذج ثنائي الحدين السالب :

$$f(y|\alpha,\mu) = \frac{\Gamma(\alpha+y)}{\Gamma(\alpha)\Gamma(y+1)}\left(\frac{\alpha}{\mu+\alpha}\right)^{\alpha}\left(\frac{\mu}{\mu+1}\right)^{y} \quad \ldots\ldots(20-2)$$

يمتاز الإنموذج : $\alpha = {}^{1}/_{\sigma^2}$ كذلك $\mu_i = \exp(x_i\beta)$ ، التباين الشرطي قيمتة اكبر من الوسط الحسابي الشرطي في هذا الإنموذج اي يعد إنموذجا" لفرط التشتت (over dispersion) . فان وجدت علاقة تربيعية بين المتوسط – التباين يعرف بإنموذج من النوع (Neg–Bin II)(NB2) , اما اذا كانت العلاقة خطية للمتوسط – التباين يعرف بإنموذج (Neg–Bin I)(NB1) نوظف هذة العلاقة الخطية لنحصل على $\alpha_i = \sigma^{-2}\exp(x_i\beta)$ وكذلك $\mu_i = \exp(x_i\beta)$.

دالة الكثافة الاحتمالية لل NB2 يمكن ان تكتب :

$$f(y_i|.) = \frac{\Gamma(\sigma^{-2} + y_i)}{\Gamma(\sigma^{-2})\Gamma(y_i + 1)}[\sigma^2\exp(x_i\beta) + 1]^{\sigma^2}\left(\frac{\sigma^2}{\sigma^2\exp(x_i\beta) + 1}\right)^{y_i} \qquad \dots(21-2)$$

اذ ان $\sigma \to 0$ سوف يتحول هذا الإنموذج الى إنموذج بواسون , واذا كان $\sigma \geq 0$ إنموذج بواسون يمتلك حدود فضاء المعلمة . بافتراض ان العينة مستقلة :

$$L(\sigma^2, \beta, y, x) = \prod_{i=1}^{n} \frac{\Gamma(\sigma^{-2} + y_i)}{\Gamma(\sigma^{-2})}(\sigma^2\exp(x_i\beta) + 1)^{\sigma^2}\left(\frac{\sigma^2}{\sigma^2\exp(x_i\beta) + 1}\right)^{y_i} \qquad \dots(22-2)$$

اذ ان : $\alpha = \sigma^{-2}\exp[(1-k)x_i\beta]$ وكذلك $\mu_i = \exp(x_i\beta)$.

K هي المعلمة اللاخطيه تسمح بأخذ اي قيمة مستمرة . عند مقارنة هذا الإنموذج بإنموذج بواسون , اثنين من المعلمات الاضافية يمكن تقديرها في هذا الإنموذج ويسمى NEGBIN_K يمكن ان يفسر على انه hyper – model لثنائي الحدين السالب , كذلك تتداخل معه نماذج NB2, NB1 عن طريق محددات المعالم عندما k=1 , k=0 على الترتيب . دالة الكثافة الاحتمالية للNEGBIN_K يمكن ان تكتب :

$$f(y_i|\mu_i, \sigma^2, k) = c_i \times \begin{cases} \prod_{j=1}^{y_i}\left[\frac{\mu_i + \sigma^2(j-1)\mu_i^k}{[1 + \sigma^2\mu_i^k]j}\right] & \text{for } y_i = 1,2,\dots \\ 1 & \text{for } \quad y_i = 0 \end{cases} \qquad \dots(23-2)$$

اذ ان : $c_i = [1 + \sigma^2\mu_i^k]^{-\mu_i^{1-k}/\sigma^2}$

$$\mu_i = \exp(x_i\beta) \quad , \quad \sigma^2 \geq 0$$

(7-2-2) انحدار ثنائي الحدين السالب [31][32](negative binomial regression)

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 صحيحة غير سالبة .... , 3, 2, 1 , 0 انحدار ثنائي الحدين السالب هو إعمام لانحدار بواسون , لكن في إنموذج 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 الصيغة شائعة لانها تسمح لاستعمال توزيع كاما لعدم تجانس بواسون poisson heterogeneity .

إنموذج ثنائي الحدين السالب (negative binomial model) هو نوع من النماذج الخطية العامة (generalized linear models) يحوي المتغير المعتمد (y) ياخذ ارقاما" قابلة للعد لاي ظاهرة او حدث ما , المعلمات الملائمة لتوزيع ثنائي الحدين السالب هي:

$$\theta = \alpha\mu$$

$$p(y) = p(Y = y) = \frac{\Gamma(y + 1/\alpha)}{\Gamma(y + 1)\Gamma(1/\alpha)}\left(\frac{1}{1 + \alpha\mu}\right)^{1/\alpha}\left(\frac{\alpha\mu}{1 + \alpha\mu}\right)^{y} \qquad \ldots(24-2)$$

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 عن طريق عدد حالات الفشل قبل $(1/\alpha)^{th}$ من حالات النجاح , علما ان $1/\alpha$ ليس بالضروري ان تكون عددا صحيحا" .

إنموذج ثنائي الحدين السالب التقليدي (NB2) هو :

$$\ln\mu = \beta_0 + \beta_1x_1 + \beta_2x_2 + \cdots + \beta_px_p \qquad \ldots\ldots(25-2)$$

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 يأخذ متغيرا "معتمدا" (y) 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} لل x ليكون x_i :

$$p(y_i) = \frac{\Gamma(y_i + 1/\alpha)}{\Gamma(y_i + 1)\Gamma(1/\alpha)} \left(\frac{1}{1 + \alpha e^{x_i \beta}}\right)^{\frac{1}{\alpha}} \left(\frac{\alpha e^{x_i \beta}}{1 + \alpha e^{x_i \beta}}\right)^{y_i}, i = 1,2, \dots, n \qquad \dots (26-2)$$

ويمكن تقدير المعلمات α و β باستعمال تقدير الامكان الاعظم , فتصبح دالة الامكان الاعظم :

$$L(\alpha, \beta) = \prod_{i=1}^{n} P(y_i) = \prod_{i=1}^{n} \frac{\Gamma(y_i + 1/\alpha)}{\Gamma(y_i + 1)\Gamma(1/\alpha)} \left(\frac{1}{1 + \alpha e^{x_i \beta}}\right)^{\frac{1}{\alpha}} \left(\frac{\alpha e^{x_i \beta}}{1 + \alpha e^{x_i \beta}}\right)^{y_i} \qquad \dots (27-2)$$

ودالة الامكان اللوغارتمية تصبح :

$$\ln L(\alpha, B) = \sum_{i=1}^{n} \left(y_i \ln\alpha + y_i (x_i . \beta) - \left(y_i + \frac{1}{\alpha}\right) \ln(1 + \alpha e^{x_i \beta}) \right.$$
$$\left. + \ln\Gamma\left(y_i + \frac{1}{\alpha}\right) - \ln\Gamma(y_i + 1) - \ln\Gamma\left(\frac{1}{\alpha}\right) \right) \qquad \dots (28-2)$$

قيم α و β التي تعظم $\ln L(\alpha, B)$ سوف تكون مقدرات الامكان الاعظم . ومصفوفة التباين والتباين المشترك للمقدرات هي $\Sigma = -H^{-1}$ اذ ان : H هي مصفوفة حسن (Hessian matrix) للمشتقة الثانية لدالة الأمكان اللوغارتمية .

اذ ان مصفوفة التباين والتباين المشترك يمكن استعمالها في إيجاد حدود الثقة (Wald confidence intervals) وقيم (p-values) لتقديرات المعاملات (coefficient estimates)

وتم ذكرها تفصيلا" في طرائق التقدير .

ويمكن استعمال الإنموذج التقليدي NB2 للمعدلات (rates) وذلك باستعمال الأزاحه (offset) , اذ ان المتغير المعتمد (y) هو رقم عدد الحوادث خلال مدة زمنيه معينة (t) اذ ان معدل المشاهدة Y|t يمكن نمذجته باستعمال إنموذج ثنائي الحدين السالب المذكور آنفا" بتعديل بسيط (t), هي المساحة او جزء من المجتمع . $E(Y/t) = \mu / t$

$$\ln(\mu/t) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p \qquad \ldots\ldots(29-2)$$

ويمكن كتابته بالصيغة :

$$\ln\mu = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \ln t \qquad \ldots\ldots(30-2)$$

اذ ان : $\ln t$ يسمى الأزاحة (offset) .

في دالة الامكان اللوغارتمية استبدلنا μ بال $e^{x_i\beta}$, وهنا نستبدل μ بال $e^{x_i\beta+\ln t}$ وينتج عنه :

$$\ln L(\alpha, B) = \sum_{i=1}^{n}\left(y_i\ln\alpha + y_i(x_i.\beta + \ln t_i) - \left(y_i + \frac{1}{\alpha}\right)\ln(1 + \alpha e^{x_i\beta+\ln t_i}) + \ln\Gamma\left(y_i + \frac{1}{\alpha}\right) - \ln\Gamma(y_i + 1) - \ln\Gamma\left(\frac{1}{\alpha}\right)\right) \qquad \ldots(31-2)$$

بتعظيم دالة الأمكان اللوغارتمية الجديدة لتقدير المعلمات .

(3-2) اشتقاق توزيع ثنائي الحدين السالب [31]:

حسب تعريف (2) ان الدالة الاحتمالية لتوزيع ثنائي الحدين السالب يمكن ايجادها من :

$$f(y; \lambda, u) = \frac{e^{-\lambda_i u_i}(\lambda_i u_i)^{y_i}}{y_i!} \qquad \ldots \qquad (32-2)$$

هنا نموذج بواسون مع عدم تجانس كاما (gamma heterogeneity) بمتوسط كاما =1 , توزيع الy مشروط على ال(x,u) هو المتوسط والتباين الشرطيين لبواسون معطاة بالμ التوقع الشرطي للy مع عدم تجانس كاما هو للمعالجه بتعبير λu مقارنة بـλ فقط , نتيجة لذلك التوزيع اللاشرطي للy مشتق من الصيغ الاتية:

$$f(y;x,u)=\int_0^{\infty}\frac{e^{-(\lambda_i u_i)}(\lambda_i u_i)^{y_i}}{y_i!}g(u_i)\,\partial u_i \qquad \ldots \quad (33-2)$$

التوزيع اللاشرطي للy يخصص كيف نعرف g(u) لأنموذج توزيع كاما هذا المعطى $u=\exp(\varepsilon)$ اذ ان $\ln(\mu)=xb+\varepsilon$ علما ان متوسط كاما = 1.

$$f(y;x,u)=\int_0^{\infty}\frac{e^{-(\lambda_i u_i)}(\lambda_i u_i)^{y_i}}{y_i!}\frac{v^v}{\Gamma(v)}u_i^{v-1}e^{-vu_i}du_i \qquad \ldots\ (34-2)$$

مقلوب الv معلمة القياس لكاما وهي α وتعتبر معلمة عدم التجانس او التشتت لثنائي الحدين السالب , كذلك لدينا $v=\frac{1}{\alpha}$ و $\mu=\lambda$, طبيعة كاما للu هي دليل للاشتقاق من (34-2)

$$=\frac{\lambda_i^{y_i}}{\Gamma(y_i+1)}\frac{v^v}{\Gamma(v)}\int_0^{\infty}e^{-(\lambda_i+v)u_i}u_i^{(y_i+v)-1}du_i \qquad \ldots\ (35-2)$$

سحب المقدار $\frac{\lambda_i^{y_i}}{\Gamma(y_i+1)}\frac{v^v}{\Gamma(v)}\frac{\Gamma(y_i+v)}{(\lambda_i+v)^{y_i+v}}$ الى يسار التكامل مع بقاء المقدار تحت التكامل =1

$$=\frac{\lambda_i^{y_i}}{\Gamma(y_i+1)}\frac{v^v}{\Gamma(v)}\Gamma(y_i+v)\left(\frac{v}{\lambda_i+v}\right)^v\frac{1}{v^v}\left(\frac{\lambda_i}{\lambda_i+v}\right)^{y_i}\frac{1}{\lambda_i^{y_i}} \qquad \ldots (36-2)$$

$$=\frac{\Gamma(y_i+v)}{\Gamma(y_i+1)\Gamma(v)}\left(\frac{v}{\lambda_i+v}\right)^v\left(\frac{\lambda_i}{\lambda_i+v}\right)^{y_i}$$

$$=\frac{\Gamma(y_i+v)}{\Gamma(y_i+1)\Gamma(v)}\left(\frac{1}{1+\frac{\lambda_i}{v}}\right)^v\left(1-\frac{1}{1+\frac{\lambda_i}{v}}\right)^{y_i} \qquad \ldots\ldots\ldots \ (37-2)$$

تصبح لدينا دالة الكتله الاحتماليه PMF لثنائي الحدين السالب هي :

$$f(y;\mu,\alpha)=\frac{\Gamma\left(y_i+\frac{1}{\alpha}\right)}{\Gamma(y_i+1)\Gamma\left(\frac{1}{\alpha}\right)}\left(\frac{1}{1+\alpha\mu_i}\right)^{\frac{1}{\alpha}}\left(1-\frac{1}{1+\alpha\mu_i}\right)^{y_i} \qquad \ldots .(38-2)$$

وبالتعبير عن :

$$\Gamma(y+1)=y!\,,\Gamma(y+1/\alpha)=(y+1/\alpha-1)!\ ,(1/\alpha)=(1/\alpha-1)!$$

وتكون :

$$\frac{\Gamma(y_i+1/\alpha)}{\Gamma(y_i+1)\Gamma(1/\alpha)}=\frac{(y_i+1/\alpha)!}{y_i!\,(1/\alpha-1)!}=\binom{y_i+1/\alpha-1}{1/\alpha-1} \qquad \dots(39-2)$$

دوال كاما التي من اليسار يمكن ان تحول الى شكل توافيق combination وكذلك الجانب الايمن يمكن ان يحول الى كسر مفرد وبالنتيجه نحصل على تعبير شائع للتوزيع الاحتمالي لثنائي الحدين السالب .

$$f(y;\mu,\alpha)=\binom{y_i+\frac{1}{\alpha}-1}{\frac{1}{\alpha}-1}\left(\frac{1}{1+\alpha\mu_i}\right)^{\frac{1}{\alpha}}\left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right)^{y_i} \qquad \dots(40-2)$$

هنا صيغة لدالة الكتله الاحتماليه لثنائي الحدين السالب توجد باشكال مختلفه بعض المؤلفين يستخدم v لمعلمة كاما , او يستعمل رمز r, k لتمثيل معلمة ثنائي الحدين السالب , والتي تعد معلمة عدم التجانس ذي الحدين السالب , في هذا الشكل معلمة عدم التجانس ترتبط عكسيا بكمية فرط تشتت بواسون (poison overdispersion) في البيانات . α ,y يفترض ان تكون اعداد صحيحه ولكن هذا الافتراض لايمكن الحصول عليه الا عندما تكون قواعد او اسس التوزيع لإنموذج انحدار , كإنموذج قابل للعد الـy يجب ان تتكون من قيما" صحيحة غير سالبة , اما α ليس بالضروره كذلك ولكن القيد الوحيد على α هي تاخذ قيما" نسبية نادرا ماتكون اكبر من (4) فضلا" عن ذلك لاتدمج معلمة عدم تجانس ثنائي الحدين السالب مع احصاءة التشتت للنموذج العديد من الباحثين يشيرون الى ان معلمة عدم التجانس هي معلمة التشتت , اذ ترتبط بكمية فرط التشتت في البيانات ولكن احصاءة التشتت هي ليست معلمه . بالحقيقه هي احصاءة تولدت بواسطة انقسام احصاءه بيرسون مربع كاي عن طريق درجات الحريه لإنموذج البواقي , وهذا مايميزها عن الاحصاءه التي هي مقدار التكامل للتوزيع الاحتمالي لثنائي الحدين السالب . دالة الامكان هي اعادة تعبير للداله الاحتماليه لثنائي الحدين السالب .

(4-2) طرائق التقدير (Methods of estimation)

(1-4-2) طريقة الامكان الاعظم (maximum likelihood)[31][3]

هي الطريقه الاكثر شيوعا والمألوفة في التقدير , تعد مقدرات الامكان الاعظم ثابتة وتعطي مقدرات الامكان الاعظم معيارا عاليا" للكفاءه وكذلك بخاصية الاتساق اي ان عملية التقدير عن طريق جعل دالة الامكان للمتغيرات العشوائية اعظم مايمكن .

$$L(\mu;y,\alpha)=\prod_{i=1}^{n}\exp\left\{y_i\ln\left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right)-\frac{1}{\alpha}\ln(1+\alpha\mu_i)+\ln\Gamma\left(y_i+\frac{1}{\alpha}\right)-\ln\Gamma(y_i+1)-\ln\Gamma\left(\frac{1}{\alpha}\right)\right\} \qquad \ldots\ldots(41-2)$$

دالة الامكان اللوغارتمية يمكن ان نحصل عليها وذلك باخذ اللوغارتم الطبيعي لكلا الجانبين للمعادله , فتصبح الداله جمعية اكثر ما هي ضربية :

$$L(\mu;y,\alpha)=\sum_{i=1}^{n}y_i\ln\left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right)-\frac{1}{\alpha}\ln(1+\alpha\mu_i)+\ln\Gamma\left(y_i+\frac{1}{\alpha}\right)-\ln\Gamma(y_i+1)-\ln\Gamma\left(\frac{1}{\alpha}\right) \qquad \ldots\ldots(42-2)$$

ان الامكان اللوغارتمي لثنائي الحدين السالب له معلمة β ومعاملات الإنموذج كما يعبر عنها بالمعادله :

$$L(\beta_j;y,\alpha)=\sum_{i=1}^{n}y_i\ln\left(\frac{\alpha\exp(x_i'\beta)}{1+\alpha\exp(x_i'\beta)}\right)-\frac{1}{\alpha}\ln[1+\alpha\exp(x_i'\beta)]+\ln\Gamma\left(y_i+\frac{1}{\alpha}\right)-\ln\Gamma(y_i+1)-\ln\boldsymbol{\Gamma}\left(\frac{1}{\alpha}\right) \qquad \ldots(43-2)$$

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

إن مصفوفة المشتقات الجزئيه الثانيه للامكان اللوغارتمي غالبا ماتعرف باسم حسن ومصفوفة المعلومات المشاهدة (observed information matrix) تعرف بسالب حسن (Hessian negative matrix) تقيم عن طريق مقدر الامكان الاعظم .الاخطاء القياسيه للإنموذج يمكن الحصول عليها من الجذر التربيعي للعناصر القطرية لمصفوفة التباين والتباين المشترك , والتي تعد معكوس المعلومات لذلك الاخطاء القياسيه للإنموذج هي الجذر التربيعي للحدود القطرية لمصفوفه معكوس سالب حسن (negative inverse Hessian matrix) .

ونيوتن- رافسون تعتمد على تقدير الامكان الاعظم ويستلزم مصفوفة معلومات المشاهده , المشتقتان الاولى والثانيه لدالة الامكان الاعظم اللوغارتميه بالنسبه الى α , β كما يأتي :

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

NB GRADIENT – β

$$\frac{\partial L}{\partial \beta} = \sum_{i=1}^{n} \frac{x_i(y_i - \mu_i)}{1 + \alpha\mu_i} \quad \dots\dots\dots (44-2)$$

المشتقه الاولى بالنسبه الى μ (فقط في حدود الμ)

$$\frac{\partial L}{\partial \mu} = \sum_{i=1}^{n} \frac{y_i - \mu_i}{\mu_i(1 + \alpha\mu_i)} \quad \dots\dots.(45-2)$$

NB GRADIENT – α

$$\frac{\partial L}{\partial \alpha} = \sum_{i=1}^{n} \left[\frac{1}{\alpha^2}\left(\ln(1 + \alpha\mu_i) + \frac{\alpha(y_i - \mu_i)}{1 + \alpha\mu_i}\right) + \Psi\left(y_i + \frac{1}{\alpha}\right) - \Psi\left(\frac{1}{\alpha}\right)\right] \quad \dots.(46-2)$$

NB –HESSIAN –β

$$\frac{-\partial^2 L}{\partial \beta\, \partial \beta'} = \sum_{i=1}^{n} \frac{\mu_i(1 + \alpha y_i)}{(1 + \alpha\mu_i)^2} x_i x_i' \quad \dots\dots..(47-2)$$

او

$$\frac{\partial^2 L}{\partial \beta_j'\, \partial \beta_j} = \sum_{i=1}^{n} \left[-x_{ij}x_{ij}' \frac{\mu_i(1 + \alpha\mu_i)}{(1 + \alpha\mu_i)^2}\right] \quad \dots\dots.(48-2)$$

NB –HESSIAN –β ; α

$$\frac{\partial^2 L}{\partial \beta\, \partial \alpha} = E\left[-\sum_{i=1}^{n} \frac{\mu_i(y_i - \mu_i)x_{ij}}{(1 + \alpha\mu_i)^2}\right] \quad .\dots\dots.(49-2)$$

NB –HESSIAN –α

$$\frac{\partial^2 L}{\partial \alpha^2} = \sum_{i=1}^{n} \left[-\frac{1}{\alpha^3}\left(\frac{\alpha(1 + 2\alpha\mu_i)(y_i - \mu_i) - \alpha\mu_i(1 + \alpha\mu_i)}{(1 + \alpha\mu_i)^2}\right.\right.$$
$$\left.\left. + 2\ln(1 + \alpha\mu_i) + \Psi'\left(y_i + \frac{1}{\alpha}\right) - \Psi'\left(\frac{1}{\alpha}\right)\right] \quad \dots\dots\ (50-2)$$

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

$$\mathrm{DIGAMMA} = \Psi_1 = \Psi(x) = \frac{\ln\Gamma(x+0.0001) - \ln\Gamma(x-0.0001)}{0.0002} \qquad \dots\dots (51-2)$$

وكذلك

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

او

$$= \frac{\left\{\ln\Gamma\left(\left({}^{1}/_{\alpha}\right) + 0.0001\right) - \ln\Gamma\left(\left({}^{1}/_{\alpha}\right) - 0.0001\right)\right\}}{0.0002} \qquad \dots\dots. \quad (52-2)$$

(2-4-2) طريقة المربعات الصغرى معادة الوزن التكرارية (الحصينة)[31]:-

Iteratively re – weighted least squares (IRLS)

تعتمد هذه الطريقه على ترميز فشر (fisher scoring) وتعد هذه الطريقة كجزء من طريقة الأمكان الاعظم والتي يمكن استعمالها لتقدير الإنموذج الخطي بالنسبة لمعلمات الإنموذج تسمى مصفوفة حسن Hessian matrix .

ومن المعلوم ان دالة الكثافة الاحتمالية للعائلة الاسية (exponential family) :-

$$f(y;\theta,\Phi) = \exp\left[\frac{y_i\theta_i - b(\theta_i)}{\alpha_i(\Phi)} + c(y_i;\Phi)\right] \qquad \dots\dots (53-2)$$

اذ ان :

θ_i هي المعلمه القانونية او دالة الربط (link function)

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

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

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

شكل العائلة الاسية يعد وحيدا" (unique) في المشتقتين الاولى والثانيه لمعلمة الموقع التراكميه (cumulant) بالنسبه الـθ فينتج دوال المتوسط والتباين على الترتيب ,

$$b'(\theta_i) = mean$$

$$b''(\theta_i) = variance$$

ان الداله الاحتماليه للإنموذج الخطي العام GLM :-

$$f(y;\theta,\Phi) \qquad \ldots\ldots\ldots.(54-2)$$

إذ ان :

Y متغير الاستجابه (response variable)

θ معلمة الموقع (location parameter)

Φ معلمة القياس (scale parameter)

في النماذج العددية count models لدينا قيم متغير الاستجابه y لها خواص توزيعية تتلائم مع التوزيع المستعمل في التقدير , الدوال الاحتمالية تحدد خواص الاستجابة , طريقة التقدير بالامكان الاعظم تعتمد على تعظيم الاحتمال , دالة الامكان هي متممة للدالة الاحتمالية . وبالرغم من ان البيانات هي التي تحدد قيم المتوسط وقيم القياس لكن يبقى الهدف المهم للامكان هو تقدير قيم المعالم المجهولة.

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 مفكوك تايلر لدالة الامكان اللوغارتميه :

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

أول حدين تختزل الى :

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

ويمكن كتابتها

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

المشتقه الاولى لدالة الامكان اللوغارتمية تدعى ايضا بدالة فشر fisher score , اذا كانت دالة الامكان اللوغارتمية مقعرة , سوف نجعل المشتقة مساوية للصفر ونحل بالنسبة الى β لتقدير معلمات الامكان الاعظم .

المشتقه الثانيه لدالة الامكان اللوغارتمية التي تسمى مصفوفة حسن (Hessian matrix), عندما يكون شكل دالة الامكان اللوغارتمية محدبا" اكثر ما هو مفلطح , معكوس حسن السالب(negative inverse Hessian matrix) يعطي مصفوفة التباين والتباين المشترك .

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

$$U = \partial L \qquad \text{and} \qquad H = \partial^2 L \qquad \ldots\ldots(55-2)$$

اذ ان:

U : هي المشتقة الاولى لداله الامكان , H : المشتقة الثانية لدالة الامكان .

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

$$\beta_r = \beta_{r-1} - H^{-1}U \qquad \ldots\ldots\ldots\ldots(56-2)$$

عندما

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

r : رقم التكرار وهو عدد المعالم المطلوب تقديرها .

عن طريق خوارزمية نيوتن – رافسون نحصل على β_r وهي مقدرات معلمة الإنموذج , و بتكرار ايجاد الحلول للH يصبح لدينا مصفوفة المعلومات المشاهدة

(Observed information matrix), كذلك مصفوفة حسن المستعملة في IRLS تسمى مصفوفة المعلومات المتوقعة (EIM) (expected information matrix)

ولايجاد المشتقة الاولى the gradient(U)

في شكل العائلة الاسية دالة الامكان اللوغارتمية تكون :

$$L(\theta; y, \Phi) = \sum_{i=1}^{n} \frac{y_i\theta_i - b(\theta_i)}{\alpha_i(\Phi)} + c(y_i; \Phi) \qquad \ldots\ldots(57-2)$$

نشتق الدالة بالنسبة الى β باستعمال قانون السلسلة :

$$\frac{\partial L}{\partial \beta_j}=\sum_{i=1}^{n}\left(\frac{\partial l_i}{\partial \theta_i}\right)\left(\frac{\partial \theta_i}{\partial \mu_i}\right)\left(\frac{\partial \mu_i}{\partial \eta_i}\right)\left(\frac{\partial \eta_i}{\partial \beta_i}\right) \qquad \dots\dots(58-2)$$

نحل كل حد

$$\frac{\partial L}{\partial \beta_j}=\sum_{i=1}^{n}\frac{y_i\theta_i-b'(\theta_i)}{\alpha_i(\Phi)}+\sum_{i=1}^{n}\frac{y_i-\mu_i}{\alpha_i(\Phi)} \qquad \dots\dots..(59-2)$$

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

$$\frac{\partial \mu_i}{\partial \theta_i}=\frac{\partial b'(\theta_i)}{\partial \theta_i}=b''(\theta_i)=V(\mu_i)\,,\frac{\partial \theta_i}{\partial \mu_i}=\frac{1}{V(\mu_i)} \qquad \dots\dots..(60-2)$$

كذلك

$$\frac{\partial \eta_i}{\partial \beta_j}=\frac{\partial\left(x_i\beta_j\right)}{\partial \beta_j}=x_{ij}\ ,\text{since}\quad \eta_i=x_i\beta_j \qquad \dots\dots\dots.(61-2)$$

كذلك

$$\frac{\partial \mu_i}{\partial \eta_i}=[g^{-1}(\eta_i)]'=\frac{1}{\partial\eta_i/\partial\mu_i}=\frac{1}{g'(\mu_i)} \qquad \dots\dots\dots.(62-2)$$

هذه مشتقة دالة الربط بالنسبه الى μ , η معكوس دالة الربط . استبدال التعابير لمقدر الامكان الاعظم لل β هو الحل للمعادلة التقديرية .

$$\sum_{i=1}^{n}\frac{(y_i-\mu_i)x_i}{\alpha_i(\Phi)V(\mu_i)g'(\mu_i)}=\sum_{i=1}^{n}\frac{(y_i-\mu_i)x_i}{\alpha_i(\Phi)V(\mu_i)}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)=0 \qquad \dots\dots\dots(63-2)$$

عندما y تمثل الاستجابة , μ تمثل المتغيرات الملائمة (fitted variables) , x متجه صفي , المجموع الناتج متجه عمودي .

ولايجاد المشتقة الثانية نستبدل H بال I

$$I=-E\left[\frac{\partial^2 L}{\partial \beta_j\,\partial \beta_k}\right]=E\left[\frac{\partial L}{\partial \beta_j}\frac{\partial L}{\partial \beta_k}\right] \qquad \dots\dots\dots(64-2)$$

$$I=\frac{\partial}{\partial \beta_j}\left[\frac{(y_i-\mu_i)x_j}{\alpha_i(\Phi)V(\mu_i)}\left(\frac{\partial}{\partial \eta}\right)_i\right]*\frac{\partial}{\partial \beta_k}\left[\frac{(y_i-\mu_i)x_k}{\alpha_i(\Phi)V(\mu_i)}\left(\frac{\partial}{\partial \eta}\right)_i\right] \qquad \dots(65-2)$$

$$I=\frac{(y_i-\mu_i)^2x_jx_k}{\{\alpha_i(\Phi)V(\mu_i)\}^2}\left(\frac{\partial\mu}{\partial\eta}\right)_i^2 \qquad \ldots\ldots\ldots(66-2)$$

اذ ان :

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

نفرض ان :

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

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

نضع المعادلات مع بعضها سوف نحصل :

$$\beta_r=\beta_{r-1}-\left[\frac{x_jx_k}{V(y_i)}\left(\frac{\partial\mu}{\partial\eta}\right)_i^2\right]^{-1}\left[\frac{(y_i-\mu_i)x_k}{V(y_i)}\left(\frac{\partial\mu}{\partial\eta}\right)_i\right] \qquad \ldots\ \ldots(67-2)$$

بضرب طرفي المعادله في I

$$\left[\frac{x_jx_k}{V(y_i)}\left(\frac{\partial\mu}{\partial\eta}\right)_i^2\right]\beta_r=\left[\frac{x_jx_k}{V(y_i)}\left(\frac{\partial\mu}{\partial\eta}\right)_i^2\right]\beta_{r-1}+\left[\frac{(y_i-\mu_i)x_k}{V(y_i)}\left(\frac{\partial\mu}{\partial\eta}\right)_i\right] \qquad \ldots(68-2)$$

نفرض ان w يساوي

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

ومع التركيبه الخطيه η يعطي

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

إذ ال η_i يبلغ قيمته في التكرار (n-1)

$$\left[\frac{x_jx_k}{V(y_i)}\left(\frac{\partial\mu}{\partial\eta}\right)_i^2\right]\beta_r=[X'WX]\beta_r \qquad \ldots\ldots\ldots(69-2)$$

نعرف w, V(y)

$$\frac{(y_i-\mu_i)x_k}{V(y_i)}\left(\frac{\partial\mu}{\partial\eta}\right)_i=\frac{(y_i-\mu_i)x_k}{\frac{1}{w}\left(\frac{\partial\mu}{\partial\eta}\right)_i^2}\left(\frac{\partial\mu}{\partial\eta}\right)_i \quad \dots\dots\dots(70-2)$$

وهنا

$$\left[\frac{x_jx_k}{V(y_i)}\left(\frac{\partial\mu}{\partial\eta}\right)_i^2\right]\beta_{r-1}=x'w\eta_i \quad \dots\dots\dots(71-2)$$

بدمج التركيبين

$$[x'wx]\beta_r=x'w\eta_i+\left[\frac{(y_i-\mu_i)x_k}{\frac{1}{w}\left(\frac{\partial\mu}{\partial\eta}\right)_i^2}\left(\frac{\partial\mu}{\partial\eta}\right)_i\right] \quad \dots\dots\dots(72-2)$$

$$[x'wx]\beta_r=x'w\eta_i+\left[x_kw(y_i-\mu_i)\left(\frac{\partial\eta}{\partial\mu}\right)_i\right] \quad \dots\dots\dots(73-2)$$

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

$$\beta_r=[x'wx]^{-1}x'wz \quad \dots\dots\dots(74-2)$$

(3-4-2) طريقة الامكان الموزونة [6] (weighted likelihood estimation)

طريقة الامكان الموزونة تعطينا مقدرات ذات كفاءة عالية في حالات ومواقف تكون فيها الاخطاء متغيرات (iid) اي تكون مستقلة ومتماثلة التوزيع . الاوزان تتكون من مقارنة التوزيع التجريبي للبواقي والتوزيع النظري .

هذه الطريقة يمكن تطبيقها في حالات عندما توزيع الاخطاء يكون معتمدا" على المتغيرات المستقلة (covariates) مثل انحدار بواسون او انحدار ثنائي الحدين السالب .

نفرض $NB_{\alpha,\mu}$ عائلة توزيعات ثنائي الحدين السالب كذلك

$$Y_{\alpha,\mu} \sim NB_{\alpha,\mu}$$

اذ ان : $E(Y_{\alpha,\mu}) = \mu \quad \text{and} \quad var(Y_{\alpha,\mu}) = \mu + \alpha\mu^2$

متغير الاستجابة (Y) $\quad Y_{\alpha_0,\mu_0(x)} \sim NB_{\alpha_0,\mu_0(x)}$

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

الغرض الرئيسي من هذه 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) :

$$\sum_{i=1}^{n} w(z_i, h) s(h, z_i) = 0 \qquad \dots\dots\dots (75-2)$$

اذ ان :

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

خلال عملية تكوين الاوزان : يمكن تعريف الاحتمالات :

$$P_h(z_i) = P(Y_{\alpha,\mu(x_i)} \le y_i) - u_i P(Y_{\alpha,\mu(x_i)} = y_i) \qquad \dots (76-2)$$

(2−4−4) طريقة المربعات الصغرى الموزونة (weighted least squares)[2]

تستعمل هذه الطريقة لزيادة كفاءة مقدرات المربعات الصغرى, ويمكن ان نحصل على تقدير لمعلمات الانحدار عن طريق أوزان تعطى لكل مشاهدة وذلك حسب تباين البواقي اذ ان الفرق بين قيم الحد الثابت والميل الحدي المقدره للإنموذج 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 ماذكر آنفا" بشكل مختصر وكالاتي :

$$p^{-1}Y = p^{-1}X\beta + p^{-1}U \qquad \ldots\ldots\ldots (77-2)$$

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

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

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

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

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

$$E(U.U') = \sigma_u^2 I_n I_n = \sigma_u^2 I_n \qquad \dots\dots\dots (78-2)$$

اي تم تحقق الفروض الخاصة بتجانس تباين الخطا وعدم وجود الارتباط الذاتي , ان الفرضيات الخاصة بأنموذج الانحدار قد تحققت ومن ثم يمكن تطبيق طريقة المربعات الصغرى الاعتيادية في تقدير موجه معالم الإنموذج وكالاتي :

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

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

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

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

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

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

$$= Y'W^{-1}Y - 2\beta'X'W^{-1}Y + \beta'X'W^{-1}X\beta \qquad \dots\dots\dots (79-2)$$

وباخذ المشتقة بالنسبة الى β' نحصل على :

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

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

$$b_{WLS} = (X'W^{-1}X)^{-1}X'W^{-1}Y \qquad \dots\dots\dots (80-2)$$

(5−2) معايير المقارنة (comparative criterions)[40]

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

$$MSE = \frac{SSE}{n-p} \qquad \ldots\ldots\ldots(81-2)$$

اذ ان : n حجم العينة , p عدد معلمات الأنموذج .

$$sse = \sum_{i=1}^{n} u_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \qquad \ldots\ldots\ldots(82-2)$$

❖ معامل التحديد (coefficient of determination)

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

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(y_i - \hat{\theta}_i)^2 + \sum_{i=1}^{n}(\hat{\theta}_i - \bar{y})^2 + 2\sum_{i=1}^{n}(y_i - \hat{\theta}_i)(\hat{\theta}_i - \bar{y}) \quad \ldots(83-2)$$

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 $0 \le R^2 \le 1$

الفصل الثالث

الجانب العملي

المبحث الأول: الجانب الطبي

(3-1) التشوهات الخلقية

(3-1-1) مفهوم التشوهات الخلقية (concept of congenital anomalies)[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 عن طريق وجود خلل جيني اثناء المراحل المبكرة لتطوير الخلية التكوينية بسبب تعرضها لبعض التأثيرات والعوامل الشاذة عن طبيعة تطور الجنين .وتعد التشوهات الخلقية اما ظاهرية او داخلية ويمكن التاكد منها بالفحوصات المختبرية والفحص بالأشعة والسونار .

من أهم المسببات لموت الجنين داخل الرحم أو بعد الولادة هي التشوهات الخلقية للجنين , وربما يبقى على قيد الحياة بعوق جسدي أو وظيفي فضلا" عن ماتتركة من آثار على نفسية وصحة الأم ويحدث التشوة للجنين في مدة تكوين الأعضاء والأنسجة داخل الرحم ويمكن تشخيصة اثناء مدة الحمل بالموجات فوق الصوتية أو التشخيص بعد الولادة أو بعد أيام من الولادة أو بعد أسابيع , ويمكن تصنيف التشوهات الخلقية الى ثلاثة أصناف:

(3-1-1-1) التشوهات النسيجية (الداخلية)[1] وتشمل الأنواع الاتية :

- التشوهات الخلقيه في القلب وجهاز الدوران (congenital anomalies of heart & circulatory system) هي تشوهات متعددة الأنواع والشدة , أما ان تصيب الاوردة أو الشرايين الداخلة الى القلب أو الخارجة منه , كذلك من المحتمل ان تصيب صمامات القلب أو حواجز القلب إذ يعد هذا النوع من التشوهات هي الأكثر أنتشارا" وشيوعا" وتظهر الأعراض بعد الولادة مباشرة , تشوهات القلب وجهاز الدوران تحدث لدى الجنين في الأسابيع الأولى من الحمل ومن أهم أسباب حصول هذا النوع من التشوهات هو خلل في الكروموسومات أو تعاطي الأم أدوية أو اصابة الأم بالحصبة الألمانيه أو بداء السكري ويمكن علاج هذه التشوهات بعمليات جراحية .

- التشوهات الخلقية في الجهاز الهظمي (congenital anomalies of digestive system) وجود الفتحة بين المرئ والقصبة الهوائية , خروج الامعاء وبعض اعضاء الجهاز الهضمي خارج البطن او انسداد في الأمعاء .
- عيوب خلقية كروموسومية (chromosomal anomalis) عند الشخص الطبيعي يبلغ عدد الكروموسزمات (46) وتحصل زيادة او نقصان في عدد الكروموسومات نتيجة طفرات وراثية (genetic mutation) غير معروفة السبب , ومنها :

❖ المنغولية (mongolism) او متلازمة داون ظاهرة كثيرة الانتشار في العالم بشكل واسع وتحصل نتيجة زيادة عدد الكروموسومات (47) كروموسوما" وتحدث في زوج الكروموسوم (21) وسمي بالمنغولي للشبه الكبير للمولود بأقوام المغول , وعندما يصبح عمر الأم أكثر من (35) عام فأن نسبة الأنجاب لطفل منغولي تزداد فضلا عن انها معرضة ان تنجب المنغولي بكل الأعمار .

❖ متلازمة كلاين فلتر : تحدث في زوج الكروموسومات (23) اي تصبح 47 كروموسوما" ويصاب بهذا التشوه فقط الذكور , اي ان الأعضاء التناسلية للذكر صغيرة ولاتنتج الحيوانات المنوية فضلا" عن كبر الصدر وأتساع الحوض مثل الأناث .

❖ متلازمة تيرنر : تحدث نتيجة نقص في زوج الكروموسومات الجنسية (23) ليكون عدد الكرموسومات (45) يحدث هذا النوع من التشوه في الاناث فقط إذ تبدو الانثى قريبه الى الرجل ذات حوض ضيق ومنكبين عريضين وقصيرة القامه عند مقارنتها بالنساء الطبيعيات.

(3-1-1-2) العيوب الظاهرية [1]:

- التشوهات الخلقية في الوجه ومنها :

❖ الشق الخلقي بالشفه (شق الارنب) هو تشوه خلقي ناتج عن عدم اتصال الانسجه في الشفه (منتصف الوجه) في المدة الاولى من الحمل (اول ثلاثة اشهر) اي مدة تكوين الجنين والسبب وراء هذا التشوه لحد الان غير معروف ولكن هناك عوامل مهيئة بيئية مثل تناول ادوية مثل الكورتزون او ادوية الصرع او زواج الاقارب او وجود اصابة في العائلة او اصابة الام بالحمى او نقص حامض الفولك folic acid .

- ❖ الشق الخلقي بالحنك : ينتج هذا التشوه من عدم التحام الانسجه في سقف (داخل الفم) واسباب هذا التشوه تتشابه مع شق الارنب .
- ❖ الشق الخلقي بالحنك والشفه : هو عدم اتصال انسجة الشفه وتمتد للهاة .
- ❖ العيوب الخلقيه في العين : اي عيوب في عدسة العين وعتامتها،شكل العين كذلك اختفاء او ضمور فتحة العين او دمج العينين بعين واحدة في الجبهة وتسمى مزج العينين من اسبابها وراثية او اصابة الام بجراثيم مثل التوكسوبلازما (toxoplasmosis) .
- ❖ صيوان الاذن الاضافي : وهو نتوء لحمي غضروفي ويعد قريبا من صيوان الاذن .
- التشوهات الخلقية في القوائم ومنها :
  - ❖ الصلب الاشرم (الظهر المشقوق) : وهو عبارة عن خلل في النمو للحبل الشوكي والدماغ وتشوه في فقرات الظهر فلا يوجد عظم يحمي الحبل الشوكي من التأثيرات الخارجية واسبابه هي تعرض الام خلال المدة الاولى من الحمل الى اشعاع او تناول أدوية ضد الصرع أو حمى أو نقص حامض الفوليك أو عوامل وراثية .
  - ❖ العيوب الخلقية في الطرف العلوي (congenital anomalies Of upper limb) وتشمل ضمور في الأطراف العليا أو ألتصاق في اللأصابع أو زيادة عدد الأصابع لليد .
  - ❖ العيوب الخلقية في الأطراف السفلى (congenital Anomalies Of lower limb) ويشمل تغيير في شكل القدم أو ألتصاق الأصابع أو خلع في مفصل الورك .
  - ❖ ضمور الاطراف السفلى والعليا معاً (congenital anomalies of lower & upper limbs)
  - ❖ التشوهات الخلقية في الجلد(congenital anomalies of skin) : مثل اختفاء صبغة المبلاثين او خشونة جلد المولود مثل صدف السمكه والجلد الفقاعي الولادي .
- العيوب الخلقية في الاعضاء البولية والتناسلية :
  - ❖ الخصيه غير النازل : او تسمى بعدم وجود خصية او خصية مهاجرة .

❖ الاعضاء التناسلية الخارجية المبهمة : وهنا الأعضاء التناسلية غير معروفة هل هي ذكرية أم انثوية بالنسبة للمولود ومن اسبابها زيادة افراز الغدة الكظرية للجنين او تناول هرمون التيستستيرون من الام او زيادة في عدد كروموسومات الجنين .

❖ فتحه أحليلية تحتانية أو فوقانية : عيب خلقي يحصل في قضيب الطفل حيث تكون فتحة الاحليل أما في أسفل أو في أعلى القضيب للطفل بسبب تعرض الأم الحامل لمواد هرمونية او كيمياوية .

❖ القيلة المائية : تجمع مائي يحدث حول احدى الخصيتين ويؤدي الى ورم في كيس الصفن عند الذكر والسبب لهذا التشوه غير معروف .

- استسقاء الرأس الخلقي (congenital hydrocephalus) كبر حجم الرأس نتيجة لزيادة السائل الدماغي الشوكي المحيط بالمخ وزيادة الضغط على انسجة الدماغ والسبب الدقيق غير معروف وتعتقد التهاب الدماغ بسبب داء المقوسات (toxoplasmosis) او نزف دموي في الدماغ .
- انعدام فتحة الشرج (imperforated anus): فتحة المخرج غير موجودة .

(3-1-1-3) عيوب داخلية وخارجية معاً (مزدوجه) [1]

❖ صغر الرأس (microcephalus): ناتج عن التصاق عضام الجمجمة وضمور الدماغ فيصغر حجم الرأس بالنسبة لحجم الجسم بسبب تعرض الام لالتهابات فيروسية مثل جدري الماء او الحصبة الالمانية او تعرض الام لبعض الادوية والسموم في المدة الاولى من الحمل .

❖ التشوهات الخلقيه في الدماغ والحبل الشوكي مثل انعدام الدماغ (anencephaly) يولد الطفل بدون عظام الجمجمة او بدون الدماغ بسبب نقص حامض الفوليك .

(3-1-2) اسباب التشوهات الخلقية وتشخيصها :

- اسباب التشوهات الخلقية :

❖ اختلال الجينات : لكل فرد 46 كروموسوما" او 23 زوجا" من الكروموسومات لكل زوج من هذه الكروموسومات يحمل العديد من الجينات وهذه بدورها مسؤولة عن الصفات الوراثية

وكل جين يحمل صفتين واحده من الأم والاخرى من الاب وان أي سبب او نقص في ترتيبها او تكوينها يؤدي الى تشوهات خلقية ويوجد نوعين هما :

❖ العيوب الخلقية الناتجة عن الامراض الوراثية السائدة : وتحدث نتيجة عيب خلقي عند احد الابوين بسبب خلل في احد الصفتين مثل قصر القامه الشديد وتورث للجنين كذلك او يولد على احد الابوين سليم فيكون أحتمالية ولادته سليما" 50%وغير سليم 50%.

❖ العيوب الخلقية الناتجة عن الامراض الوراثية المتنحية : وفي هذه الحالة يظهر التشوه الخلقي في حالة النسختين غير سليمتين او يكون احد الوالدين حامل للتشوه لكن لم تظهر عليه علامات التشوه فقط يحمل جين العوق لكن في حالة تزاوج الاحفاد تظهر هذه التشوهات تتجمع هذه الجينات لتكون أزواجا من المورثات المعطوبة وتكون احتمالية ظهور المرض في الاحفاد 25% لكل مولود .

❖ توافق عامل (Rh) للابوين : نتيجة اختلاف الدم وعدم تطابقه بين الام والاب يؤدي تشوهات في الجنين او موت الجنين ولاسيما ان الاب يحمل عاملا" ريسيا" موجب +Rh والام عاملا" ريسيا" سالبا" –Rh وتوجد طريقة علمية اي حقن الام بمضادات الاختلاف بعد 72 من الولادة الاولى لتفادي التشوهات .

❖ عمر الام : تحدث التشوهات الخلقيه بسبب كبر عمر الام وربما تكون هذه التشوهات كبيرة ما يؤدي الى الاجهاض او موت الجنين بسبب زيادة عدد الكرموسومات مع زيادة عمر الام واختلال في عدد وتركيب كرموسومات يؤدي الى تشوهات ولادية .

❖ الاسباب البيئيه : نتيجة عوامل بيئيه خارج جسم الام اثناء مدة الحمل وسيما المدة الاولى من الحمل يؤدي الى التشوهات الخلقية ومن اهمها المايكروبات مثل الحصبة الالمانية او داء القطط تسبب تشوهات خلقية احياناً صغر حجم الرأس أو أمراض القلب أو تضخم الطحال والكبد .

❖ أمراض الأم المزمنة : مثل السكر وأخذ الأنسولين بجرعات عالية قبل الحمل وخلال الحمل يؤدي الى حدوث تشوهات خلقية مثل الشفة الارنبية وأمراض القلب .

❖ النقص في حامض الفوليك folic acid يؤدي الى تشوهات خلقية متكررة وأجهاض متكرر نتيجة تشوة الدماغ والحبل الشوكي .

- تناول العقاقير لاسيما في الفتره الاولى من الحمل مثل أدوية الصرع أو ضد السرطان .
- تعرض الام لارتفاع في درجة حرارة الجسم اكثر من 39 قد يصاب الجنين تشوة الانبوب العصبي .
- التعرض للاشعة السينيه (X rays) ، تعاطي المشروبات الكحولية والتدخين وانواع المخدرات يؤدي لحدوث تشوهات خلقية .
- تعرض الابوين الى الاشعاع لاسيما اليورانيوم المنضب نتيجة الحروب .
- الاخصاب الخارجي وطرائق الانجاب الحديثة فالاطفال المولودين بهذه الطرائق عرضة للاصابة بالتشوهات الخلقية .
- الرش بالمبيدات الحشرية يؤدي الى خطر الاصابة بالاجهاض والتشوهات الخلقية لدى المواليد .

المبحث الثاني : الجانب التطبيقي

(1-2-3) 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 (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 الانجليزية	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

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 الاذن 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 بالوجه 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

(3-2-2) متغيرات الدراسة

وشملت الدراسة (14) متغيرا" ورمزنا للمتغير المعتمد (y) وهو متغير الاستجابة العددية (count response variable) ويأخذ رقم الرمز من صفر الى (27) اي نوع العوق (التشوه الخلقي) ,

اي ان المتغير المعتمد يأخذ اعدادا" رقمية ومن هنا جاءت تسمية نماذج الاستجابة العددية , المتغيرات المستقلة 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

(3-2-3) معرفة التوزيع الأحتمالي الملائم للبيانات

لمعرفة التوزيع الأحتمالي لبيانات التشوهات الخلقية لحديثي الولادة تم تطبيق اختبار حسن المطابقة **(goodness of fit)** للبيانات عن طريق البرنامج الأحصائي **(easy fit Fit1)** 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 مؤشر بأن البيانات تتبع لتوزيع ثنائي الحدين السالب .

(3-2-4) تقدير معلمات أنموذج انحدار ثنائي الحدين السالب :-

توجد العديد من الطرائق الأحصائية لتقدير معلمات إنموذج انحدار ثنائي الحدين السالب ومنها طرائق الأمكان الأعظم (MLE) وطريقة المربعات الصغرى المعادة الوزن التكرارية (IRLS) وطريقة الأمكان الموزونة (WLE) وطريقة المربعات الصغرى الموزونة(WLS) .

(3-2-4-1) طريقة الامكان الاعظم (maximum likelihood estimation)

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

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

$$H_1 : \text{at least one of them not equal zero}$$

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

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

المتغيرات المستقلة	معاملات الأنحدار (b)	الخطأ المعياري	P- value
X1 (عمر الأم)	-0.0076193	0.0095585	0.425
X2 (مهنة الأم)	-0.2640292	0.3827613	0.490
X3 (عمرالأب)	0.0024065	0.00763	0.752
X4 (مهنة الاب)	-0.0790277	0.0426701	0.064
X5 (درجة القرابة)	-0.0517542	0.1166073	0.657
X6 (الولادات السابقة)	0.2388692	0.118685	0.044
X7 (نوع الولادة)	-0.5673884	0.2730285	0.038

X8 (الولادة الحاليه)	-0.3568640	0.127135	0.005
X9 (تعرض الأم)	-0.1255543	0.0389791	0.001
X10 (عدد الأسقاط)	-0.0488763	0.094799	0.606
X11 (نوع السكن)	0.0274382	0.1042203	0.792
X12 (جنس الطفل)	0.0646894	0.0958825	0.500
X13 (وزن الطفل)	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 p − value اقل من 0.05 اي ان العوامل المؤثره في نوع العوق هي وجود العوق في الولادات السابقة ونوع الولادة مفردة ام متعدده كذلك الولادة الحالية حية ام ميتة واخيرا" تعرض الأم الى حمى او اشعاع او تناول ادويه تساعد في حصول التشوة الخلقي لاسيما في المرحلة الأولى من الحمل. اما بقية العوامل فليس لها تأثير في نوع العوق .

(3-2-4-2) طريقة المربعات الصغرى معادة الوزن التكرارية (IRLS)

يستعمل فيها البرنامج الأحصائي (STATA 14) لتقدير وأختبار معلمات إنموذج انحدار ثنائي الحدين السالب علما ان فرضية العدم هي :

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

$$H_1 : \text{at least one of them not equal zero}$$

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

جدول (3-4) 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

(درجة القرابة) X5	-0.6986716	1.236127	0.527
(الولادات السابقة) X6	3.061872	1.279343	0.017
(نوع الولادة) X7	-5.0951	2.809135	0.071
(الولادة الحاليه) X8	-3.5937	1.361783	0.009
(تعرض الأم) X9	-1.433926	0.4265287	0.001
(عدد الأسقاط) X10	-0.4209224	1.074158	0.969
(نوع السكن) X11	0.3859155	1.142004	0.736
(جنس الطفل) X12	0.7500499	1.014451	0.460
(وزن الطفل) X13	0.0011191	0.0007725	0.149
الحد الثابت	23.76104	7.469959	0.002

من الجدول (3-4) تبين وجود خمسة عوامل أظهرت تأثيرأ معنويا في المتغير المعتمد(نوع العوق) اي ان قيم $\text{p} - \text{value}$ 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 مدة الحمل x_9 , أما بقية العوامل فكانت غير معنوية أي ليس لها تأثير في المتغير المعتمد (y) .

(3-2-4-3) طريقة الامكان الموزونة (WLE)

في هذه الطريقة تم استعمال البرنامج الأحصائي(STATA 14) لتقدير وأختبار معلمات إنموذج انحدار ثنائي الحدين السالب , علما ان فرضية العدم :

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

$$H_1 : \text{at least one of them not equal zero}$$

وباستعمال البرنامج الأحصائي (mini tab 17) إذ كان معامل التحديد (%14.30) ومتوسط مربعات الخطأ (69.82) وكانت النتائج مبينة في الجدول الأتي :

جدول (3-5) 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 (3-5) اي عامل من العوامل (المتغيرات المستقلة) تأثيرا" معنويا واضحا" على المتغير التابع (نوع العوق) اي ان جميع المتغيرات المستقلة كانت قيم p-value اكبر من 0.05 ونستنتج عدم رفض فرضية العدم ولجميع المتغيرات المستقله .

(3-2-4-4) طريقة المربعات الصغرى الموزونة (WLS)

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

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

$$H_1 : \text{at least one of them not equal zero}$$

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

جدول (3-6) 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 الأنموذج , ويتضح من الجدول (3-6) 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 .

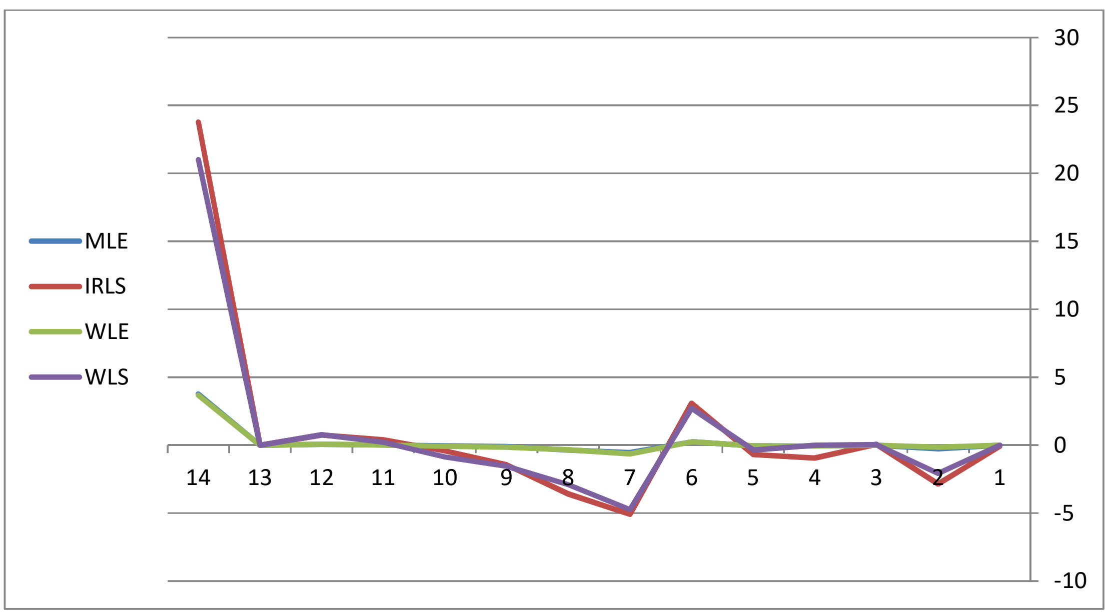
(3-2-5) المقارنة بين طرائق التقدير المختلفة :

ويمكن ان نقارن قيم معاملات الأنحدار بأربع طرائق مختلفة :

جدول (3-7) يبين قيم معاملات الأنحدار للطرق WLS, WLE, IRLS, MLE

معاملات الأنحدار(b)	MLE	IRLS	WLE	WLS

−0.066	−0.0076094	−0.0829436	−0.0076193	عمر الام
−2.087	−0.1626012	−2.869986	−0.2640292	مهنة الام
0.037	0.0019222	0.0552933	0.0024065	عمر الاب
−0.672	−0.066428	−0.9558305	−0.0790277	مهنة الاب
−0.357	−0.0388292	−0.6986716	−0.0517542	درجة القرابة بين الابوين
2.724	0.2415253	3.061872	0.2388692	الولادات السابقة
−4.752	−0.6431148	−5.0951	−0.5673884	نوع الولادة
−2.903	−0.337303	−3.5937	−0.3568640	الولادات الحالية
−1.545	−0.1550309	−1.433926	−0.1255543	تعرض الام
−0.869	−0.090717	−0.4209224	−0.0488763	عدد الاسقاط السابق
0.206	0.020613	0.3859155	0.0274382	نوع السكن
0.762	0.0696726	0.7500499	0.0646894	جنس الطفل
0.001	0.0001332	0.0011191	0.0001033	وزن الطفل
21.011	3.666405	23.76104	3.759862	الحد الثابت



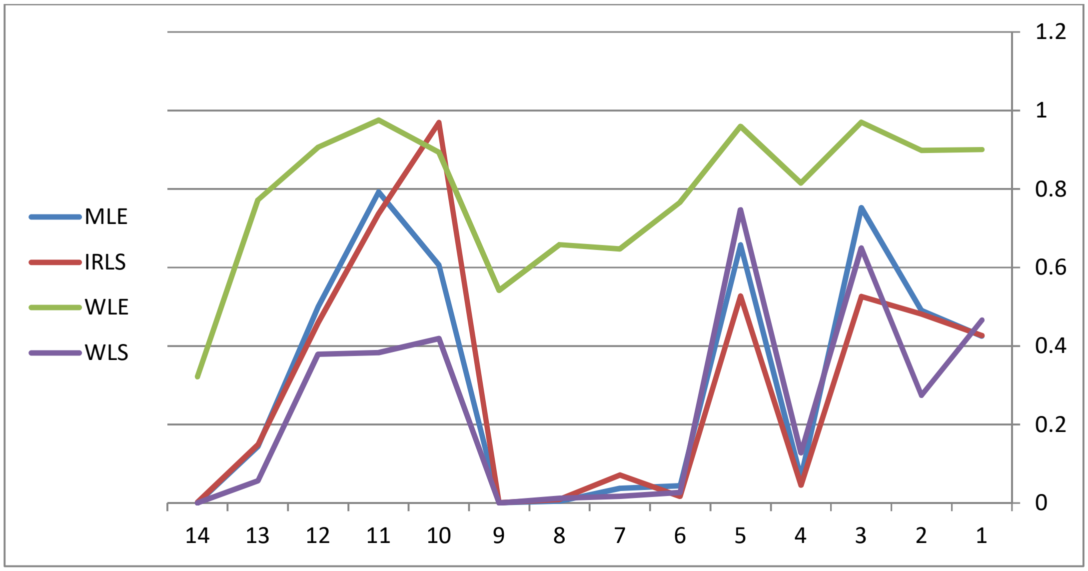
الشكل (3−1) يوضح معاملات الأنحدار للطرائق MLE ,IRLS ,WLE ,WLS

بالأعتماد على الجدول (3-7) نلحظ ان قيم معاملات الأنحدار في الطريقتين (MLE , WLE) تكون متقاربة جدا وتظهر وكأنها على خط واحد وأختفاء اللون الأزرق , اما قيم معاملات الأنحدار بالطريقتين (IRLS , WLS) فتكون قريبة لحد ما فيما بينها ومتباعدة عن الطريقتين السابقتين .

وعند مقارنة قيم p-value لأربع طرائق مختلفة كانت النتائج كما يأتي :

جدول (3-8) يبين قيم p-value لطرق التقدير MLE ,IRLS ,WLE ,WLS

p-value	MLE	IRLS	WLE	WLS
عمر الام	0.425	0.427	0.900	0.466
مهنة الام	0.490	0.481	0.898	0.274
عمر الاب	0.752	0.526	0.970	0.650
مهنة الاب	0.064	0.045	0.815	0.128
درجة القرابة بين الابوين	0.657	0.527	0.959	0.747
الولادات السابقة	0.044	0.017	0.766	0.027
نوع الولادة	0.038	0.071	0.647	0.017
الولادات الحالية	0.005	0.009	0.658	0.012
تعرض الام	0.001	0.001	0.542	0.000
عدد الاسقاط السابق	0.606	0.969	0.893	0.419
نوع السكن	0.792	0.736	0.975	0.383
جنس الطفل	0.500	0.460	0.906	0.379
وزن الطفل	0.144	0.149	0.772	0.057
الحد الثابت	0.000	0.002	0.322	0.000



الشكل (3-2) يوضح قيم p-value لطرائق التقدير MLE,IRLS,WLE,WLS

من الجدول (3-8) نلحظ تشابه نتائج طريقة MLE مع نتائج طريقة IRLS مع أختلاف بمتغير واحد في طريقة IRLS

وبالأعتماد على الجدول (3-9) لمعامل التحديد ومتوسط مربعات الخطأ لكل طريقه من طرائق التقدير المذكوره :

الجدول (3-9) يبين معامل التحديد ومتوسط مربعات الخطأ لطرائق التقديرMLE,IRLS,WLE,WLS

المعيار	MLE	IRLS	WLE	WLS
MSE	66.32	54.73	69.82	62.76
R^2	%15.25	%19	%14.3	%17.5

يظهر ان طريقة المربعات الصغرى المعادة الوزن التكرارية **IRLS** هي أفضل طريقة للتقدير وعن طريق النتائج التي تطرقنا لها إذ تمتلك أقل متوسط مربعات للخطأ **(54.73)** وأعلى معامل تحديد أذ بلغ **(%19)** ثم طريقة المربعات الصغرى الموزونة WLS بمتوسط مربعات الخطأ (62.76) ومعامل تحديد (%17.5) ومن ثم طريقة الأمكان الأعظم MLE بمتوسط مربعات الخطأ(66.32)

ومعامل تحديد (15.25%) واخيرا" طريقة الأمكان الأعظم الموزونة بمتوسط مربعات الخطأ (69.82) ومعامل تحديد (14.3%) .

إذ ان قيمة معامل التحديد ظهرت منخفضه لان عدد المتغيرات المدروسة قليلا" وأصبح غير كافي لدراسة التشوهات الخلقية .

الاستنتاجات والتوصيات

أولا" : الاستنتاجات

1- توصل البحث الى أن أفضل توزيع لبيانات حديثي الولادة الذين يعانون من التشوهات الخلقية هو توزيع ثنائي الحدين السالب .

2- عن طريق تقدير معلمات إنموذج انحدار ثنائي الحدين السالب وعن طريق ما قدمتا طريقة المربعات الصغرى المعادة الوزن التكرارية (الحصينة) IRLS أداء" هو الأفضل من بين الطرائق الأربع ومن ثم طريقة المربعات الصغرى الموزونة (WLS) وبعدها طريقة الامكان الاعظم (MLE) جميعها قدمت اداء افضل من طريقة الامكان الموزونة WLE .

3- وعن طريق معامل التحديد R^2 يتضح ان افضل طريقة لتقدير معلمات الإنموذج هي طريقة المربعات الصغرى المعادة الوزن التكرارية لانها تمتلك أعلى معامل تحديد R^2 . وكذلك عن طريق متوسط مربعات الخطأ MSE ان افضل طريقة لتقدير معلمات الإنموذج هي طريقة المربعات الصغرى المعادة الوزن التكرارية لانها تمتلك اقل متوسط مربعات خطأ MSE . وبهذا تصبح افضل طريقة لتقدير معلمات الإنموذج هي طريقة المربعات الصغرى المعادة الوزن التكرارية وتأتي بعدها طريقة المربعات الصغرى الموزونة .

ثانيا" : التوصيات

1. ضرورة استعمال إنموذج انحدار ثنائي الحدين السالب عند دراسة البيانات التي تحتوي خاصية فرط التشتت overdispertion اي يكون فيها التباين اعلى من الوسط الحسابي
2. ضرورة تطبيق طريقتي المربعات الصغرى المعادة الوزن التكرارية (الحصينة) والمربعات الصغرى الموزونة بسبب ان مقدراتها افضل من الطريقتين الباقيتين عند تطبيقها على إنموذج انحدار ثنائي في الحدين السالب .
3. ضرورة البحث في البيانات العددية count data وبناء النماذج الخاصة بها لاسيما البيانات الطبية منها لافتقار البحوث الحديثة لهذه الموضوعات والظواهر المتعلقة بها .
4. زيادة الوعي الصحي عن طريق الأعلانات والنشرات الصحية , كذلك تحديث بيانات دوائر الصحة وعدم الأعتماد على السجلات القديمة التي أصبح دورها ضعيفا" جدا في

حدوث التشوهات الخلقية , وتفعيل دور فرق البيئة ومكافحة كل أشكال التلوث البيئي والحياتي .

5. – لظهور قيمة معامل التحديد R^2 منخفض وعن طريق تعريف معامل التحديد هو مدى مساهمة العوامل (المتغيرات المستقلة) في التغيرات التي تطرأ على المتغيير المعتمد اي نوع العوق ، في عراق ما بعد 1980 والحروب التي عصفت بالعراق بدات هذه المتغيرات المستقل (العوامل) دورها ينحسر بسبب سقوط اطنان من المتفجرات والاشعاع والتلوث البيئي بكل اشكال (شحة المياه وتلوثها ، استعمال المواد الكيمياوية ، المبيدات الزراعية في المواد الغذائي وغيرها) التطور التكنولوجي الكبير والموجات الكهرومغناطيسية واجهزة الاتصال وابراج البث وانتشارها في المدن .مع تدهور الوضع النفسي وانخفاض المستوى الثقافي والوعي الصحي كلها عوامل لها دور كبير في حصول التشوهات الخلقية واصبح دور هذه العوامل اكبر من العوامل والمتغيرات المستقلة التي تناولناها في دراستنا , اي نوصي بأضافة متغيرات مستقلة جديدة .

6. نوصي بدراسة تقدير معلمات ثنائي الحدين السالب المقطوع في مجالات العلوم الطبية والهندسية والاجتماعية .

المصادر

أولا : المصادر العربية

- القرآن الكريم

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 الادارة 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 الادارة والأقتصاد جامعة بغداد (2017)

ثانيا" : المصادر الأجنبية

4. Abel , G . J ., " International Migration Flow Table Estimation , Ph.D. thesis for faculty of Low , Arts & Social Sciences , University of Southampton (2009)
5. Allison , D ,R.Waterman ,fixed effect negative binomial regressionmodel.(2002)
6. Amiguet , M ."Weighted likelihood Negative Binomial Regression , Institute for social and preventive medicine , University of Lausanne , Saint Petersburg (2013)
7. Anscombe , J .,the transformation binomial , poison ,negative binomial data , biometrika 35 ,pp(246−254)(1948)
8. Anscombe , J. , sampling theory for the negative binomial and logarithmic series distribution (1972)

9. Anscombe , J. ,the statistical analysis of insect counts based on the negative binomial distribution , biometrika 5 : pp(165−173) (1949)

10. Anscombe, F. J ., sampling theory for the negative binomial and logarithmic series distribution , Biometrika 37 (3/4):pp (368-382)(1950).

11. Anscombe, F. J .,The statistical analysis of insect counts based on the negative binomial distribution , Biometrika 5 :pp(165−173) (1949).

12. Bartlett , M . S ."The use of transformations , Biometrics 3 :pp(39-52)(1947)

13. Beall , G ."The transformation of data from entomological field experiments so that analysis of variance becomes applicable , Biometrika 29 :pp(243-262)(1942)

14. Bulmer, M.G. " On fitting the poisson lognormal distribution to Species – Abundance Data ", Biometrics, Vol.30, No.1, PP. 101–110.(1974)

15. Cameron ,A . C .and P . K . Trivedi ,"Regression analysis of count data , Econometric society monographs No.30 , Cambridge University Press (2013)

16. Dey, K. D.,& Chung, Y. " compound poisson distiribution : properties and estimation " , Commun. Statist. – Theory Meth ., Vol.21, No. 11, PP.(3097–3121)(1992).

17. Eggenberger , F . and G .Polya ,Uber die statistik ver ketteter vorgange , Journal of applied Mathematics and Mechanics 1: pp(279-289)(1923)

18. Emmanual , C . "Modelling and data analysis , Acomparison of Poisson or Negative Binomial Regression and Lee-Carter Models of Forcasting Norwegian male mortality , Master thesis for faculty of Mathematics and Natural Sciences , University of Oslo (2015)

19. Evans , D . A .,Experimental evidence concerning contagious distributions in ecology , Biometrika 40 :pp(186-211)(1953)

20. Faraway , J.,extending the linear model with R ,Boca Raton , chapman and hall /CRC Press (2006).

21. Fisher, R.A. , Corbet , S. ,& Williams , C.B. " The relation between The number of individuals In A Random Sample of An Anmal Population " , Journal of Animal Ecology Vol. 12, No.1, PP. 4–58.(1943)

22. Ghitany, M,& AL- Mutairi , D.K., " estimation methods for the discrete poisson –lindley distribution " , Journal of Statistical computation and simulation, Vol. 79, No. 1, PP.(1–9)(2009).

23. Ghitany, M. , Al – Mutairi, D. K. ,& Nadarajah, Zero – truncated Poisson – Lindley distribution and its application " , mathematics and computers in simulation, No. 79, PP. (279–287).(2008)

24. Gosset, K.Pearson , aderivation of negative binomial in Biometric Laboratory in London (student ,1907) .

25. Green , W . H ., Accounting for excess zeros and sample selection in Poisson and Negative Binomial Regression models , EC -94-10(1994)

26. Greenwood ,M .and G .U .Yule , an inquiry into the nature of frequency distributions of multiple happenings , with particular references to the occurrence of multiple attacks of disease or repeated accidents , Journal of the Royal statistical society A ,83 :pp(255-279)(1920)

27. Gurland, J. "Some Interrelations among compound and generalized distributions" , Biometrika,Vol. 44, No.1/2, PP.(256 –268)(1957).

28. Hauptfleish , F . "Mixed Poisson Models for claim counts ,Faculty of Mathematics and Physics , Charles university , Prague (2017)

29. Hilbe , J . M ." a generalized linear model in international incyclopedia of statistical sciences , ed . M . lovric ,NewYork, Springer(2010_b)

30. Hilbe , J . M ."Censored Negative Binomial Regression , research papers in economics , Boston school of economics (2005c)

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

32.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).

33.Jain ,G . C.,and P. Cosul , A generalized negative binomial distribution , Journal of applied mathematics , 21(4) ,pp(501-513)(1959).

34.Jones ,R . ,Maillardet and A . Robinson , scientific programing and simulation using R , Boca ,Raton , (2009).

35.Jung , J. ,"Using generalized linear models with mixed random components to analyze count data , master thesis , in mathematics , B.S Kyungpook National University (1999)

36.Karlis, D. " EM algorithm for mixed poisson And Other discrete distribution " , Astin Bulletin, Vol. 35, No. 1, PP.(3–24)(2005).

37.Kocherlakota, S"On the compounded bivariate poisson distribution: A unified treatment" . Ann. Inst. Statist. Math.,Vol. 40, No.1, PP.(61–76)(1988).

38.McCullagh ,P . , and J . A . Nelder , generalized linear models , second edition ,New York : chapman & hall (1989).

39.McCullagh ,P . , and J . A . Nelder , generalized linear models , first edition ,New York : chapman & hall (1982).

40.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)

41.Nelder , J . A .,Generalize linear models with negative binomial or beta binomial error , (1994) .

42.Nelder , J . A .and R . W . M .Wedderburn , generalized linear models , journal of the Royal statistical society , A 135 (3) pp(370-384) (1972)

43.Nelder , J .A .and D .Pregibon "an extended quasi-likelihood function , Biometrika 74 :pp(221-232)(1987)

44. Okech ,N . O ." Negative Binomial distributions for fixed and random parameters , Schools of Mathematics , College of Biological & Physical Sciences , University of Nairobi (2011)

45. Özel, G., & Inal, C. " The probability function Of a geometric poisson distribution " , Journal of statistical computation and simulation, Vol. 80, No. 5 , PP.(479–487)(2010).

46. Sankaran, M. " The discrete poisson – Lindley distribution, biometrics,Vol. 26, No.1, PP.(145–149)(1970)."

47. Shaban, S.A" On the discrete poisson – inverse gaussian distribution " , Biometric, Vol.23, No. 3, PP.(297–303)(1981).

48. Simon , L ." Fitting negative binomial distribution by the method of maximum likelihood , proceedings of casuality actuarial society 48 pp(45-53)(1961).

49. Williamson , E. AND M.H. Bretherton " tables of the negative binomial distribution , New York : Wily (1963) .

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

51. Zaky, B. , Mouba , A. , & Kotb,.S. " AN emprical study For compound distributions " , Journal Of Faculty Of Commerce – Al Azhar University . Girl’s Branch , Vol. 10,(PP. 1–17)(1993).

52. Zwilling , M . L .," Negative Binomial Regression , Department of Mathematics , University of Mount Union , the Mathematica Journal (2013)

Abstract

Consider the negative binomial regression model one of the count models , that used to represent many phenomena and cases , these can not be expressed by ordinary models , necessarily we must use a count models generally and negative binomial regression spetially to represent these phenomena and cases , as well as lack of the researches and modern studies for these threads particularly health topics . Have been estimated the parameters of negative binomial regression model by four different methods : [maximum likelihood estimation (MLE) ,iteratively re-weighted least square (IRLS) , weighted likelihood estimation (WLE) , weighted least square (WLS) . The goal of the study to reach the best method for estimation , where taked a random sample (257) patients are newborn suffering from congenital anomalies , they registered in the department of health Babylon . by using statistical programs (stata , spss , minitab) to estimate the parameters of the models and determine the best method. The show results that the method of (IRLS)is the best because it has the lowest mean square error MSE and the highest R^2

Republic of Iraq
Ministry of higher Education and Scientific Research
University of karbala
Faculty of Administration and Economics
Department of statistics

Choice of best estimate of regression model to negative binomial distribution with practical application

A Thesis Submitted to
Council of The Administration and Economics / Karbala University
as Partial fulfillment of the Requirements for the Degree of Master of Science in Statistics

Researcher

Adnan Fadhel Touma

Supervised By

Assistant Prof. Dr.Shrooq Abd–ALRida Saeed

2018 **1439**