تقدير معلمات Log-Logistic بأستخدام الخوارزميه الجينيه مع تطبيق تجربة المحاكاة

أ.م.د مشتاق كريم عبد الرحيم
كلية الادارة والاقتصاد/ جامعة كربلاء
mushtag.k@uokerbala.edu.iq

الباحث حسين خليل عبيد
كلية الادارة والاقتصاد/ جامعة كربلاء
hussain.kh@s.uokerbala.edu.iq

ملخص

في هذا البحث تم تقديم احد أهم نماذج الانحدار غير الخطية التي له استعمالات عديدة في كثير من التطبيقات الاحصائية هو انموذج الانحدار (log–logistic) الثنائي وتم تقدير معلمات هذا الانموذج وذلك عن طريق استعمال الطريقة الاحصائية الكلاسيكية وكذلك تم توظيف بعض الطرائق الاعتيادية التي تم تحسينها عن طريق الخوارزمية الجينية في التقدير وذلك لكي تلائم تقدير معلمات هذا النوع من نماذج الانحدار الغير خطية، ومن ثم تمت المقارنة بين هذه الطرائق وقد تضمنت المقارنة بين الطرائق الاعتيادية هي طريقة الامكان الاعظم، وطريقة المربعات الصغرى الموزونة و طريقة تصغير مربع كأي أما طرائق التقدير المحسنة التي تضمنت الخوارزمية الجينية بالاعتماد على تقنية طريقة الامكان الاعظم وطريقة الخوارزمية الجينية بالاعتماد على تقنية تصغير مربع كأي. ولكي نبين الفضلى بينهما من حيث تقدير الانموذج من خلال افتراض عدد من النماذج واحجام عينات مختلفة من خلال تجربة المحاكاة وذلك عن طريق المعيار الاحصائي (MSE) وذلك لغرض تقدير انموذج الانحدار (log–logistic).

الكلمات المفتاحية: الانحدار (log–logistic) الثنائي، الخوارزمية الجينية، أنموذج اللوجت، اختبار هوزمر– ليمشو لجودة المطابقة طريقة الإمكان الاعظم (MLE)..

"Estimation of Log-Logistic Parameters Using Genetic Algorithm with Simulation Experiment Application

Hussein Khalil Obaid
College of Administration and Economics/University of Karbala
hussain.kh@s.uokerbala.edu.iq

Prof. Mushtaq Karim Abdul Rahim
College of Administration and Economics/University of Karbala
mushtag.k@uokerbala.edu.iq

Abstract

In this research، one of the most important non-linear regression models that has many uses in many statistical applications was presented، which is the binary (log-logistic) regression model. The parameters of this model were estimated by using the classical statistical method. It was improved by the genetic algorithm in estimation in order to fit the estimation of the parameters of this type of non-linear regression models، and then a comparison was made between these methods. The improved estimation that included the genetic algorithm based on the technique of the greatest possibility method and the genetic algorithm method based on the square minimization technique as any. In order to show the best between them in terms of model estimation by assuming a number of models and different sample sizes through the simulation experiment through the statistical standard (MSE) for the purpose of estimating the log-logistic regression model.

Keywords: Binary log-logistic regression، genetic algorithm، log model، Hozmer-Lemshaw test for quality of fit، greatest possibility method (MLE).

المقدمة

يعرف الانحدار بانه طريقة احصائية تستخدم في التمويل والاستثمار والتخصصات الاخرى التي تحاول تحديد قوة شخصية العلاقة بين متغير تابع واحد (يشار إليه عادة بواسطة (**Dependent Variable (Y)**) وسلسلة من متغيرات الاخرى تسمى المتغيرات المفسرة (المستقلة) (**Independent Variable**) يرمز له ب (X). وبذلك يتم التقدير او التنبؤ بمتوسط قيمة المتغير التابع بمعلومية المتغيرات المستقلة. بناء على ذلك فان أسلوب الانحدار يستخدم للتوصيل الى أنموذج رياضي يوضح العلاقة الكمية بين المتغير التابع والمتغيرات المستقلة، مثال على ذلك أرتفاع ضغط دم شخص متمثل ب (**Y**) وعلى عمر الشخص نفسه متمثل ب(**X**) هذا الارتباط و التابعية بين (**X**) و(**Y**) هي ما ندعوه بالانحدار. ينقسم الانحدار بصورة عامة الى الانحدار الخطي والذي يكون على نوعين الانحدار الخطي بسيط والانحدار الخطي المتعدد، وقسم الثاني هو الانحدار اللاخطي.سوف يتم دراسة طريقة الإمكان الأعظم (**Maximum Likelihood Method**) لتقدير معلماته وسوف يتم تحسين هذه الطريقة بأستعمال الخوارزمية الجينية (**Genetic Algorithm**) حيث ان الخوارزمية الجينية سوف تعمل هنا (**Optimization**) لتقدير معلمات أنموذج الانحدار (**log–logistic**).

٢-مشكلة البحث:

تكمن مشكلة البحث في وجود المعادلات غير الخطية في انموذج الانحدار (**log–logistic**) والتي تؤثر على معلمات الانموذج عند استعمال طريقة التقدير الامكان الاعظم (**MLE**) وبذلك تكون المعلمة غير دالة احصائياً عند تحليل البيانات التي يكون فيها متغير الاستجابة متقطع والمتغيرات التوضيحية كمية واخرى نوعية ولكي تكون المعلمة ذات دلالة احصائية جيدة يتم تحسينها باستعمال الخوارزمية الجينية (**GA**).

٣-هدف البحث:

ان الهدف من هذا البحث هو تحسين طرائق التقدير الاعتيادية عن طريق الخوارزمية الجينية واستعمال المعيار الاحصائي متوسط مربعات الخطأ لغرض التوصل الى اقل تأثير للمتغيرات المستقلة.

٤-الجانب النظري

٤-١ مفهوم الانحدار اللوجستي الثنائي (The Concept of Log-Logistic Regression) [6]

هو انموذج احصائي يستخدم للبيانات الفئوية (Categorical data) وينتمي الى نماذج الانحدار اللاأخطية ويعد احد الاساليب الاحصائية المهمة التي تستخدم للتنبؤ باحتمالية وقوع حدث معين وذلك من خلال ترتيب البيانات والمعلومات على المنحى اللوجستي. يعمل انحدار اللوجستي اللوغاريتمي لوصف العلاقة بين متغير الاستجابة والمتغيرات التوضيحية. لذلك يعد أنموذج الانحدار (Log–Logistic) حالة خاصة من نماذج الانحدار الاعتيادية بسبب طبيعته الاسمية التي يحملها متغير الاستجابة مثل (نجاح، فشل) (موفق، غير موفق) ويستخدم الانحدار اللوجستي بشكل كبير في كثير من التطبيقات الاحصائية وتحليل البيانات في معظم الدراسات الطبية والهندسية والاقتصادية والغاية منه هو تقدير الانموذج الذي يمثل العلاقة الغير خطية بين المتغيرات لكي يتم تطبيقها في التنبؤ الاحصائي كالتنبؤ بوقوع حدث معين او عدم وقوعه. ويكون على عدة انواع الانحدار اللوجستي الثنائي (Binary Logistic Regression) ويستخدم هذا النوع عندما يأخذ متغير الاستجابة قيمتين هما (١، ٠). أما الانحدار اللوجستي المتعدد الحدود (Multinomial Logistic Regression) يتم استعماله عندما يكون متغير الاستجابة متعدد القيم (اكثر من قيمتين) وكذلك الانحدار اللوجستي الترتيبي (Ordinal Logistic Regression) وفيه تكون متغيرات الاستجابة رتبويا. في دراستنا سوف نركز ونقتصر في العرض النظري على الانحدار (Log-Logistic Regression) الثنائي.

٤-٢ الدالة الوجستية (Logistic function)

دائما ما يعرف انموذج الانحدار (Log–Logistic) الثنائي على انه أحد نماذج الانحدار اللاأخطية حيث يبنى على فروض اساسية الذي يكون فيها متغير الاستجابة الذي نهتم بدراسته يتبع توزيع برنولي (Bernoulli)، وعليه يكون أحتمال النجاح عندما (y=١) واحتمال فشل ١-عندما (y=٠) بذلك تكون صيغة دالة الكثافة الاحتمالية تكتب كالآتي

$$P(y_i \backslash X_i) = [\pi(X_i)]^{y_i} \; [1-\pi(X_i)]^{1-y_i} \ldots\ldots \qquad (1)$$

وعند تعويض قيم0،1 = yi سيكون P(y)، فان

$$P(Y) = \begin{cases} \pi_i & \text{عندما } Y = 1 \text{ عند حدوث الاستجابة} \\ 1-\pi_i & \text{عندما } Y = 0 \text{ عند عدم حدوث الاستجابة} \end{cases} \quad ..(2)$$

π_i: يمثل احتمال النجاح (Success)

$1 - \pi\_i$: يمثل احتمال الفشل (Failure)

وعندما yi =1 نحصل على

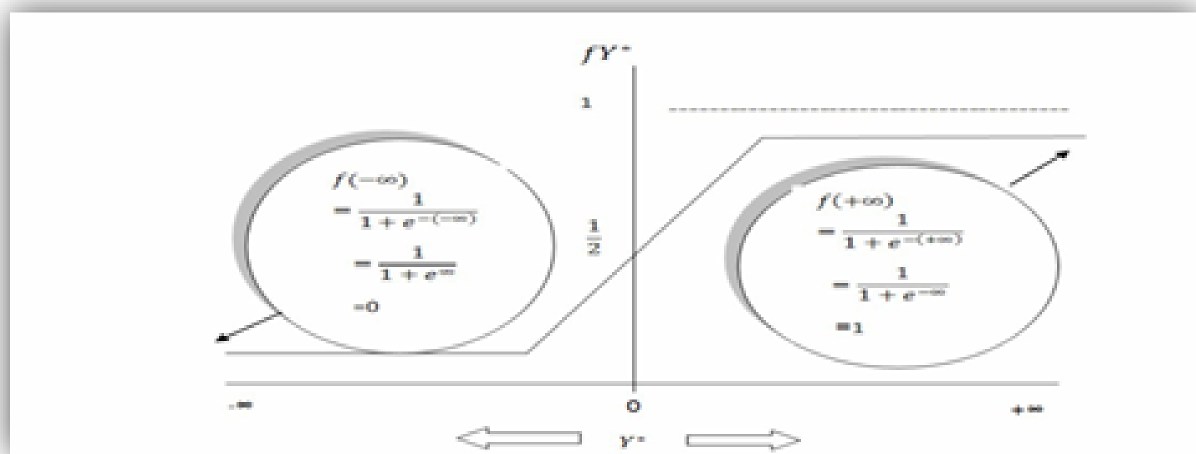
$$\pi i = \frac{e^{B_0 + B_1 X_1}}{1+e^{B_0 + B_1 X_1}} \quad \ldots\ldots\ldots \quad (3)$$

وعندما yi= 0 نحصل على :

$$1 - \pi i = 1 - \frac{e^{B_0 + B_1 X_1}}{1 + e^{B_0 + B_1 X_1}}$$

$$1 - \pi i = \frac{1}{1 + e^{B_0 + B_1 X_1}} \quad \ldots\ldots\ldots \quad (4)$$

الشكل (١) الشكل العام للدالة اللوجستية



من الشكل أعلاه، يتضح بِأنَّ الدالة اللوجستية f (y*) هي دالة احتمالية تكون محصورة ضمن القيم (0،1)

٤-٣ أنموذج اللوجت (logit model) (٢)

يعرف بانه في الانحدار اللوجستي هو حالة خاصة لوظيفة الارتباط في نموذج خطي معمم (انها وظيفة الارتباط المتعارف عليها لتوزيع برنولي) . اذن هو عبارة عن التحويل الخطي لدالة الانحدار اللوجستي (دالة **logit**) حيث تستخدم لإزالة الانحناءات الموجودة بالأنموذج اللوجستي لتأثيرها السلبي على مقدرات المعلمات من خلال اخذ اللوغاريتم لمعاملات المفاضلة لتحويل العلاقة بين احتمالية حدوث الاستجابة والمتغيرات التوضيحية الى علاقة خطية بتطبيق أنموذج الانحدار الخطي ضمن المجال {-∞،∞}حيث تأخذ الدالة **logit** القيم بين اللانهاية السالبة والموجبة. وحسب الصيغة الرياضية التالية :

$$\mathrm{logit}\,(\pi_i) = \ln \frac{\pi_i}{1-\pi_i} \qquad \ldots (5)$$

$$= \ln(\mathrm{odds})$$

$$= \ln(e^{\beta_\circ+\beta_1 X_1+\cdots+\beta_p X_p})$$

$$= \ln(e^{\boxed{}})$$

$$= \boxed{}$$

$$Z = \qquad \ldots (6)$$

وعليه ان التحويل الخطي لدالة الانحدار اللوجستي (Z_i) يتوزع توزيعا طبيعيا تقاربيا

(*Asymptotically Normal distribution*) اي ان:

$$Z_i \sim \mathrm{Asym.N.}\left\{X_i'\beta,\ \left[\pi_i\left(1-\pi_i\right)\right]^{-1}\right\} \qquad \ldots (7)$$

4-4 طرائق تقدير معلمات أنموذج (Log-Logistic) الثنائي:

(Methods Of Estimating Parameters Of The Binary Logistic Model)

4-4-1طريقة الامكان الاعظم [Maximum Likelihood Method (MLM)][5]

تقوم هذه الطريقة على مبدأ ايجاد مقدرات للمعلمات عن طريق جعل دالة الامكان الاعظم في نهايتها العظمى، وعليه فان أنموذج الانحدار اللوجستي الثنائي يكون فيه متغير الاستجابة (Y_i) الثنائي والذي يعد أنموذج اللوجستك أحدها يتوزع كما ذكرنا حسب توزيع برنولي اي ان له مستويان هما الصفر والواحد حسب الصيغة (1) وكما يلي:

$$P(Y_i/X_i) = [\pi(X_i)]^{Y_i}[1 - \pi(X_i)]^{1-Y_i}$$

وللحصول على تقديرات المعلمة بواسطة الامكان الاعظم (MLE) يتم ضرب الحدود في الصيغة $(1$ لعينة حجمها (n) تعتمد على (X)من المتغيرات التوضيحية (التفسيرية) ومجموعة متغير الاستجابة (Y) حسب الصيغة الرياضية الآتية:

$$P(Y/X) = \prod_{i=1}^{n}[\pi(X_i)]^{Y_i}[1 - \pi(X_i)]^{1-Y_i} \qquad \dots(8)$$

وان $\pi(X_i)$ و $P(Y/X)$ تعتمد على المعلمات وهدفنا هو تقدير المعلمات الغير معلومة فبإمكاننا تحديد دالة الإمكان $l(\beta)$ للكشف عن التبعية.

$$l(\beta) = P(Y/X)$$

وان $l(\beta)$: يمثل الامكان (الاحتمال) لمشاهدات العينة Y.

والغرض من (MLE) هو ايجاد التقدير الأفضل لمتجه $(\hat{\beta})$ الذي يزيد من احتمال المشاهدات (Y) وحسب ما

وبالتالي ان دالة الامكان الاعظم في الانموذج اللوجستي الثنائي هي:

$$l(\beta)=\prod_{i=1}^{n}[\pi(X_i)]^{Y_i}[1-\pi(X_i)]^{1-Y_i}$$

وبأخذ (ln) على الطرفين للحصول (MLE) لتسهيل عملية الحل

$$ln[l(\beta)]=\sum_{i=1}^{n}Y_i ln[\pi(X_i)] + (1-Y_i)ln[1-\pi(X_i)] \qquad ...(9)$$

وبالتعويض عن $\pi(X_i)$ و $1-\pi(X_i)$ بما يساويهم كالاتي:

$$ln[l(\beta)]=\sum_{i=1}^{n}\left[Y_i ln\left(\frac{e^{\underline{\acute{X}_i\beta}}}{1+e^{\underline{\acute{X}_i\beta}}}\right) + (1-Y_i)ln\left(1-\frac{e^{\underline{\acute{X}_i\beta}}}{1+e^{\underline{\acute{X}_i\beta}}}\right)\right]$$

$$=\sum_{i=1}^{n}\left[Y_i ln\left(e^{\underline{\acute{X}_i\beta}}\right)-Y_i ln\left(1+e^{\underline{\acute{X}_i\beta}}\right)+ln\left(\frac{1}{1+e^{\underline{\acute{X}_i\beta}}}\right) - Y_i ln\left(\frac{1}{1+e^{\underline{\acute{X}_i\beta}}}\right)\right]$$

$$=\sum_{i=1}^{n}\left[Y_i\left(\underline{\acute{X}_i\beta}\right)-ln\left(1+e^{\underline{\acute{X}_i\beta}}\right)\right] \qquad \dots(10)$$

ومن اجل الحصول على تقديرات (β) وهي $(\hat{\beta})$ لتعظيم لوغارتم دالة الامكان $ln[l(\beta)] = L\left(\underline{\beta}\right)$ تؤخذ المشتقات من الدرجة الاولى ومساواة الدالة الناتجة بالصفر لكل j من الصيغ و j من المعلمات اي نستخرج المشتقة الجزئية الاولى لكل β وحسب ما يأتي:

$$\acute{L}\left(\underline{\beta}\right) = l\acute{n}[l(\beta)] = \frac{\partial L\left(\underline{\beta}\right)}{\partial \underline{\beta}} = \frac{\partial l\acute{n}[l(\beta)]}{\partial \underline{\beta}}$$

$$= \sum_{i=1}^{n}\left[Y_i \frac{\partial}{\partial \beta_j}\left(\underline{\acute{X}_i}\underline{\beta}\right) - \frac{\partial}{\partial \beta_j} ln\left(1 + e^{\underline{\acute{X}_i}\underline{\beta}}\right)\right]$$

$$\frac{\partial}{\partial \beta_j}\left(\underline{\acute{X}_i}\underline{\beta}\right) = \frac{\partial}{\partial \beta_j}\left(\beta_\circ + \beta_1 X_{i1} + \cdots + \beta_p X_{ip}\right)$$

$$= \acute{X}_{ij}$$

and $\quad X_{i0} = 1$

$$\frac{\partial}{\partial \beta_j} ln\left(1 + e^{\underline{\acute{X}_i}\underline{\beta}}\right) = \frac{1}{1+e^{\underline{\acute{X}_i}\underline{\beta}}} \; \frac{\partial}{\partial \beta_j}\left(1 + e^{\underline{\acute{X}_i}\underline{\beta}}\right)$$

$$= \frac{1}{1 + e^{\underline{\acute{X}_i}\underline{\beta}}} \; e^{\underline{\acute{X}_i}\underline{\beta}} \; \frac{\partial}{\partial \beta_j}\left(\underline{\acute{X}_i}\underline{\beta}\right)$$

$$= \pi(X_i)\acute{X}_{ij}$$

$$\therefore \acute{L}\left(\underline{\beta}\right) = \sum_{i=1}^{n}\left[Y_i \acute{X}_{ij} - \pi(X_i)\acute{X}_{ij}\right] = 0 \qquad \ldots(11)$$

$$\sum_{i=1}^{n}\left[\left(Y_i - \pi(X_i)\right)\acute{X}_{ij}\right] = 0$$

ونعوض قيمة $\pi(X_i)$ فينتج:

$$\acute{L}\left(\underline{\beta}\right) = \sum_{i=1}^{n}\left[\left(Y_i - \frac{e^{\underline{\acute{X}}_i\underline{\beta}}}{1+e^{\underline{\acute{X}}_i\underline{\beta}}}\right)\acute{X}_{ij}\right] = 0$$

نستنتج بانه تكونت لدينا $(p + 1)$ من المعادلات غير الخطية ونقوم بحلها بإحدى الطرائق التكرارية التقليدية كطريقة نيوتن رافسون (NR) لإيجاد تقديرات دالة الامكان الاعظم في المشتقة من الدرجة الاولى ومساواتها للصفر

ولذلك فان خوارزمية نيوتن رافسون التكرارية لإيجاد قيم (β) التقديرية لدالة الامكان الاعظم في الانموذج اللوجستي ستكون في $(m + 1)$ من التكرارات وكالآتي:

$$\hat{\beta}^{(m+1)} = \hat{\beta}^{(m)} + (\acute{X}\,\mathrm{V}^{(\mathrm{m})}\,X)^{-1}\,\acute{X}\left(Y - P^{(m)}\right) \qquad \dots(12)$$

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} , \qquad P^{(m)} = \begin{bmatrix} \pi_1{}^{(m)} \\ \pi_2{}^{(m)} \\ \vdots \\ \pi_n{}^{(m)} \end{bmatrix} , \qquad X = \begin{bmatrix} \underline{\acute{X}}_1 \\ \underline{\acute{X}}_2 \\ \vdots \\ \underline{\acute{X}}_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1p} \\ 1 & X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{np} \end{bmatrix}$$

$$\mathrm{V}^{(\mathrm{m})} = \begin{bmatrix} \pi_1{}^m(1-\pi_1{}^m) & 0 & \cdots & 0 \\ 0 & \pi_2{}^m(1-\pi_2{}^m) & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \pi_n{}^m(1-\pi_n{}^m) \end{bmatrix}$$

حيث ان:

Y: يمثل متجه متغير الاستجابة ذو رتبة (n * 1)

$P^{(m)}$: يمثل القيم الاحتمالية لحدوث متغير الاستجابة ذو رتبة (n * 1) للتكرار m

X: تمثل مصفوفة المتغيرات التوضيحية ذو رتبة (n * P + 1)

$V^{(m)}$: مصفوفة مربعة للتباينات عناصر قطرها الرئيسي $[\pi_i{}^m(1-\pi_i{}^m)]$ مكتسبة من التكرار (m)

وكل حل يحدد نقطة حرجة اما عظمى او صغرى وتكون النقطة الحرجة عظمى عندما تكون مصفوفة المشتقات الجزئية من الدرجة الثانية سالبة اي ان كل عنصر من عناصر القطر اصغر من الصفر ومن خصائصها انها تشكل مصفوفة التباين والتباين المشترك لتقديرات المعلمة بإيجاد المشتقة الجزئية الثانية لكل $(p+1)$ من المعادلات في المعادلة (9) وكما يلي:

$$\acute{\acute{L}}\left(\underline{\beta}\right) = \frac{\partial^2 L(\beta)}{\partial\beta_j \partial\acute{\beta}_j} = \frac{\partial}{\partial\acute{\beta}_j}\left[\sum_{i=1}^{n}\left(Y_i\acute{X}_{ij} - \pi(X_i)\acute{X}_{ij}\right)\right]$$

$$= \frac{\partial}{\partial\acute{\beta}_j}\left[\sum_{i=1}^{n} -\pi(X_i)\acute{X}_{ij}\right]$$

$$= -\sum_{i=1}^{n}\acute{X}_{ij}\,\frac{\partial}{\partial\acute{\beta}_j}\left(\frac{e^{\acute{X}_i\underline{\beta}}}{1+e^{\acute{X}_i\underline{\beta}}}\right)$$

$$= \frac{e^{\acute{X}_i\underline{\beta}}\,\dfrac{\partial}{\partial\acute{\beta}_j}\left(\acute{X}_i\underline{\beta}\right)}{\left(1+e^{\acute{X}_i\underline{\beta}}\right)^2}$$

$$= \pi(X_i)\,[1-\pi(X_i)]X_{ij}$$

وبالتعويض نحصل على المشتقة الجزئية من الدرجة الثانية وكالآتي.

$$\acute{L}\left(\underline{\beta}\right) = -\sum_{i=1}^{n} \acute{X}_{ij} \left[\pi(X_i)\left(1 - \pi(X_i)\right)\right] X_{ij} \qquad \ldots(13)$$

ونستمر بتطبيق الصيغة (10) حتى يتحقق معيار التوقف وهو تجانس المعالم المقدرة اي ان التقديرات تبدأ بالأستقرار اي يبدأ التغيير في المقدرات يتلاشى واحياناً يمكن ان تتوقف دون الوصول الى النقاط العالمية ومن فرضيات هذه الخوارزمية هو تحديد نقاط البداية بالعملية التكرارية.

5-4 الاختبارات المتعلقة بأنموذج الانحدار (Log-Logistic) الثنائي

(Tests related to the binary logistic regression model)

هناك العديد من الاختبارات التي يتم فيها تقيم جودة التوفيق ومعيار التقييم لأنموذج الانحدار اللوجستي وكذلك حسن المطابقة للانموذج ومن هذه الاختبارات اختبار والد (*wald*) واختبار معامل التحديد (R^2) واختبار نسبة الامكان الاعظم واختبار نسبة مربع كاي كذلك واختبار هوزمر__ليمشو (*Hosmer- Lemeshow*) وغيرها من المعاير الاخرى . وذلك لكي يتم معرفة الى أي مدى يتناسب انموذج المقدر مع البيانات، وعليه سوف نختصر على بعض الاختبارات التالية:

1-5-4 اختبار هوزمر - ليمشو لجودة المطابقة (Hosmer- Lemeshow Quality of Conformance Test)[4]

يقوم هذا الاختبار بتجميع حالات العينة بناء على قيم الاحتمالات المتوقعة حيث يعتبر اختبار هوزمر-ليمشو أحد الاختبارات التي تستخدم في جوده التوفيق لأنموذج الانحدار اللوجستي الوغارتيمي. ويعتمد هذا الاختبار على مدى قرب الاحتمالات المتوقعة والاحتمالات المشاهدة. حيث يعمل على أساس تجميع حالات قيم الاحتمالات المتوقعة ، لذا يعد هذا الاختبار مشابه لحد ما لاختبار ($\mathcal{X}^2$) لحسن المطابقة حيث يستخدم لتقيم حسن المطابقة للأنموذج وهو يسمح بأي عدد من المتغيرات التوضيحية سواء كانت متقطعة أو مستمرة، وإنَّ هذه الاحصاءة تتوزع توزيع ($\mathcal{X}^2$) بدرجة حرية (*c-1*)(*r*-1)، واذا كانت قيمتها أكبر من مستوى المعنوية فإنَّ هذا يؤكد جودة التوفيق للأنموذج بالكامل، حيث يقوم هذا الاختبار على أساس تقسيم الحالات المدروسة الى عشر مجموعات على شكل اعمده ،ما صفوف تقسم على أساس القيم المشاهدة لمتغير المعتمد وهما الصفر والواحد. تكون الفرضية الإحصائية للانموذج كالاتي:

H_0: الأنموذج المقدر يوافق البيانات بشكل جيد.

H_1: الأنموذج المقدر لا يوافق البيانات بشكل جيد.

ويمكن ترميز احصاءة الاختبار هذه بالرمز Y ويتم احتسابها وفقا للصيغة الأتية: -

$$Y = \sum_{i=1}^{k} \frac{\left(O_i - n_i'\overline{P_i}\right)^2}{n_i'\overline{P_i}\left(1-\overline{P_i}\right)} \quad \ldots\ldots\ldots (14)$$

إذ إنَّ:

$O_i = \sum_{j=1}^{n_i'} Y_j$: تمثل القيم المشاهدة للمجموعة

$\overline{P_i} = \sum_{j=1}^{n_i'} \frac{p_i}{n_i'}$: تمثل القيم المتوقعة للمجموعة

4-5- 2 اختبار والد (Wald Test)[7]:

يقيم اختبار *Wald* (الذي سمي على اسم ابراهام والد) القيود المفروضة على معلمات الإحصائية بناء على المسافة الموزونة بين التقدير غير المقيد وقيمته المفترضة في ظل الفرضية الصفرية والتي تنص على (إنَّ تأثير معامل لوجت ما يساوي صفرا)، ويكون ذات اهمية ومعنوية للمتغيرات والثوابت التوضيحية وتأثيرها على متغير الاستجابة في النماذج اللوجستية ، لذا يعد أحد الأساليب الكلاسيكية (الثلاثة)لاختبار الفرضيات، وميزة هذا الاختبار عن الاختبارات الأخرى هو أنه لا يتطلب سوى تقدير أنموذج غير المقيد، وإنَّ الفرضية الإحصائية للانموذج كالاتي:

H_0: تأثير معامل لوجيت يساوي صفر.

H_1:تاثير معامل لوجيت لا يساوي صفر.

ويمكن حساب احصاءه الاختبار حسب الصيغة الأتية

$$\text{wald} = \frac{b}{SE_b} \quad \ldots\ldots\ldots (15)$$

حيث إنَّ:

B: هي قيمة معامل الانحدار اللوجستي للمتغير المستقل.

SEb: هي قيمة الخطأ المعياري لمعامل الانحدار اللوجستي للمتغير المستقل.

علما إنَّ الاحصاءه *Wald* تتبع توزيع $(\mathcal{X}^2)$ بدرجة حرية $(c\text{-}1)(r\text{-}1)$، كما إنَّ هذا الاختبار هو اختبار من الطرفين ويجب إنَّ تكون قيمة معنوية المعلمات المناظرة لقبول أو رفض فرضية العدم باستعمال الاحتمالات التي تكون أقل من (*0.05*) لكي يتم رفض الفرضية وقبول الفرضية البديلة والتي تنص على (إنَّ تأثير المعامل لوجت لا يساوي صفر)، أي إنَّ المتغير المستقل له تأثير بالتنبؤ بقيمة المتغير التابع. ويرمز لإحصائه اختبار والد بالحرف (W) وتكون عبارة عن مربع اختبار (T) وتحسب حسب الصيغة الاتية:

$$W = T^2 = \left[\frac{\hat{\beta}_j}{S.E\,(\hat{\beta}_j)}\right]^2 \quad \dots(16)$$

حيث ان:

$\hat{\beta}_j$: تقدير معلمة الانحدار اللوجستي للمتغير التوضيحي.

$S.E\,(\hat{\beta}_j)$: تقدير الخطأ المعياري لمعلمة الانحدار اللوجستي للمتغير التوضيحي.

إذا كانت قيمة الاختبار الاحتمالية $(P - value)$ اقل من (0.05) نرفض فرضية العدم وهنا يعني ان معاملات المتغير التوضيحي معنوية ذات دلالة احصائية.

وفي اختبار $(wald)$ تستعمل احصاءه (Z) وهي عبارة عن الجذر التربيعي لاختبار والد وتتوزع توزيعاً طبيعياً قياسياً حسب الصيغة الاتية:

$$Z = \frac{\hat{\beta}_j}{S.E\,(\hat{\beta}_j)} \quad , \; j = 1,2,\dots.p \quad , \quad Z \sim N(0,1) \quad \dots(17)$$

وتقارن قيمة الاحصاءه (Z) مع القيم الجدولية $Z_{\left(1-\frac{\alpha}{2}\right)}$ ، $Z_{\left(\frac{\alpha}{2}\right)}$

فتقبل فرضية العدم في حالة وقوع قيمة (Z) بينهما وبمستوى معنوية (α).

6-4 معيار المقارنة بين طرائق التقدير المستخدمة: - (Comparison Criterion Between Used Estimation Methods)

في تحليل الانحدار يعد التخطيط طريقة طبيعية أكثر لعرض الاتجاه العام للبيانات بأكلها يمكن حساب متوسط المسافة من كل نقطة الى انموذج الانحدار المتوقع واظهاره على انه متوسط الخطأ التربيعي ، أحد الامثلة على الانحدار الخطي باستخدام MSE) طريقة المربعات الصغرى التي تبين وتقيم مدى ملاءمة أنموذج الانحدار الخطي لنمذجة مجموعة البيانات ثنائية المتغير، فيتم الاعتماد على مقياس متوسط مربعات الخطأ $Mean\ Square\ Error\ (MSE)$ لكل محاولات (تكرارات) الظاهرة قيد الدراسة مقسوما على عدد تلك المحاولات كمقياس مقارنة بين طرائق التقدير المستعملة يشير الى مدى دقة التقدير وان تناقص قيمته تشير الى جودة ودقة المقدرات والهدف الذي يراد تحقيقه هنا هو الحصول على المقدر $[(\widehat{Y}_i) = \hat{\pi}(X_i)]$

في الصيغة (18) $\pi(X_i) = \frac{e^{\square}}{1+e^{\square}}$

ويتم احتسابه كالآتي:

$$MSE = \frac{1}{n-p}\sum_{i=1}^{N}(Y_i - \widehat{Y}_i)^2 \quad \dots(19)$$

اذ ان:

$\widehat{Y}_i$: تمثل القيمة التقديرية للمشاهدات.

n: تمثل حجم العينة.

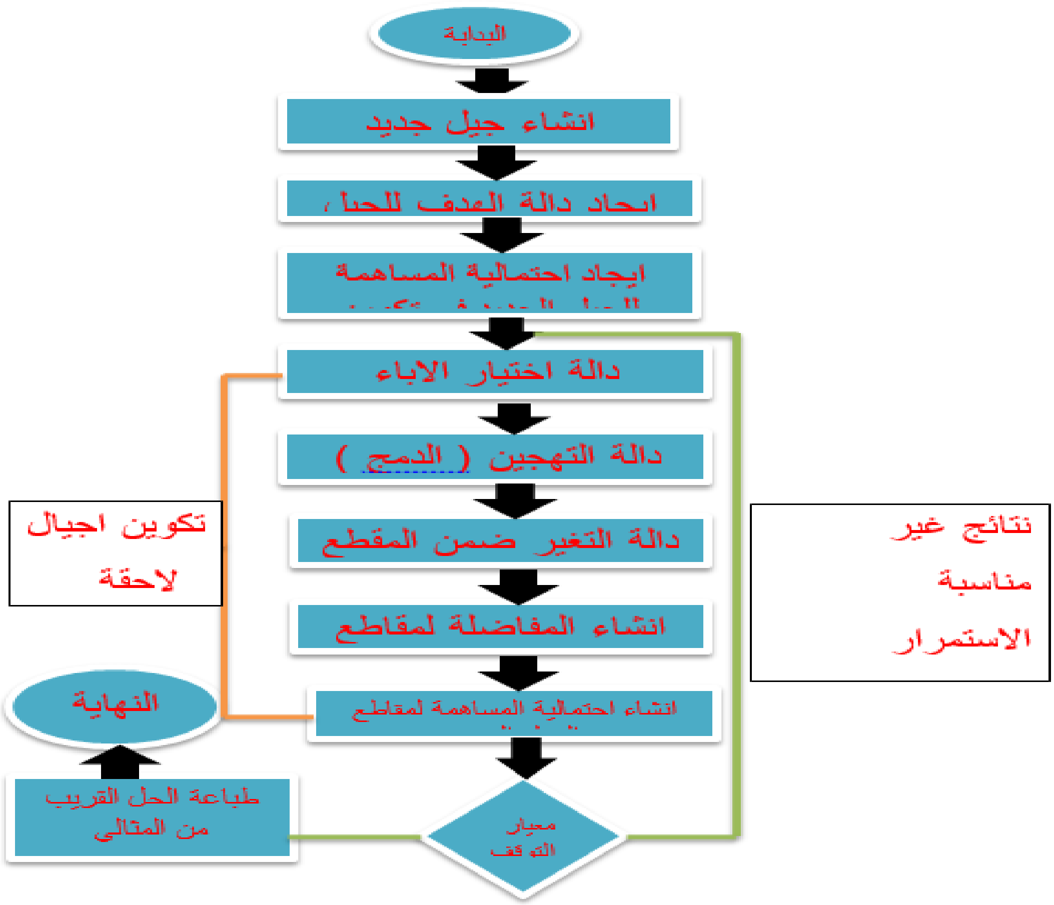
p: تمثل عدد المعلمات

٤-٧ مفهوم الخوارزمية الجينية (Genetic algorithm concept)

٤-٧-١ الخوارزمية (algorithm) [١،٣]:

وتعني اصطلاحاً؛ مجموعة خطوات، وقواعد رياضية ، يتم برمجة الحواسيب عليها لتتعامل مع المسائل المختلفة، وايجاد الحلول المناسبة لها.

وتستعمل الخوارزمية الجينية الوراثية تقنية بحث لايجاد حلول مضبوطة او تقريبية تحقق الامثلية، تصنف الخوارزميات الوراثية على انها من طرق البحث الشامل الاستدلالي (**Global search heuristics**). لذلك تكمن فلسفة الخوارزمية الجينية بشكل عام على توليد عدد كبير من الحلول المتوفرة لمشكلة معينة ومن ثم يتم تقييم كل حل من هذه الحلول وصولاً الى الحل الافضل وتكون فرصته اكبر لتوليد حل اخر في تحسين نقل فرصة الحل السيء وبتكرار هذه العملية تتطور نوعية الحلول الموجودة وتصل او تقترب من الحل الامثل وتطبيقها بالشكل الصحيح تكون فعالة في حل المشكلات المعقدة التي تعجز الطرائق الاخرى على حلها، ونظراً للافتراضات المقيدة الخاصة بخوارزمية (**NR**) سوف ندرس كفاءة خوارزمية اخرى كالخوارزمية الجينية (**GA**) لتقدير المعلمة ويمكن ان ينظر لها على انها بديل جيد ينفذ بشكل متكرر في تحسين الدوال غير الخطية والخوارزمية الجينية تفعل ذلك بطريقة سريعة جداً وقريبة من الحل الدقيق.



الشكل (2) المخطط العام للخوارزمية الجينية من

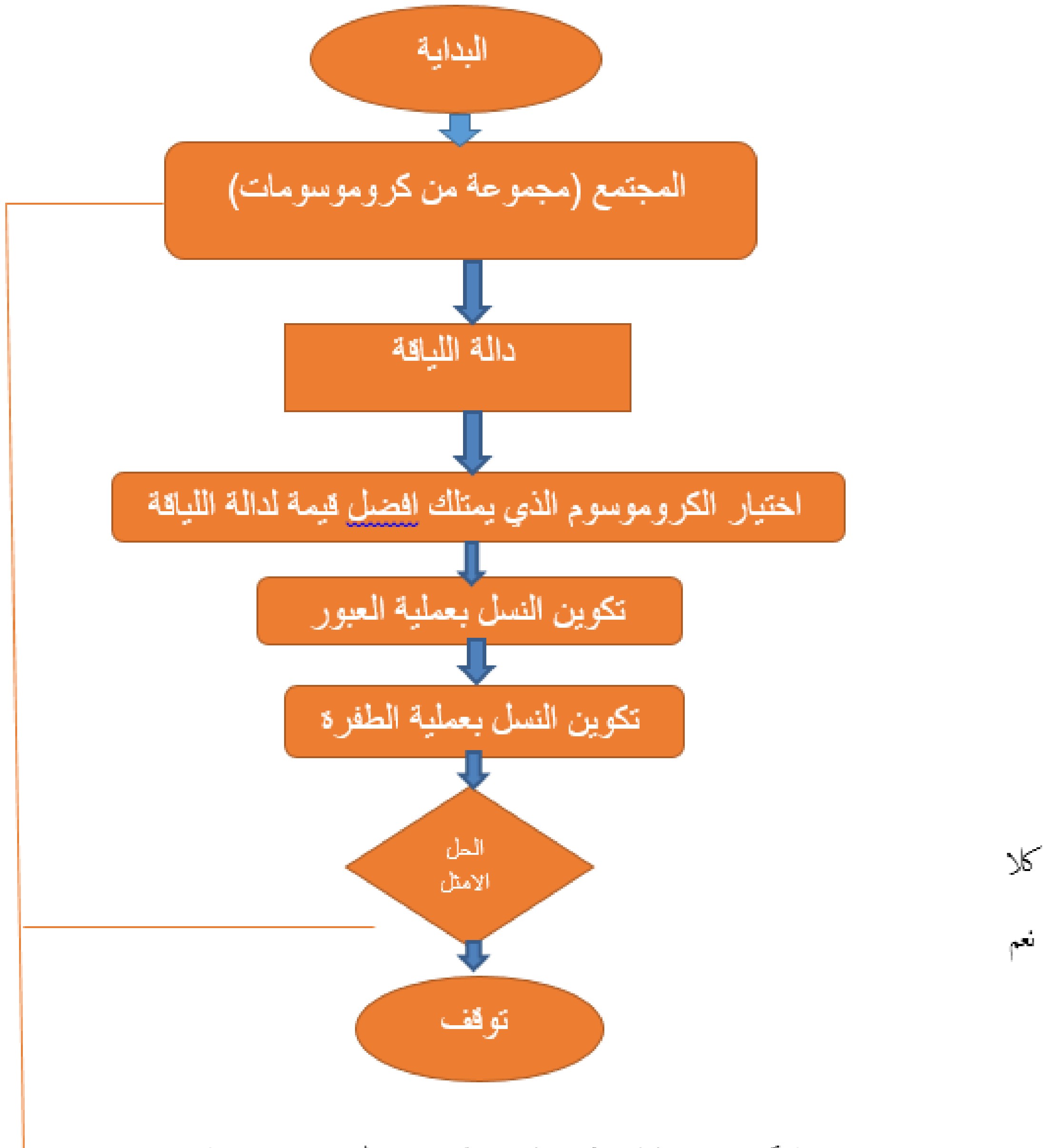
الخطوات التي تعمل بها الخوارزمية الجينية:

1- يتم تقييم الابناء الجدد بالاعتماد على الدالة الأصلية.

2- تتغير الكروموسومات الاصلية بالاعتماد على تقييم الابناء وان كل كروموسوم يتكون من عدد من القيم ويتم تحديد عدد هذه القيم حسب المسألة وقيمة البداية.

3- يتم اختيار أفضل الأباء لكي تتم عملية انتاج الابناء.

4- يتم توليد جيل جديد باستخدام الطفرة والعبور.



الشكل (3) مخطط يوضح خطوات الخوارزمية الجينية من تصميم الباحث

5-تجربة المحاكاة:

في هذا الجزء من البحث سيتم توليد عشرة متغيرات توضيحية وذلك عن طريق استخدام اسلوب مونت كارلو وذلك من خلال التوزيع المنتظم (Uniform Distribution) أما في ما يتعلق بتوزيع الخطأ فأنه يتبع توزيع برنولي (Bernoulli Distribution) وكذلك في هذه المرحلة يتم توليد قيم المتغيرات التوضيحية وقيم متغير الاستجابة، سوف نستعمل طريقة الرفض والقبول لاحتساب متغيرات ثنائية الاستجابة وتحديد القيم الاحتمالية وفق الاتي:

$$Y = \begin{Bmatrix} 1 \; if \; \pi(X_i) \geq 0.5 \\ 0 \; if \; \pi(X_i) < 0.5 \end{Bmatrix}$$

5-1التقديرات

تم أخذنا (Standardized) للمعلمات في الانموذج الثاني وفق البرنامج الجاهز (SPSS) وكما مبين في الجدول رقم 1

Para \ Mod.	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}
1	0.816	0.95	-0.073	-0.002	0.013	0.004	-1.031	-0.085	-0.0415	0.001	-0.002
2	-0.046	0.091	-0.088	-1.031	-0.803	-0.002	0.150	0.092	-0.004	0.060	-0.013

جدول رقم (1) طريقة تقدير المعلمات وللانموذجين الاول والثاني

في هذه المرحلة يتم تقدير معلمات انموذج الانحدار **(Log-Logistic)** الثنائي المعطى في المعادلة (10) وفق الطريقة الاعتيادية، وأيضا وفق توظيف الخوارزمية الجينية مع طريقة التقدير الاعتيادية التي تم ذكرها في الجانب النظري (MLE) للبحث. (طريقة الخوارزميـة الجينيـة بالاعتماد على تقنية تقديــرات الامكـان الاعظم (MLE.GA).

5-2المقارنة بين طرائق التقدير الاعتيادية والحديثة

وهي المرحلة الأخيرة من مراحل وصف تجربة المحاكاة وفي هذه المرحلة تم المقارنة بين طريقة التقدير بعد أن تم إيجاد تقديرات المعلمات باستخدام المقاييس الاحصائية المستخدمة لغرض الحصول على أفضل طريقة تقدير للانموذج قيد الدراسة ، لذلك سوف يتم المقارنة بين طريقة تقدير معلمات أنموذج الانحدار **(Log-Logistic)** الثنائي والتي تكون على اساس (الاعتيادية والمحسنة ونبين من هي الافضل) لطريقة المستخدمة وذلك عن طريق استعمال احد المعايير الاحصائية المهمة وهو متوسط مربعات الخطأ للأنموذج المدروس حسب المعادلة (19) الاتية:

$$MSE = \frac{1}{R}\sum_{i=1}^{R} MSE_i = \frac{1}{R}\sum_{i=1}^{R}\left[\frac{1}{N-P}\sum_{i=1}^{N}\left(Y_i - \widehat{Y_i}\right)^2\right]$$

حيث (**R**) تمثل عدد مرات تكرار التجربة، وقد تم تكرار تجربة المحاكاة (R=1000)وذلك بهدف الحصول على أفضل نتائج.

5-3تحليل نتائج المحاكاة (Analyze the Simulation Results)

سوف يتم عرض نتائج عملية المحاكاة ومن ثم تحليلها للوصول الى ان طريقة MLE كطريقة فضلى لتقدير معلمات أنموذج الانحدار **(Log-Logistic)** الثنائي بالاعتماد على المقياس الاحصائي متوسط مربعات الخطأ (MSE) لمقدرات الأنموذج.

سوف نقوم بعرض نتائج المحاكاة التي تم الحصول عليها من خلال برنامج (MATLAB) وفيما يلي النتائج التي سيتم تحليلها حسب تسلسل الجداول الاتية:

جدول (٢) متوسط مربعات الخطأ (MSE) لمقدرات الانموذج الانحدار (Log–Logistic) الثنائي في الانموذج(١) للمعلمات باستعمال طريقة MLE الاعتيادية والجينية وكافة احجام العينات.

sizes Sample	Methods		Best
	Classic MLE	Genetic MLE	
N = 30	0.231	0.0057	Genetic
N = 60	0.1952	0.0024	Genetic
N = 90	0.0101	0.0023	Genetic

جدول (٣) متوسط مربعات الخطأ (MSE) لمقدرات الانموذج الانحدار(Log–Logistic) الثنائي في الانموذج (٢) للمعلمات باستعمال طريقة الاعتيادية والجينية وكافة احجام العينات.

sizes Sample	Methods		Best
	Classic MLE	Genetic MLE	
N = 30	0.1351	0.0056	Genetic
N = 60	0.1839	0.0023	Genetic
N = 90	0.1029	0.0022	Genetic

نلاحظ من الجداول اعلاه ان طريقة الامكان الاعظم المحسنة (MLE.GA) بأستعمال الخوارزمية الجينية أفضل من طريقة الامكان الاعظم الاعتيادية (MLE) بأستعمال خوارزمية نيوتن رافسون

الاستنتاجات والتوصيات

عند احجام العينة (n = 30 , 60 ,90) بالنسبة للقيم الافتراضية للمعلمات والانموذج الاول والثاني في حالة الطريقة الاعتيادية نلاحظ (MLE) عند توظيف واستعمال الخوارزمية الجينية في التحسين في هذه الطرائق وللنماذج نفسها ان طريقة الامكان الاعظم عند التحسين هي الافضل لامتلاكها اصغر متوسط مربعات الخطأ من متوسط مربعات الخطأ للطريقة الاعتيادية وعليه فان مقدرات الانموذج (log-logistic) الثنائي تحسن اداء الطرائق عند استعمال (Genetic Algorithm) وذلك من خلال ايجاد افضل مقدر للمعلمة المجهولة في انموذج الانحدار (log-logistic) الثنائي.

نوصي باستعمال اكثر من طريقة تقدير اعتيادية وتحصنها باستعمال الخوارزمية الجينية ونوصي بتطبيق النتائج على بيانات حقيقية .

المصادر

1. Akkus ,Ö. , Demir , E. , (2016), " Comparison Som Classical And Meta-Heuristic Optimazation Techniques in The Estimation Of The Logit Model Parameters", IJAR, pp.1026-1042.
2. Chatterjee, S., Hadi, A., (2012) ,"Regression Analysis By Example", John Wiley ,INC. "Zuhair A. Al-Hemyari Abstract." 2009.
3. Demir , E. , Akkus , Ö., (2015), " An Introductory Study on How the Genetic Algorithm Works in the Parameter Estimation of Binary Logit Model", IJS:BAR, pp.162-180.
4. Hosmer, D., Lemeshow, S. & Sturdivant , R. ,(2013)," Applied Logistic Regression", 3 rd edition ,New York: willey,WSIPS, http : // ihmsi.org.
5. Menard ,S.,(2002),"Applied Logistic Regression Analysis",2nd Edition Thousand Oaks , CA : SAGE Publications , Series Quantitative Applications in the Social Sciences.
6. Rodriguez ,G.,(2007), "Logit Models for Binary Data" ,Chapter(3) ,Retrieved from,http://data.princeton.edu/wws509/notes/c3.pdf
7. Wuensch, K. , (2014) , " Binary Logistic Regression with SPSS" , Retrieved from WWW . Care . ecu . edu / psyc/ Wuensch / MV /Logistic SPSS, pp.1-29.