

## Comparison of some methods for estimating the parameters of the binary logistic regression model using the genetic algorithm with practical application

مقارنة بعض الطرائق لتقدير معالمات انموذج الانحدار اللوجستي الثنائي باستعمال الخوارزمية الجينية مع تطبيق عملي

أ.م.د. رباب عبد الرضا صالح البكري جامعة بغداد /كلية الادارة والاقتصاد /mah\_2008@yahoo.com  
الباحث / ساره عادل مظلوم الرديني /sarahstatistic88@gmail.com

OPEN ACCESS

P - ISSN 2518 - 5764  
E - ISSN 2227 - 703X

Received:4/12/2018

Accepted :7/1/2018

### المستخلص

يعاني الانسان بسبب ضغوطات الحياة الطبيعية من تعرضه الى عدة انواع من امراض القلب وذلك نتيجة لعوامل مختلفة، وبهدف معرفة حالة حدوث الوفاة من عدمه يتم نمذجتها باستعمال أنموذج الانحدار اللوجستي الثنائي، لذا تم في هذا البحث استعمال أحد أهم نماذج الانحدار غير الخطية الواسعة الاستعمال في نمذجة التطبيقات الاحصائية، من حيث الإصابة بأمراض القلب وهو انموذج الانحدار اللوجستي الثنائي. ومن ثم تقدير معالمات هذا الأنموذج باستعمال طرائق التقدير الاحصائية ولكن اثناء استعمال هذا الأنموذج تواجهنا مشكلة في تقدير معالمته وذلك عندما يكون عدد المعالمات  $(P + 1)$ ، وان ايجاد تقدير المعالمات باستعمال الطرائق العددية احيانا لا تعطي الحل الامثل لأنها تعتمد على المقدرات البدائية، باستعمال بعض الطرائق الاعتيادية بعد تحسينها من خلال اتباع منهجية الخوارزمية الجينية في التقدير لتلائم تقدير معالمات هذا النوع من نماذج الانحدار غير الخطية، ومن ثم المقارنة بين طرائق التقدير، وقد شملت المقارنة نوعين من طرائق التقدير المهمة وهي طرائق التقدير الاعتيادية التي تضمنت طريقة الامكان الاعظم، وطريقة تصغير مربع كاي، وطرائق التقدير المحسنة التي تم تطويرها من الباحثة والتي تضمنت طريقة الخوارزمية الجينية بالاعتماد على تقنية تقديرات الامكان الاعظم MLE، وطريقة الخوارزمية الجينية بالاعتماد على تقديرات تصغير مربع كاي MCSE، من أجل اختيار الطريقة الأفضل في التقدير وذلك من خلال القيم الافتراضية لتقدير معلمة انموذج الانحدار الخطي المتعدد بطريقة المربعات الصغرى الاعتيادية ols وكذلك تقدير المعلمة بتحويل القيم الحقيقية الى القياسية Standardized وبأحجام عينات مختلفة خلال المحاكاة وباستعمال المعيار الاحصائي متوسط مربعات الخطأ (MSE) لمقدرات الانموذج اللوجستي لغرض المقارنة بين أفضلية طرائق تقدير معالمات الأنموذج، وقد تم التوصل بشكل عام الى أن طريقة (Mcs) هي الأفضل بالمرتبة الاولى من بين طرائق التقدير الاعتيادية، وطريقة (Mcs. GA) هي الأفضل من بين طرائق التقدير المحسنة لغرض تقدير المعالمات للأنموذج اللوجستي الثنائي وذلك لأنها تمتلك اقل (MSE) للمقدرات، وقد تم في الجانب التطبيقي استعمال هذا الأنموذج لنمذجة البيانات الخاصة بالمصابين بأمراض القلب وتقدير المعالمات باستعمال طريقة (Mcs. GA)، وتم التوصل فيه من خلال مقارنة اسباب حالات حدوث الوفاة الحقيقية مع اسباب حالات حدوث الوفاة المقدرة الى مدى ملائمة الأنموذج في نمذجة هذا النوع من البيانات واستخلاص السبب الرئيسي لحدوث الوفاة هو التدخين، وكذلك دقة الطريقة (Mcs. GA) في تقدير معالمات الأنموذج .

**المصطلحات الرئيسية للبحث /** تصغير مربع كاي، انموذج الانحدار اللوجستي الثنائي، المعالمات، الخوارزمية الجينية، خوارزمية نيوتن\_ رافسون.



## 1-1 المقدمة

### (Introduction)

ترتبط قيمة النتائج العلمية المستخرجة بشكل مباشر بمدى دقة الأساليب والتجارب المستخدمة في موضوع البحث وصحته، ومن أجل صياغة تلك الأساليب والتجارب في صورة يمكننا من الحصول على نتائج أكثر وضوح وذات دقة عالية مقبولة ومرضية للتوصية منها، يفضل استخدام الأساليب الإحصائية المناسبة والأكثر دقة وعلى وفق أسس علمية تكون حصيلتها اتخاذ القرار العلمي بقناعة علمية وثقة كافية.

يعتبر نموذج الانحدار اللوجستي الثنائي من النماذج اللاخطية التي تصف العلاقة بين متغير تابع ثنائي القيمة أي يأخذ قيمتين هما الصفر لاحتفال عدم حدوث حدث معين والواحد الصحيح لاحتفال حدوث ذلك الحدث والمتغيرات المستقلة تأخذ قيم وصفية أو كمية، ولذلك يستعمل استراتيجية تحويل الاستجابات على فئتي أو فئات المتغير التابع الى وحدات لوغاريتمية من نوع لوجت وحينئذ فقط تتحول العلاقات غير الخطية الى علاقات خطية بين متغير الاستجابة (التابع) والمتغيرات التوضيحية (المستقلة)، ومن اهم الطرائق التي تستخدم في تقدير معالمه هي طريقة تقدير الامكان الاعظم (MLE) ويتم الحصول على معادلات الامكان بأخذ المشتقات من الدرجة الاولى لمعاملات دالة الامكان في النموذج اللوجستي الثنائي مما يزيد من احتمالية مشاهدة الفرد وتكون غير خطية وبالتالي لا يمكن الحصول تحليليا على القيم المثلى للمعاملات فنستخدم اساليب تكرارية مثل نيوتن رافسون أو غاوس نيوتن وغيرها من الأساليب الأخرى ضمن فئة تقنيات التحسين الكلاسيكية وقد استخدمنا خوارزمية نيوتن رافسون (NR) كطريقة لحل مشكلة الغير خطية واستخدمنا طريقة تقدير تصغير مربع كاي (MCSE) من اجل الحصول على قيم المعلمة المثلى (افضل نتيجة تقدير) من خلال تقليل مجموع مربعات الخطأ العشوائي باستخراج المشتقة الاولى ومساواتها للصفر فنستنتج بوجود علاقة غير خطية فنستخدم خوارزمية نيوتن رافسون لحلها ومن شروطها ان تكون دالة الهدف مستمرة ويجب تحديد نقاط البداية المناسبة للمعاملات ونظراً لما تتمتع بها خوارزمية نيوتن رافسون من القيود والافتراضات بتحديد لها نقطة البداية وقد تكون هذه النقطة غير جيدة فلانستطيع ايجاد القيم المثلى فكانت هناك حاجة لتطوير طريقة بديلة، لحل هذه المشاكل رغم القيود المفروضة فنستخدم منهجية الخوارزمية الجينية (GA) وهي طريقة محسنة للحل لاحتوائها على صفات وخصائص مهمة مثلاً ليس من الضروري وجود دالة هدف تفاضلية وتشكيل افضل مجموعة حل للمعلمة التي من شأنها تقليل مقدار الخطأ الى اصغر ما يمكن.

ومن الواضح صعوبة الحصول على افضل تقدير لمعاملات نموذج الانحدار اللوجستي الثنائي بالطرائق الاعتيادية عندما يكون عدد المعاملات  $(P + 1)$  اذ ظهرت الحاجة لتطويرها، في اقتراح وتوظيف الطرائق الاعتيادية بعد تحسينها باتباع الخوارزمية الجينية في التقدير لغرض تقدير معالم هذا النموذج.

ومن ثم تتم المقارنة بين الطرائق التي ذكرناها في اعلاه باستخدام تقنية التحسين (GA) على نتائج نموذج الانحدار اللوجستي الثنائي فنستخدم لذلك برنامج MATLAB ومحاكاة مونت كارلو مع تفسيراتهم بالتفصيل وبعد المقارنة نحصل على التوصيات من النتائج التي توصلنا اليها من الطرائق المستخدمة في التقدير.

## 2-1 هدف البحث

### (Research Aim)

ان الهدف من هذا البحث هو توظيف الخوارزمية الجينية في تقدير معالم أنموذج الانحدار اللوجستي الثنائي بعد تحسينها للحل من خلال استعمال طرائق التقدير الاعتيادية وهي طريقة تقدير الامكان الاعظم (MLE) وطريقة تقدير تصغير مربع كاي (MCS) من اجل الوصول الى أفضل الطرائق في التقدير، ومن ثم المقارنة بين هذه الطرائق من خلال القيم الاولية للمعلمة، مرة بطريقة المربعات الصغرى الاعتيادية ols واخرى بالطريقة القياسية Standardized خلال المحاكاة، وباستعمال المعيار الاحصائي متوسط مربعات الخطأ (MSE) لغرض المقارنة والتوصل الى الطريقة الأفضل، ومن ثم بناء أنموذج الانحدار اللوجستي في الجانب التطبيقي لبيان اهم العوامل المؤثرة على امراض القلب وتقدير معالمه باستعمال افضل الطرائق التي تم التوصل اليها في الجانب التجريبي.

Binary Logistic (BLRM)

3-1 نموذج الانحدار اللوجستي الثنائي

(Regression Model)

يعد هذا النموذج من نماذج الانحدار اللاخطي ويتصف بان متغير الاستجابة (Y) يتبع توزيع برنولي (Bernoulli distribution) يأخذ القيم (0) و (1) [10;2013;p.8] اي ان متغير الاستجابة الفئوي (Y) له حالتين تتمثل الحالة الاولى وقوع حدث معين عندما (Y = 1) والحالة الثانية بعدم وقوع ذلك الحدث عندما (Y = 0) باحتمال وقوع الحدث (النجاح) هو  $[\pi(X_i)]$  اعتماداً على قيم المتغيرات التوضيحية للمشاهدات واحتمال عدم وقوع الحدث (الفشل) هو  $[1 - \pi(X_i)]$  وبذلك تكون دالة الكثافة الاحتمالية بالصيغة الاتية [20;2007;p.3]:

$$P(Y_i/X_i) = [\pi(X_i)]^{Y_i} [1 - \pi(X_i)]^{1-Y_i} ; Y_i = 0,1 \quad \dots (1)$$

اما اذا كان نموذج الانحدار اللوجستي يحتوي على اكثر من متغير توضيحي واحد، فيعبر عن توقعه الشرطي لاحتمال متغير الاستجابة (وقوع الحدث) حسب الصيغة الرياضية الاتية [3;2017;p.40]:

$$\pi(X_i) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip})}} \quad \dots (2)$$

$$= \frac{e^{(\beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip})}}{1 + e^{(\beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip})}} \quad \dots (3)$$

وان احتمال عدم وقوع الحدث هو:

$$1 - \pi(X_i) = \frac{1}{1 + e^{(\beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip})}} \quad \dots (4)$$

يحيث ان:

$$j = 0, 1, \dots, p$$

$$; i = 1, 2, \dots, n$$

$X_1, X_2, \dots, X_p$ : المتغيرات التوضيحية التي تكون المصفوفة X وعددها (P).

$\beta_0, \beta_1, \dots, \beta_p$ : متجه للمعاملات المراد تقديرها وعددها (P).

ويتم تحويل هذا النموذج الى شكل خطي يتمثل بعلاقة خطية مع لوجت الاحتمال  $[\text{logit } \pi(X_i)]$  وحسب الصيغة الرياضية الاتية [1;2002;p.166]:

$$Z = \text{logit } \pi(X_i) = \ln \left[ \frac{\pi(X_i)}{1 - \pi(X_i)} \right] \quad \dots (5)$$

$$= \beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip} \quad \dots (6)$$

$$= [1 \quad X_{i1} \quad \dots \quad X_{ip}] \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} = \underline{X'_i} \underline{\beta} \quad \dots (7)$$

يتم تعويض المعادلة (7) في المعادلتان (3) و (4) وكالاتي [4;2003;p.12]:

$$\pi(X_i) = \frac{e^{\underline{X'_i} \underline{\beta}}}{1 + e^{\underline{X'_i} \underline{\beta}}} \quad \dots (8)$$

$$1 - \pi(X_i) = \frac{1}{1 + e^{\underline{X'_i} \underline{\beta}}} \quad \dots (9)$$

#### 4-1 طريقة الامكان الاعظم (Maximum Likelihood Method (MLM))

للحصول على تقديرات المعلمة بواسطة الامكان الاعظم (MLE) يتم ضرب الحدود في المعادلة (1) لعينة حجمها (N) تعتمد على (X) من المتغيرات التوضيحية (التفسيرية) ومجموعة متغير الاستجابة (Y)، وبالتالي ان دالة الامكان الاعظم في النموذج اللوجستي الثنائي لـ n من المشاهدات هي [10;2013;p.9]:

$$l(\beta) = P(Y/X) = \prod_{i=1}^n [\pi(X_i)]^{Y_i} [1 - \pi(X_i)]^{1-Y_i} \quad \dots (10)$$

وبأخذ اللوغارتم (Log) على الجانبين للحصول (MLE) لتسهيل عملية الحل

$$\ln[l(\beta)] = \sum_{i=1}^n Y_i \ln[\pi(X_i)] + (1 - Y_i) \ln[1 - \pi(X_i)] \quad \dots (11)$$

وبالتعويض عن  $\pi(X_i)$  و  $1 - \pi(X_i)$  بما يساويهم حسب المعادلتان (8) و (9) وكالاتي:

$$\ln[l(\beta)] = \sum_{i=1}^n \left[ Y_i \ln \left( \frac{e^{\underline{X}_i \underline{\beta}}}{1 + e^{\underline{X}_i \underline{\beta}}} \right) + (1 - Y_i) \ln \left( 1 - \frac{e^{\underline{X}_i \underline{\beta}}}{1 + e^{\underline{X}_i \underline{\beta}}} \right) \right] \quad \dots (12)$$

$$= \sum_{i=1}^n \left[ Y_i \ln(e^{\underline{X}_i \underline{\beta}}) - Y_i \ln(1 + e^{\underline{X}_i \underline{\beta}}) + \ln \left( \frac{1}{1 + e^{\underline{X}_i \underline{\beta}}} \right) - Y_i \ln \left( \frac{1}{1 + e^{\underline{X}_i \underline{\beta}}} \right) \right] \quad \dots (13)$$

$$= \sum_{i=1}^n \left[ Y_i (\underline{X}_i \underline{\beta}) - \ln(1 + e^{\underline{X}_i \underline{\beta}}) \right] \quad \dots (14)$$

ومن اجل الحصول على تقديرات المعلمات وهي ( $\hat{\beta}$ ) لتعظيم لوغارتم دالة الامكان  $\ln[l(\beta)] = L(\underline{\beta})$  تؤخذ المشتقات من الدرجة الاولى ومساواة الدالة الناتجة بالصفر

$$\hat{L}(\underline{\beta}) = \sum_{i=1}^n \left[ \left( Y_i - \frac{e^{\underline{X}_i \underline{\beta}}}{1 + e^{\underline{X}_i \underline{\beta}}} \right) \underline{X}_{ij} \right] = \hat{X}(Y - P^{(m)}) = 0 \quad \dots (15)$$

$$\hat{L}(\underline{\beta}) = - \sum_{i=1}^n \underline{X}_{ij} [\pi(X_i) (1 - \pi(X_i))] \underline{X}_{ij} = - \hat{X} V^{(m)} X \quad \dots (16)$$

نستنتج بانه تكونت لدينا (p + 1) من المعادلات غير الخطية ونقوم بحلها باحدى الطرائق التكرارية التقليدية كطريقة نيوتن رافسون (NR)، وان خوارزمية نيوتن رافسون التكرارية لإيجاد قيم ( $\hat{\beta}$ ) التقديرية لدالة الامكان الاعظم في النموذج اللوجستي ستكون في (m + 1) من التكرارات وكالاتي [4;2003;pp.36,42]:

$$\hat{\beta}^{(m+1)} = \hat{\beta}^{(m)} + (\hat{X} V^{(m)} X)^{-1} \hat{X} (Y - P^{(m)}) \quad \dots (17)$$

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad P^{(m)} = \begin{bmatrix} \pi_1^{(m)} \\ \pi_2^{(m)} \\ \vdots \\ \pi_n^{(m)} \end{bmatrix}, \quad X = \begin{bmatrix} \underline{X}_1 \\ \underline{X}_2 \\ \vdots \\ \underline{X}_n \end{bmatrix} = \begin{bmatrix} \underline{X}_{10} \\ \underline{X}_{21} \\ \vdots \\ \underline{X}_{np} \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & \dots & X_{1p} \\ 1 & X_{21} & X_{22} & \dots & X_{2p} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & X_{n1} & X_{n2} & \dots & X_{np} \end{bmatrix}$$

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

$$V^{(m)} = \text{diag} [\pi_i^m (1 - \pi_i^m)] \quad \dots (18)$$

حيث ان:

$\hat{\beta}^{(m)}$  هي تقديرات الامكان الاعظم ذو رتبة (P + 1 \* n) للتكرار (m).

$\hat{\beta}^{(m+1)}$  المقدرات الجديدة بالتكرار اللاحق (m + 1).

$Y$ : يمثل متجه متغير الاستجابة ذو رتبة  $(n * 1)$  لل تكرار  $m$ .  
 $P^{(m)}$ : يمثل القيم الاحتمالية لحدوث متغير الاستجابة ذو رتبة  $(n * 1)$  لل تكرار  $m$ .  
 $X$ : 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 السابق  $(m)$ .

5-1 طريقة تصغير مربع كاي (MCSM) (Minimum Chi - Square Method)  
وهي من الطرائق الشائعة الاستخدام في التقدير وتعتمد على تصغير احصاء مربع كاي لبيرسون المعروفة حسب الصيغة الرياضية الآتية: [11;2015;p.2]

$$x^2 = R(\beta) = \sum_{i=1}^N \frac{(\theta_i - \varepsilon_i)^2}{\varepsilon_i} \quad \dots (19)$$

حيث ان:

$\theta_i$ : تمثل القيمة المشاهدة عند المستوى  $i$

$\varepsilon_i$ : تمثل القيمة المتوقعة عند المستوى  $i$

$R(\beta)$ : تمثل احصاء مربع كاي لبيرسون

وفي حالة النموذج اللوجستي الثنائي الاستجابة فإن: [10;2013;p.136]

$$R(\beta) = \sum_{i=1}^n \frac{(Y_i - \pi_i)^2}{\pi_i} + \frac{[(1 - Y_i) - (1 - \pi_i)]^2}{1 - \pi_i} \quad \dots (20)$$

ويتم اختصارها الى:

$$R(\beta) = \sum_{i=1}^n \frac{(Y_i - \pi_i)^2}{\pi_i (1 - \pi_i)} \quad \dots (21)$$

وبعد تعويض قيمة  $(\pi_i, 1 - \pi_i)$  في المعادلة اعلاه نستنتج:

$$R(\beta) = \sum_{i=1}^n \left[ \left( Y_i - \frac{e^{\beta X_i}}{1 + e^{\beta X_i}} \right)^2 \frac{(1 + e^{\beta X_i})^2}{e^{\beta X_i}} \right] \quad \dots (22)$$

وبعد التبسيط تصبح المعادلة كالاتي:

$$R(\beta) = \sum_{i=1}^n \left[ Y_i^2 e^{-\beta X_i} + (1 - Y_i)^2 e^{\beta X_i} - 2Y_i(1 - Y_i) \right] \quad \dots (23)$$

لإيجاد  $\beta$  التي تعطي اقل  $R(\beta)$  يتطلب ايجاد المشتقة الاولى ومساواتها بالصفر كما يأتي:

$$\frac{\partial R(\beta)}{\partial \beta} = \sum_{i=1}^n X_{ij} \left[ (1 - Y_i)^2 \left( \frac{\pi_i}{1 - \pi_i} \right) - (Y_i)^2 \left( \frac{1 - \pi_i}{\pi_i} \right) \right] \quad \dots (24)$$

وهي علاقة غير خطية وعند حلها يتطلب استخدام احدى الطرق التكرارية كما هو الحال مع طريقة الامكان الاعظم نستخدم طريقة نيوتن \_ رافسون لإيجاد تقديرات  $\beta$  حيث يتم حساب المشتقة الثانية بالإضافة الى المشتقة الاولى وكما يأتي:

$$\frac{\partial^2 R(\beta)}{\partial \beta_j \partial \beta_j} = \sum_{i=1}^n X_{ij} X_{ij} \left[ (1 - Y_i)^2 \left( \frac{\pi_i}{1 - \pi_i} \right) + (Y_i)^2 \left( \frac{1 - \pi_i}{\pi_i} \right) \right] \quad \dots (25)$$

ونطبق طريقة نيوتن رافسون حسب الصيغة الرياضية وكالاتي:

$$\hat{\beta}^{(m+1)} = \hat{\beta}^{(m)} - \left[ \hat{L}(\hat{\beta})^{(m)} \right]^{-1} \left[ \hat{L}(\hat{\beta})^{(m)} \right] \quad \dots (26)$$

$$\hat{L}(\hat{\beta})^{(m)} = \frac{\partial R(\beta)}{\partial \beta_j} \quad ; \quad \hat{L}(\hat{\beta})^{(m)} = \frac{\partial^2 R(\beta)}{\partial \beta_j \partial \beta_j} \quad \dots (27)$$

$$V^{(m)} = \begin{bmatrix} (Y_i)^2 \left( \frac{1-\pi_i}{\pi_i} \right) + (1-Y_i)^2 \left( \frac{\pi_i}{1-\pi_i} \right) & 0 & \dots & 0 \\ 0 & (Y_i)^2 \left( \frac{1-\pi_i}{\pi_i} \right) + (1-Y_i)^2 \left( \frac{\pi_i}{1-\pi_i} \right) & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & (Y_i)^2 \left( \frac{1-\pi_i}{\pi_i} \right) + (1-Y_i)^2 \left( \frac{\pi_i}{1-\pi_i} \right) \end{bmatrix}$$

$$V^{(m)} = \text{diag} \left[ (Y_i)^2 \left( \frac{1-\pi_i}{\pi_i} \right) + (1-Y_i)^2 \left( \frac{\pi_i}{1-\pi_i} \right) \dots \right] \quad (28)$$

حيث ان

Y: يمثل متجه متغير الاستجابة تم ذكره.

X: تمثل مصفوفة المتغيرات التوضيحية تم ذكرها.

$V^{(m)}$ : مصفوفة مربعة عناصر قطرها الرئيسي المعادلة اعلاه مكتسبة من التكرار السابق (m).

### 6-1 منهجية الخوارزمية الجينية (Genetic Algorithm Methodology)

بهدف الحصول على الحلول المثلى للمسائل الرياضية تم استخدام الخوارزميات الجينية كإحدى الطرق التكرارية المستخدمة حديثاً من أجل اتخاذ القرارات الصحيحة [21;2005;p.97].

ان اهم ما يميز هذه الطريقة هو ايجاد علاقة بين المشكلة قيد الدراسة وبين الخوارزمية الجينية وتنشأ هذه العلاقة عن طريق الترميز (Encoding) ودالة التقييم (Evaluation Function) ويكون ترميز الحلول باستخدام سلسلة من الارقام الثنائية (Binary Numbers) (0,1) وهو من افضل الحلول او استعمال رموزا اخرى كأرقام حقيقية وتسمى هذه الارقام كروموسومات [16;2011;p.1284].

امادالة التقييم فتأخذ كل كروموسوم على جانب وتقييم ادائه في حل المشكلة بإعطائه قيمة معينة وكلما كانت هذه القيمة كبييرة كلما كان الكروموسوم ذات كفاءة عالية (Fitness Function) وتسمى دالة المفاضلة ثم تطبق عملية التهجين والطفرة للحصول في النهاية على مجموعة الكروموسومات التي تمثل الجيل الاخير وباختيار الافضل وصولاً للحل الامثل وتنتهي عنده [22;2012;pp.1-2].

### 7-1 تطبيق مراحل الخوارزمية الجينية في نموذج الانحدار اللوجستي

(Application of the Stages of the Genetic Algorithm in the Logistic Regression Model)

نطبق مراحل الخوارزمية الجينية في معادلة دالة الهدف لكل طريقة لإيجاد تقديرات معاملات نموذج الانحدار اللوجستي وفقاً لما يأتي [4-3;2012;pp.3-4], [7-1;2008;pp.1-7].

1- البداية: تكوين الكروموسوم من خلال قيم  $\beta_p$  التي تشكل جينات الكروموسوم حيث ان  $(P = 0,1, \dots, p)$  هي حدود عظمى ضمن الاعداد الحقيقية.

2- التهينة: انشاء الجيل الابتدائي عن طريق ايجاد قيمة اولية للجينات مع القيم العشوائية لمجموعة القيود الاخرى.

3- في دالة الهدف يتم تقييم الكروموسوم من حيث الكفاءة وصولاً إلى الحل الأفضل بتحديد قيمة  $\beta_p$ .

4- اجراء عملية الاختيار للكروموسوم الذي يمتلك قيمة دالة هدف صغيرة باختيار الاحتمال الكبير لها وايجاد دالة التقييم له من خلال المعادلة التالية :

$$\text{fitness function} = \frac{1}{1 + \text{objective function}} \quad \dots (29)$$

$\text{fitness function}$  : تمثل دالة التقييم. ;  $\text{objective function}$  : تمثل دالة الهدف. ومن خلال معادلة دالة التقييم نستطيع ايجاد احتمالية هذه الدالة ( افضل القيم) حسب الصيغة الرياضية الآتية [9;2015;p.777].

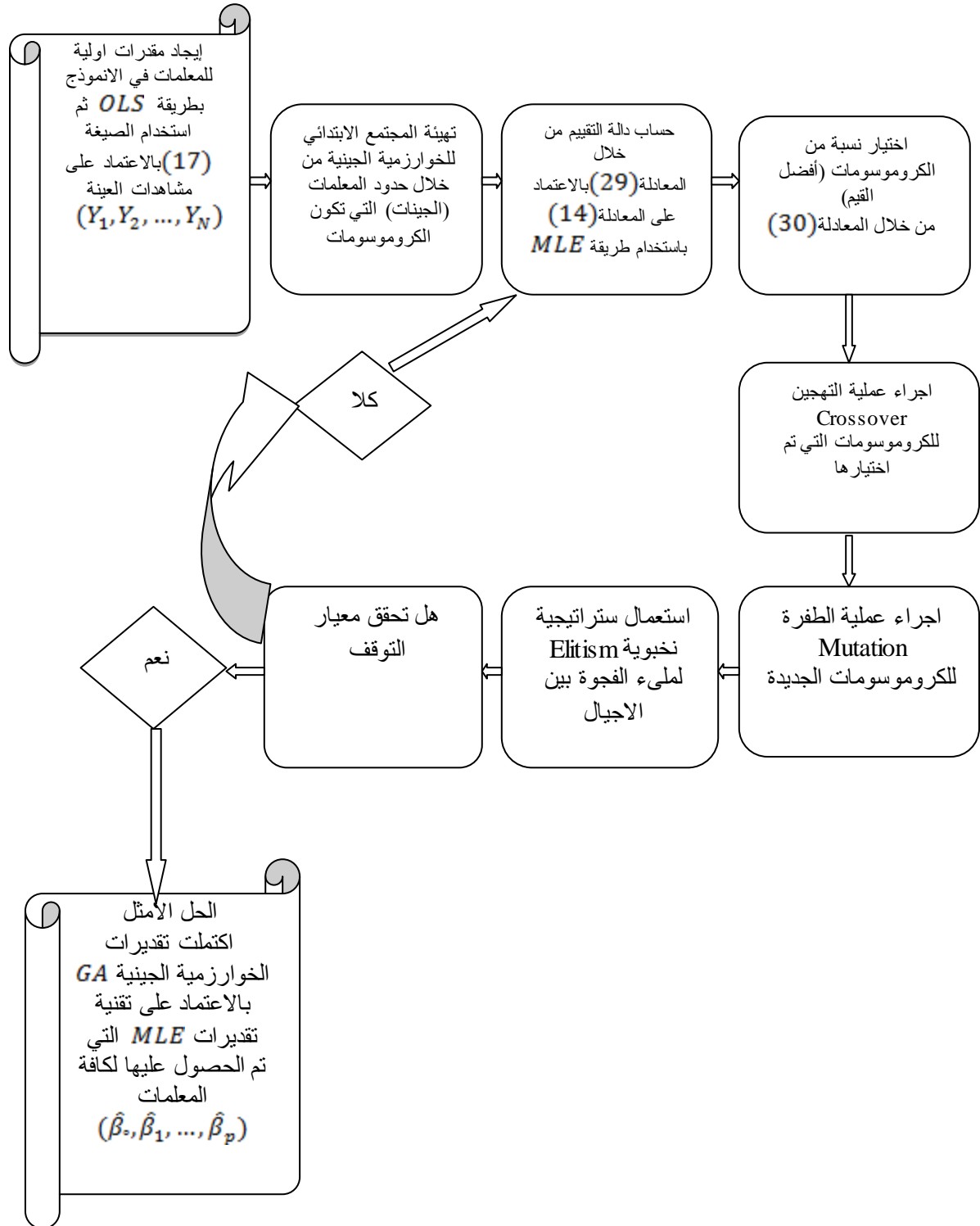
$$C_{(i)} = \frac{f_{(i)}}{\sum_{i=1}^N f_{(i)}} \quad \dots (30)$$

$C_{(i)}$ : تمثل احتمالية الفرد (الكروموسوم)  $i$ . ;  $f_{(i)}$ : تمثل دالة التقييم للفرد. ;  $N$ : يمثل حجم المجتمع  
وباستخدام احد معايير الاختيار (roulette wheel) عجلة الروليت بتوليد رقم عشوائي  $R_{(c)}$   
محصور في المجال [1,0] فإذا كان  $R_{(c)} < C_{(1)}$  سوف يتم اختيار الكروموسوم الاول (كالأم) او  
يتم الاختيار بحيث يكون الاحتمال محصور وفق  $R_{(c)} < C_{(p)} < C_{(p-1)}$  او يكون الرقم  
العشوائي محصور وفق  $R_{(c)} < C_{(p)} < C_{(p-1)}$  وفي كل مرة يتم تحديد كروموسوم واحد  
للمجتمع الجديد بالاعتماد على دالة المفاضلة.

5- بعد اتمام عملية الاختيار تأتي بعدها عملية التهجين للكروموسومات الجيدة في صفاتها عن  
طريق التزاوج بين كل كروموسومين وبتطبيق احد معاييرته وهو التهجين المنظم بالاعتماد على  
احتمالية التهجين  $P_c$  وتقارن هذه القيمة مع قيمة الجينات للكروموسومين ( الاباء) لتكوين الجيل  
الجديد(الابناء)ويحدث التبادل عندما تكون قيمة الجين اكبر او تساوي القيمة  
الاحتمالية [13;2008;p.505].

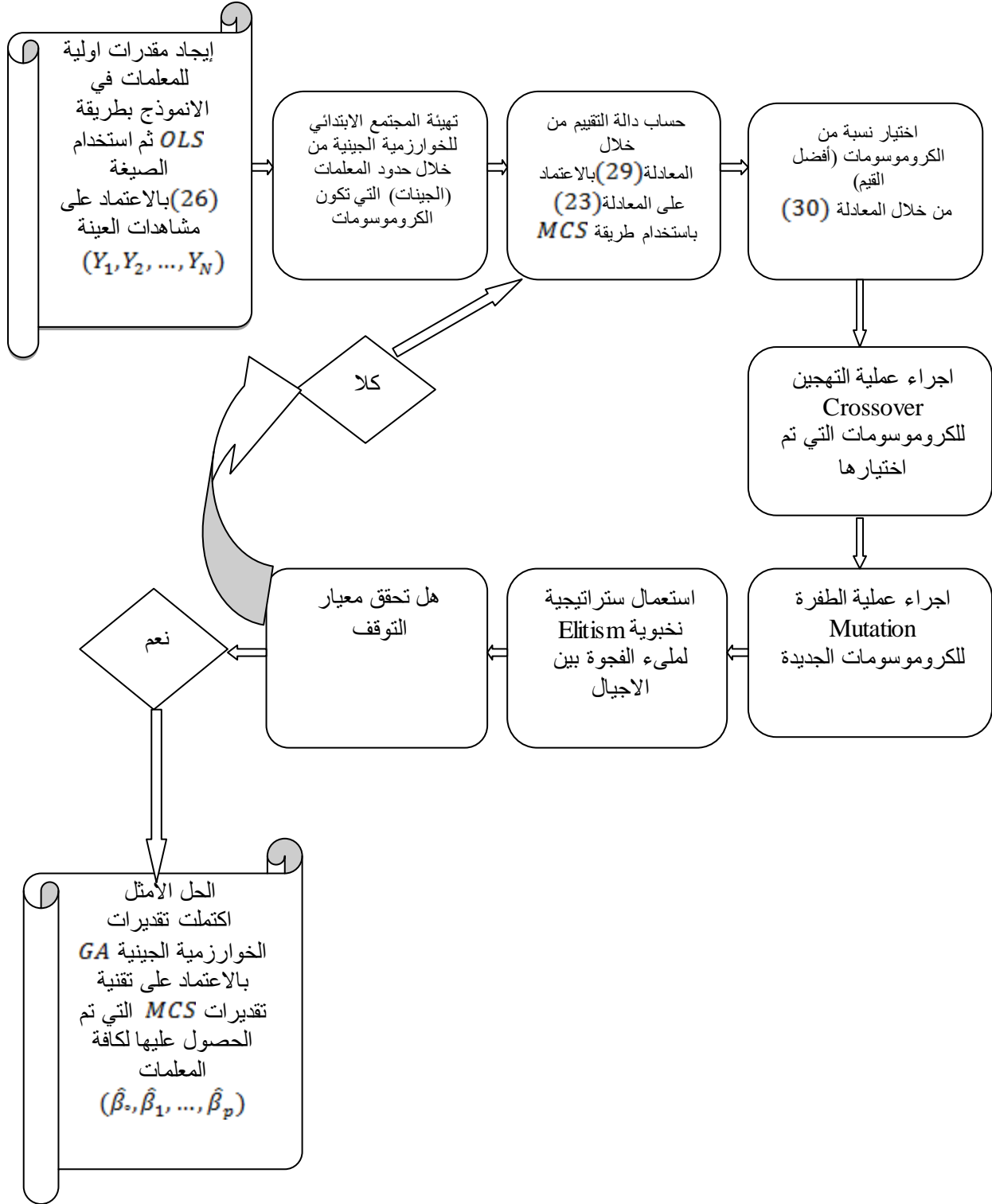
6- آخر خطوة ممكن ان تمر بها الكروموسومات هي عملية الطفرة وايضاً تعتمد على مقدار احتمالي ( $P_m$ )  
للمعلمت باستبدال جينات منتقاة عشوائياً مع قيمة جديدة ايضاً حصلنا عليها بشكل عشوائي [8;1991;p.89]  
بتطبيق المعادلة: مجموع الجينات = عدد الجينات في الكروموسوم \* مجموع السكان ... (31)  
وان المخططات التالية (1)، (2) تمثل تطبيق مراحل الخوارزمية الجينية في نموذج الانحدار اللوجستي وفق  
الطرائق (MCS, MLE) على التوالي التي عرضناها في البحث.

الشكل (1) يمثل المخطط الانسيابي للخوارزمية الجينية بالاعتماد على تقنية تقديرات MLE





الشكل (2) يمثل المخطط الانسيابي للخوارزمية الجينية بالاعتماد على تقنية تقديرات MCS



## 8-1 المحاكاة (Simulation)

يمكن استعمال اسلوب المحاكاة لإجراء مقارنة ما بين طرائق التقدير المدروسة أو المقترحة في البحث، من خلال المؤشر الاحصائي متوسط مربعات الخطأ للتوصل الى الطريقة الأفضل، وهذا هو المحور الذي انصب عليه اهتمامنا، إذ تم صياغة أنموذج محاكاة يحاكي العديد من الحالات المفترضة التي من الممكن وجودها فعلا في الواقع العملي من حيث (عدد المشاهدات الكلية، وقيم المعلمات، وعدد المتغيرات التوضيحية) بغية تحقيق الهدف الأساس المتمثل في إيجاد أفضل الطرائق المدروسة لتقدير معاملات انموذج الانحدار اللوجستي الثنائي وذلك من خلال ما تبين من تأثير طرائق التقدير عندما يتغير حجم العينة (عدد المشاهدات الكلية) وايضا عندما يكون التغير في قيم معلمات الانموذج.

وتتم المقارنة بين طرائق التقدير الاعتيادية ( $MLE$ ) و ( $MCSE$ ) وتحديد الافضل، اما المقارنة الاخرى فكانت بين طرائق التقدير المحسنة بالخوارزمية الجينية ( $MLE.GA$ ) و ( $MCSE.GA$ ) وتحديد الافضل، أن تكوين تجربة المحاكاة التي سيتم الحصول عليها تعتمد على عدد من المراحل، والتي سوف يتم توضيحها في انموذج الانحدار اللوجستي الثنائي الذي تم تعريفه سابقا في معادلة (8).

## 9-1 مراحل بناء تجربة المحاكاة (Stages of Building Simulation Experiment)

ان بناء تجربة المحاكاة تتضمن اربع مراحل مهمة وهي كالاتي:

### 1-9-1 المرحلة الأولى- تحديد القيم الافتراضية:

في هذه المرحلة يتم تحديد القيم الافتراضية للمعلمة، وذلك من خلال انموذج الانحدار الخطي المتعدد من تقدير البيانات الحقيقية للبحث بطريقة  $OLS$  ومرة اخرى بطريقة  $Standardized$  وفيها يتم تحديد اربع اجسام مختلفة للعينات، فتكون عدد المشاهدات ( $N = 50,100,150,200$ ) مشاهدة على التوالي، وتعد هذه المرحلة من أهم المراحل التي تعتمد عليها بقية مراحل المحاكاة، وكما مبين في الجدول (1) :

جدول (1) القيم الافتراضية للمعلمات في انموذج الانحدار اللوجستي الثنائي

Para	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$	$\beta_6$	$\beta_7$	$\beta_8$	$\beta_9$	$\beta_{10}$
..											
OLS	-1.031	-0.073	0.095	-0.055	-0.000415	0.013	-0.002	0.816	-0.000039	0.001	0.004
Sta.	-1.031	-0.088	0.091	-0.046	-0.013	0.060	-0.004	0.803	-0.002	0.150	0.092

### 2-9-1 المرحلة الثانية- توليد البيانات:

في هذه المرحلة يتم توليد قيم المتغيرات التوضيحية وقيم متغير الاستجابة حيث تم توليد عشر متغيرات توضيحية كما في الجانب التطبيقي باستخدام اسلوب مونت كارلو وذلك من خلال التوزيع المنتظم  $Uniform distribution$ ، اما في ما يتعلق بقيم متغير الاستجابة تحت افتراض انه يتبع توزيع برنولي  $[0,1]$ ، يتم احتسابها وفق طريقة الرفض والقبول وكما يلي:

$$Y = \begin{cases} 1 & \text{if } \pi(X_i) > 0.5 \\ 0 & \text{if } \pi(X_i) < 0.5 \end{cases}$$

### 3-9-1 المرحلة الثالثة- إيجاد التقديرات

يتم في هذه المرحلة تقدير معالم نموذج الانحدار اللوجستي الثنائي المعطى في الصيغة (8) وفق الطرائق الاعتيادية، وايضا وفق توظيف الخوارزمية الجينية مع طرائق التقدير الاعتيادية التي تم ذكرها في البحث، وهذه الطرائق هي:

#### A. طرائق التقدير الاعتيادية

1-طريقة تقديرات الامكان الاعظم (MLE).

2-طريقة تقديرات اصغر مربع كاي (MCSE).

#### B. طرائق التقدير المحسنة

1-طريقة الخوارزمية الجينية بالاعتماد على تقنية تقديرات الامكان الاعظم (MLE.GA).

2-طريقة الخوارزمية الجينية بالاعتماد على تقنية تقديرات تصغير مربع كاي (MCSE.GA).

### 4-9-1 المرحلة الرابعة- المقارنة بين طرائق التقدير:

يتم في هذه المرحلة المقارنة ما بين طرائق تقدير معالم نموذج الانحدار اللوجستي الثنائي والتي تكون على اساس (الاعتيادية مع المحسنة والاعتيادية من هي الافضل والمحسنة من هي الافضل) لكافة الطرائق قيد البحث، وذلك باستعمال احد المعايير الاحصائية المهمة وهو متوسط مربعات الخطأ لمقدرات النموذج المدروس، والهدف الذي يراد تحقيقه هو الحصول على المقدر  $[(\hat{Y}_i) = \hat{\pi}(X_i)]$  حسب صيغة المعادلة الآتية [15;2002;p.19].

$$MSE = \frac{1}{R} \sum_{i=1}^R MSE_i = \frac{1}{R} \sum_{i=1}^R \left[ \frac{1}{N} \sum_{i=1}^N (Y_i - \hat{Y}_i)^2 \right] \quad \dots (32)$$

R : يمثل عدد مرات تكرار التجربة، وقد تم تكرار تجربة المحاكاة عدد  $R = 1000$  للحصول على نتائج الجانب التجريبي.

جدول (2) متوسط مربعات الخطأ (MSE) لمقدرات نموذج الانحدار اللوجستي الثنائي في الحالة الاولى

للمعلمت باستعمال الطرائق (Mcs و Mle) الاعتيادية والجينية ومختلف احجام العينات

Sample size	Methods	Classic	Genetic	Best
50	Mle	0.231229552871543	0.005659256776011	Genetic
	Mcs	0.215989821797131	0.000970068091206	Genetic
100	Mle	0.195214313074604	0.002374785723625	Genetic
	Mcs	0.155371787851206	0.000468182866024	Genetic
150	Mle	0.010072144667444	0.002319596320434	Genetic
	Mcs	0.038498319485976	0.000312568716347	Genetic
200	Mle	0.105158079021595	0.001984539745217	Genetic
	Mcs	0.073204095012192	0.000241971930941	Genetic
	Best	Mcs	Mcs.GA	

في الجدول رقم (2) عند احجام العينات ( $N = 50,100,150,200$ ) نلاحظ تفوق طريقة الامكان الاعظم المحسنة (Mle.GA) على طريقة الامكان الاعظم الاعتيادية (Mle) وطريقة تصغير مربع كاي المحسنة (Mcs.GA) على طريقة تصغير مربع كاي الاعتيادية (Mcs) في تقدير المعلمت من حيث امتلاكها اقل (MSE) للمقدرات، وايضا نميز ان طريقة (Mcs.GA) قد احتلت المرتبة الاولى للأفضلية في تقدير المعلمت من حيث امتلاكها اقل (MSE) للمقدرات مقارنة بطريقة (Mle.GA) بالنسبة للطرائق الجينية وتصدرت طريقة (Mcs) المرتبة الاولى مقارنة بطريقة (Mle) بالنسبة للطرائق الاعتيادية وذلك لكافة القيم الافتراضية للمعلمت والنموذج الاول الذي تم افتراضه.



## مقارنة بعض الطرائق لتقدير معلمات نموذج الانحدار اللوجستي الثنائي باستعمال الخوارزمية الجينية مع تطبيق عملي

جدول (3) متوسط مربعات الخطأ (MSE) لمقدرات نموذج الانحدار اللوجستي الثنائي في الحالة الثانية للمعلمات باستعمال الطرائق (Mcs و Mle) الاعتيادية والجينية ومختلف احجام العينات

Sample size	Methods	Classic	Genetic	Best
50	Mle	0.135106928071695	0.005582329026557	Genetic
	Mcs	0.090487296781982	0.000989744128643	Genetic
100	Mle	0.183950622761129	0.002338797164337	Genetic
	Mcs	0.139472754614625	0.000497968581145	Genetic
150	Mle	0.102916386500496	0.002237460217365	Genetic
	Mcs	0.073973528994289	0.000332468599320	Genetic
200	Mle	0.124285386959490	0.001934958542496	Genetic
	Mcs	0.093994577706719	0.000249435210068	Genetic
	Best	Mcs	Mcs.GA	

في الجدول رقم (3) عند احجام العينات ( $N = 50, 100, 150, 200$ ) نلاحظ تفوق طريقة الامكان الأعظم المحسنة ( $Mle.GA$ ) على طريقة الامكان الأعظم الاعتيادية ( $Mle$ ) وطريقة تصغير مربع كاي المحسنة ( $Mcs.GA$ ) على طريقة تصغير مربع كاي الاعتيادية ( $Mcs$ ) في تقدير المعلمات من حيث امتلاكها اقل ( $MSE$ ) للمقدرات، وايضا نميز ان طريقة ( $Mcs.GA$ ) قد احتلت المرتبة الاولى للأفضلية في تقدير المعلمات من حيث امتلاكها اقل ( $MSE$ ) للمقدرات مقارنة بطريقة ( $Mle.GA$ ) بالنسبة للطرائق الجينية وتصدرت طريقة ( $Mcs$ ) المرتبة الاولى مقارنة بطريقة ( $Mle$ ) بالنسبة للطرائق الاعتيادية وذلك لكافة القيم الافتراضية للمعلمات والانموذج الثاني الذي تم افتراضه.

### 10-1 وصف البيانات

اعتمدت الدراسة على الاحصاءات الطبية لمحافظة ذي قار لسنة 2016 م الصادرة من مركز امراض القلب لـ (110) شخص مريض بالقلب وتم استخدام احدى عشر متغير منها متغير واحد تابع (متغير الاستجابة والذي يأخذ قيمتين هما الصفر لاحتمال عدم حدوث الوفاة والواحد الصحيح لاحتمال حدوث الوفاة وعشرة متغيرات توضيحية، تم تعريف كل متغير في الجدول (4) حيث اخذت المتغيرات الاتية:

جدول (4) تعريف المتغيرات التوضيحية ومتغير الاستجابة

رمز المتغير	تمثيل المتغير	القيم التي يأخذها	تمثيل كل قيمة
Y	متغير الاستجابة	0	احتمال عدم حدوث الوفاة
		1	احتمال حدوث الوفاة
X <sub>1</sub>	متغير وصفي يمثل نوع الردهة التي يرقدها المريض	1	ردهة القلب المفتوح
		2	ردهة انعاش القلب
		3	ردهة القلب و الاوعية
		4	ردهة التمريض الخاص
		1	ذكر
X <sub>2</sub>	متغير وصفي يمثل جنس المريض الراقده	1	ذكر
		2	انثى



## مقارنة بعض الطرائق لتقدير معاملات نموذج الانحدار اللوجستي الثنائي باستعمال الخوارزمية الجينية مع تطبيق عملي

اعزب	1	متغير وصفي يمثل الحالة الزوجية	X <sub>3</sub>
متزوج	2		
ارمل او ارملة مطلق او مطلقة	3		
متغير كمي يمثل عمر المريض بالسنوات			X <sub>4</sub>
ذبحة صدرية	1	متغير وصفي يمثل نوع المرض الذي يعاني منه المريض الراقد	X <sub>5</sub>
احتشاء واعتلال عضلة القلب	2		
اضطرابات في توصيل الدم	3		
توقف او تسارع او رجفان القلب	4		
عجز القلب	5		
جلطة دماغية	6		
انسداد الشرايين	7		
اضطرابات في جهاز الدوران	8		
وذمة رئوية	9		
تشوهات خلقية ولادية	10		
عدم اجراء عملية	1	متغير وصفي يمثل التداخل الجراحي للمريض الراقد	X <sub>6</sub>
اجراء عملية وسطى	2		
اجراء عملية كبرى	3		
اجراء عملية فوق الكبرى	4		
مدخن	1	متغير وصفي يمثل التدخين	X <sub>7</sub>
غير مدخن	2		
متغير كمي يمثل ضغط المريض			X <sub>8</sub>
متغير كمي يمثل سكر المريض			X <sub>9</sub>
متغير كمي يمثل وزن المريض			X <sub>10</sub>

جدول (5) المعلمات المقدرة وقيم الخطأ المعياري لكافة المتغيرات التوضيحية بطريقة (Mcs.GA)

Coeffi. Para. of Variables (X <sub>i</sub> )	Estimator of Parameter ( $\hat{\beta}_i$ )	Error Standard $SE(\hat{\beta}_i)$	Ratio ( $Z$ ) $\frac{\hat{\beta}_i}{SE(\hat{\beta}_i)}$	Signific.
$\beta_0$	-4.35763006967658	16.0595143437140	-0.271342580878372	Non-sig
$\beta_1$	-0.289810441734439	0.498757009846207	-0.581065400612219	Non-sig
$\beta_2$	0.338465641827383	0.560370005529038	0.604003851897538	Non-sig
$\beta_3$	-0.247076451451670	0.742557349011530	-0.332737197713511	Non-sig
$\beta_4$	-0.00590462410339408	0.0990339755267670	-0.0596222061367028	Non-sig
$\beta_5$	0.0526362472209598	0.249747027437004	0.210758253105682	Non-sig
$\beta_6$	0.0496411594287042	0.388458496016308	0.127790124138822	Non-sig
$\beta_7$	2.11351984839268	0.550704829928840	3.83784512778975	Sig
$\beta_8$	0.00583105067242122	0.0857054893811672	0.0680359066207320	Non-sig
$\beta_9$	0.000458356744008137	0.0397746063623089	0.0115238536827478	Non-sig
$\beta_{10}$	0.0218092350856305	0.120615630956456	0.180815993024188	Non-sig

تتم معالجة البيانات التي تعاني من وجود عشر عوامل تسبب حالة الوفاة للأشخاص الذين يعانون من امراض القلب بتطبيق افضل الطرائق التي تم الحصول عليها من نتائج المحاكاة وهي ( $Mcs.GA$ )، حيث نلاحظ في الجدول (5) تقديرات المعلمة وقيم الخطأ المعياري لكل معلمة مقدره بطريقة تصغير مربع كاي الجينية ( $Mcs.GA$ )، ان تقدير المعلمات ضروري لاختبار اهمية التأثير الكلي على المتغيرات التوضيحية، من خلال النتائج التي حصلنا عليها نجد نسبة ( $\bar{Z}$ ) التي تتبع التوزيع الطبيعي القياسي وبمستوى دلالة ( $\alpha = 0.05$ ) وبمقارنتها مع قيمة ( $Z$ ) الجدولية  $Z_{\frac{\alpha}{2}(0.05)} = 1.96$  نحصل على العوامل (المتغيرات المؤثرة وغير المؤثرة في النموذج).

جدول (6) المعلمات المقدره وقيم الخطأ المعياري للمتغيرات التوضيحية المؤثرة فقط (المعنوية) بطريقة ( $Mcs.GA$ )

Coeffi. Para. of Variables ( $X_i$ )	Estimator of Parameter ( $\hat{\beta}_i$ )	Error Standard $SE(\hat{\beta}_i)$	Ratio ( $\bar{Z}$ ) $\frac{\hat{\beta}_i}{SE(\hat{\beta}_i)}$	Significance
$\beta_7$	2.11351984839268	0.550704829928840	3.83784512778975	Sig.

نلاحظ في الجدول (6) ان معامل انحدار متغير التدخين  $\hat{\beta}_7$  فقط يكون معنوي ويؤثر في النموذج بشكل كبير وتم ازالة المعلمات المقدره  $\beta_0$  و  $\hat{\beta}_1$  و  $\hat{\beta}_2$  و  $\hat{\beta}_3$  و  $\hat{\beta}_4$  و  $\hat{\beta}_5$  و  $\hat{\beta}_6$  و  $\hat{\beta}_8$  و  $\hat{\beta}_9$  و  $\hat{\beta}_{10}$  لأنها لا تؤثر على النموذج المقدر وكان من الضروري اعادة تقدير المعلمات وقيم الانحراف المعياري ونسبة ( $\bar{Z}$ ) كما في جدول (7) اي ان المتغيرات (الردهة  $X_{i1}$  والجنس  $X_{i2}$  والحالة الاجتماعية  $X_{i3}$  والعمر  $X_{i4}$  وسبب الرقود  $X_{i5}$  ودرجة العملية  $X_{i6}$  والضغط  $X_{i8}$  والسكر  $X_{i9}$  والوزن  $X_{i10}$ ) ليس لها تأثير على حالة حدوث الوفاة بالنسبة للمصابين بأمراض القلب، ونلاحظ ايضا اشارة معامل انحدار متغير التدخين ( $\hat{\beta}_7$ ) موجبة وهذا يعني ان كلما اقترب متغير التدخين من القيمة (0) غير مدخن يتجه متغير الاستجابة لأخذ القيمة (0) عدم حدوث الوفاة ومن خلال القيمة يتضح انه كلما زاد متغير التدخين واقترب من القيمة (1) أدى ذلك الى زيادة احتمال حدوث الوفاة بمعدل (2.114)، ونحصل على القيم التقديرية لأنموذج الانحدار اللوجستي الثنائي التي تم استخراجها حسب المعادلة الآتية:

$$\hat{\pi}(X) = \frac{e^{(\hat{\beta}_7 X_{i7})}}{1 + e^{(\hat{\beta}_7 X_{i7})}} = \frac{e^{(2.11351984839268 X_{i7})}}{1 + e^{(2.11351984839268 X_{i7})}} \quad \dots (28)$$

جدول (7) تصنيف بيانات العينة باستخدام النموذج المقدر بطريقة ( $Mcs.GA$ )

المجموع	التنبؤ		حالة المريض
	حدوث الوفاة 1 $\hat{\pi} \geq 0.5$	عدم حدوث الوفاة 0 $\hat{\pi} < 0.5$	
46	5	41	عدم حدوث الوفاة Y=0
64	63	1	حدوث الوفاة Y=1
110	68	42	المجموع
	89.1		دقة النموذج
	98.4		حساسية النموذج
	94.5		نسبة التصنيف الصحيح

نلاحظ في الجدول (7) ان النموذج الانحدار اللوجستي قام بتصنيف المتغير المعنوي ( $X_7$ ) من خلال ايجاد قيم النموذج الانحدار اللوجستي الثنائي حيث صنف النموذج 41 مصاب من اصل 46 ممن لا توجد لديهم حالة الوفاة تصنيفاً صحيحاً، وبلغت نسبة التصنيف الصحيح لحالة عدم حدوث الوفاة 89% للمصابين بأمراض القلب، و صنف 63 مصاب من اصل 64 ممن حدثت معهم حالة الوفاة حيث بلغت نسبة التصنيف الصحيح لحالة حدوث الوفاة 98% من المصابين بأمراض القلب ويعزى ذلك الى حساسية طبيعة الدراسة (ما يتعلق بحالة حدوث الوفاة) اي انه يستطيع التنبؤ بطريقة صحيحة بناء على المتغيرات التوضيحية المدخلة فيه للذين يعانون من امراض القلب وحدثت معهم حالة الوفاة، وقد كانت نسبة التصنيف الصحيح بصورة عامة مرتفعة بلغت 95% دل ذلك على جودة النموذج المقدر وتعتبر هذه النسبة مقبولة جداً ، اي ان نسبة الخطأ تساوي 5% تم تصنيفهم بصورة خاطئة، دل ذلك على جودة النموذج المقدر.

### 11-1 الاستنتاجات (Conclusions)

بتطوير وتحسين استعمال الطرائق الاعتيادية ( $MCS$ ) ( $MLE$ )، من قبل الباحثة من خلال توظيف الخوارزمية الجينية في التقدير وهي ( $GA.MLE$ )، ( $GA.MCS$ )، وبناءً على ما تم استخراجها من تجارب المحاكاة وما تم تحليله من نتائج في الجانب التطبيقي فقد تم التوصل الى الاستنتاجات الآتية:

- 1- تفوق طرائق التقدير المحسنة على طرائق التقدير الاعتيادية في تقدير المعلمات لأنموذج الانحدار اللوجستي الثنائي وذلك لكافة حجوم العينات التي تم افتراضها.
- 2- تفوقت طريقة التقدير الاعتيادية ( $MCS$ ) على طريقة التقدير الاعتيادية ( $MLE$ ) إذ حققت المرتبة الأولى من الأفضلية في تقدير المعلمات لأنموذج الانحدار اللوجستي الثنائي وذلك لكافة حجوم العينات والنماذج.
- 3- تفوقت الطريقة المحسنة بالخوارزمية الجينية ( $GA.MCS$ ) التي تم تطويرها من الباحثة على طريقة التقدير المحسنة ( $GA.MLE$ ) إذ حققت الأفضلية في تقدير المعلمات لأنموذج الانحدار اللوجستي الثنائي وذلك لكافة حجوم العينات ولكلا الأنموذجين المفترضين.
- 4- توصلنا الى انه عند احجام العينة ( $N = 50,100,150,200$ ) بالنسبة للقيم الافتراضية للمعلمات والنماذج المفترضة في حالة الطرائق الاعتيادية (الامكان الاعظم وتصغير مربع كاي) وعند توظيف الخوارزمية الجينية في هذه الطرائق ولنفس النماذج المفترضة نلاحظ تفوق ( $GA.MCS$ ) وهي الأفضل، وهذا يدل على تحسن أداء طريقة مقدرات النموذج اللوجستي الثنائي عند استعمال الخوارزمية الجينية في إيجاد افضل مقدر للمعلمة المجهولة.
- 5- تتناقص قيمة ( $MSE$ ) لمقدرات الأنموذج اللوجستي باستعمال كافة طرائق التقدير الاعتيادية والمحسنه بازدياد حجم العينة، وهذا يطابق النظرية الاحصائية.
- 6- استعمال الطريقة الأفضل التي تم الحصول عليها من نتائج تجربة المحاكاة بالنسبة لطرائق التقدير الاعتيادية والمحسنه ( $GA.MCS$ ) في تقدير معلمات الانموذج اللوجستي الثنائي لحالات حدوث الوفاة اذ يمكن اعادة بناء الانموذج من خلال المعلمات المقدره ، فيتم التعبير عن حالة المريض بواسطة المعلمات المقدره.

- 7- أشارت النتائج التي حصلنا عليها من اختبار ( $z$ ) الى ان عامل التدخين هو الاكثر تأثيراً وبمعنوية اقل من 5% على متغير الاستجابة (حالة المصابين بأمراض القلب) باستخدام نموذج الانحدار اللوجستي الثنائي بطريقة ( $MCS.GA$ ).
- 8- أثبتت الطريقة ( $MCS.GA$ ) ان متغير التدخين يعتبر أهم متغير لتحديد حالة حدوث الوفاة وهذا ما تم تأكيده من قبل الأطباء المتخصصين، وأشارت النتائج التي توصلنا اليها في دراستنا انها متفقة مع الدراسات السابقة في المتغيرات التوضيحية الاكثر تأثيراً على حالة المريض.
- 9- من خلال اختبار جدول التصنيف رقم (7) نجد ان نسبة التصنيف الصحيح بلغت باستخدام نموذج الانحدار اللوجستي الثنائي المقدر بطريقة ( $MCS.GA$ ) (94.5) , أي انه يمكن استعمال هذه الطريقة لتصنيف الحالات الجديدة للمرضى المصابين بأمراض القلب الى (حدوث وفاة ، عدم حدوث وفاة) اعتماداً على قيم المتغيرات التوضيحية لتلك الحالات.
- 10- من خلال البيانات الحقيقية التي تمثل المصابين بأمراض القلب من حيث حالات الوفيات الحقيقية والمقدرة، تم التوصل الى مدى ملائمة نموذج الانحدار اللوجستي الثنائي في نمذجة هذا النوع من البيانات، والذي كان واضحاً من خلال حالات الوفيات المقدرة في الجانب التطبيقي والتي لا يمكن تمييز الفرق بينها وبين حالات الوفيات الحقيقية، والذي يدل ايضاً على مدى دقة الطريقة المقترحة ( $MCS.GA$ ) في تقدير معلمات هذا النموذج.

## 12-1 التوصيات

- 1- نوصي باستعمال طريقة ( $MCS.GA$ ) لتقدير معلمات نموذج الانحدار اللوجستي الثنائي في حالة توزيع *Bernoulli* للبيانات.
- 2- تطوير الطريقة التي تم عرضها من الباحثة لتقدير معلمات نموذج الانحدار اللوجستي المتعدد والترتيبي والذي يمكن استعماله لنمذجة البيانات في المجالات الطبية.
- 3- اجراء الطريقة التي تم عرضها باستخدام الخوارزمية الجينية في مجالات اخرى مثل الآلات التكنولوجية التي تنطبق عليها المواصفات المذكورة والتي يمكن للنموذج أن يمثلها او على البيانات الاجتماعية.
- 4- تحديث وتطوير قاعدة جمع البيانات على الكومبيوتر في المؤسسات الصحية من خلال استعمال أنظمة برمجية حديثة بغية الحصول على بيانات حقيقية ونتائج اكثر دقة.
- 5- اقتراح استخدام خوارزميات تطويرية اخرى مع الطرائق الاعتيادية الكلاسيكية كالطريقة التي تم استعمالها في البحث كخوارزمية مستعمرة النمل (*Ant Colony Optimization (ACO)*) و خوارزمية محاكاة التلدين (*Simulated Annealing Algorithm (SA)*) والمقارنة بينهم من خلال المعيار الاحصائي ( $MSE$ ) للتوصل الى أفضل طريقة في التقدير.

## المصادر

1. Agresti, A., (2002) , "Cartegorical Data Analysis", 2th edition, Jhon Wiley & sons Inc , , Hoboken, New Jersey.
2. Akkus ,Ö. , Demir , E. , (2016), " Comparison Som Classical And Meta-Heuristic Optimazation Techniques in The Estimation Of The Logit Model Parameters", International Journal of Advanced Research, ISSN, pp.1026-1042.





3. Alrahamneh, A. & Hawamdeh, O. ,(2017) , "The Factors Affecting Eye Patients (Cataract) In Jordan by Using the Logistic Regression Model", Modern Applied Science, ISSN, pp. 38-42.
4. Cramer, J. S., (2003),"Logit Models From Economics and other fields", Cambridge University Press cape Town, New York, ISBN, pp.33-45.
5. Demir , E. , Akkus , Ö., (2015), " An Introductory Study on How the Genetic Algorithm Works in the Parameter Estimation of Binary Logit Model", International Journal of Sciences : Basic and Applied Research, ISSN , pp.162-180.
6. Gayou, O., & et al., (2008)," A genetic algorithm for variable selection in logistic regression analysis of radiotherapy treatment outcomes", American Association of Physicists in Medicine, pp. 5426 – 5433.
7. Gelfand, A., & Smith, A., (1990)," Sampling Based Approaches to Calculating Marginal Densities", JASA, pp. 398-409.
8. Goldberg, D. E., & Deb, K., (1991)," A Comparative Analysis of Selection Schemes Used in Genetic Algorithms ,Foundation of Genetic Algorithms", San Francisco: CA: Morgan Kaufmann, pp. 69-93.
9. Hadji, S. & et al. ,(2015)," Theoretical and experimental analysis of genetic algorithms based mppt for PV systems", EnergyProcedia,pp.772-787.
10. Hosmer, D., Lemeshow, S. ,Sturdivant , R. ,(2013)," Applied Logistic Regression", 3<sup>rd</sup> edition ,New York: wiley [http : // ihmsi.org](http://ihmsi.org).
11. Hussain, J. N., Nassir, A. J., (2015)," Cluster Analysis as a Strategy Of Grouping to Construct Goodness – Of – Fit Tests when the Continuous Covariates Present in the Logistic Regression Models", BJMCS, pp. 1-16.
12. Lee, K.H., Kim, K.W.,(2015),"Performance comparisons of particle swarm optimization and genetic algorithm for inverse surface radiation problem", International Journal of Heat and Mass Transfer, PP. 330-337.
13. Liu, H., Ong, C., (2008), "Variable selection in clustering for marketing segmentation using genetic algorithms", Expert Systems with Applications, pp. 502-510.
14. Mahdavi, I., Paydar, M., & et al. ,(2008)," Genetic alogorithm approach for solving acell formation problem in cellubar manufacturing", Expert systems with Applications, pp. 1-7.
15. Menard ,S.,(2002),"Applied Logistic Regression Analysis",2<sup>nd</sup> Edition Thousand Oaks Edition Thousand Oaks , CA : Sage Publications , Series Quantitative Applications in the Social Sciences , PP. 1 – 111.



16. Meng ,Q., Weng, J., (2011) , " A genetic Algorithm approach To assessing Work Zone casualty Risk " , Safety Science ,49, pp. 1283-1288.
17. Misra, A. ,(2013)," Portfolio optimization of commercial Banks- An Application of Genetic Algorithm", European Journal of Business and management, ISSN, pp. 120-129.
18. Pasia, J., Hermosilla, A., & et al., (2005)," A useful tool for statistical estimation genetic algorithm", Journal Of Statistical Computation and Simulation , (Nov. 2014) , pp. 237 – 251.
19. Raghupathikumar, D., & Raja, K., (2012)," genetic algorithm based scheduling of an input queued switch", International Journal Of Computer Applications, pp.37-42.
20. Rodriguez , G. ,(2007), "Logit Models for Binary Data" ,Chapter(3) ,Retrieved from, <http://data.princeton.edu/wws509/notes/c3.pdf> ,pp.1-50.
21. Sastry K., & Goldberg .D., (2005), "Genetic Algorithm",Retrieved from Internet:www.citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.61.6575&rep=rep &type=Pdf ,Jon 25, (May,2005).
22. Sefian , S., Benbouziane , M. ,(2012) ,"Portfolio Selection Using Genetic Algorithm" , MPRA , Journal of Applied Finance & Banking, pp.143-154.
23. Yoder, S. E., (2009)," An Investigation on the use and flexibility of Genetic Algorithm for Logistic Regression", Clemson University , All Dissertations, pp. 1-77.



---

**Comparison of some methods for estimating the parameters of the binary logistic regression model using the genetic algorithm with practical application**

**Abstract**

Suffering the human because of pressure normal life of exposure to several types of heart disease as a result of due to different factors. Therefore, and in order to find out the case of a death whether or not, are to be modeled using binary logistic regression model

In this research used, one of the most important models of nonlinear regression models extensive use in the modeling of applications statistical, in terms of heart disease which is the binary logistic regression model. and then estimating the parameters of this model using the statistical estimation methods, another problem will be appears in estimating its parameters, as well as when the number parameters  $(P + 1)$ , and to find estimate the parameters using the numerical methods, sometimes does not give optimum solution because it depends on the initial estimators.

Some standard methods have been proposed and employed after modifying them by using the genetic algorithm approach in estimation to suit the estimation of the parameters of this of nonlinear regression models, and then making a comparison between two types of the important estimation methods including the standard estimation methods which included the maximum likelihood method, minimum chi-square method, and improved estimation methods developed which by the researcher which included genetic algorithm method depending on the technique estimates  $MLE$ , genetic algorithm method depending on the technique estimates  $MCSE$ , to choose the best method of estimation by default values to estimate parameter multi-linear regression model a method ols and then convert values the real to standardized and different samples sizes during simulation and by using the statistical criteria Mean Squares Error (MSE) for estimators.

The  $(Mcs)$  method is found to be the best one in the first place one among the standard estimation methods, and  $(Mcs.GA)$  method is the best among the important estimation methods for the purpose of estimating the parameters for binary logistic regression model because it has less (MSE) for estimators compared to other methods.

In the practical side of this study, this model has been used for modeling the own data infected heart disease and estimating the parameters using the  $(Mcs.GA)$  method, reached in it by comparing reasons for cases of occurrence death the real with reasons for cases of occurrence death for the estimated to the appropriate model in the modeling of this type of data and extraction the main cause of death is smoking and also the accuracy of the  $(Mcs.GA)$  method in estimating the parameters of the model.

**Keyword:** Minimum Chi-Square, Parameters, Genetic Algorithm, Newton-Raphson Algorithm.