مقارنة بين الخوارزمية الجينية والمربعات الصغرى الموزونة لتقدير أنموذج انحدار لوجستي متعددة الاستجابة (باستخدام المحاكاة Simulation)

A comparison between genetic algorithm and weighted least squares for estimating a multiple response logistic regression model using simulation

أ.د عبد الحسين حسن حبيب
Prof. Abdul hussian.H.Habeab.AL.Tai
كلية الادارة والاقتصاد جامعة كربلاء
College of Administration and Economics
University of Karbala
@s.uokerbala.edu.iq
07714875859

تقى عبد محمد الخزعلي
Tuqa. Abid. Mohammed
كلية الادارة والاقتصاد جامعة كربلاء
College of Administration and Economics
University of Karbala
tuqaa.a@s.uokerbala.edu.iq
07735206665

المستخلص:

يعد الانموذج الانحدار اللوجستي واحد من اهم نماذج تحليل الانحدار في حالة البيانات الوصفية. حيث تضمن هذا البحث مقارنة بين طريقة الخوارزمية الجينية (**Genetic algorithm**)مع طريقة المربعات الصغرى الموزونة(**Weighted Least Squares**)، حيث تم اجراء تجربة محاكاة(**Simulation**) لحجوم عينات مختلفة مختلفة(صغيرة ،متوسطة، كبيرة) (125،75،25) لثلاثة حالات من القيم الافتراضية تم وضعها من قبل الباحثة للقيم معلمات الانموذج(B_0،B_1،B_2)باستعمال طريقة مونت كارلو(Monte-Carlo) لغرض الوصول الى طريقة المثلى لتقدير انموذج الانحدار اللوجستي متعدد المستويات، وأظهرت نتائج ان طريقة الخوارزمية الجينية (GA) افضل واكثر كفاءة من طريقة المربعات الصغرى الموزونة (WLS) عند حجوم عينات الصغيرة (25).

الكلمات المفتاحية: الخوارزمية الجينية، المربعات الصغرى الموزونة ، نموذج انحدار لوجستي متعدد، المحاكاة .

ABSTRACT

The logistic regression model is one of the most important regression analysis models in the case of metadata. Where this research included a comparison between the genetic algorithm method with the weighted least squares method, where a simulation experiment was conducted for different sample sizes (small, medium, large) (125,75,25). For three cases of default values that were set by the researcher for the values of the model parameters (B2, B1, B0) using the Monte-Carlo method for the purpose of arriving at the optimal method for estimating a multi-level logistic regression model, and the results showed that the genetic algorithm (GA) method It is better and more efficient than the weighted least squares (WLS) method at small sample sizes (25).

Keywords: genetic algorithm, weighted least squares, multiple logistic regression model, simulation.

1-المقدمة:

يعد الانموذج الانحدار اللوجستي واحد من اهم الاساليب الإحصائية التي تنتمي الى النماذج الانحدار اللاخطية، حيث يتم نمذجة البيانات من خلال مجموعة من معلمات غير الخطية التي تعتمد على المتغير التابع ذي المستوى الاسمي ومتغير واحد او اكثر من المتغيرات المستقلة (المفسرة).

لذلك يعتبر أنموذج الانحدار اللوجستي حالة خاصة من نماذج الانحدار الاعتيادية بسبب طبيعته الاسمية التي يحملها متغير الاستجابة مثل (الحياة، الموت). ويستخدم أنموذج الانحدار اللوجستي بشكل واسع في تحليل البيانات والتطبيقات الإحصائية. ويكون الانحدار اللوجستي على عدة أنواع منها الانحدار اللوجستي الثنائي، ويستخدم هذا النوع عندما يأخذ متغير الاستجابة قيمتين وهما (0،1) وكذلك الانحدار اللوجستي متعدد الاستجابة وهو امتداد لأنموذج الانحدار اللوجستي الثنائي عندما يقع المتغير التابع في أكثر من فئتين، أما الانحدار اللوجستي الرتبي او يدعى الرتبوي ويستخدم هذا النوع من الانحدار في حالة متغير الاستجابة يكون متغير رتبياً، سوف يتم في هذا البحث بشكل أساس على الأنموذج اللوجستي المتعدد.

2-هدف البحث:

مقارنة بين طريقة الخوارزمية الجينية (GA) وطريقة المربعات الصغرى الموزونة (WLS) لتقدير الأنموذج الانحدار اللوجستي متعدد المستويات.

3-مفهوم الأنموذج الانحدار اللوجستي:

يعد الأنموذج الانحدار اللوجستي هو أنموذج احصائي شائع ينتمي التي نماذج الانحدار اللاخطية الذي يمتاز بالمرونة العالية، ويعرف بانة احد الأساليب الإحصائية المهمة التي تستخدم لتنبؤ باحتمالية وقوع حدث ما وذلك من خلال ملاءمة البيانات والمعلومات على المنحى اللوجستي. حيث يعمل الانحدار اللوجستي عاده لوصف طبيعة العلاقة بين الاستجابة وبين متغير او مجموعة من المتغيرات المستقلة.

4-أنموذج الانحدار اللوجستي المتعدد

يعتبر أنموذج الانحدار اللوجستي متعدد الاستجابة من النماذج الاحصائية ذات أهمية كبيرة في تحليل البيانات المصنفة، حيث يستخدم بصورة عامة في حالة كون متغير الاستجابة يعود الى المتغيرات التي تكون من النوع الاسمي أو الرتبي والمكونة من تصنيفين أو مستويين فأكثر[4].

ويعد أنموذج الانحدار اللوجستي المتعدد الاستجابة امتداد بسيطا لأنموذج الانحدار اللوجستي ثنائي الاستجابة. ويعتمد بصورة الأساسية على التوزيع المتعدد الحدود وان دالة الكثافة الاحتمالية تكتب بصورة الاتية:

$$P_r(r_{i1} = r_{i1}, \ldots\ldots, r_{ij} = r_{ij}) = \binom{ni}{r_{ij},\ldots\ldots,r_{iJ}} \pi_{i1}^{r_{i1}} \ldots \pi_{iJ}^{r_{iJ}} \quad \ldots \quad (1)$$

و الصيغة العامة لأنموذج الانحدار اللوجستي المتعدد الاستجابة كالاتي:

$$= \frac{e^{B_0 + \sum_i^n B_i X_i}}{1 + e^{B_0 + \sum_i^n Bi X_i}} \quad \ldots \quad (2)$$

π_i

$i=1،2\ldots\ldots n$

$B_1، B_2 \ldots\ldots\ldots\ldots\ldots B_i$ معلمات الأنموذج متعدد الاستجابة.

5-طرائق تقدير أنموذج الانحدار اللوجستي

بعد ان تم وصف الأنموذج وتحويله من اللاخطي الى خطي من خلال نماذج التحويل الخاصة به (أنموذج لوغاريتم النسبة المضافة)، بعد ذلك تأتي مرحلة تقدير الأنموذج حيث يتم تقدير الأنموذج ببعض الطرائق التي تكون خاصة بالمتغيرات متعدد المستويات، والتي تعد احد أهم الجوانب في الاستدلال الإحصائي لأغلب النماذج، ومن هذه طرائق التقدير طريقة المربعات الصغرى الموزونة (Weighted least squares Method)، الخوارزمية الجينية (Genetic Algorithm).

1-5طريقة المربعات الصغرى الموزون :

في هذه الطريقة يتم تحويل معادلة المنحني اللوجستي الى معادلة خطية يمكن منها تقدير معلمات الأنموذج المختار. ان معادلة المنحني اللوجستي هي[11]

$$\pi_i=\frac{e^{B_0+B_1X_{1i}+B_2X_{2i}+\cdots\ldots+B_kX_{ki}}}{1+e^{B_0+B_1X_{1i}+B_2X_{2i}+\cdots\ldots+B_kX_{ki}}} \quad \ldots. \quad (3)$$

حيث πتمثل احتمال النجاح الحقيقي والتي يمكن التعبير عنه بالدالة التجمعية (C.D.F) ،وبذلك فان احتمال نجاح العينة المشاهدة P_i هو عبارة عن احتمال النجاح مضافاً اليه الأخطاء العشوائية (u_i)

$$P_i = \pi_i + u_i \quad \ldots. \quad (4)$$

وعلى فرض ان الأخطاء العشوائية u_i مستقلة (independent) و تتوزع توزيعا ذي الحدين (binomial distribution) بمتوسط صفر وتباين $(\frac{\pi_i(1-\pi_i)}{n_i})$ وان $N=\sum n_i$ يمثل المجموع الكلي لعدد الحالات ،فان احتمال فشل استجابة العينة المشاهدة $(1-P_i)$ سيكون

$$1 - P_i = 1 - \pi_i - u_i \quad \ldots.(5)$$

وبقسمة المعادلة(4)على المعادلة (5) نحصل على

$$\frac{P_i}{1-P_i} = \frac{\pi_i}{1-\pi_i}\left(\frac{1+u_i/\pi_i}{1-u_i/Q_i}\right)$$

حيث

$$Q_i = 1 - \pi_i$$

$$\ln\frac{P_i}{1-P_i} = ln\,\frac{\pi_i}{1-\pi_i} + ln\left(\frac{1+u_i/\pi_i}{1-u_i/Q_i}\right) \quad \ldots.(6)$$

حيث $\ln\left(\frac{P_i}{1-P_i}\right)$ يطلق عليها $(\text{logit}(P_i))$ للقيم المشاهدة بينما $\left(ln\,\frac{\pi_i}{1-\pi_i}\right)$ يطلق عليها $(\text{logit}(\pi_i))$ للقيم الحقيقة .

ومن معادلة (3) فان

$$\text{E}(P_i) = \pi_i \qquad ، \text{Var}(P_i)=\frac{\pi_i(1-\pi_i)}{n}$$

أي ان :

$$P_i \sim bino\left(\pi_i, \frac{\pi_i(1-\pi_i)}{n}\right)$$

وللبحث عن علاقة بين القيم الحقيقة والمشاهدة للوغاريتم النسبة المضافة (logit)سنفرض

$$f(P_i)=\ln\left(\frac{P_i}{1-P_i}\right) \qquad ،f(P_i)=\ln\left(\frac{\pi_i}{1-\pi_i}\right)$$

وباستخدام صيغة تايلر من درجة الأولى حول π_i للمعادلة (3)فان :

$$f(P_i)=f(\pi_i+u_i)$$

$$f(P_i)=f(\pi_i)+u_i\frac{\partial f\pi_i}{\partial P_i}$$

وبمقارنة هذه المعادلة بالمعادلة (5) ومساواه المشتقة الى صفر نحصل على :

$$u_i\frac{\partial f\pi_i}{\partial P_i}=ln\left(\frac{1+u_i/\pi_i}{1-u_i/Q_i}\right)$$

وباشتقاق المعادلة (5) نحصل على

$$\frac{\partial f\pi_i}{\partial P_i}=\frac{1}{\left(\frac{\pi_i}{1-\pi_i}\right)}\cdot\frac{(1-\pi_i)(1)-\pi_i(-1)}{(1-\pi_i)^2}$$

$$\therefore \frac{\partial f\pi_i}{\partial P_i}=\frac{1}{\left(\frac{\pi_i}{1-\pi_i}\right)(1-\pi_i)^2}=\left(\frac{1}{\pi_i(1-\pi_i)}\right)$$

وبذلك فان معادلة (5) ستكون:

$$\text{Ln}\frac{P}{1-P}=\ln\frac{\pi_i}{\pi_i}+\frac{u_i}{\pi_i(1-\pi_i)}$$

على فرض ان Logitللقيم الحقيقة هو $y^*{}_i$ أي ان $ln\frac{\pi_i}{1-\pi_i}=\ y^*{}_i$ للقيم الحقيقة فأن

$$y^*{}_i=B_0+\sum_{i=1}^{n}B_iX_i \qquad \ldots. \quad (7)$$

وبذلك تم تحويل المعادلة (3)غير الخطية الى معادلة خطية يمكن تقدير معاملات الانحدار منها وان المتغير الذي تم تحويلة $y^*{}_i$ له متوسط مساو تقريبا الى $\ln\left(\frac{\pi_i}{1-\pi_i}\right)$ بعبارة أخرى مساو الى (X_iB) وتباين $\sigma^2{}_i$ حيث ان:

$$\sigma^2{}_i=\frac{1}{n\pi_i(1-\pi_i)} \qquad \ldots. \quad (8)$$

ان انحدار المتغير الذي تم تحويله $y^*{}_i$ يعتمد على X_i له تباين غير ثابت (غير متجانس) لإنَّه يعتمد على قيمتي $(n،P_i)$ وطالما إنَّ قيمة P_i تعتمد على X_i وب الأتي فإنَّ $y^*{}_i$ يعتمد على X_i بذلك سيفقد أنموذج أغلب افتراضاته

الخاصة ومن هذه الافتراضات سيكون تباين الخطأ العشوائي غير ثابت وب الأَتي قد يعاني من مشكلة عدم تجانس التباين وعندها لا يمكن تقدير معلمات أنموذج باستعمال طريقة المربعات الصغرى الاعتيادية (OLS) وبذلك تم اللجوء الى طريقة المربعات الصغرى الموزونة (WLS) لتقدير معلما أنموذج اللوجستي المتعدد ليتم الحصول على أفضل تقدير. لذلك لابد من ضرب المعادلة الانحدار (7) بالجذر التربيعي للوزن المرافق $\sqrt{W_i}$ لكي نحصل على تجانس للتباين وان[1] :

$$\mathrm{W_i} = n_i P_i (1 - P_i)$$

$$\sqrt{\mathrm{W_i}} = \sqrt{n_i P_i (1 - P_i)}$$

وبذلك تصبح المعادلة (7) بعد ضربها بالوزن كالاتي:

$$\sqrt{W_i} y^*{}_i = B_0 + \sqrt{W_i} X_i B_i + u_i \qquad \ldots\ldots (9)$$

بذلك يمكن تقدير معلمات أنموذج $(\mathrm{B_0}، \mathrm{B_1}، \ldots\ldots، \mathrm{B_p})$ في المعادلة (9) لإيجاد افضل المعالم التي تبحث عن القيم التي تجعل الفرق بين الاستجابة المشاهدة والاستجابة المقدرة أقل ما يمكن أي بمعنى تصغير مجموع مربعات الخطأ الموزون[8].

$$SSE_{\ i} = \sum W_i (Y_i - \hat{Y}_i)^{\,2} \qquad \ldots. (10)$$

$$SSE = \sum W_i \left(Y_i - \hat{B}_0 - \hat{B}_1 X_{1i} - \cdots \hat{B}_1 X_{pi}\right)^{\,2} \qquad \ldots. (11)$$

وبحل التفاضلات الجزئية بالنسبة $(\mathrm{B_0}، \mathrm{B_1}، \ldots\ldots، \mathrm{B_p})$ ومساواة الناتج الى الصفر نحصل على ماياتي:

$$\frac{\partial \mathrm{SSE}}{\partial \mathrm{B_0}} = 0 = \mathrm{B_0} \sum \mathrm{W_i} + \mathrm{B_1} \sum \mathrm{W_i\, X_{1i}} + \cdots + \mathrm{B_p} \sum \mathrm{W_i\, X_{pi}} = \sum \mathrm{W_i\, Y_i}$$

$$\frac{\partial \mathrm{SSE}}{\partial \mathrm{B_1}} = 0 = \mathrm{B_0} \sum \mathrm{W_i\, X_{1i}} + \mathrm{B_1} \sum \mathrm{W_i\, X^{\,2}{}_{1i}} + \cdots + \mathrm{B_p} \sum \mathrm{W_i\, X_{1i} X_{np}} = \sum \mathrm{W_i\ X_{1i} Y_i}$$

$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots$$

$$\frac{\partial SSE}{\partial B_p} = 0 = B_0 \sum W_i\, X_{pi} + B_1 \sum W_i\, X_{pi} X_{1i} + \cdots + B_p \sum W_i\, X^{\,2}{}_{pi} = \sum W_i\ X_{ip} Y_n$$

عند اتباع أسلوب المصفوفات تكتب المعادلات بصورة الشكل الاتي:

$$\underline{y}=\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} ، W=\begin{bmatrix} W_1 & 0 & \dots & 0 \\ 0 & W_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & W_n \end{bmatrix} ،X=\begin{bmatrix} 1 & X_{11} & \cdots & X_{1i} \\ 1 & X_{21} & \cdots & X_{2i} \\ \vdots & \vdots & & \vdots \\ 1 & X_{n2} & \cdots & X^2_{ni} \end{bmatrix} ،B=\begin{bmatrix} B_0 \\ B_1 \\ \vdots \\ B_p \end{bmatrix}$$

$$X'WX=\begin{bmatrix} \sum W_i & \sum W_i X_{i1} & \sum W_i X_{i2} & \cdots & \sum W_i X_{ip} \\ \sum W_i X_{i1} & \sum W_i X^2_{i1} & \sum W_i X_{i1}X_{i2} & & \sum W_i X_{i1}X_{np} \\ \vdots & & \ddots & & \vdots \\ \sum W_i X_{ip} & \sum W_i X_{ip}X_{i1} & \sum W_i X_{ip}X_{i2} & \cdots & \sum W_i X^2_{ip} \end{bmatrix} ، \quad X'WY$$

$$=\begin{bmatrix} \sum W_i Y_i \\ \sum W_i X_{i1} Y_i \\ \vdots \\ \sum W_i X_{ip} Y_n \end{bmatrix}$$

$X'WX\ \beta = \acute{X}W\mathrm{Y}$

حيث أن:

Y: يمثل متجه التحويل الخطي (اللوجت) لنموذج ذو رتبة (nx1)

W: يمثل مصفوفة مربعة عناصر قطرها الرئيسي يمثل الاوزان ذو رتبة (nxn)

$X'WX$: تمثل مصفوفة مربعة حصيلة ضرب مصفوفة الاوزان مع مصفوفة المتغيرات التوضيحية ذو رتبة [(p+1) (p+1)]

XWY تمثل متجه مبدلة مصفوفة متغيرات التوضيحية ومصفوفة الاوزان مع متجه التحويل الخطي ذو رتبة [(p+1)x n]

وبضرب طرفي المعادلة بXWX^{-1} وبذلك سوف نصل الى قيمة$\hat{B}$المقدرة والتي تكون:

$$\hat{B}=(X'WX)^{-1}X'WY \qquad \dots \quad (12)$$

2-5طريقة تقدير الخوارزمية الجينية :

تعد الخوارزمية الجينية أحد أساليب الذكاء الاصطناعي فهي من طرق البحث العشوائي التي تعالج مشكلة ما من اجل التوصل الى أفضل النتائج الممكنة حيث تعتمد على الية الانتقاء الطبيعي ونظام الجينات الطبيعية لذا تمتلك كم هائلا من الحلول البديلة، والحل الناتج من تطبيق الخوارزمية يكون في اغلب الأحيان حلاً قريبا الى الجينات الطبيعية لذلك برزت اهميتها في حل المسائل المعقدة.

وان فكرة عمل الخوارزمية الجينية تعتمد على أفكار الهندسة الوراثية، التي تتميز بالإنتاج المقصود للمجموعات المورثة لكي يتم تكوين افراد جديدة ذات صفات جيدة على هذا الأساس تقوم الخوارزمية الجينية باختيار الحلول الأفضل من بين عدد كبير من الحلول وتقوم بأجراء بعض التبديلات والتدخلات بين هذه الحلول بهدف تكوين حل أفضل فهي طريقة الية.

1-2-5مفهوم الخوارزمية الجينية :

تصنف الخوارزمية الجينية بانها أحد طرق البحث العشوائي حيث برزت تقنيه الخوارزمية الجينية كأداة حقيقة وقوية لإيجاد الحل الأمثل مستخدمة الجينات بالاعتماد على مبدا داروين في التطور الذي ينص على ان كل كائنات الحية على مر الزمان تنحدر من اسلاف مشتركة، ومن خلال تطبيق مبدأ الانتخاب الطبيعي (البقاء للأفضل) يتم الحصول على الافراد ذوي التقييم الأفضل، والقادرين على التكيف مع البيئة المتجددة. وتمتاز الخوارزمية الجينية بسمتين أساسيتين وهما[2]:

- العشوائية: ان القاعدة الأساسية في الخوارزمية الجينية هي العشوائية حيث ان كلاعمليتي اعاده الإنتاج والاختيار تم بطريقة عشوائية.
- فضاء البحث: ان الخوارزمية الجينية تعتمد على فضاء البحث او ما يسمى بالمجتمع الذي يحتفظ عند عملية إعادة الإنتاج وتشكيل مجتمع جديد بالحلول التي تنتج أفضل الصفات[3].

2-2-5 خطوات عمل الخوارزمية الجينى(Steps Work Genetic Algorithms):

هناك عدة خطوات تجرى داخل الخوارزمية الجينية يمكن حصرها بالنقاط الآتية[13],[10]:

1- يتم تغيير الكروموسومات الأصلية بالاعتماد على تقييم الأبناء وأن كل كروموسوم يتكون من عدد من القيم ويتم تحدد عدد هذه القيم حسب المسالة وقيمة البداية، وتكون عشوائية ضمن المحددات من المسالة المراد حلها.

2- يجب ان يتم اختيار أفضل الأباء لكي تتم عملية الإنتاج الابناء.

3- يتم توليد جيل جديد باستخدام الطفرة والعبور.

4- بالاعتماد على الدالة الأصلية يتم تقيم الأبناء الجدد.

مكونات الخوارزمية الجينية:

Components A Genetic Algorithm

تتألف الخوارزمية الجينية من عدة عناصر وهي:

1. **المجتمع (Population)**[2].
2. **دالة اللياقة (التقييم) (Fitness Function (Evaluation)**[6].
3. **التشفير (Encoding)**[11],[8],[7].
4. **الانتقاء (Selection)** [3],[9],[16].
5. **العبور (Crossover)**[3].
6. **الطفرة (Mutation)**[6],[14].

3-2-5 خطوات خطوات إجراء الخوارزمية الجينية لتقدير المعلمات الأنموذج الانحدار اللوجستي المتعدد:

يتم توليد مجتمع ابتدائي بشكل عشوائي يحتوي على N من الكروموسومات وكل كروموسوم يكون طوله (n)إذ أن n تمثل عدد المعلمات الموجودة في الأنموذج النهائي لدالة اللوجستية.

1. يتم استخدام تشفير حقيقي، أي بمعنى يتم تشفير الكروموسوم وذلك من خلال تشفير المعلومات الموجودة ضمن كل كروموسوم لكي يتم تميز الكروموسومات.

2. دالة اللياقة: إن قيمة دالة اللياقة في هذه الخوارزمية تمثل قيمة معيار متوسط مربعات الخطأ.
3. الأنتقاء: هي عملية اختيار الاباء في المجتمع لأجل التزاوج وإنتاج جيل جديد، وبذلك يتم اختيار الآباء ذلك من خلال انتقاء عجلة.
4. العبور: تتم اختيار العبور بنقطتي القطع، أي بمعنى يتم العبور بأجراء تبديل بين قيم متقابلة من كروموسومين الابوين المنتقيين لتكوين الفرد الجديد.
5. الطفرة: تم عملية الطفرة باستخدام الطفرة المنتظمة. أي ان يتم تبديل بين قيم الفرد الواحد لتكوين أفراد تعطي حلولاً جديدة إلى الجيل اللاحق لم يسبق تكوينها في الأجيال السابقة وذلك بهدف توسيع الحلول الممكنة.
6. شرط التوقف: إن المقياس الذي تم اعتماده في الخوارزمية الجينية لتحقيق شرط التوقف في حالة عدم الوصول للحل المثالي أو الأمثل، هو (nx100) حيث أن n تمثل عدد المعلمات في الأنموذج النهائي لدالة اللوجستية.

6 مفهوم المحاكاة:

هي طريقة علمية عددية تعطي القدرة على التجريب والاختيار من خلال تكرار العملية لمرات عديدة. اي بمعنى ان المحاكاة هي عبارة عن محاولة إيجاد صورة طبق الأصل من أي أنموذج ونظام بدون اخذ او الرجوع الى ذلك النظام والأنموذج ذاته. لذا يتم استخدم أسلوب المحاكاة بواسطة مجموعة من العينات ويتم توليد هذه العينات من مجتمع افتراضي بعد ذلك يتم تطبيق النماذج والأساليب الإحصائية والرياضية عليها لغرض اجراء المقارنة بين الطرائق المعتمدة في تقدير الأنموذج المدروس من خلال المقاييس الاحصائية وعندها يتم اتخاذ القرار المناسب على ضوء هذه النتائج. حيث كلما كانت نتائج المحاكاة واقعية ودقيقة كلما كانت التجربة أقرب الى الحقيقة[15].

6-1مراحل بناء ووصف تجربة المحاكاة:

تتضمن مرحلة بناء تجربة المحاكاة أربعة مراحل لغرض تحليل البيانات وهي:

المرحلة الأولى : وهي من اهم المراحل التي تعتمد عليها بقية مراحل المحاكاة حيث يتم اختيار ثلاثة حجوم للعينات تمثل العينات (صغيرة ،متوسط، كبيرة) وكانت على توالي (N=25,75,125) وبتكرار تجربة (R=1000) مرة عند كل حجم ، حيث تم اختيار القيم الافتراضية استناداً للدراسات السابقة القريبة من هذه الدراسة وكذلك من خلال خبرة الطالب بهذا الموضوع كما موضح الجدول الاتي :

جدول (1) يوضح القيم الافتراضية

Model	b_0	b_1	b_2
1	0.5	-0.4	0.3
2	0.2	-0.3	0.4
3	0.6	0.5	-0.2

المرحلة الثانية : وهي المرحلة التي يتم فيها توليد قيم عشوائية للمتغيرات المستقلة تتبع توزيع ثنائي الحدين (Binomial) للنموذج افتراضي الذي يعطي استجابة مناضرة للمتغير المعتمد(y) و يخضع هذا المتغير الى التوزيع متعدد الحدود(Multinomial) .

المرحلة الثالثة : وهي المرحلة التي يتم فيها إيجاد تقدير معلمات أنموذج الانحدار اللوجستي متعدد الاستجابة التي تم التطرق اليها في جانب النظري وفق الطرق التي تم استخدامها وذكرها في الجانب النظري المتمثلة بـ (الشبكات العصبية، الإمكان الأعظم، الخوارزمية الجينية، المربعات الصغرى الموزونة).

المرحلة الرابعة : وهي مرحلة الأخيرة من مراحل وصف تجربة المحاكاة وفي هذه المرحلة تم المقارنة بين طرائق التقدير بعد إيجاد المقدرات في المرحلة الثالثة باستخدام المقاييس الإحصائية المستعملة لغرض الحصول على أفضل طريقة تقدير للنموذج قيد الدراسة ومن هذه المقاييس:

1- معيار متوسط مربعات الخطأ (MSE) وحسب الصيغة الاتية:

$$\mathbf{MSE}=\frac{\mathbf{1}}{\boldsymbol{n-k}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{N}}\left(\boldsymbol{Y_i}-\widehat{\boldsymbol{Y}}_{\boldsymbol{i}}\right)^{\mathbf{2}}$$

حيث ان :

$\widehat{Y}_i$ **: تمثل القيمة التقديرية لمشاهدات**

n : تمثل حجم العينة، K : تمثل عدد المعلمات.

2- معيار متوسط مطلق الخطأ (MAE)وحسب الصيغة الاتية:

$$\mathrm{MAE}=\frac{1}{\mathrm{n}}\sum_{\mathrm{t}=1}^{\mathrm{n}}|\mathrm{e_t}|$$

6-2تحليل نتائج تجربة المحاكاة:

سيتم تحليل وعرض نتائج تجربة المحاكاة والوصول الى افضل مقدرات للنموذج اللوجستي متعدد الاستجابة حسب ثلاثة حالات من القيم الافتراضية التي تم افترضها بالاعتماد على المقاييس الإحصائية و يتم الحصول على كافة نتائج تجربة المحاكاة من خلال استخدم برنامج لغة (R) وحسب طرائق التقدير المستعملة (المربعات الصغرى الموزونةWLS، الخوارزمية الجينية).

الحالة الأولى : من القيم الافتراضية (0.5,-0.4,0.3)

تم اختيار قيم افتراضية لمعالم النموذج بناءاًعلى دمج خبرة الباحثة مع الدراسات السابقة وقد استقرت الباحثة ان تكون قيم هذه المعلمات $(B_0=0.5, B_1=-0.4,\ B_2=0.3)$حيث تم استخدامها في كل من طريقتي التقدير وكما يأتي :

***طرائق التقدير(GA,WLS,MLE):**

تم الحصول على مقدرات معالم أنموذج الانحدار اللوجستي متعدد الاستجابة باستعمال طرائق التقدير المتمثلة بـ(الإمكان الأعظم Maximum Likelihood Method، المربعات الصغرى الموزونة Weighted Least Squares Method، الخوارزمية الجينية Genetic Algorithm) باستخدام برنامج لغة (R)عند حجوم عينات (25,75,125) حسب القيم الافتراضية التي تم افتراضها على توالي (0.5,-0.4,0.3) لمعلمات أنموذج ($B_0, B_1,\ B_2$)وبتكرار التجربة (1000) مرة وكما موضح بالجدول الاتي:

جدول (2) يوضح قيم معلمات الانحدار اللوجستي متعدد الاستجابة المقدرة باستعمال (المربعات الصغرى الموزونة، الخوارزمية الجينية) فضلا عن معايير المفاضلة:

الحجوم	المعلمات / الطرائق	b_0	b_1	b_2	MSE	MAE
25	WLS	-0.1861	-0.1541	0.4390	0.9068	0.7559
	GA	-0.1469	-0.1618	0.4183	0.9002	0.7541
75	WLS	0.2389	-0.2349	0.3055	0.9588	0.7853
	GA	0.2789	-0.2613	0.3045	0.9582	0.7805
125	WLS	0.3387	-0.3589	0.3015	0.9785	0.7897
	GA	0.2224	-0.3382	0.3583	0.9784	0.7887

من خلال النتائج الموضحة في الجدول (2) تم تقدير أنموذج الانحدار اللوجستي متعدد الاستجابة لجميع الطرائق المستعملة(WLS,GA). وحسب الحالة الأولى من القيم الافتراضية (0.5,-0.4,0.3)تبين ان :

عند جميع حجوم العينات (صغيرة، متوسطة، كبيرة)(25،75،125) ان الأفضلية كانت لطريقة الخوارزمية الجينية اعتماداً على المقاييس الإحصائية (MSE) و(MAE).

الحالة الثانية : من القيم الافتراضية(0.2,-0.3,0.4)

تم اختيار قيم افتراضية لمعالم النموذج بناءاًعلى دمج خبرة الباحثة مع الدراسات السابقة وقد استقرت الباحثة ان تكون قيم هذه المعلمات $(B_0 = 0.2, B_1 = -0.3, B_2 = 0.4)$حيث تم استخدامها في كل من طريقتي التقدير وكما يأتي :

***طرائق التقدير(GA, WLS)**

تم الحصول على مقدرات معالم أنموذج الانحدار اللوجستي متعدد الاستجابة باستعمال طرائق التقدير المتمثلة بـ(المربعات الصغرى الموزونة Weighted Least Squares Method، الخوارزمية الجينية Genetic Algorithm) باستخدام برنامج لغة (R)عند حجوم عينات (25,75,125) حسب القيم الافتراضية التي تم افتراضها على توالي(0.2,-0.3,0.4) لمعلمات أنموذج (B_0, B_1, B_2) وبتكرار التجربة (1000) مرة وكما موضح بالجدول الاتي:

جدول (3) يوضح قيم معلمات الانحدار اللوجستي متعدد الاستجابة المقدرة باستعمال (المربعات الصغرى الموزونة، الخوارزمية الجينية) فضلا عن معايير المفاضلة:

الحجوم	المعلمات / الطرائق	b_0	b_1	b_2	MSE	MAE

25	WLS	-0.1971	-0.1731	0.5590	0.8990	0.7449
	GA	-0.1589	-0.1917	0.3193	0.8003	0.7336
75	WLS	0.3369	-0.3889	0.5698	0.9577	0.7815
	GA	0.3478	-0.3995	0.5743	0.9373	0.7812
125	WLS	-0.0674	-0.1634	0.4916	0.9811	0.7884
	GA	-0.0987	-0.1600	0.5055	0.9804	0.7883

من خلال النتائج الموضحة في الجدول (3) تم تقدير أنموذج الانحدار اللوجستي متعدد الاستجابة لجميع الطرائق المستعملة(WLS,GA). وحسب الحالة الأولى من القيم الافتراضية (0.2,-0.3,0.4) تبين ان :

عند جميع حجوم العينات (صغيرة، متوسطة، كبيرة)(25،75،125) ان الأفضلية كانت لطريقة الخوارزمية الجينية اعتماداً على المقاييس الإحصائية (MSE) و(MAE).

الحالة الثانية : من القيم الافتراضية(0.6,0.5,-0.2)

تم اختيار قيم افتراضية لمعالم النموذج بناءاًعلى دمج خبرة الباحثة مع الدراسات السابقة وقد استقرت الباحثة ان تكون قيم هذه المعلمات $(B_0 = 0.2, B_1 = -0.3,\ B_2 = 0.4)$حيث تم استخدامها في كل من طريقتي التقدير وكما يأتي:

***طرائق التقدير(GA,WLS,MLE):**

تم الحصول على مقدرات معالم أنموذج الانحدار اللوجستي متعدد الاستجابة باستعمال طرائق التقدير المتمثلة بـ(الإمكان الأعظم Maximum Likelihood Method، المربعات الصغرى الموزونة Weighted Least Squares Method، الخوارزمية الجينية Genetic Algorithm) باستخدام برنامج لغة (R)عند حجوم عينات (25,75,125) حسب القيم الافتراضية التي تم افتراضها على توالي (0.6,0.5,-0.2) لمعلمات أنموذج ($B_0, B_1,\ B_2$) وبتكرار التجربة (1000) مرة وكما موضح بالجدول الاتي:

جدول (4) يوضح قيم معلمات الانحدار اللوجستي متعدد الاستجابة المقدرة باستعمال (المربعات الصغرى الموزونة، الخوارزمية الجينية) فضلا عن معايير المفاضلة:

الحجوم	المعلمات / الطرائق	b_0	b_1	b_2	MSE	MAE
25	WLS	10.181	101.246	-12.504	0.900	0.755
	GA	-9.973	0.315	0.278	0.880	0.740
	WLS	0.507	-0.390	0.280	0.943	0.772

75	GA	0.404	-0.315	0.278	0.942	0.772
125	WLS	0.936	-0.546	0.238	0.987	0.794
	GA	0.918	-0.535	0.231	0.977	0.794

من خلال النتائج الموضحة في الجدول (4) تم تقدير أنموذج الانحدار اللوجستي متعدد الاستجابة لجميع الطرائق المستعملة(WLS,GA). وحسب الحالة الأولى من القيم الافتراضية (0.6,0.5,-0.2) تبين ان :

عند جميع حجوم العينات (صغيرة، متوسطة، كبيرة)(25،75،125) ان الأفضلية كانت لطريقة الخوارزمية الجينية اعتماداً على المقاييس الإحصائية (MSE) و(MAE)

7 -الاستنتاجات:

1- إمكانية الاستفادة من تطبيق أنموذج الانحدار اللوجستي متعدد الاستجابة على البيانات المصنفة لدراسة تأثير جميع العوامل في أن واحد.
2- اتضح من خلال الجانب التجريبي تفوق طريقة الخوارزمية الجينية على طريقة المربعات الصغرى الموزونة المستعملة في تقدير أنموذج الانحدار اللوجستي متعدد الاستجابة اعتمادا على المقاييس الإحصائية (MSE)و(MAE).
3- اتضحت امكانية تحويل أنموذج الانحدار اللوجستي متعدد الاستجابة اللاخطي الى أنموذج خطي وان متوسط مربعات الخطأ تزداد بزيادة عدد المعالم، وبالتالي يمكن تطبيق الأساليب الإحصائية التي تنطبق على النماذج الخطية الأخرى.
4- اتضح من خلال الجانب التجريبي عند جميع حجوم العينات (كبيرة - متوسطة - صغيرة) تفوق طريقة تقدير الخوارزمية الجينية على المربعات الصغرى الموزونة في تقدير أنموذج الانحدار اللوجستي متعدد الاستجابة اعتمادا على المقاييس الإحصائية متوسط مربعات الخطأ(MSE)و متوسط الخطأ المطلق(MAE).
5- نستنتج من خلال نتائج الجانب التجريبي بأن لحجم العينة تأثيراً واضحاً حيث نلاحظ كلما كانت حجوم العينات صغيرة كلما كانت قيمة متوسط مربعات الخطأ (MSE) وقيمة متوسط الخطأ المطلق (MAE) أقل من الحجوم الكبيرة.

8-التوصيات

1. نوصي باستخدام طريقة الخوارزمية الجينية لتقدير الأنموذج اللوجستي متعدد الاستجابة لكونها تعطي نتائج تقدير أفضل من طريقة المربعات الصغرى الموزونة.
2. نوصي باستعمال طرائق تقدير ومعاير مفاضلة أخرى على الأنموذج اللوجستي متعدد الاستجابة.
3. نوصي باستعمال نماذج جديدة غير النماذج اللوجستية مثلا نماذج التربيعية ونماذج التمييزية عند تطبيقهما على البيانات المصنفة .

المصادر العربية :

1- الطائي ، عبد الحسين حسن ،2000**"تقدير وتحليل معادلة الانحدار المتعدد في حال كون المتغيرات المعتمدة وصفية ومحددة"** أطروحة دكتوراه، كلية الإدارة والاقتصاد، جامعة بغداد.

المصادر الأجنبية:

2 - Bala، J.، Huang، J. ، Vafaie، H.، Dejony، k. and Wechsler، H.، (1995)، " Hybrid learning using Genetic Algorithms and decision trees for pattern classification"، IJCAI Conference، Montreal، PP.19،25.

3- Davis، R.، Lee، T.، and Rodriguez، G.، (2006)،" Structural Break Estimation for Nonstationary Time Series Models"، Journal of the American Statistical Association، Vol. 101، No. 473، PP. 229–239.
4- El‹Habil‹ A. M. (2012). An application on multinomial logistic regression model. Pakistan journal of statistics and operation research‹ 271‹291.
5- Goldberg ، D.، (1989)،" Genetic Algorithm in Search Optimization & Machine Learning"، Addison‹Wesley Publishing company، Inc.
6- Guo، C. and Yang، X.، (2011)،" A Programming of Genetic Algorithm in Matlab7.0"، Modern Applied Science، Vol. 5، No.1
7- Haupt، R. and Haupt، S.، (2004)،" Practical genetic Algorithms"، John Wiley and Sons، Inc.
8- Hosmer Jr، D. W.، Lemeshow، S.، & Sturdivant، R. X. (2015). Applied logistic regression (Vol. 398). John Wiley & Sons.
9- Jones، A.، (1993)،" Genetic Algorithms and Their Application to The Design of Neural Network"، Neural Computing and Application، Vol.1، No.1، PP.32‹45.
10- Mitra، S.، Pal، S. and Mitra، p.، (2002)،"Data Mining in Soft Computing Framework: A Survey"، IEEE Transactions on neural networks، Vol. 13، No. 1، PP.3‹14.
11- Muller، Marlene ، (2004) ،" Generalized Linear Models"، Fraunhofer Institute for Industrial Mathematics (ITWM) ، (Germany) ،WWW. Marlenmuller . ed / publication / hand book CS. Pdf.
12- Pao، Y.، (1989)،"Adaptive Pattern Recognition and Neural Networks"، Publishing Company، Inc.، New York.
13- Prebys، E.، (2007)،" The Genetic Algorithm in Computer Science"، MIT Undergraduate Journal of Mathematics، PP.165‹170.
14- Ruan، D. ، (1997)، "Intelligent Hybrid Systems : Fuzzy Logic، Neural Networks and Genetic Algorithms"، Springer Science Business Media New York.
15- Shreider A.(1966) "The Monte Carlo Method‹the method of statistics"Volume 2‹fourth edition‹London.
16-Sivanandam، S. and Deepa، S.، (2008)،"Introduction to Genetic Algorithms"، Springer Berlin Heidelberg New York.