# The use of the Biz method and classical methods in estimating the parameters of the binary logistic regression model

# استعمال طريقة بيز و الطرائق الكلاسيكية في تقدير معلمات انموذج الأنحدار اللوجستي الثنائي

أ.م.د. رباب عبد الرضا صالح البكري / جامعة بغداد / كلية الادارة والاقتصاد rnah_2008@yahoo.com
الباحث/ ساره عادل مظلوم الرديني sarahstatistic88@gmail.com

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

**Received:5/12/2018**
**Accepted: 22/1/2019**

## المستخلص

يعد انموذج الانحدار اللوجستي الثنائي الاسلوب المستعمل في تصنيف البيانات وهو الاداة الاكثر قوة ومرونة في حالات دراسة متغير الاستجابة الثنائي عند مقارنته بالانحدار الخطي ، في هذا البحث تم استعمال بعض الطرائق الكلاسيكية لتقدير معلمات انموذج الانحدار اللوجستي الثنائي تضمنت طريقة تقدير الامكان الاعظم (MLE) وطريقة تقدير تصغير مربع كاي (MCSE) وطريقة تقدير المربعات الصغرى الموزونة (WLSE) مع طريقة تقدير بيز (BE)، من أجل اختيار الطريقة الأفضل في التقدير وذلك من خلال القيم الافتراضية لتقدير المعلمات وفق انموذجين مختلفين من نماذج الانحدار الخطي العام وبأحجام عينات مختلفة ( $N = 50,100,150,200,250,300$ ), وبناء تجربة المحاكاة الخاصة بالبحث ومن ثم عرض نتائجا وتحليلها لغرض الوصول الى أفضل الطرائق الاعتيادية باستعمال المعيار الاحصائي متوسط مربعات الخطأ Mean Square Error(MSE) لمقدرات الانموذج اللوجستي الثنائي لغرض المقارنة بين أفضلية طرائق تقدير معلمات الأنموذج، وقد تم التوصل بشكل عام الى أن طريقة المربعات الصغرى الموزونة (WLS) هي الأفضل بالمرتبة الأولى من بين طرائق التقدير الاعتيادية، وذلك لأنها تمتلك اقل (MSE) للمقدرات, وهدا يدل على دقة الطريقة (WLS) في تقدير معلمات الأنموذج.

**المصطلحات الرئيسة للبحث**/ انموذج الانحدار اللوجستي الثنائي ، طريقة الامكان الاعظم ،خوارزمية نيوتن رافسون ، طريقة تصغير مربع كاي، طريقة المربعات الصغرى الموزونة، طريقة بيز ،خوارزمية جبس.

*البحث مستل من رسالة الماجستير

Journal of Economics and Administrative Sciences
2019; Vol. 25, No.113
Pages: 543- 556

## 1- المقدمة (Introduction)

ان الهدف الاساسي هو تحليل العلاقة بين مجموعة من المتغيرات و الحصول على افضل الصيغ لوصف الانموذج، في حالة ان يكون متغير الاستجابة من النوع المتقطع لا يمكننا تطبيق الانموذج الخطي وانما نلجأ الى نماذج بديلة أخرى، نذكر منها انموذج الانحدار اللوجستي وهو من النماذج الاحصائية المهمة وهو حالة خاصة من حالات نماذج الانحدار الخطية العامة (generalized linear regression models) ويسمى احياناً انموذج لوجستي (logistic model) او انموذج لوجت (logit model) [12;p.1] والذي لا يتطلب افتراضات كثيرة لا يوجد تعدد خطي بين المتغيرات التوضيحية وان تكون حجم المشاهدات في كل مجموعة اكبر من خمس مرات عدد المعلمات[17].

ويكون على عدة انواع: الانحدار اللوجستي الثنائي (Binary Logistic Regression) ويستعمل هذا النوع عندما يأخذ متغير الاستجابة قيمتين مثلا المصاب بالمرض(حدوث الوفاة ،عدم حدوث الوفاة) والانحدار اللوجستي المتعدد (Multinomial Logistic Regression) يتم استعماله عندما يكون متغير الاستجابة متعدد القيم (اكثر من قيمتين) مثلا ممارسة الرياضة (دائما، احيانا ،نادرا) والانحدار اللوجستي الترتيبي(Ordinal Logistic Regression) وفيه يكون متغير الاستجابة رتبوي مثلا مرحلة الدراسة الابتدائية (الاول ، الثاني، ... ،السادس) وفي هذا البحث سنركز على النوع الاول[6;p.13].

وتوجد عدة طرائق لتقدير معلماته منها طريقة تقدير الامكان الاعظم (MLE) ويتم الحصول على معادلات الامكان بأخذ المشتقات من الدرجة الاولى لمعلمات دالة الامكان في الانموذج اللوجستي الثنائي واستعملنا 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 وكذلك استعملنا طريقة المربعات الصغرى الموزونة (WLSE)التي لا تحتاج الى قيم اولية وايضاً تم استعمال طريقة بيز(BE) في تقدير المعلمات كطريقة حديثة بأسلوب معاينة جبس (Gibbs)، لتشكيل افضل مجموعة حل للمعلمة التي من شأنها تقليل مقدار الخطأ الى اصغر ما يمكن وتعظيم دالة الاحتمال اللوغارتمي.

سيتم الاعتماد على الطرائق اعلاه والمقارنة بينها باستعمال المحاكاة في برنامج (MATLAB) للحصول على افضل طريقة في تقدير المعلمات.

## 2- هدف البحث (Research Aim)

ان الهدف من هذا البحث هو الحصول على افضل المقدرات لأنموذج الانحدار اللوجستي الثنائي من خلال استعمال طرائق التقدير الاعتيادية (الامكان الاعظم Maximum Likelihood Estimation (MLE) وتصغير مربع كاي Minimum Chi − Square Estimation (MCS) والمربعات الصغرى الموزونة (Weighted Least Squares Estimation (WLSE) طريقة بيز Bayes Estimation (BE) ، والمقارنة بين هذه الطرائق من خلال افتراض عدد من النماذج خلال المحاكاة باستعمال المعيار الاحصائي متوسط مربعات الخطأ(MSE) لغرض المقارنة والتوصل الى الطريقة الأفضل.

## 3- انموذج الانحدار اللوجستي الثنائي (Binary Logistic Regression (BLRM) Model)

يعد هذا الانموذج من نماذج الانحدار اللاخطي ويتصف بان متغير الاستجابة ($Y$) يتبع توزيع برنولي (Bernoulli distribution) يأخذ القيم ($0$) و ($1$) [4;p.40] اي ان متغير الاستجابة ($Y$) 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[14;p.3]:

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

$$P(Y=1/X_i)=\pi(X_i) \qquad \dots(2)$$
$$P(Y=0/X_i)=1-\pi(X_i) \qquad \dots(3)$$

وان توقع متغير الاستجابة يمثل احتمال النجاح $E(Y_i)=\pi(X_i)$ وايضا تباين متغير الاستجابة $v(Y_i)=\pi(X_i)(1-\pi(X_i))$

اما اذا كان انموذج الانحدار اللوجستي يحتوي على اكثر من متغير توضيحي واحد، فيعبر عن توقعه الشرطي لاحتمال متغير الاستجابة (وقوع الحدث) حسب الصيغة الرياضية الاتية[2;p.24] :-

$$\pi(X_i)=\frac{e^{\underline{X}_i'\underline{\beta}}}{1+e^{\underline{X}_i'\underline{\beta}}} \qquad \dots(4)$$

وان احتمال عدم وقوع الحدث هو:

$$1-\pi(X_i)=\frac{1}{1+e^{\underline{X}_i'\underline{\beta}}} \qquad \dots(5)$$

اذ ان:

$\pi(X_1)$: هو التوقع الشرطي $E(Y/X)$ (conditional mean) لمتغير الاستجابة $(Y)$ عند قيمة معينة لـ $(X)$ (احتمال حدوث الاستجابة).

$1-\pi(X_1)$: هو احتمال عدم حدوث الاستجابة.

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

$\beta_\circ,\beta_1,\dots,\beta_p$: متجه للمعلمات المراد تقديرها وعددها **(P)**.

ويتم تحويل هذا الانموذج الى شكل خطي يتمثل بعلاقة خطية من خلال الموجه الصفي $(\acute{X}_i)$ من المتغيرات التوضيحية مع لوجت الاحتمال $[logit\,\pi(X_i)]$ وحسب الصيغة الرياضية الاتية[3;p.166]:

$$logit\,\pi(X_i)=ln\left[\frac{\pi(X_i)}{1-\pi(X_i)}\right] \qquad \dots(6)$$
$$=\beta_\circ+\beta_1X_{i1}+\dots+\beta_pX_{ip} \qquad \dots(7)$$
$$=[1\quad X_{i1}\quad \dots\quad X_{ip}]\begin{bmatrix}\beta_\circ\\ \beta_1\\ \vdots\\ \beta_p\end{bmatrix}=\underline{X}_i'\underline{\beta} \qquad \dots(8)$$

اذ ان:

$n$:عدد المشاهدات. ; $i=1,2,\dots,n$

$p$:عدد المتغيرات التوضيحية والمعلمات المجهولة. ; $j=0,1,\dots,p$

## 4- طرائق تقدير معلمات انموذج الانحدار اللوجستي الثنائي

**(Methods Estimating Parameters The Model Regression Logistic Binary)**

**1-4 طريقة الامكان الاعظم (Maximum Likelihood Method (MLM))** للحصول على تقديرات المعلمات 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 الانموذج اللوجستي الثنائي لـ n من المشاهدات هي[9;p.9]:

$$l(\beta)=P(Y/X)=\prod_{i=1}^{n}[\pi(X_i)]^{Y_i}[1-\pi(X_i)]^{1-Y_i} \qquad \dots(9)$$

وبأخذ اللوغارتم $(Log)$ للصيغة (9) وكالاتي:

$$ln[l(\beta)] = \sum_{i=1}^{n} Y_i ln[\pi(X_i)] + (1 - Y_i) ln[1 - \pi(X_i)] \quad ...(10)$$

$$= \sum_{i=1}^{n} \left[ Y_i ln\left(e^{\acute{X}_i \underline{\beta}}\right) - Y_i ln\left(1 + e^{\acute{X}_i \underline{\beta}}\right) + ln\left(\frac{1}{1 + e^{\acute{X}_i \underline{\beta}}}\right) - Y_i ln\left(\frac{1}{1 + e^{\acute{X}_i \underline{\beta}}}\right) \right] \quad ...(11)$$

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

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

$$\acute{L}\left(\underline{\beta}\right) = \sum_{i=1}^{n} \left[ \left( Y_i - \frac{e^{\underline{X}_i \underline{\beta}}}{1 + e^{\underline{X}_i \underline{\beta}}} \right) \acute{X}_{ij} \right] = 0 \quad ...(13)$$

نستنتج بانه تكونت لدينا $(p + 1)$ من المعادلات غير الخطية ونقوم بحلها بإحدى الطرائق التكرارية التقليدية كطريقة نيوتن رافسون $(NR)$،وان خوارزمية نيوتن رافسون التكرارية لإيجاد قيم (β) التقديرية لدالة الامكان الاعظم في الانموذج اللوجستي ستكون في $(m + 1)$ من التكرارات وكالآتي[5;pp.36,42]:

$$\hat{\beta}^{(m+1)} = \hat{\beta}^{(m)} + (\acute{X}\, V^{(m)}\, X)^{-1}\, \acute{X}\, (Y - P^{(m)}) \quad ...(14)$$

$$V^{(m)} = diag\, [\pi_i{}^{m}(1 - \pi_i{}^{m})] \quad ...(15)$$

اذ ان:

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

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

$Y$: يمثل متجه متغير الاستجابة ذو رتبة $(n * 1)$ للتكرار $(m)$.

$P^{(m)}$: 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 السابق $(m)$.

## 2-4 طريقة تصغير مربع كاي *(Minimum Chi - Square Method ) (MCSM)*

وهي من الطرائق الشائعة الاستعمال في التقدير وتعتمد على تصغير احصاءة مربع كاي لبيرسون المعروفة حسب الصيغة الرياضية الآتية:[10;p.2]

$$x^2 = R(\beta) = \sum_{i=1}^{N} \frac{(\theta_i - \varepsilon_i)^2}{\varepsilon_i} \quad ...(16)$$

اذ ان:

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

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

$R\left(\underline{\beta}\right)$: تمثل احصاءة مربع كاي لبيرسون

وفي حالة الانموذج اللوجستي الثنائي الاستجابة فإن:[9;p.136]

$$R\left(\underline{\beta}\right) = \sum_{i=1}^{n} \frac{(Y_i - \pi_i)^2}{\pi_i} + \frac{[(1 - Y_i) - (1 - \pi_i)]^2}{1 - \pi_i} \quad ...(17)$$

ويتم اختصارها الى:

$$R\left(\underline{\beta}\right) = \sum_{i=1}^{n} \frac{(Y_i - \pi_i)^2}{\pi_i (1 - \pi_i)} \quad ...(18)$$

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

$$R\left(\underline{\beta}\right) = \sum_{i=1}^{n}\left[\left(Y_i - \frac{e^{\underline{X}_i\underline{\beta}}}{1+e^{\underline{X}_i\underline{\beta}}}\right)^2 \frac{\left(1+e^{\underline{X}_i\underline{\beta}}\right)^2}{e^{\underline{X}_i\underline{\beta}}}\right] \quad \dots(19)$$

وبعد التبسيط تصبح المعادلة كالآتي:

$$R\left(\underline{\beta}\right) = \sum_{i=1}^{n}\left[Y_i^2\, e^{-\underline{\acute{X}_i\beta}} + (1-Y_i)^2\, e^{\underline{\acute{X}_i\beta}} - 2Y_i(1-Y_i)\right] \quad \dots(20)$$

لإيجاد $\underline{\beta}$ التي تعطي اقل $R\left(\underline{\beta}\right)$ يتطلب ايجاد المشتقة الاولى ومساواتها بالصفر كما يأتي:

$$\frac{\partial R\left(\underline{\beta}\right)}{\partial \underline{\beta}} = \sum_{i=1}^{n} \acute{X}_{ij}\left[(1-Y_i)^2\left(\frac{\pi_i}{1-\pi_i}\right) - (Y_i)^2\left(\frac{1-\pi_i}{\pi_i}\right)\right] \quad \dots(21)$$

وهي علاقة غير خطية وعند حلها يتطلب استعمال احدى الطرائق التكرارية كما هو الحال مع طريقة الامكان الاعظم نستعمل طريقة نيوتن _ رافسون لإيجاد تقديرات $\boldsymbol{\beta}$ حسب الصيغة **(14)** وان [2;p.26]:

$$V^{(m)} = diag\left[(Y_i)^2\left(\frac{1-\pi_i}{\pi_i}\right) + (1-Y_i)^2\left(\frac{\pi_i}{1-\pi_i}\right)\right] \quad \dots(22)$$

## 3-4 طريقة المربعات الصغرى الموزونة (Weighted Least Squares Method(WLSM))

يعتبر انموذج الانحدار اللوجستي الثنائي حالة خاصة من النماذج الخطية العامة التي تكون امتداد للانموذج الخطي البسيط[13;p.1] .

يمكن كتابة انموذج الانحدار الخطي العام كالآتي[11;p.76]:

$$y_i = \beta_\circ + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi} + \varepsilon_i \quad \dots(23)$$

$$y_i = \beta_\circ + \sum_{j=1}^{p} \beta_j x_{ij} + \varepsilon_i \qquad ,i = 1,2,\dots,n\ ;\ j = 1,2,\dots,p \quad \dots(24)$$

ويتم تقدير معلمات نموذج الانحدار اللوجستي الثنائي باستعمال طريقة المربعات الصغرى الموزونة ($WLS$) للحصول على افضل تقدير.

$$W_i = \pi_i(1-\pi_i) \quad \dots(25)$$

اذ ان:

$W_i$ : تمثل مصفوفة التباينات وهي الاوزان المختارة للمستوى $\boldsymbol{i}$

فيكون تقدير المعلمات $(\beta_\circ, \beta_1, \dots, \beta_p)$ حسب ( $WLS$) بإيجاد القيم التي تجعل الفرق بين الاستجابة المشاهدة والاستجابة المقدرة اقل ما يمكن اي تصغير مجموع مربعات الخطأ ($SSE$) [9;p.135].

$$sse_i = \sum w_i\left(Z_i - \hat{Z}_i\right)^2 \quad \dots(26)$$

$$Z_i = ln\frac{\pi_i}{1-\pi_i} \quad \dots(27)$$

$$sse_i = \sum w_i\left(Z_i - \hat{\beta}_\circ - \hat{\beta}_1 x_{1i} - \cdots - \hat{\beta}_p x_{pi}\right)^2 \quad \dots(28)$$

اذ ان:

Z: يمثل متجه التحويل الخطي (اللوجت) للنموذج ذو رتبة ($n*1$).

W: تمثل مصفوفة مربعة عناصر قطرها الرئيسي يمثل التباينات ذو رتبة **(n*n)**.

وبذلك نصل الى قيمة $\widehat{\beta}$ المقدرة والتي تكون[2;p.20]:

$$\underline{\widehat{\beta}} = (\acute{X}WX)^{-1}\,\acute{X}WZ \quad \dots(29)$$

**4-4 طريقة بيز (Bayes Method (BM ))**

ان الاسلوب البيزي يتكون من ثلاث اقسام هي التوزيع السابق والتوزيع اللاحق ودالة الامكان الاعظم ووفقاً لهذا الاسلوب يمكن كتابة دالة التوزيع اللاحق لمعلمات الأنموذج اللوجستي حسب الصيغة الرياضية الآتية[8;pp.193-197]:

$$\pi(\beta \backslash Y,Z) = f(Y \backslash \beta,Z)\, f(\beta \backslash Z) \qquad \ldots(30)$$

أو

$$\pi(\beta \backslash Y,Z) = l(\beta)\pi(\beta) \qquad \ldots(31)$$

اذ ان:

$\pi(\beta \backslash Y,Z)$: تمثل دالة التوزيع اللاحق للمعلمات المراد تقديرها.

$l(B) = f(Y \backslash \beta,Z)$: تمثل دالة الامكان الاعظم وقد ذُكرت مسبقاً.

$\pi(\beta) = f(\beta \backslash Z)$: تمثل دالة التوزيع السابق للمعلمات.

ويكون التوزيع المشروط لجميع المعلمات $(\beta_j)$ مع $(Z)$ هو التوزيع المنتظم ($Uniform$ $Distribution$) كما في الصيغة الرياضية الآتية[7;p.860]:

$$\beta_j \backslash \beta_{j-p}, Y, Z \sim U(a_j, b_j) \qquad ; j = 0,1,\ldots,P \qquad \ldots(32)$$

اذ ان:

$U(a_j, b_j)$: تمثل التوزيع المنتظم للفترة $(a)$ و $(b)$

$$a_j = \max\left[\frac{1}{X_{ip}}\log\left(\frac{Z_i}{1-Z_i}\right) - \sum_{j\neq p}^{N} BX_{ij}\right] \qquad \ldots(33)$$

$$b_j = \min\left[\frac{1}{X_{ip}}\log\left(\frac{Z_i}{1-Z_i}\right) - \sum_{j\neq p}^{N} BX_{ij}\right] \qquad \ldots(34)$$

تطبق معاينة جبس من خلال التوزيع الشرطي لـ $(Z_i)$ وبالتكرار يمكن توليد بيانات عشوائية من التوزيع المنتظم حسب الصيغة الرياضية الآتية[7;p.860]:

$$Z_i \backslash \beta, Y \sim U \begin{Bmatrix} \left(0, \dfrac{e^{X_i'\beta}}{1+e^{X_i'\beta}}\right) & if\ Y_i = 1 \\ \left(\dfrac{e^{X_i'\beta}}{1+e^{X_i'\beta}}, 0\right) & if\ Y_i = 0 \end{Bmatrix} ; i = 1,2,\ldots,N \qquad \ldots(35)$$

يتم توليد $(\overset{(i)}{\beta})$ باستعمال قيم اولية $(\overset{(\circ)}{Z})$ في الصيغة $(32)$ ليكتمل تكرار واحد ثم توضع في المعادلة $(35)$ لتوليد $(\overset{(1)}{Z})$ وبعد $(n)$ من التكرارات يتم تحديث التوزيعات الشرطية لتوليد $(\overset{(n)}{Z}, \overset{(n)}{\beta})$ الى ان نصل لمرحلة الاستقرار ، وان التوزيع المشترك للعينة يقترب بمعدل أسي من التوزيع اللاحق المشترك الأصلي $P(\beta \backslash Z, Y)$ كلما زاد عدد مرات توليد عينات جديدة الى ما لانهاية $(n \to \infty)$[16;pp.79-86] .

فنحصل على قيم $(\beta)$ التقديرية بأخذ متوسط العينة المسحوبة (المتولدة) من التوزيع اللاحق حسب الصيغة الرياضية الآتية[15;p.169]:

$$\overset{(i)}{\beta} \sim \hat{\pi}_i(\beta \backslash Y) = \frac{1}{N}\sum_{j=1}^{N} P\left(\beta / Y, \overset{(i-1)}{Z_j}\right) \qquad \ldots(36)$$

## 5- المحاكاة (Simulation)

تتضمن تجربة المحاكاة عدة مراحل وهي :

**1-5 المرحلة الأولى- تحديد القيم الافتراضية:**

يتم خلال هذه المرحلة تحديد القيم الافتراضية للمعلمات $(P = 11)$، وفيها تم تحديد ستة احجام مختلفة للعينات، وبذلك تكون عدد المشاهدات الكلية ( $N = 50,100,150,200,250,300$ ) مشاهدة على التوالي، وتعد هذه المرحلة من أهم المراحل التي تعتمد عليها بقية مراحل المحاكاة، وان هذه القيم تم تحديدها في الانموذج الاول من تقديرات المعلمات الابتدائية للبيانات الحقيقية قيد البحث لأنموذج الانحدار الخطي العام اذ ان انموذج الانحدار اللوجستي حالة خاصة منه ثم اخذنا ( Standardized) للمعلمات في الانموذج الثاني وفق البرنامج الجاهز (SPSS) وكما في الجدول(1) :

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

| Para. \ Mod. | $\beta_0$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ | $\beta_5$ | $\beta_6$ | $\beta_7$ | $\beta_8$ | $\beta_9$ | $\beta_{10}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | -1.031 | −0.073 | 0.095 | −0.055 | −0.000415 | 0.013 | −0.002 | 0.816 | −0.000039 | 0.001 | 0.004 |
| 2 | -1.031 | −0.088 | 0.091 | −0.046 | −0.013 | 0.060 | −0.004 | 0..803 | −0.002 | 0..150 | 0.092 |

وبالتالي تكون النماذج المفترضة:

$$y_1 = -1.031 + -0.073x_{1i} + 0.095\, x_{2i} + \cdots + 0.004x_{pi}$$
$$y_2 = -1.031 + -0.088x_{1i} + 0.091\, x_{2i} + \cdots + 0.092x_{pi}$$

**2-5 المرحلة الثانية- توليد البيانات:**

فيها يتم توليد عشرة متغيرات توضيحية$(X = 10)$ كما في الجانب التطبيقي باستعمال اسلوب مونت كارلو وذلك من خلال التوزيع المنتظم (Uniform distribution)، اما في ما يتعلق بقيم متغير الاستجابة فيتم احتسابه تحت افتراض انه يتبع توزيع برنولي Bernoulli distribution [0,1] وتحديد القيم الاحتمالية حسب طريقة الرفض والقبول وكما يلي:

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

**3-5 المرحلة الثالثة- إيجاد التقديرات**

يتم في هذه المرحلة تقدير معلمات انموذج الانحدار اللوجستي الثنائي بحسب الصيغة (4) وفق الطرائق الكلاسيكية وهي [طريقة تقديرات الامكان الاعظم $(MLE)$ و طريقة تقديرات اصغر مربع كاي $(MCSE)$ و طريقة تقديرات المربعات الصغرى الموزونة $(WLSE)$ ]، و طريقة تقديرات بيز $(BE)$.

**4-5 المرحلة الرابعة- المقارنة بين طرائق التقدير:**

يتم في هذه المرحلة المقارنة بين طرائق التقدير الاعتيادية (الكلاسيكية) وطريقة بيز لمعلمات انموذج الانحدار اللوجستي الثنائي، وذلك باستعمال متوسط مربعات الخطأ للأنموذج ومن هي الافضل لكافة الطرائق، وقد تم تكرار تجربة المحاكاة عدد $(1000)$ للحصول على النتائج وبحسب الصيغة الاتية[1;p.433] :

$$MSE = \frac{1}{R}\sum_{i=1}^{R} MSE_i = \frac{1}{R}\sum_{i=1}^{R}\left[\frac{1}{N-P}\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2\right] \quad \ldots (37)$$

اذ ان:

$R$:تمثل عدد مرات تكرار تجربة المحاكاة.

$[(\hat{Y}_i) = \hat{\pi}(X_i)]$:يتم الحصول عليه من تقدير احتمال الاستجابة في الصيغة(4).

## 6- تحليل نتائج المحاكاة (Analysis of Simulation Results)

تم الحصول على كافة نتائج عملية المحاكاة باستعمال برنامج كتب بلغة [MATLAB,2017]، وتم عرض هذه النتائج في الجداول كالاتي:

جدول (2) متوسط مربعات الخطأ (MSE) لمقدرات انموذج الانحدار اللوجستي الثنائي في الانموذج الاول للمعلمات باستعمال كافة الطرائق عند حجم العينة 50

| MSE | Methods | Sample size |
|---|---|---|
| 0.231229552871543 | Mle | N=50 |
| 0.215989821797131 | Mcs | |
| 0.019099986421551 | Wls | |
| 0.896063717954890 | Bayes | |
| Wls | Best | |

في الجدول رقم (2)عند حجم العينة $(N = 50)$ نلاحظ تفوق طريقة المربعات الصغرى الموزونة $(Wls)$ في تقدير المعلمات من حيث امتلاكها اقل $(MSE)$ للمقدرات مقارنة بطرائق التقدير الاخرى، وايضا نميز ان طريقة تصغير مربع كاي $(Mcs)$ قد احتلت المرتبة الثانية للأفضلية وذلك لكافة القيم الافتراضية للمعلمات والانموذج الاول المفترض.

جدول (3) متوسط مربعات الخطأ (MSE) لمقدرات انموذج الانحدار اللوجستي الثنائي في الانموذج الاول للمعلمات باستعمال كافة الطرائق عند حجم العينة 100

| MSE | Methods | Sample size |
|---|---|---|
| 0.195214313074604 | Mle | N=100 |
| 0.155371787851206 | Mcs | |
| 0.017568718160916 | Wls | |
| 0.687385040696479 | Bayes | |
| Wls | Best | |

في الجدول رقم (3)عند حجم العينة $(N = 100)$ نلاحظ تفوق طريقة المربعات الصغرى الموزونة $(Wls)$ في تقدير المعلمات من حيث امتلاكها اقل $(MSE)$ للمقدرات مقارنة بطرائق التقدير الاخرى، وايضا نميز ان طريقة تصغير مربع كاي $(Mcs)$ قد احتلت المرتبة الثانية للأفضلية وذلك لكافة القيم الافتراضية للمعلمات والانموذج الاول المفترض.

جدول (4) متوسط مربعات الخطأ (MSE) لمقدرات انموذج الانحدار اللوجستي الثنائي في الانموذج الاول للمعلمات باستعمال كافة الطرائق عند حجم العينة 150

| MSE | Methods | Sample size |
|---|---|---|
| 0.010072144667444 | Mle | N=150 |
| 0.038498319485976 | Mcs | |
| 0.023542680540880 | Wls | |
| 0.458073693010931 | Bayes | |
| Mle | Best | |

في الجدول رقم (4)عند حجم العينة $(N = 150)$ نلاحظ تفوق طريقة الامكان الاعظم $(Mle)$ في تقدير المعلمات من حيث امتلاكها اقل $(MSE)$ للمقدرات مقارنة بطرائق التقدير الاخرى، وان طريقة المربعات الصغرى الموزونة $(Wls)$ قد تصدرت المرتبة الثانية للأفضلية وذلك لكافة القيم الافتراضية للمعلمات والانموذج الاول المفترض.

**جدول (5) متوسط مربعات الخطأ (MSE) لمقدرات انموذج الانحدار اللوجستي الثنائي في الانموذج الاول للمعلمات باستعمال كافة الطرائق عند حجم العينة 200**

| Sample size | Methods | MSE |
|---|---|---|
| N=200 | Mle | 0.105158079021595 |
| | Mcs | 0.073204095012192 |
| | Wls | 0.006287675940011 |
| | Bayes | 0.472422431130471 |
| | Best | Wls |

في الجدول رقم (5)عند حجم العينة $(N = 200)$ نلاحظ تفوق طريقة المربعات الصغرى الموزونة $(Wls)$ في تقدير المعلمات من حيث امتلاكها اقل $(MSE)$ للمقدرات مقارنة بطرائق التقدير الاخرى، و ان طريقة تصغير مربع كاي $(Mcs)$ قد احتلت المرتبة الثانية للأفضلية وذلك لكافة القيم الافتراضية للمعلمات والانموذج المفترض.

**جدول (6) متوسط مربعات الخطأ (MSE) لمقدرات انموذج الانحدار اللوجستي الثنائي في الانموذج الاول للمعلمات باستعمال كافة الطرائق عند حجم العينة 250**

| Sample size | Methods | MSE |
|---|---|---|
| N=250 | Mle | 0.118088251341391 |
| | Mcs | 0.084338380647715 |
| | Wls | 0.015775518249187 |
| | Bayes | 0.479378222933252 |
| | Best | Wls |

في الجدول رقم (6)عند حجم العينة $(N = 50)$ نلاحظ تفوق طريقة المربعات الصغرى الموزونة $(Wls)$ في تقدير المعلمات من حيث امتلاكها اقل $(MSE)$ للمقدرات مقارنة بطرائق التقدير الاخرى، وايضا نميز ان طريقة تصغير مربع كاي $(Mcs)$ قد احتلت المرتبة الثانية للأفضلية وذلك لكافة القيم الافتراضية للمعلمات والانموذج الاول المفترض.

**جدول (7) متوسط مربعات الخطأ (MSE) لمقدرات انموذج الانحدار اللوجستي الثنائي في الانموذج الاول للمعلمات باستعمال كافة الطرائق عند حجم العينة 300**

| Sample size | Methods | MSE |
|---|---|---|
| N=300 | Mle | 0.082075989273375 |
| | Mcs | 0.071375768123519 |
| | Wls | 0.005883737386357 |
| | Bayes | 0.287783785987223 |
| | Best | Wls |

في الجدول رقم (7) عند حجم العينة $(N = 300)$ نلاحظ تفوق طريقة المربعات الصغرى الموزونة $(Wls)$ في تقدير المعلمات من حيث امتلاكها اقل $(MSE)$ للمقدرات مقارنة بطرائق التقدير الاخرى، و ان طريقة تصغير مربع كاي $(Mcs)$ قد احتلت المرتبة الثانية للأفضلية وذلك لكافة القيم الافتراضية للمعلمات والانموذج الاول المفترض.

جدول (8) متوسط مربعات الخطأ (MSE) لمقدرات انموذج الانحدار اللوجستي الثنائي في الانموذج الثاني للمعلمات باستعمال كافة الطرائق عند حجم العينة 50

| MSE | Methods | Sample size |
|---|---|---|
| 0.135106928071695 | Mle | 05N= |
| 0.090487296781982 | Mcs | |
| 0.036680094541704 | Wls | |
| 0.909697598068564 | Bayes | |
| Wls | Best | |

في الجدول رقم (8)عند حجم العينة $(N = 50)$ نلاحظ تفوق طريقة المربعات الصغرى الموزونة $(Wls)$ في تقدير المعلمات من حيث امتلاكها اقل $(MSE)$ للمقدرات مقارنة بطرائق التقدير الاخرى، و ان طريقة تصغير مربع كاي $(Mcs)$ قد احتلت المرتبة الثانية للأفضلية وذلك لكافة القيم الافتراضية للمعلمات والانموذج الثاني المفترض.

جدول (9) متوسط مربعات الخطأ (MSE) لمقدرات انموذج الانحدار اللوجستي الثنائي في الانموذج الثاني للمعلمات باستعمال كافة الطرائق عند حجم العينة 100

| MSE | Methods | Sample size |
|---|---|---|
| 0.183950622761129 | Mle | 010N= |
| 0.139472754614625 | Mcs | |
| 0.015568965561645 | Wls | |
| 0.721198879994797 | Bayes | |
| Wls | Best | |

في الجدول رقم (9)عند حجم العينة $(N = 100)$ نلاحظ تفوق طريقة المربعات الصغرى الموزونة $(Wls)$ في تقدير المعلمات من حيث امتلاكها اقل $(MSE)$ للمقدرات مقارنة بطرائق التقدير الاخرى، و ان طريقة تصغير مربع كاي $(Mcs)$ قد احتلت المرتبة الثانية للأفضلية وذلك لكافة القيم الافتراضية للمعلمات والانموذج الثاني المفترض.

جدول (10) متوسط مربعات الخطأ (MSE) لمقدرات انموذج الانحدار اللوجستي الثنائي في الانموذج الثاني للمعلمات باستعمال كافة الطرائق عند حجم العينة 150

| MSE | Methods | Sample size |
|---|---|---|
| 0.102916386500496 | Mle | 015N= |
| 0.073973528994289 | Mcs | |
| 0.005753260067777 | Wls | |
| 0.480883954286149 | Bayes | |
| Wls | Best | |

في الجدول رقم (10) عند حجم العينة $(N = 150)$ نلاحظ تفوق طريقة المربعات الصغرى الموزونة $(Wls)$ في تقدير المعلمات من حيث امتلاكها اقل $(MSE)$ للمقدرات مقارنة بطرائق التقدير الاخرى، و ان طريقة تصغير مربع كاي $(Mcs)$ قد احتلت المرتبة الثانية للأفضلية وذلك لكافة القيم الافتراضية للمعلمات والانموذج الثاني المفترض.

جدول (11) متوسط مربعات الخطأ (MSE) لمقدرات انموذج الانحدار اللوجستي الثنائي في الانموذج الثاني للمعلمات باستعمال كافة الطرائق عند حجم العينة 200

| Sample size | Methods | MSE |
|---|---|---|
| 020N= | Mle | 0.124285386959490 |
| | Mcs | 0.093994577706719 |
| | Wls | 0.009937727506664 |
| | Bayes | 0.485372652284942 |
| | Best | Wls |

في الجدول رقم (11) عند حجم العينة $(N = 200)$ نلاحظ تفوق طريقة المربعات الصغرى الموزونة $(Wls)$ في تقدير المعلمات من حيث امتلاكها اقل $(MSE)$ للمقدرات مقارنة بطرائق التقدير الاخرى، و ان طريقة تصغير مربع كاي $(Mcs)$ قد احتلت المرتبة الثانية للأفضلية وذلك لكافة القيم الافتراضية للمعلمات والانموذج الثاني المفترض.

جدول (12) متوسط مربعات الخطأ (MSE) لمقدرات انموذج الانحدار اللوجستي الثنائي في الانموذج الثاني للمعلمات باستعمال كافة الطرائق عند حجم العينة 250

| Sample size | Methods | MSE |
|---|---|---|
| N=250 | Mle | 0.132436701475596 |
| | Mcs | 0.099210780250428 |
| | Wls | 0.018785545076914 |
| | Bayes | 0.505246854363138 |
| | Best | Wls |

في الجدول رقم (12) عند حجم العينة $(N = 250)$ نلاحظ تفوق طريقة المربعات الصغرى الموزونة $(Wls)$ في تقدير المعلمات من حيث امتلاكها اقل $(MSE)$ للمقدرات مقارنة بطرائق التقدير الاخرى، و ان طريقة تصغير مربع كاي $(Mcs)$ قد احتلت المرتبة الثانية للأفضلية وذلك لكافة القيم الافتراضية للمعلمات والانموذج الثاني المفترض.

جدول (13) متوسط مربعات الخطأ (MSE) لمقدرات انموذج الانحدار اللوجستي الثنائي في الانموذج الثاني للمعلمات باستعمال كافة الطرائق عند حجم العينة 300

| Sample size | Methods | MSE |
|---|---|---|
| N=300 | Mle | 0.056465091643550 |
| | Mcs | 0.038712584446633 |
| | Wls | 0.010499748027347 |
| | Bayes | 0.295309927444674 |
| | Best | Wls |

في الجدول رقم (13)عند حجم العينة $(N = 300)$ نلاحظ تفوق طريقة المربعات الصغرى الموزونة $(Wls)$ في تقدير المعلمات من حيث امتلاكها اقل $(MSE)$ للمقدرات مقارنة بطرائق التقدير الاخرى، و ان طريقة تصغير مربع كاي $(Mcs)$ قد احتلت المرتبة الثانية للأفضلية وذلك لكافة القيم الافتراضية للمعلمات والانموذج الثاني المفترض.

**جدول (14) يمثل أفضلية كافة طرائق التقدير حسب (MSE) لمقدرات الانموذج اللوجستي الثنائي**

| طرائق التقدير | عدد مرات الأفضلية | النسبة |
|---|---|---|
| Wls | 11 | 91.7 |
| Mle | 1 | 8.3 |

الجدول(14) يلخص نسب الأفضلية لطرائق التقدير المتنافسة حيث نلاحظ ان الطرائق (Wls) حققت أعلى نسبة أفضلية على كافة طرائق التقدير الاخرى في تقدير المعلمات، وذلك لكافة القيم الافتراضية للمعلمات، واحجام العينات المختلفة، وللنموذجين الأول والثاني.

## 7- الاستنتاجات والتوصيات

1-تفوق طريقة $(Wls)$ على كافة طرائق التقدير الاعتيادية ، في تقدير المعلمات لأنموذج الانحدار اللوجستي الثنائي إذ تتصدر هذه الطريقة المرتبة الأولى من حيث عدد مرات امتلاكها اقل $(MSE)$ لمقدرات الانموذج ولكونها الطريقة الوحيدة التي لا تحتاج الى قيم بدائية لغرض تقدير معلمات الانموذج ، وهذا ما يميزها على كافة طرائق التقدير الاخرى.وبعدها جاءت في المرتبة الثانية طريقة (Mcs) وذلك لأغلب احجام العينات والقيم الافتراضية للمعلمات والنماذج التي تم افتراضها .

2- عند حجم العينة $(N = 150)$ حققت طريقة (Mle) المرتبة الاولى واحتلت طريقة $(Wls)$ المرتبة الثانية، واحتلت طريقة $(Mcs)$ المرتبة الثانية للأفضلية في تقدير المعلمات بعد طريقة $(Wls)$ وذلك لكافة احجام العينات وللأنموذج المفترض الأول مقارنة بطرائق التقدير الاخرى.

3-تناوبت طرائق التقدير $(MLE)(MCS)$, على احتلال المرتبة الثانية من الأفضلية حيث اظهرت نتائج تقديراتهم بأنها متقاربة .

4-اثبتت طريقة بيز$(BE)$ انها اقل كفاءة في ايجاد تقديرات المعلمات لأنموذج الانحدار اللوجستي الثنائي عند مقارنتها بالطرائق الكلاسيكية كونها تملك اكبر $(MSE)$ لمقدرات الانموذج بالنسبة الانموذج الاول و الانموذج الثاني.

5-نوصي باستعمال طريقة $(Wls)$ لتقدير معلمات انموذج الانحدار اللوجستي الثنائي في حالة توزيع $Bernoulli$ للمتغير Y.

6-يمكن التوسع بالدراسة باستعمال اسلوب الخوارزمية الجينية ومقارنتها مع الطرائق الكلاسيكية لمعرفة مدى كفاءة الطرائق في تقدير معلمات انموذج الانحدار اللوجستي الثنائي.

## المصـــادر

1. احمد، ايمان حسن وشهاب، ضمياء حامد،(2018)،"مقارنة بعض المقدرات الحصينة لتقدير معلمات انموذج الانحدار اللوجستي ثنائي الاستجابة باستعمال المحاكاة"، بغداد، *مجلة العلوم الاقتصادية والادارية*، الصفحات 423 – 440.

2. العزاوي ، احمد ذياب، (2005)،" المقارنة بين بعض طرائق تقدير انموذج انحدار اللوجستك والطرائق الحصينة للتجارب الحياتية ذات الاستجابة الثنائية باستخدام اسلوب المحاكاة"، رسالة ماجستير في الاحصاء، كلية الادارة والاقتصاد، جامعة بغداد.

3. Agresti, A., (2002) , "Cartegorical Data Analysis",2th edition, Jhon Wiley & sons Inc , , Hoboken, New Jersey.

4. 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.

5. Cramer, J. S., (2003),"Logit Models From Economics and other fields", Cambridge University Press cape Town, New York, *ISBN*, pp.33-45.

6. Garson, D.,(2014),"Logistic Regression :Binary and Multinomial" Statistical Associates Publishing. All rights reserved worldwide in all media ,*ISBN* , Retrieved from ,http : *www.statisticalassociates.com*.

7. Groenwald , P. & Mokgatlhe, L., ( 2005) , " Bayesian Computation for Logistic Regression", Computaion Statistics & Data Analysis 48 , University of Bostswana , pp. 857-868.

8. Henry , A. ,( 2013), "Bayesian Logistic Regression Modelling Via Markov Chain Monte Carlo Algorithm ",University of Cape Coas , Ghana,pp.193– 197.

9. Hosmer, D., Lemeshow, S. & Sturdivant , R. ,(2013)," Applied Logistic Regression", 3rd edition ,New York: willey *http : // ihmsi.org.*

10. 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.

11. Mccullagh, P., & Nelder, J., (1983)," Generalized Linear Models", London: Chapman and Hall.

12. Menard ,S.,(2002),"Applied Logistic Regression Analysis",2nd Edition Thousand Oaks Edition Thousand Oaks , CA : Sage Publications , *Series Quantitative Applications in the Social Sciences* , PP. 1 – 111.

13. Muller, Marlene , (2004) ," Generalized Linear Models", Fraunhofer Institute for Industrial Mathematics ( ITWM ) , ( Germany ) ,*WWW. Marlenmuller . ed / publication / hand book CS. Pdf.*

14. Rodriguez , G. ,(2007), "Logit Models for Binary Data" ,Chapter(3) ,Retrieved from, *http://data.princeton.edu/wws509/notes/c3.pdf* ,pp.1-50.

15. Tektas, D., & Günay, S. ,(2008)," A Bayesian Approach to Parameter Estimation in Binary Logit and Probit Models", *Hacettepe Journal of Mathematics and Statistics*, pp. 167-176.

16. Zager, S., & Karim, R., (1991),"Generalized Linear Models with Random Effects : A Gibbs Sampling Approach", *JASA* , pp. 79-86.

17. https://www.statisticssolutions.com/assumptions-of-logisticregression/

## The use of the Biz method and classical methods in estimating the parameters of the binary logistic regression model

**Abstract**

**Binary logistic regression model used in data classification and it is the strongest most flexible tool in study cases variable response binary when compared to linear regression. In this research, some classic methods were used to estimate parameters binary logistic regression model, included the maximum likelihood method $(MLEM)$, minimum chi-square method $(MCSM)$, weighted least squares $(WLSM)$, with bayes estimation $(BE)$, to choose the best method of estimation by default values to estimate parameters according two different models of general linear regression models ,and different sample sizes $(N = 50,100,150,200,250,300)$,and building an experiment simulation experience then displaying the results and the analysis using the statistical criteria Mean Squares Error (MSE),to choose the best standard methods for estimators the binary logistic regression model.**

**Generally, The $(WLS)$ method was found to be the best one among the standard estimation methods, for the purpose of estimating the parameters for binary logistic regression model because it has the less (MSE) for estimators compared to other methods, which indicates the accuracy of the $(WLS)$ method in estimating the parameters of the model.**

**Keywords:** **Binary Logistic Regression Models (BLRM), Maximum Likelihood Method (MLM) , Newton-Raphson Algorithm(NR), Minimum Chi-Square (MCSM), Weighted Least Squares Method (WLSM), Bayes Method(BM) , Gibbs Algorithm.**