ISSN (1681- 6870)

استخدام اسلوب الانحدار اللوجستي الثنائي لتحليل العوامل المؤثرة على وفيات مرضى القلب دراسة تطبيقية على عينة من المرضى في محافظة ذي قار

م . حيدر رائد طالب[1]،
[1] كلية الادارة والاقتصاد / قسم الاحصاء، جامعة سومر

ساره عادل مظلوم[2]
[2] المعهد التقني – الصويرة

1-1 المستخلص

تناول هذا البحث أحد أهم نماذج الانحدار غير الخطية الواسعة الاستعمال في نمذجة التطبيقات الاحصائية وهو انموذج الانحدار اللوجستي الثنائي ,ومن ثم تقدير معلمات هذا الأنموذج باستعمال طريقة المربعات الصغرى الموزونة ,وقد تم في الجانب التطبيقي استعمال هذا الأنموذج لنمذجة البيانات الخاصة بالمصابين بأمراض القلب، 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 الرئيس لحدوث الوفاة هو التدخين, وكذلك دقة الطريقة (WLSE) في تقدير معلمات الأنموذج .

الكلمات المفتاحية: الانحدار اللوجستي الثنائي، المربعات الصغرى الموزونة، النمذجة

The Use of Binary Logistic Regression Method to Analyze The Factors Affecting Heart Disease Deaths: An Applied Study on A Sample of Patients in Dhi Qar Governorate

Haider Raed Talib
Sumer University / College of Administration and Economics / Department of Statistics / Iraq.
. h_r_t86@yahoo.com

Sarah Adel Mazloum
AL-Suwaira Technical Institute

Abstract:

This research dealt with one of the most important non-linear regression models used in the modeling of statistical applications, which is the model of binary logistic regression, and then estimate the parameters of this model using the least-squares method. By comparing the causes of actual deaths with the estimated causes of death to the extent to which the model is appropriate in modeling this type of data and extracting the main cause of death is smoking, as well as the accuracy of the method (WLSE) in estimating Models.

Keywords: Non-Linear Regression Models, WLSE.

ISSN (1681- 6870)

2-1 مشكلة البحث (Research Problem)

ان مشكلة البحث تكمن في وجود المعادلات غير الخطية في انموذج الانحدار اللوجستي الثنائي التي تؤثر على مقدرات معلمات الانموذج عند تحليل البيانات التي يكون فيها متغير الاستجابة من النوع المتقطع(الفئوي او الرتبي)والمتغيرات التوضيحية بعضها كمي وبعضها الآخر نوعي ,ولكي تكون المعلمة ذات دلالة احصائية جيدة يتم تقديرها باستعمال المربعات الصغرى الموزونة (WLS).

3-1 هدف البحث (Research Aim)

ان الهدف من هذا البحث هو توظيف طريقة المربعات الصغرى الموزونة في تقدير معلمات أنموذج الانحدار اللوجستي الثنائي، ومن ثم بناء أنموذج الانحدار اللوجستي في الجانب التطبيقي لبيان اهم العوامل المؤثرة على وفاة مرضى القلب في محافظة ذي قار.

أولا :الجانب النظري

4-1 انموذج الانحدار اللوجستي الثنائي (Binary Logistic Regression) (Model

يعد هذا النموذج من نماذج الانحدار اللاخطية ويتصف بان متغير الاستجابة (Y) يتبع توزيع برنولي بأخذ القيم (0)و(1)[6;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 الحث (النجاح) هو$[\pi(X_i)]$ اعتمادا على قيم المتغيرات التوضيحية للمشاهدات واحتمال عدم وقوع الحدث (الفشل) $[1-\pi(X_i)]$ وبذلك تكون دالة الكثافة الاحتمالية بالصيغة الاتية [10;2007;P.3]:

$$p\left({}^{Y_i}/_{X_i}\right) = [\pi(X_i)]^{Y_i}[1-\pi(X_i)]^{1-Y_i} \qquad \ldots(1)$$

$$Y_i = 0,1$$

$$P(Y=1\backslash X_i) = \pi(X_i) \qquad \ldots(2)$$

$$P(Y=0\backslash X_i) = 1-\pi(X_i) \qquad \ldots(3)$$

وان توقع متغير الاستجابة يمثل احتمال النجاح

$$\mathrm{E(Y_i) = \pi(X_i)}$$

وايضا تباين متغير الاستجابة يمثل احتمال حاصل ضرب النجاح مع احتمال الفشل

$$\mathrm{V(Y_i) = \pi(X_i)(1-\pi(X_i))}$$

ويتم تحويل هذا النموذج الى شكل خطي يتمثل بعلاقة خطية من خلال الموجه الصفي ($\underline{X'_i}$) من المتغيرات التوضيحية مع لوجت الاحتمال $[logic\ \pi(X_i)]$ وحسب الصيغة الرياضية الاتية :

ISSN (1681- 6870)

$$Z = \text{logic}\,\pi(X_i) = \ln\left[\frac{\pi(X_i)}{1-\pi(X_i)}\right] = B_0 + B_1X_{i1} + \ldots + B_pX_{ip}$$

$$Z = \underline{X'_i}\ \underline{B} \qquad \ldots(4)$$

$$\therefore Z = \underline{X'_i}\ \underline{B} + \varepsilon_i \qquad \ldots(5)$$

حيث ان متغير الاستجابة في الانحدار الخطي يفترض [6;2013;p.7]:

$$Y_i = E\left(\frac{Y}{X}\right) + \varepsilon_i$$

$$Y_i = \pi(X_i) + \varepsilon_i \rightarrow Y_i = \frac{e^{(B_0+B_1X_{i1}+\ldots+B_pX_{ip})}}{1+e^{(B_0+B_1X_{i1}+\ldots+B_pX_{ip})}} + \varepsilon_i$$

وان التوزيع المشروط يكون طبيعي بمتوسط E(Y/X) وتباين ثابت وان $\varepsilon_i = Y_i - E(Y/X)$ وعند تعويض قيمة Y=1 ينتج $\varepsilon_i = 1 - \pi(X_i)$ عندما Y=0ينتج $\varepsilon_i = -\pi(X_i)$ وبذلك فان حد الخطأ (ε_i) في هذه الحالة يتبع توزيع برنولي بمتوسط مقداره صفر وتباين مقداره $[\pi(X_i)][1-\pi(X_i)]$ وهذا يدل ان تباين حد الخطا يكون غير متجانس ، ويمكن التعبير عن انموذج الانحدار اللوجستي بالصيغة الرياضية الاتية [10;2007;P.8]:

$$\pi(X_i) = \frac{e^{\underline{X'_i}\ \underline{B}}}{1+e^{\underline{X'_i}\ \underline{B}}} \qquad \ldots(6)$$

$$1-\pi(X_i) = \frac{1}{1+e^{\underline{X'_i}\ \underline{B}}} \qquad \ldots(7)$$

اذ ان

$\pi(X_i)$ هو التوقع الشرطي E(Y/X) (Conditional Mean) لمتغير الاستجابة (Y)عند قيمة معينة لـ (X) (احتمال حدوث الاستجابة).

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

$X_1, X_2, \ldots, X_p$: المتغيرات التوضيحية التي تكون المصفوفة $X=(X_{ij})$ برتبة (n*p)

i=1,2,…n n:عدد المشاهدات .

j=0,1,2,…,p p: عدد المتغيرات التوضيحية والمعلمات المجهولة.

$B_0, B_1, \ldots, B_p$: متجه للمعلمات المراد تقديرها .

ISSN (1681- 6870)

5-1 تقدير معلمات الانموذج اللوجستي الثنائي باستعمال طريقة المربعات الصغرى الموزونة

(Estimation Of The Binary Logistic Model Using The Weighted Last Squares Method (WLSM))

ان طريقة المربعات الصغرى الاعتيادية (OLS) لا يمكن استعمالها في تقدير معلمات الانموذج اللوجستي لعدم تحقق احد شروطها وهو عدم تجانس تباين الخطأ [3;2012;P336].

يعتبر انموذج الانحدار اللوجستي الثنائي حالة خاصة من النماذج الخطية العامة التي تكون امتداد للأنموذج الخطي البسيط [9;2004;P.1].

يمكن كتابة انموذج الانحدار الخطي العام كالاتي [8;1983;P76].

$$y_i = b_\circ + b_1 x_{1i} + b_2 x_{2i} + \cdots + b_p x_{pi} + \varepsilon_i \qquad \dots (8)$$

$$y_i = b_\circ + \sum_{j=1}^{p} b_j x_{ij} + \varepsilon_i \qquad ,i = 1,2,\dots,n\ ;\ j = 1,2,\dots,p$$

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

$$W_i = \pi_i(1 - \pi_i) \qquad \dots (9)$$

و ان:

W_i : تمثل مصفوفة التباينات وهي الاوزان المختارة للمستوى i

فيكون تقدير المعلمات $(B_0,B_1,\dots,B_P)$ حسب (WLS) بايجاد القيم التي تجعل الفرق بين الاستجابة المشاهدة والاستجابة المقدرة اقل ما يمكن اي تصغير مجموع مربعات الخطا (SSE)[6;2013;P.135].

$$SSE_i = \sum w_i\left(Z_i - \hat{Z}_i\right)^2 \qquad \dots (10)$$

$$Z_i = ln\frac{\pi_i}{1 - \pi_i}$$

(Z_i) تم ذكرها مسبقا في معادلة (4) تمثل العلاقة الخطية الناتجة من اخذ اللوغاريتم الطبيعي لمعاملات المفاضلة.

$$SSE_i = \sum w_i\left(Z_i - \hat{\beta}_\circ - \hat{\beta}_1 x_{1i} - \cdots - \hat{\beta}_p x_{pi}\right)^2 \qquad \dots (11)$$

وبحل التفاضلات الجزئية بالنسبة الى $(B_0,B_1,\dots,B_P)$ ومساواة الناتج الى الصفر نحصل على ما يأتي :

ISSN (1681- 6870)

$$\beta_0 \sum w_i + \beta_1 \sum w_i x_{1i} + \cdots + \beta_p \sum w_i x_{pi} = \sum w_i Z_i$$
$$\beta_0 \sum w_i x_{1i} + \beta_1 \sum w_i x^2_{1i} + \cdots + \beta_p \sum w_i x_{1i} x_{pn} = \sum w_i x_{i1} Z_i$$
$$\beta_0 \sum w_i x_{pi} + \beta_1 \sum w_i x_{pi} x_{1i} + \cdots + \beta_p \sum w_i x^2_{pi} = \sum w_i x_{ip} Z_n$$

وباتباع اسلوب المصفوفات تكتب هذه المعادلات كما يلي:

$$\underline{Z} = \begin{bmatrix} ln\frac{\pi_1}{1-\pi_1} \\ ln\frac{\pi_2}{1-\pi_2} \\ \vdots \\ ln\frac{\pi_n}{1-\pi_n} \end{bmatrix} = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{bmatrix}, W = \begin{bmatrix} W_1 & 0 & \dots & 0 \\ 0 & W_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & W_n \end{bmatrix}$$

و ان:

$$\acute{X}W\underline{\beta} = \acute{X}W\underline{Z}$$

$$\acute{X}WX = \begin{bmatrix} \sum W_i & \sum W_i X_{i1} & \sum W_i X_{i2} & \dots & \sum W_i X_{ip} \\ \sum W_i X_{i1} & \sum W_i X^2_{i1} & \sum W_i X_{i1} X_{i2} & \dots & \sum W_i X_{i1} X_{np} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \sum W_i X_{ip} & \sum W_i X_{ip} X_{i1} & \sum W_i X_{ip} X_{i2} & \dots & \sum W_i X^2_{ip} \end{bmatrix}$$

$$\acute{X}WZ = \begin{bmatrix} \sum W_i Z_i \\ \sum W_i X_{i1} Z_i \\ \vdots \\ \sum W_i X_{ip} Z_n \end{bmatrix}$$

وان

Z: يمثل متجه التحويل الخطي (اللوجت) للنموذج ذو الرتبة (n*1)

W: مصفوفة مربعة عناصر قطرها الرئيسي يمثل التباينات ذو الرتبة (n*n)

X'WX: مصفوفة مربعة حصيلة ضرب مصفوفة المتباينات مع مصفوفة المتغيرات التوضيحية ذو الرتبة [(P+1)(P+1)]

X'WZ : تمثل متجه مبدلة المصفوفة المتغيرات التوضيحية ومصفوفة التباينات مع متجه التحويل الخطي ذو رتبة [(P+1)*n]

وبذلك نصل الى قيمة $\hat{B}$ المقدرة والي تكون [9;2004;P;8]

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

ISSN (1681- 6870)

6-1 الاختبارات المتعلقة بأنموذج الانحدار اللوجستي الثنائي

(The Test For Binary Logistic Regression Model)

هناك الكثير من معايير التقييم المتعلقة بأنموذج الانحدار اللوجستي الثنائي المتعلقة بجودة التوفيق وحسن المطابقة للأنموذج نذكر منها (اختبار الدرجة او المعيارية و اختبار نسبة الامكان الاعظم و اختبار هوزمر-ليمشو ومعيار المعلومات لأكايكي)وغيرها من المعايير الاخرى وقد تم التركيز في البحث على الاختبارات الاتية:

1-6-1 اختبار والد

يستعمل هذا الاختبار لمعرفة اهمية ومعنوية الثوابت للمتغيرات التوضيحية وتأثيرها على متغير الاستجابة في النماذج اللوجستية ويختبر الفرضية التي تنص على مساواة معلمة الانحدار اللوجستي للمتغير التوضيحي بالصفر وكما يأتي [6;2013;P.4].

$H_0: B_j = 0$

$H_0: B_j \neq 0$

اذا كانت قيمة الاختبار الاحتمالية $p - value$ اقل من (0.05)(ترفض فرضية العدم) وهذا يعني ان معاملات المتغير التوضيحي معنوية ذات دلالة احصائية[13;2014;p.13].

وفي اختبار (wald) تستعمل إحصاءه (Z) وهي عبارة عن الجذر التربيعي لاختبار والد وتتوزع توزيعا طبيعيا قياسيا حسب الصيغة الاتية [10;2007;P.12].

$$Z = \frac{\hat{\beta}_j}{S.E\,(\hat{\beta}_j)} \quad , j = 1,2,\dots.p \quad , \quad Z \sim N(0,1) \qquad \dots(13)$$

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

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

ISSN (1681- 6870)

2-6-1 احصاءات معامل التحديد (R^2)

(The Cofficient Of Determination (R^2) Statistics)

ان قيمة معامل التحديد (R^2) تبين القوة التفسيرية لانموذج الانحدار اللوجستي بين المتغيرات التوضيحية ومتغير الاستجابة ويمكن تحديدها باستعمال الصيغة الرياضية الاتية [12;2014;P.51].

$$R^2 = 1 - exp\left[-\left(\frac{l^2}{n}\right)\right] \quad \dots (14)$$

و ان :

l: تمثل قيمة لوغارتم دالة الامكان الاعظم بأنموذج الانحدار اللوجستي.

n: تمثل حجم العينة بأنموذج الانحدار اللوجستي.

وقيمته بين [1,0] وكلما اقتربت (R^2) من (1) دل ذلك على قوة العلاقة بين المتغير التوضيحي ومتغير الاستجابة وعلى العكس عندما تقترب (R^2) من (0) يدل ذلك على ضعف العلاقة بين المتغير التوضيحي ومتغير الاستجابة .[5;2014;P.50]

3-6-1 معيار جدول التصنيف (Classification Table Criterion (CTC))

يوضح جدول التصنيف مدى ملائمة الانموذج ويتمثل في مقارنة النتائج المتوقعة من خلال الانموذج مع النتائج المشاهدة من خلال تصنيف القيم المتوقعة على اساس قيمة احتمالية معينة تسمى نقطة القطع (Cut Value) يعبر عنها بالحرف (C) وان نقطة القطع محصورة بين الصفر والواحد ($0 \leq C \leq 1$) وهي في الاغلب تساوي (0.5) فاذا تجاوزت الاحتمالات المتوقعة نقطة القطع (C) اعطيت تلك الحالة تصنيفا متوقعا يساوي واحد وما عدا ذلك فان الحالة يعطى لها تصنيف متوقع يساوي صفر[11;1997'P1-25].

$$Y = \begin{cases} 1 & \text{if } \pi \geq C \\ 0 & \text{if } \pi < C \end{cases}$$

يتكون جدول التصنيف من اربع خلايا تحتوي على عدد الحلات التي تمتلك الصفة او الحلات التي لاتمتلك الصفة محل دراستنا والتي تم تصنيفها بطريقة صحيحة او بطريقة خاطئة وعليه فان الانموذج اذا توقع تصنيف الحالات بشكل صحيح دل ذلك على ان الانموذج مطابق للبيانات وكما في الجدول (1) الاتي [4;1999;P.1-23].

ISSN (1681- 6870)

الجدول (1) جدول التصنيف

Observed / Predicted	Pos. (P)	Neg. (N)	Total
Pos. (P)	(a) الموجب الصحيح	(b) الموجب الخاطئ	(a + b)
Neg. (N)	(c) السالب الخاطئ	(d) السالب الصحيح	(c + d)
Total	(a + c)	(b + d)	(a + b + c + d)

من خلال جدول التصنيف (1) اعلاه نستطيع حساب قيمتين رئيسيتين هما الحساسية والدقة لتقدير ملائمة النموذج وهي [7;2013;pp.2955-2961].

1-الحساسية (Sensitivity(SE))

وتعرف الحساسية بانها قيمة الاحتمال بان يكون التصنيف المتوقع موجبا (1) للحالة التي تكون موجبة اي ان

$$SE = \frac{a}{a+c} \quad \ldots (15)$$

2-الدقة (Accuracy (AC))

وتعرف بانها قيمة احتمال ان يكون التصنيف المتوقع سالبا (0) للحالة التي تكون سالبة اي ان :

$$AC = \frac{d}{b+d} \quad \ldots (16)$$

ويمكن ان نحسب نسبة التصنيف الصحيح (Hit Ratio) للانموذج ككل وتعرف ايضا بنسبة الكفاءة [Efficiency (EF)=a+d] فتكون نسبة التصنيف حسب الصيغة [1;2015;p.38].

$$\text{Hit Ratio} = \frac{\text{الكفاءة}}{\text{المجموع الكلي}} = \frac{EF}{Total} = \frac{a+d}{a+b+c+d} \quad \ldots (17)$$

و أن :

a: الحالة المشاهدة تساوي واحد والمتوقعة أكبر من أو يساوي (C)

b: الحالة المشاهدة تساوي صفر والمتوقعة أكبر من أو تساوي (C)

c: الحالة المشاهدة تساوي واحد والمتوقعة أصغر من (C)

d: الحالة المشاهدة تساوي صفر والمتوقعة أصغر من (C)

وكلما كانت نسبة التصنيف مرتفعة دل ذلك على جودة الانموذج المقدر .

ISSN (1681- 6870)

ثانيا : الجانب التطبيقي

تمهيد

نظرا لأهمية الجانب الصحي وتأثيره على حياة الانسان تم اختيار التطبيق في هذا الجانب حيث سيتم التركيز على امراض القلب وتأثيرها على حياة الناس لكونها من الامراض واسعة الانتشار في المجتمع والكثير من الوفيات تكون اسبابها هذه الامراض , يبحث هذا الجانب في وصف البيانات المتمثلة بالإصابة بأمراض القلب من خلال اهم العوامل(المتغيرات التوضيحية) التي تسبب هذه الامراض ومعرفة وتقدير مدى تأثيرها على حالة حدوث الوفاة او عدم حدوثه (متغير الاستجابة) باستعمال انموذج الانحدار اللوجستي الثنائي الذي اعتمد على المربعات الصغرى الموزونة بتقديرها لمعلمات الانموذج ,وسيتم التطبيق على بيانات المرضى الراقدين في مركز الناصرية للقلب لسنة 2016 الذين كان عددهم (110) لمعرفة العوامل التي لها اكثر تأثير على وفيات هؤلاء المرضى واخذت المعلومات من طبلات المرضى الراقدي وسجل المرضى الراقدين اضافة الى برنامج الراقدين المصمم من وزارة الصحة العراقية. وتتم معالجة البيانات التي تعاني من وجود عشرة عوامل تسبب حالة الوفاة للأشخاص الذين يعانون من امراض القلب بتطبيق الاختبارات الخاصة بالأنموذج التي تحدد اكثر عامل مؤثر في حدوث الوفاة.

وصف البيانات

اعتمدت الدراسة على الاحصاءات الطبية لمحافظة ذي قار لسنة 2016 م الصادرة من مركز امراض القلب وتم استعمال احد عشر متغير منها متغير واحد تابع(متغير الاستجابة y) والذي يأخذ قيمتين هما الصفر لحالة عدم حدوث الوفاة والواحد الصحيح لحالة حدوث الوفاة وعشرة متغيرات توضيحية ,تم تعريف كل متغير في الجدول (2)حيث اخذت المتغيرات الاتية:

ISSN (1681- 6870)

X_3	متغير وصفي يمثل الحالة الزوجية	2	متزوج
		3	ارمل او ارملة مطلق او مطلقة
X_4	متغير كمي يمثل عمر المريض بالسنوات		
X_5	متغير وصفي يمثل نوع المرض الذي يعاني منه المريض الراقد	1	ذبحة صدرية
		2	احتشاء واعتلال عضلة القلب
		3	اضطرابات في توصيل الدم
		4	توقف او تسارع او رجفان القلب
		5	عجز القلب
		6	جلطة دماغية
		7	انسداد الشرايين
		8	اضطرابات في جهاز الدوران
		9	وذمة رئوية
		10	تشوهات خلقية ولادية
X_6	متغير وصفي يمثل التداخل الجراحي للمريض الراقد	1	عدم اجراء عملية
		2	اجراء عملية وسطى
		3	اجراء عملية كبرى
		4	اجراء عملية فوق الكبرى
X_7	متغير وصفي يمثل التدخين	1	مدخن
		2	غير مدخن
X_8	متغير وصفي يمثل ضغط المريض	1	(100 – 149)
		2	(150 – 199)
		3	(200 – 249)
X_9	متغير وصفي يمثل سكر المريض	1	(75 – 179)
		2	(180 – 284)
		3	(285 – 389)
		4	(390 – 494)
X_{10}	متغير كمي يمثل وزن المريض		

جدول (2)

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_1	متغير وصفي يمثل نوع الردهة التي يرقد بها المريض	1	ردهة القلب المفتوح
		2	ردهة انعاش القلب
		3	ردهة القلب و الاوعية
		4	ردهة التمريض الخاص
X_2	متغير وصفي يمثل جنس المريض الراقد	1	ذكر
		2	انثى
		1	اعزب

ISSN (1681- 6870)

تحليل النتائج التطبيقية (Analysis of Applied Results)

سوف يتم عرض نتائج الجانب التطبيقي ومن ثم تحليلها للوصول الى مدى ملائمة البيانات الحقيقية مع انموذج الانحدار اللوجستي الثنائي المقدر من خلال اجراء الاختبارات الخاصة بمقدرات الانموذج اللوجستي .

تم الحصول على كافة نتائج النتائج التطبيقية باستعمال برنامج كتب بلغة [MATLAB, 2017] وفيما يلي سوف يتم عرض هذه النتائج في الجداول التي سيتم تحليلها حسب تسلسل الجداول وكالاتي:

جدول (3) تقدير معلمات انموذج الانحدار اللوجستي الثنائي بطريقة المربعات الصغرى الموزونة

Para. / Meth.	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}
$WLSE$	−1.011	−0.071	0.096	−0.095	−0.001	0.013	−0.007	0..816	0.026	0.078	0.004

نلاحظ في الجدول (3) تنفيذ مقدرات انموذج الانحدار اللوجستي الثنائي من خلال تحديد متجه مقدر المعلمات $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \ \dots \ , \hat{\beta}_{10})$ وتعويضها مع المتغيرات التوضيحية نحصل على مقدرات انموذج الانحدار اللوجستي الثنائي و بدورها تقلل مجموع مربعات الخطأ الى أصغر ما يمكن حسب طريقة المربعات الصغرى الموزونة الاعتيادية $(WLSE)$.

جدول (4) المعلمات المقدرة و الخطأ المعياري لكافة المتغيرات التوضيحية بطريقة$(WLSE)$

المعلمات $(\hat{\beta}_i)$	المعلمات المقدرة	الخطأ المعياري $SE(\hat{\beta}_i)$	نسبة(z) $\frac{\hat{\beta}_i}{SE(\hat{\beta}_i)}$	المعنوية
β_0	-1.011	2.031	-0.498	Non-sig
β_1	-0.072	0.031	-2.284	Sig.
β_2	0.096	0.039	2.426	Sig.
β_3	-0.096	0.069	-1.365	Non-sig
β_4	-0.001	0.001	-0.440	Non-sig
β_5	0.013	0.008	1.596	Non-sig
β_6	-0.007	0.019	-0.383	Non-sig
β_7	0.816	0.038	21.285	Sig.
β_8	0.026	0.036	0.702	Non-sig
β_9	0.078	0.021	3.819	Sig.
β_{10}	0.004	0.002	2.367	Sig.

ISSN (1681- 6870)

نلاحظ في الجدول (4) تقديرات المعلمة وقيم الخطأ المعياري لكل معلمة مقدرة. وان تقدير المعلمات ضروري لاختبار اهمية التأثير الكلي على المتغير التوضيحي من خلال الفرضيات العدم والبديلة على التوالي:

$$H_0: \hat{\beta}_1 = \hat{\beta}_2 = \cdots = \hat{\beta}_{10} = 0$$
$$H_1: \hat{\beta}_1 \neq \hat{\beta}_2 \neq \cdots \neq \hat{\beta}_{10} \neq 0$$

ويتم توليد اختبار والد باستخدام نسبة (Z) في المعادلة (13) ومن خلال النتائج التي حصلنا عليها في الجدول (4) وهي تتبع التوزيع الطبيعي القياسي وبمستوى دلالة $(\alpha =$ 0.05) وتقارن مع قيمة(Z) الجدولية حيث ان $Z_{\frac{1}{2}(1-0.05)} = -1.96$ و $Z_{\frac{1}{2}(0.05)} =$ 1.96 وان $(-1.96 \leq Z \leq 1.96)$، والعمود الاخير يمثل معنوية المتغيرات التوضيحية على حالة المصاب ،(أقل من 0.05 المتغير قيد الاختبار يكون معنويا).

وعليه تتم ازالة المقدرات $\hat{\beta}_0$ و $\hat{\beta}_3$ و $\hat{\beta}_4$ و$\hat{\beta}_5$ و $\hat{\beta}_6$ و $\hat{\beta}_8$ لأنها لا تؤثر بشكل كبير على الانموذج المقدر اي ان المتغيرات (الحالة الاجتماعيةX_{i3} والعمر X_{i4} وسبب الرقود X_{i5} ودرجة العمليةX_{i6} والضغطX_{i8}) ليس لها تأثير معنوي والمتغيرات(الردهة والجنس والتدخين والسكر والوزن) لها تأثير على حالة الوفاة بالنسبة للمرضى الذين يعانون من امراض القلب.

علاوة على ذلك، لمعرفة قوة العلاقة بين المتغيرات التوضيحية ومتغير الاستجابة من خلال احتساب قيمة احصاءة (R^2) في المعادلة (14) و ان $R^2 = 0.22$ وهذا يعني ان المتغيرات (الردهة والجنس والتدخين والسكر والوزن) تفسر الاختلافات في الاستجابات المشاهدة لحالات عدم حدوث الوفاة بنسبة 22% ، ويتضح ايضا ان اشارة معامل تأثير متغير الردهة سالب وهذا يعني ان كلما زاد متغير الردهة بمعدل ردهة واحدة تقل قيمة متغير الاستجابة وتقترب من القيمة(0) اي عدم حدوث الوفاة أدى ذلك الى نقصان حدوث الوفاة بمعدل(0.072)، اما معامل انحدار متغير الجنس نلاحظ ان الاشارة موجبة وهذا يعني ان كلما زاد متغير الجنس (ذكر=1 ، انثى=0) بمعدل ذكر واحد اي اقترب من القيمة (1) زاد حدوث الوفاة بمعدل(0.096)، اما معامل انحدار متغير التدخين (مدخن=1 ، غير مدخن=0) فتظهر الاشارة موجبة وهذا يعني انه كلما زاد متغير التدخين اي اقترب من القيمة(1) مدخن أدى ذلك الى زيادة حدوث الوفاة بمعدل(0.816)، اما اشارة معامل انحدار متغير السكر تكون موجبة وهذا يعني انه كلما زاد متغير السكر أدى ذلك الى زيادة حدوث الوفاة بمعدل(0.078)، اما اشارة معامل انحدار متغير الوزن فتكون موجبة وهذا يعني انه كلما زاد متغير الوزن أدى ذلك الى زيادة حدوث الوفاة بمعدل(0.004)، وبالتالي ان قيم الانموذج اللوجستي الثنائي المقدر يكون حسب الصيغة الاتية:

ISSN (1681- 6870)

$$\hat{\pi}(X) = \frac{e^{(\bar{\beta}_1 x_{i1} + \bar{\beta}_2 x_{i2} + \bar{\beta}_7 x_{i7} + \bar{\beta}_9 x_{i9} + \bar{\beta}_{10} x_{i10})}}{1 + e^{(\bar{\beta}_1 x_{i1} + \bar{\beta}_2 x_{i2} + \bar{\beta}_7 x_{i7} + \bar{\beta}_9 x_{i9} + \bar{\beta}_{10} x_{i10})}}$$

$$\hat{\pi}(X) = \frac{e^{\begin{pmatrix} -0.072X_{i1} + 0.096X_{i2} \\ +0.816X_{i7} + 0.078X_{i9} \\ +0.004X_{i10} \end{pmatrix}}}{1 + e^{\begin{pmatrix} -0.072X_{i1} + 0.096X_{i2} \\ +0.816X_{i7} + 0.078X_{i9} \\ +0.004X_{i10} \end{pmatrix}}}$$

جدول (5) تصنيف بيانات العينة باستعمال الانموذج المقدر بطريقة $(WLSE)$

حالة المريض		التنبؤ		
		عدم حدوث الوفاة 0 $\hat{\pi} < 0.5$	حدوث الوفاة 1 $\hat{\pi} \geq 0.5$	المجموع
المشاهدة	عدم حدوث الوفاة $Y = 0$	43	3	46
	حدوث الوفاة $Y = 1$	2	62	64
	المجموع	45	65	110
دقة النموذج		93.5		
حساسية النموذج		96.9		
نسبة التصنيف الصحيح		95.5		

يتضح في الجدول (5) ان انموذج الانحدار اللوجستي الثنائي قام بتصنيف البيانات(المتغيرات المعنوية) المؤثرة على الوفاة فقط من خلال ايجاد القيم الاحتمالية المقدرة لأنموذج الانحدار اللوجستي الثنائي و انه صنف43 مصاب من اصل 46 ممن لا توجد لديهم حالة الوفاة تصنيفا صحيحا, وبلغت نسبة التصنيف الصحيح لحالة عدم حدوث الوفاة 94% للمصابين بأمراض القلب، و صنف 62 مصاب من اصل 64 ممن حدثت معهم حالة الوفاة حيث بلغت نسبة التصنيف الصحيح لحالة حدوث الوفاة 97% 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 96% وتعد هذه النسبة مقبولة جدا ، اي ان نسبة الخطأ تساوي 0.04 (4% تم تصنيفهم بصورة خاطئة).

ISSN (1681- 6870)

يتضح في الجدول (5) ان انموذج الانحدار اللوجستي الثنائي قام بتصنيف البيانات(المتغيرات المعنوية)المؤثرة على الوفاة فقط من خلال ايجاد القيم الاحتمالية المقدرة لأنموذج الانحدار اللوجستي الثنائي و انه صنف 43 مصاب من اصل 46 ممن لا توجد لديهم حالة الوفاة تصنيفا صحيحا ,وبلغت نسبة التصنيف الصحيح لحالة عدم حدوث الوفاة94% للمصابين بأمراض القلب، و صنف 62 مصاب من اصل 64 ممن حدثت معهم حالة الوفاة حيث بلغت نسبة التصنيف الصحيح لحالة حدوث الوفاة % 97 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 % 96 ، وتعد هذه النسبة مقبولة جدا ، اي ان نسبة الخطأ تساوي0.04 (4%تم تصنيفهم بصورة خاطئة).

الاستنتاجات (Conclusions)

1- العوامل التي كان لها التأثير الأكبر وبمعنوية اقل من 5% على متغير الاستجابة (حالة المصابين بأمراض القلب) هي (الردهة، الجنس، التدخين، السكر، الوزن) بالنسبة لأنموذج الانحدار اللوجستي الثنائي وهذا ما تم تأكيده من الأطباء المتخصصين ,وأشارت النتائج التي تم التوصل اليها في الدراسة انها متفقة مع الدراسات السابقة في المتغيرات التوضيحية الاكثر تأثيرا على حالة المريض .

2- نستنتج ان الطريقة (WLSE) لها القدرة على التصنيف ,أي انه يمكن استعمالها لتصنيف الحالات الجديدة للمرضى المصابين بأمراض القلب الى (حدوث وفاة ، عدم حدوث وفاة) اعتمادا على قيم المتغيرات التوضيحية لتلك الحالات.

3- من خلال البيانات الحقيقية والتي تمثل اسباب حدوث الوفاة لمرضى القلب في الجدول(2) والبيانات المقدرة الموضحة وكذلك من خلال قيم معامل التحديد والتي تم تقديرها باستعمال(WLSE) لأنموذج الانحدار اللوجستي الثنائي نلاحظ ان الانموذج ملائم لنمذجة حالات الاشخاص المصابين بأمراض القلب من حيث اسباب وفاتهم او بقائهم على قيد الحياة, اذ تتطابق القيمة المقدرة مع القيمة الحقيقية بشكل معقول ,كما يدل ذلك على دقة (WLSE) في تقدير معلمات هذا الانموذج ,وأن الانموذج من الممكن تقديره بدرجة عالية من الدقة.

المصادر

[1] Abdulqader, Q. M., (2015),"Comparison Of Discriminant Analysis and Logistic Regression Analysis: An Application or Caesarean Births and Natural Births Data" ,Zakho Tecnical Institute Duhok, Polytechnic University Duhok, Duhok , Iraq, *INAS*, pp.34-46.

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

[3] Chatterjee, S., Hadi, A., (2012) ,"Regression Analysis By Example", John Wiley ,INC.

[4] Ferrer, J. ,Wang, L., (1999)," Comparing the Classification Accuracy among ,Parametric Nonparametric, Discriminant Analysis and Logistic Regression Methods", ERIC , Quebec, Canada , pp.1-23

[5] Garson, D.,(2014),"Logistic Regression :Binary and Multinomial" ,ISBN , Retrieved from http: www.statisticalass ociates.com.

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

[7] Karim , A. J., (2013),"Statistical Study to Classify β-Thalassaemia diseases in Erbil City at Thalassaemia Center by Using ROC Curve Analysis", IJSR, pp.2955-2961.

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

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

[10] Rodriguez ,G.,(2007), "Logit Models for Binary Data" ,Chapter(3) ,Retrieved from,http://data.princeton.edu/wws509/notes/c3.pdf

[11] Soderstrom, R. , Leitner، W. (1997)." The Effects of Base Rate, Selection Ratio, Sample Size and Reliability of Predectors on Predective Efficiency Indic. Associated with Logistic Regression Models", ERIC, PP.1-25.

[12] Sukono, Sholahuddin, A. & et al., (2014)," Credit Scoring for Coopera of Financial Services Using Logistic Regression Estimated by Genetic Algorithm", AMS, pp. 45-57.

[13] Wuensch, K. , (2014) , " Binary Logistic Regression with SPSS" , Retrieved from WWW . Care . ecu . edu / psyc/ Wuensch / MV /Logistic SPSS, pp.1-29.