مجلة قەڵای زانست العلمية
مجلة علمية دورية محكمة تصدر عن الجامعة اللبنانية الفرنسية – اربيل، كوردستان، العراق
المجلد(٥) – العدد (٤)، شتاء ٢٠٢٠
رقم التصنيف الدولي: ISSN 2518-6558 (Print) - ISSN 2518-6566 (Online)

إستخدام أنموذج اللوجستي لدراسة أهم العوامل المؤثرة على مرض السكري حسب نوع المرض

- بحث مشترك -

كوثر سعيد طه
قسم الاحصاء و المعلوماتية, كلية الادارة والاقتصاد, جامعة صلاح الدين، أربيل، اقليم كورستان، العراق
kawther.hameen@gmail.com

درخشان جلال حسن
قسم الاحصاء و المعلوماتية, كلية الادارة والاقتصاد, جامعة صلاح الدين، أربيل، اقليم كورستان، العراق
كلية الادارة والاقتصاد, جامعة بيان، أربيل، اقليم كورستان، العراق
drakhshan.hassan@su.edu.krd

معلومات البحث

تاريخ البحث:
الاستلام:2019/8/5
القبول:2020/9/10
النشر:شتاء 2020

الكلمات المفتاحية:
Logistics, dependent Parametric, Regression, categorical, Polychromous, Multinomial, classification

Doi:

10.25212/lfu.qzj.5.4.21

الملخص

تم في هذا البحث أستخدام تحليل التمييز أنموذج اللوجستي وطريقة الإمكان الأعظم لتقدير معلمات الأنموج اللوجستي من خلال بيانات تمثل مجموعة من مرضى السكري النوع الاول و النوع الثاني تم الحصول عليها بواسطة أستمارة الاستبيان تم ملأها عن طريق المقابلات الشخصية مع المرضى و السجلات من مستشفى ليلى قاسم في مدينة أربيل للسنة 2018م ، والهدف الاساسي من هذا البحث هو دراسة أهم العوامل التي تمييز بين النوع الاول والثاني للمصابين بمرض السكري من خلال أستخدام أنموذج اللوجستي ومدى تأثير النوع الاول والثاني في زيادة نسبة الاصابة بمرض السكري.

وتوصلت الدراسة الى ملائمة أنموذج التمييز اللوجستي ومعنوية معلماته المقدرة في دراسة العوامل المؤثرة على مرض السكري النوع الاول والنوع الثاني اعتماداً على أختبار (Wald). وتبين بأن المتغيريين (العمر في بداية الاصابة و نسبة السكر) تعد مؤثرين في المتغير المعتمد والذي يمثل نوع المرض (السكري) وأن بقية المتغيرات المدروسة ليس لها تأثير. وأن نسبة التصنيف الصحيح هي (83.6%).

مجلة قەڵای زانست العلمية
مجلة علمية دورية محكمة تصدر عن الجامعة اللبنانية الفرنسية – اربيل، كوردستان، العراق
المجلد(٥) – العدد (٤)، شتاء ٢٠٢٠
رقم التصنيف الدولي: ISSN 2518-6558 (Print) - ISSN 2518-6566 (Online)

1- الفصل الاول:

1-1 المقدمة:-

يعد نموذج اللوجستك أمتداد لتحليل البيانات المصنفة و يفترض النموذج أن يكون حجم العينة كافيا بحيث تكون جميع الخلايا لها تكرار متوقع واحد على الأقل، والتحليل والتفسير للنموذج مشابه تماماً لأسلوب الأنحدار المتعدد، وأستخدمت دالة تمييز اللوجستك لتمثيل الأحتمال اللاحق كأساس للتمييز بين مجتمعين ثم عممت الى حالة أكثر من مجتمعين و يعتبر التمييز اللوجستي وسط بين الطرق المعلمية (Parametric) المعتمدة على التوزيع الطبيعي متعدد المتغيرات والطرق اللامعلمية (Non Parametric) حرة التوزيع لهذا السبب تدعى هذه الطريقة بالطريقة الشبه المعلمية (Semi-Parametric) مما جعل هذا النموذج ذو أهمية كبيرة في حل مسائل التمييز والأنحدار.

2-1 مرض السكري:

مرض السكري مرض مزمن يحدث عندما يعجز البنكرياس عن إنتاج الإنسولين بكمية كافية، أو عندما يعجز الجسم عن الاستخدام الفعال للإنسولين الذي ينتجه. والإنسولين هو هرمون ينظّم مستوى السكر في الدم .ويُعد فرط سكر الدم أو ارتفاع مستوى السكر في الدم من الآثار الشائعة التي تحدث جرّاء عدم السيطرة على داء السكري، ويؤدي مع الوقت إلى حدوث أضرار وخيمة في العديد من أجهزة الجسم، ولاسيما الأعصاب والأوعية الدموية. وينقسم الى قسمين وهما:

1-2-1 مرض السكري من النوع (النمط) الاول:

يتسم داء السكري من النمط الاول (الذي كان يُعرف سابقاً باسم داء السكري المعتمد على الإنسولين أو داء السكري الذي يبدأ في مرحلة الشباب أو الطفولة) بنقص إنتاج الإنسولين، ويقتضي تعاطي الإنسولين يومياً. ولا يُعرف سبب داء السكري من النمط الأول، ولا يمكن الوقاية منه باستخدام المعارف الحالية.

وتشمل أعراض هذا الداء فرط التبوّل، والعطش، والجوع المستمر، وفقدان الوزن، والتغيرات في البصر، والإحساس بالتعب. وقد تظهر هذه الأعراض فجأة.

2-2-1 مرض السكري من النوع (النمط) الثاني:

يحدث هذا النمط (الذي كان يُسمى سابقاً داء السكري غير المعتمد على الإنسولين أو داء السكري الذي يظهر في مرحلة الكهولة) بسبب عدم فعالية استخدام الجسم للإنسولين. وتحدث في معظمها نتيجة لفرط الوزن والخمول البدني.

مجلة قەڵای زانست العلمية
مجلة علمية دورية محكمة تصدر عن الجامعة اللبنانية الفرنسية – اربيل، كوردستان، العراق
المجلد(٥) – العدد (٤)، شتاء ٢٠٢٠
رقم التصنيف الدولي: ISSN 2518-6558 (Print) - ISSN 2518-6566 (Online)

وقد تكون أعراض هذا النمط مماثلة لأعراض النمط الاول، ولكنها قد تكون أقل وضوحاً في كثير من الأحيان. ولذا فقد يُشخّص الداء بعد مرور عدة أعوام على بدء الأعراض، أي بعد حدوث المضاعفات.

وهذا النمط من داء السكري لم يكن يُصادف إلا في البالغين حتى وقت قريب، ولكنه يحدث الآن في صفوف الأطفال أيضاً. (World Health Organization2018)

2- الفصل الثاني الجانب النظري:

2-1 التميز اللوجستي:

تحليل التمييز اللوجستي يعد من الاساليب الاحصائية المهمة التي يمكن استخدامها في كثير من مجالات الحياة مثلاً في مجال الطب والبايولوجيا والجيولوجيا والزراعة وعندما يكون المتغير التابع ذات طبيعة ثنائية أو مصنفة ، شكل العلاقة يأخذ صيغة الدالة التوزيعية لنموذج اللوجستي وتستخدم في حالات كثيرة وخاصة في الابحاث الحياتية لتمثل العلاقة بين أحتمال الأستجابة ومقياس الجرعة وذلك لشموليتها في التطبيق، أما المتغيرات التوضيحية فتأخذ القيم الكمية أو النوعية أو كلاهما معاً. (مولود 2000، Menard 2002)

أسباب أهمية التمييز اللوجستي في مسائل التمييز والانحدار:-

1- يتطلب القليل من الأفتراضات.
2- أن الصيغ اللوجستية تنتج عن تنويع واسع من الأفتراضات التحتية حول المتغيرات التوضيحية.

أ- القيمة تكون ثنائية تأخذ القيمة صفر أو واحد.

ب- بعضها طبيعية متعددة المتغيرات.

ت- بعضها متغيرات متصلة أو منفصلة.

ث- بعضها متغيرات مرتبة حسب درجة الخطورة.

ج- بعضها متغيرات مصنفة.

3- سهولة أستخدامه: فبعد تقدير معلمات النموذج فأن تصنيف مفردة جديدة يتطلب حساب دالة خطية واحدة.

مجلة قەڵای زانست العلمية
مجلة علمية دورية محكمة تصدر عن الجامعة اللبنانية الفرنسية – اربيل، كوردستان، العراق
المجلد(٥) – العدد (٤)، شتاء ٢٠٢٠
رقم التصنيف الدولي: ISSN 2518-6558 (Print) - ISSN 2518-6566 (Online)

4- يتحقق تحت عائلة واسعة من التوزيعات منها الطبيعية ذات تباينات متساوية أو متغيرات ذات توزيعات ثنائية الحدين وغيرها.
5- قابليته على التقدير بغض النظر عن أسلوب المعاينة سواء كان أستطلاعيا (Prospective) أو أسترجاعيات (Retrospective) ويتمثل أسلوب المعاينة الأستطلاعية بعدد كبير من المتطوعين من المفردات أو الحالات المرضية ليتم متابعتها لمدة طويلة يمكن من خلالها أن يقع عدد كبير نسبيا من المفردات ضحايا للمرض أو المشكلة موضوع الدراسة. أما أسلوب المعاينة الأسترجاعية فهي تكون من خلال سجلات المفردات أو الحالات المطلوب دراساتها خلال مدة معينة.

2-2 نماذج الاختيار المتقطع Discrete choice models:-

تطبق هذه النماذج عندما تكون المتغيرات المعتمدة (متغيرات الاستجابة) و صفية ومحددة (Qualitative and limited dependent variables) اي تلك التي يتم التعبير عنها بمجموعة محددة من الصفات أو الحقائق (characteristic) أو انها محددة (مقيدة) بقيود لايمكن تجاهلها، اذ ليس دائما يكون توقع المتغير المعتمد قيمة عددية ويأخذ اي قيمة محصورة بين $(-\infty,\infty)$ كما في المنحني الطبيعي، وهنالك العديد من التطبيقات التي تستخدم كل من المتغيرات المعتمدة المحددة والوصفية ونماذج الاختيار المتقطع في مجالات عدة تتضمن الاقتصاد، المال، التسويق، العلوم الاجتماعية،...الخ.

وتعد نماذج الاستجابة الثنائية والمتعددة حالة خاصة من هذه النماذج، فمثلاً أنموذج ثنائي الاستجابة يطبق عندما يأخذ المتغير المعتمد قيمتين فقط $[0,1]$ والتي قد تشير الى نمط الاستجابة لمؤشرما (كالادوية و المنشطات) اي انه محدد بحالة النجاح و الفشل على التوالي.(Cox 1970 ، البياتي 2005)

اذ يتبع احتمال حدوث الاستجابة توزيع برنولي (Bernoulli distribution) وكما يلي :

$$Y \sim B(1,P)$$
$$E(Y_i) = \Pr[Y_i=1] = P; \Pr[Y_i=0] = 1-P$$
$$V(Y_i) = E[(Y_i-P)]^2 = P(1-P)$$

وتعد نماذج الاختيار الثنائية (binary choice models) كأحد الطرق المهمة في تحليل متغيرات الاستجابة الثنائية(binary or dichotomous response variable) ، بحيث تكون استجابة المشاهدة(i) ، ويمكن التعبير عنها بمتغير متقطع (Discrete variable) كما ان القيمة المتوقعة للمتغير المعتمد لاتعد دالة خطية بالنسبة للمتغيرات التوضيحية (كما في معادلة الانحدار الكمية) (Quantitative

مجلة قەڵای زانست العلمية
مجلة علمية دورية محكمة تصدر عن الجامعة اللبنانية الفرنسية – اربيل، كوردستان، العراق
المجلد(٥) – العدد (٤)، شتاء ٢٠٢٠
رقم التصنيف الدولي: ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

(Regression equation) بل تكون على شكل دالة احتمالية خطية محددة بالفترة [0,1] وللتعبير عن أنموذج التمييز لمثل هكذا متغيرات يكون بالصيغة الاتية:

$$y_i^*=X_i'\beta+\varepsilon_i \quad .....(2-1) \qquad i=1,2,.....,n$$

y_i^* : يمثل المتغير الوهمي (الكامن) (Latent variable) ليعبر عن حالة الاستجابة للمتغير المعتمد y_i (واحد عند حدوث الاستجابة وصفر في حالة عدم حدوث الاستجابة)

X_i' : يمثل متجه المتغيرات التوضيحية.

β : يمثل متجه المعلمات الخاص بالأنموذج.

ε_i : يمثل الخطأ العشوائي.

عندما يأخذ المتغير المعتمد القيم الاتية:

$$y_i=\begin{cases}1 & y_i^*>o \\ & if \\ 0 & o.w\end{cases}$$

وان احتمالية ان يكون قيمة المتغير المعتمد الواحد الصحيح لأنموذج اللوجستي الثنائي هو

$$P(y_i=1)=\frac{\exp(X_i'\beta)}{1+\exp(X_i'\beta)} \qquad ;i=1,2,....,n \quad .....(2-2)$$

حيث β ، X_i' معرفة كما في (2-1)

وان احتمالية الاختيار البديل نحصل عليها من خلال $P(y_i=0)=1-P(y_i=1)$

اما عندما يأخذ المتغير المعتمد (متغير الاستجابة) قيما متقطعة متعددة، اي عندما تكون استجابة المتغير المعتمد تصنف ضمن اكثر من مجموعتين فان أنموذج اللوجستي متعدد الحدود يستخدم لتحليل متغيرات الاستجابة المصنفة غير المرتبة (Un ordered categorical response variable)

مجلة قەڵای زانست العلمية
مجلة علمية دورية محكمة تصدر عن الجامعة اللبنانية الفرنسية – اربيل، كوردستان، العراق
المجلد(٥) – العدد (٤)، شتاء ٢٠٢٠
رقم التصنيف الدولي: ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

ولحساب احتمالية الاختيار (الاستجابة) (j) خلال (g) من البدائل فأن أنموذج اللوجستي متعدد الحدود (اوالشرطي) يأخذ الصيغة الاتية : (Jarajda 2003)

$$p(y_i = j) = \frac{\exp(X_{ij}'\beta)}{1+\exp(X_{ij}'\beta)} \qquad j = 1,2,.....,g-1 \qquad .....(2-3)$$

X_{ij} : يمثل متجه المتغيرات التوضيحية اللاستجابة رقم.(j)

β : يمثل متجه المعلمات الخاص بالأنموذج.

وفيما يلي سنتطرق بالتفصيل الى استخدام أنموذج (اللوجستي) متعدد الحدود لتحليل المتغيرات المصنفة المتعددة الاستجابة (Polychromous variable) .

3-2 طرائق تقدير المعالم في النموذج:-

يتم الحصول على تقديرات معلمات النموذج بثلاثة طرائق هي-:

1- طريقة تقديرات الامكان الأعظم ML .
2- طريقة تقديرات المربعات الصغرى التعاقبية MLS .
3- طريقة تقديرات تصغير مربع كاي MCS .

تم التطرق الى طريقة تقديرات الامكان الاعظم فقط لأنها أستخدمت في تقدير معلمات النموذج اللوجستي في هذا البحث .

4-2 تقديرات الامكان الأعظم Maximum Likelihood Method:-

ليكن لدينا Z من المتجهات المستقلة ذات توزيع متعدد الحدودMultinomial :

موجه ينتمي الى واحدة من g من مجاميع الاستجابة Response Categories نرمز للمشاهدات بـ $(Z_1,\ldots,Z_n)$ حيث كل مشاهدة Z_i محدودة بواحدة من g من المجاميع ($Z_i = Z_{i1},.....Z_{ig}$) وأن $\sum_j X_{ij} = n_i$ يكون محدد لكل i تكون دالة الأمكان الاعظم الشرطية لتوزيع متعدد الحدود بالشكل الآتي:-

$$L = \prod_{i=1}^{N} \sum_{j=1}^{g} [P(X_i)]^{z_{ij}} \qquad .....(2-4)$$

حيث أن-:

مجلة قەڵای زانست العلمية
مجلة علمية دورية محكمة تصدر عن الجامعة اللبنانية الفرنسية – اربيل، كوردستان، العراق
المجلد(٥) – العدد (٤)، شتاء ٢٠٢٠
رقم التصنيف الدولي: ISSN 2518-6558 (Print) - ISSN 2518-6566 (Online)

$P_j(X_i)$ = نسبة أو احتمال الانتماء إلى المجموعة j.

بأخذ المشتقة الاولى للوغاريتم الاحتمال الاعظم للمعادلة (2- 4) و مساواتها بالصفر نحصل على المعادلات الطبيعية الاتية:-

$$\sum_{i}^{N}\left[Z_{ij}-n_i\hat{P}_j(X)\right]X_{ik}=0$$

و أن جذور هذه المعادلات تمثل قيم المعالم التقديرية الناتجة عن عملية التعظيم – وتكون هذه المعادلات غير خطية في المعالم و لذلك نلجأ إلى أستخدام طريقة نيوتن رافسون التكرارية وبعد بضع دورات تعاقبية تنتج تقديرات مناسبة وتكون تقديرات المعالم بطريقة نيوتن رافسون وفق الصيغة الآتية:-

$$B(t+1)=B(t)+\{X'v(t)X\}^{-1}X'r(t) \qquad .....(2-5)$$

حيث أن:-

t :(عدد الدورات التعاقبية).

r : (هو متجه البواقي ذو بعد $[N(g-1),1]$) .

v : هو $Diag(v_i.....v_n)$ وان كل v_i هو عبارة عن مصفوفة مربعة تكتب بالشكل الآتي:-

$$v_i=n_i\{P_s(X_i)(\delta_{st}-P_i(X_i))\}St$$

وان δ_{st} هو دلتا كرونكر Kronecker Delta ويعرف بالآتي:-

$$\delta_{st}=\begin{bmatrix}1 & S=t\\ 0 & S\neq t\end{bmatrix}$$

ويمكن أن نبدأ من قيم أبتدائية بمساواة موجه المعالم بالصفر وفي الحالة الثنائية يمكن أستخدام الرسم البياني أو طريقة المربعات الصغرى الأعتيادية أو أستخدام تقديرات دالة التمييز الخطية كقيم أبتدائية في تقدير المعالم حيث أن أستخدامها سوف يقلل من عدد الدورات التعاقبية و عند الحصول على التقارب المطلوب بين الدورات، تكون هذه التقديرات المثلى المطلوبة. (عباس 2012 ، Menard 2002)

مجلة قەڵای زانست العلمية
مجلة علمية دورية محكمة تصدر عن الجامعة اللبنانية الفرنسية – اربيل، كوردستان، العراق
المجلد(٥) – العدد (٤)، شتاء ٢٠٢٠
رقم التصنيف الدولي: ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

5-2 تقييم أداء الآنماذج -: Evaluating the Performance of the models

لقد تم استخدام الاختبارات الآتية:

1-5-2 اختبار مربع كاي لبيرسون Pearson chi-square:-

اقترح العالم Pearson (1900) استخدام هذا الاختبار كأحد مقاييس جودة التوفيق ويعني مدى اقتراب قيم المشاهدة من خط التقدير، إذ ان مدى التوافق بين الجزء المشاهد والجزء المتوقع يزودنا بدليل الملائمة أو عدم النموذج لاختبار الفرضية الاتية:

H_0: الأنموذج ملائم للبيانات

H_1: الأنموذج غير ملائم للبيانات

وللعينات ذات الحجوم الكبيرة فان توزيع الاحصاءة يقترب من توزيع χ^2 (approximate χ^2) وبتطبيق هذا الاختبار لتحديد مدى ملائمة أنموذج اللوجستي متعدد الحدود باستخراج قيمة الاحصاءة وحسب الصيغة الآتية:

$$\chi^2 = \sum_{i=1}^{n} \sum_{j=1}^{g} \frac{\left(Y_{ij} \hat{P}_j(X_{ij})\right)^2}{\hat{P}_j(X_{ij})(1-\hat{P}_j(X_{ij}))} \quad .....2-6$$

حيث $P_j(X_{ij})$ معرفة كما في المعادلة (2-4).

2-5-2 احصاءة نسبة الامكان Likelihood ratio statistics:-

وتعد هذه الاحصاءة احد اختبارات حسن المطابقة ويستخدم لتحديد معنوية الأنماذج احصائياً لاختبار الفرضية التالية:

$$H_0 : \beta_1 = \beta_2 = ... = \beta_k = 0 \quad .....(2-7)$$

H_1: *at least two of them are not equal zero.*

مجلة قەڵای زانست العلمية
مجلة علمية دورية محكمة تصدر عن الجامعة اللبنانية الفرنسية – اربيل، كوردستان، العراق
المجلد(٥) – العدد (٤)، شتاء ٢٠٢٠
رقم التصنيف الدولي: ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

والصيغة العامة اللاحصاءة هي:

$$LR(k) = -2\left[LnL(\alpha) - LnL(\alpha,\beta)\right]$$

إذ ان $LnL(\alpha)$: يمثل لوغاريتم دالة الامكان اللأنموذج المختزل (Reduced model) الذي يحوي على معلمة التقاطع (Intercept parameter) فقط.

وان $LnL(\alpha,\beta)$: يمثل لوغاريتم دالة الامكان اللأنموذج الكلي (Final model) .

وهذه الاحصاءة (LR) تتوزع χ^2 بدرجة حرية مساوية لـــ K إذ ان K تمثل عدد المتغيرات التوضيحية في الأنموذج. (البياتي، 2005)

3-5-2 اختبار معنوية معلمات الأنموذج المقدر:-

في عام (1943) اقترح Wald استخدام هذه الاحصاءة لبيان معنوية المعلمة المقدرة بطريقة الامكان الاعظم لاختبار الفرضية الاتية:

$$H_0: \beta_k = \beta_{k0}$$
$$H_1: \beta_k \neq \beta_{k0}$$

ويتم اختبار الفرضية القائلة بعدم وجود تأثير لمعلمات الأنموذج $H_0: \beta_k = 0$ باستخراج قيمة الاحصاءة وحسب الصيغة الاتية:

$$W = \left(\hat{\beta}_k - \hat{\beta}_{ko}\right)^2 / H_{kk} \qquad .....(2-8)$$

k : تمثل عدد المتغيرات التوضيحية في الأنموذج.

$\hat{\beta}_k$: تمثل المعلمة المقدرة بطريقة الامكان الاعظم.

H_{kk} : يمثل مربع الخطأ المعياري المقدرASE .

مجلة قەڵای زانست العلمية
مجلة علمية دورية محكمة تصدر عن الجامعة اللبنانية الفرنسية – اربيل، كوردستان، العراق
المجلد(٥) – العدد (٤)، شتاء ٢٠٢٠
رقم التصنيف الدولي: ISSN 2518-6558 (Print) - ISSN 2518-6566 (Online)

وهذه الاحصاءة هي عبارة عن مربع نسبة المعلمة المقدرة الى الخطأ المعياري المقدر ويفضل استخدامها للعينات ذات الحجوم الكبيرة وتتوزع χ^2 وبدرجة حرية تساوي واحد. (البياتي 2005 ، Agresti 1990)

4-5-2 اختبارQ:-

اقترح اختبار Q من قبل (Liu & Dyer) عام (1988) لأنموذج الانحدار اللوجستي وهي شبيهة بمعامل التحديد R^2 في أنموذج الانحدار الخطي وتحسب كالاتي:

بعد تقدير الأنموذج يتم ترتيب المشاهدات تنازلياً وفق الاحتمالات اللاحقة للاستجابة (Yi=1) ويمكن التعبير عن Q كالاتي:

$$Q = \frac{2(\overline{R}_1 - \overline{R}_0)}{n} \qquad \text{..... } (2-9)$$

حيث ان:

$\overline{R}_1$: متوسط n_1 من الرتب ذات الاستجابة $y=1$.

$\overline{R}_0$: متوسط $(n-n_1)$ من الرتب ذات الاستجابة $y=0$

يكون الحد الأعلى إلى Q مساوياً لـ(واحد) عندما يكون n_1 من المشاهدات ذات الرتب الاعلى جميعها ذات الاستجابة $y=1$. ويكون الحد الادنى الى Q ذات توقع يساوي صفر عندما تتوزع المشاهدات بشكل عشوائي بين (0,1) وفقاً لرتبتها. أن استخدامها الرئيسي يكون في المساعدة لاختيار أنموذج بين نماذج مختلفة تحتوي على عدد مختلف من المتغيرات التوضيحية.

ويمكن استخراج قيمة موزونة لـ Q بأعطاء اوزان للرتب لكل حالة. وان قيمة Q تتأثر كثيراً بالحالات الشاذة التي تكون القيم التنبؤية لها واطئة وخاصة عندما يكون عدد الحالات $y=1$ صغيراً نسبياً. (طه2006، Hamad 2016)

مجلة قەڵای زانست العلمية
مجلة علمية دورية محكمة تصدر عن الجامعة اللبنانية الفرنسية – اربيل، كوردستان، العراق
المجلد(٥) – العدد (٤)، شتاء ٢٠٢٠
رقم التصنيف الدولي: ISSN 2518-6558 (Print) - ISSN 2518-6566 (Online)

2 -6 خطأ التصنيف Miss classification :-

أن تقييم تحليل نموذج اللوجستك عند استخدام تقديرات الامكان الأعظم لدالة الاحتمال الشرطية يتم عن طريق احصاءات جودة التوفيق للأحتمالات اللاحقة المقدرة من النموذج أو عن طريق تقدير احتمال خطأ التصنيف وتوجد عدة طرق لتقدير احتمال خطأ التصنيف منها.

طريقة اعادة التعويض Re-Substitution method:-

عندما تكون n_j هي المشاهدات التي تعود الى المجموعة j حيث أن $j=1,...g$ وهو عدد المجاميع تحت الدراسة وعلى افتراض أن n_{ij} هو عدد المشاهدات في المجموعة (Y_i, j) وصنفت على انها تعود الى المجموعة Y_i فان تقدير احتمال خطأ التصنيف مساوي: (مولود 2000، Menard 2002)

$$\hat{P}_{ij}=\frac{n_{ij}}{n_j}$$

ويكون احتمال خطأ التصنيف الكلي:-

$$\hat{P}=\frac{(n_{ij}+n_{ji})}{(n_j+n_i)}$$

3- الفصل الثالث الجانب التطبيقي:

3-1 المقدمة:

لقد تم جمع البيانات بواسطة أستمارة الأستبيان و معلومات داخل السجلات للمصابين بالمرض السكري لكلا النوعين من مستشفى ليلى قاسم في مدينة أربيل للعام 2018م من (506) مريض وكانت طبيعة المعلومات بعضها كمية وبعضها نوعية، وقد تم أستخراج نتائج التحليلات بالأعتماد على البرنامج الإحصائي الجاهزة (Statistical Package for Social Sciences). وتضمن هذا الفصل الجانب التطبيقي وهو أهم جزء من البحث وتمت تطبيق أنموذج اللوجستي للبيانات التي تم الحصول عليها للمصابين بالمرض السكري لكلا النوعين.

3-2 وصف البيانات (الاحصاء الوصفي):-

ويتم تسمية (وصف) المتغيرات المستخدمة بالشكل الآتي:-

مجلة قەڵای زانست العلمية
مجلة علمية دورية محكمة تصدر عن الجامعة اللبنانية الفرنسية – اربيل، كوردستان، العراق
المجلد(٥) – العدد (٤)، شتاء ٢٠٢٠
رقم التصنيف الدولي: ISSN 2518-6558 (Print) - ISSN 2518-6566 (Online)

X_1 : العمر.

X_2 : الجنس: 1: الذكر.

2: الانثى.

X_3 : الوزن.

X_4 : نوع العمل: 1: ربت بيت (بدون عمل).

2: كاسب.

3: موظف (سائق).

4: طالب.

5: متقاعد.

6: معلم (مشرف، مدرب).

X_5 : الحالة الاجتماعية: 1: متزوج.

2: أعزب.

3: مطلق.

X_6 : العامل الوراثي: 0: لايوجد.

1: يوجد.

X_7 : العمر في بداية الاصابة.

X_8 : نسبة السكر.

X_9 : أمراض أخرى قبل المرض: 0: لايوجد.

1: ضغط الدم العالي.

مجلة قەڵای زانست العلمية
مجلة علمية دورية محكمة تصدر عن الجامعة اللبنانية الفرنسية – اربيل، كوردستان، العراق
المجلد(٥) – العدد (٤)، شتاء ٢٠٢٠
رقم التصنيف الدولي: ISSN 2518-6558 (Print) - ISSN 2518-6566 (Online)

2: كولسترول.

Y_i: نوع المرض السكري: 1: نوع الثاني (الاقراص).

2: نوع الاول (الانسولين).

الجدول (3-1) يوضح جزء من البيانات الخام للعينة قيد الدراسة

ت	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	y_i
1	54	2	75	1	1	2	4	100	0	1
2	38	2	74	1	1	1	15	213	1	2
3	52	2	88	1	1	2	2	218	2	1
4	60	1	79	5	1	2	5	182	0	1
5	65	2	96	1	1	1	18	270	0	2
6	53	2	106	1	1	2	3	295	1	1
7	72	1	80	5	3	2	2	196	1	1
8	57	2	86	1	3	2	6	122	2	1
9	50	2	76	1	3	2	3	270	2	1
10	57	2	160	1	3	2	1	130	1	1

3-3 تطبيق التمييز اللوجستي للبيانات:

يقوم برنامج (SPSS) بتحليل بيانات العينه بخطوات متعددة بدءاً بالخطوه الصفرية التي يكون فيها النموذج خالياً من المتغيرات التوضحيه. وادناه يوضح تلك النتائج.

مجلة قەڵای زانست العلمية
مجلة علمية دورية محكمة تصدر عن الجامعة اللبنانية الفرنسية – اربيل، كوردستان، العراق
المجلد(٥) – العدد (٤)، شتاء ٢٠٢٠
رقم التصنيف الدولي: ISSN 2518-6558 (Print) - ISSN 2518-6566 (Online)

الجدول(3-2) المتغيرات الداخلة في المعادلة

	B	S.E.	Wald	d.f	P-Value	Exp(B)
Step 0 Constant	-1.586	0.118	179.537	1	0.000	0.205

الجدول (3-2) يبين المتغيرات الداخلة للمعادلة في الخطوة الصفرية التي يكون فيها النموذج خالياً من المتغيرات التوضحية.

الجدول(3-3) يوضح المتغيرات التي أستبعدت من نموذج اللوجستي في المرحلة الصفرية

الخطوة الصفر للمتغييرات	Score	d.f	P-Value
العمر	4.296	1	.038
الجنس	.120	1	.729
الوزن	.945	1	.331
نوع العمل	.333	1	.564
الحالة الاجتماعية	.030	1	.862
العامل الوراثي	1.657	1	.198
العمر في بداية الاصابة	10.792	1	.001
نسبة السكر	13.889	1	.000
أمراض أخرى قبل المرض	1.778	1	.182
Overall Statistics	30.798	9	.000

الجدول (3-3) يبين فيها المتغيرات المستقلة التي أستبعدت من نموذج اللوجستي في المرحلة الصفرية.

مجلة قەڵای زانست العلمية
مجلة علمية دورية محكمة تصدر عن الجامعة اللبنانية الفرنسية – اربيل، كوردستان، العراق
المجلد(٥) – العدد (٤)، شتاء ٢٠٢٠
رقم التصنيف الدولي: ISSN 2518-6558 (Print) - ISSN 2518-6566 (Online)

الجدول (3-4) يوضح عدد الدورات التكرارية

تكرار		-2 Log likelihood	المعاملات									
			الثوابت	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
Step 1	1	443.656	-2.771	.010	.031	-.005	.023	.046	.204	.032	.003	-.119
	2	431.039	-3.989	.018	.035	-.008	.035	.079	.341	.051	.005	-.204
	3	430.610	-4.272	.020	.031	-.009	.037	.089	.379	.056	.005	-.229
	4	430.610	-4.284	.020	.031	-.010	.037	.089	.381	.056	.005	-.230
	5	430.610	-4.284	.020	.031	-.010	.037	.089	.381	.056	.005	-.230

الجدول (3-4) يوضح عدد الدورات التكرارية، نلاحظ من الجدول قد توصلنا الى نموذج اللوجستي الأفضل في الخطوة الأولى بخمسة الدورات التكرارية.

الجدول(3-5) اختبار كاي سكوير للمرحلة الاولى

		قيمة كاي سكوير	الدرجة الحرية	P-Value
Step 1	Step	30.684	9	.000
	Block	30.684	9	.000
	Model	30.684	9	.000

الجدول (3-5) يوضح معنوية النموذج من خلال احصاءة كاي سكوير، وبما ان القيم P-Value اقل من ($\alpha = 0.05$) فإن النموذج معنوي في الخطوة الاولى , أي المتغير(نوع المرض السكري) قيد الاختبار يكون معنوياً.

ولمعرفة قوة الأنموذج نستخدم R^2 لأنموذج اللوجستي الذي يفسر قوة تأثير المتغيرات التوضيحية على المتغير التابع.

الجدول(3-6) قيم R^2 لأنموذج اللوجستي

مجلة قەڵای زانست العلمية
مجلة علمية دورية محكمة تصدر عن الجامعة اللبنانية الفرنسية – اربيل، كوردستان، العراق
المجلد(٥) – العدد (٤)، شتاء ٢٠٢٠
رقم التصنيف الدولي: ISSN 2518-6558 (Print) - ISSN 2518-6566 (Online)

Step	-2 Log likelihood	Cox & Snell R^2	Nagelkerke R^2
1	430.610a	2890.	3970.

نلاحظ في الجدول (3-6) وبالاعتماد على احصاءة (Cox & Snell R^2)، وكذلك احصاءة (Nagelkerke R^2)، واختبار النسبة الامكان الاعظم ، ولمعرفة ان هذا الفرق معنوي ام لا يستخدم اختبار كاي سكوير للنموذج اللوجستي كالآتي:

H_0: الأنموذج ملائم

H_1: الأنموذج غير ملائم

الجدول(3-7) اختبار كاي سكوير للنموذج اللوجستي

Hosmer and Lemeshow Test			
Step	Chi-square	Df	P-Value
1	10.120	8	0.257

من الجدول (3-7) نلاحظ بأن قيمة P-Value=0.257 أكبر من ($\alpha = 0.05$) أذن لن نرفض فرضية العدم وهذا يعني بأن النموذج ملائم للبيانات في الخطوة الاولى.

الجدول(3-8) القيم المشاهدة والمتوقعة للنموذج اللوجستي

	النوع الاول (الانسولين)		النوع الثاني (الحبوب)		Total
	قيمة المشاهدات	القيمة المتوقعة	قيمة المشاهدات	القيمة المتوقعة	
1	1	2.880	50	48.120	51
2	3	4.265	48	46.735	51
3	5	5.193	46	45.807	51
4	7	6.060	44	44.940	51

مجلة قەڵای زانست العلمية
مجلة علمية دورية محكمة تصدر عن الجامعة اللبنانية الفرنسية – اربيل، كوردستان، العراق
المجلد(٥) – العدد (٤)، شتاء ٢٠٢٠
رقم التصنيف الدولي: ISSN 2518-6558 (Print) - ISSN 2518-6566 (Online)

5	12	6.976	39	44.024	51
6	9	8.206	42	42.794	51
7	5	9.307	46	41.693	51
8	11	11.115	40	39.885	51
9	17	13.741	34	37.259	51
10	16	18.257	31	28.743	47

من الجدول (3-8) نلاحظ القيم المشاهدة والمتوقعة للنموذج اللوجستي لكلا النوعين من المرض السكري.

الجدول(3- 9)المتغيرات الداخلة في النموذج

Variables	B	S.E	Wald	Df	P-Value	Exp(B)
العمر	.020	.012	2.736	1	.098	1.020
الجنس	.031	.318	.009	1	.923	1.031
الوزن	-.010	.010	.946	1	.331	.991
نوع العمل	.037	.093	.161	1	.688	1.038
الحالة الاجتماعية	.089	.156	.329	1	.566	1.093
العامل الوراثي	.381	.256	2.209	1	.137	1.463
العمر في بداية الاصابة	.056	.018	9.672	1	.002	1.058
نسبة السكر	.005	.002	11.495	1	.001	1.005
أمراض أخرى قبل المرض	-.230	.161	2.050	1	.152	.794
Constant	-4.284	1.477	8.413	1	.004	.014

مجلة قەڵای زانست العلمية
مجلة علمية دورية محكمة تصدر عن الجامعة اللبنانية الفرنسية – اربيل، كوردستان، العراق
المجلد(٥) – العدد (٤)، شتاء ٢٠٢٠
رقم التصنيف الدولي: ISSN 2518-6558 (Print) - ISSN 2518-6566 (Online)

الجدول(3-9) يبين المتغيرات التوضيحية الداخلة في النموذج، اذ نلاحظ المتغيرات يعرض نتائج تقدير أنموذج التمييز اللوجستي وأيظاً نلاحظ:

ان قيم اختبار(wald) يمثل قيم اختبار المعلمات للنموذج ويظهر بأن المتغير x7 (العمر في بداية الاصابة) والمتغير X_8 (نسبة السكر) هما المتغيرين المعنوين في الدراسة وذلك من خلال مقارنة القيم P-Value مع مستوى المعنوي (0.05) ، ان العمود P-Value يمثل معنوية تأثير المتغيرات على حالة المصاب، (عندما تكون قيم P-Value اقل من 0.05 لمتغير قيد الاختباريكون معنوياً)، لذلك فأن المتغيرين (عمر في بداية الاصابة ونسبة السكر) تعد مؤثرين على المتغير المعتمد (الاستجابة) والذي يمثل حالة المصاب بمرض السكري النوع الاول والثاني وان بقية المتغيرات الاخرى ليس لهم التأثيرعلى المرض حسب عينة البحث ، لذلك فان المتغيرين (العمر في بداية الاصابة ونسبة السكر) سوف تبقى في النموذج وبقية المتغيرات سوف تحذف من معادلة الانحدار (الانموذج اللوجستي).

الجدول (3-10) التصنيف للمرحلة الأخيرة لأنموذج اللوجستي

Classification Table

المشاهدات		Predicted		
		نوع المرض		percentage Correct
		الاول (الانسولين)	الثاني (الحبوب)	
نوع المرض	الاول (الانسولين)	4	82	4.7
	الثاني (الحبوب)	1	419	99.8
Overall Percentage				83.6

الجدول (3-10) هو جدول التصنيف للمتغير المعتمد إذ ان النسبه المئوية الكلية للدقة تنبؤ للمتغير المعتمد بالاعتماد على النموذج هي (83.6%).

4- الفصل الرابع الاستنتاجات:

4-1 الاستنتاجات:

أهم الاستنتاجات في البحث هو كالآتي:-

مجلة قەڵای زانست العلمية
مجلة علمية دورية محكمة تصدر عن الجامعة اللبنانية الفرنسية – اربيل، كوردستان، العراق
المجلد(٥) – العدد (٤)، شتاء ٢٠٢٠
رقم التصنيف الدولي: ISSN 2518-6558 (Print) - ISSN 2518-6566 (Online)

1- أن نموذج التمييز اللوجستي ملائم في دراسة العوامل المؤثرة على مرض السكري النوع الاول والنوع الثاني وتبين ان المعلمات المقدرة لنموذج اللوجستي معنوية في الخطوة الاولى وبخمسة الدورات التكرارية.
2- تبين أن المتغيرين (العمر في بداية الاصابة و نسبة السكر) تعد مؤثريين في المتغير المعتمد (الاستجابة) والذي يمثل حالة المصاب بالمرض السكري من نوع الاول ونوع الثاني وأن بقية المتغيرات ليس لهم تأثير.
3- أن نسبة التصنيف الصحيح المئوية الكلية لدقة التنبؤ للمتغير المعتمد (نوع مرض السكري) بالاعتماد على النموذج المقدرهي (83.6%).

المصادر:

أ- المصادر العربية:

1- البياتي، هبة أبراهيم صالح(2005): *" تحليل المسار في نموذج الانحدار اللوجستي مع تطبيق "* رسالة ماجستير في علوم الاحصاء، جامعة المستنصرية، كلية الادارة والاقتصاد.
2- مولود، كوردستان أبراهيم(2000):*"استخدام التحليل المميز لتشخيص أهم العوامل المؤثرة في التصنيف السريري لمرض القلب"* رسالة ماجستير في علوم الاحصاء، جامعة صلاح الدين/ أربيل، كلية الادارة والاقتصاد.
3- طه، كوثر سعيد(2006):*"استخدام أنموذج الانحدار اللوجستي في تحديد المتغيرات المؤثرة على بعض الامراض القلبية"* رسالة ماجستير في علوم الاحصاء، جامعة صلاح الدين/أربيل، كلية الادارة والاقتصاد.
4- عباس، علي خضر(2012):*"أستخدام نموذج الانحدار اللوجستي في التنبؤ بالدوال ذات المتغيرات الاقتصادية التابعة النوعية"*مجلة جامعة كركوك للعلوم الادارية والاقتصادية، المجلد(2) العدد(2).

ب- المصادر الانكليزية:

1- Agresti, A(1990):" *Categorical Data Analysis* " John Wiley Sons. Inc. New York.
2- Cox, D.R.(1970):" *The Analysis of Binary Data* " Chapman and Hall, London.
3- Hamad , Karzan Faidh. (2016):" *Applying Logistic Regression Model to Study Some Statistical Measurements in Distinguishing Shapes of Mass in Medical Images* " B.Sc. Statistical Sciences- Salahaddin University.
4- Menard, Scott. (2002):" *Applied Logistic Regression Analysi s* " Second Edition, Sage Publication, Inc.
5- Jarajda, Stepan and Box.(2003):"Econometrics of Panel Data and Limited Dependent Model" ecture Notes Series of Academy of Sciences of Th Czech Republic Economics Institut. http://www Cerge-ei-cs.

مجلة قەڵای زانست العلمية
مجلة علمية دورية محكمة تصدر عن الجامعة اللبنانية الفرنسية – اربيل، كوردستان، العراق
المجلد(٥) – العدد (٤)، شتاء ٢٠٢٠
رقم التصنيف الدولي: ISSN 2518-6558 (Print) - ISSN 2518-6566 (Online)

6- World Health Organization(30\10\2018): مركز وسائل \الصفحة الرئيسية : داء السكري \صحائف الوقائع \الإعلام Detail. www.who.int/ar/news-room/fact-sheets/detail/diabetes. Downloading From Internet [10/6/2019]

پوختە:

لەم توێژینەوەیەدا شیکردنەوەی جیاکردنەوەی لۆجستی بەکارهێنراوە هەروەها ڕێگای توانی بەرز بەکارهێنراوە بۆ خەملاندنی پارامیتەری مۆدێلی لۆجیستی، بۆ ئەم مەبەستە داتای پێویست وەرگیراوە بە شێوەی فۆرمی ڕاپرسی لە هەر نەخۆشێك بە چاوپێکەوتن لەگەڵ نەخۆش وە تۆماری هەر نەخۆشێك لە کۆمەڵێك نەخۆش کە نەخۆشی شەکرەیان هەیە لە جۆری یەکەم وە جۆری دووەم لە نەخۆشخانەی لەیلا قاسم لە پارێزگای هەولێر لە ساڵی 2018م، ئامانجی سەرەکی لەم توێژینەوەیەدا بریتیە لە لێکۆلینەوەی گرنگترین فاکتەرەکانی کە دەبێتە هۆی جیاکردنەوە لە نێوان هەردوو جۆری نەخۆشی شەکرە ئەمەش بە هۆی بەکارهێنانی مۆدێلی جیاکەرەوەی لۆجستی و کاریگەری جۆری یەکەم و دووەمی نەخۆشی شەکرە لە سەر زیادبوونی ڕێژەی تووش بوون بە نەخۆشی شەکرە.

گرنگترین ئەنجام بۆ توێژینەوەکە ئەوەیە مۆدیلی لۆجیکی گونجاوە لە لێکۆلینەوەی فاکتەرە کاریگەرەکانی سەر جۆری یەکەم و دووەمی نەخۆشی شەکرە وە بەهۆی تاقیکردنەوەی والد دەردەکەوێت خەملێنەرەکانی مۆدیلی لۆجستی معنەوین، وە دەرکەوت کە دوو گۆڕاو (تەمەن لە سەرەتای تووش بوون ، ڕێژەی شەکرە) بە کاریگەر هەژمار دەکرێن بۆ گۆڕاوی پشت پێ بەستراو کە ئەویش تووش بوانی نەخۆشی شەکرە جۆری یەکەم و دووەم وە گۆڕاوەکانی تر کاریگەریان نی یە لە سەر نەخۆشی شەکرە. وە ڕێژەی پۆلێنکردنی ڕاست یەکسان بوو بە (83.6%).

مجلة قەڵای زانست العلمية
مجلة علمية دورية محكمة تصدر عن الجامعة اللبنانية الفرنسية – اربيل، كوردستان، العراق
المجلد(٥) – العدد (٤)، شتاء ٢٠٢٠
رقم التصنيف الدولي: ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

Using logistics models to study the most important factors affecting diabetes by type of disease

Kawther Saeed Taha
Department Of Statistics & Informatics, College Of Administration& Economic, University Of Salahaddin, Erbil, Kurdistan Region, Iraq
kawther.hameen@gmail.com

Drakhshan Jalal Hassan
Department of Statistics & Informatics, College Of Administration& Economic, University Of Salahaddin, Erbil, Kurdistan Region, Iraq
College Of Administration& Economic, University Of Bayan, Erbil, Kurdistan Region, Iraq
Drakhshan.Hassan@Su.Edu.Krd

Keywords: *Logistics, dependent Parametric, Regression, categorical, Polychromous, Multinomial, classification*

Abstract

In this research, analysis of the discrimination of logistics is used, and the maximum likelihood method was used to estimate the parameters of the logistic model. This study has been done on diabetic patents visiting laila kassim diabetes center in hawler city in 2018. The data has been collected from both types of diabetes (type I and type II) via filing on questionnaire during personal interview with every patients and historical records. The main objective of this research is to study the main risk factors that differentiate between type I and type II and through using statistical analysis of the discrimination of logistics we can discover the effect of either factor on increasing risk (incidence) of diabetic.

We concluded in this study, the model of logistic discrimination is appropriate in the study of the factors affecting for both types of diabetes and found that the estimated parameters of the logistic model are significant through the test

رقم التصنيف الدولي: ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

(Wald). The two variables (the age of incidence of diabetes and blood sugar level) are significant factors on dependable variable which represent the incidence of diabetes while the other variables have no effect. And the correct classification ratio is (83.6%).