دراسة أهم العوامل المؤثرة التي تميّز بين الجنسين المصابين بمرض السكري بإستخدام أنموذج التمييز اللوجستي

(PP 281 - 291)

ID No. 2101

https://doi.org/10.21271/zjhs.22.4.17

بيخال صمد صديق

كلية الإدارة والاقتصاد/ جامعة صلاح الدين-اربيل

bikhal.sedeeq@su.edu.krd

الاستلام: 2017/03/01

القبول: 2018/04/16

النشر: 2018/09/10

ملخص

تم في هذا البحث استخدم تحليل التمييز اللوجستي وطريقة الامكان الأعظم لتقدير معلمات الأنموذج اللوجستي من خلال بيانات تمثل مجموعة من مرضى السكري والمؤخوذة من السجلات التأريخية في مستشفى ليلى قاسم/اربيل ، والهدف الاساسي من هذا البحث هو دراسة أهم العوامل التي تميز بين الجنسين المصابين بمرض السكري من خلال استخدام أنموذج التمييز اللوجستي ومدى تأثير الجنس في زيادة نسبة الاصابة بمرض السكري ودراسة وجود العلاقة من عدمها بين متغيرالجنس وأربعة من المتغيرات التي اعتمدها البحث وهي (نسبة السكر في الدم، نسبة الكوليسترول في الدم، نسبة اليوريا في الدم، نسبة التراكليسرايد).

وتوصلت الدراسة الى ملائمة أنموذج التمييز اللوجستي ومعنوية معلماته المقدرة في دراسة العوامل المؤثرة على مرض السكري اعتمادا على اختبار (Wald). وتبين بأن المتغيرات (مستوى الكرياتين في الدم، نسبة كوليسترول الجيد، مستوى هيموكلوبين) تعد مؤثرة في المتغير المعتمد والذي يمثل حالة المصاب بمرض السكري وان بقية المتغيرات المدروسة ليس لها تأثير، وايضا تبين بأنه توجد علاقة بين متغيرالجنس والمتغيرين (نسبة السكر في الدم ونسبة الكولسترول في الدم) بالإضافة الى عدم وجود علاقة بين متغيرالجنس والمتغيرين (نسبة اليوريا في الدم ونسبة التراكليسرايد) وان نسبة التصنيف الصحيح هي(78%).

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 تماما لأسلوب الأنحدار المتعدد ، وأستخدمت دالة تمييز اللوجستك لتمثيل الأحتمال اللاحق كأساس للتمييز بين مجتمعين ثم عممت الى حالة أكثر من مجتمعين و يعد التمييز اللوجستي وسطة 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 جعلها ذو أهمية كبيرة في حل مسائل التمييز والأنحدار

مرض السكري هو أحد اكثر الأمراض انتشاراً حول العالم، ويتزايد عدد المرضى كل عام بنسبة كبيرة، ويرجح أطباء منظمة الصحة العالمية أن هذا المرض مرشح لأن يكون السبب السابع للوفاة في العالم خلال العقدين القادمين ويرجع سبب هذا المرض الى عدم مقدرة البنكرياس على افراز الكمية المطلوبة من الأنسولين ويؤدي الى إضطرابات ومن أهمها ارتفاع نسبة السكر في الدم مما يؤدي الى أن الكلى لا تستطيع المحافظة على السكرالمطلوب وبالنتيجة يظهر في التحاليل الخاصة بالبول السكر بالنسبة معينة للمصاب بمرض السكري.

1- الجانب النظري

1-1 اختبار كاى سكوير للاستقلالية (Test for Independence)

يعتبر هذا الأختبار أحد تطبيقات توزيع مربع كاي المهمة جداً في الحياة العملية ، إن هذا الأختبار يختص في الكشف عما إذا كانت مستويات عاملين (صفتين) او أكثر مسقلة عن بعضها . وغالباً ما تكون هذه العوامل متغيرات نوعية ولمعرفة الاستقلالية بين المتغيرين نستخدم اختبار الفرضيات بشكل الأتي: (الهيتي 2006)

فرضية العدم: لاتوجد علاقة بين المتغيرين

الفرضية البديلة: توجد علاقة بين المتغيرين

والصيغة العامة للإختبار هي:

$$cal\,\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{m} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad \ldots.(1.1)$$

2-1 التميز اللوجستي

تحليل التمييز اللوجستي يعد من الأساليب الاحصائية المهمة التي يمكن استخدامها في كثير من مجالات الحياة مثلا في مجال الطب والجيولوجيا والبايولوجيا والزراعة وعندما يكون المتغير المعتمد ذات طبيعة ثنائية أو مصنفة شكل العلاقة يأخذ صيغة الدالة التوزيعية لنموذج اللوجستي وتستخدم في حالات كثيرة وخاصة في الأبحاث الحياتية لتمثيل العلاقة بين احتمال الأستجابة ومقياس الجرعة وذلك لشموليتها في التطبيق، اذا تستخدم في حالة كون المتغيرات كمية أو نوعية.

(مولود 2000 ، Menard 2002)

اسباب اهمية 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 وتأخذ 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.
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 فتكون من خلال سجلات المفردات أو الحالات المطلوب دراساتها خلال مدة معينة.

3-1 نماذج الاختيار المتقطع Discrete choice models:-

تطبق هذه النماذج عندما تكون المتغيرات المعتمدة (متغيرات الاستجابة) وصفية ومحددة (Qualitative and limited dependent variables) اي تلك التي يتم التعبير عنها بمجموعة محددة من الصفات أو الحقائق (characteristic) أو انها محددة (مقيدة) بقيود لا يمكن 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 المتعمدة 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 انموذج ثنائي الاستجابة يطبق عندما يأخذ المتغير المعتمد قيمتين فقط $[0,1]$ والتي قد تشير الى نمط الاستجابة لمؤثر ما (كالادوية و المنشطات) اي انه محدد بحالة النجاح و الفشل على التوالي.(Cox 1970 , البياتي 2005)

اذ يتبع احتمال حدوث الاستجابة توزيع برنولي (Bernolli 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) أحد 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 لا تعد دالة خطية بالنسبة للمتغيرات التوضيحية (كما في معادلة الانحدار الكمية) (Quantitative Regression equation) بل تكون على شكل دالة احتمالية خطية محددة بالفترة $[0,1]$ و للتعبير عن أنموذج التمييز لمثل هكذا متغيرات يكون بالصيغة الاتية:

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

y_i^* :يمثل المتغير الوهمي ليعبر عن حالة الاستجابة للمتغير المعتمد y_i (واحدعند حدوث الاستجابة وصفر في حالة عدم حدوث الاستجابة).

X_i' : يمثل موجه المتغيرات التوضيحية.

β : يمثل موجه المعلمات الخاص بالأنموذج.

ε_i : يمثل الخطأ العشوائي.

عندما يأخذ المتغير المعتمد القيم الاتية:

$$y_i = \begin{cases} 1 & y_i^* > o \\ 0 \quad \text{if} & O.W \end{cases}$$

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

$$P(y_i=1) = \frac{\exp(X_i'\beta)}{1+\exp(X_i'\beta)} \qquad ;i = 1,2,\ldots,n \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(1-3)$$

حيث X_i' ، β 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)

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

$$p(y_i=j) = \frac{\exp(X_{ij}'\beta)}{1+\exp(X_{ij}'\beta)} \qquad j = 1,2,\ldots\ldots,g-1 \quad \ldots\ldots\ldots\ldots\ldots\ldots(1-4)$$

X_{ij} :يمثل موجه المتغيرات التوضيحية الأستجابة رقم.(j)

β :يمثل موجه المعلمات الخاص بالأنموذج.

طرائق تقدير المعالم في النموذج:-

يتم الحصول على تقديرات معلمات النموذج بثلاثة طرائق هي:-

طريقة تقديرات الاحتمال الأعظم ML.

طريقة تقديرات المربعات الصغرى التعاقبية MLS.

طريقة تقديرات تصغير مربع كاي MCS.

1- 4 تقديرات الاحتمال الأعظم Maximum Likelihood Method:

ليكن لدينا Z من المتجهات المستقلة ذات توزيع متعدد الحدود Multinomial:

موجه ينتمي الى واحدة من g من مجاميع الاستجابة Response Categories نرمز للمشاهدات بـ ($Z_1,\ldots\ldots Z_n$)

حيث كل مشاهدة Z_i محدودة بواحدة من g من المجاميع ($Z_i=Z_{i1},\ldots\ldots Z_{ig}$) و أن $\sum_j X_{ij}=n_i$ يكون محدد لكل i تكون دالة الأمكان الاعظم الشرطية لتوزيع متعدد الحدود بالشكل الآتي:-

$$L=\prod_{i=1}^{N}\coprod_{j=1}^{g}\left[P(\mathrm{X}_i)\right]^{z_{ij}} \qquad \ldots\ldots\ldots\ldots(1-5)$$

حيث أن:-

$P_j(\mathrm{X}_i)$ = نسبة أو احتمال الانتماء إلى المجموعة j

بأخذ المشتقة الاولى للوغاريتم الاحتمال الاعظم و مساواتها بالصفر نحصل على المعادلات الطبيعية الاتية:-

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

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:-

$$B(t+1)=B(t)+\left\{\mathrm{X}'v(t)\mathrm{X}\right\}^{-1}\mathrm{X}'r(t) \qquad \ldots\ldots\ldots\ldots\ldots(1-6)$$

حيث أن:-

t (عدد الدورات التعاقبية)

r (هو موجه البواقي ذو بعد $[N(g-1),1]$

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

$$v_i=n_i\left\{P_s(\mathrm{X}_i)(\delta_{st}-P_i(\mathrm{X}_i))\right\}St$$

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

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

ويمكن أن نبدأ من قيم ابتدائية 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)

5-1 تقييم أداء الآنماذج: Evaluating the Performance of the models

لقد تم استخدام الاختبارات الآتيه:

(1-5-1) 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 الملائمة 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 (χ^2 approximate) وبتطبيق هذا الاختبار لتحديد مدى ملائمة أنموذج اللوجستي متعدد الحدود باستخراج قيمة الاحصاءة وحسب الصيغة الآتية:

$$\chi^2=\sum_{i=1}^{n}\sum_{j=1}^{g}\frac{\left(Y_{ij}-\hat{P}_j(X_{ij})\right)^2}{\hat{P}_j(X_{ij})(1-\hat{P}_j(X_{ij}))} \qquad .....(1-7)$$

حيث $P_j(X_{ij})$ معرفة كما في المعادلة (5-1).

وهناك العديد من الباحثين الذين ناقشوا امكانية ان يكون توزيع هذه الاحصاءة هو التوزيع الطبيعي بدلاً من تقريب مربع كاي. (Weesie 2002)

$$\frac{\chi^2-ng}{\sqrt{n}\sigma_n}\to N(0,1) \qquad ----------(1-8)$$

عندما

$$\sigma_n^2=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{g}\left\{\frac{1-2P_j(X_{ij})}{P_j(X_{ij})}q_j(X_{ij})-\sum_{h\neq j}^{g}q_j(X_{ij})q_h(X_{ih})\right\}-v'_n\Omega_n^{-1}v_n$$

$$v_n=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{g}q_j(X_{ij})(X_{ij}-\sum_{h=1}^{g}P_h(X_{ih})X_{ih})$$

$$q_j(X_{ij})=\frac{1-2P_j(X_{ij})}{1-P_j(X_{ij})}$$

و ان Ω_n تمثل مصفوفة المشتقات الثانية للوغاريتم دالة الامكان.

(2-5-1) احصاءة نسبة الامكان: Likelihood ratio statistic

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

$$H_0:\beta_1=\beta_2=.....=\beta_k=0 \qquad .....(1-9)$$

H_1: at Least two of them are not equal zero.

والصيغة العامة للاحصاءة 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)

(3-5-1)ً 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:

$$W=\left(\hat{\beta}_k-\hat{\beta}_{ko}\right)^2/H_{kk} \qquad .....(1-10)$$

k : تمثل عدد المتغيرات التوضيحية في الأنموذج.

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

H_{kk} : يمثل مربع الخطأ المعياري المقدر ASE.

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

(5-1-4) اختبار Q:

اقترح إختبار Q من قبل (Liu & Dyer) عام (1988) لأنموذج الانحدار اللوجستي و هي شبيهة بمعامل التحديد R^2 في أنموذج الانحدار الخطي و تحسب كالآتي:

بعد تقدير الأنموذج يتم ترتيب المشاهدات تنازلياً وفق الاحتمالات اللاحقة الأستجابة ($Y_i=1$) ويمكن التعبير عن Q كالآتي:

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

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 الى Q مساوياً لـ (واحد) عندما يكون n_1 من المشاهدات ذات الرتب الاعلى جميعها ذات الاستجابة $y=1$ ، ويكون الحد الادنى Q ذات توقع يساوي صفر عندما تتوزع المشاهدات بشكل عشوائي بين (1,0) وفقاً لرتبتها،ان استخدامها الرئيسي يكون في المساعدة لاختبار 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.(Hamad 2016)

6-1 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 أوعن طريق تقدير احتمال خطأ التصنيف وتوجد عدة طرق لتقديراحتمال 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:-

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

2- الجانب التطبيقي

2-1المقدمة

لقد تم جمع البيانات من (100) مريض في المستشفى ليلى قاسم/اربيل، وقد تم استخراج نتائج التحليلات بالاعتماد على البرنامج الإحصائي الجاهز SPSS، يمكن تقسيم هذا الفصل الى جزئين، الاول يتعلق بمعرفة وجود العلاقة من عدمها بين متغيرالجنس وأربعة من المتغيرات التي اعتمدها البحث وهي (نسبة السكر في الدم، نسبة الكوليسترول في الدم، نسبة اليوريا في الدم، نسبة التراكليسرايد) ،أما الثاني فيتم تطبيق الانحدار اللوجستي للبيانات التي تم الحصول عليها .

2-2 وصف البيانات

ويتم وصف المتغيرات البحث بالشكل الأتي:

X1: يمثل الجنس

X2: (s.CR) يمثل مستوى الكرياتين في الدم.

X3: (H.D.L) يمثل نسبة الكوليسترول الجيد.

X4: (HBA1C) يمثل مستوى الهيموكلوبين A1C(نسبة تراكمية).

X5: (b.su) يمثل نسبة السكر في الدم.

X6: (S.CHo) يمثل نسبة الكولسترول في الدم.

X7: (B.Ur) يمثل نسبة اليوريا في الدم.

X8: (T.G) يمثل نسبة التراكليسرايد في الدم.

Yi : حالة المصاب بمرض السكري.

2-3 تطبيق اختبار كاي سكويرللاستقلالية

تم تطبيق اختبار كاي سكوير لمعرفة وجود العلاقة بين متغيرالجنس وأربعة من المتغيرات التي اعتمدها البحث وهي (نسبة السكر في الدم، نسبة الكوليسترول في الدم، نسبة اليوريا في الدم، نسبة التراكليسرايد)، وبالاعتماد على الجدول الآتي:

الجدول(1-2)اختبار كاي سكوير لمعرفة وجود العلاقة بين متغيرالجنس والمتغيرات التي اعتمدها البحث

Variables	Chi-square value	d.f	P-value
نسبة السكر في الدم	4.028	1	0.045
نسبة الكولسترول في الدم	7.086	2	0.029
نسبة اليوريا في الدم	1.961	2	0.375
نسبة التراكليسرايد	3.099	2	0.212

نلاحظ من الجدول أعلاه بالاعتماد على قيمة P-value ومقارنتها مع قيمة المستوى المعنوي (0.05). ان قيمة P-value للمتغيرين (نسبة السكر في الدم، نسبة الكولسترول في الدم) اقل من 0.05 وهذا يدل على وجود علاقة بين المتغيرين المذكورة والجنس، وبما ان قيمة P-value للمتغيرين (نسبة اليوريا في الدم، نسبة التراكليسرايد) اكبر من 0.05 وهذا يدل على عدم وجود علاقة بين المتغيرين المذكورة والجنس.

2-4 تطبيق التمييز اللوجستي للبيانات

يقوم برنامج SPSS بتحليل بيانات العينة بخطوات متعددة بدءا" بالخطوة الصفرية التي يكون فيها النموذج خاليا من المتغيرات التوضحية. وكالتالي

الجدول(2-2) المتغيرات الداخلة في المعادلة

	B	S.E.	Wald	d.f	Sig.	Exp(B)
Step Constant 0	-0.40	0.200	0.040	1	0.841	0.961

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

الجدول(2-3) يوضح المتغيرات التي استبعدت من نموذج اللوجستيك في المرحلة الصفرية

الخطوة الصفر للمتغييرات	score	d.f	p- Value
مستوى الكرياتين في الدم.	4.781	1	0.029
نسبة كوليسترول الجيد.	9.118	1	0.003
مستوى هيموكلوبين A1C (النسبة التراكمية).	9.714	1	0.002
نسبة السكر في الدم.	4.028	1	0.045
نسبة الكولسترول في الدم.	6.827	1	0.009
نسبة اليوريا في الدم.	0.873	1	0.350
نسبة التراكليسرايد في الدم.	3.096	1	0.078
Overall Statistics	25.815	7	0.001

الجدول (2-4) عدد الدورات التكرارية

تكرار		المعاملات			
	-2 Log likelihood	الثوابت	مستوى هيموكلوبين	نسبة كوليسترول الجيد	مستوى الكرياتين في الدم
Step 1 1					
	128.586	2.531-	0.285		
2	128.545	2.704-	0.305		
3	128.545	2.706-	0.305		
4	128.545	2.706-	0.305		
Step 2 1					
	117.024	3.817-	0.299	0.021	
2	115.314	4.72-	0.335	0.032	
3	115.184	4.996-	0.342	0.037	
4	115.184	5.018-	0.343	0.037	
5	115.184	5.018-	0.343	0.037	
Step 3 1					
	111.904	2.622-	0.28	0.021	1.548-
2	109.381	3.033-	0.325	0.033	2.459-
3	109.212	3.223-	0.336	0.038	2.676-
4	109.211	3.244-	0.337	0.038	2.688-
5	109.211	3.244-	0.337	0.038	2.688-

الجدول(2-5)اختبار كاي سكوير للمراحل الثلاثة

		قيمة كاي سكوير	درجة الحرية	p-Value
Step 1	Model	10.044	1	0.002
Step 2	Model	23.406	2	0
Step 3	Model	29.378	3	0

الجدول (5-2) يوضح معنوية النموذج من خلال احصاءة كاي سكوير، وبما ان القيم P-Value اقل من 0.05 فإن النموذج معنوي في الخطوات الثلاثة اي المتغير قيد الاختبار يكون معنويا.

ولمعرفة قوة الأنموذج نستخدم R^2 لأنموذج اللوجستك الذي يفسر تأثير المتغير التوضيحي على التمتغير المعتمد.

الجدول(6-2) قيم R^2 لأنموذج اللوجستك

Step	-2 Log likelihood	Cox & Snell R^2	Nagelkerke R^2
1	128.545	0.096	0.127
2	115.184	0.209	0.278
3	109.211	0.255	0.339

نلاحظ في الجدول(6-2) و بالاعتماد على احصاءة Cox & Snell R^2 أن ازدياد قيمة هذه الاحصاءة بالخطوة الاخيرة بمقدار 0.159 عن الخطوة الاولى، وكذلك احصاءة Nagelkerke R^2 انها ازدادت بالخطوة الاخيرة بمقدار 0.212 عن الخطوة الاولى، واختبار النسبة الامكان الاعظم انخفضت قيمه بالخطوة الاخيرة بمقدار 19.334 عن الخطوة الاولى، فبهذا يكون الانموذج الذي له اقل قيمة هو الافضل، ولمعرفة ان هذا الفرق معنوي ام لا يستخدم اختبار كاي سكوير للنموذج اللوجستيك كالآتي:

الجدول(7-2)اختبار كاي سكوير للنموج اللوجستيك

Hosmer and Lemeshow Test

Step	Chi-square value	d.f	p-Value
3	8.279	8	0.407

الجدول(7-2) نلاحظ أن الأنموذج الذي حصلنا عليه في الخطوة الأخيرة هو الأنموذج الملائم للبيانات .

الجدول(8-2) القيم المشاهدةوالمتوقعة للنموذج اللوجستيك

Contingency Table for Hosmer and Lemeshow Test

	ذكر		انثى		Total
	قيمة المشاهدات	القمة المتوقعة	قيمة المشاهدات	القمة المتوقعة	
1	8	9.323	2	0.677	10
2	9	7.949	1	2.051	10
3	8	7.169	2	2.831	10
4	6	6.466	4	3.534	10
5	7	5.847	3	4.153	10
6	6	4.951	4	5.049	10
7	5	3.933	5	6.067	10
8	1	2.988	9	7.012	10
9	1	1.815	9	8.185	10
10	0	0.56	10	9.44	10

الجدول (9-2) التصنيف للمرحلة الأخيرة لأنموذج اللوجستيك

Classification Table[a]

المشاهدات	الجنس		percentage Correct
	ذكر	انثى	

الجنس	ذكر	42	9	82.4
	انثى	13	36	73.5
Overall Percentage				78

الجدول(10-2)المتغيرات الداخلة في النموذج

Variables	B	S.E	Wald	df	p-Value	Exp(B)
مستوى الكرياتين في الدم.	-2.951	1.333	4.9	1	0.027	0.052
نسبة الكوليسترول الجيد.	0.046	0.017	7.311	1	0.007	1.047
مستوى الهيموكلوبين (نسبة التراكمية).	0.376	0.141	7.141	1	0.008	1.456
نسبة السكر في الدم.	-0.337	0.316	1.132	1	0.287	0.714
نسبة الكولسترول في الدم.	0.116	0.294	0.155	1	0.694	1.123
نسبة اليوريا في الدم.	0.054	0.301	0.032	1	0.859	1.055
نسبة التراكليسرايد في الدم.	0.378	0.317	1.422	1	0.233	1.46
Constant	-3.832	1.667	5.282	1	0.022	0.022

الجدول(10-2) يبين المتغيرات التوضيحية الداخلة في النموذج ، اذ نلاحظ المتغيرات يعرض نتائج تقدير أنموذج التمييز اللوجستي وايظا نلاحظ:

ان قيم اختبار (wald) يمثل قيم اختبار المعلمات للنموذج ويظهر بأن المتغير X2 (مستوى الكرياتين في الدم) والمتغير X3 (نسبة الكوليسترول الجيد) والمتغير X4 (مستوى الهيموكلوبين) يمثل المتغيرات المعنوية في الدراسة وذللك من خلال مقارنة القيم P-value مع مستوى المعنوي (0.05) ، ان العمود P-value يمثل معنوية تأثير المتغيرات على حالة المصاب، (عندما تكون قيمة P-value اقل من 0.05 لمتغير قيد الاختبار يكون معنويا)، لذلك فأن المتغيرات الثلاثة (مستوى الكرياتين في الدم، نسبة الكوليسترول الجيد، مستوى الهيموكلوبين) تعد المؤثرة على المتغير المعتمد (الاستجابة) والذي يمثل حالة المصاب بمرض السكري وان بقية المتغيرات الاخرى ليس لهم التأثير، لذلك فان المتغيرات (مستوى الكرياتين في الدم، نسبة الكوليسترول الجيد، مستوى الهيموكلوبين) سوف تبقى في النموذج وبقية المتغيرات سوف تحذف من معادلة الأنحدار (الانموذج اللوجستي).

3- الاستنتاجات

أهم الاستنتاجات في البحث هو كالآتي:-

1- من خلال تطبيق اختباركاي سكويرللاستقلالية وبالاعتماد على قيمة P-value تبين بأنه توجد علاقة بين متغيرالجنس والمتغيرين (نسبة السكر في الدم ونسبة الكولسترول في الدم) ولا توجود علاقة بين متغيرالجنس والمتغيرين (نسبة اليوريا في الدم ونسبة التراكليسرايد).
2- ان نموذج التمييز اللوجستي ملائم في دراسة العوامل المؤثرة على مرض السكري وتبين ان المعلمات المقدرة لنموذج اللوجستي معنوية.
3- تبين بأن المتغيرات الثلاثة (مستوى الكرياتين في الدم، نسبة الكوليسترول الجيد، مستوى الهيموكلوبين) تعد مؤثرة في المتغير المعتمد (الاستجابة) والذي يمثل حالة المصاب بمرض السكري وان بقية المتغيرات ليس لهم التأثير.
4- نسبة التصنيف الصحيح هى (78%).

(5) المصادر

1- الهيتي، صلاح الدين حسين(2006):"***الأساليب الاحصائية في العلوم الادارية***" جامعة مؤتة، عمان-الأردن.

2- البياتي، هبة ابراهيم صالح (2005):"***تحليل المسار في نموذج الانحدار اللوجستي مع تطبيق***" رسالة ماجستيرفي علوم الاحصاء، جامعة المستنسرية، كلية الادارة والاقتصاد.

3- عباس، علي خظر(2012):"***استخدام نموذج الانحدار اللوجستي في التنبؤ بالدوال ذات المتغيرات الاقتصادية التابعة النوعية***" مجلة جامعة كركوك للعلوم الادارية والاقتصادية، المجلد (2) العدد(2).

4- مولود ، كوردستان ابراهيم (2000):"***استخدام التحليل المميز لتشخيص اهم العومل المؤثرة في التصنيف السريري لمرض القلب*** " رسالة ماجستير علوم في الاحصاء، جامعة صلاح الدين/ اربيل، كلية الادارة والاقتصاد.

5- Agresti, A(1990):"***Categorical Data Analysis***" John Wiley Sons. Inc. New York.

6- Cox, D.R.(1970):"***The Analysis of Binary Data***" Chapman and Hall, London.

7- 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.

8- Jarajda, Stepan and Box.(2003):"***Econometrics of Panel Data and Limited Dependent Model***" Lecture Notes Series of Academy of Sciences of Th Czech Republic Economics Institut. http://www Cerge-ei-cs.

9- Menard, Scott. (2002):"***Applied Logistic Regression Analysis***" Second Edition, Sage Publication, Inc.

10- Weesie. (2002)"***On Pearson*** χ^2 ***for Categorical Response Variable***" German Stat User's Meeting.

پوختە

لەم توێژینەوەدا شیکردنەوەی جیاکردنەوەی لۆجستیك بەکارهێنراوە هەروەها رێگای تواني بەرز بەکارهێنراوە بۆ خەملاندنی پارامیتەری مۆدێلی لۆجیکی، بۆ ئەم مەبەستەش داتای پێویست وەرگیراوە لە کۆمەڵێك نەخۆش کە نەخۆشی شەکرەیان هەیە لە تۆماری مێژویی لە نەخۆشخانەی لیلی قاسم لە هەولێر، ئامانج لەم توێژینەوەدا بریتیە لە لێکۆلینەوەی گرینگترین فاکتەرەکانی کە دەبێتە هۆی جیاکردنەوە لە نێوان هەردوو رەگەز بۆ ئەوانەی کە توشی نەخۆشی شەکرە هاتوون ئەمەش بە هۆی بەکارهێنانی مۆدێلی جیاکەرەوەی لۆجیکی و کاریگەری رەگەز لە سەر زیادەی رێژەی توش بوون بە نەخۆشی شەکرە و زانینی بوونی یان نەبوونی پەیوەندی لە نێوان گۆراوی رەگەز لەگەڵ هەر چوار گوراوی (رێژەی شەکر لە خوێن، رێژەی کولیسترۆل لە خوێن، رێژەی یۆریا لە خوێن، رێژەی ترکلیسراید).

گرینگترین ئەنجام بۆ توێژینەوەکە ئەوەیە مۆدێلی جیاکردنەوەی لۆجیکی گونجاوە لە لێکۆلینەوەی فاکتەرە کاریگەرەکانی سەر نەخۆشی شەکرە وە بە هۆی تاقیکردنەوەی والد دەردەکەوێت خەملێنەرەکانی مۆدێلی لوجیکی معنەوین. وەدەرکەوت کە هەرسێ گۆراوی (ئاستی کریاتین لە خوێن، رێژەی کۆلیسترۆل باش، ئاستی هیموگلۆبین) بە گاریگەر هەرژماردەکرێ بۆ گۆراوی پشت پێ بەستراو کە ئەویش توش بوانی نەخۆشی شەکرەن وە گۆراوەکانی تر کاریگەریان نی یە. وە هەروەها دەرکەوت کە پەیوەندی هەیە لە نێوان رەگەز و هەردوو گۆراوی ریژەی شەکر لە خوێن و رێژەی کولسترول لە خوێن بەڵام پەیوەندی نی یە لە نێوان گۆراوی رەگەز و هەردوو گۆراوی رێژەیی یوریا لە خوێن و رێژەی تراکلیسراید. وە رێژەی پۆلین کردنی راست بریتیە لە (78%)

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.

Data were collected of a group of diabetic patients through historical records at Laila Qassim / Erbil Hospital. The main objective of this research is studying the most important factors that distinguish between the sexes of people with diabetes through the use of the model of the distinction of logistics and the extent of the impact of sex in increasing the incidence of diabetes and knowledge of the existence of the relationship or lack of relationship between sex variable and four variables (blood sugar , Cholesterol ratio, urea ratio in the blood, ratio of triglyceride).

The most important conclusion of the research is that the model of logistic discrimination is appropriate in the study of the factors affecting diabetes and found that the estimated parameters of the logistic model are significant through the test (Wald). The three variables (serum creatinine, HDL, hemoglobin level) were significant in the adopted variable, which represents the condition of the diabetic patient and the remaining variables have no effect. It also shows that there is a relationship between sex and the variables of blood sugar and cholesterol in the blood, but the lack of relationship between sex and variables ratio of urea in the blood and the proportion of triglyceride. The correct rating was 78%.