بناء نموذج انحدار لوجستي للاصابة بمرض السكري من النوعين الاول والثاني

(تطبيق في مستشفى البصره العام في عام ٢٠١٦)

م. ناديه علي عايد

جامعة البصرة / كلية الادارة والاقتصاد /قسم الاحصاء

Building a logistic regression model for type I and II diabetes
(Application in the General Hospital in 2016)

Nadia Ali Ayed

AL -Basrah University ,Administration &Economics College , Department of

بناء نموذج انحدار لوجستي للاصابة بمرض السكري من النوعين الاول والثاني

م. ناديه علي عايد

الملخص :

يعد مرض السكري من اكثر الامراض شيوعا ومن أكثرها خطر إذ يترتب على هذا المرض الكثير من المضاعفات التي تؤدي الى الوفاة. يهدف البحث الى بناء نموذج انحدار لوجستي للإصابة بمرض السكري من النوعين الاول والثاني، وقد اعتمد البحث على بيانات اوليه مأخوذة من عينه بحجم (٢١٦)مريضاً بالسكري من كلا الجنسين وبفئات عمريه مختلفة من مستشفى البصره العام .أستخدمت الباحثة المنهجين الوصفي والتحليلي إذ تم بناء نموذج الانحدار اللوجستي الثنائي ثم تحليل البيانات باستخدام برنامج spss ومن اهم النتائج التي اظهرتها الدراسة أن نموذج الانحدار اللوجستي الثنائي المستخدم له قدره تفسيريه وتصنيفيه عاليه وبشكل دال احصائيا . نجحت المتغيرات المستقلة الداخلة في النموذج في تفسير ما نسبته (٤٥%) الى(٧٧%) من التغيرات في المتغير التابع وهذا ما اكدته قيم R^2 ولكن بحسب هذه النتيجة توجد عوامل اخرى لم تدرج في النموذج ،قد يكون لها تأثير وهذا العوامل قد تكون نفسيه او وراثية وان نموذج متغيرات الدراسة اظهر تأثيراً معنوياً لكل من (العمر، الوزن)على تصنيف المريض سواء كان مصاباً بالسكري من النوع الاول و الثاني اذ بلغت نسبة التصنيف الاجمالي الصحيح في النموذج(٩٤%) و التصنيف الصحيح بمرض السكري من النوع الاول(٧٥%) اما المصابين بمرض السكري من النوع الثاني فأن نسبة التصنيف الصحيح لهم(٩٨%).

Abstract

Diabetes is considered is one of the most common and most dangerous disease, as the disease has complication that lead to death. The research aims to build a logistic regression model for diabetes type I and II .The research was based on preliminary data taken from the same size of 216 patients with diabetes of both sexes and different age groups from Al Basra General Hospital. The researcher used the descriptive and analytical methods as the logistic regression model was constructed and the results were analyzed using the SPSS program, and the most important result were that the logistic regression model is highly explanatory and descriptive. The independent variables in the model succeeded in interpreting 45% to 77% of the changes in the dependent variable. This was confirmed by the values of R^2,according to

This result , other factors not included in the model have an effect .These factors may be psychological or hereditary .The study model showed a significant effect on age and weight on patient classification .The correct overall rating of model was 94%. And the correct classification of diabetes type I was 75%. People with type II diabetes have the right classification rate of 98%.

المقدمـــة

استخدام الباحثون خلال فترة طويلة تحليل الانحدار لدراسة اثر المتغيرات المستقلة المفسرة قي المتغير التابع من خلا ل بناء معادلة لغرض التفسير او التقدير او التحكم بقيمه المتغير التابع بدلالة متغير او اكثر من هذه المتغيرات المستقلة.

اذ ان المتغيرات التابعة في الدراسات الطبية والاجتماعية والإنسانية عادة ما تكون ذات طبيعة تصنيفية وليست كمية مما لا تتحقق معه فرضيات ومتطلبات نماذج الانحدار .لذا برزت الحاجة لاستكشاف اساليب احصائية بديلة تتيح التعامل مع مثل هذه المتغيرات.

اهداف الدراسة:-

بناء نموذج رياضي باستخدام الانحدار اللوجستي الثنائي للتعرف على اهم المتغيرات المؤثرة في الاصابه بمرض السكري من النوع الاول والنوع الثاني.

المنهج وعينه ألدراسة

١- منهج الدراسة

اعتمدت الدراسة على المنهج الوصفى التحليلي،اذ اعتمدت على الجانب الوصفي في استخدام الجداول التكراريه اما الجانب التحليلي فتم استخدامه في بناء نموذج الانحدار اللوجستي الثنائي وفحص جودة توفيقه للبيانات.

٢- عينه الدراسة

سحبت عينه بواقع ٢١٦ مريضاً مصابين بمرض السكري النوعين الاول والثاني ولكلأ الجنسين وبفئات عمريه مختلفة من مستشفى البصره لعام ٢٠١٦.

فرضيات الدراسة:-

قمنا باختيار فرضيات العدم (H_0) للمتغيرات المستقلة جميعها على وفق ما ياتي

$H_0=b_i=0 \qquad i=1,2,.....216$

وتشير (b_i)الى معاملات الانحدار للمتغيرات المستقلة . .

المفاهيم الاساسيه:-

تعريف مرض السكري:

مرض السكري من الأمراض الشائعة على مستوى العالم, و يقدر المصابون به ب ١٢٠ مليون شخص و يتوقع أن يصل العدد إلى ٢٢٠ مليون شخص مصاب بحلول سنة ٢٠٢٠.

و مرض السكري هو مرض إستقلابي (أيضي) مزمن Metabolic Disease يتميز بزيادة مستوى السكر في الدم Hyperglycemia نتيجة لنقص نسبي أو كامل في الأنسولين Insulin في الدم أو لخلل في تأثير الأنسولين بالأنسجة , مما ينتج عنه مضاعفات مزمنة في أعضاء مختلفة من الجسم.ثلث الأشخاص المصابين بمرض السكري لا يُشخصون.

الأنسولين هو هرمون يُفرز من خلايا بيتا في جزر لانجرهانس Islets of Langherhans Beta Cells في البنكرياس Pancreas و يتكون من سلسلتين من الأحماض الأمينية مرتبطتين بروابط كيميائية بعد أن تنفصل

منه سلسلة سي C Peptide حتى يصبح فعالاً و يمر في الكبد إذ يُدمر ٥٠% من الأنسولين المفروز.و الأنسولين ضروري للجسم كي يتمكن من الاستفادة و استخدام السكر و الطاقة في الطعام.

انواع مرض السكري :-

هنالك تقسيمات مختلفة للمرض فبعض الأطباء يتحدثون عن تقسيمات وتسميات مختلفة منهم من كان يقسم المرض الى نوع أو نمط يصيب صغار السن .ونمط يصيب كبار السن أو الكهول.

النوع أو النمط الأول، أو النوع الذي يصيب الأطفال واليافعين ويشخص قبل سن ٣٠ عاما، وهذا غالبا يعتمد على الإنسولين Insulin Dependent Diabetes Mellitus أو اختصار IDDM ، وسببه في الغالب وراثي وهو مسؤول عن ١٠-١٥% من جميع حالات داء السكري، ويعد أشد خطورة ويحتاج إلى عناية خاصة. يتصف هذا النوع بضآلة إفراز الأنسولين الداخلى أو غيابه، لذلك يتعرض هؤلاء المرضى لحدوث المضاعفات الحادة للسكري مثل كثرة التبول والعطش والنهم وحموضة الدم الكيتونية ketoacidosis أو حموضة الدم الديابيتي diabetic acidosisوهي حالة غير سوية تقل فيها قلوية الدم أو تكثر حموضته بسبب إحدى النتائج الثانوية لعملية أيض المواد الدهنية التي تولد مواد أسيتونية acetone substances أو أجسام الكيتونات ketone bodiesوتؤدي بالتالي إلى حالة غيبوبة سكرية diabetic coma قد تكون مميتة. ولمنع حدوث حموضة الدم الكيتونية والموت يلزم الاستعانه بالأنسولين الخارجى. قبل اكتشاف العلاج المناسب بواسطة الأنسولين والحقن الوريدية الأخرى كانت حالة حموضة الدم Acidosis يعد السبب الرئيس لحالات الوفاة بين المصابين بداء البول السكري.من أسباب ذلك قد يكون إصابة الشخص في الصغر بأمراض تؤثر في البنكرياس كالأمراض الفيروسية والنكاف أو وجود حالات من الحساسية ضد خلايا البنكرياس بحيث يعاملها الجسم على أنها أجسام غريبة فيحطمها وبعض الحالات وجد فيها نقص أو خطأ في بعض الجينات أو المورثات بحيث لا تستطيع خلايا البنكرياس إنتاج الأنسولين. هذه الحالات يجب أن تعالج بالحمية الغذائية ويجب إعطاء المريض حقن إنسولين. إن ذروة بدء النمط الأول للسكري بين عمر ١١ و ١٣ سنة وتتواقت مع بدء البلوغ، لكنه قد يبدأ في أي عمر بما فيها الشيخوخة.

2- النوع الآخر والمعروف بالنوع أو النمط رقم اثنين، او الذي نراه يصيب كبار السن أكبر من ٣٠ عاما، وغالبا لا يعتمد على الأنسولين Non-Insulin Dependent Diabetes Mellitus أو اختصار .NIDDM والسبب في هذا النوع هو نقص في إفراز الإنسولين إستجابة للكلكوز أو في فعاليته (مقاومة الإنسولين insulin resistance)بسبب خلل في مستقبلات الإنسولين أو نقص عددها أو بسبب خلل في تمثيل السكر داخل الخلايا يمنع حرق السكر بداخلها بشكل طبيعي. وعادة يتم علاجه بالحمية الغذائية. أحيانا يحتاج المريض بعض العقاقير التي تعطى بالفم للسيطرة على ارتفاع مستوى السكر في الدم. لهذا سمي هذا النوع بالنوع الذي لا يعتمد على الأنسولين.

يتصف هذا النوع بغياب الأعراض الحادة لداء السكري إلا في حالات الارتفاع الشديد في نسبة السكر بالدم. لذلك يجب على من هو معرض للإصابة بهذا النوع مداومة الكشف الدوري على نسبة السكر بالدم سنويا .

الانموذج ونتائج التحليل الاحصائي

١- نموذج الانحدار اللوجستي

يرى الباحثون ان الغرض الرئيس لأغلب البحوث هو الوصول إلى صيغة توضح العلاقة بين مجموعة من المتغيرات وان الانحدار يهدف الى تحديد ووصف تلك العلاقة من خلال صياغة معادلة رياضية تربط جميع تلك المتغيرات كما ان طرائق تحليل الانحدار تعد الجزء الاساسي في تحليل البيانات وذلك عن طريق وصف العلاقة بين مجموعه من المتغيرات المستقلة والمتغير التابع.

وقد جاء تحليل الانحدار لتحقيق ثلاثة اغراض اساسية هي وصف شكل العلاقة بين المتغيرات والتقدير والتنبؤ بقيمة المتغير التابع عند أي مستوى للمتغيرات المفسرة وأخيرا التحكم بقيم المتغير التابع تبعا للتغير في قيم المتغير المفسر عند ثبات بقية المتغيرات المفسرة.

وبالرغم من ان تحليل الانحدار يحقق اغلب اهداف البحث العلمي وان أساليبه تعد الجزء الاساسي لأي تحليل للبيانات يهدف الى دراسة وتفسير العلاقة بين المتغير التابع والمتغيرات المستقلة إلا انه يعجز عن وصف وتفسير العلاقات بين المتغيرات المفسرة والمتغير التابع اذا كان الاخير ثنائي القيمة حيث يشترط ان تكون طبيعة المتغير التابع كمية متصلة وليست تصنيفية مع الاخذ بعين الاعتبار شيوع هذا النوع من المتغيرات التابعة في حالات دراسة الحالات الطبية والإنسانية والاجتماعية على نحو كبير ولهذا ظهرت الحاجة الى تطوير اساليب احصائية جديدة لها قوة الانحدار الخطي في التوصيل لأفضل المعادلات توفيقا وتعالج بالوقت ذاته مشكلة تعذر تطبيق نماذج تحليل الانحدار الخطي الاعتيادي في حالة المتغيرات التابعة ثنائية القيمة.

ولزمن طويل كان الباحثون يلجؤ ون الى استخدام الانحدار الخطي حتى في حالات المتغير التابع الثنائي نظرا للفراغ الاحصائي انذاك ويدعم هذا القول استخدام الباحثين للانحدار الخطي المتعدد مع المتغيرات الثنائيه قبل ظهور الانحدار اللوجستي إذ يعد تحليل الانحدار الخطي أقل ملاءمة للاستخدام من الانحدار اللوجستي في حالات التنبؤ والتصنيف للمتغيرات التابعة ثنائية القيمة.

وإذا ما اتجهنا الى المقارنة بين كل من الانحدار اللوجستي والانحدار الخطي فانه ومن الوجهة التطبيقية فان كلا الاسلوبين مماثل إلا انهما يختلفان من الوجهة الاحصائية حيث ان الصياغة الرياضية والحسابات التفصيلية لكليهما مختلفة تماما اذ يرى الباحثون ان طريقة تقديم واختيار ومن ثم فحص النموذج تختلف تماما في كل من التحليلين فضلا عن الاختلاف في افتراضات التحليل وتقدير المعاملات وتفسير النتائج مما يجعل التحليلين مختلفين تماما.

وبناء على ماسبق فان العديد من الباحثين يرون عدم ملاءمة الانحدار الخطي في حالات دراسة المتغير التابع ثنائي القيمه وبالتالي فان الانحدار اللوجستي يعد الاداة الاكثر ملاءمة في مثل تلك الحالات لمايتمتع به من مرونه تفوق بها على الانحدار الخطي في كثير من الاحيان.

يستخدم الانحدار اللوجستي عندما يكون المتغير التابع (Y) متغيرا ثنائيا يأخذ قيمتين فقط يرمز للاولى وهي وقوع الحدث بالرمز (١) وذلك باحتمال قدره (P) في حين يرمز للثانية وهي عدم وقوع الحدث بالرمز(0) وذلك باحتمال يساوي (1−P) فيما لا يضع قيودا على انواع المتغيرات المستقلة x_i والتي يمكن لها ان تكون متصلة أو فئوية أوخليطاً من الاثنين كما انة لا يشترط اعتدالية توزيعها .

وحيث ان معادلة الانحدار الخطي البسيط تكون على الصورة:

$$Y|X = \beta_0 + \beta_1 X + e \qquad (1)$$

إذ يعني الرمز Y/X: المتغير التابع Y بشرط حدوث المتغير المستقل X.

وبافتراض ان الخطأ العشوائي (e) يتبع التوزيع الطبيعي بمتوسط ٠ وانحراف معياري مقداره $\sigma_{Y/X}$ اي أن $e \sim N(0, \sigma_{Y/X})$ فان المتغير التابع Yيتبع التوزيع الطبيعي بمتوسط $\mu_{Y/X}$ وانحراف معياري $\sigma_{Y/X}$ أي أن $Y \sim N(\mu_{Y/X}, \sigma_{Y/X})$ وذلك لكل قيمة من قيم المتغير المستقل X .

ونظرا لان E(e) =٠ لذا فإن القيمة المتوقعة للمتغير التابع Y عند قيمه معينة للمتغيرالمستقل Xتكون على الشكل التالي:

$$E(Y/X) = \hat{\beta}_0 + \hat{\beta}_1 X \qquad (2)$$

إلا إنه لعدم امكانية تطبيق الانحدار الخطي البسيط في حالة المتغير التابع ثنائي القيمة (Y) نتيجة لما يلي:

– ان تباين المتغير التابع (Y) يتغير بتغير قيم المتغير المستقل (X).

–ان تباين الخطأ لا يتوزع على وفق التوزيع الطبيعي.

–ان القيم المقدرة لا يمكن تفسيرها بوصفها احتمالات ذلك لأن قيمها لا تتراوح بين (١,٠).لذا يتم استخدام النموذج اللوجستي الذي يعالج المشكلات السابقة إذ يمكن كتابته في حالة وجود متغير مستقل واحد كالتالي:

$$Log_e\left(\frac{P}{1-p}\right) = \hat{\beta}_0 + \hat{\beta}_1 X \qquad (3)$$

وبصورة أخرى:

$$\therefore \left(\frac{P}{1-p}\right) = e^{\hat{\beta}_0 + \hat{\beta}_1 X} \quad \rightarrow \quad P = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 X}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 X}}$$

حيث :

P: هواحتمال وقوع الحدث محل الاهتمام اي احتمال النجاح.

1–P: هو احتمال وقوع الحدث ليس محل الاهتمام اي احتمال الفشل.

$\left(\frac{P}{1-P}\right)$: نسبة الترجيح للحدث محل الاهتمام (odds).

$\log_e$: هو اللوغاريتم الطبيعي e=٢.٧١٨٢٨١٨٢٨٤

$\log_e\left(\frac{P}{1-P}\right)$: اللوغاريتم الطبيعي لنسبة الترجيح او اللوجيت.

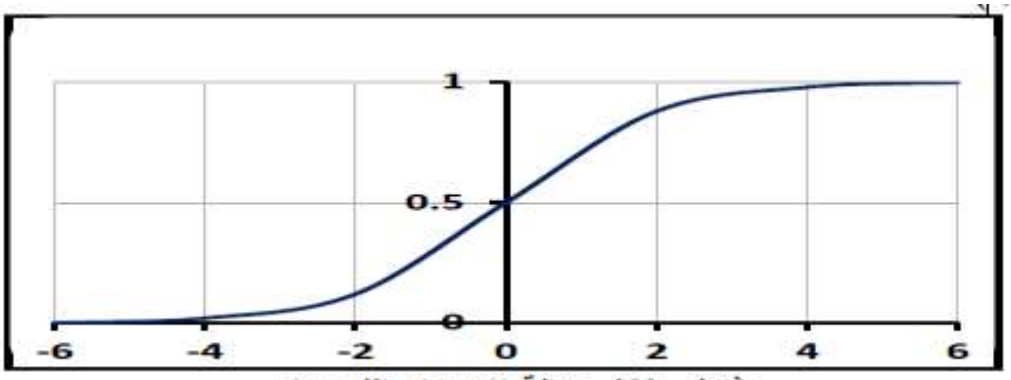
وبذلك يمكن كتابة معادلة الانحدار في حالة وجود عدد k من المتغيرات المستقله على الصورة:

$$Log_e\left(\frac{P}{1-p}\right) = \hat{\beta}_0 + \sum_{i=1}^{k} \hat{\beta}_i X_i = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \cdots + \hat{\beta}_k X_k$$

$$\therefore Logit = Log_e(odds) = Log_e\left(\frac{P}{1-p}\right) = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \cdots + \hat{\beta}_k X_k \qquad (4)$$

وتسمى المعادلة (٤) نموذج الانحدار اللوجستي.

هذا وسيتم اختصارا استخدام الرمز log من الان فصاعدا وذلك للتعبير عن اللوغاريتم الطبيعي $\log_e$ والدالة اللوجستية كما يتضح في الشكل (١) هي دالة متصلة يتراوح مداها بين (١,٠) إذ تقترب من الصفر كلما اقترب الطرف الأيمن للدالة من (∞ –) كما تقترب من الواحد كلما اقترب هذا الطرف من (+∞)



شكل (1) : دالة الانحدار اللوجستي

ويتم تقدير معالم نموذج لوجيستي باستخدام طريقة الارجحية العظمى (Maximum Likelihood) وهي من طرائق التقدير الاحصائية المعروفة إذ تقيس دالة الارجحية العظمى للاحتمالات المشاهدة لعدد K من المتغيرات المستقلة التي تتم دراستها ولتكن P_1 , P_2,.....P_k

وتكون هذة الدالة هي حاصل ضرب هذة الاحتمالات .

تحليل البيانات ومناقشة النتائج

التوصيف الاحصائي لمتغيرات الدراسة:-

تم الحصول على البيانات المستخدمه في هذا البحث من مستشفى البصره العام أذا تم تجميع عينه عشوائيه حجمها (٢١٦) مفردة من المرضى المصابين بالمرض السكري النوع الاول والنوع الثاني لسنه (٢٠١٦) .

وقد حددت المتغيرات المستقله بالأتي:-

١- الجنس (x_1) وصنف بالشكل التالي : 0 يمثل ذكر 1 يمثل انثى

٢- العمر (x_2) وصنف حسب الفئات بالشكل التالي:-

0 :(١٠- ١٩)

١. (٢٠ - ٢٩)

٢. (٣٠- ٣٩)

٣. (٤٠ فأكثر)

٣- الطول (x_3)

٤- الوزن (x_4)

اما المتغير المعتمد او متغير الاستجابة ويمثل نوع الاصابة بالمرض السكري وأعطيت الفيمة (0) للنوع الاول والقيمة (١) للنوع الثاني.

التحليل الاحصائي:- من خلال تحليل البيانات التي حصلنا عليها تم الحصول على النتائج الاتيه:-

جدول(١)

أعداد ونسب المصابين بالسكري من النوع الاول والثاني حسب الجنس.

الجنس	عدد المصابين بالنوع الاول	عدد المصابين بالنوع الثاني	المجموع	نسبه الاصابة بالنوع الاول	نسبه الاصابة بالنوع الثاني
ذكر	14	70	84	16.6	83.3
انثى	26	106	132	19.69	80.30
المجموع	40	176	216	18.5	81.48

المصدر :اعداد من الباحثه باستخدام بيانات الدراسة.

ومن الجدول (١) نلاحظ نسبة الاصابة بالسكري من النوع الثاني اكثر من النوع الاول للذكور والإناث في عينه الدراسة ،اذ بلغت نسبة الاصابة بالنوع الاول والثاني (18.5)،(81.48) وعلى التوالي.

جدول (2)

أعداد المصابين بمرض السكري من النوع الاول والنوع الثاني حسب الفئات العمرية

الفئات العمريه	عدد المصابين بالنوع الاول	عدد المصابين بالنوع الثاني	المجموع	نسبه المصابين بالنوع الاول	انسبه المصابين بالنوع الثاني
(10-19)	14	3	17	%82.35	%17.65
(20-29)	9	2	11	%81.81	%18.18
(30-39)	9	39	48	%18.75	%81.25
40 فاكثر	0	140	140	0	%100

المصدر :من اعداد الباحثة باستخدام بيانات الدراسة

نلاحظ من جدول (2) ان نسبه المصابين في عينه الدراسة بالنوع الاول في الفئات العمريه (19−10)،(29−20) أكثر من غيرها أذ بلغت 82.35%،81.81%. وان نسبه المصابين بالنوع الثاني (39−30)، (40 – فأكثر) اكثر من الفئتين الاولى والثانية اذ بلغت81.25%،100%وعلى التوالي .

جدول (٣)

يوضح عدد الدورات التكراريه لمشتقات دالة الامكان الاعظم

Iteration History[a,b,c,d]

Iteration		-2 Log likelihood	Coefficients				
			Constant	x1	x2	x3	x4
Step 1	1	104.803	2.197	-.347-	1.260	.007	-.026-
	2	73.965	7.041	-.772-	1.985	.013	-.064-
	3	64.546	12.294	-1.157-	2.651	.017	-.103-
	4	62.447	15.801	-1.383-	3.169	.017	-.129-
	5	62.251	17.078	-1.459-	3.396	.016	-.139-
	6	62.249	17.230	-1.468-	3.427	.016	-.140-
	7	62.249	17.233	-1.469-	3.427	.016	-.140-
	8	62.249	17.233	-1.469-	3.427	.016	-.140-

a. Method: Enter
b. Constant is included in the model.
c. Initial -2 Log Likelihood: 194.642
d. Estimation terminated at iteration number 8 because parameter estimates changed by less than .001.

المصدر :من اعداد الباحثه باستخدام بيانات الدراسه وبرنامج SPSS

من جدول(٣) نلاحظ أننا حصلنا في الدوره الثامنه لمشتقه سالب ضعف داله الامكان الاعظم على اقل قيمه لها وهي تساوي (62.249) اي (: 62.249 =2 Log Likelihood) وتوقفنا عند هذه الدوره لان التغير في معاملات (X_1, X_2 ,X_3,X_4) أصبح اقل من 0.001 . وفي حقيقه الامر فان التغير في معالم ا لمقدره اصبح بطيئا جدا بعد الدوره الخامسه لذلك يمكن القول ان مقدرات المعالم في الدورا ت (6 ،7، 8)هي متشابه مع فروقا ت بسيطه جدا وتوقفنا عند الدوره الثامنه واعتبرنا معالمها افضل نتيجه يمكن الحصول عليها للمعالم اذ ان سالب ضعف لوغارتيم دالة الامكان الاعظم هي في نهايتها الصغرى عند هذه الدوره و يلخص الجدول (4) معالم النموذج الامثل التي حصلنا عليها في الدوره الثامنه في الجدول(٣)

جدول(٤)

يوضح معالم النموذج المقدر

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1	x1	-1.469-	.816	3.237	1	.072	.230	.046	1.140
	x2	3.427	.639	28.744	1	.000	30.787	8.796	107.761
	x3	.016	.018	.786	1	.375	1.016	.980	1.054
	x4	.140	.042	10.981	1	.001	.869	.800	.944
	Constant	17.233	6.581	6.857	1	.009	30478625.225		

a. Variable(s) entered on step 1: x1, x2, x3, x4.

SPSS المصدر :من اعداد الباحثه باستخدام بيانات الدراسه وبرنامج

الجدول(٤) ويتضمن جميع معالم النموذج المقدرة والخطأ المعياري لكل معلمة . وإحصاءه(wald) لكل معلمة من معالم النموذج وعدد درجات الحرية ومعنوية المعالم التي سنقوم بتفسيرها بشكل مفصل لاحقا .

التحقق من ملاءمة النموذج ككل:

تم التحقق من ملاءمه النموذج بشكل كلي باستخدام عدة مقاييس مهمه تساعد في تقويم النموذج النهائي الذي ،هي اختبار x^2 وأحصاءات R^2 واختبار هوزمو– ليمشو لجوده المطابقه وجداول التصنيف.

وجد ان مربع كاي لدلالة الفرق بين قيمتي لوغاريتم دالة الترجيح لنموذج الانحدار اللوجستي بالمتغيرات المستقلة وبدون المتغيرات المستقلة موضع الفحص على النحو التالي:

جدول(٥)

يوضح أختبار معنوية النموذج المقدر

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	132.394	4	.000
	Block	132.394	4	.000
	Model	132.394	4	.000

SPSS المصدر :من اعداد الباحثه باستخدام بيانات الدراسه وبرنامج

جدول(٥) يبين قيمهx^2 تساوي (132.394) وهي معنوية عند مستوى دلاله اقل من (0.001) حيث Sig = .000 وبذلك نرفض العدم ونقبل الفرض البديل اي ان النموذج معنوي ويمثل البيانات بشكل جيد.

جدول(٦)

يوضح قيمه التكرار الثامن لدالة الارجحيه واختبار R^2

Model Summary

Step	-2 Log likelihood	R^2 Cox & Snell	R^2 Nagelkerke
1	62.249[a]	.458	.772

a. Estimation terminated at iteration number 8 because parameter estimates changed by less than .001.

المصدر: من اعداد الباحثه باستخدام بيانات الدراسه وبرنامج spss

الجدول (٦) يوضح القيمه الثامنه لمشتقه دالة الامكان الاعظم التي حصلنا عليها من جدول(٣) وهي تقريبا تساوي(62.249) وقيمه المقياس

(R^2 Cox & Snell = .458) و قيمه المقياس المعدل(R^2 Nagelkerke=.772) حيث هذه المقاييس تهدف الى تحديد نسبه التباين المفسر

في نموذج الانحدار اللوجستي اي توضح لنا ماتفسره المتغيرات التوضيحيه من المتغير المعتمد.وهذا يدل على انه ماتزال هناك نسبة من التغيرات في المتغير التابع تعود لمتغيرات اخرى غير مدرجه في النموذج

وباستخدام اختبار (Hosmer and lemeshow) لجودة مطابقة النموذج وذلك لاختبار فرضية العدم إن النموذج الذي تم توفيقه ملائم للبيانات مقابل الفرضية البديلة ، كانت النتائج كما في جدول(٧)

جدول (٧)

يوضح اختبارHosmer and Lemeshow TestHosmer and lemeshow

Step	Chi-square	df	Sig.
1	١٠.١٧٠	8	١٩٠.

المصدر:من اعداد الباحثه باستخدام بيانات الدراسه وبرنامجspss

ونلاحظ من الجدول(٧) التالي ان قيمه H-statistic = ١٠.١٧٠ ودرجة الحرية تساوي ٨ ومستوى معنويه (190.)اكبر من (٠.٠٥) في الخطوة الاولى وهذا يعني انه لايوجد دليل كاف لرفض فرضية العدم وبالتالي فان النموذج النهائي يعد مناسبا للبيانات.

جدول (٩)

يبين النسبه المئويه للتصنيف الصحيح[a] Classification Table

		Predicted		
	Observed	y		Percentage Correct
		.00	1.00	
Step 1	y .00	27	9	75.0
	1.00	3	177	98.3
	Overall Percentage			94.0

a. The cut value is .500

المصدر :من اعداد الباحثه باستخدام بيانات الدراسه وبرنامج spss

الجدول (٩) نجد ان نسبه التصنيف الصحيح للعوامل الموثرة في الاصابة بالسكري من النوع الاول بلغت 75.0 % اما المصابين بالسكر من النوع الثاني نجد نسبه تصنيفهم الصحيح للعوامل بلغت 98.3 %. وان نسبه التصنيف الصحيح في النموذج الكلي للعوامل الموثرة في الاصابة بمرض السكري من النوع الاول والنوع الثاني(94.0 %= ٢١٦ /(٢٧ +١٧٧)) ونسبه التصنيف الخطأ ٦ % وتعد هذه النسبه جيده جد

تفسير معالم النموذج :-

بالنظر الى جدول(٤) الذي يوضح معالم النموذج الامثل التي حصلنا عليها بالدورة الثامنة اذ نلاحظ العمود B يحتوي على معاملات النموذج المقدر وهي بوحدات Log- odd فضلا عن الى الخطأ المعياري لكل معلمه وإحصائية (Wald) لكل معلمه مع عدد درجات الحرية ودلالتها الاحصائية .

واما فرضيه اختبار العدم (H0:bi=0) ضد الفرضية البديلة (H1:bi=0) وهي اختبار من طرفين واما بالنسبة sig فهو عمود يمثل معنوية المعاملات المناظرة للقبول او للرفض باتخاذ مستوى المعنوية (0.05) و(0.05>sig) يتم رفض فرضية العدم اي ان المعامل معنوي وهو لا يساوي صفراً في المجتمع الذي سحبت منه العينة . وأما العمود Exp(B) odds Ratioفيوضح قيمة الدالة الاسية لمعامل الانحدار وهو يعبر عن المضاعف الذي تتغير به نسبة الترجيح (احتمال وقوع الحدث p(y) الى احتمال عدم وقوعه (p-1)) . ومن خلال جدول (٤) تبين معنوية المتغيرات (العمر(X_2) ،الوزن(X_4)) وتأثيرهما على الاصابة بمرض السكر من النوعين الاول والثاني ونلاحظ ارتفاع قيمه (Wald) للمتغيرات المعنوية.ونموذج الدراسة بالشكل التالي

$$\log(Y) = 17.233+3.427x^2+.140\ x^3$$

١-احتل متغير العمر (x_2) المرتبة الاولى في التأثير بالمتغير التابع (Y) اذ ان معامل الانحدار لهذا المتغير (b_2=3.427) وان هذا المعامل اظهر معنويته على المتغير التابع عند مستوى دلاله(0.00) من اجل (df=1) وان إحصاءه wald=(28.744) وان الخطأ المعياري لجميع المعاملات بحدود متقاربة .

2 - احتل متغير الوزن (x_4) المرتبة الثانيه في التأثير بالمتغير التابع (Y) اذ ان معامل الانحدار لهذا المتغير

(b_4=0.140) وان هذا المعامل اظهر معنويته على المتغير التابع عند مستوى دلاله(0.001) من اجل (df=1) وان إحصاءه (10.981)=wald وان الخطأ المعياري لجميع المعاملات بحدود متقاربة

الاستنتاجات :-

١- يتضح في عينه الدراسة ان نسبه الاصابة بالسكري بالنوع الاول كانت بالفئات العمريه(١٠-١٩),(٢٠-٢٩) اكثر من الفئات العمريه الاخرى ونسبه ألإصابة بالسكري من النوع الثاني كانت بالفئات العمريه (٣٠-٣٩),(٤٠ فأكثر) اكثر من غيرها.

٢- كفاءة النموذج المستنتج بشكل كلي في التنبؤ بالإصابة بمرض السكري من النوعين الاول والثاني على التوالي.

٣- نجحت المتغيرات المستقلة الداخلة في النموذج في تفسير ما نسبته (٤٥%) الى(٧٧%) من التغيرات في المتغير التابع وهذا ما اكدته قيم R^2 ولكن وبحسب هذه النتائج فان هناك عوامل اخرى لم تدرج في النموذج قد يكون لها تأثير وهذا العوامل قد تكون نفسيه او وراثية.

٤-ان نموذج متغيرات الدراسة اظهرت تأثيراً معنوياً لكل من (العمر، الوزن)في تصنيف المريض سواء كان مصاباً بالسكري من النوع الاول إم الثاني أو بلغت نسبة التصنيف الاجمالي الصحيح في النموذج(٩٤%) و التصنيف الصحيح بمرض السكر من النوع الاول(٧٥%) اما المصابون بمرض السكري من النوع الثاني فأن نسبة التصنيف الصحيح لهم(٩٨%).

التوصيات :-

١- نوصى بأجراء دراسات مماثله لأنواع مختلفة من الامراض ومعرفه العوامل المؤثره في كل نوع من تلك الانواع.

٢- نوصي بالدقة في تسجيل البيانات من خلال وضع ملاكات احصائية في المستشفيات . لان البيانات المسجله بشكل دقيق هي العامل الاساسي في الدراسات والبحوث. لذلك نوصي بتوخي الدقه في اختبار المتغيرات التوضيحيه.

٣- توسيع استخدام الانحدار اللوجستي ثنائي الاستجابة في الدراسات الطبية والتربوية اذا انحصرت استخداماته السابقه في المجالين الاقتصادي و الاجتماعي.

المصادر

المصادر العربيه

1.محمد عبد المنعم محمد الحسن،" استخدام الانحدار اللوجستي ثنائي الإستجابة لتحديد أهم العوامل الاقتصادية والاجتماعية والديمغرافية المؤثرة على الاصابة بالايدز ولاية الجزيرة ".(السودان ٢٠١٠- ٢٠١١ م)، جامعة الجزيرة ، ٢٠١٣م .

2.محمد عبدالرحمن إسماعيل ، " تحليل الإنحدار الخطي " معهد الإدارة العامة مركز البحوث ٢٠٠١م- السعودية.

3.ناريمان أحمد ، نجاة عبدالقادر ، ندى الطيب ، ندى عبدالله ، ندى عبد الماجد ،" استخدام أسلوب الانحدار اللوجستي الثنائي لتحديد أهم العوامل المؤثرة على الإصابة بمرض سرطان الثدي لدى الإناث دراسة حالة ولاية الخرطوم ٢٠١٥م" جامعة أمدرمان الإسلامية ، ٢٠١٥م.

ثانيا: المصادر الانكليزية

4- Binary Logistic Regression with PASW/SPSS, available at :

http://core.ecu.edu/psyc/wuenschk/MV/Multreg/Logistic-SPSS.PDF

5-Cramer , J. S. (2002) , The Origins of Logistic Regression , Tinbergen Institute Discussion Paper TI 2002-119/4 , available at:
http://www.tinbergen.nl/discussionpapers/02119.pdf

6- Garson , David (2006), Logistic Regression , available at: