مجلة قەڵای زانست العلمية
مجلة علمية دورية محكمة تصدر عن الجامعة اللبنانية الفرنسية – اربيل، كوردستان، العراق
المجلد(٣) – العدد (٣)، صيف ٢٠١٨
رقم التصنيف الدولي: ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

إستخدام أنموذج اللوجستي لدراسة أهم العوامل المؤثرة التي تمييز بين المصابين

وغير المصابين بالأورام الخبيثة

Using the logistic model to study the most important factors that distinguish between the infected And not infected with malignancies

مدرس مساعد / زينب عبدالله محمد

قسم الإحصاء / كلية الادارة والاقتصاد/ جامعة صلاح الدين – أربيل

zainab.Muhammad@su.edu.krd

معلومات البحث

تاريخ البحث:
الاستلام: ٢٠١٨/٣/٢٥
القبول: ٢٠١٨/٥/٤
النشر: **صيف ٢٠١٨**

Doi: 10.25212/lfu.qzj.3.3.19

الكلمات المفتاحية:
Malignant tumors, infected, logistics, discrimination of the infected, triglycerides, influencing factors, urea

الملخص

أستخدم في هذا البحث تحليل التمييز أنموذج اللوجستي التي تمييز بين المصابين وغير المصابين بالأورام الخبيثة , واستخدمت طريقة الإمكان الأعظم لتقدير معلمات أنموج المستخدم لتحديد مدى تأثير المتغيرات التوضيحية.

وقد تم إستخدام أختبار (Wald) لمعريفة معنوية المتغيرات التوضيحية وأختبار حسن المطابقة (Person Chi - Square) لمعرفة ملائمة الأنموذج , ولمعرفة قوة الأنموذج نستخدم (R^2) لأنموذج اللوجستي الذي يفسر تأثير المتغير التوضيحي على التمتغير المعتمد. ولتحقيق أهداف البحث قسم البحث الى الجانب النظري فيركز على بيان وتوضيح المفاهيم الأساسية لتمييز اللوجستي ومكوناته المختلفة مع الإشارة الى أهداف وإستخدامات التمييز اللوجستي في التحليل الإحصائي والفرضيات التي يستند عليها باستخدام أنموذج اللوجستي ، أما الجانب التطبيقي فقد تضمن بيانات للمرضى المصابين وغير المصابين بالأورام الخبيثة تم الحصول عليها من مستشفى رزطاري ومستشفى نانةكةلي في أربيل , وتبين ان نموذج التمييز اللوجستي ملائم في دراسة العوامل المؤثرة على المصابين بأورام الخبيثة وأن المتغيرات (يوريا (X_2) والدهون الثلاثية (X_5)) هم المتغيرات المعنوية في الدراسة أو تعد مؤثرة في المتغير المعتمد..

مجلة قەڵای زانست العلمية
مجلة علمية دورية محكمة تصدر عن الجامعة اللبنانية الفرنسية – اربيل، كوردستان، العراق
المجلد(٣) – العدد (٣)، صيف ٢٠١٨
رقم التصنيف الدولي: ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

المقدمة Introduction [10] [4] [3]:

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 و التحليل و التفسير للنموذج مشابها تماما لأسلوب الأنحدار المتعدد.

أستخدمت دالة تمييز اللوجستي لتمثيل الأحتمال اللاحق كأساس للتمييز بين مجتمعين ثم عممت الى حالة أكثر من مجتمعين ويعتبر التمييز اللوجستي وسط بين الطرق المعلمية (Parametric) المتعمدة على التوزيع الطبيعي متعدد المتغيرات والطرق اللامعلمية (Non Parametric) حرة التوزيع لهذا السبب تدعى هذه الطريقة بالطريقة شبه المعلمية (Semi - Parametric) مما جعل هذا النموذج ذو أهمية كبيرة في حل مسائل التمييز والأنحدار وذلك اللأسباب الآتية:-

1 - يتطلب القليل من الأفتراضات.
2 - أن الصيغ اللوجستية تنتج عن تنويع واسع من الأفتراضات التحتية حول المتغيرات التوضيحية.
   أ - القيمة تكون ثنائية تأخذ القيمةصفر أو واحد.
   ب - بعضها طبيعية متعددة المتغيرات.
   ت - بعضها متغيرات متصلة أو منفصلة.
   ث - بعضها متغيرات مرتبة حسب درجة الخطورة.
   ج - بعضها متغيرات مصنف.
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.

2- الجانب النظري:

(1-2) أنماذج الاختيار المتقطع Discrete choice models [5] [4] [7]:

تطبق هذه الأنماذج عندما تكون المتغيرات المتعمدة (متغيرات الاستجابة) وصفية ومحددة (Qualitative and limited dependent variables) أي تلك التي يتم التعبير عنها بمجموعة محددة من الصفات أو الحقائق

مجلة قەڵای زانست العلمية
مجلة علمية دورية محكمة تصدر عن الجامعة اللبنانية الفرنسية – اربيل، كوردستان، العراق
المجلد(٣) – العدد (٣)، صيف ٢٠١٨
رقم التصنيف الدولي: ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

(characteristic) أو انها محددة (مقيدة) بقيود لايمكن تجاهلها، اذ ليس دائماً يكون توقع المتغير المعتمد قيمة عددية يأخذ اي قيمة محصورة بين $(-\infty,\infty)$ كما في المنحني الطبيعي. وهنالك العديد من التطبيقات التي تستخدم كل من المتغيرات المتعمدة المحددة والوصفية وأنماذج الاختيار المتقطع في مجالات عدة تتضمن (الاقتصاد، المال، التسويق، العلوم الاجتماعية،...الخ).

وتعد أنماذج الاستجابة الثنائية والمتعددة حالة خاصة من هذه الأنماذج، فمثلاً الأنموذج ثنائي الاستجابة يطبق عندما يأخذ المتغير المعتمد قيمتين فقط $[0,1]$ التي قد تشير الى نمط الاستجابة لمؤشرما (كالادوية و المنشطات) اي انه محدد بحالة النجاح والفشل على التوالي.

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

$$f(Y)=P^{Y}(1-P)^{1-Y} \quad ,Y=0,1$$
$$\mathrm{E}(\mathrm{Y}_i)=\Pr[\mathrm{Y}_i=1]=P;\Pr[\mathrm{Y}_i=0]=1-P$$
$$V(\mathrm{Y}_i)=\mathrm{E}[(\mathrm{Y}_i-P)]^2=P(1-P)$$

وتعد أنماذج الاختيار الثنائية (binary choice models) كأحد الطرق المهمة في تحليل متغيرات الاستجابة الثنائية (binary or dichotomous response variable)، بحيث تكون استجابة المشاهدة (i) يمكن التعبير عنها بمتغير متقطع (Discrete variable) كما ان القيمة المتوقعة للمتغير المعتمد لاتعد دالة خطية بالنسبة للمتغيرات التوضيحية (كما في معادلة الانحدار الكمية) (Quantitative Regression equation) بل تكون على شكل دالة احتمالية خطية محددة بالفترة [0,1] وللتعبير عن أنموذج التميز لمثل هكذا متغيرات يكون بالصيغة الآتية:

$$y_i^{*}=\mathrm{X}_i'\beta+\varepsilon_i \quad i=1,2,\ldots,n \quad \ldots(1-2)$$

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 موجه المتغيرات التوضيحية.

β: يمثل موجه المعلمات الخاص بالأنموذج.

ε_i: يمثل الخطأ العشوائي.

عندما يأخذ المتغير المعتمد القيم الاتية:

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

مجلة قەڵای زانست العلمية
مجلة علمية دورية محكمة تصدر عن الجامعة اللبنانية الفرنسية – اربيل، كوردستان، العراق
المجلد(٣) – العدد (٣)، صيف ٢٠١٨
رقم التصنيف الدولي: ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

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

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

حيث X_i'، β معرفة كما في (2 -1).

وان احتمالية الاختيار البديل نحصل عليها من خلال

$$P(y_i=0)=1-P(y_i=1)$$

أما عندما يأخذ المتغير المعتمد (متغير الاستجابة) قيماً 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 (أو الشرطي) يأخذ الصيغة الاتية:

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

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

β :يمثل موجه المعلمات الخاص بالأنموذج.

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

(2-2) طرائق تقدير المعالم في النموذج[4][10][6]:

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.

مجلة قەڵای زانست العلمية
مجلة علمية دورية محكمة تصدر عن الجامعة اللبنانية الفرنسية – اربيل، كوردستان، العراق
المجلد(٣) – العدد (٣)، صيف ٢٠١٨
رقم التصنيف الدولي: ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

تم التطرق الى هذه الطريقة فقط لأنها أستخدمت في تقدير معلمات النموذج اللوجستي في هذا البحث:

تقديرات الإمكان الأعظم Maximum Likelihood Method[8][10][11]:

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 ($Z_1,...,Z_n$) حيث كل مشاهدة Z_i محدودة بواحدة من g من المجاميع ($Z_i=Z_{i1},.....Z_{ig}$) و أن $\sum_j X_{ij}=n_i$ يكون محدد لكل i تكون دالة الأمكان الاعظم الشرطية لتوزيع متعدد الحدود بالشكل الآتي:-

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

حيث أن:

$P_j(X_i)=$ نسبة أو احتمال الانتماء إلى المجموعة j وهي معرفة كما في المعادلة (2 -2) بأخذ المشتقة الاولى للوغاريتم الاحتمال الاعظم للمعادلة (2 -4) و مساواتها بالصفر نحصل على المعادلات الطبيعية الاتية:-

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

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

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

حيث أن:-

t (عدد الدورات التعاقبية)

r (هو موجه 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:

مجلة قەڵای زانست العلمية
مجلة علمية دورية محكمة تصدر عن الجامعة اللبنانية الفرنسية – اربيل، كوردستان، العراق
المجلد(٣) – العدد (٣)، صيف ٢٠١٨
رقم التصنيف الدولي: ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

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

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.

(3-2) تقييم أداء الآنماذج: Evaluating the Performance of the models [4] [10] [3] [6]:

اولاً: اختبارات حسن المطابقة: Goodness of Fit

لقد تم استخدام احصائتين لتقييم مدى حسن مطابقة الأنماذج وهي:

(2-3-1) اختبار مربع كاي لبيرسون Pearson chi - square [8] [2]:

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

H_0:

H_1: الأنموذج ملائم للبيانات

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 اللوجت متعدد الحدود باستخراج قيمة الاحصاءة وحسب الصيغة الآتية:

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

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

(2-3-2) احصاءة نسبة الامكان: Likelihood ratio statistics [4] [7] [9]:

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

$$H_0: \beta_1=\beta_2=.....=\beta_k=0 \qquad ...(6-2)$$

مجلة قەڵای زانست العلمية
مجلة علمية دورية محكمة تصدر عن الجامعة اللبنانية الفرنسية – اربيل، كوردستان، العراق
المجلد(٣) – العدد (٣)، صيف ٢٠١٨
رقم التصنيف الدولي: ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

at Least two of themare not equal zero.H_1:

والصيغة العامة اللاحصاءة هي:

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

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.

ثانياً: اختبار معنوية معلمات الأنموذج المقدر[10] [9] [5]:

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:

$$W = \left(\hat{\beta}_k - \hat{\beta}_{ko}\right)^2 / H_{kk} \qquad ...(7-2)$$

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.

مجلة قەڵای زانست العلمية
مجلة علمية دورية محكمة تصدر عن الجامعة اللبنانية الفرنسية – اربيل، كوردستان، العراق
المجلد(٣) – العدد (٣)، صيف ٢٠١٨
رقم التصنيف الدولي: ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

(4-2) اختبار Q [8] [4] [6]:

اقترح Q من قبل (Liu & Dyer ; 1988) لأنموذج الانحدار اللوجستي وهي شبيهة بمعامل التحديد R^2 في أنموج الإنحدار الخطي وتحسب كالآتي:

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

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

حيث ان:

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

$\bar{R}_0$: 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 (1,0) وفقاً لرتبتها. أن استخدامها الرئيسي يكون في المساعدة لاختبار أنموذج بين نماذج مختلفة تحتوي على عدد مختلف من المتغيرات التوضيحية.

يمكن 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.

(5-2) خطأ التصنيف Mss classification [10] [8] [1]:

أن تقييم تحليل نموذج اللوجستك عند استخدام تقديرات الامكان الأعظم لدالة الاحتمال الشرطية يتم عن طريق احصاءات جودة التوفيق للإحتمالات 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:

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

مجلة قەڵای زانست العلمية
مجلة علمية دورية محكمة تصدر عن الجامعة اللبنانية الفرنسية – اربيل، كوردستان، العراق
المجلد(٣) – العدد (٣)، صيف ٢٠١٨
رقم التصنيف الدولي: ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

ويكون احتمال خطأ التصنيف الكلي:

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

3- الجانب التطبيقي:

يتضمن هذا الفصل الجانب التطبيقي وهو أهم جزء من البحث وتمت تطبيق أنموذج اللوجستي للبيانات التي تم الحصول عليها من مستشفى رزطاري ومستشفى نانةكةلي في أربيل تحت عنوان (إستخدام أنموذج اللوجستي لدراسة أهم العوامل المؤثرة التي تمييز بين المصابين وغير المصابين بالأورام الخبيثة) , وقد تم تحليل النتائج عن طريق ادخال البيانات في البرنامج الإحصائي الجاهزة (Statistical Package for Social Sciences) (SPSS- V:22).

أولاً- وصف البيانات

تم جمع البيانات من كل مستشفى رزطاري ومستشفى نانةكةلي في أربيل بالاعتماد على مجموعة من الفحوصات الطبية حيث سحبت عينتين عشوائية مكونة من (75) و (60) مريضا يمثلون مجموعتين من المرضى المصابين وغير المصابين بالاورام الخبيثة أي ان مجموعة الكلية للعينة تساوي (135) المريض , حيث تمثل المجموعة الأولى فحوصات المرضى غير المصابين و المجموعة الثانية تمثل فحوصات المرضى المصابين بالاورام الخبيثة وأعتبرت الفحوصات كمتغير إستجابة فقد أعطيت القيمة (1) للمجموعة الأولى والقيمة (2) للمجموعة الثانية أما عدد المتغيرات الإضافية وهي (7) متغيرات يمكن توضيحها بالشكل الآتي :

1. (كلوكوز) نسبة السكر في الدم (X_1).
2. يوريا (X_2).
3. مستوى الكرياتينين (X_3).
4. نسبة الكولسترول في الدم (X_4).
5. الدهون الثلاثية (X_5).
6. الدهون العالية (X_6).
7. الدهون الواطئة (X_7).

ثانياً - تطبيق التمييز اللوجستي للبيانات:

يقوم برنامج (SPSS) بتحليل البيانات العينه بخطوات متعددة بدأ بالخطوه الصفرية التي يكون فيها النموذج خالياً من المتغيرات التوضحيه. وأدناه يوضح تلك النتائج.

مجلة قەڵای زانست العلمية
مجلة علمية دورية محكمة تصدر عن الجامعة اللبنانية الفرنسية – اربيل، كوردستان، العراق
المجلد(٣) – العدد (٣)، صيف ٢٠١٨
رقم التصنيف الدولي: ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

الجدول (2-1) المتغيرات الداخلة في المعادلة

	B	S.E.	Wald	d.f	Sig.	Exp(B)
Step 0 Constant	-0.223	0.173	1.660	1	0.198	0.800

الجدول (2-1) يوضح فيها النموذج خالياً من المتغيرات التوضحية. وحسب أختبار (Wald) ان النموذج المقدر غير معنوية لأن ($\alpha = 0.05$) < (0.198) = P-value .

الجدول (2-2) يوضح المتغيرات التي أستبعدت من نموذج اللوجستي في المرحلة الصفرية

الخطوة الصفر للمتغييرات	score	d.f	p-Value
(كلوكوز) نسبة السكر في الدم (X_1)	0.289	1	0.591
يوريا (X_2)	5.655	1	0.017
مستوى الكرياتينين (X_3)	1.270	1	0.260
نسبة الكولسترول في الدم (X_4)	0.780	1	0.377
الدهون الثلاثية (X_5)	11.448	1	0.001
الدهون العالية (X_6)	0.201	1	0.654
الدهون الواطئة (X_7)	0.155	1	0.694
Overall Statistics	16.025	7	0.025

الجدول (2-3) يوضح تأريخ التكرارات

مجلة قەڵای زانست العلمية
مجلة علمية دورية محكمة تصدر عن الجامعة اللبنانية الفرنسية – اربيل، كوردستان، العراق
المجلد(٣) – العدد (٣)، صيف ٢٠١٨
رقم التصنيف الدولي: ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

تكرار	-2 Log likelihood	المعاملات							
		الثوابت	X_1 نسبة السكر في الدم	X_2 يوريا	X_3 الكرياتينين	X_4 الكولسترول في الدم	X_5 الدهون الثلاثية	X_6 الدهون العالية	X_7 الدهون الواطئة
Step 1	168.482	2.205	.010	-1.469	-.001	-.010	-1.490	.003	-.003
2	168.040	2.604	.011	-1.761	.000	-.014	-1.731	.003	-.003
3	168.038	2.632	.011	-1.781	.000	-.015	-1.748	.003	-.003
4	168.038	2.632	.011	-1.781	.000	-.015	-1.748	.003	-.003
Step 2 1	168.481	2.194	.010	-1.473		-.010	-1.492	.003	-.003
2	168.040	2.605	.011	-1.761		-.014	-1.731	.003	-.003
3	168.038	2.634	.011	-1.780		-.015	-1.748	.003	-.00

مجلة قەڵای زانست العلمية
مجلة علمية دورية محكمة تصدر عن الجامعة اللبنانية الفرنسية – اربيل، كوردستان، العراق
المجلد(٣) – العدد (٣)، صيف ٢٠١٨
رقم التصنيف الدولي: ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

4	168.038	2.634	.011	-1.780		-.015	-1.748	.003	-.003
Step 3 1	168.716	1.702	.011	-1.425		-.010	-1.530	.002	
2	168.263	2.115	.011	-1.708		-.014	-1.775	.003	
3	168.260	2.147	.011	-1.727		-.014	-1.793	.003	
4	168.260	2.147	.011	-1.728		-.014	-1.793	.003	
Step 4 1	169.050	1.992		-1.329		-.011	-1.463	.002	
2	168.580	2.418		-1.602		-.015	-1.723	.003	
3	168.578	2.451		-1.622		-.016	-1.742	.003	
4	168.578	2.451		-1.623		-.016	-1.742	.003	
Step 5 1	169.683	1.673		-1.244			-1.534	.002	
2	169.328	1.960		-1.476			-1.783	.003	

مجلة قەڵای زانست العلمية
مجلة علمية دورية محكمة تصدر عن الجامعة اللبنانية الفرنسية – اربيل، كوردستان، العراق
المجلد(٣) – العدد (٣)، صيف ٢٠١٨
رقم التصنيف الدولي: ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

3	169.327	1.976		-1.490			-1.797	.003	
4	169.327	1.976		-1.490			-1.797	.003	
Step 6 1	170.367	1.907		-1.149			-1.546		
2	170.052	2.208		-1.340			-1.792		
3	170.052	2.224		-1.351			-1.805		
4	170.052	2.224		-1.351			-1.805		

الجدول (2-4) اختبار كاي سكوير للمراحل السادسة

		قيمة كاي سكوير	الدرجة الحرية	p- Val ue
Step 1	Model	17.442	7	0.015
Step 2	Model	17.442	6	0.008
Step 3	Model	17.219	5	0.004
Step	Model	16.902	4	0.002

مجلة قەڵای زانست العلمية
مجلة علمية دورية محكمة تصدر عن الجامعة اللبنانية الفرنسية – اربيل، كوردستان، العراق
المجلد(٣) – العدد (٣)، صيف ٢٠١٨
رقم التصنيف الدولي: ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

4				
Step 5	Model	16.153	3	0.001
Step 6	Model	15.428	2	0.000

الجدول (2-4) يوضح معنوية النموذج من خلال احصاءة كاي سكوير، وبما ان القيم (P- value) أقل من ($\alpha = 0.05$) فإن النموذج معنوي في الخطوات السادسة , أي المتغير (المصابين وغير المصابين بالاورام الخبيثة) يكون معنوياً.

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

الجدول (2-5) قيم R^2 لأنموذج اللوجستي

Step	-2Log likelihood	Cox & Snell R^2	Nagelkerke R^2
1	168.038	0.121	0.162
2	168.038	0.121	0.162
3	168.260	0.120	0.160
4	168.578	0.118	0.158
5	169.327	0.113	0.151
6	170.052	0.108	0.145

نلاحظ في الجدول (2-5) وبالاعتماد على احصاءة (Cox &Snell R^2) أن انخفاض قيمة هذه الاحصاءة بالخطوة الاخيرة بمقدار (0.013) عن الخطوة الاولى، وكذلك احصاءة (Nagelkerke R^2) انها انخفضت بالخطوة الاخيرة بمقدار (0.017) عن الخطوة الأولى، واختبار النسبة الامكان الاعظم ازدادت قيمه بالخطوة الاخيرة بمقدار (2.014) عن الخطوة الأولى،

مجلة قەڵای زانست العلمية
مجلة علمية دورية محكمة تصدر عن الجامعة اللبنانية الفرنسية – اربيل، كوردستان، العراق
المجلد(٣) – العدد (٣)، صيف ٢٠١٨
رقم التصنيف الدولي: ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

فبهذا يكون الانموذج الذي له اقل قيمة هو الافضل، ولمعرفة ان هذا الفرق معنوي ام لا نستخدم اختبار كاي سكوير للنموذج اللوجستي كالآتي:

H_0 : الأنموذج ملائم

H_1 :الأنموذج غير ملائم

الجدول (2-6) اختبار كاي سكوير للنموج اللوجستي

Hosmer and LemeshowTest			
Step	Chi - square	df	Sig.
1	7.493	8	0.484
2	7.492	8	0.485
3	9.572	8	0.296
4	6.575	8	0.583
5	2.982	8	0.935
6	1.118	8	0.997

من الجدول (2-6) نلاحظ بأن الانموذج الملائم للبيانات في كل الخطوات.

الجدول(2-7) القيم المشاهدة والمتوقعة للنموذج اللوجستي

Contingency Table for Hosmer and LemeshowTest

	Group = 1	Group = 2	Total

مجلة قەڵای زانست العلمية
مجلة علمية دورية محكمة تصدر عن الجامعة اللبنانية الفرنسية – اربيل، كوردستان، العراق
المجلد(٣) – العدد (٣)، صيف ٢٠١٨
رقم التصنيف الدولي: ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

	قيمة المشاهدات	القمة المتوقعة	قيمة المشاهدات	القمة المتوقعة	
1	12	11.936	2	2.064	14
2	10	10.291	4	3.709	14
3	10	9.401	4	4.599	14
4	8	8.611	6	5.389	14
5	8	7.909	6	6.091	14
6	8	7.040	6	6.960	14
7	7	6.210	7	7.790	14
8	5	5.747	9	8.253	14
9	5	5.194	9	8.806	14
10	2	2.660	7	6.340	9

الجدول (2-8) المتغيرات الداخلة في النموذج

Variables	B	S.E	Wald	df	p-Value	Exp(B)
(كلوكوز) نسبة السكر في الدم (X_1)	0.011	0.020	0.268	1	0.605	1.011
يوريا (X_2)	-1.781	0.848	4.413	1	0.036	0.168
مستوى الكرياتينين (X_3)	0.000	0.031	0.000	1	0.993	1.000
نسبة الكولسترول في الدم	-	0.024	0.386	1	0.534	0.985

مجلة قەڵای زانست العلمية
مجلة علمية دورية محكمة تصدر عن الجامعة اللبنانية الفرنسية – اربيل، كوردستان، العراق
المجلد(٣) – العدد (٣)، صيف ٢٠١٨
رقم التصنيف الدولي: ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

(X_4)	0.015					
(X_5) الدهون الثلاثية	-1.748	0.634	7.596	1	0.006	0.174
(X_6) الدهون العالية	0.003	0.003	1.009	1	0.315	1.003
(X_7) الدهون الواطئة	-0.003	0.006	0.221	1	0.639	0.997
Constant	2.632	1.547	2.895	1	0.089	13.900

الجدول(2-8) يبين المتغيرات التوضيحية الداخلة في النموذج ، اذ نلاحظ المتغيرات ويعرض نتائج تقدير أنموذج التمييز اللوجستي ومنه نلاحظ:

ان قيم اختبار (wald) يمثل قيم اختبار المعلمات للانموذج ويظهر بأن المتغير يوريا (X_2) والمتغير الدهون الثلاثية (X_5) هم المتغيرات المعنوية في الدراسة وذلك من خلال مقارنة القيم (P-value) مع مستوى المعنوي ($\alpha = 0.05$) ، ان العمود (P-value) يمثل معنوية تأثير المتغيرات على حالة المصاب ، (عندما تكون قيم P-value أقل من ($\alpha = 0.05$) لمتغير معتمد يكون معنويا) ، لذلك فأن المتغيرين (يوريا , الدهون الثلاثية) تعد مؤثرة في المتغير المعتمد (الاستجابة) والذي يمثل حالة المصاب بمرض بالأورام الخبيثة وان بقية المتغيرات الأخرى ليس لها تأثير ، لذلك فان المتغيرين (يوريا , الدهون الثلاثية) سوف تبقى في النوذج وبقية المتغيرات سوف تحذف من الأنموذج.

الجدول (2-9) التصنيف للمرحلة الأخيرة لأنموذج اللوجستي

Classification Table

المشاهدات		Predicted		
		Group		percentage Correct
		1	2	
Group	1	53	22	70.7
	2	24	36	60.0

مجلة قەڵای زانست العلمية
مجلة علمية دورية محكمة تصدر عن الجامعة اللبنانية الفرنسية – اربيل، كوردستان، العراق
المجلد(٣) – العدد (٣)، صيف ٢٠١٨
رقم التصنيف الدولي: ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

Overall Percentage			65.9

الجدول (2-9) هو جدول التصنيف للمتغير المعتمد إذ ان النسبه المئوية الكلية للدقة تنبؤ للمتغير المعتمد بالاعتماد على النموذج هي (65.9% وهي نسبة مقبولة أو جيدة في التصنيف.).

4 / الأستنتاجات:

1. إن أنموذج اللوجستي ملائمه لإستخدام دراسة العوامل المؤثرة التي تمييز بين المصابين وغير المصابين بالأورام الخبيثة.
2. ان نسبة التصنيف الصحيح للدالة التميز الخطية هي (65.9%) وهي نسبة مقبولة أو جيدة في التصنيف.
3. نلاحظ في الجدول (2-4) لمعنوية النموذج من خلال احصاءة كاي سكوير، بما ان P- value اقل من ($\alpha = 0.05$) فإن النموذج معنوي في الخطوات السادسة.
4. لاحظنا في الجدول (2-5) بأن قيمة احصاءة (Cox & Snell R^2) انخفض بالخطوة الاخيرة بمقدار (0.013) عن الخطوة الاولى، وكذلك احصاءة (Nagel kerke R^2) انها انخفضت بالخطوة الاخيرة بمقدار (0.017) عن الخطوة الأولى، واختبار النسبة الامكان الاعظم ازدادت قيمه بالخطوة الاخيرة بمقدار (2.014) عن الخطوة الأولى، فبهذا يكون الانموذج في الخطوة السادسة هو الأفضل.
5. تبين بأن اختبار كاي سكوير للأنموج اللوجستي الملائم للبيانات في كل الخطوات.
6. تبين ان معلمات المقدرة لنموذج اللوجستى معنويه و ذلك من خلال قيمة (P- value) لأختبار (Wald) ومقارنته مع مستوى المعنوي ($\alpha = 0.05$).
7. تتبين بأن المتغيرين (يوريا , الدهون الثلاثية) تعد مؤثرة في المتغير المعتمد (الاستجابة) والذي يمثل حالة المصاب بمرض بالأورام الخبيثة وان بقية المتغيرات الأخرى ليس لها تأثير.

5 – المصادر:

1. بابان ، ياسين عبدالقادر جميل (1988): "إستخدام طريقة الإمكان الأعظم في تقدير معلمات أنموذج اللوجستي مع تطبيق عملي". رسالة ماجستير، قسم الإحصاء كلية الادارة والاقتصاد، جامعة بغداد- بغداد، العراق.
2. البلداوي , تنسيم حسن (1996) , "مقارنة تحليلية بين نماج اللوجستي ونماذج الدوال التمييزية" , رسالة دكتوراه , كلية الادارة والاقتصاد، جامعة بغداد- بغداد، العراق.
3. البياتي، هبة ابراهيم صالح (2005): "تحليل المسار في نموذج الانحدار اللوجستي مع تطبيق". رسالة ماجستير ، قسم الإحصاء كلية الادارة والاقتصاد، جامعة المستنصرية - بغداد ، العراق.
4. الجبوري , شلال حبيب وعبد ,صلاح حمزة (2000) , "متعدد المتغيرات" , دار الكتب للطباعة والنشر.
5. داود، ألاء عبدالستار (2005): "استخدام الدالة التميزية وطرق تحديد متغيراتها". رسالة ماجستير، كلية علوم الحاسبات والرياضيات، جامعة الموصل، العراق.

6. مولود ، كوردستان ابراهيم (2000): "استخدام التحليل المميز لتشخيص اهم العومل المؤثرة في التصنيف السريري لمرض القلب ". رسالة ماجستير قسم الإحصاء كلية الادارة والاقتصاد، جامعة صلاح الدين/ أربيل، كوردستان العراق.
7. Afifi, A. and Clark, V (1984): "Computer Aided Multivariate Analysis". Life time Learning Publications, USA.
8. Anderson, T. W. (1984): "An Introduction to Multivariate Statistical Analysis". 2nd - Edition, John Wiley and Sons, New York-USA.
9. 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, Erbil, Iraq.
10. Menard, Scott.(2002): "Applied Logistic Regression Analysis". Second Edition, Sage Publication, Inc.
11. Walter , W.& Hauck, J.R.& Donner (1970): "Walds Test as Applied to Hypothesis in Logic Analysis" . JASA. Vol. 36, pp. 851-853.

مجلة قەڵای زانست العلمية
مجلة علمية دورية محكمة تصدر عن الجامعة اللبنانية الفرنسية – اربيل، كوردستان، العراق
المجلد(٣) – العدد (٣)، صيف ٢٠١٨
رقم التصنيف الدولي: ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

پوختە

لەم توێژینەوەیەدا شیکردەوەی جیاکردنەوەی مۆدێلی لۆجستی بەکارهێنراوە بۆ کەسێك کە توشی نەخۆشیەکە بووە یان توش نەبووە (نەخۆشە تووشبووەکان وتووشنەبووەکان بە (شێرپەنجە پیسەکان)) , وە ڕێگای گەورەترین پێوەر (maximum likelihood method) بەکارهێنراوە بۆ خەملاندنی پارامیتەری مۆدیلی لۆجستی بۆ دیاری کردنی ڕادەی کاریگەری گۆڕاوە ڕوونکراوەکان. وە تاقیکردنەوەی (Wald) بەکارهێنراوە بۆ زانینی گونجاندنی گۆڕاوە ڕوونکراوەکان , وە بۆ زانی هێزی مۆدیلەکە ئەوا (R^2) بەکاردەهێنین بۆ مۆدیلی لۆجستیك لەکاتێك کاریگەری گۆڕاوە ڕوونکراوەکان لەسەر گۆڕاوە پەیوەستەکان ڕوونەکاتەوە, وە بۆ جێبەجێکردنی ئامانجی توێژینەوە دابەشکراوە بۆ لایەنی تیۆری کە ئاماژە دەکات بۆ نیشاندان وڕوونکردنەوەی پێوەکە سەرەکیەکانی کە لە جیاکردنەوەی مۆدێلی لۆجستی بەکاردێت وەپێکهاتە جیاوازەکان لەگەڵ دیاری کردنی ئامانجەکان وبەکارهێنانەکانی جیاکردنەوەی مۆدێلی لۆجستی لەشیکەرەوەی ئاماری وگریمانەکان کە پەیوەستە بە بەکارهێنانی جیاکردنەوەی مۆدێلی لۆجستی، بەلام لایەنی پراکتیکی بریتیە لە داتای نەخۆشە تووشبووەکان وتووشنەبوەکان بە (شێرپەنجە پیسەکان) کە دەست کەوت لە نەخۆشخانەی ڕزگاری ونەخۆشخانەی نانەکەلی لە هەولێر وە ئەنجامەکانمان دەست کەوت کە بە پێزانینی ئێمە گرنگە ودەبێت تێبینی بکرێت لە لایەن لایەنی تەندروستیەکان.

Abstract:

In this Research analysis of the discrimination of Logistics model is used, which discriminately between patients infected and not infected with malignant tumoursby, and the maximum likelihood method was used to estimate the parameters of the logistic model to determine the most important factors affecting.

And (Wald) test was used to determine the significance of the explanatory variables, (Person Chi-Square) used for goodness of fit Model, and to know power of model we use (R^2) for the logistic model to determine the explanatory variable effect on dependent variable.

And to achieve the search of objectives the research partition on the theoretical side was addressed to the analysis of the logistic model, the applied side included data for patients with infected and non-tumours infected, obtained from (Rizgarry Hospital) and (Nana Kali Hospital) in Erbil have been reached to conclusions that we believe are important and should be noted by health providers.