مجلة

كلية التـــراث الجامعة

مجلة علمية محكمة

متعددة التخصصات نصف سنوية

العدد السابع والثلاثون

15 حزيران 2023

ISSN 2074-5621

رئيس هيئة التحرير

أ.د. جعفر جابر جواد

مدير التحرير

أ. م. د. حيدر محمود سلمان

رقم الايداع في دار الكتب والوثائق 719 لسنة 2011

مجلة كلية التراث الجامعة معترف بها من قبل وزارة التعليم العالي والبحث العلمي بكتابها المرقم (ب 3059/4) والمؤرخ في (2014/ 4/7)

استعمال خوارزمية الامكان الموضعي (Logit) لتقدير الانحدار اللوجستي ثنائي الاستجابة

الباحثة نرجس باسم خلف **أ.د. لقاء علي محمد**
جامعة بغداد - كلية الادارة والاقتصاد

الخلاصة:
في هذا البحث تم الاعتماد على طرائق المقدر اللبي في تقدير الانحدار اللوجستي ثنائي الاستجابة، حيث تم استعمال خوارزمية الامكان الموضعي اللوجت (Logit)، تم تقدير معلمة تمهيد (λ) مثلى بطريقتين العبور الشرعي والعبور الشرعي المعمم، وان معلمه التمهيد لها تأثير واضح في عملية التقدير ايضا، ولها دور اساسي في تقريب وتنعيم المنحني واقترابه من المنحنى الحقيقي، وان الهدف من استعمال المقدرات اللبية هو تعديل المشاهدات لكي نتمكن من الحصول على مقدرات ذات صفات قريبة من خواص المعلمات الحقيقية، وبالاعتماد على بيانات طبية تم جمع بيانات عن المصابين بأمراض سرطان الدم الليمفاوي المزمن (تمثل المصابين والغير مصابين) ومن خلال استعمال دالة (Gaussian) وبالاعتماد على معيار المقارنة (MSE)، تبين ان طريقة خوارزمية الامكان الموضعي اللوجت (Logit) هي الافضل عند احجام عينات (50,100,150) وعند قيمة تباين (0.1) وذلك لحصولها على اقل قيمة لهذا المعيار.
الكلمات المفتاحية : انموذج الانحدار اللوجستي الثنائي، خوارزمية اللوجت الامكان الموضعي، طريقة العبور الشرعي، طريقة العبور الشرعي المعمم، مرضى سرطان الدم الليمفاوي المزمن.

1- المقدمة:

يعد الانحدار الخطي من ابسط وأهم تطبيقات الانحدار ويتطلب ان يكون المتغير التابع متصلا و غير محدود ولكن هنالك الكثير من الحالات التي يكون فيها متغير الاستجابة نوعيا ثنائي القيمة (1,0) بدلا من ان يكون متصلا كما في العلوم الانسانية والاجتماعية والطبية وغيرها، فان العديد من الظواهر الطبيعية عند دراسة سلوكها تجدها تسلك سلوكاً غير خطي ولتحليل تلك الظواهر يستعمل النماذج اللاخطية لوصف وتحليل تلك الظواهر ومن هذه النماذج انموذج الانحدار اللوجستي الثنائي.

2- مشكلة البحث :

العديد من الدراسات وتحديدا الطبية منها يكون متغير الاستجابة عبارة عن توزيع ثنائي الحدين (برنولي) نلجأ حينها إلى استعمال التقدير عن طريق الانحدار اللوجستي الذي يتعامل مع هكذا بيانات بشرط توفر كافة شروط التقدير اضف الى ذلك ان المتغيرات التوضيحية ليس لها توزيع محدد او ذات علاقة لاخطية مع متغير الاستجابة الامر الذي يكون فيه انموذج الانحدار اللوجستي المعلمي غير مجدي ويدفعنا للبحث عن اسلوب يتناسب مع المشكلة الحالية .

3-الجانب النظري:

3-1 أنموذج الانحدار اللوجستي (Logistic regression Model)

يعد أنموذج الانحدار اللوجستي من اهم النماذج المستعملة في التطبيقات الاحصائية وتحليل المشاهدات ومن اكثر النماذج المستعملة في دراسة العلوم (الاجتماعية، الطبية، الاقتصادية) بشكل عام يستعمل لتفسير وتحليل العلاقة بين متغير مستقل واحد او اكثر والمتغير التابع[Shim and Seok ,2012,pp:5].

في نموذج الانحدار اللوجستي الثنائي تكون العلاقة بين المتغير الاستجابة والمتغيرات التوضيحية غير خطية حيث يكون المتغير الاستجابةY ثنائي الاستجابة يأخذ إحدى القيمتين (اما 0 او 1) أما النجاح (success) حدوث الاستجابة باحتمال (p) أو الفشل (failure) عدم حدوث الاستجابة باحتمال $(1-p)$ و علية فأنه المتغير التابع Y يتوزع توزيع برنولي[العزاوي pp:16-17 ,2005,]

$$y_i \sim Ber(p) \quad \ldots(1)$$

بدالة كتلة احتمالية

$$p(Y = y_i) = p^{y_i}(1-p)^{1-y_i} \quad \ldots(2)$$

اذ إن:

y_i : متغير تابع ثنائي الاستجابة.

p : احتمال حدوث الاستجابة عندما($y_i = 1$).

لذلك فان توقع المتغير التابع يمثل احتمال حدوث الاستجابة.

$$E(y_i) = p(Y = 1) = p \quad \ldots(3)$$

أما تباين المتغير التابع حسب توزيع برنولي فيكون

$$V(y_i) = pq \quad \dots (4)$$

وللانحدار اللوجستي عدة انواع: منها انموذج الانحدار اللوجستي الثنائي الذي سيتم دراسته في هذا البحث وانموذج الانحدار اللوجستي المتعدد الذي يكون فيه المتغير التابع اكثر من قيمتين وانموذج الانحدار اللوجستي الرتبوي أو الترتيبي الذي يكون فيه المتغير التابع متغيراً رتبوياً [شهاب :2016, pp:9-11]

2-3Local Likelihood Estimation تقدير الإمكان الموضعي

بشكل عام تتمثل طريقة التقدير الإمكان الموضعي في الحصول على $\hat{\beta}$ حيث ان:[pp:2,2020,Rifada]

$$\hat{\beta} = \beta^{\arg max} \sum_{i=1}^{n} lnf\big(y_i, g(X_i,\beta)\big) K_\lambda(X_i - x) \quad \dots (5)$$

وان $lnf\big(y_i, g(X_i,\beta)\big)$ هو log-pdf من المشاهدات (Y_i, X_i)، و $g(X_i,\beta) = E(Y/X_i)$ تم استعمال التقدير الموضعي من اجل تقدير دالة الكثافة والخطورة، لكن نادراً ما يتم استعمال وتطبيق لتقدير دالة الانحدار مع الاستجابة الثنائية.

3-3 تقدير اللوجت الإمكان الموضعي Local Likelihood Logit Estimation [pp:4,2020,Rifada]

بالنظر الى البيانات (Y_i, X_i) حيث $(i = 1,2, \dots \dots, n)$ مع متغير الاستجابة Y'_i له فئتان، ومتغيرات توضيحية $X_i = (1, X_{1i}, X_{2i}, \dots, X_{pi})$ وهي من النوع المستمر. من المفترض ان البيانات المزدوجة تتوافق مع انموذج الانحدار اللوجستي الثنائي اللامعلمي، فأن طريقة تقدير لوجت الإمكان الموضعي التي تم الحصول عليها عن طريق استبدال لوغاريتم الاحتمالية لدالة الكثافة من (Y_i, X_i) وبذلك نحصل على تقدير لوجت الإمكان الموضعي وكلاتي:

$$L = \sum_{i=1}^{n} \big[Y_i \, ln\, \pi(X_i, x_0) + (1 - Y_i) \ln\big(1 - \pi(X_i, x_0)\big)\big] K_\lambda(X_i - x_0) \quad \dots (6)$$

حيث $K_\lambda(X_i - x_0)$ هو مضاعف دالة النواة (Kernel) لمتغيرات التوضيحية من النوع المستمر، أي ان :

$$K_\lambda(X_i - x_0) = \prod_{j=1}^{p} K\left(\frac{X_{ji} - x_j}{\lambda_j}\right) \quad \dots (7)$$

X_{ji} و x_j :كل منهما هو قيمة المشاهدة لـj لمتغير التوضيحي X_j والنقطة الافتراضية الثابتة لـj .

K:هي دالة النواة (Kernel) الأحادية المتماثلة ، و λ_j هي معلمة التمهيد لمتغير التوضيحي X_j.

4-3 اختيار معلمة التمهيد المثلى Selection of Optimal Bandwidth [pp:5,2020,Rifada]

يعد اختيار معلمة التمهيد (λ) مهماً جداً في الحصول على مقدر دالة الانحدار بناءً على نهج اللامعلمي، وان معلمة التمهيد هو وحدة التحكم في التوازن بين سلاسة دالة البيانات، اذا كانت(λ) صغيرة جداً فستكون الدالة المقدرة التي تم الحصول عليها خشنة للغاية وستنتقل للبيانات، بينما اذا كانت (λ) كبيرة جداً فأن الدالة المقدرة التي تم الحصول عليها ستكون سلسة للغاية وفي اتجاه متوسط متغير الاستجابة.

لذلك عند اختيار (λ) من المتوقع ان تكون القيمة هي المثلى، وهناك طريقتان للحصول على (λ) المثلى، وهما طريقة العبور الشرعي (CV) وطريقة العبور الشرعي المعمم (GCV).

اذا كان معيار مفكوك (λ) مقتصراً على فئة مقدر خطي فهناك مصفوفة$A(\lambda)$ بحجم $(n \times n)$ بحيث

$$\hat{m}(x) = A(\lambda)Y \quad \dots (8)$$

ويتم الحصول على القيمة المثلى من خلال تقليل GCV الى الحد الأدنى كالاتي:

$$GCV(\lambda) = \frac{MSE(\lambda)}{\left(\frac{1}{n} tr[I - A(\lambda)]\right)^2} \quad \dots (9)$$

وان

$$MSE(\lambda) = \frac{1}{n}\sum_{i=1}^{n}[y_i - \hat{f}(x_i)]^2 \quad \dots (10)$$

بالنسبة لطريقة تقدير الإمكان الموضعي يتم استعمال معاير $CV_{ML}(\lambda)$ التي تتوافق مع دالة الإمكان كالاتي:

$$CV_{ML}(\lambda) = \sum_{i=1}^{n} Y_i \ln g\big(X_i, \hat{\beta}(x_0)_{-x_i/\lambda}\big) + (1 - Y_i) \ln\big[1 - g\big(X_i, \hat{\beta}(x_0)_{-x_i/\lambda}\big)\big] \quad \dots (11)$$

وان $\hat{\beta}(x_0)_{-x_i/\lambda}$ هو تقدير معامل (leave-one-out) لتقدير $E[Y/X=X_i]$ الذي تم الحصول عليه من بيانات العينة دون مراقبة i.

5-3 *تقدير انموذج الانحدار اللوجستي الثنائي اللامعلمي* Estimation of Nonparametric Binary Logistic Regression Model

العينة المستقلة (Y_i, X_i) وان $(i = 1,2, \dots \dots, n)$، حيث يفترض ان متغير الاستجابة $Y = (0,1)$ ومتجه المتغيرات التوضيحية $X_i = (1, X_{1i}, X_{2i}, \dots, X_{pi})$ يتوافق مع انموذج الانحدار اللوجستي الثنائي اللامعلمي كالاتي:

[pp:6,2020,Rifada]

$$G(X_i, x_0) = X_i\beta(x_0) \quad ; i = 1,2, \dots, n \qquad \dots (12)$$

وان دالة الارتباط اللوجت هي

$$G(X_i, x_0) = \ln\left[\frac{\pi(X_i, x_0)}{1 - \pi(X_i, x_0)}\right] \qquad \dots (13)$$

$\beta(x_0)$:عبارة عن متجه من المعلمات التي تتوافق مع المتغير التوضيحي X_i .

من الصيغة (*12*) يتم الحصول على احتمالية النجاح في التجربة i والتي تتوافق مع المتغير التوضيحي X_i وهو:

$$g\big(X_i, \beta(x_0)\big) = E[Y/X = X_i] = \pi(X_i, x_0) = \frac{exp\big(X_i, \beta(x_0)\big)}{1 + exp\big(X_i, \beta(x_0)\big)} \quad \dots (14)$$

لتقدير $\beta(x_0)$ في الصيغة (14) يتم استعمال طريقة الإمكان الموضعي، ويتم الحصول على دالة اللوجت للإمكان الموضعي كالاتي:

$$L = \sum_{i=1}^{n}\{Y_i X_i\beta(x_0) - \ln\big[1 + exp\big(X_i, \beta(x_0)\big)\big]\}K_\lambda(X_i - x_0) \quad \dots (15)$$

حيث ان الشرط الكافي لدالة لوجت الإمكان الموضعي في الصيغة (15) تصل الى القيمة القصوى وهي $\frac{dL}{d\beta(x_0)} = 0$

لذلك نحصل على :

$$\frac{dL}{d\beta_j(x_0)} = \sum_{i=1}^{n}\big[X_{ji}\big(Y_i - \pi(X_i, x_0)\big)\big]K_\lambda(X_i - x_0) = 0 \qquad \dots (16)$$

$$j = 0,1, \dots p \quad ; \quad X_{0i} = 1$$

ان الصيغة (16) تمثل مقدرات $\beta(x_0)$ ، لا يمكن حلها بالطرق التحليلية لذلك يتم حلها بطرق عددية باستعمال خوارزمية نيوتن رافسون، بعد ذلك يتم تحديد متجه الانحدارات حيث يتم الحصول على عناصره من معادلة (*16*) كالاتي:

$$f\big(\beta(x_0)\big) = \left(\frac{dL}{d\beta_0(x_0)}, \frac{dL}{d\beta_1(x_0)}, \dots \dots, \frac{dL}{d\beta_p(x_0)}\right)' \qquad \dots (17)$$

والمشتقة الثانية لدالة لوجت الإمكان الموضعي للصيغة (15) كالاتي:

$$\frac{d^2L}{d\beta_j(x_0)^2} = -\sum_{i=1}^{n}\left[\frac{X^2{}_{ji}exp(X_i\beta(x_0))}{1+exp(X_i\beta(x_0))}\right]K_\lambda(X_i - x_0) \qquad \dots (18)$$

$$j = 1,2, \dots, p; X_{0i} = 1$$

$$\frac{d^2L}{d\beta_j(x_0)\beta_k(x_0)} = -\sum_{i=1}^{n}\left[\frac{X_{ji}X_{ki}exp\big(X_i\beta(x_0)\big)}{1 + exp\big(X_i\beta(x_0)\big)}\right]K_\lambda(X_i - x_0) \quad \dots (19)$$

$$for j \neq k \; ; j, k = 1,2, \dots p \; ; \; X_{0i} = 1$$

ويتم تحديد مصفوفة Hessian التي تم الحصول على عناصرها من الصيغتين (18) و(19) كالاتي:

$$H\big(\beta(x_0)\big) = \frac{d^2L}{d\beta(x_0)\beta(x_0)'} \qquad \dots (20)$$

ان الشرط الضروري لـ$\beta(x_0)$ لتعظيم دالة لوجت الامكان الموضعي لصيغة (*15*) هو مصفوفة Hessian في الصيغة (20) وهي سالبة مؤكدة.

6-3 خوارزمية لوجت الامكان الموضعي Algorithm Estimation of Nonparametric Binary Logistic Regression [pp:7,2020,Rifada]

لتقدير متجه المعلمة $\beta(x_0)$ في الصيغة (15) يتم تلخيص عمل الخوارزمية بالخطوات الأتية:

1. تحديد القيمة الابتدائية لمعلمه التمهيد لكل متغير توضيحي.
2. تقدير $\hat{\beta}(x_0)_{-x_i/\lambda}$ وان ($i = 1,2 \dots n$) باستعمال التحليل العددي لخوارزمية نيوتن–رافسون للبيانات العينة بدون المراقبة i كلاتي:

a) تحديد القيم الابتدائية لمتجه المعلمة $\beta(x_0)^{(0)}_{-X_i/\lambda}$ باستعمال انموذج الانحدار لوجستي الثنائي اللامعلمي مع دالة ارتباط لوجت.

b) حساب متجه الانحدارات $f(\beta(x_0)^{(0)}_{-X_i/\lambda})$ من الصيغة (17) ومصفوفة Hessian $H(\beta(x_0)^{(0)}_{-X_i/\lambda})$ في الصيغة (20) وان ($i = 1,2 \dots n$).

c) يتم حساب

$$\beta(x_0)^{(v+1)}_{-X_i/\lambda} = \beta(x_0)^{(v+1)}_{-X_i/\lambda} - \left[H(\beta(x_0)^{(0)}_{-X_i/\lambda})\right]^{-1} f(\beta(x_0)^{(0)}_{-X_i/\lambda})$$
$$for\ v = 1,2,..,n$$

d) أذا كان الحد الأقصى

$$|\beta(x_0)^{(v+1)}_{-X_i/\lambda} - \beta(x_0)^{(v)}_{-X_i/\lambda}| < \varepsilon$$

وان ε هو عدد حقيقي صغير موجب، ثم ننتقل الى خطوة (e)، لكن أذا لم يكن متقاربة نعود للخطوة (c)

e) يتم الحصول على المقدر $\hat{\beta}(x_0)_{-x_i/\lambda}$.

3. يتم حساب قيمة $g(X_i, \hat{\beta}(x_0)_{-x_i/\lambda})$ وان ($i = 1,2,..n$).
4. أحسب قيمة $CV_{ML}(\lambda)$ من الصيغة (11).
5. يتم التكرار من الخطوة (1) الى الخطوة (4) من خلال تكرار القيم الابتدائية لمعلمة التمهيد لكل متغير توضيحي
6. يتم الحصول على قيمة معلمة تمهيد المثلى بناءً على القيمة العظمى لـ $CV_{ML}(\lambda)$
7. تقدير متجه المعلمة$\beta(x_0)$ في الصيغة (15) باستعمال خوارزمية نيوتن– رافسون بناءً على أقل قيمة مثلى لمعلمة التمهيد (λ).

يتم حساب دقة التصنيف لأنموذج الانحدار اللوجستي الثنائي اللامعلمي.

7-3 معيار المقارنة(Comparison standard) :

يتم المقارنة بين طرائق التقدير من خلال متوسط مربعات الخطأ (MSE) للأنموذج .

$$\text{MSE} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-k-1} \quad \dots\dots\dots (21)$$

4- الجانب العملي :

إن أسلوب جمع البيانات تم عن طريق الاعتماد على طبلات المرضى (المصابين وغير المصابين) والتي تضمنت دراسة حول مرض (سرطان الدم الليمفاوي المزمن) في مستشفى الديوانية التعليمي في محافظة القادسية (وحدة استشارية الأورام) وقد شملت الدراسة 100 مشاهدة من الذكور والإناث (المصابين وغير المصابين) وقد قسمت المشاهدات إلى مجموعتين كالأتي:

1. المجموعة الأولى: قد شملت الأشخاص المصابين بالمرض بحجم (53) عينة ورمزنا لهم بالرمز(0).
2. المجموعة الثانية: قد شملت الأشخاص غير المصابين بالمرض بحجم(47) عينة ورمزنا لهم بالرمز(1).

1-4 العوامل المؤثرة في المرض :

هنالك العديد من العوامل المؤثرة في الحد من الاصابة بأمراض سرطان الدم الليمفاوي المزمن ذكرت اغلب الدراسات الطبية التي تناولت مرض سرطان الدم الليمفاوي المزمن عدة عوامل تأثر على الإصابة بالمرض فضلا عن المتغيرات التي تم اقتراحها وهي كالأتي :

Y (متغير الاستجابة) : وهو متغير معتمد يأخذ قيمتين فقط هما (0) إذا كان الشخص مصاب بالمرض، (1) إذا كان الشخص غير المصاب بالمرض .

اما المتغيرات التوضيحية فهي :

X_1 : يمثل نوع الجنس (Sex) حيث يرمز (1) للذكور و (2) للإناث.

X_2 : يمثل العمر (age) .
X_3 : يمثل خلايا الدم البيضاء WBC(White Blood Cells).
X_4 : يمثل نسبة هيموجلوبين الدم HGB (Hemoglobin Blood).
X_5 : يمثل نسبة الهيماتوكريت HCT (Haematocrit)
X_6 : يمثل نسبة الصفائح الدموية PLT (Blood Platelets) .

2-4 بناء الانموذج :

من اهم مراحل بناء الانموذج هو تحديد المتغير الاستجابة (Dependent Variable) والمتغيرات التوضيحية (Variable Exploratory) بناءً على طبيعة ونوع البيانات قيد الدراسة والتي سيتم توضيحها كما في ادناه :

المتغير الاستجابة y (يمثل معدلات الاصابة بمرض سرطان الدم اللمفاوي من عدمها) : وكون التغير يمثل الاصابة من عدمها لذلك سيكون Y=0 ويمثل الاصابة بالمرض و Y=1 يمثل غير المصابين بالمرض.

المتغيرات التوضيحي xi تم توضيحها في (4-2) .

3-4 تحليل النتائج (the Results Of Analysis)

قبل البدء بعملية التقدير تم اجراء اختبار (VIF) لاختبار وجود مشكلة ارتباط خطي بين المتغيرات المستقلة اذ تبين انه لا توجد علاقة ارتباط خطية بيت المتغيرات المستقلة اذ ان جميع متغيرات التوضيحية كانت قيمة VIF اقل من 5 وكما مبين في الجدول ادناه :

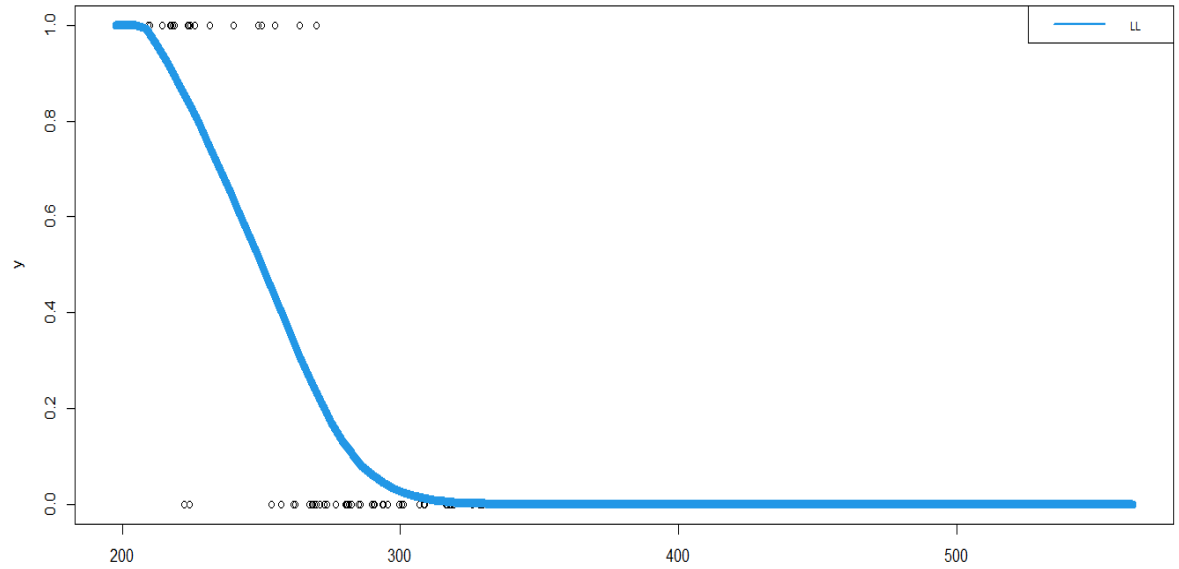
جدول رقم (1) يبين عامل التضخم التباين للبيانات الحقيقة التي تمثل الاصابة بسرطان الدم اللمفاوي

المتغير	SEX	Age	WBC	HGB	HCT	PLT
VIF	1.164	1.484	1.128	4.57	4.11	1.049

جدول رقم (2) يبين قيم متوسطات مربعات الخطأ (MSE) لمعدل الاصابة سرطان الدم اللمفاوي

Estimation Method d	LL L
MSE E	0.3210946

تم احتساب قيم متوسط مربعات الخطأ (**MSE**) حيث بين النتائج ان طريقة **LL** هي افضل الطرائق اللامعلمية .



شكل رقم (1) يوضح البيانات الحقيقية والتقديرية للمتغير المعتمد y (معدل الاصابة)لأفضل طرائق التقدير

5- التوصيات والمقترحات :

1. استعمال طرائق تقدير لامعلمية اخرى وبعرض حزمة اخرى لبيان امكانية زيادة كفاءة الانموذج المستعمل .

2. توظيف الطرائق المشار اليها بالأفضلية في حال انموذج الانحدار اللوجستي رتبي وبيان مدى ملائمتها لذلك .
3. اهتمام الجهات الصحية المختصة بتوفير بيانات دقيقة حول المرض من خلال اعداد صفحة خاصة لكل مريض تتضمن تاريخه الطبي متضمنا كافة المعلومات التي من شئنها دراسة المرض بشكل اكثر دقة وتحدي العوامل التي من المحتمل ان تودي الى حدوث المرض .
4. توظيف الطرائق المستخدم والانموذج المستعمل لدراسة ظواهر صحية اخرى لما له من كفاءة في عملية دراسة احتمالية الاصابة والعوامل المؤثرة في المرض
5. بضرورة العمل الجاد على نشر ثقافة العناية بإجراء الفحص الدوري (ألمختبري) للكشف المبكر عن إصابة أو عدم إصابة الشخص بمرض ما عن طريق وسائل الإعلام المختلفة مثل التلفاز، الراديو، الانترنيت أو إقامة الندوات.

6-الخاتمة:

من خلال استعمال خوارزمية لوجت الإمكان الموضعي وهي من الطرائق اللامعلمية وتم تطبيقها باستعمال الانحدار اللوجستي اللامعلمي الثنائي وكذلك استعمال دالة Gaussian ومعلمة تمهيد مثلى وبالاعتماد على معيار المقارنة MSE نلاحظ ان خوارزمية لوجت الإمكان الموضعي عند احجام العينات الثلاثة (50,100,150) وقيمة تباين 0.1 كانت هي الأفضل .

7- المصادر:

1. العزاوي، احمد ذياب احمد، (2005) ، "المقارنة بين بعض طرائق تقدير انموذج انحدار اللوجستك والطرائق الحصينة للتجارب الحياتية ذات الاستجابة الثنائية باستخدام اسلوب المحاكاة" رسالة ماجستير، كلية الادارة والاقتصاد ،جامعة بغداد.
2. شهاب ، ضمياء حامد، (2017) ،"مقارنة بعض طرائق التقدير الحصينة مع اسلوب بيز في تقدير دالة الانحدار اللوجستي مع تطبيق عملي" رسالة ماجستير ،كلية الادارة والاقتصاد، جامعة بغداد
3. صالح، عائدة هادي، (2011)، "تحليل الانحدار اللوجستي لدراسة زمن البقاء لمرضى سرطان الدم"، رسالة ماجستير، قسم الاحصاء، كلية الادارة والاقتصاد، جامعة المستنصرية.
4. عباس، مؤيد سلمان، (2017)، "استخدام المحاكاة لتقدير انموذج الانحدار اللامعلمي الضبابي"، اطروحة دكتوراه، قسم الاحصاء، كلية الادارة والاقتصاد، جامعة بغداد.
5. Czepiel, S. A. (2002).Maximum likelihood estimation of logistic regression models: theory and implementation. Available at czep. net/stat/mlelr. Pdf.
6. Liu, H., 2008. Generalized additive model. Department of Mathematics and Statistics University of Minnesota Duluth: Duluth, MN, USA, 55812.
7. Shim, J. Y., and Seok, k. H.(2012) "Semiparametric kernel logistic regression with longitudinal datac".Journal of the Korean Data and information Science Society, 23 (2) ,385-392.
8. Suliyanto, Rifada, M., & Tjahjono, E. (2020, September). Estimation of nonparametric binary logistic regression model with local likelihood logit estimation method (case study of diabetes mellitus patients at Surabaya Hajj General Hospital). In AIP Conference Proceedings (Vol. 2264, No. 1, p. 030007). AIP Publishing LLC.

Using the Logit algorithm to estimate two-response logistic regression

Narjes Bassem Khalaf [1]
University of Baghdad, College of Administration and Economics, Baghdad, Iraq
narjes.bassem1201a@coadec.uobaghdad.edu.iq

Lekaa Ali Mohammed[2]
University of Baghdad, College of Administration and Economics, Baghdad, Iraq
lekaa.a@coadec.uobaghdad.edu.iq

Abstract:
In this research, the methods of the kernel estimator were relied on in estimating the Binary Local Likelihood Logit Estimation, where the logit positional possibility algorithm was used, an optimal smoothing parameter (ג) was estimated by two methods of the Cross Validation and

(1)
(2)

the Generalized Cross Validation, and that the smoothing parameter has a clear effect on the process Estimation also has a fundamental role in approximating and smoothing the curve and bringing it closer to the real curve.The aim of using kernel estimators is to modify the observations so that we can obtain estimators with characteristics close to the properties of the real parameters. Depending on medical data, data has been collected on patients with chronic lymphocytic leukaemia (representing infected and non-infected) and through the use of the (Gaussian) function and based on the comparison standard (MSE), it was found that method of the Local Likelihood Logit algorithm (Logit) is the best at sample sizes (50,100,150) and at a variance value (0.1) because it obtains The lowest value for this criterion.
Keywords: The Binary logistic regression model, the Local Likelihood Logit algorithm, the Cross Validation and the Generalized Cross Validation, patients with chronic lymphocytic leukaemia.