

استخدام أسلوب البوتستراب في تقدير أنموذج الانحدار اللوجستي بطريقة الإمكان الأعظم - بحث تطبيقي

مصطفى نصيف جاسم**

أ.د. حمزة إسماعيل شاهين*

المستخلص

يتناول هذا البحث اجراء تطبيق عملي على بيانات حقيقية تخص العملية الجراحية لورم الدماغ ولعينة حجمها (130) مريض لعام (2017) لتقدير أنموذج الانحدار اللوجستي الثنائي الاعتيادي وباستخدام أسلوب البوتستراب وفق طريقة تقدير الإمكان الأعظم إذ تمت المقارنة بين أنموذج الانحدار اللوجستي الاعتيادي وباستخدام أسلوب البوتستراب بالاعتماد على معايير المقارنة متوسط مربعات الخطأ (Mean Square Error (MSE) ومتوسط الخطأ المطلق (Mean Absolute Error (MAE) وتم الحصول على النتائج باستخدام برنامج (Xplore) وتبين افضلية أسلوب البوتستراب في تقدير الأنموذج.

الكلمات المفتاحية : أنموذج الانحدار اللوجستي الثنائي ، أسلوب البوتستراب

The use of the bootstrap method in estimating the logistic regression model in the greatest possible way: Applied research

Abstract

This study deals with the practical application of real data on the operation of brain tumor and a sample size (130) patients (2017) to estimate the model of the logistic regression binary and using the method of bootstrap according to the method of estimating the greatest possible, as was compared between the standard logistic regression model and using the method of bootstrap based On the comparison criteria, Mean Square Error (MSE) and Mean Absolute Error (MAE) were obtained using the Xplore program.

Keywords: Binary Logistic Regression Model, Bootstrap Method.

1 . المقدمة ومشكلة وهدف البحث

يعد أسلوب البوتستراب (Bootstrap procedure) أحد التقانات الإحصائية المفيدة لما له أهمية كبيرة للحصول على تقديرات أكثر دقة حيث يعتمد عليه للاستدلال لأخذ العينات من خلال سحب عينات متكررة جزئية مع الارجاع مستمدة من العينة الاصلية للمجتمع بدلاً من استخدام بيانات المجتمع كله أي انه يعتمد على أسلوب المعاينة مع الارجاع، وهو أسلوب حسابي يستعمل لتحديد قياس الدقة لتقدير الاحصاءة ومن هنا تبرز اهمية استعماله من اجل الحصول على تقديرات أكثر دقة ومرونة، عند محاولة الباحثين استخدام تحليل الانحدار الخطي (البسيط او المتعدد) لتوفيق البيانات ذات المتغير التابع الوصفي فإن هذا يشكل تحدياً كبيراً بالنسبة للباحثين نتيجة لذلك جعل المختصين في ميدان النمذجة الإحصائية يبحثون عن بدائل للانحدار الخطي (البسيط او المتعدد) ومن اشهر هذه البدائل هي تقانة الانحدار اللوجستي (Logistic Regression Technique)، كذلك عند تقدير أنموذج الانحدار اللوجستي فإن طريقة المربعات الصغرى (Least Square Method) تكون غير مناسبة والسبب في ذلك عدم تحقق الفروض الخاصة بها وفي هذه الحالة لابد من البحث عن طرائق تقدير مناسبة لتقدير أنموذج الانحدار اللوجستي ومن هذه الطرائق طريقة تقدير الإمكان الأعظم.

يهدف البحث الى استخدام طريقة تقدير الإمكان الأعظم في تقدير أنموذج الانحدار اللوجستي الثنائي الاعتيادي وباستخدام أسلوب البوتستراب واجراء المقارنة بينهما من خلال معايير المقارنة متوسط مربعات الخطأ (MSE) ومتوسط الخطأ المطلق (MAE) بالاعتماد على بيانات تخص العملية الجراحية لورم الدماغ لعينة من المرضى بحجم (130) مريض.

2. أنموذج الانحدار اللوجستي ثنائي الاستجابة: (6)، (4)، (3)، (1)

Logistic Regression Model (Binary Response)

إن أنموذج الانحدار اللوجستي يقوم على فرض أساسي حيث ينص هذا الفرض على إن المتغير الذي نهتم بدراسته وهو متغير الاستجابة (Response Variable) (y) يكون متغيراً منفصلاً وهذا المتغير يتوزع توزيع بيرنولي (Bernoulli) حيث يأخذ متغير الاستجابة القيمة (1) في حالة النجاح (Success) باحتمال مقداره (π) ويأخذ القيمة (0) في حالة الفشل (Failure) باحتمال مقداره (1-π) أي إن

π : هو احتمال حدوث الاستجابة عندما (y=1)

1-π : هو احتمال عدم حدوث الاستجابة عندما (y=0)

إن المتغيرات التوضيحية وكذلك متغير الاستجابة في الانحدار الخطي البسيط تأخذ قيماً مستمرة لذلك فإن الأنموذج الآتي هو الذي يربط بين هذه المتغيرات:

$$Y = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i=1,2,\dots,p \quad \dots\dots\dots(1)$$

إذ إن (y) هو المتغير المعتمد الذي يكون متغيراً مستمراً (Continuous Variable) وعلى فرض إن متوسط قيم المتغير (y) الفعلية أو الملاحظة عند قيمة معينة من قيم المتغير التوضيحي (x) تمثل E(y) لذلك فإن الأنموذج يمكن كتابته كالآتي:

$$E(y|x) = \beta_0 + \beta_1 x_i, \quad i=1,2,\dots,p \quad \dots\dots\dots(2)$$

من المعروف إن الطرف الأيمن لأنموذج الانحدار الخطي أعلاه يأخذ قيماً من سالب ما لانهاية (-∞) الى موجب ما لانهاية (∞) فإن استخدام هذا الأنموذج مع البيانات الوصفية لا يكون ملائماً والسبب في ذلك لأن:

$$E(y|x) = p(y=1) = \pi \quad \dots\dots\dots(3)$$

لذلك فإن قيمة الطرف الأيمن تكون محصورة بين (1,0) وعليه فإن أنموذج الانحدار الخطي من خلال وجهة نظر الانحدار يكون غير قابل للتطبيق لأن القيم المتوقعة لمتغير الاستجابة E(y|x) تتجاوز الرقمين (1,0) لذلك في هذه الحالة نستخدم الانحدار اللوجستي الذي يأخذ شكل حرف الـ (S) ويتضح من الشكل أن العلاقة بين المتغيرين (y,x) غير خطية ولصعوبة استخدام النماذج اللاخطية لوصف وتحليل البيانات هذا ما دفع الباحثون الى تحويل أنموذج الانحدار اللوجستي الى الشكل الخطي هذا ما اقترحه الباحث (Berkson) عام 1944 (2) من خلال ادخال تحويل رياضي تعرف بتحويل اللوجيت (Logit Transformation) على المتغير المعتمد (y)، من المعروف إن قيمة π محصورة بين (1,0) وبعد ذلك النسبة π/(1-π) وهي محصورة بين (0, ∞) هذا يعني انها مقدار موجب حيث نقوم بأخذ اللوغاريتم الطبيعي (Normal Logarithm) للأساس (e) للنسبة π/(1-π) أي إن $(-\infty \leq \log_e \left(\frac{\pi}{1-\pi}\right) \leq \infty)$ هذا يعني مجال قيمة تكون محصور بين (-∞, ∞)، وعليه فإن أنموذج الانحدار اللوجستي البسيط ثنائي الاستجابة (Binary) الخطي في حالة متغير توضيحي واحد ويمكن كتابته كالآتي:

$$\log_e \left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x \quad \dots\dots\dots(4)$$

إن أنموذج الانحدار اللوجستي الخطي ثنائي الاستجابة (Binary) في حالة أكثر من متغير واحد من المتغيرات التوضيحية فيمكن كتابته كالآتي:

$$\log_e \left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots\dots\dots + \beta_k X_p \quad \dots\dots\dots(5)$$

إذ إن:

(p) عدد المتغيرات التوضيحية.

(k) عدد المعلمات.

$X_1, X_2, \dots, X_p$: تمثل المتغيرات التوضيحية.

$\beta_0, \beta_1, \dots, \beta_k$: معلمات النموذج التي يجب تقديرها.

$(\pi/(1-\pi))$: تسمى نسبة الأفضلية أو افضلية النجاح (odds of success) وكذلك يمكن إن تسمى افضلية الفشل (odds of failure).

$\log_e (\pi/(1-\pi))$: يسمى اللوجيت (Logit) أو لوغاريتم نسبة الأفضلية (odds ratio log).

وباختصار فإن أنموذج الانحدار اللوجستي هو ببساطة تحويلة لوغاريتمية للانحدار الخطي إلا أننا في الانحدار اللوجستي يكون هدفنا ليس تفسير التغير في قيم متغير الاستجابة وإنما تفسير احتمال حدوث وعدم حدوث الظاهرة محل الدراسة.

من مميزات أنموذج الانحدار اللوجستي إن الافتراضات الآتية لا يشترط توافرها:

1. توزيع المتغيرات إن يكون توزيعاً طبيعياً (Normal Distribution).

2. وجود علاقة خطية ما بين المتغير المعتمد والمتغيرات المفسرة.

3. تجانس التباين (Homoscedasticity) (تحقق خاصية ثبات التباين).

3. طريقة تقدير الإمكان الأعظم: (7) (10)

Maximum Likelihood Estimation Method

من أجل تقدير معلمات أنموذج الانحدار اللوجستي يتم اللجوء الى طريقة تقدير الإمكان الأعظم وهي من الطرائق المعلمية الأكثر شيوعاً في تقدير معلمات أنموذج الانحدار اللوجستي ($\hat{\beta}$) التي تعظم دالة الإمكان وتجعلها أعظم ما يمكن وطريقة الإمكان الأعظم هي طريقة تكرارية يتم فيها تكرار العمليات الحسابية أكثر من مرة (عدة مرات) وبذلك يتم الحصول على أفضل قيم لتقدير المعلمات.

وعندما (y) يتبع توزيع ثنائي الحدين فإن دالة الكثافة الاحتمالية لـ (y) هي:

$$P(y = y_i) = C_{y_i}^{n_i} \pi_i^{y_i} (1-\pi_i)^{n_i-y_i} \dots\dots\dots (6)$$

إذ إن:

$$i=1,2,\dots,n$$

إن دالة الكثافة الاحتمالية لـ (y) لدالة الإمكان هي:

$$L(\beta, y) = \prod_{i=1}^n C_{y_i}^{n_i} \pi_i^{y_i} (1-\pi_i)^{n_i-y_i} \dots\dots\dots (7)$$

وبتبسيط المعادلة (7) نحصل على:

$$L(\beta, y) = \prod_{i=1}^n C_{y_i}^{n_i} \left[\frac{\pi_i}{1-\pi_i} \right]^{y_i} (1-\pi_i)^{n_i} \dots\dots\dots (8)$$

وبأخذ اللوغاريتم للمعادلة (8) نحصل على لوغاريتم دالة الإمكان وكالاتي:

$$\text{Ln}L(\beta, y) = \sum_{i=1}^n [\ln C_{y_i}^{n_i} + y_i \text{Ln} \left(\frac{\pi_i}{1-\pi_i} \right) + n_i \text{Ln}(1-\pi_i)] \dots\dots\dots (9)$$

إذ إن (π_i) هو احتمال النجاح وتكون صيغته على النحو الآتي:

$$\pi_i = \frac{e^{\beta_0 + \sum \beta_j X_{ji}}}{1 + e^{\beta_0 + \sum \beta_j X_{ji}}} \dots\dots\dots (10)$$

وعند تعويض المعادلة (10) في المعادلة (9) نحصل على الآتي:

$$x_1 + \beta_2 x_2 + \dots + \beta_k x_p) \quad {}_1 \text{LnL}(\beta, y) = \sum_{i=1}^n [\text{Ln } C_{yi}^{ni} + y_i (\beta_0 + \beta_1 \text{Ln}(\frac{1}{1 + e^{\beta_0 + \sum \beta_j x_{ji}}}))] \quad (11)$$

إن تقديرات الإمكان الأعظم هي قيم المعلمات المقدرة ($\hat{\beta}$) التي تعظم دالة الإمكان ولإيجاد تقديرات الإمكان الأعظم نقوم بأشتقاق المعادلة (11) للمعلمة المراد تقديرها وبعدها نقوم بمساواتها للصفر لنحصل على (k+1) من المعادلات حيث نستخدم طريقة نيوتن-رافسون (Newton-Raphson) لحلها.

Bootstrap procedure

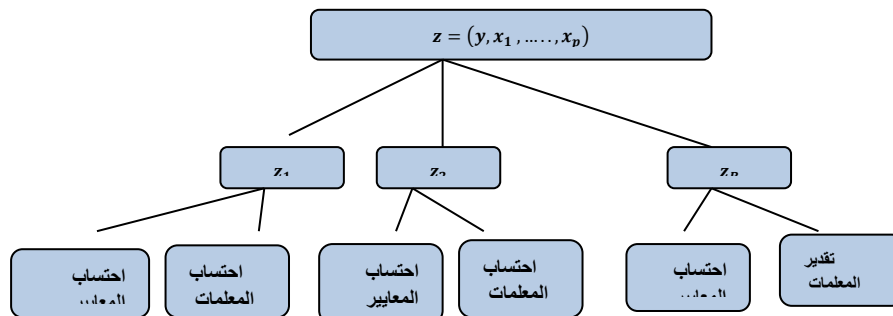
4. أسلوب البوتستراب: (5)، (6)

طور العالم (Efron) عام (1977) طريقة جديدة لإعادة المعاينة بالإرجاع سميت بأسلوب البوتستراب (Bootstrap procedure) وهي إعادة عينة تشتمل على (b) من العناصر المسحوبة بالإرجاع بشكل عشوائي من (n) من البيانات الأصلية، وإن أسلوب البوتستراب يزودنا بتقديرات للخطأ المعياري والتحيز عن طريق (b) من العينات المسحوبة المأخوذة بالإرجاع من العينة الأصلية.

الخطوات المتبعة في استعمال أسلوب البوتستراب وهي على النحو الآتي:-

- 1- من العينة الأصلية $[Z = (y, x_1, x_2, \dots, x_p)]$ نسحب عينة عشوائية بالإرجاع وبحجم معين نرمز له (b) $[Z^* = (y^*, x_1^*, x_2^*, \dots, x_p^*)]$ بدلاً من استخدام بيانات المجتمع كله كعينة بوتسترابية.
- 2- حساب تقدير أنموذج الانحدار اللوجستي بطريقة الإمكان الأعظم للعينة المسحوبة وكذلك حساب معايير المقارنة للطريقة لنفس العينة.
- 3- نكرر الخطوات (1) و (2) — (B) من المرات إذ إن $(B \geq 1000)$ ومن ثم يتم حساب المتوسط العام للحصول على تقدير الأنموذج البوتسترابي للطريقة.

ويمكن توضيح الخطوات حسب المخطط الآتي:



Chi Square Test

5. اختبار مربع كاي: (2)، (9)

يعد اختبار مربع كاي أحد الاختبارات المهمة المستعملة لمعرفة معنوية الأنموذج وذلك من خلال اختبار الفرضيتين الآتيتين وعلى النحو الآتي:

فرضية العدم (H_0): - لا يوجد تأثير للمتغيرات المستقلة على المتغير المعتمد (الأنموذج غير معنوي).

الفرضية البديلة (H_1): - يوجد تأثير للمتغيرات المستقلة على المتغير المعتمد (الأنموذج معنوي).

كذلك يوضح لنا هذا الاختبار كيف يصف الأنموذج متغير الاستجابة، والذي يساوي مجموع مربعات الفروق بين القيم المشاهدة والمتوقعة مقسوماً على القيم المتوقعة، وتقدير معنوية الأنموذج يتضمن تقارب القيم المتوقعة من القيم المشاهدة وعلى النحو الآتي:

$$= \sum_{i=1}^N \frac{(O_i - E_i)^2}{E_i} \dots \dots \dots (12) \chi^2$$

إذ ان (O_i) تعني القيم المشاهدة و (E_i) تعني القيم المتوقعة، وتقارن قيمة إحصاء الاختبار مع قيمة الجدولية لمربع كاي (χ^2) بدرجة حرية تساوي عدد المتغيرات التوضيحية ولمستوى معنوية (α) .

6. اختبار (Wald): (2) (9)

من الاختبارات المهمة التي تستعمل لبيان أهمية ومعنوية تأثير المتغير المستقل على المتغير المعتمد إذ يمكن استخدام الاختبار للعديد من النماذج المختلفة بما فيها النماذج مع المتغيرات الثنائية أو المتغيرات المستمرة. وفي حالة أنموذج (Logit) يتم استخدام إحصاء (Wald) لغرض احتساب معنوية المعلمات المقدرة لكل متغير توضيحي من أجل اختبار الفرضيتين الآتيتين:

H_0 : إن تأثير معاملات الانحدار اللوجستي الثنائي المرتبط بالمتغير المستقل تساوي صفراً.

H_1 : إن تأثير معاملات الانحدار اللوجستي الثنائي المرتبط بالمتغير المستقل لا تساوي صفراً.

إن صيغة إحصاء (Wald) تكون على النحو الآتي:

$$Wald = \left(\frac{\hat{\beta}}{SE(\hat{\beta})} \right)^2 \dots\dots\dots (13)$$

إذ إن:

$\hat{\beta}$: قيمة معامل الانحدار اللوجستي المقدر للمتغير التوضيحي (x).

$SE(\hat{\beta})$: قيمة الخطأ المعياري لمعامل الانحدار اللوجستي المقدر للمتغير التوضيحي (x).

إن اختبار (Wald) هو اختبار من طرفين ويجب إن تكون معنوية المعلمات المناظرة لقبول أو رفض فرضية العدم باستعمال الاحتمالات أقل من (0.05) لكي يتم رفض فرضية العدم (H_0) وقبول الفرضية البديلة (H_1).

7. تقييم القوة التفسيرية للأنموذج: (2) (9)

من المعلوم أن معامل التحديد R^2 هو مقياس يوضح نسبة التغير في المتغير المعتمد (y) الذي سببها التغير في المتغير المستقل (x) و في أنموذج الانحدار اللوجستي فإن هنالك العديد من المقاييس لحساب معامل التحديد R^2 اقترحت لقياس حسن المطابقة وتسمى أحياناً شبه معامل التحديد ومنها:-

1 – مؤشر *Cox and Snell* :-

إن صيغة $R^2_{Cox \& snell}$ تكون على النحو الآتي:

$$R^2_{Cox \& snell} = 1 - \left[\frac{L_0}{L_1} \right]^{2/n} \dots\dots\dots (14)$$

إذ إن:

L_1 : تمثل قيمة الامكان الاعظم لأنموذج الانحدار اللوجستي الثنائي عندما يحتوي الأنموذج على كل المتغيرات

L_0 : تمثل قيمة الامكان الاعظم لأنموذج الانحدار اللوجستي الثنائي عندما يحتوي الأنموذج على الحد الثابت فقط.

n: تمثل حجم المجتمع.

2 – مؤشر *Nagelkerke* :

ويعد هذا المؤشر مصححاً أو معدلاً لـ $(R^2_{Cox \& snell})$ وهو يعد مقياساً لحسن المطابقة العامة وصيغته:

$$R^2_{Nagelkerke} = R^2_{Cox \& snell} / R^2_z \dots\dots\dots (15)$$

إذ إن:

$$R^2_z = 1 - (L_0)^{2/n} \dots\dots\dots (16)$$

إن مؤشر Nagelkerke يكون دائماً أكبر من $R^2_{Cox \& Snell}$, لأنه في الحقيقة مصحح له وإن أعلى قيمة لهما هي الواحد.

وبشكل عام فإن قيمة R^2 في الانحدار اللوجستي تعد مقياساً للنماذج الجيدة، وتكون قيمها أقل من R^2 في النماذج الخطية المقدرية بطريقة المربعات الصغرى الاعتيادية وذلك لأنها تعتمد على المدى وتوزيع المتغيرات المستقلة وإيضاً لأن المتغيرات الثنائية تميل إلى أن يكون R^2 منخفضاً حتى في علاقة الانحدار المثالية.

8. متوسط مربعات الخطأ: (2) (9) Mean Square Error (MSE)

وهو من المعايير المهمة التي تستخدم للمقارنة بين النماذج وطرائق التقدير ويكون النموذج أو الطريقة المثلى الذي يعطي أقل قيمة لمتوسط مربع الخطأ (MSE) وتختلف صيغته من طريقة إلى أخرى وعلى النحو الآتي:

1- صيغة معيار متوسط مربع الخطأ (MSE) حسب الطريقة المعلمية تكون كالآتي

$$MSE = \frac{1}{n-K} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad \dots\dots\dots (17)$$

2- صيغة معيار متوسط مربعات الخطأ (MSE) حسب الطريقة اللامعلمية تكون كالآتي

$$MSE = \frac{1}{n-P} \sum_{i=1}^n (y_i - \hat{m}_{NW}(x))^2 \quad \dots\dots\dots (18)$$

إذ إن:

K : تمثل عدد معاملات النموذج.

n : تمثل حجم المجتمع.

P : تمثل عدد المتغيرات التوضيحية في النموذج.

تمثل دالة الانحدار المعلمي التقديرية بطريقة الإمكان الأعظم $\hat{y}_i$:

تمثل دالة الانحدار اللامعلمي التقديرية بطريقة مقدر النواة $\hat{m}_{NW}(x)$:

9. متوسط الخطأ المطلق: (2) (9) Mean Absolute Error (MAE)

إن معيار متوسط الخطأ المطلق هو مشابه لمعيار متوسط مربع الخطأ من حيث اختيار النموذج أو الطريقة المثلى التي تعطي أقل قيمة لمتوسط الخطأ المطلق (MAE)، أما صيغته الرياضية تختلف من طريقة إلى أخرى وعلى النحو الآتي:

1- صيغة متوسط الخطأ المطلق (MAE) حسب الطريقة المعلمية تكون كالآتي

$$MAE = \frac{1}{n-K} \sum_{i=1}^n |y_i - \hat{y}_i| \quad \dots\dots\dots (19)$$

2- صيغة متوسط الخطأ المطلق (MAE) حسب الطريقة اللامعلمية تكون كالآتي

$$MAE = \frac{1}{n-P} \sum_{i=1}^n |y_i - \hat{m}_{NW}(x)| \quad \dots\dots\dots (20)$$

11. الجانب التطبيقي:

11-1 وصف العينة والمتغيرات المستخدمة

تم الحصول على بيانات الدراسة من وزارة الصحة / مركز السرطان شعبة الإحصاء لعام (2017) حيث تم اختيار عينة حجمها (130) مريض تم إجراء عمليات جراحية لهم خاصة بورم الدماغ وبشكل عشوائي من أصل (376) عملية جراحية لعام (2017) وذلك من خلال الاستمارة الإحصائية للأورام السرطانية.

y_i : المتغير المعتمد ويمثل العملية الجراحية الخاصة بورم الدماغ ومصنف إلى:

- (1) نجاح العملية الجراحية، (0) فشل العملية الجراحية.
 X_1 : يمثل حجم الورم ومصنف الى:
 (1) حجم الورم متقدم، (2) حجم الورم غير متقدم.
 X_2 : يمثل درجة الورم ومصنف الى:
 (1) ورم من الدرجة الأولى، (2) ورم من الدرجة الثانية، (3) ورم من الدرجة الثالثة، (4) ورم من الدرجة الرابعة، (5) ورم متقدم الدرجة غير معروف الدرجة.
 X_3 : يمثل مدى انتشار الورم ومصنف الى:
 1- (1) في الموقع الأصلي، (2) متمركز في موضعه، (3) امتداد محلي، (4) امتداد محلي للعقد اللمفاوية، (5) منتشر في موضعه، (6) انتشار بعيد، (7) غير قابل للانتشار، (8) انتشار كبير غير معروف المدى.
 X_4 : يمثل الجنس ومصنف الى: (1) ذكر، (2) انثى.

كما موضح في الجدول (1)

الجدول (1) يبين لنا الإحصاءات الوصفية كافة لبيانات البحث

المتغيرات	عدد الحالات	النسبة المئوية
العملية الجراحية (Y)	130	100 %
نجاح العملية (1)	95	73.1 %
فشل العملية (0)	35	26.9 %
حجم الورم (X_1)	130	100 %
متقدم (1)	46	35.4 %
غير متقدم (2)	84	64.6 %
درجة الورم (X_2)	130	100 %
من الدرجة الأولى (1)	25	19.2 %
من الدرجة الثانية (2)	43	33.1 %
من الدرجة الثالثة (3)	23	17.7 %
من الدرجة الرابعة (4)	8	6.2 %
متقدم الدرجة غير معروف (5)	31	23.8 %
مدى انتشار الورم (X_3)	130	100 %
في الموقع الأصلي (1)	12	9.2 %
متمركز في موضعه (2)	64	49.2 %
امتداد محلي (3)	33	25.4 %
امتداد محلي للعقد اللمفاوية (4)	3	2.3 %
منتشر في موضعه (5)	2	1.5 %
انتشار بعيد (6)	5	3.8 %
غير قابل للانتشار (7)	2	1.5 %
انتشار كبير غير معروف (8)	9	6.9 %
الجنس (X_4)	130	100 %
ذكر (1)	69	53.1 %
انثى (2)	61	46.9 %

2-11 اختبار مشكلة التعدد الخطي (8)، (3)

قبل البدء في تطبيق تحليل أنموذج الانحدار اللوجستي لابد من الكشف عن مشكلة التعدد الخطي للتأكد من وجودها أو عدم وجودها وذلك من أجل ضمان ملائمة البيانات لافتراضات تحليل الانحدار اللوجستي وذلك من خلال معامل تضخم التباين (Variance Inflation Factor) إذ إن صيغته تكون على النحو الآتي:

$$VIF = \frac{1}{Tolerance} \dots\dots\dots (22)$$

إذا أتضح إن قيمة معامل تضخم التباين أكبر من عشرة ($VIF > 10$) فهذا يدل على وجود هذه المشكلة وهذا يعني إن المتغيرات المستقلة تكون مرتبطة مع بعضها البعض ارتباطاً عالياً والعكس بالعكس وكما موضح في الجدول.

الجدول (2) حساب معامل تضخم التباين (VIF)

المتغيرات المستقلة	Tolerance	VIF
X ₁	0.791	1.264
X ₂	0.814	1.229
X ₃	0.709	1.411
X ₄	0.987	1.013

نلاحظ من نتائج الجدول (2) إن قيم اختبار معامل تضخم التباين (VIF) لجميع المتغيرات المستقلة تقل عن الـ (10) ويعد هذا مؤشراً على عدم وجود ارتباط عالي بين المتغيرات المستقلة.

3-11 اختبار معنوية النموذج

لمعرفة تأثير المتغيرات المستقلة في المتغير المعتمد يتم ذلك من خلال حساب معاملات الانحدار (β) ولمعرفة ذلك نفرض الفرضيتين الآتيتين وعلى النحو الآتي:

فرضية العدم (H_0): لا يوجد تأثير للمتغيرات المستقلة في المتغير المعتمد (النموذج غير مناسب)

الفرضية البديلة (H_1): يوجد تأثير للمتغيرات المستقلة في المتغير المعتمد (النموذج مناسب)

ولاختبار معنوية النموذج أي لاختبار الفرضيتين أعلاه نستخدم اختبار مربع كاي (Chi-Square)، كما موضح في الجدول (3) وعلى النحو الآتي:-

جدول (3) نتائج اختبار (Chi-Square) لمعنوية النموذج

Model	Chi-Square	df	Sig.
	104.626	4	.000

ومن خلال نتائج الجدول (3) نتضح معنوية الاختبار وذلك من خلال القيمة الاحتمالية عند درجة حرية (4) ومستوى معنوية (0.05) والتي تساوي ($P\text{-Value}=0.000$) وبما إن قيمتها أقل من مستوى المعنوية لذا نرفض فرضية العدم ونقبل الفرضية البديلة أي إن النموذج بالكامل معنوي ويمثل البيانات تمثيلاً جيداً.

4-11 اختبار ملائمة النموذج لبيانات الدراسة

ولمعرفة مدى ملائمة نموذج الانحدار اللوجستي الثنائي لبيانات الدراسة من خلال الجدول الآتي:-

جدول (4) ملائمة النموذج

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	46.822	0.553	0.803

ومن خلال نتائج الجدول (4) يتضح لنا الآتي:

1- ($R^2_{Cox \& Snell}$) التي تساوي (0.553) وهذا يعني إن (55.3%) من التغيرات الحاصلة بين نجاح العملية أو فشلها يمكن تفسيرها من خلال نموذج الانحدار اللوجستي الثنائي.

2- ($R^2_{Nagelkerke}$) التي تساوي (0.803) وهذا يعني إن (80.3%) من التغيرات الحاصلة بين نجاح العملية أو فشلها يمكن تفسيرها من خلال نموذج الانحدار اللوجستي الثنائي.

5-11 تقدير نموذج الانحدار اللوجستي الثنائي بالطريقة الاعتيادية حسب طريقة تقدير الإمكان الأعظم

لبيان تأثير المتغيرات المستقلة في المتغير المعتمد يتم ذلك من خلال حساب معاملات الانحدار (β) حيث تم تقدير معاملات نموذج الانحدار اللوجستي الثنائي حسب طريقة الإمكان الأعظم وكما موضح في الجدول (5) إذ تم الحصول على المعلمات المقدرة في التكرار السادس وعلى النحو الآتي:-

جدول (5) نتائج نموذج الانحدار اللوجستي حسب طريقة تقدير الإمكان الأعظم

المتغيرات التوضيحية	معاملات الانحدار اللوجستي β	الأخطاء المعيارية S.E	احصاءة Wald	درجات الحرية df	القيمة الاحتمالية Sig.	نسبة الافضلية $\text{Exp}(\beta)$
حجم الورم X_1	2.057	0.881	5.445	1	0.020	7.825
درجة الورم X_2	-1.441	0.326	18.77	1	0.000	0.244
مدى الانتشار X_3	-1.455	0.395	13.601	1	0.000	0.233
الجنس X_4	-1.93	0.925	4.357	1	0.037	0.145
الحد الثابت	9.73	2.9	11.258	1	0.001	16814.49

من خلا نتائج الجدول (5) يمكن كتابة معادلة نموذج الانحدار اللوجستي الثنائي وعلى النحو الآتي

$$\left(\frac{\pi}{1-\pi}\right) = 9.73 + 2.057X_1 - 1.441X_2 - 1.455X_3 - 1.93X_4 \quad \dots\dots\dots(23) \log_e$$

كذلك إن القيم الإحصائية الخاصة بأنموذج الانحدار اللوجستي الثنائي حسب طريقة الإمكان الأعظم تم توضيحها في الجدول (6) إذ تم حساب متوسط مربعات الخطأ (MSE) ومتوسط الخطأ المطلق (MAE) وعلى النحو الآتي:-

جدول (6) المعايير الإحصائية الخاصة بأنموذج الانحدار اللوجستي الثنائي حسب طريقة الإمكان الأعظم (MLE)

MAE	MSE
2.5815	8.0766

6-11 تقدير نموذج الانحدار اللوجستي حسب طريقة تقدير الإمكان الأعظم باستخدام أسلوب البوتستراپ

من أجل الحصول على نتائج تقدير نموذج الانحدار اللوجستي باستخدام أسلوب البوتستراپ تم سحب عينة عشوائية جزئية من العينة الأصلية وبحجم (50) مشاهدة وبالإرجاع للمتغير المعتمد والمتغيرات المستقلة وتقدير نموذج الانحدار اللوجستي الثنائي بطريقة تقدير الإمكان الأعظم وتكرار سحب العينات البوتستراپية (1000) مرة وحساب المتوسط العام للتقدير.

لبيان تأثير المتغيرات المستقلة على المتغير المعتمد لأنموذج الانحدار اللوجستي يتم ذلك من خلال حساب معاملات الانحدار (β) البوتستراپية حيث تم تقدير معاملات نموذج الانحدار اللوجستي الثنائي حسب طريقة الإمكان الأعظم باستخدام أسلوب البوتستراپ وكما موضح في الجدول (7) حيث تم الحصول على المعلمات المقدرة البوتستراپية في التكرار الثالث وعلى النحو الآتي:-

جدول (7) نتائج نموذج الانحدار اللوجستي حسب طريقة الإمكان الأعظم باستخدام أسلوب البوتستراپ (Bootstrap procedure)

المتغيرات التوضيحية	معاملات الانحدار اللوجستي β	الأخطاء المعيارية S.E	احصاءة Wald	درجات الحرية df	القيمة الاحتمالية Sig.	نسبة الافضلية $\text{Exp}(\beta)$
حجم الورم X_1	4.890	0.238	422.14	1	0.001	0.008
درجة الورم X_2	-0.047	0.058	0.656	1	0.003	0.954
مدى الانتشار X_3	-0.142	0.043	10.905	1	0.000	0.868
الجنس X_4	-8.150	0.018	205004.19	1	0.020	0.0002

3.975	0.0015	1	6.136	0.557	1.380	الحد الثابت
-------	--------	---	-------	-------	-------	-------------

من خلا نتائج الجدول (7) يمكن كتابة معادلة أنموذج الانحدار اللوجستي الثنائي باستخدام أسلوب البوتستراب وعلى النحو الآتي

$$\left(\frac{\pi}{1-\pi}\right) = 1.380 - 4.890X_1 - 0.047X_2 - 0.142X_3 - 8.15X_4 \dots\dots\dots(24) \log_e$$

كذلك إن القيم الإحصائية الخاصة بأنموذج الانحدار اللوجستي الثنائي حسب طريقة الإمكان الأعظم وباستخدام أسلوب البوتستراب (Bootstrap procedure) تم توضيحها في الجدول (8) إذ تم حساب لكل عينة بوتسترابية متوسط مربعات الخطأ (MSE) و متوسط الخطأ المطلق (MAE) وعلى النحو الآتي:-

جدول (8) المعايير الإحصائية الخاصة بأنموذج الانحدار اللوجستي حسب طريقة الإمكان الأعظم (MLE) باستخدام أسلوب البوتستراب

MSE	MAE
2.3899	1.4542

7-11 المقارنة بين تقدير أنموذج الانحدار اللوجستي الثنائي الاعتيادي وباستخدام أسلوب البوتستراب

الجدول (9) يبين لنا نتائج المقارنة لطريقة تقدير الإمكان الأعظم في تقدير أنموذج الانحدار اللوجستي الثنائي الاعتيادي وباستخدام أسلوب البوتستراب حسب المعايير الإحصائية للمقارنة (MSE ، MAE) لاختيار الأفضل ، وبملاحظة نتائج الجدول (9) تبين لنا إن أفضل أسلوب في تقدير أنموذج الانحدار اللوجستي ثنائي الاستجابة هو أسلوب البوتستراب لحصوله على أقل قيم لجميع المعايير الإحصائية للمقارنة (MSE ، MAE) جدول (9) المعايير الإحصائية للمقارنة بين الطريقتين في تقدير أنموذج الانحدار اللوجستي الثنائي الاعتيادي والبوتسترابي

المعيار المقارنة	باستخدام أسلوب البوتستراب	الطريقة الاعتيادية
	(MLE)	(MLE)
MSE	2.3899	8.0766
MAE	1.4542	2.5815

12- الاستنتاجات والتوصيات

1-12 الاستنتاجات

Conclusions

في ضوء تحليل النتائج والتطبيق العملي تم التوصل الى الاستنتاجات الآتية:-

1- بينت نتائج أنموذج الانحدار اللوجستي إن المتغيرات التوضيحية (حجم الورم، درجة الورم ومدى انتشار الورم) لهم تأثير معنوي على متغير الاستجابة (نجاح او فشل العملية الجراحية لورم الدماغ) إلا إن متغير الجنس ليس له تأثير معنوي على متغير الاستجابة.

2- تبين إن اعتماد أسلوب البوتستراب في تقدير أنموذج الانحدار اللوجستي ثنائي الاستجابة له اثر واضح من خلال معايير المفاضلة (MSE ، MAE) لمقدر الإمكان الأعظم.

4- على ضوء نتائج معايير المقارنة نستنتج إن استخدام أسلوب البوتستراب بطريقة الإمكان الأعظم في تقدير أنموذج الانحدار اللوجستي هي أفضل من الطريقة الاعتيادية في التقدير.

The Recommendations

2-12 التوصيات

- 1- بالنظر لتفوق أسلوب البوتستراب في تقدير أنموذج الانحدار اللوجستي نوصي بدراسة هذا الأسلوب لتقدير أنموذج الانحدار اللوجستي متعدد الاستجابة في الدراسات القادمة.
- 2- مقارنة أنموذج الانحدار اللوجستي باستخدام أسلوب البوتستراب مع أساليب إحصائية أخرى مثل تحليل النماذج اللوغاريتمية الخطية أو تحليل الدوال التمييزية.
- 3- إجراء دراسة مستقبلية تتناول طرائق تقدير أخرى كطريقة العزوم وطريقة تحليل الموجه الصغير Wavelet وطريقة انحدار الشرائح (Spline Regression method).
- 4- من الضروري تشخيص المرض عن طريق مراجعة المختصين لمرض ورم الدماغ وإجراء العملية الجراحية اللازمة في وقت مبكر، كذلك زيادة الوعي الصحي للأمراض المتعلقة بالأورام ومنها مرض ورم الدماغ عن طريق المراجعة المنتظمة للطبيب وإجراء التحاليل اللازمة.

المصادر

- 1- الجاعوني، فريد وغانم، عدنان، (2011)، "استخدام تقنية الانحدار اللوجستي ثنائي الاستجابة في دراسة أهم المحددات الاقتصادية والاجتماعية لكفاية دخل الأسرة دراسة تطبيقية على عينة عشوائية من الاسر في محافظة دمشق"، بحث منشور مجلة جامعة دمشق للعلوم الاقتصادية والقانونية، المجلد (27)، العدد (1).
- 2- النصراوي، نور عباس عمران، (2017) "استخدام أسلوب البوتستراب في تحليل النماذج المعلمية وشبه المعلمية والمقارنة بينهما" رسالة ماجستير مقدمة الى قسم الإحصاء، كلية الادارة والاقتصاد، جامعة كربلاء.
- 3- جبارة، أزهار كاظم، (2014) "تحليل البيانات متعددة الاستجابة لتشخيص أمراض العيون باستخدام الدالة التمييزية والانحدار اللوجستي (دراسة مقارنة"، رسالة ماجستير مقدمة إلى قسم الإحصاء، كلية الإدارة والاقتصاد، الجامعة المستنصرية.
- 4- عبدالرضا، علي يحيى، (2015) "استعمال أنموذج الانحدار اللوجستي لدراسة ظاهرة البطالة عند الشباب في محافظة بغداد"، بحث دبلوم عالي إلى قسم الاحصاء، كلية الإدارة والاقتصاد جامعة بغداد.
- Adjei, I. A., & Karim, R. (2016). An Application of Bootstrapping in Logistic 5 Regression Model. Open Access Library Journal, 3(09),1.
- 6- Fitrianto, A., & Cing, N. M. (2014). Empirical Distributions of Parameter Estimates in Binary Logistic Regression Using Bootstrap. International Journal of Mathematical Analysis, 8, 721-726.
- 7- Czepiel, S. A. (2002). Maximum likelihood estimation of logistic regression models: theory and implementation. Available at czep.net/stat/mlelr. Pdf.
- 8- Gujarati, Damodar N. (1988). "Basic Econometrics", McGraw-Hill Book Company, New York.
- 9- Jassim, Hussain N., And, Low, Heng Chin , And ,AL-Karkhi , Abbas F.MK (2007) "An Overview Of Evaluation Criteria In Logistic Regression Model ", International Conference on Mathematical Sciences, University Teknologi , Malaysia.
- 10- Liang, H., & Du, P. (2012). Maximum likelihood estimation in logistic regression models with a diverging number of covariates. Electronic Journal of Statistics, 6, 1838-1846

