



## استعمال اسلوب الجاكناييف لتقدير نموذج الانحدار اللوجستي لمرض سرطان الثدي

أ.د. محمد جاسم محمد  
جامعة بغداد /كلية الإدارة والاقتصاد /قسم  
الإحصاء  
[m.jasim@coadec.uobaghdad.edu.iq](mailto:m.jasim@coadec.uobaghdad.edu.iq)

أ.د. محمود مهدي البياتي  
جامعة بغداد /كلية الإدارة  
والاقتصاد /قسم الإحصاء

الباحث/ نبال جعفر عبد  
طالبة /جامعة بغداد /كلية الإدارة  
والاقتصاد /قسم الإحصاء  
nabaajafer204@gmail.com

Received: 21/10/2020 Accepted :22/11/2020 Published :FEBRUARY / 2021

هذا العمل مرخص تحت اتفاقية المشاع الابداعي نسب المُصنّف - غير تجاري - الترخيص العمومي الدولي 4.0

[Attribution-NonCommercial 4.0 International \(CC BY-NC 4.0\)](https://creativecommons.org/licenses/by-nc-sa/4.0/)



### مستخلص البحث:

يعتبر أنموذج الانحدار اللوجستي من اقدم النماذج الانحدار واكثرها شيوعا ، وكما يعرف على انه واحد من الاساليب الاحصائية التي تستعمل لوصف وتقدير العلاقة بين متغير عشوائي تابع (Y) ومتغيرات عشوائية توضيحية (X) ، والذي لا يفترض افتراضات الانحدار الخطي وهما افتراضات رئيسيان الاول التوزيع الطبيعي والثاني تجانس التباين ، وكما ويمكن يكون فيها المتغير التابع ثنائي الاستجابة يأخذ قيمتين ( واحد عند وقوع حدث معين و صفر عند عدم وقوع ذلك الحدث ) مثل (مصائب وغير مصاب ، متزوج وغير متزوج ) وان كثرة المتغيرات التوضيحية ادى الى ظهور مشكلة التعدد الخطي التي تجعل التقديرات غير دقيقة ، وتم استخدام طريقة الامكان الاعظم وطريقة انحدار الحرف في تقدير أنموذج الانحدار اللوجستي ثنائي الاستجابة باعتماد أسلوب الجاكناييف واجراء المقارنة بين المقدرات وفق معيار المعلومات (AIC) .

وتم تطبيق اسلوب الجاكناييف والمقدرات الاحصائية المذكورة لدراسة العلاقة بين متغير الاستجابة (اصابة وعدم الاصابة بسرطان الثدي) لعينة حجمها (100) عينة لعام (2020) وبين المتغيرات التوضيحية (نسبة الهيموغلوبين الموجودة في الكريات الحمراء في الدم، كريات الدم الحمراء، كريات الدم البيضاء، الصفائح الدموية ، نسبة الهيموغلوبين الموجودة في الدم ، نسبة اللمفاويات في الدم ، نسبة كريات الدم الوحيدة ، نسبة كريات الحمضة، نسبة كريات القعدة) واتضح من خلال المقارنة ان طريقة انحدار الحرف في تقدير أنموذج الانحدار اللوجستي ثنائي الاستجابة هي الافضل في تقدير معلمات أنموذج الانحدار اللوجستي في حالة وجود مشكلة التعدد الخطي.

نوع البحث: ورقة بحثية.

**المصطلحات الرئيسية للبحث:** الانحدار اللوجستي الثنائي ، طريقة الامكان الاعظم ، طريقة انحدار الحرف اللوجستي ، اسلوب الجاكناييف

بحث مستل من رسالة ماجستير

## المقدمة

يعد أنموذج الانحدار اللوجستي الأسلوب الإحصائي المستخدم لتوفيق العلاقة بين المتغير التابع ثنائي القيمة وعدة متغيرات أيا كان نوعها ، ويسمى الانموذج في هذه الحالة أنموذج الانحدار اللوجستي الثنائي ، وتعتبر نماذج الانحدار اللوجستي حالة خاصة من حالات الانحدار العام. (Generalized linear models) ويعد أسلوب الجاكنيف من الأساليب اللامعلمية وطريقة من طرائق إعادة معاينة التي لا تحتاج الى افتراضات ، وفكرتها الأساس هي معاملة العينة كمجتمع ويركز هذا البحث توضيح تقنية الجاكنيف بحذف حالة واحدة في ظل وجود مشكلة تعدد الخطي.

ويعد الانحدار الخطي الاعتيادي احد انواع نماذج الانحدار اذ انه يحقق اهداف البحث العلمي ، ويعد اساليبه الجزء المهم والاساس لأي تحليل للبيانات ويهدف الى دراسة ووصف وتفسير العلاقة بين المتغير التابع (y) والمتغيرات التوضيحية (X) ، وبالرغم من ذلك فهو يعجز عن دراسة وتفسير العلاقة بين المتغيرات التفسيرية او التوضيحية عندما يكون المتغير التابع ثنائي الاستجابة ولهذا برزت الحاجة الى اساليب انحدار اخرى كالانحدار اللوجستي الثنائي. ويهدف البحث الى استخدام مقدرات الإمكان الأعظم و مقدرات انحدار الحرف في تقدير الأنموذج الانحدار اللوجستي الثنائي ، من خلال استخدام أسلوب الجاكنيف واجراء المقارنة بينهما لمعرفة الطريقة الفضلى، باعتماد على معيار المعلومات (AIC).

## أنموذج الانحدار اللوجستي (الثنائي الاستجابة)

### Logistic Regression Model (Binary Response)

يعتبر أنموذج الانحدار اللوجستي ثنائي الاستجابة الذي يقوم على فرض أساسي وهو إن متغير الاستجابة (y) يكون متغير وصفي يتوزع توزيع بيرنولي (Bernoulli) حيث يأخذ القيمة (1) في حالة النجاح (Success) باحتمال مقداره (π) والقيمة (0) في حالة الفشل (Failure) باحتمال مقداره (1-π) [Gebara, 2014]

إن المتغيرات التوضيحية تأخذ قيمة مستمرة لذلك فإن الأنموذج الاتي هو الذي يربط بين هذه المتغيرات:

$$Y = \beta_0 + \beta_1 x_i + \epsilon_i \quad \dots\dots\dots(1)$$

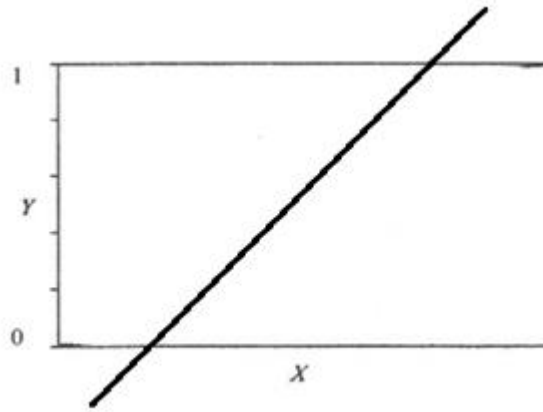
ويعد (y) هو المتغير المعتمد الذي يكون متغيراً عشوائياً (Random Variable) وعلى فرض إن متوسط قيم المتغير (y) الفعلية او المشاهدة عند قيمة معينة من قيم المتغير التوضيحي (x) تمثل E(y) ، لذا فإن الأنموذج يمكن كتابته كالآتي:

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

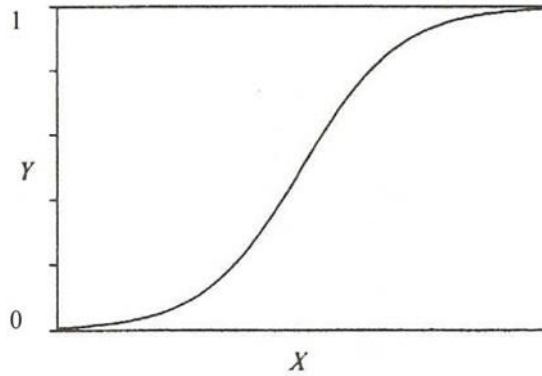
من المعروف إن الطرف الأيمن لأنموذج الانحدار الخطي أعلاه يأخذ قيمة من سالب ما لانهاية (-∞) الى موجب ما لانهاية (∞) ، لذا فإن استخدام هذا الأنموذج مع البيانات الوصفية لا يكون ملائماً والسبب في ذلك لأن:

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

و بما أن قيمة الطرف الأيمن تكون محصورة بين (1,0) ، ولهذا السبب 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) [Al-Jaouni, 2011] كما في الشكل (1) ادناه.



الشكل (1) يمثل خط توفيق العلاقة بين متغير ثنائي القيمة وآخر مستمر لذلك في هذه الحالة لا يمكننا استخدام أنموذج الانحدار الاعتيادي وإنما نستخدم أنموذج الانحدار اللوجستي الذي يأخذ شكل حرف الـ (S) كما في الشكل (2)



الشكل (2) العلاقة بين متغير الاستجابة الوصفي (y) الثنائي والمتغير التوضيحي (x) ويتضح من الشكل رقم (2) أن العلاقة بين المتغيرين (y,x) غير خطية ، ولصعوبة استخدام النماذج اللاخطية لوصف وتحليل البيانات تم تحويل أنموذج الانحدار اللوجستي الى الشكل الخطي هذا ما اقترحه الباحث (Berkson) عام (1944) من خلال ادخال تحويل رياضية تعرف بتحويل اللوجيت (Logit Transformation) على المتغير التابع.

و من المعروف إن قيمة (π) محصورة بين (1,0) أي إن (0 ≤ π ≤ 1) ولذلك فإن النسبة (π / (1-π)) ستكون محصورة بين (∞ , 0) أي إن (∞ ≤ π / (1-π) ≤ 0) هذا يعني انها مقدار موجب ،هذا سيتيح لنا أخذ اللوغاريتم الطبيعي (Normal Logarithm) للأساس (e) للنسبة (π / (1-π))، أن مجال لهذه القيمة

سوف تكون (∞ ≤ log<sub>e</sub> (π / (1-π)) ≤ ∞) أي محصور بين (-∞ , ∞). [Berkson,1944]. وعليه فإن أنموذج الانحدار اللوجستي البسيط ثنائي الاستجابة (Binary) الخطي في حالة متغير توضيحي واحد ويمكن كتابته بالصيغة الآتية :

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

إن الصيغة (4) اعلاه تمثل انموذج الانحدار اللوجستي البسيط ثنائي الاستجابة بصيغته خطية تعرف بتحويل اللوجيت (Logit Transformation)، حيث إن أنموذج الانحدار اللوجستي هو عبارة عن تحويله لوغاريتمية لئلا يتأثر الخطي إلا أننا في الانحدار اللوجستي يكون هدفنا ليس تفسير التغير في قيم متغير الاستجابة وإنما تفسير احتمال حدوث وعدم حدوث الظاهرة محل الدراسة. [Bertsimas,2017] طريقة تقدير الامكان الاعظم

تستخدم طريقة الامكان الأعظم لايجاد تقديرات معالم نماذج الانحدار، وإن هذه التقديرات لمعالم نماذج الانحدار  $\beta$  القيم التي تعظم دالة الامكان. وعندما يتبع متغير الاستجابة  $(y)$  يتبع توزيع ثنائي الحدين فإن دالة الكثافة الاحتمالية لـ  $(y)$  هي كما يأتي: [Liang,2012]

$$P(y = y_i) = C_{y_i}^{n_i} \pi_i^{y_i} (1-\pi_i)^{n_i-y_i} \dots\dots\dots (5)$$

ان دالة الكثافة الاحتمالية لـ  $(y)$  لدالة الامكان هي

$$L(\beta, y) = \prod_{i=1}^n C_{y_i}^{n_i} \pi_i^{y_i} (1-\pi_i)^{n_i-y_i} \dots\dots\dots(6)$$

وعند تبسيط المعادلة رقم (6) واخذ اللوغاريتم نحصل على :

$$\text{Ln}L(\beta, y) = \sum_{i=1}^n [ \ln C_{y_i}^{n_i} + y_i \text{Ln}(\frac{\pi_i}{1-\pi_i}) + n_i \text{Ln}(1 - \pi_i) ] \dots\dots\dots(7)$$

إذ أن  $(\pi_i)$  هو احتمال النجاح وتكون صيغته على نحو الآتي:

$$\pi_i = \frac{e^{\beta_0 + \sum \beta_j x_i}}{1 + e^{\beta_0 + \sum \beta_j x_i}} \dots\dots\dots (8)$$

وعند تعويض المعادلة (8) في المعادلة (7) نحصل على الآتي:

$$\text{Ln}L(\beta, y) = \sum_{i=1}^n [ \text{Ln} C_{y_i}^{n_i} + y_i (\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_p) + n_i \text{Ln}(\frac{1}{1 + e^{\beta_0 + \sum \beta_j x_i}}) ] \dots\dots\dots(9)$$

وبما إن تقديرات الامكان الأعظم هي قيم المعلمات المقدرة  $\beta$  التي تعظم دالة الامكان، لهذا السبب يمكن ايجادها عن طريق اشتقاق المعادلة (9) للمعلمات المراد تقديرها، ثم مساوات المعادلات الناتجة بالصفر لنحصل على  $(k+1)$  من المعادلات وبما ان هذه المعادلات غير خطية لذلك يتم استعمال طريقة نيوتن- رافسون (Newton-Raphson) لايجاد حل لهذه المعادلات [Czepiel,2002].

### اختبار وجود مشكلة التعدد الخطي

لاختبار وجود او عدم وجود مشكلة التعدد الخطي نستعمل اختبار الدليل الشرطي (condition Index) إذ انه يساوي حاصل قسمة أكبر جذر مميز على كل من الجذور المميزة، فإذا كانت نتيجة قيمة الدليل اكبر من (15) فهو مؤشر على إمكانية وجود مشكلة التعدد الخطي، ويستفاد منه في معرفة درجة التعدد الخطي والمتغيرات المرتبطة مع بعضها البعض [Lin,2007]

### Logistic Ridge Regression

### طريقة انحدار الحرف اللوجستي

لمعالجة مشكلة التعدد الخطي في الانحدار اللوجستي كبديل لطريقة الامكان الاعظم نستخدم طريقة انحدار الحرف اللوجستي، حيث قدم في عام 1962 الباحثين Arthur E. Hoerl & Robert W. Kennard مقالين في مجلة Technometrics قدما فيهما انحدار الحرف (Ridge Regression) كحل متكامل لمشكلة التعدد الخطي. [Mansson & Shakur, 2011] واقتراح العالم (Shaefer et al) في عام 1984 صيغة انحدار الحرف اللوجستي الآتية: [Kibria & Mansson & Shakur, 2012]

$$\hat{B}_{LRR} = (W^T \hat{W} X + K I_p)^{-1} (X^T \hat{W} X) \hat{B}_{ML} \dots\dots\dots (10)$$

$\hat{W}$ : تمثل اوزان

$k$ : تمثل قيمة ثابت الحرف وهي معلمة غير عشوائية عندما  $k=0$  فان مقدرات طريقة انحدار الحرف هي نفسها مقدرات طريقة الامكان الاعظم، ويمكن ايجاد قيمة  $k$  من خلال الصيغة الآتية: (Hoerl & Kennard)

$$K_1 = \frac{\hat{\sigma}^2}{\hat{\alpha}_{max}^2} \dots\dots\dots (11)$$

حيث ان:

$\hat{\alpha}_{max}^2$ : العنصر الاكبر من المتجه  $(Y \beta_{ML})$ ،  
 $Y$ : مصفوفة المتجهات المميزة للمصفوفة  $(X^T \hat{W} X)$ ،  
 $\hat{\sigma}^2$ : تباين البواقي ويمكن ايجادها بالصيغة الآتية:

$$\hat{\sigma}^2 = \frac{Y^T \hat{W} Y - b_{ML}^T \hat{W} Y}{n-p-1} \dots\dots\dots(12)$$

وإن  $p$ : تمثل عدد المتغيرات المستقلة.

## Jackknife procedure

## أسلوب الجاكنيف

اقترح (Quenouille) في عام (1949) أسلوب الجاكنيف (Jackknife) وهو أسلوب لا معلمي وطريقة من طرق اعادة المعاينة يستعمل لتقدير التحيز والتباين للاحصاءات المختلفة، ويهدف إلى تقليل التحيز في تقدير المعلمات بنماذج الانحدار إلى أقل ما يمكن. تعرف طريقة (الجاكنيف) على أنها تلك الطريقة التي تقوم اساساً على استبعاد جزء من البيانات في كل مرة عند إجراء عملية التقدير. وللجزء المستبعد حالتان: إما اهمال الجزء المستبعد من البيانات بشكل كلي عند إجراء عملية التقدير، أو أخذ الجزء المستبعد من البيانات بنظر العناية عند إجراء عملية التقدير إذ يتم احتساب وزن لهذا الجزء وادخاله في عملية التقدير. وبالإمكان تطبيق أسلوب (الجاكنيف) في تقدير معلمات نماذج الانحدار الخطية وغير الخطية- [Obiora-Houno,2012,p.75].

وان احد اهم الخوارزميات المستعملة في تقدير انموذج الانحدار اللوجستي هي (JACKKNIFE DELETE-ONE ALGORITHM) اي خوارزمية الجاكنيف لحذف مشاهدة واحدة وان خطوات هذه الخوارزمية هي على النحو الاتي: [Ali,p.90]

الخطوة الاولى: سحب عينة عشوائية بحجم  $n$  من المجتمع وتسمية عناصر المتجه  $Z_i = (Y_i, X_{ji})'$  متجه  $Y_i = (y_1, y_2, \dots, y_n)^T$  ، المصفوفة  $X_{ji} = (x_{j1}, x_{j2}, \dots, x_{jn})^T$  حيث ان:

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

الخطوة الثانية: حذف الصف الاول للمتجه  $Z_i = (Y_i, X_{ji})'$  وتسمية مجموعة المشاهدات المتبقية ب  $(n - 1)$  اي:

$$Z_1^{(j)} = (y_2^{(j)}, y_3^{(j)}, \dots, y_n^{(j)})^T \text{ و } X_{ji}^{(j)} = (x_{j2}^{(j)}, x_{j3}^{(j)}, \dots, x_{jn}^{(j)})'$$

وبعد هذا يتم تقدير معلمات الانحدار اللوجستي باستعمال طريقة انحدار الحرف لعينة الجاكنيف، الخطوة الثالثة: نستمر بنفس الاسلوب نحذف كل مرة حالة واحدة من مجاميع المشاهدات ونقدر معلمات الانحدار اللوجستي ثنائي الاستجابة  $\hat{\beta}^{(j)}$  حيث ان  $\hat{\beta}^{(j)}$  معلمات الانحدار المقدره بطريقة الجاكنيف بعد حذف مجموعة المشاهدات  $i$  من  $Z_i$ .

الخطوة الرابعة: الحصول على التوزيع الاحتمالي  $F(\hat{\beta}^{(j)})$  من تقديرات المعلمات (jackknife)  $(\hat{\beta}^{(j_1)}, \hat{\beta}^{(j_2)}, \dots, \hat{\beta}^{(j_n)})$

الخطوة الخامسة: حساب معاملات الانحدار jackknife الذي يمثل الوسط الحسابي لتوزيع  $F(\hat{\beta}^{(j)})$  وعلى النحو الاتي:

$$\hat{\beta}^{(j)} = \frac{\sum_{i=1}^n \hat{\beta}^{(j_i)}}{n} = \bar{\beta}^{(j_i)} \quad \dots \dots \dots (2 - 13)$$

حيث ان  $\hat{\beta}^{(j)}$  تقديرات غير متحيزة لمتجه المعلمات  $\beta$ .  
الخطوة السادسة: حساب معادلة الانحدار التقديرية jackknife لحذف حالة واحدة لانموذج الانحدار اللوجستي وعلى النحو الاتي:

$$\hat{Y} = [1 + \exp(-\hat{\beta}^{(j)})X]^{-1} \quad \dots \dots \dots (2-9)$$

## سرطان الثدي

يعد سرطان الثدي من السرطانات المنتشرة عند الاناث في العالم بصورة عامة والعراق بصورة خاصة ، إذ انه يمثل ما يقارب ثلث نسبة السرطانات التي أصيبت بها المرأة العراقية ، حسب ما مدون في سجل السرطان العراقي الاخير الذي تصدره وزارة الصحة العراقية، ووضحت الصحة ان سرطان الثدي يحتل المرتبة الاولى نسبة الى السرطانات التي يصاب به الفرد العراقي ، و من اسباب الاصابة بسرطان الثدي ما يأتي:

- 1- التقدم في عمر الانسان:  
يندر الإصابة بسرطان الثدي قبل خمس وعشرين (25) سنة ، حيث أن خطر الإصابة يزيد في الفئة العمرية بين (44-50) سنة. [ Saadi & Badarina,2008 ]
- 2- التاريخ العائلي:  
نسبة الإصابة للأقرباء من الدرجة الأولى بسرطان الثدي (15-20) وتزيد النسبة إذا كانت الإصابة في كلا الثديين [ American Cancer Society,2013]
- 3- العوامل الوراثية:  
تمثل العوامل الوراثية احتمالية الإصابة بسرطان الثدي 5% من عدد الحالات وخصوصاً إذا كانت الام مصابة أو احدى الأخوات ، وتكون اعلى في النساء اللاتي لديهن اقارب من الدرجة الأولى (الام - الاخت - الابنة) مصابات بهذا المرض ، اما إذا كانت الاقارب من الدرجة الثانية (الجدة - العمّة - الخالة) سواء من ناحية الام او الاب فإن نسبة الإصابة ترتفع ولكن اقل من الحالة الأولى [ American Cancer Society,2013]
- 4- العوامل الغذائية:  
اتباع نظاماً غذائياً سيئاً وخاصة إذا كانت نسبة الشحوم (الدهون) فيه مرتفع يرافقه زيادة الوزن (السمنة) كل هذا يزيد من احتمالية الإصابة , اي إذا زاد وزن المرأة عن 40% من الوزن المثالي . [Saadi & Badarina,2008]
- 5- حبوب منع الحمل:  
استخدام حبوب منع الحمل بشكل مستمر يسبب في زيادة احتمالية الإصابة وخاصة قبل الحمل الاول. [ . ]
- 6- التعرض الزائد للهرمونات النسائية:  
ان استخدام العلاج الهرموني البديل يزيد من احتمالية الإصابة بسرطان الثدي ، في حين انه يختفي هذا الخطر بعد توقف العلاج ، و اظهرت الدراسات انه لا يوجد أي ترابط بين استخدام السابق للعلاج الهرموني لمدة تزيد عن (10) سنوات ، ويزيد خطر الإصابة عندما يتجاوز مدة العلاج (5) سنة . [ Saadi & Badarina,2008 ] و [ Al-Adaini & Al-Najjar,2012 ]
- 7- عوامل اخرى:

### الإشعاعات :

ان التعرض للإشعاع وخاصة بين سن البلوغ وسن (30) سنة ، يزيد من خطر الإصابة بسرطان الثدي بدرجة كبيرة ، وقد يصاب الاشخاص اللذين يتعرضون لإشعاعات معينة ولمدة طويلة بامراض السرطان . [ . ]

Saadi & Badarina,2008

### التلوث البيئي بمواد كيميائية :

في كثير من الحالات يدخل المسرطن الكيميائي الاغذية عبر استعماله للزراعة ، وعند تفرغ النفايات تنشر بعض المصانع مواد مسرطنة كيميائية في البيئة وقد تتلوث هذه المواد الهواء او مياه الشرب .

[ Saadi & Badarina,2008 ]

ويمكن تحديد مراحل الورم السرطاني وفقاً لحجم الورم و إذا كان انتشر إلى العقد اللمفاوية أو غيرها من أجزاء الجسم [Lancet,1996,p1713],ومن مراحل سرطان الثدي :

- 1- المرحلة الصفرة: وتعرف بسرطان الأقفنية الموضوع وهو يبدأ في القنوات اللبنية ولا ينتقل إلى الأنسجة المحيطة
- 2- المرحلة الأولى: وفيها يكون الورم في حجم 2 سنتيمتر ولا يؤثر على الغدد اللمفاوية بعد.
- 3- المرحلة الثانية: وفي هذه المرحلة يكون الورم لا يزال حجمة 2 سنتيمتر، ولكن يبدأ ينتشر إلى العقد اللمفاوية القريبة.
- 4- المرحلة الثالثة: الورم ازداد حجمه اذ يبلغ حجمة 5 سنتيمتراً ، ويبدأ ينتشر في بعض الغدد اللمفاوية.
- 5- المرحلة الرابعة: وفيها يكون الورم قد بدأ في الانتشار إلى الأعضاء البعيدة خصوصاً العظام والكبد والدماغ والرئتين

### عينة الدراسة

تم الحصول على بيانات الدراسة من مستشفى بعقوبة العام /قسم المختبرات للمدة الزمنية (يناير 2020 – يونيو 2020)، وتم اختيار عينة حجمها (100) من خلال الاستمارة الإحصائية لسرطان الثدي .  
أذ إن متغيرات البحث كما يلي :

Y : يمثل متغير معتمد ويمثل نتيجة لمرض سرطان الثدي ويأخذ القيم:  
(y=1) يمثل مصاب, (y=0) يمثل غير مصاب.

اما متغيرات التوضيحية فتتمثل كما يأتي:

X<sub>1</sub> : يمثل نسبة الهيموغلوبين الموجودة في الكريات الحمراء في الدم

X<sub>2</sub>: يمثل كريات الدم الحمراء

X<sub>3</sub> : يمثل كريات الدم البيضاء

X<sub>4</sub> : يمثل الصفائح الدموية

X<sub>5</sub> : يمثل نسبة الهيموغلوبين الموجودة في الدم

X<sub>6</sub> : يمثل نسبة اللمفاويات في الدم

X<sub>7</sub> : يمثل نسبة كريات الدم البيضاء الوحيدة

X<sub>8</sub> : يمثل نسبة كريات الحمضة

X<sub>9</sub> : يمثل نسبة كريات القعدة

تم تحليل البيانات بالاعتماد على برنامج لغة (R) ، حيث تم الحصول على النتائج الخاصة بالأنموذج الانحدار اللوجستي الثنائي بطريقة الامكان الاعظم وطريقة انحدار الحرف باستخدام اسلوب الحاكنايف التقدير باستعمال طريقة الامكان الاعظم

لمعرفة مدى تأثير المتغيرات التوضيحية في المتغير المعتمد يتم تقدير معاملات الانحدار ( $\hat{\beta}$ ) كما موضح في الجدول الآتي :

جدول رقم (1) نتائج تقديرات طريقة مقدر الامكان الاعظم لأنموذج

الانحدار اللوجستي الثنائي

القيمة الجدولية $\chi^2(1)$	درجة الحرية Df	احصاءة Wald	الأخطاء المعيارية S.E	معاملات الانحدار $\hat{\beta}$ اللوجستي	المتغيرات التوضيحية
3.841	1	0.001	4469.708	-125.897	x <sub>1</sub>
3.841	1	1.572	3.600	-4.514	x <sub>2</sub>
3.841	1	1.656	0.635	-0.818	x <sub>3</sub>
3.841	1	0.176	0.013	0.006	x <sub>4</sub>
3.841	1	0.001	4469.709	127.900	x <sub>5</sub>
3.841	1	0.615	0.168	-0.132	x <sub>6</sub>
3.841	1	0.036	0.394	0.075	x <sub>7</sub>
3.841	1	0.048	1.004	-0.220	x <sub>8</sub>
3.841	1	2.069	68.381	98.363	x <sub>9</sub>
3.841	1	0.008	17.295	1.549	الحد الثابت

الجدول رقم (1) يوضح ان قيمة الاحصاءة Wald لجميع المعالم اقل من القيمة الجدولية لمربع كاي التي تساوي 3.841 عند درجة حرية (1) ومستوى معنوي (0.05) لذا نقبل فرضية العدم، اي ان متغيرات ليس له تأثير معنوي في متغير الإستجابة .

قيمة معيار المعلومات (AIC) تساوي (13.3105)

### تقدير حسب طريقة الامكان الاعظم باستخدام أسلوب الجاكنيف

من اجل الحصول على نتائج تقدير نموذج الانحدار اللوجستي باستخدام أسلوب الجاكنيف (Jackknife) وهي احد طرق اعادة المعاينة حيث يتم حذف عينة واحدة مع الاسترجاع في كل مرة يتم تقدير المعلمة فيها. ولمعرفة مدى تأثير المتغيرات التوضيحية في المتغير المعتمد يتم تقدير معاملات الانحدار ( $\beta$ ) ، وكما موضح في الجدول الآتي :

جدول رقم (2) نتائج تقديرات طريقة مقدر الامكان الاعظم باستخدام أسلوب الجاكنيف لأنموذج الانحدار اللوجستي الثنائي

القيمة الجدولية	درجة الحرية	احصاءة Wald	الأخطاء المعيارية S.E	معاملات الانحدار اللوجستي $\beta$	المتغيرات التوضيحية
$\chi^2(1)$	Df				
3.841	1	3.1877	73784.1850	-416.5839	$x_1$
3.841	1	1.4339	299.1478	-3.5821	$x_2$
3.841	1	3.8836	301.2786	-5.9372	$x_3$
3.841	1	2.0751	156.7695	0.7141	$x_4$
3.841	1	3.2566	93836.3608	535.4989	$x_5$
3.841	1	1.0734	303.7928	-3.1474	$x_6$
3.841	1	1.4672	607.7583	0.7361	$x_7$
3.841	1	3.0332	188.4916	-1.0381	$x_8$
3.841	1	1.2636	1896.6119	67.4193	$x_9$
3.841	1	9.6293	6063.6631	59.5023	الحد الثابت

جدول رقم (2) اعلاه يوضح ان اشارة معامل تأثير ان اشار معامل التأثير المتغير ( $x_3$ ) سالبة ، هذا يعني وجود علاقة عكسية بين المتغير ( $x_3$ ) والمتغير التابع التي تزيد من احتمالية عدم الاصابة بسرطان الثدي ، ويتضح ايضاً ان متغير ( $x_3$ ) له تأثير معنوي من خلال قيمة الاحصاءة Wald والتي هي اكبر من القيمة الجدولية لمربع كاي التي تساوي (3.841) عند درجة حرية (1) ومستوى معنوي (0.05) ، لذا نرفض فرضية العدم، وله تأثير معنوي ، والمتغيرات ( $x_1, x_2, x_4, x_5, x_6, x_7, x_8, x_9$ ) ليس لهم تأثير معنوي لان قيمة الاحصاءة Wald اقل من القيمة الجدولية لمربع كاي التي تساوي 3.841 عند درجة حرية (1) ومستوى معنوي (0.05) لذا نقبل فرضية العدم، اي ان متغيرات ( $x_1, x_2, x_4, x_5, x_6, x_7, x_8, x_9$ ) ليس له تأثير معنوي في متغير الاستجابة .  
قيمة معيار المعلومات (AIC) تساوي (14)

### نتائج اختبار وجود مشكلة التعدد الخطي

اذ ان من خلال ملاحظة الجدول رقم (3) توجد قيم الدليل الشرطي أكثر من (15) ، وكذلك يوجد في نسب التباين قيم أكثر من (0.70) ، وهذا يشير الى وجود ارتباط بين اثنين او اكثر من المتغيرات التوضيحية (X) ومن ثم وجود مشكلة التعدد الخطي .



جدول رقم (3) اختبار مشكلة التعدد الخطي ونسب التباين

Dimension	Condition Index	Variance Proportions									
		(Constant)	x1	x2	x3	x4	x5	x6	x7	X8	X9
1	1.000	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
2	2.802	.00	.00	.00	.00	.00	.00	.00	.78	.00	.05
3	2.971	.00	.00	.00	.01	.00	.00	.00	.14	.01	.36
4	4.143	.00	.00	.00	.02	.00	.00	.00	.01	.85	.00
5	6.522	.00	.00	.00	.27	.04	.01	.13	.01	.11	.12
6	10.419	.00	.01	.00	.09	.04	.07	.71	.03	.00	.10
7	11.383	.00	.00	.00	.50	.87	.00	.03	.01	.00	.09
8	18.664	.02	.03	.14	.03	.04	.77	.04	.00	.00	.02
9	30.220	.50	.05	.68	.01	.00	.11	.00	.01	.00	.02
10	36.055	.48	.90	.17	.06	.00	.03	.09	.02	.03	.24

### معالجة مشكلة التعدد الخطي

من طرق معالجة مشكلة التعدد الخطي هي طريقة انحدار الحرف وكما يأتي:

### تقدير طريقة انحدار الحرف

تعتبر طريقة انحدار الحرف (Ridge Regression) احد طرق المعالجة عند وجود مشكلة التعدد الخطي في الانحدار اللوجستي كبديل لطريقة الامكان الاعظم ، لذلك تم استخدامها في تقدير الأنموذج الانحدار اللوجستي الثنائي حسب طريقة انحدار الحرف وكما موضح في الجدول الآتي :

جدول رقم (4) نتائج تقديرات طريقة انحدار الحرف لأنموذج الانحدار اللوجستي الثنائي

القيمة الجدولية $\chi^2(1)$	درجة الحرية Df	احصاءة Wald	الأخطاء المعيارية S.E	معاملات الانحدار اللوجستي $\hat{\beta}$	المتغيرات التوضيحية
3.841	1	-1.685	1.9634	-3.3082	$x_1$
3.841	1	-2.164	2.0513	-4.4386	$x_2$
3.841	1	-3.284	2.1904	-7.1926	$x_3$
3.841	1	-0.635	2.0728	-1.3165	$x_4$
3.841	1	0.947	1.9929	1.8864	$x_5$
3.841	1	-2.147	2.0924	-4.4923	$x_6$
3.841	1	-0.369	2.3320	-0.8609	$x_7$
3.841	1	-0.797	2.0068	-1.6001	$x_8$
3.841	1	6.765	2.2722	15.3716	$x_9$
				7.1005	الحد الثابت

جدول رقم (4) اعلاه يوضح ان اشار معامل التأثير المتغير ( $x_9$ ) موجبة ، هذا يعني وجود علاقة طردية بين المتغير ( $x_9$ ) والمتغير التابع التي تزيد من احتمالية الإصابة بسرطان الثدي ، ويتضح ايضاً ان متغير ( $x_9$ ) له تأثير معنوي من خلال قيمة الاحصاءة Wald والتي هي اكبر من القيمة الجدولية لمربع كاي التي تساوي (3.841) عند درجة حرية (1) ومستوى معنوي (0.05) ، لذا نرفض فرضية العدم، وله تأثير معنوي، والمتغيرات ( $x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$ ) ليس لهم تأثير معنوي لان قيمة الاحصاءة Wald اقل من القيمة الجدولية لمربع كاي التي تساوي 3.841 عند درجة حرية (1) ومستوى معنوي (0.05) لذا نقبل فرضية العدم، اي ان متغيرات ( $x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$ ) ليس له تأثير معنوي في متغير الاستجابة .

وان قيمة معطمة الحرف (Ridge paramter) تساوي (0.02235696)

قيمة معيار المعلومات (AIC) تساوي (-42)

### تقدير طريقة انحدار الحرف باستخدام اسلوب الجاكناف

ولمعرفة مدى تأثير المتغيرات التوضيحية في المتغير المعتمد يتم تقدير معاملات الانحدار ( $\hat{\beta}$ ) وكما موضح في الجدول الآتي :

جدول رقم (5) نتائج تقدير طريقة انحدار الحرف باستخدام اسلوب الجاكناف لأنموذج الانحدار اللوجستي الثنائي

القيمة الجدولية $\chi^2(1)$	درجة الحرية Df	احصاءة Wald	الأخطاء المعيارية S.E	معاملات الانحدار اللوجستي $\hat{\beta}$	المتغيرات التوضيحية
3.841	1	1.9548	-4.0942	7.1048	$x_1$
3.841	1	2.1799	-4.6721	-0.1810	$x_2$
3.841	1	5.5310	-8.2005	-0.6898	$x_3$
3.841	1	7.8186	-1.5439	-0.1928	$x_4$
3.841	1	9.0116	2.7230	-0.0013	$x_5$
3.841	1	5.4305	-5.3657	0.0817	$x_6$
3.841	1	2.1273	-0.8895	-0.0395	$x_7$
3.841	1	1.8850	-1.5865	-0.0004	$x_8$
3.841	1	3.5287	20.0588	-0.0688	$x_9$
				3.7680	الحد الثابت

المتغيرات ( $x_3, x_4, x_5$ ) سالبة ، هذا يعني وجود علاقة عكسية بين المتغيرات ( $x_3, x_4, x_5$ ) والمتغير التابع التي تزيد من احتمالية عدم الإصابة بسرطان الثدي ، ويتضح ايضاً ان متغيرات ( $x_3, x_4, x_5$ ) لهم تأثير معنوي من خلال قيمة الاحصاءة Wald والتي هي اكبر من القيمة الجدولية لمربع كاي التي تساوي (3.841) عند درجة حرية (1) ومستوى معنوي (0.05) ، لذا نرفض فرضية العدم، ولهم تأثير معنوي ، ان اشار معامل التأثير المتغير ( $x_6$ ) موجبة ، هذا يعني وجود علاقة طردية بين المتغير ( $x_6$ ) والمتغير التابع التي تزيد من احتمالية الإصابة بسرطان الثدي ، ويتضح ايضاً ان متغير ( $x_6$ ) له تأثير معنوي من خلال قيمة الاحصاءة Wald والتي هي اكبر من القيمة الجدولية لمربع كاي التي تساوي (3.841) عند درجة حرية (1) ومستوى معنوي (0.05) ، لذا نرفض فرضية العدم، وله تأثير معنوي، والمتغيرات ( $x_1, x_2, x_7, x_8, x_9$ ) ليس لهم تأثير معنوي لان قيمة الاحصاءة Wald اقل من القيمة الجدولية لمربع كاي التي تساوي 3.841 عند درجة حرية (1) ومستوى معنوي (0.05) لذا نقبل فرضية العدم، اي ان متغيرات ( $x_1, x_2, x_4, x_5, x_8$ ) ليس لهم تأثير معنوي في متغير الاستجابة. قيمة معيار المعلومات (AIC) تساوي (-41)

### الاستنتاجات:

- 1- اظهرت النتائج ان طريقة انحدار الحرف لأنموذج الانحدار اللوجستي الثنائي هي الافضل من استخدام طريقة انحدار الحرف باستخدام اسلوب الجاكناف لأنموذج الانحدار اللوجستي الثنائي حيث تمتلك اقل قيمة لمعيار المعلومات (AIC) .
- 2- اوضحت النتائج ان المتغيرات التوضيحية لطريقة الامكان الاعظم باستخدام اسلوب الجاكناف لأنموذج الانحدار اللوجستي الثنائي ( $x_3$ ) لها تأثير معنوي، والمتغيرات ( $x_1, x_2, x_4, x_5, x_6, x_7, x_8, x_9$ ) ليس لها تأثير معنوي.
- 3- وبينت النتائج ان المتغيرات التوضيحية لطريقة انحدار الحرف باستخدام الجاكناف لأنموذج الانحدار اللوجستي الثنائي ( $x_3, x_4, x_5, x_6$ ) لها تأثير معنوي، والمتغيرات ( $x_1, x_2, x_7, x_8, x_9$ ) ليس لها تأثير معنوي.

## التوصيات

- 1 – دراسة إمكانية إدخال متغيرات توضيحية جديدة قد تكون مهمة ولها تأثير معنوي في الإصابة وعدم الإصابة بسرطان الثدي , ويتم عن طريق مراجعة المختصين لمرض سرطان الثدي لتكون أكثر دقة في اختيار المتغيرات التي لها تأثير معنوي لأنها تؤثر على دقة النتائج التي نحصل عليها.
- 2- التوصية بدراسة نموذج الانحدار اللوجستي المتعدد الاستجابة بالاعتماد على اسلوب الجاكنيف في الدراسات القادمة.
- 3 – من الضروري زيادة الوعي الصحي عن طريق المراجعة المنتظمة للطبيب المختص واجراء الفحوصات اللازمة في وقت مبكر .

## References

- 1- Al-Adaini, Nujoud and Al-Najjar, Sanaa (2012) (Women's attitudes to prevent breast cancer and its impact on psychological compatibility)
- 2-Al-Jaouni, Farid and Ghanem, Adnan, (2011), "The Use of Dual-Response Logistic Regression Technique in the Study of the Most Important Economic Determinants of Family Sufficiency An Applied Study A Random Sample of Households in Damascus Governorate", Research of the Journal of the University of Economic and Legal Sciences, Volume (27), Issue ( 1).
- 3-Ali, Fayyad Abdullah, "Efficient Regression Capabilities for Regular Least Squares and Jackknife Based on Jackknife's Regular Algorithm by Deleting One Case" Research published in the Journal of Statistical Sciences, College of Administration and Economics, Wasit University. Seventh Issue, Page 90
- 4- American Cancer Society (2013), All You Want To Know About Cancer Series, Jarir Bookstore - Riyadh
- 5- Bertsimas, D., & King, A. (2017). Logistic regression: From art to science. *Statistical Science*, 32(3), 367-384.
- 6-6-Berkson, J. (1944) "Application of the Logistic Function to Bio – Assay" *JASA*, Vol. 39, pp. 357 – 365
- 7-Saadi, Warda and Badarina, Muhammad Al-Arabi (2008) (a comparative study of breast cancer in women and its relationship to psychological and social compatibility, comparison strategies).
- 8- Collaborative Group on Hormonal Factors in Breast Cancer. Breast cancer and hormonal contraceptives: collaborative reanalysis of individual data on 53 297 women with breast cancer and 100 239 women without breast cancer from 54 epidemiological studies. *Lancet*. 1996 Jun 22;347(9017):1713-1727
- 9- Gebara, Azhar Kazem, (2014) "Multiple response data analysis for diagnosing eye diseases using discriminatory function and logistic regression, a comparative study", Master Thesis submitted to the Department of Statistics, College of Business and Economics, Al-Mustansiriya University.
- 10- Gabriel, Muhammad Salman Muhammad, 2014, "Linear polarity, its causes, its effects and the main treatment with application to hypothetical data", Ph.D. in Statistics, College of Graduate Studies, Sudan University of Science and Technology, page 73.
- 11- Kibria, B. M. G., K. Mansson, and G. Shukur, (2012). "Performance of some logistic ridge regression estimators". *Compute Econ* 40, 401–414.
- 12- Mansson, K. and G, Shakur, (2011)."On ridge parameters in logistic regression". *Communications in Statistics-Theory and Methods* 40, 3366–3381
- 13-Obiora-IIouno .H.O and Mbegbu .J.I,2012 "JACKKNIFE ALGORITHM FOR THE ESTIMATION OF LOGISTIC REGRESSION PARAMETERS "Vol.2,No.4.

## Using jack knife to estimation logistic regression model for Breast cancer disease

**Naba'a Jaafar Abd**

/ pupils of University Baghdad / College of  
Administration & Economics / Dept of  
Statistics

[nabaajafer204@gmail.com](mailto:nabaajafer204@gmail.com)

**A.D. Mahmoud Mahdi Al-  
Bayati**

University of Baghdad / College of  
Administration & Economics / Dept of  
Statistics

**A.D. Muhammad Jassim  
Muhammad**

University of Baghdad / College of  
Administration & Economics / Dept of  
Statistics

[m.jasim@coadec.uobaghdad.edu.iq](mailto:m.jasim@coadec.uobaghdad.edu.iq)

Received: 21/10/2020

Accepted :22/11/2020

Published :FEBRUARY / 2021



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International \(CC BY-NC 4.0\)](https://creativecommons.org/licenses/by-nc/4.0/)

### Abstract:

It is considered as one of the statistical methods used to describe and estimate the relationship between randomness (Y) and explanatory variables (X). The second is the homogeneity of the variance, in which the dependent variable is a binary response takes two values (One when a specific event occurred and zero when that event did not happen) such as (injured and uninjured, married and unmarried) and that the large number of explanatory variables led to the emergence of the problem of linear multiplicity that makes the estimates inaccurate, and the method of greatest possibility and the method of declination of the letter was used in estimating A double-response logistic regression model by adopting the Jacknaev method and comparing the capabilities according to the information standard (AIC)

The Jackknife method and the aforementioned statistical capabilities were applied to study the relationship between the response variable (incidence and absence of breast cancer) for a sample size of (100) samples for the year (2020) and the explanatory variables (the percentage of hemoglobin present in red cells in the blood, red blood cells, white blood cells, Platelets, the percentage of hemoglobin in the blood, the percentage of lymphocytes in the blood, the percentage of monocytes, the percentage of eosinophils, the percentage of basophils) And it was evident through comparison that the character regression method in estimating the two-response logistic regression model is the best in estimating the parameters of the logistic regression model in the case of a problem of linearity.

Research Type: Research Paper

**Search keywords** / binary logistic regression, greatest possibility method, logistic regression method, jackknife method