JRUCS
Journal of AL-Rafidain University College for Sciences

PISSN: (1681-6870); EISSN: (2790-2293)
مجلة كلية الرافدين الجامعة للعلوم

Available online at: https://www.jrucs.iq

AL- Rafidain University College

استعمال نموذج الانحدار اللوجستي الرتبي المتعدد في تحديد أهم العوامل الديموغرافية والبيئية المؤثرة على الاصابة بأمراض السرطان لدى الاطفال في محافظة البصرة

م.م. شيرين علي حسين	م.د. سعد كاظم حمزة
shreen.a@uobaghdad.edu.iq	saad.hamza@coadec.uobaghdad.edu.iq
قسم ضمان الجودة والاداء الجامعي - جامعة بغداد، بغداد، العراق.	قسم الإحصاء- كلية الإدارة والاقتصاد- جامعة بغداد، بغداد، العراق.

المستخلص

يمتاز الانحدار اللوجستي بالمرونة العالية اذ يعرف بانه احد الاساليب الاحصائية المهمة التي تستخدم للتنبؤ باحتمالية وقوع حدث ما وذلك من خلال ملائمة البيانات والمعلومات على المنحنى اللوجستي حيث يعمل عادة لوصف طبيعة العلاقة بين متغيرين الاستجابة والذي يكون عادة فئوي او رتبي ومجموعة من المتغيرات المستقلة. لذلك يعد انموذج الانحدار اللوجستي حالة خاصة من نماذج الانحدار الاعتيادي بسبب طبيعته الاسمية ويكون الانحدار اللوجستي على عدة انواع بحسب متغير الاستجابة منها الانحدار اللوجستي الثنائي (Binary Logistic) ويستخدم عادة عندما يكون متغير الاستجابة يأخذ قيمتين وهي (0,1) وكذلك متعدد الاستجابة (Multinomial Logistic) عندما يأخذ اكثر من فئتين، اما الانحدار اللوجستي الرتبي (Ordinal Logistic) اذ يستخدم في حالة متغير الاستجابة يكون رتبياً .

تم في هذا البحث تحديد اهم العوامل الديمغرافية والبيئية المؤثرة على الاصابة بأمراض السرطان لدى الاطفال في محافظة البصرة وكون متغير الاستجابة عبارة عن شدة الاصابة ابتداءً من عدم الاصابة الى اصابة شديدة لذلك تطلب الامر استعمال نماذج انحدار لوجستي رتبي تمثلت بيانات البحث بعينة من الاطفال بحجم (295) طفلاً من مستشفى الطفل التخصصي في البصرة من سنة 2019 وتوصلنا الى المتغيرات التي تؤثر في تحديد الاصابة ومن بين متغيرات (الجنس، العمر، موقع السكن، مهنة الاب، مستوى تعليم الاب، المستوى المعيشي، نوع عمل الاب، الوراثة، موقع السكن بالنسبة لحقول النفط، مدى القرب من الاماكن الملوثة بالإشعاع، نوع السرطان، العلاج، نوع السكن، مدة الاصابة).

اذ توصل الباحث الى نتيجة انه يوجد عدة متغيرات كان لها تأثير مباشر على الاصابة بمرض السرطان منها (موقع السكن وخاصة اذا كان قريبا من ابار ومصافي النفط، عمل الاب وخاصة اذا كان رب الاسرة كاسبا، العلاج فيما اذا كان متوفر من عدمه) حيث كانت المعنوية اقل من 0.05 وهي ذات تأثير معنوي على الاصابة بالمرض.

معلومات البحث

تواريخ البحث
تاريخ تقديم البحث: 2023/1/11
تاريخ قبول البحث: 2023/3/3
تاريخ رفع البحث على الموقع: 2023/12/31

الكلمات المفتاحية
نماذج الانحدار، الانحدار اللوجستي الرتبي المتعدد، العوامل الديموغرافية والبيئية، سرطانات الاطفال.

للمراسلة:
م.م. شيرين علي حسين
shreen.a@uobaghdad.edu.iq

https://doi.org/10.55562/jrucs.v54i1.601

1. المقدمة Introduction

وردت عدة تعاريف للانحدار اللوجستي منها الانموذج الاكثر ملائمة لبيانات الاستجابة الفئوية .فهو يدل على نحو متزايد في مجموعة واسعة من التطبيقات وكانت الاستعمالات المبكرة له في الدراسات الطبية الحيوية ولكن خلال العقدين الماضيين كان الاستعمال الواسع له في ابحاث العلوم الاجتماعية والتسويق.

وكذلك يعرف بانه تحليل الانحدار المناسب عندما يكون المتغير التابع وصفيا. وهو عبارة عن تحليل تنبؤي يستعمل لوصف البيانات وشرح العلاقة بين المتغير المعتمد وواحد او اكثر من المتغيرات المستقلة الاسمية او الرتبية او الفئوية .

لذلك فهو يعد من الاساليب الاحصائية التي تهتم بتحليل البيانات التي يكون فيها متغير الاستجابة من النوع الوصفي المتقطع وثم يستعمل لوصف العلاقة بين متغير الاستجابة والمتغيرات التوضيحية ويكون الانحدار اللوجستي على ثلاثة انواع اعتمادا على نوع متغير الاستجابة (Y) ،ابسطها الانحدار اللوجستي الثنائي هو النوع الاكثر شيوعا من انواع الانحدار اللوجستي ،اذ يستعمل عندما تكون الاستجابة لكل موضوع نجاحا او فشلا. فندما يكون لمتغير الاستجابة اثنان فقط من القيم المحتملة (0 او 1) يستعمل هذا النوع من الانحدار اللوجستي عادة لتمثيل العلاقة بين متغير الاستجابة من حيث نسبة الارجحية .

والنوع الثاني هو الانحدار اللوجستي متعدد الحدود يستعمل هذا النوع في الحالات التي يكون فيها متغير الاستجابة فئويا ولديه اكثر من قيمتين ممكنتين وتكون قيم هذا المتغير غير قابلة للترتيب .والثالث هو الانحدار اللوجستي الرتبي هو نوع خاص من الانحدار اللوجستي متعدد الحدود والذي يمكن استعماله عندما يكون متغير الاستجابة وصفيا رتبيا ويعرف هذا النوع من المتغيرات بانه شبه كمي خاص من المتغيرات الفئوية حيث يتم ترتيب القيم من الناحية المفهومية ، مثال على ذلك يمكن تصنيف درجة الالم الى (لا شيء، خفيفة، معتدلة، شديدة) وهو الانموذج الذي تم استعماله في هذه الرسالة لتحليل بيانات الدخل الشهري للأسرة. والذي سيتم توظيفه في البحث لغرض تحدي اهم العوامل المؤثرة بأمراض سرطان الاطفال في محافظة البصرة اذ يعتبر مرض السرطان من الامراض المستعصية الخطرة, قد يترتب عليها كثير من المضاعفات التي قد تؤدي الى الوفاة ، وهذه الامراض منتشرة في دول متقدمة ودول نامية بصورة ملحوظة وخطرة نتيجة لتطورات الحياة الحديثة ، فالملوثات الناتجة من ابخرة عوادم السيارات ودخان المعامل والمصافي وحقول النفط ومصانع البتروكيمياويات والحروب التي استخدمت فيها قنابل مصنوعة من اليورانيوم المنضب الذي يحتوي على عناصر مشعة والتي عند استخدامه في الحروب تلوث المكان الذي تم قصفه وتسبب السرطان اوقد تكون مسببات المرض وراثية. وتعد محافظة البصرة من اكثر محافظات العراق التي تعرضت لهذه الاسلحة كما تعد من اكثر المحافظات المنتجة للنفط والتي تحتوي على حقول نفطية عملاقة ومصافي تكرير النفط الخام وهذه المنشآت تصدر بدورها ابخرة ملوثة بشكل يومي مستمر وهي منتشرة في اكثر المناطق البصرة .

تم تحديد عوامل الخطر لأمراض السرطان من خلال استعمال تحليل الانحدار اللوجستي المتعدد الرتبي ، كما هو معروف ومثل أي تقنية أخرى لبناء النموذج فإن الهدف من تحليل الانحدار اللوجستي هو العثور على النموذج الأنسب والأكثر كفاءة وهو الانموذج الانسب لوصف العلاقة بين النتيجة (متغير تابع أو متغير استجابة بأكثر من مستويين) ومجموعة متغيرات مستقلة (أو تفسيرية).

2. نماذج الانحدار اللوجستي الرتبي Ordinal logistic models

هناك العديد من النماذج التي يمكن استعمالها لمثل هكذا متغيرات والتي تعد اشكالا مختلفة ولكل منها خصائصه واستعمالاته حسب طبيعة البيانات والهدف من الانموذج. اما فيما يخص هذا البحث فان طبيعة البيانات تكون من النوع الرتبي وهناك العديد من النماذج اللوجستية التي تستعمل لتحليل هذا النوع من البيانات منها (انموذج الارجحية النسبية ، انموذج الارجحية النسبية الجزئية، انموذج لوجيت للفئات المتجاورة، انموذج لوجيت لنسبة الاستمرارية المقيدة) ولكن اشهرها هو انموذج الارجحية النسبية[6] .

3. انموذج الارجحية النسبية The Proportional Odds Model (POM)

يعد انموذج الارجحية النسبية الذي وصفه Mc Cullagh هو الانموذج الاكثر شيوعا من نماذج الانحدار اللوجستي الرتبي .ويشار احيانا الى انموذج الارجحية النسبية بأنموذج اللوجيت التراكمي The cumulative logit model. ان الصفة المميزة لهذا الانموذج هي ان نسبة الارجحية للمؤشر يمكن تفسيرها على انها ملخص لنسب الارجحية التي تم الحصول عليها من انحدارات لوجستية ثنائية منفصلة باستعمال جميع الفئات المحتملة للمتغير الرتبي.

وبما ان فئات متغير الاستجابة تأخذ الشكل الرتبي فيجب استعمال نماذج لوجيت الرتبية حيث تكون مناسبة لها .ويعد انموذج الارجحية النسبية من نماذج اللوجيت الرتبية المناسبة .ويمكن تمثيله بالأنموذج الاتي [7] :

$$Logit[Pr(Y \le j|X)] = \alpha j + \beta' X \qquad i=1,2,.....J-1 \qquad (1)$$

حيث ان αj هو متجه لمعلمات الحد الثابت intercept parameters (ويمكن ان تختلف بحسب مستويات متغير الاستجابة) و $\beta' = (\beta_1, \beta_2, \ldots, \beta_p)$ هو متجه معلمات الانحدار slope parameter $X = (X_1, \ldots, X_p)$ يمثل المتغيرات التوضيحية وان y يمثل متغير الاستجابة و j يمثل عدد الفئات ويكون الانموذج بالشكل الاتي

$$\begin{aligned} \text{logit}[\Pr(y = j|X)] &= \log\left(\frac{\Pr(y = j|X)}{1 - \Pr(y = j|X)}\right) \\ &= \alpha_j + \beta_1 X_1 + \cdots + \beta_P X_P \end{aligned} \qquad (2)$$

$$\begin{aligned} \text{logit}[\Pr(y = 1|X)] &= \log\left(\frac{\Pr(y = 1|X)}{1 - \Pr(y = 1|X)}\right) \\ &= \alpha_1 + \beta_1 X_1 + \cdots + \beta_P X_P \end{aligned} \qquad (3)$$

$$\text{logit}[\Pr(y = J-1|X)] = \log\left(\frac{\Pr(y = J-1|X)}{1-\Pr(y = J-1|X)}\right) \quad (4)$$

$$= \alpha_{J-1} + \beta_1 X_1 + \cdots + \beta_P X_P \qquad \text{j=1, … ,J-1}$$

وهذا ما يسمى بأنموذج الارجحية النسبية لأنه يقوم على افتراض ان الاثار المترتبة على المتغيرات التوضيحية $X = (X_1, .., X_p)$ هي نفسها لجميع الفئات، على مقياس لوغاريتمي ويدعى انموذج الارجحية النسبية بافتراض الخطوط المتوازية Parallel lines assumption حيث انه يعتمد على الاحتمالات التراكمية لفئات الاستجابة لذلك يجب اختبار هذ الافتراض لكل متغير بشكل منفصل وفي الانموذج النهائي، وذلك باستعمال اختبار برانت Brant test على سبيل المثال عندما يتناسب انموذج الارجحية النسبية جيدا يتطلب معلمة مفردة بدلا من المعلمات $j\text{-}1$ لوصف تأثير X_i . وهذا يعني انه اذا استعمل باحثان استجابة مختلفة الفئات في دراسة تأثير عامل التنبؤ يجب ان يصل كل باحث الى استنتاجات مماثلة ومن ثم فان اختيار عدد الفئات لن يكون له تأثير في النتائج [7] .

ويمكن اجراء اختبارات لفرضيات نماذج الانحدار اللوجستي باستخدام إحصاءه والد Wald استنادا الى تقديرات المعلمات. اما في حالة عدم تحقق فرض الارجحية النسبية لبعض المتغيرات عند اجراء اختبار Brant فننتقل الى نموذج اخر يضم المتغيرات التي تحقق فرض الارجحية النسبية والمتغيرات التي لا تحقق هذا الفرض ويسمى بأنموذج الارجحية النسبية الجزئية .

4. تقدير معلمات الانحدار اللوجستي

يبنى نموذج انحدار اللوجستي على فرض اساسي هو ان متغير الاستجابة او متغير تابع (y) هو متغير ثنائي الاستجابة. ويأخذ الرتبة اما (1) باحتمال (π) او (0) باحتمال ($1\text{-}\pi$) اي الى حدوث الاستجابة وعدم حدوثها .

وتأخذ دالة الاستجابة (الدالة اللوجستية) الشكل الاتي:

$$F(Z) = E(Y/Z) = \frac{e^z}{1-e^z} \quad (5)$$

وهي مثل نظرية الاحتمالات تأخذ القيم اما (0 او1) وتكون عملية تصنيف خاضعة لمخرجات هذه الدالة فاذا كانت قيمتها تساوي (0) يعني ان المشاهدة تنتمي للمجموعة الاولى واذا كانت قيمتها تساوي (1) يعني ان المشاهدة تنتمي للمجموعة الثانية وتأخذ دالة الاستجابة مدخلات من (∞,∞-) لكن المجموعات دائما بين (1, 0) ويمثل المتغير (z) المتغيرات التوضيحية ان F(Z) تمثل الاحتمال لمخرج معين لمجموعة من متغيرات توضيحية ويقيس متغير Z مجموع مساهمات جميع المتغيرات التوضيحية المستخدم في هذا الانموذج والتي تعرف باللوجت (logit) ويعرف المتغير Z كالاتي [8] [4] :

$$Z = \beta_0 + \beta_1 X_1$$

اما في حالة انحدار لوجستي متعدد فيكتب بالشكل الاتي:

$$Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots\ldots + \beta_K X_k \quad (6)$$

حيث ان: $\beta_0, \beta_1, \ldots\ldots\ldots\ldots, \beta_k$

ولتحويل الصيغة (6) الى الشكل الخطي يتم استعمال ما يعرف بتحويلة لوجت (Loget Transformation) الاتية :

$$g(z) = Ln\left[\frac{f(z)}{1-f(z)}\right] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_K X_K \quad (7)$$

اذ ان:

P (Y=1/X) الاحتمال الشرطي (Conditional Probability) بان تكون Y=1 عند قيمة (X)

P (Y= 0/x) =1-f(x) عندما Y=0 وتمثل P (Y=0/X) والاحتمال الشرطي (Conditional Probability) بان تكون Y =0 عند قيمة معينة (X) .

ويمكن اثبات ذلك من خلال الخطوات التالية:

$$g(z) = Ln\left[\frac{\frac{e^z}{1+e^z}}{1+(\frac{e^z}{1+e^z})}\right]$$

$$g(z) = Ln\left[\frac{\frac{e^z}{1+e^z}}{\frac{1+e^z-e^z}{1+e^z}}\right]$$

$$g(z) = \text{Ln}\left[\frac{\frac{e^z}{1+e^z}}{\frac{1}{1+e^z}}\right]$$

$$g(z) = Ln\left[\frac{e^z}{1+e^z} * \frac{1+e^z}{1}\right]$$

$$g(z) = \text{Ln}[e^z] \rightarrow g(z) = z$$

$$z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_K X_K$$

$$g(z) = Ln\left[\frac{f(z)}{1-f(z)}\right] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_K X_K$$

او يمكن كتابته بالصيغة الاتية **[14]:**

$$\boldsymbol{\pi_i} = \frac{e^{\beta + \Sigma_{j=1}^{p} \beta_j X_{ij}}}{1 + e^{\beta_0 + \Sigma_{j=1}^{p} \beta_{ij}}}$$

$$(\mathbf{1} - \boldsymbol{\pi_i}) = \frac{\mathbf{1}}{e^{\beta_0 + \Sigma_{j=1}^{p} \beta_{ij}}} \quad (8)$$

اذ ان :

$\beta_i (\text{j} = 0,1, \ldots \text{p})$ معلمات مجهولة يتم تقديرها

$X_i (\text{j} = 1,2, \ldots \text{b})$: متغيرات مستقلة(توضيحية)

n: عدد المشاهدات

ويمكن تحويل دالة اللوجستك الى دالة خطية الذي اقترحها (Berkson) عام 1944 وحسب الصيغة الاتية:

$$\frac{\pi_i}{(1-\pi_i)} = \frac{p(w=1)}{1-p(w=1)} = e^{\beta_0 + \Sigma_{j=1}^{p} \beta_j X_{ij}}$$

$$z_i = \text{Ln}\frac{\pi_i}{(1-\pi_i)} = \beta_0 + \Sigma_{j=1}^{p} \beta_j X_{ij} \quad (9)$$

اذ ان :

(z_i) تمثل العلاقة الخطية الناتجة من اخذ اللوغارتيم الطبيعي لـ ($\frac{\pi_i}{1-\pi_i}$) والذي يتبع التوزيع الطبيعي بمتوسط ($X_i\beta$)وتباين $[n_i\pi_i(1-\pi_i)]^{-1}$ اي ان

$$Z_i \sim (X_i\beta), [n_i\pi_i(1-\pi_i)]^{-1}$$

لتمثيل علاقة خطية بين متغيرات توضيحية ودالة اللوجت Logit π_i والتي يمكن ان تمثل بالرسم [13].

5. السرطان Cancer:

تعريف السرطان هو مرض ينجم عن حدوث طفرات في المادة الوراثية "للحمض النووي DNA"، مما تؤدي إلى تحول الخلية إلى خلية غير طبيعية تنقسم باستمرار ولا تموت وتنتقل وتغزو المناطق المجاورة والبعيدة ويصيب السرطان الأطفال والبالغين ولكن سرطان الأطفال يختلف بشكل عام عن سرطان البالغين في ثلاثة أمور أساسية:

1. يحدث سرطان الأطفال في كثير من الأحيان فجأة.
2. لا يسبق ظهور المرض أعراض مبكرة.
3. معدل الشفاء من سرطان الأطفال مرتفع، وهذه نقطة مشروطة بمدى توفر رعاية صحية، ولذلك فإن نسبة نجاة الأطفال في الدول المتقدمة تبلغ 80%، ولكنها تنخفض بشكل حاد إلى 20% أو حتى 10% في الدول النامية والفقيرة.

ويلعب التشخيص المبكر للعلاج وتلقي العلاج الملائم دورا كبيرا في زيادة احتمال الشفاء ومعدل النجاة، وقد شهدت السنوات الماضية قفزة كبيرة في علاج سرطانات الأطفال، ويظهر هذا في ارتفاع معدلات الشفاء والنجاة من الموت لدى الأطفال المصابين بالسرطان في دول متقدمة. [2]

6. التلوث Pollution :[29]

هو وجود مواد غريبة على البيئة بحيث تجعلها غير قابلةً للاستعمال وتكون هذه مواد إمّا مشعة أو طاقة أو مايكروبات تؤذي الإنسان والكائنات الحية والعالم وتؤدّي به إلى الهلاك.

7. أنواع التلوث:

1. **التلوث المائي:** ينتج بسبب الأنشطة التي يقوم بها الإنسان في المسطحات المائية، ومن أخطر نوع التلوّث تسرب النفط إلى البحار ممّا يؤذي ,(الكائنات البحريّة) ويؤدّي إلى نفوقها وهلاكها.
2. **التلوث الجوي:** يحدث بسبب تلوّث الهواء الجوي بعوادم السيارات، وأبخرة المصانع، وغيرها، ويعدّ من الاخطر أنواع التلوث
3. **التلوث الكيمائي:** يحدث بسبب التجارب العلمية، الكيمائية، المفاعلات النووية، التجارب الذرية التي تؤدي إلى تلوث الغلاف الجوي وتنتج عنها أمراض مزمنة وخطرة مثل(السرطان)
4. **التلوّث البيولوجي:** ويكون في الدول النامية بسبب الجهل والتخلّف العلمي والتكنولوجي وزيادة السكان وزيادة المخلّفات ممّا يؤدي إلى الإصابة بأمراض طفيلية وأوبئة وأمراض معدية
5. **التلوث الضوضائي أو السمعي:** ويعدّ من انواع التلوث التي تؤذي الإنسان بسبب ارتفاع الأصوات وحدّتها لأوقات طويلة ومتواصلة مما يؤدي إلى إصابة الإنسان بالأمراض العصبية والنفسية.
6. **تلوّث التربة:** أهداف المحافظة على البيئة من التلوث، تقليل الاستنزاف من موارد الطبيعية، معالجة التلوث الناتج عن الأنشطة البشرية وبخاصة الصناعية منها، رفع الإنتاج الزراعي، فعندما تزداد إنتاجية الزراعة والمساحات الخضراء زادت النّظافة وجماليّة الأشياء، خلق الوعي البيئي بين الأجيال، تبادل الخبرات مع الدول المتقدّمة، استعمال المصادر البديلة للطاقة كالشمس والماء والرياح.
7. **تلوث الهواء** من أهم أسباب تلوث الهواء هي العمليات الطبيعية وإزالة الغابات محطات توليد الطاقة والصناعة وانبعاثات المركبات وحرق الأخشاب

أبرز النقاط التي تتسبب بتلوث بيئة محافظة البصرة، منها [1]

1. وجود (18) موقعا للمواد المشعة لم تعالج من قبل الجهات المعنية حتى الآن.
2. يوجد مليار وثلاثمئة مليون متر مربع، غير مطهر من المخلفات الحربية وهذا مثبت بالأرقام والوثائق.
3. الغاز المصاحب لعملية الاستكشاف النفطي، فمع ازدياد الكميات المصدرة من النفط ولم تراعى المحددات البيئية للسلامة، مما انعكس سلبا على أبناء محافظة البصرة، وهذا مثبت بتقارير لكل المواقع النفطية.
4. الغوارق في شط العرب بسبب الحرب مع إيران وحرب الخليج، وكذلك أيضاً التراكمات للإهمال الوزاري بهذا الجانب.
5. تلوث الأنهار وخصوصاً مركز محافظة البصرة، ووجود (7000) نهر تلوث أغلبها بسبب استخدام هذه الأراضي كمكب للنفايات.

8. تصنيف البيانات Classification of data

جمعت البيانات الخاصة بموضوع البحث من, (مستشفى الطفل التخصصي في البصرة) ، وقد تم الاستعانة بمجموعة من المختصين بأمراض السرطان من اجل الوقوف على اهم العوامل المؤثرة على المرض ، اذ تم جمع البيانات الخاصة لسنة 2019 وبعينة حجمها (295) طفلا، ولغرض تحليل البيانات فقد تم اعتبار متغير الاستجابة (y) هو $(y = 0)$ غير مصاب، $(y = 1)$ اصابة خفيفة بالمرض $(y = 2)$ اصابة شديدة بالمرض بالاعتماد على المتغيرات التوضيحية.

جدول (1): يمثل اصابة من عدمها

مرض سرطان اطفال	التكرارات	النسبة
غير مصاب	61	20.7%
اصابة خفيفة	135	45.8%
اصابة شديدة	99	33.6%
المجموع	250	100.0%

9. اختبار وجود مشكلة التعدد الخطي

لاختبار الفرضية التي تنص على عدم وجود مشكلة التعدد الخطي يتم استعمال معامل تضخيم التباين VIF وكما موضح في الجدول ادناه

جدول (2): يمثل اختبار وجود مشكلة التعدد الخطي

المتغير	الجنس	العمر	الموقع	مستوى تعليم الاب	نوع السكن	المستوى المعاشي	مهنة الاب	الوراثة	القرب من مصفى	القرب من اماكن السكراب	العلاج
VIF	1.03	1.11	1.22	1.4	1.07	1.26	1.28	1.30	1.51	1.45	1.11

من الجدول رقم (2) يوضح الاختبار لفرضيات الارتباطات بين المتغيرات المستقلة وهو مهم جداً ومن نتائج هذا الجدول عن طريق معامل التضخم VIF حيث ان جميع المتغيرات اقل من 5 فلا توجد مشكلة تعدد خطي

10. تقدير انموذج الانحدار اللوجستي الرتبي

لغرض تقدير هذا الانموذج نستعمل طريقة الامكان الاعظم المذكورة في الجانب النظري وعند التقدير تكون لدينا معادلتان كل معادلة خاصة بمستوى معين من مستويات متغير الاستجابة الثلاثة اي ان عدد المعادلات تكون (عدد المستويات -1) وبالنتيجة

هناك معادلتين تختلف عن هذه المعادلات بنقاط القطع مع المحور الرأسي فقط وان تقديرات المعلمات تبقى ثابتة لكل معادلة وتكون تقديرات المعلمات واختبار معنوياتها كما موضحة في الجدول (3) .

جدول (3): تقديرات المعلمات لانموذج الانحدار اللوجستي الرتبي واختبار معنوية كل معلمة

Effect	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	482.910[a]	.000	0	.
جنس الطفل	484.227	1.317	2	.518
موقع سكن الطفل	483.673	.763	2	.683
مستوى تعليم الاب	486.961	4.051	2	.132
المستوى المعاشي لأهل الطفل	488.292	5.382	2	.068
مهنة الاب	491.897	8.987	2	.011
الوراثة	485.310	2.400	2	.301
نفط او حقل القرب من مصفى	498.959	16.049	2	.000
القرب من اماكن الحديد الملوثة	483.288	.378	2	.828
العلاج	498.313	15.403	2	.000
نوع السكن	483.890	.981	2	.612
بالسنوات عمر الطفل	515.839	32.929	28	.238

يتضح من الجدول اعلاه ومن خلال مقارنة مستوى المعنوية لكل معلمة مع قيمة (α=0.05) وكذلك القيمة المحسوبة قيمة (wald) مع القيمة الجدولية ان المعلمات المقدرة ذات مستوى المعنوية اقل من 0.05 والقيمة المحسوبة اكبر هي معلمات معنوية اي انها مؤثرة في الانموذج .

ونلاحظ ان المعلمات المعنوية هي الخاصة بالمتغيرات التوضيحية (مهنة الاب *X6* ,القرب من حقول النفط *X8* ,*X10* توفر العلاج) حيث يكون تأثيرها على متغير الاستجابة معنوي. اذ تشير المعلمة المقدرة الى ان المتغير (مهنة الاب *X6*) لكل وحدة زيادة في هذا المتغير، نتوقع نقصان قدره 0.618. في لوغاريتم نسبة الأرجحية الخاصة بالمتغير التابع (الاصابة بأمراض السرطان) وعليه يتم التفسير انه كلما كان وظيفة الاب جيدة انخفضت أرجحية الاصابة بأمراض السرطان بمقدار 0.618. اما ما يخص المتغير الاخر(القرب من حقول النفط *X8*) لكل وحدة زيادة في هذا المتغير ، نتوقع زيادة قدرها 0.445 في لوغاريتم نسبة الأرجحية الخاصة بالمتغير التابع (الاصابة بأمراض السرطان) انها كلما ازداد القرب من حقول النفط في م بوحدة واحدة ازدادت أرجحية الاصابة بأمراض السرطان بمقدار 0.445. اما ما يخص المتغير الاخر(توفر العلاج *X10*) لكل وحدة نقصان في هذا المتغير، نتوقع زيادة قدره 0.145 في لوغاريتم نسبة الأرجحية الخاصة بالمتغير التابع (الاصابة بأمراض السرطان) اما عند المقارنة من خلال نسبة الارجحية تلاحظ انه كلما يقل العلاج بمقدار بوحدة واحدة تزداد أرجحية الاصابة بأمراض السرطان بمقدار 0.145.

11. اختبار معنوية انموذج الانحدار اللوجستي الرتبي

لمعرفة مدى معنوية نموذج الانحدار اللوجستي الرتبي ككل نستعمل اختبار chi-square الموضح في الجدول (4).

جدول (4): يوضح اختبار معنوية الانموذج

Model	chi-square	df	sig	chi-square table
Final	91.363	48	0.000	25.123

وبمقارنة قيمة مربع كاي المحسوبة مع القيمة الجدولية نلاحظ ان القيمة المحسوبة اكبر من الجدولية وعليه نرفض فرضية العدم وتقبل الفرضية البديلة اي ان الانموذج ككل هو انموذج معنوي .

12. اختبار جودة توفيق الانموذج

لاختبار جودة توفيق الانموذج نستعمل اختبار Pearson اذ يعتمد على إحصاءه chi-square لاختبار الفرضية الاتية:

H_o : البيانات مطابقة لأنموذج اللوجستي الرتبي

H_1 : البيانات غير مطابقة لأنموذج اللوجستي الرتبي

وبمقارنة قيمة X^2 المحسوبة مع قيمة X^2 الجدولية بدرجة حرية k - (I-1)(j-1) حيث ان (I) يمثل عدد الصفوف ,(j) يمثل عدد مستويات متغير الاستجابة و (k)يمثل عدد المتغيرات عند مستوى معنوية 0.05 اذ يتم قبول او رفض فرضية العدم من خلال مقارنة قيمة X^2 المحسوبة مع الجدولية فاذا كانت اكبر نرفض فرضية العدم والعكس صحيح .

اما بالنسبة لاختبار Deviance لجودة التوفيق حيث يمثل الاختبار نسبة الارجحية بين الانموذج الموفق عندما تأخذ التوزيع اللوجستي وبين الانموذج الحر free distribution اي عندما تأخذ قيمة حرة بين الصفر والواحد وعلى هذا الاساس تكون فرضية كالاتي [5]:

H_o : p تتوزع حسب التوزيع اللوجستي

H_1 : p لتتوزع بشكل حر بين الصفر والواحد

وبمقارنة قيمة X^2 المحسوبة مع الجدولية يتم قبول او رفض فرضية العدم حسب ما تم ذكره سابقا والجدول رقم (5) يوضح الاختبارات اعلاه

جدول (5): يوضح اختبار جودة توفيق الانموذج باستعمال اختباري (pearson, deviance)

test	Chi-square	sig	df	Chi-square table
pearson	603.486	0.13.	528	803.362
deviance	480.137	0.933	528	803.362

وبمقارنة قيمة كاي سكوير المحسوبة مع الجدولية لاختبار person بدرجة حرية 528 ومستوى معنوية 0.05 نلاحظ ان قيمة كاي سكوير محسوبة اصغر مما يؤدي الى قبول فرضية العدم مما يؤكد ان البيانات مطابقة لأنموذج اللوجستي الرتبي كذلك الحال بالنسبة لاختبار deviance فان كاي محسوبة اصغر مما يودي الى قبول فرضية العدم والتي تشير الى ان قيم الاحتمالات P تتوزع وفق التوزيع اللوجستي .

مما تقدم ومن خلال الاختبارين اعلاه يكون انموذج الانحدار اللوجستي الرتبي ذات جودة توفيق ومطابقة للبيانات قيد الدراسة .

13. تفسير المتغيرات التوضيحية الداخلة في الانموذج

لقياس نسبة تفسير المتغيرات التوضيحية من متغير الاستجابة الرتبي في انموذج الانحدار اللوجستي الرتبي نستعمل معامل Cox and Snell و Nagelkerke ومعامل McFadden كما موضح في الجدول (6)

جدول (6): يوضح تفسير المتغيرات التوضيحية الداخلة في الانموذج

Pseudo R-square	
Cox and Snell	.266
Nagelkerke	.310
McFadden	.158

جدول (6) يوضح ان المتغيرات الداخلة في الانموذج انها قد فسرت.26.6 % باستخدام معامل (R2 Cox & Snell) وهذا يشير الى ان المتغيرات التوضيحية تفسر نسبة 26.6% من التغير في الترجيح للتعرض للإصابة بأحد امراض السرطان (المتغير التابع) و31% باستخدام (Nagelkerke) والذي يفسر ان المتغيرات التوضيحية تفسر 31% من في حين ان مقياس McFadden بين ان المتغيرات التوضيحية تفسر 15.8% من التغير الحاصل في ترجيح متغير الاستجابة الرتبي للإصابة، وان افضل معامل تحديد يمكن الاعتماد عليه هو (Nagelkerke) وكذلك يدل على انه ما تزال هناك نسبة من التغيرات في المتغير التابع تعود لمتغيرات اخرى غير مدرجة في الانموذج[10} .

14. اختبار الخطوط المتوازية

في انموذج الانحدار اللوجستي الرتبي واحد من اهم الافتراضات الخاصة به هو ان قيم المعلمات يجب ان لا تتغير باختلاف الفئات او المستويات لمتغير الاستجابة ويمكن اختبار الفرض بوضع الفرضية الاتية .

H_o: $B1j=B2j=\ldots\ldots=B7j$

H_o: $B1j=B2j=\ldots\ldots=B7j$

لاختبار هذه الفرضية نستعمل اختبار chi-square بدرجة حرية $k(j-2)$ حيث ان k يمثل عدد المتغيرات التوضيحية ، j يمثل عدد مستويات متغير الاستجابة والجدول (7) يوضح هذا الاختبار **[9]**

جدول (7): يوضح اختبار الخطوط المتوازية

Model	Chi-square	df	sig	Chi-square table
Null Hypothesis				
General	18.34	24	0.78	26.29

وعند مقارنة قيمة chi-square المحسوبة (18.34) مع القيمة الجدولية بدرجة حرية.(26.29) نلاحظ ان القيمة المحسوبة اصغر من الجدولية وهذا يعني قبول فرضية العد وتحقق الفرض القائل ان قيم المعلمات لا تتغير باختلاف مستويات متغير الاستجابة .

15. اختبار كفاءة تصنيف الانموذج

تم اختبار كفاءة تصنيف الانموذج والذي يعد احد طرق فحص جودة مطابقة الانموذج للبيانات، ظهرت النتائج كما هو موضح في الجدول رقم (8).

جدول (8): كفاءة تصنيف الانموذج في الخطوة Classification Table 5

المشاهدات observed	غير مصاب	اصابة خفيفة	اصابة شديدة	النسبة المصححة Percent correct
غير مصاب	20	10	6	16.7%
اصابة خفيفة	36	85	1	69.7%
اصابة شديدة	107	28	2	78.1%
Overall percentage	3.1%	41.7%	55.3%	67.1%

ويتضح من الجدول رقم (8) ان الانموذج عند ادخال متغيرات المتنبئة قد حقق نسبة تصنيف كلية صحيحة وهي عبارة عن عدد التنبؤات الصحيحة على العدد الكلي لعينة الدراسة بلغت 67.1% وهي نسبة مرتفعة .[11].

$$Logit = Log(odds) = Log\left(\frac{p}{1-p}\right) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_k x_k$$

$$Log(odds) = 23.435 + 0.488X6 + 0.0.445X8 - 0.140\ X10 \qquad (10)$$

16. الاستنتاجات Conclusions

على ضوء ما تم التوصل اليه من نتائج في الجانبين النظري والتطبيقي يمكن استخلاص اهم الاستنتاجات وكالآتي:

1. كفاءة الانموذج المستنتج بشكل كلي في التنبؤ بالإصابة بمرض السرطان لدى الاطفال دون سن الخامسة عشر.
2. ان العوامل ذات التأثير المعنوي على الاصابة بالمرض كانت الاتية : مهنة الاب مما يشير ان الاشخاص ذات الدخول الجيدة او لديهم مصدر رزق ثابت يساعدهم في ايجاد وسائل تحول دون اصابة اولادهم واهمهما السكن وغيرها من الظروف المعاشية.
3. وكذلك الحال بالنسبة للعامل المؤثر الاخر وهو حالة توفر العلاج كون توفر العلاجات يحول دون وقوع المرض وتقلل من احتمال الاصابة او تطورها.
4. ان السكن بالقرب من مصافي النفط هو عامل اخر ساعد في تعرض الاطفال للإصابة بالمرض وكما ذكر سابقا ان ضعف الحالة المعاشية دفع هذه الاسر للعيش بقرب المصافي وغيرها من المواقع الموبوءة.
5. نجحت المتغيرات المستقلة (عوامل المؤثرة) الداخلة في الانموذج في تفسير ما نسبته 31% من التغير الحاصل في المتغير التابع (الاصابة بالمرض من عدمها) وهذا ما اكدته قيمة R2 المستخلصة، ولكن وبحسب النتائج، فان هنالك عوامل اخرى لم تدرج في الانموذج لها تأثير ايضا على الاصابة بالمرض، وهذه العوامل عوامل قد تعذر دراستها لأسباب عديدة لم تشخص او مؤثرات اخرى قد تكون طارئة الحدوث.
6. لوحظ ان هنالك بعض العوامل غير مؤثرة في حدوث المرض مثل الشهادة، نوع الجنس من عدمه والعمر.

17. التوصيات The Recommendations

من خلال الاستنتاجات التي تم التوصل اليها سيتم استعراض المجالات المختلفة للإفادة من البحث قيد الدراسة :

1. بالإمكان الاستفادة من توظيف انحدار اللوجستي الرتبي كأسلوب فاعل, في التنبؤ بالإصابة بأمراض السرطان وغيرها من الامراض لإعداد خطط ودراسات مستقبلية تساعد في تقليل الاصابة بهذا المرض وغيره من الامراض عبر تحديد اهم العوامل المؤثرة ومحاولة معالجتها.
2. توسعة البحث الحالي ليضم عوامل اخرى تمت الاشارة اليها في الاستنتاجات، ولم يتناولها البحث في الدراسة الحالية ليكون لدى متخذ القرار رؤية اكثر وضوحاً.
3. اهمية توفير البيانات الوافية لكل مريض، وذلك من خلال اعداد (طبلة) منتظمة تحتوي على المعلومات الخاصة بالمريض كافة وبشكل ميسر لأغراض البحث العلمي، ليتسنى للمختصين توظيف الاساليب الاحصائية المناسبة للوصول الى نتائج تفيد الجهات المعنية في بناء استراتيجياتهم المستقبلية للوقاية من المرض او تقليل الاصابة به.
4. ضرورة قيام الجهات المختصة بتوعية المراجعين حول المخاطر الاصابة بالمرض عن طريق اعلانات ورقية وتوعوية وتلفزيونية ودفع الاهالي الى متابعة اطفالهم والاماكن التي يرتادونها.
5. التنسيق ضمن المحافظة على ايجاد حلول للعوائل الساكنة بالقرب من المصافي وايجاد سكن لهم بعيد من مناطق التلوث.
6. التنسيق ما بين المنظمات الانسانية ووزارة العمل لمساعدة العوائل المتضررة وذات الدخل المتدني.
7. استعمال انظمة البرمجيات الحديثة عبر تطوير قاعدة جمع البيانات لغرض تبويب وارشفة البيانات الطبية، التي تخص المرضى في المؤسسات الصحية. بهدف الحصول على بيانات واقعية وفي غاية الدقة يتم الاعتماد عليها للحصول على نتائج جيدة.
8. توفير العلاجات المناسبة للحيلولة دون تطور العرب لحالات دون وقوعه او تطروه في الشدة

المصادر

[1] الثعلبي، ساهرة حسين، (2008)، "تحليل البيانات الثنائية لدراسة العوامل المؤثرة في حدوث التشوهات الولادية في مستشفى البصرة"، رسالة ماجستير في الاحصاء، كلية الادارة والاقتصاد، جامعة بغداد.

[2] سعيد، رشا عادل (2015)، "استخدام انموذج انحدار اللوجستي في دراسة العوامل المساعدة على تشخيص حالات الاصابة بسرطان المثانة "، مجلة العلوم الاقتصادية والادارية، المجلد 21 العدد 83، 345-357 .

[3] يس، امل حسن محمد، (2008)، "استخدام الانحدار اللوجستي لمعرفة مسببات الإصابة بسرطان الثدي عند الإناث"، رسالة ماجستير، جامعة السودان للعلوم والتكنولوجيا.

[4] Adeleke, K.A, Adeboju, A. (2010) "Ordinal Logistic Regression Model : An Application to Pregnancy Outcomes" Journal of Mathematics and Statistics, 6(3) PP 279-285.

[5] Cramer, J.S. (2002). " The origins of Logistic Regression", Tinbergen Institute Working paper No. 2022-119/4, University of Amsterdam .

[6] Das, S, Rahman, R.M (2011) "Application of Ordinal Logistic Regression Analysis in Determining Risk Factors of Child Malnutrition in Bangladesh" Das & Rahman Nutrition Journal 10,124. PP 1-11.

[7] Hosmer . D. W، and Sturdivant. R.X(2013), Applied Logistic Regression, 3rd edition, John Wiley & Sons, Inc.

[8] Irfan a , P. and Hemant, D. (2006), "A Logistic Regression Analysis of the Ischemic heart disease risk", Journal of Applied Sciences. Vol. (6), No. (4). Pp. 785-788.

[9] Juan, M.J. Puchades. (2010), "Evolution of kidney function and progression factors in nephrictomised patients", Nefrologia, Vol. (30), No. (10).

[10] Menard, S. (2002), Applied logistic Regression Analysis"2nd edition, SAGE Publications, Inc.

[11] Mohammed Moniru jjaman Khan. (2021),"Comparative Analysis for Prediction of Kidey Disease Using Intelligent Machine Learning Methods", Computational and Mathematical Methods in Medicine, Volume 2021, Article ID: 6141470.

[12] Muhammad Abdul Mabood Khalil.(2018)." Factors Affecting Inpatient Mortality in Elderly People with Acute Kidney Injury". The Scientific World Journal, Vol. 2018, doi: 10.1155/2018/2142519

[13] Reddy, O.C and Alemayehu, E (2015) "Ordinal Logistic Regression Analysis to Assess the Factors that Affect Health the Status of Students in Ambo University : A case of Natural and Computational Science College" International Journal of Modern Chemistry and Applied Science, 2(3), PP 153-163.

Journal of AL-Rafidain University College for Sciences (2023); Issue 54; 309 – 318

AL- Rafidain University College

PISSN: (1681-6870); EISSN: (2790-2293)

Journal of AL-Rafidain University College for Sciences

Available online at: https://www.jrucs.iq

JRUCS

Journal of AL-Rafidain University College for Sciences

Using the Multiple Ordinal Logistic Regression Model to Identify the Most Important Demographic and Environmental Factors Affecting the Incidence of Cancer Among Children in Basra Governorate

Shereen A. Hussain	Dr. Saad K. Hamza
shreen.a@uobaghdad.edu.iq	saad.hamza@coadec.uobaghdad.edu.iq
Department of Statistics, College of Administration and Economics, University of Baghdad, Baghdad, Iraq	Department of Quality Assurance and University Performance, University of Baghdad, Baghdad, Iraq

Article Information

Article History:
Received: January, 11, 2023
Accepted: March, 3, 2023
Available Online: December, 31, 2023

Keywords:
Regression models, multiple logistic regression, demographic and environmental factors, childhood cancers

Abstract

High flexibility characterizes logistic regression, which is regarded as one of the significant statistical techniques for estimating the probability of an event occurring by matching data and information on the logistic curve, as it usually works to describe the nature of the relationship between the two response variables, which are usually categorical or ordinal and a group of independent variables. Because of its nominal nature, the logistic regression model expresses a special case of the normal regression model. Depending on the response variable, there are several types of logistic regression, including binary logistic regression, which is typically used when the response variable has two values (0,1), as well as multi-response (multinomial logistic) when the response variable has more than two categories, as for ordinal logistic regression, as it is used in the case of the response variable, it is ordinal. This study used rank- order logistic regression models to identify the most significant environmental and demographic factors affecting the incidence of cancer in children in the Basra Governorate. The response variable was the severity of the injury, starting from no injury to a severe injury. We arrived at the variables that affect the determination of infection in a child at the 2019 Children's Specialist Hospital in Basra. These variables include sex, age, residence location, father's occupation, father's educational attainment, standard of living, father's type of work, heredity, housing location in relation to oil fields, proximity to radiation-contaminated areas, cancer type, treatment, type of housing, and duration of injury. The researcher reached the conclusion that there are several variables that have a direct impact on the incidence of cancer, including the location of residence, particularly if it is close to oil wells and refineries, the father's work, particularly if the family head is employed, and the availability of treatment, where the significance was lower than 0.05, which has a significant effect on the incidence of the disease.

Correspondence:
Shereen A. Hussain
shreen.a@uobaghdad.edu.iq

https://doi.org/10.55562/jrucs.v54i1.601