IRAQI Academic Scientific Journals

University of Kirkuk Journal For Administrative and Economic Science (2025) 15 (4) Part (1): 1-15

UKJAES

University of Kirkuk Journal For Administrative and Economic Science

ISSN:2222-2995 E-ISSN:3079-3521
University of Kirkuk Journal For Administrative and Economic Science

Rashid Amal Hadi. Addressing multicollinearity in logistic regression using partial least squares (An applied study on patients with viral hepatitis). *University of Kirkuk Journal For Administrative and Economic Science* (2025) 15 (4) Part (1):1-15.

Addressing multicollinearity in logistic regression using partial least squares (An applied study on patients with viral hepatitis)

Amal Hadi Rashid [1]

[1] *Department of Statistics-College of Administration and Economics-University of Diyala, Ba'aqubah-Diyala, Iraq*

amal@uodiyala.edu.iq [1]

Abstract: The logistic regression model is one of the nonlinear regression models aimed at obtaining estimators with high efficiency, which takes a more advanced form in the process of statistical analysis as it is suitable for binary data. Real data was used, and data was collected on viral hepatitis, where there are two types of viruses (c) and (b). A sample consisting of 119 individuals infected with this disease was taken during the year 2023, and the data was collected from Diyala Health Hospital and Khalis Health Hospital in Diyala Governorate for the year 2023, with two variables: (0) type c virus and (1) type b virus.

In addressing the problem of multicollinearity in the data that arises when using some estimation methods, where there are some observations that deviate significantly and exhibit high correlations between variables, it becomes difficult to use the logistic regression model unless this issue is addressed. Therefore, the method of partial least squares regression was resorted to using the (NIPALS (PLS1)) algorithm and the ridge regression method to reduce the high correlations between the variables used and address this problem, allowing for the use of the logistic regression model, which is considered one of the most important statistical models used in classification. The study concluded that the gender variable does not have a significant effect on the incidence of the disease, while the age variable and the LDL variable have a very significant effect on the incidence of viral hepatitis.

Keywords: logistic regression, multicollinearity problem, partial least squares method, (NIPALS (PLS1)) algorithm, ridge regression, viral hepatitis.

معالجة التعدد الخطي في الانحدار اللوجستي باستخدام طريقة المربعات الصغرى الجزئية دراسة تطبيقية على المرضى المصابين بالتهاب الكبد الفيروسي

م.م. أمل هادي رشيد [١]

[١] *جامعة ديالى-كلية الإدارة والاقتصاد-قسم الاحصاء، بعقوبة-ديالى، العراق*

المستخلص: يعد انموذج الانحدار اللوجستي من النماذج الانحدار اللاخطية الذي يهدف للحصول على مقدرات تمتلك كفاءة عالية، والذي يأخذ طابعا اكثر تقدم في عملية التحليل الاحصائي لكونه من النماذج الملائمة للبيانات الثنائية. حيث تم استعمال بيانات حقيقة وتم جمع بيانات عن مرض التهاب الكبد الفيروسي حيث يوجد نوعين من الفيروسات(c) والنوع الثاني (b) اذا اخذ عينة مكونة من (١١٩) شخص مصاب بهذا المرض خلال سنة ٢٠٢٣ وتم اخذ البيانات من مستشفى صحة ديالى ومستشفى صحة الخالص في محافظة ديالى لسنة ٢٠٢٣، بمتغيرين هما (٠) نوع الفايروس c و(١) نوع الفايروس b.

وفي معالجة مشكلة التعدد الخطيبين البيانات التي تظهر عند استعمال بعض طرائق التقدير اي وجود بعض المشاهدات التي تنحرف بشكل ملحوظ ويوجد فيها ارتباطات عالية بين المتغيرات ويصبح من الصعب استعمال انموذج الانحدار اللوجستي الا بعد معالجة هذه المشكلة فقد تم اللجوء الى طريقة انحدار المربعات الصغرى الجزئية باستعمال خوارزمية (NIPALS(PLS1)) وطريقة العبور الشراعية وذلك لتقليل الارتباطات العالية بين المتغيرات المستخدمة ومعالجة هذه المشكلة واستعمال انموذج الانحدار اللوجستي الذي يعد من اهم النماذج الاحصائية المستخدمة بالتصنيف وقد توصلت الدراسة الى ان متغير الجنس ليه له تاثير معنوي على الاصابة بالمرض وان متغير العمر ومتغير LDL لهما تاثير كبير جدا على الاصابة بمرض الكبد الفيروسي .

الكلمات المفتاحية: الانحدار اللوجستي، مشكلة التعدد الخطي، طريقة المربعات الجزئية، خوارزمية (NIPALS(PLS1))، العبور الشراعي، التهاب الكبد الفيروسي.

Corresponding Author: E-mail: amal@uodiyala.edu.iq

المقدمة

تنوعت الدراسات والبحوث في اظهار اهمية الاساليب الاحصائية المستعملة في تحليل البيانات ضمن الظواهر المدروسة، ويعد تحليل الانحدار اللوجستي واحد من اهم الاساليب الاحصائية المستعملة في التحليلات ، اذا ان الانحدار بمجمله يهدف الى وصف وتحديد العلاقة من خلال صياغة معادلة رياضية تربط جميع تلك المتغيرات ،اذ تعد طرائق تحليل الانحدار الجزء الاساس في تحليل البيانات وتمت دراسة تحليل الانحدار اللوجستي من خلال تعريف نظريات الانحدار في الاحصاء وهو انموذج إحصائي ينتمي لنماذج الانحدار اللاخطي يحلل ويفسر العلاقة الموجودة بين المتغير التابع (y) والمتغير التوضيحي (x) ،من خلال نمذجة متغير ثنائي بدلالة مجموعة من المتغيرات العشوائية المتوقعة، رقمية كانت أو فئوية، ويستعمل الانحدار اللوجستي للتنبؤ باحتمالية وقوع حدث ما بمعرفة إضافية لقيم متغيرات يمكن أن تكون مفسرة أو مرتبطة بهذا الحدث، وقد تم ربط تحليل الانحدار اللوجستي ببيانات التهاب الكبد الفيروسي التي والتي كانت تعاني من مشكلة التعدد الخطي.

المبحث الأول: الجانبي النظري

اولاً: مشكلة البحث

ان مشكلة البحث تكمن في الكشف عن مشكلة التعدد الخطي للبيانات المدروسة الذي يعتمد على احجام العينات لمرضى المصابين بالتهاب الكبد الفيروسي من حيث أن يوجد هنالك عدة أنواع من هذا المرض فبالتالي سوف يعاني من مشكلة التعدد الخطي ولمعالجة هذا المشكلة تم استعمال طريقة المربعات الجزئية لأنموذج الانحدار اللوجستي بعد تقدير معالمه بطريقة دالة الامكان الاعظم واعتماد خوارزمية (NIPALS(PLS1)) وطريقة العبور الشرعي باعتماد عرض حزمة خاصة.

ثانياً: هدف البحث

يهدف البحث الى معرفة وتحديد اهم العوامل الإصابة بالتهاب الكبد الفيروسي باختيار أفضل أنموذج انحدار لوجستي وبأقل عدد ممكن إذ أنه يتبع أسلوب إحصائي ويهدف أيضا عن طريق منظمة الصحة العالمية بشأن التهاب الكبد الفيروسي لعام (٢٠١٧) إلى توفير نقطة انطلاق عن طريق الإشارة إلى إحصاءات ومعرفة اهم العوامل المؤثرة على الإصابة، وذلك باستعمال اسلوب الانحدار اللوجستي من خلال معالجة مشكلة التعدد الخطي باستعمال طريقة المربعات الصغرى الجزئية.

ثالثاً: أنموذج الانحدار اللوجستي Logistic regression model

يعد الانحدار اللوجستي من النماذج الإحصائية المهمة في تحليل البيانات والهدف الأساسي في معظم الدراسات هو التحليل والتقويم العلاقات بين مجموعة من المتغيرات للحصول عليها من صيغة نستطيع من خلالها أن تصف الأنموذج ويستعمل لوصف العلاقة بين متغير الاستجابة من النوع المتقطع والمتغيرات التوضيحية، وهو حالة خاصة من النموذج الخطي المعمم [5].
إذا يتصف بان متغير الاستجابة (Y) يتوزع توزيع برنولي ثنائي يأخذ القيم) صفر، واحد)، أي ان متغير الاستجابة له حالتين الحالة الاولى ووقوع حدث معين (Y=1) والحالة الثانية بعدم وقوع ذلك الحدث (Y=0) وباحتمال قدره (πi) من النجاح واحتمال قدره ($1-\pi i$) من الفشل وبمعنى اخر (١) في حالة حدوث الاستجابة و(٠) في حالة عدم حدوث الاستجابة [3].

$$y_i \sim Ber\pi_i \quad , i = 0,1$$

وان دالة الكثافة الاحتمالية تكون بالشكل الاتي

$$\pi_i^{y_i}(1-\pi_i)^{1-y_i} \quad (1)$$

Yi= متغير تابع ثنائي الاستجابة يأخذ قيم (0,1).
πi=احتمال حدوث الاستجابة عندما (Yi=1).
1-πi= احتمال عدم حدوث الاستجابة عندما (Yi=0).
لذلك فان التوقع والتباين للمتغير (Yi) يمثل حدوث الاستجابة وكما هو موضوح

$$E(Y_i) = p_r(Y=1) = \pi_i, \quad V(Y_I) = \pi_i(1-\pi_i) \quad (2)$$

وليكن ($X_1,X_2,\ldots,X_P$) مجموعة من المتغيرات التوضيحية ولتكن (n) تمثل عدد المشاهدات لهذا المتغيرات التي تكون المصفوفة (x)

$$X = (X_{ij})_{nxp} \ , i,j = 1,2,..,n \quad (3)$$

اذا ان
n: تمثل حجم العينة ، p: تمل عدد المتغيرات التوضيحية.
فان $Yi=[y_1,y_2,..,y_n]$ عينة عشوائية من المتغير ثنائي الاستجابة وبالتالي فان انموذج الانحدار اللوجستي يكون كما في الشكل ادناه [12]

$$Y_i = \pi_i + \varepsilon_i \quad (4)$$

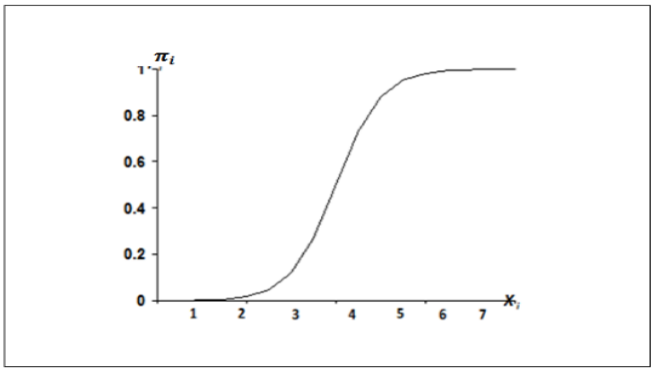
اذا ان (πi) تمثل دالة الانحدار اللوجستي (احتمالية الاستجابة)

$$\pi_i = \frac{e^{(\beta_0+\sum_{j=1}^{p} B_j X_{ij})}}{1+e^{(\beta_0+\sum_{j=1}^{p} B_j X_{ij})}} \quad (5)$$

وان

$$1-\pi_i = \frac{1}{1+e^{(\beta_0+\sum_{j=1}^{p} B_j X_{ij})}} \quad (6)$$

β_0,β_1,β_p : معلمات أنموذج الانحدار اللوجستي ، وان ε_i يمثل الخطأ العشوائي بمتوسط يساوي صفر وتباين $\pi_i(1-\pi_i)$.
ومن معادلة رقم (٥) نلاحظ ان شكل العلاقة بين المتغيرات التوضيحية (X_{ij}) واحتمالية الاستجابة (π_i) وبالتالي فانه لا يمكن ان يكون خطيا وهي تأخذ شكلا منيحا والمخطط في شكل رقم (١) يوضح ذلك



شكل (١): العلاقة بين المتغيرات التوضيحية (X_{ij}) واحتمالية الاستجابة (π_i) [1]

ويمكن تحويل دالة الانحدار اللوجستي الى دالة خطية باستعمال اقتراح الباحث (Berkson) الذي اقترحه في عام ١٩٤٤ وحسب الصيغة الاتية [9]

$$\frac{\pi_i}{1-\pi_i} = e^{(\beta_0+\sum_{j=1}^{p} B_j X_{ij})} \quad (7)$$

وبأخذ اللوغاريتم الطبيعي لكل الطرفين نحصل على

$$z_i = \ln\left(\frac{\pi_i}{1-\pi_i}\right) = (\beta_0 + \sum_{j=1}^{p} B_j X_{ij}) \quad (8)$$

اذا ان z_i تمثل العلاقة الخطية الناتجة من اخذ اللوغاريتم الطبيعي للمعادلة رقم (٧) الذي يتبع التوزيع الطبيعي[7]

$$z_i \sim N\left((X_i,\beta),[,[n_i,\pi_i(1-\pi_i)]^{-1}\right)$$

رابعاً: تقدير معلمات أنموذج الانحدار اللوجستي

يبني الأنموذج على فرض أساسي هو ان المتغير (التابع) (y) متغير الاستجابة الذي تهتم بدراسته وهو متغير ثنائي يتبع توزيع (Bernoulli) يأخذ الرتبة (1)باحتمال (p) والرتبة (صفر) باحتمال(q=1-p) اي حدوث "الاستجابة " او "عدم حدوثه. وتوجد عدة طرق لتقدير معالم انموذج الانحدار اللوجستي منها[13]:

١- طريقة الامكان الاعظم (Maximum Likelihood Estimation Method)

تعتبر طريقة الامكان الاعظم طريقة تكرارية في تقدير معالم انموذج الانحدار اللوجستي $(\hat{B}_i)$ تعتمد على تكار العمليات الحسابية لأكثر من مرة حتى يتم الوصول الى افضل تقدير لمعلمات أنموذج الانحدار اللوجستي والتي من خلالها يتم التحليل والتفسير وباتباع الخطوات الاتية ابتداء من دالة الكثافة الاحتمالية لـ(y) لدالة الامكان الاعظم [10]

$$L(P_i) = \prod_{i=1}^{n} P_i(Y_i = y_i) \qquad (9)$$

$$L(P_i) = \prod_{i=1}^{n} C_{y_i}^{n_i} P_i^{y_i} (1-P)^{n_i - y_i} \qquad (10)$$

اذ ان

n : حجم العينة .

(i=1,2,…,n)

وعند اخذ اللوغاريتم للمعادلة اعلاه نحصل على [13]

$$Ln\ L(P_i) = \prod_{i=1}^{n} [C_{y_i}^{n_i} + y_i LnP_i + n_i Ln(1-P_i) - y_i Ln(1-P_i)] \qquad (11)$$

وبتبسيط المعادلة (١١) نحل على

$$Ln\ L(P_i) = \sum_{i=1}^{n} [LnC_{y_i}^{n_i} + y_i Ln(\frac{P_i}{1-P_i}) + Ln(1-P_i)^{n_i}] \qquad (12)$$

وبما ان تقديرات طريقة الامكان الاعظم هي قيم المعلمات المقدرة التي تعظم دالة الامكان ويمكن ايجادها عن طريق اشتقاق المعادلة اعلاه للمعلمة المراد تقديرها ويتم سماوتها بالصفر لنحلى على المعادلات وبما ان هذه المعادلات تكون غير خطية لذا نلجا الى طريقة نيوتن –رافسون لحلها باتباع الخطوات الاتية [2]

$$\beta_{s+1} = \beta_s - G^{-1} g(s) \qquad (13)$$

حيث إن :

β_{s+1}: يمثل متجه المعلمات المراد تقديرها.

β_s: يمثل متجه القيم الاولية للمعلمات .

$g(s)$: يمثل متجه المشتقة الاولى الى لوغاريتم دالة الامكان.

$$g_{(s)} = \frac{\partial LnL}{\partial \hat{B}_S} = \begin{bmatrix} \frac{\partial Ln}{\partial \beta_0} \\ \frac{\partial Lnl}{\partial \beta_1} \\ \vdots \\ \frac{\partial LnL}{\partial \beta_k} \end{bmatrix} = \begin{bmatrix} \sum (y_i - \hat{y}_i) \\ \sum x_{i1}(y_i - \hat{y}_i) \\ \vdots \\ \sum x_{ik}(y_i - \hat{y}_i) \end{bmatrix}$$

S: يمثل عدد الدورات.

G: يمثل مصفوفة المعلومات لفشر.

$$G = \left[\frac{\partial^2 LnL}{\partial \hat{\beta}_j \partial \hat{\beta}_j}\right] = \begin{bmatrix} -E\frac{LnL}{\partial {\beta_0}^2} & -E\frac{\partial^2 LnL}{\partial \beta_0 \partial \beta_1} & -E\frac{\partial^2 LnL}{\partial \beta_0 \partial \beta_K} \\ -E\frac{\partial^2 LnL}{\partial \beta_0 \partial \beta_1} & -E\frac{LnL}{\partial {\beta_1}^2} & -E\frac{\partial^2 LnL}{\partial \beta_1 \partial \beta_K} \\ -E\frac{\partial^2 LnL}{\partial \beta_0 \partial \beta_K} & -E\frac{\partial^2 LnL}{\partial \beta_1 \partial \beta_K} & -E\frac{\partial^2 LnL}{\partial {\beta_K}^2} \end{bmatrix}$$

$$G^{-1} = (\acute{X}VX)^{-1}$$

ان مصفوفة فشر تحتوي في داخلها على التباين والتباين المشترك اذا ان عناصر القطر الرئيس تمثل التباين وعناصر خارج القطر تمثل التباين المشترك وبالتعويض في معادلة (١٣) نحصل على [10]

$$\hat{\beta}_{s+1} = \hat{\beta}_s + \left[\frac{\partial^2 LnL}{\partial\hat{\beta}_j \partial\hat{\beta}_j}\right]^{-1} \frac{\partial LnL}{\partial\hat{B}_S} \quad (14)$$

ونستمر بتطبيق المعادلة اعلاه لحين ما يكون الفرق بين الدورة السابقة واللاحقة صغير جدا ويقترب من الصفر.

خامساً: مشكلة التعدد الخطي

ان مصطلح التعدد الخطي او الارتباط الخطي المتعدد هو مصطلح مركب من (multi) متعدد و(co) مشترك او متداخل او مرتبط (Linearly) خطي ويشير المصطلح إلى وجود علاقة خطيه تامة او غير تامة بين اثنين أو أكثر من المتغيرات التوضيحية.
حيث يعتبر الانحدار الخطي المتعدد من التقنيات الإحصائية الأكثر استعمالا بين الباحثين في مختلف المجالات، اذ ان مصفوفة (X)التي تحتوي على المتغيرات التوضيحية قد تنتج عند اجراء عملية حسابية عليها قيم مفردة يحدث هذا عندما تكون بعض المجموعات الخطية لاعمدة (x) تكون مساوية تمام للصفر. يصبح الامر أكثر وضوح عندما يتم حساب تحليل المربعات الصغرى، وتحصل مشكلة التعدد الخطي عندما يرتبط متغيران توضيحيان أو أكثر بعالقة خطية قوية، بحيث يصعب فصل تأثير كل متغير توضيحي عن المتغير المعتمد في الواقع التطبيقي، كما تحدث مشكلة التعدد الخطي حينما تكون قيم أحد المتغيرات متساوية لكافة المشاهدات، أو عندما تعتمد قيمة أحد المتغيرات التوضيحية على واحد أو أكثر من المتغيرات التوضيحية الاخرى في النموذج قيد الدراسة [6].
والتعدد الخطي عدة أنواع هي التعدد الخطي (Exact or perfect multicollinearity) ويحدث في حالة وجود ارتباط تام بين متغيرين تفسيريين او أكثر والنوع الثاني التعدد الخطي الغير تام (Non-Exact or perfect multicollinearity)

سادساً: طريقة المربعات الصغرى الجزئية

تعد هذه طريقة اكثر الطرق أهمية في الانحدار فهي تستعمل لتقليص عدد المتغيرات التوضيحية المرتبطة بالأنموذج إلى مركبات غير مرتبطة (خطية، متعامدة) عندما يكون عدد المتغيرات اكثر من المشاهدات في التجربة فهذه الطريقة هي مشابهة لطريقة المركبات الرئيسية وطريقة انحدار الحرف في معالجة مشكلة التعدد الخطي. حيث ان طريقة المركبات الرئيسية تأخذ بنضر الاعتبار التباين بين المتغيرات التوضيحية، وتقوم بتحويل المتغيرات التوضيحية المرتبطة الى مركبات رئيسة تختلف في الحسابات. ففي حالة تساوي عدد المركبات مع عدد المتغيرات التوضيحية فان النتائج ستكون متطابقة مع طريقة المربعات الصغرى فهي تعتمد على خطوتين اساسيتين لإيجاد المتغيرات بين X,Y من خلال تعظيم مصفوفة التباين والتباين المشترك اما الخطوة الثانية فهي انحدار Y على المركبات [6].

سابعاً: انحدار المربعات الصغرى الجزئية

ان طريقة المربعات الصغرى الجزئية تعتمد على خطوتين أساسيتين لإيجاد المتغيرات الكامنة بين X، Y ومن خلال تعظيم مصفوفة التباين والتباين المشترك أما الخطوة الثانية فهي انحدار Yعلى المركبات، فلو فرضنا أن لدينا مصفوفة Xn-p والمتجه Yn-1 فطريقة المربعات الصغرى الجزئية تعتمد على الأنموذج الثنائي بين XوYالاتي وتعد طريقة المربعات الصغرى الجزئية مشابه الى من الطرائق (Canonical Correlation) و(Principal Component) و(Discriminate Analysis) [8]

$$\mathbf{X = T'P + E} \ldots\ldots\ldots (\mathbf{15})$$

$$\mathbf{Y = U'q + F} \ldots\ldots\ldots (\mathbf{16})$$

حيث ان :
T : مصفوفة درجات X — X-score ببعد n*r
U: مصفوفة درجات Y — Y-score ببعد n*r
P : مصفوفة تحميلات X — X-loding ببعد p*r
q: متجه تحميلات Y — X-loading ببعد r *1
E: متجه البواقي X — X-residual ببعد n*p
F : متجه البواقي Y — Y-residual ببعد n*1
والمصفوفة p′والمتجه qله rمن الأعمدة ويكون عدد بما يأتي

$$\mathbf{r < min(n, p)} \ldots\ldots\ldots (\mathbf{17})$$

حيث ان :
P: عدد المتغيرات
n: عدد المشاهدات
T: عدد المركبات
والعلاقة الداخلية تكون كالاتي :

$$U = TD + H \ldots\ldots (18)$$

حيث أن :
D:هي مصفوفة قطرية ذات بعد r*r
H: مصفوفة البواقي ذات بعد n*r
ملخص طريقة المربعات الصغرى الجزئية هي إيجاد wمن مجال X والمتجه C من مجال Y

$$||Y_c|| = 1 \dots \dots \dots \dots (19) \text{Max cov}(X_w, Y_c) \quad \text{with} \quad ||X_w|| = 1$$

حيث ان $\text{cov}(X_w)(Y_c)$هو تقدير التباين والتباين المشترك وتنفذ الطريقة بصورة تكرارية متسلسلة وهذا يعمي إيجاد المتجهات Scoresالواحدة بعد الأخرى حتى يتم استخراج كافة المتجهات إلى rتحت قيد هدم الارتباط بين المتجهات r.
ولحل المعادلة رقم (١٩) نستعمل خوارزمية التكرار غير الخطي للمربعات الصغرى الجزئية (**Non-linear Iterative partial least squares NIPALS(PLS1)**)

خطوات خوارزمية NIPALS(PLS1) [14]

الخطوة الاولى : يتم تهيئة U1 عن طريق $\underline{Y}$ بحيث

$$U_1 = \underline{Y} \quad (20)$$

اذا ان U1 متجه ببعد n*1.
الخطوة الثانية :حساب اوزان (X-Weight)

$$\underline{W_1} = \frac{\acute{X}U_1}{\acute{U}_1 U_1} \quad (21)$$

وان W1 متجه ببعد (p*1).
الخطوة الثالثة: W1 تكون normalized وعلى النحو الاتي :

$$W_1 = \frac{W_1}{\|W_1\|} \quad (22)$$

الخطوة الرابعة: نبدأ بحساب درجات (x- score) وهي اسقاطات للبيانات او المشاهدات X على اوزان (X- weight)

$$t_1 = XW_1 \quad (23)$$

حيث ان t_1 متجه ببعد (n*1).
الخطوة الخامسة : حساب اوزان (y-weight)

$$\underline{C_1} = \frac{\acute{U}t_1}{\acute{t}_1 t_1} \quad (24)$$

حيث ان C_1 متجه ببعد (1*1). ويكون (normalized) على النحو الاتي :

$$C_1 = \frac{C_1}{\|C_1\|} \quad (25)$$

الخطوة السادسة: حساب درجات (y- score) وهي تراكيب خطية لمتغير الاستجابة وهي استقطاعات y على اوزان (Y-weight)

$$U^*{}_1 = YC_1 \quad (26)$$

حيث ان $U^*{}_1$ متجه ببعد (n*1).
الخطوة السابعة : نجد U وعلى النحو الاتي

$$U = U^*{}_1 - U_1 \quad (27)$$
$$\Delta U = (U\Delta)' * (U\Delta) \quad (28)$$

فاذا كانت $\Delta U < \varepsilon$ وجدنا اول مركبة نتوقف حيث ان قيمة ε تكون صغيرة جدا ، عدا ذلك نذهب الى الخطوة الاولى ونستعمل $U^*{}_1$ بدل من U_1 ونتسمر بالخطوات .
الخطوة الثامنة: ايجاد تحميلات (X-loading) وهي معالم خطية تربط المتغيرات التوضيحية X مع درجات (X-score)

$$P_1 = \frac{\acute{X}t_1}{\acute{t}t_1} \quad (29)$$

الخطوة التاسعة: ايجاد تحميلات (Y-loading) وهي معالم خطية تربط متجه الاستجابات الى درجات (Y-Sscore) وهي بأبعاد (1*1) ويتم حسابها من الصيغة التالية:

$$q = \frac{\acute{Y}U_1}{\acute{U}_1 U_1} \quad (30)$$

حيث ان q متجه ببعد (١*١)
الخطوة العاشرة : بعد ذلك يتم ايجاد التداخل الخطي للمعالم بواسطة انحدار (OLS) بالصيغة الاتية :

$$d_1 = \frac{\acute{U}_1 t_1}{t_1 t_2} \quad (31)$$

حيث ان d_1متجه ببعد (١*١).
الخطوة الحادية عشر عمل تفريغ الى بيانات x,y

$$X_1 = X_1 - t_1 \acute{p}_1 \quad (32)$$
$$Y_1 = Y - d_1 t_1 \acute{c}_1 \quad (33)$$

يتم الاستمرار بالخطوة المذكورة مرات عدة باستعمال البيانات المفرغة الى Y,X حتى نحصل على عدد المركبات المحددة، ونستطيع ايجاد معاملات الانحدار من خلال المعادلة الاتية:

$$\beta = W(\acute{P}W)^{-1}C \quad (34)$$

حيث ان W مصفوفة ببعد (p*r) وان Pمصفوفة ببعد(p*r) وان Cمصفوفة ببعد (r*r).

ثامناً: العبور الشرعي cross validation [4]

تعد طريقة العبور الشرعي (التقاطع) من الطرائق المهمة في اختيار قيمة معلمة عرض الحزمة (Bandwidth) اذ يتم تقدير هذه المعلمة من خلال استعمال طريقة (Nadaraya-Watson) مع استبعاد مشاهدة واحدة، وتسمى ايضا (leave-one-out-cross-validation) حيث تؤدي هذه المعلمة دورا في تباين وتحفيز المقدر، ويزداد التباين بنقصان قيمة معلمة عرض الحزمة، ويزداد التحيز بتزايد قيمة هذه المعلمة والتي تكون قيمتها صغيرة جدا وقريبة من الصفر، ويتم عمل هذه الطريقة وفق الشكل الاتي:

الخطوة الاولى: استبعاد قيمة واحدة من قيم المتغير التوضيحي.
الخطوة الثانية: يتم حساب دالة الوزن وفق الصيغة الاتية:

$$W_i(x) = \frac{K(\frac{x - x_i}{h})}{\frac{1}{nh}\sum_{i=1}^{n} K(\frac{x - x_i}{h})} \quad (35)$$

اذ ان$K(\frac{x-x_i}{h})$ دالة (kernel) .

الخطوة الثالثة: نقوم بحساب معيار العبور الشرعي بعد استبعاد قيمة واحدة من قيم المتغير التوضيحي وكما يلي :

$$CV(h) = \frac{1}{n}\sum_{i=1}^{n}\left[Y_i - \hat{f}(y_i, h)\right]^2 W_i(x) \quad (36)$$

حيث ان $\hat{f}(y_i, h$ هي القيمة تقديريه إلى yالحاصل عليها من البيانات بدون المشاهدة المحذوفة Iمن أنموذج المربعات الصغرى الجزئية PLSللمركبة h

الخطوة الرابعة: اعادة الخطوة الاولى والخطوة الثانية والخطوة الثالثة لجميع المشاهدات حيث يتم استبعاد مشاهدة كل مرة .
الخطوة الخامسة: يتم حساب قيمة معلمة التمهيد المثلى (h_{CV}) وذلك على النحو الاتي :

$$h_{CV} = \arg\min CV(h) \quad (37)$$

اذا ان المعادلة اعلاه تمثل معلمة التمهيد (عرض الحزمة) والتي تقابل اصغر قيمة للعبور الشرعي.

تاسعاً: اختبار Wald'

وهو من الاختبارات الاحصائية المهمة الذي يستعمل لبيان أهمية وتأثير كل متغير التوضيحي (x) في المتغير التابع (Y)، ويتم استبعاد المتغيرات التي لا يكون لها تاثير على الانموذج المدروس وتستعمل احصاءة هذا الاختبار لاختبار معنوية كل معلمة من معلمات انموذج الانحدار اللوجستي والتي تتبع توزيع مربع كاي وبدرجة حرية مساوية الى الواحد (df=1) وتكون فرضية الاختبار بالشكل ادناه:

$H0: \beta_k = 0$
$H1: \beta_k \neq 0$

$$\text{wald}' = t^2 = \left(\frac{\hat{\beta}_i}{\text{S.E.}(\hat{\beta}_i)}\right)^2 \quad (38)$$

$\hat{\beta}_i$:- تمثل قيمة معلمة معامل الانحدار اللوجستي المقدر للمتغيرات التوضيحية.
S.E$(\hat{\beta}_i)$:- تمثل قيمة الخطأ القياسي للمعلمة.
ان اختبار (wald) يتبع توزيع مربع كاي ويتم مقارنة القيمة المحسوبة لاختبار (wald) مع القيمة الجدولية لمربع كاي فاذا كانت القيمة المحسوبة اكبر من الجدولية نرفض فرضية العدم وتكون معلمة المتغير التوضيحية معنوية[١١].

المبحث الثاني: الجانب التطبيقي

ان التهاب الكبد الفيروسي (Hepatitis) هو من الامراض الشائعة ويحدث نتيجة اصابة الكبد بضرر او التهاب ويمكن ان يكون سببه الفايروسات او بكتريا او كائنات حية وفي حالة اصابة الكبد باي ضرر فان خلايا الكبد قادرة على تجديد نفسها بشكل سريع وعودة الكبد الى وضعة، اذ انه يتم انتقال فيروسات هذا المرض عن طريق سوائل الجسم او عن طريق الماء، او عن طريق الادوات الملوثة بهذا الفايروس، او من الام الى جنينها[11] .

اذا تم اخذ عينة مكونة من (١١٩) شخص تتضمن اهم العوامل المؤثرة والمسببة للالتهاب وتقدير مدى تأثير هذه العوامل على متغير الاستجابة (وفاة او شفاء المريض) للأشخاص المصابين بحسب نوع الفيروس من حيث العمر والجنس والسكن باستعمال انموذج الانحدار اللوجستي حيث تم جمع البيانات من مستشفى صحة ديالى ومستشفى صحة الخالص.

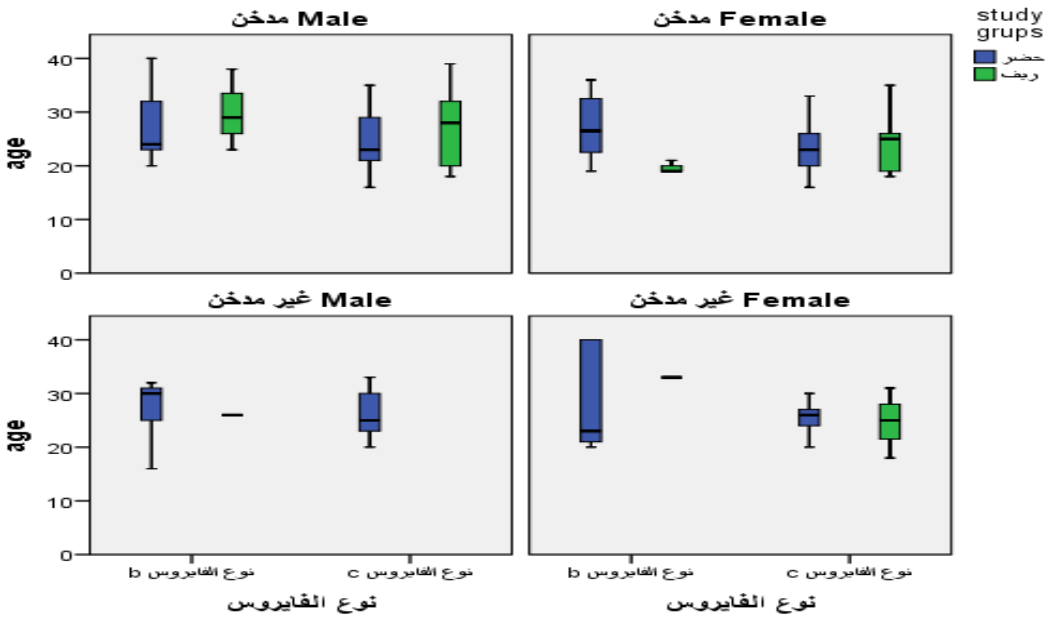
اولاً: متغيرات الدراسة

لنفرض ان :

١. **نوع الفايروس:** - المتمثل بالمتغير(Y).
٢. **الجنس:** المتمثل بالمتغير(X1).
٣. **السكن:** المتمثل بالمتغير(X2).
٤. **التدخين:** المتمثل بالمتغير(X3).
٥. **العمر:** المتمثل بالمتغير(X4) ان للعمر تأثير كبير على مرض التهاب الكبد الفيروسي ففي الغالب يصاب الكبير بالسن والشباب سواء كان ذكر او انثى بسبب مرض وراثي او عن طريق سوائل الجسم.
٦. **نسبة السكر:** المتمثلة بالمتغير (X5).
٧. **نسبة LDL:** المتمثلة بالمتغير(X6) والذي يوضح نسبة البروتين الشحمي المنخفض الكثافة ويعتبر الكوليسترول الضار في دم الإنسان.
٨. **نسبة VLDL:** المتمثلة بالمتغير(X7) والذي يوضح نسبة البروتين الدهني منخفض الكثافة وهو أحد أنواع البروتينات الحمية التي تصنع في الكبد.
٩. **نسبة TRI:** المتمثلة بالمتغير(X8) والذي يوضح المادة الشحمية مكونة من ثلاثة أحماض دسمة عبارة عن شحوم يأتي مصدرها من الغذاء ويتم تركيبها في الكبد.

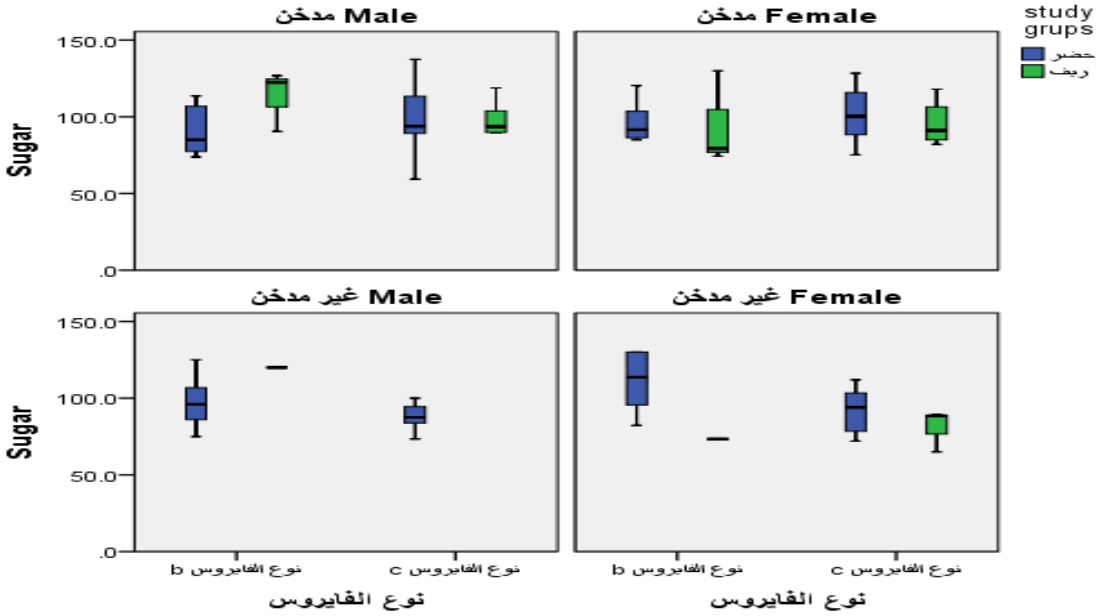
ثانياً: طريقة المربعات الصغرى الجزئية

قبل البدء بتطبيق طريقة الانحدار المربعات الصغرى الجزئية (partial last square regression) نتأكد من عدم وجود قيم شاذة وذلك باختبار (Box plot) حيث ان طريقة المربعات الصغرى الجزئية لا يمكن ان تطبق في حالة وجود قيم شاذة نلاحظ ان الشكل رقم (1) المتغيرات التوضيحية لا تحتوي على اي قيمة شاذة خارج حدود (box plot) التي تمثل متغير العمر بثبوت الجنس ونوع الفايروس ومتغير التدخين والسكن.



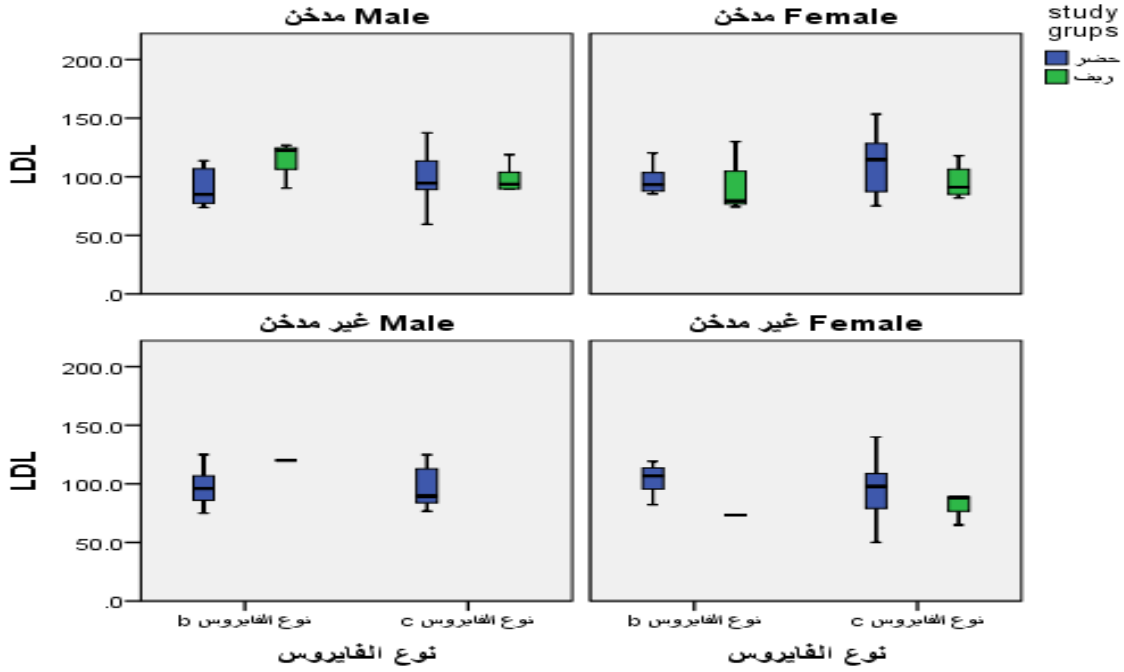
شكل (٢): الكشف عن القيم المتطرفة لمتغير العمر

يبين من الشكل اعلاه عدم وجود قيم متطرفة لمتغير العمر بالاعتماد على متغيرات الوصفية المتمثلة بنوع الفايروس المصنف الى نوع (b,c) والجنس (ذكر وانثى) والتدخين (مدخن وغير مدخن) وسكن المريض (حضر وريف).



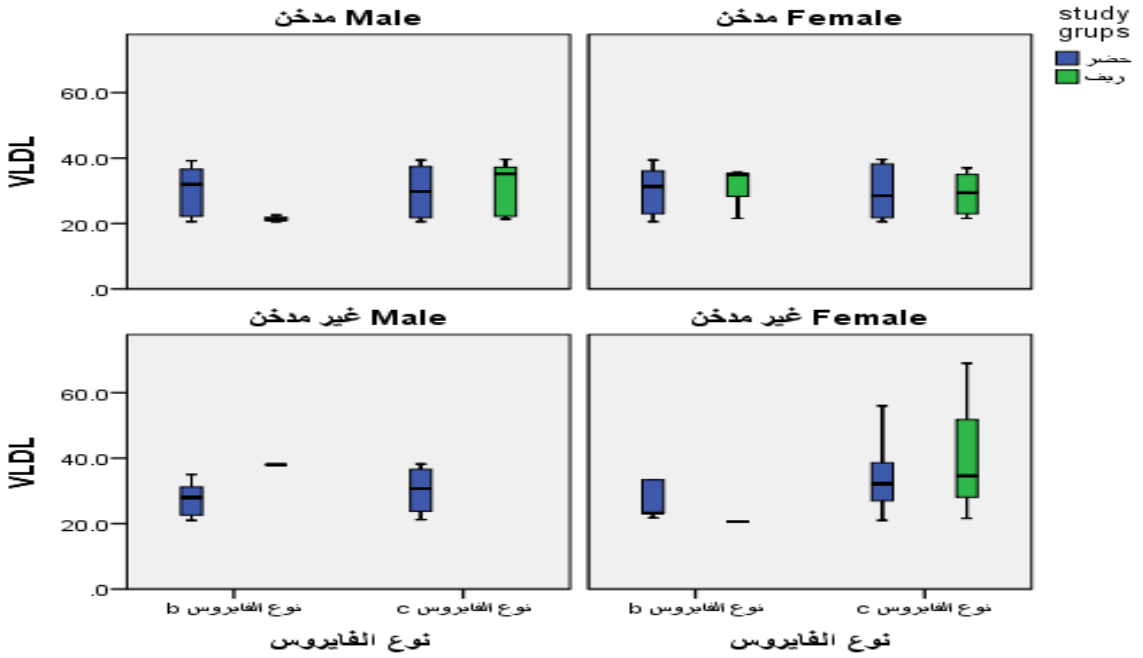
شكل (٣): الكشف عن القيم المتطرفة لمتغير السكر

يبين من الشكل اعلاه عدم وجود قيم متطرفة لمتغير السكر بالاعتماد على متغيرات الوصفية المتمثلة بنوع الفايروس المصنف الى نوع (b,c) والجنس (ذكر وانثى) والتدخين (مدخن وغير مدخن) وسكن المريض (حضر وريف).



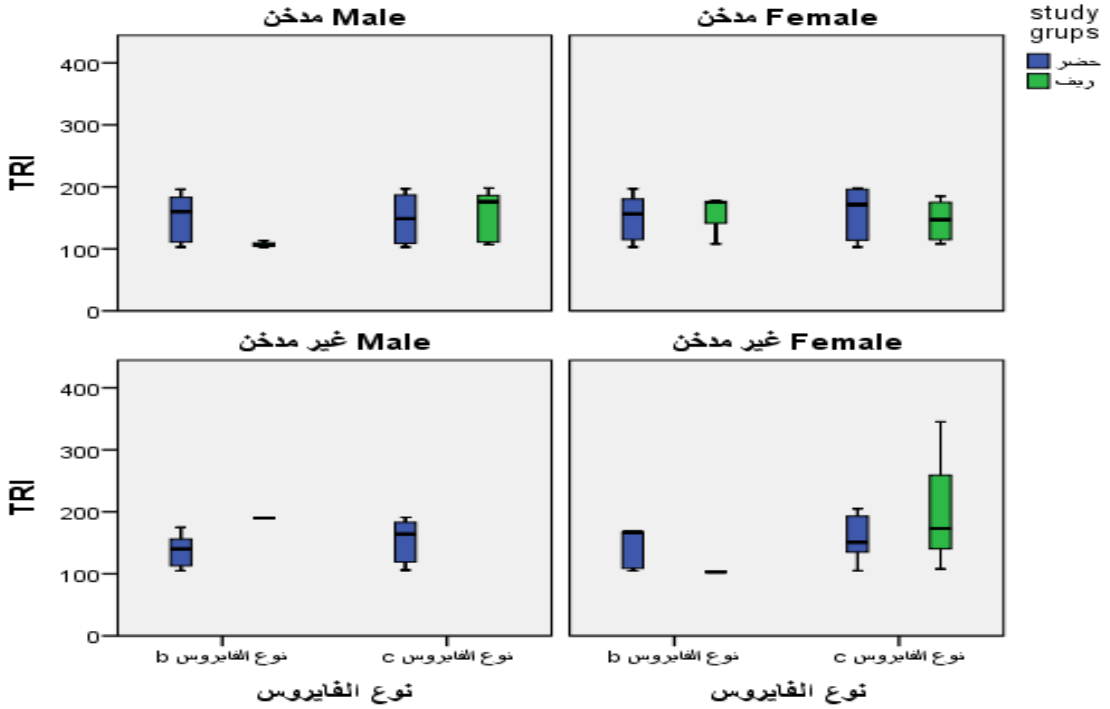
شكل (٤): الكشف عن القيم المتطرفة لمتغير LDL

يبين من الشكل اعلاه عدم وجود قيم متطرفة لمتغير LDL بالاعتماد على متغيرات الوصفية المتمثلة بنوع الفايروس المصنف الى نوع (b,c) والجنس (ذكر وانثى) والتدخين (مدخن وغير مدخن) وسكن المريض (حضر وريف).



شكل (٥): الكشف عن القيم المتطرفة لمتغير VLDL

يبين من الشكل اعلاه عدم وجود قيم متطرفة لمتغير VLDLبالاعتماد على متغيرات الوصفية المتمثلة بنوع الفايروس المصنف الى نوع (b,c) والجنس (ذكر وانثى) والتدخين (مدخن وغير مدخ) وسكن المريض (حضر وريف).



شكل (٦): الكشف عن القيم المتطرفة لمتغير TRI

يبين من الشكل اعلاه عدم وجود قيم متطرفة لمتغير TRIبالاعتماد على متغيرات الوصفية المتمثلة بنوع الفايروس المصنف الى نوع (b,c) والجنس (ذكر وانثى) والتدخين (مدخن وغير مدخ) وسكن المريض (حضر وريف).

ولتطبيق طريقة انحدار المربعات الصغرى الجزئية (partial last square regression) نقوم بالتأكيد من وجود ارتباط ذاتي بين المتغيرات من خلال اختبارVIF كما موضح في الجدول (١) الذي يبين مشكلة التعدد الخطي ونلاحظ ان قيمةVIF > 5 فقط للمتغيرات VLDLوالمتغيرTRL (X7,X8) اما بقية المتغيرات فان قيمة VIF < 5 (X1,X2,X3,X4,X5,X6) وهذا يدل على وجود مشكلة التعدد الخطي، فلذلك سوف نلجأ الى حل هذه المشكلة استخدم طريقة المربعات الصغرى الجزئية (Partial least square method).

جدول (١): يوضح مشكلة التعدد الخطي

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig	Collineanty Statistics	
	B	Std. Error	Beta			Tolerance	VIF
Constant	1.043	0.729		1.431	1.155		
X1	0.001	0.091	0.001	0.012	0.990	0.953	1.049
X2	0.100	0.117	0.078	0.851	0.397	0.950	1.053
X3	-0.041	0.095	-0.040	-0.429	0.668	0.944	1.059
X4	-0.016	0.008	-0.189	-2.057	0.042	0.951	1.052
X5	-0.011	0.005	-0.356	-2.278	0.052	0.329	3.043
X6	0.008	0.004	0.328	1.939	0.055	0.281	3.561
X7	0.006	0.019	0.103	0.335	0.738	0.084	11.849
X8	0.001	0.004	0.047	0.159	0.874	0.094	10.644

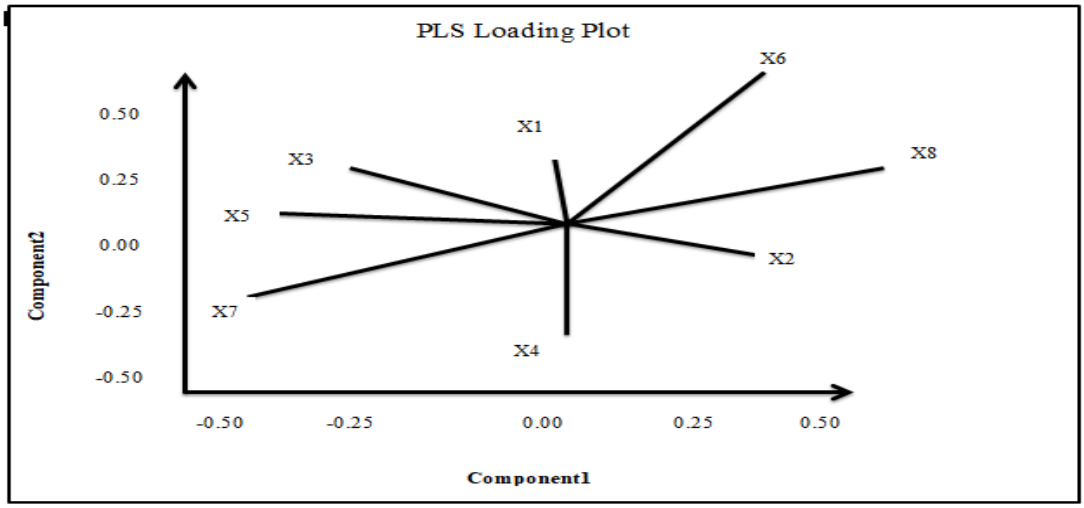
جدول (٢): يبين اختبار لمعنوية المربعات الصغرى الجزئية

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	3.437	8	٠.430	1.838	0.0077
Residual	25.706	110	٠.234		
Total	29.143	118			

ونلاحظ ان من الجدول اعلاه جدول تحليل التباين لمتغير الاستجابة والذي يتضمن الانموذج الذي يحتوي على (8) مركبات حيث ان قيمة p-valueالمحسوبة لمتغير الاستجابة قد بلغت قيمتها (0.0077) وهي اقل من قيمة ال (0.05) وهذا يدل على ان المركبات الثمانية معنوية.

وتم استعمال طريقة المربعات الصغرى الجزئية للتخلص من مشكلة التعدد الخطي باستعمال طريقة العبور الشرعي لتحديد عدد المركبات وقد بلغ عدد المركبات داخلة في الانموذج المربعات الصغرى الجزئية (8) مركبات وكما موضح في جدول رقم (٣) فيكون النموذج بالشكل الاتي

$$y = 1.043 + 0.408x_1 + 0.465x_2 + 0.601x_3 + 5.875x_4 + 0.481x_5 + 0.469x_6 + 2.364x_7 + 2.647x_8$$



شكل (٧): علاقة المتغيرات المستقلة بالمتغير التوضيحي

من الشكل اعلاه وهو الشكل البياني لمركبات التحميل (PLS Loading) المتعلقة بخوارزمية المربعات الصغرى الجزئية حيث ان x1 والمتمثل بمتغير الجنس يمتلك خط قصير جدا وهذا يدل على امتلاكها (X-Loading) منخفضة وهي ليست ذات علاقة مع متغير الاستجابة (نوع الفقر) اما باقي المتغيرات والمتمثلة بـ(x2,x3,x4,x5,x6,x7,x8) تمتلك خطوط طويلة اي ان لها تحميلات عالية اي ترتبط بعلاقة معنوية مع متغير الاستجابة.

جدول (٣) : عدد المركبات الداخلة في الانموذج

Variables	Score	df	Sig
x1	0.408	1	0.023
x2	0.465	1	0.009
x3	0.601	1	0.038
x4	5.875	1	0.015
x5	0.481	1	0.008
x6	0.469	1	0.03
x7	2.364	1	0.004
x8	2.647	1	0.004
Overall Statistics	14.034	8	0.081

علما ان اختبار (Wald) حسب جدول رقم (٤) كانت قيمته (2.412) وقيمة معامل التحديد من جدول رقم (٥) كانت قيمته (0.343)

جدول (٤): يبين اختبار ولد Wald

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	0.288	0.185	2.412	1	0.020	1.333

ومن الجدول اعلاه نلاحظ بان قيمة اختبار (Wald) كانت معنوية كونها اقل من (0.05)

جدول (٥): يبين قيم معامل التحديد

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	0.343	0.118	0.054	0.48341

ثالثاً: الانحدار اللوجستي باستعمال المربعات الصغرى الجزئية

ان اول خطوة في التحليل الاحصائي للانحدار اللوجستي هو ان نقوم بتضمين الحد الثابت فقط في الانموذج الخطوة الصفرية ثم بعدها يتم اضافة المتغيرات التوضيحية وذلك لتحديد كفاءة الانموذج:

١- لا جراء اختبار معنوية النموذج بوجود علاقة بين المتغير المعتمد والمركبات كانت النتائج قيم chi-square بان قيمة Sig.اصغر من 0.05 اي ان النموذج ملائم عند ادخال المركبات والحد الثابت وكما مبين في الجدول التالي

جدول (٦): يبين معنوية النموذج

Step 1	Chi-square	df	Sig.
Step	15.771	8	0.046
Block	15.771	8	0.046
Model	15.771	8	0.046

٢- ان النسب المئوية للتصنيف الصحيح لمشاهدات المتغير المعتمد في الخطوة الصفرية حيث بلغت (57.1%) وان جميع المشاهدات ضمن المجموعة الاولى صنفت بشكل خاطئ اي بمعنى ان نسبة التصنيف الصحيحة (0. %) وان معيار الخطأ التصنيف) 100%) وان جميع المشاهدات صنفت بشكل صحيح خلال المجموعة الثانية حيث بلغت نسبة التصنيف الصحيحة 100%وان معيار خطأ التصنيف (٠%) مبين بالجدول التالي

جدول (٧): يبين نسبة التصنيف للخطوة الصفرية

Observed		Predicted		
		نوع الفايروس		Percentage Correct
		الفايروس من نوع(B)	الفايروس من نوع (C)	
نوع الفيروس	الفايروس من نوع(B)	0	51	0.0
	الفايروس من نوع(C)	0	68	100
Overall Percentage				57.1

٣- ان قيمة سالب ضعف لوغاريتم دالة الامكان الاعظم في تقدير معالم الانحدار اللوجستي بلغت (146.761) وهي اقل من متجه الانموذج الذي يتضمن الحد الثابت وهذا يدل على جودة النموذج. واتضح ايضا قيمة معامل التحديد (R^2) والتي شكلت (٠,١٢٤) من التباين في متغير الاستجابة الذي تم تفسيره من قبل باقي المتغيرات كما موضح في الجدول ادناه.

جدول (٨): يبين معنوية النموذج (سالب لوغاريتم دالة الامكان الاعظم)

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	146.761a	0.124	0.167

٤- ان النسبة المئوية للتصنيف الصحيح لمشاهدات المتغير المعتمد في الخطوة الاولى حيث بلغت (69.7%) وان (28) من المشاهدات ضمن مجموعة الاولى صنفت بشكل صحيح وان (23) صنفت بشكل خاطئ. وان (55) من البيانات ضمن المجموعة الثانية صنفت بشكل صحيح و (13) صنفت بشكل خاطئ. وكما مبيت بالجدول ادناه :

جدول (٩): يبين نسبة التصنيف للخطوة الاولى

Observed		Predicted		
		نوع الفايروس		Percentage Correct
		الفايروس من نوع(B)	الفايروس من نوع (C)	
نوع الفيروس	الفايروس من نوع(B)	28	23	54.9
	الفايروس من نوع(C)	13	55	80.9
Overall Percentage				69.7

جدول (١٠): يبين المعلمات المقدرة للنموذج

Variable	B	S.E.	Wald	df	Sig.	Exp(B)
x1	0.028	0.407	0.005	1	0.944	1.029
x2	0.442	0.532	0.692	1	0.405	1.556
x3	-0.243	0.423	0.330	1	0.566	0.784
x4	-0.070	0.035	4.078	1	0.043	0.932
x5	-0.057	0.029	3.792	1	0.051	0.945
x6	0.058	0.027	4.522	1	0.033	1.060
x7	0.055	0.083	0.430	1	0.512	1.056
x8	0.001	0.017	0.001	1	0.980	1.000
Constant	0.126	2.369	0.003	1	0.958	1.134

ونلاحظ ان المعلمات المقدرة باستعمال دالة الامكان الاعظم وإحصاءة Wald لهذه المعلمات ان اهم المركبات المعنوية هي مركبه الرابعة والسادسة، وكما مبين في الجدول (١٠)

فيكون النموذج الرياضي للانحدار اللوجستي بالشكل الاتي

$$log(\frac{p}{1-p}) = 0.126 + 0.028px_1 + 0.442px_2 - 0.243px_3 - 0.070px_4 - 0.057px_5 + 0.058px_6 + 0.055x_7 + 0.01px_8$$

الاستنتاجات

١. وجد ان المتغيرات (X7,X8 حيث كانت قيمة (VIF>5) ام المتغيرات (X1,X2,X3,X4,X5,X6) فقد كانت قيمة (VIF<5) أي انها تعاني من مشكلة التعدد الخطي عند استخدام طريقة المربعات الصغرى الجزئية فان معامل التضخيم (VIF) أصبح اقل من (٥) وبذلك تم التخلص من مشكلة التعدد الخطي للمتغيرات التي عانت من هذه المشكلة.
٢. من خلال تطبيق خوارزمية المربعات الصغرى الجزئية والذي اتضح في الشكل البياني لمركبات التحميل (PLS Loading) المتعلقة بخوارزمية المربعات الصغرى الجزئية حيث ان x1 والمتمثل بمتغير الجنس يمتلك خط قصير جدا وهذا يدل على امتلاكها (X-Loading) منخفضة وهي ليست ذات علاقة مع متغير الاستجابة (نوع الفقر) اما باقي المتغيرات والمتمثلة بـ(x2,x3,x4,x5,x6,x7,x8) تمتلك خطوط طويلة اي ان لها تحميلات عالية اي ترتبط بعلاقة معنوية مع متغير الاستجابة.
٣. وجد عند تضمين الحد الثابت بان نسبة التصنيف الصحيحة بلغت (57.1%) وان جميع المشاهدات ضمن المجموعة الاولى صنفت بشكل خاطئ اي بمعنى ان نسبة التصنيف الصحيحة (0 %). وان معيار الخطأ التصنيف (100%) وان جميع المشاهدات صنفت بشكل صحيح خلال المجموعة الثانية حيث بلغت نسبة التصنيف الصحيحة 100%وان معيار خطأ التصنيف (0%).
٤. ان قيمة سالب ضعف لوغاريتم دالة الامكان الاعظم في تقدير معالم الانحدار اللوجستي بلغت (146.761) وهي اقل من متجه الانموذج الذي يتضمن الحد الثابت وهذا يدل على جودة النموذج. واتضح ايضا قيمة معامل التحديد (R^2) والتي شكلت (٠,١٢٤) من التباين في متغير الاستجابة الذي تم تفسيره من قبل باقي المتغيرات
٥. من خلال النتائج اوضحت ان متغيرات (X4,X6) والتي تمثلت بمتغير العمر ومتغير LDL له تأثير كبير على مرض التهاب الكبد الفيروسي.
٦. -استنتجنا من الرسم من عدم وجود قيم متطرفة بالنسبة لكل متغير بالاعتماد على نوع الفايروس بالإضافة الى الجنس ونوع السكن.

التوصيات

١. نوصي باستخدام خوارزمية تجزئة القيم المفردة في معالجة مشكلة التعدد الخطي.
٢. نوصي الى استخدام طريقة المربعات الصغرى الجزئية في التخلص من مشكلة التعدد الخطي لأنها عالجت الارتباط الخطي بين المتغيرات.
٣. اجراء دراسات احصائية في حالة وجود مشكلة غير مشكلة تعدد خطي مثل مشكلة الارتباط الذاتي
٤. نوصي بتوسيع انحدار المربعات الصغرى الجزئية وذلك في حالة عدد المتغيرات التوضيحية اكبر من عدد المشاهدات ومن ثم استعمالها في المقارنة بين الانحدار اللوجستي والدالة المميزة الخطية سواء كانت خطية او تربيعية.
٥. تطوير قاعدة جمع البيانات الاحصائية في وزارة الصحة للحصول على بيانات صحيحة وواقعية وحقيقية وفي غاية الدقة لكي تكون النتائج جيدة ومرضية تفيدنا في تطوير مجال الوصول الى الهدف المنشود.
٦. استعمال انظمة البرمجيات الحديثة من خلال تطوير قاعدة جمع البيانات لغرض تبويب وارشفة البيانات الطبية التي تخص المرضى.

المصادر

اولاً: المصادر العربية

١- احمد، ايمان حسن وشهاب، ضمياء حامد، (٢٠١٨)،"مقارنة بعض المقدرات الحصينة لتقدير معلمات انموذج الانحدار اللوجستي ثنائي الاستجابة باستعمال المحاكاة"، بغداد، مجلة العلوم الاقتصادية والادارية، الصفحات (٤٢٣-٤٤٠).
٢- البحيري، نجوى محمد والهربيطي ، يحيى السيد وباشا، امينة محمد. (٢٠٢٤).،"استخدام الانحدار اللوجستي وانحدار بيتا في تقدير فجوة الفقر مع التطبيق علي بيانات لمصر، المجلة العلمية للبحوث والدراسات التجارية، المجلد ٣٨، العدد ٣.
٣- البكري، رباب عبد الرضا (٢٠١٥)، مقارنة بعض الطرائق الخطية لمعالجة مشكلة التعدد الخطي في النماذج مع تطبيق عملي، رسالة دكتوراه– كلية الادارة والاقتصاد –جامعة بغداد.
٤- الحسناوي، زينب جواد كاظم ،(٢٠١٧)" بناء انموذج الانحدار اللوجستي للولادات المشوهة في العراق "، بحث دبلوم عالي، جامعة بغداد، كلية الادارة والاقتصاد، قسم الاحصاء.
٥- صالح، رباب عبد الرضا. (٢٠١٦)،"مقارنة بين طرائق المربعات الصغرى الجزئية والمركبات الرئيسة باستعمال المحاكاة "، مجلة العلوم الاقتصادية والادارية، المجلد ٢٢، العدد ٨٧.
٦- الكرخي ، امل هادي رشيد ،(٢٠١٨)"مقارنة طريقتي Turnbull وBerans اللامعلمية مع طرائق اخرى في تقدير دالة البقاء مع تطبيق عملي"، رسالة ماجستير، جامعة بغداد، كلية الادارة والاقتصاد، قسم الاحصاء.
٧- محمد، لقاء على وخلف، نرجس باسم. (٢٠٢٤)، "دراسة احصائية لبعض المقدرات اللامعلمية لأنموذج الانحدار اللوجستي "، مجلة كلية الرافدين الجامعة للعلوم، العدد ٥٦.

ثانياً: المصادر العربية المترجمة

1- Ahmed, Iman Hassan and Shehab, Dhumia Hamed, (2018), "Comparison of Some Robust Estimators for Estimating Parameters of a Two-Response Logistic Regression Model Using Simulation," Baghdad, Journal of Economic and Administrative Sciences, pp. 423-440.
2- Al-Bahiri, Najwa Muhammad, Al-Harbiti, Yahya Al-Sayed, and Basha, Amina Muhammad (2024). "Using Logistic Regression and Beta Regression in Estimating the Poverty Gap with Application to Egyptian Data," Scientific Journal of Business Research and Studies, Volume 38, Issue 3.
3- Al-Bakri, Rabab Abdul-Ridha (2015). "Comparison of Some Linear Methods for Addressing the Problem of Multicollinearity in Models with a Practical Application," PhD Dissertation, College of Administration and Economics, University of Baghdad.
4- Al-Hasnawi, Zainab Jawad Kadhim (2017). "Constructing a Logistic Regression Model for Malformed Births in Iraq," Higher Diploma Research, University of Baghdad, College of Administration and Economics, Department of Statistics.
5- Al-Karkhi, Amal Hadi Rashid (2018). "Comparison of Turnbull Methods "Nonparametric Berans Estimators with Other Methods for Estimating the Survival Function with Practical Application," Master's Thesis, University of Baghdad, College of Administration and Economics, Department of Statistics.
6- Mohammed, Liqaa Ali and Khalaf, Narjis Basim. (2024), "A Statistical Study of Some Nonparametric Estimators of the Logistic Regression Model," Al-Rafidain University College of Science Journal, Issue 56.
7- Saleh, Rabab Abdul-Ridha (2016). "Comparison Between Partial Least Squares and Principal Component Methods Using Simulation," Journal of Economic and Administrative Sciences, Volume 22, Issue 87.

ثالثاً: المصادر الاجنبية

1- Abdi, hervi, 2010," Partial least squares regression and projection on latent structure regression (PLS Regression)", John Wiley & Son
2- Berkson, J. (1994),"Application of the logistic to Bio-Assay", JASA Vol. 39,No.227,pp.357- 465
3- Czepiel,S.A.(2002),"Maximum likelihood estimation of logistic regression". theory and implementation. Available at czep. net/stat /mlelr .pp78-97.
4- Erik Brorson, Asterios Geroukis-2014, "A Compqrison bletween discriminan tanalysis. And logistic regression using principal components, "Department of statistics, uppsala university, Uppsala University.
5- Hosmer, David W.& Leme show, stanley, (2013), "Applied Logistic Reqression ", 3rd editon, Newyourk ; wiley.
6- John, C., Ekpenyong, E. J., & Nworu, C. C. (2019). Imputation of missing values in economic and financial time series data using five principal component analysis approaches. CBN Journal of Applied Statistics (JAS), 10(1), 3.
7- Kerkri, A., Zarrouk, Z., & Allal, J. (2015). A comparison of NIPALS algorithm with two other missing data treatment methods in a principal component analysis, University Mohamed.