

استخدام اسلوب بيز والانحدار المتدرج في تقدير معلمات الانحدار اللوجستي الثنائي الاستجابة للمصابين بأمراض فقر الدم الحاد والمزمن

Comparison between Bayes and the stepwise regression of the logistic regression with application

م.م. أحمد نزيه الخطيب

م.د. مزاحم محمد الهاشمي

Ass. Lec: Ahmed nazieh AL-Khateeb

Lect. DR.: Muzahem M. AL-Hashimi

كلية علوم الحاسوب والرياضيات/

كلية علوم الحاسوب والرياضيات

جامعة الموصل

جامعة الموصل

e-mail: ahmednazih17@gmail.com

e-mail: muzahim63@gmail.com

تاريخ استلام البحث 2019/6/13 تاريخ قبول النشر 2019/8/7 تاريخ النشر 2019/10/28

المستخلص:

يتضمن البحث استخدام نموذج الانحدار اللوجستي الثنائي الاستجابة المتدرج كأسلوب تقليدي في نمذجة وتحليل أثر مجموعة من المتغيرات التوضيحية على فقر الدم (يتضمن نوعين، فقر الدم المزمن والحاد) والذي يمثل متغير نوعي ثنائي الاستجابة في بعض المستشفيات العراقية. تم التحليل على اعتبار ان الاناث هي المستوى المرجعي. تم استخدام طريقة بيز كأسلوب حديث في تقدير معلمات نموذج الانحدار ثنائي الاستجابة. وتم استخدام كلا من الاحتمال الاولي غير المعلوماتي والاحتمال الاولي ذو المعلومات لإيجاد التقديرات البيزية. تمت المقارنة بين نتائج التقديرات البيزية وفقا لدالتي الاحتمال. كما تمت المقارنة بين الطريقة التقليدية والنتائج التي تم التوصل اليها من طريقة بيز. تم استخدام معايير دقة التوفيق AIC و SC و $-2 \log L$ ، كما تم استخدام كلا من منحني خصائص عمليات المُستقبل (Receiver Operating Characteristic) وجدول التصنيف في التعرف على قدرة الانموذج على التصنيف. توصل البحث الى ان التقديرات البيزية هي الافضل في توفيق البيانات قيد الدراسة. وكانت اهم المتغيرات المؤثرة في فقر الدم عند الذكور (باعتبار ان الاناث هي المستوى المرجعي)، هي الجنس ونسبة التعداد الشبكي فضلا عن نقص الحديد في الدم.

الكلمات المفتاحية: فقر الدم الحاد والمزمن، الانحدار اللوجستي الثنائي الاستجابة المتدرج، الانحدار اللوجستي الثنائي الاستجابة البيزي، الاحتمال الاولي ذو المعلومات القليلة، الاحتمال الاولي ذو المعلومات.

Abstract

This paper involves applied a binary logistic regression as a traditional method for modeling and analyzing the effect of a set of explanatory variables on anemia (including two types, chronic and acute anemia), which represents a categorical binary response variable in some of Iraqi hospitals. The analysis was considered that females are the reference level. The Bayes method was applied as a modern method to estimating the parameters of a binary response regression model. Both the non-information and information prior probability distributions were used to find the Bayesian estimates. The results of the Bayesian estimates were compared according to probability function. A comparison was also made between the traditional method and the results obtained from the Bayes method. The AIC, SC, and 2 Log L as a model selection criterion were used. Both the Receiver Operating Characteristic and the classification table were used also to identify the model's accuracy and classification capability. The paper found that the parameters estimates of the binary logistic regression has similar to the non-informational prior distribution with a very slight superiority of the latter method according to the standard deviations of the estimated parameters as well as the model selection criterion. The analysis showed a clear superiority of the Bayes method according to the information prior distribution based on estimates of the logistic regression parameters on the other estimates. The analysis showed the high classification capacity of the Bayes method according to the prior information distribution. One of the most important variables affecting anemia in males (considering that females are the reference level), is sex and the Reticulocyte count as well as iron deficiency in the blood.

Keywords: *Acute and chronic anemia, Stepwise binary logistic regression, Bayesian binary Logistic Regression, Non-informative prior distribution, Informative prior distribution.*

المقدمة:

تُعرف الطرائق الإحصائية الأكثر استخدامًا بالطرائق التقليدية، إذ تفترض هذه الطرائق أن المعلمات هي كميات ثابتة يتطلب تقديرها إما بتقدير نقطة أو فترة ثقة من خلال بيانات عينة عشوائية مسحوبة من مجتمع معين. وعليه فإن هذه المعلمات لا يمكن تمثيلها بتوزيع احتمالي أو اعتبارها على أنها متغير عشوائي له توزيع احتمالي معين. أما طرائق بيز فهي تقدم نهجا بديلا، إذ يتم التعامل مع المعلمات على أنها متغيرات عشوائية ولها توزيعات احتمالية، وبالتالي، فإنه يمكن إجراء استدلال حول هذه المعلمات.

اصبحت أساليب بيز من أكثر الأساليب شيوعا في التحليل الإحصائي الحديث، إذ يتم تطبيقها في العديد من المجالات العلمية لما تمتاز به من كفاءة ومرونة في التحليل. إن جوهر التحليل البيزي يتمثل في استخدام المعلومات المتاحة والخبرة عن القيم الممكنة لمعلمات الأنموذج، كما يشمل أيضا المعرفة التي تم الحصول عليها من خلال التجارب السابقة.

يستخدم نموذج الانحدار اللوجستي الثنائي الاستجابة (Binary logistic regression model) الذي يتبع توزيع بيرنولي (Bernoulli distribution) في تحليل أثر المتغيرات التوضيحية من خلال دالة ربط معينة (Link function)، والتي يمكن ان تكون Logit او Probit. ويتم تقدير معلمات الانحدار اللوجستي باستخدام طريقة الامكان الاعظم (maximum likelihood) واسلوب بيز.

يعد فقر الدم (Anemia) مشكلة عالمية للصحة العامة تؤثر على البلدان النامية والمتقدمة على حد سواء، مع عواقب وخيمة على صحة الإنسان، كما انها تمثل عبأ اجتماعيا واقتصاديا. يعرف فقر الدم بانه نقص في عدد خلايا الدم الحمراء أو في الجودة، مما يقلل القدرة على حمل ما يكفي من الأوكسجين لتلبية احتياجات أجهزة الجسم المختلفة، والتي تختلف بحسب العمر والجنس وحالة الحمل، وغيرها. يُعتقد أن نقص الحديد هو السبب الأكثر شيوعاً لفقر الدم على مستوى العالم، على الرغم من أن حالات أخرى يمكن أن تسبب فقر الدم، مثل نقص حمض الفوليك وفيتامين (B12) وفيتامين (A) والالتهابات المزمنة والالتهابات الطفيلية والاضطرابات الوراثية (Bekele et al., 2016: 2). تساعد الاعراض وشدتها في تحديد نوع فقر الدم، وتشمل الأعراض الشائعة لفقر الدم، الخمول وعدم انتظام دقات القلب وآلام في الصدر وضيق التنفس والشحوب والإرهاق والصداع وعدم التركيز والخمول. وقد يصاب الأطفال بالتهدج وضعف تناول الطعام عن طريق الفم (Fishbane, S., & Spinowitz, B. 2018: 427).

يصنف فقر الدم على أنه حاد (Acute anemia) أو مزمن (Chronic anemia). يحدث فقر الدم الحاد عندما يكون هناك انخفاض مفاجئ في كريات الدم الحمراء، والتي تحدث في معظم الأحيان بسبب انحلال الدم (hemolysis) أو النزيف الحاد (acute hemorrhage). اما فقر الدم المزمن، فهو انخفاض تدريجي في كرات الدم الحمراء، بسبب نقص الحديد أو نقص التغذية أو الأمراض المزمنة وغيرها من الأسباب (Alder, L., & Tambe, (A. (2019).

يحدث فقر الدم في جميع الفئات العمرية، ولكنه أكثر انتشاراً عند النساء الحوامل والأطفال (Mettananda et al., 2018: 1). وبموجب احصاءات منظمة الصحة العالمية، يعاني 1.62 مليار شخص (24.8%) على مستوى العالم من فقر الدم، من بينهم 56 مليون امرأة حامل. وان أعلى معدلات الإصابة بفقر الدم هو بين الأطفال في سن ما قبل المدرسة (من صفر الى 4.9 سنة) (47.4%) (Kaur, 2014: 32). كما ينتشر فقر الدم بين النساء 21-80% في جميع أنحاء العالم الرجال، ووفقاً لمنظمة الصحة العالمية، يبلغ معدل انتشار فقر الدم بين النساء 21-80% في جميع أنحاء العالم (Saydam, et al., 2017: 433).

هدف البحث:

ان الهدف من البحث، هو المقارنة بين الانحدار اللوجستي التقليدي والانحدار اللوجستي البيزي في ظل نوعين من الاحتمالية الاولى للأسلوب البيزي، وهما الاولى ذات المعلومات القليلة وذات المعلومات، وذلك لأجل الوصول الى أفضل تمثيل لمرض فقر الدم المزمن عند الذكور مقارنة بالإناث في بعض المستشفيات العراقية. وكذلك ايجاد منحى خصائص عمليات المُستقبل (Receiver Operating Characteristic) (ROC) لتحديد أي من هاتين الطريقتين أكثر دقة وقدرة على التصنيف. كما تم حساب حدود الثقة وجدول التصنيف والاحتمالية التنبؤية لمتغير الاستجابة.

الجانب النظري

تستند الاستدلالات الاحصائية عادة على تقديرات الامكان الاعظم (estimation maximum likelihood) (MLE) والتي تفترض ان المعلمات المجهولة والتي يتم تقديرها هي ثابتة. وهذا ما سيتم مناقشته لاحقا بطريقة الانحدار اللوجستي الثنائي الاستجابية. اما اسلوب بيز فيتم التعامل مع المعلمات على انها متغيرات عشوائية لها توزيع احتمالي معين، يسمى بالتوزيع الاحتمالي الاولي (Prior Distribution). ويقترن هذا التوزيع بالمعلومات المتاحة عن الظاهرة قيد الدراسة او من خلال التجارب السابقة. ويتم الحصول على التوزيع اللاحق (Posterior Distribution)، والذي يستخدم للاستدلال عن معلمات الأنموذج من خلال دمج التوزيع الاحتمالي الاولي بدالة الامكان الاعظم للملاحظات (Bolstad and others, 2016: 73) و (الدريني واخرون، 2017: 51).

1.1 الانحدار اللوجستي الثنائي الاستجابية المتدرج (Stepwise Logistic Regression)

يعد الانحدار اللوجستي من الطرائق الاكثر شيوعا في نمذجة وتحليل بيانات الاستجابية الثنائية بين متغير الاستجابية والعديد من المتغيرات التوضيحية التي قد تكون مستمرة أو فئوية (Categories) او خليط من النوعين معا، ويستخدم على نطاق واسع في التطبيقات الطبية عندما يكون متغير الاستجابية ثنائي القيمة، كالنجاح والفشل. وعليه فان متغير الاستجابية يأخذ قيمتين، على سبيل المثال، $Y = 1$ في حالة النجاح، و $Y = 0$ في حالة الفشل. او $Y = 1$ في حالة الإصابة بمرض معين، $Y = 0$ في حالة عدم الإصابة.

على فرض أن x متجه للمتغيرات التوضيحية وان $\pi = P(Y = 1|x)$ هي احتمالية الاستجابية المطلوب نمذجتها (احتمال وقوع الحدث محل الاهتمام). فان الأنموذج اللوجستي الخطي يأخذ الصيغة الآتية (Hosmer and others, 2013: 6

$$\text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta'x \quad \dots(1)$$

اذ ان β_0 تمثل معلمة الحد الثابت و $\beta = (\beta_1, \beta_2, \dots, \beta_k)'$ هو متجه لـ k من معلمات الانحدار. وان $\text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right)$ هو اللوغيت او اللوغاريتم الطبيعي لنسبة الترجيح (Odds ratio).

يعد الأنموذج اللوجستي أحد النماذج الخطية المعممة، اذ يفترض أن دالة الوسط لمتغير الاستجابية مرتبطة خطياً بالمتغيرات التوضيحية، وبما أن الوسط يعتمد ضمناً على السلوك العشوائي لمتغير الاستجابية فضلا عن المتغيرات التوضيحية، فان الدالة توفر ربط بين المكون العشوائي (Random Component)¹، والمكون المنتظم (Systematic

¹ ويسمى في بعض الأحيان بأنموذج الضوضاء (Noise model) او أنموذج الخطأ (error model) ، وهو يشير الى توزيع متغير الاستجابية Y .

(Component)². ولهذا السبب، تستخدم دالة الربط $g(\mu)$ ، والتي تتمثل بدالة اللوغيت logit، التي هي معكوس الدالة اللوجستية التجميعية (inverse of the cumulative logistic distribution function) $F(X) = \frac{\exp(X)}{1+\exp(X)}$.

(a) طريقة التقدير:

عادة ما يتم استخدام طريقة الامكان الاعظم التكرارية (iterative maximum likelihood method) ممثلة بطريقة فيشر (Fisher scoring method) وطريقة نيوتن رافسن (Newton-Raphson) في تقدير معاملات النماذج غير الخطية ومنها معاملات الأنموذج اللوجستي. في هذا البحث، تم استخدام طريقة فيشر لأنها تعد من الطرق الحصينة (robust) وتتقارب بسرعة اكبر من طريقة نيوتن رافسن (Schworer and others, 2004: 10).

• طريقة فيشر (Fisher scoring):

افرض ان $Z_j = (Z_{1j}, Z_{2j}, \dots, Z_{(k+1)j})'$ متغير متعدد الحدود يأخذ القيم الاتية (SAS Institute, 2010):

$$Z_j = \begin{cases} 1 & \text{if } Y_j = i \\ 0 & \text{otherwise} \end{cases} \dots(2)$$

باحتمالية π_{ij} ، وان القيمة المتوقعة لـ Z_j هي $\pi_j = (\pi_{1j}, \pi_{2j}, \dots, \pi_{(k+1)j})'$ ، اذ ان $\pi_{(k+1)j} = 1 - \sum_{i=1}^k \pi_{ij}$ وان مصفوفة التباين-التباين المشترك لـ Z_j هي V_j . ولتكن β موجه معاملات الانحدار، اي ان:

$$\beta = (\alpha_1, \alpha_2, \dots, \alpha_k, \beta_1, \beta_2, \dots, \beta_s)'$$

ولتكن D_j مصفوفة المشتقات الجزئية لـ π_j بالنسبة لـ β . عليه، فان معادلة التقدير لمعاملات الانحدار هي:

$$\sum_j D_j' W_j (D_j - \pi_j) = 0 \dots(3)$$

اذ ان $W_j = w_j f_j V_j^{-1}$ ، وان w_j و f_j هما الوزن والتكرار للملاحظات j على الترتيب، وان V_j^{-1} هي معكوس المصفوفة القطرية التي قطرها π_j .

تبدأ خطوات الحل بقيمة اولية، ولتكن $\beta^{(0)}$ وبعملية تكرارية يتم الوصول الى تقريبات متعاقبة لتقدير الامكان

الاعظم لـ β وكالاتي:

$$\beta^{(m+1)} = \beta^{(m)} + (\sum_j D_j' W_j D_j)^{-1} \sum_j D_j' W_j D_j (Z_j - \pi_j) \dots(4)$$

فاذا كانت $\beta^{(m+1)}$ اقل من $\beta^{(m)}$ ، تعاد الاجراءات الحسابية مرة اخرى لحين الحصول الى التقارب، بمعنى ان

$\beta^{(m+1)}$ تقترب حد الكفاية من $\beta^{(m)}$. عندها يتم اعتماد قيم $\beta^{(m+1)}$ على انها القيم التقديرية لـ β ، اي ان

$$\hat{\beta} = \beta^{(m+1)}$$

²، وهو مجموعة المعلمات β والمتغيرات التوضيحية X_1, X_2, \dots, X_k في الأنموذج، والتي تمثل التركيبية او التوافق

الخطية (Nelder and others, 1972: 27).

يتم تقدير مصفوفة التباين-لتباين المشترك لـ $\hat{\beta}$ كالآتي:

$$\widehat{cov}(\hat{\beta}) = (\sum \hat{D}_j' \hat{W}_j \hat{D}_j)^{-1} \quad \dots(5)$$

اذ ان \hat{D}_j و \hat{W}_j هما مقدر لـ D_j و W_j على الترتيب. يمكن تحديد القيم الابتدائية لمعاملات الانحدار من قبل الباحث.

(b) اختيار أفضل نموذج انحدار:

تبدأ اجراءات الاختيار المتدرج بشكل مشابه لطريقة الاختيار الأمامي. وابتداء، تكون بدون أي متغيرات في النموذج، وفي كل خطوة من خطوات التحليل يتم اختبار النموذج، فاذا كان المتغير في النموذج يفشل في تلبية معيار البقاء في النموذج، فيتم استبعاد هذا المتغير. وبخلافه، يتم إدخال المتغير. وعندما تفي جميع المتغيرات في النموذج بمعيار البقاء ولا يفي أي من المتغيرات الأخرى بمعيار الإدخال، تتوقف عملية الاختيار المتدرج.

(c) معايير توفيق الانموذج:

من المعايير التي تستخدم في الوصول الى أفضل توفيق للبيانات باستخدام الانحدار اللوجستي، هي الآتي:

• 2- لوغاريتم الامكان (-2 Log Likelihood):

إذا كان لدينا z من المشاهدات، فان صيغة المعيار هي الآتي:

$$-2 \log L = -2 \sum_j \frac{w_j}{\sigma^2} f_j \log(\hat{\pi}_j) \quad \dots(6)$$

اذ ان w_j و f_j هما قيم الوزن والتكرار على الترتيب لـ z من المشاهدات، و σ^2 تمثل معلمة التشتت، وان $\hat{\pi}_j$ تمثل الاحتمالية التقديرية لمشاهدات متغير الاستجابة. وبموجب هذا المعيار يتم اختيار أفضل توفيق للبيانات وفق الانموذج الذي له اقل قيمة $-2 \log L$.

• معيار معلومات اكاكي (Akaike Information Criterion):

ويأخذ الصيغة الآتية:

$$AIC = -2 \log L + 2p \quad \dots(7)$$

اذ ان p تمثل عدد المعلمات في الانموذج. وبموجب هذا المعيار يتم اختيار أفضل توفيق للبيانات وفق الانموذج الذي له اقل قيمة (AIC).

• معيار شوارز (معلومات بيز) (Schwarz (Bayesian Information Criterion):

ويأخذ الصيغة الآتية:

$$SC = -2 \log L + p \log(\sum_j f_j n_j) \quad \dots(8)$$

اذ ان p تمثل عدد المعلمات في الانموذج.

وبموجب هذا المعيار يتم اختيار أفضل توفيق للبيانات وفق الانموذج الذي له اقل قيمة (SC).

• مربع كاي للباقي (Residual Chi-Square):

على فرض ان S تمثل المتغيرات التوضيحية محل الاهتمام، فان المتجه لكل المعلمات لانموذج الاستجابة هو:

$$\beta = (\alpha_1, \alpha_2, \dots, \alpha_k, \beta_1, \beta_2, \dots, \beta_s)' \quad \dots(9)$$

اذ ان $\alpha_1, \alpha_2, \dots, \alpha_k$ هي معاملات الحد الثابت، وان $\beta_1, \beta_2, \dots, \beta_s$ هي معاملات الميل لـ S من المتغيرات التوضيحية ذات التأثير على متغير الاستجابة. ان نموذج اللوغيت العام الكلي يتضمن متجه المعلمات الاتية:

$$\beta = (\alpha_1, \alpha_2, \dots, \alpha_k, \beta_1', \beta_2', \dots, \beta_s')'$$

اذ ان:

$$\beta_{ij}' = (\beta_{i1}, \beta_{i2}, \dots, \beta_{is}) ; i = 1, 2, \dots, k \quad \dots(10)$$

اذ ان β_{ij} هي معلمة الميل للمتغير التوضيحي j في انموذج اللوغيت i ، وتأخذ الصيغة الاتية:

$$g'(\hat{\beta}_{H_0})I^{-1}(\hat{\beta}_{H_0})g(\hat{\beta}_{H_0}) \quad \dots(11)$$

$$\hat{\beta}_{H_0} = (\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_k, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k, 0, 0, \dots, 0)'$$

1.2. الاسلوب البيزي

الإحصاءات البيزية هي فرع من الإحصاءات التي نشأت من نظرية بيز التي وضعها توماس بيز (1763)، ولقد ظهرت فكرة نظرية بيز لأول مرة من خلال مقالة نشرت في عام (1763) بعنوان "رسالة إلى جون كانتون" (A letter to John Canton)، اذ أثبت المؤلف حالة خاصة لما يسمى الآن باسم نظرية بيز. ومن خلال هذه الفكرة، قدم بيير سيمون لابلاس (Pierre-Simon Laplace) تعميم لنظرية بيز، واستخدم هذه النظرية في حل العديد من المسائل المتعلقة بالإحصاءات الطبية (medical statistics) والمعملية (reliability) والتشريع (jurisprudence) (Stigler, 1986: 98).

في الوقت الحاضر، تعد أساليب بيز من أكثر اساليب التحليل الإحصائي الحديث شيوعا في تقدير معاملات نماذج مختلفة، وتستند في ذلك الى نظرية بيز. وعليه، فان التحليل البيزي في تحليل الانحدار يفترض أن كل معاملات الأنموذج هي متغيرات عشوائية، وبالتالي يمكن أن تتضمن معرفة مسبقة ممثلة بالخبرة السابقة او نتائج لبحوث سابقة او أي معلومات عن المعلمة المجهولة. هذا الافتراض يتناقض مع الاستدلال الإحصائي التقليدي، والذي يعتبر ان جميع المعلمات المجهولة هي كميات محددة.

ان التحليل البيزي يتبع قاعدة احتمالية، وهي قاعدة بيز، التي توفر صيغة لدمج المعلومات الاولية مع معطيات البيانات الموجودة للحصول على التوزيع اللاحق. بمعنى اخر، ان التحليل البيزي يتضمن الانتقال من التوزيع الاولي الى التوزيع اللاحق، والذي يستخدم للاستدلال عن معاملات الأنموذج، بما في ذلك تقديرات النقطة مثل الوسط والوسيط والنسب المئوية وتقديرات الفترة، كما يمكن التعبير عن جميع الاختبارات الإحصائية حول معاملات الأنموذج كتعبير احتمالية على أساس التوزيع اللاحق التقديري.

يمكن التعبير عما جاء اعلاه، بالآتي (Stokes and others, 2014: 1):

بالنسبة للحادثتين A و B ، فانه يمكن صياغة الاحتمال الشرطي للحادثة A بمعلومية الحادثة B وفقا لنظرية بيز، كالاتي:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad \dots(12)$$

والتي يمكن كتابتها بالصيغة الاتية:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A)+P(B|\bar{A})P(\bar{A})} \quad \dots(13)$$

اذ ان \bar{A} مكمل A . وإذا افترضنا ان A على انها المعلمة θ وان B على انها البيانات y ، عليه فان:

$$P(\theta|y) = \frac{P(y|\theta)P(\theta)}{P(y)} = \frac{P(y|\theta)P(\theta)}{P(y|\theta)P(\theta)+P(y|\theta)P(\theta)} \quad \dots(14)$$

اذ ان $P(y)$ هي الاحتمالية الهامشية (marginal probability) وهي بمثابة ثابت لن يكون له تأثير في التوزيع اللاحق، وبالتالي يمكن تجاهله. يمكن كتابة التوزيع اللاحق بصيغة التناسب (α) (قاسم، فتحي، 2008: 15):

$$P(\theta|y) \propto P(y|\theta)P(\theta) \quad \dots(15)$$

والتي تعني ان التوزيع اللاحق $P(\theta|y)$ يتناسب مع حاصل ضرب دالة الامكان $P(y|\theta)$ بالتوزيع الاولي $P(\theta)$. فاذا أردنا تقدير المعلمة θ من البيانات $y = \{y_1, y_2, \dots, y_n\}$ باستخدام الأنموذج الاحصائي الموضح بدالة الكثافة $P(y|\theta)$ فانه، وبحسب الفلسفة البيزية، لا يمكن تحديد قيمة θ بالضبط ولكن يمكن وصف عدم اليقين (uncertainty) باستخدام صيغ احتمالية وتوزيعات للتعبير عن نقص المعرفة عن قيم المعلمات. فاذا كانت صيغة التوزيع الاولي $\pi(\theta)$ معبرة عن المعلومات السابقة عن θ . فانه يمكن تحديث هذه المعلومات وذلك بدمج المعلومات من التوزيع الاولي بالبيانات الموصوفة وفق الأنموذج الاحصائي $P(y|\theta)$. لتوليد التوزيع اللاحق $P(\theta|y)$.

$$P(\theta|y) = \frac{P(\theta,y)}{P(y)} = \frac{P(y|\theta)\pi(\theta)}{P(y)} = \frac{P(y|\theta)\pi(\theta)}{\int P(y|\theta)\pi(\theta)d\theta} \quad \dots(16)$$

اذ ان الكمية $P(y)$ تسمى بالاحتمالية الهامشية وهي ثابت للتوزيع اللاحق، وعادة ما تهمل. عليه يمكن كتابة $P(\theta|y)$ بالصيغة الاتية:

$$P(\theta|y) \propto P(y|\theta)\pi(\theta) = L(\theta)\pi(\theta) \quad \dots(17)$$

اذ ان L هي دالة الامكان وتعرف على انها دالة متناسبة مع $P(y|\theta)$. يمكن كتابة $P(\theta|y)$ بالصيغة الاتية:

$$P(\theta|y) = \frac{P(\theta,y)}{P(y)} = \frac{P(y|\theta)\pi(\theta)}{P(y)} = \frac{P(y|\theta)\pi(\theta)}{\int P(y|\theta)\pi(\theta)d\theta} \quad \dots(18)$$

على فرض ان المطلوب هو تقدير المعلمة θ من البيانات $y = \{y_1, y_2, \dots, y_n\}$ باستخدام الأنموذج الاحصائي الموضح بدالة الكثافة الاحتمالية $P(y|\theta)$. بموجب النظرية البيزية فانه لايمكن تحديد قيمة المعلمة θ

بالضبط، وإنما يمكن القول مثلا بان المعلمة تتبع التوزيع الطبيعي بوسط وتباين معينين، وذلك في حالة ان هذا التوزيع يصف عدم اليقين المرتبط بالمعلمة على أفضل وجه.

الخطوات الاتية تصف العناصر الأساسية للاستدلال البيزي (Gunes and others, 2014: 2):

1. صياغة التوزيع الاحتمالي ل θ على انه $\pi(\theta)$ ، والذي يعرف بالتوزيع الاولي (prior distribution)، اذ يعبر التوزيع الاولي عن المعلومات السابقة المتوفرة عن المعلمات (على سبيل المثال، الوسط، الالتواء، وما إلى ذلك).
2. اختيار الأنموذج الاحصائي $P(\mathbf{y}|\theta)$ وفقا للبيانات المشاهدة \mathbf{y} وذلك لوصف توزيع \mathbf{y} بمعلومية θ .
3. تحديث المعلومات عن θ من خلال حساب التوزيع اللاحق $P(\theta|\mathbf{y})$ المتضمن دمج المعلومات من التوزيع الاولي بالبيانات.

يتم تنفيذ الخطوة الثالثة باستخدام نظرية بيز، والتي تمكن من الدمج بين التوزيع الاولي والأنموذج وفقا للاتية:

$$P(\theta|\mathbf{y}) = \frac{P(\theta,\mathbf{y})}{P(\mathbf{y})} = \frac{P(\mathbf{y}|\theta)\pi(\theta)}{P(\mathbf{y})} = \frac{P(\mathbf{y}|\theta)\pi(\theta)}{\int P(\mathbf{y}|\theta)\pi(\theta)d\theta} \quad \dots(19)$$

اذ ان المقدار:

$$P(\mathbf{y}) = \int P(\mathbf{y}|\theta)\pi(\theta)d\theta \quad \dots(20)$$

هو ثابت طبيعية (normalizing constant) للتوزيع اللاحق، وهو ايضا التوزيع الهامشي (marginal distribution) ل \mathbf{y} ، ويطلق عليه في بعض الاحيان بالتوزيع الهامشي للبيانات.

ان دالة الامكان ل θ هي اي دالة تتناسب مع $P(\mathbf{y}|\theta)$ ، بمعنى اخر، ان:

$$L(\theta) \propto P(\mathbf{y}|\theta) \quad \dots(21)$$

كما يمكن كتابة نظرية بيز بالصيغة الاتية:

$$P(\theta|\mathbf{y}) = \frac{L(\theta)\pi(\theta)}{\int L(\theta)\pi(\theta)d\theta} \quad \dots(22)$$

كما يمكن كتابة $P(\theta|\mathbf{y})$ بصيغة التناسب وعلى النحو الاتي:

$$P(\theta|\mathbf{y}) \propto L(\theta)\pi(\theta) \quad \dots(23)$$

من الناحية النظرية، تقدم أساليب بيز بدائل بسيطة للاستدلال الإحصائي، إذ يتم استنتاج الاستدلالات من التوزيع اللاحق.

1.1.2 التوزيعات السابقة (Prior Distributions):

التوزيع السابق للمعلمة هو توزيع احتمالي يمثل حالة عدم اليقين حول المعلمة المراد تقديرها قبل تحليل البيانات قيد الدراسة. وإن عملية ضرب التوزيع السابق بدالة الامكان يؤدي إلى التوزيع اللاحق للمعلمة، مما يسمح بأجراء الاستدلالات حول المعلمة، إذ لا يمكن تنفيذ أي استدلال بيزي دون استخدام التوزيع السابق.

• التوزيعات الاحتمالية السابقة غير المعلوماتية (noninformative prior distribution)

في حالة عدم توفر معلومات مسبقة عن المعلمات، فإنه يتطلب تحديد توزيع اولي، وليكن $\pi(\theta)$. ويعرف على انه غير معلوماتي إذا كان له تأثير محدود على التوزيع اللاحق لـ θ . أن التوزيعات السابقة غير المعلوماتية لا تعني الجهل التام عن المعلمة، وإنما تعني ان المعرفة حول هذه المعلمات غامضة، وأنه يتطلب التعبير عن هذا الغموض بتوزيع احتمالي مناسب. في بعض الحالات، يمكن أن تؤدي التوزيعات غير المعلوماتية السابقة غير الملائمة إلى توزيعات لاحقة غير ملائمة، مما يتطلب دمجها مع دالة الامكان للحصول على توزيع احتمالي لاحق ملائم. إذ لا يمكن اجراء الاستدلالات في ظل التوزيعات اللاحقة غير الملائمة.

التوزيعات السابقة ذات المعلوماتية (Informative Priors)

التوزيعات السابقة ذات المعلوماتية هي توزيعات سابقة لها تأثير على التوزيع اللاحق. فإذا كان التوزيع السابق هو الذي يهيمن على دالة الامكان الاعظم، فهو توزيع سابق معلوماتي، وعليه، يجب تحديد هذه الأنواع من التوزيعات بعناية في التطبيقات العملية. من ناحية أخرى، يوضح الاستخدام الصحيح للتوزيعات السابقة قوة أسلوب بيز. إذ يمكن الجمع بين المعلومات التي تم جمعها من الدراسة السابقة ومن الخبرة السابقة أو رأي الخبراء مع المعلومات الحالية بصورة طبيعية.

اقترح كل من سوليفان وغريند (Sullivan and others, 2012: 309) تحليل بيزي وذلك بزيادة البيانات. وتمثل العملية المقترحة في تكوين بيانات سابقة من التوزيع الطبيعي لـ θ ضمن الفترة الاولى (OR_{upper}, OR_{lower}) التي تمثل حدود الثقة 95% والتي يفترض ان تقع ضمنها قيم β ، كما انها تقابل نسب الاحتمالية التي تم الحصول عليها من البيانات الأصلية من خلال الانحدار اللوجستي التقليدي، أي ان $OR = e^{\beta}$ (Oliveira and others, 2017: 2).

التوزيع الطبيعي السابق لـ θ_{prior} يعرف بالصيغة الآتية:

$$\theta_{prior} = \ln(OR_{prior}) = \frac{\ln(OR_{upper}) + \ln(OR_{lower})}{2} \quad \dots(24)$$

وان التباين التوزيع الطبيعي السابق هو: (Sullivan and others, 2012: 309)

$$V_{prior} = \left[\frac{\ln(OR_{upper}) - \ln(OR_{lower})}{2 \times 1.96} \right]^2 \quad \dots(25)$$

2. الجانب التطبيقي:

لغرض اجراء المقارنة بين الانحدار اللوجستي التدريجي التقليدي والانحدار اللوجستي البيزي، فقد تم دراسة العلاقة بين فقر الدم بنوعيه الحاد والمزمن عند الذكور مقارنة بالاناث، كمتغير استجابة ثنائي وبين مجموعة من المتغيرات التوضيحية الموضحة ادناه لـ 140 مشاهدة من الذين تم اجراء الفحوص عليهم في كل من مستشفى بعقوبة العام في محافظة ديالى ومستشفى كلار العام في محافظة السليمانية ومستشفى الطوارئ المركزي في محافظة أربيل (العزي، البكري 2017: 386)، باستخدام كل من الانحدار اللوجستي التقليدي والانحدار اللوجستي البيزي في ظل نوعين من الاحتمالية السابقة، وهما ذات المعلومات القليلة وذات المعلومات، وكذلك ايجاد منحنى خصائص عمليات المُستقبل (ROC) (Receiver Operating Characteristic) للوصول الى افضل تمثيل للظاهرة قيد الدراسة، فضلا عن تحديد أي من هاتين الطريقتين أكثر دقة وقدرة في التصنيف. كما تم حساب حدود الثقة وجدول التصنيف والاحتمالية التنبؤية لمتغير الاستجابة. تم استخدام البرنامج الاحصائي SAS 9.4 في التحليل.

متغيرات البحث:

1. Y متغير استجابة وهمي يأخذ القيمة 0 لتمثل فقر الدم الحاد و 1 لتمثل فقر الدم المزمن.
2. الجنس المتمثل بالمتغير X_1 ، وهو متغير وهمي يأخذ القيمة 1 للإناث و 2 للذكور.
3. العمر المتمثل بالمتغير X_2 .
4. نسبة الهيموغلوبين المتمثل بالمتغير X_3 .
5. نسبة الفيريتين (Ferritin) في الدم والمتمثل بالمتغير X_4 ، وهو بروتين ينتج في عملية التمثيل الغذائي ويوجد داخل خلايا الجسم ويتحكم في تخزين وإطلاق الحديد بحسب الحاجة (d'Estaintot and others, 277).
6. نسبة التعداد الشبكي (Reticulocyte count) والمتمثل بالمتغير X_5 ، وهو مؤشر مهم في تشخيص وتصنيف ومراقبة مرضى فقر الدم، ويستخدم لتقييم استجابة نخاع العظام لعلاج فقر الدم (Wollmann and others, 2017:26).
7. نسبة الحجم الكروي الوسطي ((MCV) mean corpuscular volume) المتمثل بالمتغير X_6 ، وهو متوسط كمية الهيموغلوبين في خلية الدم الحمراء الواحدة (Stengle and others, 1965: 4).

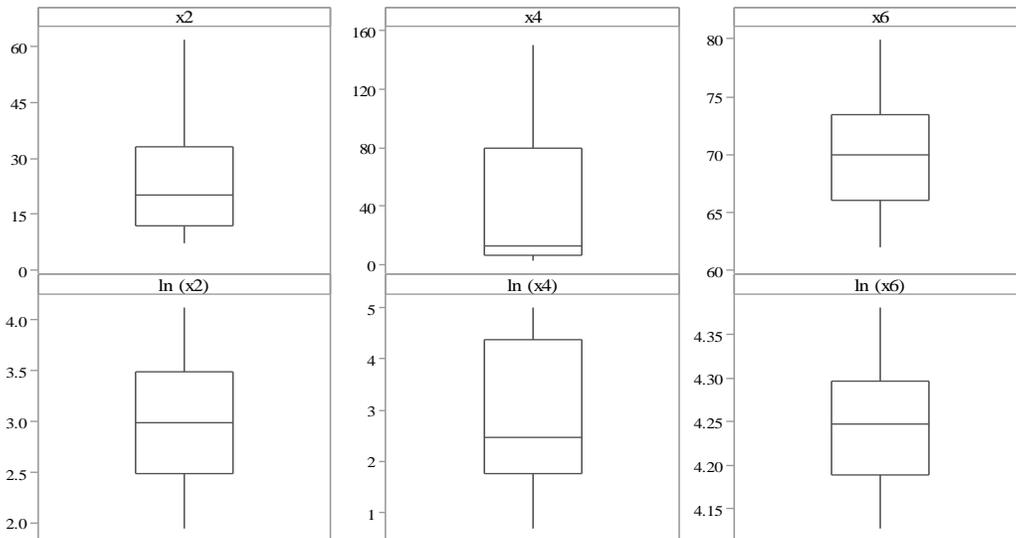
8. نقص الحديد في الدم Iron deficiency Anemia المتمثل بالمتغير X_7 ، وهو يشير الى نقصان عدد خلايا الدم الحمراء او الهيموغلوبين في الدم (Camaschella, 2015: 1833).
9. نسبة الترانسفيرين Transferrin في الدم والمتمثل بالمتغير X_8 . وهو البروتين الرئيس في الدم المسؤول عن نقل الحديد الى جميع اجزاء الجسم (Tolosano, 2015: 565).
10. سبب فقر الدم هو نزف الدم المتمثل بالمتغير X_9 .
11. فقر دم الامراض المزمنة ممثلا بالمتغير X_{10} .
12. فقر الدم هو نقصان في كريات الدم الحمر المتمثل بالمتغير X_{11} .
- تتضمن البيانات 73 مشاهدة تمثل المرضى مصابين بفقر الدم المزمن ($Y = 1$)، و67 مشاهدة تمثل المرضى المصابين بفقر الدم الحاد ($Y = 0$). علما ان عدد الذكور هو 57 وعدد الاناث هو 83 (العزي، البكري 2017: 387) و (العزي، 2017: 64).

يمكن كتابة الانموذج المطلوب تقديره بالشكل الاتي:

$$\text{logit } P(Y = 1 | \mathbf{X}) = \beta_0 + \beta_1 x_1 + \dots + \beta_{11} x_{11}$$

ان الخطوة الاولى في عملية التحليل هو في الكشف عن القيم الشاردة في المتغيرات المستمرة. وعليه فقد تم رسم الشكل الصندوقي للمتغيرات المستمرة الثلاثة (X_2, X_4, X_6). يلاحظ من الشكل (1-أعلى) بان البيانات لا تحتوي على بيانات شاردة. ثانيا، ولأجل ان تكون المتغيرات المستمرة متجانسة التباين، فقد تم استخدام التحويل اللوغاريتمي للمتغيرات الثلاثة، وتم رسم الشكل الصندوقي، وكما موضح ذلك في الشكل (1-أسفل)، ويلاحظ ان البيانات للمتغيرات الثلاثة المستمرة اصبحت بتباين أكثر تجانسا.

Boxplot of x2, x4, x6, ln (x2), ln (x4), ln (x6)



الشكل (1): الرسم الصندوقي للمتغيرات المستمرة (X_2, X_4, X_6) قبل (الرسم التي في الاعلى) وبعد اجراء التحويل اللوغاريتمي (الرسم التي في الاسفل)

تم استخدام تحليل الانحدار الخطي لتشخيص مشكلة تعدد العلاقات الخطية بين المتغيرات التوضيحية، وفقا لمعامل تضخم التباين (VIF) (Variance Inflation Factor) ³ وقيمة التسامح (TOL4) (Tolerance). يلاحظ من الجدول (1)، التحليل (Diagnostic Multicollinearity (1))، ان المتغيرات X_5 و X_6 و X_8 و X_9 و X_{10} لها VIF أكبر من 5، وبالاقتران مع الـ TOL، يلاحظ ان المتغير X_9 لها $TOL < 0.1$. عليه، ولغرض معالجة مشكلة تعدد العلاقات الخطية، فقد تم حذف هذا المتغير. تم اعادة تحليل الانحدار بعد استبعاد المتغير X_9 ، وتم اجراء الكشف عن تعدد العلاقات الخطية. من الجدول (1) (Diagnostic Multicollinearity (2))، يلاحظ ان الـ VIF أكبر من 5 للمتغيرين X_6 و X_8 ، وعند ملاحظة الـ TOL، تبين انه لا يوجد متغير له $TOL < 0.1$ ، عليه، فقد تم اللجوء الى القيم الكامنة (Eigenvalue) ومؤشر الحالة (Condition Index) ⁵، ووجد انهما (0.00002652) (617.23083) على الترتيب (لم تدرج في الجدول)، فضلا عن نسب التباين (Proportion of Variation) المقترنة بهما والموضحة في الجدول (1) (Diagnostic Multicollinearity (2))، والتي كانت (0.98198) للمتغير X_6 ، عليه فقد تم حذف

³ إذا كانت $VIF > 5$ ، فهذا مؤشر على وجود مشكلة تعدد العلاقات الخطية (Montgomery and others, 2012: 527).

⁴ إذا كان $TOL < 0.1$ ، فهذا مؤشر على وجود مشكلة تعدد العلاقات الخطية.

⁵ إذا كان واحد أو أكثر من القيم الكامنة (Eigenvalue) صغيراً (قريب من الصفر) وكان مؤشر الحالة (condition index) في المقابل كبيراً، فهذا مؤشر على وجود مشكلة تعدد العلاقات الخطية.

هذا المتغير. تم اعادة التحليل وتبين انه لا توجد مشكلة لتعدد العلاقات الخطية بين المتغيرات، لاحظ الجدول (1)
(3) (Diagnostic Multicollinearity).

جدول (1): الكشف عن مشكلة تعدد العلاقات الخطية

Diagnostic Multicollinearity (1)											
	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉	X ₁₀	X ₁₁
VIF	3.47 926	4.04 392	1.69 005	2.39 587	5.49 536	5.95 367	4.23 539	8.68 468	13.17 281	7.34 669	3.70 929
TOL	0.28 742	0.24 729	0.59 170	0.41 738	0.18 197	0.16 796	0.23 611	0.11 515	0.075 91	0.13 612	0.26 959
Diagnostic Multicollinearity (2)											
VIF	3.34 261	3.99 952	1.66 626	1.99 157	3.66 389	5.67 422	3.65 069	8.24 421	-	4.95 651	2.85 993
TOL	0.29 917	3.99 952	0.60 014	0.50 212	0.27 293	0.17 624	0.27 392	0.12 130	-	0.20 175	0.34 966
Proportion of Variation	0.18 866	0.15 352	0.01 048	0.14 580	0.09 763	0.98 198	0.01 547	0.43 096	-	0.01 522	0.03 519
Diagnostic Multicollinearity (3)											
VIF	2.95 329	3.60 537	1.66 626	1.64 349	3.29 047	-	3.64 046	4.60 101	-	4.77 674	2.77 423
TOL	0.33 861	0.27 736	0.60 015	0.60 846	0.30 391	-	0.27 469	0.21 734	-	0.20 935	0.36 046

تطبيق اسلوب الانحدار اللوجستي المتدرج:

لغرض استخدام الاختيار المتدرج في تحديد العوامل المؤثرة في متغير الاستجابة. تم تحديد مستوى المعنوية 0.05 لإدخال المتغير الى الانموذج، ومستوى معنوية 0.05 لإبقاء المتغير في الانموذج. تم استخدام اختبار Hosmer

and Lemeshow لدقة التوفيق في اختيار الانموذج النهائي. تم اختيار متغير الاستجابة $I = 1$ (المرضى الذكور المصابين بفقر الدم المزمن) كحدث لغرض نمذجة احتمال الاصابة بمرض فقر الدم المزمن.

في عملية الاختيار المتدرج، يتم حذف المتغيرات غير المعنوية من الانموذج قبل اضافة المتغير المعنوي الى الانموذج، فضلا عن ان حذف او اضافة متغير توضيحي يعتمد على قيمة مربع كاي للباقي. اذ يتم التوقف عن الحذف والاضافة عندما تكون قيمة مربع كاي للباقي أكبر من قيمة يتم تحديدها مسبقا. (في هذا البحث تم تحديد 0.05 لحذف او لإبقاء المتغير في الانموذج). الجدول (2) يبين ملخص لإضافة المتغيرات الى الانموذج، اذ يلاحظ في الخطوة الاولى انه تم تقدير الحد الثابت بمعزل عن بقية المتغيرات التوضيحية، ويلاحظ ان قيمة مربع كاي للباقي 69.4372 بمعنوية احصائية عالية، فضلا عن ان AIC و SC و -2 Log L كانت 195.824 و 198.766 و 193.824 على الترتيب. في الخطوة الثانية، تم اختيار المتغير X_5 في الانموذج باعتباره الأكثر أهمية، ولكنه معنويا احصائيا فقد تم ابقائه في الانموذج. كما يلاحظ بان قيمة مربع كاي للباقي انخفضت الى أكثر من النصف 31.4776، الا انها ذات معنوية احصائية عالية ($p=0.0005$)، فضلا عن ان AIC و SC و -2 Log L كانت 145.084 و 150.968 و 141.084 على الترتيب، وهي اقل مما كانت عليه في الخطوة الاولى. في الخطوة الثالثة، تم اضافة المتغير X_1 الى الانموذج، وعليه أصبح الانموذج بالمتغيرين X_5 و X_1 فضلا عن الحد الثابت. ونظرا لكون المتغيرين معنويين احصائيا، فقد تم ابقائهما في الانموذج. يلاحظ ان قيمة مربع كاي للباقي كانت 29.1556 بمعنوية احصائية عالية، فضلا عن ان AIC و SC و -2 Log L كانت 139.689 و 148.513 و 133.689 على الترتيب، وبتحسن طفيف مما كان عليه في الخطوة الثانية. في الخطوة الرابعة، تم إضافة المتغير X_7 إلى الانموذج. وعليه أصبح الانموذج يحتوي على الحد الثابت والمتغيرات X_1 و X_5 و X_7 وبمعنوية جميع المتغيرات. في هذه الخطوة، يلاحظ ان قيمة مربع كاي للباقي كانت 14.9261 غير معنوية (0.0606)، فضلا عن ان AIC و SC و -2 Log L كانت 127.836 و 139.602 و 119.836 على الترتيب. وعليه، فان الانموذج الاخير تفوق على ما سبق. وبما انه لا يوجد أي من المتغيرات المتبقية خارج الانموذج تفي بمعيار الدخول، فضلا عن ان قيمة مربع كاي للباقي غير معنوية (0.0606). عليه، تم إنهاء الاختيار المتدرج.

الجدول (2) التحليل اللوجستي المتدرج ومعايير المفاضلة بين النماذج التقديرية

Step	entered	Removed effects for the model in Steps	Residual χ^2 Test	Pr > ChiSq	AIC	SC	-2 Log L
1	Intercept		69.437 2	<.0001	195.82 4	198.76 6	193.82 4
2	Intercept and X_5	No effects are removed	31.477 6	0.0005	145.08 4	150.96 8	141.08 4
3	Intercept and X_1, X_5	No effects are removed	29.155 6	0.0006	139.68 9	148.51 3	133.68 9
4	Intercept and X_1, X_5, X_7	No effects are removed	14.926	0.0606	127.83 6	139.60 2	119.83 6

يوضح الجدول (3) القيم التقديرية لمعاملات الانموذج، اذ يلاحظ المعنوية العالية للمتغيرات $X_1 = -2.2354$ (الجنس) (الاناث هي المستوى المرجعي) والمتغير $X_5 = -1.6688$ (نسبة التعداد الشبكي) فضلا عن المتغير $X_7 = -0.5754$ (نقص الحديد في الدم). الجدول (4)، يبين نسبة الترجيح (Odd ratio) وحدود الثقة لكل من X_1 و X_5 و X_7 ، اذ بلغت نسب الترجيح (0.107 و 0.188 و 0.562) على الترتيب. بالنسبة لـ $X_1 = 0.107$ ، وبما ان الإناث هي المستوى المرجعي، فإن نسبة الأرجحية هذه، تعني أن الإصابة بفقر الدم المزمن عند الذكور تبلغ بحدود 1/10 (عُشر) الإصابة عند الاناث، بمعنى اخر، ان الإصابة عند الاناث تبلغ بحدود 10 مرات الإصابة عند الذكور. وبالنسبة لـ $X_5 = 0.188$ فتعني ان التغير في التعداد الشبكي بمقدار وحدة واحدة يؤدي الى زيادة في ارجحية ان المريض مصاب بفقر الدم المزمن بمقدار 0.188، وكذلك $X_7 = 0.562$ تدل على ان التغير في نقص الحديد في الدم بمقدار وحدة واحدة يؤدي الى زيادة ارجحية ان المريض مصاب بفقر الدم المزمن بمقدار 0.562. الشكل (2-a) يوضح نسب الارجحية للمتغيرات المعنوية وحدود الثقة 95%.

الجدول (3) القيم التقديرية لمعاملات الانموذج

Parameter	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	11.1882	2.5534	19.1986	<.0001
X_1 2 vs 1	-2.2354	0.6209	12.9641	0.0003
X_5	-1.6688	0.2658	39.4132	<.0001
X_7	-0.5754	0.1754	10.7640	0.0010

الجدول (4) تقديرات نسبة الترجيح

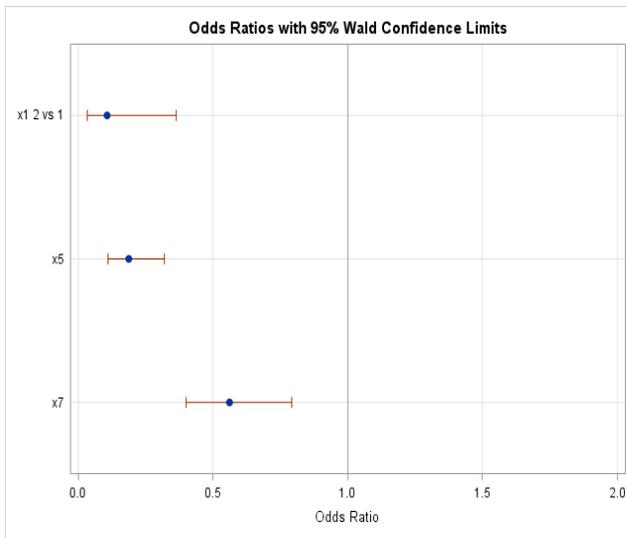
Explanatory variables	Point Estimate	95% Wald Confidence Limits	
X_1 2 vs 1	0.107	0.032	0.361
X_5	0.188	0.112	0.317
X_7	0.562	0.399	0.793

الجدول (5) يبين اختبار Hosmer-Lemeshow (HL) والتي تستند إلى الانموذج المقدر. إذ يوضح الجدول إحصاءة مربع كاي لمقارنة التكرارات المتوقعة مع التكرارات المشاهدة، والتي تبدو فيها التكرارات المتوقعة متقاربة مع التكرارات المشاهدة. من الجدول، يلاحظ ان قيمة χ^2 المستخرجة ($\chi^2 = 12.3685$) غير معنوية عند مستوى المعنوية 0.08907، مما يدل على عدم وجود اختلاف معنوي بين توزيع البيانات قيد الدراسة والتوزيع اللوجستي.

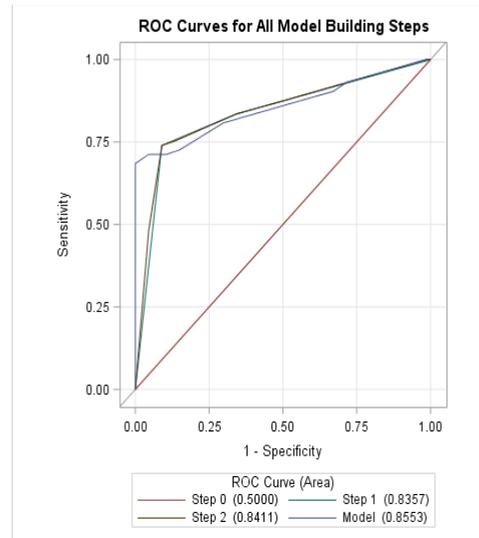
الجدول (5) اختبار Hosmer-Lemeshow (HL) لمقارنة التكرارات المتوقعة مع التكرارات المشاهدة

Partition for the Hosmer and Lemeshow Test					
		Y = 0		Y = 1	
Group	Total	Observed	Expected	Observed	Expected
1	10	2	0.76	8	9.24
2	14	3	1.97	11	12.03
3	5	2	1.03	3	3.97
4	32	7	7.17	25	24.83
5	16	6	7.44	10	8.56
6	13	3	8.34	10	4.66
7	7	7	5.75	0	1.25
8	25	25	23.29	0	1.71
9	18	18	17.23	0	0.77
		Chi-Square	DF	Pr >	ChiSq
		12.3685	7	0.08907	

يبين الشكل (2-b) منحنى خصائص عمليات المُستقبل (ROC) (Receiver Operating Characteristic)، والغرض منه تقييم القدرة التنبؤية للأنموذج فضلا عن تقييم قدرة الانموذج في التصنيف بين الحالات التي تمتلك صفة معينة قيد الدراسة (الحدث-Event) والحالات التي لاتمتلك هذه الصفة (عدم وقوع الحدث-Nonevent). وتقاس هذه الدقة بواسطة المساحة تحت المنحنى، وهي المنطقة الواقعة بين المنحنى ومحور السينات. اذ كلما اقترب المنحنى من أعلى الزاوية اليسرى، بمعنى ان المساحة تقترب من الـ 1، كلما كانت قدرة الانموذج التنبؤية والتصنيفية عالية، بينما تشير المساحة 0.5 الى ان القدرة التنبؤية والتصنيفية للأنموذج ضعيفة⁶. من الشكل (2-b)، يلاحظ خط القطر (45 درجة) والذي يمثل منحنى الـ ROC المتوقع للأنموذج مع الحد الثابت فقط (ويساوي دائما 0.5، ويمثل نصف الرسم)، اي بدون اي قوة تنبؤية. في الخطوة الاولى من عملية الانحدار اللوجستي التدريجي، يلاحظ ان المساحة قد زادت الى 0.8357 بوجود المتغير X_5 وفي الخطوة الثانية ازدادت المساحة بوجود المتغيرين X_1 و X_5 الى 0.8411، واخيرا في الخطوة الثالثة وصلت المساحة تحت المنحنى الى 0.8553 بوجود المتغيرات الثلاث X_1 و X_5 و X_7 ، وهذا يدل على ان قدرة الانموذج التنبؤية والتصنيفية عالية.



(a)



(b)

الشكل (2): (a) وحدود الثقة 95% لنسب الاحرجية، (b) منحنى خصائص عمليات المُستقبل

⁶ بحسب Hosmer-Lemeshow (Hosmer and others, 2013: 162)، إذا كانت المساحة تحت المنحنى أكبر او يساوي الـ 80 وأصغر من 90، فهذا يعني ان قدرة الانموذج على التصنيف ممتازة.

تطبيق الاسلوب البيزي:

يختلف الاسلوب البيزي عن الطرائق التقليدية. اذ ان الاخيرة تقترض أن البيانات هي عينة عشوائية وان المعلمات هي ثابتة. بينما يفترض الاسلوب البيزي أن البيانات والمعلمات عشوائية. وعليه، فان التحليل البيزي يستخدم المعرفة المسبقة للمعلمات فضلا عن التوزيع الشرطي للبيانات لإيجاد التوزيعات الخلفية لمعلمات البيانات المشاهدة. لغرض تقدير المعلمات وفق الاسلوب البيزي، فانه يتطلب تحديد توزيع اولي لكل معلمة من المعلمات غير المعلومة. وعليه، فقد تم استخدام التوزيعات الاولية غير معلومانية والمعلوماتية للحصول على التوزيعات اللاحقة للمتغيرات الثلاثة X_1 و X_5 و X_7 التي تم تحديدها في الانحدار المتدرج لغرض بناء انموذج انحدار لوجستي بيزي. بما ان Y_i متغير ثنائي الاستجابة، يتبع توزيع بيرنولي باحتمالية (P_i) ، اذ ان P_i هي احتمالية ان المتغير Y_i تساوي الـ 1، ويرتبط بمتغيرات الانحدار التوضيحية x بدالة ربط، هي دالة التحويل لوجيت (logit transformation). فان انموذج بيز يمكن كتابته بالشكل الاتي:

$$Pr(\beta | \text{logit}(P_i), x) \propto Pr((\text{logit}(P_i) | x, \beta)) \times Pr(\beta)$$

فيما يتعلق بالتوزيع الاولي غير المعلوماتي، فقد تم افتراض ان المعلمات غير المعلومة تتبع التوزيع المنتظم، وبذلك نضع احتمال متساو لجميع القيم الممكنة التي يمكن أن تأخذها معاملات الانحدار. والافتراض، هو:

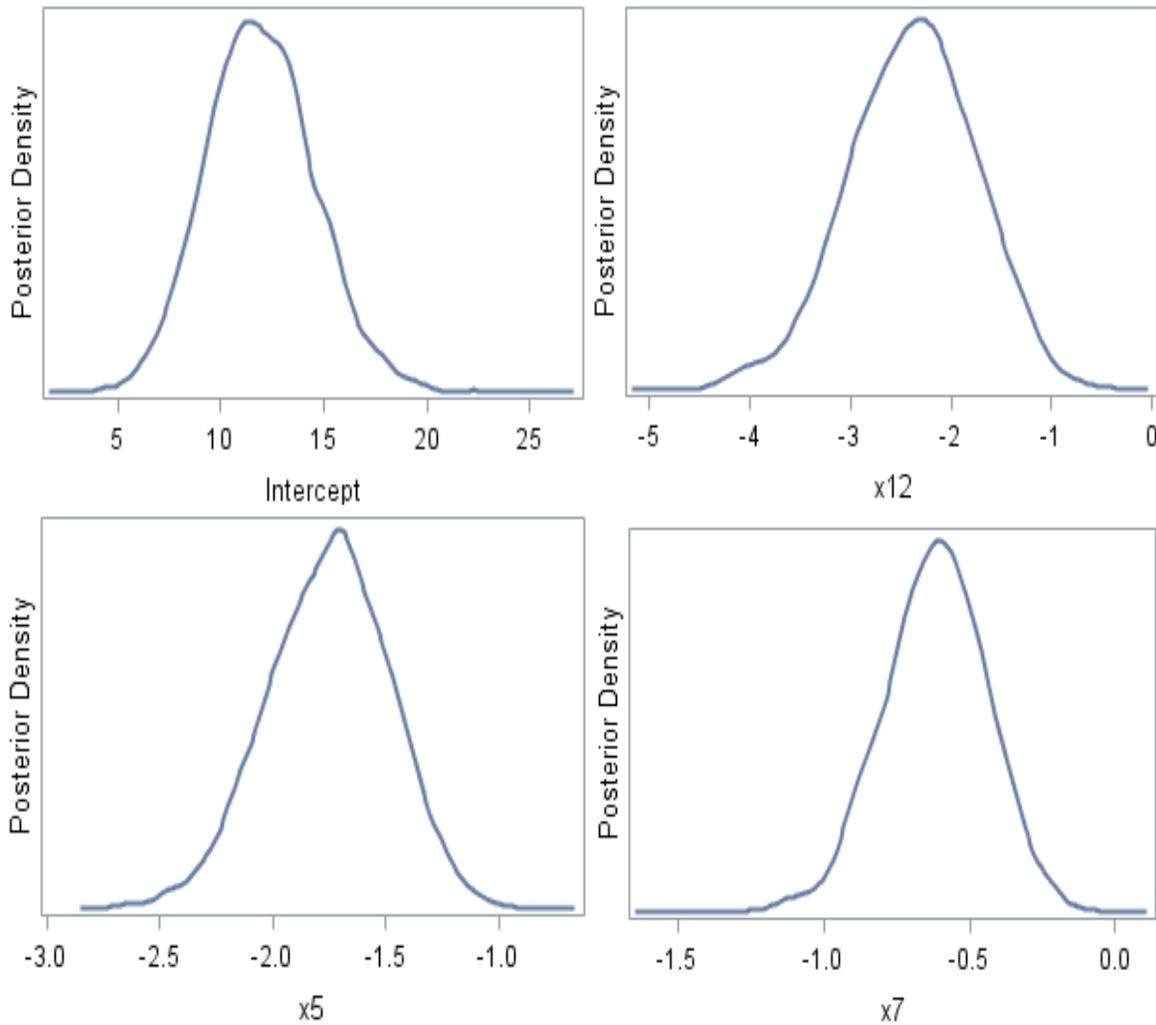
$$\beta_0 \sim U(-10,10), \beta_1 \sim U(-10,10), \beta_5 \sim U(-10,10), \beta_7 \sim U(-10,10)$$

(ترقيم المعلمات بموجب المتغيرات التوضيحية X_1 و X_5 و X_7) نظراً لاستخدام توزيعات سابقة غير معلوماتية لمعاملات الانحدار، فيلاحظ من الجدول (6)، ان الاوساط والانحرافات المعيارية لمعاملات الانموذج هي قريبة من تقديرات الامكان الاعظم. كما يلاحظ من الشكل (3)، ان الرسوم البيانية لدوال الكثافة اللبية الاحتمالية (kernel density plot) المقدره للتوزيع الهامشي الخلفي لمعاملات الانحدار تدعم الاوساط المقدره للمعلمات، اذ يلاحظ ان التوزيع الخلفي للمعلمات لها منوال واحد وشكل التوزيع الطبيعي، وان الاوساط لهذه التوزيعات هي مقارنة للمعلمات المقدره، وهي بذلك تدل على تقارب التوزيع مع المعلمات المقدره.

الجدول (6) اختبار

Parameter	ML		Noninformative Prior	
	Estimate	Standard Error	Estimate	Standard Error
Intercept	11.1882	2.5534	11.9663	2.5503
X_1 2 vs 1	-2.2354	0.6209	-2.3903	0.6178
X_5	-1.6688	0.2658	-1.7584	0.2627

X ₇	-0.5754	0.1754	-0.6207	0.1698
AIC	127.836		127.301	
SC	139.602		138.806	
-2 Log L	119.836		119.301	

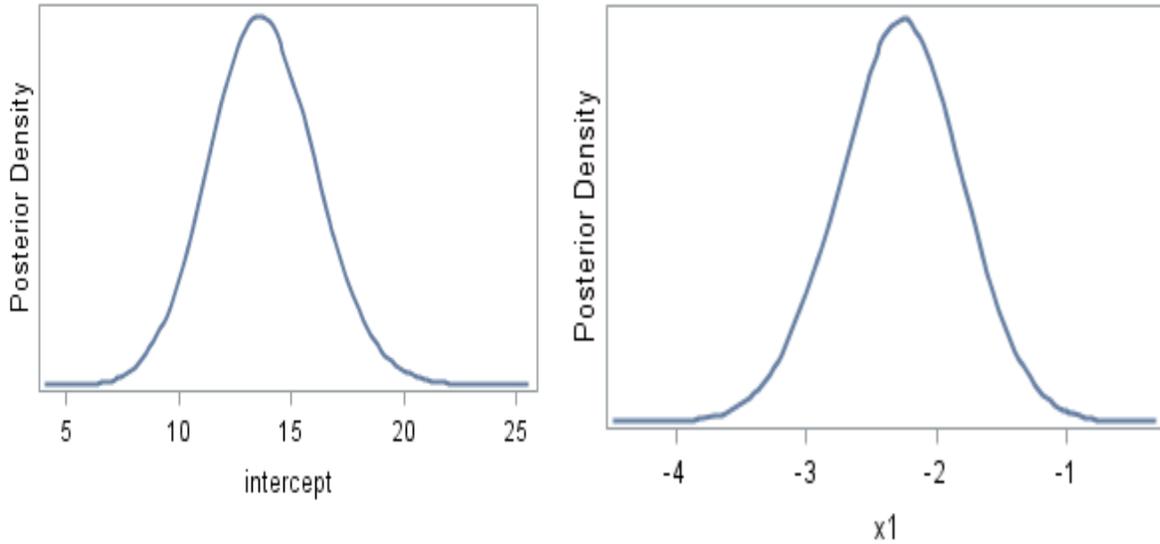


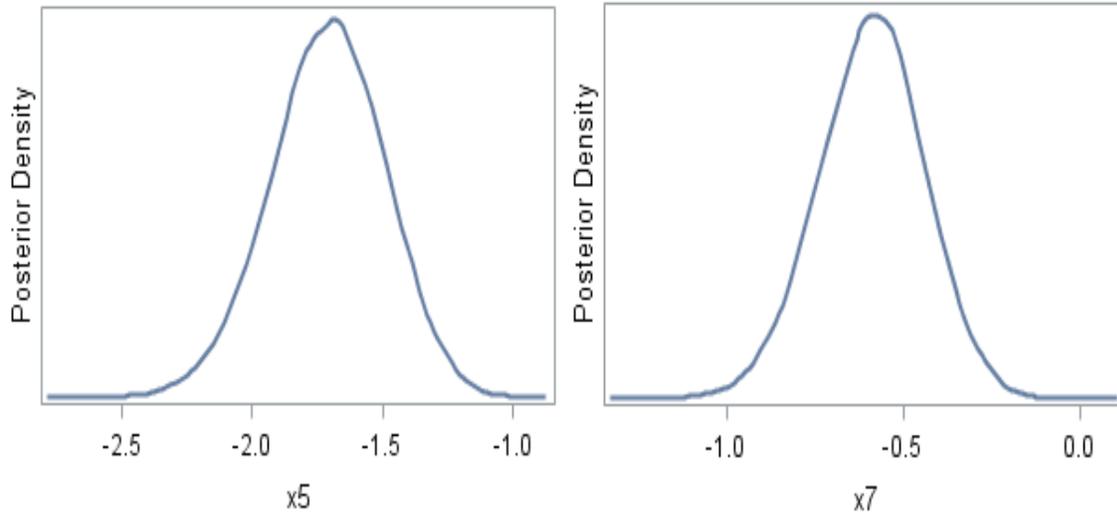
الشكل (3): التمثيل البياني لدوال الكثافة الاحتمالية لمعاملات الانموذج البيزي باستخدام التوزيع الاول غير المعلوماتي فيما يتعلق بالتوزيع الاول المعلوماتي، فقد تم افتراض ان المعلمات غير المعلومة تتبع التوزيع الطبيعي (Normal prior) المقترح من قبل (Sullivan and Greeland). عليه، فقد تم افتراض الاتي:

$$\beta_0 \sim Normal(0,1000), \beta_1 \sim Normal(-2.24,0.621), \beta_5 \sim Normal(-1.67,0.27), \beta_7 \sim Normal(-0.57,0.157)$$

اذ ان معلمات التوزيع هي التي حصلنا عليها من استخدام الانحدار اللوجستي ثنائي الاستجابة، وبالنسبة للمعلمة β_0 فلها توزيع طبيعي بوسط مقداره الصفر وتباين كبير، وتم افتراضه على انه 1000.

من الجدول (7)، يلاحظ ان الانحرافات المعيارية لتقديرات التوزيع الاولي المعلوماتي اقل من مثيلاتها باستخدام الامكان الاعظم، فضلا عن ان AIC و SC و $-2 \log L$ كانت الاقل عند استخدام التوزيعات المعلوماتية مما هو عليه عند استخدام التوزيع الاولي غير المعلوماتي والامكان الاعظم. كما يلاحظ من الشكل (4)، ان الرسوم البيانية لدوال الكثافة اللبية الاحتمالية المقدره للتوزيع الهامشي الخلفي لمعلمت الانحدار تدعم الاوساط المقدره للمعلمت، كما يلاحظ ان التوزيع الخلفي للمعلمت لها منوال واحد وشكل التوزيع الطبيعي، وان شكل التوزيع الطبيعي بدا أكثر نعومة مما كان عليه عند استخدام التوزيع الاولي غير المعلوماتي، وتفسير ذلك، لأنها نتاج دمج معلومات إضافية في التحليل، والذي هو أحد الاختلافات الرئيسة بين الطرائق الكلاسيكية والاساليب البيزية.





الشكل (4): التمثيل البياني لدوال الكثافة الاحتمالية لمعاملات الانموذج البيزي باستخدام التوزيع الاولي المعلوماتي

الجدول (7)

المعاملات التقديرية والانحراف المعياري باستخدام توزيعات احتمالية اولية غير المعلوماتية والمعلوماتية فضلا عن المقدره بطريفة ML

Parameter	ML		Noninformative Prior		Informative Prior	
	Estimate	Standard Error	Estimate	Standard Error	Estimate	Standard Error
Intercept	11.1882	2.5534	11.9663	2.5503	13.7631	2.3562
X ₁ 2 vs 1	-2.2354	0.6209	-2.3903	0.6178	-2.2831	0.4714
X ₅	-1.6688	0.2658	-1.7584	0.2627	-1.7131	0.2191
X ₇	-0.5754	0.1754	-0.6207	0.1698	-0.5906	0.1496
AIC	127.836		127.301		122.077	
SC	139.602		138.806		131.485	
-2 Log L	119.836		119.301		114.077	

الجدول (8) تقديرات نسبة الترجيح وفق التوزيعات الاحتمالية الاولى المعلوماتية

Explanatory variables	Point Estimate	95% Wald Confidence Limits	
X_1 2 vs 1	0.101968	0.074	0.210
X_5	0.180306	0.144	0.269
X_7	0.553995	0.425	0.687

لغرض تقييم الانموذج المقدر بالأسلوب البيزي من حيث مدى دقته وقدرته على التصنيف، فقد تم رسم منحني الـ (ROC)، وكما موضح ذلك في الشكل (5-a). اذ يلاحظ من الشكل ان القدرة التنبؤية والتصنيفية للانموذج اكبر مما كانت عليه باستخدام طريقة الامكان الاعظم، اذ بلغت المساحة تحت المنحني 0.8928 بالأسلوب البيزي، بينما كانت 0.8553 باستخدام طريقة الامكان الاعظم.

كما تم قياس دقة الانموذج وقدرته على التصنيف من خلال جدول التصنيف، وكما موضح ذلك في الجدول (9). اذ يمثل العمود "Prob Level" قيم القطع (Cutoff) للتنبؤ فيما إذا كانت المشاهدة تمتلك الصفة. وتمثل الـ (Event) في العمود "الصحيح-Correct" عدد المشاهدات التي تم التنبؤ بها بشكل صحيح على أنها تمتلك الصفة، وتمثل الـ (Nonevent) عدد المشاهدات التي تم التنبؤ بها بشكل صحيح على انها لا تمتلك الصفة. بينما تمثل الـ (Event) في العمود "غير الصحيح-Incorrect" عدد المشاهدات التي تمتلك الصفة وتم التنبؤ بها بشكل غير صحيح على انها لا تمتلك الصفة، وتمثل الـ (Nonevent) عدد المشاهدات التي لا تمتلك الصفة وتم التنبؤ بها بشكل غير صحيح على أنها تمتلك الصفة. وتعتبر الحساسيه⁷ (Sensitivity) عن دقة التصنيف (القدرة على التنبؤ بحدث ما بشكل صحيح)، كما تعبر النوعية⁸ (Specificity) عن دقة الانموذج (القدرة على التنبؤ بعدم وقوع الحدث بشكل صحيح). وتشير كل من الإيجابية الخاطئة (False positive) عن قياس الخطأ في تصنيف عدم وقوع الحدث، والسلبية الخاطئة (False Negative) عن قياس الخطأ في تصنيف وقوع الحدث.

يلاحظ من الصف الاول من الجدول (9)، ان هناك 73 حالة مرضية بفقر الدم المزمن و67 حالة مرضية بفقر الدم الحاد (وهي اعداد المصابين قيد الدراسة). تم ايجاد نقطة القطع المثالية، ومن الشكل (5-b)، يلاحظ أن نقطة القطع المثالية هي النقطة 0.54. وعند هذه النقطة (انظر الجدول (9))، يلاحظ ان من بين 73 مريضاً مصاباً بفقر الدم المزمن، تم تصنيف 64 منهم بشكل صحيح على أنهم يمتلكون الصفة و9 تم تصنيفهم بشكل غير صحيح على

⁷ الحساسيه (Sensitivity): وهي نسبة المشاهدات في متغير الاستجابة من التي تمتلك الصفة وتم توقعها بشكل صحيح على انها تمتلك الصفة.

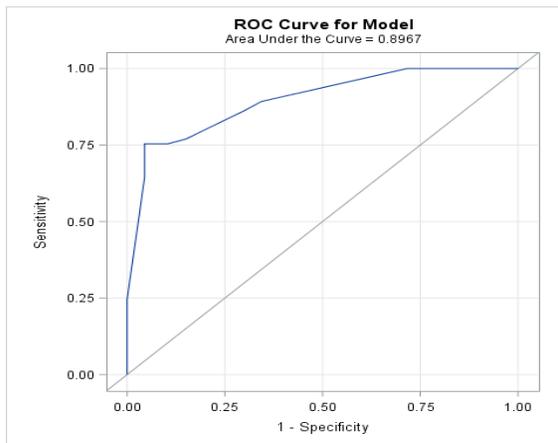
⁸ النوعية (Specificity): وهي نسبة المشاهدات في متغير الاستجابة من التي لا تمتلك الصفة وتم توقعها بشكل صحيح على انها لا تمتلك الصفة. (Hosmer and others, 2013: 160).

انهم لايمتلكون الصفة. كما تم تصنيف 61 مريضا من 67 مريضا مصابا بفقر الدم الحاد بشكل صحيح على انهم لايمتلكون الصفة، وتم تصنيف 6 مرضى بشكل غير صحيح على انهم لايمتلكون الصفة. وعليه، فقد تم تصنيف 125 من أصل 140 مشاهدة بشكل صحيح، مما ينتج عنه نسبة مئوية صحيحة، هي $(64 + 61)/140 = 89.2857\% = 0892857$ (وهي نفس المساحة تحت منحنى الـ ROC).

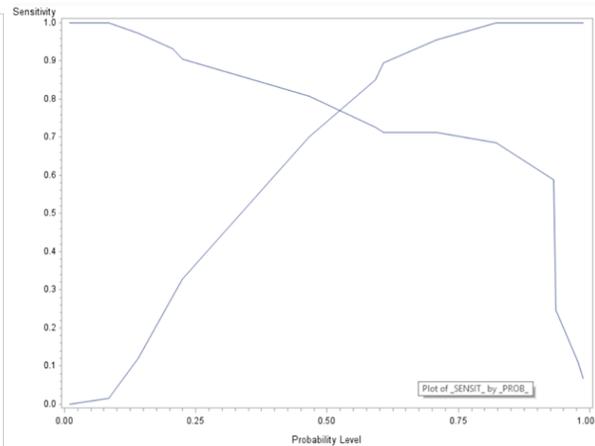
الحساسية هي نسبة عدد المشاهدات التي تمتلك الصفة والمصنفة بشكل صحيح على إجمالي عدد المشاهدات التي تمتلك الصفة، ويلاحظ عند مستوى القطع 0.54، ان الحساسية تبلغ 87.7% (أي 64/73)، اذ تم تصنيف 64 من أصل 73 ممن يمتلك الصفة على أنها تمتلك الصفة. والدقة هي نسبة عدد المشاهدات التي لا تمتلك الصفة والمصنفة بشكل صحيح على إجمالي عدد المشاهدات التي لا تمتلك الصفة. ويلاحظ عند مستوى القطع 0.54 ان الدقة هي 91.0% (أي 61/67)، بمعنى اخر، تم تصنيف 61 من أصل 67 من التي لا تمتلك الصفة وبشكل صحيح على انها لا تمتلك الصفة. كما يلاحظ انه قد تم تصنيف 6 مشاهدات بشكل غير صحيح على انهم مرضى بفقر الدم المزمن بنسبة خطأ إيجابي قدره 8.5% (أي 6/(6+64))، وتم تصنيف 9 مرضى بشكل غير صحيح على أنهم مرضى بفقر الدم الحاد بنسبة خطأ سلبى قدره 12.8% (أي 9/(61+9))، ويلاحظ انهما نسب منخفضة. عليه، يمكن القول وبموجب المعطيات اعلاه، ان الانموذج فعال جدا للتنبؤ والتصنيف.

جدول (9): جدول التصنيف

Prob Level	Correct		Incorrect		Percentages				
	Event	Non-Event	Event	Non-Event	Correct	Sensitivity	Specificity	False POS	False NEG
0.000	73	0	67	0	52.1	100.0	0.0	47.9	.
0.540	64	61	6	9	89.2857	87.7	91.0	8.5	12.8



(a)



(b)

الشكل (5): (a) منحني خصائص عمليات المُستقبل; (b) الحساسية والنوعية مقابل جميع نقاط القطع المحتملة

الجدول (10) يوضح الاحتمالية التنبؤية (\hat{P}) لمتغير الاستجابة وفقا للمتغيرات التوضيحية المؤثرة في التحليل. بالنسبة للصف الأول، يلاحظ بان القيمة الاحتمالية التنبؤية (\hat{p}) كانت 0.519184، هذا يعني أن الاحتمالية التقديرية بان المريض مصاب بفقر الدم المزمن ($Y = 1$) هو 0.519184 وفق قيم المتغيرات التوضيحية ذات العلاقة. وبالمقابل فان الاحتمالية التقديرية بان المريض مصاب بفقر الدم الحاد ($Y = 0$) هو 0.480816، وهكذا بالنسبة لبقية القيم.

الجدول (10) الاحتمالية التنبؤية (\hat{P}) لمتغير الاستجابة وفقا للمتغيرات التوضيحية المؤثرة في التحليل

\hat{P}_i	\hat{P}_i	\hat{P}_i	\hat{P}_i	\hat{P}_i	\hat{P}_i
0.519184	0.075724	0.169245	0.167105	0.069924	0.927941
0.855631	0.069910	0.800818	0.782407	0.309632	0.168232
0.775608	0.853900	0.777576	0.519268	0.017675	0.069833
0.777532	0.923891	0.776176	0.061446	0.523115	0.779959
0.780615	0.374410	0.061928	0.310730	0.167711	0.417653
0.858576	0.307737	0.072188	0.777473	0.775971	0.854742
0.773839	0.524471	0.060681	0.072037	0.170811	0.075045
0.521261	0.780996	0.783404	0.855773	0.062918	0.922892
0.306809	0.016174	0.522310	0.070933	0.995694	0.069906
0.069160	0.776741	0.779021	0.523634	0.071399	0.023077
0.074848	0.778912	0.018377	0.069743	0.926089	0.783019
0.922435	0.853263	0.921120	0.777292	0.069346	0.074128
0.058785	0.802020	0.853011	0.419159	0.858772	0.024616
0.518881	0.071560	0.779531	0.014119	0.798648	0.857634
0.017992	0.921604	0.073192	0.853297	0.062981	0.060033
0.420268	0.518552	0.780152	0.779638	0.371540	0.775565
0.799681	0.778734	0.070604	0.167386	0.519494	0.419753
0.520718	0.308362	0.798432	0.069426	0.070262	0.856703
0.170232	0.925675	0.521582	0.518819	0.777452	0.726328

0.775773	0.060117	0.855428	0.780751	0.060193	0.026445
0.920790	0.070740	0.072367	0.776533	0.025776	0.896343
0.781307	0.522606	0.781569	0.060891	0.859972	0.921390
0.522447	0.777736	0.074816	0.856200	0.519993	0.783836

3. الاستنتاجات:

- من خلال النتائج التي تم التوصل اليها من التطبيق العملي، يمكن استنتاج الآتي:
- (a) تعطي طريقة الانحدار اللوجستي ثنائي الاستجابة تقديرات مشابهة لمعاملات التوزيع الاولي غير المعلوماتي مع تفوق بسيط للطريقة الاخيرة من ناحية الانحرافات المعيارية للمعاملات المقدره.
- (b) تفوق التوزيع الاولي المعلوماتي المستند الى تقديرات معاملات الانحدار اللوجستي بشكل كبير على التقديرات الاخرى من ناحية الانحرافات المعيارية للمعاملات المقدره ومعايير دقة التوفيق.
- (c) تعطي التقديرات البيزية على اختلاف التوزيعات الاولية المستخدمة نتائج أفضل من تقديرات الامكان الاعظم.
- (d) ان اهم المتغيرات المؤثرة في فقر الدم عند الذكور (باعتبار ان الاناث هي المستوى المرجعي)، هي الجنس ونسبة التعداد الشبكي فضلا عن نقص الحديد في الدم.

المصادر:

أولاً. المصادر العربية

البحوث والمجلات

- 1- قاسم، محمد نذير إسماعيل ايمان طارق فتحي "حول تقدير بيز في النماذج الخطية المختلطة باستخدام معاينة جيبس"، المجلة العراقية للعلوم الاحصائية السنة 2008: المجلد 8: العدد: 13: ص 34-10.
- 2- العزي، محمد شاكر (2017)، مقارنة عملية التصنيف بأسلوب الدالة المميزة الخطية والانحدار اللوجستي بوجود مشكلة التعدد الخطي مع تطبيق، رسالة ماجستير-كلية الادارة والاقتصاد-جامعة بغداد.
- 3- العزي، محمد شاكر والبكري، رباب عبد الرضا صالح، (2017)، "التحليل المميز والانحدار اللوجستي بوجود مشكلة التعدد الخطي (دراسة تطبيقية على مرض فقر الدم)"، مجلة العلوم الاقتصادية والإدارية، العدد 29، المجلد 23، 373-397.
- 4- علي اسماعيل عبد الصمد، الدريني، محمود محمد، "التنبؤ بالرضا الوظيفي لدى العاملين في محطة كهرباء غرب القاهرة باستخدام النموذج اللوجستي بطريقة بيز"، الإداري، 2017، العدد 150، ص 47-98.

ثانياً. المصادر الأجنبية

1. الكتب

- 1- Alder, L. and A. Tambe (2019). Acute Anemia. StatPearls [Internet], StatPearls Publishing.
- 2- Bolstad, W. M. and J. M. Curran (2016). Introduction to Bayesian statistics, John Wiley & Sons.
- 3- Box, G. E. and G. C. Tiao (2011). Bayesian inference in statistical analysis, John Wiley & Sons.
- 4- Hosmer Jr, D. W., S. Lemeshow and R. X. Sturdivant (2013). Applied logistic regression, John Wiley & Sons.
- 5- Montgomery, D. C., E. A. Peck and G. G. Vining (2012). Introduction to linear regression analysis, John Wiley & Sons.
- 6- Stigler, S. M. (1986). The history of statistics: The measurement of uncertainty before 1900, Harvard University Press.

2. البحوث و المجلات

- 1- Bekele, A., et al. (2016). "Prevalence of anemia and Its associated factors among pregnant women attending antenatal care in health institutions of Arba Minch town, Gamo Gofa Zone, Ethiopia: A Cross-sectional study." Anemia 2016.
- 2- Camaschella, C. (2015). "Iron-deficiency anemia." New England journal of medicine 372(19): 1832-1843.

- 3- Che, X. and S. Xu (2010). "Bayesian data analysis for agricultural experiments." Canadian journal of plant science **90**(5): 575-603.
- 4- d'Estaintot, B. L., P. Santambrogio, T. Granier, B. Gallois, J. M. Chevalier, G. Precigoux, S. Levi and P. Arosio (2004). "Crystal structure and biochemical properties of the human mitochondrial ferritin and its mutant Ser144Ala." Journal of molecular biology **340**(2): 277-293.
- 5- Fishbane, S., & Spinowitz, B. (2018). Update on anemia in ESRD and earlier stages of CKD: core curriculum 2018. American Journal of Kidney Diseases, 71(3), 423-435.
- 6- Geiziane Oliveira, SAS Institute, Brazil; George von Borries, Universidade de Brasília, Brazil; Priscila Zaczuk Bassinello, Embrapa Rice and Beans, Brazil." Comparing Priors in Bayesian Logistic Regression for Sensorial Classification of Rice".
- 7- Gelman, A., A. Jakulin, M. G. Pittau and Y.-S. Su (2008). "A weakly informative default prior distribution for logistic and other regression models." The Annals of Applied Statistics **2**(4): 1360-1383.
- 8- Gunes, F. and F. Chen Getting started with the MCMC procedure, Citeseer, 2014.
- 9- Ibrahim, J. G. and M.-H. Chen (2000). "Power prior distributions for regression models." Statistical Science **15**(1): 46-60.
- 10- Kaur, K. (2014). "Anaemia 'a silent killer' among women in India: Present scenario." European Journal of Zoological Research 3(1): 32-36.
- 11- Mettananda, S., et al. (2018). "Anaemia among females in child-bearing age: Relative contributions, effects and interactions of α - and β -thalassaemia." PloS one 13(11): e0206928.
- 12- Nelder, J. A. and R. W. Wedderburn (1972). "Generalized linear models." Journal of the Royal Statistical Society: Series A (General) **135**(3): 370-384.
- 13- Oliveira, G., Borries G. and Bassinello P. "Comparing Priors in Bayesian Logistic Regression for Sensorial Classification of Rice". SAS Institute Inc, Brazil, (2017).
- 14- Sas Institute. SAS/STAT 9.22 User's Guide: The LOGISTIC Procedure (book Excerpt). Books24x7. com, 2010.
- 15- Schworer, A. and P. Hovey (2004). "Newton-raphson versus fisher scoring algorithms in calculating maximum likelihood estimates."
- 16- Stengle, J., M. Strumia, T. J. Liddy and G. Brecher (1965). "Mean corpuscular Haemoglobin concentration (MCHC) and mean corpuscular volume (MCV) as biologic constants." Bibliotheca haematologica **21**: 2014.
- 17- Stokes, Maura, Fang Chen, and Funda Gunes. "An introduction to Bayesian analysis with SAS/STAT® software." Proceedings of the SAS Global Forum 2014 Conference, SAS Institute Inc, Cary, USA .
- 18- Sullivan, S. G. and S. Greenland (2012). "Bayesian regression in SAS software." International journal of epidemiology **42**(1): 308-317.
- 19- Tolosano, E. (2015). Increasing serum transferrin to reduce tissue iron overload due to ineffective erythropoiesis, Haematologica.
- 20- Wollmann, M., B. M. C. Gerzson, V. Schwert, R. W. Figuera and G. de Oliveira Ritzel (2014). "Reticulocyte maturity indices in iron deficiency anemia." Revista brasileira de hematologia e hemoterapia **36**(1): 25-28.