العلوم الاقتصادية
العدد الخامس والسبعون/ المجلد التاسع عشر/ كانون الاول 2024

اختيار المتغيرات لبيانات البقاء ضمن توزيع ويبل باستخدام خوارزمية السنجاب

المدرس الدكتورعمر رمزي جاسم
قسم ادارة الاعمال، كلية الادارة والاقتصاد، جامعة الحمدانية، العراق
المدرس الدكتور سرمد عبدالخالق صالح
قسم الرياضيات، كلية التربية للعلوم الصرفة، جامعة الحمدانية، العراق.

تواريخ البحث
تاريخ تقديم البحث : 2024/7/25
تاريخ قبول البحث :2024/8/18
تاريخ رفع البحث على الموقع:
2024/12/15

المستخلص :

يُعد نموذج انحدار وايبل أحد أكثر أشكال نموذج الانحدار المعلمية شيوعاً، حيث يوفر تقديرًا لدالة المخاطرة الأساسية، بالإضافة إلى معاملات المتغيرات المشتركة. نظرًا للصعوبات الفنية الناتجة من استخدام هذا النموذج لذلك نادرًا ما يتم استخدام نموذج الانحدار وايبل في الأدبيات الطبية مقارنة بنموذج الانحدار اللامعلمي. تهدف هذه الدراسة الى تقديم بعض الجوانب النظرية المهمة لنموذج انحدار وايبل وكيفية مطابقة البيانات لهذا النموذج بالاضافة الى استخدام خوارزمية بحث السنجاب ومقارنتها مع طرائق اخرى في إختيار المتغيرات في نموذج انحدار وايبل باستخدام البيانات الحقيقة التي تم جمعها من مصابين بمرض العجز الكلوي المزمن والذين يتعالجون بالغسيل الكلوي المستمر، وقد شخصت حالة المرضى من قبل اطباء مختصين بالتعاون مع مستشفى ابن سينا التعليمي – وحدة الكلية الاصطناعية. فقد أظهرت النتائج تفوق خوارزمية بحث السنجاب مقارنةً بطرائق اختيار المتغيرات الأخرى.

الكلمات المفتاحية: خوارزمية بحث السنجاب، نموذج انحدار وايبل، اختيار المتغيرات، بيانات البقاء.

Variable Selection for Survival Data Under Weibull Distribution by Using Squirrel Algorithm

Lecturer, PhD Omar Ramzi Jassim

Department of Business Administration, College of Administration and Economics

Lecturer, PhD Sarmad Abdulkhaliq Saleh

College of Admin & Economic / Banking and Financial Sciences Dep./ **University of Basrah**

Abstract :

Weibull Regression Model (WRM) is one of the most common forms of parametric regression model, as it provides an estimate of the basic hazard function, In addition to the coefficients of the covariates. Due to the technical difficulties resulting from using this model, the Weibull regression model is rarely used in the medical literature compared to the nonparametric regression model. This study aims to present some important theoretical aspects of the Weibull regression model and how to fit data to this model, in addition to using the Squirrel Search Algorithm (SSA) and comparing it with other methods in selecting variables in the Weibull regression model using real data collected from patients with chronic kidney disease who are treated with continuous hemodialysis, and the patients' condition was diagnosed by specialized doctors in cooperation with Ibn Sina Teaching Hospital - Artificial Kidney Unit. The results showed the superiority of the squirrel search algorithm compared to other methods of selecting variables.

Keywords: Squirrel search algorithm, Weibull regression model, variable selection, survival data.

المقدمة:

يُعد تحليل الانحدار احد الادوات الاحصائية المهمة في بناء النموذج الاحصائي، أذ يتم من خلاله تقدير العلاقة بين متغير واحد يُدعى بالمتغير التابع ومتغير اخر او عدة متغيرات اخرى تُدعى بالمتغيرات التوضيحية (التفسيرية)، بحيث ينتج من خلالها معادلة احصائية توضح العلاقة بين متغيرات الدراسة[15]. لقد إحتل تحليل الانحدار بنماذجه المختلفة مكانة متميزة في توجهات العديد من علماء الإحصاء، ونالت نصيبها الوافر عبر المؤلفات الاحصائية المختلفة، واصبح دورها مهم جداً في تطبيقات علوم الحياة المتنوعة خصوصاً في المجال الاقتصادي الذي أخذ على عاتقه اعتماد نماذج الانحدار بالدرجة الاساس لتكون ابرز وسائل الدعم العملي للنظريات الاقتصادية، فضلاً عن العلوم الاخرى كالصحية والحياتية والاجتماعية وغيرها [4].

يفترض أنموذج الانحدار الخطي الاعتيادي أن متغير الاستجابة يعتمد على مجموعة من المتغيرات التوضيحية، بحيث يمكن أن تكون هذه المتغيرات عبارة عن متغيرات مستمرة أو متغيرات متقطعة قابلة للعد، ومع ذلك، عندما يكون متغير الاستجابة بشكل متغيرات قابلة للعد مثل عدد المرضى، فأنه سوف لن تتحقق افتراضات الانحدار الخطي. لذلك تم اقتراح أنموذج انحدار وايبل كأحد نماذج الانحدار اللوغاريتمية الخطية التي تتوافق مع هكذا حالات[3,4].

عليه فأن أنموذج انحدار وايبل من النماذج اللوغاريتمية الخطية، وبنفس نماذج الانحدار الخطية الاخرى فأن نموذج انحدار وايبل قد يحتوي على العديد من المتغيرات المستقلة التي تؤثر سلباً على دقة النموذج وبساطته في تفسير النتائج، وبناءاً على ذلك تُعد عملية اختيار المتغيرات أحد أصعب جوانب بناء النموذج. غالباً ما يُنصح بأن يكون الاختيار للمتغير أكثر تركيزًا على المعرفة السريرية والأدبيات السابقة من التركيز على طرق الاختيار الإحصائية وحدها[17,16].

وغالباً ما تحتوي البيانات على العديد من المتغيرات الإضافية التي لا تُستخدم في نهاية المطاف في تطوير النموذج الخاص بالدراسة. لذلك يجب اختيار المتغيرات المناسبة بعناية لتجنب تضمين متغيرات الضوضاء (المتغيرات عديمة الفائدة) في النموذج النهائي، حيث اصبحت الاساليب التقليدية لاختيار المجموعات الجزائية مثل طريقة الاختيار الامامية (Forward Selection) و طريقة الاختيار الى الخلف (Backward Elimination) و طريقة الاختيار التدريجية (Stepwise Selection) غير جيدة في اداء وظيفتها حيث اصبحت اكثر تكلفة في حسابها، اضافة الى ذلك فان معايير المعلومات لاختيار المتغيرات مثل معيار أكاكي للمعلومات ((Akaike Information Criterion (AIC) ومعيار بيز للمعلومات ((Bayesian Information Criterion (BIC) اصبحت غير عملية في اختيار المتغيرات التوضيحية وذلك بسبب تعقيدها الحسابي الذي ينمو بشكل طردي مع ازدياد عدد المتغيرات التوضيحية [12-10].

ولأهمية كل من أنموذج الانحدار الشبه المعلمي وأنموذج الانحدار المعلمي والذين يتميز احداهما على الاخر بطبيعة معاملاتهم، أذ يركز ألانموذج شبه المعلمي على تأثير المخاطرة للمتغيرات المشتركة، كذلك في حساب كل من شكل نموذج توزيع وقت البقاء ((survival time (ST) [9-7]. في حين يمكننا وبطريقة اسهل استخدام الانموذج المعلمي لكافة المتغيرات وحساب شكل التوزيع وقت البقاء، أذ يتميز النموذج المعلمي لتحليل وقت البقاء على قيد الحياه بمجموعة من المميزات يمكن توضيحها بالاتي:

- يمكن من خلاله تحديد كل من التوزيع وقت البقاء.[6]
- يمكن استخدام الطرق المعلمية في تقدير ومن اشهرها طريقة الامكان الاعظم (Maximum Likelihood Method) لتقدير معلمات النموذج قيد الدراسة.
- يمكن تمثيل الاخطاء (Residuals) بالفرق بين قيم الوقت المرصودة والمقدرة.[6]

واخيراً هناك مجموعة كبيرة ومتنوعة من النماذج التي يتم من خلالها تحديد وقت الفشل المتسارع (Accelerated Failure Time Model (AFT)) بما في ذلك نماذج الانحدار الأسي ((ERM) Exponential Regression Models) ونموذج الانحدار اللوجستي ((LLRM) Log-Logistic Regression Models) وغيرها من النماذج الاخرى.[5,15]

لقد تناولت الدراسة الحالية أنموذج انحدار وايبل ((WRM) Weibull Regression Model) الذي يُعتبر أحد النماذج الأكثر شعبية بين النماذج التي لديها متغير استجابة قابل للعد, أذ يُعد أحد أكثر نماذج الانحدار المعلمية شيوعًا، أذ أنه يوفر تقديراً اساسياً لدالة المخاطرة ((Hazard Function (HF)، بالإضافة إلى كافة معاملات التغاير المشترك (Coefficients for Covariates).[5] ونظراً للصعوبات التقنية التي يمكن أن نواجهها من استخدام هذا النموذج في الادبيات الطبية، لذلك نادراً ما يتم استخدام هذا النموذج ويتم الاتجاه الى بعض النماذج اخرى اكثر بساطتاً لتجاوز هذا الاشكال.[3,4]

1. أنموذج انحدارو ايبل: Weibull regression model

يُعد أنموذج انحدار وايبل أحد أهم نماذج الانحدار اللوغاريتمية الخطية, وهو الأداة التي يتم من خلالها نمذجة المتغير المعتمد عندما تكون قيم ذلك المتغير على شكل قيم قابلة للعد.[16] وكغيره من سائر نماذج الانحدار, قد يحتوي النموذج على متغيرات مستقلة كثيرة ما يؤثر سلباً على دقة النموذج وبساطته في تفسير النتائج.[5]

يُستخدم نموذج انحدار وايبل في مجموعة متنوعة من الحقول العلمية بما في ذلك الطب، والهندسة، والعلوم الاجتماعية. ففي مجال الطب، يُمكن استخدام نموذج انحدار وايبل لتحليل بيانات الأبحاث السريرية وتقدير معدلات الفشل للعلاجات أو الأدوية. كما يُمكن استخدامها في تحليل بيانات التجارب السريرية لتقدير تأثير العوامل المستقلة على نتائج الدراسات.[2]

في مجال الهندسة، يُمكن استخدام نموذج انحدار وايبل لتحليل بيانات الاختبارات التجريبية وتقدير المعدلات المتوقعة للفشل في المواد أو الهياكل. كما يُمكن استخدامها في تقدير عمر الخدمة المتوقع للأنظمة والأجهزة.

وفي مجال العلوم الاجتماعية، يُمكن استخدام نموذج انحدار وايبل لتحليل بيانات استطلاعات الرأي والتوقعات وتقدير تأثير المتغيرات المستقلة على سلوك المجتمع.

في بداية الامر يجب التطرق على كافة الخصائص والمميزات للنموذج المفترض لهذة الدراسة, أذ يمكن تمثيل الهيكل الاساسي لنموذج الانحدار وايبل عندما يفترض بان توزيع الحدث (T) هو دالة ذات متغير مشترك واحد ويُعبر عن ذلك بالشكل الاتي:[16,17]

$$\ln(T) = \beta_0 + \beta_1 x + \sigma_\varepsilon \quad \dots (1)$$

حيث أن:

β_0 : يمثل الوسط الحسابي للبيانات.

β_1 : هو معامل المتغير المشترك المقابل.

ε : يمثل الخطأ الناتج ويتبع توزيع وايبل أو توزيع القيم المتطرفة الادنى (extreme minimum value distribution) وبالمعلمات $(G(0,\sigma))$ أذ أن (σ) تمثل معلمة الشكل للنموذج (Shape Parameter of modle).

وكذلك يُسمى أنموذج انحدار وايبل في بعض المصادر بنموذج وقت الفشل المتسارع (Accelerated Failure Time Model) وذلك لأن تأثير المتغير المشترك يتضاعف على النطاق الزمني ويقال إنه "يسرع" من وقت البقاء. في المقابل، فإن تأثير المتغير المشترك مضاعف على مقياس المخاطرة (Hazard Scale) في ايجاد المخاطرة النسبية (Proportional Hazards Model).[2]

وعليه يمكن تمثيل دالة المخاطرة باستخدام نموذج انحدار وايبل على شكل المخاطرة المتناسبة كالاتي:[16]

$$h(t,x,\beta,\lambda) = \lambda t^{\lambda-1} e^{-\lambda(\beta_0+\beta_1 x)} \quad (2)$$

$$h(t,x,\beta,\lambda) = \lambda t^{\lambda-1} e^{-\lambda\beta_0} * e^{-\lambda\beta_1 x} \quad (3)$$

$$h(t,x,\beta,\lambda) = \lambda\gamma t^{\lambda-1} e^{-\lambda\beta_1 x} \quad (4)$$

$$h(t,x,\beta,\lambda) = h_0(t) * e^{\theta_1 x} \quad (5)$$

حيث أن $\gamma = e^{\frac{-\beta_0}{\sigma}} = e^{\theta_0}$ وأن $\theta_1 = \frac{-\beta_1}{\sigma}$ ، في حين أن دالة المخاطرة الاساسية (Baseline Hazard Function) يُمكن أن تُمثل بالشكل الاتي:

$$h_0(t) = \lambda\gamma t^{\lambda-1} \quad (6)$$

وأن (σ) هو معامل شبيه بالتباين على مقياس الوقت اللوغارتمي.

$\gamma = \frac{1}{\sigma}$ وعالباً ما تسمى بمعلمة القياس (Measurement Parameter). وأن (λ) عبارة عن معلمة الشكل(Shape Parameter). أما (θ_1) تمثل نسبة المخاطرة (Hazard Ratio).

وعليه يُمكن كتابة دالة المخاطرة لوقت الفشل المتسارع على النحو الاتي:

$$h(t,x,\beta,\lambda) = \lambda t^{\lambda-1} e^{-\lambda(\beta_0+\beta_1 x)}$$

$$h(t,x,\beta,\lambda) = \lambda\gamma\left(t * e^{-\beta_1 x}\right)^{\lambda-1} * e^{-\beta_1 x} \quad (7)$$

من خلال ذلك يمكن كتابة نموذج انحدار وايبل في كل من الأشكال المتسارعة والمتناسبة، مما يسمح بالوصف المتزامن لتأثير دالة المخاطرة (HR) والتغير النسبي في وقت البقاء او ما يُسمى بنسبة وقت الحدث ((Event Time Ratio (ETR)).

2. ملاءمة البيانات لنموذج انحدار وايبل:

الخطوة الاولى لتحليل البيانات باستخدام انموذج انحدار وايبل هو مطابقة البيانات لتوزيع وايبل، أذ يجب أن تتوزع بيانات متغير الاستجابة توزيع وايبل (Weibull distribution)، ولتنفيذ هذه العملية يتم استخدام اختبار حسن المطابقة المتوفر في البرمجية الجاهزة (matlab v.2024). كذلك يمكن التحقق من ملاءمة أنموذج انحدار وايبل مع المتغيرات التصنيفية من خلال منحنيات كابلان ماير الطبقية (Stratified Kaplan-Meier Curves). أذ سيظهر الرسم بياني لوقت البقاء اللوغاريتمي مقابل $[\log[-\log(KM)]]$ خطوطًا خطية ومتوازية إذا كان النموذج ملائماً.[16]

3. خوارزمية السنجاب: Squirrel algorithm

تعتبر خوارزمية السنجاب ((Squirrel algorithm (SA) أحد الخوارزميات التطورية الحديثة والتي يمكن أن تستخدم في اختيار المتغيرات الافضل من بين مجموعة من المتغرات. تقوم هذه الخوارزمية بتحليل البيانات واختيار المتغيرات الأكثر أهمية بناءً على مجموعة من القيم المحسوبة والمعايير المحددة مسبقاً.[2] أذ تُعد خوارزمية السنجاب من الخوارزمية غير الحدسية (Non- Heuristic Algorithms) في حل مشاكل الامثلية العالمية (Global Optimization)، فكرة الخوارزمية مستوحاه من السلوك الطبيعي للسنجاب في الية البحث عن الغذاء. أثناء الخريف، تبحث السناجب عن مصادر الغذاء من خلال الزحف من شجرة إلى أخرى. وفي الوقت نفسه، تغير منطقتها وتستكشف مناطق مختلفة من الغابات. أذ تقوم الخوارزمية بتحديث أماكن السناجب وفقًا لموسم المد والجزر ونوع السناجب وكالاتي:[13]

1.4: تكوين المجتمع:

بافتراض أن عدد المجتمع هو (N)، فإن الحدود العليا والسفلى لمساحة المطاردة يمكن تمثيلها ب(X_U, X_L). علية يمكن توليد مجتمع عشوائي ذو حجم (N) من السناجب على النحو التالي:[13]

$$X_i = X_L + rand(1, D) * (X_U - X_L) \qquad (8)$$

حيث يُشير (X_i) إلى السنجاب ذو التسلسل (i^{th}) وأن $(i = 1,2, \dots, N)$ و(rand) دالة تعمل على توليد رقم عشوائي بين (0 و1) في حين أن (D) هو قياس المشكلة.

2.4: تجميع المجتمع:

تتطلب خوارزمية بحث السنجاب (SSO) وجود سنجاب واحد فقط عند كل شجرة، مع افتراض أن العدد الكلي للسناجب هو (N)، وبالتالي، يوجد (N) شجرة في الغابة. تحتوي الغابة على انواع مختلفة من الاشجار لذلك يمكن تقسيم السناجب إلى ثلاثة أنواع:[13]

- السناجب الموجودة عند أشجار جوز $(\mathrm{W_h})$.
- السناجب الموجودة عند أشجار بذور البلوط (W_a).
- السناجب الموجودة عند الأشجار العادية $(\mathrm{W_n})$.

3.4: تحديث موقع السناجب:

تقوم السناجب بتحديث أوضاعها من خلال التوجه إلى أشجار الجوز أو أشجار بذور البلوط على النحو التالي:

$$X_i^{t+1} = \begin{cases} X_i^t + d_g G_c (X_{ai}^t - X_i^t) & if\ r_1 \geq P_{dg} \\ Random\ Location & otherwise \end{cases} \qquad (9)$$

$$X_i^{t+1} = \begin{cases} X_i^t + d_g G_c (X_h^t - X_i^t) & if\ r_2 \geq P_{dg} \\ Random\ Location & otherwise \end{cases} \qquad (10)$$

عادتاً ما يتم افتراض أن قيمة احتمالية المطاردة $(P_{dg} = 0.1)$، فأذا كانت $(r > P_{dg})$ معنى ذلك بعدم وجود مطاردة للسنجاب وامكانية حركتها بحرية للبحث عن الطعام، في حين اذا كانت $(r < P_{dg})$ عند ذلك يكون هنالك مطاردة للسنجاب، وتضطر السناجب إلى الحد من نطاق الأنشطة، وتكون السناجب معرضة للخطر، ويتم نقل مواقعها بشكل عشوائي. أما قيمة (d_g) يمكن تحديده بواسطة المعادلة التالية:[13]

$$d_g = \frac{h_g}{tan\emptyset} \qquad (11)$$

حيث أن(h_g) هو الرقم المقدر الثابت والذي يساوي(8)، في حين أن $(tan\emptyset)$ يشير إلى نقطة الانحدار التي يمكن تحديدها بواسطة $(tan\emptyset = D/L)$، ويمكن تقدير قوة السحب وقوة الرفع على النحو التالي:

$$D = \frac{1}{2\rho V^2 S C_D} \qquad and \qquad L = \frac{1}{2\rho V^2 S C_L} \qquad (12)$$

4.4: حكم التغيير العرضي والتجديد العشوائي:

في بداية كل جيل، تتطلب خوارزمية بحث السنجاب الى أن يكون المجتمع بالكامل في فصل الشتاء، مما يعني أن جميع السناجب يتم تجديدها وفقًا للمعادلتين (9) و(10). عند هذه النقطة يتم تجديد السناجب، بغض النظر عما إذا كان التغيير الموسمي او لا ويتم تحديده بواسطة الصيغ التالية:

$$S_C^t = \sqrt{\sum_{k=1}^{d} \left(X_{ai,k}^t - X_{h,k}^t\right)^2} \qquad ; \qquad i = 1,2,\dots,N_a \qquad (13)$$

$$S_{min} = \frac{10e^{-6}}{(365)^{t/(t_{max}/2.5)}} \qquad (14)$$

فإذا كان$(\ ^t_C < S_{min})$ ، فإن الشتاء قد انتهى ويتحول الموسم إلى الصيف، ما عدا ذلك يبقى موسم الشتاء وتبقى السناجب في حالة سكون.

عندما يتحول الموسم إلى موسم الصيف، عندها تتحرك السناجب على وفق المعادلة التالية:

$$X_{inew}^{t+1} = X_L + Le'vy(x) * (X_U - X_L) \qquad (15)$$

حيث أن:

$$Le'vy(x) = 0.01 * \frac{\alpha * r_a}{|r_b|^{\frac{1}{\beta}}} \qquad (16)$$

أذ تُسمى $(Le'vy(x))$ بنموذج المشي التعسفي (arbitrary walk model). وأن قيمة (α) يتم حسابها من خلال المعادلة التالية:[13]

$$\alpha = \left[\frac{\Gamma(1+\beta) * \sin(\frac{\pi\beta}{2})}{\Gamma\left(\frac{1+\beta}{2}\right) * \beta * 2^{\left(\frac{\beta-1}{2}\right)}}\right]^{\frac{1}{\beta}} \qquad (17)$$

4. تطبيق الخوارزمية:

في هذه البحث، تم استخدام وتطوير خوارزمية بحث السنجاب لاختيار افضل المتغيرات لبيانات البقاء. وعلى النحو الاتي:

الخطوة 1: توليد قيم بشكل عشوائي ضمن الحدود المحددة لكل وحدة.

الخطوة 2: تطبيق استراتيجية التجديد لتلبية قيد توازن الطاقة.

الخطوة 3: تقييم قيمة دالة الهدف لجميع السناجب باستخدام معيار مجموع مربعات الخطأ.

الخطوة 4: ترتيب قيم دالة اللياقة لكل سنجاب بشكل تصاعدي.

الخطوة 5: تحديد قيمة السنجاب ذو قيمة اللياقة الدنيا موجود على شجرة جوز(مصدر الغذاء الأمثل)، والسناجب الثلاثة الأفضل الموجودة على شجرة البلوط (مصدر الغذاء الطبيعي)، والسناجب المتبقية الموجودة على الأشجار الطبيعية (لا يوجد مصدر غذاء).

الخطوة 6: تحديث موقع السناجب الموجودة على شجرة البلوط والأشجار الطبيعية كما هو موضح في الفقرة (3.4).

الخطوة 7: نقل مواقع بعض السناجب عشوائياً عند استيفاء شرط المراقبة الموسمية.

الخطوة 8: في حالة الوصول إلى الحد الأقصى لعدد التكرارات، فقم بإخراج موقع السنجاب على شجرة جوز باعتباره الحل الأمثل. وإلا ارجع إلى الخطوة رقم (2).

6. الجانب التطبيقي:

وكجانب عملي للبحث، تم التطبيق على بيانات تتبع توزيع وايبل والتي أخذت من بيانات استخدمت من قبل (لقاء سعيد واخرون،2011) [1] حول مرض الفشل الكلوي المزمن حيث تم جمع عينة مقدارها (73) نموذج دم لأشخاص مصابين بمرض العجز الكلوي المزمن والذين يتعالجون بالغسيل الكلوي المستمر، وتم سحب نماذج الدم لمجموعة المرضى قبل اجراء عملية الغسيل الكلوي التي تستغرق (4-3) ساعات، وقد شخصت حالة المرضى من قبل اطباء مختصين بالتعاون مع مستشفى ابن سينا التعليمي– وحدة الكلية الاصطناعية، تراوحت اعمارهم بين(80-20) سنة، وتتضمن العينة ما مقداره (38) نموذجاً للذكور و (35) نموذجاً للإناث، ودونت المعلومات الخاصة بالمرضى على وفق استمارة استبيان خاصة لكل مريض اعدت لهذا الغرض لسنة 2013، وقد تم تسجيل ودراسة ثمانية متغيرات توضيحية والتي يعتقد بان لها تأثير في متغير الاستجابة الذي يمثل تكرار الغسيل الكلوي بالشهر. ويبين الجدول رقم (1) وصف للمتغيرات التوضيحية المستخدمة في الدراسة.[1]

جدول (1) : وصف المتغيرات التوضيحية المستخدمة في الدراسة.

رمز المتغير التوضيحي	وصف المتغير التوضيحي	وحدة القياس
X_1	الجنس	(ذكر=1، انثى =2)
X_2	العمر	سنوات
X_3	مدة المرض	الايام
X_4	الوراثة	(نعم=1، كلا=2)
X_5	نسبة اليوريا	(ملي مول/لتر)
X_6	نسبة البروتين الكلي	غرام/100 ميليلتر
X_7	نسبة الالبومين	غرام/100 ميليلتر
X_8	نسبة الكلوبيولين	غرام/100 ميليلتر

يتم تقدير معلمات نموذج انحدار وايبل بواسطة مقدر الامكان الاعظم (MLE) بغض النظر عن تقدير(B_0)، ثم يتم ايجاد قيم$(\hat{Y})$ لحساب متوسط مربعات الخطأ (MSE) للنموذج، ومن خلال ملاحظة الجدول (2) الذي يوضح

نتائج متوسط مربعات الخطأ للنموذج المقدر نلاحظ تفوق خوارزمية بحث السنجاب (SSA) على باقي طرائق التقدير المستخدمة الاخرى كما في المصدر[4]. حيث انها اعطت اقل قيمة لمتوسط مربعات الخطأ مما يجعلها افضل طريقة للتقدير، بالمقارنة مع كل من طريقة (LASSO) والتي تمثل مختصر (Least Absolute Shrinkage and Selection Operator) وكذلك طريقة (SCAD) والتي تمثل مختصر (Smoothly Clipped Absolute Deviation) . أذ تأتي بالمرتبة الثانية طريقة (SCAD) من حيث قيمة متوسط مربعات الخطأ, وكذلك كانت طريقتي (MLE , LASSO) أسوأ طريقتين كونها اعطت اعلى قيم لمتوسط مربعات الخطأ.

جدول (2) : نتائج الطرق المستخدمة بالاعتماد على معيار(MSE)

لبيانات مرضى العجز الكلوي.

Methods	MSE
MLE	**9.358487**
LASSO	**7.877187**
SSA	5.9634
SCAD	**6.9741**

7. الاستنتاجات:

من خلال النتائج التي تم التوصل اليها تبين بان خوارزمية بحث السنجاب من الخوارزميات التطويرية المهمة والتي تستخدم في حل مشاكل الامثلية وكذلك فأن نموذج انحدار وايبل من النماذج اللوغارتمية الخطية المهمة في التنبؤ وتحديد المتغيرات ذات التأثير العالي. بالاضافة الى أن البحث يهدف إلى تقديم مساهمة جديدة في مجال تحليل بيانات البقاء باستخدام خوارزمية بحث السنجاب، وإلى تعزيز فهمنا لكيفية تحديد المتغيرات المؤثرة.

المصادر:

[1] عبدالله، لقاء سعيد وعلوش، ذكرى علي و الجراح، إسراء عبد الحق ،(2011)، "**دراسة إنزيم ميتالو إندوبيتايديز وعلاقته بمرض العجز الكلوي المزمن**"، مجلة علوم الرافدين، المجلد (22)، العدد (4)، ص (71-87).

[2] Ahmed, S. E., Hossain, S., & Doksum, K. A. (2012). LASSO and shrinkage estimation in Weibull censored regression models. *Journal of Statistical Planning and Inference*, *142*(6), 1273-1284.

[3] Alangood, H. N. A., Algamal, Z. Y., & Khaleel, M. A. (2024). Variable selection in Poisson regression model based on chaotic meta-heuristic search algorithm. In *BIO Web of Conferences* (Vol. 97, p. 00161). EDP Sciences.

[4] Algamal, Z. Y. (2019). Variable selection in Poisson regression model using invasive weed optimization algorithm. *Iraqi Journal of Statistical Sciences*, *16*(3), 39-54.

[5] Carroll, K. J. (2003). On the use and utility of the Weibull model in the analysis of survival data. *Controlled clinical trials*, *24*(6), 682-701.

[6] Collett, D. (2023). *Modelling survival data in medical research*. Chapman and Hall/CRC.

[7] Famoye, F., & Rothe, D. E. (2003). Variable selection for Poisson regression model. *Journal of Modern Applied Statistical Methods*, *2*, 380-388.

[8] Hashimoto, E. M., Ortega, E. M., Cancho, V. G., & Cordeiro, G. M. (2010). The log-exponentiated Weibull regression model for interval-censored data. *Computational statistics & data analysis*, *54*(4), 1017-1035.

[9] Kalktawi, H. S. (2017). *Discrete Weibull regression model for count data* (Doctoral dissertation, Brunel University London).

[10] Klein, J. P., & Moeschberger, M. L. (2003). Survival analysis: techniques for censored and truncated data (Vol. 1230). New York: Springer.

[11] Muche, R. (2001). Applied Survival Analysis: Regression Modeling of Time to Event Data. DW Hosmer, Jr., S Lemeshow. New York: John Wiley, 1999, pp. 386, US $89.95. ISBN: 0-471-15410-5.

[12] Smith, R. L. (1991). Weibull regression models for reliability data. *Reliability Engineering & System Safety*, *34*(1), 55-76.

[13] Suman, M., Sakthivel, V. P., & Sathya, P. D. (2020). Squirrel Search Optimizer: Nature Inspired Metaheuristic Strategy for Solving Disparate Economic Dispatch Problems. *International Journal of Intelligent Engineering & Systems*, *13*(5).

[14] Valenzuela, T. D., Roe, D. J., Cretin, S., Spaite, D. W., & Larsen, M. P. (1997). Estimating effectiveness of cardiac arrest interventions: a logistic regression survival model. *Circulation*, *96*(10), 3308-3313.

[15] Vigas, V. P., Silva, G. O., & Louzada, F. (2017). The Poisson-Weibull Regression Model. *Journal of Statistics*, *8*(1), 25-51.

[16] Zhang, Z. (2016). Parametric regression model for survival data: Weibull regression model as an example. *Annals of translational medicine*, *4*(24).

[17] Zhang, Z. (2016). Model building strategy for logistic regression: purposeful selection. *Annals of translational medicine*, *4*(6).