AL KUT JOURNAL OF ECONOMIC AND ADMINISTRATIVE SCIENCES
Publisher: College of Economics and Management - Wasit University

مقارنة بين طريقتي الانحدار الخطّي الموضعي والجوار الأقرب في تقدير أنْموذج الانحدار اللامعلمي دراسة محاكاة مع التطبيق

A Comparison Between Local Linear Regression and K- Nearest Neighbor Methods in Estimating the Nonparametric Regression Model with Simulation and Application

https://doi.org/10.29124/kjeas.1650.07

حوراء ضياء جبوري الخطّيب[1] أ . م . د سعد صبر محمد[2]

hawraajabbori523@uowasit.edu.iq ssabr@uowasit.edu.iq

جامعة واسط ـ كُلّية الإدارة والاقتصاد / قسم الإحصاء

المستخلص:

تَمّ في هذا البحث تقدير أنْموذج الانحدار اللامعلمي بطريقتين من طرائق التقدير الشائعة المتَمثّلة بكُلّ من (مقدّر الانحدار الخطّي الموضعي Local Linear Regression Estimator (LLRE)، ومقدّر الجوار الأقرب K- Nearest Neighbor Estimator (KNNE)). وقد أجريت عملية المحاكاة باستعمال لغة البرمجة الإحصائية R لهاتين الطريقتين، وتَمّتْ المقارنة بينهما باستعمال معيار معدّل جذر متوسّط مربعات الخطأ Average Root Mean Square Error (ARMSE)، وبأحجام ثلاث عينات (50,100,150) وثلاثة مستويات لتباين الخطأ (0.3, 0.7,1) ولأنْموذجين غير خطيين، إذْ أظهرت النتائج بالنسبة للأنْموذج الأول الأفضلية لطريقة مقدّر الجوار الأقرب (KNNE) في معظم الحالات. وللأنْموذج الثاني تفوق مقدّر الانحدار الخطّي الموضعي (LLRE)عند جميع حجوم العينات جميعها ولمستويات التباين جميعها. أمّا في الجانْب التطبيقي فقد تَمّتْ دراسة اثر الهيموغلوبين Hemoglobin (Hb) على حجم خلايا الحمر المكدسة Packed Cell Volume (PCV) للمرضى المصابين بأمراض الكُلّى المزمنة والبالغ عددهم (150)، إذْ تَمّتْ عملية التقدير باستعمال الطريقتين، وأجريت المقارنة بينها باستعمال معيار جذر متوسّط مربعات الخطأ Root Mean Square Error (RMSE) وقد تبين من خلال نتائج الجانب التطبيقي أنْ الأفضلية كانت لطريقة مقدّر الجوار الأقرب (KNNE).

الكُلّمات المفتاحية: الانحدار اللامعلمي، مقدّر الانحدار الخطّي الموضعي، مقدّر الجوار الأقرب، طريقة التقاطع الشرعي.

Abstract:

In this paper, the nonparametric regression model was estimated using two common estimation methods (Local Linear Regression Estimator (LLRE) and K-Nearest Neighbor Estimator (KNNE). The simulation process was conducted using the statistical programming language, R, for these methods, and they were compared using the Average Root Mean Squares Error (ARMSE) criterion, with three sample sizes (50, 100, 150), three levels of error variance (0.3, 0.7, 1) and two non-linear models. The results for the first model showed the superiority of the nearest neighbor estimator (KNNE) method in most cases, and for the second model the local linear regression estimator (LLRE) was superior at all sample sizes and for all levels of variance. On the applied side, the effect of hemoglobin (Hb) on the packed cell volume (PCV) of 150 patients with chronic kidney disease was studied. The estimation process was made using the two methods, and the comparison between them was made using the root mean square error (RMSE) criterion. It was shown through the results of the applied side that the preference was for the nearest neighbor estimator (KNNE) method.

Key words: Non parametric Regression, Local Linear Regression Estimator, K- Nearest Neighbor Estimator, Cross Validation.

1- المقدّمة Introduction

من المعلوم لدراسة أيّة ظاهرة (اقتصادية، اجتَمّاعية ، علمية، وغيرها) يتطلّب تحليل الظاهرة وايجاد العوامل التي ترتبط بها وتؤثّر فيها، وذلك من خلال ايجاد أنْموذج يعبر عنها ويكون مدخلا رئيسيا لفهمها ، اذ يُعدّ هذا من اهم اهداف نماذج الانحدار إذْ يعرف تحليل الانحدار بأنّه اسلوب إحصائي يستعمل لمعرفة العلاقة بين متغيّر يعرف بالمتغيّر المعتَمّد (متغيّر الاستجابة) ومتغيّر واحد أو متغيّرات عدّة تعرف بالمتغيّرات التفسيرية أو (التوضيحية)، فطرائق الانحدار المعلمية Parametric Regression Methods كأنْت في السنوات الماضية موضوع نقاش الكثير من الباحثين والتي تفترض أنْ العيّنة تأتي من مجتَمّع معروف التوزيع مثل عائلة كامّا أو العائلة الطبيعية ومن ثُمّ العمل على تقدير المعالم المجهولة من خلال طرائق التقدير المتَمّثّلة بـ (المربعات الصغرى الاعتيادية، العزوم، الإمكانْ الأعظم، بيز) وغيرها من طرائق التقدير أو عن طريق الاختبارات أو الاشتقاقات لحدود الثقة للمعالم غير المعلومة بالاعتَمّاد على العيّنة، وفي معظم الأحيانْ تكون تلك الافتراضات صارمة ومن ثُمّ فأنْ التوزيع المعلمي ليس بالضرورة أنْ يكون التوزيع المعني للمسألة المراد حلها لكون الافتراض الخاطئ للتوزيع المعلمي وللمسألة المراد حلها قد تؤدي بالطرائق الإحصائية المستعملة إلى مقدّرات غير متسقة و نتائج غير صحيحة، ما يعني عدم امتلاك البيانات عائلة معلمية، وكون الطرائق المعلمية لا تتلاءم مع احجام العينات الصغيرة و كذلك العينات غير معلومة التوزيع، وعلى الرغم أنْ الأنْماذج المعلمية تُعدّ أحد أهم أدوات تحليل البيانات، لأنّها فعالة و يمكن تفسيرها بسهوله، إلا أنّها لم تستطع أنْ

تكون كافية في بعض الحالات ولا بُدّ من التطرق إلى الطرائق اللامعلمية ، ونظراً للأسباب المذكورة آنفاً أجبرت الباحثين اللجوء إلى الطرائق الإحصائية الحديثة والمتَمثّلة بالطرائق اللامعلمية والشبه المعلمية في تحليل وتفسير البيانات.

تَمّتْ دراسة طرائق التقدير اللامعلمي في العديد من البحوث والدراسات من ابرزها في عام (1990) قَدّمَ الباحث (Hardle)[17] كتابا قَدّمَ فيه طرائق لتقدير دالّة الانحدار اللامعلمي وايضا ناقش فيه طرائق لتقدير المعلمة التَمّهيدية واسلوب اختيارها ومدى تأثيرها على اداء دالّة الانحدار اللامعلمي، وفي عام (2007) قام الباحث [5] (Aydin)بتقديم بحثاً تضمن استعمال الأنْماذج اللامعلمية لبيانات الناتج القومي الإجمالي في تركيا وبيانات زراعة القلب في ستانفورد، إذْ ناقش طريقتين لامعلمية الأولى تَمّهيد الشريحة والثانْية انحدار kernel، في عام (2011) قَدّمَ الباحثان (الصفاوي ومتي) [4] بحثا تضمن مقارنة لثلاث طرائق تَمّهيدية (طريقة الجوار الأقرب (KNN) K- Nearest Neighbor Method وطريقة الدالّة اللبية(K) Kernel Function Method وطريقة الشرائح التَمّهيدية (Smoothing Splines Method (SS) باستعمال اسلوب المحاكاة بنموذجين مختلفين وثلاث قيم للتباين. وفي عام (2016) قَدّمَ الباحثانْ (حسن ومجيد)[3] بحثا تضّمن استعمال بعض طرائق التَمّهيد منها انحدار الشريحة Spline Regression (SR)، والشريحة الجزائية Penalized Spline (PS)، والتَمّهيد اللبي مثُعدّد الحدود الموضعي Local Polynomial Kernel (LPK)، وفي عام (2019) تطرق الباحثان (بدر وعبد)[2] إلى مقارنة مقدّرات أنْموذج الانحدار اللامعلمي من خلال المقدّر الانحدار الخطّي الموضعي (LLRE) Local Linear Regression Estimator ومقدّر الانحدار التربيعي الموضعي (LQRE) Local Quadratic Regression Estimator باستعمال معيار التقاطع الشرعي Cross Validation (CV)، وايضا استعمال مقدّر الشرائح التَمّهيدية الحصينة Robust Smoothing Splines (RSS) وباستعمال المعيار الحصين (RCV) Robust Cross Validation وايضا تَمّ استعمال اسلوب المحاكاة كأداة مثلى لدراسة احجام العينات والتوزيعات المختلفة للخطأ العشوائي منها وجود بيانات شاذة ومنها حالتها الطبيعية وبمستويات لتباين ونماذج مختلفة، وفي عام (2023) قَدّمَ الباحث (الاشقر) [1] بحثا تضمن فيه تقديم بعض من طرائق شرائح التَمّهيد ودوال الاساس والمتَمّثّلة بالشرائح الجزائية (PS) والشرائح التكعيبية الطبيعية (NCS) والشرائح التكعيبية (CS) وشرائح الأساس (BS) ثُمّ مقارنة هذه الطرائق فيما بينها للوصول إلى معدّلات الوفاة بشركات التامين المصرية خلال المدّة (2007-2010).

وسيَتَمّ تقسيم ما تبقّى من هذا البحث إلى فقرات عدّة والمتَمّثّلة بأنْموذج الانحدار اللامعلمي وطرائق التقدير، يليه الجانب التجريبي (المحاكاة)، ثُمّ الجانْب التطبيقي (البيانات الحقيقة للمرضى المصابين بأمراض الكُلّى المزمنة)، والاستنتاجات فالتوصيات، وأخيرا المصادر.

2- أنْموذج الانحدار اللامعلمي[18]، [19] Nonparametric Regression Model

يَتَمّ تعريف أنْموذج الانحدار اللامعلمي بما في ذلك متغيّر التنبؤ ومتغيّر الاستجابة على أنّه :

$$y_i = m(x_i) + \epsilon_i \qquad i = 1, 2, \dots, n \qquad \dots (1)$$

y_i **:** يمثّل قيمة المتغيّر المعتَمّد (الاستجابة) للمشاهدة i.

$m(x_i)$: تَمّثل دالّة التَمّهيد غير المعروفة.

x_i **:** تَمّثل قيمة المتغيّر المستقل للمشاهدة i.

ϵ_i**:** قيمة الخطأ العشوائي ذات التوزيع الطبيعي المستقل والمتَمّاثل $\epsilon_i \sim N(0, \sigma^2)$.

ومن خلال الأنْموذج المذكور انفاً بامكأنْنا استعماله كأداة لوصف اتجاه و تأثير المتغيّر التفسيري في متغيّر الاستجابة وذلك من خلال استعمال اسلوب التَمّهيد (Smoothing) والذي يمكن تَمّثيله كدالّة بدلالة متغيّر واحد، أو أكثر من المتغيّرات التفسيرية لمنحنى الأنْموذج $m(x_i)$ إذْ يُعدّ اسلوباً بيانْياً لتقدير اي دالّة من الدوال الإحصائية، أنْ الهدف من هذا الأسلوب تقريب دالّة الانحدار اللامعلمي التقريبية لدالّة الانحدار اللامعلمية الحقيقة ثُمّ تُأخذ البيانات بعد ذلك في تحديد ما يُكون دالّة الانحدار الحقيقية، وفي حالة غياب اي من المعلومات عن الدالّة الحقيقية يعتبر تحليل الانحدار اللامعلمي هو الأفضل دائما، إذْ تَمّثل آلية الانحدار اللامعلمي بأنّها تعتَمّد كثيرا على البيانات في تحديد تلك الدالّة أكثر من اعتَمّادها على الافتراضات لغرض الحصول على معلومات عن الـ m بدلا من الصيغة الدالية، واحياناً تحتاج إلى افتراضات أقل والتي تجعل من أنْموذج الانحدار خطّي، غير خطي، لوجستي أو احتَمّالي ... الخ ولا سيّما عندما تكون بيانات العيّنة نسبوية أو رتيبة أو بيانات وصفية، ومن المعروف أنْ الانحدار اللامعلمي يجعل البيانات أكثر تأقلماً وأقل تعقيدا من اسلوب الانحدار المعلمي لذلك تَمّ اللجوء لهذا الأسلوب كبديل ملائم لمعالجة البيانات، ويُعدّ أداة مرغوبة لدى معظم الباحثين كونه له القابلية في كشف تركيبة البيانات ولا سيّما عندما يكون حجم العيّنة صغيرا من النماذج المعلمية للاستدلال عن البيانات كون البيانات الفعلية ليست دوما تَمّلك وصفاً ملائماً، كما يُعدّ أنْموذج الانحدار اللامعلمي اداة لإعطاء وصف عام للظاهرة وليس لدراسة التفاصيل الدقيقة لها.

3- طرائق تقدير أنْموذج الانحدار اللامعلمي

Methods of Estimating the Nonparametric Regression Model

سيَتَمّ في هذا البحث استعمال طريقتين لتقدير أنْموذج الانحدار اللامعلمي وكالاتي:

3-1 مقدّر الانحدار الخطّي الموضعي[8]،[9]،[11]،[12]،[16] Local Linear Regression Estimator

اقترحت فكرة الانحدار الخطّي الموضعي Local Linear Regression (LLR) في عام 1979من قبل Cleveland، كطريقة لامعلمية تستعمل لنمذجة العلاقة بين متغيّر مستقل ومتغيّر معتَمّد، اذ تُعدّ تقنّية مرنة تسمح بتقدير العلاقة الموضعية عند كُلّ نقطة في البيانات، بدلاً من افتراض وجود علاقة عامة بين المتغيّرات، يقوم المقدّر الخطّي الموضعي بتقدير العلاقة عند كُلّ نقطة عن طريق ملاءمة أنْموذج انحدار خطي مع مجموعة موضعية بجوار النقاط. يُعّرفَ هذا الجوار بأنّه معلمة التَمّهيد (عرض الحزمة) والتي تحدد عدد النقاط القريبة المستعملة في هذا التقدير. يستعمل هذا المقدّر أنْموذج انحدار خطّي داخل الجوار الموضعي، وذلك من خلال تصغير مجموع مربعات البواقي الموزونة بدالّة kernel، ومن أهم مزايا هذا الأسلوب اكتشاف العلاقات غير الخطّية بين المتغيّرات ولا سيّما عندما تكون هنالك تباينات كبيرة.

يُعدّ مقدّر الانحدار الخطّي الموضعي نسخة لامعلمية من طريقة المربعات الصغرى الموزونة، إذْ يمكن التعبير عنها على شكْلّ تصغير للصيغة الآتية:

$$\sum_{i=1}^{n}\{y_i-\beta_0-\beta_1(x_i-x)\}^2K_\lambda(x_i-x) \qquad \dots(2)$$

ويمكن كتابتها على شكْلّ مصفوفات كما يأتي:

$$(\mathbf{y}-\mathrm{X}\boldsymbol{\beta})'\mathrm{W}(\mathbf{y}-\mathrm{X}\boldsymbol{\beta}) \qquad \dots(3)$$

إذْ أنْ $\boldsymbol{y}$ هو متجه قيم المتغيّر المعتَمّد، وإنْ:

$$\boldsymbol{X}=\begin{bmatrix}\mathbf{1} & (x_1-x)\\ \mathbf{1} & (x_2-x)\\ \vdots & \vdots\\ \mathbf{1} & (x_n-x)\end{bmatrix}$$

$$\boldsymbol{W}=\begin{bmatrix}K_\lambda(x_1-x) & 0 & \cdots & 0\\ 0 & K_\lambda(x_2-x) & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & K_\lambda(x_n-x)\end{bmatrix}$$

وباستعمال طريقة المربعات الصغرى الموزونة فأنْ مقدّر المَعلَمات التي يقوم بجعل المعادلة (3) أصغر ما يمكن هو ما يأتي:

$$\widehat{\boldsymbol{\beta}}=(\boldsymbol{X'WX})^{-1}\boldsymbol{X'Wy}=(\boldsymbol{QX})^{-1}\boldsymbol{Qy} \qquad \dots(4)$$

إذْ أنْ $Q=X'W$، وتعرف كما يأتي:

$$\mathbf{Q}=\begin{bmatrix}\mathbf{1} & \mathbf{1} & \cdots & \mathbf{1}\\ (\mathbf{x_1}-\mathbf{x}) & (\mathbf{x_2}-\mathbf{x}) & \cdots & (\mathbf{x_n}-\mathbf{x})\end{bmatrix}\begin{bmatrix}\mathbf{K_\lambda(x_1-x)} & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{0} & \mathbf{K_\lambda(x_2-x)} & \cdots & \mathbf{0}\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{K_\lambda(x_n-x)}\end{bmatrix}$$

$$=\begin{bmatrix}\mathbf{K_\lambda}(x_1-x) & \mathbf{K_\lambda}(x_2-x) & \cdots & \mathbf{K_\lambda}(x_n-x)\\ (x_1-x)\mathbf{K_\lambda}(x_1-x) & (x_2-x)\mathbf{K_\lambda}(x_2-x) & \cdots & (x_n-x_0)\mathbf{K_\lambda}(x_n-x)\end{bmatrix}$$

ويمكن إيجاد QX كما يأتي:

$$\boldsymbol{QX}$$

$$= \begin{bmatrix} \mathbf{K_\lambda}(x_1 - x) & \mathbf{K_\lambda}(x_2 - x) & \dots & \mathbf{K_\lambda}(x_n - x) \\ (x_1 - x)\mathbf{K_\lambda}(x_1 - x) & (x_2 - x)\mathbf{K_\lambda}(x_2 - x) & \dots & (x_n - x_0)\mathbf{K_\lambda}(x_n - x) \end{bmatrix} \begin{bmatrix} 1 & (x_1 - x) \\ 1 & (x_2 - x) \\ \vdots & \vdots \\ 1 & (x_n - x) \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{n} \mathbf{K_\lambda}(x_i - x) & \sum_{i=1}^{n} (x_i - x)\mathbf{K_\lambda}(x_i - x) \\ \sum_{i=1}^{n} (x_i - x)\mathbf{K_\lambda}(x_i - x) & \sum_{i=1}^{n} (x_i - x)^2\mathbf{K_\lambda}(x_i - x) \end{bmatrix}$$

كما يمكن إيجاد $\boldsymbol{Qy}$ كما يأتي:

$$\boldsymbol{Qy}$$

$$= \begin{bmatrix} \mathbf{K_\lambda}(x_1 - x) & \mathbf{K_\lambda}(x_2 - x) & \dots & \mathbf{K_\lambda}(x_n - x) \\ (x_1 - x)\mathbf{K_\lambda}(x_1 - x) & (x_2 - x)\mathbf{K_\lambda}(x_2 - x) & \dots & (x_n - x_0)\mathbf{K_\lambda}(x_n - x) \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{n} y_i\mathbf{K_\lambda}(x_i - x) \\ \sum_{i=1}^{n} y_i(x_i - x)\mathbf{K_\lambda}(x_i - x) \end{bmatrix}$$

ومعكوس المصفوفة QX يمكن كتابتها كما يأتي:

$$(\boldsymbol{QX})^{-1} = \frac{1}{\boldsymbol{det(QX)}} \begin{bmatrix} \sum_{i=1}^{n} (x_i - x)^2\mathbf{K_\lambda}(x_i - x) & -\sum_{i=1}^{n} (x_i - x)\mathbf{K_\lambda}(x_i - x) \\ -\sum_{i=1}^{n} (x_i - x)\mathbf{K_\lambda}(x_i - x) & \sum_{i=1}^{n} \mathbf{K_\lambda}(x_i - x) \end{bmatrix}$$

إذْ $det(QX)$ تَمثل محددة QX وتعطى كما يأتي:

$$\boldsymbol{det(QX)} = \sum_{i=1}^{n} \boldsymbol{K_\lambda}(x_i - x) \sum_{i=1}^{n} (x_i - x)^2\boldsymbol{K_\lambda}(x_i - x) - \left\{\sum_{i=1}^{n} (x_i - x)\boldsymbol{K_\lambda}(x_i - x)\right\}^2$$

$$= \boldsymbol{S_0S_2 - S_1^2}$$

إذْ أنْ:

$$S_r = \sum_{i=1}^{n} (x_i - x)^r \mathbf{K}_\lambda (x_i - x)\,; \qquad r = 0, 1, 2$$

وبناءً على ما سبق يمكن إعادة كتابة المعادلة (4) كما يأتي:

$$\hat{\beta} = \frac{1}{S_0 S_2 - S_1^2} \begin{bmatrix} S_2 & -S_1 \\ -S_1 & S_0 \end{bmatrix} \begin{bmatrix} \sum_{i=1}^{n} y_i \mathbf{K}_\lambda (x_i - x) \\ \sum_{i=1}^{n} y_i (x_i - x) \mathbf{K}_\lambda (x_i - x) \end{bmatrix} \qquad \dots (5)$$

$$= \frac{1}{S_0 S_2 - S_1^2} \begin{bmatrix} S_2 \sum_{i=1}^{n} y_i \mathbf{K}_\lambda (x_i - x) - S_1 \sum_{i=1}^{n} y_i (x_i - x) \mathbf{K}_\lambda (x_i - x) \\ S_0 \sum_{i=1}^{n} y_i (x_i - x) \mathbf{K}_\lambda (x_i - x) - S_1 \sum_{i=1}^{n} y_i \mathbf{K}_\lambda (x_i - x) \end{bmatrix}$$

ويكون مقدّر β_0 معرفاً وفقاً للصيغة الآتية:

$$\hat{\beta}_0 = \frac{S_2 \sum_{i=1}^{n} y_i \mathbf{K}_\lambda (x_i - x) - S_1 \sum_{i=1}^{n} y_i (x_i - x) \mathbf{K}_\lambda (x_i - x)}{S_0 S_2 - S_1^2}$$

$$= \frac{\sum_{i=1}^{n} [S_2 - (x_i - x) S_1] \mathbf{K}_\lambda (x_i - x)\, y_i}{S_0 S_2 - S_1^2}$$

والمعادلة (6) تَمّثل مقدّر الانحدار الخطّي الموضعي للدالّة m(x)، بمعنى أنْ:

$$\hat{m}(x) = \frac{\sum_{i=1}^{n} [S_2 - (x_i - x) S_1] \mathbf{K}_\lambda (x_i - x)\, y_i}{S_0 S_2 - S_1^2} \qquad \dots (6)$$

وتقدر معلمة التَمّهيد (λ) باستعمال معيار التقاطع الشرعي (CV) إذْ يتلخص عمل هذه الطريقة أنْ في كُلّ مرة تستبعد قيمة واحدة من قيم المتغيّرات، وأنْ عملية تقدير هذه المعلمة تكون كالآتي:

1. يَتَمّ حساب تقدير الدالّة $\hat{m}_i(x_i)$ لكُلّ مشاهدة (i)، إذ أنْ $\hat{m}_i(x_i)$ هي استبعاد مشاهدة واحدة في كُلّ مرة، وأنْ المقدّر $\hat{m}_i(x_i)$ يحسب وفق المعادلة (6)، أي أنْ:

$$\hat{m}_i(x_i) = \frac{\sum_{j \neq i}^{n} [S_{2j} - (x_i - x_j) S_{1j}] \mathbf{K}_\lambda (x_i - x_j)\, y_i}{S_{0j} S_{2j} - S_{1j}^2} \qquad \dots (7)$$

$$i = 1\,, 2, \dots, n-1 \quad , \quad j = 2,3, \dots, n$$

إذْ أنْ:

$$\boldsymbol{S_{rj} = \sum_{i=1}^{n} (x_i - x_j)^r K_\lambda (x_i - x_j)\,; \qquad r = 0, 1, 2}$$

2. يَتَمّ حساب دالّة العبور الشرعي (CV) باستبعاد مشاهدة واحدة في كُلّ مرة وبالشكْلّ الآتي:

$$\mathbf{CV_{(h)} = \frac{1}{n}\sum_{i=1}^{n} [y_i - \hat{m}_i(x_i)]^2} \qquad \dots (8)$$

3. يَتَمّ حساب قيمة (λ_{cv}) المثلى من خلال:

$$\boldsymbol{\lambda_{cv} = \arg\min CV_{(\lambda)}} \qquad \dots (9)$$

2-3 مقدّر الجوار الأقرب (K-NNE)[6] ،[13] ،[14] ،[15] ،[21] K- Nearest Neighbor Estimator

والتي تُعدّ واحدة من أشهر الطرائق أو التقنيات اللامعلمية التقليدية والمستعملة لحلّ مشاكُلّ التصنيف والانحدار، إذْ تطوّرتْ هذه الطريقة لأول مرة من قبل Joseph Hodges & Evelyn Fipbx فى عام (1951)، و التي لها القابلية على توليد التقديرات المحلية للنقطة ، ويُعدّ هذا الأسلوب غير معقد في معظم الأحيانْ في حالة وجود قاعدّة بيانات مثلى عند توفر قاعدّة بيانات تصنف الطريقة من ضمن الطرائق الدقيقة لمعرفة مقدّر الجوار الأقرب بسهولة، وعندما تكون العملية معقدة ينتج قاعدّة البيانات كبيرة جدا، يقيس الجوار الأقرب المسافة الأقليدية بين كُلّ نقطة والنقطة الأقرب اليها، اذ أنْ المسافة الأقليدية تكون قليلة جدا عندما يكون أسلوب البيانات بشكُلّ متجمع ، وكُلّما تبعثرت البيانات وتباعدت تصبح المسافة بين النقاط كبيرة، كما تَمّ توسيعها من قبل Thomas Cover ، وايضا استخدم هذا المقدّر من قبل Royal فى عام (1966) و Stone فى عام (1977) و Mach لعام (1981) بموجب الأنْموذج العشوائي بعد احتساب $W_{ni}(x)$.

$$\mathbf{W_{ni}(x) = \frac{k^{\frac{(xi-x)}{kn}}}{\sum_{i=1}^{n} k^{\frac{(xi-x)}{kn}}}}$$

K(.): تَمّثل دالّة لبية وغير سالبة وتحقّق $K(u) = 0$, for all $|u| > 1$

حيث يعرف مقدّر **(K-NNE)** بالصيغة الآتية:

$$\widehat{\boldsymbol{m}}_{\boldsymbol{KNN}}(\boldsymbol{X}) = \frac{\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{y_i}\, \boldsymbol{k}\left(\left(\boldsymbol{H_n}, \boldsymbol{k}(\boldsymbol{x})\right)^{-\mathbf{1}} \boldsymbol{d}(\boldsymbol{x_i}, \boldsymbol{x})\right)}{\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{k}\left(\left(\boldsymbol{H_n}, \boldsymbol{k}(\boldsymbol{x})\right)^{-\mathbf{1}} \boldsymbol{d}(\boldsymbol{x_i}, \boldsymbol{x})\right)} \quad \boldsymbol{x} \in \boldsymbol{F} \qquad \ldots(\mathbf{10})$$

K(.) **:** تَمّثل دالّة kernel بالفترة (∞**0,**] .

$H_n, k(x)$: تَمّثل متغيّراً عشوائياً موجب يعتَمّد على قيم ($x_1, x_2, \ldots, x_n$)

ويُعّرفَ كالآتي:

$$\mathbf{H_n, k(x) = min}\left\{\mathbf{h} \in \mathbf{R_+} : \sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{IB_{(x,h)}}\,(\mathbf{x_i}) = \mathbf{k}\right\}$$

$\boldsymbol{d}(\boldsymbol{x_i}, \boldsymbol{x_j})$: المسافة الأقليدية Euclidian Distance بين x و k من بين قيم الـ xi.

ويَتَمّ حساب المسافة الأقليدية وفق الصيغة الآتية:

$$\boldsymbol{d}(\boldsymbol{x_i}, \boldsymbol{x_j}) = \sqrt{\sum_{\boldsymbol{k=1}}^{\boldsymbol{n}} \left(\boldsymbol{x_{i\,k}} - \boldsymbol{x_{j\,k}}\right)^2} \qquad ; \boldsymbol{i, j} = \mathbf{1, 2, \ldots, n} \qquad \ldots(\mathbf{11})$$

ويكون المقدّر متسقاً بموجب الأنْموذج العشوائي بشرط (k_n= k) بحيث أنْ ($k_{n\to\infty}$) و ($\frac{\boldsymbol{k_n}}{\boldsymbol{n}} \to \mathbf{0}$) هذا يعني أنْ ($\boldsymbol{k} = \sqrt{\boldsymbol{n}}$).

4- الجانْب التجريبي The Experimental Side

4-1 مفهوم المحاكاة Concept of Simulation

يَتَمّ استعمال الجانْب التجريبي من خلال تطبيق اسلوب المحاكاة Simulation لغرض تطبيق طرائق التقدير اللامعلمية، والتي تَمّ التطرق اليها في الجانْب النظري، اذ يعّرف هذا الأسلوب على أنّه تقليد للواقع العملي إذْ يقوم بتوظيف نماذج تتضمّنَ العديد من الحالات الافتراضية لضمانْ الحصول على نتائج أكثر شمولية، كما يستعمل في حالة عجز البرهانْ الرياضي في بيانْ أفضلية طرائق تقدير معينة على حساب طرائق أخرى وأيضا في الحالات التي يصعب الحصول على بيانات دقيقة وواقعية لظاهرة ما.

4-2 مراحل تجربة المحاكاة Simulation Experiment Stages

تَمّ استعمال لغة البرمجة الإحصائية R بإصدارها (4.3.1) والتي تُعدّ واحدة من أكثر لغات البرمجة شعبية وأداة قوية كونها توفر وظائف وأدوات متُعدّدة لتحليل البيانات في برمجة طرائق التقدير.

المرحلة الأولى: مرحلة تحديد دوال الانحدار اللامعلمي

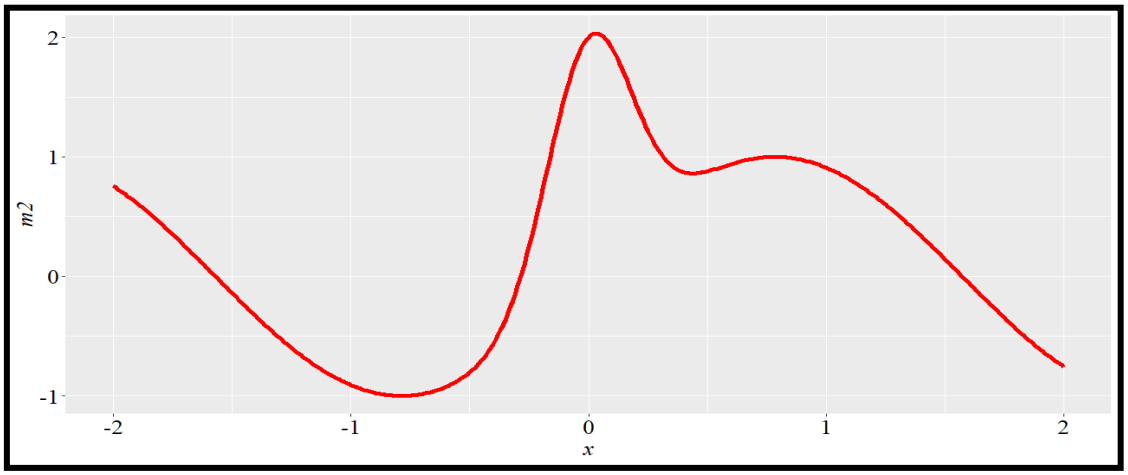
The Stage of Determining Nonparametric Regression Functions

إذْ تَمّ اختيار دالتين من دوال انحدار اللامعلمي (غير الخطّي) وهما كما يأتي:

الأنْموذج الأول [10]

$$m_1(x) = \sin(2x) + 2exp^{-16x^2} \quad ,x \in [-2, 2] \qquad \dots (12)$$

والشكْلّ الآتي يبيّن رسم هذه الدالّة:

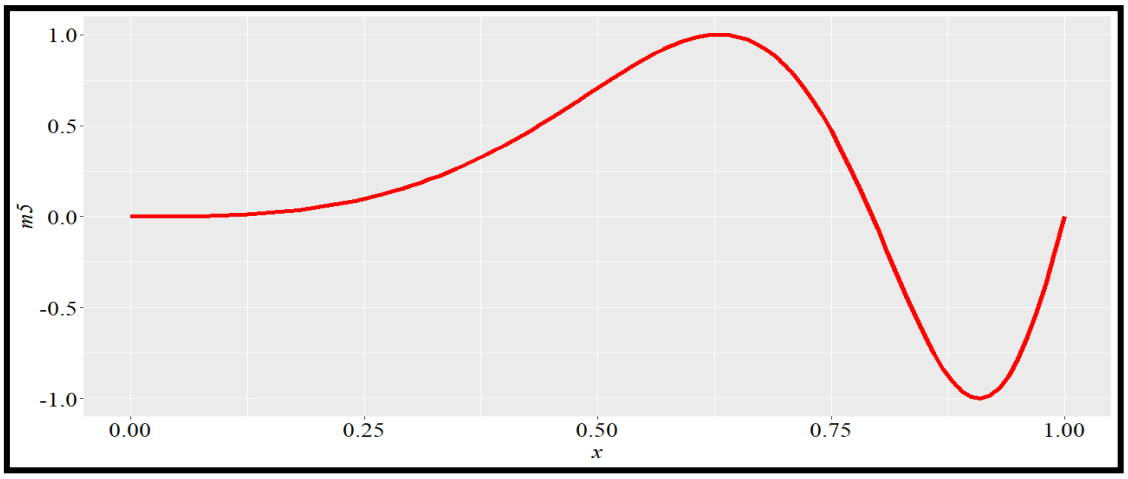


الشكْلّ (1): رسم دالّة الأنْموذج الأول

الأنْموذج الثاني[20]

$$m_2(x) = \sin(2\pi x^3) \ , \ x \in (0, 1) \qquad \dots (13)$$

والشكْلّ الآتي يعرض رسم هذه الدالّة:



الشكل (2): رسم دالّة الأنموذج الثاني

المرحلة الثانية: مرحلة تحديد القيم الافتراضية **Default Values Determination Phase**

وقد تَمّ اختيار قيم افتراضية لتباين الخطأ العشوائي ($\sigma^2 = 0.3, 0.7, 1$)، وأحجام عيّنات هي (50, 100, 150)، وقد تَمّ تكرار كُلّ تجربة 1000 مرة.

المرحلة الثالثة: توليد البيانات **Data Generation**

تَمّ توليد قيم المتغيّر المستقل من التوزيع المنتظم ضمن الفترة [a,b] بالاعتِمّاد على مدى كُلّ دالّة، كما تَمّ توليد الخطأ العشوائي (ε) من التوزيع الطبيعي بمتوسّط يساوي صفر، وتباين يساوي σ^2. وبناءً على المعادلة (1) تَمّ توليد المتغيّر المعتَمّد.

المرحلة الرابعة: التقدير **Estimation**

تَمّ تقدير أنموذج الانحدار اللامعلمي باستعمال طريقتين للتقدير وكما يأتي:

1. طريقة مقدّر الانحدار الخطّي الموضعي (LLRE).
2. طريقة مقدّر الجوار الأقرب (KNNE).

المرحلة الخامسة: مرحلة المقارنة بين الطرائق[7] **Comparison Phase Between Methods**

تَمّ استعمال معيار معدّل جذر متوسّط مربعات الخطأ Average Root Mean Squares Error (ARMSE) ، لغرض المقارنة بين الطريقتين إذْ أنّ الطريقة التي تَمتْلك أقل قيمة (ARMSE) تعدّ الأفضل.

$$ARMSE = \frac{1}{r}\sum_{j=1}^{r}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_{ij}(x_i) - \hat{m}_j(x_i)\right)^2} \quad \dots (14)$$

إذ إنْ :

$y_{ij}(x_i)$: تَمّثل المشاهدة i من التكرار j.

$\widehat{m}_j(x_i)$: دالّة الانحدار اللامعلمي المقدّرة.

n: تَمّثل عدد المشاهدات.

r: تَمّثل عدد التكرارات.

3-4 نتائج عمليات المحاكاة **Results of Simulations**

لغرض تطبيق الطريقتين لتقدير أنْموذج الانحدار وتحديد الطريقة الأفضل، إذْ سيَتَمّ عرض النتائج التي تَمّثل قيم (ARMSE) لكُلّ طريقة وفقاً لأحجام العينات والقيم المختلفة لمعلمة σ^2 وكما يأتي:

الجدول (1): معيار المقارنة (ARMSE) للطريقتين وباختلاف التباين مع توليد (50) مشاهدة بالنسبة للأنْموذج الأول ($m_1(x) = \sin(2x) + 2exp^{-16x^2}$)

Methods / σ^2	KNNE	LLRE
0.3	0.64871	0.76604
0.7	0.97057	0.97511
1	1.15521	1.10659

يتبين من الجدول آنفاً عند حجم عيّنة (50) مشاهدة وتباين الخطأ ($\sigma^2 = 0.3$) فأنْ أفضل طريقة كانْت هي طريقة مقدّر الجوار الأقرب (KNNE) لامتلاكها أقل قيمة لمعيار معدّل جذر متوسّط مربعات الخطأ (ARMSE) والمساوية إلى (0.64871).

أمّا عندما تكون قيمة تباين ($\sigma^2 = 0.7$) الأفضلية كانْت لطريقة مقدّر الجوار الأقرب (KNNE) لامتلاكها أقل قيمة لمعيار (ARMSE) والمساوية إلى (0.97057).

وفي حالة ($\sigma^2 = 1$) كانْت طريقة مقدّر الانحدار الخطّي الموضعي (LLRE)، هي الأفضل، وذلك لامتلاكها أقل قيمة لمعيار (ARMSE) والمساوية إلى (1.10659).

أظهرت النتائج المذكورة آنفاً أنْ أفضل طريقة كانْت بالنسبة للأنْموذج الأول عند حجم عينة (50) مشاهدة وقيم تباين خطأ ($\sigma^2 = 0.3,0.7$) هي مقدّر الجوار الأقرب (KNNE)، وذلك لامتلاكها أقل قيمة بالنسبة لمعيار

المقارنة (ARMSE)، ماعدا ($\sigma^2 = 1$) تكون الأفضلية لطريقة مقدّر الانحدار الخطّي الموضعي (LLRE) لكونها تَمتْلك أقل قيمة لمعيار (ARMSE).

الجدول (2): معيار المقارنة (ARMSE) للطريقتين وباختلاف التباين مع توليد (100) مشاهدة بالنسبة للأنْموذج الأول($m_1(x) = \sin(2x) + 2exp^{-16x^2}$)

Methods / σ^2	KNNE	LLRE
0.3	0.63548	0.73584
0.7	0.96408	0.97438
1	1.13622	1.10462

يتضح من الجدول (2) عند حجم عينة (100) مشاهدة وتباين الخطأ ($\sigma^2 = 0.3$) فإنْ أفضل طريقة كأنْت هي طريقة مقدّر الجوار الأقرب (KNNE) لامتلاكها أقل قيمة لمعيار معدّل جذر متوسّط مربعات الخطأ (ARMSE) والمساوية إلى (0.63548).

أمّا عندما تكون قيمة تباين ($\sigma^2 = 0.7$) الأفضلية كأنْت لطريقة مقدّر الجوار الأقرب (KNNE) لامتلاكها أقل قيمة لمعيار (ARMSE) والمساوية إلى (0.96408).

وفي حالة ($\sigma^2 = 1$) كأنْت طريقة مقدّر الانحدار الخطّي الموضعي (LLRE) هي الأفضل، وذلك لامتلاكها أقل قيمة لمعيار (ARMSE) والمساوية إلى (1.10462).

تَمّ التوصل إلى أنْ الأفضلية كأنْت بالنسبة للأنْموذج الأول عند حجم عينة (100) مشاهدة وقيم تباين خطأ ($\sigma^2 = 0.3,0.7$) هي مقدّر الجوار الأقرب (KNNE)، وذلك لامتلاكها أقل قيمة بالنسبة لمعيار المقارنة (ARMSE)، ماعدا ($\sigma^2 = 1$) تكون الأفضلية لطريقة مقدّر الانحدار الخطّي الموضعي (LLRE)، لكونها تَمتْلك أقل قيمة لمعيار (ARMSE).

الجدول (3): معيار المقارنة (ARMSE) للطريقتين وباختلاف التباين مع توليد (150) مشاهدة بالنسبة للأنْموذج الأول ($m_1(x) = \sin(2x) + 2exp^{-16x^2}$)

Methods / σ^2	KNNE	LLRE
0.3	0.62980	0.70836
0.7	0.95244	0.95438
1	1.12036	1.00974

يتبين من الجدول (3) عند زيادة حجم العيّنة إلى (150) مشاهدة وبتباين خطأ ($\sigma^2 = 0.3$) فإنْ أفضل طريقة كانْت هي طريقة مقدّر الجوار الأقرب (KNNE)، لامتلاكها أقل قيمة لمعيار معدّل جذر متوسّط مربعات الخطأ (ARMSE) والمساوية إلى (0.62980).

أمّا عندما تكون قيمة تباين الخطأ ($\sigma^2 = 0.7$) الأفضلية كانْت لطريقة مقدّر الجوار الأقرب (KNNE) لامتلاكها أقل قيمة لمعيار (ARMSE) والمساوية إلى (0.95244).

وفي حالة ($\sigma^2 = 1$) كانْت طريقة مقدّر الانحدار الخطّي الموضعي (LLRE)، هي الأفضل وذلك لامتلاكها أقل قيمة لمعيار (ARMSE) والمساوية إلى (1.00974).

أظهرت النتائج المذكورة انفاً أنْ أفضل طريقة كانْت بالنسبة للأنْموذج الأول عند حجم عينة (150) مشاهدة وقيم تباين خطأ ($\sigma^2 = 0.3,0.7$) هي مقدّر الجوار الأقرب (KNNE)، لامتلاكها أقل قيمة بالنسبة لمعيار المقارنة (ARMSE)، ماعدا ($\sigma^2 = 1$) الأفضلية كانْت لطريقة مقدّر الانحدار الخطّي الموضعي (LLRE)، لكونها تَمتْلك أقل قيمة لمعيار (ARMSE).

الجدول (4): معيار المقارنة (ARMSE) للطريقتين وباختلاف التباين مع توليد (50) مشاهدة بالنسبة للأنْموذج الثاني($m_2(x) = \sin(2\pi x^3)$)

Methods / σ^2	KNNE	LLRE

0.3	0.63650	0.58854
0.7	0.96891	0.85082
1	1.13960	0.98893

يتضح من الجدول (4) عند حجم عينة (50) مشاهدة وتباين الخطأ ($\sigma^2 = 0.3$) أنّ الأفضلية كانْت لطريقة مقدّر الانحدار الخطّي الموضعي (LLRE) لامتلاكها أقل قيمة لمعيار معدّل جذر متوسّط مربعّات الخطأ (ARMSE) والمساوية إلى (0.58854).

أمّا عندما تكون قيمة تباين الخطأ ($\sigma^2 = 0.7$) فأنْ أفضل طريقة كانْت هي مقدّر الانحدار الخطّي الموضعي (LLRE) لامتلاكها أقل قيمة لمعيار (ARMSE) والمساوية إلى (0.85082).

وفي حالة ($\sigma^2 = 1$) كانْت طريقة مقدّر الانحدار الخطّي الموضعي (LLRE) هي الأفضل وذلك لامتلاكها أقل قيمة لمعيار (ARMSE) والمساوية إلى (0.98893).

تَمّ التوصل إلى أنْ أفضل طريقة بالنسبة للأنْموذج الثاني عند حجم عينة (50) مشاهدة كانْت هي طريقة مقدّر الانحدار الخطّي الموضعي (LLRE) لكونها تَمتْلك أقل قيمة لمعيار (ARMSE) قيم تباين الخطأ جميعها.

الجدول (5): معيار المقارنة (ARMSE) للطريقتين وباختلاف التباين مع توليد (100) مشاهدة بالنسبة للأنْموذج الثانْي($m_2(x) = \sin(2\pi x^3)$)

Methods / σ^2	KNNE	LLRE
0.3	0.63371	0.57812
0.7	0.96585	0.84724
1	1.13117	0.91912

يتبين من الجدول (5) عند حجم عينة (100) مشاهدة وتباين الخطأ ($\sigma^2 = 0.3$) فإنْ أفضل طريقة كانْت هي طريقة مقدّر الانحدار الخطّي الموضعي (LLRE)، وذلك لامتلاكها أقل قيمة لمعيار معدّل جذر متوسّط مربعات الخطأ (ARMSE) والمساوية إلى (0.57812).

أمّا عندما تكون قيمة تباين ($\sigma^2 = 0.7$) الأفضلية كأنْت طريقة مقدّر الانحدار الخطّي الموضعي (LLRE) لامتلاكها أقل قيمة لمعيار (ARMSE) والمساوية إلى (0.84724).

وفي حالة ($\sigma^2 = 1$) كأنْت طريقة مقدّر الانحدار الخطّي الموضعي (LLRE) هي الأفضل لكونها تَمتْلك أقل قيمة لمعيار (ARMSE) والمساوية إلى (0.91912).

أظهرت النتائج المذكورة آنفاً أنْ أفضل طريقة بالنسبة للأنْموذج الثانْي عند حجم عينة (100) مشاهدة كأنْت هي طريقة مقدّر الانحدار الخطّي الموضعي (LLRE) لامتلاكها أقل قيمة لمعيار (ARMSE) ولجميع قيم تباين الخطأ.

الجدول (6): معيار المقارنة (ARMSE) للطريقتين وباختلاف التباين مع توليد (150) مشاهدة بالنسبة للأنْموذج الثانْي($m_2(x) = \sin(2\pi x^3)$)

Methods / σ^2	KNNE	LLRE
0.3	**0.62840**	**0.57684**
0.7	**0.96458**	**0.84150**
1	**1.13059**	**0.90578**

يتضح من الجدول (6) عند زيادة حجم العيّنة إلى (150) مشاهدة وتباين خطأ ($\sigma^2 = 0.3$) فإنْ الأفضلية كأنْت لطريقة مقدّر الانحدار الخطّي الموضعي (LLRE)، وذلك لامتلاكها أقل قيمة لمعيار معدّل جذر متوسّط مربعات الخطأ (ARMSE) والمساوية إلى (0.57684).

أمّا عندما تكون قيمة تباين الخطأ ($\sigma^2 = 0.7$) كأنْت الأفضلية لطريقة مقدّر الانحدار الخطّي الموضعي (LLRE)، لكونها تَمتْلك أقل قيمة لمعيار (ARMSE) والمساوية إلى (0.84150).

وفي حالة ($\sigma^2 = 1$) كأنْت طريقة مقدّر الانحدار الخطّي الموضعي (LLRE) هي الأفضل لامتلاكها أقل قيمة لمعيار المقارنة (ARMSE) والمساوية إلى (0.90578).

تَمّ التوصل إلى أنْ أفضل طريقة بالنسبة للأنْموذج الثاني عند حجم عينة (150) مشاهدة كأنْت هي طريقة مقدّر الانحدار الخطّي الموضعي (LLRE) وذلك لامتلاكها أقل قيمة لمعيار (ARMSE) قيم تباين الخطأ جميعها.

5- التطبيق Application

في هذا الجانْب سيَتَمّ دراسة اثر الهيموغلوبين Hemoglobin (Hb) والمتَمّثل بالمتغيّر المستقل على حجم الخلايا الحمر المكدسة Packed Cell Volume (PCV)، والمتَمّثل بالمتغيّر التابع للمرضى المصابين بأمراض الكُلّى المزمنة، والتي تَمّ الحصول عليها من مستودع التعلم الآلي (UCI Machine Learning Repository) [1]، إذْ تَمّ استعمال معيار للمقارنة بين الطريقتين، والمتَمّثل بمعيار جذر متوسّط مربعات الخطأ (RMSE)، إذْ أنْ الطريقة التي تَمّثلك أقل قيمة RMSE تعدّ أفضل والمعرفة بالمعادلة رقم (15):

RMSE

$$= \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i(x_i) - \hat{m}\,(x_i)\right)^2} \qquad \ldots (15)$$

إذْ أنْ:

$y_i(x_i)$: تَمّثل المتغيّر التابع للمريض i.

$\hat{m}\,(x_i)$: دالّة الانحدار اللامعلمي المقدّرة عند قيمة المتغيّر المستقل للمريض i.

n: تَمّثل عدد المشاهدات.

وقد تَمّ تقدير أنْموذج الانحدار اللامعلمي للبيانات الحقيقية بطريقين للتقدير والتي تَمّ ذكرهما سابقاً لمعرفة سلوكها عند أنماذج وأحجام عينات مختلفة وقيم مختلفة للانحراف المعياري للخطأ، إذْ وضعت قيم المتغيّر التابع المقدّرة باستعمال هذه الطرائق في الجدول (7)، في حين كأنْت نتائج معايير المقارنة كما في الجدول (8)، ورسمت هذه القيم مع القيم الحقيقية كما في الأشكال (3) إلى (4).

الجدول (7): البيانات المقدّرة بطريقتي التقدير (LLRE و KNNE)

No.	KNNE	LLRE	No.	KNNE	LLRE	No.	KNNE	LLRE
1	31.67	34.49	51	35.67	39.15	101	44.00	45.66
2	38.67	38.13	52	26.00	26.03	102	44.33	44.40
3	29.33	28.69	53	33.67	32.40	103	46.67	46.21
4	32.33	33.09	54	23.00	24.71	104	41.33	45.49

[1]. عبارة عن موقع يحتوي على عدد كبير من مجاميع البيانات ويمكن الوصول إليه من خلال الرابط الآتي:
https://archive.ics.uci.edu/

5	30.67	29.35	55	30.33	30.34	105	43.00	44.51
6	22.00	16.16	56	33.00	34.85	106	48.33	46.64
7	23.00	22.45	57	29.33	28.69	107	47.00	45.94
8	37.67	39.82	58	29.67	30.01	108	46.00	44.63
9	24.33	22.77	59	23.67	24.06	109	47.00	45.84
10	29.00	29.68	60	31.67	34.49	110	45.33	46.44
11	29.67	30.01	61	26.33	26.36	111	42.33	44.74
12	34.67	35.95	62	36.67	37.43	112	46.00	44.87
13	35.67	39.15	63	23.33	21.49	113	42.33	46.51
14	30.33	30.68	64	32.33	33.44	114	41.67	44.14
15	33.00	34.85	65	32.33	33.44	115	48.33	46.57
16	43.00	40.83	66	38.67	38.13	116	45.67	45.75
17	30.67	29.35	67	43.00	40.83	117	47.33	42.42
18	24.00	23.42	68	35.33	35.58	118	48.00	45.10
19	35.67	38.81	69	24.00	23.42	119	44.33	44.99
20	30.33	30.34	70	30.00	29.02	120	48.33	46.64
21	34.00	32.05	71	33.00	31.71	121	39.33	41.18
22	32.67	33.79	72	29.67	28.36	122	44.33	46.76
23	50.67	45.21	73	45.33	42.13	123	50.67	45.21
24	29.00	29.68	74	38.67	38.13	124	46.00	42.68
25	27.33	27.36	75	24.33	25.70	125	46.67	43.40
26	37.00	37.08	76	37.00	39.48	126	44.67	46.82

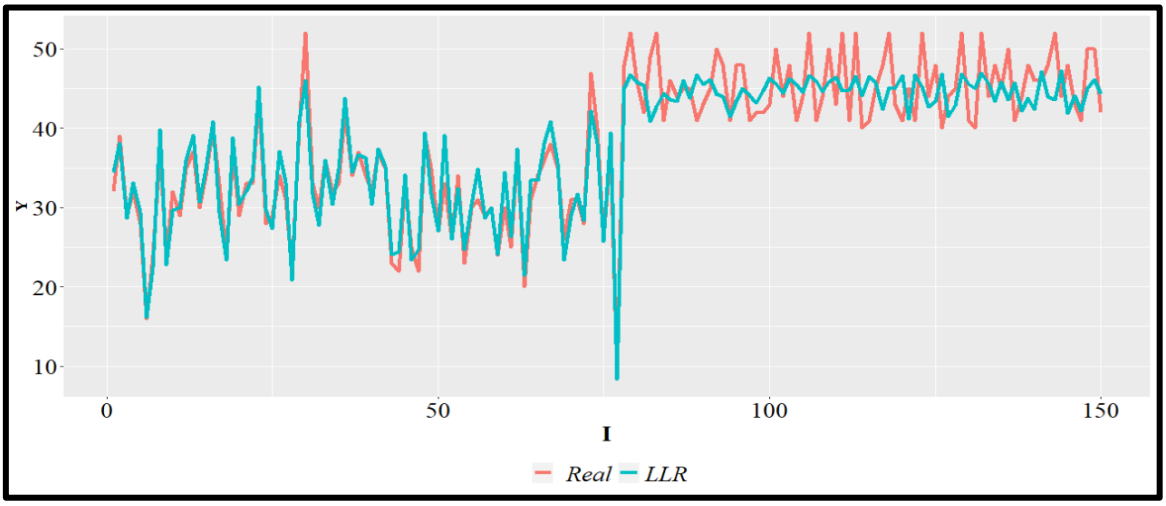
27	32.33	33.09	77	19.33	8.32	127	45.67	41.51
28	22.67	20.86	78	44.33	44.99	128	46.00	42.94
29	43.00	40.83	79	44.67	46.70	129	44.33	46.88
30	49.00	46.04	80	47.00	45.84	130	45.67	45.58
31	33.00	31.71	81	41.67	45.40	131	44.33	44.99
32	28.00	27.70	82	43.00	40.83	132	44.33	46.94
33	34.67	35.95	83	46.00	42.68	133	44.00	45.66
34	30.33	30.34	84	44.33	44.40	134	46.67	43.40
35	33.00	34.85	85	50.00	43.62	135	47.00	45.84
36	45.33	43.82	86	46.67	43.40	136	50.00	43.62
37	31.67	34.49	87	49.00	46.04	137	45.67	45.75
38	35.00	36.71	88	45.33	43.82	138	45.33	42.13
39	36.00	36.33	89	44.67	46.70	139	45.33	43.82
40	30.33	30.34	90	41.33	45.49	140	47.33	42.42
41	36.67	37.43	91	47.33	46.13	141	49.33	47.15
42	34.00	35.21	92	42.33	44.27	142	45.00	43.99
43	23.67	24.06	93	45.00	43.99	143	50.00	43.62
44	23.33	24.39	94	45.67	41.51	144	50.00	47.22
45	31.67	34.14	95	46.67	43.40	145	44.00	41.83
46	24.00	23.42	96	44.33	44.99	146	41.67	44.14
47	23.00	24.71	97	41.67	44.14	147	45.33	42.13
48	37.00	39.48	98	46.67	43.17	148	44.33	44.99

49	33.00	31.71	99	42.33	44.74	149	47.33	46.13
50	28.00	27.03	100	47.67	46.37	150	42.33	44.27

الجدول (8): يبين قيمة معيار المقارنة (RMSE) للبيانات الحقيقية

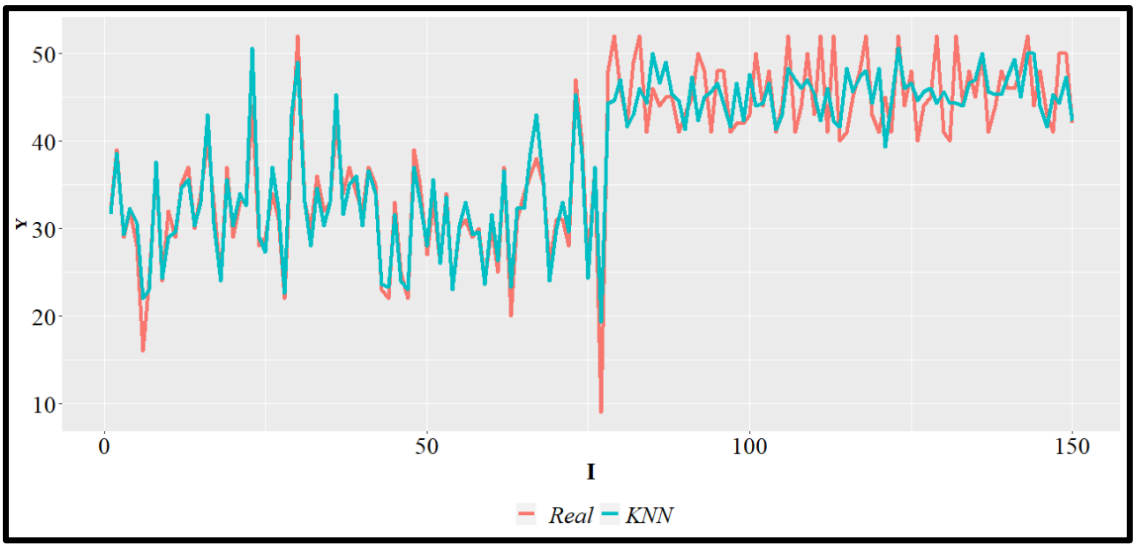
Methods	KNNE	LLRE
RMSE	3.280137	3.317077

تبين من الجدول (8) أنْ أفضل طريقة كانَتْ هي طريقة مقدّر الجوار الأقرب (KNNE) لامتلاكها أقل قيمة لمعيار جذر متوسّط مربعات الخطأ (RMSE) والمساوية إلى (3.280137).



الشكْلّ (3): حجم الخلايا الحمر المكدسة الحقيقية مع المقدّرة بطريقة مقدّر الانحدار الخطّي الموضعي

يتبين من الشكْلّ (3) تقارب سلوك البيانات المقدّرة بطريقة الانحدار الخطّي الموضعي (LLRE) مع البيانات الحقيقة، والمتَمَثّلة بحجم الخلايا الحمر المكدسة (PCV).



الشكْل (4): حجم الخلايا الحمر المكدسة الحقيقية مع المقدّرة بطريقة مقدّر الجوار الأقرب

يتبين من الشكْل (4) تقارب سلوك البيانات المقدّرة بطريقة الجوار الأقرب (KNNE) مع البيانات الحقيقة والمتَمَثّلة بحجم الخلايا الحمر المكدسة (PCV).

6- الاستنتاجات Conclusions

بناءً على ما تَمّ التوصل اليه من خلال نتائج تجارب المحاكاة (الجانْب التجريبي) والجانْب التطبيقي نستنتج ما يأتي:

6-1 استنتاجات الجانْب التجريبي Conclusions of the Experimental Side

- بالنسبة للأنْموذج الأول كانْت الأفضلية لطريقة مقدّر الجوار الأقرب (KNNE)عند حجوم العينات (50,100,150) وبمستوى تباين خطأ $(\sigma^2 = 0.3,0.7)$، باستثناء قيمة $(\sigma^2 = 1)$، ولأحجام العينات (50,100,150) كانْت الطريقة الأفضل هي مقدّر الانحدار الخطّي الموضعي (LLRE).
- بالنسبة للأنْموذج الثاني تفوق مقدّر الانحدار الخطّي الموضعي (LLRE) عند حجوم العينات (50,100,150) مستويات التباين المختلفة جميعها.

6-2 استنتاجات الجانْب التطبيقي Conclusions of the Applied Side

- من خلال النتائج للبيانات الحقيقة والمتَمَثّلة بالمصابين بأمراض الكُلّى المزمنة تَمّ التوصل إلى أنْ أفضل طريقة كانْت هي طريقة مقدّر الجوار الأقرب (KNNE).

7- التوصيات Recommendation

- اعتَمَاد طريقة مقدّر الجوار الأقرب (KNNE) كطريقة لتقدير أنْموذج الانحدار اللامعلمي، كونها أثبتتْ جدارتها في نتائج المحاكاة والتطبيق.
- إجراء مقارنة بين هذه الطرائق المعتَمّدة بأنْموذج شبة معلمي.
- اعتَمَاد طرائق أخرى في تقدير أنْموذج الانحدار اللامعلمي.

المصادر **References**

أولاً: المصادر العربية.

1. الاشقر، السيد الشربيني، (2023)، " استعمال بعض الشرائح لتَمّهيد معدّلات الوفاة الخام بجدول الاكتواري المصري "، المجلّة العلمية للدراسات والبحوث المالية والتجارية، كُلّية التجارة، جامعة دمياط، 4 (1) 3، 512-547.
2. بدر، دريد حسين، وعبد، هيفاء طه، (2019)،" مقارنة بعض طرائق تقدير أنْموذج دالّة الانحدار اللامعلمي باستعمال المحاكاة "، مجلّة الادارة والاقتصاد، العدد (121).
3. حسن، فارس طاهر، ومجيد، غياث حميد، (2016)، "تحديد أفضل أسلوب تَمّهيدي لتقدير أنْموذج انحدار لامعلمي مع تطبيق عملي" .مجلّة كُلّية التربية الاساسية ، 22 (93 /علمي)،1-22.
4. الصفاوي، صفاء يونس، ومتي، نور صباح ، (2011). " تقدير دوال الانحدار اللامعلمي باستعمال بعض أساليب التَمّهيد ".المجلّة العراقية للعلوم الإحصائية ،العدد 20 (373- 392).

ثانيا: المصادر الإنكليزية.

5. Aydin, D., (2007), "A comparison of the nonparametric regression models using smoothing spline and kernel regression", World Academy of Science, Engineering and Technology, 36, 253-257
6. Burba, F., Ferraty, F., & Vieu, P. (2009). k-Nearest Neighbour method in functional nonparametric regression. Journal of Nonparametric Statistics, 21(4), 453-469.
7. Chai, T., & Draxler, R. R. (2014). "Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature". Geoscientific model development, 7(3), 1247-1250.
8. Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. Journal of the American statistical association, 74(368), 829-836.
9. Denis H. and Yan L. (2005), "Cross-Validation in Nonparametric Regression with Outliers", The Annals of Statistics, Vol. 33, No. 5, pp. 2291-2310.
10. Denison, D.G.T., Mallick, B.K. & Smith, A.F.M., (1998), "Automatic Bayesian Curve Fitting", J.R. Statist. Soc. B, Vol. 60, 333-350.
11. Eguasa, O., Edionwe, E., & Mbegbu, J. I. (2023). Local Linear Regression and the problem of dimensionality: a remedial strategy via a new locally adaptive bandwidths selector. Journal of Applied Statistics, 50(6), 1283-1309.
12. Fan, J. (1993). Local linear regression smoothers and their minimax efficiencies. The annals of Statistics, 196-216.

13. Ferraty, F., & Vieu, P. (2004)." Nonparametric models for functional data, with application in regression, time series prediction and curve discrimination". Nonparametric Statistics, 16(1-2), 111-125.
14. Fix, E. ; Hodges , J. (1989)." Discrimintory Analysis . Nonparametric Discrimintion :consistency properties ". Inter.Stat. Re.,57, 238 -247.
15. Györfi, L. Kohler, M. Krzyzak, A. and Walk, H. (2002). Distribution-Free Theory of Nonparametric Regression, New York: Springer.
16. Härdle, W., Müller, M., Sperlich, S., & Werwatz, A. (2004). Nonparametric and semiparametric models (Vol. 1). Berlin: Springer.
17. Hardle.W.(1990),"Applied nonparametric regression",Cambridge- MA: cambirg university press.
18. Linton,O. Xiao, Z.(2001)." A Nonparametric Regression Estimator That Adapts to Error Distribution of Unkown From" Working Paper . o.linton @ Ise.ac. uk.
19. Mundra, K. (2005). "Nonparametric slope estimation for fixed effects paneldata. Working Paper" , Department of Economics, San Diego State University.
20. Smith, M., Sheather, S. & Kohn, R., (1996), " Finite Sample Performance of Robust Bayesian Regression", Computational Statistics, Vol. 11, 269-301.
21. Theodoridis S. and Koutroumbas K.,(2003),"Pattern Recognition " 2nd ed. ,Elsevier(USA).