تقدير أنموذج الانحدار اللامعلمي لبيانات الفترة باستعمال المحاكاة

يسرى جاسم سهيل
سعد كاظم حمزة

جامعة بغداد /كلية الادارة والاقتصاد/ قسم الإحصاء

المستخلص

ان استعمال النماذج المعلمية يتطلب توافر العديد من الشروط الاولية لكي تكون قراءة هذه النماذج بصورة صحيحة الامر الذي دفع الباحثون الى البحث عن نماذج اقل تعقيدا من النماذج المعلمية تمثلت هذه النماذج بالنماذج اللامعلمية. وان احد الاساليب المستعملة في تقدير دالة الانحدار اللامعلمي هو التقدير اللبي(kernel estimation) ,كما ان احد المشاكل الاخرى التي تواجهنا هي وجود بيانات على شكل فترات (intervals) حيث تمتلك هذه البيانات تبايناً داخلياً غير موجود في البيانات التقليدية وبسبب هذا التباين تطلب تطوير طرائق جديدة لتحليل هذه البيانات وتمثلت هذه الطرائق ب (N-W:CR N-W:C, L-L-R:CR ,) حيث تم تطبيق هذه الطرائق ومقارنتها من خلال اسلوب محاكاة ,فضلا عن استعمال انموذجين وثلاثة حجوم عينات (100,200,300) وانحرافات مختلفة .ومن ملاحظة نتائج المحاكاة تبين ان افضل طريقة هي CK+RL .

1- المقدمة:

ان التقدم العلمي في المجالات كافة تطلب ان يكون هناك تطورا في علم الاحصاء ,هذا العلم الذي يطوع نظرياته لخدمة العلوم الاخرى ,سواء كانت انسانية ام اجتماعية ام علمية وغيرها ,لان جميعها تعتمد على جمع المعلومات وتبويبها وتحليلها للحصول على النتائج المطلوبة .
عندما تكون هناك معرفة بالظاهرة المطلوب دراستها وتحليل بياناتها ,كأن تكون دالة سببية تصف علاقة ضغط الدم بعمر الشخص ,لذا يكون من المعقول تكوين صيغة رياضية بشكل دالّي للظاهرة والاعتماد على هذا الشكل من خلال بعض المعلمات التي من الممكن ان تلخص هذا الشكل.وينصب العمل الاحصائي باهتمام على هذه المعلمات كملخص وجيز عن المشاهدات ,ويسمى هذا بالتحليل المعلمي .
اما في حالة غياب المعرفة عن الظاهرة ,كان تكون تجربة تقام لاول مرة او ظاهرة لايمكن تحديد العلاقة السببية التي تربط متغيراتها بصورة قاطعة مع بعض ضمن صيغة رياضية مسبقة ,لذلك ستكون الحاجة للتحرر من استخدام الشكل الداليّ .فبدلاً من ذلك, يتم استخدام اسلوب اكثر مرونة وهو ان ترك البيانات تفصح عن المعلومات التي تحملها دون فرض قيود عليها لانه ليس هناك معلمات تنوب عن المشاهدات ,ويسمى هذا بالتحليل اللامعلمي .[8]
ولذا برز استعمال طرائق التمهيد كاداة فعّالة تعتمد على البيانات بدلاً من الصيغة الدالّية المسبقة للبيانات.وأحد اهم انواع الممهدات وابرزها هي الممهد اللبي kernel وان هذا الممهد يمكن استعماله لتقدير اي دالة احصائية ,اما خصائصه فهي سهولة برمجته على الحاسب الالكتروني ورسم مقدراته بيانيا وكذلك مقدراته تكون غير متحيزة بالمحاذي Asymptotically Unbiased ومتسقة بالمحاذي Asymptotically Consistent ويعمل على تعديل المشاهدات وعلى تقريب دالة الانحدار اللامعلمي الى دالة الانحدار اللامعلمي الحقيقية .[3] [5] [7]
وسيتم التركيز في هذه الرسالة على الممهدات اللبية في حالة وجود بيانات فترة ,حيث يتمتع الانحداراللبي kernel لبيانات الفترة بنفس مزايا وعيوب انحداراللبي kernel للبيانات التقليدية ومع ذلك فان له القدرة على نمذجة العلاقة بين المتغير المعتمد والمتغيرات المستقلة التي تسمح باخذ التباين الداخلي في الاعتبار.
ومن مميزات استخدام الانحداراللبي kernel لمدى البيانات هي التاكد من ان تقدير الحد الاعلى اكبر من تقدير الحد الادنى لكل متغير معتمد.

2- البيانات الرمزية symbolic data

الاحصاء هو علم وفن التعامل مع البيانات . ومع ذلك,فان البيانات التي نتمكن عادة من تحليلها هي بيانات تقليدية حيث تكون كل مشاهدة هي نقطة واحدة . ولكن قد لاتكون البيانات التي نريد تحليلها عبارة عن نقطة واحدة , بدأ Diday في عام (1987) مفهوم البيانات الرمزية للتعامل مع هذه الظاهرة .يمكن ان تكون البيانات الرمزية عبارة عن فترات (intervals) , قوائم (lists), رسوم بيانية (histograms) وهذه كلها امثلة على البيانات الرمزية . تختلف البيانات الرمزية عن البيانات التقليدية , ان اكبر اختلاف بينهما هو ان قيم البيانات التقليدية ليس لها تباين داخلي بينما في المقابل فان قيم البيانات الرمزية لها تباين داخلي .[5]

سوف ناخذ بيانات الفترة كمثال لبيان الفرق بينهما كما في الجدول التالي:

الجدول (1) الفرق بين البيانات التقليدية والبيانات الرمزية

البيانات	القيم الاصلية	البدائل التقليدية		البدائل الرمزية	
		القيم	التباين	القيم	التباين
المشاهدة 1	45	45	0	45	0
المشاهدة 2	[30,60]	45	0	[30,60]	75
المشاهدة 3	[20,70]	45	0	[20,70]	208.3

في حالة عدم المعرفة في كيفية التعامل مع بيانات الفترة , ياخذ المحللون عادة بديل تقليدي مثل المتوسط للقيم الاصلية . في حالة افتراض ان القيم تتبع توزيع منتظم داخل الفترة x~uniform(a,b).حيث ان((a) يمثل الحد الادنى و (b) يمثل الحد الاعلى) بالنسبة للبديل التقليدي ناخذ فقط نقطة المتوسط لإجراء التحليل مع اهمال قيم التباين الداخلي .وبالتالي بالنسبة للبيانات (45) و (30,60) و(20,70), سنحصل على نفس المعلومات باستخدام تحليل البيانات التقليدي , ولكن في الواقع هناك اختلاف كبير بين القيم . يتم حساب التباين حسب المعادلة التالية:

$$\mathrm{Var}(x)=\frac{(b-a)^2}{12} \qquad \dots(2.1)$$

تعطي البدائل التقليدية للمشاهدات الاصلية الثلاثة المختلفة نفس القيم والتباين . اما البدائل الرمزية فهي تعكس القيم الحقيقية والتباين للمشاهدات الاصلية .[12,31]

3- الانحدار اللبي kernel Regression

وهو طريقة احصائية لا معلمية تهدف الى ايجاد علاقة لاخطية بين المتغير المعتمد والمتغيرات المستقلة كما انه طريقة مبسطة تستخدم لايجاد هيكلية او نمط مجموعة من البيانات دون الحاجة الى ضرورة وجود انموذج معلمي .

ولمعرفة طريقة عمل ممهدات kernel بصورة عامة فان معدلات موزونة مختلفة يمكن استعمالها كمقدرات لدالة الانحدار الغير معروفة m(x) حيث يتم تعريف هذه الطريقة فقط من المشاهدات المجاورة لقيمة ثابتة x طالما ان مشاهدات y من النقاط البعيدة عن x سوف تمتلك قيم متوسطات مختلفة.

حيث ان هذه الطريقة تكون مشابهة الى طريقة المربعات الصغرى الموزونة وكما ياتي :

$$\widehat{m}(x)=n^{-1}\ \sum_{i=1}^{n} W_{hi}(x)Y_i \qquad \ldots(2.2)$$

حيث ان $W_{hi}(x)$ تشير الى سلسلة من الاوزان وان دوال الوزن تكون طبيعية اذا كان $n^{-1}\ \sum_{i=1}^{n} W_{hi}(x) = 1$ وتكون غير سالبة اذا كانت $W_h \geq 0$ وتكون دالة وزن احتمالية اذا كانت غير سالبة وطبيعية .[25] [78] [3] [35][40]

بفرض ان دالة الوزن W_h (.) تمتلك الخصائص التالية:

1. $(x)W_h$ for $0< |x| <1$
2. $(x)= W_h\ (-x)W_h$
3. $x\geq 0$ ان دالة الوزن تكون دالة غير متزايدة
4. $(x)=0W_h$ for $|x|\geq 1$

حيث ان كمية المعدل نسيطر عليها من خلال سلسلة الاوزان $W_{hi}(x)$ والتي تلائم بواسطة معلمة التمهيد h (عرض الحزمة) حيث ان هذه المعلمة هي التي تتحكم بعرض الجوارحول x اي تسيطرعلى حجم الاوزان .

يتم تحديد شكل الاوزان من خلال دالة kernel او تسمى (دالة الوزن او دالة النافذة) .وان هذه الدالة تصل الى اعظم ما يمكن عندما x_i يقترب الى x وتتناقص عندما x_i يبتعد عن x. [12] [8] [6]

ان الوزن يتم الحصول عليه عادة من خلال تطبيق دالة kernel القائمة على مربع المسافة الاقليدية Euclidean (Squared distance) بين x و x_i والتي تعطى كما ياتي :

$$d(x,x_i) =\sum_{i=1}^{n}(x-x_i)^2 \qquad \ldots (2.3)$$

4- الدوال اللبية kernel functions

عند استخدام الطريقة اللبية kernel لتقدير دالة الانحدار اللامعلمي يجب تحديد كلاً من دالة kernel و عرض الحزمة (معلمة التمهيد) وقد اكد الباحث (Hardel 1990) على ان اختيار هذه الدالة يكون اقل اهمية من اختيار عرض الحزمة. وان هذه الدالة يمكن اخذها مع اسناد محدود [-1,1] , اذ يوجد سلسلتان من دوال kernel وهما دوال kernel ذات اقل تباين (Minimum Variance) والتي تعمل على تقليل التباين المحاذي , ودوال kernel المثلى (Optimal kernel) والتي تعمل على تقليل MISE المحاذي , اي اشتقاق MISE بالنسبة لدالة kernel.حيث ان دالة kernel ذات اقل تباين من رتبة 2 تمثل حل لمسألة تقليل التباين التالية :

$$C_k = \int_{-1}^{1} k^2(u)du \qquad \ldots(2.4)$$

حيث ان$(\frac{x-x_i}{h})$ = u

وان هذه الدالة تمثل الدالة المستطيلة : $I[-1,1]$, $K(u)=\frac{1}{2}$

اما التحيز المحاذي d_k وMISE المحاذي الامثل فتم تعريفهما كما ياتي :

$$d_k = \int_{-1}^{1} u^2 k(u) du \qquad \dots (2.5)$$

$$\text{MISE}=C_k^2\, d_k \qquad \dots (2.6)$$

وان دالة kernel المثلى تعمل على تقليل MISE المحاذي وان الدالة الناتجة من اشتقاق MISE تسمى بدالة Epanchnikov kernel وان استخدام هذه الدالة يؤدي الى الحصول على خصائص مثلى محاذية

وغالبا ما تفضل في التطبيق .الا ان ذلك لا يمنع من استخدام دوال اخرى تبعاً لطبيعة البيانات المستخدمة في البحث [14][18].

وسيتم في هذا الرسالة استخدام اللبة الكاوسية Gaussian kernel والتي تعطى كما ياتي :

$$k_h\big(d(x,x_i)\big) = \frac{1}{\sqrt{2\pi}}\frac{1}{h}e^{-\frac{d(x,x_i)}{2h^2}} \qquad \dots (2.7)$$

والتي ينتج عند استخدامها تاثيرات مرئية جيدة لممهدات kernel ولكن بجهد حسابي اكبر.

في دالة kernel هذه تم تقدير معلمة التمهيد h بالاعتماد على الانحراف المعياري للتوزيع الطبيعي الى x_i^c.[21]

5- تقدير دالة الانحدار اللامعلمي

لتكن (x_n, y_n),….., (x_1, y_1) تمثل عينة عشوائية ماخوذة من مجتمع معين له دالة كثافة احتمالية f(x) ودالة كثافة احتمالية مشتركة f(x,y).

وللتعبير عن انموذج الانحدار العام تم استعمال الصيغة الاتية :

$$y_i = m(x_i)+e_i \qquad \dots\ (2.8) \qquad i=1,2\dots\dots\dots n$$

حيث ان :-

y_i: تمثل قيمة المتغير المعتمد للمشاهدة i.

$m(x_i)$: تمثل دالة الانحدار غير المعروفة والمطلوب تقديرها.

x_i: تمثل قيمة المتغير المستقل للمشاهدة i.

e_i: تمثل قيمة الخطا العشوائي $e_i \sim N(0,\sigma^2)$

وان تقدير دالة الانحدار $\hat{y} = \hat{m}(x)$

وان دالة انحدار Y على X تكون كما ياتي:

$$m(x)=E(Y/X=x)=\int y\, f(y/x)dy$$
$$= \int y\, \frac{f(x,y)}{f(x)} dy$$

ومن هنا يتضح انه في حالة افتراض (x_i) متغير عشوائي مستقل ومتماثل التوزيع (i,i,d) على الفترة (0,1) مع دالة كثافة احتمالية $f(x)$, فان دالة الانحدار يمكن ايجادها بسهولة اذا كانت دالة الكثافة الاحتمالية المشتركة f(x,y) معروفة .[46][11].

ويمكن ان تكون البيانات على شكل فترات حيث القيم المشاهدة للمتغيرات x_i و y_i لها حد ادنى(Lower) bound وحد اعلى (Upper bound) كما ياتي :

$$x_i = [x_{Li}, x_{Ui}] \in £=\{[x_L, x_U]: x_L, x_U \in R \ \ x_L \le x_U\}$$

$$\text{and} \quad y_i=[y_{Li}, y_{Ui}] \in £$$

ولقد تم اختيار كل من Nadaraya-Watson Smoother و Local Linear Regression من بين ممهدات kernel . لإيجاد دالة الانحدار التقديرية $\widehat{m}(x)$.

6- ممهد نادرايا واتسون Nadaraya-Watson Smoother (N-W)

ويعد من اقدم الممهدات اللامعلمية واكثرها شيوعاً واستعمالا وتم تسميته اعتماداً على اسماء الباحثين اللذين اقترحا هذا الممهد وهما كل من(Nadaraya و Watson) في عام 1964 مع العلم ان هذه الطريقة تعود الى تقديم سابق من قبل الباحث Tukey (1961) لما يسمى (Regressogram) حيث تم الاعتماد على استخدام فكرة الرسم البياني لتقدير دالة الكثافة . وان هذا الممهد يكون معدل موزون محدد لقيم Y_i حيث ان صيغته تكون كما في المعادلة (2.2) والتي هي :[37][30] [87]

$$\widehat{m} = n^{-1} \sum_{i=1}^{n} W_{hi}(x) Y_i$$

وان سلسلة الوزن تكون كما ياتي :

$$W_{hi}(x) = k_h(x - x_i) / f_n(x) \qquad \dots(2.9)$$

وان $f_n(x)$ يمثل مقدر كثافة Parzen لدالة الكثافة الاحتمالية ل x ويكون كما ياتي :

$$f_n(x) = n^{-1} \sum_{i=1}^{n} k_h(x - x_i) \qquad \dots(2.10)$$

$k_h(u) = h^{-1}\, k(u/h)$ مع عرض الحزمة h ويتضح من الصيغة السابقة ان هذا الممهد يكون خطياً وان دالة kernel تكون دالة حقيقية متماثلة محددة مستمرة وتكاملها مساو الى الواحد اي انها تحقق الشروط الاتية :- [72][26][46][49]

1- $\sup |k(u)| < \infty \quad , \int_{-\infty}^{\infty} k(u)du = 1 \quad , \quad \int_{-\infty}^{\infty} |k(u)|du < \infty \quad ,$

$\lim_{|u| \to \infty} |u||k(u)| = 0$

2- $\int_{-\infty}^{\infty} u^q k(u)du = 0 \qquad , q = 1,2, \dots \dots , z - 1 \qquad \dots(2.11)$

3- $\int_{-\infty}^{\infty} |u^q k(u)|du < \infty \qquad , q = z$

يمكن الاستدلال من الشروط العزم (2 و 3) المذكورة اعلاه ان دالة kernel تكون من رتبة z . وتم ملاحظة ان هذه الشروط تتحقق عند z=2 اي تكون الدوال من الرتبة الثانية.

ان هذا الممهد يمكن اعتباره ايضا متوسط متحركاً للعينة y والتي اوزانها تعتمد على دوالkernel , اي ان التقدير عند نقطة معينة x يمثل المعدل لمشاهدات y_i طبقاً الى x_i الذي ينتمي الى النافذة[x-h,x+h] , اي بالاعتماد على عرض الحزمة h والتي تحدد عدد النقاط القريبة المستخدمة في التقدير. حيث اذا كانت h→0 اي عرض الحزمة صغير فان المقدر يكون كما ياتي:

$$\widehat{m}(x) = Y_i \qquad \dots(2.12)$$

اي ان المنحنى المقدر يمر بجميع نقاط البيانات ويكون له تباين عالي وتحيز قليل ويسمى منحنى تحت التمهيد under (smooth). اما اذا كانت h→∞ اي ان عرض الحزمة كبير فان المقدر يكون كما ياتي :

$$\hat{m}(x) = n^{-1}\sum_{i=1}^{n} Y_i \qquad \ldots(2.13)$$

والذي يمثل المعدل لجميع المشاهدات وان المنحنى المقدر يكون قريب من الخط المستقيم وله تباين قليل وتحيز عالي ويسمى منحنى فوق التمهيد (Over smooth). لذلك يجب اختيار عرض الحزمة h بصورة مثلى ليوازن بين التحيز والتباين .ومن صفات هذا الممهد انه يعطي تقديراً مستمراً للانحدار وخاصة عندما تكون دالة kernel المستعملة مستمرة .ومن الجدير بالذكر ان هذا المقدر تمت دراسته بالنسبة للتصميم العشوائي والتصميم الثابت (يكون انموذج التصميم عشوائي ,إذ تم افتراض ان x_i يكون متغيراً عشوائياً .ويكون ثابت اذ تم افتراض ان x_i يكون متغيراً غير عشوائيا) ولكنه قليلاً ما يدرس في حالة التصميم الثابت.ونلاحظ ان هذا الممهد يعاني من التحيزحيث ان التحيز فيه يعتمد على مواقع نقاط البيانات x_i,كما انه يعاني من تاثير التحيز عند حدود الفترة (Boundary Bias) والذي ينتج بسبب ان اغلب الاوزان تكون غير معلومة وتقع خارج الحد.لذلك يتطلب تعديلات على الحد حيث تم اللجوء الى بتر فترة المشاهدات للتخلص من تأثيرالحد وذلك لكون الفترة الكلية للمشاهدات تتأثر بتأثرات الحد.[22] [17] [4] [1].

اما لاشتقاق التحيز والتباين لهذا الممهد في حالة التصميم العشوائي فان الشروط الاتية يجب توفرها :[21]

1 - ان دالة الانحدار تكون قابلة للاشتقاق مرتين .

2 -1<|u| for K(u)=0

$$\ldots(2.14)\int_{-1}^{1} u^q k(u)du = \begin{cases} 1 & for \quad q=0 \\ 0 & for \quad q=1 \\ \neq 0 & for \quad q=2 \end{cases}$$

3 – $| k(u) - k(v) | \leq | u - v|^{\gamma}$, for some $\gamma > 0$ and some $u, v \in [-1,1]$

اي ان دالة kernel تحقق شرط Holder في حالة $\gamma = 1$

4 - $| f(u) - f(v) | \leq | u - v|^{B}$, for some $B > 0$ and for $u, v \in [0,1]$

-5 $h \to 0$, $\frac{nh}{\log(n)} \to \infty$ عندما $n \to \infty$

حيث ان التحيز Bias والتباين Variance لممهد Nadaraya-Watson في حالة التصميم العشوائي Random تكون كما ياتي:

$$\text{Bias} = \left(\frac{1}{2}m''(x) + \frac{m'(x)f'(x)}{f(x)}\right)d_k h^2 \qquad \ldots(2.15)$$

$$\text{Variance} = \frac{\sigma^2(x)}{nh\, f(x)} C_k \qquad \ldots(2.16)$$

حيث ان التباين الشرطي ل$(Y/X = x)$

$$\sigma^2(x) = E(y^2/X = x) - m^2(x) \qquad \ldots(2.17)$$

وفيما يخص متوسط مربعات الخطا المرجح MISE الشرطي المحاذي في الفترة [a,b] في حالة التصميم العشوائي فيكون كما ياتي :[77]

$$\text{Condi MISE}(\hat{m}(x)) = h^4 d_k^2 \int_a^b \left(\frac{(m''(x)}{2} + \frac{m'(x)f'(x)}{f(x)}\right)^2 W(x)dx + \frac{C_k}{n\,h}\int_a^b \frac{\sigma^2(x)}{f(x)} W(x)dx + o_p\left(h^4 + \frac{1}{n\,h}\right)(2.18)$$

اذ ان W(x) تمثل دالة وزن موجبة محددة مستندة الى الفترة [a,b] لدالة الكثافة f(x).

7- ممهد الانحدار الخطي الموضعي **Local Linear Regression Smoother (L-L-R)**

يعد هذا الممهد من افضل طرائق التمهيد حيث تم اقتراحه من قبل Cleveland عام (1979) كطريقة لا معلمية تستخدم لنمذجة العلاقة بين متغير مستقل ومتغير معتمد , اذ تُعد طريقة مرنة تسمح بتقدير العلاقة الموضعية عند كل نقطة في البيانات , بدلاً من افتراض وجود علاقة عامة بين المتغيرات وذلك عن طريق ملاءمة انموذج انحدار خطي مع مجموعة بيانات موضعية بجوار النقطة x. ويعرف هذا الجوار بانه معلمة التمهيد h والتي تحدد عدد النقاط القريبة المستخدمة في التقدير اما دالة kernel فهي تحدد الكيفية التي تساهم بها المشاهدات لملائمة هذا الممهد عند النقطة x. ويتم ذلك من خلال تصغير مجموع مربعات البواقي الموزونة بدالة kernel . حيث يعمل هذا الممهد على اكتشاف العلاقات غير الخطية بين المتغيرات وخصوصاً عندما تكون هنالك تباينات كبيرة . ويتمتع هذا الممهد بقابلية التكيف للتصميم ,حيث يمكن ان يستخدم مع التصميمين الثابت والعشوائي, وان هذا الممهد لا يحتاج الى معالجة للتحيز قرب حدود الفترة (Boundary Bias) لانه يتكيف بصورة مباشرة عكس ممهد Nadaraya-Watson حيث ان معدل التقارب لا يتأثر بمواقع نقاط البيانات.يعد هذا الممهد نسخة لا معلمية من طريقة المربعات الصغرى الموزونة اذ يمكن التعبير عنها على شكل تصغير للصيغة الاتية [22][13][11].

$$\sum_{i=1}^{n}\{y_i-\beta_0-\beta_1(x-x_i)\}^2 K_h\,(x-x_i) \qquad \ldots(2.19)$$

ويمكن التعبير عنها على شكل مصفوفات كما ياتي :

$$(Y-X\,\beta)'\;W\;(Y-X\,\beta) \qquad \ldots(2.20)$$

اذ ان $Y=(y_1, y_2, y_3, \ldots\ldots y_n)$ هو متجه قيم المتغير المعتمد من الدرجة n*1 ,وإن مصفوفة المتغيرات المستقلة تكتب بالصيغة $X=(1 \quad (x-x_i))$ من درجة n*2 , وان مصفوفة الاوزان تكتب بالصيغة الاتية

$W=\mathrm{diag}\,(K_h\,(x-x_1), \ldots\ldots, K_h\,(x-x_n))$ وتكون من درجة n*n. وباستخدام طريقة المربعات الصغرى الموزونة فان مقدر المعلمات الذي يجعل المعادلة السابقة اصغر ما يمكن هو كما ياتي:

$$\hat{\beta}=(X'WX)^{-1}X'WY \qquad \ldots(2.21)$$

اذ ان

$$X'WX = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ (x-x_1) & (x-x_2) & \cdots & (x-x_n) \end{bmatrix} \begin{bmatrix} K_h\,(x-x_1) & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & K_h\,(x-x_n) \end{bmatrix} \begin{bmatrix} 1 & (x-x_1) \\ \vdots & \vdots \\ 1 & (x-x_n) \end{bmatrix}$$

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

$$X'WY = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ (x-x_1) & (x-x_2) & \cdots & (x-x_n) \end{bmatrix} \begin{bmatrix} K_h\,(x-x_1) & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & K_h\,(x-x_n) \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{n} y_i K_h\,(x-x_i) \\ \sum_{i=1}^{n} y_i(x-x_i)K_h\,(x-x_i) \end{bmatrix}$$

ومعكوس المصفوفة $X'WX$ يمكن كتابتها كما ياتي:

$$(X'WX)^{-1} = \frac{1}{\det(X'WX)} \begin{bmatrix} \sum_{i=1}^{n}(x-x_i)^2 K_h\,(x-x_i) & -\sum_{i=1}^{n}(x-x_i)K_h\,(x-x_i) \\ -\sum_{i=1}^{n}(x-x_i)K_h\,(x-x_i) & \sum_{i=1}^{n} K_h\,(x-x_i) \end{bmatrix}$$

اذ ان $\det(X'WX)$ تمثل المحددة وتعرف كما ياتي :

$$\sum_{i=1}^{n} K_h\,(x-x_i) \sum_{i=1}^{n}(x-x_i)^2 K_h\,(x-x_i) - \{\sum_{i=1}^{n}(x-x_i)K_h\,(x-x_i)\}^2 = \det(X'WX)$$

$$= S_0 S_2 - S_1^2$$

اذ ان : $S_\alpha = \sum_{i=1}^{n}(x-x_i)^\alpha K_h\,(x-x_i)$ $\alpha=0,1,2$

وبناءً على ما سبق يمكن اعادة كتابة المعادلة (2.21) كما ياتي :

$$\hat{\beta} = \frac{1}{S_0 S_2 - S_1^2}\begin{pmatrix} S_2 & -S_1 \\ -S_1 & S_0 \end{pmatrix}\begin{pmatrix} \sum_{i=1}^{n} y_i K_h\,(x-x_i) \\ \sum_{i=1}^{n} y_i (x-x_i) K_h\,(x-x_i) \end{pmatrix} \qquad \ldots(2.22)$$

$$= \frac{1}{S_0 S_2 - S_1^2}\begin{pmatrix} S_2 \sum_{i=1}^{n} y_i K_h\,(x-x_i) - S_1 \sum_{i=1}^{n} y_i (x-x_i) K_h\,(x-x_i) \\ S_0 \sum_{i=1}^{n} y_i (x-x_i) K_h\,(x-x_i) - S_1 \sum_{i=1}^{n} y_i K_h\,(x-x_i) \end{pmatrix}$$

ويكون مقدر β_0 والذي يمثل مقدر الانحدار الخطي الموضعي للدالة $\widehat{m}(x)$ معرفاً كما في الصيغة الاتية:

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

$$= \frac{\sum_{i=1}^{n}\{S_2 - (x-x_i)S_1\} K_h\,(x-x_i)\, y_i}{S_0 S_2 - S_1^2} \qquad \ldots(2.23)$$

اما التحيز Bias والتباين Variance لممهد (L-L-R) في حالة التصميم العشوائي Random يكون كما ياتي :[34]

$$\text{Bias} = \frac{h^2}{2} m''(x) d_k \qquad \ldots(2.24)$$

$$\text{Variance} = \frac{\sigma^2(x)}{nh\, f(x)} C_k \qquad \ldots(2.25)$$

اما الخصائص المحاذية لهذا المقدر وبافتراض تحقق الشروط الاتية :[3][13][11]

1- ان دالة الانحدار m(.) تمتلك مشتقة ثانية .

2- ان دالة الكثافة الاحتمالية f(.) ل x تحقق شرط Holder

$|f(x)-f(y)| \le c\,|x-y|^\alpha$, $0<\alpha<1$, $f(x)>0$ اذ ان c يمثل ثابتا موجبا .

3- التباين الشرطي $\sigma^2(x) = var\,(y/X=x)$ يكون محددا ومستمرا.

4- ان دالة kernel تكون دالة كثافة مستمرة ومحددة .

الخاصية 1 :

اذا كانت $h=dn^{-\theta}$, $0<\theta<1$ فان MSE غير الشرطي لهذا الممهد في حالة التصميم العشوائي يكون كما ياتي :

$$E([\widehat{m}(x)-m(x)]^2=\frac{h^4}{4}(m''(x)d_k)^2+\frac{1}{nh}f^{-1}(x)\sigma^2(x)C_k+o\left(h^4+\frac{1}{nh}\right) \qquad \dots(2.26)$$

الخاصية 2 :

ان MISE غير الشرطي في حالة التصميم العشوائي يكون كما ياتي :

$$E\int_a^b[\widehat{m}(x)-m(x)]^2W(x)\,dx$$
$$=\frac{h^4}{4}d_k^2\int_a^b(m''(x))^2W(x)\,dx+\frac{C_k}{n\,h}\int_a^b\frac{\sigma^2(x)}{f(x)}\,W(x)dx+o(h^4+\frac{1}{n\,h}) \qquad \dots(2.27)$$

8- ممهد نادرايا واتسون بالاعتماد على معلومات المركز

Nadaraya-Watson based on Center information: (N-W: C)

الفكرة لهذه الطريقة هي استخدام معلومات المركز للفترات لبناء علاقة لا خطية بين المتغير المعتمد Y و المتغير المستقل X, حيث تم تمثيل المتغير المعتمد Y بواسطة زوج من المتغيرات الكمية $(Y_L\ ,Y_U)$عندما تكون وصف للحد الادنى والحد الاعلى لهذا المتغير على التوالي . وان المتغير المستقل تم تمثيله بواسطة متغير كمي X^c الذي يصف مركز هذا المتغير هذا يعني X يتمثل بـ(X^c) .

في طريقة الانحدار هذه , لاستكشاف Y بواسطة X هذا يعادل استكشاف Y_L بواسطة X^c و Y_U بواسطة X^c بشكل منفصل [9,14,19].

حيث العلاقة اللاخطية بين Y و X تعطى كما ياتي :

$$E(Y/X)=[E((Y_L/X^c),E((Y_U/X^c)] \qquad \dots(2.28)$$
$$=[m_L^c(X^c),m_U^c(X^c)]$$

عندما تكون m_L^c, m_U^c هي دوال غير معروفة وان

$$x^c=\frac{x_L+x_U}{2}$$

بالاعتماد على الممهد Nadaraya-Watson, يتم تقدير دالة الوزن للمتغير المعتمد و كما ياتي :

$$\widehat{m}_L^c(x^c)=\sum_{i=1}^{n}w_i^c\,Y_{Li} \qquad \dots(2.29)$$
$$\widehat{m}_U^c(x^c)=\sum_{i=1}^{n}w_i^c\,Y_{Ui} \qquad \dots(2.30)$$

عندما يعرف الوزن باستخدام دالة kernel كما ياتي :

$$w_i^c = \frac{k_h(d(x^c, x_i^c))}{\sum_{i=1}^{n} k_h(d(x^c, x_i^c))} \qquad \ldots(2.31)$$

ان دالة Gaussian kernel لمركز الفترة تعطى كما ياتي :

$$k_h\left(d(x^c, x_i^c)\right) = \frac{1}{\sqrt{2\pi}}\frac{1}{h}e^{-\frac{d(x^c,x_i^c)}{2h^2}} \qquad \ldots(2.32)$$

عندما تكون $d(x^c, x_i^c)$ هي مربع المسافة الاقليدية بين x^c و x_i^c والتي تعطى كما ياتي :

$$d(x^c, x_i^c) = \sum_{i=1}^{n}(x^c - x_i^c)^2 \qquad \ldots(2.34)$$

9- ممهد نادرايا واتسون بالاعتماد على معلومات المركز والمدى

Nadaraya-Watson Based on Center and Range information:(N-W:CR)

الفكرة لهذه الطريقة هي استخدام معلومات المركز و المدى للفترات لبناء علاقة لا خطية بين المتغير المعتمد Y و المتغير المستقل X . حيث تم تمثيل المتغير المعتمد بواسطة زوج من المتغيرات الكمية (Y^c, Y^r)عندما تكون وصف لمركز ومدى الفترات, على التوالي . كذلك تم تمثيل المتغير المستقل X بواسطة زوج من المتغيرات الكمية (X^c, X^r)عندما تكون وصف لمركز ومدى الفترات .[32]

في طريقة الانحدار هذه , لاستكشاف Y بواسطة X هذا يعادل استكشاف Y^c بواسطة X^c و Y^r بواسطة X^r بشكل منفصل.

ان العلاقة اللاخطية بين Y و X تعطى كما ياتي :

$$\mathrm{E}(Y/X)=[\mathrm{E}(Y^c/X^c) - 0.5\,E(Y^r/X^r), \mathrm{E}(Y^c/X^c) + 0.5\,E(Y^r/X^r)]$$

$$=[m^c(X^c) - 0.5m^r(X^r), m^c(X^c) + 0.5m^r(X^r)]$$

$$where\ \ x^c = \frac{x_L + x_U}{2} \qquad و \qquad y^c = \frac{y_L + y_U}{2}$$

$$\text{And} \quad x^r = x_U - x_L \qquad و \qquad y^r = y_U - y_L$$

بالاعتماد على الممهد Nadaraya-Watson , يتم تقدير دوال الوزن للمتغير المعتمد وكما ياتي :

$$\widehat{m}^c(X^c) = \sum_{i=1}^{n} w_i^c\ y_i^c \qquad \ldots(2.35)$$

$$\widehat{m}^r(X^r) = \sum_{i=1}^{n} w_i^r\ y_i^r \qquad \ldots(2.36)$$

عندما يعرف الوزن w_i^c كمافي المعادلة (2.31) السابقة وان w_i^r يعرف باستخدام دالة kernel كما ياتي :

$$w_i^r = \frac{k_h(d(x^r, x_i^r))}{\sum_{i=1}^{n} k_h(d(x^r, x_i^r))} \qquad \ldots(2.37)$$

ان دالة Gaussian kernel لمدى الفترة تعطى كما ياتي :

$$k_h\left(d(x^r, x_i^r)\right) = \frac{1}{\sqrt{2\pi}}\frac{1}{h}e^{-\frac{d(x^r,x_i^r)}{2h^2}} \qquad \ldots(2.38)$$

عندما تكون $d(x^r, x_i^r)$ هي مربع المسافة الاقليدية بين x^r و x_i^r والتي تعطى كما ياتي :

$d(x^r, x_i^r) = \sum_{i=1}^{n}(x^r - x_i^r)^2$ ...(2.39)

10-ممهد الانحدار الخطي الموضعي بالاعتماد على معلومات المركز والمدى

Local Linear Regression Based on Center and Range information: (L-L-R:CR)

الفكرة لهذه الطريقة هي استخدام معلومات المركز والمدى للفترات لبناء علاقة لا خطية بين المتغير المعتمد Y و المتغير المستقل X . حيث تم تمثيل المتغير المعتمد بواسطة زوج من المتغيرات الكمية (Y^c, Y^r) عندما تكون وصف لمركز ومدى الفترات, على التوالي . كذلك تم تمثيل المتغير المستقل X بواسطة زوج من المتغيرات الكمية (X^c, X^r) عندما تكون وصف لمركز ومدى الفترات .[28,30]

في طريقة الانحدار هذه , لاستكشاف Y بواسطة X هذا يعادل استكشاف Y^c بواسطة X^c و Y^r بواسطة X^r بشكل منفصل.

ان العلاقة اللاخطية بين Y و X تعطى كما ياتي :

$$E(Y/X)=[E(Y^c/X^c) - 0.5\, E(Y^r/X^r), E(Y^c/X^c) + 0.5\, E(Y^r/X^r)]$$

$$=[m^c(X^c) - 0.5m^r(X^r), m^c(X^c) + 0.5m^r(X^r)]$$

بالاعتماد على ممهد الانحدار الخطي الموضعي (L-L-R) يمكننا الحصول على تقدير دوال الانحدار ($\hat{m}^c, \hat{m}^r$) كما في المعادلات التالية :[24]

$$\hat{m}^c(x) = n^{-1}\frac{\sum_{i=1}^{n}\{S_2 - (x^c - x^c{}_i)S_1\}k_h(d(x^c, x^c{}_i))\, y^c{}_i}{S_0S_2 - S_1^2} \qquad ...(2.40)$$

$$S_\alpha = n^{-1}\sum_{i=1}^{n}(x^c - x^c{}_i)^\alpha\, k_h(d(x^c, x^c{}_i))$$

$$\hat{m}^r(x) = n^{-1}\frac{\sum_{i=1}^{n}\{S_2 - (x^r - x^r{}_i)S_1\}k_h(d(x^r, x^r{}_i))\, y^r{}_i}{S_0S_2 - S_1^2} \qquad ...(2.41)$$

$$S_\alpha = n^{-1}\sum_{i=1}^{n}(x^r - x^r{}_i)^\alpha\, k_h(d(x^r, x^r{}_i)) \quad ,\alpha=0,1$$

11-المقارنة بين الطرائق

تم استخدام المعيار الاحصائي متوسط مربعات الخطا للحد الادنى والحد الاعلى على التوالي MSE_L, MSE_U,وكذلك تم استخدام معيار متوسط حجم الخطا النسبي (MMRE), ثم تم اخذ المتوسط (Mean) والانحراف المعياري (Standard deviation) (Std) لهذين المعيارين, لغرض المقارنة بين الطرائق واختيار الطريقة الافضل وحسب الصيغ الاتية : [32][54]

$$MSE_L = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_{Li} - y_{Li})^2 \qquad ....(3-3)$$

$$MSE_U = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_{Ui} - y_{Ui})^2 \qquad ....(3-4)$$

$$MMRE=\sum_{i=1}^{n}\frac{1}{2n}\left\{\left|\frac{y_{Li}-\hat{y}_{Li}}{y_{Li}}\right| + \left|\frac{y_{Ui}-\hat{y}_{Ui}}{y_{Ui}}\right|\right\} \qquad ....(3-5)$$

12- النماذج المستخدمة في تجارب المحاكاة

ان النماذج المقترحة في هذه الدراسة فهي انموذجين تم اقتراحهما لمحاكاة واقع معين ومن المفترض انها كانت اكثر تمثيلاً للمجتمع المراد التعرف على خصائصه وكما ياتي:

1- الانموذج الاسي للحد الادنى والحد الاعلى وصيغته :[20]

$y_L = e^{x_L}$
$y_U = e^{x_U}$ ....(3 − 1)

2- الانموذج الجيبي للحد الادنى والحد الاعلى وصيغته [32]:

$y_L = sin(x_L * \pi) + 1$
$y_U = sin(x_U * \pi) + 1$ ....(3 − 2)

13-توليد المتغيرات العشوائية

تم تنفيذ تجارب المحاكاة باستخدام ثلاث حجوم عينات (n=100,200,300) وبتكرار مقداره 1000 مرة لكل تجربة وكما ياتي :

-1 لمتغيرات المستقلة : حيث تم توليد المتغيرات المستقلة للحد الادنى والحد الأعلى بأستخدام التوزيع المنتظم اي ان ا
x_L~unif (0.1,0.3)
x_U~unif (0.4,0.6)

-2 الاخطاء العشوائية : تتوزع توزيعا طبيعيا بمتوسط صفر وتباين σ^2 اي ان $j = 1,2,3$ $e \sim N(0, \sigma^2{}_j)$
$\sigma_j = 0.3 , 0.6 , 0.9$ اذ ان σ يمثل الانحراف المعياري للخطا e

-3 المتغير المعتمد : ويتم توليده من خلال النماذج المستخدمة في تجارب المحاكاة وذلك باستخدام المتغيرات المستقلة التي تم توليدها في الفقرة 1,مضافاً اليها الاخطاء العشوائية المولدة في الفقرة 2,وللانموذجين المدروسة .

14- مرحلة التقدير :

هذه المرحلة تتم فيها عملية التقدير لدالة الانحدار اللامعلمي باستخدام الطرائق الموجودة في الجانب النظري.

15- نتائج المحاكاة :

لغرض تطبيق الطرائق لتقدير دالة الانحداراللامعلمي وتحديد الطريقة الافضل , اذ سيتم عرض وتحليل نتائج تجارب المحاكاة التي تمثل قيم المتوسط والانحراف المعياري للمعيارين MSE_L , MSE_U و MMRE لكل طريقة وفقاً لجميع حجوم العينات وقيم الانحراف المعياري للخطا وكما يأتي :

جدول رقم (2) يشير الى المتوسط و(الانحراف المعياري) لمعيار MSE_L , MSE_U للطرائق مصنف حسب حجم العينة والانحراف المعياري للخطا بالاعتماد على الانموذج الاول

الطريقة	حجم العينة	σ =0.3		σ =0.6		σ =0.9	
		MSE_L	MSE_U	MSE_L	MSE_U	MSE_L	MSE_U
N-W: C	100	0.0156 (0.0075)	0.0277 (0.0103)	0.0183 (0.0153)	0.0304 (0.0203)	0.0228 (0.0243)	0.0349 (0.0313)
	200	0.0079 (0.0038)	0.0140 (0.0052)	0.0092 (0.0077)	0.0154 (0.0103)	0.0115 (0.0123)	0.0176 (0.0159)
	300	0.0053 (0.0025)	0.0094 (0.0035)	0.0062 (0.0052)	0.0103 (0.0069)	0.0077 (0.0082)	0.0118 (0.0106)
N-W :CR	100	0.0156 (0.0075)	0.0277 (0.0103)	0.0183 (0.0153)	0.0304 (0.0203)	0.0228 (0.0243)	0.0349 (0.0313)

	200	0.0079 (0.0038)	0.0140 (0.0052)	0.0092 (0.0077)	0.0154 (0.0103)	0.0115 (0.0123)	0.0176 (0.0159)
	300	0.0053 (0.0025)	0.0094 (0.0035)	0.0062 (0.0052)	0.0103 (0.0069)	0.0077 (0.0082)	0.0118 (0.0106)
L-L-R:CR	100	0.0160 (0.0077)	0.0283 (0.0105)	0.0187 (0.0156)	0.0310 (0.0207)	0.0232 (0.0248)	0.0356 (0.0319)
	200	0.0080 (0.0038)	0.0142 (0.0052)	0.0093 (0.0078)	0.0155 (0.0104)	0.0116 (0.0124)	0.0178 (0.0160)
	300	0.0053 (0.0026)	0.0094 (0.0035)	0.0062 (0.0052)	0.0103 (0.0069)	0.0077 (0.0083)	0.0118 (0.0107)

ومن الجدول (2) أظهرت النتائج كالآتي :

1-عند جميع حجوم العينات و قيم الانحراف المعياري للخطا كانت طريقة (N-W: C) هي الافضل لامتلاكها أقل قيمة لمعيار MSE .

2- قيم MSE تتناقص عند تزايد حجوم العينات ولجميع الطرائق وقيم الانحراف المعياري للخطا المستعملة, مما يشير الى تاثر الطرائق بقلة حجم العينة وازدياد قدرة ودقة تلك الطرائق عند تزايد حجم العينة.

3- قيم MSE تتزايد عند تزايد قيم الانحراف المعياري للخطا ولجميع الطرائق وحجوم العينات المستعملة ,مما يشير الى تاثر تلك الطرائق بتشتت البيانات وقلة قدرتها عند تزايد تشتت البيانات.

4- عند حجمي العينة (200,100) ولجميع قيم الانحراف المعياري للخطا كانت طريقة (L-L-R:CR)هي الاسوأ لامتلاكها أكبر قيمة لمعيار MSE.اما عند حجم العينة (300) كانت النتائج متشابهة بين جميع الطرائق.

جدول رقم (3) يشير الى المتوسط و(الانحراف المعياري) لمعيار(MMRE) للطرائق مصنف حسب حجم العينة والانحراف المعياري للخطا بالاعتماد على الانموذج الاول

الطريقة	حجم العينة	σ =0.3	σ =0.6	σ =0.9
		MMRE	MMRE	MMRE
N-W: C	100	0.0099 (0.0000)	0.0099 (0.0000)	0.0099 (0.0000)
	200	0.0050 (0.0000)	0.0050 (0.0000)	0.0050 (0.0000)
	300	0.0033 (0.0000)	0.0033 (0.0000)	0.0033 (0.0000)
N-W :CR	100	0.0099 (0.0000)	0.0099 (0.0000)	0.0099 (0.0000)
	200	0.0050 (0.0000)	0.0050 (0.0000)	0.0050 (0.0000)
	300	0.0033 (0.0000)	0.0033 (0.0000)	0.0033 (0.0000)

L-L-R:CR	100	0.0100 (0.0000)	0.0100 (0.0000)	0.0100 (0.0000)
	200	0.0050 (0.0000)	0.0050 (0.0000)	0.0050 (0.0000)
	300	0.0033 (0.0000)	0.0033 (0.0000)	0.0033 (0.0000)

ومن الجدول (3) أظهرت النتائج كالآتي:

1عند جميع حجوم العينات و($\sigma = 0.3,0.6$) كانت طريقة (N-W: C) هي الافضل لأمتلاكها اقل قيمة لمعيار MMRE . اما عند ($\sigma = 0.9$) وحجم العينة (100) فان الطريقتين (N-W:C) و (N-W :CR) كانتا الافضل ,وعند حجمي العينة (200,300) كانت النتائج متشابهة بين جميع الطرائق.

2- قيم MMRE تتناقص عند تزايد حجوم العينات ولجميع الطرائق وقيم الانحراف المعياري للخطا المستعملة, مما يشير الى تاثر الطرائق بقلة حجم العينة وازدياد قدرة ودقة تلك الطرائق عند تزايد حجم العينة.

3- ان قيم MMRE للطرائق (L-L-R:CR) و (N-W: C) و(N-W:CR) كانت ثابتة عند تزايد قيم الانحراف المعياري للخطا و لجميع حجوم العينات المستعملة .

جدول رقم (4) يشير الى المتوسط و(الانحراف المعياري) لمعيار MSE_L , MSE_U للطرائق مصنف حسب حجم العينة والانحراف المعياري للخطا بالاعتماد على الانموذج الثاني

الطريقة	حجم العينة	σ =0.3		σ =0.6		σ =0.9	
		MSE_L	MSE_U	MSE_L	MSE_U	MSE_L	MSE_U
N-W: C	100	0.0256 (0.0104)	0.0395 (0.0117)	0.0282 (0.0198)	0.0422 (0.0238)	0.0327 (0.0304)	0.0467 (0.0366)
	200	0.0129 (0.0052)	0.0199 (0.0059)	0.0142 (0.0100)	0.0213 (0.0120)	0.0165 (0.0154)	0.0236 (0.0186)
	300	0.0086 (0.0035)	0.0133 (0.0040)	0.0095 (0.0067)	0.0142 (0.0081)	0.0110 (0.0103)	0.0157 (0.0124)
N-W: CR	100	0.0256 (0.0104)	0.0395 (0.0117)	0.0282 (0.0198)	0.0422 (0.0238)	0.0327 (0.0304)	0.0467 (0.0366)
	200	0.0129 (0.0052)	0.0199 (0.0059)	0.0142 (0.0100)	0.0213 (0.0120)	0.0165 (0.0154)	0.0236 (0.0186)
	300	0.0086 (0.0035)	0.0133 (0.0040)	0.0095 (0.0067)	0.0142 (0.0081)	0.0110 (0.0103)	0.0157 (0.0124)

L-L-R:CR	100	0.0261 (0.0106)	0.0403 (0.0120)	0.0288 (0.0202)	0.0430 (0.0243)	0.0334 (0.0310)	0.0476 (0.0374)
	200	0.0130 (0.0053)	0.0201 (0.0060)	0.0144 (0.0101)	0.0215 (0.0122)	0.0167 (0.0155)	0.0238 (0.0187)
	300	0.0087 (0.0035)	0.0134 (0.0040)	0.0096 (0.0067)	0.0143 (0.0081)	0.0111 (0.0103)	0.0158 (0.0125)

ومن الجدول (4) أظهرت النتائج كالآتي:

1-عند جميع حجوم العينات ولجميع قيم الانحراف المعياري للخطا كانت طريقة (N-W: C) هي الافضل لامتلاكها أقل قيمة لمعيار MSE .

2- قيم MSE تتناقص عند تزايد حجوم العينات ولجميع الطرائق وقيم الانحراف المعياري للخطا المستعملة, مما يشير الى تاثر الطرائق بقلة حجم العينة وازدياد قدرة ودقة تلك الطرائق عند تزايد حجم العينة.

3- قيم MSE تتزايد عند تزايد قيم الانحراف المعياري للخطا ولجميع الطرائق وحجوم العينات المستعملة ,مما يشير الى تاثر تلك الطرائق بتشتت البيانات وقلة قدرتها عند تزايد تشتت البيانات.

4-عند جميع حجوم العينات ولجميع قيم الانحراف المعياري للخطا كانت طريقة (L-L-R:CR) هي الاسوأ لامتلاكها أكبر قيمة لمعيار MSE .

جدول رقم (5) يشير الى المتوسط و(الانحراف المعياري) لمعيار (MMRE) للطرائق مصنف حسب حجم العينة والانحراف المعياري للخطا بالاعتماد على الانموذج الثاني

الطريقة	حجم العينة	σ =0.3	σ =0.6	σ =0.9
		MMRE	MMRE	MMRE
N-W :C	100	0.0099 (0.0000)	0.0099 (0.0000)	0.0099 (0.0000)
	200	0.0050 (0.0000)	0.0050 (0.0000)	0.0050 (0.0000)
	300	0.0033 (0.0000)	0.0033 (0.0000)	0.0033 (0.0000)
N-W :CR	100	0.0099 (0.0000)	0.0099 (0.0000)	0.0099 (0.0000)
	200	0.0050 (0.0000)	0.0050 (0.0000)	0.0050 (0.0000)
	300	0.0033 (0.0000)	0.0033 (0.0000)	0.0033 (0.0000)

L-L-R:CR	100	0.0100 (0.0000)	0.0100 (0.0000)	0.0100 (0.0000)
	200	0.0050 (0.0000)	0.0050 (0.0000)	0.0050 (0.0000)
	300	0.0033 (0.0000)	0.0033 (0.0000)	0.0033 (0.0000)

ومن الجدول (5) أظهرت النتائج كالآتي:

1-عند جميع حجوم العينات ولجميع قيم الانحراف المعياري للخطا كانت طريقة (N-W: C) هي الافضل لامتلاكها أقل قيمة لمعيار MMRE .

2- ان قيم MMRE للطرائق (L-L-R:CR) و (N-W: C) و(N-W:CR) كانت ثابتة عند تزايد قيم الانحراف المعياري للخطا و لجميع حجوم العينات .

3- قيم MMRE لجميع الطرائق تتناقص عند تزايد حجوم العينات ماعدا عند طريقة (CK+RL) حيث تزايدت قيم MMRE عند ($\sigma = 0.9$) عند تزايد حجم العينة من (100) الى (200).

4 – قيم MMRE لجميع الطرائق تناقصت عند تزايد قيم الانحراف المعياري للخطا من (0.3) الى (0.6),في حين ان هذه القيم ازدادت عند تزايد قيم الانحراف المعياري من (0.6) الى (0.9) عند حجم العينة (100).اما عند حجمي العينة (300,200) تزايدت هذه القيم عند تزايد قيم الانحراف المعياري للخطا .

16- الاستنتاجات

1- بصورة عامة فان طريقة (N-W:C) هي افضل طرائق التقدير باختلاف حجم العينة ,وقيم الانحراف المعياري تلاها بعد ذلك بعد طريقة (N-W:CR) .

2- يصورة عامة ان قيم معياري المقارنة MSE و MMRE تتناقص عند تزايد حجوم العينات ولجميع الطرائق وقيم الانحراف المعياري وللانموذجين المستعملة .

3- بصورة عامة ان قيم MSE تتزايد عند تزايد قيم الانحراف المعياري ولجميع الطرائق ,اما قيم MMRE فكانت ثابتة ولجميع الطرائق,ولجميع حجوم العينات وللنموذجين المستعملة .

4- اظهرت النتائج ان افضل طريقة كانت هي (N-W: C) لامتلاكها اقل قيم لمعياري المقارنة MSE و MMRE .

17- التوصيات

1- استعمال طريقة (N-W: C) لتقدير دالة الانحدار اللامعلمي في حالة البيانات على شكل فترات والعلاقة لاخطية .

2- التثقيف على الاهتمام بصحة القلب من خلال التمارين الرياضية, وتقليل الوزن ,وتجنب كل ما يقلل نسب الأكسجين .

3- توفيرالدعم اللازم للمشاريع البحثية التي تستعمل الممهدات اللبية في عمليات التقدير .

4- نوصي بتوسيع نطاق تطبيق طريقة (N-W: C) في مجالات اخرى مثل العلوم الاقتصادية والاجتماعية وغيرها وذلك للاستفادة من امكانياتها التقديرية .

18-المصـــــادر

1. حمزة,سعد كاظم ,(2009), "مقارنة بعض الطرائق اللبية في تقدير نماذج الانحدار اللامعلمية بوجود بيانات تامة وغير تامة " رسالة ماجستير في الاحصاء,كلية الادارة والاقتصاد,جامعة بغداد.

2. علي,عمر عبد المحسن (2007),"مقارنة مقدرات النماذج التجميعية االمعممة باستخدام الشرائح التمهيدية عند تحليل الانحدار اللامعلمي وشبه المعلمي " اطروحة دكتوراه في الاحصاء, كلية الادارة والاقتصاد ,جامعة بغداد.
3. محمد ,سعد صبر والخطيب ,حوراء ضياء جبوري,(2024) , "مقارنة بين طريقتي الانحدار الخطي الموضعي والجوار الاقرب في تقدير أنموذج الانحدار اللامعلمي دراسة محاكاة مع التطبيق ".مجلة الكوت للعلوم الاقتصادية والادارية,(50) pp.111-133,16.
4. Ahn, J., Peng, M., Park, C., and Jeon, Y. (2012). A resampling approach for interval-valued data regression. Statistical Analysis and Data Mining, 5:336–348..
5. Billard, L. and Diday, E. (2007). Symbolic Data Analysis: Conceptual Statistics and Data Mining. Wiley, Chichester.
6. Billard, L. and Diday, E.(2000), Regression Analysis for Interval-Valued Data. In: Data Analysis, Classification and Related Methods: Proceedings of the Seventh Conference of the International Federation of Classification Societies (IFCS'00), Springer-Verlag, Belgium, 369-374.
7. Birke, M., & Dette, H. (2007). Estimating a convex function in nonparametric regression. *Scandinavian Journal of Statistics*, 34(2), 384-404.
8. Chacón, J.E. and O. Rodríguez,(2021), Regression Models for Symbolic Interval-Valued Variables. Entropy. 23(4): p. 429.
9. Chuang, CC., Jeng, JT., Lin, WY. Et al,(2020). Interval Fuzzy c-Regression Models with Competitive Agglomeration for Symbolic Interval-Valued Data. Int. J. Fuzzy Syst. 22, 891–900 .
10. Cleveland ,W.S. "Robust locally weighted regression and smoothing scatter plots ",(1979),JASA,Vol.74,No.368,pp 829-836.
11. D. Ruppert,(1996), "Local polynomial regression and its applications in environmental statistics," Cornell University Operations Research and Industrial Engineering.
12. Diday, E. (1987). Introduction `a l'Approache Symbolique en Analyse des Donn´ees. Premi`eres Journ´ees Symbolique – Num´erique. CEREMADE, Universit´e Paris, 21-56.
13. 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.
14. Fagundes RAA, De Souza RMCR, and Cysneiros FJA (2014). Interval kernel regression, Neurocomputing, 128, 371–388.
15. Fan .J and Gijbels.I.(1996)"Astudy of veriable bandwidth selection for local polynomial regression "statistca sinica -pp 113-127.
16. Fan, J. (1993). Local linear regression smoothers and their minimax efficiencies. The annals of Statistics, 196-216.
17. G. S. Watson,(1964), "Smooth regression analysis," Sankhyā Indian J. Stat. Ser. A, pp. 359–372.
18. Gasser,T.,Muller,H.G. and Mammitzsch,R.(1985)"Kernels for Nonparametric curve estimation "J.Royal S.S.,Series.B,Vol.47,No.2,pp 238-252.
19. Gasser,T.and Muller , H.G.(1984),"Estimation regression functions and their derivatives by kernel method ",Scandinavian journal of Stat.,Vol.11,pp 171-1859.

20. Guan, L., Li, M.,(2024), Interval-valued linear regression model with an asymmetric Laplace distribution. J. Korean Stat. Soc.
21. Hardle, W.,(1994), Applied Nonparametric Regression ,Institute fur Statistic und Okonometrie, Berlin.
22. Härdle, W., Müller, M., Sperlich, S., & Werwatz, A. (2004). Nonparametric and semiparametric models (Vol. 1). Berlin: Springer..
23. Jang, J. and K.-H. Kang,(2020), Local linear regression analysis for interval-valued data. Communications for Statistical Applications and Methods. 27(3): p. 365-376.
24. Lima Neto, E., , Cordeiro, G., and de Carvalho, F. (2011). Bivariate symbolic regression models for interval-valued variables. Journal of Statistical Computation and Simulation, 81:1727–1744.
25. Lima Neto, E.d.A.; de Carvalho, F.d.A.T.(2018), An exponential-type kernel robust regression model for interval-valued variables. In Information Sciences; Elsevier: Amsterdam, The Netherlands; Volume 454, pp. 419–442
26. Muller, H.G,(1987), "Weighted local regression and Kernel methods for Nonparametric curve fitting ",JASA,VOI.82,No.397,pp231-238..
27. Phadkantha, R., Yamaka, W., & Sriboonchitta, S. (2022, January). A Bayesian approach to quantile regression for interval-valued data: application to CAPM. In *International Conference of the Thailand Econometrics Society* (pp. 313-324). Cham: Springer International Publishing.
28. R.A.A. Fagundes, R.M.C.R. Souza, and F.J.A. Cysneiros.(2013), Robust Regression with Application to Symbolic Interval Data, vol. 26,564-573.
29. Wang, H., Guan, R. & Wu, J.,(2012), Linear regression of interval-valued data based on complete information in hypercubes. J. Syst. Sci. Syst. Eng. 21, 422–442 .
30. Xu, M., & Qin, Z. (2022). A bivariate Bayesian method for interval-valued regression models. *Knowledge-Based Systems*, *235*, 107396.
31. Xu, W. (2010). Symbolic Data Analysis: Interval-Valued Data Regression. PhD thesis, University of Georgia .
32. Yongho Jeon, Jeongyoun Ahn & Cheolwoo Park (2014): A Nonparametric Kernel Approach to Interval-valued Data Analysis.
33. Yuan Wei, Shanshan Wang & Huiwen Wang (2017): Interval-valued data regression using partial linear model, Journal of Statistical ,Computation and Simulation .
34. Zhao, Q., Wang, H. & Wang, S.(2023), Robust regression for interval-valued data based on midpoints and log-ranges. Adv Data Anal Classif 17, 583–621.
35. Zhong, Y., & Li, S. (2019). Interval-Valued Nonparametric Regression With Generalized Fourier Series Estimators. International Journal of Intelligent Technologies and Applied Statistics, 12(1), 1-15.
36. Zhong, Y., Zhang, Z. & Li, S.,(2020), A Constrained Interval-Valued Linear Regression Model: A New Heteroscedasticity Estimation Method. J Syst Sci Comple