

## مقارنة بعض طرائق التقدير الحصينة مع طرائق الانحدار العميق

## Comparison of some robust estimation methods with deep regression methods

أ.م. د أيمن محمد عبد الله

Iman mohammed abdullah<sup>(2,b)</sup>

كلية الإدارة والاقتصاد بغداد, بغداد, العراق

Iman MO@uobaghdad.edu.iq

نور حميد زويد

Noor hammed zwaïd<sup>(1,a)</sup>

كلية الإدارة والاقتصاد بغداد, بغداد, العراق

zwaïd2201m@coadec.uobaghdad.edu.iq

**المستخلص:** أن عملية الحصول على مقدرات قريبة جدا من معالمها هو هدف أي أسلوب تحليلي ليمثل المجتمع الذي سحبت منه العينة تمثيلا جيدا. فعندما يكون توزيع ظاهرة ما توزيعا طبيعيا أي أنه يخضع للافتراضات الموضوعية فان الطرائق الاعتيادية مثل طريقة المربعات الصغرى وطريقة الإمكان الأعظم تعدّ طرائق كفؤة في الحصول على مقدرات جيدة ولكن ما يعاب على الوسط الحسابي هو أنه يتأثر بنحو كبير في حالة وجود الشواذ (Outliers) بين البيانات المأخوذة وفي مثل هذه الحالة ينبغي الاعتماد على توسيط أحر أو إجراء عملية تهذيب للبيانات من خلال تحديد المشاهدات الشاذة والتعويض عنها بقيم بديلة وتعد وظيفة طرائق العمق الاحصائي في الانحدار من الطرائق الحديثة التي يمكن استخدامها لتطوير الطرائق الحصينة وخصوصا في حالات عدم تغيير قيم المشاهدات الشاذة بناء على تقديرات حصينة لمصفوفة الموقع والتشتت وذلك لتوفير دوال تقدير قوية في التقدير واعمق لموقع البيانات والتي تتوفر بها خصائص المتانة المرغوب بها مثل التحيز ودالة التأثير ونقطة الانهيار العالية التي تعمل في البيانات احادية المشكلة وتكون غير حساسة اتجاه وجود القيم الشاذة وتعطينا مقدرات كفؤة وفي هذا البحث تم الاعتماد على طريقة M الحصينة و طريقة المربعات الصغرى الموزونة المكررة (IRLS) , لغرض مقارنة كفاءتها مع طرائق عمق الانحدار وهي كل من طريقة المربعات الصغرى المقطوعة للبواري (LTS) و طريقة عمق الانحدار للوسيط (RDM) لغرض الحصول على افضل مقدر لنموذج الانحدار بالاعتماد على المعيار الاحصائي متوسط مربعات الخطاء (MSE).

**الكلمات المفتاحية:** عمق الانحدار ، الانحدار الحصين ، طريقة عمق الانحدار للوسيط ، طريقة M الحصينة.

**ABSTRACT:** The goal of any analytical method to obtain estimates that are very close to their parameters is to represent the population from which the sample was drawn well. When the distribution of a phenomenon is a normal distribution, i.e. it is subject to the established assumptions, then the usual methods such as the least squares method and the maximum likelihood method are efficient methods for obtaining good estimates, but the drawback of the arithmetic mean is that it is greatly affected in the case of the presence of outliers among the data taken, and in such a case, it is necessary to rely on another mediation or to carry out a data refinement process by identifying the outlier observations and compensating for them with alternative values. The function of statistical depth methods in regression is one of the modern methods that can be used to develop robust methods, especially in cases where the values of the outlier observations are not changed based on robust estimates of the location and dispersion matrix, in order to provide robust estimation functions in the estimation and deeper for the location of the data, which have the desired robustness properties such as bias, the influence function, and the high breaking point that works in single-problem data and is insensitive to the presence of outliers and gives us efficient estimates. In this research, we relied on the robust M method and the repeated weighted least squares method. IRLS)), for the purpose of comparing its efficiency with the regression depth methods, which are the least squares method for residuals (LTS) and the regression depth method for mediators (RDM) for the purpose of obtaining the best estimator for the regression model based on the statistical criterion of the mean square error (MSE).

**Keywords :**depth regression, regression hop, median regression depth method, M-method.

### 1. المقدمة.

يعد تحليل الانحدار من أهم الأساليب الإحصائية المستخدمة في دراسة العلاقة بين المتغيرات ، حيث يتمحور حول تقدير معاملات النموذج الخطي لتفسير البيانات أو التنبؤ بها ، وإن دراسة سلوك أي ظاهرة قد تكون إنتاج أو استهلاك أو عدد سكان ... إلخ ومدى تغير هذه الظاهرة خلال الفترات الزمنية المتعاقبة والنموذج الذي يمثل السلسلة الزمنية يساعد في عمليات التقدير والتنبؤ والتخطيط.

يعد عمق الانحدار (RD) طريقة حصينة مقارنة بطريقة المربعات الصغرى عند وجود قيم شاذة في مجموعة البيانات . ولإظهار حصانة عمق الانحدار (RD) ، فأنا نركز على قيمة الانهيار ، وهي القيمة التفصيلية للنموذج وتعرف القيمة التفصيلية بأنها أصغر جزء من مجموعة البيانات يجب استبداله بقيم عشوائية .

وفي هذا البحث تسعى الدراسة وفي هذا البحث الى استعمال نظرية عمق الانحدار باشتقاق بعض طرائق التقدير وهي كل من طريقة M الحصينة و طريقة المربعات الصغرى الموزونة المكررة (IRLS) ، لغرض مقارنة كفاءتها مع طرائق عمق الانحدار وهي كل من طريقة المربعات الصغرى المقطوعة للبواري (LTS) و طريقة عمق الانحدار للوسيط (RDM) من خلال تطبيق اسلوب المحاكات بالاعتماد على المعيار الاحصائي متوسط مربعات الخطاء (MSE) للمقارنة بين افضلية هذه طرائق المستعملة.

### 2. مشكلة البحث.

ان المقدر الحصين بأنه المقدر الذي يتصف باحتفاظه بالعديد من الخصائص المرغوب فيها عند انتهاك إحدى فرضيات التوزيع الطبيعي المتعدد المتغيرات أو وجود قيم شاذة عندها تكون المقدرات الناتجة ذات اداء ضعيف بسبب كبر حجم التباين والتحيز لها . ولذلك قد يتعذر اجراء الطرائق الحصينة في تقدير المعلمات المجهولة الانموذج كونها لا تساعد في الحصول على نتائج دقيقة , الامر الذي يستوجب استخدام اساليب المعق في الانحدار حيث ان دوال التقدير الناتجة من العمق تحتوي على الكثير من خصائص المتانة المرغوب بها المرغوب مثل ( التحيز ودالة التائي ونقطة الانهيار ) الامر الذي يجعلها لا تتأثر كثيرا بوجود القيم الشاذة ، اذ يتم الحصول من خلالها على مقدرات حصينة لمصفوفة الموقع والتشتت و تمتلك كفاءة عالية أي الحصول على أقل تباين ممكن للمعلمات المقدر في الانموذج .

### 3. هدف البحث.

- ❖ استعمال الطرائق الحصينة لتقدير معاملات أنموذج الانحدار الخطي والتي تشمل طريقة M الحصينة و طريقة المربعات الصغرى للوسيط (LMS) .
- ❖ تقدير معاملات أنموذج الانحدار العميق والتي تشمل طريقة المربعات الصغرى المقطوعة للبواري (LTS) و طريقة عمق الانحدار للوسيط (RDM) .
- ❖ مقارنة اداء طرائق تقدير انحدار العمق الاحصائي مع طرائق التقدير الحصين بالاعتماد على المعيار الاحصائي متوسط مربعات الخطاء (MSE) .

### 4. عمق الانحدار.

انحدار العمق (Depth Regression) هو مفهوم إحصائي مهم تم تقديمه من قبل (Rousseeuw and Hubert, 1999) ، والذي يشار إليه بالرمز (RH99) . يعد هذا النموذج امتداد مباشر لنموذج انحدار العمق الخاص Tukey .

يعد عمق الانحدار (RD) طريقة حصينة مقارنة بطريقة المربعات الصغرى عند وجود قيم شاذة في مجموعة البيانات . ولإظهار حصانة عمق الانحدار (RD) ، فأنا نركز على قيمة الانهيار ، وهي القيمة التفصيلية للنموذج . تعرف القيمة التفصيلية بأنها أصغر جزء من مجموعة البيانات يجب استبداله بقيم عشوائية لتفعيل الطريقة . [4]

ويعتبر النموذج ذا عمق الانحدار الأعلى هو الأكثر ملائمة للبيانات ، مقارنة بالنماذج ذات عمق الانحدار الأقل . بالنسبة لأكثر عمق انحدار ضمن مجموعة بيانات معينة  $Z_n$  ، يرمز إليه ب  $\frac{n}{3}$  على الأقل . يتم تحقيق هذا العمق عند إنشاء خط يمثل النموذج الخطي بشكل ثابت ومتمين .

يتسم النموذج الخطي بأعلى مستوى من عمق الانحدار عندما تكون البيانات مستقيمة أو قريبة من الخطية ، بينما يكون العمق في أدنى مستوياته عندما تقع البيانات على منحنيات غير خطية مثل القطع المكافئ أو المنحنيات المحدبة أو المقعرة تماما . وبالتالي ، يعكس عمق الانحدار الأعظم درجة الخطية في البيانات ومدى تطابقها مع النموذج الخطي . [6]

يتفوق عمق الانحدار ، في ظل الإجراءات الحصينة ، على إجراءات الانحدار التقليدية عند وجود مشاهدات شاذة في البيانات أو حتى في غيابها مما يعزز دقته وفعاليتها في التحليل الإحصائي ويتم تعريف مفهوم انحدار العمق باستخدام الصيغة التالية :

$$RD(\beta; Pr) = \inf_{D \in \mathcal{D}} \{Pr((r(\beta) \geq 0) \cap D) + Pr((r(\beta) \leq 0) \cap D^c)\} \quad (1)$$

يمتاز مقدر عمق الانحدار الاعظم بصفة الحصانة ، على غرار مقدر الموقع (المعلمة) الأحادي . ويلاحظ أن مقدر العمق الأعظم ، كما يقدمه (RD) ، يتمتع بمقاومة عالية تصل إلى 33% للتلوث في البيانات دون انهيار ، مقارنة ب 0% للمقدرات التقليدية .

**1-4 خصائص عمق الانحدار . Properties of regression depth**

و يعتبر دالة العمق جيد يجب أن يحقق الخصائص الثبات الإحصائية الأربعة التالية :

أولا : ثبات بالنسبة للانحدار (Regression invariance) : [5]

$$G(\beta + b; F_{(y+x'b, x)}) = G(\beta; F_{(y, x)}) \quad \forall b \in R^p \quad (2)$$

يشير ذلك إلى أن دالة العمق تبقى ثابتة عند إضافة معامل انحدار جديد .

ثانيا : ثبات بالنسبة للمقياس (Scale invariance) :

$$G(s\beta; F_{(sy, x)}) = G(\beta; F_{(y, x)}) \quad \forall s (\neq 0) \in R \quad (3)$$

يشير ذلك إلى أن مضاعفة البيانات أو النموذج بمعامل قياس لا تؤثر على قيمة العمق .

ثالثا : ثبات بالنسبة لتحويلات الخطية غير منفردة (Affine invariance) : [5]

$$G(A^{-1}\beta; F_{(y, A'x)}) = G(\beta; F_{(y, x)}) \quad \forall \text{ nonsingular } p \text{ by } p \text{ matrix } A$$

يعني ذلك أن أي تحويل خطي غير منفرد على البيانات أو النماذج لا يغير قيمة العمق .

$$G(\lambda\beta^* + (1 - \lambda)\beta; F_{(y, x)}) \geq G(\beta; F_{(y, x)})$$

رابعا - التلاشي عند اللانهاية الدالة G تتلاشى عندما  $\|\beta\| \rightarrow \infty$  . وهذا يعني أن :

$$\lim_{\|\beta\| \rightarrow \infty} G(\beta; F_{(y, x)}) = 0$$

**Depth in multiple regression**

**4-2 دوال العمق في الانحدار المتعدد.**

**❖ دالة عمق توكي: Tukey depth method**

هي واحدة من أكثر الطرق شيوعاً لحساب عمق الانحدار والتي تم اقتراحها من قبل (Tukey, 1975) ، ويعرف بعمق الموقع (location Depth) ، حيث تعتمد على فكرة إزالة مجموعة من النقاط وفحص ما إذا كان النموذج لا يزال مناسباً . ويتم حساب عمق كل نقطة على حدة ، وهو يمثل عدد المرات التي يمكن فيها إزالة هذه النقطة دون تغيير ترتيب النقاط الأخرى .

ويعرف عمق موقع النقطة  $x$  إذا  $x = (x_1, x_2, \dots, x_p)$  والتي تنتمي إلى مجموعة البيانات للمتغيرات التوضيحية الموصفة بالاتي:  $\{x_i = (x_{i1}, \dots, x_{ip}); i = 1, \dots, n\} \in R^n$  يمكن تعريفها بأنه يمثل اصغر عدد من نقاط البيانات في نصف مساحة مغلقة الحدود للنقطة  $x$  ويقاس من خلال المعادلة (5) . [7]

$$\text{Depth}(\beta) = \min_{\{e_i; i=1,2,\dots,n\}} \left( \frac{\text{no. } (e_i > 0)}{n}, \frac{\text{no. } (e_i < 0)}{n} \right) \quad (5)$$

حيث أن :

$e_i$  : تمثل البواقي لنموذج الانحدار الخطي .

**( Projection depth )**

**❖ دالة العمق الإسقاطي**

يعتمد على فكرة إسقاط البيانات على خطوط مختلفة وفحص كثافة النقاط في كل إسقاط . حيث كلما كانت نقطة ما محاطة بكثافة أكبر من النقاط الأخرى ، كان عمقها أكبر .

**( Mahalanobis )**

**❖ دالة عمق الماهالانوبيس**

ويعتمد على مسافة الماهالانوبيس ، التي هي عبارة عن مقياس للمسافة بين نقطة ما ومركز التوزيع . حيث كلما كانت نقطة ما أقرب إلى المركز، كان عمقها أكبر، وقد تم اقتراحه من قبل (Liu and Singh , 1993) بالاعتماد على مسافة ماهالانوبيس ، وتعرف مسافة ماهالانوبيس التربيعية الصيغة الرياضية له الآتية : [6]

$$D(Y, \hat{Y}, S) = (Y - \bar{Y}) S^{-1} (Y - \bar{Y})$$

اما مسافة ماهالانوبيس فتعرف بالمعادلة (6) الآتية :

$$MD(Y, \hat{Y}, S) = (1 + (Y - \bar{Y}) S^{-1} (Y - \bar{Y}))^{-1} \quad (6)$$

وتعرف مسافة ماهالانوبيس بصيغة اخرى كالآتي :

$$MD(x, S) = (1 + (x - \bar{x})' S^{-1} (x - \bar{x}))^{-1} \quad (7)$$

حيث ان :

$\bar{x}, S$  : متجه المتوسطات و مصفوفة التباين والتباين المشترك للمتغيرات التوضيحية على التوالي .

ويعاب على هذا المقياس في قابليتها في الفشل نتيجة اعتمادها على مقاييس غير حصينة والمتمثلة بالمتوسطات و العزوم من الدرجة الثانية في صيغتها الرياضية .

#### ❖ دالة عمق الانحدار التقريبي باستخدام مسافة كوك (Cook's distance)

مسافة كوك هي مقياس للتأثير الذي تحدثه نقطة معينة على النموذج . حيث يمكن استخدام مسافة كوك كتقريب لعمق الانحدار، حيث أن النقاط ذات مسافة كوك الكبيرة تكون أقل عمقاً . [4]

$$D_i = \frac{\sum_{j=1}^n \hat{y}_j - \hat{y}_{j(i)}}{p * MSE} \quad (8)$$

حيث ان :

$\hat{y}_j$  : قيمة المشاهدة  $j$  المقدره بوجود جميع البيانات .

$\hat{y}_{j(i)}$  : قيمة المشاهدة  $j$  المقدره بعد حذف المشاهدة  $i$  .

MSE : متوسط مربعات الخطأ .

وتحدد قوة تأثير المشاهدة  $i$  على التقدير وفق الآتي :

#### (Method of estimation regression depth)

يتم توضيح مفهوم عمق الانحدار بشكل أبسط ، لنفترض ان لدينا مجموعة من البيانات ونرغب بإيجاد خط انحدار يمثل العلاقة بين متغيرين . فإن عمق الانحدار هو عدد أدنى من النقاط التي يجب إزالتها من البيانات حتى يتغير خط الانحدار بشكل كبير لجعل النموذج المقدر مناسباً او غير مناسب كلما زاد هذا العدد كلما كان عمق الانحدار أكبر وبالتالي كان النموذج أكثر حصانة وأقل تأثراً بالبيانات الشاذة و كان خط الانحدار أكثر استقراراً وأقل تأثراً بإزالة أي نقطة من البيانات ، وتوجد عدة طرائق لتقدير معاملات عمق الانحدار وفي دراستنا تم التركيز على :

❖ طريقة المربعات الصغرى المقطوعة للبواقي (LTS) .

❖ طريقة عمق الانحدار للوسيط (RDM) .

#### 1-5 طريقة المربعات الصغرى المقطوعة للبواقي. Least squares of trimmed residuals regression (LTS)

تعد طريقة المربعات الصغرى المقطوعة هي إحدى طرائق التقدير لانحدار العمق (depth regression) وتم اقتراحها من قبل (Yijun Zuo and Hanwen Zuo, 2023) إذ يتم في هذه الطريقة استبدال مجموع مربعات الخطأ في طريقة المربعات الصغرى الاعتيادية بالقيمة المطلقة (L1), للحصول على أقل تقدير للانحرافات المطلقة وان المقدرات في طريقة (LTS) تحتوي أيضاً على نقطة انهيار عالية للعينة ويمكنه مقاومة التلوث بنسبة تصل إلى 50% دون انهيار وان فكرة (LTS) بسيطة، وهي تقليل المشاهدات التي لها بواق كبيرة البواقي التربيعية ثم تقليل البواقي الأكبر

والاحتفاظ بما لا يقل عن  $[n/2]$  من المتبقيات التربيعية، حيث يطلق على مجموع مربعات البواقي المتبقية مقدر LTS وأن المقدر بطريقة LTS هو: [4]

إن فكرة طريقة التقدير LTS نشأت من خلال تحويل الأسلوب المتبع في طريقة المربعات الصغرى إذ يتم استبدال مجموع مربعات البواقي للمقدار الآتي:

$$\hat{\beta}_{LTS} = \arg \min_{\beta \in R^p} \sum_{i=1}^h (r^2)_{i:n} \quad (9)$$

حيث ان :

أي يبدأ التقدير باستخدام مقدر المربعات الصغرى للحصول على مقدر LTS ويرمز للبواقي  $(r_i)$  وتساوي المقدار الآتي:

$$r_i(\beta) = y_i - \hat{w}_i \beta$$

حيث ان:

$$r: \text{رتبة مربعات البواقي وان } (r^2)_{1:n} \leq (r^2)_{2:n} \leq \dots (r^2)_{i:n}$$

وان  $D(x, X^{(n)})$  تمثل دالة عمق انحدار (Depth based trimming) المقترح من قبل الباحث (Zuo, Y. (2003)) ويستخدم لتحديد القيم الشاذة بناء على الكثافة المحلية للبيانات ، حيث يتم مقارنة الكثافة المحلية لنقطة معينة بالكثافة المحلية لنقاط الجوار أن الصيغة هي التي تم الاعتماد عليها في هذه الدراسة. تكتب المعادلة بالصيغة الآتية:

$$D(x, X^{(n)}) = \frac{|x - \text{Med}(X^{(n)})|}{\text{MAD}(X^{(n)})} \quad (10)$$

حيث ان :

Med : متوسط نقاط البيانات ويكتب بالصيغة التالية :

$$\text{Med}(X^{(n)}) = \text{median} X^{(n)} \quad (11)$$

MAD: متوسط الانحرافات المطلقة ويكتب بالصيغة التالية :

$$\text{MAD}(X^{(n)}) = \text{Med}\{|x_i - \text{Med}(X^{(n)})| \mid i \in \{1, 2, \dots, n\}\}$$

بالاعتماد على ما سبق وباستخدام المتجه  $Z_n = (\hat{x}_i, y_i); i \in 1, \dots, n \in R^p$  يمكن تعريف الدوال التالية :

$$m(Z^{(n)}, \beta) = \text{Med}_i \{r_i\}$$

$$\sigma(Z^{(n)}, \beta) = \text{MAD}_i \{r_i\}$$

بهدف هو ملائمة النموذج  $y_i$  باستخدام هذه البيانات للوصول إلى تمثيل إحصائي يعكس العلاقة بين المتغيرات المستقلة والتابعة بشكل دقيق تقليل البواقي  $(r_i)$  وتحليل العلاقة حيث يعتمد ذلك على  $Z^{(n)}$  و  $\beta, \alpha$  لذلك يمكن التعبير رياضياً بالصيغة التالية :

$$Q(Z^{(n)}, \beta, \alpha) = \sum_{i=1}^n r_i^2 \left( \frac{|r_i - m(Z^{(n)}, \beta)|}{\sigma(Z^{(n)}, \beta)} \leq \alpha \right) = \sum_{i=1}^n r_i^2 \quad (12)$$

من خلال المعادلات السابقة يمكن صياغة مقدر Lst ويعتبر مقدر إحصائي يعتمد على تقليل البواقي  $(r_i)$  بالصيغة الآتية :

$$\hat{\beta}_{Lst}^n = \hat{\beta}_{Lst}(Z^{(n)}, \alpha) \arg \min_{\beta \in R^p} Q(Z^{(n)}, \beta, \alpha) \quad (13)$$

(Regression depth median method (RDM))

2-5 طريقة عمق الانحدار للوسيط .

(في عام (2000) اقترح الباحثان (Stefan Van Aelst) و اخرون طريقة (RDM) لمعالجة وجود قيم شاذة في مجموعة البيانات لتقدير انحدار أعمق انحدار متعدد الحدود يتميز بقيمة انهيار  $\frac{1}{3}$  مما يعزز دقته وفعاليته في التحليل الإحصائي وإن طريقة (RDM) مبنية في المرحلة الاولى يتم تحديد م بيانات على الشكل التالي: [7]

من قبل الباحثين ((RD Martin, RH Zamar, 1989)) ، ، حيث تعتمد على فكرة إزالة مجموعة من النقاط وفحص ما إذا كان النموذج لا يزال مناسباً وتم تعريف مفهوم انحدار العمق باستخدام الصيغة التالية :

$$DR(Z_n) = \text{med} \frac{y_i}{x_i} \quad (14)$$

المرحلة الاولى :

1- يتم إيجاد متجه مصفوفة المتغيرات الشاملة وكالتالي :

$$\text{sweeping variables} = \{x_1^s \dots \dots x_{p-1}^s\}, \{x_1^s \dots \dots x_{j-1}^s\}$$

2-بعد ذلك يتم حساب تقدير مصفوفة المعلمات  $\beta$  بالاعتماد على  $x_k^s$  وكالتالي:

$$\hat{\beta}_{lk} = \text{med}_{j \in J} \frac{x_{jl} - \text{med}_{il} x_{il}}{x_{jk}^s - \text{med}_{ik} x_{ik}^s} \quad (15)$$

حيث ان :

J :هي مجموعة المؤشرات التي يكون المقام فيها لا يساوي صفر .

4-ثم نستبدل قيم  $x_l$  بالصيغة التالية: [12]

$$x_l - \hat{\beta}_{lk} x_k; k < l - 1 \quad (16)$$

ومن الجدير بالذكر أننا استبدلنا متجه  $x_l$  بمتجه  $x_l - \hat{\beta}_{lk} x_k; k < l - 1$  في جميع المعادلات بعد ترتيبها تصاعدياً أو تنازلياً - من المقاييس الحصينة قليلة الحساسية تجاه القيم الشاذة وأفضل من وتم استخدامه في جميع الطرائق، ويكون متجه  $x_l^s$  بالصيغة: [9] [13]

$$x_1^s = x_1$$

$$x_2^s = x_2 - \hat{\beta}_{21} x_1^s$$

.

$$x_{p-1}^s = x_{p-1} - \hat{\beta}_{p-1,1} x_1^s - \hat{\beta}_{p-1,p-2} x_{p-2}^s$$

5- استبدلنا متجه  $x_l$  بمتجه القيم  $[x_1^s, \dots \dots x_{p-1}^s]$  خارج  $y = y^0$  بحيث  $y = y^0$  وان  $k = \{1 \dots \dots, p - 1\}$

يتم بعد ذلك حساب  $\hat{\beta}_k$  و حسب الصيغة الآتية:

$$\hat{\beta}_k = \text{med}_{j \in J} \frac{y_j^{k-1} - \text{med}_i y_i^{k-1}}{x_{jk}^s - \text{med}_i x_{ik}^s} \quad (17)$$

وكذلك نستبدل متجه  $y^{k-1}$  بمتجه  $y^k = y^{k-1} - \hat{\beta}_{lk} x_k^s; k < l - 1$  في جميع المعادلات بعد ترتيبها تصاعدياً أو تنازلياً - من المقاييس الحصينة قليلة الحساسية تجاه القيم الشاذة وأفضل من وتم استخدامها في جميع الطرائق، ويكون متجه  $y^s$  بالصيغة:

$$y^s = y - \hat{\beta}_1 x_1^s \dots \dots \dots - \hat{\beta}_{p-1} x_{p-1}^s \quad (18)$$

- 1- يتم تكرار المعادلة 1 و 2 من المرات وذلك لضمان استقرار النموذج حتى يتم الوصول إلى التقارب. في كل خطوة تكرر يتم تغيير جميع المعاملات.
- 2- بعد ذلك يتم أخذ معدلات لـ  $r$  من المرات وبذلك يتم تقدير مصفوفتي المعلمات بطريقة MCD. وقد أجرى الباحث تعديلات على خوارزمية طريقة (MCD) لتكون بدقة عالية ووقت أقل للتنفيذ

### 6. طرائق التقدير الحصينة. robust Classic Methods Of Estimation

أن الغرض الأساسي من طرائق التقدير هو الحصول على مقدرات معالم الإنموذج المراد دراسته ويجب أن تتوفر فيه المواصفات الجيدة التي تجعله إنموذجاً يمكن الاستفادة منه وتكون طرائق التقدير مختلفة باختلاف الأفكار والأساليب لتحقيق غرضين مهمين ، فالأول هو احتواء المقدرات على مواصفات جيدة ، والغرض الآخر هو ظهور الطريقة بأسلوب سهل التنفيذ ، وسنتطرق إلى طرائق التقدير الحصينة وفي دراستنا تم الاعتماد على الطرائق الحصينة التالية: [1]

- ❖ طريقة المربعات الصغرى الموزونة المكررة (IRLS).
- ❖ طريقة M الحصينة.

#### 1-6 طريقة المربعات الصغرى الموزونة المكررة (IRLS).

يتم استعمالها في حل بعض المشاكل بطريقة تكرارية تتضمن كل خطوة حلاً لمشكلة المربعات الصغرى المرجحة ، إذ يتم استعمال طريقة المربعات الصغرى الموزونة المكررة (IRLS) للعثور على تقديرات الاحتمال القصوى لإنموذج خطي معمم ، وفي الانحدار الحصين للعثور على مقدر كوسيلة للتخفيف من تأثير القيم المتطرفة وقد تم تطوير هذه الطريقة لحل مشكلة طول المتجه  $L_2$  التي يمكن الحصول عليها بواسطة حل  $L_s$  الاعتيادية، بينما طول المتجه  $L_p$  هو حلول أصغر ما يمكن إذ  $(1 \leq p \leq 2)$  ، والخوارزميات المكررة تسمى المربعات الصغرى الموزونة المكررة (IRLS) تطورت لحل هذه المشاكل والتي تقع بين مشكلة القيم المطلقة الصغرى وحلول المربعات الصغرى الاعتيادية ، الفائدة الرئيسية لطريقة المربعات الصغرى الموزونة المكررة (IRLS) لتزويدنا بطريقة سهلة لحساب حل طول المتجه  $L_1$  التقريبي ، حل طول المتجه  $L^1$  هي تعرف لتكون أكثر حصينة من حلول طول المتجه  $L^2$  ، والمشكلة تحل بطريقة المربعات الصغرى الموزونة المكررة ، وهو تصغير البواقي الموزونة في المربعات الصغرى الاعتيادية . [2]

$$r = W(Lm - d)$$

$r$  : يمثل متجه البواقي الموزونة .

$W$  : يمثل الوزن .

$Lm$  : طول المتجه ذو البعد  $m$

$d$  : متجه البيانات *data vector*

المشتقة الأولى لطول المتجه للبقايا الموزونة في المربعات الصغرى الاعتيادية تصبح :

$$L'W = \frac{\partial}{\partial m} (Lm - d)' W' W (Lm - d) \quad (19)$$

الاختيار الدقيق للوزن  $W$  هو واحد من نتائج تصغير طول المتجه  $L_p$  للبواقي , اختبار العنصر القطري  $i^{th}$  للوزن  $W$  يكون دالة مركبة  $i^{th}$  لمتجه البواقي كما يأتي :

$$\text{diag}(W) = |r|^{(P-2)/2}$$

طول المتجه للبواقي الموزونة يصبح :

$$r'W'Wr = |r|^{(P-2)} r = |r|^P$$

لذلك يمكن عدّ هذا كطريقة لمقدرات المشتقة الاولى في طول المتجه  $L_p$  للبواقي .

هذه الطريقة جيدة لأطوال المتجهات عندما  $1 \leq P \leq 2$  وعندما طول المتجه  $L_1$  مرغوب به الوزن يصبح :

$$\text{diag}(W) = |r|^{-1/2}$$

$\text{diag}(W)$ : يمثل مصفوفة الأوزان القطرية

هذه سوف تقلل مساهمة البواقي الكبيرة وتحسن مطابقة البيانات فتصبح جاهزة للتقدير . طول المتجه  $L_1$  يعتمد على التقليل هو الحصين , وهو أقل حساسية للقيم الشاذة .

اقترح هوبر إنموذج (1973) Huber  $L_1/L_2$  norm طريقة هجينة التي تعالج البواقي الصغيرة في طول المتجه  $L_2$  والبواقي الكبيرة في طول المتجه  $L_1$ , إذ يتعامل هذا النهج مع كل من المتقطعة و Gaussian , ويمكن تحقيقه عن طريق الوزن الآتي :

$$\text{diag}(W) = \begin{cases} |r|^{-1/2} & |r| > \varepsilon \\ 1 & |r| \leq \varepsilon \end{cases} \quad (20)$$

إذ  $\varepsilon$  هي القيمة التي يمكن أن تستعمل كعتبة ( threshold ) بين طول المتجه  $L_1$  وطول المتجه  $L_2$  .

لإيجاد مقدرات المربعات الصغرى الموزونة التكرارية ( IRLS ) , نقوم بالتقدير بطريقة المربعات الصغرى التقليدية الاعتيادية OLS, والمشاهدات التي تمتلك تأثيراً كبيراً مثل ( قيم مسافة كوك لها Cook's أكبر من 1 ) . يتم استبعادها تلقائياً بعد الخطوة الأولى , ثم بعد ذلك يتم حساب الأوزان لكل مشاهدة باستعمال دالة هوبر Huber function ( وهذه الدالة تقلل المشاهدات التي لها بواق كبيرة ) ثم يتم حساب المربعات الصغرى الموزونة . بعد تكرار حساب المربعات الصغرى الموزونة عدة مرات تنتقل دالة الترجيح أو الوزن إلى وزن توكي الثنائي Tukey , ويتحول إلى توزيع طبيعي بكفاءة 95%.

## 2-6 طريقة M الحصينة.

وهي إحدى طرائق التقدير لانحدار الحصين وأن الحرف ( M ) يعني هي طريقة تقدير الإمكان الأعظم ( Maximum Likelihood ) وأن المقدر هنا بطريقة M هو

$$\hat{\beta} = [\beta_n(x_1, x_2, \dots, x_n)] = \beta \quad (21)$$

وبأخذ التوقع سيكون [3]

$$E[\beta_n(x_1, x_2, \dots, x_n)] = \beta$$

ومن المعادلة (19 – 2) تبين أن المقدر

$$\hat{\beta} = [\beta_n(x_1, x_2, \dots, x_n)]$$

هو تقدير غير متحيز والتباين أقل ما يمكن وكذلك مقدر M صغير جدا بالنسبة للتباين وأقل ما يمكن مقارنة بتباين بقية المقدرات

$$\text{var}(\hat{\beta}) \geq \frac{[\hat{\beta}]^2}{nE\left[\left(\frac{\partial}{\partial \beta} \ln f(x_i; \beta)\right)^2\right]} \quad (22)$$

إذ أن  $(\hat{\beta})$  تقدير خطي آخر غير متحيز ل  $\beta$

أن مبدأ طريقة M في التقدير هو تقليل أو تصغير دالة البواقي P :

$$\hat{\beta}_m = \min_{\beta} P(y_i - \sum_{j=0}^k x_{ij} \beta_j) \quad (23)$$

علينا أن نحل

$$\begin{aligned} \min_{\beta} \sum_{i=1}^n P(u_i) &= \min_{\beta} \sum_{i=1}^n P\left(\frac{e_i}{\sigma}\right) \\ &= \min_{\beta} \sum_{i=1}^n P\left(\frac{y_i - \sum_{j=0}^k x_{ij} \beta_j}{\sigma}\right) \end{aligned}$$

وللحصول على معادلة ( 24 ) , وضعنا مقدر ل  $\sigma$

$$\hat{\sigma} = \frac{\text{MAD}}{0.6745} = \frac{\text{median}|e_i - \text{median}(e_i)|}{0.6745} \quad (24)$$

بالنسبة للدالة  $\rho$  تستعمل Tukeys biasquare دالة الهدف

$$\rho(ui) = \begin{cases} \frac{u_i^2}{2} - \frac{u_i^4}{2c^4} + \frac{u_i^6}{6c^4}, & |u_i| \leq c \\ \frac{c^2}{6}, & |u_i| > c \end{cases} \quad (25)$$

فضلاً عن ذلك نحن نبحث عن أول مشتق جزئي  $\beta_m$  ل  $\beta$  لذلك

$$\sum_{i=1}^n x_{ij} \psi\left(\frac{y_i - \sum_{j=0}^k x_{ij} \beta}{\hat{\sigma}}\right) = 0, \quad j = 0, 1, 2, 3, \dots, k \quad (26)$$

إذ أن  $\psi = \rho'$  ,  $x_{ij}$  هي الملاحظة الأولى على المتغير المستقل  $\sum_{i=1}^n x_{i0} = 1$

$$w(e_i) = \frac{\psi\left(\frac{y_i - \sum_{j=0}^k x_{ij} \beta}{\hat{\sigma}}\right)}{\left(\frac{y_i - \sum_{j=0}^k x_{ij} \beta}{\hat{\sigma}}\right)} \quad (27)$$

$$W_i = \begin{cases} \left[1 - \left(\frac{u_i}{c}\right)^2\right]^2, & |u_i| \leq c \\ 0, & |u_i| > c \end{cases}$$

نحن نأخذ  $c=4.685$  لدالة Tukeys biasquare المرجحة لذا فان المعادلة (24) تصبح

$$\sum_{i=1}^n x_{ij} w_i (y_i - \sum_{j=0}^k x_{ij} \beta_j) = 0, \quad j=0,1,\dots,k \quad (28)$$

في هذه الطريقة نفترض وجود مقدر أولي  $\beta$  و  $\hat{\sigma}_i$  هو أرقام وأعداد المعلمات ثم [4]:

$$\sum_{i=1}^n x_{ij} w_i^0 (y_i - \sum_{j=0}^k x_{ij} \beta_j^0) = 0, \quad j=0,1,\dots,k \quad (29)$$

ويمكن كتابة معادلة (29) كما يأتي :

$$X'W_i X \beta = X'W_i Y \quad (30)$$

إذ إن  $W_i$  هو مصفوفة من الدرجة  $n \times n$  والوزن لها عناصر القطر الرئيس وأن معادلة (28-2) تعرف بالمربعات الصغرى الموزونة  $WLS$  وان الحل لهذه المعادلة يعطى تقدير لـ  $\beta$  , وبكلام اخر :

$$\hat{\beta} = (X'W_i X)^{-1} (X'W_i Y) \quad (31)$$

### (Simulation)

### 7. الجانب التجريبي.

تعد المحاكاة أسلوب رقمي يستعمل من خلال الحاسوب الالكتروني في إيجاد نموذج مماثل الى الانموذج الحقيقي من دون المحاولة للحصول على الانموذج الحقيقي نفسه، ويمكن القول ان أسلوب المحاكاة هي أنواعا من العمليات المنطقية والرياضية تمثيل وتقليد للواقع الحقيقي لوصف سلوك الظواهر الحقيقية الواقعية المعقدة صعبة الفهم والتحليل، ووصف سلوكها خلال فترة زمنية معينة، من خلال الحصول على بيانات تقريبية لدراسة تلك الظواهر في حال استحالة الحصول على تلك البيانات او عدم توفرها بشكل كافي ويمكن تلخيص اسلوب المحاكات بالخطوات الآتية :

أولاً- تحديد القيم الافتراضية: تم اختيار اربع حجوم للعينات وهي ( $n = 25,50,75,100$ ) واستخدمت قيم افتراضية للمعلمات افترضنا ان عدد المعالم  $m=5$  فكانت كما في الجدول الآتي: [6] [1]

جدول رقم (1) القيم الافتراضية لمعلمات الانموذج

Experiment	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$
1	1.25	3.5	0.75	-1.5	-0.1

ثانياً - توليد بعض المتغيرات العشوائية المستمرة: توجد عدة طرائق لتوليد المتغيرات العشوائية التي تتبع التوزيع الطبيعي متعدد المتغيرات وغيره منها طريقة (Box – Muller) والطريقة التقاربية (Asymptotically) والطريقة القطبية (Polar). وتعتبر طريقة (Box – Muller) أكفأ الطرائق وأكثرها استخداماً وأسهلها في التطبيق البرمجي. ولتوليد المتغير العشوائي ( $Z$ ) الذي يتبع التوزيع الطبيعي القياسي يجب أن نتبع الخطوات الآتية وفق طريقة (Box – Muller): في البدء، يتم توليد رقمين عشوائيين مستقلين  $U_1$  و  $U_2$  بحيث يتبعان التوزيع المنتظم المستمر باستعمال الأمر RND وباستعمال لغة البرمجة (ماتلاب) يتم توليد متجه من هذين الرقمين بحجم العينة المطلوبة ( $n$ ) وكالاتي: [4] [9]

$$U_i = \text{RND}(1, n) \quad i = 1,2 \quad (32)$$

1- تحويل هذه الرقمين (متجه الأرقام) إلى التوزيع الطبيعي القياسي حسب الصيغة الآتية:

$$\begin{aligned} Z_1 &= \sqrt{-2 \ln(U_1)} \cos(2\pi U_2) \\ Z_2 &= \sqrt{-2 \ln(U_1)} \sin(2\pi U_2) \end{aligned} \quad (33)$$

حيث أن  $Z_1$  و  $Z_2$  متغيران عشوائيان طبيعيين قياسييان مستقلان.

2- لتحويل المتغير العشوائي  $Z$  من توزيع  $N(0, I)$  إلى توزيع  $N_p(\underline{\mu}, S)$  يتم ذلك من خلال الصيغة الآتية:

$$\underline{X} = \underline{\mu} + gZ \quad (34)$$

حيث أن:

$g$ : مصفوفة مثلثية سفلى من الدرجة  $p \times p$  بحيث أن  $gg' = S$  يتم الحصول عليها باستخدام طريقة كلاوسكي وكالاتي [3]:  
للحصول على مصفوفة مثلثية سفلى Lower Triangular نتبع قاعدة Cholesky وفق الخطوات الآتية:  
إذا كانت لدينا  $M_{p \times p}$  مصفوفة متماثلة فإن المصفوفة  $S$  هي مصفوفة مثلثية عليا Upper Triangular، تعرف كما يلي:

$$\begin{aligned} S_{11} &= \sqrt{m_{11}} \\ S_{1j} &= m_{1j}/S_{11}, \quad j = 2, 3, \dots, p \\ S_{ii} &= \left( m_{ii} - \sum_{k=1}^{i-1} S_{ik}^2 \right)^{1/2}, \quad i > 1 \\ S_{ij} &= \left( m_{ij} - \sum_{k=1}^{i-1} S_{ik}S_{kj} \right) / S_{ii}, \quad j > i \\ S_{ij} &= 0, \quad i > j \end{aligned}$$

وبالتالي يمكن الحصول على المصفوفة المثلثية السفلى كما يلي:

$$T = S'$$

❖ توليد متغيرات عشوائية ذات توزيع طبيعي متعدد ملوث:

إن عملية توليد متغيرات عشوائية ملوثة تعني أنه توجد حالة وجود نسبة معينة ( $\varepsilon$ ) من التلوث في البيانات، إن البيانات الملوثة تتبع توزيعاً آخر وتشارك مع بقية البيانات في معلمة التباين ولكن يوجد اختلاف في معلمة الوسط لذلك يسمى هذا النموذج بنموذج إزاحة الوسط وهذا يعني أن المتغير العشوائي  $X$  يمكن أن يكون بالصيغة الآتية:

$$X \sim \left[ (1 - \varepsilon)N_p(\underline{\mu}, S) + \varepsilon N_p(\underline{\mu}_1, S) \right] \quad (35)$$

حيث أن:  $\underline{\mu} \neq \underline{\mu}_1$ .

ولتوليد مثل هكذا نوع من المتغيرات نتبع الآتي:

- 1- تحديد نسبة احتمال ثابت ( $\varepsilon$ ) يقع ضمن الفترة (0,1).
- 2- توليد ثلاثة أرقام عشوائية ( $U_1, U_2, U_3$ ) تتبع التوزيع المنتظم ضمن الفترة (0,1).
- 3- إذا كان ( $U_3 > \varepsilon$ ) يتم توليد المتغيرين العشوائيين ( $x_1, x_2$ ) وذلك باستخدام طريقة (Box - Muller) اللذين يتبعان التوزيع الطبيعي بمتجه أوسط حسابية  $\underline{\mu}$  ومصفوفة تباين مشترك  $S$ .
- 4- إذا كان ( $U_3 \leq \varepsilon$ ) يتم توليد المتغيرين العشوائيين ( $x_1, x_2$ ) اللذين يتبعان التوزيع الطبيعي بمتجه أوسط حسابية  $\underline{\mu}_1$  ومصفوفة تباين مشترك  $S$ .
- 5- اعتماد نسب مختلفة لتلوث البيانات وذلك بافتراض ثلاثة مستويات لهذه النسب وهي ( $T = 25\%$ ) ،  $T$  تمثل نسبة التلوث المفترضة في هذا البحث
- 6- تكرار التجربة مساوياً إلى (1000) مرة.

رابعا - تقدير المعالم للنموذج اعلاه باستعمال طرائق التقدير التي تم ذكرها سابقا .

سادسا- حساب معيار المتوسط لمتوسط مربعات الأخطاء (MSE) لجميع تجارب المحاكاة ومقارنة النتائج على أساسها وفق الصيغة الآتية: -

$$MSE = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n - k + 1} \quad (36)$$

وفي ما يلي نتائج تجارب المحاكات:

جدول (2) قيم متوسط مربعات الخطاء للمعالم ولأنموذج الانحدار الخطي العام عند استعمال (n=25) عندما تكون نسبة التلوث (T=25%)

P=2,q=4					
S(1, 1) = 0.3, S(2, 2) = 0.3, S(3, 3) = 0.3, S(4, 4) = 0.3					
MSE of est Parameter	IRLS	M	LTS	RDM	Best
b0	0.327989119	0.121594687	0.068248056	0.048730031	RDM
b1	0.172986948	0.043317278	0.008142015	0.011769876	LTS
b2	0.149207496	0.030383903	0.010954304	0.008088624	RDM
b3	0.330657586	0.107757698	0.007117296	0.010378073	RDM
b4	0.1330801	0.032089814	0.010150804	0.007801851	RDM
MSE of Model	0.22278425	0.067343581	0.022300223	0.015975963	RDM

جدول (3) قيم متوسط مربعات الخطاء للمعالم ولأنموذج الانحدار الخطي العام عند استعمال (n=50) عندما تكون نسبة التلوث (T=25%)

P=2,q=4					
S(1, 1) = 0.3, S(2, 2) = 0.3, S(3, 3) = 0.3, S(4, 4) = 0.3					
MSE of est Parameter	IRLS	M	LTS	RDM	Best
b0	0.080049923	0.037810022	0.024869879	0.028803498	LTS
b1	0.058582772	0.015552678	0.004127028	0.0036499	RDM
b2	0.029375392	0.016247189	0.003988331	0.003365972	RDM
b3	0.114749952	0.015863613	0.003030984	0.004323985	LTS
b4	0.031423762	0.016257261	0.004216279	0.003733609	RDM
MSE of Model	0.06283636	0.019972799	0.0090465	0.007775393	RDM

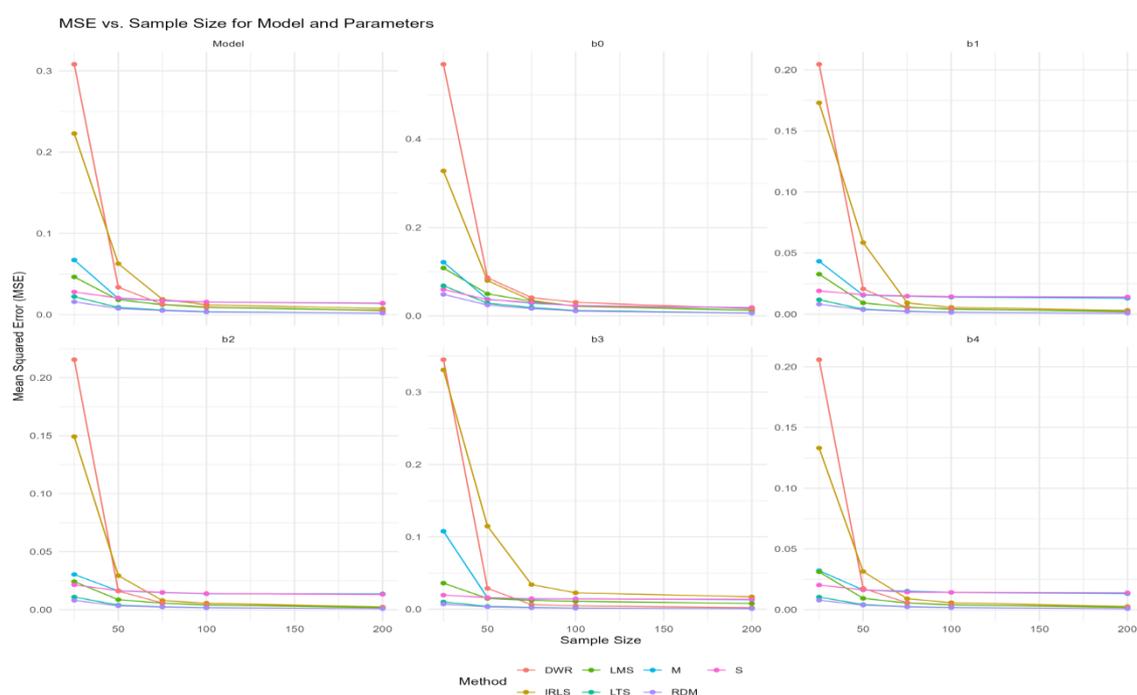
جدول (4) قيم متوسط مربعات الخطاء للمعالم ولأنموذج الانحدار الخطي العام عند استعمال (n=75) عندما تكون نسبة التلوث (T=25%)

<b>P=2,q=4</b>					
<b>S(1, 1) = 0.3, S(2, 2) = 0.3, S(3, 3) = 0.3, S(4, 4) = 0.3</b>					
MSE of est Parameter	IRLS	M	LTS	RDM	Best
b0	0.021873084	0.023480648	0.012280415	0.011114168	RDM
b1	0.005559284	0.013964711	0.001664186	0.001488637	RDM
b2	0.005649261	0.013753352	0.001773923	0.00163484	RDM
b3	0.022716657	0.014445208	0.001769616	0.001510284	RDM
b4	0.005759179	0.014220173	0.001552694	0.001815407	LTS
MSE of Model	0.012311493	0.015695414	0.003848167	0.003472667	RDM

جدول (5) فيم متوسط مربعات الخطاء للمعالم ولانموذج الاتحدار الخطي العام عند استعمال (n=100) عندما تكون نسبة التلوث (T=25%)

<b>P=2,q=4</b>					
<b>S(1, 1) = 0.3, S(2, 2) = 0.3, S(3, 3) = 0.3, S(4, 4) = 0.3</b>					
MSE of est Parameter	IRLS	M	LTS	RDM	Best
b0	0.012459431	0.017931153	0.006565351	0.005988132	RDM
b1	0.00309308	0.013003315	0.000757327	0.000718667	RDM
b2	0.002375795	0.013678995	0.0009719	0.000896642	RDM
b3	0.017254527	0.013257263	0.00082453	0.000792858	RDM
b4	0.00260486	0.013180822	0.000843038	0.000812888	RDM
MSE of Model	0.007557539	0.013976791	0.001992429	0.001841838	RDM

شكل رقم (1) يبين فيم متوسط مربعات الخطاء للموديل والمعلمات ولانموذج الانحدار الخطي العام عند استعمال (n=25,50,75,100)



عند افتراض قيم المعالم:  $(\beta_0 = 1.25, \beta_1 = 3.5, \beta_2 = 0.75, \beta_3 = -1.5, \beta_4 = -0.1)$

❖ اظهرت نتائج المحاكاة في الجدول (2) و الجدول (3) و الجدول (4) و الجدول (5) عند كافة حجوم العينات ان الطريقة الافضل للتقدير هي طريقة عمق الانحدار الوسيط (RDM) كونها تمتلك (MSE) ولكافة معلمات النموذج المقدره.

**1.5.الاستنتاجات.**

1. نستنتج من خلال نتائج المحاكاة بان أفضل طريقة لتقدير معلمات نماذج الانحدار الخطية هي طريقة RDM وطريقة (LTS) كونهما يمتلكان اقل (MSE).
2. من خلال الرسوم البيانية في الشكل رقم (1) تتناقص القيم الخاصة بالمقياس الإحصائي متوسط مربعات التكامل (MSE) كلما زاد حجم العينة وهذا يطابق النظرية الإحصائية لهذا المؤشر.
3. نلاحظ من خلال نتائج المحاكاة ارتفاع قيمة (MSE) لمقدر IRLS عندما يكون حجم العينة صغير.
4. من خلال نتائج الجانب العملي وجدنا بان طرائق عمق الانحدار هما الأفضل وتعطي للمقدرات (MSE) اقل.

## 16. التوصيات.

1. استعمال طرائق عمق الانحدار بوجود مشكلة القيم الشاذة بدلا من المقدرات الاخرى. من مرونة وكفاءة عالية.
2. يوصي الباحث باستعمال طريقة **RDM** وطريقة **(LTS)** معالم انموذج الانحدار الخطي بوجود مشكلة الشوارد والقيم الشاذة .
3. يمكن التوصية بمعرفة مدى تأثير زيادة نسبة التلوث (زيادة عدد الشوارد وزيادة مستوى شرودها) على كفاءة الطريقة الحصينة.
4. توسيع استعمال عمق الانحدار في الدراسات الاقتصادية والاجتماعية والطبية.

## المصادر العربية والاجنبية .

- 1- Al-Qahwani, Wasfi Taher Saleh, (2000). "Some robust estimation methods in improving digital images", a specialized doctoral thesis submitted to the Board of Directors of the College of Economics, University of Baghdad, Iraq.
- 2- Ali, Fayyad Abdullah, (2000). "Rigid Estimation in Simultaneous Equation Models", PhD Thesis, College of Administration and Economics, University of Baghdad.
- 3- Saleh, Dhikaa Yousef Aziz, (2001). "Comparison of some robust methods for estimating the parameters of the multiple linear regression model", Master's thesis submitted to the College of Computer Science and Mathematics, University of Mosul.
4. Al-Guraibawi, M., & Mohammed, B. (2023). Regression Depth for Statistical Depth Function. *Journal of Al-Rafidain University College For Sciences* (Print ISSN: 1681-6870, Online ISSN: 2790-2293), (1), 539-546.
5. Carvalho, M., Le Saux, B., Trouvé-Peloux, P., Almansa, A., & Champagnat, F. (2018, October). On regression losses for deep depth estimation. In 2018 25th IEEE International Conference on Image Processing (ICIP) (pp. 2915-2919).
6. Dovoedo, Yinaze Herve. Contributions to outlier detection methods: Some theory and applications. The University of Alabama, 2011.
7. Eldesokey, A., Felsberg, M., & Khan, F. S. (2019). Confidence propagation through cnns for guided sparse depth regression. *IEEE transactions on pattern analysis and machine intelligence*, 42(10), 2423-2436.
8. Fu, H., Gong, M., Wang, C., Batmanghelich, K., & Tao, D. (2018). Deep ordinal regression network for monocular depth estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2002-2011).
9. Mosler, Karl. "Depth statistics." *Robustness and complex data structures*. Springer, Berlin, Heidelberg, 2013. 17-34.
10. Rousseeuw, P. J., and A. M. Leroy. "Robust Regression and Outlier Detection. New York: John Wiley & Sons." (1987).
11. Zuo, Y., & Zuo, H. (2023). Least sum of squares of trimmed residuals regression. *Electronic Journal of Statistics*, 17(2), 2416-2446.
12. Zuo, Yijun, and Robert Serfling. "General notions of statistical depth function." *Annals of statistics* (2000): 461-482..
13. Zuo, Y., & Zuo, H. (2023). Computation of least squares trimmed regression--an alternative to least trimmed squares regression. arXiv preprint arXiv:2312.05077.