1- المقدمـة وهدف البحث *:

ان التطور الحاصل في مجال البحث العلمي فرض نفسه في موضوع تقدير قيم معالم النموذج الاحصائي وباتت الأفكار والأساليب المقترحه على طرق التقدير كثيره جدا حيث صمم بعضها لمعالجة حالات تخص بعض الظواهر وقسم آخر وضع عندما تسلك بعض قيم الظاهره سلوكاً خاصاً يختلف عن أقرانها وهو مايسمى في وقتنا هذا بالشواذ(Outliers) .

حيث كانت أول فكره أو مقاله تنشر عن الازعاجات التي تسببها القيم الشاذه لطريقة المربعات الصغرى على يد الباحث Legendre عام 1805 .

ثم توالت البحوث التي تهتم بكيفية التصرف مع القيم الشاذه ما إن وجدت في البيانات وخاصه في النصف الثاني من القرن العشرين حيث أصبحت للطرق التي تتعامل مع بيانات قد تمتلك قيم شاذه ،فئه خاصه من طرق التقدير وأطلق عليها تسمية طرائق التقدير الحصينه (Robust estimation methods) حيث تكون هذه الطرائق غير حساسة تجاه القيم الشاذه إن وجدت في البيانات . وكما هو معروف فإن النموذج الاحصائي يختلف من حيث الشكل والهدف فقد يكون نموذج سلسله زمنيه أو نموذج توزيع إحتمالي أو نموذج تصميم أو نموذج إنحدار أو غير ذلك . حيث حظي كل نموذج من النماذج بمجموعة ليست بالقليله من البحوث التي تبين أفكاراً وإقتراحات لطرق حصينه في تقدير معالمه . ولكن من أقل النماذج الإحصائيه حظا في ذلك هي النماذج اللاخطيه والتي منها على سبيل المثال النماذج الأُسية والنموذج اللوجستي وغيرها من النماذج الاخرى .و معروف ان العديد من الظواهر الطبيعية عند دراسة سلوكها نجدها تسلك سلوكا غير خطي ولتحليل تلك الظواهر تستخدم النماذج اللاخطية لوصف تلك الظواهر وفي الواقع وبالنظر لصعوبة استخدام هذة النماذج فانه غالبا ما تستخدم النماذج الخطية لوصف وتحليل هذه الظواهر .ومن هنا جاء الاهتمام في هذا البحث على استخدام الطرق الحصينة في تقدير معالم النماذج اللاخطيه حيث يهدف هذا البحث الى مقارنة بعض الطرق التي تهتم في ايجاد مقدرات حصينة لمعالم النماذج اللاخطية فضلا عن اسلوبين مقترحين من قبل الباحث لتحديد قيمة الثابت k (ثابت التشذيب) المستخدم في دالة هوبر وتطبيق الطرق قيد البحث على نموذج غير خطي وبنسب تلوث مختلفه وبأحجام عينات مختلفه للوصول الى تقديرات كفوءه وحصينه لمعالم النماذج اللاخطيه وذلك باستخدام اسلوب المحاكاة .

---

* البحث مستل من اطروحة دكتوراه .

2- طرق التقدير Methods Of Estimation :

ان الهدف من طرق التقدير هوايجاد مقدرات لمعالم النموذج المدروس تتصف بمواصفات جيدة بحيث تؤهلها الى تكوين نموذج تقديري يمكن الاعتماد عليه في اغراض مختلفة .
وفيما يخص موضوع دراستنا فيمكن تقسيم طرق التقدير الى طرق تقدير كلاسيكية وكذلك طرق تقدير حصينة لاتتأثر الى حدٍ ما بالقيم الشاذة إن وجدت .

2-1 طرق التقدير الكلاسيكية Classic Methods Of Estimation:

توجد عدة طرق لتقدير معالم نموذج الانحدار اللاخطي إلا أن هذه الطرق تعتبر غير حصينة اي أن وجود الشواذ في البيانات المستخدمة يؤثر على خواص مخرجاتها ومن هذه الطرق ، هي طريقة steepest Descent[3] وطريقة Marquardts Compromise[1] وطريقة المربعات الصغرى Least Square[3] والتي تعد من اشهر الطرق واكثرها استخداماً في تقدير عالم نموذج الانحدار اللاخطي ،فأذا كان لدينا النموذج الاتي:-

$$y_i = f(\underline{x}_i , \underline{\theta}) + \in_i \qquad (1)$$

حيث ان :-

y_i :- تمثل قيمة المشاهدة i ويسمى متغير الاستجابة .

$\underline{x}_i$:- تمثل موجه مجموعة المتغيرات التوضيحية للمشاهدة رقم i ،ويمكن القول كذلك انه يمثل الصف رقم i من مصفوفة المتغيرات التوضيحية .

$\underline{\theta}$:- تمثل موجه المعالم المجهولة والمطلوب تقديرها في النموذج .

$f(.)$:- تمثل دالة النموذج والتي يتحدد بموجبها شكل الانحدار .

$\in i$:- الخطأ العشوائي.

وان تقدير موجه المعالم $\underline{\theta}$ بأستخدام طريقة (OLS) بتطبيق طريقة نيوتن رافسون التكرارية بهدف تصغير مربعات الخطأ يكون وفق الصيغة الاتية :-

$$\hat{\underline{\theta}}_{j+1} = \hat{\underline{\theta}}_j + (Z'_j Z_j)^{-1} Z'_j (Y_i - f_j(\underline{x}_i, \hat{\underline{\theta}}_j)) \qquad (2)$$

إذ أن :-

Z_j :- هي المصفوفة Z في الدورة j وهي تمثل مصفوفة مشتقات الدالة $f(), \underline{\theta}\ \underline{x}_i$ نسبةً الى $\underline{\theta}$ والتي بدورها تستبدل بالقيم الحالية لـ $\underline{\theta}$.

أي أن

$$Z_{ik}^{j} = \left[\frac{\partial f(\underline{x}_i, \underline{\theta})}{\partial \theta_k} \right] \qquad \text{where} \qquad \underline{\theta} = \hat{\underline{\theta}}_j \qquad (3)$$

ومن المفروض أن تتقارب قيم $\hat{\underline{\theta}}_{j+1}$ من $\hat{\underline{\theta}}_j$ بأستمرار تطبيق الصيغة (2) حيث يتم التوقف عن تكرار العمل بعد الاقتناع بقيمة خطأ متحقق من الصيغة الآتية :-

$$\left| \hat{\underline{\theta}}_{j+1} - \hat{\underline{\theta}}_j \right| < \underline{\varepsilon} \qquad (4)$$

حيث أن ε هي الفرق بين تكرارين متتاليين.

5- الجانب التجريبي :-

تم في هذة الفقرة تطبيق طريقة M ببعض إحصائياتها التي تم التطرق اليها في الفقرة (٢) من هذا البحث بتوظيف اسلوب المحاكاة وذلك بتقدير معالم نموذج لا خطي وهو نموذج لوجستك المستمر continuous Logistic model [9] والذي يحتوي على معلمتين ويأخذ الشكل الاتي:-

$$y_i = \frac{1}{1+e^{-(\theta_1+\theta_2 x_i)}} + e_i \qquad i = 1,2,\ldots,n \qquad (18)$$

إذ أن :-

θ_1, θ_2 :- هي معلمتا النموذج المراد تقديرهما .

y :- متغير الاستجابة .

x_i :- المتغير التوضيحي .

e_i :- الخطأ العشوائي .

وفيما يأتي مراحل دراسة هذا النموذج .

5-1 توليد بيانات النموذج Generating data of model :-

يعد اسلوب المحاكاة من الاساليب العلمية الرصينة التي تقوم على اعطاء صورة طبق الاصل لظاهرة حقيقية ليتسنى الاستفادة من هذه الصورة في دراسة خواص تلك الظاهرة المدروسة ومميزاتها،و يمكن ان يُعتمد أسلوب المحاكاة في اثبات صحة طريقة معينة يكون من الصعب اثبات صحتها نظرياً.

وقد تم توليد البيانات بموجب الخطوات الآتية بعد تثبيت حجم العينة المراد توليدها وشكل النموذج المطلوب .

5-1-1 المتغيرات التوضيحية Independent variables :-

عند توليد المتغيرات التوضيحية يجب ان يؤخذ بنظر الاعتبار طبيعة كل متغير والظاهرة التي يمثلها ذلك المتغير فهناك ظواهر لا تقبل الكميات السالبة أو الاعداد العشرية أو يستبعد عن متغيراتها انها تأخذ قيماً معينة من أجل ان تكون البيانات المولدة تطابق الواقع الذي تُحاكيه .

حيث تم توليد المتغير التوضيحي x_i بموجب الصيغة الاتية:-

$$x_i = a + INT((b - a + 1) * u_i) \qquad (19)$$

وقد تم اختيار a = - 4 و b = 8 اي ان x_i المولدة محصورة بين -4 , 8 وتأخذ التوزيع المنتظم Uniform distribution .

5-1-2 الاخطاء العشوائية Random errors :-

يتم توليد الاخطاء العشوائية على اساس الغاية المرجوة من البيانات حيث يمكن توفير اي مشكلة في البيانات بالاخطاء العشوائية والمشكلة التي نحن بصدد دراستها هي احتواء البيانات على القيم الشاذة (outliers) وعليه نحتاج الى توليد اخطاء عشوائية تأخذ التوزيعات الاتية :-

. التوزيع الطبيعي Normal distribution :- تم الاعتماد على طريقة Box-Muller في توليد متغيرات التوزيع الطبيعي وفق الخطوات الاتية (٥).:-

الخطوة الاولى :- توليد متغيرين عشوائيين U1 و U2 يتبعان التوزيع المنتظم U(0,1) بأستخدام الامر RND المتوفر في لغة Qbasic .

الخطوة الثانية :- تحويل المتغيرين في الخطوة الاولى الى متغيرين عشوائيين مستقلين Z1 و Z2 يتبعان التوزيع الطبيعي القياسي وفقا للصيغة الاتية :

$$Z1 = SQR(-2 \ln U1) \sin(2\pi U2) \qquad (20)$$

$$Z2 = SQR(-2 \ln U1) \cos(2\pi U2) \qquad (21)$$

6-الاستنتاجات والتوصيات

6-1 الاستنتاجات Conclusions

في ضوء تحليل نتائج تجارب المحاكاة ، تم التوصل الى الاستنتاجات الاتية:

1. يلاحظ وجود تكافؤ بين طريقة OLS وطريقة Huber عند عدم وجود قيم شاذة وخاصة عندما يكون عدد المعالم قليلاً.
2. تعتبر الطريقة المقترحة الاولى والطريقة المقترحة الثانية طريقتان يمكن اعتمادهما لانهما امتازتا بمايلي:

a. حققت MSE اقل من طريقة L1-norm و Huber حتى في حالة عدم احتواء البيانات على قيم شاذة وذلك لان التشذيب فيها سيتم مهما كانت درجة الشذوذ بالقيم ضئيلة وخاصة في الطريقة المقترحة الاولى فأنه سيستمر مع كل دورة تكرارية تحدث في تقدير المعلمه .

b. أن عدد الدورات التكرارية التي تحدث للوصول الى قيمة المعلمة المقدرة أقل عند استخدام الطريقة المقترحة الاولى واكثر بقليل عند استخدام الطريقة المقترحة الثانية ، من التي تحدث عند استخدام الطرق المدروسة المتبقية.

c. **أن** الحكم على شذوذ اي قيمة يعتمد على مدى تجانسها مع بقية قيم العينة من حيث مقدار القيمة ومن حيث توزيعها وعليه فأن تحديد ثابت التشذيب (K) بشكل مطلق كما يحدث في طريقة Huber لا يمد بصلة الى طبيعة قيم العينة وهذا عكس ما يحدث في الطريقتين المقترحتين حيث يؤخذ الثابت K من العينة نفسها المراد تشذيب شواذها مما يضمن أولا حدوث عملية التشذيب وثانيا عدم المساس بالقيم غير الشاذة في العينة (وخاصة في الطريقة المقترحة الثانية).

3. تبرز طريقة L1-norm عندما يكون حجم العينة ومقدار التلوث كبيرين .
4. تكون الطرق المقترحة افضل من بقية الطرق المدروسة عندما يكون التلوث من جانبين .
5. من متابعة عمل البرنامج في التقدير يمكن تحديد القيم الشاذة في العينة إن وجدت وخاصة عند استخدام طريقة Huber .

المصادر References :-

1 Drapper N.R. and Smith (1981) ."**Applied Regression Analysis** "John Wilely and Sons .

2 Harter,H.M.,Moore,A.H.,and Curry,T.F.(1979)."**Adaptive robust estimation for location and scale parameters of symmetric population**" .Comm.Statist.8,1473-1492.

3 Hodges, J.L.Jr. and Lehmann,E.L.(1963)."**Estimation of location based on rank test**".Ann.Math.Statist.,34,598-611 .

4 Huber,P.J. (1981) ."**Robust Statistics** "Wiely,New York .

5 Jame.H et al,(2000)"**Simulation modeling methods** "McGrawhill .

6 Janne Heikkila . (1981) ."**Robust regression** " .jth@ee.oulu.fi .

7 Lenth ,R.V. (1981) ."**Robust measures of location for Directional Data** ".Technometrics ,23,77-81 .

8 Muller,C.(2002)."**Robust estimation of discontinuous function** " . http://www.search for science .com.

9 Rousseeuw ,P.J. and Croux,C. (1993) ."**Altelnatives to the Median Absolute deviation** ".J.Amer.Statist.Assn.,88,1273-1283 .

10 Sarhan ,A.J. and Greenberg,B.G.(1958)."**Estimation of Location and Scale parameters by order statistics from Singly and Doubly Censored Sample** ".Part 2 ,Ann,Math,Statist.,29,79-105 .

11 Shawn Patrick Neugebauer ,(1996) ."**Robust Analysis of M-Estimators of nonlinear models** "Thesis submitted to the faculty of the Virinia Polytechnic Institute and state University in Partial fulfillment of the requirements for the degree of MASTER OF SCIENCE in Electrical Engineering .

12 Welsh ,A.H. and Morrison ,H.L. (1990) ."**Robust L-Estimation of scale with an application in astronomy** " .J. Amer,Statist. Assn.,85,729-743 .