



كلية الدراسات العليا

جامعة السودان للعلوم و التكنولوجيا

كلية الدراسات العليا



بحث مقدم لنيل درجة الدكتوراة في الإحصاء بعنوان:

طرق إزالة أثر القيم المتطرفة على نموذج
الانحدار الخطي

**Ways to Eliminate the Outliers Effect on the Linear
Regression Model**

إشراف

د/ أحمد محمد عبدالله حمدي

إعداد الدارس

حامد حسين على حمد

المشرف المعاون

د. امل السر الخضر عبد الرحيم

يوليو 2019م



صفحة الموافقة

اسم الباحث : حامد حسن محمد

عنوان البحث : طريقة ازالة اثر القيم المتطرفة على نموذج الانحدار الخطي

Ways to Eliminate the Outliers Effect on the Linear Regression Models

موافق عليه من قبل :

الممتحن الخارجي

الاسم : د. محمد عبد الرحمن

التوقيع : [Signature] التاريخ : 03/11/19

الممتحن الداخلي

الاسم : خالد ركلة الله حسن قساوي

التوقيع : [Signature] التاريخ : 2019/8/11 م

المشرف

الاسم : د. محمد عبد الرحمن (المشرف المساعد)

التوقيع : [Signature] التاريخ : 2019/8/11 م

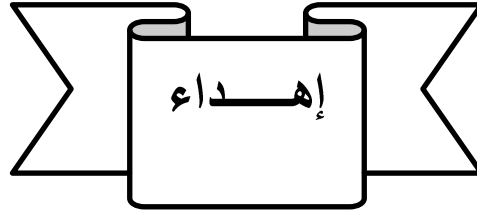
الآية

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

{ وَيَسْأَلُونَكَ عَنِ الرُّوحِ قُلِ الرُّوحُ مِنْ أَمْرِ رَبِّي وَمَا أُوتِيتُمْ مِنَ الْعِلْمِ إِلَّا قَلِيلًا }

صدق الله العظيم

{ سورة الإسراء: الآية 85 }



إلى وطني الحبيب

إلى روح أمي و روح أبي

الى روح استاذي البروفسور / احمد محمد عبد الله حمدي

لهم الرحمة جميعاً

إلى كل من ساندني وشد على يدي

اهدي جهدي المتواضع

الباحث

الشكر و التقدير

الحمد لله حمداً كثيراً طيباً غير مكفي ولا مستغنى عنه والصلاة والسلام على نبينا محمد وعلى آله وصحبه وسلموبعد.

فالشكر لله الذي من علينا بعظيم فضله وأجل نعمه، حيث هدانا للعلم وبلغنا مناهله، ومن ثم فإن وافر شكري وكثير امتناني أقدمه إلى من مد لي يده داعماً جهودي.

وأخص بالشكر الدكتور / احمد محمد عبد الله حمدي الذي تفضل بالإشراف على هذا البحث، و الدكتور / امل السر الخضر عبد الرحيم المشرف المعاون، و التي منحتني جهداً و وقتاً، ولم تدخر وسعاً في توجيهي ومتابعتي. كما أتقدم ببالغ الشكر والامتنان إلى أساتذة قسم الإحصاء بجامعة السودان للعلوم و التكنولوجيا و اخص منهم بالشكر الدكتور / خالد رحمة الله الذي لم يبخل بنصائحه رداً على استشاراتي الملحة و المتكررة.

كما أتقدم ببالغ الشكر وكثير الامتنان إلى كل من أعانني وقدم دعماً أو تسهيلات لهذه الدراسة وأخص بالشكر الزملاء بجامعة السودان التقنية الدكتور/ مالك احمد محمد الشيخ و الدكتور / صلاح الدين الخليفة لتقتهم بي و تعباً استمارات التذكية، و الدكتور / الشاذلي احمد عمر لما قدمه من تسهيلات جمة.

كما أتقدم ببالغ الشكر للجهات التي ساهمت في توفير بيانات الدراسة و اخص منها بالشكر مراكز البحوث الزراعية، و الجهاز المركزي للإحصاء، و إدارة التخطيط التجاري و الاقتصادي، و هيئة الأرصاد الجوية السودانية، و جامعة السودان التقنية.

الباحث

المستخلص

تناول هذا البحث دراسة إزالة تأثير المشاهدات الشاذة على نتائج توفيق نموذج الانحدار الخطي. و يهدف الى التوصل لطرق تشخيص و معالجة يُعتمد عليها لإزالة تأثير المشاهدات الشاذة تتسم بالموضوعية و الكفاءة و سهولة التطبيق. حيث تم اقتراح طرق جديدة تتضمن طريقة لاكتشاف المشاهدات الشاذة مصحوبة بمقياس بياني لتحديد عدد المشاهدات الشاذة في النموذج، و طريقة لمعالجة القيم الشاذة باستبدالها بقيم جديدة. تم تطبيق هذه الطرق على 35 نموذج انحدار خطي بسيط، و 31 نموذج انحدار خطي متعدد، من اجل اختبار الافتراضات المتعلقة بالكفاءة النسبية للطرق المقترحة و موضوعية هذه الطرق في إزالة اثر المشاهدات الشاذة على نتائج تحليل الانحدار، و الافتراض المتعلق بتقدير عدد المشاهدات الشاذة لنموذج الانحدار الخطي. واعتمد هذا البحث المنهج الاستدلالي التحليلي في تقدير متوسط الكفاءة النسبية و مقارنة متوسطات المؤشرات التي تم التوصل اليها من خلال التطبيق العملي باستخدام برنامج التحليل الاحصائي في توفيق النماذج ، و برنامج الجداول الالكترونية لحساب الصيغ الرياضية للطرق المقترحة. و من اهم نتائج هذا البحث : هو ان الطرق المقترحة أحرزت كفاءة عالية من حيث التشخيص و المعالجة و ذلك من خلال التحسن الكبير في مؤشرات نموذج الانحدار الخطي بارتفاع قيم F و R^2 و انخفاض قيمة MSE بالإضافة الي ان هذه الطرق أظهرت كفاءة و موضوعية اكثر من الطرق السابقة، و من اهم توصيات هذه الدراسة: ضرورة البحث في إمكانية تضمين هذه الطرق المقترحة في النظام الاحصائي SPSS في الجزء المتعلق بأسلوب تحليل الانحدار الخطي.

ABSTRACT

The research dealt with the elimination of the outliers effect on results of fitting the linear regression models. Also the research aims to arrive at a reliable detecting and treating ways which described as competency, unbiased, and usability. So a proposed new ways, which included a way for detecting outliers observation followed by a graphical measure so as to determine the outliers size, and a way to treat the outliers values substituted them with new ones. These ways has been applied on 35 simple linear regression models, and 31 multi linear regression models for testing the hypothesis concerning in the proficiency ratio and the biasness naught of the proposed ways in eliminating the outliers effect on results of fitting the linear regression models, and the hypothesis concerning of estimating the outliers numbers in the linear regression model. The research depended on the inference method to estimate the rational proficiency means, and the comparing method for the parameters means that witch obtained from among the practice of model fitted by SPSS, and the Microsoft Exile Program in calculating the mathematical functions of the proposed ways. The important results in this research showed the proposed ways netted a high proficiency in detecting and treating outliers, noted the enhancement of the linear model parameters as R^2 & F value raised and MSE value diminishes, with addition to that proposed ways appears more efficiency and more unbiased than the other previous ways. The important recommendation of this research is to see if it is available to include the proposed ways in SPSS program to be as an updating for the part concerning in the linear regression analysis method.

فهرست المحتويات

رقم الصفحة	المحتوى	الرقم
ج	الآية	
د	إهداء	
هـ	شكر وتقدير	
و	مستخلص	
ز	Abstract	
ك	فهرس الجداول	
ل	فهرس الأشكال	
الفصل الأول : المقدمة		
14	تمهيد	0-1
14	مشكلة البحث	1-1
15	أهمية البحث	2-1
15	أهداف البحث	3-1
16	مجتمع و عينة البحث	4-1
17	فرضيات البحث	5-1
17	منهجية البحث	6-1
18	الدراسات السابقة	7-1
30	التعليق على الدراسات السابقة	8-1
35	هيكلية البحث	9-1
الفصل الثاني : المشاهدات الشاذة في نموذج تحليل الانحدار الخطي		
38	تمهيد	0-2
38	الانحدار الخطي	1-2
39	مفهوم القيم الشاذة	2-2
40	اسباب ظهور القيم الشاذة	3-2
41	طرق الكشف عن القيم الشاذة	4-2

50	المشاهدات الشاذة المؤثرة و طرق الكشف عنها	5-2
56	بعض الحلول المقترحة لمعالجة القيم الشاذة	6-2
58	بعض الحلول المقترحة لمعالجة القيم الشاذة	7-2
الفصل الثالث: الطرق المقترحة لإزالة اثر المشاهدات الشاذة		
63	تمهيد	0-3
63	الطريقة المقترحة لتشخيص القيم الشاذة	1-3
64	المقياس البياني لتحديد عدد المشاهدات الشاذة	2-3
66	الطريقة المقترحة لمعالجة المشاهدات الشاذة	3-3
الفصل الرابع: التطبيق العملي		
73	تمهيد	0-4
73	التطبيق العملي بغرض المقارنة	1-4
75	التطبيق العملي علي نماذج الانحدار الخطي البسيط	2-4
85	اختبارات الفروض فيما يتعلق بنماذج الانحدار الخطي البسيط	3-4
91	التطبيق العملي علي نماذج الانحدار الخطي المتعدد	4-4
100	اختبارات الفروض فيما يتعلق بنماذج الانحدار الخطي المتعدد	5-4
الفصل الخامس: النتائج و التوصيات		
108	النتائج	1-5
120	التوصيات	2-5
121	المراجع	

فهرست الجداول

الرقم الصفحة	اسم الجدول	الرقم
70	نتائج توفيق بيانات نموذج الانحدار الخطي بالنسبة للاشكال الثلاث (Scatter Plots)	1 - 3
جداول التطبيق العملي علي نماذج الانحدار الخطي البسيط		
74	نتائج المقارنة لقيم المؤشرات و الكفاءة النسبية للطرق	1 - 4
76	قيم المؤشرات $(R^2; F; t_{\beta_0}; t_{\beta_1}; MSE)$ للبيانات الاصلية	2 - 4
78	قيم المؤشرات $(R^2; F; t_{\beta_0}; t_{\beta_1}; MSE)$ للبيانات بعد التشخيص و الحذف	3 - 4
79	قيم المؤشرات $(R^2; F; t_{\beta_0}; t_{\beta_1}; MSE)$ للبيانات بعد التشخيص و المعالجة	4 - 4
82	قيم مؤشرات الكفاءة النسبية $(R^2; F; t_{\beta_0}; t_{\beta_1}; MSE)$ للبينات بعد التشخيص و المعالجة الى البيانات الاصلية	5 - 4
83	قيم عدد المشاهدات الكلي n و عدد المشاهدات الشاذة m و ربع عدد المشاهدات للنموذج	6 - 4
85	نتائج اختبار الفروق بين متوسطات المؤشرات للبيانات الاصلية و للبيانات بعد حذف الشواذ	7 - 4
86	نتائج اختبار الفروق بين متوسطات المؤشرات للبيانات الاصلية و للبيانات بعد معالجة الشواذ	8 - 4
88	نتائج اختبار الفروق بين متوسطات المؤشرات للبيانات بعد حذف الشواذ و للبيانات بعد معالجة الشواذ	9 - 4
89	نتائج اختبار متوسطات الكفاءة النسبية للمؤشرات	10 - 4
89	نتائج اختبار اختلاف متوسطات الكفاءة النسبية للمؤشرات عن الواحد الصحيح	11 - 4
90	نتائج اختبار الفروق بين متوسطات القيم التي تمثل عدد	12 - 4

	المشاهدات الشاذة (m) و متوسطات القيم التي تمثل ربع عدد المشاهدات	
جداول التطبيق العملي علي نماذج الانحدار الخطي المتعدد		
92	قيم المؤشرات ($R^2; F; t_{\beta_0}, t_{\beta_1}, t_{\beta_2}, t_{\beta_3}; MSE$) للبيانات الاصلية	13 - 4
93	قيم المؤشرات ($R^2; F; t_{\beta_0}, t_{\beta_1}, t_{\beta_2}, t_{\beta_3}; MSE$) للبيانات بعد التشخيص و الحذف	14 - 4
95	قيم المؤشرات ($R^2; F; t_{\beta_0}, t_{\beta_1}, t_{\beta_2}, t_{\beta_3}; MSE$) للبيانات بعد التشخيص و المعالجة	15 - 4
97	قيم مؤشرات الكفاءة النسبية ($R^2; F; t_{\beta_0}, t_{\beta_1}, t_{\beta_2}, t_{\beta_3}; MSE$) للبيانات بعد التشخيص و المعالجة الي البيانات الاصلية	16 - 4
99	قيم عدد المشاهدات الكلي n و عدد المشاهدات الشاذة m و ربع عدد المشاهدات - (عدد المتغيرات - 2) للنموذج	17 - 4
100	نتائج اختبار الفروق بين متوسطات المؤشرات للبيانات الاصلية و للبيانات بعد حذف الشواذ	18 - 4
101	نتائج اختبار الفروق بين متوسطات المؤشرات للبيانات الاصلية و للبيانات بعد معالجة الشواذ	19 - 4
103	نتائج اختبار الفروق بين متوسطات المؤشرات للبيانات بعد حذف الشواذ و للبيانات بعد معالجة الشواذ	20 - 4
104	نتائج اختبار متوسطات الكفاءة النسبية للمؤشرات	21 - 4
104	نتائج اختبار اختلاف متوسطات الكفاءة النسبية للمؤشرات عن الواحد الصحيح	22 - 4
106	نتائج اختبار الفروق بين متوسطات القيم التي تمثل عدد المشاهدات الشاذة (m) و متوسطات القيم التي تمثل ربع عدد المشاهدات - (عدد المتغيرات - 2) للنموذج	23 - 4

فهرست الاشكال

رقم الصفحة	اسم الشكل	الرقم
42	مخطط الرسم الصندوقي	1 - 2
65	تصميم المقياس البياني	1 - 3
68	انتشار البيانات في حال ظهور مشاهدة شاذة ذات تأثير جامح سلبي	2 - 3
69	انتشار البيانات في حال ظهور مشاهدة شاذة ذات تأثير جامح ايجابي	3 - 3
70	شكل انتشار البيانات بعد إعطاء القيم ذات الشذوذ الجامح قيم تقريبية	4 - 3

الفصل الأول

المقدمة

0 -1 تمهيد:

1-1 مشكلة البحث:

2 -1 أهمية البحث:

3 -1 اهداف البحث:

4 -1 مجتمع و عينة البحث:

5 -1 فرضيات البحث:

6 -1 منهجية البحث:

7 -1 الدراسات السابقة:

8 -1 التعليق على الدراسات السابقة:

9 -1 هيكلية البحث:

1-0 تمهيد:

إن نموذج تحليل الانحدار من الاساليب الواسعة الانتشار و الاستخدام ، حيث انه يصف العلاقة بين المتغيرات المستقلة و المتغير التابع علي هيئة معادلة يستدل من تقدير معالمها علي قوة و اتجاه هذه العلاقة الامر الذي يبين تقدير قوة الاستجابة و التنبؤ بها بما يفيد كثيراً في عمليات التخطيط و اتخاذ القرارات حولها ، لذا فان نجاح و فشل النموذج المقدر يعتمد علي طريقة تقدير معالم النموذج . و ان أوسع الطرق انتشاراً في تقدير معالم نموذج الانحدار الخطي هي طريقة المربعات الصغرى لما لها من خواص و مزايا جيدة تميزت بها مقدراتها ، الا ان هذه الطريقة تكون غير كفوءة في حالة عدم تحقق احد افتراضاتها الأساسية او بسبب وجود القيم الشاذة التي تؤثر في مقدراتها و تجعلها غير دقيقة و بعيدة عن الواقع . عليه فان سلامة البيانات و تنقيتها من القيم الشاذة مهمة ضرورية لضمان سلامة النتائج و صحة الاستنتاجات المبنية عليها ، نجد البدايات الأولى لموضوع القيم الشاذة كانت تتعامل مع هذا النوع من القيم بحذفها و اجراء التحليل علي بقية المشاهدات الامر الذي يتسبب في نقص المعلومات خصوصا في حالة قلة العدد الكلي للبيانات او ان القيم الشاذة عددها كبير او مقارب لنصف عدد البيانات الكلي. لذا من الأفضل تقدير تلك القيم لإزالة تأثيرها على مقدرات نموذج الانحدار الخطي، الامر الذي يستدعي دراسة القيم الشاذة من حيث التشخيص و المعالجة .

1-1 مشكلة البحث:

تتمثل مشكلة طرق إزالة اثر المشاهدات الشاذة على نموذج الانحدار الخطي في محورين أساسيين، المحور الأول هو طرق التشخيص الدقيق و السليم للمشاهدات الشاذة، اذ ان هناك بعض طرق التشخيص تشخص جزء من المشاهدات الشاذة و تترك البعض من دون تشخيص، وكذلك هنالك طرق تشخيص تشخص مشاهدات غير شاذة على أساس انها شاذة و هي ليست كذلك،

المحور الثاني هو طرق المعالجة الصحيحة أي عملية تقدير قيم جديدة للملاحظات الشاذة بالوضع الذي يحد من تأثيرها السلبي على النموذج. اضع الى ذلك صعوبة تطبيق طرق التشخيص و المعالجة بالنسبة للنموذج الخطي خصوصا عند تطبيق طريقة المربعات الصغرى في توفيق النماذج.

1-2 أهمية البحث:

تاتي أهمية هذا البحث في الآتي:

- 1/ اقتراح الباحث طرق جديدة لتشخيص و معالجة مشكلة القيم الشاذة.
- 2/ مناقشة مشكلة المشاهدات الشاذة في نماذج الانحدار الخطي (البسيط و المتعدد) من خلال تناول طرق التشخيص و المعالجة بالنسبة للدراسات السابقة.
- 3/ مقارنة الطرق المقترحة للتشخيص و المعالجة الجديدة بطرق التشخيص و المعالجة في الدراسات السابقة التي أظهرت درجة من الكفاءة النسبية في معالجة مشكلة المشاهدات الشاذة.
- 4/ الرغبة الملحة للباحث في اختبار الطرق المقترحة من حيث متوسط الكفاءة النسبية، و مدى فاعلية و دقت هذه الطرق في حل مشكلة المشاهدات الشاذة.

1-3 أهداف البحث:

يهدف هذا البحث الي الآتي :-

- 1/ التأكد من وجود المشاهدات الشاذة و تحديد العدد المتوقع لها في نماذج الانحدار الخطي من خلال تشخيصها باستخدام الطرق المقترحة الجديدة و إمكانية قياس درجة تأثيرها على مؤشرات نموذج الانحدار الخطي عن طريق معالجتها بطريقة المعالجة المقترحة و ملاحظة مقدار التغير الذي يطرأ علي نتائج النموذج.

2/ قياس الكفاءة النسبية للطرق المقترحة الجديدة مقارنة بالطرق السابقة من حيث تشخيص المشاهدات الشاذة و معالجتها.

3/ اثبات موضوعية و كفاءة الطرق المقترحة لازالة اثر المشاهدات الشاذة علي نموذج الانحدار الخطي من حيث التشخيص و المعالجة و سهولة تطبيقها و مقارنة نتائج النماذج الموقفة بطريقة المربعات الصغري لثلاث أوضاع الوضع الأول نتائج توفيق البيانات الاصلية، و الوضع الثاني نتائج توفيق البيانات بعد حذف المشاهدات الشاذة، و الوضع الثالث نتائج توفيق البيانات بعد معالجة المشاهدات الشاذة.

1-4 مجتمع و عينة البحث:

يتمثل مجتمع البحث من جميع بيانات نماذج الانحدار الخطي البسيط و المتعدد، حيث تم تصنيف عدد منها باحجام مختلفة و عدد مختلف من حيث المتغيرات المستقلة، بالإضافة الى نماذج انحدار خطي مصنفة مسبقاً، يتراوح عدد المتغيرات فيها من اثنان الي اربع متغيرات، باعتبارهما عينتان لتمثيل نماذج الانحدار الخطي البسيط و المتعدد.

تم اخذ معظم هذه النماذج من صفحة (Github, 2018) على الانترنت متاحة على الموقع الالكتروني (<https://vincentarelbundock.github.io/Rdatasets/>) و منها على سبيل المثال بيانات (Mickey, et al 1967) التي استخدمها العديد من الباحثين لنفس الغرض، و البعض الآخر من أماكن إقليمية و محلية متفرقة، مثل مراكز البحوث الزراعية و و الجهاز المركزي للإحصاء و إدارة التخطيط التجاري و الاقتصادي و بعض كليات جامعة السودان التقنية مثل كلية فارس التقنية و كلية الجريف شرق التقنية، و بعضها تم تصنيفه بواسطة الباحث من البيانات الخام لتلك الجهات و البعض تم جمعه و تصنيفه بواسطة الباحث مثل بيانات استهلاك الكهرباء في المنزل و ذلك من خلال مراغبة الاستهلاك للكهرباء (متغير تابع) من عداد الدفع

المقدم لمنزل الباحث مباشرة خلال فترات زمنية غير متساوية تم تحديدها بطريقة عشوائية (متغير مستقل).

5-1 فرضيات البحث:

تأتي الافتراضات لاختبار مدى فاعلية و كفاءة الطرق المقترحة في تشخيص و معالجة المشاهدات الشاذة و ذلك من خلال مقارنة متوسطات مؤشرات النماذج الخطية البسيطة بعد توفيقها بطريقة المربعات الصغرى وفق التسلسل التالي :

1/ توجد فروق ذات دلالة معنوية بين مؤشرات النماذج للبيانات الاصلية و لنفس البيانات بعد حذف المشاهدات الشاذة.

2/ توجد فروق ذات دلالة معنوية بين مؤشرات النماذج للبيانات الاصلية و لنفس البيانات بعد معالجة المشاهدات الشاذة.

3/ لا توجد فروق ذات دلالة معنوية بين مؤشرات النماذج للبيانات بعد حذف المشاهدات الشاذة و لنفس البيانات بعد معالجة المشاهدات الشاذة.

4/ متوسط الكفاءة النسبية لمؤشرات البيانات بعد معالجة المشاهدات الشاذة الى مؤشرات البيانات الاصلية اكبر من الواحد الصحيح.

5/ يقدر عدد المشاهدات الشاذة لنموذج الانحدار بربع عدد المشاهدات الكلي للنموذج - (عدد متغيرات النموذج - 2).

6-1 منهجية البحث:

يستند هذا البحث الي المنهج التحليلي الاستدلالي القائم على اختبار الفروض بالنسبة للبيانات التي يتم تحليلها باسلوب تحليل الانحدار الخطي البسيط أو المتعدد فيما يتعلق بتشخيص و معالجة المشاهدات الشاذة بقية تعميم النتائج على جميع نماذج الانحدار الخطي، بالإضافة لاسلوب المقارنة

من حيث الكفاءة النسبية للطرق المقترحة لتشخيص و معالجة القيم الشاذة بالطرق الاخرى، و بعض أساليب المقارنة لمتوسطات المؤشرات المتحصل عليها من نتائج التحليل للأوضاع المختلفة بأسلوب اختبارات الفروض للفروق بين متوسطات المؤشرات.

تمت معالجة بيانات هذا البحث بالطرق الإحصائية المناسبة بواسطة الحاسوب باستخدام برنامج التحليل الاحصائي SPSS، و بالطرق الرياضية باستخدام صيغ رياضية بواسطة الحاسوب من خلال استخدام برنامج الجداول الالكترونية Excel، بالإضافة الى استخدام ورق الرسم البياني.

1-7 الدراسات السابقة:

هنالك العديد من الدراسات و البحوث السابقة التي تناولت موضوع طرق الكشف عن المشاهدات الشاذة و معالجتها و بيان مدى تأثيرها علي نتائج تحليل نماذج الانحدار الخطي ، و من الأهمية بمكان الوقوف على الدراسات و البحوث التي أجريت في هذا المجال للاسترشاد بنتائجها و الاخذ بتوصياتها و مقارنتها من حيث الكفاءة النسبية بنتائج الدراسة الحالية، و ملاحظة الى أي مدى تتفق او تختلف هذه الدراسات مع بعضها البعض و مع الدراسة الحالية. و فيما يلي بعض الدراسات و البحوث العلمية السابقة التي لها علاقة اصيلة بموضوع الدراسة الحالية و تأثير القيم الشاذة على نتائج تحليل الانحدار الخطي البسيط و المتعدد:

1-7-1 دراسة (الجبوري، وناسي، 2002):

عنوان الدراسة: اكتشاف وتقدير المشاهدات الشاذة باستخدام معادلة القطع الناقص Ellipse في حالة الانحدار الخطي البسيط:

هدفت هذه الدراسة الي اقتراح طريقة جديدة لاكتشاف وتقدير المشاهدات الشاذة باستخدام معادلة القطع الناقص Ellipse الذي مركزه متوسطي المتغيرين X و Y وطول محوريه يمكن

الحصول عليهما بإجراء بعض التحويلات المثلثية، وعليه فإن أية نقطة تقع خارج القطع الناقص تعتبر مشاهدة شاذة. واقترحت ثمانية طرق لإيجاد تقديرات لهذه المشاهدات بعضها اعتمد على إيجاد نقطة تقاطع المستقيم الواصل بين كل من النقطة الشاذة ومركز القطع مع معادلة القطع الناقص، وبعضها الآخر اعتمد على إضافة أو طرح كل من التباين أو التباين المشترك مضروباً بقيمة الطرف الأيسر للقطع الناقص.

و كتطبيق عملي استخدمت بيانات (Mickey, et al 1967). حيث أعطت الطرق المقترحة للاكتشاف و المعالجة نتائج جيدة، إذ رفعت قيمة معامل التحديد R^2 و خفضت متوسط مربعات الخطأ MSE. و توصلت الدراسة الى نتائج أهمها: أن جميع طرق التقدير لم تغير من شكل انتشار النقاط سوى سحب نقطة أو نقطتين نحو مركز البيانات (X,Y) باستثناء الطريقة الأولى NSR-1 التي سحبت سبعة مشاهدات نحو مركز البيانات، أن الطرق أبدت كفاءة نسبية معقولة من حيث تقليل قيمة MSE، إلا أنها لم تؤدي إلى رفع قيمة R^2 بنفس النسبة، من ناحية أخرى طريقة الكشف عن المشاهدات الشاذة لم تحدد عدد معين للمشاهدات الشاذة، فتارة تكون المشاهدات الشاذة سبعة وتارة اثنتين أو واحدة في حين أن التطبيق العملي للدراسة كان على عينة واحدة. اوصت الدراسة بضرورة البحث من اجل التوصل لطرق ذات كفاءة افضل من حيث تشخيص و معالجة القيم الشاذة.

1-7-2 دراسة (حمد، 2005) :

عنوان الدراسة: طرق بديلة لكشف و تقليل اثر المشاهدات الشاذة على نموذج الانحدار الخطي البسيط:

هدفت هذه الدراسة الي التوصل لطريقة جديدة يمكن ان يعتمد عليها الباحثين في الكشف عن المشاهدات الشاذة و تقليل اثرها علي نموذج الانحدار الخطي البسيط. و توصلت الدراسة الي

اقترح طريقة جديدة لاكتشاف و تقليل اثر المشاهدات الشاذة في بيانات نموذج الانحدار الخطي البسيط باستخدام الصيغة الرياضية $Y_i - \hat{\beta}_1 X_i = Z_i$ من خلال طريقتين، طريقة تقريبية يتم فيها إعطاء رتب Rank لقيم المتغير Z_i و يتم تحديد المشاهدات الشاذة بهذه الطريقة بتمثيل النقاط $(Rank_i, Z_i)$ بيانيا الامر الذي يجعل النقاط تنتشر علي خط مستقيم تتحرف منه نقاط المشاهدات الشاذة عند اطرافه بوضوح يتيح للباحث تحديدها بسهولة. و طريقة أخرى شمولية مصحوبة بمقياس بياني - مقترح - و يتم تحديد المشاهدات الشاذة بهذه الطريقة بتطبيق المعادلة $Y_i - \hat{\beta}_1 X_i = Z_i$ بعد توفيق النموذج و حساب قيمة Z_i و ترتيبها تصاعديا او تنازليا و من ثم حذف المشاهدة التي تناظر الانحراف المطلق الأكبر ل Z_i عن $\hat{\beta}_0$ باعتبارها مشاهدة شاذة و من ثم توفيق النموذج مرة أخرى ببقية المشاهدات و تكرار نفس العملية الي ان يحدد المقياس البياني عدد المشاهدات الشاذة. و لتقليل اثر المشاهدات الشاذة المكتشفة بأي من الطريقتين اقترح الباحث ترجيح انحرافات Z_i عن $\hat{\beta}_0$ المناظرة للمشاهدة الشاذة بالقيمة الكسرية $p/(1-p)$ حيث p هو احتمال المشاهدات الشاذة و ذلك للحصول قيم Z_i^* جديدة تناظرها قيم (X_i^*, Y_i^*) محسوبة رياضياً تعتبر بمثابة قيم بديلة للمشاهدات الشاذة. و كتطبيق عملي استخدمت بيانات (Mickey, et al) (1967)، و توصلت الدراسة الي نتائج أهمها: ان الطريقة التقريبية كشفت عن جزء من المشاهدات الشاذة بينما كشفت الطريقة الشمولية عن العدد الكلي للمشاهدات الشاذة، حيث تمت عملية الكشف من المشاهدة ذات الأثر الأكبر و التي تليها و التي تليها الي المشاهدة ذات الأثر الأقل. كان لترتيب المشاهدات وفق ترتيب Z_i و ضرب d_i (انحرافات Z_i عن $\hat{\beta}_0$) المناظرة للمشاهدات الشاذة بالقيمة $p/(1-p)$ الأثر الأكبر في تقليل اثر المشاهدات الشاذة و ذلك من غير تحيز لان القيمة $p/(1-p)$ تتناسب طردياً مع عدد المشاهدات الشاذة المكتشفة. اوصت الدراسة بالبحث في إمكانية تعميم هذه الطرق لتشمل نموذج الانحدار الخطي المتعدد.

1- 7 - 3 دراسة (مروان، و فرح، 2006) :

عنوان الدراسة: تأثير القيم الشاذة على نتائج تحليل الانحدار مع التطبيق على المواليد الخدج:

هدفت هذه الدراسة الى الكشف عن القيم الشاذة التي تشير اليها نتائج تحليل الانحدار و الرسم الصندوقي بالإضافة الى تأثير وجود و معالجة القيم الشاذة على نتائج تحليل الانحدار و توصلت الدراسة الى عدد من النتائج أهمها: تؤثر القيم الشاذة على تركيبة النموذج و نتائج تحليل الانحدار من خلال ادخال متغيرات غير مؤثرة على الظاهرة المدروسة و استبعاد بعض المتغيرات المؤثرة. لا يمكن التنبؤ بتأثير القيم الشاذة على معامل التحديد او القيم التقديرية لتباين المجتمع، أي تكون قيمتها غير حقيقية في حالة وجود القيم الشاذة. القيم الشاذة التي يشير اليها تحليل الانحدار تؤثر على نتائجه و التي يكشفها الرسم الصندوقي فقط لا تؤثر على نتائج التحليل. و يشير تحليل الانحدار الى نوعين من القيم الشاذة احدهما قيم متطرفة و هي التي يجب معالجتها و هذا النوع ايضاً يشير اليه الرسم الصندوقي و النوع الآخر شذوذ البواقي المعيارية التي يمكن التغلب عليها من خلال معالجة قيم أخرى او تغيير في تركيبة النموذج. اوصت الدراسة باستخدام طرق أخرى للكشف عن المشاهدات الشاذة في دراسات مستقبلية.

1- 7 - 4 دراسة (احمد، 2007):

عنوان الدراسة: المشاهدات غير المألوفة في نموذج الانحدار:

هدفت هذه الدراسة الى دراسة تشخيص المشاهدات غير المألوفة في نموذج الانحدار و مدى تأثيرها على معاملاته. حيث تم تشخيص تم تشخيص المشاهدات الشاذة في المتغيرات المستقلة باستخدام مصفوفة القبة، و مشاهدات المتغير التابع الشاذة باستخدام بواقي t المعيارية (Studentized Residuals) ، اما المشاهدات المؤثرة فتم الكشف عنها بتطبيق صيغة مسافة كوك (Cook's Distance). و توصلت الدراسة الي ان معاملات الانحدار تتأثر بالمشاهدات

غير المؤلفوة سواء في المتغيرات المستقلة او المتغير التابع و يختلف تأثيرها حسب موقعها من خط الانحدار بين مؤثرة و غير مؤثرة، بالإضافة الى ان بعض المشاهدات غير المؤلفوة او الشاذة تكون غير مؤثرة سلباً على معاملات النموذج حيث يلغي تأثير بعض المشاهدات البعض الآخر. اوصت الدراسة باستخدام طرق أخرى للكشف عن المشاهدات الشاذة في دراسات مستقبلية.

1- 7 - 5 دراسة (قاسم، و اسماعيل 2008) :

عنوان الدراسة: الكشف عن القيم الشاذة بأسلوب بيز معاينة جيس:

هدفت هذه الدراسة الى دراسة القيم الشاذة و الكشف عنها بطريقة بيز في نموذج تحليل الانحدار الخطي البسيط بأسلوب معاينة جيس و التي هي حالة خاصة بطريقة مونتو كارلو سلسلة ماركوف و تطبيق هذا الأسلوب على بيانات حقيقية تم الحصول عليها من احدى البحوث المنشورة لاجل الكشف عن القيم الشاذة. و توصلت الدراسة الى عدد من النتائج أهمها: ان الكشف عن القيم الشاذة بأسلوب بيز باستخدام معاينة جيس يحل الكثير من المسائل المعقدة و لا سيما ذات النماذج المعقدة و الصعبة لان معاينة جيس تقسم المسائل و النماذج المعقدة الى مجموعة مسائل بسيطة يمكن تحليلها ومعالجتها بسهولة كما ان النتائج تكون دقيقة وذلك لان هذه الطريقة تعتمد على المعلومات الأولية للمعلمات والتي غالباً ماتكون متوفرة في معظم الدراسات، كما ان هذه الطريقة تمكننا من الحصول على احتمالية كون المشاهدة او المشاهدات قيماً شاذة و هذا ما لم يمكن الحصول عليه في استخدام الطرائق الكلاسيكية في الكشف عن القيم الشاذة. ان معاينة جيس هي طريقة ذات كفاءة و دقيقة لانها تبدأ بقيم ابتدائية و من ثم يجري عليها عدد كبير من التكرارات كما ان خوارزمتها فائقة السرعة في التنفيذ. تعد معاينة جيس من الطرائق الإحصائية المهمة جداً في التطبيقات العلمية و النظرية لانها تعطينا التوزيع الاحتمالي للمتغيرات العشوائية و من ثم تستطيع الحصول على المؤشرات الرياضية لكل متغير كالتوقع و التباين و غيرها من المؤشرات

الرياضية الأخرى. عند تطبيق الطريقة الكلاسيكية للكشف عن القيم الشاذة على البيانات نفسها قورنت النتائج مع طريقة معاينة جيس فكانت متقاربة و مشجعة جداً. اوصت الدراسة باستخدام طرق أخرى للكشف عن المشاهدات الشاذة في دراسات مستقبلية.

1-7-6 دراسة (داود، 2009):

عنوان الدراسة: تأثير القيم الشاذة ونقطة الأصل على نتائج تحليل الانحدار:

هدفت الدراسة الى التعرف على تأثير وجود بعض القيم الشاذة على نتائج تحليل الانحدار مقارنة بمعالجتها ، حيث تم الكشف عن وجود القيم الشاذة بطريقة الرسم الصندوقي وعولجت بطريقة متوسطة البتر. وتوصلت الدراسة الى نتائج أهمها ان وجود القيم الشاذة في بعض المتغيرات أثر سلبا على نتائج تحليل لانحدار حيث انها قللت من قيمة F المحسوبة لمعادلة الانحدار التقديرية وكذلك قيمة معامل التحديد R^2 وزادت من قيمة متوسط مربعات الخطأ MSE مقارنة مع نتائج التحليل بعد معالجة القيم الشاذة. اوصت الدراسة بالاهتمام في اتخاذ القرار حول احتواء معادلة الانحدار التقديرية على β_0 من عدمه.

1-7-7 دراسة (يحي، 2009) :

عنوان الدراسة: المشاهدات غير العادية في الانحدار الخطي المتعدد و بعض طرق تشخيصها

مع التطبيق:

هدفت هذه الدراسة الى تناول طرائق تشخيص ثلاثة أنواع من المشاهدات غير العادية (Unusual Observation) في تحليل الانحدار المتعدد، و بيان مدى تأثير المشاهدات الشاذة في التوزيع الطبيعي للبواقي. حيث تم الكشف عن المشاهدات الشاذة (Outlier) باستخدام الرسم الصندوقي (Box Blot) و مقياس البواقي القياسية (Studentized Residuals) و الكشف عن المشاهدات ذات قوة الرفع (Leverage Observation) من خلال مصفوفة (Hat

(Matrix)، اما المشاهدات المؤثرة فتم الكشف عنها باستخدام مقياس (DFβETAS). و توصلت الدراسة الى ان حذف المشاهدات الشاذة أدى الى اقتراب توزيع البيانات من التماثل تماماً، و كذلك اثر حذف المشاهدات الشاذة في دقة معاملات النموذج بانخفاض متوسط مربعات الخط (MSE) و عامل تضخم التباين (VIF) و ارتفاع قيمة (F) و معامل التحديد (R²) و تغيير إشارات بعض المعلمات المقدرة و الوصول الى نموذج انحدار يمثل بيانات الدراسة خير تمثيل. اوصت الدراسة باستخدام طرق أخرى للكشف عن المشاهدات الشاذة في دراسات بهذا الخصوص يتم تناولها مستقبلاً.

1- 7 - 8 دراسة (جواد ، و كمال، 2009) :

عنوان الدراسة: مقارنة طرق تقدير معالم نموذج الانحدار في حالة ظهور مشكلة التعدد الخطي و القيم الشاذة:

هدفت هذه الدراسة الى دراسة بعض طرائق تقدير معالم الانحدار الخطي و المقارنة فيما بينها عند اختلاف توزيع الخطأ و حجم العينة و قوة العلاقة بين المتغيرات المستقلة. حيث استخدم في هذه الدراسة أسلوب المحاكاة لاختبار قوة و حصانة المقدرات لنموذج الانحدار الخطي المتعدد عند وجود مشاكل التداخل الخطي و الأخطاء الغير طبيعية و ذلك بتطبيق عدة طرق للتقدير منها الاعتيادية و الحصينة، و هي طريقة المربعات الصغرى (OLS) و انحدار الريدج (RID) و طريقة القيمة المطلقة الصغرى (RLAV) و طريقة مقدرات الريدج الموزون (WRID) و طريقة (MM) و مقدرات انحدار الريدج الحصينة (RMM) المعتمدة على مقدر (MM). و توصلت الدراسة الى انه عندما تكون درجة التداخل الخطي عالية فإن طريقة (RMM) هي الأفضل من بين طرق التقدير الأخرى، و عند المقارنة بين مقدرات الريدج الحصينة (RMM ، WRID ، RLAV) توصلت الدراسة الى ان مقدرات (RMM) هي الأفضل من المقدرات

RLAV ، WRID) و لمختلف افتراضات توزيع الخطأ و درجة التداخل الخطي المتعدد. اوصت الدراسة باستخدام طرق أخرى للكشف عن المشاهدات الشاذة في دراسات بهذا الخصوص يتم تناولها مستقبلاً.

1-7-9 دراسة (شاكر، 2009) :

عنوان الدراسة: تحسين أسلوب M الحصين في تقدير معلمات نموذج الانحدار الخطي المتعدد:
هدفت هذه الدراسة الى إيجاد تركيبة حصى موزونة لمقدرات نموذج الانحدار الخطي المتعدد بدمج أسلوب M الحصين و طريقة المربعات الصغرى الموزونة لمعالجة القيم الشاذة في بيانات نموذج تحليل الانحدار الخطي المتعدد للحصول على مقدرات جديدة تقاوم اثر القيم الشاذة على نتائج التحليل و مقارنة طرائق التقدير (OLS, WLS, RM, RMW) بالاعتماد على معيار متوسط مربعات الخطأ (MSE). حيث تم الكشف عن القيم الشاذة بطريقة الرسم الصندوقي و فحص عناصر قطر المصفوفة H كما استخدم مقياس مربع مسافة كوك لاختبار تأثير حذف المشاهدة رقم i على المعلمات المقدرة، و مقياس (DF) للكشف عن وجود القيم المتطرفة في متجه الأخطاء العشوائية. و توصلت الدراسة الى عدة نتائج أهمها: عدم وجود فروق في دقة نتائج الطرائق المدروسة عند عدم وجود القيم الشاذة في البيانات. لا يمكن الاعتماد على مقدرات أسلوب (M) الحصين في حالة وجود تطرف قوي في مشاهدات المتغير (المتغيرات) المستقلة او حتى المتغير المعتمد او في كليهما. و ايضاً اثبتت الدراسة ان أسلوب (M) الحصين الموزون (RMW) المقترح قد اثبت كفاءته في تقدير معلمات نموذج الانحدار الخطي و بدقة عالية عند وجود القيم الشاذة في قيم المتغير (المتغيرات) المستقلة او في المتغير المعتمد او في كليهما. اوصت الدراسة باستخدام طرق أخرى للكشف عن المشاهدات الشاذة في دراسات بهذا الخصوص يتم تناولها مستقبلاً.

1- 7 - 10 دراسة (حسن، 2009) :

عنوان الدراسة: استخدام المعايير الحصينة في تحديد المتغيرات المؤثرة في نموذج الانحدار الخطي لقياس السكر في الدم:

هدفت هذه الدراسة الى استخدام المعايير الحصينة في تحديد المتغيرات المهمة في نموذج الانحدار الخطي المتعدد في حالة وجود القيم الشاذة في البيانات و ذلك من خلال تحليل بيانات السكر في الدم بوصفه متغيراً معتمداً و متغيرات أخرى مستقلة، حيث تم تقدير جميع النماذج الجزئية الممكنة بطريقة (M) الحصينة، و تمت المقارنة بين هذه النماذج باستخدام المعايير الحصينة (RC, RSIC, RAIC, RAPE, RF, RSC, RVC) لتحديد المتغيرات المستقلة في النموذج الأفضل. و توصلت الدراسة الى عدد من النتائج أهمها ان افضل المعايير اداءً كان المعيار (RAPE) حيث ان اختيار هذا المعيار للنموذج الأفضل جاء متوافقاً مع رأي المختصين في مجال الطب. اوصت الدراسة باستخدام طرق أخرى للكشف عن المشاهدات الشاذة في دراسات مستقبلية.

1- 7 - 11 دراسة (حسن، و رضا، 2011):

عنوان الدراسة: استخدام الانحدار الحصين لايجاد انسب نموذج لتمثيل بيانات الانواء الجوية في مدينة أربيل خلال الفترة 1998 - 2010:

هدفت هذه الدراسة الى إيجاد انسب نموذج انحدار خطي متعدد لتمثيل بيانات الانواء الجوية بدراسة اهم العوامل المؤثرة في الرطوبة النسبية و لتحقيق هذا الهدف تم تطبيق بعض طرق الانحدار الحصين في حالة وجود المشاهدات الشاذة منها، طريقة المربعات الصغرى الاعتيادية (OLS) و تمت المقارنة بين هذه الطرق بالاعتماد على معيار متوسط مربعات الخطأ (MSE) و تم الكشف عن المشاهدات الشاذة باستخدام قطر عناصر المصفوفة H و طريقة بوكس ويسكر

بلوت (Box – and – whisker plot) البيانية، و المشاهدات المؤثرة تم الكشف عنها باستخدام المقياس (DFFITS). و توصلت الدراسة الى عدد من النتائج أهمها: أن طريقة مربعات الوسيط LMS هي الأفضل لتمثيل انصب نموذج انحدار لدراسة بيانات حالة الطقس. كما اوصت الدراسة باستخدام طرق أخرى للكشف عن المشاهدات الشاذة في دراسات مستقبلية مثل طريقة تقديرات (S,L,R).

1- 7 - 12 دراسة (الشميري، و آخرون، 2014):

عنوان الدراسة: اكتشاف القيم الشاذة وتقديرها في الانحدار الخطي بالتطبيق على بيانات معدلات النمو وفقاً لأعمار الأطفال في العام 1987:

هدفت الدراسة إلى اقتراح طريقة تقدير لمعالجة القيم الشاذة والمتطرفه في النماذج الخطيه, كما هدفت أيضاً إلى مقارنة الطريقة المقترحة مع بعض الطرق الموجوده في الدراسات السابقة. تقوم الطريقة على بعدين البعد الأول هو فحص النموذج الخطي لمعرفة وجود أو عدم وجود مشاهدات شاذة والبعد الثاني هو كيفية معالجة هذه المشاهدات فهل تحذف؟ أم تقدر؟ علي اعتقاد أن الحذف يؤدي إلى نقص في المعلومات ويؤثر على درجات الحرية وبالتالي على اختبار معلمات النموذج وعليه فإن اعتماد طريقة التقدير لهذه المعلمات بدلاً من حذفها يؤدي إلى المحافظة على المعلومات وجودتها. و لاكتشاف المشاهدات الشاذة استخدمت الدراسة طريقة الصندوق Tukey بعد تقدير حد الخطأ المطلق العشوائي $|e_i|$ و ترتيبه تصاعدياً و حذف المشاهدة الشاذة، و تعتمد الطريقة المقترحة للتقدير المشاهدات الشاذة بالدرجة الأساس على استخدام القيمة المطلقة لموجة حد الخطأ المقدر و تقسيمه الى مجموعات متجانسة مع إيجاد وزن لكل قيمة مطلقة داخل مجموعتها الامر الذي يقود في النهاية الى حساب قيمة مقدره جديدة للمشاهدة الشاذة و التي يتم اعادتها إلى أصل البيانات ثم يعاد التحليل للنموذج لإيجاد حد الخطأ المطلق العشوائي $|e_i|$ واختباره وفي حالة

وجود قيمة شاذة تعادة الطريقة السابقة وهكذا حتى يتم تخلص النموذج من القيم الشاذة. وكتطبيق عملي استخدمت بيانات (Mickey, et al 1967)، و توصلت الدراسة إلى الإستنتاجات التالية: الطريقة المقترحة تتميز بالبساطة و سهولة التطبيق، برهنت الطريقة المقترحة كفاءة جيدة بتقدير المشاهدات الشاذة، أظهرت الطريقة المقترحة حساسيتها بدرجة كافية في تقدير المشاهدات المشاذة، أظهرت الطريقة المقترحة تفوقا ملحوظا عند مقارنتها مع طريقة المربعات الصغرى الإعتيادية. اوصت الدراسة باستخدام طرق أخرى للكشف عن المشاهدات الشاذة في دراسات مستقبلية.

1- 7 - 13 دراسة (عبد الله، 2015) :

عنوان الدراسة: تأثير القيم الشاذة في معلمات نموذج تحليل الانحدار الخطي المتعدد:

تناولت الدراسة دراسة تأثير القيم الشاذة علي معلمات نموذج تحليل الانحدار الخطي المتعدد. وهدفت إلي عدة أهداف أهمها: دراسة تأثير وجود ومعالجة القيم الشاذة علي نتائج نموذج تحليل الانحدار الخطي المتعدد ، المقارنة بين طرق ومعايير تشخيص القيم الشاذة. و من الفرضيات التي استندت عليها الدراسة: القيم الشاذة لها تأثير معنوي علي نتائج نموذج تحليل الانحدار الخطي المتعدد. واعتمدت الدراسة علي المنهج الوصفي لوصف متغيرات الدراسة والمنهج التحليلي الاستدلالي بتطبيق نموذج تحليل الانحدار الخطي المتعدد وبعض الطرق والاختبارات الخاصة بتشخيص ومعالجة القيم الشاذة. حيث تم الكشف عن القيم الشاذة في المتغيرات المستقلة بطريقة الرسم الصندوقي وعناصر القطر الرئيسي للمصفوفة (Hat Matrix)، وكذلك تم الكشف عن القيم الشاذة في المتغير التابع باستخدام بواقي ستودنت المحذوفة وطريقة الرسم الصندوقي ، أما القيم الشاذة المؤثرة فتم الكشف عنها باستخدام الطرق (COOK'S ,DFBETAS ,DFFITS ,DISTANCE ,COVRATIO)، كما تمت معالجة القيم الشاذة بطريقة الحذف ومتوسط البتر.

ومن أهم النتائج التي توصلت إليها الدراسة: هنالك تأثيراً معنوياً من قبل القيم الشاذة علي معلمات نموذج الانحدار الخطي المتعدد حيث تعمل علي تضخم قيمة تباين الأخطاء MSE وتقلل من قيمة R^2 و قيمة F، إن وجود القيم الشاذة في بيانات نموذج الانحدار الخطي المتعدد يؤثر معنوياً وبصور كبيرة علي عدم تحقق فرضية التوزيع الطبيعي، احتمال ظهور الارتباط الذاتي لبواقي النموذج عند وجود القيم الشاذة في بيانات النموذج، معالجة القيم الشاذة بطريقة متوسط البتر أفضل من معالجتها بطريقة الحذف. و أوصت الدراسة بعدة توصيات أهمها: ضرورة تشخيص وتقليل أثر القيم الشاذة عند تطبيق نموذج تحليل الانحدار المتعدد للحصول علي نموذج أفضل.

1- 7 - 14 دراسة (الخثعمي، 2018):

عنوان الدراسة: الاخلال بافتراض خلو البيانات من القيم المتطرفة و اثره على استخدام تحليل الانحدار الخطي المتعدد في التنبؤ:

هدفت هذه الدراسة الى دراسة اثر وجود القيم المتطرفة على نموذج الانحدار الخطي المتعدد و دقة التنبؤ و معالجتها عبر دراسة اثرها على معلمات النموذج. طبق الباحث نموذج تحليل الانحدار الخطي المتعدد و بعض طرق الاختبارات الخاصة بتشخيص و معالجة القيم المتطرفة على عينة مكونة من 60 طالباً. حيث تم الكشف عن القيم المتطرفة باستخدام صندوق ويسكر (Whisker Box Plot)، كما تم استخدام عناصر القطر الرئيسي لمصفوفة القبعة (Hat Matrix) للكشف عن القيم المتطرفة في المتغيرات المستقلة، و باستخدام بواقي استيودنت المحذوفة (Studentized Deleted Residuals) للكشف عن القيم الشاذة في المتغير التابع. اما تأثير القيم المتطرفة فتم الكشف عنه باستخدام الطرق التالية: مقياس (DFFITS)، و قياس الأثر على معاملات الانحدار (DFBETAS)، و مقياس المسافة كوك (COOKS) (DISTANCE) الخاص بمقياس الأثر على كل معاملات الانحدار، و مقياس الأثر على الاخطاء

المعيارية (COVARATIO). و قد تمت معالجة القيم المتطرفة باستخدام طرق الحذف و متوسط البتر مع قياس قدرة النموذج قبل و بعد المعالجة عبر استخدام بعضاً من المؤشرات و هي: معامل التحديد R^2 و معامل التحديد المعدل R^2_a و احصاءة F و متوسط الخطأ المطلق (MAE) (Mean Absolute Deviation)، و جذر متوسط مربع الخطأ (RMSE) (Root Mean Square Error)، و متوسط الخطأ النسبي المطلق (MAPE) (Mean Absolute Percentage Error)، بالإضافة الى معامل التساوي لثايل (Theil's Inequality Coefficient). و توصلت الدراسة الي عدد من النتائج أهمها: هناك تأثير معنوي للقيم المتطرفة على معاملات نموذج الانحدار الخطي المتعدد حيث تعمل القيم المتطرفة على تضخيم قيم تباين الأخطاء MSE و تقلل من قيمة معامل التحديد R^2 و قيمة F. و ان معالجة القيم المتطرفة بمتوسط البتر افضل من حذفها. و اوصت الدراسة بعدة توصيات أهمها: ضرورة تشخيص وتقليل اثر القيم المتطرفة عند تطبيق نموذج تحليل الانحدار للحصول على نموذج افضل.

1- 8 التعليق على الدراسات السابقة:

يتضح من الدراسات السابقة التي لها صلة بموضوع الدراسة الحالية ان بعضها تناول دراسة القيم الشاذة بالتطبيق على نموذج تحليل الانحدار منها دراسة (الجبوري، و ناسي 2002م) حيث اقترحت هذه الدراسة طريقة القطع الناقص لتحديد المشاهدات الشاذة، و اقتراح ثماني طرق لمعالجة القيم الشاذة، و من اهم النتائج هو تحديد ارقام تراوحت بين واحد الي ثمانية مشاهدة شاذة و نجد ان بعض طرق المعالجة قد قللت قيمة MSE بدرجة معقولة و لم تؤدي الي رفع قيمة R^2 بنفس القدر. اما دراسة (حمد، 2005م) و التي تم فيها اقتراح طريقة تقريبية و طريقة شمولية مصحوبة بمقياس بياني لتشخيص القيم الشاذة في نموذج الانحدار الخطي البسيط و طريقة لمعالجة القيم الشاذة، اهم نتائج هذه الدراسة ان الطريقة التقريبية كشفت عن جزء من المشاهدات الشاذة

بينما كشفت الطريقة الشمولية عن العدد الكلي للملاحظات الشاذة بواسطة المقياس البياني و أحرزت الطرق كفاءة اعلى من الطرق السابقة حيث قللت طريقة المعالجة قيمة MSE و رفعت قيمة معامل التحديد البسيط بدرجة كبيرة، كانت هذه الدراسة جيدة و لكنها اغفلت جانب الملاحظات الشاذة ذات التأثير الجامح السلبي و الإيجابي و ذلك لان التطبيق كان على نموذج انحدار خطي بسيط واحد. اما دراسة (مروان، و فرح 2006م) فهدفت الي كشف القيم الشاذة التي تشير اليها نتائج تحليل الانحدار من خلال الرسم الصندوقي، حيث كانت اهم نتائجها ان وجود القيم الشاذة يؤثر على تركيبة النموذج و نتائج تحليل الانحدار من خلال ادخال متغيرات غير مؤثرة على الظاهرة المدروسة و استبعاد بعض المتغيرات المؤثرة، و كذلك لايمكن التنبؤ بتأثير القيم الشاذة على معامل التحديد او القيم التقديرية لتباين المجتمع. اما دراسة (احمد، 2007) كشفت عن القيم الشاذة في المتغيرات المستقلة باستخدام عناصر القطر الرئيسي Hat Matrix و كشفت عن مشاهدات المتغير التابع باستخدام بواقي t المعيارية و كشفت عن المشاهدات الشاذة المؤثرة باستخدام مسافة كوك، و اهم نتائج هذه الدراسة ان معاملات الانحدار تتاثر بالمشاهدات الشاذة في النوعين من المتغيرات و يختلف تأثيرها حسب موقعها من خط الانحدار. اما دراسة (قاسم، و إسماعيل 2008م) كشفت عن القيم الشاذة بطريقة بيز في نموذج الانحدار الخطي البسيط بأسلوب معاينة جيس، و اهم النتائج التي توصلت اليها الدراسة هو ان الكشف عن المشاهدات الشاذة بأسلوب بيز باستخدام معاينة جيس يحل الكثير من المشاكل المعقدة كما ان النتائج تكون دقيقة، و كذلك عند تطبيق الطرق الكلاسيكية للكشف عن المشاهدات الشاذة على البيانات نفسها قورنت النتائج مع طريقة معاينة جيس فكانت متقاربة و مشجعة جداً. اما دراسة (داوود، 2009م) كشفت عن القيم الشاذة باستخدام طريقة الرسم الصندوقي و تمت معالجتها بطريقة متوسط البتر و من اهم نتائجها ان وجود القيم الشاذة في بعض المتغيرات اثر سلبي على نتائج تحليل الانحدار حيث انها

قللت من قيمة F المحسوبة لمعادلة الانحدار التقديرية و كذلك قيمة معامل التحديد R^2 و زادت من قيمة متوسط مربعات الخطأ MSE بالإضافة الي ان طريقة الرسم الصندوقي كشفت عن بعض القيم الشاذة في المتغيرات المستقلة و هذا ما توصلت اليه الدراسة الحالية ايضاً. اما دراسة (يحي، 2009م) كشفت هذه الدراسة عن القيم الشاذة باستخدام الرسم الصندوقي و مقياس البواقي المعيارية و مصفوفة Hat Matrix اما المشاهدات المؤثرة فتم الكشف عنها باستخدام المقياس DFBETAS و اهم ما توصلت اليه الدراسة ان حذف المشاهدات أدى الي اقتراب توزيع البيانات من التماثل و كذلك اثر حذف المشاهدات الشاذة على دقة معاملات النموذج بانخفاض متوسط مربعات الخطأ MSE و ارتفاع قيمة F و ارتفاع قيمة معامل التحديد البسيط R^2 و تغيير إشارات بعض المعلمات المقدرة و الوصول الي نموذج يمثل بيانات الدراسة خير تمثيل، وهذا ما توصلت اليه الدراسة الحالية و لكن ان يؤدي حذف المشاهدات الشاذة الي ان البيانات المتبقية تمثل بيانات الدراسة خير تمثيل يخالف ما توصلت اليه الدراسة الحالية لان الحذف يؤدي الي نقص المعلومات و من ثم عدم الوثوق بالنتائج. اما دراسة (نزار، و غفران، 2009) استخدمت هذه الدراسة أسلوب المحاكاة لاختبار قوة و حصانة المقدرات لنموذج الانحدار المتعدد عند وجود التداخل الخطي و الأخطاء الغير طبيعية بتطبيقهما عدة طرق منها الاعتيادية و الحصينة، و من اهم نتائج هذه الدراسة ان مقدرات انحدار الريدج الموزون RMM هي افضل المقدرات لمختلف افتراضات توزيع الخطأ و درجة التداخل الخطي المتعدد. اما دراسة (شاكور، 2009) لجأت هذه الدراسة ايضاً الي تطبيق أسلوب الطرق الحصينة لمعالجة القيم الشاذة، حيث تم الكشف عن القيم الشاذة بطريقة الرسم الصندوقي و عناصر قطر مصفوفة Hat Matrix و مقياس مربع مسافة كوك لاختبار تأثير حذف المشاهدات الشاذة على المعلمات المقدرة و المقياس DF للكشف عن وجود القيم المتطرفة في متجه الأخطاء العشوائية، و اهم ما توصلت اليه هذه الدراسة ان أسلوب M

الحصين الموزون R.MW قد اثبتت كفاءته في تقدير معلمات نموذج الانحدار الخطي و بدقة عالية عند وجود القيم الشاذة في قيم المتغيرات المستقلة او المتغير التابع او كليهما. اما دراسة (حسن، 2009) هدفت الي تحديد المتغيرات المهمة في نموذج تحليل الانحدار الخطي المتعدد في حالة وجود المشاهدات الشاذة حيث تم تقدير جميع النماذج الجزئية الممكنة بطريقة M الحصينة و تمت المقارنة بين هذه النماذج باستخدام عدد من المعايير الحصينة أهمها معيار RAPE لاختيار النموذج الأفضل و الذي جاء متطابقاً مع رأي المختصين في مجال الطب، و هذا يتوافق مع اتجاهات الدراسة الحالية حيث ان بعض المتغيرات بسبب وجود المشاهدات الشاذة تكون غير مؤثرة و لكن بعد الكشف عن المشاهدات الشاذة و معالجتها يتضح ان هذه المتغيرات أساسية في النموذج. اما دراسة (حسن، و رضا، 2011م) طبقت ايضاً هذه الدراسة أسلوب الطرق الحصينة لتقدير نموذج الانحدار الخطي المتعدد في حالة وجود القيم الشاذة باستخدام عناصر قطر مصفوفة Hat Matrix و طريقة Box-and-Whisker plot البيانية و المشاهدات المؤثرة تم الكشف عنها باستخدام المقياس DFFITS، و اهم ما توصلت اليه هذه الدراسة ان طريقة مربعات الوسيط LMS هي الأفضل لتمثيل افضل نموذج انحدار خطي متعدد لبيانات الدراسة، في حين نجد ان الدراسة الحالية حريصة على تطبيق طريقة المربعات الصغرى الاعتيادية OLS لما لها من مقدرات ذات مزايا جيدة غير متحيزة. اما دراسة (الشميري، 2014م) كشفت هذه الدراسة عن القيم الشاذة باستخدام طريقة الصندوق Tukey بعد تقدير حد الخطأ المطلق و ترتيبه تصاعدياً و حذف المشاهدة الشاذة، و لمعالجة القيم الشاذة اعتماد تقسيم القيمة المطلقة لموجة حد الخطأ المقدر الى مجموعات متجانسة مع إيجاد وزن لكل قيمة مطلقة داخل مجموعتها مما يقود في النهاية الي حساب قيمة مقدره جديدة للمشاهدة الشاذة، و اهم نتائج هذه الدراسة ان الطريقة المقترحة برهنت كفاءة جيدة بتقدير المشاهدات الشاذة، كما أظهرت تفوقاً ملحوظاً عند مقارنة النتائج بنتائج طريقة

المربعات الصغرى الاعتيادية، و هذا ما توصلت اليه الدراسة الحالية و اكثر من ذلك، لان هذه الدراسة كشفت عن مشاهدة واحدة فقط حيث كشفت الدراسة الحالية عن ستة مشاهدات شاذة لنفس البيانات و تفوقت عليها بدرجة كبيرة في المعالجة. اما دراسة (عبد الله، 2015م) كشفت هذه الدراسة عن القيم الشاذة في المتغيرات المستقلة بطريقة الرسم الصندوقي Hat Matrix و كشفت عن القيم الشاذة في المتغير التابع باستخدام بواقي ستودنت المحذوفة و طريقة الرسم الصندوقي، اما القيم المؤثرة تم الكشف عنها باستخدام الطرق COOK'S DISTANCE, COVRATIO, DFFITS, DFBETAS, و تمت معالجة القيم الشاذة بطريقة الحذف و متوسط البتر، و اهم ماتوصلت اليه هذه الدراسة هو ان القيم الشاذة تعمل على تضخيم قيمة MSE و تقلل من قيمة R^2 و قيمة F المحسوبة، و هذا ما توصلت اليه الدراسة الحالية. اما دراسة (الخنعمي، 2018) كشفت هذه الدراسة عن القيم المتطرفة عن طريق صندوق ويسكر Whisker Box Plot، كما تم استخدام عناصر القطر الرئيسي لمصفوفة القبة Hat Matrix للكشف عن القيم المتطرفة في المتغيرات المستقلة، و باستخدام بواقي ستودنت المحذوفة للكشف عن القيم الشاذة في المتغير التابع، و قد تمت معالجة القيم الشاذة باستخدام طرق الحذف و متوسط البتر مع قياس قدرة النموذج قبل و بعد المعالجة عبر استخدام بعضاً من المؤشرات، و اهم ماتوصلت اليه هذه الدراسة ان القيم المتطرفة تعمل على تضخيم قيمة MSE و تقليل قيمة R^2 و قيمة F المحسوبة، و هذا ما توصلت اليه الدراسة الحالية. و اهم ما يميز الدراسة الحالية عن الدراسات السابقة:

1. تم تطبيق الدراسة على عدد 35 نموذج انحدار خطي بسيط و على عدد 31 نموذج انحدار خطي متعدد.
2. توصلت الدراسة لاكتشاف المشاهدات الشاذة دون سواها و تحديد عددها باستخدام المقياس البياني بالنسبة لكل نموذج انحدار خطي بسيط او متعدد.

3. نبهت الدراسة لتحديد و معالجة المشاهدات ذات التأثير الجامح السلبي و الإيجابي مسبقاً قبل البدء في عملية كشف المشاهدات الشاذة و التحري عن ظهور المشاهدات ذات التأثير الجامح الإيجابي بعد المعالجة لتضمينها مع المشاهدات الشاذة و إعادة المعالجة اذا لم يتم تشخيصها من ضمن المشاهدات الشاذة خاصة تلك التي تقع علي مقربة من خط الانحدار الا انها تظهر بصورة معزولة من المشاهدات (المشاهدات ذات التأثير الجامح الإيجابي).
4. عالجت هذه الدراسة مشكلة المشاهدات الشاذة بمعيار النسبية (يتم تقليل تأثير المشاهدات الشاذة بدرجة كبيرة اذا كان عددها صغير و بدرجة اقل اذا كان عددها كبير).
5. أظهرت الطرق المقترحة في هذه الدراسة كفاءة كبيرة من حيث التشخيص و المعالجة بعد مقارنتها بعدد من الطرق السابقة.
6. قدرت الدراسة العدد المتوسط المتوقع للمشاهدات الشاذة في نموذج الانحدار الخطي بشقيه البسيط و المتعدد.
7. أعادت هذه الدراسة الثقة في مقدرات طريقة المربعات الصغرى الاعتيادية في ظل وجود المشاهدات الشاذة (حيث انه تم توفيق جميع نماذج الانحدار الخطي المتعدد و و الانحدار الخطي البسيط بطريقة المربعات الصغرى اثناء عمليات التشخيص و بعد المعالجة باستخدام طريقة المربعات الصغرى) الامر الذي أدى الى تحسن كبير في نتائج تحليل نموذج الانحدار الخطي البسيط و المتعدد.

1-9 هيكلية البحث:

يحتوي هذا البحث خمسة فصول، الفصل الأول و يشتمل على (تمهيد، مشكلة البحث، أهمية البحث، اهداف البحث، مجتمع و عينة البحث، فرضيات البحث، منهجية البحث، الدراسات السابقة، التعليق على الدراسات السابقة، هيكلية البحث)، و الفصل الثاني و يتضمن (تمهيد،

مفهوم القيم الشاذة و أسباب ظهورها، طرق الكشف عن القيم الشاذة في المتغيرات المستقلة و المتغير التابع، المشاهدات الشاذة المؤثرة و مقياس الكشف عنها، بعض الحلول المستخدمة لمعالجة القيم الشاذة)، و الفصل الثالث و يتضمن (تمهيد، الطريقة المقترحة لتشخيص القيم الشاذة، المقياس البياني لتحديد عدد المشاهدات الشاذة، الطريقة المقترحة لمعالجة المشاهدات الشاذة، كيفية التعامل مع المشاهدات ذات التأثير الجامح السلبي، كيفية التعامل مع المشاهدات ذات التأثير الجامح الإيجابي)، و الفصل الرابع و يتضمن (تمهيد، التطبيق العملي بغرض المقارنة، التطبيق العملي علي نماذج الانحدار الخطي البسيط بغرض استخراج المؤشرات، اختبارات الفروض فيما يتعلق بنماذج الانحدار الخطي البسيط، التطبيق العملي علي نماذج الانحدار الخطي المتعدد بغرض استخراج المؤشرات، اختبارات الفروض فيما يتعلق بنماذج الانحدار الخطي المتعدد)، و اخيراً الفصل الخامس و يتضمن (النتائج والتوصيات، ثم المراجع و الملاحق).

الفصل الثاني

المشاهدات الشاذة في نموذج تحليل الانحدار الخطي

2-0 تمهيد:

2-1 الانحدار الخطي:

2-2 مفهوم القيم الشاذة:

2-3 أسباب ظهور القيم الشاذة:

2-4 طرق الكشف عن القيم الشاذة:

2-4-1 الطرق اللامعلمية:

2-4-2 اكتشاف المشاهدات الشاذة في المتغيرات المستقلة:

2-4-3 اكتشاف المشاهدات الشاذة في المتغير التابع:

2-5 المشاهدات الشاذة المؤثرة وطرق الكشف عنها:

2-5-1 قياس التأثير على القيم المقدرة - مقياس DFFITS:

2-5-2 قياس الأثر على معاملات الانحدار - مقياس DFBETAS:

2-5-3 قياس الأثر على كل معاملات الانحدار - مقياس مسافة كوك:

2-5-4 قياس الأثر على الأخطاء المعيارية - مقياس COVRATIO:

2-6 بعض الحلول المقترحة لمعالجة القيم الشاذة:

2-7 اختبار الفرضيات Test of Hypotheses

2- 0 تمهيد:

من خلال هذا الفصل سوف نتناول بشيء من التفصيل مفهوم القيم الشاذة وأسباب ظهورها في البيانات وطرق الكشف عنها وتقديرها في المتغيرات المستقلة والمتغير التابع و طريقة تشخيص المشاهدات المؤثرة على معاملات نموذج تحليل الانحدار الخطى والتي تؤثر على دقة النتائج، بالإضافة إلى بعض طرق معالجة القيم الشاذة.

2- 1 الانحدار الخطي: Linear Regression (النجار، و محمد 2007)

يعرف الانحدار الخطي بأنه علاقة رياضية تربط بين المتغيرات، ويتم التعبير عن هذه العلاقة بصيغة معادلة خطية تضم المتغير المعتمد ومتغير واحد أو أكثر من المتغيرات التوضيحية والمعادلة (5) توضح هذه العلاقة

$$y_t = B_0 + B_1x_{1t} + B_2x_{2t} \cdots + B_kx_{kt} + \varepsilon_t \quad \cdots(1- 2)$$

وبصيغة المصفوفات يتم التعبير عن الانحدار الخطي كالأتي.

$$\underline{Y} = \underline{X} \underline{B} + \underline{\varepsilon} \quad \cdots(2 - 2)$$

حيث إن:

Y متجه ($n \times 1$) لملاحظات المتغير المعتمد.

X مصفوفة ($n \times P$) لملاحظات المتغيرات التوضيحية.

B متجه ($P \times 1$) للمعاملات المجهولة في الأنموذج والمراد تقديرها.

ε متجه ($n \times 1$) للأخطاء العشوائية.

يعرف الأنموذج أعلاه بالأنموذج الخطي العام ولغرض تقدير معالمته تستخدم أكثر من طريقة، ومن أكثر هذه الطرائق استخداماً هي طريقة المربعات الصغرى، والتي تعتمد على معيار تصغير مجموع مربعات الخطأ

$$\underline{\varepsilon}' \underline{\varepsilon} = (Y - X \underline{B})' (Y - X \underline{B}) \quad \dots (3 - 2)$$

ولتصغير الدالة في المعادلة (7) نجد الاشتقاق الجزئي بالنسبة للمعلمات \underline{B} ثم مساواتها بالصفر وبالنتيجة سوف نحصل على P من المعادلات الطبيعية والتي تكون كما في المعادلة (8) الآتية:

$$X' X \underline{b} = X' Y \quad \dots (4 - 2)$$

وبحل المعادلة (8) نحصل على قيم تقديرية للمعلمات \underline{B} وهذه القيم هي \underline{b} التي يجب أن تحقق الشرط الثاني الخاص بالتصغير وهو أن تكون ألمشتقة الجزئية الثانية للمعادلة (8) بالنسبة إلى قيم المعلمات \underline{B} موجبة التحديد، وفي حالة نماذج الانحدار إذ تكون مصفوفة المتغيرات التوضيحية X ذات رتبة كاملة يمكن الحصول على تقدير المعلمات من خلال المعادلة الآتية.

$$\underline{b} = (X' X)^{-1} X' Y \quad \dots (5 - 2)$$

2-2 مفهوم القيم الشاذة:

للقيم الشاذة تعريفات عدة نذكر منها ما يلي:

عرف (Bross, 1961) المشاهدة الشاذة بأنها تلك المشاهدة التي تظهر منحرفة بشكل كبير عن سائر مكونات العينة التي وجدت فيها تلك المشاهدة. أما (Freeman, 1980) فقد عرف القيمة الشاذة بأنها اية مشاهدة لم تتولد بالطريقة العامة التي ولدت الأغلبية العظمى من مشاهدات البيانات. أما (Barnett, 1978) فقد عرفت المشاهدة الشاذة في مجموعة البيانات بأنها تلك المشاهدة التي تبدو غير منطقية اذا قورنت بسائر مجموعة البيانات. و كذلك عرف (Al-Jobouri)

(1976) القيمة الشاذة بأنها تلك القيمة التي تكون غير منسجمة مع بقية بيانات المجموعة لمتغير من المتغيرات لظاهرة معينة او مجموعة من الظواهر، او ان القيم الشاذة هي القيم التي تأتي من مجتمع يختلف عن مجتمع العينة قيد الدراسة. او كما نص تعريف (إسماعيل، 2001) علي ان المشاهدات الشاذة هي مجموعة قليلة من المشاهدات تبعد قيمها بصورة كبيرة عن بقية قيم المشاهدات في العينة. و ينص (جايترجي، 1989) ان القيم الشاذة هي نقاط البيانات ذات البواقي الكبيرة، حيث أن البيانات تصنف على أنها قيم شاذة عندما تكون لها بواقي عالية مقارنة مع بواقي المشاهدات الباقية". و يصف (الراوي، 1987) على أن الخوارج مصطلح يطلق على المشاهدات المتطرفة (extreme) التي تقع بعيداً عن خط الانحدار وعادة يكون حد الخطأ لها كبيراً (Large residuals) مقارنة ببقية المشاهدات الأخرى. كما عرف (Keller, 2000) المشاهدات الشاذة بأنها المشاهدات التي تقع بعيدة عن معادلة الانحدار و يكون لها خطأ كبير مقارنة ببقية المشاهدات الطبيعية الأخرى في مجموعة البيانات و بذلك يكون لها تأثير في النموذج الخطي و تقديراته.

2- 3 أسباب ظهور القيم الشاذة:

قد تظهر القيم الشاذة في مجموعة البيانات لاسباب عدة منها مايلي:

1. ان البيانات تعود لتوزيعات غير متماثلة أي يكون فيها التواء عالي نحو اليمين او نحو اليسار و لصياغة نموذج لهذه التوزيعات قدم Green, 1976 تصنيفاً لعوائل هذه التوزيعات، توزيعات عرضة للقيم الشاذة و هي التي تكون لها نهايات تؤول الى الصفر ببطء و هذه التوزيعات تكون عرضة للقيم الشاذة بصورة مطلقة. و توزيعات تكون مقاومة للقيم الشاذة و هي التي تؤول نهاياتها الى الصفر بشكل اسرع من سابقتها و تكون مقاومة للقيم الشاذة بصورة مطلقة.
2. تأتي البيانات من نوعين من التوزيعات احدهما التوزيع الاسي و الذي يولد مشاهدات جيدة بينما الآخر يسمى التوزيع الملوث و الذي يولد قيماً شاذة.

3. قد تحدث القيم الشاذة نتيجة لاسباب أخرى منها أخطاء يقع فيها الباحث نفسه عند رصد القياسات، كما نص (إسماعيل، 2001) "يرجع بروز البيانات الشاذة إلى أخطاء أما في مرحلة جمع البيانات، أو في مرحلة المعالجة كإدخال البيانات في الحاسب"، أو قد تحدث نتيجة الى وجود خلل في أجهزة القياس و خاصة في التجارب المختبرية، أو قد تحدث نتيجة لاختفاء في الحسابات مما يؤدي الي ظهور القيم الشاذة.

4. أو قد تكون هذه البيانات حقيقية ناتجة عن ظروف غير عادية، فمثلاً حدوث كوارث طبيعية كالزلازل، الأعاصير، الأمطار الغزيرة تؤثر على مستويات الإنتاج الزراعي ، الحيواني، والصناعي، إضراب عمال في منشأة ما يؤثر على إنتاجها، الحروب بين الدول تؤثر على اقتصاديات هذه الدول... الخ.

2-4 طرق الكشف عن القيم الشاذة:

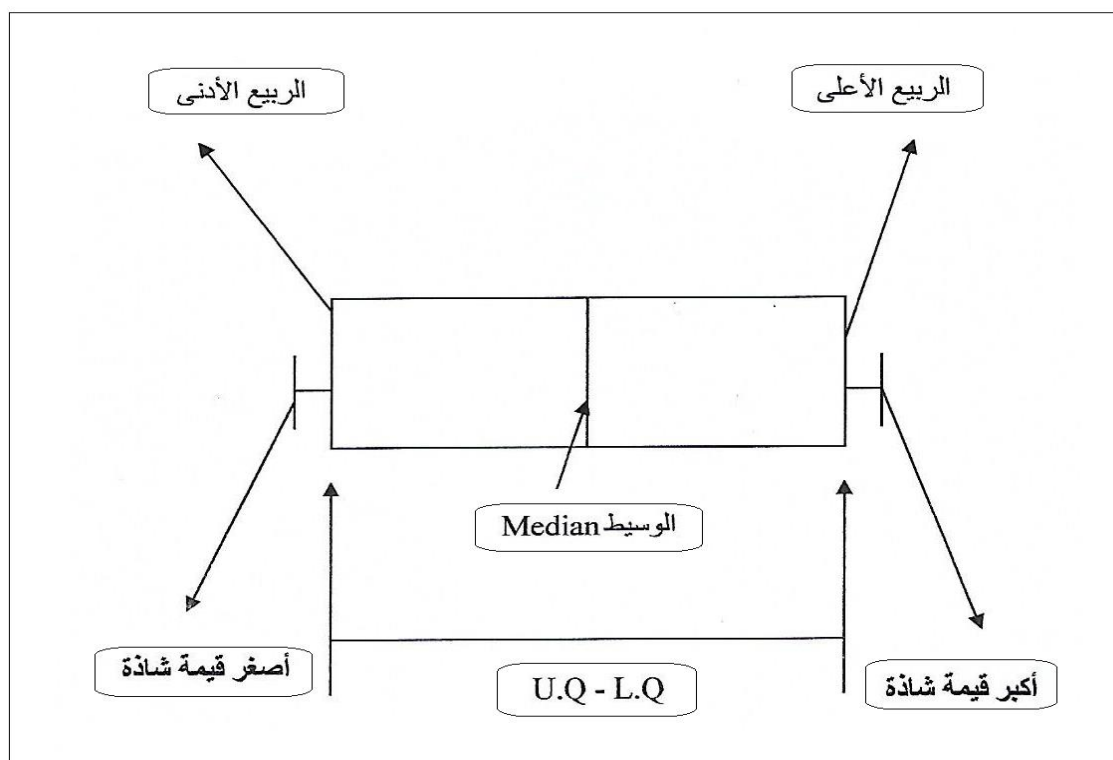
نشير الى ان نماذج الانحدار الخطي تتأثر بالقيم الشاذة و المتطرفة إن و جدت كغيرها من النماذج الإحصائية الأخرى. و مما لا شك فيه ان طريقة المربعات الصغرى و بسبب القيم الشاذة و المتطرفة تخفق في إعطاء التقديرات الدقيقة لمعالم النموذج، ذكر (Huber, 1973) في مقولته المشهورة أن وجود قيمة شاذة واحدة تهدم المزايا الجيدة لمقدرات المربعات الصغرى لذا في هذا الجزء سنركز على طرق تشخيص القيم الشاذة و تحديد مدى تأثير وجود هذه القيم على مقدرات نموذج الانحدار الخطي،، بين (Kleinbaum, 1988) ان وجود قيم شاذة في بيانات نموذج الانحدار في مشاهدات المتغير التابع او المتغيرات المستقلة يؤثر على تقديرات معالم هذا النموذج و الإحصاءات المرتبطة بها و يؤثر ايضاً على اختيار المتغيرات المؤثرة في نموذج الانحدار. و سنتناول بشيء من التفصيل استخدام بعض الطرق اللامعلمية و التحليلية

لمعالجة موضوع الحالات الشاذة من حيث تحديدها و قياس اثرها على مقدرات المربعات الصغرى في نموذج تحليل الانحدار الخطي فيما يلي:

2- 4-1 الطرق اللامعلمية:

و هي تلك الطرائق التي لا تفترض توزيع معين علي البيانات متخطية بذلك احد الافتراضات الأساسية التي تعتمد عليها الطرائق المعلمية. و تعتبر طريقة الصندوق اللامعلمية المقترحة من قبل العالم (Tukey, 1977) من اكثر هذه الطرق شيوعاً و استخداماً و التي تستخدم لفحص القيمة او القيم الشاذة في حالة المتغير الواحد و بافتراض عدم معرفة التوزيع الاحتمالي للمتغير. ويمكن توضيح طريقة الرسم الصندوقي بالشكل (1-2) ادناه:

الشكل (2 - 1): مخطط الرسم الصندوقي



و تعتبر المشاهدة او المشاهدات متطرفة اذا كانت تبعد قيمها عن قمة او قاعدة المستطيل مسافة تزيد عن ثلاثة أمثال طول الصندوق محسوبة من الربيع الأدنى او الربيع الأعلى، أي تكون المشاهدات اكبر من $Q_3 + 3(Q_3 - Q_1)$ او ان تكون المشاهدات اقل من $Q_3 - 3(Q_3 - Q_1)$ و تظهر عادةً في شكل نجمة (*)، اما المشاهدات الشاذة هي التي تبعد قيمها عن قمة او قاعدة المستطيل مسافة ما بين 1.5 و 3 أمثال طول المستطيل محسوبة من الربيع الأدنى او الربيع الأعلى و تظهر عادةً في شكل الحرف الإنجليزي (O) للإشارة علي انها شاذة.

و تصنف الطرق البيانية على انها طرق لاطمئنة كذلك خصوصاً تلك الطرق البيانية التي تعتمد في اغلب الأحيان على شكل انتشار بيانات نموذج الانحدار الخطي و على مهارة الباحث في تحديد القيم الشاذة من خلال رصد القيمة او القيم التي تظهر من خلال شكل انتشار البيانات (Scatter Plot بصورة معزولة عن اغلبية المشاهدات و من ثم استبدالها بقيم تقديرية، او كما نص (الراوى، 1987) "يمكن اكتشاف الخواارج بيانيا عندما نرسم الرسم البياني لـ e_{is} (Standardized Residual) ضد X_i فالنقاط التي تقع خارج (± 2) تعد من الخواارج"، و هنالك العديد من الطرق البيانية مثل طريقة (الجبرى، و آخرون 2002) التي اقترحا فيها معادلة القطع الناقص الذي مركزه متوسطي المتغيرين X و Y و عليه فإن النقطة التي تقع خارج القطع الناقص تعتبر شاذة اصف الى ذلك طريقة (حمد، 2005) البيانية التي كانت تعتمد في تشخيص المشاهدات الشاذة على المعادلة $Z_i = Y_i - \hat{\beta}_1 X_i$ من خلال توفيق النموذج بطريقة المربعات الصغرى و حساب قيم Z_i و ترتيبها تصاعدياً و تمثيلها بيانياً ضد رتبها من بعد ذلك تبرز المشاهدات الشاذة و يسهل التعرف عليها و من ثم يتم استبدال القيم الشاذة بقيم جديدة.

2-4-2 الطرق التحليلية :

هنالك نوعين من طرق التشخيص التحليلية للمشاهدات الشاذة:

النوع الأول:

هو طرق التشخيص التقليدية التي تعمل علي تشخيص المشاهدات الشاذة من خلال مقاييس محسوبة من البيانات و التي قسمها (الجبوري، و آخرون 2002) الي مقاييس تعتمد علي البواقي و مقاييس تعتمد على حجم القطع الناقص و مقاييس تعتمد علي دوال الأثر و مقاييس تعتمد علي الأثر الجزئي.

النوع الثاني:

هو كما ذكر (محمد، 2009) متمثل في "الطرق التي تعتمد علي الانحدار المتين (Robust Regression) التي تحاول إيجاد مقدرات لا تتأثر بالمشاهدات الشاذة بصورة كبيرة " ومن امثلة الطرق على هذا النوع طريقة وسيط المربعات الأدنى (LMS) و طريقة المربعات المشذبة الصغرى (LTS) و طريقة مقدرات (M). و سوف نتناول بعضاً من هذه الطرق التحليلية بشي من التفصيل و كالاتي:

2-4-1 اكتشاف المشاهدات الشاذة في المتغيرات المستقلة:

المشاهدات الشاذة في المتغيرات المستقلة تسمى احياناً بنقاط الجذب leverage points و هي تلك النقاط التي تقع بعيدة عن معظم القيم في المصفوفة X الموجودة في النموذج و التي تشمل واحد او اكثر من المتغيرات المستقلة في تحليل الانحدار، و هذه النقاط لها تأثير كبير و قوي في مقدرات المربعات الصغرى الاعتيادية، و كذلك يقصد بنقطة الجذب قيمة من القيم X_i شاذة الا ان قيم Y_i المقابلة لها تطابق النموذج بشكل جيد جداً و هذه النقطة تجذب تقديرات المربعات الصغرى نحوها.

ويتم الكشف عن المشاهدات الشاذة للمتغيرات المستقلة في نموذج الانحدار المتعدد من خلال

العناصر القطرية h_{ii} لمصفوفة القبعة Hat Matrix المعرفة بالعلاقة التالية:

$$H_{n \times n} = X(X^T X)^{-1} X^T = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1n} \\ h_{21} & h_{22} & \cdots & h_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ h_{n1} & h_{n2} & \cdots & h_{nn} \end{bmatrix} \quad \dots\dots\dots(6-2)$$

حيث ان:

h_{ii} : هو العنصر القطري للمصفوفة H التي لها الخواص التالية:

1. انها مصفوفة متماثلة $H = H^T$ و جامدة ايضاً $H^2 = H$.

2. العنصر القطري h_{ii} يعرف بالمعادلة التالية:

$$h_i = h_{ii} = h_i^T h_i = h_i^2 + \sum_{i \neq j} h_{ij}^2, \quad \frac{1}{n} \leq h_{ii} \leq 1$$

3. مجموع قيم العناصر القطرية يساوي عدد معالم نموذج الانحدار بما في ذلك المعامل الثابت أي:

$$\sum_{i=1}^n h_i = p \quad \dots\dots\dots(7-2)$$

4. متوسط قيم العناصر القطرية يساوي عدد معالم نموذج الانحدار على عدد المشاهدات أي:

$$\bar{h} = \frac{p}{n} \quad \dots\dots\dots(8-2)$$

5. تسمى المصفوفة H بمصفوفة التنبؤ Predictor Matrix اذ يمكن الحصول منها على القيم

التقديرية للنموذج كما يلي:

$$\hat{Y} = H'Y = \sum_{i=1}^n h_{ij} y_j \quad \dots\dots\dots(9-2)$$

تعتبر العناصر القطرية h_{ii} مقياساً مفيداً في الكشف عن الحالات الشاذة في المتغيرات المستقلة

في نموذج الانحدار الخطي المتعدد و يسمى العنصر القطري بقيمة القبعة Hat Value او

بالرافعة Leverage و تقيس الرافعة المسافة بين الحالة رقم i ($X_{1i}, X_{2i}, \dots, X_{pi}$) و

متوسط قيم كل الحالات $\bar{X}_{1i}, \bar{X}_{2i}, \dots, \bar{X}_{pi}$ ، فعندما تزداد قيم العناصر h_{ii} تقل بالمقابل قيمة تباين البواقي e_i حتى اذا أصبحت قيمة $h_{ii}=1$ سيكون تباين البواقي $\sigma_{ei}^2=1$ و نستنتج من ذلك انه اذا كانت قيمة $h_{ii}=1$ فإن قيمة المتغير التابع Y_i تساوي القيمة المقدرة \hat{Y}_i .
و لتشخيص الحالات الشاذة بالاعتماد على قيمة الرافعة هنالك عدد من الطرق سنتناول بالشرح منها مايلي:

1. طريقة بيلسلي و آخرون (Belsley, et al, 1980) :

تعد المشاهدة \bar{x} مشاهدة شاذة أي لها نقطة انعطاف و مؤثرة في نتائج النموذج اذا كانت قوة الرافعة لها اكبر من ضعف متوسط الرافعات أي عندما تكون:

$$h_{ii} \geq \frac{2(P)}{n} \quad \dots\dots\dots(10-2)$$

2. طريقة نيتير و آخرون (Neter, et al, 1990) :

اعتبر نيتير و آخرون ان الرافعة التي تزيد قيمتها عن (0.5) كبيرة و تشير الى ان الحالة شاذة و لابد من دراستها، أي:

$$h_{ii} \geq 0.5 \quad \dots\dots\dots(11-2)$$

3. طريقة جون فوكس (Fox, 1997):

ان المعايير التي ذكرت من قبل لتشخيص القيم الشاذة في المتغيرات المستقلة تكون جيدة في تحديد الحالات الشاذة في العينات الكبيرة و المتوسطة نسبة الى عدد المعالم، اما في حالة العينات الصغيرة تكون هذه المعايير ضعيفة الأداء و غير عملية و ذلك لان نسبة عدد المعالم الى حجم العينة $\frac{P}{n}$ سيكون كبيراً و بالتالي يتوقع ترشيح عدد كبير من الحالات الشاذة لفحصها لذلك اقترح جون فوكس في هذا الصدد دراسة كل الحالات التي تزيد قيم رافعتها h_{ii} عن ثلاثة اضعاف

متوسط قيم الرافعات بدلاً من دراسة الحالات التي تزيد قيم رافعتها h_i عن ضعف قيم متوسط الرافعات كما في معيار بيلسلي و آخرون السابق، أي تعتبر الحالة شاذة اذا كانت :

$$h_i \geq \frac{3(P)}{n} \quad \dots\dots\dots(12-2)$$

2-4-2 اكتشاف المشاهدات الشاذة في المتغير التابع:

للكشف عن المشاهدات في المتغير التابع تستخدم بواقي ستودنت المحذوفة (Standardized Detected Residual) التي يتم الحصول عليها بإيجاد القيمة المعيارية للباقي المحذوف، حيث يتم حسابه باتباع الخطوات الآتية:

1. يتم أولاً حذف الحالة رقم \bar{z} و من ثم يتم بناء نموذج انحدار باستخدام بقية الحالات $n-1$.
2. يتم تقدير القيمة المتوقعة رقم \bar{z} للمتغير التابع و التي يرمز لها بالرمز $\hat{Y}_{i(i)}$ و ذلك بتعويض قيم المتغيرات المستقلة المناظرة للحالة رقم \bar{z} أي $(X_{1i}, X_{2i}, \dots, X_{pi})$ في معادلة الانحدار التي تم تقديرها في الخطوة السابقة.
3. يتم حساب الباقي المحذوف بالصيغة الآتية:

$$d_i = Y_i - Y_{i(i)} \quad \dots\dots\dots(13-2) \quad \text{حيث أن:}$$

d_i : الباقي المحذوف رقم \bar{z} .

Y_i : القيمة الفعلية للمتغير التابع رقم \bar{z} .

$\hat{Y}_{i(i)}$: القيمة المقدرة رقم \bar{z} للمتغير التابع.

و يتضح من المعادلة (2 - 13) اننا نحتاج الى بناء عدد n نموذج انحدار لحساب البواقي المحذوفة لجميع المشاهدات.

و لصعوبة هذا الاجراء تم تقدير معادلة أخرى تمكن من حساب البواقي المحذوفة دون الحاجة

لبناء هذا العدد من النماذج تعطى بالصيغة التالية (Neter, et al,1990) :

$$d_i = \frac{e_i}{1-h_{ii}} \quad \dots\dots\dots(14-2)$$

حيث ان:

e_i : البواقي العادية لنموذج دالة الانحدار باستخدام جميع الحالات n .

h_{ii} : العنصر القطري رقم i من المصفوفة H باستخدام جميع المشاهدات لتقدير دالة الانحدار.

و نلاحظ من المعادلة (14 - 2) ان قيمة الباقي المحذوف d_i تكون كبيرة اذا كانت قيمة h_{ii}

كبيرة فهذا يجعل البواقي المحذوفة تحدد احياناً المشاهدات الشاذة في قيم المتغير التابع التي تفشل

البواقي العادية e_i في تحديدها، و في أحيان أخرى تفقد البواقي المحذوفة الي التشخيصات نفسها

التي تفقد اليها البواقي العادية. لذلك تستخدم بواقي ستيدنت المحذوفة d_i^* للكشف عن المشاهدات

الشاذة في المتغير التابع و التي تعطى بتطبيق المعادلة التالية:

$$d_i^* = \frac{d_i}{S.e(d_i)} \quad \dots\dots\dots(15-2)$$

حيث ان:

d_i^* : بواقي ستيدنت المحذوف رقم i .

d_i : الباقي المحذوف رقم i .

$S.e(d_i)$: الانحراف المعياري للباقي المحذوف للمشاهدة رقم i و يساوي:

$$S.e(d_i) = \sqrt{\frac{MSE_{(i)}}{1-h_{ii}}} \quad \dots\dots\dots(16-2)$$

حيث ان:

$MSE_{(i)}$: متوسط مجموع مربعات الخطأ للنموذج المقدر بعد حذف المشاهدة رقم i .

و نلاحظ من المعادلة (2 - 16) اننا نحتاج بناء عدد n نموذج انحدار لحساب d_i^* و لقد اثبت العلماء ان باقي ستيودنت المحذوف يتبع توزيع t بدرجات حرية $n - p - 2$ و توصلوا الي صيغة أخرى تمكن من حساب بواقي ستيودنت المحذوفة دون الحاجة لبناء عدد n نموذج انحدار و ذلك باستخدام المعادلة التالية:

$$d_i^* = e_i \left[\frac{n-p-2}{RSS(1-h_{ii})-e_i^2} \right]^{\frac{1}{2}} \dots\dots\dots(17-2)$$

حيث ان:

e_i : البواقي العادية ، n : عدد المشاهدات ، p : عدد معالم النموذج.
 RSS : مجموع مربعات البواقي لنموذج الانحدار المقدر باستخدام جميع المشاهدات n .

h_{ii} : العنصر القطري رقم i من المصفوفة H باستخدام جميع المشاهدات لتقدير دالة الانحدار.
و لتشخيص المشاهدات الشاذة في المتغير التابع تتم مقارنة القيمة المطلقة لباقي ستيودنت المحذوف d_i^* بقيمة t عند درجات حرية $n-p-2$ و مستوى معنوية α_0 فإذا كانت:

$$|d_i^*| \geq t_{\alpha_0, n-p-1} \dots\dots\dots(18-2)$$

تعتبر الحالة Y_i حالة شاذة لا بد من دراستها و تحديد مدى تأثيرها على مقدرات المربعات الصغرى.

2-4-3 الانحدار المتين (Rousseeuw, and Leroy, 1987):

يحاول الانحدار المتين إيجاد مقدرات لا تتأثر بالمشاهدات الشاذة بصورة كبير و يعتقد كثير من الإحصائيين أن الغرض من الإنحدار المتين هو تجاهل المشاهدات الشاذة، ولكن هذا غير صحيح بل على العكس بالنظر إلي بواقي الإنحدار المتين يمكن تحديد المشاهدات الشاذة التي يصعب تحديدها عادة بإستخدام بواقي المربعات الصغرى. ولهذا فإن الإنحدار المتين والطرق التقليدية

لتشخيص المشاهدات الشاذة كلاهما له نفس الهدف ولكن فقط بترتيب مختلف حيث يتم حذف المشاهدات الشاذة أولاً ثم توفيق الانحدار باستخدام المربعات الصغرى للبيانات الجيدة في الطرق التقليدية ، بينما يهتم الانحدار المتين بتوفيق الانحدار لغالبية البيانات ثم يكتشف المشاهدات الشاذة التي يكون لها بواقي كبيرة ناتجة عن ذلك الانحدار المتين.

2-5 المشاهدات الشاذة المؤثرة و طرق الكشف عنها:

المشاهدة المؤثرة هي التي تقع في الجهة المقابلة لوسط الخط الممثل لنموذج الانحدار اذ تكون بعيدة عن المحور Y و كذلك بعيدة عن بقية المشاهدات الأخرى، تعمل هذه المشاهدة على تحريك خط الانحدار نحو الأعلى او نحو الأدنى، بمعنى آخر باتجاه المشاهدة المؤثرة، و يمكن ان يكون لها تأثير معنوي في نتائج نموذج الانحدار وفي القدرة على التنبؤ، وتعتبر الحالة او المشاهدة مؤثرة اذا كان ادراجها او حذفها يؤدي الى احداث تغيرات جوهرية في قيم معاملات النموذج المقدره والإحصاءات المرتبطة بها (Hair, et al, 1998). وبعد الكشف عن الحالات الشاذة في المتغيرات المستقلة والمتغير المعتمد لابد من تحديد ماذا كانت هذه الحالات لها تأثير على النتائج ام لا، وكما ذكرنا سابقا تعتبر المشاهدة مؤثرة اذا كان استثناءها يسبب تغيرات جوهرية في تقدير معالم الانحدار المقدره والإحصاءات المرتبطة بها وتوجد عدة مقاييس لتحديد الحالات المؤثرة تعتمد في تحديد أثر الحالة الشاذة على قياس الفرق بين قيم مقدرات المربعات الصغرى باستخدام كل الحالات n و باسقاط حالة واحدة أي $n-1$ ، و سوف نتناول بالشرح المفصل اكثر المقاييس المستخدمة لتحديد المشاهدات المؤثرة على نتائج نموذج تحليل الانحدار الخطى فيما يلي:

2-5-1 قياس التأثير على القيم المقدره (Belsley, et al 1980):

يستخدم هذا المقياس لقياس الحالة رقم \bar{z} على القيمة المقدره \hat{Y}_i و لقد طور بيلسلي و آخرون

المعادلة التالية لقياس أثر الحالة رقم \bar{z} على القيمة المقدره \hat{Y}_i :

$$DFFITs_i = \frac{\hat{Y}_i - \hat{Y}_{i(i)}}{\sqrt{MSE_{(i)} h_{ii}}} \dots\dots\dots(19 - 2)$$

و تم تطوير الصيغة (2 - 19) الى الصيغة التالية:

$$DFFITs_i = d_i^* \sqrt{\frac{h_{ii}}{1 - h_{ii}}} \dots\dots\dots(20 - 2)$$

حيث ان:

\hat{Y}_i : القيمة المقدرة للمتغير التابع للملاحظة رقم \bar{i} عند استخدام جميع المشاهدات.

$\hat{Y}_{i(i)}$: القيمة المقدرة للمتغير التابع بعد حذف المشاهدة رقم \bar{i} .

d_i^* : باقي ستودنت المحذوف للملاحظة رقم \bar{i} باستخدام جميع المشاهدات.

h_{ii} : قيم العنصر القطري رقم \bar{i} من المصفوفة H المقدرة باستخدام جميع المشاهدات لتقدير

دالة الانحدار، $MSE_{(i)}$: متوسط مربعات الخطأ عند حذف المشاهدة رقم \bar{i} .

و نلاحظ من العلاقة (2 - 20) ان قيم $DFFITs$ تعتمد على قيم h_{ii} و d_i^* ، أي انه اذا كانت

المشاهدة رقم \bar{i} شاذة في المتغير X (قيمة h_{ii} لها كبيرة) و ايضاً اذا كانت المشاهدة رقم \bar{i}

شاذة في المتغير التابع Y (قيمة d_i^* لها كبيرة فإن القيمة المطلقة لهذا المقياس تكون كبيرة ايضاً

و تكون القيمة الكبيرة كدليل نستدل منه فيما اذا كانت المشاهدة رقم \bar{i} مؤثرة ام لا.

و لتحديد او تشخيص الحالات المؤثرة على نتائج نموذج تحليل الانحدار الخطي هناك عدد من

الطرق الرسمية التي تستخدم لنفس الغرض نذكر منها ما يلي:

1. طريقة (Belsley, et al 1980):

اعتبر بليسلي و آخرون الحالة مؤثرة على نتائج تحليل الانحدار الخطي المتعدد اذا كانت:

$$|DFFITs_i| = 2\sqrt{\frac{p}{n}} \quad \dots\dots\dots(21-2)$$

2. طريقة (Neter, et al 1990):

تستخدم هذه الطريقة في حالة العينات الكبيرة و الصغيرة و المتوسطة، ففي حالة العينات الصغيرة و المتوسطة تعتبر الحالة مؤثرة اذا كانت:

$$|DFFITs_i| \geq 1 \quad \dots\dots\dots(22-2)$$

اما في حالة العينات الكبيرة تعتبر الحالة مؤثرة اذا كانت:

$$|DFFITs_i| \geq 2\sqrt{\frac{p}{n}} \quad \dots\dots\dots(23-2)$$

3. طريقة (Chatterjee, and Hadi, 1988):

اقترح كل من شاتيرجي و هادي معيار لمقارنة القيمة المطلقة ل $DFFITs$ بقيمة اكبر قليلا من القيمة التي اقترحها بيلسلي و آخرون تعرف كالآتي:

$$|DFFITs_i| \geq 2\sqrt{\frac{p}{n-p}} \quad \dots\dots\dots(24-2)$$

2- 5- 2 قياس الأثر على معاملات الانحدار (Belsley, et al 1980):

يستخدم مقياس $DFBETAS$ الذي طوره بيلسلي و آخرون لقياس الفرق بين قيم معاملات الانحدار b_k المقدره باستخدام جميع الحالات n و قيم معاملات الانحدار $b_{k(i)}$ المقدره بعد حذف الحالة i رقم أي باستخدام $n-1$ حالة، و يتم الحصول على هذا المقياس بتطبيق المعادلة التالية:

$$DFBETAS_i = \frac{b_k - b_{k(i)}}{S_{(i)} \sqrt{(X^T X)^{-1}_{kk}}} \quad \text{for } k=0,1,2,\dots,p \quad \dots\dots\dots(25-2)$$

حيث ان:

b_k : معمل الانحدار رقم K المقدر باستخدام جميع الحالات n .

S_i : الخطأ المعياري لنموذج الانحدار الموفق باستخدام $n-1$ حالة، أي بعد اسقاط الحالة

رقم i .

$b_{k(i)}$: معامل الانحدار رقم K المقدر باستخدام $n-1$ حالة بعد حذف الحالة رقم i .

$(X^T X)^{-1}_{KK}$: العنصر القطري رقم K من المصفوفة $(X^T X)^{-1}$.

و لتشخيص الحالات المؤثرة علي قيمة معامل الانحدار رقم K اقترح (Neter, et al 1990)

معيار لتحديد الحالات المؤثرة في حالة العينات الصغيرة و المتوسطة و الكبيرة، كالاتي:

$$|DFBETAS_{k(i)}| \geq 1 \quad \dots\dots\dots(26-2)$$

اما في حالة العينات الكبيرة فتعتبر الحالة مؤثرة اذا كانت:

$$|DFBETAS_{k(i)}| \geq \frac{2}{\sqrt{n}} \quad \dots\dots\dots(27-2)$$

2- 5- 3 قياس الأثر علي كل معاملات الانحدار- مقياس مسافة كوك

:(Cook, 1980)

مقياس مسافة كوك (Cook's Distance Measure) هو مقياس اجمالي للتأثير المشترك

لقياس اثر حذف الحالة رقم i على جميع معاملات الانحدار المقدره، و يختلف مقياس كوك عن

مقياس ($DFBETAS$) في ان الأخير يقيس اثر حذف الحالة رقم i على كل معامل من

معاملات الانحدار على حده و يحسب مقياس كوك بتطبيق الصيغة التالية:

$$D_i = \frac{(b - b_{(i)})^T X^T X (b - b_{(i)})}{P \times MSE} \dots\dots\dots(28-2)$$

حيث ان:

b : متجه معاملات نموذج الانحدار باستخدام الحالات n ، $b_{(i)}$: متجه معاملات نموذج الانحدار بعد حذف الحالات رقم i ، P : عدد معالم النموذج، MSE : تباين نموذج الانحدار باستخدام كل الحالات n .

و عند حساب قيمة D_i حسب المعادلة (23 - 2) نلاحظ انه تحتاج الى بناء عدد n نموذج انحدار، و لصعوبة هذا الاجراء يمكن حساب قيمة D_i بصيغة أخرى بتطبيق المعادلة التالية:

$$D_i = \frac{e_i^2}{P \times MSE} \left(\frac{h_{ii}}{1 - h_{ii}} \right) \dots\dots\dots(29-2)$$

حيث ان:

e_i : قيمة البواقي العادية لنموذج الانحدار للمشاهدة رقم i ، h_{ii} : قيم العنصر القطري رقم i من المصفوفة H باستخدام جميع المشاهدات لتقدير دالة الانحدار.

و نلاحظ من المعادلة أعلاه (29 - 2) ان قيمة D_i تعتمد على كبر قيم الانعطاف h_{ii} او قيم البواقي e_i او كليهما، فكلما كانت قيمة أي من h_{ii} او e_i كبيرة تكون قيمة D_i ايضاً كبيرة. و لتشخيص اثر الحالة رقم i على جميع معاملات نموذج الانحدار يقترح (Fox, 1997) ان تتم مقارنة قيمة $(4/(n-p))$ فإذا كانت قيمة D_i اكبر من هذه القيمة تعتبر الحالة رقم i حالة غير مؤثرة على قيم معاملات نموذج الانحدار و بخلاف ذلك تكون الحالة غير مؤثرة، أي: تعتبر الحالة مؤثرة اذا كانت:

$$D_i \geq \frac{4}{n-p} \quad \dots\dots\dots(30-2)$$

2- 5- 4 قياس الأثر على الأخطاء المعيارية (Belsely, et al 1980):

يستخدم مقياس الأثر على الأخطاء المعيارية (Influence on Standardized Error) الذي طوره بيلسلي و آخرون لقياس اثر أي حالة على مصفوفة تباين و تغاير معاملات الانحدار المقدر، و هذا المقياس عبارة عن نسبة محددة مصفوفة تباين - تغاير معاملات الانحدار المقدر بعد حذف الحالة رقم i الى محددة مصفوفة تباين- تغاير معاملات الانحدار باستخدام جميع الحالات n و يحسب بالصيغة الآتية:

$$COVRATIO_i = \frac{\det(S_{(i)}^2 (X_{(i)}^T X_{(i)})^{-1})}{\det(S^2 (X^T X)^{-1})} \quad \dots\dots\dots(31-2)$$

حيث ان:

$S_{(i)}^2$: التباين المقدر بعد حذف الحالة رقم i .

$X_{(i)}$: مصفوفة البيانات بعد حذف المشاهدة رقم i من الدرجة $(p+1) \times (n-1)$.

و تم تطوير المعادلة في الصيغة (2- 31) الى الصيغة التالية:

$$COVRATIO_i = \frac{1}{(1-h_i) \left(\frac{n-p-1+d_i^{*2}}{n-p} \right)^p} \quad \dots\dots\dots(32-2)$$

حيث ان:

P : عدد معالم النموذج ، n : عدد المشاهدات ، h_{ii} قيم العنصر القطري رقم i من المصفوفة المقدر H باستخدام جميع المشاهدات لتقدير دالة الانحدار ، d_i^* : باقي ستيدنت المحذوف للمشاهدة رقم i باستخدام جميع المشاهدات.

و نلاحظ من من الصيغة أعلاه ان قيمة هذا المقياس تعتمد علي قيمتي الرافعة h_{ii} و باقي ستيوذنت d_i حيث تزيد قيمة المقياس بزيادة h_i و تتخفض بانخفاض d_i .
و لتشخيص الحالة رقم d_i على الأخطاء المعيارية لمعاملات الانحدار اقترح بيلسلي و آخرون مقارنة قيمة $COVRATIO$ المطلقة بالقيمة $(1+3(p)/n)$ فإذا كانت قيمة $COVRATIO$ المطلقة اكبر من هذه القيمة تعتبر الحالة رقم t مؤثرة على قيم الأخطاء المعيارية لمعاملات الانحدار، أي تعتبر الحالة مؤثرة اذا كانت قيمة $COVRATIO$ تقع خارج المدى:

$$1 - \frac{3(p)}{n} \geq COVRATIO \geq 1 + \frac{3(p)}{n} \quad \dots\dots\dots(33-2)$$

6-2 بعض الحلول المقترحة لمعالجة القيم الشاذة:

بعد التأكد من وجود القيم الشاذة في مشاهدات المتغير التابع او مشاهدات المتغيرات المستقلة لا بد من معالجة هذه القيم لمعرفة مصادرها و الأسباب التي أدت الى وجودها و ذلك من اجل الحصول على نموذج انحدار تكون نتائجه خالية من تأثير وجود القيم الشاذة على معلماته المقدره فإذا تبين بعد الفحص و الدراسة ان وجود المشاهدات الشاذة ناتج عن خطأ في مرحلة جمع البيانات او في مرحلة المعالجة فعلى الباحث تصحيح هذه الأخطاء و إعادة حل النموذج، اما اذا كانت المشاهدات الشاذة بيانات حقيقية فهناك عدد من الحلول المقترحة نذكر منها ما يلي (إسماعيل، 2001) :

1. إعادة توصيف النموذج (Model re-specification) و ذلك اما بإضافة او حذف متغيرات مستقلة.

2. اجراء تحويلات اما للمتغير التابع او للمتغيرات المستقلة او بعض منها كتحويل اللوغاريتم و المعكوس الجذر التربيعي و غيرها.

3. حذف المشاهدات الشاذة اذا كان حجم العينة كبيراً و إعادة توفيق النموذج.

4. جمع بيانات إضافية لزيادة حجم العينة و بالتالي تقليل اثر وجود المشاهدات الشاذة الا ان هذا الحل قد يصاحبه بعض المصاعب كأرتباط الظاهرة محل الدراسة بفترة زمنية محددة مما يعني الانتظار حت (ى حلول الظاهرة مرة أخرى و قد تكون هذه الفترة غير معلومة كحدوث الكوارث فضلاً عن التكلفة المضافة لجمع هذه البيانات.

5. طريقة متوسط البتر (Trimed Mean) و تتلخص خطوات اجراء هذه الطريقة بترتيب مشاهدات العمود المشخص بأنه شاذ تصاعدياً او تنازلياً، و تحذف اكبر قيمة و اصغر قيمة في بيانات العمود الشاذ، ثم إيجاد الوسط الحسابي للقيم المتبقية أي إيجاد (Trimed Mean) ل ($n-2$) من القيم و الذي يمثل تقديراً لهذه القيمة (الجبوري، 1998).

6. كذلك لمعالجة وجود المشاهدات الشاذة في مجموعة البيانات المراد تحليلها يتم استخدام طرائق التقدير الحصينة بدلا من طريقة المربعات الصغرى وذلك لوجود تأثير للشواذ على نتائج طريقة المربعات الصغرى (الدباغ، 1999). حيث تستخدم طريقة المربعات الصغرى الموزونة للحصول على التقديرات الحصينة على وفق الصيغة الآتية:

$$b_R = (X'WX)^{-1} X'WY \quad \text{و لإيجاد} \quad \dots\dots\dots(34-2)$$

التقديرات الحصينة لمعاملات المعادلة (2 - 34) تستخدم طريقة المربعات الصغرى الموزونة الضبابية (Kamile, and Aysen, 2004) و كالاتي:

$$a_R = (X'WX + C'WC + D'WD)^{-1} (X'WY + C'WE + D'WF) \quad \dots\dots\dots(35-2)$$

حيث ان W تمثل مصفوفة الأوزان التي يتم من خلالها تحديد وزن لكل مشاهدة يمثل مدى مساهمتها في المعلومات التي ستستخدم في عملية تقدير المعلمات، وكذلك يمثل هذا الوزن معيار لقياس مدى اقتراب أو ابتعاد المشاهد عن البيانات.

7-2 اختبار الفرضيات **Test of Hypotheses** (صافي، 2008):

يعتبر موضوع اختبار الفرضيات الإحصائية من أهم الموضوعات في مجال اتخاذ القرارات وسنبدأ بذكر بعض المصطلحات الهامة في هذا المجال.

1- الفرضية الإحصائية

هي عبارة عن ادعاء قد يكون صحيحاً أو خطأ حول معلمة أو أكثر لمجتمع أو لمجموعة من المجتمعات.

تقبل الفرضية في حالة أن بيانات العينة تساند النظرية، وترفض عندما تكون بيانات العينة على النقيض منها، وفي حالة عدم رفضنا للفرضية الإحصائية فإن هذا ناتج عن عدم وجود أدلة كافية لرفضها من بيانات العينة ولذلك فإن عدم رفضنا لهذه الفرضية لا يعنى بالضرورة أنها صحيحة، أما إذا رفضنا الفرضية بناء على المعلومات الموجودة في بيانات العينة فهذا يعنى أن الفرضية خاطئة، ولذلك فإن الباحث يحاول أن يضع الفرضية بشكل يأمل أن يرفضها، فمثلاً إذا أراد الباحث أن يثبت بأن طريقة جديدة من طرق التدريس أحسن من غيرها فإنه يضع فرضية تقول بعدم وجود فرق بين طرق التدريس.

إن الفرضية التي يأمل الباحث أن يرفضها تسمى بفرضية العدم (الفرضية المبدئية) ويرمز لها بالرمز H_0 ، ورفضنا لهذه الفرضية يؤدي إلى قبول فرضية بديلة عنها تسمى الفرضية البديلة ويرمز لها بالرمز H_1 .

2- مستوى المعنوية أو مستوى الاحتمال

وهي درجة الاحتمال الذي نرفض به فرضية العدم H_0 عندما تكون صحيحة أو هو احتمال الوقوع في الخطأ من النوع الأول ويرمز له بالرمز α ، وهي يحددها الباحث لنفسه منذ البداية وفي معظم العلوم التطبيقية نختار α مساوية 1% أو 5% على الأكثر.

3- دالة الاختبار الإحصائية

عبارة عن متغير عشوائي له توزيع احتمالي معلوم وتصف الدالة الإحصائية العلاقة بين القيم النظرية للمجتمع والقيم المحسوبة من العينة.

4- القيمة الاحتمالية: (Sig. or P-value)

احتمال الحصول على قيمة أكبر من أو تساوي (أقل من أو تساوي) إحصائية الاختبار المحسوبة من بيانات العينة أخذاً في الاعتبار توزيع إحصائية الاختبار بافتراض صحة فرض العدم H_0 وطبيعة الفرض البديل H_1 . ويتم استخدام القيمة الاحتمالية لاتخاذ قرار حيال فرض العدم.

2- 7 - 1 خطوات اختبار الفرضيات:

1- تحديد نوع توزيع المجتمع.

يجب تحديد ما إذا كان المتغير العشوائي الذي يتم دراسته يتبع التوزيع الطبيعي أم توزيع بواسون أم توزيع ذو الحدين أم غيره من التوزيعات الاحتمالية المتصلة أو المنفصلة، معظم التوزيعات الاحتمالية يكون توزيعها مشابهاً للتوزيع الطبيعي خاصة إذا كان حجم العينة كبيراً.

هناك نوعان من الطرق الإحصائية التي تستخدم في اختبار الفرضيات:

(أ) الاختبارات المعلمية: وتستخدم في حالة البيانات الرقمية التي توزيعها يتبع التوزيع الطبيعي.
(ب) الاختبارات غير المعلمية: وتستخدم في حالة البيانات الرقمية التي توزيعها لا يتبع التوزيع الطبيعي طبيعي، وكذلك في حالي البيانات الترتيبية والوصفية.

2- صياغة فرضيتا العدم والبديلة.

مثلاً: عند اختبار أن متوسط المجتمع μ يساوى قيمة معينة μ_0 مقابل الفرضية القائلة بأن μ لا

يساوى μ_0 ، فإن فرضية العدم H_0 والفرضية البديلة H_1 تكون على النحو التالي:

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

3- اختيار مستوى المعنوية α .

4- اختيار دالة الاختبار الإحصائية المناسبة.

5- جمع البيانات من العينة وحساب قيمة دالة الاختبار الإحصائية.

6- اتخاذ القرارات.

حيث نرفض H_0 ونقبل H_1 إذا كانت قيمة الاحتمال (Sig. or P-value) أقل من أو تساوي

مستوى المعنوية (α)، أما إذا كانت قيمة الاحتمال أكبر من α فلا يمكن رفض H_0 .

وبرنامج SPSS يعطي Sig. 2-tailed فبالتالي نرفض فرضية العدم H_0 عندما تكون

$$P - Value(Sig.) < \alpha$$

2- 7 - 2 اختبار T في حالة اختبار فرضيات متعلقة بمتوسط واحد:

إذا كان المطلوب اختبار فرضية العدم $H_0 : \mu = \mu_0$ على مستوى دلالة α مقابل

$$H_1 : \mu \neq \mu_0 \quad (1)$$

$$H_1 : \mu > \mu_0 \quad (2)$$

$$H_1 : \mu < \mu_0 \quad (3)$$

2- 7 - 3 اختبارات الفروق بين متوسطين مجتمعين مستقلين:

في هذه الحالة نأخذ عينة عشوائية من توزيع طبيعي $N(\mu_1, \sigma_1^2)$ ، وعينة عشوائية أيضاً من توزيع طبيعي $N(\mu_2, \sigma_2^2)$ ومستقل عن التوزيع الأول، وتكون $\sigma_1^2 = \sigma_2^2$ ولكنهما مجهولتان. إذا كان المطلوب اختبار فرضية العدم $H_0: \mu_1 - \mu_2 = 0$ على مستوى دلالة α مقابل

$$H_1: \mu_1 - \mu_2 \neq 0 \quad (1)$$

$$H_1: \mu_1 - \mu_2 > 0 \quad (2)$$

$$H_1: \mu_1 - \mu_2 < 0 \quad (3)$$

2-7-4 اختبارات الفروق بين متوسطي مجتمعين من عينات مرتبطة:

في هذه الحالة تكون البيانات مزدوجة، أي أن العينتين مرتبطتان حيث أن البيانات تكون على شكل أزواج وبالتالي فإن حجم العينتين لابد أن يكون متساوياً.

إذا كان المطلوب اختبار فرضية العدم $H_0: \mu_1 - \mu_2 = 0$ على مستوى دلالة α مقابل

$$H_1: \mu_1 - \mu_2 \neq 0 \quad (1)$$

$$H_1: \mu_1 - \mu_2 > 0 \quad (2)$$

$$H_1: \mu_1 - \mu_2 < 0 \quad (3)$$

الفصل الثالث

0-3 تمهيد:

1-3 الطريقة المقترحة لتشخيص القيم الشاذة:

2-3 المقياس البياني لتحديد عدد المشاهدات الشاذة:

3-3 الطريقة المقترحة لمعالجة المشاهدات الشاذة:

1-3-3 كيفية التعامل مع المشاهدات ذات التأثير الجامح السلبي:

2-3-3 كيفية التعامل مع المشاهدات ذات التأثير الجامح الايجابي:

0-3 تمهيد:

من اجل إزالة اثر المشاهدات الشاذة على نموذج الانحدار الخطي المتعدد تم اقتراح طرق جديدة تتضمن طريقة لتشخيص المشاهدات الشاذة مصحوبة بمقياس بياني لتحديد عدد المشاهدات الشاذة و طريقة لمعالجة القيم الشاذة تستبدل هذه المشاهدات الشاذة بقيم جديدة، أي انه لتشخيص القيم الشاذة الموجودة في نماذج الانحدار الخطي البسيط و المتعدد تم تصميم طريقة جديدة لتشخيص المشاهدات الشاذة و لتقدير هذه القيم الشاذة تم تصميم طريقة لتقليل اثر المشاهدات الشاذة (طريقة المعالجة) على نموذج الانحدار الخطي المتعدد، حيث تعمل طريقة التشخيص علي اكتشاف كل المشاهدات الشاذة باستخدام المقياس البياني (المقترح) و تعمل طريقة المعالجة وفق معيار نسبي أي انه اذا كان عدد المشاهدات الشاذة قليل يتم تقليل اثرها بدرجة كبيرة و اذا كان عدد المشاهدات الشاذة كبير يتم تقليل اثرها بدرجة اقل، بعبارة أخرى يفترض اذا تم تطبيق هذه الطرق فأن مؤشرات النموذج تتحسن بدرجة نسبية عكسية لنسبة المشاهدات الشاذة الي كل المشاهدات و نتناول بالشرح هذه الطرق و كالاتي:—

1-3 الطريقة المقترحة لتشخيص القيم الشاذة (حمد، 2019):

تفترض هذه الطريقة ان مؤشرات نموذج الانحدار الخطي المتعدد يطرأ عليها تحسن كبير في حال حذف كل المشاهدات الشاذة، تم تصميم هذه الطريقة بناءً على الصيغة المعممة التالية:—

$$Z_{ij} = Y_{ij} - \hat{\beta}_{1j} X_{1ij} - \hat{\beta}_{2j} X_{2ij} - \dots - \hat{\beta}_{kj} X_{kij} \rightarrow (1-3)$$

حيث $Z_{ij} = \hat{\beta}_{0j} + e_{ij}$ و $i = 1, 2, \dots, n; j = 1, 2, \dots, m$ حيث تمثل n عدد المشاهدات للنموذج و تمثل k عدد المتغيرات المستقلة و تمثل m عدد مرات توفيق النموذج، و بناءً على الصيغة (1-3) - المعممة يتم حذف مشاهدة شاذة في كل مرة يتم فيها توفيق النموذج من المشاهدة الشاذة

ذات الأثر الأكبر و حتي المشاهدة الشاذة ذات الأثر الأقل، و ذلك من خلال توفيق النموذج بالمشاهدات n_j (عدد مشاهدات النموذج بعد إعادة توفيقه للمرة رقم j) و تعوض قيم $\hat{\beta}_{1j}, \hat{\beta}_{2j}, \dots, \hat{\beta}_{kj}$ في الصيغة (3 - 1) و حساب قيمة Z_{ij} و ترتيبها تصاعدياً او تنازلياً و من ثم حذف المشاهدة المناظرة للخطأ المعياري المطلق الأكبر أي ل $|e_{ij}| = |Z_{ij} - \hat{\beta}_{0j}|$ الأكبر، و يتكرر هذا الاجراء الى ان يحدد المقياس البياني عدد المشاهدات الشاذة.

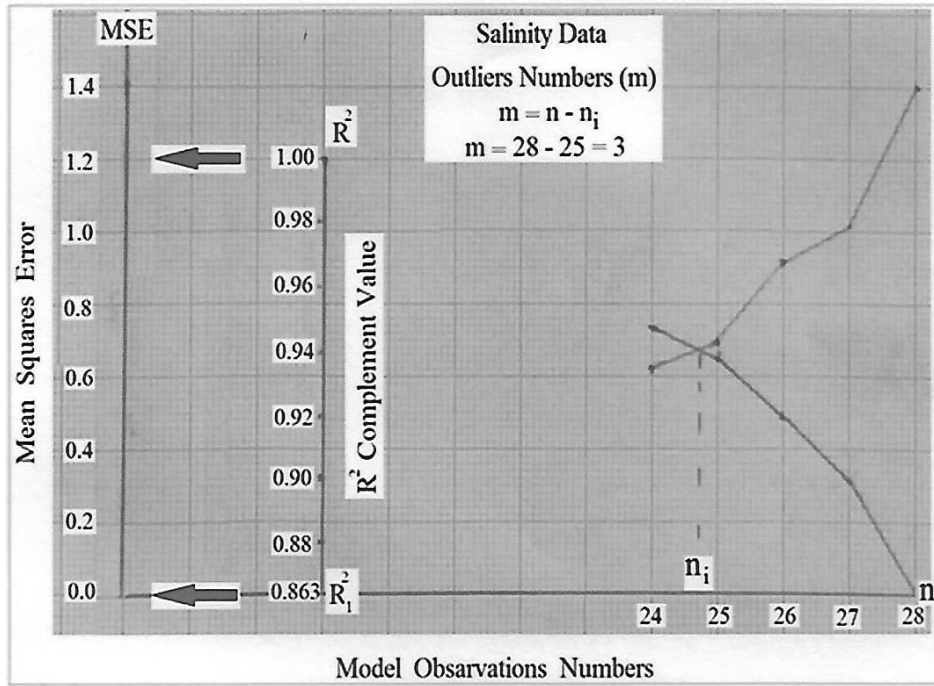
3- 2 المقياس البياني لتحديد عدد المشاهدات الشاذة (حمد، 2019):

و بما انه يتم توفيق النموذج في كل محاولة لحذف مشاهدة شاذة من المفترض ان يتناقص متوسط مربع الخطأ المعياري MSE_j (متوسط مربع الخطأ المعياري للنموذج بعد إعادة توفيقه للمرة رقم j) و يتزايد R_j^2 (معامل التحديد البسيط للنموذج بعد إعادة توفيقه للمرة رقم j) و ينقص عدد المشاهدات n_j بمعدل مشاهدة واحدة في كل مرة، و ذلك استناداً الى الحقيقة المسلم بها و حسب الدراسات السابقة انه كلما قل وجود المشاهدات الشاذة كلما قل متوسط مربعات الخطأ MSE و ارتفعت قيمة معامل التحديد البسيط R^2 .

عليه يمكن تصميم المقياس البياني عن طريق تمثيل نقاط المنحنى (n_j, MSE_j) و نقاط المنحنى (n_j, R_j^2) و نسبة لوجود محوران عموديان محور MSE و محور R^2 و محور افقي مشترك (n) ، و لكي ما يؤدي هذا المقياس هدفه المنشود يجب وضع الواحد الصحيح الذي يمثل اعلى قيمة لمحور R^2 بمحاذاة قيمة نتيجة الضرب $(MSE_1 * R_1^2)$ حيث R_1^2 معامل التحديد البسيط لبيانات النموذج الاصلية و MSE_1 متوسط مربعات الخطأ لبيانات النموذج الاصلية على محور MSE و وضع قيمة R_1^2 بمحاذاة نقطة الأصل و من ثم تقسيم المسافة $(1 - R_1^2)$ بمقياس الرسم الذي يتناسب مع مقياس الرسم الذي تم تحديده لمحور MSE اما بالنسبة لمحور (n) المشترك يتم تقسيمه بالوضع الذي يلائم نصف عدد المشاهدات الكلي او اكثر بقليل اذ لا

يعقل ان يتعدى عدد المشاهدات الشاذة نصف العدد الكلي للملاحظات، و من نقطة تقاطع المنحنيان و النزول عمودياً على محور (n) عند القيمة n_j التي تعبر عن عدد المشاهدات الغير شاذة و بعد تقريبها لاقرب عدد صحيح يكون عدد المشاهدات الشاذة مساوياً لعدد مرات توفيق النموذج m أي ان $m = n - n_j$ و الشكل (3 - 1) يوضح تصميم المقياس البياني:-

الشكل (3 - 1) : تصميم المقياس البياني



المصدر : اعداد الباحث باستخدام ورق الرسم البياني و برنامج الرسام

أي انه بالنظر الى تصميم المقياس البياني الشكل (3 - 1) نلاحظ انه صمم على مبدأ العلاقة العكسية بين مؤشر R^2 و مؤشر MSE و مبدأ النسبية لنفس المؤشرين أي ان نتيجة الضرب $(MSE_1 * R_1^2)$ تمثل نسبة من قيمة MSE ، حيث R_1^2 معامل التحديد البسيط لبيانات النموذج الاصلية.

3-3 الطريقة المقترحة لمعالجة المشاهدات الشاذة (حمد، 2019):

في هذه الطريقة يقترح الباحث ترجيح الأخطاء المعيارية e_i لكل من المشاهدات الشاذة و الغير شاذة بنسبتيهما لنحصل علي أخطاء معيارية e_i^* جديدة تقابلها منظومة مشاهدات $(x_{1i}^*, x_{2i}^*, \dots, x_{ki}^*, y_i^*)$ جديدة، أي ان $e_i^* = e_i(n_j/n)$ للغير شاذة، $e_i^* = e_i(m/n)$ للشاذة، و هذا الاجراء يغير كل مشاهدات النموذج! و بضرب النسب الترجيحية للنوعين من المشاهدات (للشاذة و الغير شاذة) في مقلوب نسبة الترجيح للمشاهدات الغير شاذة أي ان $e_i^* = e_i(n_j/n)(n/n_j) = e_i$ للغير شاذة تقابلها منظومة المشاهدات $(x_{1i}^*, x_{2i}^*, \dots, x_{ki}^*, y_i^*)$ للبيانات الاصلية، و $e_i^* = e_i(m/n)(n/n_j) = e_i(m/n_j)$ للشاذة تقابلها منظومة المشاهدات $(x_{1i}^*, x_{2i}^*, \dots, x_{ki}^*, y_i^*)$ الجديدة. هذه الطريقة تعتمد على الصيغة (1) المعممة كذلك و لكن يتم توفيق النموذج لمرة واحدة بالمشاهدات الغير شاذة n_j فقط و تعوض قيم $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$ في الصيغة (3 - 1) المعممة، و كذلك $Z_i = \hat{\beta}_0 + e_i$ أي انه يتم تعويض $\hat{\beta}_0$ للنموذج الموفق بالمشاهدات الغير شاذة n_j ايضاً، و بما ان النموذج يتم توفيقه لمرة واحدة تكتب الصيغة (3 - 1) المعممة على الصورة التالية:—

$$Z_i = Y_i - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} - \dots - \hat{\beta}_k X_{ki} \quad \rightarrow (2-3)$$

و لكن يتم حساب Z_i لكل المشاهدات بما في ذلك المشاهدات الشاذة، و بوضع $d_i = e_i$ تمثل انحرافات Z_i عن $\hat{\beta}_0$ أي ان $d_i = e_i = Z_i - \hat{\beta}_0$ ، و بوضع $\Pr = m/n_j$ و ضربها في d_i نحصل على d_i^* أي ان $d_i^* = Z_i^* - \hat{\beta}_0 = d_i^*$ و $d_i(\Pr) = e_i(m/n_j) = e_i^*$ و بما انه يمكننا حساب $Z_i^* = d_i^* + \hat{\beta}_0$ ، فإنه يمكننا بعد ترتيبها تصاعدياً او تنازلياً حساب منظومة المشاهدات الشاذة $(x_{1i}^*, x_{2i}^*, \dots, x_{ki}^*, y_i^*)$ و كالآتي:—

$$\forall (y_i < y_{i+1}) y_i^* = y_i + (y_{i+1} - y_i)(Z_i - Z_i^*) / (Z_i - Z_{i+1}) \rightarrow (3-3)$$

$$\forall (y_i > y_{i+1}) y_i^* = y_i - (y_i - y_{i+1})(Z_i - Z_i^*) / (Z_i - Z_{i+1}) \rightarrow (4-3)$$

$$\forall (x_{hi} < x_{hi+1}) x_{hi}^* = x_{hi} + (x_{hi+1} - x_{hi})(Z_i - Z_i^*) / (Z_i - Z_{i+1}) \rightarrow (5-3)$$

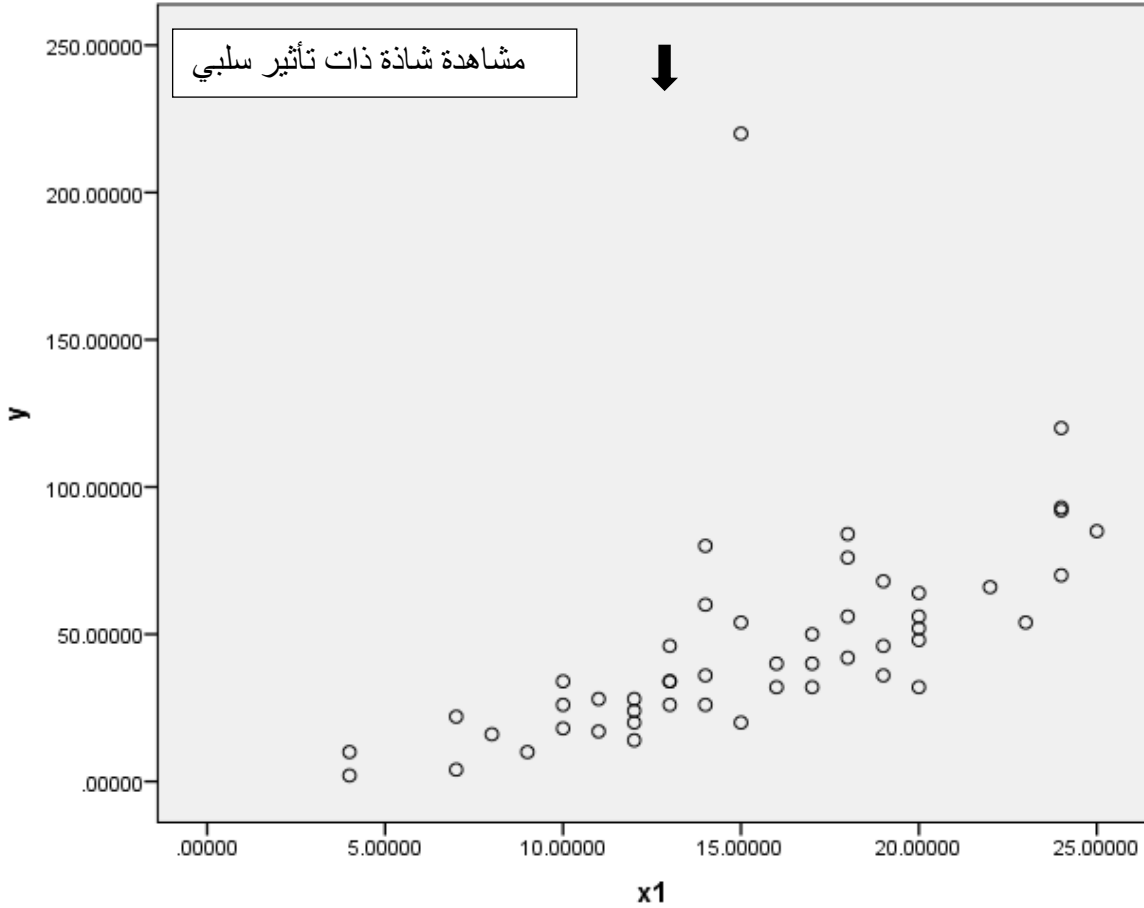
$$\forall (x_{hi} > x_{hi+1}) x_{hi}^* = x_{hi} - (x_{hi} - x_{hi+1})(Z_i - Z_i^*) / (Z_i - Z_{i+1}) \dots \rightarrow (6-3)$$

حيث $i = 1, 2, \dots, n$; $h = 1, 2, \dots, k$

1-3-3 كيفية التعامل مع المشاهدات ذات التأثير الجامح السلبي (حمد، 2019):

نود ان ننبه الى ان طريقة التشخيص قد تسلك منحى مختلف في تشخيص المشاهدات الشاذة خاصة اذا تضمنت مجموعة البيانات - عند رسم شكل انتشار البيانات لاي من المتغيرات المستقلة و المتغير التابع - مشاهدة او اكثر من المشاهدات الشاذة ذات التأثير الجامح السلبي و التي غالباً ما تظهر بصورة معزولة عن اغلب المشاهدات، لذا يجب الانتباه لمتل هذه المشاهدات من خلال شكل انتشار البيانات (Scatar Plot) قبل بداية التشخيص و اعطاءها قيم تجعلها قريبة من اغلب المشاهدات أي بالوضع الذي يجعل شذوذها معقول و ليس جامحا هذا النوع من المشاهدات يظهر من ضمن المشاهدات الشاذة بعد التشخيص، هذا النوع من المشاهدات يظهر اعلى او ادنى خط انحدار البيانات بمسافة بعيدة تختلف بدرجة كبيرة عن مسافات انتشار البيانات عن بعضها البعض و الشكل (3 - 2) يوضح مثلاً لبيانات بها مشاهدة شاذة واحدة ذات تأثير جامح سلبي .

الشكل (3 - 2) : انتشار البيانات في حال ظهور مشاهدة شاذة ذات تأثير جامح سلبي



المصدر : اعداد الباحث باستخدام برنامج SPSS

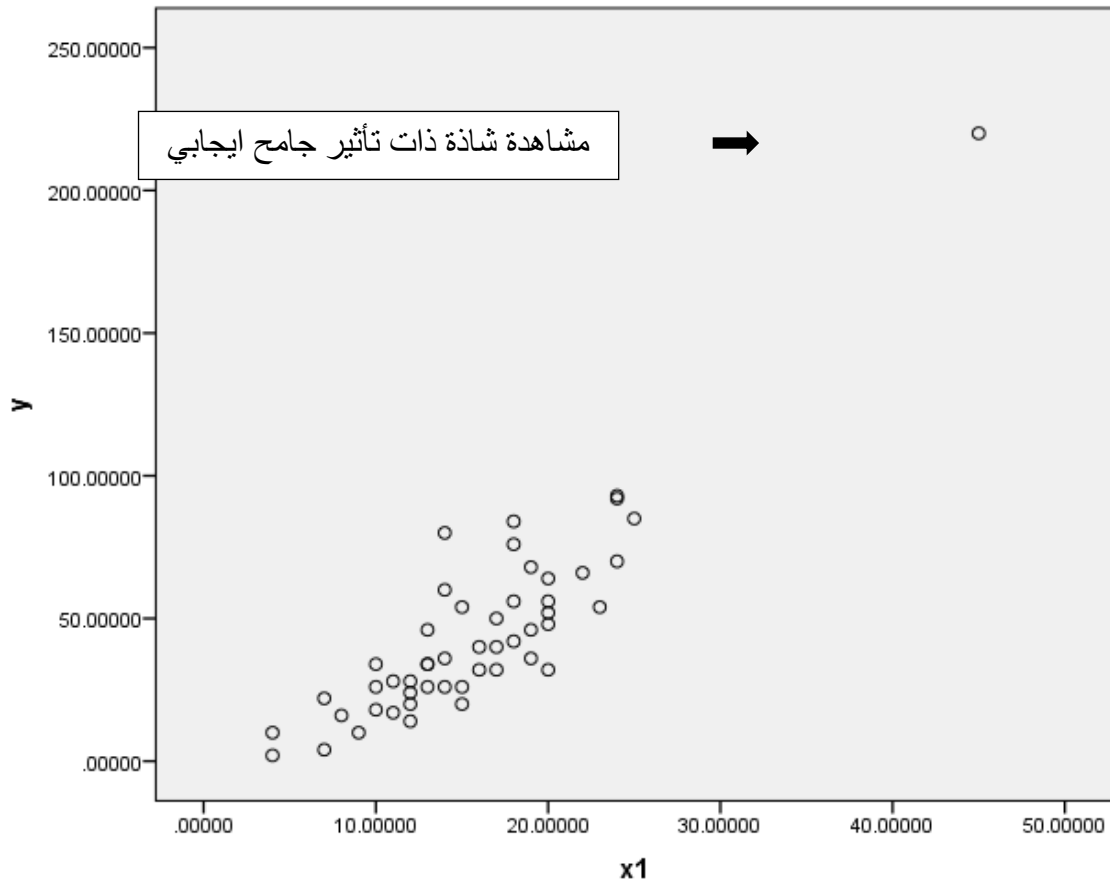
3-3-2 كيفية التعامل مع المشاهدات ذات التأثير الجامح الإيجابي (حمد،

2019):

نود ان ننبه كذلك الى ان طريقة التشخيص قد تسلك منحى مختلف في تشخيص المشاهدات الشاذة خاصة اذا تضمنت مجموعة البيانات - عند رسم شكل انتشار البيانات لاي من المتغيرات المستقلة و المتغير التابع - مشاهدة او اكثر من المشاهدات الشاذة ذات التأثير الجامح الايجابي و التي غالباً ما تظهر بصورة معزولة عن اغلب المشاهدات، و تقع علي او قريبة من خط الانحدار الامر الذي يجعل مؤشرات النموذج تبدو جيدة ظاهرياً، لذا يجب الانتباه لمثل هذه المشاهدات من خلال شكل انتشار البيانات (Scatar Plot) قبل بداية التشخيص و اعطاءها قيم تجعلها قريبة من

اغلب المشاهدات و اذا ظهرت بعد المعالجة يجب تصنيفها من ضمن الش المشاهدات الشاذة واذ و معالجتها، هذا النوع من المشاهدات يظهر في اتجاه خط انحدار البيانات بمسافة بعيدة تختلف بدرجة كبيرة عن مسافات انتشار البيانات عن بعضها البعض و الشكل (3 - 3) يوضح مثلاً لبيانات بها مشاهدة شاذة واحدة ذات تأثير جامح ايجابي.

الشكل (3 - 3) : انتشار البيانات في حال ظهور مشاهدة شاذة ذات تأثير جامح ايجابي



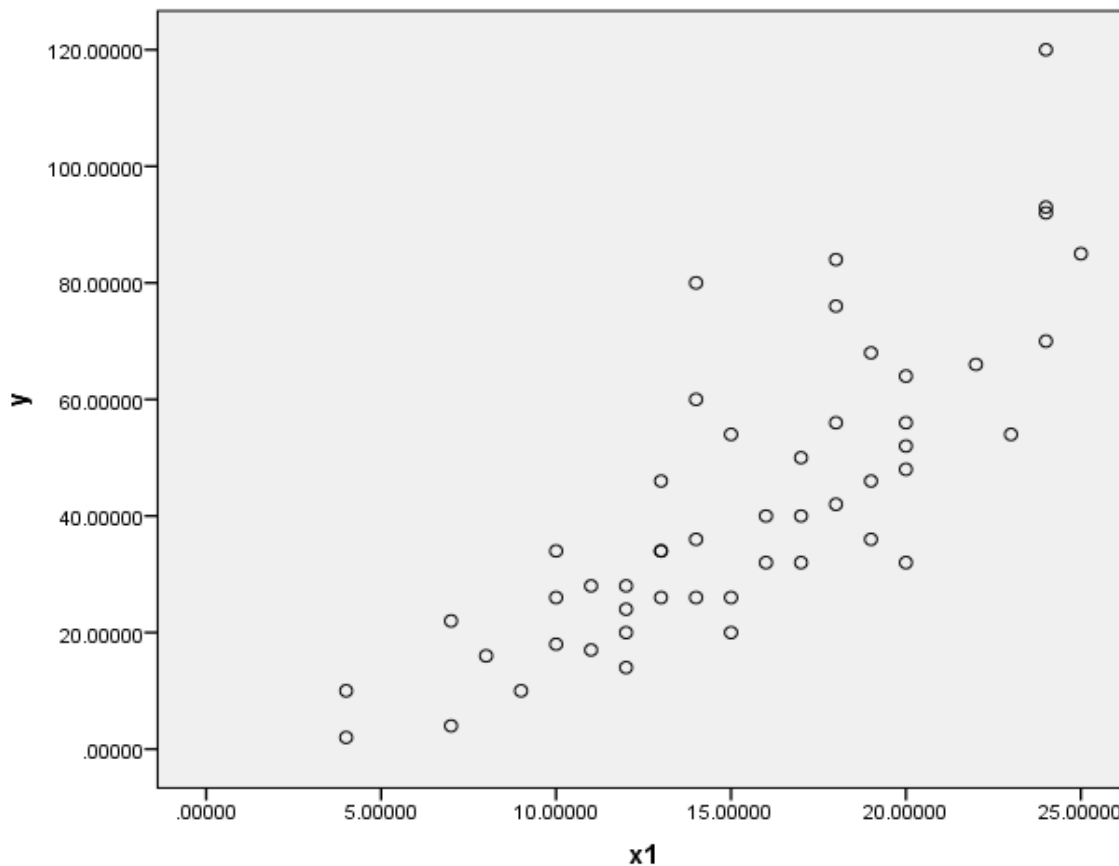
1

لمصدر : اعداد الباحث باستخدام البرنامج SPSS

و الشكل (3 - 4) يوضح شكل انتشار البيانات بعد إعطاء المشاهدات ذات التأثير الجامح السلبي و التأثير الجامح الإيجابي قيماً جعلتها تظهر بمسافة متجانسة مع مسافات انتشار البيانات من بعضها البعض، و لا يعني ذلك ان هذا الشكل الذي يمثل هذه البيانات خالي من المشاهدات

الشاذة، بل يعني ذلك انه اصبح بحالة تسمح لنا بتشخيص و معالجة المشاهدات الشاذة بالطرق المقترحة دون ان تسلك هذه الطرق منحى مختلف في تشخيص المشاهدات الشاذة.

شكل (3 - 4) : انتشار البيانات بعد إعطاء القيم ذات الشذوذ الجامح قيم تقريبية



المصدر : اعداد الباحث باستخدام البرنامج SPSS

و الجدول (3 - 1) يبين نتائج توفيق بيانات نموذج الانحدار بالنسبة للاشكال الثلاث السابقة

فيما يتعلق بقيم المؤشرات (R^2 و MSE و F و Sig_F و β_0 و Sig_{β_0} و β_1 و Sig_{β_1}).

جدول (3 - 1) : نتائج توفيق بيانات نموذج الانحدار بالنسبة للاشكال الثلاث

نتائج البيانات	R^2	MSE	F	Sig_F	β_0	Sig_{β_0}	β_1	Sig_{β_1}
شكل (3 - 2)	.328	880.302	23.378	.000b	-12.827	.330	3.876	.000
شكل (3 - 3)	.788	254.787	178.632	.000b	-27.582	.000	4.587	.000
شكل (3 - 4)	.651	236.532	89.567	.000b	-17.579	.012	.416	.000

المصدر : اعداد الباحث باستخدام البرنامج SPSS

بالنظر للجدول (3 - 1) أعلاه نلاحظ ان مؤشرات النموذج بالنسبة للشكل (3 - 3) بصورة عامة تتحسن اكثر منه في مؤشرات نموذج الاشكال (3 - 4) و (3 - 2)، و ذلك بسبب مشاهدة شاذة واحدة ظهرت في اتجاه خط الانحدار بعيدة عن اغلب المشاهدات الامر الذي حدا بالباحث لوصف هذه المشاهدة بالمشاهدة الشاذة ذات التأثير الجامح الإيجابي، و كلمة إيجابي لاتعني جودة النتائج بناءً على وجود مثل هذا النوع من المشاهدات بما انه قد تم وصفها بالشذوذ. و بالنسبة لمؤشرات نموذج الشكل (3 - 2) ظهرت بالشكل الأسوأ و ذلك بسبب مشاهدة شاذة واحدة ظهرت اعلى خط الانحدار بعيدة عن اغلب المشاهدات الامر الذي حدى بالباحث لوصفها بالمشاهدة ذات التأثير الجامح السلبي. و تعني كلمة جامح ان هذا النوع من المشاهدات له تأثير كبير و بصورة مبالغ فيها سواء كان ايجابياً او سلبياً، لذ لزم تخفيف هذا الجموح اولاً لتبدو النتائج معقولة نوعاً ما كما في نتائج الشكل (3 - 4) و التي لاتعني خلو النموذج من المشاهدات الشاذة.

الفصل الرابع

التطبيق العملي

0 -4 تمهيد:

1 -4 التطبيق العملي بغرض المقارنة:

2 -4 التطبيق العملي علي نماذج الانحدار الخطي البسيط:

3 -4 اختبارات الفروض فيما يتعلق بنماذج الانحدار الخطي البسيط:

4 -4 التطبيق العملي علي نماذج الانحدار الخطي المتعدد:

5 -4 اختبارات الفروض فيما يتعلق بنماذج الانحدار الخطي المتعدد:

4-0 تمهيد:

في هذا القسم سنبين قدرة الطرق المقترحة في تشخيص و معالجة المشاهدات الشاذة في نموذج الانحدار الخطي البسيط و ذلك من اجل المقارنة مع الطرق السابقة على نفس البيانات التي تم استخدامها في الطرق السابقة. في جانبين، الجانب الأول مقارنتها بالطرق السابقة و الجانب الثاني يتعلق باختبار كفاءتها و منهجيتها. كما سنبين قدرة الطرق المقترحة في تشخيص و معالجة المشاهدات الشاذة في اكثر من نموذج واحد من نماذج الانحدار الخطي البسيط و في اكثر من نموذج واحد من نماذج الانحدار الخطي المتعدد. بالإضافة الي نظم هذه المؤشرات للنوعين من النماذج كل على حدة، و من ثم اجراء الإختبارات المتعلقة بالكفاءة النسبية و إختبار موضوعية هذه الطرق في التعامل مع المشاهدات الشاذة، و كما سنحاول تقدير العدد المتوقع للمشاهدات الشاذة في نموذج الانحدار الخطي.

4-1 التطبيق العملي بغرض المقارنة:

تم تطبيق طرق التشخيص و المعالجة على بيانات (Mickey, et al 1967) كما وردت عند (الجبوري، و آخرون 2002) و التي استخدمها (الشميري، و آخرون 2014). عليه تتم مقارنة الطرق المقترحة بطرق الجبوري و التي استخدم فيها للتشخيص طريقة القطع الناقص و عدد من طرق المعالجة، و طريقة الشميري التي استخدم فيها للتشخيص حد الخطأ العشوائي e_i و للمعالجة موجه حد الخطأ المقدر كما نشير في جانب تطبيق الطرق المقترحة الي انه تم تطبيق خطوات طريقة التشخيص من خلال برنامج التحليل الاحصائي SPSS و المقياس البياني باستخدام ورق الرسم البياني و بالإضافة لبرنامج الجداول الالكترونية Excel الذي استوعب كل الصيغ الرياضية لطريقتي التشخيص و المعالجة، و كذلك تم مراعاة المشاهدات ذات التأثير الجامح، و الجدول (4 - 1) يوضح عملية المقارنة :-

الجدول (4 - 1): نتائج المقارنة لقيم المؤشرات (R^2 ; MSE) و الكفاءة النسبية للطرق من حيث

رفع قيمة R^2 و تقليل قيمة MSE للبيانات بعد التشخيص و المعالجة

توضيح	اسم طريقة التشخيص	معامل التحديد البسيط R^2	متوسط مربعات الخطأ MSE	الكفاءة النسبية من حيث		ارقام المشاهدات الشاذة	عدد المشاهدات الشاذة
				تقليل MSE	رفع R^2		
البيانات العادية	Or.	0.41	121.51	1.000	1.000	-	0
افضل طرق الجبوري و ناسي (2002)	NSR-1	0.565	58.674	2.071	1.378	3,13,11,14,18,19,20	7
	NSR-4	0.581	70.605	1.721	1.417	19	1
	NSR-6	0.441	72.642	1.673	1.076	18,19	2
طريقة الباحث 2005	البيانية	0.744	38.924	3.122	1.815	3,13,14,19	4
طريقة الباحث 2005 المعدلة	البيانية المعدلة	0.605	47.998	2.532	1.446	3,13,14,19,18	5
طريقة الشميري (2014)	الخطأ المقدر	0.56	73.7	1.648	1.366	19	1
طرق الباحث الحالية	الطرق المقترحة	0.711	36.995	3.284	1.734	19,13,3,14,20,18	6

المصدر: اعداد الباحث باستخدام البرنامج SPSS و برنامج الجداول الالكترونية Excel

الجدول (4 - 1) يبين نتائج المقارنة للطرق المقترحة الحالية مع الطرق السابقة من خلال

استخدام مقياس الكفاءة النسبية من حيث رفع قيمة معامل التحديد البسيط R^2 و خفض قيمة متوسط مربعات الخطأ العشوائي MSE ، كما نشير الى ان مقياس الكفاءة النسبية لرفع قيمة R^2 تم حسابه بقسمة قيمة R^2 للبيانات بعد المعالجة باي من الطرق على قيمة R^2 للبيانات الاصلية و بالنسبة لتقليل قيمة MSE تم حسابه بقسمة قيمة MSE للبيانات الاصلية على قيمة MSE للبيانات بعد المعالجة باي من الطرق لاحظ ان مقياس الكفاءة النسبية للبيانات الاصلية هو الواحد الصحيح لكلا القيمتين. الطريقة البيانية (2005) تم عليها تعديل نظراً لعدم مراعاتها المشاهدات التي تظهر بصورة معزولة من اغلب البيانات و في نفس الوقت تقع قريبة من خط انحدار البيانات (المشاهدات ذات التأثير الجامح الإيجابي) مما يجعل النموذج يظهر بمؤشرات جيدة ظاهرياً مثل

المشاهدة 18 التي اعتبرها الباحث من المشاهدات الشاذة في الدراسة الحالية، عليه تستبعد الطريقة البيانية من المقارنة و تدخل في المقارنة بعد التعديل (بعد ضم المشاهدة 18 لمشاهداتها الشاذة و إعادة المعالجة)، و علي الرقم من تطبيق كل الطرق السابقة بما فيهم الطريقة الحالية علي بيانات نموذج واحد نلاحظ ان عدد المشاهدات المشخصة للطرق السابقة تراوح بين 1 الي 7 مشاهدات شاذة الامر الذي يشير الي عدم توفر معيار ثابت بالنسبة لهذه الطرق في تحديد عدد المشاهدات الشاذة، بينما نجد ان الطريقة الحالية و من خلال استخدام المقياس البياني حددت عدد 6 مشاهدات شاذة.

ايضاً من الجدول (4 - 1) نلاحظ ان افضل طريقة من بين الطرق السابقة قللت قيمة MSE بكفاءة نسبية قدرها 2.532 و هي طريقة الباحث البيانية المعدلة بينما نجد ان الطرق المقترحة الحالية قللت قيمة MSE بكفاءة نسبية قدرها 3.284 الامر الذي يبين تفوق كفاءة الطرق المقترحة الحالية من حيث تقليل قيمة MSE ، كما نلاحظ ان افضل طريقة من بين الطرق السابقة رفعت قيمة R^2 بكفاءة نسبية قدرها 1.446 و هي طريقة الباحث البيانية المعدلة بينما نجد ان الطرق المقترحة الحالية رفعت قيمة R^2 بكفاءة نسبية قدرها 1.734 الامر الذي يبين تفوق كفاءة الطرق المقترحة الحالية من حيث رفع قيمة R^2 . عليه نستنتج ان الطرق المقترحة متفوقة من حيث الكفاءة النسبية على الطرق السابقة.

4-2 التطبيق العملي علي نماذج الانحدار الخطي البسيط:

تم تطبيق خطوات طريقة التشخيص و المعالجة من خلال برنامج التحليل الاحصائي SPSS و المقياس البياني باستخدام ورق الرسم البياني (الشكل (3 - 1)) و بالإضافة لبرنامج الجداول الالكترونية Excel الذي استوعب كل الصيغ الرياضية لطريقتي التشخيص و المعالجة المقترحتين. اي انه لهذا الغرض تم تطبيق الطرق المقترحة على 35 نموذج انحدار خطي بسيط

بما فيها البيانات التي تم استخدامها في دراسات المقارنة السابقة، و لكي ما يتم التمكن من اجراء اختبار الفروض المطروحة بالنسبة لهذا البحث و بالوضع الذي يستوفي شروط الاختبار حسب المنهجية العلمية، فقد تم استخلاص مؤشرات هذه النماذج بالنسبة للثلاث أوضاع السالفة الذكر و كالآتي:-

1/ الوضع الأول:

تم توفيق النماذج بطريقة المربعات الصغرى علي البيانات الاصلية (Original Data) و الجدول (4 - 2) يوضح مؤشرات النماذج:-

الجدول (4 - 2): قيم المؤشرات ($R^2; F; t_{\beta_0}; t_{\beta_1}; MSE$) للبيانات الاصلية (original)

لنماذج الانحدار الخطي البسيط

رقم ملف البيانات Data File No.	رقم مسلسل No.	عدد قيم النموذج n	معامل التحديد R^2	قيمة F المحسوبة F	قيمة t المحسوبة $ t_{\beta_0} $	قيمة t المحسوبة $ t_{\beta_1} $	متوسط مربعات الخطأ MSE
book001 02 012	1	12	0.746	29.301	7.756	5.413	11.619
book002 02 022	2	22	0.778	70.227	2.594	8.38	10814.373
book003 02 005	3	5	0.877	21.472	0.463	4.634	9543.214
book004 02 050	4	50	0.651	89.567	2.601	9.464	236.532
book005 02 097	5	97	0.629	160.993	9.366	12.688	0.082
book006 02 007	6	7	0.799	19.906	1.679	4.462	242.971
book007 02 012	7	12	0.576	13.56	0.586	3.682	7.226
book008 02 051	8	51	0.836	250.259	10	15.82	0.392
book009 02 025	9	25	0.54	26.954	1.367	5.192	45.783
book010 02 056	10	56	0.716	135.934	12.888	11.659	126.241
book011 02 026	11	26	0.812	103.615	0.906	10.179	0.211
book012 02 053	12	53	0.537	59.206	0.543	7.695	18.727
book013 02 013	13	13	0.915	117.795	1.73	10.853	4.826
book014 02 020	14	20	0.462	22.608	19.366	4.755	0.000
book015 02 011	15	11	0.235	2.759	3.817	1.661	909.885

book016 02 033	16	33	0.415	22.013	12.153	4.692	4.879
book017 02 039	17	39	0.313	16.875	0.552	4.108	1202.677
book018 02 067	18	67	0.863	409.836	0.877	20.244	7500
book019 02 020	19	20	0.995	500	55.835	25	1.513
book020 02 051	20	51	0.448	39.822	4.777	4.448	6500
book021 02 010	21	10	0.644	14.503	0.439	3.808	45.367
book022 02 005	22	5	0.972	105.257	67.351	10.259	0.062
book023 02 047	23	47	0.499	44.842	27.876	6.696	0.009
book024 02 030	24	30	0.677	58.573	0.327	7.653	885.685
book025 02 027	25	27	0.857	150.038	2.248	12.249	0.011
book026 02 061	26	61	0.429	44.296	57.217	6.656	3500
book027 02 053	27	53	0.352	27.646	8.985	5.258	157.017
book028 02 069	28	69	0.203	17.018	3.509	4.125	18.5
book029 02 021	29	21	0.41	13.202	21.681	3.633	121.505
book030 02 015	30	15	0.825	61.097	1.236	7.816	120.641
book031 02 020	31	20	0.872	122.208	14.511	11.055	2471.864
book032 02 022	32	22	0.245	6.507	0.339	2.551	36.551
book033 02 024	33	24	0.409	15.209	1.526	3.9	71.321
book034 02 022	34	22	0.874	139.212	0.918	11.799	0.385
book035 02 025	35	25	0.462	19.782	0.124	4.448	380.085

المصدر: اعداد الباحث باستخدام البرنامج SPSS و برنامج الجداول الالكترونية Excel

بالنظر للجدول (4 - 2) أعلاه نلاحظ ان قيمة R^2 اقل من 0.75 لاغلب النماذج أي لعدد

22 نموذج انحدار خطي بسيط من اصل 35 نموذج انحدار خطي بسيط، الامر الذي يدل على

انخفاض قيمة R^2 بالنسبة لنماذج الانحدار الخطي البسيط .

/2 الوضع الثاني:

تم توفيق النماذج بطريقة المربعات الصغرى علي البيانات بعد تشخيص المشاهدات الشاذة و

حذفها (Deleted Outliers) و الجدول (4 - 3) يوضح مؤشرات النماذج:—

الجدول (3 - 4): قيم المؤشرات ($R^2; F; t_{\beta_0}; t_{\beta_1}; MSE$) للبيانات بعد التشخيص و
الحذف (detected)

رقم ملف البيانات Data File No.	رقم مسلسل No.	عدد مشاهدات النموذج n_j	عدد القيم الشاذة m	معامل التحديد R^2	قيمة F المحسوبة F	قيمة t المحسوبة $ t_{\beta_0} $	قيمة t المحسوبة $ t_{\beta_1} $	متوسط مربعات الخطأ MSE
book001 02 012	1	10	2	0.888	63.728	11.924	7.983	5.194
book002 02 022	2	20	2	0.899	160.079	2.447	12.652	4768.627
book003 02 005	3	4	1	0.923	24.075	0.724	4.907	8696.778
book004 02 050	4	40	10	0.805	157.118	3.017	12.535	70.17
book005 02 097	5	74	23	0.83	352.596	16.095	18.778	0.029
book006 02 007	6	6	1	0.916	43.418	1.612	6.589	59.25
book007 02 012	7	9	3	0.868	46.111	1.895	6.791	2.603
book008 02 051	8	43	8	0.904	385.131	8	19.625	0.138
book009 02 025	9	19	6	0.783	61.275	3.301	7.828	12.338
book010 02 056	10	40	16	0.827	181.095	15.793	13.457	42.094
book011 02 026	11	16	10	0.892	115.221	1.313	10.734	0.064
book012 02 053	12	38	15	0.774	123.642	1.974	11.119	5.259
book013 02 013	13	11	2	0.968	273.893	2.858	16.55	1.878
book014 02 020	14	13	7	0.733	30.131	34.577	5.489	0
book015 02 011	15	6	5	0.937	59.442	12.098	7.71	128.392
book016 02 033	16	20	13	0.832	89.408	20.877	9.456	1.393
book017 02 039	17	22	17	0.565	26.007	5.373	5.1	120.667
book018 02 067	18	61	6	0.923	709.346	1.693	26.634	2800
book019 02 020	19	16	4	0.997	1000	75.337	35	0.81
book020 02 051	20	35	16	0.75	95.749	13.625	9.785	1800
book021 02 010	21	7	3	0.899	44.605	1.919	6.679	16.162
book022 02 005	22	4	1	0.992	256.889	111.375	16.028	0.009
book023 02 047	23	32	15	0.729	80.652	46.882	8.981	0.002
book024 02 030	24	26	4	0.867	155.949	1.396	12.488	341.052
book025 02 027	25	21	6	0.919	214.834	5.868	14.657	0.004
book026 02 061	26	38	23	0.798	142.565	83.537	11.94	1200
book027 02 053	27	27	26	0.354	13.687	22.288	3.7	16.388
book028 02 069	28	35	34	0.358	18.385	23.455	4.288	0.784
book029 02 021	29	15	6	0.834	70.157	39.676	8.376	33.007
book030 02 015	30	14	1	0.971	404.701	1.625	20.117	10.644

book031 02 020	31	19	1	0.931	230.737	16.154	15.19	1233.225
book032 02 022	32	14	8	0.863	75.539	0.339	8.691	6.886
book033 02 024	33	17	7	0.767	49.414	1.526	7.03	16.681
book034 02 022	34	19	3	0.956	367.889	0.918	19.18	0.07
book035 02 025	35	16	9	0.734	38.524	0.124	6.207	137.612

المصدر : اعداد الباحث باستخدام البرنامج SPSS و برنامج الجداول الالكترونية Excel

بالنظر للجدول (4 - 3) أعلاه و بعد حذف المشاهدات الشاذة من نفس النماذج السابقة نلاحظ ان قيمة R^2 أصبحت اقل من 0.75 لعدد 6 نماذج انحدار خطي بسيط فقط من اصل 35 نموذج انحدار خطي بسيط مقارنة مع الجدول (4 - 2) نلاحظ تحسن كبير في قيمة R^2 ، الامر الذي يدل على تأثر نماذج الانحدار الخطي البسيط بالمشاهدات الشاذة.

3/ الوضع الثالث:

تم توفيق النماذج بطريقة المربعات الصغرى علي البيانات بعد تشخيص المشاهدات الشاذة و معالجتها (Detected Treated Outliers) و الجدول (4 - 4) يوضح مؤشرات النماذج:—

الجدول (4 - 4): قيم المؤشرات ($R^2; F; t_{\beta_0}; t_{\beta_1}; MSE$) للبيانات بعد التشخيص و المعالجة (modulated)

رقم ملف البيانات Data File No.	رقم مسلسل No.	عدد مشاهدات النموذج n	عدد القيم الشاذة m	قيمة التوزيع الاحتمالية W Pr	معامل التحديد R^2	قيمة F المحسوبة F	قيمة t المحسوبة $ t_{\beta_0} $	قيمة t المحسوبة $ t_{\beta_1} $	متوسط مربعات الخطأ MSE
book001 02 012	1	12	2	0.2	0.899	89.081	13.3	9.438	4.424
book002 02 022	2	22	2	0.1	0.922	235.019	2.743	15.33	4356.054
book003 02 005	3	5	1	0.25	0.961	74.618	1.079	8.638	5903.076
book004 02 050	4	50	10	0.25	0.827	229.27	3.318	15.142	66.829
book005 02 097	5	97	23	0.311	0.844	513.642	20.048	22.664	0.027
book006 02 007	6	7	1	0.167	0.915	53.928	1.902	7.344	53.526
book007 02 012	7	12	3	0.333	0.871	67.817	2.499	8.235	2.466
book008 02 051	8	51	8	0.186	0.93	655.701	15	25.607	0.127
book009 02 025	9	25	6	0.316	0.789	86.207	3.038	9.285	12.196

book010 02 056	10	56	16	0.4	0.839	280.937	19.591	16.761	45.696
book011 02 026	11	26	10	0.625	0.986	1743.31	0.784	11.332	0.119
book012 02 053	12	53	15	0.395	0.703	120.625	2.497	10.983	6.511
book013 02 013	13	13	2	0.182	0.973	390.58	3.154	19.763	1.659
book014 02 020	14	20	7	0.539	0.669	36.359	34.165	6.03	0
book015 02 011	15	11	5	0.833	0.624	14.949	6.296	3.866	543.642
book016 02 033	16	33	13	0.65	0.757	96.372	21.195	9.817	2.284
book017 02 039	17	39	17	0.773	0.319	17.35	5.266	4.165	798.61
book018 02 067	18	67	6	0.098	0.931	883.03	2.113	29.716	3500
book019 02 020	19	20	4	0.25	0.998	2000	109.23	40	0.666
book020 02 051	20	51	16	0.457	0.74	139.665	16.818	11.818	2500
book021 02 010	21	10	3	0.429	0.888	63.442	3.332	7.965	15.243
book022 02 005	22	5	1	0.25	0.982	161.706	88.479	12.716	0.015
book023 02 047	23	47	15	0.469	0.759	141.653	58.27	11.902	0.003
book024 02 030	24	30	4	0.154	0.885	214.662	1.729	14.651	301.892
book025 02 027	25	27	6	0.286	0.908	245.94	6.773	15.682	0.003
book026 02 061	26	61	23	0.605	0.717	149.589	89.019	12.231	2000
book027 02 053	27	53	26	1	0.352	27.646	8.985	5.258	157.017
book028 02 069	28	69	34	1	0.203	17.018	3.509	4.125	18.5
book029 02 021	29	21	6	0.4	0.711	46.803	35.456	6.841	36.995
book030 02 015	30	15	1	0.071	0.97	426.727	1.648	20.657	10.47
book031 02 020	31	20	1	0.053	0.934	254.371	16.733	15.949	1170.016
book032 02 022	32	22	8	0.571	0.74	56.812	4.879	7.537	20.405
book033 02 024	33	24	7	0.412	0.758	68.859	3.505	8.298	21.912
book034 02 022	34	22	3	0.159	0.954	415.457	2.426	20.383	0.066
book035 02 025	35	25	9	0.563	0.624	38.137	1.016	6.176	154.594

المصد : اعداد الباحث باستخدام البرنامج SPSS و برنامج الجداول الالكترونية Excel

بالنظر للجدول (4 - 4) أعلاه و بعد معالجة قيم المشاهدات الشاذة من نفس النماذج السابقة

نلاحظ ان قيمة R^2 أصبحت اقل من 0.75 لعدد 11 نموذج انحدار خطي بسيط من اصل 35

نموذج انحدار خطي بسيط و بالمقارنة مع الجدول (4 - 2) الذي اظهر ان قيمة R^2 اقل من

0.75 لعدد 22 نموذج انحدار خطي بسيط من اصل 35 نموذج انحدار خطي بسيط أي المقارنة

بين 11 التي تمثل وسط المدى من 1 الى 22 ، نلاحظ انه قد تم معالجة قيم المشاهدات الشاذة

بدرجة متوسطة مع مراعات نسبة وجودها الممثلة بنسبة الترجيح (W Pr)، الامر الذي يدل على وسطية المعالجة لنماذج الانحدار الخطي البسيط.

و بالنظر لبيانات الجداول الثلاث أعلاه (2 - 4) و (3 - 4) و (4 - 4) و التي تمثل قيم المؤشرات لبيانات 35 نموذج انحدار خطي بسيط تم استخلاصها لثلاث أوضاع من خلال توفيق النماذج بطريقة المربعات الصغرى و تطبيق طرق التشخيص و المعالجة باستخدام برنامج التحليل الاحصائي SPSS و برنامج الجداول الالكترونية Exel مما يشير الي سهولة تطبيق هذه الطرق، بالإضافة للمقياس البياني الذي تم تصميمه على مبدأ العلاقة العكسية بين مؤشر R^2 و مؤشر MSE و مبدأ النسبية لنفس المؤشرين لذا أتت نتائج تطبيق هذا المقياس دقيقة في تحديد عدد المشاهدات الشاذة بالنسبة لنموذج الانحدار الخطي البسيط. و بمقارنة قيم المؤشرات للجدول (4 - 3) و الجدول (4 - 4) نلاحظ ان قيم المؤشرات للجدول (3 - 4) جميعها تتحسن بالمقارنة مع مؤشرات البيانات الاصلية في الجدول (2 - 4)، و لكن تتحسن قيم المؤشرات بالنسبة للجدول (4 - 4) اكثر منه في الجدول (3 - 4) اذا قل عدد المشاهدات الشاذة عن ربع عدد المشاهدات الكلي، اما اذا زاد عدد المشاهدات الشاذة عن ربع المشاهدات فأن مؤشرات النماذج في الجدول (4 - 3) تكون احسن منه كما في الجدول (4 - 4) الى ان تصل لنفس مستوى مؤشرات البيانات الاصلية في الجدول (2 - 4) اذا اقترب عدد المشاهدات الشاذة من نصف العدد الكلي تابع النموذج رقم 27 و النموذج رقم 28 بالنسبة للثلاث جداول (2 - 4)، (3 - 4)، (4 - 4).

و من اجل اختبار متوسط الكفاءة النسبية تم قسمة مؤشرات النماذج للبيانات الموفقة بعد التشخيص و المعالجة علي مؤشرات النماذج للبيانات الموفقة الاصلية باستثناء مؤشر MSE الذي تم اخذ مقلوبه بعد القسمة بصفته المؤشر الوحيد الذي تقل قيمته بعد التشخيص او بعد التشخيص و

المعالجة و ذلك حتي نضمن تحسن مؤشرات الكفاءة النسبية بالزيادة (او بالاختلاف الإيجابي عن

الواحد الصحيح) و الجدول (4 - 5) يوضح قيم الكفاءة النسبية للمؤشرات:—

الجدول (4 - 5): قيم مؤشرات الكفاءة النسبية ($R^2; F; t_{\beta_0}; t_{\beta_1}; MSE$) للبيانات بعد التشخيص و المعالجة الى البيانات الاصلية

رقم ملف البيانات Data File No.	رقم مسلسل No.	عدد مشاهدات n	عدد القيم الشاذة m	قيمة التوزيع الاحتمالية W Pr	معامل التحديد R^2	قيمة F المحسوبة F	قيمة t المحسوبة t_{β_0}	قيمة t المحسوبة t_{β_1}	متوسط مربعات الخطأ MSE
book001 02 012	1	12	2	0.2	1.205	3.040	1.715	1.744	2.626
book002 02 022	2	22	2	0.1	1.185	3.347	1.057	1.829	2.48
book003 02 005	3	5	1	0.25	1.096	3.475	2.331	1.864	1.617
book004 02 050	4	50	10	0.25	1.270	2.56	1.276	1.6	3.539
book005 02 097	5	97	23	0.311	1.342	3.191	2.141	1.786	3.037
book006 02 007	6	7	1	0.167	1.145	2.709	1.133	1.646	4.539
book007 02 012	7	12	3	0.333	1.512	5.001	4.265	2.237	2.930
book008 02 051	8	51	8	0.186	1.112	2.620	1.5	1.619	3.087
book009 02 025	9	25	6	0.316	1.461	3.198	2.222	1.788	3.754
book010 02 056	10	56	16	0.4	1.171	2.067	1.520	1.438	2.763
book011 02 026	11	26	10	0.625	1.214	16.83	0.865	1.113	1.773
book012 02 053	12	53	15	0.395	1.309	2.037	4.599	1.427	2.876
book013 02 013	13	13	2	0.182	1.063	3.316	1.823	1.821	2.909
book014 02 020	14	20	7	0.539	1.447	1.608	1.764	1.268	1
book015 02 011	15	11	5	0.833	2.655	5.418	1.65	2.328	1.674
book016 02 033	16	33	13	0.65	1.824	4.378	1.744	2.092	2.136
book017 02 039	17	39	17	0.773	1.019	1.028	9.54	1.014	1.506
book018 02 067	18	67	6	0.098	1.079	2.155	2.409	1.468	2.143
book019 02 020	19	20	4	0.25	1.003	4	1.956	1.6	2.272
book020 02 051	20	51	16	0.457	1.652	3.507	3.521	2.657	2.6
book021 02 010	21	10	3	0.429	1.379	4.374	7.59	2.092	2.976
book022 02 005	22	5	1	0.25	1.010	1.536	1.314	1.24	4.133
book023 02 047	23	47	15	0.469	1.521	3.159	2.090	1.778	3
book024 02 030	24	30	4	0.154	1.307	3.665	5.288	1.914	2.934

book025 02 027	25	27	6	0.286	1.06	1.639	3.013	1.280	3.667
book026 02 061	26	61	23	0.605	1.671	3.377	1.556	1.838	1.75
book027 02 053	27	53	26	1	1	1	1	1	1
book028 02 069	28	69	34	1	1	1	1	1	1
book029 02 021	29	21	6	0.4	1.734	3.545	1.635	1.883	3.284
book030 02 015	30	15	1	0.071	1.176	6.984	1.333	2.643	11.523
book031 02 020	31	20	1	0.053	1.071	2.082	1.153	1.443	2.113
book032 02 022	32	22	8	0.571	3.021	8.731	14.39	2.955	1.791
book033 02 024	33	24	7	0.412	1.853	4.528	2.297	2.128	71.321
book034 02 022	34	22	3	0.159	1.092	2.984	2.643	1.728	0.385
book035 02 025	35	25	9	0.563	1.351	1.928	8.194	1.389	380.085

المصدر: اعداد الباحث باستخدام البرنامج SPSS و برنامج الجداول الالكترونية Excel

بالنظر للجدول (4 - 5) أعلاه نلاحظ ان جميع قيم مؤشرات الكفاءة النسبية

($R_2; F; t_{\beta_0}; t_{\beta_1}; MSE$) تختلف عن الواحد الصحيح باستثناء النماذج التي يتساوى فيها عدد

المشاهدات الشاذة مع نصف او اقل من النصف بمشاهدة واحدة لعدد المشاهدات الكلي.

و من اجل اختبار الفرض الأخير القائل بأنه (يقدر عدد المشاهدات الشاذة لنموذج الانحدار

الخطي البسيط بربع عدد المشاهدات الكلي للنموذج) تم تكوين الجدول (4 - 6) التالي:—

الجدول (4 - 6): قيم عدد المشاهدات الكلي n و عدد المشاهدات الشاذة m و ربع عدد

المشاهدات $n/4 - [v - 2]$ للنموذج

رقم ملف البيانات Data File No.	رقم مسلسل No.	عدد مشاهدات النموذج n	عدد متغيرات النموذج v	عدد القيم الشاذة m	ربع عدد مشاهدات النموذج $n/4 - [v - 2]$	قيمة التوزيع الاحتمالية W Pr
book001 02 012	1	12	2	2	3	0.2
book002 02 022	2	22	2	2	5.5	0.1
book003 02 005	3	5	2	1	1.25	0.25
book004 02 050	4	50	2	10	12.5	0.25
book005 02 097	5	97	2	23	24.25	0.311

book006 02 007	6	7	2	1	1.75	0.167
book007 02 012	7	12	2	3	3	0.333
book008 02 051	8	51	2	8	12.75	0.186
book009 02 025	9	25	2	6	25	0.316
book010 02 056	10	56	2	16	14	0.4
book011 02 026	11	26	2	10	6.5	0.625
book012 02 053	12	53	2	15	13.25	0.395
book013 02 013	13	13	2	2	3.25	0.182
book014 02 020	14	20	2	7	5	0.539
book015 02 011	15	11	2	5	2.75	0.833
book016 02 033	16	33	2	13	8.25	0.65
book017 02 039	17	39	2	17	9.75	0.773
book018 02 067	18	67	2	6	16.75	0.098
book019 02 020	19	20	2	4	5	0.25
book020 02 051	20	51	2	16	12.75	0.457
book021 02 010	21	10	2	3	2.5	0.429
book022 02 005	22	5	2	1	1.25	0.25
book023 02 047	23	47	2	15	11.75	0.469
book024 02 030	24	30	2	4	7.5	0.154
book025 02 027	25	27	2	6	6.75	0.286
book026 02 061	26	61	2	23	15.25	0.605
book027 02 053	27	53	2	26	13.25	1
book028 02 069	28	69	2	34	17.25	1
book029 02 021	29	21	2	6	5.25	0.4
book030 02 015	30	15	2	1	3.75	0.071
book031 02 020	31	20	2	1	5	0.053
book032 02 022	32	22	2	8	5.5	0.571
book033 02 024	33	24	2	7	6	0.412
book034 02 022	34	22	2	3	5.5	0.159
book035 02 025	35	25	2	9	6.25	0.563

المصد : اعداد الباحث باستخدام البرنامج SPSS و برنامج الجداول الالكترونية Excel

بالنظر للجدول (4 - 6) أعلاه نلاحظ انه تم تضمين عدد المتغيرات في الصيغة

$(n/4 - [v - 2])$ و ذلك لان الدراسة على نموذج الانحدار الخطي البسيط و المتعدد، لاحظ

عدد المتغيرات لنموذج الانحدار الخطي البسيط هو 2 (متغير تابع و متغير مستقل) أي ان ($v=2$) و عليه فان الصيغة ($n/4 - [v-2]$) تؤول الى ($n/4$) اي ربع المشاهدات و الذي سيتم مقارنة متوسطه مع متوسط عدد المشاهدات الشاذة m ، و ذلك لان ربع عدد المشاهدات الكلي للنموذج يمثل الوسط الحسابي لعدد المشاهدات الشاذة بالنسبة لنموذج، و بافتراض خلو النموذج من المشاهدات الشاذة فإن ($m=0$) او ان يكون الحد الأقصى للملاحظات الشاذة للنموذج هو نصف عدد المشاهدات الكلي ($m=n/2$)، و بحساب متوسط القيمتين نحصل على الوسط الحسابي لعدد المشاهدات الشاذة في النموذج ($n/4$) أي ان

$$.(m = (0 + n/2) / 2 = n/4)$$

3-4 اختبارات الفروض فيما يتعلق بنماذج الانحدار الخطي البسيط:

بالنسبة للثلاث فروض التالية يتم تطبيق اختبار t لمقارنة متوسطات العينات المرتبطة باستخدام برنامج التحليل الاحصائي SPSS و كالآتي:-

1/ توجد فروق ذات دلالة معنوية بين مؤشرات النماذج للبيانات الاصلية و لنفس البيانات بعد حذف المشاهدات الشاذة. أي اختبار فرض العدم $H_0: \mu_{i_{original}} - \mu_{i_{detected}} = 0$ على مستوى دلالة $\alpha = 0.025$ مقابل الفرض البديل $H_1: \mu_{i_{original}} - \mu_{i_{detected}} \neq 0$ بعد تطبيق الاختبار كانت النتائج كما في الجدول (4 - 7) و كالآتي:-

الجدول (4 - 7): الفروق بين متوسطات المؤشرات للبيانات الاصلية (original) و للبيانات بعد حذف المشاهدات الشاذة (detected)

Paired Samples Test						
	Paired Differences			t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean			

					Lower	Upper			
Pair 1	$R^2_original - R^2_detected$	-2.03	.163	.028	-.268	-.138	-7.357	34	.000
Pair 2	$F_original - F_f_detected$	-91.71	108.32	18.31	-134.65	-48.77	-5.009	34	.000
Pair 3	$t_{\beta_0_original} - t_{\beta_0_detected}$	-6.671	9.845	1.664	-10.57	-2.768	-4.008	34	.000
Pair 4	$t_{\beta_1_original} - t_{\beta_1_detected}$	-3.867	2.786	.471	-4.971	-2.763	-8.212	34	.000
Pair 5	$MSE_original - MSE_detected$	670.06	1483.3	250.73	82.083	1258.03	2.672	34	.011

المصدر: اعداد الباحث باستخدام البرنامج SPSS

و من النتائج في الجدول (4 - 7) و بما ان $Sig.(2-tailed)$ لكل المؤشرات المختيرة اقل من $\alpha = 0.025$ عليه فإنه يمكننا رفض فرض عدم و قبول الفرض البديل القائل بأنه توجد فروق ذات دلالة معنوية بين مؤشرات النماذج للبيانات الاصلية و لنفس البيانات بعد حذف المشاهدات الشاذة ، الامر الذي يثبت مدي فعالية الطريقة الشمولية و المقياس البياني في تشخيص المشاهدات الشاذة بالضبط، أي ان المشاهدات الشاذة كان لها تأثير كبير على نتائج تحليل نموذج الانحدار.

2/ توجد فروق ذات دلالة معنوية بين مؤشرات النماذج للبيانات الاصلية لنفس البيانات بعد معالجة المشاهدات الشاذة. أي اختبار فرض عدم $H_0 : \mu_{i_original} - \mu_{i_modulated} = 0$ على مستوى دلالة $\alpha = 0.025$ مقابل الفرض البديل $H_1 : \mu_{i_original} - \mu_{i_modulated} \neq 0$ بعد تطبيق الاختبار كانت النتائج كما في الجدول (4 - 8) و كالآتي:—

الجدول (4 - 8): الفروق بين متوسطات المؤشرات للبيانات الاصلية (original) و للبيانات بعد معالجة المشاهدات الشاذة (modulated)

Paired Samples Test							
	Paired Differences				t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	97.5% Confidence Interval of the Difference			
				Lower			

Pair 1	$R^2_original - R^2_modulated$	-0.172	.123	.021	-.221	-.123	-8.24	34	.000
Pair 2	$F_original - F_modulated$	-203.0	362.9	61.35	-346.9	-59.14	-3.31	34	.002
Pair 3	$t_{\beta_0_original} - t_{\beta_0_modulated}$	-7.190	11.35	1.919	-11.69	-2.691	-3.75	34	.001
Pair 4	$t_{\beta_1_original} - t_{\beta_1_modulated}$	-5.125	3.499	.592	-6.512	-3.738	-8.67	34	.000
Pair 5	$MSE_original - MSE_modulated$	665.0	1497	253.1	71.544	1258.5	2.63	34	.013

المصدر: اعداد الباحث باستخدام البرنامج SPSS

و من النتائج في الجدول (4 - 8) و بما ان $Sig.(2-tailed)$ لكل المؤشرات المختبرة اقل من $\alpha = 0.025$ فإنه يمكننا رفض فرض عدم و قبول الفرض البديل القائل بأنه توجد فروق ذات دلالة معنوية بين مؤشرات النماذج للبيانات الاصلية و لنفس البيانات بعد معالجة المشاهدات الشاذة ، الامر الذي يشير الي كفاءة و فاعلية الطرق المقترحة في ازاله اثر المشاهدات الشاذة علي نموذج الانحدار الخطي البسيط، علي الرقم من انه تم مراعاة عدد المشاهدات الشاذة أي انه اذا كان عددها قليل يتم معالجتها بالوضع الذي يزيل اثرها بدرجة كبيرة، و اذا كان عددها كبير تتم معالجتها بالوضع الذي يزيل اثرها بدرجة اقل الي ان تترك من دون إزالة اثرها اذا ساوى عددها نصف عدد المشاهدات الكلي او نقص بمفرده واحدة، ما يدل على عدم تحيز طريقة المعالجة لاعتبارها نسبة و جود المشاهدات الشاذة الي المشاهدات الكلية، الامر الذي يجعل طريقة المعالجة تتصف بالمنهجية العلمية.

3/ لا توجد فروق ذات دلالة معنوية بين مؤشرات النماذج للبيانات بعد حذف المشاهدات الشاذة و لنفس البيانات بعد معالجة المشاهدات الشاذة. اي اختبار فرض عدم

$$H_0: \mu_{i_{detected}} - \mu_{i_{modulated}} = 0$$

على مستوى دلالة $\alpha = 0.025$ مقابل الفرض البديل

$$H_1: \mu_{i_{detected}} - \mu_{i_{modulated}} \neq 0$$

بعد تطبيق الاختبار كانت النتائج كما في الجدول (4 - 9) و

كالآتي:—

الجدول (4 - 9): الفروق بين متوسطات المؤشرات للبيانات بعد حذف المشاهدات الشاذة (*detected*) و للبيانات بعد معالجة المشاهدات الشاذة (*modulated*)

Paired Samples Test									
		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	97.5% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	$R^2_detected - R^2_modulated$.032	.082	.014	-.001	.064	2.29	34	.028
Pair 2	$F_detected - F_modulated$	-111.3	316.8	53.54	-236.86	14.28	-2.08	34	.045
Pair 3	$t_{\beta_0_detected} - t_{\beta_0_modulated}$	-92.57	380.75	64.36	-243.5	58.35	-1.438	34	.159
Pair 4	$t_{\beta_1_detected} - t_{\beta_1_modulated}$	1.172	4.741	.801	-.708	3.051	1.462	34	.153
Pair 5	$MSE_detected - MSE_modulated$	-5.052	549.13	92.82	-222.72	212.6	-.054	34	.957

المصدر : اعداد الباحث باستخدام البرنامج SPSS

و من النتائج في الجدول (4 - 9) و بما ان $Sig.(2-tailed)$ لكل المؤشرات المختبرة اكبر من $\alpha = 0.025$ فإنه يمكننا قبول فرض العدم القائل بأنه لا توجد فروق ذات دلالة معنوية بين مؤشرات النماذج للبيانات بعد حذف المشاهدات الشاذة و لنفس البيانات بعد معالجة المشاهدات الشاذة ، الامر الذي يعني ان الطرق المقترحة تقلل من اثر المشاهدات الشاذة بدرجة كبيرة اذا كان عددها قليل و العكس صحيح اذا كان عددها كبير و تترك المشاهدات الشاذة من دون تقليل لاثرها اذا تساوى او نقص عدد المشاهدات الشاذة بمفرده واحدة من نصف عدد المشاهدات الكلي للنموذج الامر الذي يبرهن منهجية هذه الطرق و يجعلها توافق المنطق، لان المشاهدات الشاذة في هذه الحالة جديرة بهذا الأثر علي النموذج.

4/ متوسط الكفاءة النسبية لمؤشرات البيانات بعد معالجة المشاهدات الشاذة الى مؤشرات البيانات الاصلية اكبر من الواحد الصحيح. أي انه من خلال تطبيق اختبار t لمتوسط العينة الواحدة نختبر فرض العدم $H_0: \mu_i = 1$ على مستوى دلالة $\alpha = 0.025$ مقابل الفرض البديل

بعد تطبيق الاختبار كانت النتائج كما في الجدول (4 - 10) و الجدول (4 - 4) - $H_1: \mu_i > 1$ و كالاتي:-

الجدول (4 - 10): متوسطات الكفاءة النسبية للمؤشرات

One-Sample Statistics				
المؤشرات	N	Mean	Std. Deviation	Std. Error Mean
R^2	35	1.3717661	.44232677	.07476687
F	35	3.6003545	2.81089946	.47512873
t_{β_0}	35	2.9579113	2.88529274	.48770349
t_{β_1}	35	1.7327529	.46821258	.07914237
MSE	35	2.8851770	1.85103194	.31288150

المصدر: اعداد الباحث باستخدام البرنامج SPSS

جدول (4 - 11): اختلاف متوسطات الكفاءة النسبية للمؤشرات عن الواحد الصحيح

One-Sample Test						
	Test Value = 1					
	t	df	Sig. (2-tailed)	Mean Difference	97.5% Confidence Interval of the Difference	
					Lower	Upper
R^2	4.972	34	.000	.37176611	.1964336	.5470986
F	5.473	34	.000	2.60035454	1.4861510	3.7145581
t_{β_0}	4.015	34	.000	1.95791132	.8142193	3.1016034
t_{β_1}	9.259	34	.000	.73275292	.5471596	.9183462
MSE	6.025	34	.000	1.88517703	1.1514523	2.6189017

المصدر: اعداد الباحث باستخدام البرنامج SPSS

من النتائج في الجدول (4 - 10) نستنتج ان متوسطات الكفاءة النسبية لمؤشرات البيانات بعد معالجة المشاهدات الشاذة الى مؤشرات البيانات الاصلية هي ($R^2 = 1.372$ ، $F = 3.6$ ، $t_{\beta_0} = 2.958$ ، $t_{\beta_1} = 1.733$ ، $MSE = 2.885$)، الامر الذي يشير الى تحسن مؤشرات

نموذج الانحدار الخطي البسيط بنسبة 37.2% لمؤشر R^2 و 260% لمؤشر F و 195.8% لمؤشر t_{β_0} و 73.3% لمؤشر t_{β_1} و 188.5% لمؤشر MSE . و من الجدول (4 - 11) و بما ان $Sig.(2-tailed)$ لكل المؤشرات المختبرة اقل من $\alpha = 0.025$ فإنه يمكننا رفض فرض عدم و قبول الفرض البديل القائل بأن متوسط الكفاءة النسبية لمؤشرات البيانات بعد معالجة المشاهدات الشاذة الى مؤشرات البيانات الاصلية اكبر من الواحد الصحيح، مما يدل علي ثبات كفاءة الطرق المقترحة بالنسبة لتشخيص و معالجة مشكلة المشاهدات الشاذة في نماذج الانحدار الخطي البسيط.

5/ يقدر عدد المشاهدات الشاذة لنموذج الانحدار الخطي البسيط بربع عدد المشاهدات الكلي للنموذج - (عدد متغيرات النموذج - 2) أي بمعنى انه لا توجد فروق ذات دلالة معنوية بين عدد المشاهدات الشاذة و ربع عدد المشاهدات الكلي للنموذج - (عدد متغيرات النموذج - 2) أي لا توجد فروق ذات دلالة معنوية بين m و $n/4 - [v - 2]$ اي اختبار فرض عدم $H_0: \mu_{i_m} - \mu_{i_{n/4-[v-2]}} = 0$ على مستوى دلالة $\alpha = 0.025$ مقابل الفرض البديل $H_1: \mu_{i_m} - \mu_{i_{n/4-[v-2]}} \neq 0$ بعد تطبيق الاختبار كانت النتائج كما في الجدول (4 - 12) و كالآتي:-

الجدول (4 - 12): الفروق بين متوسطات القيم التي تمثل عدد المشاهدات الشاذة (m) و متوسطات القيم التي تمثل ربع عدد المشاهدات - (عدد المتغيرات - 2) أي تمثل $(n/4 - [v - 2])$ للنموذج

Paired Samples Test							
Paired Differences					t	df	Sig. (2-tailed)
Mean	Std. Deviation	Std. Error	97.5% Confidence Interval of the Difference				

				Mean	Lower	Upper			
Pair 1	$\mu_{i_m} - \mu_{i_{n/4-[v-2]}}$.42857	5.94848	1.00548	-1.92933	2.78647	.426	34	.673

المصدر : اعداد الباحث باستخدام البرنامج SPSS

و من النتائج في الجدول (4 - 12) و بما ان $Sig.(2-tailed)$ للمتوسطان المختيران اكبر من $\alpha = 0.025$ فإنه يمكننا قبول فرض العدم القائل بأنه لا توجد فروق ذات دلالة معنوية بين متوسط قيم اعداد المشاهدات الشاذة للنماذج و متوسط قيم $(n/4 - (v - 2))$ اي الارباع ناقص (عدد المتغيرات ناقص اثنان) بالنسبة لاعداد مشاهدات النماذج ، عليه يمكننا ان نقول ان عدد المشاهدات الشاذة يساوي $(n/4 - (v - 2))$ و بما ان عدد متغيرات نموذج الانحدار الخطي البسيط هو 2 أي ان $(v = 2)$ و بتعويضها في الصيغة $(n/4 - (v - 2))$ نحصل على ان عدد المشاهدات الشاذة للنموذج يساوي $n/4$ الامر الذي يعني ان عدد المشاهدات الشاذة لنموذج الانحدار الخطي البسيط يساوي ربع مشاهدات النموذج.

4-4 التطبيق العملي علي نماذج الانحدار الخطي المتعدد:

تم تطبيق خطوات طريقة التشخيص و المعالجة من خلال برنامج التحليل الاحصائي SPSS و المقياس البياني باستخدام ورق الرسم البياني (الشكل (3 - 1)) و بالإضافة لبرنامج الجداول الالكترونية Excel الذي استوعب كل الصيغ الرياضية لطريقتي التشخيص و المعالجة المقترحتين على نماذج الانحدار الخطي المتعدد. أي انه و لهذا الغرض تم تطبيق الطرق المقترحة على 31 نموذج انحدار خطي متعدد من بينها عدد 20 نموذج تتضمن ثلاث متغيرات مستقلة ليظل مؤشر t_{β_3} عند حجم العينة 20 و من بينها كذلك عدد 11 نموذج تتضمن متغيرين مستقلين اثنين ليكتمل حجم العينة لباقي المؤشرات الى 31 الحجم الذي يستوفي متطلبات الاختبار بالنسبة لكل المؤشرات $(R^2; F; t_{\beta_0}, t_{\beta_1}, t_{\beta_2}, t_{\beta_3}; MSE)$ عدا مؤشر t_{β_3} الذي يظل عند حجم العينة 20. تم استخلاص

مؤشرات هذه النماذج بالنسبة للثلاث أوضاع السالفة الذكر بهدف اختبار الفروض المطروحة بالنسبة لهذا البحث، فكانت المؤشرات المستخلصة من نتائج تطبيق الطرق المقترحة على نماذج الانحدار الخطي المتعدد كالتالي:-

1/ الوضع الأول:

تم توفيق النماذج بطريقة المربعات الصغرى علي البيانات الاصلية (Original Data) و

الجدول (4 - 13) يوضح مؤشرات النماذج:-

الجدول (4 - 13): قيم المؤشرات ($R^2; F; t_{\beta_0}, t_{\beta_1}, t_{\beta_2}, t_{\beta_3}; MSE$) للبيانات الاصلية (original)

رقم ملف البيانات Data File No.	رقم مسلسل No.	عدد مشاهدات النموذج n	عدد متغيرات النموذج v	معامل التحديد R^2	قيمة F المحسوبة F	قيمة t المحسوبة $ t_{\beta_0} $	قيمة t المحسوبة $ t_{\beta_1} $	قيمة t المحسوبة $ t_{\beta_2} $	قيمة t المحسوبة $ t_{\beta_3} $	متوسط مربعات الخطأ MSE
book038 04 032	1	32	4	.32300	4.45500	0.885	1.874	0.558	3.186	2.03600
book044 04 037	2	37	4	.74400	31.93900	0.47	2.837	6.036	1.745	298.04400
book045 03 025	3	25	3	.93800	166.69700	2.28	9.089	3.258	-	10.07200
book046 04 050	4	50	4	.59100	22.18600	4.518	0.083	6.239	4.932	1637.99100
book047 03 018	5	18	3	.66600	14.95500	1.448	3.258	4.859	-	.01100
book050 03 035	6	35	3	.87200	108.63300	1.947	7.542	4.678	-	158.30100
book051 03 020	7	20	3	.55700	10.69200	1.148	1.313	1.839	-	23500.00000
book052 04 030	8	30	4	.71900	22.17500	0.73	2.824	1.252	7.762	.00100
book054 03 027	9	27	3	.94100	189.88500	0.095	3.057	4.378	-	12500.00000
book056 04 012	10	12	4	.90900	26.59200	0.613	0.481	0.478	2.106	7650.58700
book059 03 018	11	18	3	.58400	10.52700	6.598	4.192	0.281	-	152.73900
book060 03 030	12	30	3	.79500	52.39200	28.556	0.285	4.921	-	53.15700
book061 03 030	13	30	3	.84000	70.99700	14.334	11.267	7.073	-	1331.47200
book062 04 028	14	28	4	.86300	50.24400	4.273	9.325	0.808	3.998	1.40000
book063 03 024	15	24	3	.07500	.85500	3.541	0.098	1.263	-	.10200
book064 04 026	16	26	4	.91000	74.08000	2.111	5.54	9.482	4.952	357.38700
book065 04 021	17	21	4	.91400	59.90200	3.356	5.307	3.52	0.973	10.51900
book067 03 031	18	31	3	.95100	269.81900	6.452	18.482	2.43	-	12.18900
book068 04 050	19	50	4	.67200	31.42400	1.885	6.729	1.962	1.102	6.62700

book088 03 028	20	28	3	.79300	47.77300	0.251	0.467	9.769	-	36800.00000
book091 04 012	21	12	4	.72000	6.84200	7.188	3.161	1.76	1.112	654.39700
book092 04 017	22	17	4	.67900	9.16500	0.798	2.186	3.039	0.084	16500.00000
book095 04 020	23	20	4	.86000	32.81100	1.343	3.226	1.796	4.239	19.95900
book096 04 040	24	40	4	.66200	23.54400	0.532	7.888	2.051	0.797	3626.00700
book098 04 054	25	54	4	.88300	125.70600	14.398	9.222	11.196	13.76	2059.73200
book099 04 015	26	15	4	.99700	380.00000	21.463	3.25	15	7.5	24.79600
book101 04 015	27	15	4	.95100	71.13300	3.999	3.059	2.146	0.275	20.49900
book102 04 042	28	42	4	.93600	184.47500	7.972	7.394	0.308	11.26	2275.96300
book103 04 015	29	15	4	.95100	71.13300	3.999	3.059	2.146	0.275	20.49900
book104 04 022	30	22	4	.70400	14.29200	0.06	1.782	1.533	2.93	26.17300
book105 04 033	31	33	4	.37200	5.71500	13.694	0.985	0.963	1.383	10791.30900

المصدر: اعداد الباحث باستخدام البرنامج SPSS و برنامج الجداول الالكترونية Excel

بالنظر للجدول (4 - 13) أعلاه نلاحظ ان قيمة R^2 اقل من 0.75 لعدد 14 نموذج انحدار

خطي متعدد من اصل 31 نموذج انحدار خطي متعدد، الامر الذي يدل على انخفاض قيمة R^2 بالنسبة لنماذج الانحدار الخطي المتعدد.

/2 الوضع الثاني:

تم توفيق النماذج علي البيانات بعد تشخيص المشاهدات الشاذة و حذفها (Deleted Outliers)

و الجدول (4 - 14) يوضح مؤشرات النماذج:—

الجدول (4 - 14): قيم المؤشرات ($R^2; F; t_{\beta_0}, t_{\beta_1}, t_{\beta_2}, t_{\beta_3}; MSE$) للبيانات بعد

التشخيص و الحذف (detected)

رقم ملف البيانات Data File No.	رقم مسلسل N0.	عدد مشاهدات النموذج n_j	عدد متغيرات النموذج v	عدد المشاهدات الشاذة m	معامل التحديد R^2	قيمة F المحسوبة F	قيمة t المحسوبة $ t_{\beta_0} $	قيمة t المحسوبة $ t_{\beta_1} $	قيمة t المحسوبة $ t_{\beta_2} $	قيمة t المحسوبة $ t_{\beta_3} $	متوسط مربعات الخطأ MSE
book038 04 032	1	28	4	4	.8580	48.141	6.444	6.236	0.783	11.342	.43500
book044 04 037	2	34	4	3	.9030	92.805	1.672	5.964	9.12	2.85	114.15400
book045 03 025	3	20	3	5	.9630	218.886	5.978	11.982	5.717	-	3.04200

book046 04 050	4	37	4	14	.7750	37.8880	4.34	3.675	5.386	5.11	503.30600
book047 03 018	5	16	3	2	.8630	41.1120	0.915	4.877	7.971	-	.00500
book050 03 035	6	32	3	5	.9370	214.638	0.941	8.632	3.995	-	51.27600
book051 03 020	7	14	3	6	.6560	10.4910	4.094	1.261	1.99	-	12300.00
book052 04 030	8	23	4	7	.9010	57.6900	1.23	3.503	1.696	12.798	.00000
book054 03 027	9	24	3	4	.9700	334.439	0.741	4.769	4.7	-	6570.00
book056 04 012	10	11	4	1	.9490	43.1860	0.674	0.415	0.387	2.304	4711.015
book059 03 018	11	14	3	4	.7970	21.5560	7.411	5.071	1.603	-	59.53200
book060 03 030	12	25	3	5	.9090	110.343	39.455	0.967	8.08	-	22.01500
book061 03 030	13	26	3	4	.9220	136.845	20.396	14.958	10.205	-	614.78300
book062 04 028	14	25	4	3	.9380	106.697	6.533	12.885	2.228	6.136	.69200
book063 03 024	15	12	3	12	.7260	13.2490	1.488	1.144	4.64	-	.02700
book064 04 026	16	20	4	6	.9400	83.6510	2.912	8.443	12.45	5.264	113.51400
book065 04 021	17	19	4	2	.9690	158.072	5.749	10.126	2.104	1.124	3.98600
book067 03 031	18	24	3	7	.9690	330.283	8.721	22.159	4.168	-	4.68000
book068 04 050	19	39	4	11	.8390	60.6620	0.104	8.621	0.223	0.564	2.34100
book088 03 028	20	21	3	7	.8850	72.9720	0.014	0.439	11.119	-	17350.00
book091 04 012	21	10	4	2	.8370	10.2580	11.119	0.28	5.174	0.181	169.83100
book092 04 017	22	13	4	4	.7830	10.8390	0.661	1.394	3.888	0.029	8980.00
book095 04 020	23	16	4	4	.9090	39.7450	0.075	2.786	1.916	5.471	7.00200
book096 04 040	24	31	4	9	.7870	51.8130	2.261	10.321	2.351	1.851	961.66900
book098 04 054	25	48	4	6	.9240	178.388	16.972	12.18	16.014	17.876	644.84200
book099 04 015	26	12	4	3	.9990	450.000	36.355	5.598	25	12	7.53200
book101 04 015	27	13	4	2	.9800	150.454	4.68	2.606	5.612	0.854	7.21000
book102 04 042	28	37	4	5	.9630	288.896	9.876	8.39	1.178	13.478	985.57400
book103 04 015	29	13	4	2	.9800	150.454	4.68	2.606	5.612	0.854	7.21000
book104 04 022	30	17	4	5	.8680	28.5150	0.683	5.182	1.338	3.491	8.36300
book105 04 033	31	19	4	14	.7880	18.6310	15.412	2.372	1.797	3.919	2602.173

المصدر : اعداد الباحث باستخدام البرنامج SPSS و برنامج الجداول الالكترونية Excel

بالنظر للجدول (4 - 14) أعلاه و بعد حذف المشاهدات الشاذة من نفس النماذج السابقة نلاحظ

ان قيمة R^2 أصبحت اقل من 0.75 لعدد 2 نماذج انحدار خطي متعدد فقط من اصل 31 نموذج

انحدار خطي متعدد مقارنة مع الجدول (4 - 13) نلاحظ تحسن كبير في قيمة R^2 ، الامر الذي

يدل على تأثر نماذج الانحدار الخطي المتعدد بالمشاهدات الشاذة.

3/ الوضع الثالث:

تم توفير النماذج علي البيانات بعد تشخيص المشاهدات الشاذة و معالجتها (Detected Treated

Outliers) و الجدول (4 - 15) يوضح مؤشرات النماذج:ـ

الجدول (4 - 15): قيم المؤشرات ($R^2; F; t_{\beta_0}, t_{\beta_1}, t_{\beta_2}, t_{\beta_3}; MSE$) للبيانات بعد التشخيص و المعالجة (modulated)

رقم ملف البيانات Data File No.	رقم مسلسل No.	عدد مشاهدات النموذج n	عدد متغيرات النموذج v	عدد المشاهدات الشاذة m	معامل التحديد R^2	قيمة F المحسوبة	قيمة t المحسوبة $ t_{\beta_0} $	قيمة t المحسوبة $ t_{\beta_1} $	قيمة t المحسوبة $ t_{\beta_2} $	قيمة t المحسوبة $ t_{\beta_3} $	متوسط مربعات الخطأ MSE
book038 04 032	1	32	4	4	.861	57.681	7.231	6.486	1.125	12.7	.4250
book044 04 037	2	37	4	3	.905	105.20	1.894	6.496	9.654	3.09	105.34
book045 03 025	3	25	3	5	.945	189.89	6.55	10.182	5.965	-	3.0030
book046 04 050	4	50	4	14	.783	55.323	4.209	4.335	5.734	4.90	531.23
book047 03 018	5	18	3	2	.873	51.435	0.951	5.356	9.001	-	.0040
book050 03 035	6	35	3	5	.957	358.76	2.808	14.691	7.285	-	24.248
book051 03 020	7	20	3	6	.638	14.979	3.772	1.311	3.407	-	15700.
book052 04 030	8	30	4	7	.886	67.459	1.799	4.516	2.669	13.5	.0010
book054 03 027	9	27	3	4	.964	324.01	0.29	6.062	4.083	-	8950.0
book056 04 012	10	12	4	1	.957	59.793	0.728	0.425	0.381	2.67	4135.6
book059 03 018	11	18	3	4	.752	22.768	8.008	5.142	1.23	-	76.245
book060 03 030	12	30	3	5	.928	173.09	44.63	0.761	10.29	-	18.807
book061 03 030	13	30	3	4	.933	188.88	23.87	17.46	11.35	-	542.58
book062 04 028	14	28	4	3	.953	164.02	8.313	16.003	3.311	7.91	.4870
book063 03 024	15	24	3	12	.075	.85500	3.541	0.098	1.263	-	.1020
book064 04 026	16	26	4	6	.921	85.511	2.05	7.808	11.57	5.85	141.06
book065 04 021	17	21	4	2	.936	82.201	6.388	6.405	2.621	0.34	2.2320
book067 03 031	18	31	3	7	.969	441.57	11.24	25.031	5.625	-	4.2090
book068 04 050	19	50	4	11	.852	87.974	0.135	9.888	0.533	0.31	2.1200
book088 03 028	20	28	3	7	.895	106.66	0.507	2.901	11.02	-	19200.
book091 04 012	21	12	4	2	.842	14.169	11.61	0.148	6.488	0.16	160.48
book092 04 017	22	17	4	4	.876	30.563	1.014	2.08	7.044	0.31	10500.
book095 04 020	23	20	4	4	.921	62.260	0.207	3.519	2.322	7.13	6.0660
book096 04 040	24	40	4	9	.818	54.018	1.87	12.361	2.136	1.12	801.75
book098 04 054	25	54	4	6	.924	203.12	17.48	11.969	17.31	20.5	499.06
book099 04 015	26	15	4	3	.999	350.00	41.54	7.218	25	13	6.9050

book101 04 015	27	15	4	2	.982	198.79	4.918	2.661	6.889	1.13	6.4250
book102 04 042	28	42	4	5	.967	366.70	11.01	9.961	1.286	15.2	893.01
book103 04 015	29	15	4	2	.982	198.79	4.918	2.661	6.889	1.13	6.4250
book104 04 022	30	22	4	5	.839	31.179	1.49	5.719	1.825	3.33	8.6610
book105 04 033	31	33	4	14	.503	9.7720	9.267	0.054	0.219	2.52	8371.7

المصد : اعداد الباحث باستخدام البرنامج SPSS و برنامج الجداول الالكترونية Excel

بالنظر للجدول (4 - 15) أعلاه و بعد معالجة قيم المشاهدات الشاذة من نفس النماذج السابقة نلاحظ ان قيمة R^2 أصبحت اقل من 0.75 لعدد 3 نموذج انحدار خطي بسيط من اصل 31 نموذج انحدار خطي بسيط و بالمقارنة مع الجدول (4 - 13) الذي اظهر ان قيمة R^2 اقل من 0.75 لعدد 14 نموذج انحدار خطي متعدد من اصل 31 نموذج انحدار خطي متعدد، نلاحظ انه قد تم معالجة تأثير المشاهدات الشاذة بدرجة كبيرة، الامر الذي يدل على قلة عدد المشاهدات الشاذة في نماذج الانحدار الخطي بالمقارنة مع نماذج الانحدار الخطي البسيط، مما يؤيد مسلمة ان زيادة عدد المتغيرات يقلل من ظهور المشاهدات الشاذة، و ذلك لان طريقة المعالجة تقلل اثر المشاهدات الشاذة بدرجة كبيرة اذا كان عددها قليل.

و بالنظر لبيانات الجدول (4 - 13) و الجدول (4 - 14) و الجدول (4 - 15) و التي تمثل قيم المؤشرات لبيانات 31 نموذج انحدار خطي متعدد تم استخلاصها لثلاث أوضاع من خلال توفيق النماذج بطريقة المربعات الصغرى و تطبيق طرق التشخيص و المعالجة باستخدام برنامج التحليل الاحصائي SPSS و برنامج الجداول الالكترونية Excel مما يشير الي سهولة تطبيق الطرق المقترحة، بالإضافة للمقياس البياني الذي أتت نتائج تطبيقه دقيقة ايضاً في تحديد عدد المشاهدات الشاذة بالنسبة لنموذج الانحدار الخطي المنعدد. و بمقارنة قيم المؤشرات للجدول (4 - 14) و الجدول (4 - 15) نلاحظ ان قيم المؤشرات للجدول (4 - 14) جميعها تتحسن

بالمقارنة مع مؤشرات البيانات الاصلية في الجدول (4 - 13)، و لكن تتحسن قيم المؤشرات بالنسبة للجدول (4 - 15) اكثر منه في الجدول (4 - 14) اذا قل عدد المشاهدات الشاذة عن ربع عدد المشاهدات الكلي، اما اذا زاد عدد المشاهدات الشاذة عن ربع المشاهدات فأن مؤشرات النماذج في الجدول (4 - 14) تكون احسن منه كما في الجدول (4 - 15) الى ان تصل او تقترب لنفس مستوى مؤشرات البيانات الاصلية في الجدول (4 - 13) اذا اقترب عدد المشاهدات الشاذة من نصف العدد الكلي للمشاهدات.

و من اجل اختبار الافتراض الرابع و الخاص بمتوسط الكفاءة النسبية تم قسمة مؤشرات النماذج للبيانات الموقفة بعد التشخيص و المعالجة علي مؤشرات النماذج للبيانات الموقفة الاصلية باستثناء مؤشر MSE الذي تم اخذ مقلوبه بعد القسمة بصفته المؤشر الوحيد الذي تقل قيمته بعد التشخيص او بعد التشخيص و المعالجة، و ذلك حتي نضمن تحسن مؤشرات الكفاءة النسبية بالزيادة (او بالختلاف الإيجابي عن الواحد الصحيح) و الجدول (4 - 16) يوضح قيم الكفاءة النسبية للمؤشرات:-

الجدول (4 - 16): قيم مؤشرات الكفاءة النسبية ($R^2; F; t_{\beta_0}, t_{\beta_1}, t_{\beta_2}, t_{\beta_3}; MSE$) للبيانات بعد

التشخيص و المعالجة الى البيانات الاصلية

رقم ملف البيانات Data File No.	رقم مسلسل No.	عدد مشاهدات النموذج n	عدد متغيرات النموذج v	عدد المشاهدات الشاذة m	معامل التحديد R^2	قيمة F المحسوبة F	قيمة t المحسوبة $ t_{\beta_0} $	قيمة t المحسوبة $ t_{\beta_1} $	قيمة t المحسوبة $ t_{\beta_2} $	قيمة t المحسوبة $ t_{\beta_3} $	متوسط مربعات الخطأ MSE
book038 04 032	1	32	4	4	2.666	12.95	8.171	3.461	2.016	4.00	4.791
book044 04 037	2	37	4	3	1.216	3.294	4.030	2.290	1.599	1.77	2.830
book045 03 025	3	25	3	5	1.008	1.140	2.873	1.120	1.831	-	3.354
book046 04 050	4	50	4	14	1.325	2.494	.9316	7.232	.9191	.994	3.083

book047 03 018	5	18	3	2	1.311	3.440	.6568	1.644	1.852	-	2.750
book050 03 035	6	35	3	5	1.098	3.303	1.442	1.948	1.557	-	6.528
book051 03 020	7	20	3	6	1.145	1.401	3.286	.9985	1.853	-	1.497
book052 04 030	8	30	4	7	1.232	3.042	2.464	1.599	2.132	1.73	1.000
book054 03 027	9	27	3	4	1.024	1.706	3.053	1.983	.9326	.000	1.397
book056 04 012	10	12	4	1	1.053	2.249	1.188	.8836	.7970	1.27	1.850
book059 03 018	11	18	3	4	1.288	2.163	1.214	1.227	4.377	-	2.003
book060 03 030	12	30	3	5	1.167	3.304	1.563	2.670	2.089	-	2.827
book061 03 030	13	30	3	4	1.111	2.660	1.665	1.550	1.604	-	2.454
book062 04 028	14	28	4	3	1.104	3.264	1.946	1.716	4.098	1.98	2.875
book063 03 024	15	24	3	12	1.000	1.000	1.000	1.000	1.000	-	1.000
book064 04 026	16	26	4	6	1.012	1.154	.9711	1.409	1.220	1.18	2.534
book065 04 021	17	21	4	2	1.024	1.372	1.904	1.207	.7446	.347	4.713
book067 03 031	18	31	3	7	1.019	1.637	1.742	1.354	2.315	-	2.896
book068 04 050	19	50	4	11	1.268	2.800	.0716	1.470	.2717	.282	3.126
book088 03 028	20	28	3	7	1.129	2.233	2.020	6.212	1.128	-	1.917
book091 04 012	21	12	4	2	1.169	2.071	1.615	.0468	3.686	.142	4.078
book092 04 017	22	17	4	4	1.290	3.335	1.271	.9515	2.318	3.68	1.571
book095 04 020	23	20	4	4	1.071	1.898	.1541	1.091	1.293	1.68	3.290
book096 04 040	24	40	4	9	1.236	2.294	3.515	1.567	1.041	1.41	4.523
book098 04 054	25	54	4	6	1.047	1.616	1.214	1.298	1.546	1.49	4.127
book099 04 015	26	15	4	3	1.002	.9211	1.936	2.221	1.923	1.73	3.591
book101 04 015	27	15	4	2	1.033	2.795	1.230	.8699	3.210	4.12	3.191
book102 04 042	28	42	4	5	1.033	1.988	1.381	1.347	4.175	1.35	2.549
book103 04 015	29	15	4	2	1.033	2.795	1.230	.8699	3.210	4.12	3.192
book104 04 022	30	22	4	5	1.192	2.182	9.232	3.210	1.191	1.14	3.022
book105 04 033	31	33	4	14	1.352	1.710	.6767	.0548	.2274	1.82	1.289

المصدر: اعداد الباحث باستخدام البرنامج SPSS و برنامج الجداول الالكترونية Excel

بالنظر للجدول (4 - 16) أعلاه نلاحظ ان جميع قيم مؤشرات الكفاءة النسبية

$(R^2; F; t_{\beta_0}, t_{\beta_1}, t_{\beta_2}, t_{\beta_3}; MSE)$ تختلف عن الواحد الصحيح باستثناء النماذج التي يتساوى فيها

عدد المشاهدات الشاذة مع نصف او اقل من النصف بمشاهدة واحدة لعدد المشاهدات الكلي.

و من اجل اختبار الفرض الأخير القائل بأنه (يقدر عدد المشاهدات الشاذة لنموذج الانحدار

بربع عدد المشاهدات الكلي للنموذج - (عدد المتغيرات - 2)) تم تكوين الجدول (4 - 17)

التالي:-

الجدول (4 - 17): قيم عدد المشاهدات الكلي n و عدد المشاهدات الشاذة m و ربع عدد

المشاهدات - (عدد المتغيرات - 2) أي تمثل $(n/4 - [v-2])$ للنموذج

رقم ملف البيانات Data File No.	رقم مسلسل No.	عدد مشاهدات النموذج n	عدد متغيرات النموذج v	عدد القيم الشاذة m	ربع عدد مشاهدات النموذج $n/4 - [v-2]$	قيمة التوزيع الاحتمالية W Pr
book001 02 012	1	32	4	4	6.00	0.081
book002 02 022	2	37	4	3	7.25	0.2
book003 02 005	3	25	3	5	5.25	0.28
book004 02 050	4	50	4	14	10.50	0.11
book005 02 097	5	18	3	2	3.50	0.143
book006 02 007	6	35	3	5	7.75	0.3
book007 02 012	7	20	3	6	4.00	0.23
book008 02 051	8	30	4	7	5.50	0.148
book009 02 025	9	27	3	4	5.75	0.083
book010 02 056	10	12	4	1	1.00	0.222
book011 02 026	11	18	3	4	3.50	0.167
book012 02 053	12	30	3	5	6.50	0.133
book013 02 013	13	30	3	4	6.50	0.107
book014 02 020	14	28	4	3	5.00	0.5
book015 02 011	15	24	3	12	5.00	0.231
book016 02 033	16	26	4	6	4.50	0.095
book017 02 039	17	21	4	2	3.25	0.226
book018 02 067	18	31	3	7	6.75	0.22
book019 02 020	19	50	4	11	10.50	0.25
book020 02 051	20	28	3	7	6.00	0.167
book021 02 010	21	12	4	2	1.00	0.235
book022 02 005	22	17	4	4	2.25	0.2
book023 02 047	23	20	4	4	3.00	0.225

book024 02 030	24	40	4	9	8.00	0.111
book025 02 027	25	54	4	6	6.50	0.2
book026 02 061	26	15	4	3	1.75	0.133
book027 02 053	27	15	4	2	1.75	0.119
book028 02 069	28	42	4	5	8.50	0.133
book029 02 021	29	15	4	2	1.75	0.227
book030 02 015	30	22	4	5	3.50	0.424
book031 02 020	31	33	4	14	6.25	0.081

المصدر : اعداد الباحث باستخدام البرنامج SPSS و برنامج الجداول الالكترونية Excel

4- 5 اختبارات الفروض فيما يتعلق بنماذج الانحدار الخطي المتعدد:

بالنسبة للثلاث فروض التالية يتم تطبيق اختبار t لمقارنة متوسطات العينات المرتبطة باستخدام

برنامج التحليل الاحصائي SPSS و كالاتي:-

1/ توجد فروق ذات دلالة معنوية بين مؤشرات النماذج للبيانات الاصلية و لنفس البيانات بعد

حذف المشاهدات الشاذة. أي اختبار فرض العدم $H_0: \mu_{i_{original}} - \mu_{i_{detected}} = 0$ على مستوى

دلالة $\alpha = 0.025$ مقابل الفرض البديل $H_1: \mu_{i_{original}} - \mu_{i_{detected}} \neq 0$ بعد تطبيق الاختبار كانت

النتائج كما في الجدول (4 - 18) و كالاتي:-

الجدول (4 - 18): الفروق بين متوسطات المؤشرات للبيانات الاصلية (original) و

لبيانات بعد حذف المشاهدات الشاذة (detected)

Paired Samples Test									
		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	97.5% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	$R^2_{original} - R^2_{detected}$	-.133	.149	.027	-.196	-.069	-4.946	30	.000
Pair 2	$F_{original} - F_{detected}$	-44.53	36.460	6.548	-59.99	-29.08	-6.801	30	.000
Pair 3	$t_{\beta_0_{original}} - t_{\beta_0_{detected}}$	-1.989	3.515	.631	-3.478	-.499	-3.150	30	.004

Pair 4	$t_{\beta_1_original} - t_{\beta_1_detected}$	-1.632	1.792	.322	-2.391	-.872	-5.070	30	.000
Pair 5	$t_{\beta_2_original} - t_{\beta_2_detected}$	-1.659	2.283	.410	-2.626	-.691	-4.045	30	.000
Pair 6	$t_{\beta_3_original} - t_{\beta_3_detected}$	-1.656	2.252	.504	-2.881	-.431	-3.290	19	.004
Pair 7	$MSE_original - MSE_detected$	2054.7	4294.2	771.3	234.87	3874.5	2.664	30	.012

المصدر : اعداد الباحث باستخدام البرنامج SPSS

و من النتائج في الجدول (4 - 18) و بما ان $Sig.(2-tailed)$ لجميع المؤشرات اقل من $\alpha = 0.025$ فإنه يمكننا رفض فرض العدم و قبول الفرض البديل القائل بأنه توجد فروق ذات دلالة معنوية بين مؤشرات النماذج للبيانات الاصلية و نفس البيانات بعد حذف المشاهدات الشاذة ، الامر الذي يثبت مدي فعالية الطريقة الشمولية المعممة و المقياس البياني في تشخيص المشاهدات الشاذة لنموذج الانحدار الخطي المتعدد.

/2 توجد فروق ذات دلالة معنوية بين مؤشرات النماذج للبيانات الاصلية و نفس البيانات بعد معالجة المشاهدات الشاذة.

أي اختبار فرض العدم $H_0 : \mu_{i_original} - \mu_{i_modulated} = 0$ على مستوى دلالة

$\alpha = 0.025$ مقابل الفرض البديل $H_1 : \mu_{i_original} - \mu_{i_modulated} \neq 0$ بعد تطبيق الاختبار

كانت النتائج كما في الجدول (4 - 19) و كالآتي:-

الجدول (4 - 19): الفروق بين متوسطات المؤشرات للبيانات الاصلية (original) و للبيانات بعد معالجة المشاهدات الشاذة (modulated)

Paired Samples Test							
	Paired Differences				t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	97.5% Confidence Interval of the Difference			
				Lower			

Pair 1	$R^2_original - R^2_modulated$	-.105	.104	.019	-.149	-.061	-5.640	30	.000
Pair 2	$F_original - F_modulated$	-63.43	64.24	11.54	-90.66	-36.21	-5.497	30	.000
Pair 3	$t_{\beta_0_original} - t_{\beta_0_modulated}$	-2.687	4.886	.878	-4.758	-.616	-3.062	30	.005
Pair 4	$t_{\beta_1_original} - t_{\beta_1_modulated}$	-2.273	2.492	.448	-3.328	-1.217	-5.078	30	.000
Pair 5	$t_{\beta_2_original} - t_{\beta_2_modulated}$	-2.209	2.543	.457	-3.287	-1.132	-4.837	30	.000
Pair 6	$t_{\beta_3_original} - t_{\beta_3_modulated}$	-2.121	2.868	.641	-3.681	-.560	-3.307	19	.004
Pair 7	$MSE_original - MSE_modulated$	1606.6	3521	632.4	114.41	3098.7	2.540	30	.016

المصدر: اعداد الباحث باستخدام البرنامج SPSS

و من النتائج في الجدول (4 - 19) و بما ان $Sig.(2-tailed)$ لكل المؤشرات المختبرة اقل من $\alpha = 0.025$ فإنه يمكننا رفض فرض العدم و قبول الفرض البديل القائل بأنه توجد فروق ذات دلالة معنوية بين مؤشرات النماذج للبيانات الاصلية و لنفس البيانات بعد معالجة المشاهدات الشاذة ، الامر الذي يشير الي كفاءة و فاعلية الطرق المقترحة في ازاله اثر المشاهدات الشاذة علي نموذج الانحدار الخطي البسيط، علي الرقم من انه تم مراعاة عدد المشاهدات الشاذة أي انه اذا كان عددها قليل يتم معالجتها بالوضع الذي يزيل اثرها بدرجة كبيرة، و اذا كان عددها كبير تتم معالجتها بالوضع الذي يزيل اثرها بدرجة اقل الي ان يترك النموذج على حاله من دون إزالة لاثر المشاهدات الشاذة اذا ساوى عددها نصف عدد المشاهدات الكلي او نقص بمفرده واحدة، ما يدل على عدم تحيز طريقة المعالجة لاعتبارها نسبة و جود المشاهدات الشاذة الي المشاهدات الكلية، أي انه اذا كانت نسبة المشاهدات الشاذة صغيرة لا تستحق التأثير في النموذج بدرجة كبيرة و اذا كانت نسبتها كبيرة فيجب تخفيف اثرها على النموذج بدرجة قليلة الامر الذي يجعل طريقة المعالجة تتصف بالموضوعية و عدم التحيز.

3/ لا توجد فروق ذات دلالة معنوية بين مؤشرات النماذج للبيانات بعد حذف المشاهدات الشاذة و

لنفس البيانات بعد معالجة المشاهدات الشاذة. اي اختبار فرض العدم

$$H_0: \mu_{i_{\text{detected}}} - \mu_{i_{\text{modulated}}} = 0$$

على مستوى دلالة $\alpha = 0.025$ مقابل الفرض البديل

$$H_1: \mu_{i_{\text{detected}}} - \mu_{i_{\text{modulated}}} \neq 0$$

بعد تطبيق الاختبار كانت النتائج كما في الجدول (4 - 20) و

كالآتي:-

الجدول (4 - 20): الفروق بين متوسطات المؤشرات للبيانات بعد حذف المشاهدات الشاذة
(detected) و للبيانات بعد معالجة المشاهدات الشاذة (modulated)

Paired Samples Test									
		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	97.5% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	$R^2_{\text{detected}} - R^2_{\text{modulated}}$.028	.129	.023	-.027	.082	1.186	30	.245
Pair 2	$F_{\text{detected}} - F_{\text{modulated}}$	-18.9	46.242	8.305	-38.50	.699	-2.275	30	.030
Pair 3	$t_{\beta_0_{\text{detected}}} - t_{\beta_0_{\text{modulated}}}$	-.698	1.930	.347	-1.516	.120	-2.014	30	.053
Pair 4	$t_{\beta_1_{\text{detected}}} - t_{\beta_1_{\text{modulated}}}$	-.641	1.772	.318	-1.392	.110	-2.013	30	.053
Pair 5	$t_{\beta_2_{\text{detected}}} - t_{\beta_2_{\text{modulated}}}$	-.550	1.267	.228	-1.087	-.014	-2.420	30	.022
Pair 6	$t_{\beta_3_{\text{detected}}} - t_{\beta_3_{\text{modulated}}}$	-.465	.993	.222	-1.005	.076	-2.091	19	.050
Pair 7	$MSE_{\text{detected}} - MSE_{\text{modulated}}$	-443.3	1291.2	231.9	-990.5	103.9	-1.912	30	.066

المصدر: اعداد الباحث باستخدام البرنامج SPSS

و من النتائج في الجدول (4 - 20) و باستثناء مؤشر t_{β_2} و بما ان $\text{Sig.}(2 - \text{tailed})$

لجميع المؤشرات المختبرة اكبر من $\alpha = 0.025$ فإنه يمكننا قبول فرض العدم القائل بأنه لا

توجد فروق ذات دلالة معنوية بين مؤشرات النماذج للبيانات بعد حذف المشاهدات الشاذة و لنفس البيانات بعد معالجة المشاهدات الشاذة ، الامر الذي يعني ان الطرق المقترحة تقلل من اثر المشاهدات الشاذة بدرجة كبيرة اذا كان عددها قليل و العكس صحيح اذا كان عددها كبير و تترك المشاهدات الشاذة من دون تقليل لاثرها اذا تساوى او نقص عدد المشاهدات الشاذة بمفردة واحدة من نصف عدد المشاهدات الكلي للنموذج الامر الذي يبرهن منهجية هذه الطرق و يجعلها توافق المنطق.

4/ متوسط الكفاءة النسبية لمؤشرات البيانات بعد معالجة المشاهدات الشاذة الى مؤشرات البيانات الاصلية اكبر من الواحد الصحيح. أي اختبار فرض العدم $H_0: \mu_i = 1$ على مستوى دلالة $\alpha = 0.025$ مقابل الفرض البديل $H_1: \mu_i > 1$ بعد تطبيق الاختبار كانت النتائج كما في الجدول (4 - 21) و الجدول (4 - 22) و كالاتي:—

الجدول (4 - 21): متوسطات الكفاءة النسبية للمؤشرات

One-Sample Statistics				
المؤشرات	N	Mean	Std. Deviation	Std. Error Mean
R^2	31	1.1824500	.29677656	.05330264
F	31	2.5872061	2.06921081	.37164123
t_{β_0}	31	2.1175271	1.99025482	.35746032
t_{β_1}	31	1.8225867	1.50972705	.27115498
t_{β_2}	31	1.8760467	1.11064812	.19947829
t_{β_3}	20	1.8114501	1.22575697	.27408759
MSE	31	2.8981814	1.23598539	.22198953

المصدر: اعداد الباحث باستخدام البرنامج SPSS

الجدول (4 - 22): اختلاف متوسطات الكفاءة النسبية للمؤشرات عن الواحد الصحيح

One-Sample Test	
	Test Value = 1

	t	df	Sig. (2-tailed)	Mean Difference	97.5% Confidence Interval of the Difference	
					Lower	Upper
R^2	3.423	30	.002	.18245002	.0566791	.3082209
F	4.271	30	.000	1.58720614	.7102954	2.4641168
t_{β_0}	3.126	30	.004	1.11752707	.2740771	1.9609770
t_{β_1}	3.034	30	.005	.82258672	.1827796	1.4623938
t_{β_2}	4.392	30	.000	.87604675	.4053653	1.3467282
t_{β_3}	2.961	19	.008	.81145012	.1444744	1.4784259
MSE	8.551	30	.000	1.89818144	1.3743833	2.4219796

المصدر: اعداد الباحث باستخدام البرنامج SPSS

من النتائج في الجدول (4 - 21) نستنتج ان متوسطات الكفاءة النسبية لمؤشرات البيانات بعد معالجة المشاهدات الشاذة الى مؤشرات البيانات الاصلية و هي $(R^2; F; t_{\beta_0}, t_{\beta_1}, t_{\beta_2}, t_{\beta_3}; MSE)$ الامر الذي يشير الى تحسن مؤشرات نموذج الانحدار الخطي المتعدد بنسبة 18.3% لمؤشر R^2 و 158.7% لمؤشر F و 111.8% لمؤشر t_{β_0} و 82.3% لمؤشر t_{β_1} و 87.6% لمؤشر t_{β_2} و 81.2% لمؤشر t_{β_3} و 189.8% لمؤشر MSE . و من الجدول (4 - 22) و بما ان $Sig.(2 - tailed)$ لكل المؤشرات المختبرة اقل من $\alpha = 0.025$ فإنه يمكننا رفض فرض العدم و قبول الفرض البديل القائل بأن متوسط الكفاءة النسبية لمؤشرات البيانات بعد معالجة المشاهدات الشاذة الى مؤشرات البيانات الاصلية اكبر من الواحد الصحيح، مما يدل علي كفاءة الطرق المقترحة بالنسبة لتشخيص و معالجة مشكلة المشاهدات الشاذة في نماذج الانحدار الخطي المتعدد.

5/ يقدر عدد المشاهدات الشاذة لنموذج الانحدار بربع عدد المشاهدات الكلي للنموذج - (عدد متغيرات النموذج - 2) أي بمعنى انه لا توجد فروق ذات دلالة معنوية بين عدد المشاهدات الشاذة m و ربع عدد المشاهدات الكلي للنموذج - (عدد متغيرات النموذج - 2)

اي $[v-2] - n/4$ للنموذج. اي اختبار فرض العدم $H_0: \mu_{i_m} - \mu_{i_{n/4-[v-2]}} = 0$ على

مستوى دلالة $\alpha = 0.025$ مقابل الفرض البديل $H_1: \mu_{i_m} - \mu_{i_{n/4-[v-2]}} \neq 0$ بعد تطبيق

الاختبار كانت النتائج كما في الجدول (4 - 23) و كالاتي:-

الجدول (4 - 23): الفروق بين متوسطات القيم التي تمثل عدد المشاهدات الشاذة (m) و

متوسطات القيم التي تمثل ربع عدد المشاهدات - (عدد المتغيرات - 2) أي تمثل

($n/4 - [v-2]$) للنموذج

Paired Samples Test									
		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	97.5% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	$\mu_{i_m} - \mu_{i_{n/4-[v-2]}}$.31452	2.58116	0.46359	-0.77935	1.40838	.678	30	.503

المصدر: اعداد الباحث باستخدام البرنامج SPSS

و من النتائج في الجدول (4 - 23) و بما ان $Sig.(2-tailed)$ للمتوسطات المختبرة

اكبر من $\alpha = 0.025$ فإنه يمكننا قبول فرض العدم القائل بأنه لا توجد فروق ذات دلالة

معنوية بين متوسط قيم اعداد المشاهدات الشاذة للنماذج و متوسط قيم ($n/4 - [v-2]$) اي

الارباع ناقص (عدد المتغيرات ناقص اثنان) بالنسبة لاعداد مشاهدات النماذج ، الامر الذي يعني

ان عدد المشاهدات الشاذة لنموذج الانحدار الخطي المتعدد يساوي ربع مشاهدات النموذج ناقص

(عدد المتغيرات ناقص اثنان)، بعبارة أخرى ان عدد المشاهدات الشاذة لنموذج الانحدار الخطي

المتعدد يكون اقل من ربع المشاهدات الكلي للنموذج بحسب عدد المتغيرات في النموذج.

لفصل الخامس

النتائج و التوصيات

1-5 النتائج:

2-5 التوصيات:

5-1 النتائج:

في ضوء ما تم عرضه في هذه البحث نخلص الي:ـ

1/ نلاحظ ان قيمة R^2 للبيانات العادية اقل من 0.75 لاغلب النماذج أي لعدد 22 نموذج انحدار خطي بسيط من اصل 35 نموذج انحدار خطي بسيط.

2/ بعد حذف المشاهدات الشاذة من نفس النماذج نلاحظ ان قيمة R^2 تحسنت بدرجة كبيرة و أصبحت اقل من 0.75 لعدد 6 نماذج انحدار خطي بسيط فقط من اصل 35 نموذج انحدار خطي بسيط .

3/ بعد معالجة قيم المشاهدات الشاذة من نفس نماذج النحدار الخطي البسيط نلاحظ ان قيمة R^2 أصبحت اقل من 0.75 لعدد 11 نموذج انحدار خطي بسيط من اصل 35 نموذج انحدار خطي بسيط.

4/ بالنسبة لجميع نماذج الانحدار الخطي البسيط التي تمت معالجتها نلاحظ ان جميع قيم مؤشرات الكفاءة النسبية تختلف عن الواحد الصحيح باليجاب، باستثناء النماذج التي يتساوى فيها عدد المشاهدات الشاذة مع نصف او اقل من النصف بمشاهدة واحدة لعدد المشاهدات الكلي.

5/ نلاحظ ان قيمة R^2 للبيانات العادية اقل من 0.75 لعدد 14 نموذج انحدار خطي متعدد من اصل 31 نموذج انحدار خطي متعدد.

6/ بعد حذف المشاهدات الشاذة من نفس النماذج نلاحظ ان قيمة R^2 أصبحت اقل من 0.75 لعدد 2 نموذج انحدار خطي متعدد فقط من اصل 31 نموذج انحدار خطي متعدد.

7/ بعد معالجة قيم المشاهدات الشاذة من نفس النماذج نلاحظ ان قيمة R^2 أصبحت اقل من 0.75 لعدد 3 نموذج انحدار خطي متعدد من اصل 31 نموذج انحدار خطي متعدد.

8/ من النتيجة (7) نلاحظ انه قد تم معالجة تأثير المشاهدات الشاذة بدرجة كبيرة، الامر الذي يدل على قلة عدد المشاهدات الشاذة في نماذج الانحدار الخطي المتعدد بالمقارنة مع نماذج الانحدار الخطي البسيط، مما يؤيد مسلمة ان زيادة عدد المتغيرات يقلل من ظهور المشاهدات الشاذة، و ذلك لان طريقة المعالجة تقلل اثر المشاهدات الشاذة بدرجة كبيرة اذا كان عددها قليل.

9/ بالنسبة لنماذج الانحدار الخطي المتعدد التي تمت معالجتها نلاحظ ان جميع قيم مؤشرات الكفاءة النسبية تختلف عن الواحد الصحيح باستثناء النماذج التي يتساوى فيها عدد المشاهدات الشاذة مع نصف او اقل من النصف بمشاهدة واحدة لعدد المشاهدات الكلي.

10/ سهولة إمكانية تطبيق الطرق المقترحة من خلال برنامج الجداول الالكترونية (Excel) و برنامج التحليل الاحصائي (SPSS).

11/ اثبتت الطريقة المقترحة للتشخيص و المقياس البياني كفاءة عالية في تشخيص المشاهدات الشاذة و تحديد عددها.

12/ اثبتت الطريقة المقترحة للمعالجة كفاءة عالية في معالجة المشاهدات الشاذة.

13/ تعاملت الطرق المقترحة للمعالجة مع المشاهدات الشاذة بمنطق التأثير حسب نسبة وجودها في النموذج.

14/ أظهرت الطرق المقترحة للتشخيص و المعالجة تفوقاً كبيراً على طرق التشخيص و المعالجة السابقة.

15/ أظهر تطبيق الطرق المقترحة تحسناً كبيراً في مؤشرات نماذج الانحدار بشقيه البسيط و المتعدد، الامر الذي يدل على ان نماذج الانحدار الخطي كانت متاثرة بالمشاهدات الشاذة بدرجة كبيرة الى الحد الذي يجعل النموذج غير معنوي او بعض متغيرات النموذج تبدو غير معنوية.

16/ اتضح ان الطرق المقترحة بإزالتها لاثر المشاهدات الشاذة قد مهدت الطريق امام طريقة المربعات الصغري العادية - التي كانت غير موثوق فيها في ظل وجود المشاهدات الشاذة - في عملية توفيق نماذج الانحدار الخطي بشقيه البسيط و المتعدد.

17/ اتضح ان عدد المشاهدات الشاذة المتوقع لنموذج الانحدار الخطي يقدر بربع عدد المشاهدات الكلي للنموذج - (عدد متغيرات النموذج - 2).

18/ اتفقت الطرق المقترحة مع الادعاء بأن زيادة عدد المتغيرات المستقلة يقلل من عدد المشاهدات الشاذة للنموذج، و لم تتفق مع الادعاء بأن زيادة عدد المشاهدات يقلل من عدد المشاهدات الشاذة.

5- 2 التوصيات:

على ضوء ما تم التوصل اليه من نتائج في هذا البحث نوصي بالآتي:-

- 1/ عمل اجراء فحص للنموذج للتأكد من خلوه من المشاهدات الشاذة.
- 2/ تطبيق الطرق المقترحة في عمليات تشخيص و معالجة المشاهدات الشاذة في نموذج الانحدار الخطي و نسبة لما أحرزته تفوق على الطرق السابقة من حيث التشخيص و المعالجة.
- 3/ عمل دراسة مقارنة للنتائج بتطبيق الطرق المقترحة للتشخيص و المعالجة لنفس النماذج السابقة باعتبار النتيجة (8) لهذه الدراسة.
- 4/ البحث في إمكانية تضمين هذه الطرق المقترحة في النظام الاحصائي SPSS في جزئية أسلوب تحليل الانحدار الخطي لتصبح بمثابة تحديث للبرنامج الاحصائي SPSS او إمكانية تضمين طرق الكشف و المعالجة المقترحة في أي برامج اخرى.
- 5/ ضرورة عمل مزيد من الدراسة في هذا المجال.

المراجع العربية:

- 1 - الراوى، خاشع محمود، (1987)، المدخل إلى تحليل الانحدار، وزارة التعليم العالي و البحث العلمي، ط 1، جامعة الموصل، العراق.
- 2 - الجبوري، منى حسين، (1998) "دراسة تحليل للقيم الشاذة و القيم المفقودة لتصميم المربع اللاتيني و تصميم تام في حالة تكرار مشاهدات العينة"، رسالة ماجستير مقدمة الي مجلس كلية الإدارة و الاقتصاد، الجامعة المستنصرية، بغداد، العراق.
- 3 - جايترجي، سامبريت وبرائيس، بيرترام (1989)، "تحليل الانحدار بالأمثلة" ترجمة محمد مناجد الدليمي، جامعة بغداد، العراق، ص: 35.
- 4 - الدباغ، ظافر عاصم مصطفى (1999)، تحليل تباين حصين للنماذج الخطية، أطروحة دكتوراه فلسفة في الإحصاء، كلية الإدارة والاقتصاد، جامعة بغداد، 119 ص.
- 5 - إسماعيل، محمد عبد الرحمن، (2001) "تحليل الانحدار الخطي"، معهد الإدارة العامة، المملكة العربية السعودية، مركز البحوث، ص: 140-142.
- 6 - الجبوري، شلال حبيب، و آخرون (2002)، "اكتشاف و تقدير المشاهدات الشاذة باستخدام معادلة القطع الناقص Ellipse في حالة الانحدار الخطي البسيط"، المجلة العراقية للعلوم الإحصائية، العدد 4، المجلد 2، ص: 171-186.
- 7 - حمد، حامد حسين علي، (2005) "طرق بديلة لتقليل اثر المشاهدات الشاذة علي نموذج الانحدار الخطي البسيط" رسالة ماجستير، جامعة السودان للعلوم و التكنولوجيا، الخرطوم، السودان.
- 8 - دبدوب، مروان عبد العزيز، يونس، فرح عبد الغني، (2006)، " تأثير القيم الشاذة نتائج تحليل الانحدار مع التطبيق على المواليد الخدج " مجلة علوم الرافدين، المجلد 17 العدد 1، ص: 62 - 81 .

- 9 - النجار، و محمد، (2007)، "التقديرات الحصينة للانحدار الضبابي"، بحث مقدم في مؤتمر، كلية الإدارة و الاقتصاد، جامعة بغداد، 22 ص.
- 10 - احمد، محمد طه، (2007)، "المشاهدات غير المألوفة في نموذج الانحدار"، مجلة تكريت للعلوم الإدارية و الاقتصادية، العدد (55) ص ص 153 - 166.
- 11 - قاسم محمد نذير، إسماعيل، يونس حازم، (2008)، "الكشف عن القيم الشاذة بأسلوب بيز باستخدام معاينة جيس"، المجلة العراقية للعلوم الإحصائية، العدد (14) ص، ص (68 - 88).
- 12 - صافي، سمير خالد، (2008)، "دورة في البرنامج الاحصائي SPSS"، الجامعة الإسلامية، غزة، 57 ص.
- 13 - داوود، ألاء عبد الستار، (2009)، "تأثير القيم الشاذة و نقطة الأصل على نتائج تحليل الانحدار"، مجلة تكريت للعلوم الصرفة، المجلد 15، العدد 1.
- 14 - يحي، مزاحم محمد، (2009)، "المشاهدات غير العادية في الانحدار الخطي المتعدد و بعض طرائق تشخيصها مع التطبيق"، المجلة العراقية للعلوم الإحصائية، العدد (15)، ص، ص: 161 - 192 .
- 15 - شاكر، صالح مؤيد، (2009)، "تحسين أسلوب M الحصين في تقدير معاملات نموذج الانحدار المتعدد"، المجلة العراقية للعلوم الإحصائية، العدد (16)، ص:ص 219 - 242.
- 16 - حسن، صباح حسيب، (2009) "استخدام العايبير الحصينة في تحديد المتغيرات المؤثرة في نموذج الانحدار الخطي لقياس السكر في الدم"، المجلة العراقية للعلوم الإحصائية، العدد (16)، ص:ص 131 - 213.

- 17 - محمد، احمد صدقي، (2009)، "دراسة مقارنة لبعض الطرق المتينة للكشف عن المشاهدات الشاذة في نماذج الانحدار الخطي" كلية التجارة و الإحصاء التطبيقي، جامعة الإسكندرية، رسالة ماجستير منشورة في موقع مركز النظم للدراسات و خدمات البحث.
- 18 - جواد، نزار مصطفى، كمال، غفران إسماعيل، (2009)، " مقارنة طرق تقدير معالم نموذج الانحدار في حالة ظهور مشكلة التعدد الخطي و القيم الشاذة "، مجلة العلوم الإدارية و الاقتصادية، العدد (55)، ص:ص 153 - 166، العراق بغداد .
- 19 - حسن، تارا احمد، رضا، مهدي صابر، (2011)، " استخدام الانحدار الحصين لايجاد انسب نموذج لتمثيل بيانات الانواء الجوية في مدينة أربيل خلال الفترة 1998 - 2010 "، مجلة العلوم الاقتصادية و الإدارية، العدد (89)، ص:ص 195 - 213. الجامعة المستنصرية.
- 20 - الشميري، خالد سعد سلطان، و آخرون (2014) "اكتشاف القيم الشاذة و تقديرها في الانحدار الخطي بالتطبيق على بيانات النمو وفقاً لاعمار الأطفال في العام 1987" مجلة العلوم الطبيعية و الطبية، جامعة السودان للعلوم و التكنولوجيا، العدد 15، المجلد 2، ص:ص 114-123.
- 21 - عبد الله ، عصام الدين يوسف، (2015)، "تأثير القيم الشاذة في معالم نموذج الانحدار الخطي المتعدد"، رسالة دكتوراة، جامعة السودان للعلوم و التكنولوجيا، الخرطوم، السودان.
- 22 - الخثعمي، عبد العزيز سعد، (2018)، " الاخلال بافتراض خلو البيانات من القيم المتطرفة و اثره على استخدام تحليل الانحدار الخطي المتعدد في التنبؤ "، المجلة الالكترونية متعددة المعرفة لنشر الأبحاث العلمية و التربوية، العدد 5. جامعة ام القرى. المملكة العربية السعودية.

- 23 - Bross, I.D.J. (1961), "Outliers in Patternend Experiments: strategic Re-Appraisal". *Technometrics*. 3, 19 – 102.
- 24 - Huber, P. J., (1973) *Robust Regression a Symptotics Conjectures Monte Carlo*, *Ann. Statist.*, 19(2), pp. 136-140.
- 25 - AL-Joouri, S. (1976), "Test of Outliers". Unpublished, M.s.c Thesis University of Baghdad.
- 26 - Cook, R. D. (1977) "Detection of Influential Observationin Linear Regression *Technometrics*", 19: pp15 - 18.
- 27 - Tukey, j.w. (1977), "Exploratory Data Analysis". Addison – Wesley, reading, MA.
- 28 - Barnett, V. & Lewis, T. (1978), "Outliers in Statistical Data". John Wiley and Sons, New York.
- 29 - Belsley, D. et al, (1980), "Regression Diagnostics: Identifying Influential Data and Sources of Collinearity". Wiley, New York, p:17.
- 30 - Freeman, P.R. (1980), "On the Number of Outliers in Data from a linear Model in Basyesian Statistics". Ed. J. M.
- 31 - Chatterjee, S. et al (1986) "Influential Observations, High Leverage Points, and Outliers in Regression" *Journal of Statistical Science*, Vol. 1, No. 3, 379-416.
- 32 - Rousseeuw and Leroy (1987) "Robust Regression and outlier detection" John Wiley & Sons, New York.
- 33 - Kleinbaum, J, et al, (1988), "Applied Regression Analysis and other Multivaiable Methods". PWS-KENT Publishing Company, Boston, second edition, p: 210.
- 34 - Chatterjee, S. and Hadi, A. S. (1988) "Sensitivity Analysis in Linear Regression", New York: John Wiley.

- 35 - Neter, J. et al, (1990), “Applied Linear Statistical Models: Regression Analysis of Variance, and Experimental Designs”. (3rd edition). Irwin, Homewood, IL 60430, Boston, MA 02116, p:396.
- 36 - Fox, John, (1997), “Applied Regression Analysis, Linear Models, and Related Methods”. Sage publications, p:280.
- 37 - Hair, J. F. Jr., Anderson, R. E., Tatham, R. L. and Black, W. C. (1998), “Multivariate Data Analysis”. 5th edn. Prentice-Hall, New Jersey.
- 38 - Kamile. S. and Aysen. A. (2004), The Fuzzy Robust Regression Analysis, The Case Of Fuzzy Data Set Has Outlier, G.U. Journal of Science, Vol. 17(3), PP: 71-84.
- 39 - Github, (2018), “Regression datasets”, retrieved may 20, 2018 from <https://vincentarelbundock.github.io/Rdatasets/>

ملحق 1: بيانات نماذج الانحدار الخطي البسيط الاصلية و المعالجة (نماذج):-

book004 02 050 cars

Modulated Data البيانات المعالجة

Notes	No.	Y*	X1*
V Name		dist	speed
	1	2	4
	2	10	4
	3	4	7
	4	22	7
	5	16	8
	6	10	9
	7	18	10
	8	26	10
	9	34	10
	10	17	11
	11	28	11
	12	14	12
	13	20	12
	14	24	12
	15	28	12
	16	26	13
	17	34	13
	18	34	13
	19	46	13
	20	26	14
	21	36	14
	22	12.439166	7.123459
	23	64.07691	19.97116
	24	20	15
	25	23.970485	11.95573
	26	54	15
	27	32	16
	28	40	16
	29	32	17
	30	40	17
	31	50	17
	32	42	18
	33	56	18
	34	50.335257	17.01289
	35	11.142583	4.216164
	36	36	19
	37	46	19
	38	69.996878	23.9922
	39	15.528828	10.1173
	40	48	20
	41	52	20
	42	56	20

Original Data البيانات الاصلية

Notes	No.	Y	X1
V Name		dist	speed
	1	2	4
	2	10	4
	3	4	7
	4	22	7
	5	16	8
	6	10	9
	7	18	10
	8	26	10
	9	34	10
	10	17	11
	11	28	11
	12	14	12
	13	20	12
	14	24	12
	15	28	12
	16	26	13
	17	34	13
	18	34	13
	19	46	13
	20	26	14
	21	36	14
	22	60	14
	23	80	14
	24	20	15
	25	26	15
	26	54	15
	27	32	16
	28	40	16
	29	32	17
	30	40	17
	31	50	17
	32	42	18
	33	56	18
	34	76	18
	35	84	18
	36	36	19
	37	46	19
	38	68	19
	39	32	20
	40	48	20
	41	52	20
	42	56	20

	43	64	20
	44	66	22
	45	54	23
	46	70	24
	47	28.939166	10.62346
	48	29.611323	10.7546
	49	60.07691	21.47116
	50	85	25

	43	64	20
	44	66	22
	45	54	23
	46	70	24
	47	92	24
	48	93	24
	49	120	24
	50	85	25

المصدر: (<https://vincentarelbundock.github.io/Rdatasets/datasets.html>)

book010 02 056 geophones

Modulated Data البيانات المعالجة

Notes	No.	Y*	X1*
V Name		distance	thickness
	1	39.0776	278
	2	32.5776	278
	3	66.4375	251.9
	4	33.5776	278
	5	36.657129	279.9225
	6	32.157129	278.3225
	7	67.9375	251.1
	8	40.988372	273.3023
	9	49.888396	265.0446
	10	52.888396	263.8446
	11	74.960106	247.5106
	12	89.072532	240.1985
	13	50.584005	267.2664
	14	72.572532	250.1985
	15	31.25	278
	16	32.5	276
	17	25.278194	288.5549
	18	74.65204	248.9353
	19	36.25	283
	20	37.5	279
	21	38.75	279
	22	40	278
	23	41.25	276
	24	42.5	276
	25	43.75	275
	26	45	273
	27	46.25	274
	28	47.5	274
	29	48.75	275
	30	50	276
	31	51.25	275
	32	52.5	276
	33	53.75	276
	34	55	271
	35	56.25	274

Original Data البيانات الاصلية

Notes	No.	Y	X1
V Name		distance	thickness
	1	13.75	279
	2	15	278
	3	16.25	280
	4	17.5	278
	5	18.75	275
	6	20	274
	7	21.25	276
	8	22.5	277
	9	23.75	276
	10	25	275
	11	26.25	277
	12	27.5	278
	13	28.75	276
	14	30	276
	15	31.25	278
	16	32.5	276
	17	33.75	275
	18	35	273
	19	36.25	283
	20	37.5	279
	21	38.75	279
	22	40	278
	23	41.25	276
	24	42.5	276
	25	43.75	275
	26	45	273
	27	46.25	274
	28	47.5	274
	29	48.75	275
	30	50	276
	31	51.25	275
	32	52.5	276
	33	53.75	276
	34	55	271
	35	56.25	274

	36	57.5	270
	37	58.75	270
	38	61.25	270
	39	62.5	273
	40	63.75	270
	41	65	260
	42	66.25	267
	43	67.5	261
	44	68.75	260
	45	70	256
	46	71.25	251
	47	72.5	249
	48	73.75	248
	49	75	248
	50	76.25	250
	51	77.5	251
	52	78.75	253
	53	80	251
	54	81.25	251
	55	82.5	254
	56	83.75	257

	36	57.5	270
	37	58.75	270
	38	61.25	270
	39	62.5	273
	40	63.75	270
	41	65	260
	42	66.25	267
	43	67.5	261
	44	68.75	260
	45	70	256
	46	71.25	251
	47	72.5	249
	48	73.75	248
	49	75	248
	50	76.25	250
	51	77.5	251
	52	78.75	253
	53	80	251
	54	81.25	251
	55	82.5	254
	56	83.75	257

المصدر: (<https://vincentarelbundock.github.io/Rdatasets/datasets.html>)

book012 02 053 ironslag

Modulated Data البيانات المعالجة

Notes	No.	Y*	X1*
V Name		magnetic	chemical
	1	25	24
	2	22.986113	20.93057
	3	20.067977	22.97734
	4	21	18
	5	20	18
	6	13	10
	7	16	14
	8	14	16
	9	19	18
	10	15.227508	18.9545
	11	23	21
	12	20	20
	13	19	21
	14	15	15
	15	16	16
	16	16	15
	17	21.921355	24.71661
	18	15	19
	19	15	16
	20	15	15
	21	15	15
	22	17	13

Original Data البيانات الاصلية

Notes	No.	Y	X1
V Name		magnetic	chemical
	1	25	24
	2	22	16
	3	17	24
	4	21	18
	5	20	18
	6	13	10
	7	16	14
	8	14	16
	9	19	18
	10	10	20
	11	23	21
	12	20	20
	13	19	21
	14	15	15
	15	16	16
	16	16	15
	17	12	17
	18	15	19
	19	15	16
	20	15	15
	21	15	15
	22	17	13

23	25.473986	27.47977
24	21.427094	22.71872
25	18	21
26	22	24
27	20	15
28	21	20
29	21	20
30	21	25
31	25	27
32	22	22
33	18	20
34	21	24
35	21.331974	24
36	20	23
37	25	29
38	21.032357	23.90293
39	19.043519	20.91296
40	19	19
41	21.056974	25
42	16	15
43	16	16
44	26	27
45	28	27
46	28	30
47	30	29
48	21.228477	18.16617
49	28	25
50	20.963649	15.60228
51	28.36094	25.21055
52	30.125393	28.9582
53	33	25

23	18	24
24	16	22
25	18	21
26	22	24
27	20	15
28	21	20
29	21	20
30	21	25
31	25	27
32	22	22
33	18	20
34	21	24
35	18	24
36	20	23
37	25	29
38	20	27
39	18	23
40	19	19
41	16	25
42	16	15
43	16	16
44	26	27
45	28	27
46	28	30
47	30	29
48	32	26
49	28	25
50	36	25
51	40	32
52	33	28
53	33	25

المصدر: (<https://vincentarelbundock.github.io/Rdatasets/datasets.html>)

book016 02 033 medExpenses

Modulated Data البيانات المعالجة

Notes	No.	Y*	X1*
V Name		expenses	familysize
	1	8.04345	3
	2	7.65345	3
	3	7.57	3
	4	6.1	5
	5	5.205257	6.437676
	6	4.03	7
	7	12.1	2
	8	7.5	4
	9	12.1	2
	10	5.2987434	6.476807
	11	5.2185195	3.502362
	12	3.3	6

Original Data البيانات الاصلية

Notes	No.	Y	X1
V Name		expenses	familysize
	1	7.15	2
	2	6.93	3
	3	7.57	3
	4	6.1	5
	5	10.3	4
	6	4.03	7
	7	12.1	2
	8	7.5	4
	9	12.1	2
	10	8.88	5
	11	4.47	3
	12	3.3	6

13	7.35	4
14	6.78	4
15	11.1	2
16	12.226672	1.789951
17	7.53	3
18	9.3039913	3.466682
19	10.3	2
20	6.92	4
21	12.14885	2
22	4.66	5
23	7.08783	5.816478
24	6.3917836	3.341807
25	1.129757	8.387676
26	8.27	3
27	9.57	3
28	13.33389	2.930806
29	9.75	2
30	10.8	2
31	3.03	6
32	5.83605	4
33	10.35	2

13	7.35	4
14	6.78	4
15	11.1	2
16	7.54	5
17	7.53	3
18	9	4
19	10.3	2
20	6.92	4
21	12.95	2
22	4.66	5
23	13.27	3
24	4.2	4
25	5.4	7
26	8.27	3
27	9.57	3
28	10.5	6
29	9.75	2
30	10.8	2
31	3.03	6
32	5.03	4
33	10.35	2

المصدر: (<https://vincentarelbundock.github.io/Rdatasets/datasets.html>)

book017 02 039 MothEggs

Modulated Data البيانات المعالجة

Notes	No.	Y*	X1*
V Name		Eggs	BodyMass
	1	172	1.646
	2	192	2.3
	3	159	1.73
	4	106.59144	1.476657
	5	188.21255	1.838213
	6	135	1.477
	7	140	1.684
	8	140	1.666
	9	138	1.345
	10	166	1.797
	11	136	1.2
	12	131	1.744
	13	230.80427	2.338213
	14	202.76272	2.045814
	15	234.66791	2.132667
	16	180.58178	1.388081
	17	132	1.308
	18	143	1.154
	19	193.63942	1.515272
	20	127	1.248
	21	160	1.862
	22	172.53087	1.602469

Original Data البيانات الاصلية

Notes	No.	Y	X1
V Name		Eggs	BodyMass
	1	172	1.646
	2	192	2.3
	3	159	1.73
	4	125	1.92
	5	197	1.847
	6	135	1.477
	7	140	1.684
	8	140	1.666
	9	138	1.345
	10	166	1.797
	11	136	1.2
	12	131	1.744
	13	246	1.977
	14	185	1.433
	15	240	1.849
	16	203	1.616
	17	132	1.308
	18	143	1.154
	19	235	2.115
	20	127	1.248
	21	160	1.862
	22	174	1.482

	23	131.29021	1.744166
	24	165	1.733
	25	130	1.412
	26	144	1.438
	27	130.17139	0.24844
	28	149	1.496
	29	182.85735	1.382831
	30	244.61404	2.160934
	31	160	1.976
	32	155	1.589
	33	98.091435	1.136657
	34	140	1.66
	35	98.091435	1.136657
	36	148.69912	2.409048
	37	140.00218	0.836042
	38	152	1.505
	39	157.55913	2.766407

	23	124	1.74
	24	165	1.733
	25	130	1.412
	26	144	1.438
	27	212	1.785
	28	149	1.496
	29	217	1.854
	30	271	2.287
	31	160	1.976
	32	155	1.589
	33	108	1.495
	34	140	1.66
	35	109	1.573
	36	120	2.013
	37	193	1.843
	38	152	1.505
	39	95	1.668

المصدر: (https://vincentarelbundock.github.io/Rdatasets/datasets.html)

book023 02 47 starsCYG

البيانات المعالجة *Modulated Data*

Notes	No.	Y*	X1*
V Name		log.Te	log.light
	1	4.37	5.23
	2	4.56	5.74
	3	4.2928375	4.634462
	4	4.56	5.74
	5	4.3069	4.5079
	6	4.46	5.46
	7	4.4214435	4.344515
	8	4.28225	3.93125
	9	4.3465399	4.783176
	10	4.37	5.12
	11	4.3057106	4.231132
	12	4.43	5.45
	13	4.48	5.42
	14	4.1748649	4.216509
	15	4.29	4.26
	16	4.42	4.58
	17	4.23	3.94
	18	4.42	4.645139
	19	4.23	4.18
	20	4.2922803	4.375696
	21	4.29	4.38
	22	4.29	4.22
	23	4.5559185	5.701518
	24	4.5536842	5.6597
	25	4.38	5.02
	26	4.42	4.66

البيانات الاصلية *Original Data*

Notes	No.	Y	X1
V Name		log.Te	log.light
	1	4.37	5.23
	2	4.56	5.74
	3	4.26	4.93
	4	4.56	5.74
	5	4.3	5.19
	6	4.46	5.46
	7	3.84	4.65
	8	4.57	5.27
	9	4.26	5.57
	10	4.37	5.12
	11	3.49	5.73
	12	4.43	5.45
	13	4.48	5.42
	14	4.01	4.05
	15	4.29	4.26
	16	4.42	4.58
	17	4.23	3.94
	18	4.42	4.18
	19	4.23	4.18
	20	3.49	5.89
	21	4.29	4.38
	22	4.29	4.22
	23	4.42	4.42
	24	4.49	4.85
	25	4.38	5.02
	26	4.42	4.66

	27	4.29	4.66
	28	4.38	4.9
	29	4.4049432	5.290057
	30	4.3875184	4.195934
	31	4.38	4.42
	32	4.5503251	5.525695
	33	4.45	5.22
	34	4.2992876	4.357826
	35	4.4600672	5.616873
	36	4.39475	4.49375
	37	4.5864967	6.076887
	38	4.45	5.22
	39	4.6215435	6.320268
	40	4.43	5.57
	41	4.38	4.62
	42	4.45	5.06
	43	4.5	5.34
	44	4.45	5.34
	45	4.55	5.54
	46	4.45	4.98
	47	4.42	4.5

	27	4.29	4.66
	28	4.38	4.9
	29	4.22	4.39
	30	3.48	6.05
	31	4.38	4.42
	32	4.56	5.1
	33	4.45	5.22
	34	3.49	6.29
	35	4.23	4.34
	36	4.62	5.62
	37	4.53	5.1
	38	4.45	5.22
	39	4.53	5.18
	40	4.43	5.57
	41	4.38	4.62
	42	4.45	5.06
	43	4.5	5.34
	44	4.45	5.34
	45	4.55	5.54
	46	4.45	4.98
	47	4.42	4.5

المصدر: (<https://vincentarelbundock.github.io/Rdatasets/datasets.html>)

book024 02 030 TextPrices

Modulated Data البيانات المعالجة

Notes	No.	Y*	X1*
V Name		Price	Pages
	1	95	600
	2	19.95	91
	3	51.5	200
	4	167.90736	1030.518
	5	96	521
	6	48.5	315
	7	146.75	800
	8	92	800
	9	40.095266	371.5798
	10	85.5	488
	11	16.95	150
	12	9.95	140
	13	5.95	194
	14	58.75	425
	15	6.5	51
	16	62.208328	450.3671
	17	4.25	57
	18	115.25	900
	19	48.984061	316.9053
	20	6.5	104
	21	130.5	696
	22	7	294

Original Data البيانات الاصلية

Notes	No.	Y	X1
V Name		Price	Pages
	1	95	600
	2	19.95	91
	3	51.5	200
	4	128.5	400
	5	96	521
	6	48.5	315
	7	146.75	800
	8	92	800
	9	19.5	600
	10	85.5	488
	11	16.95	150
	12	9.95	140
	13	5.95	194
	14	58.75	425
	15	6.5	51
	16	70.75	930
	17	4.25	57
	18	115.25	900
	19	158	746
	20	6.5	104
	21	130.5	696
	22	7	294

	23	41.25	526
	24	169.75	1060
	25	71.25	502
	26	82.25	590
	27	12.95	336
	28	127	816
	29	41.5	356
	30	31	248

	23	41.25	526
	24	169.75	1060
	25	71.25	502
	26	82.25	590
	27	12.95	336
	28	127	816
	29	41.5	356
	30	31	248

المصدر: (https://vincentarelbundock.github.io/Rdatasets/datasets.html)

book025 02 027 toycars

البيانات المعالجة *Modulated Data*

Notes	No.	Y*	X1*
V Name		distance	angle
	1	0.43	1.3
	2	0.37	1.3
	3	0.27	1.3
	4	0.84	4
	5	0.5800656	2.389784
	6	0.69	4
	7	0.58	2.7
	8	0.64	2.7
	9	0.55	2.7
	10	0.58	2.2
	11	0.8376662	3.969994
	12	0.54	2.2
	13	0.7	3.6
	14	0.7023771	3.130943
	15	0.66	3.6
	16	1	4.9
	17	0.496191	1.38911
	18	0.95	4.9
	19	0.27	0.9
	20	0.38	0.9
	21	0.727068	3.691014
	22	0.29	1.1
	23	0.43	1.1
	24	0.6042108	3.119585
	25	0.63	3.1
	26	0.69	3.1
	27	0.6	3.1

البيانات الاصلية *Original Data*

Notes	No.	Y	X1
V Name		distance	angle
	1	0.43	1.3
	2	0.37	1.3
	3	0.27	1.3
	4	0.84	4
	5	0.92	4
	6	0.69	4
	7	0.58	2.7
	8	0.64	2.7
	9	0.55	2.7
	10	0.58	2.2
	11	0.7	2.2
	12	0.54	2.2
	13	0.7	3.6
	14	0.89	3.6
	15	0.66	3.6
	16	1	4.9
	17	1.3	4.9
	18	0.95	4.9
	19	0.27	0.9
	20	0.38	0.9
	21	0.15	0.9
	22	0.29	1.1
	23	0.43	1.1
	24	0.17	1.1
	25	0.63	3.1
	26	0.69	3.1
	27	0.6	3.1

المصدر: (https://vincentarelbundock.github.io/Rdatasets/datasets.html)

book029 02 021 #ChildSpeaks

البيانات المعالجة *Modulated Data*

Notes	No.	Y*	X1*
V Name		Gesell	Age
	1	95	15

البيانات الاصلية *Original Data*

Notes	No.	Y	X1
V Name		Gesell	Age
	1	95	15

	2	71	26
	3	68.983594	28.68854
	4	91	9
	5	102	15
	6	87	20
	7	93	18
	8	100	11
	9	104	8
	10	94	20
	11	113	7
	12	96	9
	13	101.11731	7.976538
	14	96.026958	8.995507
	15	102	11
	16	100	10
	17	105	12
	18	89.804348	17.94348
	19	114.3755	8.719379
	20	94.169	11
	21	100	10

	2	71	26
	3	83	10
	4	91	9
	5	102	15
	6	87	20
	7	93	18
	8	100	11
	9	104	8
	10	94	20
	11	113	7
	12	96	9
	13	83	10
	14	84	11
	15	102	11
	16	100	10
	17	105	12
	18	57	42
	19	121	17
	20	86	11
	21	100	10

المصدر : M.R. Mickey, O.J. Dunn, and V. Clark

book034 02 22 electricity consumption استهلاك الكهرباء بالمنزل

Modulated Data البيانات المعالجة

Original Data البيانات الاصلية

Notes	No.	Y*	X1*
V Name		consump	time
	1	1.51	2
	2	2.89	3.33
	3	1.67	2
	4	1.32	1.47
	5	1.74	2.16
	6	1.71	1.8
	7	2.99	3.56
	8	1.2	1.33
	9	0.9051822	1.592341
	10	1.11	1.5
	11	0.52	0.8
	12	1.13	1.53
	13	2.55	2.8
	14	4.42	4.92
	15	1.23	1
	16	0.59	1.67
	17	1.53	1.67
	18	0.49	0.95
	19	3.24	3.55
	20	4.68	4.87
	21	2.6051125	3.020284
	22	1.2825284	1.616952

Notes	No.	Y	X1
V Name		consump	time
	1	1.51	2
	2	2.89	3.33
	3	1.67	2
	4	1.32	1.47
	5	1.74	2.16
	6	1.71	1.8
	7	2.99	3.56
	8	1.2	1.33
	9	5.37	8.5
	10	1.11	1.5
	11	0.52	0.8
	12	1.13	1.53
	13	2.55	2.8
	14	4.42	4.92
	15	1.23	1
	16	0.59	1.67
	17	1.53	1.67
	18	0.49	0.95
	19	3.24	3.55
	20	4.68	4.87
	21	6.94	7.75
	22	3.19	6.33

المصدر : استهلاك الكهرباء

بيانات السرطان و التدخين و التدخين book035 02 025

Modulated Data البيانات المعالجة

Notes	No.	Y*	X1*
V Name		cancer	smoke
	1	84	77
	2	116	137
	3	123	117
	4	87.298994	81.45657
	5	118.27794	97.99899
	6	101	102
	7	118	111
	8	117.73344	110.0404
	9	85.505439	77.82799
	10	88	102
	11	104	91
	12	116.27474	104.9089
	13	86	107
	14	96	112
	15	88.183638	82.41397
	16	107.87114	93.10147
	17	113	125
	18	125	133
	19	110.92399	94.95657
	20	115	105
	21	79	87
	22	85	91
	23	120.54594	120.1654
	24	60	76
	25	51	66

Original Data البيانات الاصلية

Notes	No.	Y	X1
V Name		cancer	smoke
	1	84	77
	2	116	137
	3	123	117
	4	128	94
	5	155	116
	6	101	102
	7	118	111
	8	113	93
	9	104	88
	10	88	102
	11	104	91
	12	129	104
	13	86	107
	14	96	112
	15	144	113
	16	139	110
	17	113	125
	18	125	133
	19	146	115
	20	115	105
	21	79	87
	22	85	91
	23	120	100
	24	60	76
	25	51	66

المصدر: IBM 20 Spss Samples

ملحق 2: بيانات نماذج الانحدار الخطي المتعدد الاصلية و المعالجة (نماذج):-

book038 04 032 ant111

البيانات المعالجة *Modulated Data*

البيانات الاصلية *Original Data*

No.	Y*	X1*	X2*	X3*
	harvwt	id	plot	ears
1	5.16	3	3	43.5
2	2.93	40	4	40.5
3	1.73	185.5	5.5	20
4	6.79	256.5	4.5	42.5
5	3.255	219.5	3.5	31.5
6	2.65	77	5	32.5
7	5.04	115	7	43.5
8	3.519567	103.3549	31.574846	40.9725056
9	4.805	15.5	15.5	46
10	4.77	51	15	46.5
11	3.17	193.5	13.5	34
12	7.365	268	16	44
13	4.275	229.5	13.5	39
14	3.19	88.5	16.5	41.5
15	4.6	119.5	11.5	43.5
16	4.28671	230.02	13.496363	38.8763552
17	5.065	22.5	22.5	47
18	4.33	60	24	42.5
19	1.49	205	25	27
20	6.435	275	23	41.5
21	5.565	238	22	41
22	2.79	93	21	36
23	6.34	129.5	21.5	45.5
24	1.437341	206.2298	25.094597	26.2589922
25	4.51	29	29	41
26	4.8	67.5	31.5	43.5
27	1.97	212.5	32.5	25
28	7.07	283.5	31.5	45
29	6.235	249	33	46.5
30	3.515	103.5	31.5	41
31	6.125	140.5	32.5	46
32	2.785589	92.48943	20.924471	35.9425982

No.	Y	X1	X2	X3
	harvwt	id	plot	ears
1	5.16	3	3	43.5
2	2.93	40	4	40.5
3	1.73	185.5	5.5	20
4	6.79	256.5	4.5	42.5
5	3.255	219.5	3.5	31.5
6	2.65	77	5	32.5
7	5.04	115	7	43.5
8	2.02	151	7	50
9	4.805	15.5	15.5	46
10	4.77	51	15	46.5
11	3.17	193.5	13.5	34
12	7.365	268	16	44
13	4.275	229.5	13.5	39
14	3.19	88.5	16.5	41.5
15	4.6	119.5	11.5	43.5
16	2.665	158	14	56
17	5.065	22.5	22.5	47
18	4.33	60	24	42.5
19	1.49	205	25	27
20	6.435	275	23	41.5
21	5.565	238	22	41
22	2.79	93	21	36
23	6.34	129.5	21.5	45.5
24	3.16	166	22	50.5
25	4.51	29	29	41
26	4.8	67.5	31.5	43.5
27	1.97	212.5	32.5	25
28	7.07	283.5	31.5	45
29	6.235	249	33	46.5
30	3.515	103.5	31.5	41
31	6.125	140.5	32.5	46
32	3.52	177.5	33.5	45.5

المصدر : <https://vincentarelbundock.github.io/Rdatasets/datasets.html>

book044 04 037 corn

البيانات المعالجة *Modulated Data*

البيانات الاصلية *Original Data*

No.	Y*	X1*	X2*	X3*
	cornhec	soyhec	cornpix	soypix
1	165.76	8.09	374	55
2	96.32	106.03	209	218
3	104.642	118.7999	294.6296	177.909718
4	185.35	6.47	432	96

No.	Y	X1	X2	X3
	cornhec	soyhec	cornpix	soypix
1	165.76	8.0900002	374	55
2	96.32	106.03	209	218
3	76.08	103.6	253	250
4	185.35	6.4699998	432	96

5	118.205	125.6606	305.79191	259.610789
6	162.08	43.5	361	137
7	152.04	71.43	288	206
8	161.75	42.49	369	165
9	92.88	105.26	206	218
10	149.94	76.49	316	221
11	64.75	174.34	145	338
12	127.07	95.67	355	128
13	133.55	76.57	295	147
14	77.7	94.48	223	204
15	206.39	37.84	459	77
16	108.33	131.12	290	217
17	118.17	124.44	307	258
18	99.96	144.15	252	303
19	140.43	103.6	293	221
20	98.95	88.59	206	222
21	131.04	115.58	302	274
22	114.12	99.15	313	190
23	100.6	124.56	246	270
24	127.88	110.88	353	172
25	116.9	109.14	271	228
26	87.41	143.66	237	297
27	93.48	91.05	221	167
28	121	132.33	369	191
29	109.91	143.14	343	249
30	122.66	140.13	342	182
31	104.21	118.57	294	179
32	88.59	102.59	220	262
33	87.37556	146.9932	233.99372	303.129304
34	165.35	69.28	355	160
35	104	99.15	261	221
36	88.63	143.66	187	345
37	157.7	94.49	350	190

5	116.43	63.82	367	178
6	162.08	43.5	361	137
7	152.04	71.43	288	206
8	161.75	42.490002	369	165
9	92.88	105.26	206	218
10	149.94	76.489998	316	221
11	64.75	174.34	145	338
12	127.07	95.669998	355	128
13	133.55	76.57	295	147
14	77.7	94.480003	223	204
15	206.39	37.84	459	77
16	108.33	131.12	290	217
17	118.17	124.44	307	258
18	99.96	144.14999	252	303
19	140.43	103.6	293	221
20	98.95	88.589996	206	222
21	131.04	115.58	302	274
22	114.12	99.150002	313	190
23	100.6	124.56	246	270
24	127.88	110.88	353	172
25	116.9	109.14	271	228
26	87.41	143.66	237	297
27	93.48	91.050003	221	167
28	121	132.33	369	191
29	109.91	143.14	343	249
30	122.66	140.13	342	182
31	104.21	118.57	294	179
32	88.59	102.59	220	262
33	88.59	29.459999	340	87
34	165.35	69.279999	355	160
35	104	99.150002	261	221
36	88.63	143.66	187	345
37	157.7	94.489998	350	190

المصدر : <https://vincentarelbundock.github.io/Rdatasets/datasets.html>

book045 03 025 delivery

Modulated Data البيانات المعالجة

No.	Y*	X1*	X2*
	delTime	n.prod	distance
1	10.53675	3.89211	135.25563
2	11.5	3	220
3	12.03	3	340
4	17.81737	6.98715	209.44325
5	13.75	6	150
6	18.11	7	330
7	8	2	110
8	17.83	7	210
9	22.21714	8.22833	202.92243
10	21.5	5	605

Original Data البيانات الاصلية

No.	Y	X1	X2
	delTime	n.prod	distance
1	16.68	7	560
2	11.5	3	220
3	12.03	3	340
4	14.88	4	80
5	13.75	6	150
6	18.11	7	330
7	8	2	110
8	17.83	7	210
9	79.24	30	1460
10	21.5	5	605

11	18.55487	7.73715	241.94325
12	21	10	215
13	13.5	4	255
14	19.75	6	462
15	24	9	448
16	29	10	776
17	15.35	6	200
18	19	7	132
19	9.5	3	36
20	35.1	17	770
21	17.9	10	140
22	35	15	810
23	13.69851	5.96911	146.91087
24	19.83	8	635
25	10.75	4	150

11	40.33	16	688
12	21	10	215
13	13.5	4	255
14	19.75	6	462
15	24	9	448
16	29	10	776
17	15.35	6	200
18	19	7	132
19	9.5	3	36
20	35.1	17	770
21	17.9	10	140
22	52.32	26	810
23	18.75	9	450
24	19.83	8	635
25	10.75	4	150

المصدر : (1987) P. J. Rousseeuw and A. M. Leroy

book054 03 027 Metal

البيانات المعالجة *Modulated Data*

No.	Y*	X1*	X2*
	va	labor	capital
1	657.29	162.31	279.99
2	935.93	214.43	542.5
3	1110.65	186.44	721.51
4	1200.89	245.83	1167.68
5	1052.68	211.4	811.77
6	3406.02	690.61	4558.02
7	2427.89	452.79	3069.91
8	4257.46	714.2	5585.01
9	1625.19	320.54	1618.75
10	1272.05	253.17	1562.08
11	1004.45	236.44	662.04
12	598.87	140.73	875.37
13	853.1	145.04	1696.98
14	1165.63	240.27	1078.79
15	1917.55	536.73	2109.34
16	4239.106	711.408	5557.4236
17	1088.27	214.62	884.24
18	842.3882	200.5905	659.29025
19	3175.39	521.74	5686.99
20	1653.38	304.85	1701.06
21	5159.31	835.69	5206.36
22	1223.762	216.7405	993.97635
23	592.85	150.77	357.32
24	1601.98	259.91	2031.93
25	2065.85	497.6	2492.98
26	1058.247	211.6862	815.80678
27	745.67	137	768.59

البيانات الاصلية *Original Data*

No.	Y	X1	X2
	va	labor	capital
1	657.29	162.31	279.99
2	935.93	214.43	542.5
3	1110.65	186.44	721.51
4	1200.89	245.83	1167.68
5	1052.68	211.4	811.77
6	3406.02	690.61	4558.02
7	2427.89	452.79	3069.91
8	4257.46	714.2	5585.01
9	1625.19	320.54	1618.75
10	1272.05	253.17	1562.08
11	1004.45	236.44	662.04
12	598.87	140.73	875.37
13	853.1	145.04	1696.98
14	1165.63	240.27	1078.79
15	1917.55	536.73	2109.34
16	9849.17	1564.83	13989.55
17	1088.27	214.62	884.24
18	8095.63	1083.1	9119.7
19	3175.39	521.74	5686.99
20	1653.38	304.85	1701.06
21	5159.31	835.69	5206.36
22	3378.4	284	3288.72
23	592.85	150.77	357.32
24	1601.98	259.91	2031.93
25	2065.85	497.6	2492.98
26	2293.87	275.2	1711.74
27	745.67	137	768.59

المصدر : (1977) Aigner, D., K. Lovell and P. Schmidt

book061 03 030 Rubber

Modulated Data البيانات المعالجة

No.	Y*	X1*	X2*
	loss	hard	tens
1	372	45	162
2	206	55	233
3	175	61	232
4	154	66	231
5	136	71	231
6	112	71	237
7	55	81	224
8	45	86	219
9	221	53	203
10	196.075	68.02	172.95999
11	164	64	210
12	113	68	210
13	82	79	196
14	129.026	74.93588	188.0855
15	228	56	200
16	196	68	173
17	128	75	188
18	97	83	161
19	164.2373	63.94304	210.21596
20	249	59	161
21	219	71	151
22	186	80	165
23	155	82	151
24	114	89	128
25	341	51	161
26	340	59	146
27	283	65	148
28	367.6553	46.19997	161.25519
29	215	81	134
30	148	86	127

Original Data البيانات الاصلية

No.	Y	X1	X2
	loss	hard	tens
1	372	45	162
2	206	55	233
3	175	61	232
4	154	66	231
5	136	71	231
6	112	71	237
7	55	81	224
8	45	86	219
9	221	53	203
10	166	60	189
11	164	64	210
12	113	68	210
13	82	79	196
14	32	81	180
15	228	56	200
16	196	68	173
17	128	75	188
18	97	83	161
19	64	88	119
20	249	59	161
21	219	71	151
22	186	80	165
23	155	82	151
24	114	89	128
25	341	51	161
26	340	59	146
27	283	65	148
28	267	74	144
29	215	81	134
30	148	86	127

المصدر : Venables, W. N. and Ripley, B. D. (1999)

book062 04 028 salinity

Modulated Data البيانات المعالجة

No.	Y*	X1*	X2*	X3*
	Y	X1	X2	X3
1	7.6	8.2	4	23.005
2	7.7	7.6	5	23.873
3	4.3	4.6	0	26.417
4	5.9	4.3	1	24.868
5	5	5.9	2	27
6	6.5	5	3	24.2
7	8.3	6.5	4	23.215

Original Data البيانات الاصلية

No.	Y	X1	X2	X3
	Y	X1	X2	X3
1	7.6	8.2	4	23.005
2	7.7	7.6	5	23.873
3	4.3	4.6	0	26.417
4	5.9	4.3	1	24.868
5	5	5.9	2	29.895
6	6.5	5	3	24.2
7	8.3	6.5	4	23.215

9	13.2	10.1	0	22.274
10	12.6	13.2	1	23.83
11	10.4	12.6	2	25.144
12	10.8	10.4	3	22.43
13	13.1	10.8	4	21.785
14	12.3	13.1	5	22.38
15	12.62856	13.1987	1.0129819	23.8287408
16	13.50957	12.05041	3.9124872	21.3950115
17	12.89256	13.1867	1.1329819	23.8171008
18	9.5	7.7	3	22.686
19	12	10	0	21.789
20	12.6	12	1	22.041
21	13.6	12.1	4	21.033
22	14.1	13.6	5	21.005
23	13.5	15	0	25.865
24	11.5	13.5	1	26.29
25	12	11.5	2	22.932
26	13	12	3	21.313
27	14.1	13	4	20.769
28	15.1	14.1	5	21.393

9	13.2	10.1	0	22.274
10	12.6	13.2	1	23.83
11	10.4	12.6	2	25.144
12	10.8	10.4	3	22.43
13	13.1	10.8	4	21.785
14	12.3	13.1	5	22.38
15	10.4	13.3	0	23.927
16	10.5	10.4	1	33.443
17	7.7	10.5	2	24.859
18	9.5	7.7	3	22.686
19	12	10	0	21.789
20	12.6	12	1	22.041
21	13.6	12.1	4	21.033
22	14.1	13.6	5	21.005
23	13.5	15	0	25.865
24	11.5	13.5	1	26.29
25	12	11.5	2	22.932
26	13	12	3	21.313
27	14.1	13	4	20.769
28	15.1	14.1	5	21.393

المصدر (1997) Davison, A.C. and Hinkley, D.V.

book064 04 026 smoking

البيانات المعالجة *Modulated Data*

No.	Y*	X1*	X2*	X3*
	tc	qt	tt	qc
1	90	37	92	24
2	105	21	107	21
3	50	30	50	23
4	172	23	180	15
5	38	21	68	5
6	27.49417	10.15794	54.852901	2.30246746
7	37.98892	20.97784	67.912752	4.99861511
8	36	18	41	10
9	68	30	71	14
10	103	24	98	28
11	100	31	106	16
12	104.37	29.32566	105.90452	19.3716968
13	58	22	58	9
14	82	90	211	28
15	20	16	44	6
16	115.4235	38.55177	116.69585	22.4653934
17	120.27	28.72566	124.50452	22.3716968
18	127	51	146	40
19	30	13	30	5
20	89	5	84	4
21	101.1018	31.43673	106.99257	16.3374754
22	113	29	116	21
23	30	9	30	6

البيانات الاصلية *Original Data*

No.	Y	X1	X2	X3
	tc	qt	tt	qc
1	90	37	92	24
2	105	21	107	21
3	50	30	50	23
4	172	23	180	15
5	38	21	68	5
6	203	75	405	17
7	46	37	131	6
8	36	18	41	10
9	68	30	71	14
10	103	24	98	28
11	100	31	106	16
12	60	31	54	11
13	58	22	58	9
14	82	90	211	28
15	20	16	44	6
16	617	129	600	112
17	121	6	73	3
18	127	51	146	40
19	30	13	30	5
20	89	5	84	4
21	211	75	206	50
22	113	29	116	21
23	30	9	30	6

24	53	23	60	12
25	26	11	21	10
26	58	23	58	18

24	53	23	60	12
25	26	11	21	10
26	58	23	58	18

المصدر : <https://vincentarelbundock.github.io/Rdatasets/datasets.html>

book068 04 050 USArrests

البيانات المعالجة *Modulated Data*

No.	Y*	X1*	X2*	X3*
	Murder	Assault	UrbanPop	Rape
1	8.957419	191.6457	50.286216	19.560821
2	10	263	48	44.5
3	0.006155	17.92226	40.085146	4.72272109
4	8.8	190	50	19.5
5	9	276	91	40.6
6	7.9	204	78	38.7
7	3.3	110	77	11.1
8	9.703175	261.1901	49.737513	42.4222239
9	15.4	335	80	31.9
10	6.256752	87.18613	39.999298	10.0851627
11	5.3	46	83	20.2
12	2.6	120	54	14.2
13	10.4	249	83	24
14	7.2	113	65	21
15	2.2	56	57	11.3
16	6	115	66	18
17	9.01456	176.5648	69.6256	27.9524799
18	11.81615	272.289	77.870674	37.2159686
19	2.1	83	51	7.8
20	11.3	300	67	27.8
21	4.4	149	85	16.3
22	12.1	255	74	35.1
23	2.7	72	66	14.9
24	6.646321	111.4068	71.010512	14.9777455
25	9	178	70	28.2
26	6	109	53	16.4
27	4.3	102	62	16.5
28	12.2	252	81	46
29	2.1	57	56	9.5
30	7.4	159	89	18.8
31	11.4	285	70	32.1
32	11.1	254	86	26.1
33	13	337	45	16.1
34	0.8	45	44	7.3
35	7.3	120	75	21.4
36	6.6	151	68	20
37	4.9	159	67	29.3
38	6.3	106	72	14.9
39	10.38237	256.7613	41.560028	47.2168632
40	11.67526	258.358	79.375781	25.4724424

البيانات الاصلية *Original Data*

No.	Y	X1	X2	X3
	Murder	Assault	UrbanPop	Rape
1	13.2	236	58	21.2
2	10	263	48	44.5
3	8.1	294	80	31
4	8.8	190	50	19.5
5	9	276	91	40.6
6	7.9	204	78	38.7
7	3.3	110	77	11.1
8	5.9	238	72	15.8
9	15.4	335	80	31.9
10	17.4	211	60	25.8
11	5.3	46	83	20.2
12	2.6	120	54	14.2
13	10.4	249	83	24
14	7.2	113	65	21
15	2.2	56	57	11.3
16	6	115	66	18
17	9.7	109	52	16.3
18	15.4	249	66	22.2
19	2.1	83	51	7.8
20	11.3	300	67	27.8
21	4.4	149	85	16.3
22	12.1	255	74	35.1
23	2.7	72	66	14.9
24	16.1	259	44	17.1
25	9	178	70	28.2
26	6	109	53	16.4
27	4.3	102	62	16.5
28	12.2	252	81	46
29	2.1	57	56	9.5
30	7.4	159	89	18.8
31	11.4	285	70	32.1
32	11.1	254	86	26.1
33	13	337	45	16.1
34	0.8	45	44	7.3
35	7.3	120	75	21.4
36	6.6	151	68	20
37	4.9	159	67	29.3
38	6.3	106	72	14.9
39	3.4	174	87	8.3
40	14.4	279	48	22.5

41	3.8	86	45	12.8
42	12.12641	253.3916	73.639905	34.9031481
43	7.716393	178.6714	48.029806	19.0813338
44	3.2	120	80	22.9
45	2.2	48	32	11.2
46	8.5	156	63	20.7
47	4	145	73	26.2
48	5.7	81	39	9.3
49	2.6	53	66	10.8
50	6.8	161	60	15.6

41	3.8	86	45	12.8
42	13.2	188	59	26.9
43	12.7	201	80	25.5
44	3.2	120	80	22.9
45	2.2	48	32	11.2
46	8.5	156	63	20.7
47	4	145	73	26.2
48	5.7	81	39	9.3
49	2.6	53	66	10.8
50	6.8	161	60	15.6

المصدر: McNeil, D. R. (1977)

التكلفة المعدلة لهندسة التصميم book096 04 040

البيانات المعالجة Modulated Data

البيانات الاصلية Original Data

No.	Y*	X1*	X2*	X3*
	cost	sqft	inorout	yrexp
1	72.7	0.73	1	17
2	113.9693	0.676486	0	22.4401629
3	109.77	0.77	1	9
4	134.47	0.65	0	15
5	123.39	0.8	0	15
6	187.34	1.03	1	11
7	91.43	0.94	0	22
8	117.37	0.72	0	12
9	173.69	0.95	0	12
10	112.6	0.49	0	18
11	30.21642	0.08199	0	19.2884514
12	127.29	0.88	0	16
13	266.72	1.58	0	16
14	172.1	0.93	0	22
15	118.32	0.68	0	11
16	187.3139	1.029928	1.0001337	10.9998663
17	86.80862	0.622517	0	11.1217325
18	85.9061	0.357797	0	18.4174836
19	85.36	0.53	0	11
20	182.11	1.15	0	16
21	90.91351	0.367042	0	8.93068967
22	84.68029	0.549945	0.0316588	11.2532703
23	153.45	1.26	1	19
24	215.92	1.32	0	13
25	75.03	0.51	1	13
26	230.93	1.04	0	8
27	115.74	1.07	1	12
28	176.434	0.867751	0	20.4866883
29	247.6	1.38	0	16
30	61.03	0.34	0	19
31	182.0151	0.879712	0	20.1175823
32	185.46	0.89	0	20

No.	Y	X1	X2	X3
	cost	sqft	inorout	yrexp
1	72.7	0.73	1	17
2	440.48	1.92	0	20
3	109.77	0.77	1	9
4	134.47	0.65	0	15
5	123.39	0.8	0	15
6	187.34	1.03	1	11
7	91.43	0.94	0	22
8	117.37	0.72	0	12
9	173.69	0.95	0	12
10	112.6	0.49	0	18
11	281.36	1.25	0	12
12	127.29	0.88	0	16
13	266.72	1.58	0	16
14	172.1	0.93	0	22
15	118.32	0.68	0	11
16	382.82	1.57	0	12
17	97.26	1.29	0	12
18	304.42	1.44	0	15
19	85.36	0.53	0	11
20	182.11	1.15	0	16
21	139.88	1.57	1	10
22	63.89	1.16	1	19
23	153.45	1.26	1	19
24	215.92	1.32	0	13
25	75.03	0.51	1	13
26	230.93	1.04	0	8
27	115.74	1.07	1	12
28	315.28	1.21	1	13
29	247.6	1.38	0	16
30	61.03	0.34	0	19
31	419.84	1.59	0	12
32	185.46	0.89	0	20

33	269.56	1.51	0	11
34	143.06	0.95	0	17
35	187.31	1.52	0	14
36	135.52	0.67	1	9
37	302.55	1.81	0	9
38	82.97	0.34	0	11
39	105.61	0.88	0	11
40	64.75	0.38	0	18

33	269.56	1.51	0	11
34	143.06	0.95	0	17
35	187.31	1.52	0	14
36	135.52	0.67	1	9
37	302.55	1.81	0	9
38	82.97	0.34	0	11
39	105.61	0.88	0	11
40	64.75	0.38	0	18

المصدر : IBM 20 Spss Samples

الفترة التي يعيشها المريض بعد جراحة الكبد 3متغيرات 054 04 098book

البيانات المعالجة *Modulated Data*

البيانات الاصلية *Original Data*

No.	Y*	X1*	X2*	X3*
	فترة بعد الجراحة	درجة تخثر الدم	دليل الانذار	درجة اختبار الالتهاب
1	200	6.7	62	81
2	101	5.1	59	66
3	204	7.4	57	83
4	101	6.5	73	41
5	127.4676	4.443954	60.165517	75.2211492
6	80	5.8	38	72
7	80	5.7	46	63
8	127	3.7	68	81
9	70.65141	6	87.55891	18.7594928
10	203	3.7	76	94
11	329	6.3	84	83
12	65	6.7	51	43
13	171.4779	3.52269	83.664571	56.1183703
14	330	5.8	83	88
15	168	7.7	62	67
16	217	7.4	74	68
17	87	6	85	28
18	188.0732	5.946901	77.748826	60.2591546
19	215	7.3	68	74
20	172	5.6	57	87
21	109	5.2	52	76
22	136	3.4	83	53
23	70	6.7	26	68
24	220	5.8	67	86
25	276	6.3	59	100
26	144	5.8	61	73
27	181	5.2	52	86
28	180.3426	5.243954	62.040517	76.9711492
29	72	5.2	54	56
30	178	5.8	76	59
31	71	3.2	64	65
32	58	8.7	45	23
33	116	5	59	73
34	295	5.8	72	93
35	115	5.4	58	70

No.	Y	X1	X2	X3
	فترة بعد الجراحة	درجة تخثر الدم	دليل الانذار	درجة اختبار الالتهاب
1	200	6.7	62	81
2	101	5.1	59	66
3	204	7.4	57	83
4	101	6.5	73	41
5	509	7.8	65	115
6	80	5.8	38	72
7	80	5.7	46	63
8	127	3.7	68	81
9	202	6	67	93
10	203	3.7	76	94
11	329	6.3	84	83
12	65	6.7	51	43
13	830	5.8	96	114
14	330	5.8	83	88
15	168	7.7	62	67
16	217	7.4	74	68
17	87	6	85	28
18	34	3.7	51	41
19	215	7.3	68	74
20	172	5.6	57	87
21	109	5.2	52	76
22	136	3.4	83	53
23	70	6.7	26	68
24	220	5.8	67	86
25	276	6.3	59	100
26	144	5.8	61	73
27	181	5.2	52	86
28	574	11.2	76	90
29	72	5.2	54	56
30	178	5.8	76	59
31	71	3.2	64	65
32	58	8.7	45	23
33	116	5	59	73
34	295	5.8	72	93
35	115	5.4	58	70

36	184	5.3	51	99
37	118	2.6	74	86
38	120	4.3	8	119
39	151	4.8	61	76
40	148	5.4	52	88
41	95	5.2	49	72
42	75	3.6	28	99
43	483	8.8	86	88
44	153	6.5	56	77
45	191	3.4	77	93
46	123	6.5	40	84
47	311	4.5	73	106
48	256.7716	6.536415	54.744537	99.8423902
49	158	5.1	67	77
50	310	3.9	82	103
51	124	6.6	77	46
52	125	6.4	85	40
53	198	6.4	59	85
54	313	8.8	78	72

36	184	5.3	51	99
37	118	2.6	74	86
38	120	4.3	8	119
39	151	4.8	61	76
40	148	5.4	52	88
41	95	5.2	49	72
42	75	3.6	28	99
43	483	8.8	86	88
44	153	6.5	56	77
45	191	3.4	77	93
46	123	6.5	40	84
47	311	4.5	73	106
48	398	4.8	86	101
49	158	5.1	67	77
50	310	3.9	82	103
51	124	6.6	77	46
52	125	6.4	85	40
53	198	6.4	59	85
54	313	8.8	78	72

المصدر : عانض بن مشيب القحطاني، و عبدالله بن عبد الكريم الشبيحة

انتاج الذرة الرفيعة الحقيبة بعد السبعينات 04 042 book102

البيانات المعالجة *Modulated Data*

البيانات الاصلية *Original Data*

No.	Y*	X1*	X2*	X3*
	production	harvested	planted	productivity
1	150	300	310	500
2	152	300	312	507
3	280.9393	818.7464	1316.3816	341.788302
4	229	326	346	702
5	97	253	285	383
6	345	691	727	499
7	195	534	613	365
8	203	568	658	357
9	251	604	685	416
10	134	392	429	342
11	135	500	570	270
12	242	564	662	429
13	160	471	541	340
14	242	536	598	451
15	268	490	550	547
16	426	750	836	568
17	355	1079	1325	329
18	238	846	1052	281
19	446	1083	1271	412
20	347	1160	1362	299
21	281	817	1313	344
22	524	1054	1555	497
23	645	1473	1626	438
24	455	992	1187	459

No.	Y	X1	X2	X3
	production	harvested	planted	productivity
1	150	300	310	500
2	152	300	312	507
3	299	299	310	1000
4	229	326	346	702
5	97	253	285	383
6	345	691	727	499
7	195	534	613	365
8	203	568	658	357
9	251	604	685	416
10	134	392	429	342
11	135	500	570	270
12	242	564	662	429
13	160	471	541	340
14	242	536	598	451
15	268	490	550	547
16	426	750	836	568
17	355	1079	1325	329
18	238	846	1052	281
19	446	1083	1271	412
20	347	1160	1362	299
21	281	817	1313	344
22	524	1054	1555	497
23	645	1473	1626	438
24	455	992	1187	459

25	554	1295	1499	428
26	378	1126	1552	336
27	690	1405	1963	491
28	412	868	1227	475
29	483	1288	1479	376
30	388	1156	1531	336
31	449	500	509	898
32	208.455	575.8794	666.69419	358.947498
33	472	833	1124	567
34	466	1396	1474	334
35	384	477	709	805
36	572	1338	1653	428
37	555.1815	1290.909	1494.2866	431.010973
38	641	1605	1722	399
39	243.4507	538.4474	600.43378	451.276489
40	365	883	1591	413
41	579.6139	1344.192	1657.6436	431.304934
42	364	706	1241	516

25	554	1295	1499	428
26	378	1126	1552	336
27	690	1405	1963	491
28	412	868	1227	475
29	483	1288	1479	376
30	388	1156	1531	336
31	449	500	509	898
32	752	1361	1533	553
33	472	833	1124	567
34	466	1396	1474	334
35	384	477	709	805
36	572	1338	1653	428
37	647	973	1128	665
38	641	1605	1722	399
39	667	1253	1311	532
40	365	883	1591	413
41	754	1486	1764	507
42	364	706	1241	516

المصدر : وزارة الزراعة و الغابات الادارة العامة للتخطيط الاقتصاد و الزراعي