وزارة التعليم العالي والبحث العلمي
جامعة القادسية
كلية الإدارة والاقتصاد

خوارزميات ألتشخيص ألكلي لشواذ متعدد ألمتغيرات باستخدام مصفوفات تقدير الموقع و القياس الحصينة مع التطبيق

رسالة ماجستير إحصاء

إلى

مجلس كلية الإدارة والاقتصاد في جامعة القادسية وهي جزء

من متطلبات نيل شهادة الماجستير في علوم الإحصاء

تقدمت بها الطالبة

سوسن عبد الأمير حرج

بإشراف

أ.م.د. حسن سامي عريبي ألشمري

2021 A.D. 1442 A. H.

رسالة ماجستير

بعنوان

خوارزميات التشخيص الكلي لشواذ متعدد المتغيرات باستخدام مصفوفات تقدير الموقع و القياس الحصينة مع التطبيق

للطالبة

سوسن عبد الأمير حرج

بإشراف

أ.م.د. حسن سامي عريبي ألشمري

2021 A.D. 1442 A. H.

بسم الله الرحمن الرحيم

{وَمَن يَتَّقِ اللَّهَ يَجْعَل لَّهُ مَخْرَجًا وَيَرْزُقْهُ مِنْ حَيْثُ لَا يَحْتَسِبُ وَمَن يَتَوَكَّلْ عَلَى اللَّهِ فَهُوَ حَسْبُهُ إِنَّ اللَّهَ بَالِغُ أَمْرِهِ قَدْ جَعَلَ اللَّهُ لِكُلِّ شَيْءٍ قَدْرًا}

[الطلاق: 2- 3]

صدق الله العلي العظيم

ألإهـــــداء

أهدي جهدي المتواضع هذا إلى من يعنون لي كل الحياة بدءاً بأبي الذي قال لي يوما ما ستكونين ذات شأن , فكان دافعاً قوياً لي أن أواصل دراستي لأكون عند حسن ظنه ,مروراً بوالدتي سندي وعوني في هذه الحياة وصولاً لنبض قلبي ولروحي؛ ولدي ((أيهم)) لأكون له فخراً في حياته , ثم نفسي ألتي حملتها أكثر من طاقتها وأجهدتها أكثر مما يجب , فكانت نعم النفس الصابرة القوية التي ساندتي لأصل إلى ما كنت أتمنى , وستبقى بعون الله كذلك حتى أصل إلى كل ما أطمح أليه .

2021 A.D. **1442 A. H.**

شكر وامتنان

الحمد لله رب العالمين الذي هدانا لهذا وما كنا لنهتدي لو لا إذ هدانا الله , والشكر له على نعمه التي أنعم بها علينا والصلاة والسلام على أشرف الأنبياء والمرسلين محمد بن عبد الله وعلى آله الطيبين الطاهرين وصحبه ومن تبعه أجمعين .

يلزمني واجب الوفاء والعرفان بأن أتقدم بخالص الشكر والتقدير والامتنان إلى أستاذي ومعلمي المشرف على رسالتي الأستاذ المساعد الدكتور (حسن سامي عريبي) لتفضله علي بهذا الإشراف , وهذه ألمتابعه الحثيثة , ولما قدمه من جهود وتوجيهات سديدة ومتابعة مستمرة , ومنحني الكثير من وقته وجهده ووافر علمه، مما كان له بالغ الأثر في تذليل الصعوبات التي واجهتني خلال مسيرتي العلمية لأنتهل من وافر علمه لينير مسيرتي العلمية ولكي يخرج هذا البحث المتواضع كما هو عليه الآن. أسأل المولى العزيز القدير بأن ينعم عليه بالصحة والعافية والسعادة ومزيد من التقدم والنجاح , وأن يوفقه لخدمة المسيرة العلمية في بلدنا العزيز.

المستخلص :-

شهد القرن العشرين العديد من الدراسات والبحوث التي اثبتت أن وجود المشاهدات الشاذة في عينة البيانات يشكل خرقا لفرضية التوزيع الطبيعي الذي يشكل التوزيع الأساس في بناء النماذج الاحصائية. لذلك بذلت جهود كبيرة لتشخيص ألشواذ في بيانات التطبيقات العلمية المختلفة . اذ تم تصنيف الشواذ تبعا لنوع النموذج المستخدم ففي نموذج متعدد المتغيرات هي شواذ ولكن مع نموذج الانحدار الخطي المتعدد فأن مصطلح الشواذ يختص بتلك القيم التي تظهر في بواقي الانحدار أو المتغير المعتمد فقط. أما تلك القيم الشاذة التي تظهر في مصفوفة المتغيرات التوضيحية فصنفت الى Good Leverage Point و Bad Leverage point و High Leverage Point. نظراً لأهمية هذا الموضوع إقترح العديد من الباحثين طرقاً لتشخيص هذه الشواذ في النماذج الإحصائية المختلفة إلا ان هذه الطرق للأسف اعتمدت على إسلوب التشخيص الفردي للشواذ (مثل طريقة Hadi potential) لذلك اقترح عدد من الباحثين مؤخرا عدداً من طرق التشخيص الكلي Grouped Diagnostics لان طرق التشخيص الفردي أظهرت عدم قدرتها على كشف ظاهرتي الـ (Masking& Swamping): ظاهرة الــ Masking تحدث عندما تخفي قيمة شاذة قيمة أخرى على طريق التشخيص التي تكشف القيمة المؤثرة دون أن تكشف القيمة الشاذة الأخرى. اما ألــ Swamping فهو ان طريقة التشخيص تعد إحدى القيم على الأقل شاذة ولكنها في الواقع نظيفة. قدمت طريقة Imone كأول طريقة للتشخيص الكلي في الأدبيات الإحصائية لتشكل منطلقا للعديد من الباحثين اللذين سعوا لتطوير و تحسين اداء هذه الطريقة. لقد حسنت طريقة(DRGP(MVE من كفاءة طريقة Hadi في اختزال التشخيص الخاطئ Swamping و لكن إلى حد ما لأنها لم تتمكن من إختزال هذا النوع من التشخيص قياساً بطريقة(IDRGP(MVE.

لقد لاحظنا أمرين مهمين من خلال الطريقتين السابقتين أنهما لم يأخذا بنظر الاعتبار في طريقة Hadi ، الأول إن ظاهرتي (Masking& Swamping) ممكنتا الحدوث عند وجود قيم شاذة مؤثرة في حد الخطأ العشوائي و على وجه الخصوص نموذج الانحدار الخطي و بما إن Hadi وضع طريقته لمعالجة الشواذ لبيانات متعدد المتغيرات فإن طريقة DRGP(MVE) و طريقة IDRGP(MVE) تعاملتا مع مصفوفة المتغيرات التوضيحية (X)على أنها مصفوفة متعدد متغيرات لا يمكن القول عنها إنها مصفوفة متعدد متغيرات طبيعية لان فرضية التوزيع الطبيعي المشترك لهذه المتغيرات تم خرقها بسبب وجود الشواذ و إن عملية تثبيتها في نموذج انحدار خطي متعدد يفقدها صفة العشوائية و يجعلها خاضعة للقياس. اما ألأمر الثاني لاحظنا أن الأساس في بناء خوارزميات DRGP(MVE) و IDRGP(MVE) هي حساب مسافة مهنالوبيز بالاعتماد على مصفوفة الموقع والقياس MVE.

بناء على ما تقدم إقترحنا في هذه الرسالة التعامل مع مصفوفة المتغيرات التوضيحية لنموذج الانحدار الخطي المتعدد على أنها مصفوفة متعدد متغيرات تم خرق فرضية توزيعها المشترك بالشواذ. من جانب آخر وظفنا مصفوفة الموقع والقياس(RMVN) التي تعتمد على خوارزمية الخطوات الخمسة متعددة المراحل في ثلاث خوارزميات جديدة وهي ((DRGP(RMVN,

(IDRGP(MM) ,IDRGP(RMVN بهدف الحصول على إختزال جيد للتشخيص الخاطئ للطريقتين السابقتين , فضلاً عن مراعاة زمن الاحتساب وإستقرار كل طريقة.

لقد تم إخضاع الطرق كافة لعدد من دراسات المحاكاة ونوعين من البيانات الحقيقية (هندسية ومالية), وأظهرت النتائج أن طريقة DRGP(RMVN) المقترحة هي طريقة كفوءة وفعالة مع أحجام البيانات الكبيرة , أما طريقة IDRGP(RMVN) فهي طريقة كفوءة وفعالة مع أحجام البيانات الصغيرة. كما تبين أن الطريقتين المقترحتين سريعتان إلى حد ما لكنهما ليستا أسرع من طريقة IDRGP(LMS) وفقاً لنتائج المحاكاة. بالاضافة إلى ذلك لاحظنا أن طريقة IDRGP(MM) نجحت من خلال دراسة المحاكاة بالوصول إلى الشواذ كافة إلا إنها عانت من ظاهرة swamp. بالإضافة إلى ذلك لاحظنا ان طريقة IDRGP(LMS) طريقة سريعة جدا و فعالة الى حد ما خصوصا عن ارتفاع حجم العينة إلا إنها لم تكن مستقرة، علما إن كلتا الطريقتين لم تتعامل مع البيانات على انهما بيانات متعدد متغيرات لأنها ضمت المتغير المعتمد (Y) لنموذج الانحدار الخطي في مصفوفة بياناتها.

المحتويات

الفصل الأول		
1.1	تمهيد	1
2.1	المقدمة	3
3.1	**أهمية البحث**	4
4.1	**مشكلة البحث**	5
5.1	**أهداف البحث**	6
	الفصل الثاني – المبحث الاول - عرض المفاهيم النظرية و الدراسات السابقة	
1.1.2	**المقدمة**	8
2.1.2	**طريقة المربعات الصغرى الاعتيادية** Ordinary Least Squares (OLS)	8
3.1.2	**خصائص المقدرات الحصينة**	10
4.1.2	**منحنى الحساسية Sensitivity curve (SC)**	11
5.1.2	***Influence Function (IF)* دالة التأثير**	12
6.1.2	**نقطة الانهيار Beakdown Point (BP)**	12
7.1.2	**الكفاءة efficiency**	13
8.1.2	**مصفوفات معلمتي الموقع والقياس الحصينة**	13
9.1.2	**Minimum volume Ellipsoide Estimator (MVE)**	14
10.1.2	**Minimum covariance Determinan(MCD)**	16
11.1.2	**طرق تشخيص(الفردي) High Leverage Points Diagnostic Methods (HLP's)**	17
12.1.2	**Projection or Hat Matrix**	18
13.1.2	**قاعدة ثلاثة سجما Three Sigma Rule**	19
14.1.2	**طريقة مسافة مهنالوبيز Mahanalobis Distance (MD)**	20
15.1.2	**طريقة Hadi's Potential Measure**	21
16.1.2	**طرق التقدير الحصينة**	22
17.1.2	**M – Huber**	22
18.1.2	**S-Estimater**	23
19.1.2	**Least Median of Square Regression(LMS)**	24
20.1.2	**LTS- Estimator**	24
21.1.2	**MM Estimator**	25
الفصل الثاني - المبحث الثاني (خوارزميات التشخيص الكلي المقترحة)		
تمهيد		
1.2.2	**طريقة Generalized Potentials Measure (GP)**	27

2.2.2	طريقة DRGP(MVE) Measure	29
3.2.2	طريقةImprovised Diagnostic Robust Generalized Potential (IDRGP.MVE)	29
4.2.2	طريقةIDRGP.MVE based on LMS (IMLMS)	30
5.2.2	طريقة DRGP(RMVN) المقترحة	31
6.2.2	طريقة IDRGP(RMVN) المقترحة	34
7.2.2	طريقة((IDRGP(IRMM IDRGP.(RMVN) based on MM estimater	34
الفصل الثالث (النتائج)		
1.3	دراسة المحاكاة	35
2.3	الخلاصة	62
3.3	بيانات قوة انضغاط الخرسانة	66
4.3	بيانات القيمة السوقية للمصارف العراقية	69
الفصل الرابع (الاستنتاجات والتوصيات)		
1.4	المقدمة	73
2.4	مساهمة هذه الدراسة	73
1.2.4	اقتراح طريقة DRGP(RMVN) لتحسين أداء طريقة DRGP(MVE)	73
2.2.4	اقتراح طريقة IDRGP(RMVN) لتحسين أداء طريقة IDRGP(MVE)	74
3.2.4	اقتراح طريقة IDRGP(IRMM) لتحسين أداء طريقة IDRGP (LMS)	74
3.4	التوصيات	74

جدول المختصرات

الاسم باللغة الانكليزية	المختصر
Leverege Point	LP
Good Leverege Point	GLP
Bad Leverege Point	BLP
High Leverege Point	HLP
Influential Observations	IO
Ordenary Least Squares	OLS
Hat Matrix	Hat
(Best Linear Unbaised Estemater)	BLUE
Breakdown point	BP
Asymptotic Relative Efficiency	ARE
Influence Function	IF
Sensitivity curve	SC
Minimum volume Ellipsoide Estimator	MVE
Minimum covariance Determinan	MCD
Mahanalobis Distance	MD
Robust Mahanalobis Distance	RMD
Hadi's Potential Measure	Hadi
Least Median of Square Regression	LMS
M-scale M-estimator	MM
Generalized Potentials Measure	GP
Diagnostic Robust Generalized Potential (Manimum Volume Ellepoide)	DRGP(MVE)
Improvised Diagnostic Robust Generalized Potential (MVE)	IDRGP(MVE)
IDRGP(MVE) based on LMS	IMLMS
Reweighted Multivariate Normal Estimators	RMVN
Diagnostic Robust Generalized Potential(RMVN)	DRGP(RMVN)

Fast Consistent and High Breakdown	**FCH**
Reweighted FCH	**RFCH**
Median Ball	**MB**
Devlin Gnanadesikan Kettenring estimator	**DGK**
Improvised Diagnostic Robust Generalized Potential (RMVN)	**IDRGP(RMVN)**
IDRGP (RMVN)based on IDRGP(MM) estimater	**IDRGP(IRMM)**
Swamping	**Swamp**

فهرس الجداول

رقم الجدول	العنوان	الصفحة
(1)	يعرض معدلات التشخيص الصحيح والتشخيص غير الصحيح وزمن الاحتساب في LP للطرق المقارنة عندما n=50,100,300,500,1000 ونسبة تلويث المتغيرات التوضيحية 0.05 مع وجود HLP واحدة.	36
(2)	يعرض معدلات التشخيص الصحيح والتشخيص غير الصحيح وزمن الاحتساب في outlierse للطرق المقارنة عندما n=50,100,300,500,1000 ونسبة تلويث المتغيرات التوضيحية 0.05 مع وجود HLP واحدة.	39
(3)	يعرض معدلات التشخيص الصحيح والتشخيص غير الصحيح وزمن الاحتساب في LP و outlierse للطرق المقارنة عندما n=50,100,300,500,1000 ونسبة تلويث المتغيرات التوضيحية 0.05 مع وجود HLP واحدة.	42
(4)	يعرض معدلات التشخيص الصحيح والتشخيص غير الصحيح وزمن الاحتساب في LP للطرق المقارنة عندما n=50,100,300,500,1000 ونسبة تلويث المتغيرات التوضيحية 0.10 مع وجود HLP واحدة.	45
(5)	يعرض معدلات التشخيص الصحيح والتشخيص غير الصحيح وزمن الاحتساب في outlierse للطرق المقارنة عندما n=50,100,300,500,1000 ونسبة تلويث المتغيرات التوضيحية 0.10 مع وجود HLP واحدة.	48
(6)	يعرض معدلات التشخيص الصحيح والتشخيص غير الصحيح وزمن الاحتساب في LP و outlierse للطرق المقارنة عندما n=50,100,300,500,1000 ونسبة تلويث المتغيرات التوضيحية 0.10 مع وجود HLP واحدة.	51
(7)	يعرض معدلات التشخيص الصحيح والتشخيص غير الصحيح وزمن الاحتساب في LP للطرق المقارنة عندما n=50,100,300,500,1000 ونسبة تلويث المتغيرات التوضيحية 0.15 مع وجود HLP واحدة.	53
(8)	يعرض معدلات التشخيص الصحيح والتشخيص غير الصحيح وزمن الاحتساب في outlierse للطرق المقارنة عندما n=50,100,300,500,1000 ونسبة تلويث المتغيرات التوضيحية 0.15 مع وجود HLP واحدة.	56
(9)	يعرض معدلات التشخيص الصحيح والتشخيص غير الصحيح وزمن الاحتساب في LP و outlierse للطرق المقارنة عندما n=50,100,300,500,1000 ونسبة تلويث المتغيرات التوضيحية 0.15 مع وجود HLP واحدة.	59
(10)	يعرض تشخيص (masking & swamping) لطرق (Hadi ,MD ,Hat) بالمقارنة مع طريقتي DRGP(RMVN)& DRGP(MVE) لبيانات قوة انضغاط الخرسانة	67
(11)	يعرض تشخيص(masking & swamping)لطرق (DRGP.RMVN ,DRGP.MVE) بالمقارنة مع طريقتي IDRGP(RMVN)& IDRGP(MVE) لبيانات القيمة السوقية	70
(12)	يمثل خلاصة نتائج المحاكاة لبيان أفضلية الطرق المقارنة من ناحية معدل دقة التشخيص الصحيح ومعدل التشخيص الخاطئ ومعدل زمن الاحتساب باستخدام المؤشرات	ملحق
(13)	يبين فهرسة الصفوف التي تتضمن قيماً شاذة حسب طريقة التشخيص الفردي(RMD, Hat , Out) والكلي (DRGP.RMVN, DRGP.MVE, Hadi) لبيانات قوة انضغاط الخرسانة	ملحق
(14)	جدول يبين فهرسة الصفوف التي تتضمن قيماً شاذة لطرق التشخيص الكلي (DRGP.RMVN, DRGP.MVE , IDRGP.RMVN , IDRGP.MVE) لبيانات القيمة السوقية لعينة من المصارف المسجلة في سوق العراق للأوراق المالية	ملحق
(15)	يبين فهرسة الصفوف التي تتضمن قيماً شاذة حسب طرق التشخيص ألكلي (DRGP.RMVN, DRGP.MVE , IDRGP.RMVN , IDRGP.MVE) لبيانات قوة أنضغاط الخرسانة	ملحق
(16)	يبين فهرسة الصفوف التي تتضمن قيماً شاذة لطريقة التشخيص الفردي (RMD, Hat , Out) و الكلي (DRGP.RMVN, DRGP.MVE, Hadi) لبيانات القيمة السوقية لعينة من المصارف المسجلة في سوق العراق للأوراق المالية	ملحق

جدول الأشكال

رقم الشكل	ألعنوان	رقم الصفحة
(1-3)	رسم الأخطاء العشوائية باستخدام Normal Q-Q	66
(2-3)	يوضح القيم الشاذة , LP و High LP	67
(3-3)	يعرض تشخيص طريقة (DRGP(RMVN) ,Hadi DRGP(MVE) ,RMD) لبيانات قوة انضغاط الخرسانة	69
(4-3)	الشكل يعرض تشخيص طريقة (DRGP(RMVN) , DRGP(MVE) , IDRGP(MVE), IDRGP(RMVN))لبيانات القيمة السوقية للمصارف العراقية	72

جدول الملاحق

نوع الملحق	الاسم
الملحق A	يمثل خلاصة نتائج المحاكاة لبيان أفضلية الطرق المقارنة من ناحية معدل دقة التشخيص الصحيح ومعدل التشخيص الخاطئ ومعدل زمن الاحتساب باستخدام المؤشرات.
الملحق B	يبين فهرسة الصفوف التي تتضمن قيماً شاذة حسب طرق التشخيص الفردي(Out , Hat , RMD) والكلي (DRGP.RMVN, DRGP.MVE, Hadi) لبيانات قوة انضغاط الخرسانة.
الملحق C	يبين فهرسة الصفوف التي تتضمن قيماً شاذة لطرق التشخيص الكلي(DRGP.MVE , DRGP.RMVN, IDRGP.MVE , IDRGP.RMVN) لبيانات القيمة السوقية لعينة من المصارف المسجلة في سوق العراق للأوراق المالية.
الملحق D	يبين فهرسة الصفوف التي تتضمن قيماً شاذة حسب طرق التشخيص الكلي (DRGP.MVE , DRGP.RMVN, IDRGP.MVE , IDRGP.RMVN) لبيانات قوة انضغاط الخرسانة.
الملحقE	يبين فهرسة الصفوف التي تتضمن قيماً شاذة لطريقة التشخيص الفردي (Out , Hat , RMD)والكلي (DRGP.RMVN, DRGP.MVE, Hadi) لبيانات القيمة السوقية لعينة من المصارف المسجلة في سوق العراق للأوراق المالية.

1.1 تمهيد

لم يكن موضوع تشخيص الشواذ في بيانات العينات المأخوذة من مجتمعاتها الإحصائية موضع اهتمام الباحثين في المجالات العلمية المختلفة بل إن هذا الموضوع اختلف عليه الإحصائيون إلى حدما و انقسموا إلى مدرستين مختلفتين : المدرسة الكلاسيكية التي تمسكت بالأسس النظرية للتوزيع الاحتمالي الطبيعي التي أطلقها Gaussعام (1809) لجعل النماذج الإحصائية أكثر مقبولية من الناحية النظرية والحسابية. اذ ان Gauss وضع عدداً من الفرضيات لتحقيق ذلك من خلال افتراضه ألمشاهدات العينة المسحوبة عشوائيا من مجتمع الدراسة لها فرص الظهور نفسها و مسحوبة بشكل مستقل ؛ و إن هذه المشاهدات لها توزيع متماثل.لقد لاحظ Gauss أن هذه البيانات من النوع المستمر تتشكل على هيئة الشكل الناقوسي (جرس الكنيسة) أو (المنحنى) و من الممكن وصفه وفق نموذج رياضي دالي له القدرة على حساب المساحة التي يغطيها باستخدام التكاملات الرياضية.

أما مدرسة الحصانة القديمة فقد شككت في تعميم قضية التوزيع المتماثل لكل مشاهدات العينة العشوائية المسحوبة من مجتمع الدراسة ؛ اذ لاحظ Bereseعام (1852) أن هناك بعض المشاهدات تخرق هذه الفرضية التي وضعها Gaussو أن مصدرها ليس طبيعيا و مختلفاً كثيرا عن بقية المشاهدات التي من المفترض أن تتوزع طبيعيا، و لذلك من الممكن ان تأخذ إشكالاً أخرى قد تكون ملتوية إلى اليمين أو اليسار و ذات ذيل طويل. في عام (1947) أثارGeary الجدل حول فرضية التوزيع الطبيعي مرة أخرى عندما ذكر إن فكرة طبيعية المشاهدات ما هي الا خرافة و لم يكن هناك توزيعا طبيعيا و لن يكون.

على الرغم من هذه الإشكالات التي صدرت من باحثين عدة إلا إن الاعتقاد السائد كان يميل إلىإن خرق فرضية التوزيع الطبيعي سيقود بالضرورة إلى استدلال غير واقعي و غير مبني على أسس إحصائية معروفة و يقود الى عدم دقة التنبؤ.هذا السبب كان كافيا لعزوف الكثير من الباحثين عن تقبل فكرة الحاجة لوجود منحنيات أخرى غير منحنى فرضية التوزيع الطبيعي حتى نهاية القرن التاسع عشر (Pearson، 1905). لذلك حافظت الطرق الإحصائية التقليدية على شعبيتها على مر السنين على أمل أن لا تؤثر هذه الاعتراضات على الافتراضات المعيارية لهذه الطرق.

لم يقتصر التشكيك في انحراف فرضية التوزيع الطبيعي للنماذج الإحصائية على الإحصائيون فقط و انما ظهرت هذه المشكلة جليا في التطبيقات الإحصائية للعلوم المختلفة حيث عالم البيانات

الحقيقية ؛ اذ لاحظ العالم الفلكي Simone newcamb , والعالم الجغرافي Gefriey خطورة وجود ذيل طويل في منحنى توزيع الخطأ العشوائي وقدما نماذج احتمالية لنموذج الاخطاء (Gross errors) و دعوا الباحثين بجدية إلى إستخدام الطرق الحصينة للتقدير بدلا من الطرق التقليدية (Stigler،1973).

كان التركيز على إستعمال مصطلح الانحراف او اللاطبيعية في وصف الخروقات التي تحصل على فرضية التماثل للتوزيع الطبيعي حتى عام (1953) إذ استخدم Gerald Box مصطلح (robust) الحصانة في بحث نشره في العدد (40) من مجلة البايومتريكا بعنوان (- Non normality and test on variance) ترجمته إلى العربية اللاطبيعية و اختبار التباين إلاإن هذا المصطلح لم يتداول بكثرة و يكتسب صفته العلمية إلا في ستينات القرن العشرين.

مع مرور الزمن أصبح من الواضح إن إهمال افتراض انحراف فرضية التوزيع الطبيعي للنماذج الإحصائية قد يكون له عواقب وخيمة على تقدير معلمات ذلك النموذج و اختبار فرضياته.حيث أثبتت الدراسات النظرية والتطبيقية اللاحقة أن الانحرافات البسيطة لها في الغالب تأثيرات أكبر بكثير مما كان متوقعاً من قبل معظم الإحصائيين .

البحوث و الحلقات النقاشية التي نشرت في المجلات العلمية و نشرات المؤسسات الأكاديمية و المراكز البحثية في ستينات القرن العشرين شكلت نقطة التحول في تقبل أدوات الإحصاء ألحصين في الحصول على التقديرات. هذه المنشورات وضعت الأسس النظرية و التطبيقية لحقل إحصائي جديد يدعى الإحصاء الحصين. لقد بين Tukey عام (1960) الضعف في حصانة توزيعات بعض الطرق الاحصائية التقليدية و عزز هذا الرأي Huberعام (1964) باقتراح تقدير حصين لمعلمة الموقع و من ثم قدم نفس الباحث سلسلة من الأبحاث المنشورة عام (1965،1986) وصل من خلالها إلى العديد من النتائج المهمة، فمثلا في عام (1967) درس المشاكل التي تحدث في سلوك تقديرات طريقة الإمكان الأعظم عند وجود شروط غير قياسية وفي عام (1968) بين حدود الثقة للتقديرات الحصينة ثم وضع أول كتاب مرجعي يتناول الجهود المبذولة في الإحصاء الحصين عام (1973).

لقد تأثر Hampel كثيرا بطروحات Tuke و Huber فتناول أفكارهما بعمق من خلال أولاً أطروحة دكتوراه كتبت في هذا الموضوع بعنوان مساهمات في نظرية التقدير الحصين عام (1968) ثم عززها بورقة بحثية تناولت التعريف النوعي العام للحصانة عام (1971). لقد بين Hampel قواعد الرفض و التقديرات الحصينة لمعلمة الموقع في اجتماع إحصائي أوربا عام

(1974) مستخدما تحليلاً لبعض نتائج المحاكاة (المونتي كارلو) والتي أثرت على الأوساط الأكاديمية العلمية في أوربا و خصوصا في بلجيكا التي أنتجت Peter Rousseeuw و Maya Huber و Steven Van Iyliese و غيرهم من كبار الباحثين في مجال الإحصاء الحصين.

الحصانة لغوياً تعني المتانة، القوة، أوالثابت الذي لايكسر بسهولة أوالمقاوم للكسر لذا فإن استخدام مصطلح الإحصاء الحصين يشير الى ان تقديرات الطرق الحصينة لا تتأثر بالقيم الشاذة أو الخروقات الصغيرة في افتراضات النموذج .

2.1 المقدمة

أن أهمية الكشف أو تشخيص القيم الشاذة في مجموعة البيانات أمر مهم حيث أن هذه الشواذ بعيدة عن مركز تجمع البيانات من حيث الموقع بالرسم , و لا تتبع التوزيع العشوائي الذي تتبعه بقية البيانات, مما تسبب بإطالة ذيل منحنى ذلك التوزيع العشوائي. عرفها (Rousseuw and zomeren) عام (1990) بأنها مشاهدات تقع بعيدا عن معظم البيانات المتبقية و أنها تشكل ما نسبته (1%) الى (10%) من اي مجموعة بيانات حقيقية في عالمنا, ولامفر منها حتى و ان كانت هذه البيانات عالية الجودة، حسب أدعاء (Hampel) في عام (1986). اكد Huber عام (1981) أن وجود قيمة شاذة واحدة على الأقل في مجموعة البيانات يؤدي إلى انهيار المقدر الإحصائي لها.

لقد أثر ظهور القيم الشاذة على النماذج الإحصائية كافة سواء تلك التي تفترض حداً خطأ لها يتوزع طبيعيا ام لا يتوزع طبيعيا. من حيث ان القيم صنفت حسب النماذج الى عدة أنواع فتلك التي تظهر في الأخطاء العشوائية سميت بالقيم الشاذة أما تلك التي تظهر في المتغيرات التوضيحية لنموذج الانحدار الخطي المتعدد بكل أنواعه تدعى بالـ Leverage Points (LP) و يشمل هذا المفهوم نماذج متعدد المتغيرات مثل الانحدار الخطي المتعدد الاستجابة. لقد شخص Belsley et al. (1980) نوعاً آخر من المشاهدات اطلق عليها بالمشاهدات المؤثرة Influential Observations (IO) التي لها تاثير كبير لاسيما على النماذج الإحصائية التي تعتمد على تأثير متغيرات مستقلة على واحد أوأكثر من المتغيرات المعتمدة. إذإن معظم هذه النماذج تعتمد في إيجاد معالمها المقدرة على طريقة المربعات الصغرى التي هي حساسة جدا عند ظهور مثل هذه الشواذ بكل أنواعها لمزيد من المعلومات انظر (Rousseeow and Leroy, 1987; Gujarati, 2003).

لتحقيق أهداف الرسالة قسمت الى أربعة فصول, تضمن الفصل الأول التمهيد والمقدمة ومشكلة البحث وأهميته وأهداف الرسالة , أما الفصل الثاني فقد أشتمل على مبحثين , الأول تضمن عرض المفاهيم النظرية لموضوع الحصانة ومصفوفات الموقع والقياس بالإضافة الى بعض طرق التقدير الحصينة لمعلمات الانحدار وطرق التشخيص الفردي. أما المبحث الثاني فتناول الدراسات السابقة لطرق التشخيص الكلي للشواذ والطرق المقترحة (GP,DRGP(MVE),IDRGP(MVE), IMLMS) , (IDRGP(IRMM), IDRGP(RMVN), DRGP(RMVN) الفصل الثالث تضمن دراسة المحاكاة لكفاءة الطرق المقترحة بالمقارنة مع الطرق السابقة لتحقيق أهداف البحث بالإضافة إلى تطبيق الطرق المقترحة والسابقة على بيانات القيمة السوقية لسوق العراق للأوراق المالية وعلى بيانات قوة إنضغاط الخرسانة في حقل الهندسة المدنية أما الفصل الرابع فقد ذكر فيه الأستنتاجات والتوصيات التي توصل اليها الباحث .

3.1أهميةالبحث

تعد نماذج الانحدار الخطي المتعدد من أهم النماذج الإحصائية المستخدمة على نطاق واسع في العديد من المجالات العلمية و التطبيقية. إذ انه يصف علاقة أثر خطية لعدد من المتغيرات المستقلة على متغير معتمد واحد. طريقة المربعات الصغرى الاعتيادية هي من أقدم الطرق المعروفة جدا في الأوساط العلمية لسهولتها في تقدير معلمات نموذج الانحدار الخطي المتعدد من خلال تقليل مجموع مربعات بواقي الانحدار. تتميز هذه الطريقة بأنها تعطي أفضل تقديرات خطية غير متحيزة اذا تحققت فرضياتها. إذأن مقدراتها تحقق اقل تباين من تلك التي يمكن الحصول عليها من مجتمع الدراسة نفسه مع ثبات حجم العينة المسحوبة. واحدة من اهم الفرضيات ان التوزيع الاحتمالي لحد الخطأ العشوائي هو توزيع طبيعي بمتوسط قيمته صفر و تباين ثابت. من الناحية الواقعية قلما نجد بيانات حقيقية تخلو من القيمة الشاذة التي ظهورها في مجموعة البيانات يؤدي إلى خرق فرضية التوزيع الطبيعي لحد الخطأ العشوائي. اما ظهورها كـ LP فهو يصنف الى ثلاثة أنواع LP Good والتي يمكن اختصارها بـ (GLP) اوLP Bad باختصار (BLP) أوأنها High LP التي يمكن كتابتها باختصار (HLP) و هذه الأخيرة ممكن أن تكون جيدة

و مفيدة لتحسين تقديرات الانحدار و ممكن ان تكون مؤثرة. في الواقع وجد Imon and Khan (2003) أن لهذه الأخيرة لها تأثيراً اكبر من تأثير الشواذ في بواقي الانحدار و قد تكون مصدرا لخرق فرضية أخرى من فرضيات المربعات الصغرى ألا وهي فرضية استقلالية المتغيرات التوضيحية.لمزيد من المعلومات انظر: (Belsley et al., 1980; Hocking and Pendelton 1983; Rousseeow and Leroy, 1987;)

نظرا لأهمية تشخيص هذا النوع من الشواذ؛ قدم الباحث (Hadi) عام (1992) طريقته المشهورة Hadi's Potential إلا إن الباحثين Pena and Yohai أعلنا عام (1995) أن ظهور الـ HLP في مجموعة البيانات ربما تكون مصدراً لظهور مشكلة الـ Masking and Swamping في شواذ بواقي نماذج الانحدار (Mohammad , 2015). الأمر الذي دعى الى اقتراح طرق تشخيص متعددة او كلية بدلا من طريقة Hadi's Potential التي تعتمد التشخيص الفردي للـ LP. قدم Imon عام (1996) طريقة Generalized Potentials (GP) كإحدى طرق التشخيص المتعدد او الكلي. لم تكن الطريقة موفقة للأسف تماما في تشخيص العدد الكلي للـ HLPالأمر الذي دعى Habshah وآخرون عام (2009) الى اقتراح Developed the Diagnostic Robust Generalized Potential (DRGP) كطريقة أخرى لتشخيص العدد الكلي للـ HLP إلا أنها لم تقض تماما على تأثير ظاهرة Masking and Swamping لكن حجمت هذا التأثير اقل ما يمكن خصوصا عند أحجام العينات الصغيرة و تزايد نسب تلويث مجموعة البيانات. هذا السبب دفع (Mohammad *et.al* (2015 إلى اقتراح تحسين الطريقة السابقة من خلال طريقة جديدة أطلق عليها Improvised DRGP (IDRGP) التي رفعت معدل التشخيص و التقليل من معدل تاثير Masking and Swampingإلى أدنى حد ممكن.

لقد لاحظنا أن فكرة طريقة DRGP قد اعتمدت على حساب مسافة مهالانوبيز الحصينة التي اعتمدت مصفوفة موقع و قياس تدعى Minimum Volume Ellipsoid(MVE) اللتين حلتا محل متجه الأوساط الحسابية و مصفوفة التباين و التباين المشترك الكلاسيكية.

4.1 مشكلة البحث

لاحظنا الاهتمام الكبيرمن لدن الباحثين في تحسين أداء طريقة DRGP(MVE) ذلك من خلال التغلب على عيوبها و اقتراح طرق جديدة، لكن للأسف الطرق

المقترحة أيضا عانت أيضاً من بعض المشاكل مثل عدم الاستقرار في دقة التشخيص و كلفة عالية في زمن الاحتساب. لاحظنا أن بناء التشخيص الكلي يعتمد بالدرجة الأساس على تحديد مصفوفة موقع و قياس كفوءة يتم توظيفها لتحصين مسافة مهالانوبيز من خلال خوارزميات التركيز السريعة. واحدة من هذه المصفوفات التي اثبتت كفاءتها في الدراسات السابقة مصفوفة RMVN و علية يكون السؤال هل يمكن احلال RMVN محل MVE في خوارزمية DRGP سيحقق الاستقرار و سرعة في الأداء؟ و هذا السؤال ذاته يمكن أن يطرح عند استخدام طريقة IDRGP أيضا و كذلك الطريقة الأخيرة عند مقاطعتها مع بواقي LMS. أخيرا سنبحث عن كفاءة اقتراحنا مع الزيادة التدريجية في أبعاد مصفوفة البيانات من خلال مقارنتها بأداء الطرق السابقة.

5.1 أهداف البحث

يمكن تلخيص أهداف هذه الرسالة بالاتي:

1. اقتراح خوارزمية جديدة لتحسين أداء طريقة التشخيص الكلي DRGP(MVE) من خلال إستخدام مصفوفة الموقع والقياس (RMVN) تكون مستقرة أطلقنا عليها مسمى DRGP (RMVN).
2. اقتراح خوارزمية جديدة لتحسين أداء طريقة التشخيص الكلي IDRGP(MVE) من خلال إستخدام مصفوفة الموقع والقياس (RMVN) تكون مستقرة أطلقنا عليها مسمى IDRGP (RMVN).
3. اقتراح خوارزمية جديدة لتحسين أداء طريقة التشخيص الكلي IDRGP(MVE) باستخدام بواقي (LMS) من خلال أستخدام مصفوفة موقع وقياس بمسمى(IMLMS) وكذلك أستخدام مصفوفة الموقع والقياس (RMVN) مع بواقي طريقة (MM) تكون مستقرة أطلقنا عليها مسمى (IRMM) ودراسة المحاكاة بالمقارنة مع الطريقة السابقة.
4. إختبار سرعة الاحتساب الزمني لطرقنا المقترحة مع الطرق السابقة باستخدام المحاكاة .
5. اختبار كفاءة الطرق المقترحة باستخدام المحاكاة لأحجام عينة مختلفة ونسب تلويث مختلفة ومن ثم اختيار أفضل ألطرق المقترحة التي أثبتت كفاءتهما

في دراسة المحاكاة والتطبيق مع البيانات الحقيقية والمقارنة مع نظيراتها من الطرق السابقة (تشخيص فردي وكلي) .

ألمبحث الأول

عرض المفاهيم النظرية و الدراسات السابقة

1.1.2 المقدمة

يعد نموذج الانحدار الخطي المتعدد من أكثر النماذج استخداما في التطبيقات العلمية المختلفة لكونه نموذج يؤسس لعلاقة دالية لمجموعة من المتغيرات (اثنين فأكثر) على متغير معتمد واحد. يضاف إلى ذلك يميل الممارسون للإحصاء الى استخدام طريقة المربعات الصغرى الأكثر شيوعا لتقدير معالم هذا النموذج الذي يمكن استخدامه للتنبوء أيضا. هذا الفصل خصص لتناول قدر الإمكان الجهود العلمية المبذولة التي أسست لبعض المفاهيم النظرية المتناولة في الأدبيات الإحصائية التي شكلت الأساس النظري و ألمفاهيمي للدراسات السابقة للتشخيص الفردي و الكلي للقيم الشاذة و على نحو أكثر خصوصية في التشخيص الكلي لـ HLP's.

2.1.2 طريقة المربعات الصغرى الاعتيادية Ordinary Least Squares (OLS)

وهي من أكثر ألطرق استخداما وشيوعاً في تقدير معلمات نموذج الانحدار الخطي وصيغتها الرياضية كالآتي :-

$$y_{\mathrm{i}} = \beta_0 + \sum_{j}^{p} \beta_j X_{ij} + \varepsilon_i \qquad (2-1)$$

تعتمد هذه ألطريقه على فكرة (تصغير مجموع مربعات الأخطاء)، أي أن مجموع مربعات المسافة بين ألقيمة مقدرة عن نظيرتها الحقيقية هي أصغر مسافة ممكنة عن خط الانحدار، ويتم تربيع المسافة لمنع حذف المسافة الكبيرة الموجبة مع المسافات الكبيرة ألسالبه، حيث يكون خط الانحدار المقدر هو الخط المستقيم الذي يمر عبر هذه المسافات, و الذي يقطع المحور الصادي من خلال $\hat{\beta}_0$ وبميل صادي مقدر لذلك المستقيم هو$\hat{\beta}_1$، حيث يمكن كتابة معادلة الانحدار التقديرية كالآتي :

$$\widehat{y_i} = \hat{\beta}_0 + \sum_j^p \hat{\beta}_j \, x_{ij} \qquad (2-2)$$

إن الفكرة الرئيسة لهذه الطريقة هي تقليل دالة المربعات الصغرى L

$$L = \min_{\beta} \sum_{i=1}^{n} \varepsilon_i^2 \qquad (2-3)$$

و باشتقاق المعادلة أعلاه بالنسبة الى $\hat{\beta}_0$ مرة و $\hat{\beta}_1$ مرة أخرى و مساواة المعادلتين بالصفر من الممكن الوصول إلى حل أمثل وحيد لمعاملات الانحدار التقديرية و الذي يمكن كتابته بصيغة المصفوفات كالاتي:

$$\hat{\beta} = (x'x)^{-1}x'y \qquad (2-4)$$

أما القيم التقديرية فيمكن الحصول عليها من المعادلة الاتية،

$$\hat{y} = x\hat{\beta}$$
$$= x(x'x)^{-1}x'y$$
$$= Hy$$

حيث أن $H = x(x'x)^{-1}x'$ تعرف بأسم (Hat Matrix) أو (Projection Matrix).

ويعرف الفرق بين القيم الفعلية (y) والقيم المقدرة $\hat{y}$ ببواقي الانحدار ε

$$\varepsilon = y - x\hat{\beta}$$
$$= y - \hat{y}$$
$$= (I - H)y \qquad (2-5)$$

و تعد تقديرات طريقة OLS أفضل التقديرات الخطية غير المتحيزة إذا تحققت إفتراضاتها الأساسية فأن مقدرOLS سوف يكون غير متحيز ويمتلك أصغر تباين بين جميع المقدرات

الخطية الأخرى البديلة, ويطلق عليه حينئذ ب BLUE (Best Linear Unbiased Estimates) أفضل مقدر خطي غير متحيز.

أن ظهور قيمة شاذة واحدة على الأقل في الأخطاء العشوائية ستخرق فرضية التوزيع الطبيعي للأخطاء وهذه القيمة ربما تتوزع توزيعا طبيعيا لكن أما بمتوسط لا يساوي صفر او بتباين عالي او كليهما و ربما تتبع توزيعاً عشوائياً مختلفاً. على اي حال عند ظهور القيم الشاذة يتم اللجوء الى الحصول على المقدرات الحصينة (Huber 1981)

3.1.2 خصائص المقدرات الحصينة

حصانة المقدرات (Robustness) أمر في غاية الأهمية و تعد واحدة من أهم معايير المفاضلة بين المقدرات الإحصائية لمقاومة القيم الشاذة، الا ان حصانة المقدرات غير كافية مالم تقترن بالكفاءة Efficiency. مما لا يقبل الشك أن الكثير من الأدبيات الإحصائية أظهرت بشكل عام أن المقدرات الحصينة هي مقدرات أقل كفاءة قياسا بالمقدرات التقليدية في التوزيعات المتماثلة. على سبيل الفرض، إن الوسيط هو اقوى مقدر لمعلمة الموقع (Location parameter) ، متكافئ و متقارب، يملك افضل نقطة انكسار او انهيار Breakdown point، و اقل قدر من التحيز في التوزيعات المتماثلة. لكن ثمن متانة الوسيط كلفته الكفاءة المنخفضة مقارنة بالوسط الحسابي لنماذج التوزيع الطبيعي او نماذج التوزيعات ذات الذيل الخفيف. إذ إن الوسط الحسابي يتوزع توزيعا طبيعيا متماثلا، $\bar{X}\sim N(\mu,\frac{\sigma^2}{n})$ تتوزع، اما توزيع الوسيط $Med(X) \approx$ $N\left(\mu,\frac{\sigma^2}{n}\frac{\pi}{2}\right)$فهو طبيعي متماثل ايضا، لذلك فإن الكفاءة النسبية المقاربة Asymptotic Relative Efficiency (ARE) للتباين المقارب Asymptotic Variance للوسيط من التباين المقارب الوسط الحسابي هي:

$$ARE(Med,\bar{X}) = \frac{avar(\bar{X})}{avar(Med)} = \frac{2}{\pi} = 0.6366 \qquad (2-6)$$

نستدل من ذلك أن الكفاءة النسبية المقاربة للوسيط عن الوسط الحسابي هى (0.6366)، بمعنى آخر أن التباين المقارب للوسط الحسابي يكافيء ما نسبته (0.6366) من التباين المقارب للوسيط، أي أن التباين المقارب للوسيط هو أكبر من التباين المقارب للوسط الحسابي. بناءاً على

ما تقدم يمكن القول ان مقدر الوسط الحسابي أكثر كفاءة من مقدر الوسيط في التوزيعات المتماثلة.

بناء على ما تقدم وجد هيوبر (1981) منهجية الحد الأدنى (Minimax) التي اقترحها عام (1964) التي تعمل على المزاوجة بين الحصانة و الكفاءة و من ثم حدد الأهداف التي يجب إنجازها او محاولة إنجازها من هذه المنهجية و هي كالاتي:

أ. الكفاءة: الحصول على كفاءة جيدة الى حد معقول لمقدر (كفاءة مثالية او شبه مثالية) النموذج المفترض. إذ وضع Simpson عام (1995) الحصول على ما نسبته 90% الى 95% من كفاءة مقدرات طريقة المربعات الصغرى أمراً مرغوبا و هدفا للباحثين.

ب. دالة ألتأثير : يجب أن لا تؤثر الانحرافات الصغيرة عن الافتراضات النموذجية تأثيرا معنويا على ألإجراء الإحصائي المتبع.

ت. نقطة الإنهيار: لا ينبغي أن تتسبب الانحرافات الكبرى عن النموذج في حدوث انهيار للإجراء الإحصائي المتبع.

بناء على هذه المنهجية تم تحديد ثلاثة مقاييس للحصانة اشتهرت في الأدبيات الإحصائية و هي منحنى الحساسية ، دالة التأثير Influence Function (IF) و نقطة الانكسار او الانهيار Breakdown Point (BP) سيتم مناقشتها في هذا الفصل وكالاتي:

4.1.2 منحنى الحساسية Sensitivity curve(SC)

يقيس منحنى الحساسية تأثير قيمة شاذة واحدة على مقدر الإحصاءة ، و يكتب رياضياً كالاتي:

$$\mathrm{SC}_n(x,T) = n[T(x_1,x_2,\dots,x_{n-1},x_n) - T(x_1,x_2,\dots,x_{n-1})] \quad (2\text{-}7)$$

حيث إن $(x_1,x_2,\dots,x_{n-1})$ بيانات غير ملوثة و $(x_1,x_2,\dots,x_{n-1},x_n)$ بيانات ملوثة و إن T هو المقدر النموذجي للإحصاءة، مثل الوسط الحسابي او الوسيط.

افترض لدينا المشاهدات $X_{n-1} = \{x_1,x_2,\dots,x_{n-1}\}$ ، الآن دعنا نرى تأثير إضافة أي عدد حقيقي مثل x_n على الوسط الحسابي $\bar{X}_{n-1}$ باستخدام منحنى الحساسية لهذا المقدر و كالاتي:

$$\mathrm{SC}_n(x,\bar{X}) = \frac{((n-1)\bar{X}_{n-1} + x_n)/n - \bar{X}_{n-1}}{1/n} = x_n - \bar{X}_{n-1} \quad (2-7)$$

من الواضح أن القيمة المضافة x_n أما أكبر أو أقل من الوسط الحسابي و بالتالي فأن حساب الوسط الحسابي لكامل المشاهدات $\bar{X}_n$ سيختلف بالتاكيد عن الوسط الحسابي $\bar{X}_{n-1}$ بدون القيمة المضافة و لكن هذا الاختلاف بين الوسطين يعتمد على حجم القيمة المضافة ، فاذا كانت كبيرة او شاذة فسيكون الاختلاف كبيرا و إذا كانت صغيرة جدا سيكون الاختلاف كبيرا أيضا، إما إذا كانت قريبة فسيكون هناك اختلاف بسيط بينهما. مما تقدم يكشف منحنى الحساسية أن الوسط الحسابي حساس جدا للقيم الشاذة.

5.1.2 دالة التأثير(IF)Influence Function

منحنى التأثير او دالة التأثير اقترحها هامبل لأول مرة عام (1964)و عمل على تطويرها و الإستدلال (*heuristics*) بها مع مجموعة من الباحثين لتطوير اتجاهات الإحصاء الحصين الى أن تم تقديمها بصيغتها الحالية (1974). تعتبر دالة التأثير النسخة المقاربة لمنحنى الحساسية، إذ يتم من خلالها حساب المقدر T عند توزيع معين مثل F من دون الاعتماد على بيانات معينة. بمعنى أدق إن المقدر T ينبغي ان يكتب كدالة للتوزيع F، مثلاً $T(F) = E_F[X]$ للدلالة على الوسط الحسابي و $T(F) = F^{-1}(0.5)$ للدلالة على الوسيط.

مهمة دالة التأثير قياس التغيرات التي تطرأ على المقدر $T(F)$ نتيجة تلويث البيانات بـمشاهدة واحدة مثل x_0، و هو مفيد ايضا لحساب مصفوفات التباين و التباين المشترك لبعض المقدرات التي يصعب الحصول عليها بالطرق البسيطة.

6.1.2 نقطة الانهيار Beak down Point (BP)

عرف (Hambel1974) نقطة الانهيار أنها قياس لأعلى نقطة مقاومة للمقدر عندما تحتوي البيانات على نسبة عالية من الشواذ. بمعنى اخر، أنها النقطة الفاصلة ما بين انهيار دقة تقدير المقدر من عدمها. إذ إن التغيير الممكن حصوله في المقدر نتيجة تأثره بنسبة الشواذ في مجموعة البيانات بالتأكيد سيقود على استنتاجات غير دقيقة او تحليل إحصائي غير دقيق و غير كافي. لذلك عرف Donoho و Huber عام (1983) ، Rousseeuw, وLeroyعام (1987) و مارونا و آخرون عام (2006) ان الـ BP هي اصغر جزء من المشاهدات الرديئة التي يمكن ان تغير المقدر تغييرا كبيرا من خلال قيمة اعتباطية كبيرة. بناء على ذلك فان المقدر $T(D)$ لمجموعة البيانات D يمكن أن يتغير الى $T(D^*)$عند تلويث البيانات D^* بنسبة من القيم الشاذة، عليه فإن نقطة الانهيار هي

$$\mathrm{BP}(\mathrm{T}/D) = \min\left\{\frac{m}{n} : \underset{DD^*}{\mathrm{SUP}} \left\|\mathrm{T}(D) - \mathrm{T}(D^*)\right\| = \infty\right\} \qquad (2-8)$$

حيث أن D تمثل مجموعة البيانات قبل التلويث $(n-m)$ و ان D^* تمثل مجموعة البيانات بعد تلويثها بـ m من الشواذ.

وجد الباحثون أن نقطة انهيار المربعات الصغرى هي $\frac{1}{n}$ بمعنى أن مقدر هذه الطريقة ينهار بوجود قيمة شاذة واحدة في مجموعة البيانات (Huber 1973) . على العكس من ذلك بعض الطرق الحصينة تطرح مقدرات تصل نقطة انهيارها الى 50% مثل مقدرات طريقة وسيط المربعات الصغرى و مقدرات طريقة المربعات الصغرى المشذبة.

7.1.2 الكفاءة efficiency

يقال أن المقدر فعال أو كفوء غير متحيز أذا كان لاحجام عينة معينة لها تباين أصغر من أي مقدر غير متحيز أخر , فأذا كان T و T′ هما مقدرين غير متحيزين للمعلمة θ فأن :-

$$\mathrm{Var}(T) < \mathrm{Var}(T')$$

عندئذ يقال عن T أنه كفوء .

كما أن الكفاءة تعني الدرجة التي يؤديها المقدر مثل المربعات الصغرى في وجود توزيع الخطأ وعادة ما يتم التعبير عنها كنسبة مئوية ويمكن حسابها بأستخدام التباين أو متوسط مربع الخطأ (Mean square error) من المربعات الصغرى المناسبة مقسومة على تباين أو متوسط مربع الخطأ الحصين المناسب.

8.1.2 مصفوفات معلمتي الموقع والقياس الحصينة

تقدير معلمات متجه متوسط مجتمع متعدد المتغيرات و مصفوفة التباين و التباين المشترك لها من خلال سحب عينة عشوائية من مجتمع الدراسة هي الأكثر شيوعا لدى الإحصائيين و الممارسون للتطبيقات الإحصائية. لكن للاسف يصبح هذا التقدير غير موثوق به عندما تحتوي العينة المسحوبة عشوائياً على قيمة شاذة واحدة على الاقل ضمن بياناتها. لذلك إعتماد الأسلوب الاحصائي التقليدي لتقدير معلمات توزيع متعدد المتغيرات يؤدي الى عدم الدقة في تفسير العلاقة

بين المتغيرات قيد الدراسة الأمر الذي تطلب الى اللجوء الى إستخدام تقدير معلمتي الموقع والقياس الحصينتين لتمثيل هذه البيانات بشكل دقيق.

معلمة الموقع Location Parameter هي تقدير حصين لمتوسط المجتمع, فقد تكون مقدر الوسيط أو إنها وسط حسابي مرجح بإحدى الدوال الحصينة و ربما يكون وسطا حسابيا مشذبا على أي حال هو تقدير لا يتحسس و لا يتأثر من وجود القيم الشاذة في مجموعة البيانات.

افترض أن المشاهدات y_i تعتمد على معلمة غير معروفة مثل μ بالإضافة الى عملية عشوائية للخطأ فان أبسط عملية هي إضافة هذا الخطأ للمعلمة غير المعروفة وفقا لنموذج إحصائي احتمالي يطلق عليه نموذج الموقع :

$$y_i = \mu + \varepsilon_i, \qquad i = 1,2,\dots,n \qquad (2-9)$$

حيث أن ε_i متغير الخطأ العشوائي $\varepsilon \overset{iid}{\sim} N(0,\sigma^2)$ و $y_i \sim N(\mu,\sigma^2)$

وإن نموذج الموقع هذا يلخص الحصول على تقديرات النقاط وفترات الثقة لمعلمتي الموقع والقياس.

Minimum volume Ellipsoide Estimator (MVE) 9.1.2

تعد هذه الطريقة من أقدم الطرق التي قدمت بواسطة (Rousseeuw, 1984) و تمتلك نقطة انهيار breakdown point %50 تقريباً. فكرتها هي البحث عن أصغر حجم قطع ناقص من بين كل ألقطوع الناقصة (ellipsoids) التي من الممكن تشكيلها من نصف حجم المشاهدات إلاإنها تكون غير عملية في العينات الكبيرة عن عملية إيجاد حجوم كل القطوع الناقصة تأخذ وقتاً كبيراً جدا ، لذلك اقترح (Aelst and Rousseeuw, 1999) طريقة تقريبية لاختزال الوقت و الحصول على تقديرات معلمتي الموقع و القياس. فكرة الطريقة التقريبية أنها تفترض في الخطوة الاولى لها هو سحب مجموعة جزئية بحجم $p+1$ عشوائيا من مجموعة البيانات و من ثم يتم حساب المتوسط و التباين لهذه المجموعة الجزئية بالصيغ الاتية

$$\bar{T}_J = \frac{1}{p+1}\sum_{j=1}^{p+1} t_{i_j} \qquad (2-10)$$

$$C_J = \frac{1}{p}\sum_{j=1}^{p+1}\left(t_{i_j} - \bar{T}_J\right)'\left(t_{i_j} - \bar{T}_J\right) \qquad (2-11)$$

حيث أن i يمثل الصفوف

و j يمثل الاعمدة

فان C_J هي مصفوفة غير شاذة إذا كانت نقاط المجموعة الجزئية المسحوبة في الخطوة الاولى تتموضع في موقع عام خارج المستوى الزائدي hyperplane ، أما إذا كانت إحدى نقاطه على الأقل ضمن نطاق المستوى الزائدي فعندئذ ستكون شاذة و تحتاج إلى زيادة مجموعتنا الجزئية بالمشاهدات تدريجيا الى أن نحصل على مصفوفة غير شاذة وإذا لم يحصل ذلك نستمر بزيادة عدد المشاهدات إلى أن يصل حجم مشاهدات المجموعة الجزئية الى $h = \left[\frac{(n+p+1)}{2}\right]$ من المشاهدات التي عندها تكون المصفوفة C_J شاذة.

مما تقدم يتضح ان تحديد القطع الناقص يعتمد على متوسط و تباين بياناته و من ثم يتضخم و يتقلص إلى أن يضم بالضبط h من المشاهدات. مما يتطلب اخضاع ذلك الى معامل تصحيح $m_J^2/\chi_{p,\alpha}^2$،

حيث أن

$$m_J^2 = \left[median\left(t_{i_j} - \bar{T}_J\right)' C_J^{-1}\left(t_{i_j} - \bar{T}_J\right)\right]_{h:n} \qquad (2-12)$$

عليه حجم القطع الناقص حاصل ضرب معامل التصحيح بالجذر التربيعي لمحدد مصفوفة التباين والتباين المشترك، و التي تتناسب مع لمعادلة التالية :

$$\sqrt{det\left(m_J^2/\chi_{p,\alpha}^2\, C_J\right)} = \left(m_J/\sqrt{\chi_{p,\alpha}^2}\right)^p \sqrt{\det\left(C_J\right)} \quad (2-13)$$

الخوارزمية ستختار اصغر قيمة لدالة الهدف (2-13) من عدد كبير من المجموعات الجزئية

Minimum covariance Determinan(MCD) 10.1.2

قدمت هذه الطريقة بواسطة Rousseuw وهو مقدر حصين أخر لمتعدد المتغيرات للموقع والقياس ويمتلك أيضاً نسبة (50%) من نقاط الانهيار breakdown point مثل طريقة MVE (Rousseuw& leroy1987) تأخذ ال MCD المجموعة الجزئية من (Data Object) وهي (h observation out of n) وتكون هذه المجموعات الفرعية المثلى ل h من المشاهدات وعادة تكون h بين($\frac{n}{2}+1$)و $\left(\frac{n+p+1}{2}\right)$ وهي التي تعطي تباين مصفوفة العينة (sample covariance matrix) والتي لها أصغر محدد ممكن ,ويكون مقدر MCD للموقع والقياس هو متوسط ومصفوفة تباين تلك العينة .

إن أبسط طريقة حسابية للوصول إلى تقريب المقدر المضبوط لل MCD هو إعادة العينة بشكل عشوائي للمجموعات الفرعية من h من المشاهدات أو عناصر المجموعة بدون إرجاع لكل مجموعة عناصر ,ويتم حساب المتوسط الحسابي ومصفوفة التباين المشتركة كالاتي :-

1. سحب مجموعة جزئية عشوائيا H_{old}بحجم hمن مجموعة البيانات xحيث أن،

$$h = \left(\left[\frac{n}{2}\right] + 1\right) \qquad (2-14)$$

2. حساب متجه المتوسطات $\bar{T}_{H_{old}}$ و مصفوفة التباين و التباين المشترك $C_{H_{old}}$ باستخدام الصيغ الاتية:

$$\bar{T}_{H_{old}} = \frac{1}{h}\sum_{i\in J} t_i \qquad (2-15)$$

$$C_{H_{old}} = \frac{1}{h-1}\sum_{i\in J}\left(t_i - \bar{T}_{H_{old}}\right)'\left(t_i - \bar{T}_{H_{old}}\right) \qquad (2-16)$$

حيث أن $C_{H_{old}}$ ليست مصفوفة شاذة.

3. اذا كان محدد المصفوفة $C_{H_{old}}$ لا يساوي صفراً نستمر بالخوارزمية و بخلافه نتوقف
4. فاذا كان المحدد لها لا يساوي صفراً$\{\det(C_{H_{old}}) \neq 0\}$يتم حساب مربع مسافة مهالانوبيز على مجموعة البيانات الأصلية و باستخدام المقدرين أعلاه.

$$d^2_{old}(i) = \left(t_i - \bar{T}_{H_{old}}\right)' C^{-1}_{H_{old}}\left(t_i - \bar{T}_{H_{old}}\right), i = 1,2,\dots,n \qquad (2-17)$$

5. نرتب قيم مربع المسافة تصاعديا

$$d^2_{old}(\pi(1)) \le d^2_{old}(\pi(2)) \le \dots \le d^2_{old}(\pi(n))$$

حيث أن π تباديل $i = 1, 2, \dots, n$

6. الآن يتم تعريف مجموعة جزئية جديدة H_{new} من مجموعة البيانات مشروطة كالآتي

$$H_{new} = \{t_{\pi(1)}, t_{\pi(2)}, \dots, t_{\pi(h)}\} \qquad (2-18)$$

ثم يتم حساب $\bar{T}_{H_{new}}$ ، $C_{H_{new}}$و $d^2_{new}(i)$ بنفس الأسلوب السابق، فأذا كان $\det(C_{H_{new}}) = 0$ نعيد الخطوات السابقة. بخلافه اذا كانت $\det(C_{H_{new}}) <$ $\det(C_{H_{old}})$

نجعل $H_{old} := H_{new}$ ، $\bar{T}_{H_{old}} := \bar{T}_{H_{new}}$ و $C_{H_{old}} := C_{H_{new}}$ثم نعود لحساب $d^2_{old}(i)$ بنفس الأسلوب إلى أن نصل إلىأن$\det(C_{H_{new}}) = \det(C_{H_{old}})$ حيث تتوقف الخوارزمية.

11.1.2 طرق تشخيص (الفردي) High Leverage Points Diagnostic Methods (HLP's)

لقد أثبتت العديد من الدراسات أن التأثير الذي أشار إليه Huber ليس بالضرورة أن يكون سلبياً بل قد يكون تأثيراً ايجابيا. اذ يعتمد هذا الوصف على طبيعة ونوع البيانات المسحوبة او المسجلة من مجتمع الدراسة. الاعتقاد السائد لدى الباحثين أن القيم الشاذة هي مشاهدات رديئة على أي حال و هي مرفوضة و يجب استبعادها من تحليل البيانات. هذا اعتقاد خاطئ إلى حد ما، لان القيم الشاذة صنفت لاحقا تبعا لطبيعة البيانات، فإذا كانت (Univariate) فإنها ستؤثر بالتأكيد و لكن عندما تكون (Multivaraite) لنموذج انحدار مثلا فان وجودها في حد الخطأ العشوائي او البواقي مضراً و لكن ظهورها في مصفوفة المتغيرات التوضيحية يعتمد على طبيعة موقعها من خط الانحدار فقد تكون جيدة و نافعة و قد تكون ضارة. كما أن هناك نوعاً آخر يسمى بالمشاهدات المؤثرة استبعادها يؤثر في اتجاه خط الانحدار بمعنى اخر قد يغير قيمة معلمة الانحدار من الموجب الى السالب و بالعكس. مما تقدم نستنتج ان قرار الاستبعاد باستخدام الاعتقاد الشخصي غير المدروس لا يعد قرار حكيماً أو صائباً لان القيم الشاذة ليست بالضرورة قيم رديئة أو أنها توفر معلومات خاطئة بل قد تقدم معلومات خاطئة في حال استبعادها عن المشاهدات العادية. يرجع ذلك إلى حقيقة أنها تنشأ من تركيبة أو مجموعة غير

عادية من ألظروف التي قد تكون ذات أهمية حيوية تتطلب معها مزيداً من التحقق بدلاً من رفضها، وعليه فإن القيم الشاذة يجب أن لا ترفض تماماً ما لم يتم أرجاعها الى أسباب ظهورها مثل أخطاء التسجيل للمشاهدات.

مما تقدم يتضح ان عملية إجراء تشخيص القيم الشاذة و تصنيفها أمر ضروري جدا، لان وجودها يتطلب تشخيص نموذج الحصانة و قياس حصانة مقدرات ذلك النموذج. عموماً، لمعالجة مشكلة القيم الشاذة هناك اقتراحات عدة اما باستخدام دوال وزن للتقليل من تأثيرها مثل (دالة هيوبر، دالة توكي و دالة هامبل و غيرها) أو طرق التشذيب التي اقترحها Rousseuw في ثمانينات القرن العشرين من خلال دراساته لحصانة الانحدار الخطي و تشخيص الشواذ في بياناته.

بناء على ما تقدم اقترحت عدد من الطرق لتشخيص الشواذ (Rousseuw1984) في تحليل متعدد المتغيرات و اصطلح عرفا على تسمية شواذ متعدد المتغيرات بالـ Leverage Points للتوافق مع مفهوم القيمة الشاذة في مصفوفة المتغيرات التوضيحية في نموذج الانحدار و هي كالاتي :

Projection or Hat Matrix 12.1.2

وتعرف بمصفوفات الوزن وكذلك تسمى ب(Hat matrix) وتستخدم هذه المصفوفة في تحديد صفوف المشاهدات x تحتوي القيم الشاذة و تعرف تحليل الانحدار كمقياس للكشف عن وجود (LP) والصيغة الرياضية لها هي :-

$$H = x'\left(x'x\right)^{-1}x \qquad (2-19)$$

و إن عناصر القطر الرئيسي لمصفوفة الوزن H يمكن كتابتها كالاتي:

$$h_{\mathrm{ii}} = x'\left(x'x\right)^{-1}x_i \qquad (2-20)$$

هذه العناصر لها خصائص مفيدة وخصوصاً أن قيمها تتراوح بين (0 و1) وان مجموع h_{ii} يساوي p. بالإضافة الى ذلك يمكن إثبات أن h_{ii} المؤشر على احتوائها على LPللحالة i^{th} هي مقياس للمسافة بين قيم x للحالة i^{th} ومتوسط قيم x لجميع الحالات n، لذلك تشير القيمة الكبيرة

لـ h_{ii} إلى أن الحالة i^{th} بعيدة عن مركز كل مشاهدات المتغير x كلاهما وأن متوسط هذه المصفوفة يكتب بالصيغة الآتية:-

$$\bar{H} = \frac{\sum_1^n h_{ii}}{n} = \frac{p}{n} \qquad (2-23)$$

حيث أن p هو عدد المتغيرات و n هو العدد الكلي للمشاهدات .

أقترح (Hoagline and Welsch) عام 1978 ان تكون قيمة العتبة او نقطة القطع بوجود الـ LP في متغير دون سواه اذا كانت قيمة $H_{\mathrm{ii}} > \frac{2p}{n}$، و كذلك هناك قاعدة ثلاثة أضعاف المتوسط التي قدمها(Velleman and Welsch) في عام1981 لتشخيص LP عندما $H_{\mathrm{ii}} > \frac{3p}{n}$. وقدم Huber في عام 2004 نقطة قطع أخرى كمدى فاصل ثلاثي للمشاهدات< $0.2 < H_{\mathrm{ii}}$ 0.5ويجب تجنب قيم البيانات $H_{\mathrm{ii}} \geq 0.5$ حيث لاحظ Kutner*et.et al* في عام 2005 وجود فجوة بين قيم البيانات لمعظم الحالات والقيم الشاذة بشكل غير مألوف وهو دليل أخر على وجود القيم شاذة .وتجدر الإشارة إلى أن القواعد المذكورة أعلاه بأكملها لتحديد القيم الشاذة هي مخصصة لمجموعات البيانات الكبيرة الحجم نسبياً الى عدد المعلمات في نموذج الانحدار ومع ذلك أشار Hadi في عام 1992إلى إن مصفوفة الوزن قد تفشل في تحديد القيم الشاذة بسبب تأثير هذه القيم في هيكل البيانات .

13.1.2 قاعدة ثلاثة سجما Three Sigma Rule

هي إحدى الطرق المهمة في تشخيص القيم الشاذة تاخذ بنظر الإعتبار نسبة الفرق بين المشاهدة و متوسط العينة $\bar{x}$ إلى الانحراف المعياري s للعينة و كالاتي:

$$t_i = \frac{x_i - \bar{x}}{s}, i = 1,2,\dots,n \qquad (2-24)$$

المشاهدات التي قيمة القيمة المطلقة لـt_i المناظرة لها أكبر من 3 تعتبرشاذة.

الملاحظ أن هذه القاعدة تناظر تحويل المشاهدات ألى الدرجات المعيارية لها و هذا ما يبرر إستخدام القيمة (3) كعتبه لتشخيص الشواذ من عدمها، إذ إن $\mathrm{P}(|x| > 3) = 0.003$، بمعنى اخر ان احتمال تكون واقعة على مسافة تبعد ثلاث وحدات انحراف المعياري على جانبي منحنى التوزيع الطبيعي المعياري او القياسي هو ضعيف جدا او قد يكون معدوما و لذلك عند ظهور مثل

هكذا درجات تعتبر من الشواذ. لقد سجل Marona عام (2006) عدد من الملاحظات تتعلق بتطبيق هذه القاعدة كالاتي :

1. عندما تكون العينات ذات أحجام كبيرة وذات مشاهدات نظيفة فإن هذا العدد الكبير للمشاهدات سوف يؤدي إلى تشخيص قيم شاذة بشكل خاطئ و هذا ما يصطلح عليه بالـ Swamping
2. عندما تكون العينات ذات احجام صغيرة مثل $n \leq 10$ فأن هذه القاعدة سوف تكون غير فعالة.
3. إذا احتوت العينة على قيم شاذة كثيرة فإن بعض هذه الشواذ ستؤثر على غيرها مما يسبب اخفاءها وعدم تشخيصها باستخدام هذه القاعدة وهذا مايصطلح عليه بال (Masking).

هذه الأسباب دعت Marona في عام (2006) إلى الاستعانة بمعلمتي الموقع والقياس بدلا من الوسط الحسابي والانحراف المعياري في قاعدة ثلاثة سجما المذكورة في كالاتي :-

$$t_i = \frac{x_i - Med\ (x)}{MAD(x)}, i = 1,2, \dots, n \qquad (2-25)$$

حيث أن $Med\ (x)$ تشير وسيط قيم x ،

$$MAD(x) = Med\ |x_i _ Median(x_i)|\ /0.6745 \qquad (2-26)$$

اذ ان الثابت (0.6745) يستخدم لجعل الانحراف المعياري للعينة (s) مقدراً غير متحيز مقارباً لانحراف المعياري للمجتمع (S) وأن هذا الثابت هو قيمة الـ MAD لأي متغير عشوائي يتبع التوزيع الطبيعي القياسي.

14.1.2 طريقة مسافة مهنالوبيز Mahanalobis Distance (MD)

قدم Mahanalobis عام (1936) هذه الإحصاءة لقياس انحراف القيم عن مركزها مع الأخذ بنظر الاعتبار مصفوفة التباين والتباين المشترك ، إذ إنه يمكن التعبير عن المتغيرات التفسيرية للمشاهدات $\boldsymbol{i^{th}}$ في متجه الصف $\mathbf{x_i}$ حيث ان$\boldsymbol{x_i = (1, x_{i1}, x_{i2}, x_{i3}, \dots, x_{in})}$ ، عليه المعاملة ستكون صفية لـ $\mathbf{k}$ من الابعاد كالاتي:

$$MD_{\mathrm{i}} = \sqrt{(\boldsymbol{x_i} - \overline{\boldsymbol{x}})'\mathrm{c(x)}^{-1}(\boldsymbol{x_i} - \overline{\boldsymbol{x}})} \qquad \mathrm{i} = 1,2,,\ldots,\mathrm{n} \quad (2-27)$$

$$\mathrm{c(x)} = \frac{1}{\mathrm{n}-1}\sum_{\mathrm{i}=1}^{\mathrm{n}} (\boldsymbol{x_i} - \overline{\boldsymbol{x}})'\,(\boldsymbol{x_i} - \overline{\boldsymbol{x}}) \qquad (2-28)$$

حيث ان الصف $\boldsymbol{i^{th}}$ الذي يناظر قيمة مساحة مهنالوبيز اذا تجاوز القيمة الحرجة مربع كاي بدرجة حرية K ومستوى معنوية 0.95 يعد صفاً يحتوي على قيم شاذة، $MD_{\mathrm{i}} > \sqrt{\chi^2_{(K,0.95)}}$.

ان هذه الطريقة تعتمد على الوسط الحسابي و التباين اللذين كليهما حساسين جداً لوجود القيم الشاذة و لذلك تعد هذه الطريقة حساسة للقيم الشاذة ايضاً. و لتحصين هذه الطريقة لابد من استخدام معلمتي الموقع و القياس الحصينة كبدائل للوسط الحسابي و التباين. من أشهر البدائل المتوفرة في الادبيات الاحصائية مصفوفة أقل حجم للقطع الناقص Minimal Volume Ellipsoid (MVE) و أقل محدداً للتباين (Minimum Covariance Determinant) MCD و اللذان يضمنان معلمة الموقع الحصينة. بناء على ما تقدم يمكن اعادة كتابة مقياس مسافة مهالنوبيز الحصينة كالاتي:

$$RMD_{\mathrm{i}} = \sqrt{(\boldsymbol{x_i} - \overline{\boldsymbol{x}}_{\boldsymbol{Rob}})'\mathrm{c(x)}_{Rob}^{-1}(\boldsymbol{x_i} - \overline{\boldsymbol{x}}_{\boldsymbol{Rob}})} \qquad (2-29)$$

حيث أن Rob تعني Robust

و يبقى نفس الإجراء لتحديد الصف الذي يحتوي على قيم شاذة من خلال المقارنة مع القيمة الحرجة. $RMD_{\mathrm{i}} > \sqrt{\chi^2_{(K,0.95)}}$ فضلوا مقارنة قيمة مربع المسافة Midi *et.al 2009*. RMD^2 مع القيمة الحرجة لمربع كاي $\chi^2_{(\mathrm{K},0.95)}$ لتحديد الصف الذي يحتوي على القيم الشاذة.

15.1.2 طريقة Hadi's Potential Measure

في عام (1992) قدم Hadi طريقة جديدة لتشخيص (HLP) إذ قام بحذف الصفوف التي تحتوي على قيم شاذة من مصفوفة المتغيرات التوضيحية، بحذف الصف i^{th} في كل مرة و من ثم يتم ملائمة للبيانات بعد حذف الصف i^{th} لحساب قيمة p_{i} من المعادلة ادناه لكل مجموعة

فرعية محتملة من(n-1) من المشاهدات لتكوين مصفوفة الوزن والتي تعرف بـ (Potentials) كالآتي :-

$$p_{\mathrm{i}} = x_i \left(x_i^{'} x_i\right)^{-1} x_i^{'} \qquad (2-30)$$

حيث أن x_iهي مصفوفة البيانات مع حذف الصفi^{th}و التي يمكن إعادة صياغتها بالاعتماد على مصفوفة Hat matrix كالاتي،

$$p_{\mathrm{ii}} = \frac{\mathrm{w_{ii}}}{1 - \mathrm{w_{ii}}} \qquad (2-31)$$

هذه القيم تقارن بقيمة قطع لـ p_{ii} من خلال $mean(p_{ii}) + c\ SD(p_{ii})$

حيث أن c تأخذ قيمة ثابتة (2 او3) أي المشاهدات المناظرة لـقيمة p_{ii} و التي أعلى من نقطة القطع تعتبر (HLP) . صرح (Imon) في عام 2002 أن هذه الطريقة ناجحة في تشخيص نقطة (HLP) واحدة فقط في كل مرة ولكنها تفشل في تحديد مجموعة من النقاط (HLP) في المرة الواحدة.

16.1.2 طرق التقدير الحصينة

تعد تقديرات الانحدار الحصينة الاكثر كفاءةً واستقراراً ومقاومة للقيم الشاذة لذا سنتطرق إلى بعض الطرق المستخدمة في التقدير ضمن نطاق بحثنا :-

M – Huber 17.1.2

طور Huber عام (1973- 1981) مقدراً قوياً يسمى M- Estimator والذي يعتمد على فكرة أستبدال الدالة التربيعية لطريقة المربعات الصغرى بدالة هدف اخرى للحصول على مقدرات حصينة من خلال تقليل دالة الهدف الاتية :-

$$\min_{\beta} \sum_{i=1}^{n} \rho\left(\frac{r_i}{s}\right) \qquad (2-32)$$

حيث أن S هو تقدير القياس و ρ هو دالة متماثلة بحد أدنى عند الصفر .

أن أخذ المشتقة الجزئية بالنسبة لل β وجعلها مساوية للصفر ينتج عنه مجموعة من المعادلات العادية التي يمكن أن تحل مشكلة التقليل من خلال جعل ψ مشتقة لل ρ سنحصل على :-

$$\sum_{i=1}^{n} \psi \left(\frac{r_i}{s}\right) x_i = 0 \qquad (2-33)$$

يتم حساب تقديرات M بأستخدام المربعات الصغرى المرجحة تكراريا (IRLS) حيث يتم حساب البواقي r_i و تقسيمها على s و اخضاعها الى الدوال الترجيحية الحصينة مثل دالة Huber و دالة Tukey و غيرها من الدوال ثم حساب دالة ψ تمهيدا للحصول على مجموعة جديدة من الاوزانلنفس الدوال و يكرر هذا الاجراء عدة مرات حتى الحصول على تقارب على في النتائج.

S-Estimater 18.1.2

مقدر S هو مقدرحصين أخر يملك نقطة انهيار عالية اقترحه (Rosseeuw and Yohai 1984) مقدر S هو الحل الذي يوفر أصغر مقدر ممكن لتشتت البواقي كالاتي :-

$$\min \hat{\sigma}\left(r_1(\hat{\beta}), \dots, r_n(\hat{\beta})\right)$$

حيث أن $r_i(\hat{\beta})$ هو البواقي للحل المرشح لل β و بدلاً من تقليل تباين البواقي تم استخدام مقدر S لتقليل او تخفيض تقدير M لبواقي المقياس (scale)

$$\frac{1}{n}\sum_{i=1}^{n} \rho\left(\frac{r_i}{\hat{\sigma}_s}\right) = b \qquad (2-34)$$

حيث أنb هو ثابت ويتم حسابه كالتالي :-

$$b = E_{\emptyset}[\rho(r)] \qquad (2-35)$$

حيث $\emptyset$ يمثل التوزيع الطبيعي القياسي.

(Rosseeuw and Yohai 1984) أقترحوا استخدام دالة Tukey مع $c_B = 1.548 <$ b=0.1995 في المعادلة اعلاه الذي سيقدم نقطة الانهيار تصل الى (50%)

باشتقاق المعادلة (2-34) :-

$$\frac{1}{n}\sum_{i=1}^{n} \rho\left(\frac{e_i}{\hat{\sigma}_e}\right) = b \qquad \frac{1}{n}\sum_{i=1}^{n} \rho\left(\frac{u_i}{\hat{\sigma}_u}\right) = b$$

وحل المعادلة الناتجة عن الاشتقاق كما في ادناه

$$\frac{1}{n}\sum_{i=1}^{n} w\left(\frac{r_i}{\hat{\sigma_e}}\right) = b \quad (2-36)$$

حيث أن w هي دالة وزن مناسبة تؤدي الى أرتفاع نقطة أنهيار مقدر S .

Least Median of Square Regression(LMS) 19.1.2

تعتبر هذه الطريقة بديل حصين لطريقة المربعات الصغرى لتقدير معلمات الانحدار قدمها (Rousseuw) عام 1984. تسعى هذه الطريقة الى تخفيض وسيط مربعات بواقي الانحدار بدلاً من مجموع مربعات الخطأ في طريقة المربعات الصغرى (OLS) وكالاتي :-

$$\min_{\beta} Med\,(r_i^2)$$

وعلى الرغم من أن مقدر LMS يتمتع بنقطة أنهيار عالية (BP) و تساوي 50% ألا أنه يمتلك كفاءة نسبية منخفضة تبلغ 37% ومع ذلك يمكن إستخدام مقدر LMS لاكتشاف القيم الشاذة كمقدر ابتدائي للطرق الحصينة الاخرى وعلى وجه الخصوص مع مقدر MM

LTS- Estimator 20.1.2

أقترح (Rosseeuw 1984) طريقة المربعات الصغرى المشذبة باستخدام دالة الهدف الاتية :-

$$\min \sum_{i=1}^{h} r_i^2$$

حيث أن r_i^2 هو مربع البواقي وأن h يتم تحديدها كالاتي :-

$$h = [(1-\alpha)\times n] + [\alpha(p+1)] \quad (2-37)$$

حيث أن α هي النسبة المئوية للقيم الشاذة أو النسبة المئوية للمشاهدات المشذبة , n تمثل حجم العينة و p تمثل عدد معاملات الانحدار .

فكرة هذه الطريقة تقوم على سحب hمشاهدة من كافة العينات الممكن سحبها من العينة nو التي من خلالها يمكن الحصول على عينة واحدة على الاقل خالية من القيم الشاذة. لكن لايمكن أن يكون مقدر هذه الطريقة حصينا ما لم تكن النسبة المختارة α مساوية لنسبة القيم الشاذة فعلا في عينة البيانات.

تعد طريقة LTS طريقة مقاومة للقيم الشاذة وتصل قيمة نقطة أنهيارها الى (50%) (Rousseeuw and Leroy,2003) ذكر أن نقطة أنهيار الـ LTS هي :-

$$BP = \frac{\left(\left[n - \frac{p}{2}\right] + 1\right)}{n} \quad (2-38)$$

حيث أن n هو حجم العينة

و p هو عدد المعلمات

MM Estimator 21.1.2

قدم (Yohai) هذه الطريقة لتقدير معلمات الانحدار الخطي الحصين عام 1987 وتعد واحدة من أكثر التطبيقات شيوعاً في الاستخدام في المجالات العلمية التطبيقية المختلفة لأنهاتجمع بين نقاط الانهيار المرتفعه وبين الكفاءة العالية حيث تكون نقطة الانهيار ($BP = 0.5$) بينما تبلغ الكفاءة النسبية (95%) من كفاءة مقدر OLS في ظل الافتراضات الاساسية ,ويرمز له MM للدلالة على انه يتم أستخدام أكثر من عملية تقدير M لايجاد التقديرات النهائية ويتم حسابه كالاتي :-

1. يتم حساب التقديرات الابتدائية للمعاملات $T_\circ$ و $\theta_\circ$ التي تتفق مع أحدى الطرق الحصينة وتتمتع بنقطة أنهيار تساوي 50% .
2. أيجاد البواقي بألاعتماد على نتائج المرحلة الاولى وكالاتي :-

$$r(T_\circ, n) = Y_i - T_{\circ,n}^{t} X_i \qquad 1 \le i \le n \qquad (2-39)$$

ومن ثم يتم حساب البواقي

$$S_{\mathrm{n}} = \mathrm{S}\left(\mathrm{r}(\mathrm{T}_\circ, \mathrm{n})\right) \qquad (2-40)$$

بأستخدام الدالة $\rho_\circ$ التي تحقق فرضيات طريقة M-Huber بالاضافة الى أستخدام الثابت b كالاتي :-

$$\frac{b}{\mathrm{a}} = \frac{1}{2}$$

عندما $a = max\ \rho_\circ(r)$

3. أيجاد مقرات M-Huber بألاعتماد على دالة Ψ .

أفرض أن ρ_1دالة أخرى تحقق فرضيات M-Huber وبأستخدام الثابت b حيث أن

$$\rho_1(r) \le \rho_\circ(r)$$

$$\sup \rho_1(r) = \sup \rho_\circ(r) = a$$

حيث ممكن أستخدام العديد من الدوال الحصينة مثل Huber و Tukey عندها للحصول على مقدر MM , $T_{1,n}$ الذي من الممكن تعريفه على انه اي حل ممكن للدالة الاتية :-

$$\sum_{i=1}^{n} \Psi_1 \left[\frac{r_i(t)}{S_n}\right] X_i = 0 \qquad (2-41)$$

وهذه المعادلة يجب أن تحقق

$$S(T_{1,n}) \leq S(T_{\circ,n})$$

عندما

$$S(\theta) = \sum_{i=1}^{n} \rho\left(\frac{r_i(\theta)}{S_n}\right) (2-42)$$

المبحث الثاني

خوارزميات التشخيص الكلي المقترحة

تمهيد

حفز اقتراح Imon طريقة GPعام (2002) وحفز باحثين آخرين لتحسين أداء هذه الطريقة. لذلك قدمت Habshah وآخرون عام (2009) خوارزمية (DRGP) طريقة لتشخيص العدد الكلي للـHLP لكنها لم تكن دقيقة بما فيه الكفاية و خصوصا عند وجود ظاهرة Masking and Swamping حجمت تأثيرها إلى اقل ما يمكن عند أحجام العينات الصغيرة و تزايد نسب تلويث مجموعة البيانات. بناء على ذلك قدمMohammad A. Mohammad عام (2015) (IDRGP) التي رفعت معدل التشخيص و التقليل من معدل تاثير Masking and Swamping الى أدنى حد ممكن. في هذا الفصل سنستعرض طرق التشخيص الكلي و الطريقة المقترحة كالاتي:

1.2.2 طريقة Generalized Potentials Measure (GP)

فكرة هذه الطريقة التي اقترحها Imon عام (2002) تاتي بعد مرحلة التشخيص الفردي بالطرق الحصينة، إذإن الصفوف المتبقية و التي لم تكتشف الأساليب الحصينة عن وجود قيما شاذة في بياناتها تعزل في مصفوفة تدعى المتبقي (R)، أما الصفوف المشكوك بها و حذفت من المصفوفة الام فتوضع في مصفوفة تدعى(D). بمعنى ان R تحتوي على (n-d) من المشاهدات النظيفة وأن $d < (n-p)$ ، أخيرا يتم اعتبار المصفوفات السابقة بوصفها مصفوفات جزئية من مصفوفة الوزن w التي من الممكن تقسيمها كالآتي ،

$$w = \begin{bmatrix} U_R & V \\ V' & U_D \end{bmatrix} \qquad (2-43)$$

حيث أن $U_R = x_R \left(x'x\right)^{-1} x_R'$ و $U_D = x_D \left(x'x\right)^{-1} x_D'$ وهما مصفوفتان متماثلتان من (n-d) و d على التوالي $V = x_R \left(x'x\right)^{-1} x_D'$ وأستناداً الى مجموعة من الحالات المحذوفة المفهرسة بواسطة D يمكننا تحديدها من خلال :-

$$W_{ii}^{(-D)} = x_i^{'}\left(x_R^{'}x_R\right)^{-1}x_i \qquad i = 1,2,,\ldots,n \qquad (2-44)$$

وتجدر الإشارة إلى أن عناصر $W_{\mathrm{ii}}^{(-\mathrm{D})}$ لـ i^{th} من العناصر القطرية للمصفوفة $x\left(x_R^{'}x_R\right)^{-1}x^{'}$ فعندما يكون حجم R هو (p-i) و D=i فيكون :-

$$W_{ii}^{(-i)} = x_i^{'}\left(x_{(i)}^{'}x_{(i)}\right)^{-1}x_i = p_i \qquad (2-45)$$

الذي يبين أن $W_{\mathrm{ii}}^{(-\mathrm{D})}$ هو امتداد طبيعي لـ p_i ويمكن كتابة أي حالة i^{th} تمت أزالتها من المجموعة الفرعية المتبقية R والانضمام الى المجموعة الفرعية للحذف مثل :-

$$W_{\mathrm{ii}}^{-(\mathrm{D}+1)} = x_{\mathrm{i}}^{'}\left(x_{\mathrm{R}}^{'}x_{\mathrm{R}}\right)^{-1}x_{\mathrm{i}} + \frac{x_{\mathrm{i}}^{'}\left(x_{\mathrm{R}}^{'}x_{\mathrm{R}}\right)^{-1}{}^{2}x_{\mathrm{i}}}{1 - x_{\mathrm{i}}^{'}\left(x_{\mathrm{R}}^{'}x_{\mathrm{R}}\right)^{-1}x_{\mathrm{i}}} (2-46)$$

$$= \frac{W_{\mathrm{ii}}^{(-\mathrm{D})}}{1 - W_{\mathrm{ii}}^{(-\mathrm{D})}} \qquad (2-47)$$

وقدم (Imon) في عام GP 2002 كالآتي :-

$$P_{\mathrm{i}} = \begin{cases} W_{\mathrm{ii}}^{(-\mathrm{D})} & \forall \mathrm{i} \in \mathrm{D} \\ \dfrac{W_{\mathrm{ii}}^{(-\mathrm{D})}}{1 - W_{\mathrm{ii}}^{(-\mathrm{D})}} & \forall \mathrm{i} \in \mathrm{R} \end{cases} (2-48)$$

حيث أن D هي أي مجموعة من النقاط المحذوفة بشكل اعتباطي ويمكن للمرء أن يعتبر أن P_{ii} كبيرة أذا كانت فوق القيمة المقطوعة التالية :-

$$median(P_{\mathrm{ii}}) + cMAD(P_{\mathrm{ii}})$$

وأن C تأخذ على أنها عدد ثابت 2 او 3

وفي عام 2010 أقترحوا (Anwar & Habshah) قيمة القطع التالية ليتم تطبيقها عندما تكون نسبة (HLP) عالية أو متوسطة

$$median(P_{\mathrm{ii}}) + 3Q_n(P_{\mathrm{ii}}) (2-49)$$

حيث أن $Q_{\mathrm{n}} = c\{|x_i - x_J|: < j\}$

أي أنها إحصاءة الترتيب لكل المسافات الزوجية التي أقترحها (Rousseeuw) عام 1993 ويتم تحديد c=2.2219 وتعمل قاعدة القطع هذه على تحسين دقة تحديد نقاط (HLP) عندما تكون نسبة (HLP) متوسطة أوعالية .

2.2.2 طريقة DRGP(MVE) Measure

لقد أظهرت طرق التشخيص الشائعة الاستخدام فشلها في تحديد قيم (HLP) في مجموعة البيانات وذلك لظهور masking و swamping حتى عند تحصين مسافة مهالانوبيز (Robust MahanalobiseDestance RMD) أستناداً إلى حجم القطع الناقص (Minimum Volume Elepsoide بالصيغة آلاتية:

$$\boldsymbol{RMD_i(MVE) = \sqrt{(x - T_R(x))' C_R(x)^{-1}(x - T_R(x))} i = 1, 2, \dots, n(2 - 50)}$$

عام 2009 قدمت (Habshah*et. et.al*) محاولة جيدة لإيجاد حل وسط بين هذين النهجين حيث أتقرحوا طريقة تكيفيه تقوم بتحديد الـ (HLP) المشتبه بها بأساليب حصينة ثم بعدها القيام بالفحص التشخيصي ويتم إعادة (HLP) أن وجدت إلى مجموعة بيانات تقديرية مع الأخذ بفكرة Imonعام 2002 في تطوير GPواقترحوا ألخوارزمية الآتية :

1. إيجاد معلمتي الموقع والقياس باستخدام MVE
2. حساب مسافة مهالانوبيز الحصينة المعادلة (50-2) و تحديد المشاهدات المشكوك بها و وضعها في المصفوفة D الموصوفة في طريقة GP،اما بقية المشاهدات فتوضع في المصفوفة R.
3. نحسب P_{ii} كما في المعادلة (48-2) لمعرفة المشاهدات في المصفوفة D هل تحتوي على LP أم لا، فإذا تم تشخيصها تبقى في المصفوفة D و بخلافه تنقل الى المصفوفة R
4. اي قيمة من قيم P_{ii} تتجاوز قيمة القطع (49-2) تعد قيمة شاذة

3.2.2 طريقة Improvised Diagnostic Robust Generalized Potential (IDRGP(MVE)

هذه الطريقة قدمت من قبل (Mohammed A.Mohammed) عام (2015) لتصحيح بعض نقاط الضعف في طريقة DRGP(MVE) التي أشرنا إليها سابقا. إذ لاحظ (Mohammed) ان هناك تأثيراً للـ Masking and Swamping على الطريقة عندما تكون نسبة HLP

تتروح بين 5% او 10%. و قد أعزى السبب إلىإن تشخيص HLP في الخطوة الثانية من خوارزمية DRGP(MVE) التي وضعتها في مصفوفة جزئية أطلق عليها D لم يتم التحقق منها بشكل صحيح. لذلك اقترح إضافة خطوة جديدة الى الخوارزمية من خلال التشخيص باستخدام hat matrix . ومن ثم مقاطعة التشخيصين للتحقق من التشخيص الأول ,إذ يقارن فيها مابين ما تم تشخيصه كـ HLP بوصفها نتيجة نهائية للخوارزمية الأولى و التي افترضها بالمصفوفة الجزئية D_2و قارنها بما تم تشخيصه في الخطوة الثانية التي انتجت المصفوفة الجزئية D أي المقارنة ما بين D و D_2 كالاتي:

- الحالة الأولى: إذا كانت المشاهدات المشخصة كـ HLP هي ذاتها في D و D_2 فالخطوة الأخيرة للخوارزمية الإعلان عن هذا التشخيص و تتوقف.
- الحالة الثانية: إذا كان عدد المشاهدات المشخصة كـ HLP في D_2هي أكبر من عدد تلك المشخصة في D، عندئذ تعمل الخوارزمية على إعادة المشاهدات التي لم يتم تشخيصها في D_2الى المصفوفة Rواحدة تلو الأخرى على أن يتم حساب P_{ii} لكل واحدة منها في كل مرة و تقارن بقيمة نقطة القطع او العتبة في المعادلة (49-2) فإذا تجاوزت قيمة العتبة تضاف الى المشاهدة المشخصة و بخلافه تبقى مع المشاهدات النظيفة.
- الحالة الثالثة: اذا كان عدد المشاهدات المشخصة كـ HLP في D_2 هي اقل من عدد تلك المشخصة في D، أي تم تشخيص مشاهدات جديدة لم تشخص سابقا، عنئذ تعمل الخوارزمية على الدمج بين D و D_2 و من ثم يتم حساب P_{ii} لكل مشاهدة في المصفوفة الجديدة و تقارن بقيمة العتبة في المعادلة (49-2)، فاذا تجاوزت قيمة العتبة تضاف الى المشاهدة المشخصة و بخلافه تبقى مع المشاهدات النظيفة.

4.2.2 طريقة IDRGP.MVE based on LMS (IMLMS)

لتقليل تأثير HLP التي تظهر في البيانات كحالة متطرفة في البعدين X و Y تمت أضافة تشخيص القيم الشاذة لبواقي الانحدار المقدرة بواسطة طريقة LMS ومقاطعتها مع تلك القيم LP المشخصة بطريقة DRGP(MVE) فأذا كان تشخيص القيمة الشاذة واقعاً على نفس المستوى أو الصف الذي شخصت فيه LP يحذف من متجه تشخيص القيم الشاذة وبخلافه تضاف القيم الشاذة المشخصة الى المتجه النهائي للتشخيص .

5.2.2 طريقة DRGP(RMVN) المقترحة

الطريقة المقترحة تدعو الى توظيف مصفوفة الموقع و القياس Reweighted Multivariate Normal Estimators (RMVN) بدلا من MVE في خوارزمية DRGP(MVE) . إذاقترح Olive and Hawkins (2010) طريقة RMVN لإعادة ترجيح مقدر متعدد المتغيرات الطبيعي من خلال خوارزمية سريعة و متسقة و تملك نقطة انهيار عالية. لقد ركز الباحثان على خوارزمية التركيز ذات الخطوات الخمسة من خلال في المرحلتين الأولى و الثانية التي تتطلب حساب مصفوفة DGK للتباين و التباين المشترك المقترحة من لدن(Devlin et al. (1981 و مصفوفة الموقع و القياس الـ Median Ball (MB) المقترحة من قبل Olive (2004). و من ثم حساب مقدر fast consistent and high breakdown (FCH) المقترحة من قبل Olive and Hawkins (2010) و أخيرا يعاد الترجيح مرة أخرى للحصول على المقدر النهائي لهذه الطريقة. لقد وظفنا مقدرات هذه الطريقة لاقتراح خوارزمية جديدة بعنوان DRGP(RMVN)التي يمكن وصفها كالاتي:

1. حساب الوسط الحسابي و مصفوفة التباين و التباين المشترك كمقدرين تبدأ بهما الخوارزمية و لنفرض الإشارة لهما بالرمزين$(T_{0,1}, C_{0,1})$على التوالي.
2. إيجاد مصفوفة التباين والتباين المشتركDGK: يتم حساب مسافة مهالانوبيز (MD) بالاعتماد على المقدرات في الخطوة (1) و من ثم نرتب القيم MD تصاعديا للحصول على وسيط قيمMD كعتبة للحصول على مصفوفة جيدة من المشاهدات تناظر تلك الصفوف التي قيم MD لها اقل من قيمة العتبة. و تكرار الخطوتين السابقتين خمس مرات و في كل مرة تحسب قيم MD لكامل مجموعة البيانات للحصول على $(T_{5,1}, C_{5,1})$ كمقدري موضع وقياس حصينين.
3. إيجاد مقدري الموضع و القياس لمصفوفةMB: خوارزمية هذه المصفوفة تبدأ مع مقدر الوسيط و مصفوفة الوحدة كمصفوفة للتباين و التباين المشترك كمراحل بداية لها. تختلف هذه المصفوفة عن سابقتها انها تستخدم في حساب MD الوسيط و ليس الوسط الحسابي و تعيد نفس الخطوات الخمس السابقة للحصول على $(T_{5,2}, C_{5,2})$.
4. مقدرات FCH: الخطوة الأولى في خوارزمية هذه الطريقة هي حساب المسافة الاقليدية بين مقدري الموقع في الخطوتين السابقتين $(T_{5,1}T_{5,2})$ على التوالي. فاذا كانت قيمة المسافة أقل من آخر قيمة للعتبة (وسيط مسافة مهالانوبيز في الخطوة الخامسة) في مصفوفة MB تبدأ FCH لاختيار أحد هذين المقدرين كالآتي:

$$T_{FCH} = \begin{Bmatrix} T_{5,1} & if & \sqrt{|C_{5,1}|} < \sqrt{|C_{5,2}|} \\ T_{5,2} & Otherwise & \end{Bmatrix} \quad (2\text{-}51)$$

أما مقدر القياس فيتم حسابه أيضا بالاعتماد على الشرط نفسه و لكن المقدر هنا يضرب بثابت او معامل تصحيح كالاتي :

$$\mathbf{C_{FCH}} = \begin{Bmatrix} \frac{MED\left(MD_i^2((T_{5,1},C_{5,1}))\right)}{\chi^2_{(p,0.5)}} \times \boldsymbol{C_{5,1}}, & \boldsymbol{if} & \sqrt{|C_{5,1}|} < \sqrt{|C_{5,2}|} \\ \frac{MED\left(MD_i^2((T_{5,2},C_{5,2}))\right)}{\chi^2_{(p,0.5)}} \times \boldsymbol{C_{5,2}}, & \textbf{Otherwise} & \end{Bmatrix} (2\text{-}52)$$

5. مقدر RFCH: بعد الحصول على المقدرين ($T_{1,RFCH} C_{1,RFCH}$,) من الخطوة السابقة يتم حساب مسافة مهالانوبيز مرة أخرى لكامل المشاهدة و من ثم نكرر الخطوة السابقة لايجاد ($T_{2,RFCH}$) مع مراعاة معامل التصحيح و كالاتي:

$$C_{2,RFCH} = \frac{MED\left(MD_i^2((\ T_{1,RFCH},C_{1,RFCH})\ )\right)}{\chi^2_{(p,0.975)}} \times C_{1,RFCH}\ , \quad (2\text{-}53)$$

6. مقدري الـ RMVN: خوارزمية هذه الطريقة تسعى أولا الى إيجاد مصفوفة مشاهدات جديدة للمتغيرات بالاعتماد على مقدر الموقع و القياس السابقة و كالاتي:

افرض ان X_K حيث أن $S^0 = \sum_{j=1}^{n_1}$

$$X_K = \left\{X_K: MD_i(\ T_{2,RFCH},\ C_{2,RFCH}) \le MED\left(MD_i(\ T_{2,RFCH},\ C_{2,RFCH})\right)\right\} \ (2-54)$$

where $K = 1, \dots, n_1$ and $i = 1, \dots, n$

افرض ان $Q^{(1)} = min\{0.5 \times 0.975 \times n/S^0\ , 0.995\}$ للحصول على مقدر القياس الأول لـ RMVN و كالاتي

$$C_{RMVN}^{(1)} = \frac{MED\left(D_i^2(T_{RFCH},\, C_{RFCH})\right)}{\chi^2_{\left(p,Q^{(1)}\right)}} \qquad (2\text{-}55)$$

و لتحسين أداء هذا المقدر اقترح Olive and Hawkins (2010) اعادة ترجيح مقدر القياس السابق (2-52) مع الجزء الثاني من البيانات n_2 للحصول على المقدر النهائي$C_{RMVN}^{(2)}$.

مما تقدم نصل إلى أن مقدر معلمة الموقع $T_{RMVN} = T_{RFCH}$و أن مقدر معلمة القياس هو $C_{RMVN}^{(2)}$

7. حساب مسافة مهالانوبيز الحصينة

$$\mathrm{RMD_i(RMVN)} = \sqrt{(\mathrm{x} - \mathrm{T_{RMVN}(x)})'\mathrm{C_{RMVN}^{(2)}}^{-1'}(\mathrm{x} - \mathrm{T_{RMVN}(x)})} \quad (2-56)$$

8. تحديد المشاهدات المشكوك بها بالمقارنة مع العتبة ، $RMD_i(RMVN) > \sqrt{\chi^2_{(p,0.95)}}$.

و وضعها في المصفوفة D الموصوفة في طريقة GP،أما بقية المشاهدات فتوضع في المصفوفة R.

9. نحسب P_{ii} كما في المعادلة (48 - 2) لمعرفة المشاهدات في المصفوفة Dهل تحتوي على LP ام لا، فاذا تم تشخيصها تبقى في المصفوفة D و بخلافه تنقل الى المصفوفة R.

بالإضافة إلى ذلك يمكن إضافة الخطوات التي اقترحها (محمد، 2015) لتحسين أداء طريقتنا المقترحة و اقتراح طريقة أخرى تدعى IDRGP(RMVN).

6.2.2 طريقة IDRGP(RMVN) المقترحة

هذه الطريقة لاتختلف عن طريقة IDRGP(MVE) باستثناء إحلال مصفوفة الموضع والقياس RMVN بدلاً من MVE .

7.2.2 طريقة IDRGP (RMVN) based on MM Estimater

(IDRGP(IRMM))

فكرة هذه الطريقة لاتختلف عن فكرة IDRGP(MLMS) باستثناء تشخيص القيم الشاذة لبواقي الانحدار المقدرة بأستخدام طريقة MM Estimator بدلاً من طريقة LMS Estimate

الفصل الثالث
التطبيق العملي والمحاكاة

3.1 دراسة المحاكاة

ليكن لدينا نموذج الانحدار الخطي المتعدد الاتي :-

$$y = x\beta + e$$

حيث أن x هي مصفوفة المتغييرات التوضيحية ذات بعد $n \times p$ المولدة من توزيع طبيعي متعدد المتغيرات بمتوسط قيمته صفراً و تباين $\sigma = \rho^{|i-j|}$ اي أن

$$x \sim N(0, \rho^{|i-j|})$$

حيث أن $p = 15$

و$n = \{50,100,300,500,1000\}$

و $\rho = 0.5$

β هي متجه الوحدة لمعلمات هذا النموذج وكالاتي :-

$$\beta = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}_{15x1}$$

أما e فهو حد الخطأ العشوائي للنموذج بمتوسط قيمته صفر وانحراف معياري قيمته 2 .

ولغرض التأكد من قوة التشخيص للطرق المقارنة قمنا بتلويث البيانات بنسب مختلفة $= \alpha$ $(0.05, 0.10, 0.15)$من القيم الشاذة وLP وحسب الحالات التالية :-

الحالة الاولى :- تلويث مصفوفة المتغيرات التوضيحية بـα من BLP بوجود HLP واحدة لعينة بحجم n مشاهدة. وذلك من خلال ضرب الصفوف الثلاث الأولى من المتغير الثاني إلىالمتغير الخامس بالرقم 10, وضرب أعظم قيمة للمتغيرالأول بالرقم 10 وكذلك ما يناظر هذه القيمة في المتغيرy .

الحالة الثانية :- توليد عينة بحجم n مشاهدة تم تلويثها ب (α) من القيم الشاذة من خلال تلويث قيمة واحدة في المتغيير x وذلك من خلال ضرب أعظم قيمة من قيم المتغيرx ب (10) من خلال تلويث حد الخطأ العشوائي بمشاهدات تتبع توزيع χ^2 بدرجة حرية (10).

الحالة الثالثة :- توليد عينة بحجم n مشاهدة تم تلويثها ب (α) من LP وذلك بالضرب بالرقم (10).

وفي كل حالة من الحالات سعينا الى وضع HLP في البيانات المولدة لمعرفة دورها في التغطية بكشف بعض القيم الشاذة وحجم تأثيرها على هذه الطريقة او تلك من طرق التشخيص او الكشف ووفقاً للحالات التالية :-

ولأجل تحديد الطريقة الأكفأ من الطرق المقارنة في عملية التشخيص الصحيح اعتمدنا على المعايير الآتية :-

1. الطريقة الأقرب إلى التشخيص الصحيح والأقل قيمة في التشخيص غير الصحيح والأسرع زمناً تكون هي الطريقة ألأكثر كفاءة من نظيرتها.
2. ألطريقة ألأكثر كفاءة من كافة الطرق المقارنة هي الطريقة التي تتمتع بمعايير المقارنة المذكورة في (1)أو التي تتمتع بمعياري أفضلية من غيرها

جدول (1) يعرض معدلات التشخيص الصحيح والتشخيص غير الصحيح وزمن الاحتساب في LP للطرق المقارنة عندما n=50,100,300,500,1000 ونسبة تلويث المتغيرات التوضيحية 0.05 مع وجود HLP واحدة.

Case	Table(1)							
	$\alpha = 0.05$							
	n	Diagnostics	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(IMLMS)	IDRGP(IRMM)
X	50	Correct	**3.891**	**3.918**	**3.904**	**3.912**	**3.922**	**3.909**
		Swap	**2.084**	**2.413**	**3.367**	**1.710**	**3.507**	**2.692**
		Time	**0.101**	**0.077**	**0.491**	**0.046**	**0.018**	**0.118**
	100	Diagnostics	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(IMLMS)	IDRGP(IRMM)
		Correct	**5.918**	**5.919**	**5.912**	**5.912**	**5.915**	**5.915**
		Swap	**1.305**	**1.866**	**2.741**	**1.189**	**1.749**	**2.592**
		Time	**0.178**	**0.113**	**0.748**	**0.091**	**0.017**	**0.227**
	300	Diagnostics	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(IMLMS)	IDRGP(IRMM)
		Correct	**15.902**	**15.904**	**15.900**	**15.913**	**15.903**	**15.897**
		Swap	**2.297**	**2.199**	**4.562**	**2.319**	**2.237**	**4.552**
		Time	**0.527**	**0.263**	**1.956**	**0.310**	**0.023**	**0.736**
	500	Diagnostics	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(IMLMS)	IDRGP(IRMM)
		Correct	**25.7**	**25.7**	**25.7**	**25.7**	**25.7**	**25.7**
		Swap	**3.400**	**2.800**	**6.400**	**3.100**	**2.900**	**7.200**
		Time	**0.962**	**0.429**	**3.848**	**0.617**	**0.032**	**1.427**
	1000	Diagnostics	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(IMLMS)	IDRGP(IRMM)
		Correct	**50.9**	**50.9**	**50.9**	**50.9**	**50.9**	**50.9**
		Swap	**7.600**	**7.400**	**13.800**	**7.700**	**7.500**	**11.6**
		Time	**2.382**	**0.827**	**8.035**	**1.701**	**0.065**	**3.793**

في هذا الجدول نسبة التلويث كانت (0.05) في المتغيرات التوضيحية واحجام العينات كانت (50, 100 , 300, 500, 1000)

نلاحظ في هذا الجدول وعند حجم عينة 50 وجود فروقات في معدلات التشخيص الصحيح للطرق كافة و أن هناك اختلافاً في معدلات تشخيص بعض القيم النظيفة التي اعتبرت قيماً شاذة

(swap), إذ أن معدل التشخيص غير الصحيح بلغ (3.507) لطريقة IDRGP(MLMS) وهو الأعلى بين الطرق المقارنة و (1.710) لطريقة IDRGP(RMVN)الذي يعد الاقل من بين تلك الطرق, كما نلاحظ أن هذه الطريقة أيضا هي الأسرع من ناحية زمن الاحتساب.

عند مقارنة DRGP(MVE) مع DRGP(RMVN) نلاحظ أن طريقتنا المقترحة كانت الأقرب للتشخيص الصحيح ولكنها شخصت بعض القيم النظيفة على أنها شاذة ولو بفارق بسيط على الطريقة التقليدية مع ذلك فإنها كانت أسرع منها من ناحية زمن الاحتساب,أما عند مقارنة IDRGP(MVE) مع IDRGP(RMVN) من ناحية معدلات التشخيص الصحيح فإنها لا تختلف عن التفسير السابق إلا أن طريقتنا المقترحة اختزلت (swap) بفارق واضح عن طريقة IDRGP(MVE),وأنها أسرع بكثير منها.الطريقة المقترحة الثالثة IDRGP(RMM) كانت قريبة جدا من التشخيص الصحيح وأنها اختزلت بشكل جيد معدل التشخيص الخاطئ و كانت أبطأ من طريقة IDRGP(MLMS) من ناحية السرعة في زمن الاحتساب .

وعند حجم عينة 100الملاحظ أن معدلات التشخيص الصحيح لكافة الطرق متقاربة جداً ووجود اختلاف في معدلات تشخيص بعض القيم النظيفة التي اعتبرت قيماً شاذة (swap), إذ أن معدل التشخيص الغير صحيح بلغ (2.741) لطريقة IDRGP(MVE) وهو الأعلى بين الطرق المقارنة و (1.189) لطريقة IDRGP(RMVN) الذي يعد الأقل من بين تلك الطرق, كما نلاحظ طريقة IDRGP(MLMS)أيضا هي الأسرع من ناحية زمن الاحتساب.

عند مقارنة DRGP(MVE) مع DRGP(RMVN) نلاحظ أن طريقتنا المقترحة كانت الأقرب ولو بفارق بسيط للتشخيص الصحيح ولكنها لم تشخص القيم النظيفة على أنها شاذة مقارنة مع الطريقة التقليدية بفارق واضح ولكنها كانت ألأسرع عند المقارنة بالنسبة لزمن الاحتساب, أما عند مقارنة IDRGP(MVE) مع IDRGP(RMVN) من ناحية معدلات التشخيص الصحيح فأنها لاتختلف عن بعضها ألا أن طريقتنا المقترحة اختزلت (swap) بفارق واضح عن طريقة IDRGP(MVE), وأنها أسرع بكثير منها.الطريقة المقترحة الثالثة IDRGP(RMM) كانت متشابهه في التشخيص الصحيح و معدل التشخيص الخاطئ لم يسجل تقليص في النسبة كان متقاربا وكذلك هي أبطأ من طريقة IDRGP(MLMS).

وعند ارتفاع حجم العينة الى 300 نجد أن معدلات التشخيص الصحيح لكافة الطرق متقاربة ألا أن هناك اختلاف في معدلات تشخيص بعض القيم النظيفة التي اعتبرت قيماً شاذة (swap) , إذ أن معدل التشخيص الغير صحيح بلغ (4.562) لطريقة IDRGP(MVE) وهو الأعلى بين الطرق المقارنة و (2.199) لطريقة DRGP(RMVN)الذي يعد الأقل من بين تلك الطرق , كما نلاحظ طريقة IDRGP(MLMS)أيضا هي الأسرع من ناحية زمن الاحتساب.

عند مقارنة DRGP(MVE) مع DRGP(RMVN) نلاحظ أن الطريقة المقترحة رفعت من معدل نسبة التشخيص الصحيح بنسبة بسيطة أما بالنسبة لتشخيص القيم النظيفة على أنها شاذة فأنها قللت من معدل (swap) بنسبة بسيطة مقارنة مع الطريقة التقليدية أما زمن الاحتساب كان أسرع منها بفارق بسيط أيضاً . أما عند مقارنة IDRGP(MVE) مع IDRGP(RMVN) من

ناحية معدلات التشخيص الصحيح فأنها كانت متقاربة وطريقتنا المقترحة أختزلت (swap) بفارق واضح جداً عن طريقة IDRGP(MVE) وكذلك هي سجلت سرعة واضحة في زمن الاحتساب , الطريقة المقترحة الثالثة IDRGP(RMM)لم تكن لها الافضلية في التشخيص الصحيح و معدل التشخيص الخاطئ لم يسجل تقليصاً في النسبة بل كان أعلى وكذلك هي أبطأ من ناحية زمن الاحتساب مقارنة مع طريقة IDRGP(MLMS)

ونستمر بأرتفاع حجم العينة الى 500 كان عرض معدلات التشخيص الصحيح لكافة الطرق متماثلة ألا أن هناك اختلافاً في معدلات تشخيص بعض القيم النظيفة التي اعتبرت قيماً شاذة (swap) , إذ أن معدل التشخيص الغير صحيح بلغ (7.200)لطريقة IDRGP(RMM) وهو الأعلى بين الطرق المقارنة و(2.800) لطريقة DRGP(RMVN)الذي يعد الأقل من بين تلك الطرق , كما نلاحظ طريقة IDRGP(MLMS)أيضا هي الأسرع من ناحية زمن الاحتساب.

عند مقارنة DRGP(MVE) مع DRGP(RMVN) نلاحظ أن الطريقتين متماثلتين أما بالنسبة لتشخيص القيم النظيفة على أنها شاذة فأنها قللت من معدل (swap) بنسبة واضحة مقارنة مع الطريقة التقليدية, أما زمن الاحتساب فكان أسرع منها بفارق واضح أيضاً , أما عند مقارنة IDRGP(MVE) مع IDRGP(RMVN) من ناحية معدلات التشخيص الصحيح فأنها كانت متماثلة أيضاً وطريقتنا المقترحة أختزلت (swap) بفارق واضح جداً عن طريقة IDRGP(MVE) وكذلك هي سجلت سرعة واضحة في زمن الاحتساب , الطريقة المقترحة الثالثة IDRGP(RMM) كانت متماثلة في التشخيص الصحيح و معدل التشخيص الخاطئ لم يسجل تقليص في النسبة بل كان أعلى وكذلك هي أبطأ من ناحية زمن الاحتساب مقارنة مع طريقة IDRGP(MLMS).

وعند وصول حجم العينة الى 1000 نلاحظ ان معدل التشخيص الصحيح لكافة الطرق متماثلة الا ان اختزالـ (swap) كانت الأفضلية لطريقتي DRGP(RMVN)اذ بلغت (7.400) و زمن الاحتساب الأسرع لصالح طريقة IDRGP(MLMS)أيضاً .

نلاحظ عند المقارنة بين طريقتي DRGP(MVE) و DRGP(RMVN) نجد أن معدل التشخيص الصحيح متماثل ونجد أن طريقة DRGP(RMVN) أكثر اختزالا من ناحية ال swapو أن السرعة في زمن الاحتساب كانت أيضاً لصالح طريقة DRGP(RMVN) حيث بلغت (0.827) مقابل (2.382) لصالح الطريقة التقليدية , أما المقارنة الثانية بين الطريقة المقترحة الثانية والطريقة التقليدية فنجد أن التماثل مستمر من ناحية التشخيص الصحيح ولكن بينت طريقتنا المقترحة اختزالاً عالي في معدل الـ (swap) وكذلك سجلت تفوق في زمن الاحتساب , في المقارنة الثالثة بين طريقتي IDRGP(MLMS) و IDRGP(RMM) حيث سجلت تماثل بين الطريقتين في التشخيص الصحيح ولكن أختزل ال swap لم يكن لها حيث سجل معدل (11.6) مقابل (7.500) أما بالنسبة للسرعة في زمن الاحتساب كانت لصالح طريقة IDRGP(MLMS) بفارق واضح حيث سجلت الاخيرة معدل (0.065) مقابل (3.793) .

جدول (2) يعرض معدلات التشخيص الصحيح والتشخيص غير الصحيح وزمن الاحتساب في outlierse للطرق المقارنة عندما n=50,100,300,500,1000 ونسبة تلويث المتغيرات التوضيحية 0.05 مع وجود HLP واحدة.

case	Table(2)							
	$\alpha = 0.05$							
	n	dignostics	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(IMLMS)	IDRGP(IRMM)
Y	50	correct	**3.806**	**3.859**	**3.859**	**3.909**	**3.914**	**3.933**
		Swap	**2.160**	**2.641**	**3.403**	**1.639**	**3.634**	**2.599**
		Time	**0.103**	**0.078**	**0.503**	**0.047**	**0.018**	**0.121**
	100	dignostics	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(IMLMS)	IDRGP(IRMM)
		correct	**5.923**	**5.935**	**5.942**	**5.938**	**5.936**	**5.950**
		Swap	**1.297**	**2.223**	**4.088**	**1.180**	**2.062**	**2.808**
		Time	**0.182**	**0.115**	**0.764**	**0.094**	**0.018**	**0.234**
	300	dignostics	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(IMLMS)	IDRGP(IRMM)
		correct	**15.926**	**15.927**	**15.925**	**15.946**	**15.925**	**15.949**
		Swap	**2.370**	**2.336**	**11.781**	**2.297**	**2.407**	**5.383**
		Time	**0.532**	**0.261**	**1.959**	**0.316**	**0.023**	**0.750**
	500	dignostics	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(IMLMS)	IDRGP(IRMM)
		correct	**25.75**	**25.75**	**25.75**	**25.85**	**25.75**	**25.85**
		Swap	**3.150**	**2.950**	**15.850**	**3.150**	**3.001**	**7.300**
		Time	**0.962**	**0.412**	**3.422**	**0.614**	**0.036**	**1.407**
	1000	dignostics	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(IMLMS)	IDRGP(IRMM)
		correct	**50.85**	**50.85**	**50.85**	**50.85**	**50.85**	**51.01**
		Swap	**7.100**	**6.600**	**36.750**	**6.950**	**6.700**	**14.700**
		Time	**2.430**	**0.821**	**8.019**	**1.714**	**0.065**	**3.784**

يتضمن هذا الجدول نتائج دراسة المحاكاة ل (1000)عينة بأحجام مختلفة لوثت كل واحدة منها بنسبة (0.05) من القيم الشاذة في المتغير المعتمد y بوجود HLP واحدة نتائج المحاكاة في هذا الجدول عند حجم عينة 50 هي متقاربة من ناحية التشخيص الصحيح في كافة الطرق ألا أنها تظهر بشكل واضح جداً اختزال الـ (swap) في طريقة IDRGP(RMVN) وبواقع (1.639) متفوقة على باقي الطرق الأخرى كما أنها أظهرت سرعة أيضاً في زمن الاحتساب عن باقي الطرق ماعدا طريقة IDRGP(MLMS) تفوقت عليها بفارق بسيط جدا.

عند مقارنة الطرق بشكل ثنائي نجد أن طريقة DRGP(RMVN)كانت أفضل من ناحية معدل التشخيص الصحيح افضل مقارنة مع الطريقة التقليدية حيث كانت أكثر دقة وإن معدل الـ (swap) كان أعلى بقليل من طريقة DRGP(MVE) وكانت أكثر سرعه منها في زمن الاحتساب. عند مقارنة طريقتي IDRGP(RMVN) وIDRGP(MVE) نلاحظ أرتفاع في معدل التشخيص الصحيح وأختزال في حساب الـ (swap) وسرعة في زمن الاحتساب لصالح طريقة IDRGP(RMVN) .أظهرت المقارنة الأخيرة في هذا الجدول أيضا ارتفاع في نسبة احتساب معدل التشخيص الصحيح واختزال في قيمة الـ (swap) لصالح طريقة IDRGP(RMM) ولكن على حساب زمن الاحتساب حيث أظهر سرعه لصالح طريقة IDRGP(MLMS) .

وعند حجم 100 من النتائج المعروضة في الجدول وبمقارنة كل طريقة مقترحة مع نظيرتها من الطرق السابقة نلاحظ وجود تفوق في طريقة DRGP(RMVN) على طريقةDRGP(MVE)من حيث التشخيص الصحيح للقيم الشاذة وتفوق الطريقة التقليدية في أختزال الـ (swap)ألا أن طريقتنا المقترحة كانت أسرع من ناحية زمن الاحتساب. في المقارنة الثانية تتراجع طريقتنا المقترحة قليلاً في التشخيص الصحيح عن الطريقة التقليدية أما طريقة IDRGP(RMVN) أختزلت الـ (swap) بفارق واضح عن IDRGP(MVE) أذ بلغ حجم الاختزال الى (1.180) مقابل (4.088) للطريقة الاخرى وأن هذا الفارق كان واضح جداً من ناحية زمن الاحتساب أذ بلغت سرعة IDRGP(RMVN) (0.094) بينما بلغت (0.764) للطريقة التقليدية .أما الحالة الأخيرة فقد كانت متقاربة في التشخيص الصحيح مع تفوق بسيط جداً لطريقة IDRGP(RMM) ألا أنها خسرت هذا التفوق في حساب نسبة الـ swap لصالح IDRGP(MLMS) التي حققت (2.062) وكذلك هذه الأخيرة كانت أسرع من ناحية زمن الاحتساب. من الملاحظ أن طريقة IDRGP(RMVN) حققت نتائج أفضل من غيرها من حيث التشخيص والاختزال ألا أنها كانت أبطأ قليلاً من IDRGP(MLMS)

ننتقل الى عينة أخرى تبلغ (300) ونبدأ كما هو الحال لباقي العينات حسب نسبة التلويث التي بلغت (0.05) وعند عرض نتائج المحاكاة لهذا الجدول نلاحظ التشخيص الصحيح كان لصالح طريقة IDRGP(RMM)وتسجيل تفوق في معدل التشخيص الخاطئ لصالح طريقة IDRGP(RMVN) وكذلك أيضا تسجل سرعة في زمن الاحتساب .

الآن نقارن بين طريقتي DRGP(MVE) وطريقة DRGP(RMVN) وعند حجم عينة 300 حيث نجد ارتفاع في معدل التشخيص الصحيح لصالح طريقتنا ألمقترحه بلغ (15.927) مقابل (15.926) ولو بفارق بسيط وسجلت انخفاض في معدل الـ (swap) أيضاً بفارق بسيط بلغ (2.336) مقابل (2.370) وسجلت أيضاً سرعة في زمن الاحتساب بلغ (0.261) مقابل (0.532) ,

الان ننتقل للمقارنة بين طريقتي IDRGP(MVE) وطريقتنا المقترحة IDRGP(RMVN) حيث نلاحظ تفوق الطريقة المقترحة حيث بلغ معدل التشخيص الصحيح لها (15.946) مقابل (15.925) وبلغ معدل الـ (swap) لها (2.297) مقابل (11.781) وبلغ سرعة زمن الاحتساب (0.316) مقابل (1.959) , وعند المقارنة الثالثة بين الطريقة المقترحة والتقليدية نجد أرتفاع في معدل التشخيص الصحيح لصالح طريقتنا المقترحةIDRGP(RMM) بلغ (15.949) مقابل معدل (15.925) لطريقة IDRGP(MLMS) ولكن في معدل ألـ (swap) بلغت طريقتنا (5.383) مقابل معدل (2.407) لصالح الطريقة التقليدية وهو تفوق واضح لها وكذلك الحال في سرعة زمن الاحتساب حيث سجلت طريقتنا (0.750) مقابل معدل (0.023) لصالح الطريقة التقليدية .

وكذلك أيضا استمرار لباقي الحالات وعند زيادة حجم العينة الى 500 وزيادة نسبة التلويث تظهر كافة الطرق تماثل في معدل التشخيص الصحيح وفي طريقته DRGP(RMVN) تسجيل

انخفاض واضح في معدل ألـ(swap) أما السرعة في زمن الاحتساب كانت لطريقة IDRGP(MLMS).

بينما في المقارنة الثنائية هناك تماثل واضح في التشخيص العام ففي المقارنة بين طريقتي DRGP(MVE) و DRGP(RMVN) نجد تماثل في معدل التشخيص الصحيح أما معدل التشخيص الخاطئ سجل في طريقتنا معدل (2.95)مقابل (3.15) لصالح الطريقة المقترحة وزمن الاحتساب كان الأسرع في الطريقة المقترحة حيث بلغ (0.412) مقابل (0.962) , في المقارنة الثانية بين طريقتي IDRGP(MVE) و طريقة IDRGP(RMVN) كان التفوق لصالح طريقتنا ماعدا معدل التشخيص الصحيح كان لصالح الطريقة التقليدية حيث بلغ في الطريقة المقترحة (25.85) مقابل (25.75) بينما الـ (swap) بلغ في الطريقة المقترحة معدل (3.150) مقابل (15.850) وهو اختلاف واضح وكبير جداً والسرعة في زمن الاحتساب كذلك أيضاً حيث سجل معدل (0.614) مقابل (3.422) , في المقارنة الثالثة بين طريقتي IDRGP(MLMS) وطريقتنا المقترحة IDRGP(RMM) كذلك أيضا كباقي الجداول الأخرى كان التفوق فقط في التشخيص الصحيح لصالح طريقتنا حيث بلغ معدل (25.85) مقابل (25.75) وكان معدل الـ (swap) متفوق لصالح الطريقة التقليدية حيث سجل معدل (3.001) في الطريقة التقليدية مقابل معدل (7.300) والسرعة في زمن الاحتساب كانت ايضاً لصالح الطريقة التقليدية حيث بلغت معدل (0.036) مقابل معدل (1.407) لصالح الطريقة المقترحة.

الملاحظ في هذا الجدول عند حجم عينة 1000 ان معدل التشخيص الصحيح لكافة الطرق متماثلة لكن أعلى نسبة كانت لصالح طريقة IDRGP(RMM) إلا أن اختزالـ (swap) كانت الأفضلية لطريقتي IDRGP(RMVN) اذ بلغت (6.600) وزمن الاحتساب الأسرع لصالح طريقة IDRGP(MLMS) أيضاً

إذا نلاحظ عند المقارنة بين طريقتي DRGP(MVE) و DRGP(RMVN) نجد أن معدل التشخيص الصحيح متماثل ونجد أن طريقة DRGP(RMVN) أكثر اختزالا من ناحية ال swapو أن السرعة في زمن الاحتساب كانت أيضاً لصالح طريقة DRGP(RMVN) حيث بلغت (0.821) مقابل (2.430) لصالح الطريقة التقليدية , أما المقارنة الثانية بين الطريقة المقترحة الثانية والطريقة التقليدية فنجد أن التماثل مستمر من ناحية التشخيص الصحيح ولكن بينت طريقتنا المقترحة اختزال عالي في معدل ال (swap) وكذلك سجلت تفوق في زمن الاحتساب , في المقارنة الثالثة بين طريقتي IDRGP(MLMS) و IDRGP(RMM) حيث سجلت الطريقة المقترحة أفضلية في التشخيص الصحيح ولكن أختزل ال swap لم يكن لها حيث سجل معدل (14.700) مقابل (6.700) أما بالنسبة للسرعة في زمن الاحتساب كانت لصالح طريقة IDRGP(MLMS) بفارق واضح حيث سجلت الأخيرة معدل (0.065) مقابل (3.784) .

جدول (3) يعرض معدلات التشخيص الصحيح والتشخيص غير الصحيح وزمن الاحتساب في LP و outlierse للطرق المقارنة عندما n=50,100,300,500,1000 ونسبة تلويث المتغيرات التوضيحية 0.05 مع وجود HLP واحدة.

Case	Table(3)							
	$\alpha = 0.05$							
	n	dignostics	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(IMLMS)	IDRGP(IRMM)
X & Y	50	correct	**3.941**	**3.941**	**3.942**	**3.942**	**3.942**	**3.942**
		Swap	**1.986**	**2.691**	**2.429**	**1.813**	**3.593**	**2.422**
		Time	**0.101**	**0.077**	**0.494**	**0.045**	**0.018**	**0.118**
	100	dignostics	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(IMLMS)	IDRGP(IRMM)
		correct	**5.956**	**5.956**	**5.956**	**5.956**	**5.956**	**5.956**
		Swap	**1.333**	**1.879**	**2.17**	**1.194**	**1.657**	**2.417**
		Time	**0.179**	**0.113**	**0.753**	**0.091**	**0.018**	**0.228**
	300	dignostics	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(IMLMS)	IDRGP(IRMM)
		correct	**15.958**	**15.958**	**15.957**	**15.958**	**15.958**	**15.958**
		Swap	**2.320**	**2.248**	**6.487**	**2.236**	**2.333**	**3.985**
		Time	**0.527**	**0.261**	**1.955**	**0.311**	**0.023**	**0.738**
	500	dignostics	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(IMLMS)	IDRGP(IRMM)
		correct	**26**	**26**	**26**	**26**	**26**	**26**
		Swap	**4.100**	**3.600**	**11.250**	**4.050**	**3.950**	**6.600**
		Time	**0.933**	**0.406**	**3.355**	**0.587**	**0.032**	**1.354**
	1000	dignostics	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(IMLMS)	IDRGP(IRMM)
		correct	**50.9**	**50.9**	**50.9**	**50.9**	**50.9**	**50.9**
		Swap	**7.600**	**7.200**	**21.700**	**7.400**	**7.200**	**9.700**
		Time	**2.337**	**0.806**	**7.865**	**1.654**	**0.057**	**3.673**

في هذا الجدول الذي يدرس الحالة في الـ LP و outlierse معاً نجد ان معدل التشخيص الصحيح لكافة الطرق متماثلة تقريباً الا ان اختزالـ (swap) كانت الافضلية لطريقة IDRGP(RMVN) اذ بلغت (1.813) و بزمن احتساب أسرع. لم تختلف بقية النتائج على الحالات المماثلة لنسبة التلويث (5%) و حجم العينة (50) مشاهدة اذا نلاحظ عند المقارنة بين طريقتي DRGP(MVE) و DRGP(RMVN) نجد أن معدل التشخيص الصحيح متماثل ونجد أن طريقة DRGP(MVE) أكثر اختزالا من ناحية الswap ألا أن السرعة في زمن الاحتساب كانت لصالح طريقة DRGP(RMVN) حيث بلغت (0.077) مقابل (0.101) لصالح الطريقة التقليدية , أما المقارنة الثانية بين الطريقة المقترحة الثانية والطريقة التقليدية فنجد أن التماثل مستمر من ناحية التشخيص الصحيح ولكن بينت طريقتنا المقترحة اختزال عالي في معدل ال (swap) وكذلك سجلت تفوق في زمن الاحتساب , في المقارنة الثالثة بين طريقتي IDRGP(MLMS) و IDRGP(RMM) حيث استمرار التماثل في التشخيص الصحيح ولكن أختزل ال swap بمعدل (2.422) مقابل (3.593) أما بالنسبة للسرعة في زمن الاحتساب كانت لصالح طريقة IDRGP(MLMS) بفارق بسيط حيث سجلت الأخيرة معدل (0.018) مقابل (0.118) .

ننتقل إلى عينة جديدة تبلغ (100) ونبدأ بأول نسبة تلويث البالغة (0.05) الملاحظ في هذا الجدول تماثل في نسبة التشخيص الصحيح في كافة الطرق وكالعادة أفضل اختزال في معدل ألـ (swap) كان في طريقة IDRGP(RMVN) المقترحة حيث بلغ (1.194) وأيضا السرعة في زمن الاحتساب لصالح طريقة IDRGP(MLMS) التقليدية .

في المقارنة الثنائية الاولى كما قلنا يوجد تماثل في التشخيص الصحيح أما التشخيص الخاطئ كان لصالح الطريقة التقليدية بواقع معدل (1.333) مقابل (1.879) بينما السرعة في زمن الاحتساب كانت لصالح الطريقة المقترحة بواقع (0.113) مقابل (0.179) , وعند النظر في المقارنة الثانية كما قلنا أن التماثل في التشخيص الصحيح وننتقل إلى المقارنة في التشخيص الخاطئ حيث سجلت طريقتنا المقترحة اختزال عالي فيه بواقع (1.194) مقابل (2.170) وأيضا السرعة في زمن الاحتساب كان لصالح الطريقة المقترحة حيث سجل معدل (0.091) مقابل (0.753) , وفي المقارنة الثالثه يستمر التماثل في التشخيص الصحيح وتأتي الطريقة التقليدية لتسجل معدل أفضل في أختزال التشخيص الخاطئ حيث سجل(1.657) مقابل (2.417) للطريقة المقترحة وأيضاً الطريقة التقليدية تسجل سرعة في زمن الاحتساب متفوقة على الطريقة المقترحة حيث سجلت معدل (0.018) مقابل (0.228) .

هنا يستمر التماثل في كافة الطرق عند حجم عينة 300 حيث بلغ معدل (15.958) في التشخيص الصحيح ماعدا طريقة IDRGP(MVE) حيث أنخفض فيها التشخيص الصحيح بفارق بسيط بمعدل (15.957) وكان أفضل تشخيص خاطئ لصالح طريقة IDRGP(RMVN) بمعدل (2.236) وكانت اسرع طريقة هي طريقة IDRGP(MLMS)بمعدل (0.023).

في المقارنة الثنائية بين طريقتي DRGP(MVE) و DRGP(RMVN) حيث ان التماثل موجود في التشخيص الصحيح وان التشخيص الخاطئ سجل اختزالا لصالح طريقة DRGP(RMVN) بمعدل (2.248) مقابل معدل (2.320) في طريقة DRGP(MVE) وأيضا السرعة في زمن الاحتساب سجلت تفوق لصالح الطريقة المقترحة , ننتقل الى المقارنة الثانية بين طريقتي IDRGP(MVE)IDRGP(RMVN), نلاحظ ارتفاع في معدل التشخيص الصحيح لصالح الطريقة المقترحة بمعدل (15.958) على الطريقة التقليدية التي سجلت معدل (15.957) وفي التشخيص الخاطئ نلاحظ وجود تفوق كبير في الطريقة المقترحة حيث اختزلت إلى معدل (2.236) مقابل (6.487) في الطريقة التقليدية وكذلك أيضاً السرعة في زمن الاحتساب كانت لصالح الطريقة المقترحة حيث بلغت معدل (0.311) مقابل (1.955) , في المقارنة ألثالثه بين طريقتي IDRGP(MLMS) و IDRGP(RMM) نجد استمرار التماثل في معدل التشخيص الخاطئ بينما الـ (swap) والسرعة في زمن الاحتساب سجلت تفوق لصالح طريقة IDRGP(MLMS) بمعدلات (2.333) و (0.023) مقابل (3.985) و (0.738) على التوالي .

هنا يستمر التماثل في كافة الطرق عند حجم عينة 500 حيث بلغ معدل (26) في التشخيص الصحيح وكان أفضل تشخيص خاطئ لصالح طريقة DRGP(RMVN) بمعدل (3.600) وكانت أسرع طريقة هي طريقة IDRGP(MLMS)بمعدل (0.032).

في المقارنة الثنائية بين طريقتي DRGP(MVE) و DRGP(RMVN) حيث ان التماثل موجود في التشخيص الصحيح وان التشخيص الخاطئ سجل اختزالا لصالح طريقة DRGP(RMVN) بمعدل (3.600) مقابل معدل (4.100) في طريقة DRGP(MVE) وأيضا السرعة في زمن الاحتساب سجلت تفوق لصالح الطريقة المقترحة , ننتقل الى المقارنة الثانية بين طريقتي IDRGP(MVE)IDRGP(RMVN), نلاحظ التماثل مستمر في معدل التشخيص الصحيح وفي التشخيص الخاطئ نلاحظ وجود تفوق كبير في الطريقة المقترحة حيث اختزلت الى معدل (4.050) مقابل (11.250) في الطريقة التقليدية وكذلك أيضاً السرعة في زمن الاحتساب كانت لصالح الطريقة المقترحة حيث بلغت معدل (0.587) مقابل (3.355) , في المقارنة الثالثة بين طريقتي IDRGP(MLMS) و IDRGP(RMM) نجد استمرار التماثل في معدل التشخيص الخاطئ بينما الـ (swap) والسرعة في زمن الاحتساب سجلت تفوق لصالح طريقة IDRGP(MLMS) بمعدلات (3.950) و (0.032) مقابل (6.600) و (1.354) على التوالي .

هنا ايضاً يستمر التماثل في التشخيص الصحيح في كافة الطرق عند حجم عينة 1000 حيث بلغ معدل (50.9) في التشخيص الصحيح وكان افضل تشخيص خاطئ لصالح طريقة DRGP(RMVN) بمعدل (7.200) وكانت أسرع طريقة هي طريقة IDRGP(MLMS)بمعدل (0.075).

في المقارنة الثنائية بين طريقتي DRGP(MVE) و DRGP(RMVN) حيث أن التماثل موجود في التشخيص الصحيح وان التشخيص الخاطئ سجل اختزالا لصالح طريقة DRGP(RMVN) بمعدل (7.200) مقابل معدل (7.600) في طريقة DRGP(MVE) وأيضا السرعة في زمن الاحتساب سجلت تفوق لصالح الطريقة المقترحة , ننتقل الى المقارنة الثانية بين طريقتي IDRGP(MVE)و,IDRGP(RMVN) نلاحظ وجود تماثل في معدل التشخيص الصحيح وفي التشخيص الخاطئ نلاحظ وجود تفوق كبير في الطريقة المقترحة حيث اختزلت إلى معدل (7.400) مقابل (21.700) في الطريقة التقليدية وكذلك أيضاً السرعة في زمن الاحتساب كانت لصالح الطريقة المقترحة حيث بلغت معدل (1.654) مقابل (7.865) , في المقارنة ألثالثه بين طريقتي IDRGP(MLMS) و IDRGP(RMM) نجد استمرار التماثل في معدل التشخيص الصحيح بينما الـ (swap) والسرعة في زمن الاحتساب سجلت تفوق لصالح طريقة IDRGP(MLMS) بمعدلات (7.200) و (0.075) مقابل (9.700) و (3.673) على التوالي .

جدول (4) يعرض معدلات التشخيص الصحيح والتشخيص غير الصحيح وزمن الاحتساب في LP للطرق المقارنة عندما n=50,100,300,500,1000 ونسبة تلويث المتغيرات التوضيحية 0.10 مع وجود HLP واحدة.

case	Table(4)							
	$\alpha = 0.10$							
	n	dignostics	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(IMLMS)	IDRGP(IRMM)
x	50	correct	**5.778**	**5.829**	**5.821**	**5.827**	**5.836**	**5.824**
		Swap	**1.576**	**1.560**	**2.853**	**1.226**	**2.609**	**2.307**
		Time	**0.100**	**0.075**	**0.488**	**0.045**	**0.017**	**0.117**
	100	dignostics	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(IMLMS)	IDRGP(IRMM)
		correct	**10.851**	**10.854**	**10.844**	**10.850**	**10.851**	**10.842**
		Swap	**0.770**	**1.119**	**1.743**	**0.718**	**0.962**	**1.936**
		Time	**0.176**	**0.113**	**0.742**	**0.090**	**0.018**	**0.224**
	300	dignostics	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(IMLMS)	IDRGP(IRMM)
		correct	**30.753**	**30.755**	**30.726**	**30.756**	**30.754**	**30.721**
		Swap	**1.229**	**1.158**	**2.399**	**1.235**	**1.237**	**2.672**
		Time	**0.524**	**0.265**	**1.955**	**0.304**	**0.022**	**0.726**
	500	dignostics	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(IMLMS)	IDRGP(IRMM)
		correct	**50.9**	**50.9**	**50.9**	**50.9**	**50.9**	**50.8**
		Swap	**1.800**	**1.700**	**3.300**	**1.800**	**1.800**	**3.400**
		Time	**1.038**	**0.437**	**3.640**	**0.661**	**0.031**	**1.444**
	1000	dignostics	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(IMLMS)	IDRGP(IRMM)
		correct	**100.1**	**100.2**	**99.9**	**100.1**	**100.2**	**99.9**
		Swap	**2.700**	**2.600**	**3.800**	**2.700**	**2.700**	**5.400**
		Time	**2.317**	**0.822**	**7.862**	**1.611**	**0.062**	**3.582**

الملاحظ في هذا الجدول عند حجم عينة 50 أن معدلات التشخيص الصحيح للطرق كافة متقاربة الا أن هناك اختلافاً في معدلات تشخيص بعض القيم النظيفة التي اعتبرت قيماً شاذة (swap),إذإن معدل التشخيص غير الصحيح بلغ (2.853) لطريقة IDRGP(MVE) وهو الأعلى بين الطرق المقارنة و (1.226) لطريقة IDRGP(RMVN) الذي يعد الأقل من بين تلك الطرق ,كذلك هذه الطريقة هي سريعة من ناحية زمن الاحتساب مقارنة مع IDRGP(MVE) .

عند مقارنة DRGP(MVE) مع DRGP(RMVN) نلاحظ أن طريقتنا المقترحة زادت من معدل التشخيص الصحيح بفارق واضح ولكنها شخصت بعض القيم النظيفة على أنها شاذة بنسبة اقل عن الطريقة التقليدية وكذلك كانت أسرع منها من ناحية زمن الاحتساب, أما عند مقارنة IDRGP(MVE) مع IDRGP(RMVN) من ناحية معدلات التشخيص الصحيح فأنها لاتختلف عن التفسير السابق كثيراً إلا أن طريقتنا المقترحة اختزلت الـ (swap) بفارق واضح عن طريقة IDRGP(MVE), وأنها أسرع بكثير منها.الطريقة المقترحة الثالثة IDRGP(RMM) كانت قريبة من التشخيص الصحيح مع وجود فارق بسيط وأنها اختزلت بشكل جيد معدل التشخيص الخاطئ ولكنها كانت أبطأ من طريقة IDRGP(MLMS)

عند حجم عينة 100 الملاحظ أيضاً في هذا الجدول أن معدلات التشخيص الصحيح للطرق كافة متقاربة جداً إلا أن هناك اختلاف في معدلات تشخيص بعض القيم النظيفة التي اعتبرت قيماً شاذة (swap),إذ أن معدل التشخيص غير الصحيح بلغ (1.936) لطريقة IDRGP(RMM) وهو الأعلى بين الطرق المقارنة و (0.718) لطريقة IDRGP(RMVN) الذي يعد الأقل من بين تلك الطرق , كما نلاحظ طريقة IDRGP(MLMS)أيضا هي الأسرع من ناحية زمن الاحتساب.

عند مقارنة DRGP(MVE) مع DRGP(RMVN) نلاحظ أن طريقتنا المقترحة كانت الأقرب للتشخيص الصحيح ولكنها شخصت بعض القيم النظيفة على أنها شاذة مقارنة مع الطريقة التقليدية بفارق واضح بنسبة أعلى وكذلك الحال بالنسبة لزمن الاحتساب كان أسرع منها بفارق بسيط , أما عند مقارنة IDRGP(MVE) مع IDRGP(RMVN) من ناحية معدلات التشخيص الصحيح فأنها لاتختلف عن بعضها ألا أن طريقتنا المقترحة اختزلت (swap) بفارق واضح عن طريقة IDRGP(MVE),وأنها أسرع بكثير منها,الطريقة المقترحة الثالثة IDRGP(RMM) كانت متشابهه في التشخيص الصحيح و معدل التشخيص الخاطئ لم يسجل تقليص في النسبة كان أعلى وكذلك هي أبطأ من طريقة IDRGP(MLMS)

عند حجم عينة 300 الملاحظ أن معدلات التشخيص الصحيح لكافة الطرق متقاربة ألا أن هناك اختلافاً في معدلات تشخيص بعض القيم النظيفة التي اعتبرت قيماً شاذة (swap) , أذ أن معدل التشخيص الغير صحيح بلغ (2.672)لطريقة IDRGP(RMM) وهو الأعلى بين الطرق المقارنة و(1.158) بطريقة IDRGP(RMVN) الذي يعد الأقل من بين تلك الطرق , كما نلاحظ طريقة IDRGP(MLMS)أيضا هي الأسرع من ناحية زمن الاحتساب.

عند مقارنة DRGP(MVE) مع DRGP(RMVN) نلاحظ أن الطريقة المقترحة كانت متقاربة بالنسبة للتشخيص الصحيح بنسبة بسيطة أما بالنسبة لتشخيص القيم النظيفة على أنها شاذة فأنها قللت من معدل (swap) بنسبة بسيطة مقارنة مع الطريقة التقليدية, أما زمن الاحتساب فكان أسرع منها بفارق بسيط أيضاً , أما عند مقارنة IDRGP(MVE) مع IDRGP(RMVN) من ناحية معدلات التشخيص الصحيح فأن الطريقة المقترحة سجلت الأفضلية وطريقتنا المقترحة أيضاً اختزلت (swap) بفارق واضح جداً عن طريقة IDRGP(MVE) وكذلك هي سجلت سرعة واضحة في زمن الاحتساب , الطريقة المقترحة الثالثة IDRGP(RMM) كان التشخيص الصحيح أقل و معدل التشخيص الخاطئ لم يسجل تقليصاً في النسبة بل كان أعلى, وكذلك هي أبطأ من ناحية زمن الاحتساب مقارنة مع طريقة IDRGP(MLMS).

عند حجم عينة 500 يتم عرض معدلات التشخيص الصحيح لكافة الطرق حيث كانت متماثلة تقريباً ألا أن هناك اختلافاً في معدلات تشخيص بعض القيم النظيفة التي اعتبرت قيماً شاذة (swap) , إذ أن معدل التشخيص الغير صحيح بلغ (3.400) لطريقة IDRGP(RMM) وهو الأعلى بين الطرق المقارنة و (1.700) لطريقة DRGP(RMVN) الذي يعد الأقل من بين

تلك الطرق , كما نلاحظ طريقة IDRGP(MLMS)أيضا هي الأسرع من ناحية زمن الاحتساب.

عند مقارنة DRGP(MVE) مع DRGP(RMVN) نلاحظ أن الطريقتين متماثلتين في التشخيص الصحيح أما بالنسبة لتشخيص القيم النظيفة على أنها شاذة فأنها قللت من معدل (swap) مقارنة مع الطريقة التقليدية أما زمن الاحتساب فكانت طريقتنا المقترحة أسرع منها بفارق واضح أيضاً , أما عند مقارنة IDRGP(MVE) مع IDRGP(RMVN) من ناحية معدلات التشخيص الصحيح فأنها كانت متماثلة أيضاً وطريقتنا المقترحة أختزلت (swap) بفارق واضح جداً عن طريقة IDRGP(MVE) وكذلك هي سجلت سرعة واضحة في زمن الاحتساب , الطريقة المقترحة الثالثة IDRGP(RMM) كانت متماثلة تقريباً في التشخيص الصحيح و معدل التشخيص الخاطئ لم يسجل تقليص في النسبة بل كان أعلى وكذلك هي أبطأ من ناحية زمن الاحتساب مقارنة مع طريقة IDRGP(MLMS).

عند حجم عينة 1000 في هذا الجدول نجد ان معدل التشخيص الصحيح للطرق كافة متقارب لكن الأعلى نسبة كانت لصالح طريقة DRGP(RMVN) و IDRGP(MLMS) الا ان اختزال الـ (swap) كانت الأفضلية به لطريقة DRGP(RMVN) إذ بلغت (2.600) وزمن الاحتساب الأسرع أيضاً لصالح طريقة IDRGP(MLMS)أيضاً .

نلاحظ عند المقارنة بين طريقتي DRGP(MVE) و DRGP(RMVN) نجد أن معدل التشخيص الصحيح لصالح طريقة DRGP(RMVN) بفارق بسيط وهي أيضاً أكثر أختزالاً من ناحية ال swap وأن السرعة في زمن الاحتساب كانت أيضاً لصالح طريقة DRGP(RMVN) ؛ أما المقارنة الثانية بين الطريقة المقترحة الثانية والطريقة التقليدية فنجد أن الأفضلية من ناحية التشخيص الصحيح لصالح طريقتنا المقترحة حيث بلغت معدلي (100.1) وأيضاً بينت طريقتنا المقترحة أختزال عالي في معدل ال (swap) حيث سجلت معدل (2.700) مقابل (3.800) في الطريقة التقليدية وكذلك سجلت تفوق في زمن الاحتساب , في المقارنة الثالثة بين طريقتي IDRGP(MLMS) و IDRGP(RMM) حيث لم تسجل الطريقة المقترحة أفضلية في التشخيص الصحيح وأيضاً أختزل ال swap لم يكن لها حيث سجل معدل (5.400) مقابل (2.700) أما بالنسبة للسرعة في زمن الاحتساب كانت لصالح طريقة IDRGP(MLMS) بفارق واضح حيث سجلت الاخيرة معدل (0.062) مقابل (3.582) .

جدول (5) يعرض معدلات التشخيص الصحيح والتشخيص غير الصحيح وزمن الاحتساب في outlierse للطرق المقارنة عندما n=50,100,300,500,1000 ونسبة تلويث المتغيرات التوضيحية0.10 مع وجود HLP واحدة.

case	Table(5)							
	α = 0.10							
	n	dignostics	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(IMLMS)	IDRGP(IRMM)
y	50	correct	**4.846**	**5.139**	**5.624**	**5.743**	**5.768**	**5.883**
		Swap	**1.598**	**2.085**	**3.098**	**1.341**	**2.891**	**2.387**
		Time	**0.102**	**0.077**	**0.499**	**0.046**	**0.017**	**0.119**
	100	dignostics	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(IMLMS)	IDRGP(IRMM)
		correct	**9.721**	**10.287**	**10.872**	**10.877**	**10.877**	**10.908**
		Swap	**0.959**	**1.757**	**3.294**	**0.748**	**1.240**	**2.251**
		Time	**0.180**	**0.113**	**0.755**	**0.092**	**0.017**	**0.231**
	300	dignostics	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(IMLMS)	IDRGP(IRMM)
		correct	**29.349**	**30.124**	**30. 817**	**30.882**	**30.817**	**30.899**
		Swap	**1.502**	**1.308**	**11.201**	**1.211**	**1.233**	**4.203**
		Time	**0.533**	**0.262**	**1.962**	**0.317**	**0.023**	**0.747**
	500	dignostics	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(IMLMS)	IDRGP(IRMM)
		correct	**49.200**	**49.050**	**50.850**	**50.850**	**50.800**	**50.900**
		Swap	**1.900**	**1.500**	**16.850**	**1.850**	**1.650**	**6.350**
		Time	**0.977**	**0.413**	**3.474**	**0.629**	**0.032**	**1.438**
	1000	dignostics	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(IMLMS)	IDRGP(IRMM)
		correct	**99.8**	**100.4**	**100.7**	**100.4**	**100.4**	**100.6**
		Swap	**4.800**	**3.600**	**42.300**	**3.800**	**3.900**	**13.700**
		Time	**2.521**	**0.848**	**8.282**	**1.816**	**0.062**	**3.967**

في هذا الجدول التشخيص الصحيح وعند حجم عينة 50 كان متقارباً ماعدا طريقة DRGP(MVE) أظهرت انخفاض كبير في معدل التشخيص الصحيح واستمرار ألطريقة IDRGP(RMVN) في التفوق في اختزال معدل الـ (swap) والسرعة في زمن الاحتساب ولكن أيضا بفارق بسيط عن طريقة IDRGP(MLMS).

وعندما نعمل المقارنة الثنائية بين الطرق تتبين لنا النتائج كالاتي : عند المقارنة بين طريقة DRGP(MVE) و DRGP(RMVN) نجد أن الثانية قد ارتفع فيها معدل التشخيص الصحيح ارتفاعاً عالياً عن قرينتها وكذلك ارتفاعا في معدل الـ (swap) ألا أن زمن الاحتساب كان أقل منها . ارتفاعاً في معدل التشخيص الصحيح واختزال في معدل الـ (swap)وسرعه في زمن الاحتساب لصالح طريقه IDRGP(RMVN) مقارنة مع طريقة IDRGP(MVE). الطريقة الثالثة أظهرت ارتفاعاً في نسبة التشخيص الصحيح واختزال في معدل ألـ (swap) في طريقة IDRGP(RMM) ولكن زمن الاحتساب كان أقل سرعه مقارنة مع طريقة IDRGP(MLMS)

حاكى نفس الإجراء السابق ما عدا زيادة حجم العينة الى 100 وعند نسبة تلويث 0.10، الملاحظ أن هناك اختلاف واضح في التشخيص الصحيح بين طريقة DRGP(MVE) التقليدية وباقي الطرق التي شخصت ما معدله (9.721) من أصل (11) قيمة غير طبيعية التي

شخصت من قبل IDRGP(RMM) بمعدل تشخيص صحيح (10.908) وهو الأقرب إلى الصحة من بقية الطرق. طريقتا IDRGP(MLMS) و IDRGP(RMVN) كانتا متقاربتان جداً في التشخيص الصحيح ألا أن طريقة IDRGP(RMVN) كانت الأكثر اختزالا للـ (swap) وأسرع زمناً من بقية الطرق ماعدا IDRGP(MLMS).

عند حجم عينة 300 هنا وعند ارتفاع معدل التلويث لقيمة (0.10) لنفس العينة السابقة نلاحظ انخفاض في معدل التشخيص الصحيح في طريقة DRGP(MVE) حيث بلغ (29.349) مقارنة مع باقي الطرق التي كانت متقاربة لحد ما وسجلت أقل معدل (swap) في طريقة IDRGP(RMVN) وكذلك سجلت أعلى معدل في زمن الاحتساب طريقة IDRGP(MLMS) .

في المقارنة الثنائية بين طريقتي DRGP(MVE) و DRGP(RMVN) نجد تفوق واضح لصالح طريقتنا بواقع (30.124) مقابل (29.349) وكذلك سجلت تفوق في معدل أل (sawp) حيث بلغ (1.308) مقابل (1.502) أما زمن الاحتساب فقد كان أسرع هو أيضاً وبواقع (0.262) مقابل (0.533) , أما المقارنة الثنائية للطريقة الثانية أيضا سجلت تفوق لصالح طريقتنا المقترحة من ناحية التشخيص الخاطئ والسرعة في زمن الاحتساب وايضاً ارتفعت في معدل التشخيص الصحيح حيث بلغت في طريقتنا IDRGP(RMVN) معدل (30.882) مقابل معدل (30.817) لصالح طريقة IDRGP(MVE) ولكن كما اسلفنا سجلت الـ (swap) تفوق لصالح طريقتنا بلغ معدله (1.211) مقابل (11.201) وكذلك هو الحال لصالح السرعة في زمن الاحتساب حيث بلغت سرعة زمن الاحتساب معدل (0.317) مقابل معدل (1.962) , في المقارنة الثالثه بين طريقتي IDRGP(MLMS) و طريقة IDRGP(RMM) نجد الارتفاع فقط في معدل التشخيص الصحيح لصالح طريقتنا حيث بلغ (30.899) مقابل (30.817) أما معدل الـ (swap) فقد أرتفع بمعدل واضح في الطريقة المقترحة حيث بلغ (4.203) مقابل (1.233) وكذلك طريقتنا المقترحة لم تسجل فارق كبير في السرعه في زمن الاحتساب حيث كان المعدل (0.747) مقابل (0.023) .

عند حجم 500 تظهر طريقة IDRGP(RMM)أعلى معدل في التشخيص الصحيح وفي طريقتة DRGP(RMVN) تسجيل أنخفاض واضح في معدل ألـ (swap) أما السرعة في زمن الاحتساب كانت لطريقة IDRGP(MLMS) .

بينما في المقارنة الثنائية هناك تماثل واضح في التشخيص العام بين بعض الطرق ففي المقارنة بين طريقتي DRGP(MVE) و DRGP(RMVN) نجد ارتفاع في معدل التشخيص الصحيح لصالح الطريقة التقليدية أما معدل التشخيص الخاطئ سجل في طريقتنا معدل (1.500) مقابل (1.900) لصالح الطريقة المقترحة وزمن الاحتساب كان الاسرع في الطريقة المقترحة حيث بلغ (0.413) مقابل (0.977) , في المقارنة الثانية بين طريقتي IDRGP(MVE) و طريقة IDRGP(RMVN) كان التفوق لصالح طريقتنا ماعدا معدل التشخيص الصحيح كان متماثل حيث بلغ في الطريقتين (20.85) بينما الـ (swap) بلغ في الطريقة المقترحة معدل (1.850)

مقابل (16.850) وهو أختلاف واضح وكبير جداً والسرعة في زمن الاحتساب كذلك أيضاً حيث سجل معدل (0.629) مقابل (3.474) , في المقارنة الثالثة بين طريقتي)IDRGP MLMS) وطريقتنا المقترحة IDRGP(RMM) كذلك أيضا كباقي الجداول الاخرى كان التفوق فقط في التشخيص الصحيح لصالح طريقتنا حيث بلغ معدل (50.900) مقابل (50.800) وكان معدل الـ (swap) متفوق لصالح الطريقة التقليدية حيث سجل معدل (1.650) في الطريقة التقليدية مقابل معدل (6.350) والسرعة في زمن الاحتساب كانت ايضاً لصالح الطريقة التقليدية حيث بلغت معدل (0.032) مقابل معدل (1.438) لصالح الطريقة المقترحة

عند حجم 1000 في هذا الجدول نجد ان معدل التشخيص الصحيح لكافة الطرق متقارب لكن أعلى نسبة كانت لصالح طريقة IDRGP(MVE) الا ان اختزالـ (swap) كانت الافضلية لطريقتي DRGP(RMVN) اذ بلغت (3.600) و زمن الاحتساب الاسرع أيضاً لصالح طريقة IDRGP(MLMS)أيضاً .

نلاحظ عند المقارنة بين طريقتي DRGP(MVE) و DRGP(RMVN) نجد أن معدل التشخيص الصحيح لصالح طريقة DRGP(RMVN) وهي أيضاً أكثر أختزالاً من ناحية الswap و أن السرعة في زمن الاحتساب كانت أيضاً لصالح طريقة DRGP(RMVN) , أما المقارنة الثانية بين الطريقة المقترحة الثانية والطريقة التقليدية فنجد أن تقارب من ناحية التشخيص الصحيح حيث بلغتنا معدلي (100.4) و (100.7) ولكن بينت طريقتنا المقترحة أختزال عالي في معدل ال (swap) حيث سجلت معدل (3.800) مقابل (42.300) في الطريقة التقليدية وكذلك سجلت تفوق في زمن الاحتساب,في المقارنة الثالثة بين طريقتي IDRGP(MLMS) و IDRGP(RMM) حيث سجلت الطريقة المقترحة أفضلية في التشخيص الصحيح ولكن أختزل ال swap لم يكن لها حيث سجل معدل (13.700) مقابل (3.900) أما بالنسبة للسرعة في زمن الاحتساب كانت لصالح طريقة IDRGP(MLMS) بفارق واضح حيث سجلت الأخيرة معدل(0.062) مقابل (3.967) .

جدول (6) يعرض معدلات التشخيص الصحيح والتشخيص غير الصحيح وزمن الاحتساب في LP و outlierse للطرق المقارنة عندما n=50,100,300,500,1000 ونسبة تلويث المتغيرات التوضيحية0.10 مع وجود HLP واحدة.

case	Table(6)							
	$\alpha = 0.10$							
	n	dignostics	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(IMLMS)	IDRGP(IRMM)
X & y	50	correct	**5.888**	**5.894**	**5.881**	**5.895**	**5.895**	**5.896**
		Swap	**1.503**	**1.697**	**1.860**	**1.288**	**2.439**	**1.844**
		Time	**0.101**	**0.076**	**0.492**	**0.045**	**0.018**	**0.119**
	100	dignostics	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(IMLMS)	IDRGP(IRMM)
		correct	**10.899**	**10.899**	**10.897**	**10.899**	**10.899**	**10.899**
		Swap	**0.820**	**1.044**	**1.202**	**0.722**	**0.950**	**1.431**
		Time	**0.177**	**0.113**	**0.747**	**0.090**	**0.017**	**0.225**

	300	dignostics	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(IMLMS)	IDRGP(IRMM)
		correct	**30.891**	**30.891**	**30.891**	**30.891**	**30.891**	**30.891**
		Swap	**1.267**	**1.186**	**3.339**	**1.388**	**1.187**	**2.054**
		Time	**0.521**	**0.261**	**1.945**	**0.302**	**0.022**	**0.721**
	500	dignostics	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(IMLMS)	IDRGP(IRMM)
		correct	**50.85**	**50.85**	**50.85**	**50.85**	**50.85**	**50.85**
		Swap	**1.900**	**1.550**	**5.550**	**1.900**	**1.570**	**3.100**
		Time	**0.952**	**0.417**	**3.560**	**0.608**	**0.032**	**1.373**
	1000	dignostics	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(IMLMS)	IDRGP(IRMM)
		correct	**100.9**	**100.9**	**100.9**	**100.9**	**100.9**	**100.9**
		Swap	**3.700**	**3.500**	**15.7 00**	**3.700**	**3.800**	**6.100**
		Time	**2.226**	**0.974**	**7.578**	**1.546**	**0.064**	**3.433**

في هذا الجدول وفي حالة LP و outlierse معاً عند نسبة التلويث 0.10 عند حجم عينة 50 نجد أن أعلى نسبة في احتساب التشخيص الصحيح كانت لطريقة IDRGP(RMM) حيث سجلت معدل (5.896) وكان أفضل معدل لاحتساب الـ(swap) لصالح طريقة IDRGP(RMVN) حيث سجلت معدل (1.288) أما السرعة في زمن الاحتساب كانت لطريقة IDRGP(MLMS) , الان وفي المقارنة الثنائية بين الطرق نجد ارتفاع في معدل التشخيص الصحيح لصالح طريقة DRGP(RMVN) حيث بلغت (5.894) مقابل DRGP(MVE) والتي بلغت (5.888) أما الـ(swap) كان لصالح الأخيرة حيث بلغت (1.503) مقابل (1.697) للطريقة المقترحة ,وعند المقارنة بين طريقتي IDRGP(MVE) وIDRGP(RMVN) نجد ارتفاع معدل التشخيص الصحيح لصالح طريقتنا المقترحة حيث بلغت (5.895) مقابل (5.881) والـ(swap) كذلك سجل معدل في طريقتنا المقترحة (1.288) مقابل (1.860) وكذلك هو الحال من ناحية السرعة في زمن الاحتساب كانت أيضا لصالح الطريقة المقترحة بواقع (0.045) مقابل (0.492) ,في المقارنة الثالثة كان هناك تفوق بسيط جدا لصالح الطريقة المقترحة IDRGP(RMM)في معدل التشخيص الصحيح حيث بلغ (5.896) مقابل (5.895) لصالح طريقة IDRGP(MLMS) واختزالاً معدل ألـ(swap) أيضا سجل تفوق لصالح الطريقة المقترحة حيث بلغ (1.844) مقابل (2.439) أما السرعة في زمن الاحتساب كانت لصالح الطريقة التقليدية بواقع (0.018) مقابل (0.119) .

عند حجم عينة 100 هنا التماثل موجود في كل حالات التشخيص الصحيح ألا في حالة IDRGP(MVE) وأقل معدل في التشخيص الخاطئ كان لصالح طريقة IDRGP(RMVN) بواقع (0.722) وأسرع طريقة في زمن الاحتساب كانت لصالح طريقة IDRGP(MLMS) , ولدى المقارنة بطريقة ثنائية بين الطرق التقليدية والمقترحة نجد في المقارنة الأولى كما قلنا تماثل في التشخيص الصحيح بواقع (10.899) أما التشخيص الخاطئ كان لصالح طريقة DRGP(MVE) بواقع (0.820) ولكن السرعة في زمن الاحتساب كانت لصالح طريقة DRGP(RMVN) , وفي المقارنة الثانية بين طريقتي IDRGP(MVE) و IDRGP(RMVN) نجد ارتفاع معدل التشخيص الصحيح عن الطريقة التقليدية بمعدل (10.899) وفي التشخيص الخاطئ نجد اختزال عالي في طريقتنا المقترحة مقارنة بالطريقة التقليدية حيث بلغت

(0.722)مقابل (1.202) وكذلك السرعة في زمن الاحتساب كانت أيضاً لصالح طريقتنا المقترحة بمعدل (0.090) مقابل (0.747) .أما المقارنة ألثالثه كما قلنا هناك تماثل في معدل التشخيص الصحيح بينما الاختزال في التشخيص الخاطئ والسرعة في زمن الاحتساب كانت لصالح طريقة IDRGP(MLMS) .

عند حجم عينة 300 في هذا الجدول أيضا يستمر التماثل في كافة الطرق حيث بلغ معدل (30.891) في التشخيص الصحيح وان أفضل تشخيص خاطئ سجل لصالح طريقةDRGP(RMVN) بمعدل (1.186) وكانت أسرع طريقة هي IDRGP(MLMS)بمعدل (0.022).

(1.186) هو معدل الـ(swap) في طريقة DRGP(RMVN) وهو أفضل من معدل الـ(swap) في طريقة DRGP(MVE) حيث بلغ معدل (1.267) وكذلك السرعة في زمن الاحتساب أيضا أثبتت تفوق في الطريقة المقترحة حيث بلغت معدل (0.261)مقابل (0.521) , وان المعدل البالغ (30.891) هو معدل التشخيص الصحيح لكلا الطريقتين IDRGP(RMVN) و طريقة IDRGP(MVE)وأن ألـ (swap) سجل معدل متفوق لصالح الطريقة المقترحة بمعدل (1.388) مقابل معدل (3.339) في الطريقة التقليدية وان السرعة في زمن الاحتساب سجلت معدل(0.302)في طريقتنا المقترحة مقابل معدل (1.945) في الطريقة التقليدية , وفي المقارنة الثالثة نلاحظ وجود التماثل في التشخيص الصحيح ووجود تفوق في طريقة IDRGP(MLMS)من ناحية التشخيص الخاطئ والسرعة الزمنية في الاحتساب حيث بلغتا معدلات(1.187)و(0.022) مقابل معدلات (2.054)و (0.721)في طريقة IDRGP(RMM).

عند حجم عينة 500 وهنا أيضا يستمر التماثل في كافة الطرق حيث بلغ معدل (50.85) في التشخيص الصحيح وان أفضل تشخيص خاطئ سجل لصالح طريقة IDRGP(RMVN)بمعدل (1.550)وكانت أسرع هي طريقة IDRGP(MLMS) بمعدل (0.022).

(1.550) هو معدل الـ(swap) في طريقة DRGP(RMVN) وهو أفضل من معدل الـ(swap) في طريقة DRGP(MVE) حيث بلغ معدل (1.900) وكذلك السرعة في زمن الاحتساب أيضا أثبتت تفوق في الطريقة المقترحة حيث بلغت معدل (0.417)مقابل (0.952) , وان المعدل البالغ (50.85) هو معدل التشخيص الصحيح لكلا الطريقتين IDRGP(RMVN) و طريقة IRGP(MVE) وأن ألـ (swap) سجل معدل متفوق لصالح الطريقة المقترحة بمعدل (1.900) مقابل معدل (5.550) في الطريقة التقليدية وان السرعة في زمن الاحتساب سجلت معدل (0.608)في طريقتنا المقترحة مقابل معدل (3.560) في الطريقة التقليدية , وفي المقارنة الثالثة نلاحظ وجود التماثل في التشخيص الصحيح ووجود تفوق في طريقة IDRGP(MLMS) من ناحية التشخيص الخاطئ والسرعة الزمنية في الاحتساب حيث بلغتا معدلات (1.570) و (0.032) مقابل معدلات (3.100) و (1.373)في طريقة IDRGP(RMM).

عند حجم عينة 1000 نلاحظ وجود التماثل في التشخيص الصحيح لكافة الطرق وأن طريقة DRGP(RMNV) هي أفضل طريقة في تقليل معدل ال swap أما السرعة في زمن الاحتساب فهي لصالح طريقة IDRGP(IMLMS) .

وفي المقارنة الثنائية بين الطرق نجد في المقارنة الاولى أن الافضلية لصالح طريقتنا المقترحة في كافة المعايير عدا المقارنة في التشخيص الصحيح حيث كانت الطريقتين متماثلتين وفي المقارنة الثانية ايضاً الافضلية لصالح طريقتنا المقترحة حيث قللت من معدل الـ swap وايضا كانت اسرع من ناحية السرعة في زمن الاحتساب .وفي المقارنة الثالثة لم تكن طريقتنا المقترحة هي الافضل حيث لم تخفض من معدل التشخيص الخاطئ ولم تكن اسرع في زمن الاحتساب .

جدول (7) يعرض معدلات التشخيص الصحيح والتشخيص غير الصحيح وزمن الاحتساب في LP للطرق المقارنة عندما n=50,100,300,500,1000 ونسبة تلويث المتغيرات التوضيحية 0.15مع وجود HLP واحدة.

case	Table(7)							
	$\alpha = 0.15$							
	n	dignostics	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(IMLMS)	IDRGP(IRMM)
x	50	correct	**8.410**	**8.991**	**8.504**	**8.669**	**8.693**	**8.675**
		Swap	**1.060**	**0.841**	**2.359**	**0.822**	**1.699**	**1.785**
		Time	**0.099**	**0.075**	**0.487**	**0.045**	**0.017**	**0.116**
	100	dignostics	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(IMLMS)	IDRGP(IRMM)
		correct	**15.692**	**15.710**	**15.695**	**15.707**	**15.708**	**15.689**
		Swap	**0.406**	**0.584**	**1.474**	**0.334**	**0.436**	**1.507**
		Time	**0.176**	**0.113**	**0.742**	**0.089**	**0.017**	**0.223**
	300	dignostics	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(IMLMS)	IDRGP(IRMM)
		correct	**45.580**	**45.590**	**45.543**	**45.591**	**45.581**	**45.512**
		Swap	**0.591**	**0.506**	**1.985**	**0.579**	**0.507**	**2.523**
		Time	**0.508**	**0.260**	**1.915**	**0.291**	**0.022**	**0.696**
	500	dignostics	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(IMLMS)	IDRGP(IRMM)
		correct	**75.7**	**75.7**	**75.7**	**75.7**	**75.7**	**75.7**
		Swap	**1.001**	**0.800**	**2.300**	**1.001**	**0.900**	**2.500**
		Time	**0.938**	**0.415**	**3.431**	**0.589**	**0.038**	**1.341**
	1000	dignostics	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(IMLMS)	IDRGP(IRMM)
		correct	**150**	**150**	**149.5**	**150**	**1500**	**149.5**
		Swap	**2.001**	**1.800**	**5.001**	**2.001**	**1.900**	**8.300**
		Time	**2.207**	**0.809**	**7.560**	**1.538**	**0.070**	**3.435**

نستمر في رفع نسبة التلويث الى 0.15 والدراسة برفع حجم العينة فعند حجم عينة 50 الملاحظ في هذا الجدول أن معدلات التشخيص الصحيح لكافة الطرق متقاربة نوعا ما إلا أن هناك اختلافاً في معدلات تشخيص بعض القيم النظيفة التي اعتبرت قيماً شاذة (swap), أذ أن معدل التشخيص الغير صحيح بلغ (2.359) لطريقة IDRGP(MVE) وهو الأعلى بين الطرق المقارنة و (0.822) لطريقة IDRGP(RMVN) الذي يعد الاقل من بين تلك الطرق, كما

نلاحظ أن الطرق المقترحة هي الاسرع من سابقتها من ناحية زمن الاحتساب عدا طريقة IDRGP (RMM) ليست اسرع رغم كون الفارق بسيط في زمن الاحتساب

عند مقارنة DRGP(MVE) مع DRGP(RMVN) نلاحظ أن طريقتنا المقترحة كانت أفضل للتشخيص الصحيح وقللت من تشخيص بعض القيم النظيفة على أنها شاذة عن الطريقة التقليدية وكذلك فأنها كانت أسرع منها من ناحية زمن الاحتساب,أما عند مقارنة IDRGP(MVE) مع IDRGP(RMVN) من ناحية معدلات التشخيص الصحيح فأنها رفعت من معدل التشخيص الصحيح و أختزلت الكثير من (swap) بفارق واضح عن طريقة IDRGP(MVE),وأنها أسرع بكثير منها؛ الطريقة المقترحة الثالثة IDRGP(RMM) كانت قريبة جدا من التشخيص الصحيح وكذلك الحال بالنسبة لمعدل التشخيص الخاطئ ولكنها كانت أبطأ من طريقة IDRGP(MLMS)

عند حجم عينة 100 الملاحظ في هذا الجدول أن معدلات التشخيص الصحيح لكافة الطرق متقاربة جداً ألا أن هناك اختلافاً في معدلات تشخيص بعض القيم النظيفة التي اعتبرت قيماً شاذة (swap), إذإن معدل التشخيص غير الصحيح بلغ (1.507) لطريقة IDRGP(RMM) وهو الأعلى بين الطرق المقارنة و (0.334) لطريقة IDRGP(RMVN) الذي يعد الأقل من بين تلك الطرق, كما نلاحظ طريقة IDRGP(MLMS)أيضا هي الأسرع من ناحية زمن الاحتساب.

عند مقارنة DRGP(MVE) مع DRGP(RMVN) نلاحظ أن طريقتنا المقترحة زادت من معدل التشخيص الصحيح ولكنها شخصت بعض القيم النظيفة على أنها شاذة مقارنة مع الطريقة التقليدية بنسبة أعلى أما زمن الاحتساب كان أسرع منها بفارق بسيط أيضاً . أما عند مقارنة IDRGP(MVE) مع IDRGP(RMVN) من ناحية معدلات التشخيص الصحيح فإن طريقتنا المقترحة زادت من معدل التشخيص الصحيح وأن طريقتنا المقترحة أختزلت (swap) بفارق واضح عن طريقة IDRGP(MVE).وأسرع منها.الطريقة المقترحة الثالثة IDRGP(RMM)لم يكن لها الأفضلية في التشخيص الصحيح و معدل التشخيص الخاطئ لم يسجل تقليص في النسبة بل كان أعلى وكذلك هي أبطأ من طريقة IDRGP(MLMS)

عند حجم عينة 300 الملاحظ في هذا الجدول أن معدلات التشخيص الصحيح لكافة الطرق متقاربة جداً ألا أن هناك اختلاف في معدلات تشخيص بعض القيم النظيفة التي اعتبرت قيماً شاذة (swap) , إذ أن معدل التشخيص الغير صحيح بلغ (2.523) بطريقة IDRGP(RMM) وهو الأعلى بين الطرق المقارنة و (0.506) لطريقة DRGP(RMVN)الذي يعد الأقل من بين تلك الطرق , كما نلاحظ طريقة IDRGP(MLMS)أيضا هي الأسرع من ناحية زمن الاحتساب.

عند مقارنة DRGP(MVE) مع DRGP(RMVN) نلاحظ أن الطريقة المقترحة كانت أقل بالنسبة للتشخيص الصحيح بنسبة بسيطة أما بالنسبة لتشخيص القيم النظيفة على أنها شاذة فأنها قللت من معدل (swap) بنسبة بسيطة مقارنة مع الطريقة التقليدية أما زمن الاحتساب كان أسرع منها بفارق بسيط أيضاً , أما عند مقارنة IDRGP(MVE) مع IDRGP(RMVN) من ناحية معدلات التشخيص الصحيح فأن الأفضلية كانت لطريقتنا المقترحة وطريقتنا المقترحة

اختزلت (swap) بفارق واضح جداً عن طريقة IDRGP(MVE) وكذلك هي سجلت سرعة واضحة في زمن الاحتساب , الطريقة المقترحة الثالثة IDRGP(RMM) كانت أقل في التشخيص الصحيح و معدل التشخيص الخاطئ لم يسجل تقليص في النسبة بل كان أعلى, وكذلك هي أبطأ من ناحية زمن الاحتساب مقارنة مع طريقة IDRGP(MLMS).

عند حجم عينة 500 يعرض هذا الجدول أن معدلات التشخيص الصحيح لكافة الطرق حيث كانت متماثلة جميعها و أن هناك اختلافاً في معدلات تشخيص بعض القيم النظيفة التي اعتبرت قيماً شاذة (swap) , إذ أن معدل التشخيص الغير صحيح بلغ (2.500) لطريقة IDRGP(RMM) وهو الأعلى بين الطرق المقارنة و (0.008) لطريقة DRGP(RMVN)الذي يعد الأقل من بين تلك الطرق , كما نلاحظ طريقة IDRGP(MLMS) أيضا هي الأسرع من ناحية زمن الاحتساب.

عند مقارنة DRGP(MVE) مع DRGP(RMVN) نلاحظ أن الطريقتين متماثلتين في التشخيص الصحيح أما بالنسبة لتشخيص القيم النظيفة على أنها شاذة فأنها قللت من معدل (swap) بنسبة واضحة مقارنة مع الطريقة التقليدية أما زمن الاحتساب فكانت طريقتنا المقترحة أسرع منها بفارق واضح أيضاً , أما عند مقارنة IDRGP(MVE) مع IDRGP(RMVN) من ناحية معدلات التشخيص الصحيح فأنها كانت متماثلة أيضاً وطريقتنا المقترحة اختزلت (swap) بفارق واضح جداً عن طريقة IDRGP(MVE) وكذلك هي سجلت سرعة واضحة في زمن الاحتساب , الطريقة المقترحة الثالثة IDRGP(RMM) كانت متماثلة تقريباً في التشخيص الصحيح و معدل التشخيص الخاطئ لم يسجل تقليص في النسبة بل كان أعلى وكذلك هي أبطأ من ناحية زمن الاحتساب مقارنة مع طريقة IDRGP(MLMS).

عند حجم عينة 1000 في هذا الجدول سجلت الطرق تماثل في معدل التشخيص الصحيح حيث بلغ معدلات (150) و (149.5) وان اختزالـ (swap) كانت الأفضلية فيه لطريقةDRGP(RMVN) اذ بلغت (1.8) و زمن الاحتساب الاسرع أيضاً لصالح طريقة IDRGP(MLMS) أيضاً .

نلاحظ عند المقارنة بين طريقتي DRGP(MVE) و DRGP(RMVN) نجد أن معدل التشخيص الصحيح متماثل في الطريقتين , والطريقة المقترحة كانت أكثر أختزالاً من ناحية الـ swap بالمقارنة مع الطريقة المقترحة و أن السرعة في زمن الاحتساب كانت أيضاً لصالح طريقة DRGP(RMVN) ؛ أما المقارنة الثانية بين الطريقة المقترحة الثانية والطريقة التقليدية فنجد أن من ناحية التشخيص الصحيح لصالح الطريقة المقترحة حيث بلغ معدل (150) مقابل (149.5) ولكن بينت طريقتنا المقترحة اختزالا عالي في معدل ال (swap) حيث سجلت معدل (2.001) مقابل (5.001) في الطريقة التقليدية وكذلك سجلت تفوق في زمن الاحتساب , في المقارنة الثالثة بين طريقتي IDRGP(MLMS) و IDRGP(RMM) حيث لم تسجل الطريقة المقترحة أفضلية في التشخيص الصحيح وأيضاً في أختزل ال swap لم يكن لها حيث سجل

معدل (8.300) مقابل (1.900)أما بالنسبة للسرعة في زمن الاحتساب كانت لصالح طريقة IDRGP(MLMS) بفارق واضح حيث سجلت الأخيرة معدل (0.070) مقابل (3.435) .

جدول (8) يعرض معدلات التشخيص الصحيح والتشخيص غير الصحيح وزمن الاحتساب في outlierse للطرق المقارنة عندما n=50,100,300,500,1000 ونسبة تلويث المتغيرات التوضيحية 0.15 مع وجود HLP واحدة.

case	Table(8)							
	α = 0.15							
	n	dignostics	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(IMLMS)	IDRGP(IRMM)
y	50	correct	**3.815**	**4.210**	**8.025**	**7.912**	**8.032**	**8.802**
		Swap	**1.395**	**2.012**	**3.049**	**0.790**	**1.840**	**1.914**
		Time	**0.103**	**0.077**	**0.493**	**0.046**	**0.017**	**0.119**
	100	dignostics	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(IMLMS)	IDRGP(IRMM)
		correct	**9.275**	**9.831**	**15.632**	**15.761**	**15.761**	**15.832**
		Swap	**0.824**	**1.609**	**2.471**	**0.341**	**0.624**	**1.675**
		Time	**0.180**	**0.113**	**0.752**	**0.093**	**0.017**	**0.230**
	300	dignostics	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(IMLMS)	IDRGP(IRMM)
		correct	**27.452**	**28.623**	**45.708**	**45.794**	**45.705**	**45.840**
		Swap	**1.790**	**1.637**	**10.485**	**0.539**	**0.598**	**3.489**
		Time	**0.536**	**0.262**	**1.974**	**0.318**	**0.022**	**0.747**
	500	dignostics	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(IMLMS)	IDRGP(IRMM)
		correct	**38.650**	**45.300**	**75.800**	**75.800**	**75.450**	**75.750**
		Swap	**3.750**	**0.950**	**19.850**	**1.001**	**2.900**	**5.850**
		Time	**0.986**	**0.427**	**3.517**	**0.630**	**0.033**	**1.455**
	1000	dignostics	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(IMLMS)	IDRGP(IRMM)
		correct	**110.5**	**109.9**	**150.4**	**150.8**	**150.4**	**150.9**
		Swap	**9.001**	**4.100**	**32.800**	**0.800**	**0.900**	**9.100**
		Time	**2.446**	**0.818**	**8.062**	**1.770**	**0.061**	**3.870**

وفي ال outlierese عند حجم عينة 50في نسبة التلويث الثالثة ألبالغه (0.15) لنفس العينات الان نبدأ بالعينة ألبالغه (50) نلاحظ انخفاضاً معدل التشخيص الصحيح لطريقتي DRGP(MVE) و طريقة DRGP(RMVN) لمعدل (3.815) و(4.210) على التوالي من أصل (9) من القيم غير الطبيعية أما بقية الطرق فقد كانت مقاربة للتشخيص الصحيح كما سجلت طريقة IDRGP(RMVN) أقل نسبة في معدل الـ (swap) وكذلك اختزال في سرعة الاحتساب لكن وجود سرعة في زمن الاحتساب بفارق قليل جدا لطريقة IDRGP(MLMS) .

وعند المقارنة بين طرقنا المقترحة والطرق السابقة تظهر طريقة DRGP(RMVN) ارتفاعاً في نسبة التشخيص الصحيح عن طريقة DRGP(MVE) لكن هذا التشخيص هو قليل جدا مقارنة بالطرق الأخرى وكذلك مقارنة مع معدل القيم غير الطبيعية الظاهرة في الخوارزمية أما بالنسبة لل(swap) فهو أيضا كان مرتفع للطريقة المقترحة ألا أن سرعة الاحتساب كانت للطريقة المقترحة .في المقارنة الثانية لم ترتفع نسبة التشخيص الصحيح في الطريقة المقترحة على

نظيرتها السابقة حيث بلغت كل منهما (7.912) و (8.025) نسبة الـ (swap) سجلت الطريقة المقترحة انخفاضاً كبيراً به مقارنة مع الطريقة السابقة بمعدلات (0.790) مقابل (3.049) وكذلك سرعة في زمن الاحتساب حيث بلغت(0.046) مقابل (0.493) , الطريقة الثالثة سجلت زيادة في التشخيص الصحيح لطريقة IDRGP(RMM) بواقع (8.802) مقابل (8.032) أما ألـ (swap) فهو لم يسجل اختزال كبير للطريقة المقترحة بل كان بفارق بسيط بلغ (1.914) مقابل(1.840) وكذلك هو الحال بالنسبة لزمن الاحتساب حيث سجل (0.119) مقابل (0.017)

عند حجم عينة 100 عند ارتفاع نسبة التلويث في هذا الجدول الى معدل (0.15) نلاحظ استمرار التقارب في معدلات التشخيص الصحيح في كافة الطرق ماعدا طريقتي DRGP(MVE)و DRGP(RMVN) حيث انخفضت الى معدل (9.275) و (9.831) أما من ناحية الـ (swap) فقد سجلت طريقتنا ألمقترحه IDRGP(RMVN) أقل معدل لها أي أنها مازالت مستمرة في ألاختزال بأقل قيمة مقارنة مع بقية الطرق كافة وكذلك هي أيضاً سجلت سرعة في الاحتساب ولكن ليست أسرع من طريقة IDRGP(MLMS) .

وعند المقارنة بين الطرق بشكل ثنائي نجد أن طريقة DRGP(RMVN) متفوقة على طريقة DRGP(MVE) من ناحية التشخيص الصحيح حيث ارتفعت النسبة الى معدل (9.831) مقابل (9.275) وكذلك بالنسبة للزمن حيث كانت أسرع منها بمعدل (0.113) مقابل (0.180) ولكنها لم تتفوق عليها من ناحية ألـ(swap) حيث سجلت معدل (1.609) مقابل (0.824) . أما طريقتي IDRGP(MVE) و IDRGP(RMVN) نجد أن طريقتنا المقترحة أثبتت تفوق في كل شيء من ناحية ارتفاع معدل التشخيص الصحيح وكذلك الاختزال في الـ (swap) وتسجيل سرعة في زمن الاحتساب . أما المقارنة بين الطريقة الثالثة فقد تفوقت طريقة IDRGP(RMM) على طريقة IDRGP(MLMS) فقط من ناحية ارتفاع معدل التشخيص الصحيح حيث سجلت معدل (15.832) مقابل (15.761) ولكن ألـ (swap) سجل (1.675) مقابل (0.624) وكذلك لم تكن أسرع من ناحية الزمن حيث بلغت (0.230) مقابل (0.017)

عند حجم عينة 300 عند زيادة حجم العينة وزيادة نسبة التلويث تظهر طريقتي DRGP(MVE) وDRGP(RMVN) انخفاض كبير في معدل التشخيص الصحيح عن باقي الطرق الأخرى وتسجيل انخفاض واضح في طريقتي IDRGP(RMVN) و IDRGP(MLMS) في معدل ألـ (swap) أما السرعة في زمن الاحتساب كانت لطريقة IDRGP(MLMS) .

بينما في المقارنة الثنائية هناك اختلاف واضح عن التشخيص العام ففي المقارنة بين طريقتي DRGP(MVE) و DRGP(RMVN) نجد ارتفاع في معدل التشخيص الصحيح لصالح طريقتنا المقترحة بلغ (28.623) مقابل (27.452) أما معدل التشخيص الخاطئ سجل في طريقتنا معدل (1.637) مقابل (1.790) لصالح الطريقة التقليدية وزمن الاحتساب كان الأسرع في الطريقة المقترحة حيث بلغ(0.262) مقابل (0.536) , في المقارنة الثانية بين طريقتي IDRGP(MVE) و طريقة IDRGP(RMVN) كان التفوق لصالح طريقتنا في التشخيص

الصحيح حيث بلغ في الطريقة المقترحة (45.708) مقابل (45.794) بينما الـ (swap) بلغ في الطريقة المقترحة معدل (0.539) مقابل(10.485) وهو أختلاف واضح وكبير جداً والسرعة في زمن الاحتساب كان كذلك أيضاً حيث سجل معدل (0.318) مقابل (1.974) , في المقارنة الثالثة بين طريقتي IDRGP(MLMS) وطريقتنا المقترحة IDRGP(RMM) كذلك أيضا كباقي الجداول الاخرى كان التفوق فقط في التشخيص الصحيح لصالح طريقتنا حيث بلغ معدل (45.840) مقابل (45.705) وكان معدل الـ (swap) متفوق لصالح الطريقة التقليدية حيث سجل معدل (0.598) في الطريقة التقليدية مقابل معدل (3.489) والسرعة في زمن الاحتساب كانت ايضاً لصالح الطريقة التقليدية حيث بلغت معدل (0.022) مقابل معدل (0.747) لصالح الطريقة المقترحة .

عند حجم عينة 500 نتائج التشخيص الصحيح لكافة الطرق تعرض تقارب واضح بينها ماعدا طريقة DRGP(MVE) و DRGP(RMVN) حيث سجلت أنخفاض بلغا فيه (38.650) و (45.30) ونتائج الـ (swap) سجلت اقل نسبة لصالح طريقة DRGP(RMVN)وتقارب مع طريقتنا المقترحة IDRGP(RMVN) بمعدل (1) أما سرعة الاحتساب كانت لصالح طريقةIDRGP(MLMS).

وعند المقارنة الثنائية بين الطرق نأتي الى المقارنة بين طريقتي DRGP(MVE) وطريقتنا المقترحة DRGP(RMVN) نجد تفوق واضح لطريقتنا المقترحة في كافة المقارنات ففي حالة التشخيص الصحيح وصلت طريقتنا المقترحة لـ(45.300) مقابل (38.650) اما من ناحية الـ (swap)فقد بلغت (0.950) مقابل (3.750) أما الزمن فقد حصلت طريقتنا المقترحة على سرعة في زمن الاحتساب وصل ألى (0.427) مقابل (0.986) , نأتي الى المقارنة الثانية بين طريقتي IDRGP(MVE) مع طريقة IDRGP(RMVN) لنجد أيضاً تفوق لصالح طريقتنا المقترحة في كافة الامور ففي التشخيص الصحيح سجلت طريقتنا المقترحة معدل بلغ (75.800) بتماثل (75.800) أما ألـ (swap) سجل (1.001) مقابل (19.850) والسرعة الزمنية في الاحتساب سجلت معدل (0.630) مقابل (3.517) , وفي المقارنة الثالثة تستمر طريقة IDRGP(MLMS) بالتفوق على طريقتنا المقترحة في أختزال الـ (swap) والسرعة الزمنية ولكن لم تتفوق عليها في التشخيص الصحيح حيث سجلت طريقتنا المقترحة معدل (75.750) في التشخيص الصحيح مقابل (75.450) وبلغت في التشخيص الخاطئ (5.850) مقابل (2.900) والسرعة الزمنية في الاحتساب بلغت (1.455) مقابل (0.033) .

عند حجم عينة 1000 في هذا الجدول سجلت طريقة IDRGP(RMM) أعلى معدل في التشخيص الصحيح حيث بلغ معدل (150.9) وان اختزالـ (swap) كانت الأفضلية فيه لطريقة IDRGP(RMVN) اذ بلغت (0.800) و زمن الاحتساب الأسرع أيضاً لصالح طريقة IDRGP(MLMS) أيضاً .

نلاحظ عند المقارنة بين طريقتا DRGP(MVE) و DRGP(RMVN) نجد أن معدل التشخيص الصحيح لصالح طريقة DRGP(MVE) وهي ليست أكثر أختزالاً من ناحية الـ swap بالمقارنة

مع الطريقة المقترحة و أن السرعة في زمن الاحتساب كانت أيضاً لصالح طريقة DRGP(RMVN) , أما المقارنة الثانية بين الطريقة المقترحة الثانية والطريقة التقليدية فنجد أن من ناحية التشخيص الصحيح لصالح الطريقة المقترحة حيث بلغ معدل (150.8) مقابل (150.4) وأيضاً بينت طريقتنا المقترحة اختزال عالي في معدل ال (swap) حيث سجلت معدل (0.800) مقابل (32.800) في الطريقة التقليدية وكذلك سجلت تفوق في زمن الاحتساب , في المقارنة الثالثة بين طريقتي IDRGP(MLMS) و IDRGP(RMM) حيث سجلت الطريقة المقترحة أفضلية في التشخيص الصحيح ولكن أختزل ال swap لم يكن لها حيث سجل معدل (9.100) مقابل (0.900) أما بالنسبة للسرعة في زمن الاحتساب كانت لصالح طريقة IDRGP(MLMS) بفارق واضح حيث سجلت الأخيرة معدل (0.061) مقابل (3.870) .

جدول (9) يعرض معدلات التشخيص الصحيح والتشخيص غير الصحيح وزمن الاحتساب في LP و outlierse للطرق المقارنة عندما n=50,100,300,500,1000 ونسبة تلويث المتغيرات التوضيحية 0.15 مع وجود HLP واحدة.

case	Table(9)							
	$\alpha = 0.15$							
	n	dignostics	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(IMLMS)	IDRGP(IRMM)
X & y	50	correct	**8.755**	**8.832**	**8.751**	**8.796**	**8.833**	**8.834**
		Swap	**0.902**	**0.755**	**1.116**	**0.732**	**1.497**	**1.046**
		Time	**0.100**	**0.075**	**0.490**	**0.045**	**0.017**	**0.118**
	100	dignostics	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(IMLMS)	IDRGP(IRMM)
		correct	**15.847**	**15.847**	**15.837**	**15.847**	**15.847**	**15.847**
		Swap	**0.440**	**0.549**	**0.580**	**0.396**	**0.472**	**0.753**
		Time	**0.175**	**0.112**	**0.744**	**0.089**	**0.017**	**0.222**
	300	dignostics	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(IMLMS)	IDRGP(IRMM)
		correct	**45.847**	**45.847**	**45.835**	**45.847**	**45.847**	**45.847**
		Swap	**0.612**	**0.548**	**1.395**	**0.613**	**0.549**	**0.890**
		Time	**0.512**	**0.26**	**1.932**	**0.295**	**0.022**	**0.698**
	500	dignostics	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(IMLMS)	IDRGP(IRMM)
		correct	**75.8**	**75.8**	**75.8**	**75.8**	**75.8**	**75.8**
		Swap	**0.600**	**0.400**	**3.200**	**0.600**	**0.700**	**1.001**
		Time	**0.921**	**0.412**	**3.431**	**0.565**	**0.031**	**1.331**
	1000	dignostics	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(IMLMS)	IDRGP(IRMM)
		correct	**150.8**	**150.8**	**150.8**	**150.8**	**150.8**	**150.8**
		Swap	**1.500**	**1.300**	**5.800**	**1.400**	**1.500**	**1.600**
		Time	**2.158**	**0.793**	**7.446**	**1.484**	**0.062**	**3.301**

الحالة الاخيرة هي رفع نسبة التلويث الى نسبة 0.15 وفي حالتي LP و outlierse معاً

عند حجم عينة 50 كان (8.834) أعلى معدل في التشخيص الصحيح لصالح طريقتي IDRGP(RMM) و (0.732) هو أقل معدل لل (swap) في طريقة IDRGP(RMVN) و (0.017) هو أسرع زمن في الاحتساب لصالح طريقة IDRGP(MLMS) , الآن ننتقل إلى المقارنة الثنائية بين الطرق , المقارنة الأولى بين طريقتي DRGP(MVE) وDRGP(RMVN)

نجد ارتفاع في معدل التشخيص الصحيح لصالح الطريقة المقترحة بمعدل (8.832) مقابل (8.755) وكذلك اختزال بمعدل الـ(swap) بلغت (0.755) مقابل (0.902) كذلك فهي سجلت سرعة في زمن الاحتساب بمعدل (0.075) مقابل (0.100) , في المقارنة الثنائية الثانية بين طريقتي(IDRGP(MVEو IDRGP(RMVN) هناك تفوق لصالح طريقتنا المقترحة في كل الحالات حيث هناك تفوق في معدل التشخيص الصحيح بلغ (8.796) مقابل (8.751) أما ألـ (swap) سجل اختزال واضح حيث بلغ(0.732) مقابل (1.116) والسرعة في زمن الاحتساب سجلت معدل(0.045) مقابل (0.490) , في المقارنة الثالثة نلاحظ تفوق في معدل التشخيص الصحيح لصالح طريقة (IDRGP(RMM وأختزال واضح في معدل ألـ (swap) لصالح طريقتنا المقترحه بمعدل (1.046) مقابل (1.497) أما السرعة في زمن الاحتساب فهي سجلت سرعة لصالح طريقة (IDRGP(MLMS بمعدل (0.017) مقابل (0.118).

عند حجم عينة 100في هذا الجدول يستمر الحال كما هو في الجدول السابق حيث أن التماثل موجود في كل حالات التشخيص الصحيح ألا في حالة (IDRGP(MVE حيث سجلت معدل أقل وأقل معدل في التشخيص الخاطئ كان لصالح طريقة (IDRGP(RMVN بواقع (0.396) وأسرع طريقة في زمن الاحتساب كانت لصالح طريقة (IDRGP(MLMS بواقع (0.017) ,

عند المقارنة بين طريقتي (DRGP(MVE و(DRGP(RMVN نجد ان التشخيص الصحيح متماثل بينما معدل الـ(swap) فكان ألاختزال لصالح الطريقة التقليدية بواقع (0.440) مقابل (0.549) أما السرعة في زمن الاحتساب كان لصالح الطريقة المقترحة بواقع (0.112) مقابل (0.175) , في المقارنة بين طريقتي (IDRGP(MVE و (IDRGP(RMVN فعند المقارنة من ناحية التشخيص الصحيح نجد الأفضلية لصالح طريقتنا التقليدية حيث سجلت معدل (15.847) مقابل (15.837) ومعدل الـ (swap) كان أيضاً لصالح طريقتنا المقترحة حيث سجل معدل (0.396) مقابل معدل (0.580) وكذلك أيضاً السرعة في زمن الاحتساب كان لصالح طريقتنا المقترحة بمعدل(0.089) مقابل (0.744), وفي المقارنة الثالثة بين طريقتي (IDRGP(MLMS و طريقة (IDRGP(RMM نجد تماثل في التشخيص الصحيح وأن الاختزال في التشخيص الخاطئ كان لصالح الطريقة التقليدية بمعدل (0.472) مقابل (0.753) وأن السرعة في زمن الاحتساب كان أيضاً لصالح الطريقة التقليدية بمعدل (0.017) مقابل (0.222) .

عند حجم عينة 300 أيضا في هذا الجدول أيضا يستمر التماثل في كافة الطرق حيث بلغ معدل (45.847) في التشخيص الصحيح ماعدا طريقة (IDRGP(MVE حيث أنخفض فيها التشخيص الصحيح بفارق بسيط بمعدل(45.835)وان أفضل تشخيص خاطئ سجل لصالح طريقة(DRGP(RMVN بمعدل (0.548) وكانت أسرع طريقة هي طريقة (IDRGP(MLMS بمعدل (0.022).

في المقارنة الثنائية بين الطرق كل طريقتين على حدة نجد التماثل في التشخيص الصحيح في كلا طريقتي (DRGP(MVE و (DRGP(RMVN بينما التشخيص الخاطئ والسرعة في زمن الاحتساب كان التفوق فيها لطريقة (DRGP(RMVN بواقع معدلات (0.548) و (0.260)

على التوالي مقابل معدلات في الطريقة التقليدية بلغت (0.612) و (0.512) على التوالي أيضا .

في المقارنة الثانية بين طريقتي (IDRGP(MVE وIDRGP(RMVN) نلاحظ أفضلية لصالح طريقتنا المقترحة في التشخيص الصحيح حيث بلغ معدل (45.847) مقابل (45.835) وفي التشخيص الخاطئ معدل (0.613) مقابل (1.395) و في السرعة في زمن الاحتساب بمعدل (0.295) مقابل (1.932) .

وفي المقارنة الثالثة نلاحظ وجود التماثل في التشخيص الصحيح بين كلا الطريقتين المقترحة والتقليدية بمعدل (45.847) بينما سجلت الأفضلية لطريقة IDRGP(MLMS) على طريقة IDRGP(RMM)في التشخيص الخاطئ بمعدل (0.549) مقابل معدل (0.890) وفي السرعة في زمن الاحتساب بمعدلات (0.022) مقابل معدل (0.698).

عند حجم عينة 500 أيضا في هذا الجدول أيضا يستمر التماثل في كافة الطرق حيث بلغ معدل (75.8) في التشخيص الصحيح وان أفضل تشخيص خاطئ سجل لصالح طريقتي DRGP(RMVN) بمعدل (0.400) وكانت أسرع طريقة هي طريقة IDRGP(MLMS) بمعدل (0.031).

في المقارنة الثنائية بين الطرق كل طريقتين على حدة نجد التماثل في التشخيص الصحيح في كلا طريقتي DRGP(MVE) و DRGP(RMVN) بينما التشخيص الخاطئ والسرعة في زمن الاحتساب كان التفوق فيها لطريقة DRGP(RMVN) بواقع معدلات (0.400) و (0.412) على التوالي مقابل معدلات في الطريقة التقليدية بلغت (0.600) و (0.921) على التوالي أيضا .

في المقارنة الثانية بين طريقتي (IDRGP(MVE وIDRGP(RMVN) نلاحظ في التشخيص الصحيح حيث بلغ معدل (75.8) وفي التشخيص الخاطئ معدل (0.600) مقابل (3.200) و في السرعة في زمن الاحتساب بمعدل (0.565) مقابل(3.431) .وفي المقارنة الثالثة نلاحظ وجود التماثل في التشخيص الصحيح بين كلا الطريقتين المقترحة والتقليدية بمعدل (75.8) بينما سجلت الأفضلية لطريقة IDRGP(MLMS) على طريقة IDRGP(RMM) في التشخيص الخاطئ بمعدل (700 .0) مقابل معدل(1.001) وفي السرعة في زمن الاحتساب بمعدلات (0.031) مقابل معدل (1.331).

عند حجم عينة 1000 أيضا في هذا الجدول أيضا يستمر التماثل في كافة الطرق حيث بلغ معدل (150.8) في التشخيص الصحيح وان أفضل تشخيص خاطئ سجل لصالح طريقةDRGP(RMVN) بمعدل (1.300) وكانت اسرع طريقة هي طريقة IDRGP(MLMS) بمعدل (0.062).

في المقارنة الثنائية بين الطرق كل طريقتين على حدة نجد التماثل في التشخيص الصحيح في كلا طريقتي DRGP(MVE) و DRGP(RMVN) بينما التشخيص الخاطئ والسرعة في زمن الاحتساب كان التفوق فيها لطريقة DRGP(RMVN) .

في المقارنة الثانية بين طريقتي IDRGP(MVE) وIDRGP(RMVN)يستمر التماثل في التشخيص الصحيح ونلاحظ أفضلية لصالح طريقتنا المقترحة في التشخيص الخاطئ بفارق واضح جداً حيث بلغ معدل (1.400) مقابل (5.800) و في السرعة في زمن الاحتساب بمعدل (1.484) مقابل (7.446) .

وفي المقارنة الثالثة نلاحظ وجود التماثل في التشخيص الصحيح بين كلتا الطريقتين المقترحة والتقليدية بمعدل (150.8) بينما سجلت الأفضلية لطريقة IDRGP(MLMS) على طريقة IDRGP(RMM) في التشخيص الخاطئ ولو بفارق بسيط هنا بمعدل (1.500) مقابل معدل (1.600) وفي السرعة في زمن الاحتساب بمعدلات(0.062) مقابل معدل (3.301).

3.2 الخلاصة :-

يعرض الملحق (A) خلاصة دراسة المحاكاة على شكل اشارات أذ استخدمنا أشارة (√) للدلالة على أفضلية الطريقة على الطريقة المقارنة معها وعلامة (×) للدلالة على عدم افضلية الطريقة في معايير المقارنة الثلاثة ألا وهي (التشخيص الصحيح correct ,التشخيص الخاطئ swap والوقت time). أذ اخترنا اللون الأزرق للمقارنة في ألأفضلية في تقليل معدل التشخيص الخاطئ الـ swap بين الطرق كافة واللون الوردي أستخدم للمقارنة في الأفضلية في سرعة احتساب الزمن بين الطرق كافة في الخلاصة سوف نقوم بتفسير النتائج من ناحيتين الأولى عند ثبات حجم العينة في (LP و outlier و في حال كلاهما معاً) مع زيادة نسبة التلويث والحالة الثانية عند ثبات نسبة التلويث وزيادة حجم العينة وكالاتي:-

الحالة الأولى :-

1. في حالة n=50 و $\alpha = 0.05 ,0.10,0.15$

نلاحظ عند زيادة نسبة التلويث وثبات حجم العينة عند 50 مشاهدة نجد عند المقارنة بين الطرق نجد أن طريقة ألـIDRGP(RMNV) هي أفضل طريقة في اختزالاً ألـswap وأن طريقة أل(IDRGP(MLMS سجلت أقل معدل في زمن الاحتساب ولكن التذبذب كان في احتساب معدل التشخيص الصحيح ففي ألـ LP كان التشخيص الصحيح لصالح طريقة IDRGP(MLMS)وفي أل outliers كان لصالح طريقة IDRGP(RMM)وعند التقاطع بينهما كان التشخيص الصحيح يقترب إلى التماثل بين الطرق المذكورة أعلاه فعند نسبة التلويث ألبالغه0.05 يكون التماثل بين طرق IDRGP(MVE)وIDRGP(RMVN)وIDRGP(MLMS) وIDRGP(RMM)وعند ارتفاع نسبة التلويث إلى0.10 يرجع لطريقة IDRGP(RMM)وعند نسبة التلويث البالغة 0.15يتماثل في طريقتي IDRGP(MLMS) و IDRGP(RMM) .

2. في حال العينة n=100ونسبة التلويث $\alpha = 0.05,0.10,0.15$

في هذه الحالة أيضاً يكون أقل معدل swap سجل في طريقة IDRGP(RMVN)وأسرع طريقة أيضاً هي طريقة IDRGP(MLMS)والاختلاف فقط في معدل التشخيص الصحيح حيث عند نسبة تلويث 0.05 و0.10 يكون أقوى تشخيص صحيح في طريقة DRGP(RMVN)وعندما يرتفع مستوى التلويث إلى0.15 يصبح أفضل تشخيص لصالح طريقة IDRGP(MLMS)وعند ال outliers يكون التشخيص الصحيح لصالح طريقة IDRGP(RMM)عند كل مستويات نسبة التلويث وفي كلا الحالتين يكون التشخيص الصحيح متماثل في كافة الطرق ألا في حالة IDRGP(MVE)ينخفض فيها التشخيص الصحيح عن باقي الطرق .

3. عند حجم عينة n=300 و $\alpha = 0.05,0.10,0.15$

عند ارتفاع حجم العينة إلى ألـ300 في LP فأن أقل نسبة swap تسجل لصالح طريقة IDRGP(MLMS) عند نسبة تلويث 0.05 و 0.10 وعند أرتفاع نسبة التلويث ألى ألـ 0.15 فأن أقل swap يسجل لصالح طريقة DRGP(RMVN) , أما أسرع طريقة فهي بجدارة لصالح طريقة IDRGP(MLMS),أما التشخيص الصحيح في نسبة تلويث 0.05 فهو يتماثل في طريقة DRGP(RMVN) و IDRGP(RMVN) أما عند ارتفاع نسبة التلويث ألى 0.10 و0.15 فأن الأفضلية لصالح طريقة IDRGP(RMVN) , أما في ألـ outliers فأن أفضل معدل swap سجل لصالح طريقة IDRGP(MLMS) عند كافة مستويات التلويث وكذلك فأن السرعة في زمن الاحتساب لصالح طريقة ذاتها أما التشخيص الصحيح فأنه كان لصالح طريقة IDRGP(RMM) ماعدا عند ارتفاع مستوى التلويث الى نسبة 0.15 فكانت في طريقة IDRGP(MVE) وعند الانتقال إلىLP و outlierse معاً نجد تماثل في التشخيص الصحيح بين كافة الطرق ماعدا طريقة IDRGP(MVE) ينخفض فيها التشخيص الصحيح عن باقي الطرق أما التشخيص الخاطئ فعند نسبة تلويث 0.05 كان لصالح طريقة IDRGP(RMVN)وعند نسبة 0.10و0.15 لصالح طريقة DRGP(RMVN) و IDRGP(MLMS) ويبقى التفوق في احتساب السرعة الزمنية لصالح طريقة IDRGP(MLMS).

4. عند حجم عينة n=500 و $\alpha = 0.05,0.10,0.15$

أما عند أرتفاع حجم العينة الى ألـ 500 ففي الـ LP نجد هناك تماثل في افة الطرق عند معدل التشخيص الصحيح وفي الـ swap تذهب أفضلية التشخيص الخاطئ إلى طريقة DRGP(RMVN) وهذا في كافة مستويات التلويث وفي outlier نجد تذبذب في التشخيص الصحيح بين الطرق ((IDRGP(MVE و IDRGP(RMVN) و IDRGP(IRMM)) أما بالنسبة للـ swap فيبقى لصالح طريقة ((DRGP(RMVN) والسرعة في زمن الاحتساب لصالح طريقة IDRGP(MLMS))وعند التقاطع بين (LP و outlier) معاً نجد أن أيضاً نجد التماثل في التشخيص الصحيح ونجد أفضلية لصالح طريقة DRGP(RMVN) واستمرار التفوق الزمني لصالح طريقة (IDRGP(MLMS).

5. عند حجم عينة n=1000 و $\alpha = 0.05,0.10,0.15$

في هذه الحالة وعند ألـ LP نجد أيضا تماثل في التشخيص الصحيح عند كافة الطرق في نسبة تلويث (0.05) وعند ارتفاع نسبة التلويث إلى 0.10 يحدث التماثل فقط عند طريقتي DRGP(RMVN) و IDRGP(MLMS) وعند ارتفاع نسبة التلويث الى 0.15 يحدث التماثل عند كافة الطرق ماعدا طريقتي IDRGP(MVE) وIDRGP(RMM) واستمرار الأفضلية في التشخيص الخاطئ لصالح طريقة DRGP(RMVN)وكذلك أيضاً استمرار الأفضلية في السرعة في احتساب الزمن لصالح طريقة (IDRGP(MLMS). وعند الانتقال إلى الـ outliers نجد تذبذب في التشخيص الصحيح بين طرق ((IDRGP(RMM, IDRGP(MVE)وأن الـ swap أيضاً لصالح طريقتنا المقترحة DRGP(RMVN)واستمرار التفوق الزمني لصالح طريقة (IDRGP(MLMS).

أما عند الانتقال الى حالة التقاطع فنجد التماثل عند كافة الطرق في التشخيص الصحيح وعند نسب التلويث الثلاثه و أستمرارية الافضلية في التشخيص لصحيح لصالح طريقة DRGP(RMVN) وأستمرار التفوق الزمني لصالح طريقة (IDRGP(MLMS).

الحالة الثانية :- عند زيادة حجم العينة وثبات نسبة التلويث

1. عند نسبة تلويث 0.05 وأحجام عينة بلغت n=50,100,300,500,1000

في هذه الحالة وفي الـ LP نجد أن الأفضلية في السرعة الزمنية في الاحتساب لصالح طريقة IDRGP(MLMS) وأن أفضل طريقة في أختزال أل swap لصالح طريقة IDRGP(RMVN) ولكن ألاختلاف يكون فقط في التشخيص الصحيح حيث يختلف بزيادة حجم العينة من طريقة إلى أخرى فعند حجم العينة 50 يكون أفضل تشخيص صحيح لصالح طريقة IDRGP(MLMS) وعند ال 100 يكون لصالح طريقة DRGP(RMVN) وكذلك الحال عند حجم عينة 300وعند حجم عينة 500 يكون متماثل عند كافة الطرق ويكون الـ swap لصالح طريقة DRGP(RMVN) وعند الـ 1000 أيضاً يكون متماثل عند كافة الطرق وأيضاً يكون ألـ swap لصالح طريقة DRGP(RMVN), أما في الـout lierse وعند نسبة تلويث 0.05فعند احجام عينة 50 و 100 يكون أقل معدل swap لصالح طريقة IDRGP(RMVN) وأسرع طريقة في زمن الاحتساب لصالح طريقة IDRGP(MLMS) ولكن فقط عند حجم عينة 300 يكون أفضل swap وأسرع طريقة لصالح طريقة IDRGP(RMVN) وأن أفضل طريقة في التشخيص الصحيح هي لصالح طريقة IDRGP(RMM) وعند ارتفاع حجم العينة إلى500 يكون أفضل swap لطريقة DRGP(RMVN)أما عند ألـ outlierse و LP معاً فعند نسبة تلويث 0.05 وزيادة حجم العينة 50, 100, 300 نجد أن أقل معدل swap يسجل لصالح طريقة IDRGP(RMVN)وعند حجم 300,500, 1000 يكون لصالح طريقة (IDRGP(RMVNوالسرعة في زمن الاحتساب لصالح طريقة IDRGP(MLMS) أما معدل التشخيص الصحيح فيكون تقريباً متماثل في كافة الطرق .

2. عند نسبة تلويث 0.10 في LP وعند أحجام عينة 50 , 100 يكون أفضل swap لصالح طريقة IDRGP(RMVN) وعند حجم عينة 300 يكون أفضل swap لصالح طريقة MLMS وأسرع طريقة هي IDRGP(MLMS) أما التشخيص الصحيح لصالح طريقة

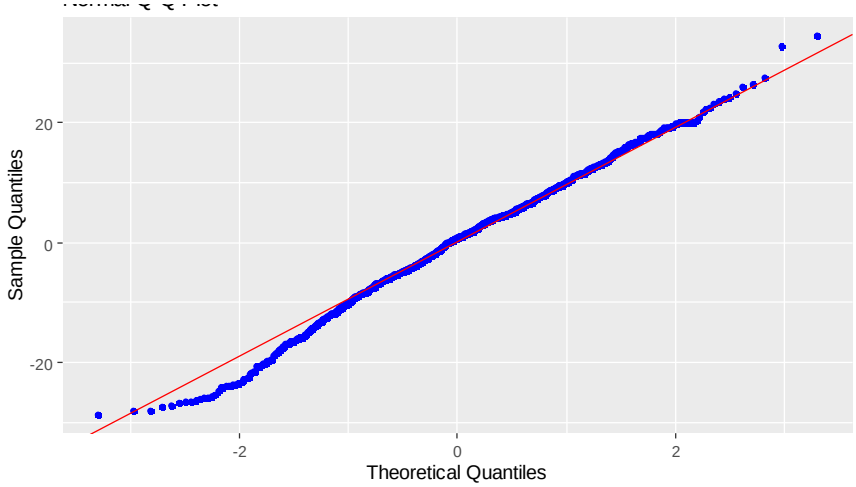
IDRGP(MLMS) عند حجم عينة 50وعند ألـ 100 يكون لصالح طريقة DRGP(RMVN) ويكون متماثل بين طريقتي DRGP(RMVN) و IDRGP(RMVN) عند حجم عينة 300 و 500 و1000 يكون لصالح طريقة IDRGP(RMVN) , وفي ألـ outliers يكون أقل معدل swap يسجل لصالح طريقة IDRGP(RMVN) في كافة أحجام العينات ماعدا حجم عينة 300و 500 و1000لصالح طريقة IDRGP(RMVN) أما الزمن فيكون لصالح طريقة IDRGP(MLMS)في كافة أحجام العينات وأن التشخيص الصحيح كان يتفوق لصالح طريقة IDRGP(RMM) , وعند الاثنين معاً نجد أن التفوق الزمني لصالح طريقة IDRGP(MLMS) أما أفضل swap يكون لصالح طريقة IDRGP(RMVN) عند أحجام عينة بلغ 50 و 100يكون لطريقة IDRGP(RMVN) وعند أل 500,300 , 1000 يكون لطريقة DRGP(RMVN) أما التشخيص الصحيح نجد أنه عند زيادة حجم العينة يقترب التشخيص الصحيح إلى ألتماثل .

3. عند نسبة تلويث 0.15فأن التشخيص الصحيح في LP لصالح طريقة IDRGP(MLMS) لغاية حجم عينة 100 وعند ألـ 300 لصالح طريقة IDRGP(RMVN) وعند أل 500 و 1000 يحدث التماثل في بعض الطرق أما أل swapفأنه عند حجم العينة فعند 300,100,50 يكون لصالح طريقة IDRGP(RMVN) وعند 500 و1000 يكون بطريقة DRGP(RMVN) أما الزمن يبقى التفوق فيه لصالح طريقة IDRGP(MLMS) وعند الـ outlier فأن التشخيص الصحيح يكون لطريقة IDRGP(RMM) عند حجم عينة 50 و 100 و أما أل 300 فأنه يذهب و لصالح طريقة IDRGP(MVE) أما أل swap فأنه لصالح طريقة IDRGP(RMVN) عند حجم عينة 50 و 100 وعند ال 300 يذهب لصالح طريقة IDRGP(MLMS) ويبقى التفوق الزمني لصالح طريقة IDRGP(MLMS) , وفي ألـ LP والـ out liers معاً نجد أن التشخيص الصحيح يذهب الى التماثل في كافة الطرق عند زيادة حجم العينة ماعدا طريقة IDRGP(MVE) فأنها تنخفض عن باقي الطرق أما أل swap فأنه يكون لصالح طريقة IDRGP(RMVN) عند حجم عينة 50 و 100 وعند ألـ 300 يتماثل في طريقتي IDRGP(MLMS) و DRGP(RMVN) والتفوق الزمني يستمر لصالح طريقة IDRGP(MLMS)

3.3 بيانات قوة انضغاط الخرسانة

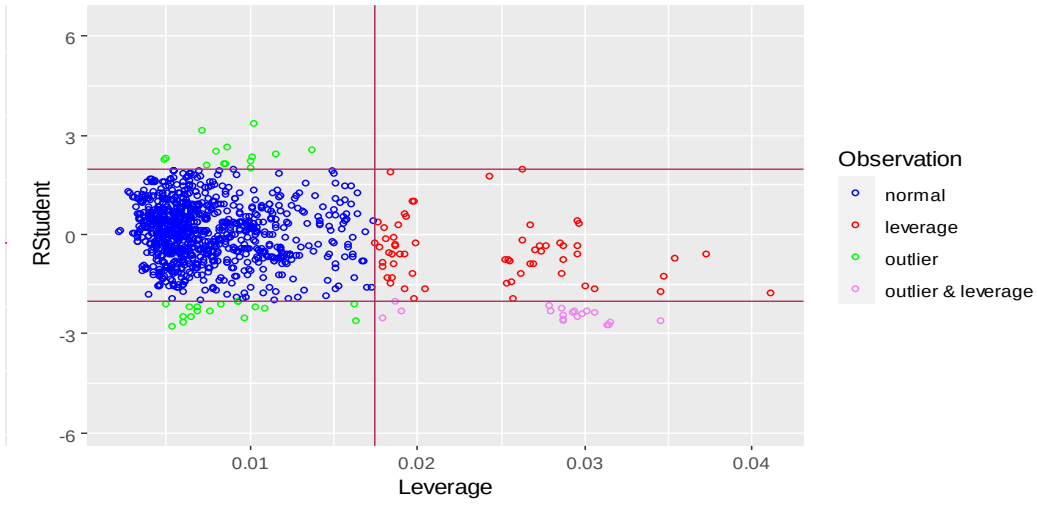
استخدمنا بيانات قوة انضغاط الخرسانة المقترحة من قبل (Yeh,1998) لغرض قياس كفاءة تشخيص القيم الشاذة لطرقنا المقترحة بالمقارنة مع الطرق السابقة. هذه البيانات تضمنت (1030) مشاهدة لتجربة قوة انضغاط الخرسانة بوجود ثمانية متغيرات وهي(كمية الماء، كمية الاسمنت، العمر الزمني للخرسانة يوم/سنة، كمية ركام المعادن المصهورة، كمية الرماد المتطاير، الملدنات الفائقة، الحشوات الخشنة، و الحشوات الناعمة)

لقد وظف (Yeh) هذه المتغيرات للتنبؤ بقوة انضغاط الخرسانة وفقاً لنموذج انحدار ألا أننا لاحظنا أن توزيع البواقي قد أنحرف عن التوزيع الطبيعي كما هو واضح في الشكل (3-1) .



الشكل (3-1) رسم الأخطاء العشوائية باستخدامNormal Q-Q

لقد لاحظنا أن المتغيرات المستخدمة في هذه البيانات تحتوي على كافة أنواع القيم الشاذة وذلك من خلال رسم العلاقة مابين مخرجات اختبار للكشف عن القيم الشاذة (R-Student) ومخرجات (Hat Matrix) كما هو واضح في الشكل (3-2)



الشكل (3-2) يوضح القيم الشاذة , LP و High LP

لمعرفة كفاءة طرقنا المقترحة قمنا بمقارنة طريقة DRGP(RMVN) مع (Hat Matrix, RMD, Hadi potential measure وDRGP(MVE), وفقاً لمعيار أن الطريقة التي تكون أكثر دقة في التشخيص الصحيح وأقل في التشخيص الخاطئ هي الأفضل . لقد استخدمنا البيانات

الهندسية المشار إليها سابقاً للمقارنة بين طريقة DRGP(MVE) و الطرق السابقة تارة والمقارنة بين طريقة DRGP(RMVE) تارة أخرى. ذلك لمعرفة التطور الذي حصل في طرق التشخيص الكلي من جانب و كفاءة طريقتنا المقترحة مع طريقة DRGP(MVE) من جانب أخر بمعنى أخر أننا وضفنا نتائج التشخيص بطريقة DRGP(MVE) كحالة قياسية للمقارنة مع (RMD, Hadi,Hat) بهدف معرفة تطور الذي حصل في فكرة التشخيص الكلي بالمقارنة مع فكرة التشخيص الفردي.من ثم قارنا طريقتنا المقترحة مع DRGP(MVE)

الجدول (10) يوضح دقة التشخيص (correct) والتشخيص الخاطئ (swamping) والفشل في التشخيص الصحيح (masking)لطرق المقارنة لطريقتنا المقترحة وطريقة DRGP(MVE) مع طرق التشخيص (Hadi, MD, Hat) لبيانات قوة انضغاط الخرسانة. من الواضح أن هناك (83) مشاهدة تم تشخيصها كـ (LP) باستخدام (Hat) ألا أن الواقع من خلال المقارنة مع طريقة DRGP(MVE) التي شخصت (107) كـ (LP) أن هناك (11) مشاهدة نظيفة تم تشخيصها على أنها LP ولكن هذا التشخيص غير صحيح , بالإضافةإلى ذلك أن هناك (35) مشاهدة عدت على أنها نظيفة باستخدام طريقة (Hat) ولكنها شخصت كـ (LP) في طريقة DRGP(MVE) . من الجدول (3-64)نلاحظ أن طريقة DRGP(RMVN) شخصت (104) مشاهدة كـ (LP) وبالمقارنة مع طريقة ألـ (Hat) فأن التشخيص الخاطئ أنخفض إلى (10) مشاهدات عما كانت عليه مع الطريقة السابقة وأن ألـ (masking) قد أنخفض إلى (31) والتشخيص الدقيق أرتفع من (72) إلى (73).

الجدول(10)يعرض تشخيص (masking & swamping) لطرق (Hadi ,MD ,Hat) بالمقارنة مع طريقتي DRGP(MVE)&DRGP(RMVN) لبيانات قوة انضغاط الخرسانة

Measure	Total	DRGP.MVE			DRGP.RMVN		
		Swamping	Masking	Correct	Swamping	Masking	Correct
Hat	83	11	35	72	10	31	73
MD	84	7	30	77	6	26	78
Hadi	93	54	68	39	55	66	38
DRGP.MVE	107				5	2	102
DRGP.RMVN	104						

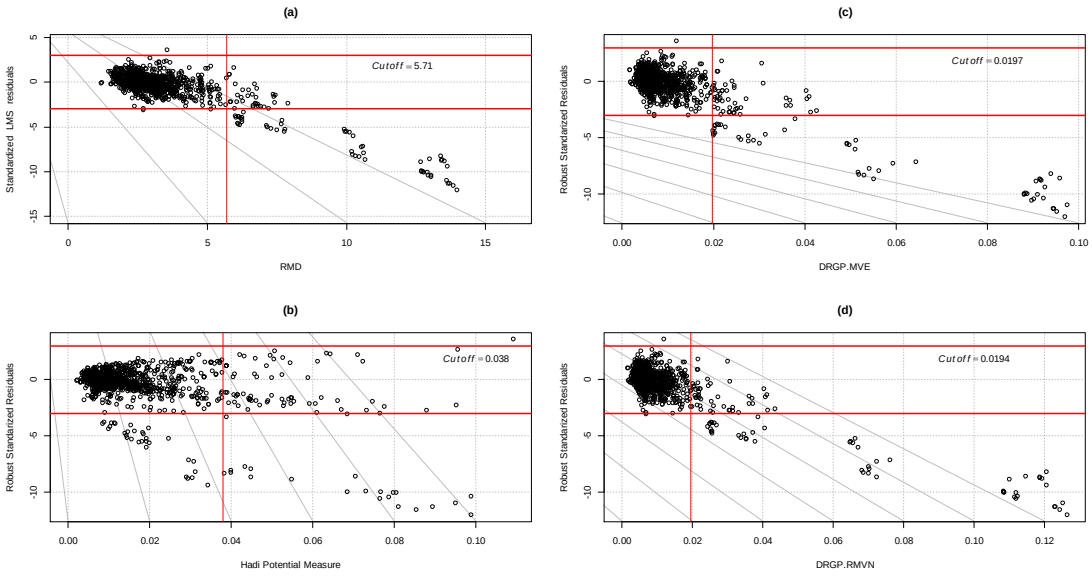
كما نلاحظ أن طريقة (MD) شخصت (84) LP تجاوزت قيمة العتبة البالغة (5.71) وهي أقل بكثير مما شخصته طريقة DRGP(MVE) (أنظر الشكل 6-a,6-c) التي كشفت عن (7) حالات كتشخيص خاطئ اعتبرتها طريقة MD قيماً شاذة وهي ليست كذلك, وأن (30) مشاهدة عدتها نظيفة ألا أنها قيماً شاذة ولهذا السبب كان الاتفاق في دقة التشخيص بين الطريقتين على (77) مشاهدة فقط. دقة التشخيص هذه ارتفعت الى (78) اتفاقاً مع طريقة DRGP(RMVN)

التي أظهرت أن هناك (6) مشاهدات نظيفة تم تشخيصها على أنها شاذة بطريقة (MD وفي المقابل قد أهملت هذه الطريقة تشخيص (26) قيمة شاذة.

عرض الجدول(10) نتائج التشخيص الكلي لطريقة (Hadi) التي لم تكن موفقه تماماً بالمقارنة مع الطريقتين السابقتين(الشكل (3-3))لا من حيث المعيار مع طريقتنا المقترحة أو طريقة DRGP(MVE) أو بالمقارنة مع طريقة Hat Matrix و MD إذ انخفضت قيمة التشخيص الصحيح إلى (39) بالمقارنة مع DRGP(MVE) فضلاً عن ارتفاع قيم ألـ swampingإلى (54) والـ maskingإلى (68) أما المقارنة مع DRGP(RMVN) أظهر ارتفاع في قيمة أل swampingإلى (55) والـ maskingإلى (66) مما يشير إلى أن هذه الطريقة هي الأسوء تشخيصاً من بين كافة الطرق.

من خلال الشكل (3-3) الذي يعرض نتائج التشخيص الواردة في الجدول (10)نلاحظ أن هناك تقارب كبير جداً بين الشكل ((c) -3-3) والشكل ((d) -3-3) والمتعلق بنتائج التشخيص لطريقتي DRGP(MVE) و DRGP(RMVN) اللتان اتفقتا على وجود (102) LP في بيانات قوة انضغاط الخرسانة وأن طريقتنا المقترحة شخصت أن هناك (5) حالات شخصتها طريقة DRGP(MVE) على أنها شواذ ولكنها ليست كذلك بالإضافة إلى تشخيص حالتين شاذتين لم تتمكن من كشفها طريقة DRGP(MVE).

الجدول (13) في الملحق (B) يظهر بشكل واضح فهرسة للصفوف التي تم تشخيص قيماً شاذة فيها باستخدام الطرق المقارنة وذلك لإظهار صورة واضحة عن كفاءة تشخيص كل طريقة للقيم الشاذة ومواقع الاختلاف والاتفاق مابين هذه الطرق.



الشكل (3-3) يعرض تشخيص طريقة(RMD, DRGP(MVE)DRGP(RMVN) ,Hadi) لبيانات قوة انضغاط الخرسانة

من الجدول (15) في الملحق (D) تم مقارنة نتائج الطريقتين مع IDRGP(MVE) , (IDRGP(RMVN) حيث وجد أن هناك (8) ثمانية قيم شاذة اكتشفت عن طريق

(DRGP(RMVN) لم تتمكن طريقة ((IDRGP(RMVN من تشخيصها بسبب حجم العينة اتفاقا مع كفاءة هذه الطرق في دراسة المحاكاة . لمزيد من التفاصيل الاطلاع على الجدول المذكور أنفاً في الملحق

4.3 بيانات القيمة السوقية للمصارف العراقية

حصل الباحث على هذه البيانات من الموقع الرسمي لسوق العراق للأوراق المالية بعد تطبيق نظام (*ISX60*), حيث تم جمع البيانات السنوية للقيمة السوقية لتسعة من المصارف المحلية الأكثر تداولاً والمسجلة في سوق العراق للأوراق المالية للمدة من (2011 – 2015). إذ تضمنت ثمانية متغيرات وهي(نسبة التداول , العائد على السهم , نسبة دوران السهم ,معدل سعر السهم السنوي , الموجودات , الأرباح المحتجزة , صافي الربح و القيمة السوقية), إذ أعتبر الباحث سبعة من هذه المتغيرات توضح أو تفسر حجم القيمة السوقية وفقاً لنموذج انحدار خطي متعدد يمكن وصفه كآلاتي :-

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7$$

حيث أن :-

y : تمثل القيمة السوقية

x_1 : تمثل نسبة التداول

x_2 : تمثل العائد على السهم

x_3 : تمثل نسبة دوران السهم

x_4 : تمثل معدل سعر السهم السنوي

x_5 : تمثل الموجودات

x_6 : تمثل الارباح المحتجزة

x_7 : تمثل صافي الربح

الملاحظ من الجدول (11)إن طريقة DRGP(MVE) شخصت (18) قيمة شاذة ألا أنها أظهرت وجود (6) مشاهدات شخصت على أنها شاذة ولكنها واقعاً نظيفة بالمقارنة مع طريقة IDRGP(MVE) و عدم وجود حالات Masking ليكون التشخيص الصحيح بين الطريقتين (12) قيمة شاذة. أما المقارنة مع طريقة IDRGP(RMVN)أظهرتإن لا وجود لحالة الـ Masking أيضاً ولكنها أخفقت في التشخيص غير الصحيح للشواذ والذي بلغ (8) مشاهدات واتفقت مع طريقتنا المقترحة ب (10) حالات تشخيص صحيح .

من النتائج المعروضة في الجدول (11)نلاحظ أيضاً إن المقارنة بين DRGP(RMVN) و IDRGP(MVE)بين عدم وجود حالات تشخيص غير صحيح swamping وظهور (3) حالات maskingواتفاق على (9) حالات تشخيص صحيح , أما المقارنة مع طريقة IDRGP(RMVN) فقد أظهرت الاتفاق على (8) حالات تشخيص صحيح (2) masking و(1) حالة واحدة تشخيص غير صحيح.

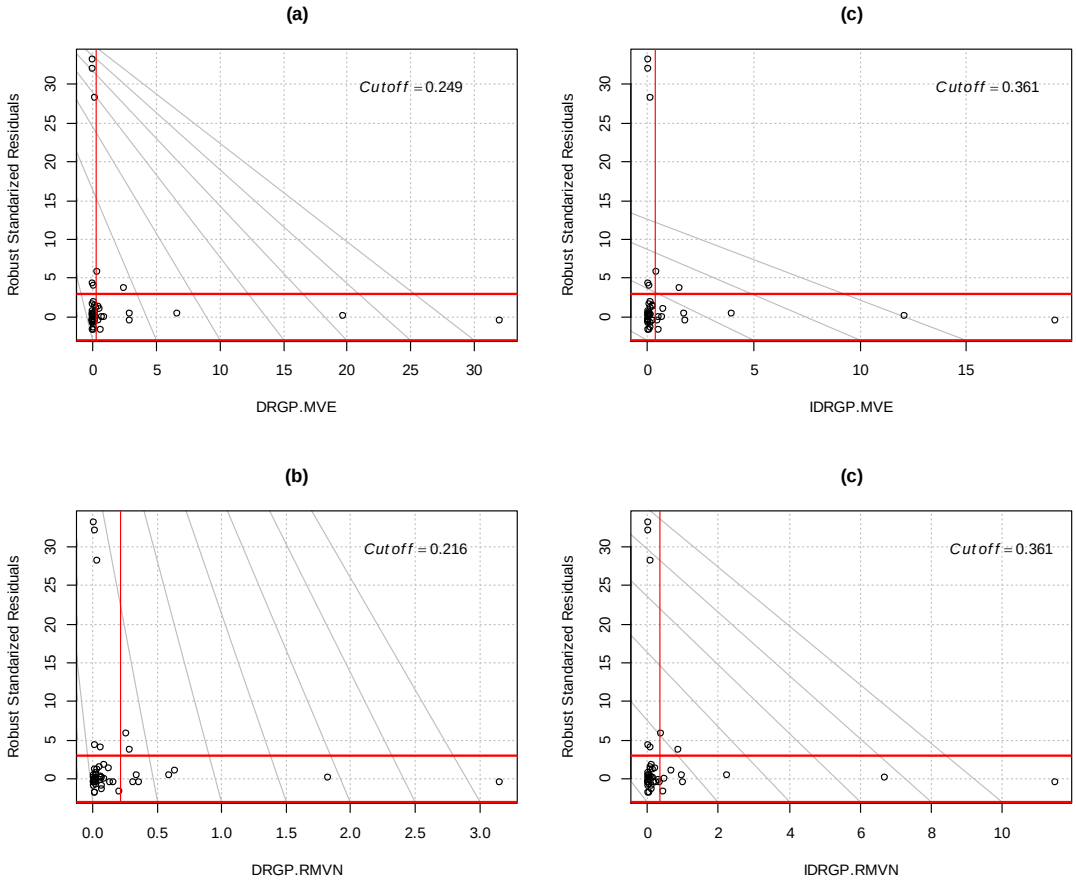
يظهر الجدول (11)نتائج المقارنة بين IDRGP(MVE) و IDRGP(RMVN) التي أظهرت أن هناك تطابق في تشخيص (10) قيم شاذة شخصت تشخيصاً صحيحاً وأن هناك (2) من القيم النظيفة شخصت على أنها شاذة في طريقة IDRGP(MVE) وعدم وجود حالات masking

من خلال الشكل (3-4) الذي يعرض نتائج التشخيص الواردة في الجدول (11)نلاحظ أن هناك تقارب كبير جدا بين الأشكال الفرعية الأربعة مع اختلاف في قيمة العتبة. لكن لا يعد تخفيض هذه القيمة معيارا للكفاءة و إنما الاختيار الأمثل لها يكون من خلال دقة تشخيص الطريقة المستخدمة ذاتها. لذلك نلاحظ إن اقل قيمة عتبة سجلت لصالح طريقة DRGP(RMVN)إلاأن هذه الطريقة سجلت وجود حالة واحدة تشخيص خاطئ لمشاهدة نظيفة عدت شاذة وحالتين لم تتمكن الطريقة من اكتشافهما . من الواضح إن اختيار عتبة امثل كان مع طريقتنا المقترحة الثانية التي أظهرت كفاءة عالية في دقة التشخيص عن غيرها.

الجدول(11) يعرض تشخيص(masking & swamping) لطرق (DRGP.MVE, DRGP.RMVN) بالمقارنة مع طريقتي IDRGP(RMVN)&IDRGP(MVE) لبيانات القيمة السوقية

Measure	Total ID.	IDRGP.MVE			IDRGP.RMVN		
		Swamping	Masking	Correct	Swamping	Masking	Correct
DRGP.MVE	18	6	0	12	8	0	10
DRGP.RMVN	9	0	3	9	1	2	8
IDRGP.MVE	12				2	0	10
IDRGP.RMVN	10						

الجدول (14) الملحق (C) يظهر مواقع الصفوف التي تحتوي على قيماً شاذة حسب طريقة التشخيص والتي تتوافق مع النتائج المعروضة في الجدول (11) ذلك لإظهار مواقع الاختلاف والاتفاق في التشخيص وعلى وجه الخصوص ذلك الاختلاف مابين طريقتي ((DRGP(MVE) , DRGP(RMVN)).علماً أن المقارنة مع نتائج التشخيص الفردي تم عرضها في الجدول (16) الملحق (E) والذي بين بشكل واضح فشل طريقة Hat و (RMD) في التشخيص الكلي للشواذ أذ لم يتجاوز (5) قيم شاذة في حين أن هناك فشل أخر حدث في طريقة ((DRGP(MVE) والتي شخصت (16) قيمة شاذة وبما يعادل ضعف القيم التي كشفتها طريقة (Hadi) ,أما DRGP(RMVN) فقد شخصت (9) قيم شاذة (5) منها هي swamp في طريقة (Hadi) و (7) هي swamp مع طريقة DRGP(MVE)



الشكل (3-4) يعرض تشخيص طريقة (DRGP(RMVN , DRGP(MVE) , IDRGP(MVE), IDRGP(RMVN))لبيانات القيمة السوقية للمصارف العراقية

-

الفصل الرابع

الاستنتاجات والتوصيات

1.4المقدمة

بينت هذه الرسالة في الفصل الأول أهمية الكشف عن القيم الشاذة في مجموعة البيانات من حيث أن وجودها يخرق فرضية التوزيع الطبيعي في النماذج الإحصائية. لذلك سعى العديد من الباحثين إلى اقتراح طرق للكشف عن القيم الشاذة بمختلف أنواعها. تميزت أغلب طرق الكشف المبكرة بالتشخيص الفردي للشواذ الامر الذي جعلها غير قادرة على كشف ظاهرة الـ Masking التي تكون عندها قيمة شاذة تغطي على قيمة شاذة واحدة او أكثر و لكن طرق الكشف تكشف القيمة التي تغطي على بقية الشواذ فقط. علاوة على ذلك ان هذه الطرق قد تعد بعض المشاهدات النظيفة قيما شاذة لسبب او لآخر Swamping. هذه الأسباب دفعت هادي عام (1996) لاقتراح طريقة Hadi Potentials Measure للتشخيص الكلي و من ثم تلتها طرق أخرى سعت إلى تحسين هذه الطريقة مثل (DRGP(MVE،IDRGP(MVE)و IDRGP(LMS)اللتين تمكنتا من تخفيض نسبة Swampingإلى أقل ما يمكن في ظل حدود معينة. لقد تناولنا طرق التشخيص الكلي السابقة بالتفصيل في الفصل الثاني المبحث الثاني من هذه الدراسة. الملاحظ ان الطرق السابقة لم تتطرق إلىأحجام العينات الكبيرة و اعتمدت كليا على مصفوفة موقع و قياس حصينة الا و هي MVE. كما لاحظنا أن زمن الاحتساب لم يدرس في الطرق السابقة.

سعينا في هذه الرسالة التركيز على ثلاث قضايا مهمة هي : تخفيض عدد المشاهدات الـ Swamping الى اقل ما يمكن في ظل وجود نسب تلويث مختلفة (0.05, 0.10,0.15) و وأحجام عينات مختلفة، ذلك من خلال استبدال مصفوفة الموقع و القياس MVE بمصفوفة موقع و قياس اخرى هي RMVN. هذا بالإضافةإلى محاولة اختزال زمن الاحتساب الى اقل ما يمكن.

2.4 مساهمة هذه الدراسة

تلخص الفقرات الآتية ستلخص مساهمتنا في هذه الدراسة

1.2.4 اقتراح طريقة DRGP(RMVN)لتحسين أداء طريقة DRGP(MVE)

في الفصل الثالث قمنا بدراسة أداء طريقتنا المقترحة DRGP(RMVN) مع DRGP(MVE) باستخدام المحاكاة والبيانات الحقيقية, وكانت النتائج لصالح طريقتنا المقترحة من ناحية دقة التشخيص, وتقليل نسب التشخيص الخاطئ ,وتقليل زمن الاحتساب وبكفاءة عالية جداً, حيث أن هذه الطريقة كانت فعالة مع أحجام البيانات العالية ,لذلك تم تطبيقها مع بيانات قوة أنضغاط الخرسانة التي حجم مشاهداتها (1030) مشاهدة.

-

2.2.4 اقتراح طريقة IDRGP(RMVN) لتحسين أداء طريقة IDRGP(MVE)

تناولنا في الفصل الثالث أيضاً اختبار كفاءة أداء طريقةIDRGP(RMVN) بالمقارنة مع طريقة IDRGP(MVE) من خلال بيانات القيمة السوقية لسوق العراق للأوراق المالية والبالغ عدد مشاهداتها (45) مشاهدة وذلك لأن دراسة المحاكاة أظهرت أن هذه الطريقة فعالة وكفوءة مع أحجام العينات الصغيرة ولذلك هي تفوقت على طريقة IDRGP(MVE) من ناحية دقة التشخيص وتقليل التشخيص الخاطئ وأختزال زمن الاحتساب.

3.2.4 أقتراح طريقة IDRGP(IRMM) لتحسين أداء طريقة IDRGP (LMS)

ناقشنا في الفصل الثالث كفاءة طريقة IDRGP(IRMM) بالمقارنة مع طريقة IDRGP(LMS) من خلال دراسة محاكاة .لقد أظهرت نتائج المحاكاة تفوق طريقة IDRGP(LMS) على طريقتنا المقترحة والطرق المقترحة الاخرى والطرق السابقة من ناحية الزمن وأختزال نسب التشخيص الخاطئ بشكل أفضل من طريقتنا المقترحة ولكن لاحظنا أن هناك حالة عدم الاستقرار في بعض الحالات خصوصاً بالمقارنة مع طرقنا المقترحة السابقة ولذلك لم يتم اختبار هاتين الطريقتين في البيانات الحقيقية بسبب عدم الاستقرار .

3.4 التوصيات

1. مما تقدم يتضح أن طريقة DRGP(RMVN) هي طريقة كفوءة وفعاله مع أحجام البيانات العالية ولذلك نوصي باستخدام هذه الطريقة للكشف عن الشواذ بأسلوب التشخيص الكلي لبيانات العلوم التطبيقية المختلفة .
2. من النتائج أتضح لنا أن طريقة IDRGP(RMVN) طريقة كفوءة وفعالة ومستقرة عندما تكون أحجام العينات صغيرة ولذلك نوصي بأستخدام هذه الطريقة في التطبيقات العلمية المختلفة .
3. لقد تناولت هذه الدراسة اقتراحين لخوارزميات التشخيص الكلي وذلك من خلال استبدال الموقع والقياس القديمة بمصفوفة RMVN ولكن هناك مصفوفات أخرى حصينة موجودة في الأدبيات الإحصائية لم تأخذها هذه الدراسة بنظر الاعتبار ولذلك نوصي الباحثيين المهتمين في الدراسات المستقبلية باختبار كفاءة توظيف مثل هذه المصفوفات بالمقارنة مع طرقنا المقترحة .
4. تناولت هذه الدراسة خوارزمية الخطوات الخمسة في إطار مصفوفة RMVN ولكن من الممكن توظيف هذه الطريقة في أطار طرق أعادة المعاينة أو طرق إعادة المعاينة الفرعية للكشف الكلي عن الشواذ .

References

1. Alguraibawi, Mohammed and Midi, Habshah and RahmatullahImon, A. H. M. (2015) *A new robust diagnostic plot for classifying good and bad high leverage points in a multiple linear regression model.* Mathematical Problems in Engineering, 2015. pp. 1-12. ISSN 1024-123X.
2. Bagheri, A., Habshah, M., & Imon, R. H. M. R. (2012). A novel collinearity-influential observation diagnostic measure based on a group deletion approach. Communications in Statistics-Simulation and Computation, 41(8), 1379-1396.
3. Belsley, D.A., Kuh, E. and Welsch, R.E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York :Wiley.
4. Box, G. E. P. (1953). Non-normality and tests on variances. *Biometrika*40,318-335.
5. Burns, P. J. (1992). A genetic algorithm for robust regression estimation. StatScience Technical Note.
6. Chatterjee, S., & Hadi, A. S. (2015). Regression analysis by example. John Wiley & Sons.
7. Croux, C., Rousseeuw, P. J., & Hössjer, O. (1994). Generalized S-estimators. Journal of the American Statistical Association, 89(428), 1271-1281
8. Devlin, Susan J, Gnanadesikan, Ramanathan, & Kettenring, Jon R. (1981). Robust estimation of dispersion matrices and principal components. *Journal of the American Statistical Association, 76*(374), 354-362.
9. Donoho, D. L. and Huber, P. J. (1983). The notion of breakdown point. In *AFestschrifrfor Erich L. Lehmann,* P. J. Bickel, K. A. Doksum, J. L. Hodges Jr.(eds.). Wadsworth, Belmont, Calif., pp. 157-184.

10. F. R. Hampel (1971), A general qualitative definition of robustness, Ann. Math. Statist., 42,pp. 1887- 1896.
11. F. R. Hampel (1968), Contributions to the theory of robust estimation, Ph. D.Thesis, University of California, Berkeley.
12. Fitrianto, Anwar and Midi, Habshah (2010) *Diagnostic-robust generalized potentials for identifying high leverage points in mediation analysis.* World Applied Sciences Journal, 11 (8). pp. 979-987. ISSN 1818-4952; ESSN: 1991-6426.
13. Geary, R. C. (1947). Testing for normality. *Biometriku 34,* 209-242. [1.3a].
14. Gujarati, Damodar N, & Porter, D. (2009). Basic Econometrics Mc Graw-Hill International Edition.
15. Habshah, M., Norazan, M.R. and Imon, A.H.M.R. (2009). The performance of Diagnostic-Robust Generalized Potentials for the identification of multiple high leverage points in linear regression. *Journal of Applied Statistics.* 36(5): 507-520.
16. Hadi, A. S. (1992). A new measure of overall potential influence in linear regression.*Computational and Statistical Data Analysis*. 14:1-27.
17. Hampel, F. R. (1968). Contributions to the theory of robust estimation. Ph.D. thesis.University of California.
18. Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*. 69: 383-393.
19. Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (1986). *Robust Statistics*. New York: Wiley.

20. Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., &Stahel, W. A. (2011). *Robust statistics: the approach based on influence functions* (Vol. 196). John Wiley & Sons.
21. Hoaglin, D. C and Welsch, R. E. (1978). The hat matrix in regression and ANOVA. *American Statistician.* 32:17-22.
22. Huber, P. J. (1964). Robust estimation of location parameters. *Annals of Mathematical Statistics.* 35:73–101.
23. Huber, P. J. (1968). Robust contidencelimits. *2. Wahrsch. uenu. Geb.* 10,269-278.
24. Huber, P. J. (1973). Robust regression: asymptotic, conjectures, and Monte Carlo. *The Annals of Statistics*, 1: 799-821.
25. Huber, P. J., & Ronchetti, E. M. (1981). Robust statistics, ser. *Wiley Series in Probability and Mathematical Statistics. New York, NY, USA, Wiley-IEEE*, *52*, 54.
26. Huber, P.J. (2004). *Robust Statistics*. New York: John Wiley & Sons.
27. Huber P. J. (1967), The behavior of maximum likelihood estimates under nonstandardconditions, in: Proc. Fifherkeley.
28. Huber P. J. (1968), Robust confidence limits,2.Wahrscheinlichkeitstheorie Vew. Gebiete,10,pp.269-278.
29. Huber P. J. (1965) A robust version of the probability ratio test, Ann. Math. Statist., 36,pp. 1753- 1758.
30. Huber, P. J. (2004). Robust statistics (Vol. 523). John Wiley & Sons.
31. Imon, A. R., &Hadi, A. S. (2008). Identification of multiple outliers in logistic regression. *Communications in Statistics-Theory and Methods*, *37*(11), 1697-1709.

32. Imon, A.H.M.R. (2002). Identifying multiple high leverage points in linear regression. *Journal of Statistical Studies*. *Special Volume in Honour of Professor Mir Masoom Ali*. 3: 207–218.
33. Imon, A.H.M.R. (2005). Identifying multiple influential observations in linear regression. *Journal of Applied Statistics*. 32: 929-946.
34. Imon, A.H.M.R. and Khan, M.A.I. (2003). A solution to the problem of multcollinearitycaused by the presence of multiple high leverage points. *International Journal of Statistical Science*. 2: 37-50.
35. Krasker, W. S., & Welsch, R. E. (1982). Efficient bounded-influence regression estimation. Journal of the American statistical Association, 77(379), 595-604.
36. Kutner, Michael H, Nachtsheim, Christopher J, Neter, John, & Li, William. (2005). Applied linear statistical models.
37. Mahalanobis, Prasanta Chandra. (1936). On the generalized distance in statistics. *Proceedings of the National Institute of Sciences (Calcutta), 2*, 49-55.
38. Maronna, R.A., Martin, R.D. and Yohai, V.J. (2006). *Robust Statistics Theory and Methods*. New York: Willy and sons.
39. Midi, H., Ramli, N and Imon, A.H.M.R. (2009). The performance of Diagnostic-Robust Generalized Potentials for the identification of multiple high leverage points in linear regression. Journal of Applied Statistics. 36(5): 507-520.
40. Midi H., Bagheri, A., &Imon, A. H. M. (2010). The Application of Robust Multicollinearity Diagnostic Method Based on Robust Coefficient

Determination to a Non-Collinear Data. *Journal of Applied Sciences*, 10(8):611-619.

41. Mohammad A. Mohammad (2015), Rbust Estimation Methods and Robust Multicollinearity Diagnostics For Multiple Regression Model in The Presence of High Leverage Collinearity -Influential Observation, Thesis submitted to the School of Graduate Studies, UPM
42. Montgomery D. C., Peck E. A., & Vining G. C.(2001) Introduction to Linear Regression Analysis, 3th ed., NewYork, John Wiely& Sons Inc, Library of congress, U.S.A.
43. Olive, David J, & Hawkins, Douglas M. (2010). Robust multivariate location and dispersion. Preprint, see (www. math. siu. edu/olive/preprints. htm).
44. Olive, David J. (2004). A resistant estimator of multivariate location and dispersion. *Computational statistics & data analysis*, *46*(1), 93-102.
45. Pearson, K. (1908). On the generalized probable error in multiple normal correlation. Biometrika. 6: 59-6.
46. Peña, Daniel, & Yohai, Victor J. (1995). The detection of influential subsets in linear regression by using an influence matrix. *Journal of the Royal Statistical Society. Series B (Methodological)*, 145-156.
47. Rousseeuw P.J. and Van Driessen, K. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics.* 41:212–223
48. Rousseeuw, P. and Van Zomeren, B. (1990).Unmasking multivariate outliers and leverage points. *Journal of American Statistical Associations*. 85: 633-639
49. Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of theAmerican Statistical Association.* 79: 871–880.

50. Rousseeuw, P. J. and Croux, C. (1993). Alternatives to the median absolute values. *Journal of the American Statistical Association*. 88: 1273-83.
51. Rousseeuw, P.J. and Leroy, A.M. (1987). *Robust Regression and Outlier Detection*. New York: Wiley
52. Simpson, J. R. (1995). *New Methods and Comparative Evaluations for Robust and Biased-Robust Regression Estimation*. Unpublished Ph.D. thesis, Arizona State University, The United States of America.
53. Stigler, *S.* M. (1973). Simon Newcomb, Percy Daniell, and the history of robustestimation 1885-1920. J. *Am. Statist. Assoc. 68,* 872-879. [1.3a].
54. Symposium on Mathematical Statistics andProbability, Vol. 1, University of California Press, Berkeley.
55. Tukey, J. W. (1960). A survey of sampling from contaminated distributions. *Contributions to probability and statistics*, 448-485.
56. Vellman, P.F. and Welsch, R.E. (1981). Efficient computing of regression diagnostics. *American Statistician*. 27: 234-242.
57. Yohai, V.J. (1987). High breakdown point and high efficiency robust estimates for regression. *The Annals of Statistics*. 15: 642-656.
58. Woodruff, D. L., & Rocke, D. M. (1994). Computable robust estimation of multivariate location and shape in high dimension using compound estimators. Journal of the American Statistical Association, 89(427), 888-896.

الملحق A

جدول رقم (12)يمثل خلاصة نتائج المحاكاة لبيان أفضلية الطرق المقارنة من ناحية معدل دقة التشخيص الصحيح ومعدل التشخيص الخاطئ ومعدل زمن الاحتساب باستخدام المؤشرات

$n = 50$	TABLE(3-1)	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(MLMS)	IDRGP(RMM)
$\alpha = 0.05$	Correct	×	√	×	√	√	×
	Swap	√	×	×	√	×	√
	time	×	√	×	√	√	×
$n = 50$	TABLE(3-2)	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(MLMS)	IDRGP(RMM)
$\alpha = 0.10$	Correct	×	√	×	√	√	×
	Swap	×	√	×	√	×	√
	time	×	√	×	√	√	×
$n = 50$	TABLE(3-3)	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(MLMS)	IDRGP(RMM)
	Correct	×	√	×	√	√	×

$\alpha = 0.15$	Swap	×	√	×	√	√	×
	time	×	√	×	√	√	×
$n = 100$	TABLE(3-7)	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(MLMS)	IDRGP(RMM)
	Correct	×	√	---	---	---	----
$\alpha = 0.05$	Swap	√	×	×	√	√	×
	time	×	√	×	√	√	×
$n = 100$	TABLE(3-8)	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(MLMS)	IDRGP(RMM)
$\alpha = 0.10$	Correct	×	√	---	----	√	×
	Swap	√	×	×	√	√	×
	time	×	√	×	√	√	×
$n = 100$	TABLE(3-9)	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(MLMS)	IDRGP(RMM)
$\alpha = 0.15$	Correct	×	√	×	√	√	×
	Swap	√	×	×	√	√	×

	time	×	√	×	√	√	×
$n = 300$	TABLE(3-13)	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(MLMS)	IDRGP(RMM)
$\alpha = 0.05$	Correct	×	√	×	√	√	×
	Swap	×	√	×	√	√	×
	time	×	√	×	√	√	×
$n = 300$	TABLE(3-14)	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(MLMS)	IDRGP(RMM)
$0\alpha = 0.1$	Correct	×	√	×	√	√	×
	Swap	×	√	×	√	√	×
		×	√	×	√	√	×
$n = 300$	TABLE(3-15)	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(MLMS)	IDRGP(RMM)
$\alpha = 0.15$	Correct	√	×	×	√	√	×
	Swap	×	√	×	√	√	×
	time	×	√	×	√	√	×

$n = 500$ $\alpha = 0.05$	TABLE(3-16)	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(MLMS)	IDRGP(RMM)
	Correct	---	---	-----	---	---	---
	Swap	×	√	×	√	√	×
	time	×	√	×	√	√	×
$n = 500$ $\alpha = 0.10$	TABLE(3-17)	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(MLMS)	IDRGP(RMM)
	Correct	√	√	√	√	√	√
	Swap	×	√	×	√	√	×
	time	×	√	×	√	√	×
$n = 500$ $\alpha = 0.15$	TABLE(3-18)	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(MLMS)	IDRGP(RMM)
	Correct	√	√	√	√	√	√
	Swap	×	√	×	√	√	×
		×	√	×	√	√	×
$n = 1000$	TABLE(3-19)	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(MLMS)	IDRGP(RMM)

$\alpha = 0.05$							
	Correct	√	√	√	√	√	√
	Swap	×	√	×	√	√	×
	time	×	√	×	√	√	×
$n = 1000$ $\alpha = 0.10$	TABLE(3-20)	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(MLMS)	IDRGP(RMM)
	Correct	×	√	×	√	√	×
	Swap	×	√	×	√	√	×
	time	×	√	×	√	√	×
$n = 1000$ $\alpha = 0.15$	TABLE(3-21)	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(MLMS)	IDRGP(RMM)
	Correct	√	√	×	√	√	×
	Swap	×	√	×	√	√	×
	time	×	√	×	√	√	×
$n = 50$	TABLE(3-22)	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(MLMS)	IDRGP(RMM)

$\alpha = 0.05$	Correct	×	√	×	√	×	√
	Swap	√	×	×	√	×	√
	time	×	√	×	√	√	×
$n = 50$	TABLE(3-23)	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(MLMS)	IDRGP(RMM)
$\alpha = 0.10$	Correct	×	√	×	√	×	√
	Swap	√	×	×	√	×	√
	time	×	√	×	√	√	×
$n = 50$	TABLE(3-24)	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(MLMS)	IDRGP(RMM)
$\alpha = 0.15$	Correct	×	√	√	×	×	√
	Swap	√	×	×	√	√	×
	time	×	√	×	√	√	×
$n = 100$	TABLE(3-28)	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(MLMS)	IDRGP(RMM)
	Correct	×	√	√	×	×	√

$\alpha = 0.05$	Swap	√	×	×	√	√	×
	time	×	√	×	√	√	×
$n = 100$	TABLE(3-29)	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(MLMS)	IDRGP(RMM)
$\alpha = 0.10$	Correct	×	√	×	√	×	√
	Swap	√	×	×	√	√	×
	time	×	√	×	√	√	×
$n = 100$	TABLE(3-30)	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(MLMS)	IDRGP(RMM)
$\alpha = 0.15$	Correct	×	√	×	√	×	√
	Swap	√	×	×	√	√	×
	time	×	√	×	√	√	×
$n = 300$	TABLE(3-34)	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(MLMS)	IDRGP(RMM)
$\alpha = 0.05$	Correct	×	√	√	×	×	√
	Swap	×	√	×	√	√	×

	time	×	√	×	√	√	×
$n = 300$	TABLE(3-35)	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(MLMS)	IDRGP(RMM)
	Correct	×	√	√	×	×	√
$\alpha = 0.10$	Swap	×	√	×	√	√	×
	time	×	√	×	√	√	×
$n = 300$	TABLE(3-36)	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(MLMS)	IDRGP(RMM)
	Correct	×	√	√	×	×	√
$\alpha = 0.15$	Swap	×	√	×	√	√	×
	time	×	√	×	√	√	×
$n = 500$	TABLE(3-37)	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(MLMS)	IDRGP(RMM)
$\alpha = 0.05$	Correct	---	---	√	×	×	√
	Swap	×	√	×	√	√	×
	time	×	√	×	√	√	×

$n = 500$ $\alpha = 0.10$	TABLE(3-38)	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(MLMS)	IDRGP(RMM)
	Correct	√	×	---	---	×	√
	Swap	×	√	×	√	√	×
	time	×	√	×	√	√	×
$n = 500$ $\alpha = 0.15$	TABLE(3-39)	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(MLMS)	IDRGP(RMM)
	Correct	×	√	×	√	×	√
	Swap	×	√	×	√	√	×
	time	×	√	×	√	√	×
$n = 1000$ $\alpha = 0.05$	TABLE(3-40)	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(MLMS)	IDRGP(RMM)
	Correct	---	---	×	√	×	√
	Swap	×	√	×	√	√	×
	time	×	√	×	√	√	×
$n = 1000$	TABLE(3-41)	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(MLMS)	IDRGP(RMM)

$\alpha = 0.10$							
	Correct	×	√	×	√	×	√
	Swap	×	√	×	√	√	×
	time	×	√	×	√	√	×
$n = 1000$ $\alpha = 0.15$	TABLE(3-42)	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(MLMS)	IDRGP(RMM)
	Correct	√	×	×	√	×	√
	Swap	×	√	×	√	√	×
	time	×	√	×	√	√	×
$n = 50$ $\alpha = 0.05$	TABLE(3-43)	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(MLMS)	IDRGP(RMM)
	Correct	----	---	√	√	√	√
	Swap	√	×	×	√	×	√
	time	×	√	×	√	√	×
$n = 50$	TABLE(3-44)	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(MLMS)	IDRGP(RMM)

$\alpha = 0.10$	Correct	×	√	×	√	×	√
	Swap	√	×	×	√	×	√
	time	×	√	×	√	√	×
$n = 50$	TABLE(3-45)	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(MLMS)	IDRGP(RMM)
$\alpha = 0.15$	Correct	×	√	×	√	√	√
	Swap	×	√	×	√	×	√
	time	×	√	×	√	√	×
$n = 100$	TABLE(3-49)	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(MLMS)	IDRGP(RMM)
$\alpha = 0.05$	Correct	√	√	√	√	√	√
	Swap	√	×	×	√	√	×
	time	×	√	×	√	√	×
$n = 100$	TABLE(3-50)	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(MLMS)	IDRGP(RMM)
$\alpha = 0.10$	Correct	√	√	×	√	√	√

	Swap	√	×	×	√	√	×
	time	×	√	×	√	√	×
$n = 100$	TABLE(3-51)	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(MLMS)	IDRGP(RMM)
$\alpha = 0.15$	Correct	√	√	×	√	√	√
	Swap	√	×	×	√	√	×
	time	×	√	×	√	√	×
$n = 300$	TABLE(3-55)	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(MLMS)	IDRGP(RMM)
$\alpha = 0.05$	Correct	√	√	×	√	√	√
	Swap	×	√	×	√	√	×
	time	×	√	×	√	√	×
$n = 300$	TABLE(3-56)	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(MLMS)	IDRGP(RMM)
$\alpha = 0.10$	Correct	√	√	×	√	√	√
	Swap	×	√	×	√	√	×

	time	×	√	×	√	√	×
$n = 300$	TABLE(3-57)	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(MLMS)	IDRGP(RMM)
	Correct	√	√	×	√	√	√
$\alpha = 0.15$	Swap	×	√	×	√	√	×
	time	×	√	×	√	√	×
$n = 500$	TABLE(3-58)	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(MLMS)	IDRGP(RMM)
$\alpha = 0.05$	Correct	√	√	√	√	√	√
	Swap	×	√	×	√	√	×
	time	×	√	×	√	√	×
$n = 500$	TABLE(3-59)	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(MLMS)	IDRGP(RMM)
$\alpha = 0.10$	Correct	√	√	√	√	√	√
	Swap	×	√	×	√	√	×
	time	×	√	×	√	√	×

$n = 500$ $\alpha = 0.15$	TABLE(3-60)	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(MLMS)	IDRGP(RMM)
	Correct	√	√	√	√	√	√
	Swap	×	√	×	√	√	×
	time	×	√	×	√	√	×
$n = 1000$ $\alpha = 0.05$	TABLE(3-61)	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(MLMS)	IDRGP(RMM)
	Correct	√	√	√	√	√	√
	Swap	×	√	×	√	√	×
	time	×	√	×	√	√	×
$n = 1000$ $\alpha = 0.10$	TABLE(3-62)	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(MLMS)	IDRGP(RMM)
	Correct	√	√	√	√	√	√
	Swap	×	√	×	√	√	×
	time	×	√	×	√	√	×
$n = 1000$	TABLE(3-63)	DRGP(MVE)	DRGP(RMVN)	IDRGP(MVE)	IDRGP(RMVN)	IDRGP(MLMS)	IDRGP(RMM)

$\alpha = 0.15$							
	Correct	√	√	√	√	√	√
	Swap	×	√	×	√	√	×
	time	×	√	×	√	√	×

جدول الاشارات الذي يوضح النتائج النهائية بالرموز وكالاتي :-

- **علامة √ : تبين الأفضلية في النتيجة عند المقارنة بين كل طريقتين**
- **علامة × : تبين عدم الافضلية في النتيجة عند المقارنة بين كل طريقتين**
- **اللون : يدل على الافضلية في الـ swap بين الطرق كافة**
- **اللون : يدل على السرعة في زمن الاحتساب بين الطرق كافة**

الملحق B

جدول (13) يبين فهرسة الصفوف التي تتضمن قيماً شاذة حسب طريقة التشخيص الفردي(Out , Hat, RMD) والكلي (Hadi, DRGP.MVE,DRGP.RMVN) لبيانات قوة انضغاط الخرسانة

Out	Hat	RMD	Hadi	DRGP.MVE	DRGP.RMVN	Out	Hat	RMD	Hadi	DRGP.MVE	DRGP.RMVN
1	--	--	1	--	--	--	32	32	--	32	32
--	3	3	3	3	3	--	--	33	--	33	33
4	4	4	4	4	4	--	34	34	34	34	34
--	5	5	5	5	5	--	35	35	--	35	35
--	--	--	6	--	--	--	36	36	36	36	36
--	7	7	7	7	7	--	--	--	--	--	39
9	--	--	9	--	--	--	--	40	--	40	40
--	13	13	13	13	13	--	--	--	41	--	--
15	--	--	15	--	--	--	42	42	42	42	42
--	--	--	16	--	--	43	43	43	43	43	43
--	18	18	--	18	18	--	--	44	--	44	44
--	--	20	--	20	20	--	--	48	--	48	48
--	--	21	--	21	21	--	--	51	--	51	51
--	24	24	--	24	24	--	--	52	--	52	52
--	25	25	--	25	25	--	--	--	54	--	--
--	26	26	--	26	26	57	57	57	57	57	57
--	27	27	27	27	27	--	--	59	--	59	59
--	--	28	--	28	28	--	--	61	--	61	61
--	31	31	--	31	31	--	--	62	--	62	62
--	64	64	--	64	64	--	--	--	131	--	--
--	--	65	--	65	65	--	132	--	--	132	132

--	--	66	--	66	66	--	144	--	--	144	144
--	67	67	67	67	67	--	146	146	--	146	146
--	--	--	69	--	--	--	149	149	--	149	149
--	70	--	70	70	70	--	--	--	153	--	--
--	--	--	72	--	--	--	--	--	154	--	--
--	--	--	73	--	--	--	155	--	--	155	155
75	75	--	75	75	75	--	--	--	156	--	--
77	77	77	77	77	77	--	--	--	158	--	--
--	--	--	78	--	--	--	--	--	161	--	--
--	--	--	79	--	--	--	167	--	--	167	167
--	80	80	80	80	80	--	169	169	--	169	169
--	--	--	81	--	--	--	172	172	172	172	172
82	--	--	82	--	--	--	177	--	--	177	177
--	--	--	85	--	--	--	--	--	--	182	182
--	86	--	86	86	86	225	225	225	225	225	225
--	--	--	88	--	--	--	226	226	226	226	226
--	--	--	91	--	--	--	227	227	227	227	227
--	--	--	93	--	--	--	228	228	228	228	228
--	98	--	--	98	98	--	229	229	229	229	229
--	100	100	100	100	100	--	--	--	295	--	--
--	103	103	103	103	103	--	--	--	334	--	--
--	109	--	--	109	109	--	--	--	357	--	--
115	--	--	115	--	--	--	--	--	358	--	--
--	121	--	--	121	121	--	381	381	--	381	381
--	123	123	--	123	123	382	--	--	382	382	--
--	126	126	--	126	126	384	--	--	384	--	--
--	385	385	--	--	--	--	--	--	585	--	--

--	--	--	396	396	--	--	--	604	--	604	604
--	397	397	--	--	--	--	605	605	605	605	605
--	--	--	401	--	--	--	--	610	--	610	610
--	--	--	405	--	--	611	611	611	611	611	611
--	406	--	406	--	--	--	--	616	--	616	616
--	--	--	--	459	459	--	617	617	617	617	617
--	--	--	477	--	--	--	--	620	--	620	620
478	--	--	478	--	--	--	621	621	621	621	621
--	--	--	489	--	--	--	--	622	--	622	622
--	--	--	490	--	--	--	623	623	623	623	623
--	500	500	--	500	--	--	--	--	667	--	--
--	501	--	--	501	501	--	--	--	687	--	--
--	502	502	502	502	502	--	--	--	700	--	--
--	503	503	503	503	--	747	--	--	747	--	--
--	504	504	--	504	--	--	--	--	--	755	755
--	505	--	505	505	505	--	756	756	--	756	756
--	506	--	--	506	506	--	757	757	--	757	757
--	507	--	--	507	507	--	--	763	--	763	763
--	--	--	510	--	--	--	--	--	764	--	--
--	--	--	513	--	--	--	--	769	--	769	769
--	--	--	515	--	--	--	770	770	770	770	770
--	--	--	526	--	--	--	--	--	775	--	--
--	--	--	527	--	--	--	--	792	--	792	792
--	554	--	554	554	554	--	793	793	793	793	793
--	560	--	--	560	560	--	--	795	--	795	795
--	--	--	564	--	--	--	796	796	--	796	796
--	--	--	582	--	--	--	798	798	--	798	798

--	799	799	799	799	799						
--	--	814	--	814	814						
--	815	815	815	815	815						
--	821	821	--	821	821						
--	--	824	--	824	824						
--	828	828	--	828	828						
--	829	829	829	829	829						
--	--	--	835	--	--						
--	874	874	--	874	874						
--	--	--	--	--	891						
--	892	--	--	--	--						
--	909	--	--	--	--						
--	933	--	--	933	933						
--	937	937	--	937	937						
--	--	--	--	--	954						
--	955	--	--	--	--						
--	972	--	--	--	--						

الملحق C

جدول (15) يبين فهرسة الصفوف التي تتضمن قيماً شاذة لطرق التشخيص الكلي (DRGP.MVE, DRGP.RMVN, IDRGP.MVE, IDRGP.RMVN) لبيانات القيمة السوقية لعينة من المصارف المسجلة في سوق العراق للأوراق المالية

DRGP.MVE	DRGP.RMVN	IDRGP.MVE	IDRGP.RMVN
1	1	1	1
4	--	--	--
6	6	6	6
7	7	7	7
11	--	11	--
13	13	13	--
16	16	16	16
23	--	23	23
25	--	--	--
29	--	--	--
30	--	--	--
31	--	--	--
33	33	33	33
34	34	34	34
35	35	35	35
36	--	36	36
37	--	--	--
40	40	40	40

الملحق D

جدول (16) يبين فهرسة الصفوف التي تتضمن قيماً شاذة حسب طرق التشخيص الكلي(DRGP.MVE, DRGP.RMVN ,IDRGP.MVE, IDRGP.RMVN)
لبيانات قوة انضغاط الخرسانة

DRGP.MVE	DRGP.RMVN	IDRGP.MVE	IDRGP.RMVN	DRGP.MVE	DRGP.RMVN	IDRGP.MVE	IDRGP.RMVN
3	3	3	3	42	42	42	42
4	4	4	4	43	43	43	43
5	5	5	5	44	44	44	44
7	7	7	7	48	48	48	48
9	--	--	--	51	51	51	51
13	13	13	13	52	52	52	52
15	--	--	--	57	57	57	57
18	18	18	18	59	59	59	59
20	20	20	20	61	61	61	61
21	21	21	21	62	62	62	62
24	24	24	24	64	64	64	64
25	25	25	25	65	65	65	65
26	26	26	26	66	66	66	66
27	27	27	27	67	67	67	67
28	28	28	28	70	70	70	70
31	31	31	31	75	75	75	75
32	32	32	32	77	77	77	77
33	33	33	33	80	80	80	80
34	34	34	34	86	86	86	86
35	35	35	35	98	98	98	98

36	36	36	36	100	100	100	100
--	39	39	--	103	103	103	103
40	40	40	40	109	109	109	109
121	121	121	121	504	--	--	--
123	123	123	123	505	505	--	505
126	126	126	126	506	506	506	506
132	132	132	132	507	507	507	507
144	144	144	144	554	554	--	--
146	146	146	146	560	560	--	--
149	149	149	149	604	604	604	604
155	155	155	155	605	605	605	605
167	167	167	167	610	610	610	610
169	169	169	169	611	611	611	611
172	172	172	172	616	616	616	616
177	177	177	177	617	617	617	617
182	182	182	182	620	620	620	620
225	225	225	225	621	621	621	621
226	226	--	--	622	622	622	622
227	227	--	--	623	623	623	623
228	228	228	--	755	755	755	755
229	229	229	229	756	756	756	756
381	381	381	381	757	757	757	757
382	--	--	--	763	763	763	763
384	--	--	--	769	769	769	769
459	459	459	459	770	770	770	770
500	--	--	--	792	792	792	792

501	501	--	501	793	793	793	793
502	502	--	--	795	795	795	795
503	--	--	--	796	796	796	796
798	798	798	798				
799	799	799	799				
814	814	814	814				
815	815	815	815				
821	821	821	821				
824	824	824	824				
828	828	828	828				
829	829	829	829				
874	874	--	874				
891	891	--	--				
933	933	--	--				
937	937	--	937				
954	954	--	--				

الملحقE

جدول (17) يبين فهرسة الصفوف التي تتضمن قيماً شاذة لطريقة التشخيص الفردي (RMD, Hat , Out) و الكلي (DRGP.RMVN,DRGP.MVE, Hadi)
لبيانات القيمة السوقية لعينة من المصارف المسجلة في سوق العراق للأوراق المالية

Out	Hat	RMD	Hadi	DRGP.MVE	DRGP.RMVN
--	--	--	--	1	1
--	--	--	--	4	--
--	--	--	--	6	6
--	--	--	--	7	7
--	--	--	--	11	--
--	--	--	--	13	13
--	--	16	--	16	16
--	--	--	--	23	--
--	24	--	24	--	--
25	--	--	25	25	--
29	--	--	29	29	--
30	--	--	30	30	--
--	31	--	31	--	--
--	--	33	--	33	33
--	34	34	34	34	34
--	35	35	35	35	35
--	--	--	--	36	--
--	40	40	40	40	40

Abstract

The statistical literature is worth by the research papers, technical reports and books that discussed the effect of the presence of outliers in dataset. That becauseof its presence violates the hypothesis of the normal distribution, which constitutes the basic distribution in the construction of statistical models. Therefore, great efforts have been made to diagnose outliers in the datasets of various scientific applications. Since the outliers were classified according to the type of the model used, in the multivariate model they are outliers, but with the multiple linear regression model, the term of outliers is concerned only with those values that appear in the residuals of the regression or the dependent variable. As for those outliers that appear in the matrix of explanatory variables, they were classified into Good Leverage Point and Bad Leverage point and High Leverage Point. Given the importance of this topic, many researchers have suggested ways to diagnose these outliers in different statistical models, but unfortunately, these methods depended on the individual diagnosis method. Therefore, a number of researchers recently proposed a number of grouped diagnostics, because individual diagnosis methods showed their inability to detect the two phenomena of masking and swamping. The masking phenomenon occurs when an outlier hides another outlier on the path of diagnosis that reveals the effective value without revealing the other one. As for the swamping means the specific method of diagnosis identifying at least one of cleaning observation as outliers. Hadi Potentials Measure method was presented as the first method for Grouped diagnoses in the statistical literature, forming a basis for many researchers who sought to develop and improve the performance of this method. DRGP (MVE) has improved the Hadi method's efficiency in reducing swamping, but only to some extent. However, this type of diagnosis could not be reduced by an IDRGP (MVE) method.

We have noticed two important matters through the previous two methods that were not taken into consideration in the Hadi method. The first is that the phenomenon of (Masking & Swamping) is possible when there are outliers affecting the random error term and in particular the linear regression model and since Hadi developed his method for treating the outliers In multivariate data, the DRGP (MVE) method and the IDRGP (MVE) method treat the matrix of explanatory variables (X) as a multivariate matrix that cannot be said about a normal multivariate matrix because the hypothesis of the common normal distribution of these variables has been violated due to the presence of outliers and

the process of fixing it in a multiple linear regression model makes it random and subject to measurement. The second thing we noticed is that the basis for building the DRGP (MVE) and IDRGP (MVE) algorithms is to calculate the Mahalanobis distance from the location and scale MVE matrix.

Based on the foregoing, we proposed in this thesis to deal with the matrix of explanatory variables of the multiple linear regression model as a matrix of multiple variables whose common distribution hypothesis was violated by outliers. On the other hand, we employed the position and measurement matrix (RMVN) that depends on the five-step multi-stage algorithm in three new algorithms which are (DRGP (RMVN), IDRGP (RMVN), and IDRGP (MM)) in order to obtain a good reduction in the diagnosis of incorrect diagnosis of the previous two methods, with respect to take into account the calculation time and the stability of each method.

All methods have been subjected to a number of simulation studies and two types of real data (engineering and financial), and the results showed that the proposed DRGP (RMVN) method is an efficient and effective method with large data volumes, while the IDRGP method (RMVN) is an efficient and effective method with small data volumes. It also turns out that the two proposed methods are somewhat quick, but they are not faster than the IDRGP (LMS) method according to the simulation results.In addition, we noticed that the IDRGP (MM) method succeeded through studying the simulation to reach all the correct outliers, but it suffered the first type of error a lot when adding an unacceptable number of misdiagnoses to its correct diagnosis.Furthermore it, we noticed that the IDRGP method (LMS) is a very fast and somewhat effective method, especially for the high sample size, but it was not stable, noting that both methods were not used with the data as being multivariate data because it included the dependent variable (Y) of the linear regression model in its data matrix.

Ministry of Higher Education & Scientific Research

University of Al-Qadisiyah

College of Administration & Economics

The Algorithem of Grouped Diagnostics for Multivariate Outliers by Using Location and Scale Matrix with Application

M.Sc. Thesis Statistics

Introduced by

Sawsan Abdulameer Haraje

Supervisor

Assist. Prof.Hassan S. Uraibi

April /2021

2021 A.D. **1442 A. H.**

Grouped Diagnostics for Multivariate Outliers by Using Location and Scale Matrix with Application

A thesis submitted to the Council of College of Administration & EconomicsUniversity of Al-QadisiyahIn Partial Fulfillment of the Requirement for the Degreeof Master of Science in Statistics

By

Sawsan Abdul Ameer

Supervised by

Assist. Prof. Hassan S. Uraibi (PhD)

2021 A.D. **1442 A. H.**