

الإخلال بافتراض خلو البيانات من القيم المتطرفة وأثره على استخدام تحليل الانحدار الخطي المتعدد في التنبؤ

إعداد

الباحث الرئيسي: عبد العزيز سعد عبد العزيز الخنعمي

E-MAIL: a210066449@gmail.com

الباحث المساعد: عبد الرحمن عبد الله احمد النفيعي

E-MAIL: aanofei@uqu.edu.sa

جامعة أم القرى، المملكة العربية السعودية، كلية التربية قسم علم النفس، تخصص إحصاء وبحوث

العام 1438 - 1439 هـ الموافق 2018م

الملخص

هدف البحث الحالي لدراسة أثر وجود القيم المتطرفة على نموذج الانحدار الخطي المتعدد ودقة التنبؤ ومعالجتها عبر دراسة أثرها على معلمات النموذج. استخدم الباحث المنهج الوصفي على عينة مكونة من 60 طالباً مأخوذة من مجتمع بلغ 204 طالباً هم طلاب كلية الهندسة للمرحلة الجامعية 1436-1437هـ. تم تطبيق نموذج تحليل الانحدار الخطي المتعدد وبعض الطرق والاختبارات الخاصة بتشخيص ومعالجة القيم المتطرفة. فقد تم الكشف عن القيم المتطرفة بالنظر عبر استخدام صندوق ويسكر Box Plot، كما تم استخدام عناصر القطر الرئيسي للمصفوفة Hat Matrix للكشف عن القيم المتطرفة في المتغيرات المستقلة، وباستخدام بواقي ستودنت المحذوفة Studentized Deleted Residuals للكشف عنها في المتغير التابع. أما تأثير القيم المتطرفة فتم الكشف عنه باستخدام الطرق التالية: مقياس DFFITS، ومقياس الأثر على معاملات الانحدار DFBEATS، ومقياس المسافة كوك COOK'S DISTANCE الخاص بقياس الأثر على كل معاملات الانحدار،



www.mecsaj.com/ar

ومقياس الأثر على الأخطاء المعيارية COVRATIO. قد تمت معالجة القيم المتطرفة باستخدام طرق الحذف ومتوسط البتر مع قياس قدرة النموذج قبل وبعد المعالجة عبر استخدام بعضاً من المؤشرات وهي: معامل التحديد R^2 ومعامل التحديد المعدل R_a^2 واحصاءة F ومتوسط الخطأ المطلق (MAE) Mean Absolute Deviation، وجذر متوسط مربع الخطأ Root Mean Square Error (RMSE)، ومتوسط الخطأ النسبي المطلق Mean Absolute Percentage Error (MAPE)، بالإضافة إلى معامل التساوي لثايل Theil's Inequality Coefficient. من أهم نتائج الدراسة ما يلي: هنالك تأثير معنوي للقيم المتطرفة على معاملات نموذج الانحدار الخطي المتعدد حيث تعمل على تضخيم قيم تباين الأخطاء MSE، وتقلل من قيمة معامل التحديد R^2 وقيمة F . كما توصل البحث إلى أن معالجة القيم المتطرفة بمتوسط البتر أفضل من حذفها. أوصت الدراسة بعدة بتوصيات أهمها ضرورة تشخيص وتقليل أثر القيم المتطرفة عند تطبيق نموذج تحليل الانحدار للحصول على نموذج أفضل.

الكلمات المفتاحية: الانحدار الخطي المتعدد- القيم المتطرفة- حذف القيم المتطرفة- متوسط البتر- مصفوفة القبة- بواقي ستودنت المحذوفة.

Abstract

The current research objective is to study the presence of outliers on multiple-linear regression model and its impact on prediction's accuracy and it show they are treated via the model's parameters. The qualitative approach was used in this study. A sample of 60 students selected from population of 204 students of Faculty of Engineering for academic year 1436-1437 AH. Multi-linear regression model and some methods and tests of diagnosis and processing the outliers were applied. Accordingly, outlier values were detected visually using Box and Whisker Plot or Box Plot.



www.mecs.com/ar

Main diagonal elements of the Hat Matrix were also used to detect outlier values in the independent variables, and in addition to that using Studentized Deleted Residuals to detect them in the dependent variable. For the effect of outliers, it was detected using the following methods: Difference in fits (DFFITS) measure, Difference of Beta Values (DFBEATS) measure (effect on regression coefficients), Cook's Distance measure for measuring the effect on all regression coefficients, and Covariance ratio (COVRATIO) to measure (effect on standard errors). Outlier values were processed using the methods of deleting and amputation average, with measuring the model leverage pre-and post-processing via using some indicators such as: R Square (R^2), R Square Adjusted and F Test, Mean Absolute Deviation (MAE), Root Mean Square Error (MRSE), and Mean Absolute Percentage Error (MAPE), in addition to Theil's Inequality Coefficient. The research concluded with various findings, the most important ones are: There is a significant impact of outlier values on parameters of the multi-linear regression model, where they amplify the Mean Squared Error (MSE), and reduce R^2 value and F value. The research has concluded also that the outlier values processing by amputation average is better than deleting them. The study recommended that: It is necessary to diagnosis and reduce the outliers effect when applying the regression analysis model to obtain better model.

Keywords: multi-linear regression- outliers - deleting the outliers- amputation average- Hat Matrix- Studentized Deleted Residuals.

1. المقدمة:

في العصر الحديث ومع تطور العلوم وتقدم الخبرة والمعرفة الإنسانية واستخدام الأساليب الكمية الإحصائية المتقدمة، أصبح التنبؤ Prediction أداة فعالة أكثر علمية ودقة في توقع الأحداث المستقبلية مما ساعد على زيادة استعداد الأفراد، والهيئات للتغيرات المتوقعة في الظواهر، والمجالات المختلفة والتي من ضمنها المجالات التربوية والنفسية. حيث ذكر (Montgomery, et al, 2008) أن التنبؤ له أهمية تمتد لكثير من المجالات التي تتمثل في الأعمال والصناعة والاقتصاد والبيئة والعلوم الطبية والعلوم الاجتماعية والسياسة "كما أشار إلى أن التنبؤ مهم جداً وخاصة في مجال التنبؤ بالأحداث المستقبلية وبالتالي يعد مدخل حاسم في العديد من أنواع عمليات التخطيط وصنع القرار.

و للتنبؤ أساليب عدة، ويعتبر نموذج الانحدار واحداً من أهمها، والذي يبنى على أن المتغير التابع يعتمد على متغيرات مستقلة تفسر سلوكه، حيث يستخدم الانحدار في دراسة العلاقة بين متغير تابع ومتغير مستقل (انحدار خطي بسيط) Simple Linear Regression، أو متغير تابع ومتغيرات مستقلة (انحدار خطي متعدد) Multiple Linear Regression والذي يعتبر من أهم أدوات الإحصاء التي تقوم ببناء نماذج إحصائية لتقدير العلاقة بين عدة متغيرات بحيث يعطينا نموذج عبارة عن معادلة رياضية توضح العلاقة بين هذه المتغيرات، وهذا ما تؤكد عليه موسوعة علم النفس التربوي حيث ذكر (Salkind & Rasmussen, 2008) أن أبحاث العلوم النفسية والتربوية عادة ما تكون مهمة باختبار العلاقة بين متغير أو أكثر ومن ثم استخدام هذه العلاقة في التنبؤ بسلوك مستقبلي والانحدار هو معالج نمذجة هذه العلاقة بين المتغيرات.

وبالتالي فإن تحليل الانحدار المتعدد هو واحد من أهم الموضوعات والأساليب الإحصائية والتي لا غنى عنها للباحثين في المجالات العلمية المختلفة، ومن ثم فإنه يتعين على المهتمين ببناء النماذج بواسطة الانحدار المتعدد الإلمام الكافي بافتراضات نموذج الانحدار الخطي المتعدد، وهو ما يحاول الباحث طرقة وتوضيحه في هذا البحث.

2. مشكلة البحث وتساؤلاته

تتبع أهمية التنبؤ من كونه محاولة لتوقع الأحداث المستقبلية بالشكل الذي يعمل على زيادة استعداد الأفراد والمنظمات للتغيرات المتوقعة في الظواهر والمجالات المختلفة،



www.mecsaj.com/ar

المجلة الإلكترونية الشاملة متعددة المعرفة لنشر الأبحاث العلمية والتربوية (MECSJ)

العدد الخامس (أيلول) 2018

والتي من ضمنها المجالات التربوية والتعليمية. وبالتالي يتطلب ذلك أن تكون لأساليب القياس والتنبؤ، أساليباً علمية كمية وهي تلك التي يزودنا بها الإحصاء كعلم للقياس، وتحديدًا قياس التنبؤ بالتغيرات التي على ضوءها تتخذ قرارات المؤسسة، وهنا تحديدًا تبرز أهمية التنبؤ والتي تتمثل في ضمان كفاءة المؤسسة (التعليمية / التربوية في مجال البحث) ومرونتها لتحقيق وظائفها التي يتطلع إليها المجتمع، بالإضافة إلى معرفة احتياجاتها، كما إن التنبؤ يسهم في الحد من المخاطر والمعوقات التي تواجه المؤسسات التربوية والتعليمية أو العاملين بها والمنتمين إليها.

وأحد أهم وسائل التنبؤ بالتغير في الظواهر هو أسلوب تحليل الانحدار في دراسة العلاقة بين متغير تابع والمتغيرات المؤثرة فيه، والانحدار المتعدد هو أحد أوسع الطرق انتشاراً في دراسة الظواهر النفسية والتربوية، وذلك لأن السلوك الإنساني على درجة عالية من التعقيد بحيث تتدخل وتؤثر عليه عدة متغيرات في آن واحد. ولأن استخدام الانحدار الخطي المتعدد يتطلب توفر عدد من الافتراضات والشروط الواجب التي يجب التحقق منها، ومن هذه الشروط شرط خلو البيانات من القيم المتطرفة والتي تمثل نطاق هذا البحث ومجاله. وبصورة أكثر دقة فإن البحث الحالي يسعى إلى الإجابة عن التساؤل الرئيس التالي: هل تتأثر دقة تنبؤ الانحدار المتعدد باحتواء البيانات على قيم متطرفة؟، وهل تتحسن الدقة عند معالجة القيم المتطرفة؟

و الذي يتفرع عنه التساؤلات التالية:

1. هل تختلف قيم احصاءات ومؤشرات دقة تنبؤ الانحدار المتعدد عن القيم المعيارية عند احتواء البيانات على قيم متطرفة؟
2. هل تتحسن قيم مؤشرات دقة تنبؤ الانحدار المتعدد وتصبح ضمن مدى القيم المعيارية بعد معالجة القيم المتطرفة؟

أهداف البحث

1. دراسة تأثير وجود القيم المتطرفة على مؤشرات فاعلية التنبؤ لنموذج الانحدار وتأثيرها على مصداقيته.
2. دراسة وضع نموذج الانحدار الخطي المتعدد ومؤشرات فاعلية التنبؤ قبل معالجة القيم المتطرفة وبعد معالجتها ومقارنة النتائج.

أهمية البحث

تنبثق أهمية هذا البحث مع بروز نموذج الانحدار المتعدد كوسيلة تنبؤ إحصائية متقدمة تعطي وتقدم خدمات للباحثين في شتى ومختلف المجالات الحياتية، ولهذا فإن الباحث يأمل أن تكون لهذا البحث أهمية نظرية وأهمية تطبيقية تتمثل في ما يلي:

الأهمية النظرية:

تتمثل الأهمية النظرية للبحث في كونها تتعلق بأحد محددات بناء نموذج الانحدار الخطي المتعدد ألا وهو خلو البيانات من القيم المتطرفة، وما تأثيرها -إن وجدت- على نموذج الانحدار الخطي المتعدد؟، وهل سيكون التأثير إيجابي أم سلبي؟، وماذا سيتغير في بنية النموذج وقدرته التنبؤية بعد معالجة القيم المتطرفة؟. كما يأمل الباحث في أن يكون هذا البحث إضافة علمية تعمل على توجيه انتباه الباحثين إلى ضرورة أخذ الاحتياطات اللازمة التي تضمن بناء نموذج انحدار دقيق في تنبؤاته.

الأهمية التطبيقية:

يقدم هذا البحث جانباً تطبيقياً مهماً يتمثل في تناوله لطرق اكتشاف القيم المتطرفة وطرق معالجتها ومن ثم المفاضلة بينها ومعرفة أي هذه الطرق أفضل من غيرها.

حدود البحث

الحدود الموضوعية: أثر انتهاك افتراض خلو البيانات من القيم المتطرفة على دقة التنبؤ باستخدام الانحدار الخطي المتعدد مع حلول مقترحة

الحدود المكانية: طلاب كلية الهندسة للمرحلة الجامعية بجامعة أم القرى عام 1436-1437 هـ

الحدود الزمانية: العام 1438-1439/2018م.

- أسلوب تحليل الانحدار المتعدد لبناء نموذج الانحدار بحسب متغيرات البحث وهي x_1 ، x_2 ،

x_3 ، x_4



www.mecsaj.com/ar

المجلة الإلكترونية الشاملة متعددة المعرفة لنشر الأبحاث العلمية والتربوية (MECSJ)
العدد الخامس (أيلول) 2018

- مؤشرات فاعلية التنبؤ المتمثلة في: متوسط الخطأ المطلق، والجذر التربيعي لمتوسط مربع الخطأ، ومتوسط الخطأ النسبي المطلق، ومعامل التساوي لثايل.
- طرق الكشف عن القيم المتطرفة تمثلت في: صندوق ويسكر ومصفوفة القبة للمتغيرات المستقلة وبواقي ستودنت المحذوفة للمتغير التابع.
- أساليب معالجة القيم المتطرفة تمثلت في طريقتي: الحذف ومتوسط البتر.

فروض البحث:

- يسعى البحث الحالي للتحقق من الفرضيتين الرئيسيتين التاليتين من خلال الفرضيات الفرعية المرتبطة بكل فرضية:
- الفرضية الأولى:** تختلف قيم مؤشرات دقة تنبؤ الانحدار المتعدد عن القيم المعيارية عند احتواء البيانات على قيم متطرفة.
- الفرضية الثانية:** تتحسن قيم مؤشرات دقة تنبؤ الانحدار المتعدد وتصبح ضمن مدى القيم المعيارية بعد معالجة القيم المتطرفة.

مصطلحات البحث

القيم المتطرفة Outliers: يعرفها إسماعيل (2001: 251) بأنها مجموعة قليلة من المشاهدات تبعد قيمها بصورة كبيرة عن باقي المشاهدات في العينة.

الانحدار الخطي المتعدد **Multiple Linear Regression:**

يعرف زايد (2011: 286) نماذج الانحدار المتعدد بأنها "نماذج تصف العلاقة بين متغير ما يطلق عليه المتغير التابع ومتغير أو أكثر يطلق عليها المتغيرات المستقلة أو المفسرة".

التنبؤ Prediction، كما يعرفه عودة والخليلي (2000: 460) بأنه طريقة علمية في البحث للحصول على المعرفة، أي أنه يهتم بتقدير بيانات غير معروفة مبنية على بيانات معروفة وذات صلة بالظاهرة المدروسة.

إجراءات البحث وطرائقه

يتناول الباحث في هذا الجزء الإجراءات المنهجية التي تتمثل في منهج البحث، مجتمع البحث وعينته وكيفية جمع البيانات والأساليب الإحصائية المستخدمة في تحليل وتفسير بيانات الدراسة بغرض الإجابة عن أسئلتها البحثية وفرضياتها، كما يلي:

منهج البحث

استخدم الباحث المنهج الوصفي الذي يفسر الظواهر المعاصرة للإجابة على تساؤلات بحثه والمتعلقة بدراسة كيفية اكتشاف القيم المتطرفة ودراسة أثرها على نموذج الانحدار الخطي المتعدد، وذلك لأن الهدف الرئيسي من هذا المنهج هو وصف بيانات وخصائص ما هو قيد الدراسة. ويأتي استخدام المنهج الوصفي كونه المنهج المناسب لدراسة أثر القيم المتطرفة على نموذج الانحدار ثم مقارنة هذا التأثير مع النتائج بعد معالجة القيم المتطرفة بالأسلوب الأمثل.

مجتمع وعينة البحث

تكون مجتمع البحث من طلاب المرحلة الجامعية في كلية الهندسة بجامعة أم القرى للعام الدراسي 1436-1437هـ، والذين بلغ عددهم 204 طالباً توزعوا في تخصصات الهندسة المدنية والهندسة الميكانيكية والعمارة الإسلامية، وقد تم الحصول على بياناتهم عن طريق عمادة تقنية المعلومات، والمتمثلة في تخصصاتهم ومعدلاتهم الفصلية لأربعة فصول دراسية هي المعدلين الفصليين الأول والثاني من العام الدراسي 1436-1437هـ، والمعدل الفصليين الأول والثاني من العام 1437-1438هـ، والمعدل التراكمي لكل طالب.

وقد تم استخدام برنامج SPSS في اختيار عينة عشوائية من مجتمع البحث تمثلت في عدد 60 مفردة بحيث يضمن الباحث وجود قيم شاذة في العينة بما يحقق أهداف البحث، كما تم بواسطة برنامج SPSS الحصول على بعض المقاييس الوصفية التي تصف طبيعة بيانات البحث والعينة في جدول (1):

جدول (1) بعض المقاييس الإحصائية الوصفية لبيانات البحث

المعدل التراكمي Y	معدل الفصل الثاني عام 1438-1437 X ₄	معدل الفصل الأول عام 1438-1437 X ₃	معدل الفصل الثاني عام 1437-1436 X ₂	معدل الفصل الأول عام 1437-1436 X ₁	المتغيرات المقاييس
2.96	3.15	2.98	2.945	2.98	المتوسط الحسابي
2.99	3.18	3.11	3.08	3	الوسيط
0.41	0.49	0.54	0.52	0.57	الانحراف المعياري
1.9	2.05	1.72	1.48	1.76	أقل قيمة
3.93	4	3.96	3.95	3.97	أكبر قيمة

اجراءات البحث

تم الحصول على بيانات العينة المستخدمة في هذا البحث من عمادة تقنية المعلومات بجامعة أم القرى وذلك لتطبيق الأسس النظرية والأدبيات المتعلقة بهذا البحث، حيث بدأ الباحث في اختيار عينة عشوائية حجمها 60 طالباً ثم بدأ في التحقق من توفر افتراضات نموذج الانحدار الخطي المتعدد في البيانات باستثناء القيم المتطرفة، ومن ثم تقدير معالم نموذج الانحدار الخطي المتعدد، ثم قام الباحث بتقييمها في ضوء القيم المعيارية والمؤشرات التي في التساؤل الأول، ثم بعد ذلك الكشف عن القيم المتطرفة بالأساليب المختلفة، وبعد تحديدها تمت معالجتها ومن ثم التحقق من توفر افتراضات نموذج الانحدار الخطي المتعدد في البيانات ومن ثم تقدير معالم النموذج مجدداً وتقييمها في ضوء القيم المعيارية السابقة ثم قارن الباحث النتائج قبل المعالجة بها بعد المعالجة.

المعالجة الإحصائية لبيانات البحث:

استخدم الباحث البرامج SPSS، Eviews، Excel لتحليل بيانات البحث وذلك بغرض الاستفادة من كامل امكانيات كل برنامج مما لا يتوفر في غيره حتى يتم إظهار نتائج البحث



www.mecs.com/ar

بصورة وافية. وقد تم استخدام أسلوب الانحدار الخطي المتعدد عبر إدخال كافة المتغيرات دفعة واحدة ومن ثم تطبيق المعالجات الإحصائية التالية:

- معامل التحديد ويتم حسابه عبر المعادلة التالية:

$$R^2 = \frac{SST-SSE}{SST} = 1 - \frac{SSE}{SST} \dots\dots\dots(1)$$

- معامل التحديد المعدل R_a^2 ويتم حسابه عبر المعادلة التالية:

$$R_a^2 = 1 - \left[\frac{(n-1)}{n-(k+1)} \right] \left(\frac{SSE}{SS_{yy}} \right) \dots\dots\dots(2)$$

$$= 1 - \left[\frac{(n-1)}{n-(k+1)} \right] (1 - R^2) = 1 - \left[\frac{(1 - R^2)(n-1)}{n-k-1} \right]$$

- جولد فيلد كواندت للكشف عن تجانس تباين حد الخطأ العشوائي عبر حساب F

- اختبار درين-واتسون (D-W) للاستقلال الذاتي للبواقي بالقانون

$$d = \frac{2 \sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2} \dots\dots\dots(3)$$

- اختبار جارك بييرا ويحسب من القانون التالي

$$\text{Jarque-Bera} = \frac{n}{6} \left(s^2 + \frac{(k-2)^2}{4} \right) \dots\dots\dots(4)$$

ولقياس مؤشرات فاعلية التنبؤ تم استخدام المؤشرات والأساليب التالية

- متوسط الخطأ المطلق (MAE) Mean Absolute Error:

$$MAE = \frac{\sum_{i=1}^t |y - \hat{y}_i|}{n} \dots\dots\dots(5)$$

- الجذر التربيعي لمتوسط مربع الخطأ Root Mean Square Error:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y - \hat{y}_i)^2} \dots\dots\dots (6)$$

- متوسط الخطأ النسبي المطلق (MAPE) mean absolute percentage error:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \dots\dots\dots (7)$$

- معامل التساوي لثايل Theil's Inequality Coefficient:

$$U = \sqrt{\frac{\sum_{i=1}^{n-1} (P_i - A_i)^2}{\sum_{i=1}^{n-1} A_i^2}} \dots\dots\dots (8)$$

و لاكتشاف القيم المتطرفة تم استخدام الأساليب التالية

- الرسم الصندوقي

- مصفوفة Hat Matrix لاكتشاف القيم المتطرفة في المتغيرات المستقلة

- بواقي ستودنت لاكتشاف القيم المتطرفة في المتغير التابع بالقانون

$$d^* = e_i \left(\frac{n-P-2}{RSS(1-h_{ii})-e_i^2} \right)^{1/2} \dots\dots\dots (9)$$

و لقياس أثر القيم المتطرفة على القيم المقدرة تم استخدام المقاييس التالية

- COVRATIO ،DFBETAS ،DFFITS ،مقياس كوك،

4. النتائج

تناول الباحث دراسة تأثير وجود القيم المتطرفة على مؤشرات فاعلية التنبؤ لنموذج الانحدار الخطي المتعدد وتأثيرها على مصداقيته، بالإضافة الى دراسة وضع نموذج الانحدار الخطي المتعدد ومؤشرات فاعلية التنبؤ قبل معالجة القيم المتطرفة وبعد معالجتها ومقارنة النتائج. وذلك عبر صياغة عدد من التساؤلات والفروض.

أولاً- عرض نتائج الإجابة عن التساؤل الأول

1. هل تختلف قيم احصاءات ومؤشرات دقة تنبؤ الانحدار المتعدد عن القيم المعيارية عند احتواء البيانات على قيم متطرفة؟، واختبار فرضيته التي تنص على "تختلف قيم مؤشرات دقة تنبؤ الانحدار المتعدد عن القيم المعيارية عند احتواء البيانات على قيم متطرفة"

قام الباحث بإجراء تحليل الانحدار المتعدد لتقدير معاملات النموذج وتم استخراج معادلته، كما سيرد فيما يلي:

تحليل الانحدار المتعدد قبل معالجة القيم المتطرفة:

تقدير معاملات النموذج:

باستخدام برنامج SPSS تم إجراء تحليل الانحدار المتعدد على بيانات العينة وبموجبه تم الحصول على قيم المعلمات $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$ ، وبالتالي تكون معادلة الانحدار على الصورة التالية:

$$\hat{Y} = 0.257 + 0.248 X_1 + 0.142 X_2 + 0.258 X_3 + 0.246 X_4$$

و تم الحصول على قيم معامل الارتباط R ومعامل التحديد R^2 ومعامل التحديد المعدل R^2_a وإحصاء F ومربع مجموع الأخطاء MSE في جدول (2) التالي:

جدول (2) معاملات النموذج قبل المعالجة

الاحصاءة - المؤشر	قيمتها
معامل التحديد R	0.96
معامل التحديد R^2	0.935
معامل التحديد المعدل R^2_a	0.93
F	197.101
مربع مجموع الأخطاء MSE	0.012

اختبار المعنوية الكلية للنموذج:

تم اختبار معنوية النموذج المقدر كاملاً باستخدام اختبار $F(ANOVA)$ حيث بلغت قيمتها (197.101) ودالة إحصائياً عند مستوى المعنوية (0.01)، وبناءً على ذلك نرفض الفرض الصفري القائل بأن نموذج الانحدار غير معنوي، وهذا يدل على أن هنالك تأثيراً معنوياً من المتغيرات المستقلة على المتغير التابع.

اختبار معنوية معاملات الانحدار β_i :

جدول (3) قيم المعنوية لمعاملات الانحدار

مستوى المعنوية	قيم المعاملات	معاملات الانحدار
0.015	0.257	β_0
0.000	0.248	β_1
0.001	0.142	β_2
0.000	0.258	β_3
0.000	0.246	β_4

ويلاحظ من الجدول أعلاه أن معنوية المعلمات $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$ دالة لأنها جميعاً أصغر من 0.05 وهذا يعني أن جميع المتغيرات المستقلة كلاً على حده يؤثر على المتغير التابع Y .

اختبار القدرة التفسيرية للنموذج:

للحكم على القدرة التفسيرية للنموذج تم حساب معامل التحديد R^2 ومعامل التحديد المعدل R^2_a من برنامج PSSS حيث بلغت قيمتها: (0.935) و (0.93) على التوالي، مما يعني أن المتغيرات المستقلة المضمنة في النموذج تفسر ما نسبته تزيد على 93% من التغيرات في المتغير التابع وأن أقل من 6% من التغيرات ترجع لعوامل أخرى منها الخطأ العشوائي.

اختبار النموذج من الناحية القياسية:

بعد تقدير النموذج واختباره احصائياً ومعرفة قدرته التفسيرية تم اختباره من الناحية القياسية التي توضح سلامة الاختبارات الاحصائية وتحدد درجة الثقة بها ويشمل هذا الجزء اختبار النموذج من عدة نواحي تتمثل في اختبار تجانس تباين الخطأ العشوائي، واختبار الارتباط الخطي المتعدد بين المتغيرات المستقلة، واختبار الاستقلال الذاتي للبواقي وذلك بحسب الاجراءات التالية:

اختبار تجانس تباين حد الخطأ العشوائي: تم استخدام اختبار جولدفيلد - كواندت للكشف عن تجانس تباين حد الخطأ العشوائي واختبار الفرضية الصفرية التي تنص على أن هنالك تجانس في تباين الخطأ كما يلي:

1. تم ترتيب المشاهدات تصاعدياً حسب المتغير X_1 لأنه الأكثر معنوية وتأثيراً بحسب قيمة
Beta = 0.339

2. تم استبعاد 20% من المشاهدات في المنتصف

3. تبقى لدينا عدد $(n-m)/2 = 24$ في كل مجموعة

4. باستخدام برنامج SPSS تم حساب مجموع مربعات الخطأ SSE لكل مجموعة فكانت النتيجة كالتالي:-

$$SSE_1 = 0.165, SSE_2 = 0.304$$

5. نحسب قيمة F:

$$V_1 = V_2 = \frac{(n-d)}{2} - p - 1 = \frac{60-12}{2} - 4 - 1 = 19$$

$$F = \frac{0.304/19}{0.165/19} = 1.8$$

وهي قيمة F المحسوبة، وبمقارنتها بـ $F_{0.05,19,19}$ الجدولية $2.12 =$ نجد أن F المحسوبة أقل من F الجدولية وبموجبه نقبل الفرض الصفري القائل بأن هنالك تجانس أو ثبات في تباين الأخطاء.

اختبار الارتباط الخطي المتعدد بين المتغيرات المستقلة

تم إيجاد قيم معامل تضخم التباين VIF وقيم التسامح Tolerance وقيمة رقم الحالة من SPSS، وعرضت النتائج في جدول (4) التالي:

جدول (4) نتائج اختبارات الكشف عن الارتباط الخطي المتعدد بين المتغيرات المستقلة للنموذج المقدر قبل معالجة القيم المتطرفة

معدل الفصل الثاني عام 1438-1437 X ₄	معدل الفصل الأول عام 1438-1437 X ₃	معدل الفصل الثاني عام 1437-1436 X ₂	معدل الفصل الأول عام 1437-1436 X ₁	الثابت	المتغيرات Variables
2.211	3.047	2.394	1.931	-	قيم معامل التضخم VIF
0.452	0.328	0.426	0.518	-	قيم التسامح Tolerance
0.006	0.013	0.015	0.019	4.94	القيم الكامنة Eigenvalues
29.8	19.28	18.15	16.2	1	رقم الحالة CI
$CN = \sqrt{\frac{4.955}{0,006}} = 28.69$					مؤشر الحالة CN

يلاحظ من الجدول (4) أن قيم VIF لجميع المتغيرات المستقلة أقل من 10، كما إن قيم Tolerance لجميع المتغيرات المستقلة أكبر من 0.1، كذلك رقم الحالة Condition Number أعطى القيمة CN=28.69 وهي قيمة أقل من 30 إلى 100، وهذه جميعها مؤشرات دالة على عدم وجود ارتباط خطي بين المتغيرات.

ثالثاً: اختبار الاستقلال الذاتي للبواقي

تم استخدام اختبار (Durbin-Watson) للحكم على مدى وجود استقلال ذاتي بين بواقي النموذج، ومن خلال برنامج SPSS تم استخراج قيمة $D-W = 2.095$ وبالرجوع إلى جدول Durbin-Watson بعدد متغيرات مستقلة $P=4$ و $n=60$ مشاهدة كانت قيمتي $D-W$ العليا



www.mecs.com/ar

المجلة الإلكترونية الشاملة متعددة المعرفة لنشر الأبحاث العلمية والتربوية (MECSJ)
العدد الخامس (أيلول) 2018

والدنيا $d_U=1.74$ و $d_L=1.53$. وحيث أن قيمة D-W المحسوبة = 2.095 وتقع في منطقة القبول فإننا نقبل الفرض الصفري القائل بعدم وجود ارتباط ذاتي، أي أن البواقي مستقلة ذاتياً.

اختبار التوزيع الطبيعي للبواقي

أولاً: اختبار جارك بيرا Jarque-Bera

تم استخراج قيمة هذا الاختبار بواسطة برنامج EView9 حيث بلغت قيمة اختبار جارك بيرا قبل معالجة القيم المتطرفة 4.067 عند مستوى المعنوية (0.13) وهي أكبر من مستوى المعنوية المعتمد في هذا البحث وهو 0.05 وعليه نقبل الفرض الصفري القائل بأن البواقي تتفق مع معامل الالتواء.

ثانياً: اختبار كولمجروف سميرونوف & اختبار ويلك شابيرو:

في هذا الاختبار تم دراسة توزيع البواقي للنموذج المقدر قبل اكتشاف البيانات المتطرفة ومعالجتها، باستخدام برنامج SPSS حيث تم الحصول على النتائج المعروضة في جدول (5) التالي:

جدول (5) نتيجة اختباري كولمجروف سميرونوف & ويلك شابيرو

اختبار ويلك شابيرو			اختبار كولمجروف سميرونوف		
مستوى المعنوية	درجات الحرية	الاحصاء	مستوى المعنوية	درجات الحرية	الاحصاء
Sig	df		Sig	df	
0.00	60	0.945	0.02	60	0.124

ويلاحظ من الجدول أعلاه أن مستويات الدلالة لكلا الاختبارين أقل من 0.05 وبالتالي فإننا نرفض الفرض الصفري ونقبل الفرض البديل والذي يعني أن البواقي لا تتوزع طبيعياً.

والنتيجة أعلاه تختلف عن اختبار جارك بيرا، ومن الجدير ذكره أن أدبيات الإحصاء تشير إلى أنه ليس بالضرورة أن تتفق جميع الاختبارات على نتيجة واحدة،



www.mecsj.com/ar

المجلة الإلكترونية الشاملة متعددة المعرفة لنشر الأبحاث العلمية والتربوية (MECSJ)
العدد الخامس (أيلول) 2018

وإنما يمكن استخدام نتائجها ليدعم بعضها بعضاً، وهذا ما سوف نلاحظه لاحقاً حيث سنتحسن قيمة اختبار جارك بيرا إيجاباً بارتفاعها بعد معالجة القيم الشاذة.

خامساً: اختبار فاعلية التنبؤ للنموذج:

لحساب مؤشرات فاعلية التنبؤ المتمثلة في متوسط الخطأ المطلق Mean Absolute Error (MAE) من الصيغة (2-23)، والجذر التربيعي لمتوسط مربع الخطأ Root Mean Square Error من الصيغة (2-24)، ومتوسط الخطأ النسبي المطلق mean absolute percentage error (MAPE) من الصيغة (2-25)، ومعامل التساوي لثايل Theil's Inequality Coefficient من الصيغة (2-26)، وذلك لمقارنتها لاحقاً مع النموذج الجديد بعد كشف القيم المتطرفة ومعالجتها، ولتحقيق هذا الغرض نستخدم برنامج Eview9 للحصول على القيم التالية للمؤشرات المذكورة:

$$\text{Root Mean Square Error (RSME)} = 0.106097$$

$$\text{Mean Absolute Error (MAE)} = 0.084503$$

$$\text{Mean absolute percentage error (MAPE)} = 2.8917$$

$$\text{Theil's Inequality Coefficient} = 0.017748$$

وكل هذه القيم تمثل مؤشرات مأخوذة من معادلة النموذج الأساسية قبل اجراءات الكشف عن القيم المتطرفة وطرق حلها. وهذه النتائج أجابت عن التساؤل الأول حيث نلاحظ اختلاف قيم المؤشرات المختلفة عن القيم المعيارية لكل مؤشر.

ثانياً - عرض نتائج الإجابة عن التساؤل الثاني

للإجابة عن هذا التساؤل والذي ينص على "هل تتحسن قيم مؤشرات دقة تنبؤ الانحدار المتعدد وتصبح ضمن مدى القيم المعيارية بعد معالجة القيم المتطرفة؟" واختبار فرضه الذي ينص على "تتحسن قيم مؤشرات دقة تنبؤ الانحدار المتعدد وتصبح ضمن مدى القيم المعيارية بعد معالجة القيم المتطرفة" تم تشخيص القيم الشاذة في المتغيرات المستقلة والمتغير التابع، وتم الكشف عن تأثيرها ومعالجتها.

تشخيص القيم المتطرفة في المتغيرات المستقلة والمتغير التابع وأثرها ومعالجتها:

استخدام صندوق ويسكر Pox Plot في اكتشاف القيم المتطرفة:

وهو عبارة عن رسم صندوق يشتمل على حدود للقيم الطبيعية فإذا تخطت القيم هذه الحدود تعتبر قيم شاذة ويعتمد هذا الصندوق في رسمه على الربيعات المتمثلة في الوسيط (الربيع الثاني) Q_2 والربيع الأدنى (الربيع الأول) Q_1 والربيع الأعلى (الربيع الثالث) Q_3 ويتم اضافته قيمة الي الربيع الثالث وهي (Q_3-Q_1) وطرح هذه القيمة من الربيع الأدنى وبذلك نكون كونا الصندوق، ويمكن الحصول على رسم المتغيرات كلا على حدة مباشرة من SPSS كما يلي: معدل الفصل الأول (X_1) 1437-1436، ومعدل الفصل الثاني (X_2) 1437-1436، ومعدل الفصل الأول (X_3) 1438-1437، ومعدل الفصل الثاني (X_4) 1438-1437، والمعدل التراكمي (Y) .

وتبين من الأشكال اكتشاف قيمة متطرفة واحدة هي القيمة رقم 39 في المتغير المستقل X_2 وهذا يدل على ضعف صندوق ويسكر في اكتشاف القيم المتطرفة، ولهذا سنحتاج إلى طرق تشخيص أخرى مساندة لهذا الأسلوب.

استخدام مصفوفة Hat Matrix في اكتشاف القيم المتطرفة في المتغيرات المستقلة:

بواسطة برنامج EView تم استخراج قيم مصفوفة القبعة Hat Matrix ثم عن طريق بعض المعالجات والعمليات الحسابية بواسطة برنامج Excel وفقاً لمعيار بيلسلي المقدرة قيمته بـ (0.166667) ، حيث تم إيجاد الفروق بينه وبين قيم مصفوفة القبعة، يتضح وجود (5) من القيم المتطرفة في المتغيرات المستقلة كما موضح بالجدول رقم (7) في ملاحق الدراسة الرئيسية: وبذلك تكون القيم المتطرفة في المتغير المستقل هي الحالات 13، 38، 39، 47، 49.

استخدام بواقي ستودنت لاكتشاف القيم المتطرفة في المتغير التابع:

تم إيجاد القيم المتطرفة في المتغير التابع باستخدام بواقي ستودنت بالصيغة عن طريق برنامج Eview 9، كما أشرنا إليها في ملاحق الدراسة الرئيسية جدول رقم (8)، حيث تمت مقارنة القيمة المطلقة لبواقي ستودنت مع قيمة t بإيجاد الفرق بينهما ثم تحديد القيمة الموجبة والتي مثلت الحالة المتطرفة، وهي الحالة 34.

قياس أثر القيم المتطرفة على القيم المقدرة قبل المعالجة:

تمت دراسة تأثير هذه القيم باستخدام مقاييس DFBETAS، DFFITS، مقياس كوك، COVRATIO، ومن خلال التطبيق تبين أن القيم المتطرفة تؤثر على المقاييس كما في جدول رقم (9) المرفق في جداول الدراسة الرئيسية. كما يتضح النتائج بالجدول كثرة نتائج المقاييس المتأثرة بالقيم المتطرفة، حيث نلاحظ عمق التأثير للحالات 13، 38، 39، 47، 49 وبالأخص الحالات 34 و38 و47 والتي سنشاهد عمق تأثيرها بعد المعالجة لاحقاً في تحسين المعاملات والمؤشرات المختلفة.

معالجة القيم المتطرفة:

تم معالجة هذه القيم بطريقتين: أولاً إما بحذف القيمة المتطرفة، أو استخدام متوسط البتر

أولاً - معالجة القيم المتطرفة بواسطة الحذف

تم معالجة القيم المتطرفة بواسطة الحذف لكل من عبارة من العبارات الموضحة أعلاه وجاءت النتائج كما هو مبين بالجدول رقم (7) أدناه، حيث بينا فيه المفاضلة بين نتائج النماذج المقدرة بعد الحذف:

جدول (7) المقارنة بين النماذج المقدرة بعد حذف القيم المتطرفة

MSE	R ²	F	β_4	β_3	β_2	β_1	β_0	الحالة المحذوفة
0.012	0.93	197.101	0.246	0.258	0.142	0.248	0.257	قبل الحذف
0.012	0.934	191.752	0.250	0.253	0.145	0.248	0.250	حذف مشاهدة 13
0.012	0.936	198.448	0.211	0.293	0.133	0.259	0.251	حذف مشاهدة 38
0.012	0.932	185.073	0.246	0.258	0.148	0.246	0.244	حذف مشاهدة 39
0.012	0.935	197.101	0.246	0.258	0.142	0.248	0.257	حذف مشاهدة 47
0.012	0.935	193.515	0.250	0.251	0.153	0.243	0.244	حذف مشاهدة 49
0.011	0.94	223.381	0.249	0.257	0.136	0.262	0.233	حذف مشاهدة 34
0.011	0.95	228.237	0.194	0.290	0.141	0.286	0.216	حذف مشاهدة 34 و38 و47



www.mecs.com/ar

المجلة الإلكترونية الشاملة متعددة المعرفة لنشر الأبحاث العلمية والتربوية (MECSJ)

العدد الخامس (أيلول) 2018

ومن النتائج بالجدول أعلاه يتضح أنه وباستخدام طريقة الحذف كان حذف المشاهدات 34 مؤثر بشكل كبير حيث أصبحت قيمة $F = 223.381$ وكان $R^2 = 0.94$ و $MSE = 0.011$.

أما عندما تم حذف المشاهدات 34 و 38 و 47 معاً كان التأثير عميقاً بشكل واضح حيث أصبحت قيمة $F = 228.237$ و $R^2 = 0.95$ و $MSE = 0.011$.

ونستنتج من هذا أن حذف المشاهدات 34 و 38 و 47 كان الأكثر تأثيراً وأدى إلى تحسن كبير في نتائج النموذج وذلك بتقليل قيمة MSE والذي كان قبل الحذف يساوي 0.012 وأصبح بعد الحذف يساوي 0.011، وهو فرق قد يبدو ضئيلاً ولكنه في الحقيقة فرق جوهري لأن هذا البحث يبحث في معدلات طلابية يؤثر فيها الجزء من مئة.

كذلك أدى حذف المشاهدات المذكورة إلى رفع قيمة F حيث كانت قبل الحذف تساوي 197.101 وأصبحت بعد الحذف تساوي 228.237، بالإضافة إلى رفع قيمة R^2 من 0.93 إلى 0.95.

ومما تقدم نستنتج أن أفضل النماذج المقدرية بعد معالجة القيم المتطرفة بطريقة الحذف هو النموذج الممثلة معاملاته في جدول (7) كما جاءت ذي المعادلة التالية:

$$\hat{Y} = 0.216 + 0.286 X_1 + 0.141 X_2 + 0.290 X_3 + 0.194 X_4$$

ثانياً- معالجة القيم المتطرفة بواسطة متوسط البتر

تم معالجة القيم المتطرفة بواسطة البتر لكل من عبارة من العبارات الموضحة في السابق وجاءت النتائج كما هو مبين بالجدول رقم (8) أدناه، حيث بينا فيه المفاضلة بين نتائج النماذج المقدرية بعد البتر:

جدول (8) المقارنة بين النماذج المقدرية بعد معالجة القيم المتطرفة بمتوسط البتر

MSE	R^2	F	β_4	β_3	β_2	β_1	β_0	الحالة المحذوفة
0.012	0.93	197.101	0.246	0.258	0.142	0.248	0.257	قبل المعالجة
0.012	0.934	195.213	0.250	0.253	0.145	0.248	0.250	معالجة مشاهدة 13
0.012	0.936	202.058	0.211	0.293	0.133	0.259	0.250	معالجة مشاهدة 38



www.mecsaj.com/ar

0.012	0.932	188.421	0.246	0.258	0.148	0.246	0244	معالجة مشاهدة 39
0.012	0.935	199.369	0.234	0.253	0.153	0.275	0.250	معالجة مشاهدة 47
0.012	0.935	197.012	0.250	0.251	0.153	0.243	0.244	معالجة مشاهدة 49
0.011	0.95	227.302	0.249	0.257	0.136	0.262	0.233	معالجة مشاهدة 34
0.010	0.95	241.049	0.193	0.290	0.141	0.286	0.216	معالجة مشاهدة 34 و38 و47

ومن النتائج بالجدول أعلاه نلاحظ أنه وباستخدام طريقة متوسط البتر كانت معالجة المشاهدات 34 مؤثرة بشكل كبير حيث $F = 227.302$ وكان $R^2 = 0.952$ و $MSE = 0.011$.

أما عندما تم حذف المشاهدات 34 و38 و47 معاً كان التأثير عميقاً بشكل جلي حيث أصبحت قيمة $F = 241.049$ و $R^2 = 0.95$ و $MSE = 0.010$.

ونستنتج من هذا أن معالجة المشاهدات 34 و38 و47 كان الأكثر تأثيراً وأدى إلى تحسن كبير في نتائج النموذج وذلك بتقليل قيمة MSE والذي كان قبل المعالجة يساوي 0.012 وأصبح بعدها يساوي 0.010، وهو فرق قد يبدو ضئيلاً ولكنه في الحقيقة فرق جوهري لأن هذا البحث يبحث في معدلات طلابية يؤثر فيها الجزء من مئة.

كذلك أدى حذف المشاهدات المذكورة إلى رفع قيمة F حيث كانت قبل المعالجة تساوي 197.101 وأصبحت بعد المعالجة تساوي 241.049، بالإضافة إلى رفع قيمة R^2 من 0.93 إلى 0.95.

ومما تقدم نستنتج أن أفضل النماذج المقدره بعد معالجة القيم المتطرفة بطريقة متوسط البتر هو النموذج التالي:

$$\hat{Y} = 0.216 + 0.286 X_1 + 0.141 X_2 + 0.290 X_3 + 0.193 X_4$$

جدول (9) أفضل نموذج مقدر بعد معالجة القيم المتطرفة بمتوسط البتر

قيمتها	الإحصاء
0.97	معامل الارتباط R



0.95	معامل التحديد R^2
0.942	معامل التحديد المعدل R^2_a
241.049	F
0.010	مربع مجموع الأخطاء MSE

وتم عرض نتائج المقارنة بين نتائج الجداول السابقة لاختيار أفضل نموذج في جدول (10):

جدول (10) المقارنة بين النماذج لاختيار أفضل نموذج

MSE	R^2	F	β_4	β_3	β_2	β_1	β_0	الحالة
0.012	0.93	197.101	0.246	0.258	0.142	0.248	0.257	قبل المعالجة بالحذف أو متوسط البتر
0.011	0.95	228.237	0.194	0.290	0.141	0.286	0.216	حذف مشاهدة 34 و38 و47
0.010	0.95	241.049	0.193	0.290	0.141	0.286	0.216	البتر مشاهدة 34 و38 و47

ومن الجدول السابق يتضح أن أفضل نموذج يمكن استخدامه للتنبؤ بقيم المتغير التابع هو النموذج الذي تم الحصول عليه بعد معالجة القيم المتطرفة بطريقة متوسط البتر للملاحظات رقم 34 و38 و47 كما يلي:

$$\hat{Y} = 0.216 + 0.286 X_1 + 0.141 X_2 + 0.290 X_3 + 0.193 X_4$$

وقيم معاملاته في جدول (28) السابق، حيث نلاحظ في النموذج المذكور تسجيل أعلى قيمة لإحصاءة $F = 241.049$ ، كما يلاحظ ارتفاع قيمة $R^2 = 0.95$ ، وانخفاض قيمة $MSE = 0.010$.

تشخيص أفضل نموذج تم تقديره:

بعد أن تم اختيار أفضل نموذج تم تقديره وقبل استخدامه للتنبؤ بقيم المتغير التابع لابد من القيام باختبار هذا النموذج من الناحية الاحصائية والقياسية،



www.mecs.com/ar

المجلة الإلكترونية الشاملة متعددة المعرفة لنشر الأبحاث العلمية والتربوية (MECSJ)
العدد الخامس (أيلول) 2018

وذلك للتأكد من أن معالجة المشاهدات الشاذة لم تخلق أي مشكلات إحصائية أو قياسية جديدة في النموذج المقدر، وللتأكد من ملاءمة النموذج قياسيًّا وإحصائيًّا.

اختبار معنوية النموذج إحصائياً:

إن مدى ملائمة نموذج الانحدار واستخدامه للتنبؤ يعتمد على أهمية (معنوية) المقدرات β_0 ، β_1 ، β_2 ، β_3 ، β_4 ويتم لهذا الغرض استخدام اختبارات والذي يستخدم عندما يكون تباين المتغير التابع مجهولاً ويمكن الحصول على قيمه عبر SPSS كما هي في جدول (11)

جدول (11) نتائج اختبار t لأفضل نموذج مقدر

مستوى المعنوية Sig.	قيم اختبار "ت" المحسوبة	قيم المعاملات Beta	
0.025	2.306	-	β_0
0.000	8.399	0.372	β_1
0.001	3.522	0.167	β_2
0.000	6.22	0.364	β_3
0.000	4.081	0.225	β_4

ويتضح من الجدول أعلاه أن كل معاملات النموذج معنوية حيث أنها أقل من مستوى المعنوية 0.05 المعتمد في هذا البحث بما في ذلك معنوية الحد الثابت β_0 .

اختبار تجانس تباين حد الخطأ العشوائي:

باستخدام اختبار جولدفيلد - كواندت للكشف عن تجانس تباين حد الخطأ العشوائي وتطبيق الخطوات السابق ذكرها عبر ترتيب المشاهدات تصاعدياً حسب المتغير X_2 لأنه الأكثر معنوية وتأثيراً بحسب قيمة بيتا، ومن ثم استبعاد 20% من المشاهدات في المنتصف ومن ثم حساب مجموع مربعات الخطأ SSE للمجموعتين تم الحصول على

$$SSE_2 = 0.164, SSE_1 = 0.145$$

$$F = (0.145/19)/(0.146/19) = 0.993$$



www.mecsj.com/ar

وهي قيمة F المحسوبة، وبمقارنتها بـ $F_{19,19,0.05}$ الجدولية $= 2.12$ نجد أن F المحسوبة أقل من F الجدولية وبموجبه نقبل الفرض الصفري القائل بأن هنالك تجانس أو ثبات في تباين الأخطاء

اختبار الارتباط الخطي المتعدد بين المتغيرات المستقلة:

باستخدام برنامج SPSS تم حساب قيم التسامح وقيم عامل التضخم وذلك لاكتشاف وجود الارتباط الخطي المتعدد، وكانت النتائج كما في جدول (12)

جدول (12) اكتشاف وجود الارتباط الخطي المتعدد

معدل الفصل الثاني عام 1438-1437 X_4	معدل الفصل الأول عام 1438-1437 X_3	معدل الفصل الثاني عام 1437-1436 X_2	معدل الفصل الأول عام 1437-1436 X_1	الثابت	المتغيرات Variables
3.097	3.494	2.294	2.005	-	قيم معامل التضخم VIF
0.323	0.286	0.436	0.499	-	قيم التسامح Tolerance
0.006	0.010	0.014	0.016	4.955	القيم الكامنة Eigenvalues
29.8	19.28	18.15	16.2	1	رقم الحالة CI
$CN = \sqrt{\frac{4.955}{0.006}} = 28.7$					مؤشر الحالة CN

يتضح من الجدول السابق (12) أعلاه أن قيم VIF لجميع المتغيرات المستقلة أقل من 10 وبالتالي لا يوجد ارتباط خطي متعدد بين المتغيرات المستقلة، كما إن قيم Tolerance لجميع المتغيرات المستقلة أكبر من 0.1 وهذا مؤشر أيضاً ودليل على عدم وجود الارتباط الخطي المتعدد بين المتغيرات المستقلة، كذلك رقم الحالة Condition Number أعطى القيمة $CN=28.7$ وهي قيمة أقل من 30 إلى 100 وبالتالي فإن جميع هذه المؤشرات تشير إلى عدم وجود ارتباط خطي.

اختبار الاستقلال الذاتي للبواقي:

باستخدام اختبار (Durbin-Watson) للحكم على مدى وجود استقلال ذاتي بين بواقي النموذج، ومن خلال برنامج SPSS تم استخراج قيمة $D-W = 2.11$ وبالرجوع إلى جدول Durbin-Watson بعدد متغيرات مستقلة $P=4$ و $n=50$ مشاهدة كانت قيمتي $D-W$ العليا والدنيا $d_U = 1.67$ و $d_L = 1.42$ ومن ثم ستكون مناطق القبول والرفض كما في جدول (23) التالي:

جدول (13) قرار دربن واتسن

0	d_L	d_U	$4-d_U$	$4-d_L$	4
0	1.14	1.74	2.26	2.86	4
رفض فرض العدم القائل بعدم وجود ارتباط ذاتي موجب	منطقة عدم قرار		قبول فرض عدم	منطقة عدم قرار	
0	1.14	1.74	2.26	2.86	4

وحيث أن قيمة $D-W = 2.11$ المحسوبة تقع داخل المنطقة المظللة على الشكل أعلاه فإننا نقبل الفرض الصفري القائل بعدم وجود ارتباط ذاتي، أي أن البواقي مستقلة ذاتياً.

اختبار التوزيع الطبيعي للبواقي:

بتطبيق اختبار جارك بيررا على النموذج المقدر نلاحظ أن قيمة الاختبار هي 3.7913 وأن مستوى المعنوية لها $= 0.150$ وهو أكبر من مستوى المعنوية 0.05 المعتمد في هذا البحث، وبموجبه نقبل الفرض الصفري القائل بأن البواقي تتوزع طبيعياً، وتجدر الملاحظة هنا بأن القيمة السابقة لهذا الاختبار قبل معالجة القيم المتطرفة كانت 4.067 وهي أكبر من القيمة بعد المعالجة وكانت بمستوى معنوية $= 0.13$ قبل المعالجة وهو أصغر منه بعد المعالجة وهذا يعطي مؤشر على أن قيمة الاختبار تتحسن بعد المعالجة حيث تصغر قيمته ويتضخم مستوى المعنوية، وهذا يعني أن القيم المتطرفة تؤثر في توزيع البواقي طبيعياً.

اختبار القدرة التنبؤية للنموذج:

مجدداً و عبر Eviews تم إيجاد قيم مؤشرات فاعلية التنبؤ التالية

$$\text{Root Mean Square Error (RSME)} = 0.09$$

$$\text{Mean Absolute Error (MAE)} = 0.07$$

$$\text{Mean absolute percentage error (MAPE)} = 2.5$$

$$\text{Theil's Inequality Coefficient} = 0.01$$

وبالمقارنة مع المؤشرات قبل المعالجة يلاحظ أن المؤشرات زاد اقترابها من الصفر بعد معالجة القيم المتطرفة وهذا مؤشر على تحسن النموذج

جدول (14) المقارنة بين مؤشرات فاعلية التنبؤ قبل وبعد المعالجة بمتوسط البتر

مؤشر فاعلية التنبؤ	قبل المعالجة	بعد المعالجة
الجذر التربيعي لمتوسط مربع الخطأ RMSE	0.106	0.09
متوسط الخطأ المطلق MAE	0.084	0.07
متوسط الخطأ النسبي المطلق MAPE	2.98	2.5
معامل التساوي لتايل	0.02	0.01

وهذا يدل على تحسن فاعلية النموذج بعد معالجة القيم المتطرفة حيث يلاحظ انخفاض في قيم المؤشرات نحو المعدلات المعيارية بعد المعالجة عنها قبل المعالجة.

قياس أثر القيم المتطرفة على القيم المقدرة بعد المعالجة بمتوسط البتر:

نلاحظ عدم ظهور قيم شاذة في المتغير التابع بعد المعالجة كما نلاحظ انخفاض عدد القيم المتطرفة في المتغيرات المستقلة إلى ثلاثة قيم فقط وهي 2، 16، 49 وهي قيم غير مؤثرة بحسب ما تم اجراءه من حسابات قبلية على النموذج قبل المعالجة، بينما اختفت الشواذ من المتغير التابع.



www.mecsj.com/ar

هذا بدوره أدى إلى انخفاض عدد القيم المقدرّة المتأثرة إلى عدد أقل منه قبل المعالجة، ونلاحظ هذا من المقارنة بين مجموع حالات التأثير في الجدولين، بالإضافة إلى أنه وكما تقدم ليست عملية التأثير بتلك الحدة لأن الشواذ المذكورة ليست عميقة التأثير.

جدول (15) المقارنة بين حالات التأثير بحسب المقياس قبل وبعد المعالجة

مقياس DFBETAS لدراسة تأثير القيم الشاذة على معاملات الانحدار كلا على حدة					مقياس COVRATIO لدراسة تأثير القيم الشاذة على الأخطاء المعيارية	مقياس كوك Cook لدراسة تأثير القيم الشاذة على معاملات الانحدار	مقياس DFFITS لدراسة تأثير القيم الشاذة على القيمة المقدرّة	القيم الشاذة في المتغير التابع	القيم الشاذة في المتغير المستقل	النموذج
β_4	β_3	β_2	β_1	β_0						
4	3	3	5	4	8	3	3	1	5	قبل المعالجة
3	3	2	3	4	5	1	1	0	3	بعد المعالجة

يوضح الجدول رقم (15) أعلاه انخفاض عدد القيم المتأثرة بالقيم المتطرفة بعد المعالجة، وهذا يدل كذلك على تأثير القيم المتطرفة على المعاملات المختلفة للنموذج وكيف يقل هذا التأثير بعد المعالجة، فمن الجدول نشاهد على سبيل المثال أن تأثير مقياس COVRATIO الذي يدرس تأثير القيم الشاذة على الأخطاء المعيارية قد أعطى قراءة ل 8 حالات قبل المعالجة وانخفضت بعد المعالجة إلى 5 حالات وهكذا في باقي المقاييس.

ثالثاً- اختبار فروض البحث:

باستعراض كل ما تقدم من أرقام ومعالجات خلص الباحث إلى الإجابات التالية على فروض البحث:

أولاً: اختلاف قيم مؤشرات دقة تنبؤ الانحدار المتعدد عن القيم المعيارية عند احتواء البيانات على قيم متطرفة:

أ- تختلف قيمة إحصاء F عن القيمة المعيارية عند احتواء البيانات على قيم متطرفة.

ب- تختلف قيمة معامل التحديد R^2 عن القيمة المعيارية عند احتواء البيانات على قيم متطرفة.



www.mecs.com/ar

المجلة الإلكترونية الشاملة متعددة المعرفة لنشر الأبحاث العلمية والتربوية (MECSJ)

العدد الخامس (أيلول) 2018

ج- تختلف قيمة معامل التحديد المعدل R_a^2 عن القيمة المعيارية عند احتواء البيانات على قيم متطرفة.

د- تختلف قيمة إحصاءة MSE عن القيمة المعيارية عند احتواء البيانات على قيم متطرفة.

هـ- تختلف قيمة معامل التساوي لثايل Theil's Inequality Coefficient عن القيمة المعيارية $U \rightarrow 0$ عند احتواء البيانات على قيم متطرفة.

و- تختلف قيم الجذر التربيعي لمتوسط مربع الخطأ Root Mean Square Error عن القيمة المعيارية $RMSE < 0.05$ عند احتواء البيانات على قيم متطرفة.

ز- تختلف قيم متوسط الخطأ المطلق Mean Absolute Error عن القيمة المعيارية $MAE \rightarrow 0$ عند احتواء البيانات على قيم متطرفة.

ثانياً: تتحسن قيم مؤشرات دقة تنبؤ الانحدار المتعدد وتصبح ضمن مدى القيم المعيارية بعد معالجة القيم المتطرفة:

أ- تتحسن قيمة إحصاءة F وتصبح ضمن مدى القيم المعيارية بعد معالجة القيم المتطرفة

ب- تتحسن قيم مؤشر معامل التحديد R^2 وتصبح ضمن مدى القيم المعيارية بعد معالجة القيم المتطرفة

ج- تتحسن قيم مؤشر معامل التحديد المعدل R_a^2 وتصبح ضمن مدى القيم المعيارية بعد معالجة القيم المتطرفة

د- تتحسن قيم إحصاءة MSE وتصبح ضمن مدى القيم المعيارية بعد معالجة القيم المتطرفة

هـ- تتحسن قيمة مؤشر معامل التساوي لثايل Theil's Inequality Coefficient وتصبح ضمن مدى القيم المعيارية $U \rightarrow 0$ بعد معالجة القيم المتطرفة

و- تتحسن قيمة الجذر التربيعي لمتوسط مربع الخطأ Root Mean Square Error وتصبح ضمن مدى القيمة المعيارية $RMSE < 0.05$ بعد معالجة القيم المتطرفة



www.mecsjs.com/ar

المجلة الإلكترونية الشاملة متعددة المعرفة لنشر الأبحاث العلمية والتربوية (MECSJ)

العدد الخامس (أيلول) 2018

ي- تتحسن قيمة متوسط الخطأ المطلق Mean Absolute Error وتصبح ضمن مدى القيمة المعيارية $MAE \rightarrow 0$ بعد معالجة القيم المتطرفة

ويوضح جدول (36) التالي المقارنة بين المؤشرات قبل وبعد المعالجة باستخدام الطريقة الأفضل للمعالجة التي توصل لها هذا البحث، وهي متوسط البتر

جدول (16) مقارنة بين احصاءات ومؤشرات النموذج قبل وبعد المعالجة

الإحصاءة – المؤشر	قبل المعالجة	بعد المعالجة
معامل التحديد R^2	0.93	0.95
معامل التحديد المعدل R^2_a	0.93	0.942
F	197.101	241.049
مربع مجموع الاخطاء MSE	0.012	0.010
الجذر التربيعي لمتوسط مربع الخطأ RMSE	0.106	0.09
متوسط الخطأ المطلق MAE	0.084	0.07
متوسط الخطأ النسبي المطلق MAPE	2.89	2.5
معامل التساوي لتايل	0.02	0.01

ومنه نلاحظ ازدياد قيمة معامل التحديد R^2 ومعامل التحديد المعدل R^2_a وقيمة F بعد المعالجة عنها قبل المعالجة، كما نلاحظ انخفاض مربع مجموع الاخطاء MSE بعد المعالجة عنه قبل المعالجة، كما إن كافة مؤشرات فاعلية التنبؤ المستخدمة والمتمثلة في الجذر التربيعي لمتوسط مربع الخطأ RMSE ومتوسط الخطأ المطلق MAE ومعامل التساوي لتايل قد تحسنت قيمها واقتربت من القيم المعيارية بعد المعالجة عنها قبل المعالجة، ومن الجدير ذكره أنه قد يبدو للناظر بأن الفروقات طفيفة ولا تذكر، لكنه في الحقيقة فرق جوهري لأن هذا البحث يبحث في معدلات طلابية يؤثر فيها الجزء من عشرة وكذلك الجزء من مئة.

5. مناقشة النتائج

النتائج التي خلصت إليها الدراسة الحالية تتفق مع نتائج الدراسات السابقة التي خلص إليها كلا من جيلبرت (1992) Gilbert في كون القيم المتطرفة ذات تأثير على خصائص الانحدار، كما يتفق أيضا مع دراسة دبوب ويونس (2006) التي خلصت إلى أن القيم المتطرفة تؤثر على تركيبية النموذج ونتائج تحليل الانحدار، كما يتفق كذلك مع ما قدمه الغنام (2007) والذي توصل إلى أن معاملات الانحدار تتأثر بالملاحظات الغير مألوفة (القيم المتطرفة)، كما اتفق مع ما قدمه تشوي (2009) Choi في أن وجود القيم المتطرفة يمكن أن يعطي نتائج مضللة، كما أضاف هذا البحث الكشف عن الشواذ بطريقة بواقي ستودنت بالإضافة إلى ما أوصت به دراسة حسن ورضا (2011) التي استخدمت طريقة البوكس بلوت ومصفوفة القبة H. بينما خلص هذا البحث إلى نتيجة مختلفة نسبياً عن ما توصل إليه يحي (2009) وقويدر (2016) في بحثيهما كلا على حدة والذين خلصا إلى نتيجة متقاربة هي أن حذف المشاهدات المتطرفة أثر في دقة معاملات النموذج بانخفاض متوسط مربعات الخطأ وعامل تضخم التباين وارتفاع قيمة F ومعامل التحديد R^2 وتغيير إشارات بعض المعلمات المقدرة والوصول إلى نموذج انحدار يمثل بيانات الدراسة خير تمثيل لدى يحي (2009)، وفي أن حذف القيم المتطرفة قد أثر بشكل إيجابي على نتائج التحليل وذلك من حيث تقليل قيمة الخطأ المعياري في التقدير وزيادة قيمة معامل التحديد لدى قويدر (2016)، غير أن البحث الحالي توصل إلى أن طريقة متوسط البتر أفضل من طريقة حذف المشاهدات المتطرفة وتثبت هذا النتائج المتوصل إليها والمقارنة المعقودة بين الطريقتين التي قدمها هذا البحث، كما تختلف مع أمانثونج وسوتي & Ampanthong (2009) Suwattee في أن جميع المقاييس التي تم استخدامها لقياس أثر القيم المتطرفة في المعاملات قد تفاوتت في اكتشاف الأثر بشكل متقارب.

الاستنتاجات:

وبالعودة إلى النتائج والأرقام المعروضة آنفا يستنتج الباحث ما يلي:

1. وجود تأثير معنوي من المتغيرات المستقلة مجتمعة على المتغير التابع وهذا يتفق مع الدراسات السابقة



www.mecsjs.com/ar

2. أظهر الرسم الصندوقي Pox Plot قدرة ضعيفة على اكتشاف القيم المتطرفة في متغيرات البحث ويظهر للباحث أن السبب في ضعف هذا الأسلوب هو كونه أسلوب بصري بالدرجة الأولى ولعدم شمول العمليات الحسابية فيه كل بيانات العينة وهو ما لم تأخذه بعض الدراسات السابقة في الاعتبار.

3. أظهرت عناصر القطر الرئيسي للمصفوفة Hat Matrix وبواقي ستودنت المحذوفة قدرات أكبر في اكتشاف القيم المتطرفة في المتغيرات المستقلة والمتغير التابع على التوالي، وذلك لأنها تعتمد طرق حسابية تأخذ في الاعتبار جميع مفردات بيانات العينة بدون استثناء.

4. للقيم المتطرفة تأثير معنوي على معاملات نموذج الانحدار الخطي المتعدد حيث تعمل على تضخيم قيم تباين الأخطاء MSE وتقلل من قيمة R^2 وقيمة F وهذا يعني أن القيم المتطرفة تقلل من جودة النموذج وتزيد من أخطاء بياناته.

5. لاحظ الباحث في بعض العينات التي أجرى عليها بحثاً إضافياً أن وجود القيم المتطرفة يعطي نتائج مظلمة فيما يخص الارتباط الخطي حيث أظهر التحليل الإحصائي للعينة محل البحث عدم وجود ارتباط خطي في النموذج قبل المعالجة، بخلاف ما أظهره فحص النموذج المقدر بعد المعالجة والذي أظهرت نتائجه وجود ارتباط خطي، وهو ما لم تشر له أياً من الدراسات السابقة.

6. تتحسن مؤشرات فاعلية التنبؤ بعد معالجة القيم المتطرفة عنها قبل المعالجة وهذا يتفق مع الدراسات السابقة.

7. يقل كثيراً حجم التأثير بعد معالجة القيم المتطرفة ويتضح هذا من خلال مقاييس الأثر المختلفة والتي توضح قلة عدد المفردات المتأثرة بعد المعالجة عنها قبل المعالجة.

8. توصل الباحث إلى أن معالجة القيم المتطرفة بمتوسط البتر أفضل من حذفها ويظهر للباحث أن السبب هنا هو أن استبدال القيم المتطرفة بقيم متوسطات البتر تحفظ حجم العينة كعدد مفردات مما يزيد قيمة F و R^2 ، بخلاف الحال عند حذف القيم المتطرفة التي تقلل مفردات العينة وبالتالي تقلل قيم F و R^2 .

6. الخلاصة

تناول الباحث في هذه الورقة البحثية الإخلال بافتراض خلو البيانات من القيم المتطرفة وأثر ذلك على استخدام الانحدار الخطي المتعدد في التنبؤ. ونظراً لأهمية التنبؤ بالأحداث المستقبلية في كافة المجالات التربوية والاجتماعية والاقتصادية في حياة الناس وأثر ذلك في اتخاذ القرارات الصحيحة وعمليات التخطيط، فقد جاءت هذه الدراسة لمناقشة مدى أهمية التخلص من القيم المتطرفة في البيانات المتعلقة بالأحداث المستقبلية والتوقعات بشأنها. تم في هذا البحث استخدام العديد من الأساليب الإحصائية من أجل التخلص من ظاهرة القيم المتطرفة في البيانات، حيث بينت نتائج الدراسة مدى دقة التنبؤ باستخدام تحليل الانحدار بعد معالجة القيم المتطرفة، حيث تحسنت قوة المؤشرات التفسيرية لنموذج الانحدار المتعدد.

7. التوصيات

اعتماداً على النتائج التي تم التوصل إليها في هذا البحث يمكن التوصية بما يلي:

1. ضرورة استخدام الطرق المختلفة للكشف عن وجود القيم المتطرفة قبل استخدام نموذج الانحدار.
2. ضرورة تقليل أثر القيم المتطرفة عند تطبيق نموذج تحليل الانحدار للحصول على نموذج أكثر دقة وفاعلية.
3. يفضل استخدام طريقة متوسط البتر لمعالجة القيم المتطرفة في بيانات نموذج الانحدار الخطي المتعدد.
4. عدم الاعتماد على الرسم الصندوقي Pox Plot في تشخيص القيم المتطرفة كلياً وإنما يكتفى الاستئناس به كنظرة مبدئية لاستشفاف وجود قيم شاذة من عدمه.
5. استخدام عناصر القطر الرئيسي للمصفوفة Hat Matrix وبواقي ستودنت المحذوفة Studentized Deleted Residuals في اكتشاف القيم المتطرفة في المتغيرات المستقلة والمتغير التابع.



www.mecsj.com/ar

المجلة الإلكترونية الشاملة متعددة المعرفة لنشر الأبحاث العلمية والتربوية (MECSJ)
العدد الخامس (أيلول) 2018

6. تقديم المزيد من الدراسات حول موضوع القيم المتطرفة وتأثيراتها المختلفة في بيانات نموذج الانحدار الخطي المتعدد والنماذج الاحصائية الأخرى.

8. البحوث المقترحة

بالنظر إلى حدود البحث ونتائجه فإن الباحث يقترح البحوث التالية:

1. اجراء البحث الحالي باستخدام طرق أخرى لاكتشاف القيم المتطرفة مثل مقياس مسافة مهلانوبيس (Mahalanobis distance) ومقارنته مع الطرق المختلفة: الرسم الصندوقي PoX Plot ومصفوفة القبة (Hat Matrix) للمتغيرات المستقلة و بواقي ستيودينت المحذوفة Studentized Deleted Residuals للمتغير التابع.

المصادر والمراجع:

- إسماعيل، محمد عبدالرحمن (2001). تحليل الانحدار الخطي. الرياض: معهد الإدارة العامة.
- الغنام، محمد طه (2007). المشاهدات غير المألوفة في نموذج الانحدار. تكريت: مجلة تكريت للعلوم الادارية والاقتصادية.
- دبوب، مروان عبدالعزيز؛ يونس، فرح عبدالغني (2006). تأثير القيم الشاذة على نتائج تحليل الانحدار مع التطبيق على الموالييد الخدج. الموصل: جامعة الموصل، مجلة علوم الراقدين.
- حسن، تارا احمد؛ رضا، مهدي صابر (2011). استخدام الانحدار الحصين لإيجاد أنسب نموذج لتمثيل بيانات الأنواء الجوية في مدينة أربيل خلال الفترة 1998-2010. أربيل: مجلة العلوم الاقتصادية والادارية.
- يحي، مزاحم محمد (2009). المشاهدات غير العادية في الانحدار الخطي المتعدد وبعض طرائق تشخيصها مع التطبيق.
- الموصل: المجلة العراقية للعلوم الاحصائية.



www.mecsaj.com/ar

المجلة الإلكترونية الشاملة متعددة المعرفة لنشر الأبحاث العلمية والتربوية (MECSJ)
العدد الخامس (أيلول) 2018

قويدر، عمر قاسم (2016). تحديد القيم الشاذة في بيانات تيمس TIMSS وأثر أسلوب التعامل معها في نتائج التحليلات الإحصائية. رسالة ماجستير غير منشورة. إربد: جامعة اليرموك.

زايد، مصطفى (2011). موسوعة الإحصاء. القاهرة: مطابع الدار الهندسية.

عودة، أحمد سليمان؛ الخليلي، خليل يوسف (2000). الإحصاء للباحث في التربية والعلوم الإنسانية، إربد: دار الأمل.

المراجع الأجنبية

Ampanthong, P., & Suwattee, P. (2009). A comparative Study of Outlier Detection Procedures in Multiple Linear Regression. Hong: Proceedings of the International MultiConference of Engineers and Computer Scientists , Vol.1.1 , IMECS.

Choi, S. (2009). The effect of outliers on regression analysis. Chicago: Regime Type and Foreign direct Investment Quarterly Journal of Political Science , (4): 153-65.

Montgomery, D., Jennings, C., & Kulahci, M. (2008). Introduction to Time Series Analysis and Forecasting. New Jersey: John Wiley & Som. Inc.

Salkind, N. J., & Rasmussen, K. (2008). Encyclopedia of Educational Psychology. Sage Publications, Inc: Los Angeles.