

”خطوات استخدام RUF للحصول على افضل نموذج (جزئي)

”للانحدار الخطي

"RUF procedures forgetting the best subset linear regression model"

م. د. صباح فرج عبد الحسين
معهد الادارة/ الرصافة

المستخلص

ان الغرض من بناء نموذج الانحدار الخطي هو وصف العلاقة الخطية (الحقيقية) ما بين كل متغير تفسيري في النموذج والمتغير المعتمد، على اساس ان المتغير المعتمد هو دالة خطية للمتغيرات التفسيرية، وبما يسمح باستخدام هذه الدالة في التنبؤ والسيطرة. وهذا الغرض لا يتحقق من دون الحصول على مقدرات معنوية ومستقرة ومعقولة لمعاملات النموذج الخطي. وقد وجد الباحث ان المعيار الذي سبق وان اقترحه واسماه RUF هو المعيار الدقيق والكافي للوصول الى هذا الغرض في حالة وجود مشكلة التعدد الخطي، وان ذلك مرتبط بتحديد النموذج الوافي الذي يحقق الفرضيات القياسية لحد الخطأ وليس بتجاوزها والاعتماد مباشرة على معايير (اصغر متوسط مربعات البواقي MSE واصغر مجموع مربعات البواقي التنبؤية PRESS واكبر قيمة لمعامل التحديد المعدل R^2) فذلك من شأنه الحصول على نموذج يحقق مطابقة ظاهرية للبيانات تضلل الباحث وتوهمه بانه قد حصل على نموذج الانحدار الخطي الافضل خلافا للحقيقة التي تشير الى ان مقدرات معاملات الانحدار في هذا النموذج (غير مستقرة وغير معقولة) لانها تعاني من قوة العلاقات الخطية فيما بينها وتأثير حد الخطأ في قوتها التفسيرية والتنبؤية. وعلى هذا الاساس حدد الباحث عدة خطوات لاستخدام RUF في الحصول على افضل نموذج (جزئي) للانحدار الخطي.

Abstract:

The purpose behind building the linear regression model is to describe the real linear relation between any explanatory variable in the model and the dependent one, on the basis of the fact that the dependent variable is a linear function of the explanatory variables and one can use it for prediction and control. This purpose does not come true without getting significant, stable and reasonable estimates for the parameters of the model, specifically regression-coefficients. The researcher found that "RUF" the criterion that he had suggested accurate and sufficient to accomplish that purpose when multicollinearity exists provided that the adequate model that satisfies the standard assumptions of the error-term can be assigned. It is wrong to ignore the assumptions and depend directly on the least "MSE & PRESS" and greatest " R^2 " because it satisfies the model with false fit to data, whereas the regression coefficients are still unstable and unreasonable because of the multicollinearity and the effect of the error-term on the explanatory and predicted power. So the researcher has made procedures for using his criterion "RUF" to get the real best subset linear model.

Key-words: multicollinearity; ill-conditioned matrix ; adequate model; ; best subset linear model:

الكلمات العربية

1. التعدد الخطي
2. النموذج الوافي
3. المصفوفة الشاذة (تقريباً)
4. النموذج الجزئي الخطي الافضل



مجلة العلوم

اقتصادية وإدارية

المجلد 18

العدد 66

الصفحات 357 - 386



خطوات استخدام RUF للحصول على أفضل نموذج (جزئي) للانحدار الخطي

1. المقدمة والهدف The preface and the purpose

ان هذا البحث هو استكمال لموضوع بحثي السابق الذي اقترحت فيه معيارا جديدا للاستدلال على عدم (استقرار ومعنوية ومعقولة) مقدرات معاملات الانحدار، ومعالجتها في نموذج الانحدار الخطي المتعدد الذي يعاني من مشكلة التعدد الخطي، وصولا الى النموذج (الجزئي) الافضل للانحدار. وقد اسميت هذا المعيار RUF اختصارا لعبارة (عامل عدم الاستقرار النسبي (Relative unstability factor) [1] وتم اشتقاق صيغته الرياضية من المعادلات الطبيعية $[X'XB=X'Y]$ التي يتم الحصول عليها بطريقة المربعات الصغرى فكانت:

$$RUF_{(i)} = \frac{Bi - riy}{riy} \quad \forall i = 1, 2, 3, \dots, m \quad (1)$$

حيث ان:

Bi: معامل الانحدار الجزئي (المعياري) في نموذج الانحدار الخطي المتعدد.

riy: معامل الانحدار الخطي (المعياري) في نموذج الانحدار الخطي البسيط (معامل ارتباط بيرسون).

m: عدد المتغيرات التفسيرية في النموذج.

وقد ربط الباحث استخدام هذا المعيار باختبار t لقياس معنوية معامل انحدار كل متغير تفسيري وذلك عندما يكون مدى عدم الاستقرار النسبي بين $[-1, -0.50]$ أي عندما تكون علاقته المباشرة بالمتغير المعتمد (متغير الاستجابة) اكبر من علاقته غير المباشرة، لكي يمكن ان ينسب عدم معنوية تقدير معامل الانحدار وعدم استقراره الى تأثير علاقته غير المباشرة بمتغير الاستجابة في النموذج. والحقيقة التاريخية هي ان العديد من الباحثين قد اقترحوا معايير لتقييم النماذج الجزئية الخطية للانحدار وصولا الى النموذج الجزئي الافضل مثل R_a^2 , MSE , CP , $PRESS$ ورسموا طرقا بخطوات محددة لعل اشهرها ما يلي:

1. طريقة كل الانحدارات الممكنة All possible Regressions

2. طريقة الاختيار الامامي Forward Selection

3. طريقة الاختيار المتدرج (الحكيم) Step-Wise Selection

4. طريقة الحذف التراجعي (الخلفي) Back ward Elimination

وفي مقدمة هؤلاء الباحثين Efroymsen في العام 1962 [2] Hooking في الاعوام 1967 و 1972 [3, 5] و Allen في العام 1971 [4] Mckay في العام 1979 [6] و Rencher في العام [7] 1980. وباستثناء الطريقة الاولى فان الطرق الباقية هي طرق مختصرة للوصول الى الهدف المذكور، الا انها - مع الاسف - قد تصل بالباحث الى نموذج جزئي بتقديرات (غير مستقرة) و (غير معقولة) في حالة كون مصفوفة المتغيرات التفسيرية هي مصفوفة شاذة (تقريبا) ill-conditioned matrix حيث تعاني من مشكلة التعدد الخطي، ولذلك اقترحت طرق اخرى تعتمد اساسا على التقدير المتحيز لكافة معاملات الانحدار في النموذج: مثال ذلك: طريقة المكونات الرئيسية principal componants [8] وطريقة انحدار الـ [9] Ridge. اما الباحث فقد اقترح معيارا وطريقة جديدة كما اشرنا للحصول على النموذج الجزئي الافضل للانحدار بمقدرات المربعات الصغرى نفسها وبوجود مشكلة التعدد الخطي، بهدف تجاوز العيوب التي تصاحب مقدرات معاملات الانحدار في جميع الطرق المذكورة آنفا، مع رسم خطوات الوصول الى النموذج الجزئي الافضل الذي يتمتع بمقدرات المربعات الصغرى (غير المتحيزة) او (قليلة التحيز) لكنها معنوية ومستقرة ومعقولة في ان معا باستخدام معياره المقترح RUF. وقد حدد الباحث المدييات المؤثرة وغير المؤثرة لقيم RUF على استقرار ومعقولة ومعنوية مقدرات معاملات الانحدار في نموذج الانحدار الخطي المتعدد، واثبت بالتطبيق على ثلاثة امثلة متنوعة ان استخدام RUF وفق الخطوات التي اقترحها توصل دائما الى نموذج جزئي بمقدرات معنوية ومستقرة لمعاملات الانحدار وباقبل متوسط لتضخم تبايناتها VIF واصغر MSE و $PRESS$ واقل تحيز مع اكبر R_a^2 مقارنة بالنماذج الجزئية الاخرى التي تحقق فرضيات حد الخطاء.



خطوات استخدام RUF للحصول على أفضل نموذج (جزئي)

الانحدار الخطي

في حين يكون استخدام الطرق الأخرى^(*) لاختيار النموذج الجزئي الأفضل (غير مضمون النتائج) فقد توصل الباحث إلى نموذج جزئي (جيد) ولكنه ليس النموذج الأفضل، وقد يبتعد كلياً عن النموذج الجيد نحو نموذج لا يحقق الفرضيات القياسية لحد الخطأ ويعاني في الوقت نفسه من مشكلة التعدد الخطي فتبدو معاملات الانحدار فيه غير مستقرة وغير معقولة وإن كان لها أصغر MSE و PRESS واكبر R^2 ، وفي بعض الحالات الأخرى تطابق هذه الطرق الطريقة المقترحة فتوصل الباحث إلى النموذج الجزئي الأفضل.

2. مستوى المعنوية الملائم لاختبار معنوية معاملات الانحدار؛ ومؤشرات التعدد الخطي والقيم المتطرفة

Appropriate significance-level and multicollinearity and extreme values indicators

يتحدد مستوى المعنوية الملائم لاختبار معنوية معاملات انحدار المتغيرات التفسيرية في نموذج الانحدار الخطي المتعدد بعاملين هما: (1) مدى تغير قيم هذه المتغيرات وارتباطها الخطي بمدى تغير قيم متغير الاستجابة. (2) قوة العلاقة الخطية ما بين هذه المتغيرات. فكلما كان العامل الأول كبيراً والثاني ضعيفاً كلما أمكن إيجاد تقديرات دقيقة باستخدام مستوى معنوية دقيق مثل 0.01 أو 0.05 والعكس صحيح حيث يجب زيادة مستوى المعنوية لكي نسمح ببقاء عدد كافٍ من المتغيرات التفسيرية في النموذج تؤدي إلى مقدر أفضل لتباين الخطأ وزيادة قدرة النموذج التفسيرية والتنبؤية ومطابقته للبيانات التي من أهم مؤشراتهما C_p , R^2 , PRESS, MSE وعادة ما يكون الحد الأعلى لمستوى المعنوية المقبول هو 0.10 وذلك لأن أكثر من ذلك يعني إبقاء متغيرات تفسيرية في النموذج لا قدرة حقيقية لها (في المجتمع) على تفسير الاختلافات في قيم الظاهرة. ولذلك فإن من الأهمية بمكان أن يلجأ الباحث—بعد أن حدد متغيراته التفسيرية ومتغيرات الاستجابة، وجمع البيانات عن طريق عينة عشوائية—إلى حساب مصفوفة الارتباط الخطية بين المتغيرات التفسيرية ومتغير الاستجابة وما بين المتغيرات التفسيرية نفسها، فذلك يعطيه فكرة أولية عن قوة متغيراته المستقلة وتأثير العلاقات الخطية فيما بينها ومن ثم يمكنه من تحديد مستوى المعنوية الملائم لاختبار معنوية تقديرات معاملات الانحدار في النموذج. وعلى أية حال فإن مستوى المعنوية (0.10) هو مستوى المعنوية (الأمين) للباحث—حسب تقديري—إذا لم يستطع أن يحدد المستوى الملائم في ضوء مصفوفة الارتباطات الخطية، وهو ما لاحظناه في بحثنا السابق عن هذا الموضوع حيث توصلت طريقة step-wise إلى نموذج الانحدار الخطي (الجزئي) المناسب عند هذا المستوى من المعنوية بينما فشلت عند مستويات المعنوية الأدنى (0.01, 0.05)^(**) وواقع الحال في كل البرامج الإحصائية ومنها برنامج SPSS الذي اعتمدناه في رسم خطوات الوصول إلى النموذج الجزئي الأفضل للانحدار الخطي—وهو أن الدورة الأولى لهذا البرنامج (انحدار y على جميع المتغيرات المستقلة) تعرض لنا ليس مصفوفة الارتباطات الخطية ما بين متغيرات النموذج وحسب، وإنما تكشف عن مدى وكيفية انتشار البواقي المعيارية حول متوسطها والتي يمكن أن يستدل عليها الباحث من خلال اختبارات توزيع البواقي. واحصاءه Cook أو صندوق Tukey (Box-Plot) الذي يكشف عن قيم البواقي المتطرفة والمؤثرة على القوة التفسيرية والتنبؤية للمتغيرات التفسيرية، هذا فضلاً عن بروز مؤشرات التعدد الخطي (إن وجد في النموذج) والتي يمكن تلخيصها بما يلي:

(*) استخدمت في هذا البحث طريقة step-wise كنموذج للطرق الأخرى المعروفة لاختيار أفضل نموذج جزئي للانحدار الخطي، إلى جانب الطريقة المقترحة RUF.

(**) لاحظ - رجاء - البحث الموسوم (طريقة ومعيان جديان للكشف عن مشكلة التعدد الخطي ومعالجتها واختيار متغيرات الانحدار) والمنشور في مجلة الإدارة والاقتصاد - الصادرة من كلية الإدارة والاقتصاد، الجامعة المستنصرية- العدد 86، العام 2011.



خطوات استخدام RUF للحصول على أفضل نموذج (جزئي)

الانحدار الخطي

- * تغير إشارة معامل الانحدار المعياري (Bi) Beta coefficient لمتغير تفسيري في نموذج الانحدار الخطي المتعدد عنه في نموذج الانحدار الخطي البسيط (riy) Pearson coefficient مما يشير بصورة قاطعة الى عدم استقراره ويجعل تفسير علاقة المتغير التفسيري بالمتغير المعتمد غير معقولة وغير مقبولة unreasonable & unacceptable.
- * عدم معنوية معامل انحدار المتغير التفسيري (المهم) في النموذج وذلك نتيجة تضخم تباينه حيث يكون في الغالب $VIF(i) \geq 10$ [10].
- * بلوغ او تجاوز القيمة المطلقة لمركبة العلاقة غير المباشرة (di) للمتغير التفسيري بالمتغير المعتمد، القيمة المطلقة لمركبة علاقته المباشرة (riy) أي ان $|di| \geq |riy|$ وهذا يظهر بوضوح وببساطة عندما تكون قيمة المعيار الذي اقترحه الباحث $RUF(i) \geq 1$ حيث RUF يحسب لكل متغير تفسيري - كما ذكرنا - وفق العلاقة البسيطة $RUF(i) = \frac{Bi - riy}{riy}$.

ان اكتشاف أي واحد من هذه المؤشرات كاف للاستدلال على وجود علاقات خطية قوية ما بين المتغيرات التفسيرية تؤثر في تقديرات معاملات الانحدار ومجمل تحليل الانحدار مما يستوجب المعالجة. وجدير بالذكر ان العديد من الباحثين يلجأ الى حساب قيم الجذور المميزة (λis) characteristic roots لمصفوفة معاملات الارتباط ما بين المتغيرات المستقلة لاكتشاف التعدد الخطي، وهي ولاشك مؤشر قاطع ولكنها تصبح صعبة المنال وتتطلب الكثير من الوقت عندما يكون عدد المتغيرات كبيرا (نسبيا) ولا يتوفر لدى الباحث برنامج خاص لحل المعادلات متعددة الحدود polynomial equations.

3. مديات RUF المؤثرة وغير المؤثرة في استقرار مقدرات معاملات الانحدار الخطي

RUF-Effective & Ineffective Ranges on stability of L.S.E of Regression coefficients

تتمثل أهمية كل متغير مستقل في نموذج الانحدار الخطي بإمكانيته (الحقيقية) على تفسير الاختلافات في قيم المتغير المعتمد، وهذه الامكانية تظهر ابتداءً في مدى اختلافات قيمه العينية المرافقة لاختلافات القيم العينية للمتغير المعتمد، وتتجسد اخيرا في الميل الحدي او ما نسميه عادة في الانحدار بالتقدير المعنوي لمعامل الانحدار الذي عندما يكون معياريا فهو مرادف لمعامل الارتباط البسيط (riy) ما بينه وبين المتغير المعتمد، وهو معامل مستقر في نموذج الانحدار الخطي البسيط وفي نماذج الانحدار الخطي المتعدد (الجزئية) ايضا اذا ما انتفت او ضعفت كثيرا العلاقات الخطية ما بين المتغيرات التفسيرية. اما اذا كانت العلاقات الخطية بين بعض او جميع المتغيرات التفسيرية، قوية، فان هذا الاستقرار يهتز ويتمثل هذا الاهتزاز غالبا في خسارة المتغير التفسيري النسبية لبعض امكانيته في تفسير الاختلافات في قيم المتغير المعتمد ويظهر ذلك في القيمة الجديدة (تقدير) معامل الانحدار المعياري (Bi) ويتم قياس هذا التغير النسبي في الامكانية (الأهمية) باستخدام عامل عدم الاستقرار النسبي RUF، الذي اقترحه الباحث في بحثه المشار اليه انفا حيث:

$$RUF(i) = \frac{Bi - riy}{riy} = \frac{\Delta riy}{riy} = \frac{di}{riy} \quad \forall i = 1, 2, \dots, m$$

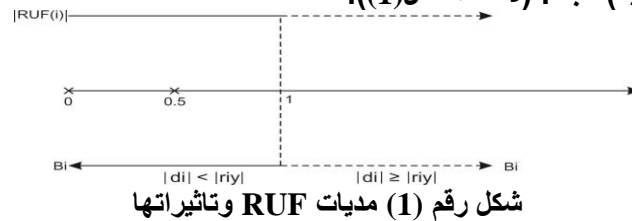


خطوات استخدام RUF للحصول على أفضل نموذج (جزئي)

الانحدار الخطي

ويتضح من صيغة حساب قيمة RUF ان الاستقرار التام لمقدرات معامل انحدار المتغير التفسيري تتحقق عندما تكون قيمة RUF تساوي صفر أي عندما تكون مركبة علاقته غير المباشرة (di) تساوي صفر، كما يتضح ايضا ان المتغير ضعيف الاهمية يكون الاقل امكانية على الاستقرار في نموذج الانحدار الخطي المتعدد الذي يعاني من العلاقات الخطية ما بين متغيراته التفسيرية. اما الخسارة النسبية التي يشير اليها RUF فهي التغير (التناقض) النسبي في امكانية المتغير التفسيري التي يمثلها التقدير المعنوي لمعامل انحداره المعياري والتي تصل حدها الاقصى عندما تصبح قيمة معامل الانحدار المعياري تساوي صفر وعند ذلك يزول المتغير المستقل من النموذج وتصبح قيمة RUF لذلك المتغير في حدها الأدنى حيث تساوي سالب واحد وهي تعني ان قوة مركبة العلاقة الخطية غير المباشرة لذلك المتغير تساوي في المقدار قوة مركبة علاقته الخطية المباشرة بالمتغير المعتمد وتعاكسها في الاتجاه، وبذلك لا يعد هذا المتغير متغيرا تفسيريًا في النموذج، والدليل على ذلك ان حذفه لا يؤدي الى زيادة تقدير تباين الخطأ من النموذج. اما اذا تجاوزت قيمة هذه الخسارة النسبية التي يشير اليها RUF(i) السالب واحد فذلك يعني تشوه علاقة الانحدار ما بين المتغير المستقل X_i والمتغير المعتمد اذ تصبح هذه العلاقة باتجاه معاكس للاتجاه (المعقول) الذي يشير اليه المنطق (النظرية) والواقع (المشاهدات)، ويتحقق ذلك عندما يكون تقدير معامل الانحدار المعياري (Bi) في النموذج باشارة مخالفة لاشارة معامل الانحدار المعياري (riy) في نموذج الانحدار الخطي البسيط أي عندما تكون قوة مركبة العلاقة الخطية غير المباشرة للمتغير المستقل (di) اكبر في المقدار من قوة مركبة علاقته الخطية المباشرة (riy) ومعاكسة لها في الاتجاه.

ويمكن تشبيه حالة المتغير المستقل X_i عندما $RUF(i) \geq -1$ بحالة النابض الحلزوني الذي تؤثر فيه محصلة قوى تتجاوز (امكانيته) على المحافظة على طبيعته الحلزونية أي تتجاوز الحد الاقصى لمرونته فعند ذلك ستتسوه طبيعته المعروفة ويصبح غير صالح للاستخدام، كذلك يصبح المتغير (التفسيري) في النموذج في هذه الحالة غير صالح للبقاء في النموذج حتى وان بدا تقدير معامل انحداره (معنويًا) فهي معنوية مضللة للباحث لانها جاءت من العلاقة الخطية غير المباشرة للمتغير المستقل أي من علاقته الخطية بالمتغيرات التفسيرية الاخرى وهو ما يخالف مفهوم معامل الانحدار باعتباره الميل الحدي لتغير المتغير المعتمد حيث بقية المتغيرات (التفسيرية) ثابتة. (لاحظ الشكل(1)).



وهذا الامر هو ما يشير اليه Draper & Smith في مخطط بناء نموذج الانحدار الخطي حيث ان واحدة من من اهم خطوات بناء النموذج هو التحقق Verification من استقرار تقديرات معاملات النموذج وكون معاملات الانحدار (معقولة reasonable) [11]، والباحث في طريقته المقترحة باستخدام RUF يربط ما بين مفهوم (استقرار تقديرات معاملات الانحدار) وبين كونها (معقولة) بمعنى موافقتها للمنطق والواقع وبين معنويتها الحقيقية، وهو ما سوف نبينه لاحقا. وعلى هذا الاساس يمكن تحديد المديات المؤثرة وغير المؤثرة لقيم RUF على استقرار ومعقولية ومعنوية تقديرات معاملات الانحدار كما يلي:

1. المدى غير المؤثر $0 \leq RUF < 0.5$ - حيث تكون التقديرات مستقرة ومعقولة ومعنوية.
2. المدى المؤثر $-0.5 \leq RUF < -1$ - وتظهر قوة تأثيره كلما اقتربت من -1 - وكلما كانت علاقة المتغير المستقل الخطية (المباشرة) بالمتغير المعتمد (متغير الاستجابة) ضعيفة نسبيا.
3. المدى المؤثر جدا (تزلزل او تشوهه في علاقة الانحدار) $RUF \geq -1$ وفيه تكون المقدرات غير مستقرة وغير معقولة وقد تبدو (بمعنوية زائفة).



خطوات استخدام RUF للحصول على أفضل نموذج (جزئي)

الانحدار الخطي

وتعتمد قوة تأثير RUF لكل متغير تفسيري ليس فقط على قوة علاقته الخطية مع المتغيرات التفسيرية الأخرى أي ليس على مقدار تضخم تقدير معامل انحداره VIF فحسب وإنما أيضاً على الضعف النسبي لقوة علاقته الخطية المباشرة بالمتغير المعتمد مقارنةً بغيره من المتغيرات التفسيرية (لاحظ العلاقة (1)) ولذلك اختلف الباحثون في المدى المؤثر لـ VIF على تقديرات معاملات الانحدار فبعضهم قال أكبر من 5 وبعضهم قال أكبر من 8 وآخرون لم يرضوا بأقل من 10 فأكثر [12] والواقع أن ذلك مرتبط بقوة علاقة المتغير التفسيري المباشرة بالمتغير المعتمد (riy) فقد تكون من الضعف النسبي بحيث أن $[VIF < 3]$ كاف لعدم استقرار تقدير معامل انحداره وقد يكون من القوة النسبية بحيث $[VIF \geq 10]$ غير كاف للتأثير في معنويته كما سنلاحظ ذلك في التطبيق، ولذلك فإن المعيار والمؤشر الصحيح هو RUF وليس VIF لأنه يتضمن قوة علاقات المتغير الخطية مع غيره من المتغيرات التفسيرية زانداً قوة علاقته الخطية المباشرة بالمتغير المعتمد (متغير الاستجابة).

4. العلاقة بين معنوية معامل الانحدار الخطي واستقراره:

The relation between significance of linear regression coefficient and its stability

تعتمد معنوية تقدير معامل الانحدار الخطي للمتغير التفسيري على مقدار الخطأ المعياري لذلك التقدير فقط - في حالة استقرار تقدير المعامل - إذ تتناسب معه عكسياً، فكلما قل الخطأ المعياري زادت معنوية التقدير. وحيث أن الخطأ المعياري لكل تقدير في النموذج يعتمد على متوسط مربعات البواقي للنموذج

كله MSE كما نلاحظ من العلاقة: $S_{bi} = \left[\frac{MSE}{S_{ii}} \cdot VIF(i) \right]^{\frac{1}{2}}$ التي يكون فيها S_{ii}

(مجموع مربعات انحراف قيم المتغير التفسيري X_i عن وسطه الحسابي) كمية ثابتة في كل نموذج انحدار خطي ويكون VIF (عامل تضخم تباين تقدير معامل الانحدار للمتغير المستقل) يساوي واحد عند استقراره، ولذلك تكون معنوية تقدير معامل الانحدار الخطي معتمدة كلياً على MES في هذه الحالة. ولتقدير المعنوية الحقيقية لمعامل انحدار متغير تفسيري نحسب معنويته في دالة انحدار خطي متعدد تضم (جميع) المتغيرات التفسيرية التي تشترك في تفسير هذه الدالة ويسمى عند ذلك نموذج الانحدار الخطي المتعدد بالنموذج الخطي (غير المتحيز) وذلك لأنه يستخدم بيانات (جميع) المتغيرات التفسيرية في تقدير التباين الحقيقي للخطأ، وهو التقدير الذي نشير إليه بالرمز (MSE). ولكن هذه الفائدة في تقدير التباين التي تنعكس في زيادة معنوية تقديرات معاملات الانحدار نتيجة لتناقص أخطائها المعيارية قد لا يظهر أثرها إذا كانت هناك علاقات خطية قوية ما بين المتغيرات التفسيرية تؤدي إلى تضخم أخطائها المعيارية وعدم استقرار تقديرات معاملات انحدارها الذي قد يصل إلى درجة بحيث يكون اتجاه الميل الحدي لمتغير الاستجابة على المتغير التفسيري (اتجاه معامل الانحدار) معاكس لاتجاهه الذي ينسجم مع المنطق (النظرية) والواقع (المشاهدات) فيعطي تفسيراً غير معقول لعلاقة المتغير التفسيري بمتغير الاستجابة (المتغير المعتمد).

وحيث أن العلاقات الخطية ما بين المتغيرات التفسيرية التي تسبب تضخم الأخطاء المعيارية لتقديرات معاملات انحدارها هي نفسها التي تسبب عدم استقرار تلك التقديرات، لذلك تكون العلاقة ما بين قوة معنوية تقديرات معاملات الانحدار واستقرارها علاقة طردية في حدود المدى المؤثر في استقرار ومعنوية تقديرات معاملات الانحدار $[-1 < RUF \leq -0.5]$ والمشار إليه انفاً. ويتضح ذلك من صيغة اختبار t لمعنوية تقدير معامل انحدار X_i تحت فرضية العدم:

$$t = \frac{b_i}{S_{b_i}}$$

تتناقص قيمته
تأثير العلاقات الخطية
تضخم قيمته
مابين المتغيرات التفسيرية
ضعف معنوية وعدم استقرار تقدير
معامل الانحدار



خطوات استخدام RUF للحصول على أفضل نموذج (جزئي)

الانحدار الخطي

وعلى هذا الأساس تكون الخسارة النسبية المؤثرة- في المدى المذكور- هي التي تؤدي الى عدم معنوية واستقرار) تقدير معامل انحدار المتغير التفسيري مما يستوجب حذفه من النموذج للحصول على تقديرات مستقرة ومعنوية لمعاملات الانحدار الأخرى. وفي حالة وجود أكثر من متغير تفسيري تعرض الى خسارة نسبية مؤثرة بسبب علاقاته الخطية، فيحذف من بينها المتغير التفسيري الذي تعرض لأكبر خسارة نسبية اي المتغير الأقل استقرار الذي تكون قيمة RUF له هي الأكبر سالبة أو الأقل قيمة. أما إذا تجاوزت الخسارة النسبية المدى المذكور بحيث أصبحت $RUF(i) > 1$ - فعند ذلك- وكما ذكرنا انفا- سوف لا يكون تقدير معامل انحدار المتغير التفسيري X_i غير مستقر وحسب وإنما يكون (غير معقول) أيضا لان اتجاهه يكون مخالفا لاتجاه العلاقة ما بينه وبين متغير الاستجابة، وان بدأ (معنويا) في بعض الاحيان نتيجة قوة مركبة علاقته الخطية غير المباشرة بمتغير الاستجابة.

5. الفرضيات القياسية لحد الخطأ العشوائي في النموذج ودقة تقديراته

The standard assumptions of the error-term and the accuracy of the model-estimators

تتأثر كافة مقدرات نموذج الانحدار الخطي في اقتربها او ابتعادها عن قيمتها الحقيقية في المجتمع بمقدار تحقيقه للفرضيات القياسية لحد الخطأ فيه، والتي وضعها Ostrom في قائمة من ست نقاط هي - اضافة الى خطية العلاقة ما بين المتغيرات التفسيرية و متغير الاستجابة (المتغير المعتمد) - ما يلي: [13] اولاً: ان قيم المتغيرات التفسيرية غير تصادفية $Nonstochastic X \Rightarrow E(eixi) = 0$ أي ان قيم حد الخطأ العشوائي غير مرتبطة بقيم المتغيرات التفسيرية.

ثانياً: متوسط قيم حد الخطأ العشوائي يساوي صفر $E(ei) = 0$ وهذا يعني ان معادلة الانحدار الخطي تعطي متوسط قيمة المتغير المعتمد (متغير الاستجابة y).

ثالثاً: تباين قيم حد الخطأ العشوائي: ثابت $E(e_i^2) = \sigma^2$ وذلك لكل قيمة من قيم المتغيرات التفسيرية.

رابعاً: عدم وجود ارتباط ذاتي بين قيم حد الخطأ العشوائي $Non\ autoregression \Rightarrow E(e_i, e_{i-m}) = 0$ حيث $m \neq 0$ ، وهذه الفرضية تكفل ان القيمة المتوسطة للمتغير المعتمد تعتمد فقط على المتغيرات التفسيرية وليس على قيم حد الخطأ [14]، وهي الأكثر أهمية للباحث عندما يريد تقييم نموده على اساس متوسط مربعات الخطأ فيه (MSE) ومعامل التحديد المعدل (R_a^2)، والاختفاء المعياري لتقديرات معاملات الانحدار في النموذج (Sbi)، وذلك لان عدم تحقيقها يؤدي الى نقص خطير في تقدير تباين الخطأ تبعده عن حقيقة قيمته الفعلية مما ينعكس في تضخم قيمة (R_a^2) وتقلص قيم الاختفاء المعياري لمعاملات الانحدار، مما يعطي الباحث فكرة مضللة عن مطابقة النموذج للبيانات ودقة تقديراته ومعنويتها [15, 16] ويتضح ذلك من علاقة MSE بـ R_a^2 حيث: $R_a^2 = 1 - \frac{MSE}{MST}$

اذ نلاحظ من هذه العلاقة ان المتغير الوحيد الذي يؤثر في قيمة R_a^2 هو MSE فإذا كان هذا التقدير لتباين الخطأ اقل كثيراً من حقيقته فان ذلك سينعكس في اقتراب قيمة R_a^2 من الواحد الصحيح فيبدو النموذج بأنه الأكثر مطابقة للبيانات والأكثر قدرة على تفسير الاختلافات في قيم الظاهرة موضوع البحث والشئ نفسه يصح

على الاخطاء المعيارية لتقديرات معاملات الانحدار حيث ان: $S_{bi} = \sqrt{\frac{MSE}{S_{ii}}}$

(على فرض عدم وجود علاقات خطية قوية بين المتغيرات التفسيرية تضخم من قيم هذه الاخطاء)، ولذلك فقد اعتمد الباحث في التاكيد من توفر هذه الفرضية على وجه التحديد، على ما يلي:

- أ. اجراء اختبار التعاقبات Runs test للتأكد من عشوائية توزيع حد الخطأ.
 - ب. اجراء اختبار Durbin-Watson للارتباط الذاتي (من الدرجة الاولى) بين قيم حد الخطأ.
 - ت. رسم قيم حد الخطأ مقابل القيم التنبؤية لمتغير الاستجابة. وهو ما يعطي صورة شاملة ليس عن تحقيق هذه الفرضية وحسب، وإنما جميع الفرضيات القياسية المذكورة انفا.
- خامساً: طبيعية توزيع حد الخطأ العشوائي في النموذج Normality of error-term: وهذه الفرضية ضرورية لضمان صحة كافة اختبارات المعنوية.



خطوات استخدام RUF للحصول على أفضل نموذج (جزئي)

الانحدار الخطي

6. النموذج الوافي واحصاءه (Cp) Adequate Model and Mallows-Statistic (Cp) Mallows

يسمى نموذج الانحدار الخطي الذي يحقق كافة الفرضيات القياسية لحد الخطأ ويتمتع باصغر متوسط مربعات للبواقي MSE أي يتضمن جميع المتغيرات التفسيرية، بالنموذج الوافي Adequate Model (*) وذلك لأنه يفي بكل متطلبات التقدير غير المتحيز والذي يمكن اعتماده كتقدير دقيق لتباين الخطأ وهذا يعني ان $E(S^2) \cong \sigma^2$ حيث $MSE = S^2$ (متوسط مربعات البواقي للنموذج الوافي). ان اهمية تحديد النموذج الوافي تكمن في الحاجة الى الحصول على التقدير الدقيق وغير المتحيز لتباين الخطأ كمعيار مرجعي (مهم) كما اشرنا انفا في تقييم مدى مطابقة النموذج للبيانات. ولكن العلاقات الخطية القوية مابين المتغيرات التفسيرية تجب اهمية هذه المتغيرات في دالة النموذج اذ تكون تقديرات معاملات انحدارها (غير معنوية) او باتجاه معاكس لاتجاه العلاقة المعقولة مابين هذه المتغيرات ومتغير الاستجابة (المتغير المعتمد). وان طرق معالجتها الشائعة تؤدي الى دالة خطية بمعادلة متحيزة بمقدار ما يختلف من طريقة لاخرى. وهذا التحيز يظهر في تقدير تباين الخطأ بواسطة متوسط مربعات البواقي MSE للنموذج حيث ان:

$$MSE = \sigma^2 + (Bias)^2$$

وقد اقترح Mallows C.L. الاحصاء (Cp) في العام 1973 والتي صيغة حسابها: [17]

$$Cp = \frac{SSEP}{s^2} - (n - 2p)$$

ليبيان مقدار التحيز في تقدير تباين الخطأ من معادلة الانحدار النهائية قياسا بمعادلة الانحدار التي تمثلها دالة النموذج الوافي الذي فيه كما ذكرنا:

$$E(S^2) \cong \sigma^2 \text{ \& } E(SSEP) \cong (n - p)\sigma^2 \Rightarrow \therefore \frac{SSEP}{S^2} = \frac{(n - p)\sigma^2}{\sigma^2} = n - p$$

وبالتعويض في صيغة حساب احصاء Mallows اعلاه يكون: $Cp = p$ في نموذج الانحدار الخطي الوافي

ان النموذج الذي نحصل عليه باستخدام طريقة RUF (المقترحة) قد يكون هو نفسه النموذج الوافي (غير المتحيز) او يكون نموذجا اخر يحقق فرضيات حد الخطأ ولكنه (متحيز قليلا) لانه يسقط احد المتغيرات التفسيرية التي تعاني من العلاقات الخطية المؤثرة مع غيرها من المتغيرات التفسيرية، وبذلك تكون قيمة Cp فيه تبتعد قليلا ولكنها الاقرب الى P، [18] ويتميز عن غيره من النماذج الجزئية الاخرى بان تقديرات معاملات الانحدار فيه معنوية ومستقرة (معقولة)، واذا اردنا حساب قيمة Cp بدقة فعلياً ان نحدد اولاً وبصورة دقيقة (النموذج الوافي) وفقاً لمواصفاته المذكورة انفا (يحقق فرضيات حد الخطأ وله اصغر MSE) وليس بالضرورة - كما اشرنا - ان يضم جميع المتغيرات المستقلة لانه قد يكون بعض هذه المتغيرات (غير تفسيرية) حيث لا يؤثر حذفها من النموذج على قوته التفسيرية والتنبؤية أي لا يؤدي الى زيادة MSE او PRESS كما لا يؤدي الى خرق فرضية من فرضيات الخطأ، فهي بهذا المعنى تعد (متغيرات زائدة) استخدمت في تعيين النموذج فاصبح النموذج بها over specified مما يتطلب حذفها للوصول الى النموذج الوافي الذي قد يكون هو النموذج الافضل اذا ما كانت جميع تقديرات معاملات انحداره معنوية ومستقرة وذلك يتحقق بدقة باستخدام RUF الذي يوشر هذه المتغيرات الزائدة او غيرها من المتغيرات التي معاملات انحدارها (غير مستقرة).

(*) ليس بالضرورة ان تكون جميع المتغيرات المستقلة تفسيرية في النموذج - كما يعتقد البعض - فذلك يتوقف على صحة اختيار الباحث للمتغيرات المستقلة (المؤثرة) في الظاهرة موضوع البحث.



خطوات استخدام RUF للحصول على أفضل نموذج (جزئي)

للاحدار الخطي

7. التطبيق Application: تم اختيار ثلاثة أمثلة متنوعة تعاني فيها المتغيرات التفسيرية من تأثير العلاقات الخطية فيما بينها، وتصل قوة هذا التأثير في أحدها إلى درجة العلاقة التامة تقريبا بين كل متغير والمتغيرات المستقلة الأخرى. وقد كان الهدف من هذا الاختيار هو اختبار المعيار المقترح RUF في الوصول إلى نموذج الانحدار الخطي (الجزئي) الأفضل، وذلك نزولا عند رغبة بعض أساتذة الإحصاء المشاركين في المحاضرة التي أقيمت في الحلقة النقاشية التي أقامها معهد الإدارة/ الرصافة لمناسبة يوم الهيئة السنوي حول المعيار المذكور^(*).

المثال التطبيقي الأول: دراسة أداء (25) مندوب مبيعات لأحدى الشركات، اختيروا بصورة عشوائية من (25) منطقة (أقليما) في الولايات المتحدة الأمريكية، وقد توفرت البيانات كما في الجدول الآتي:

جدول رقم (1) بيانات دراسة أداء مندوبي المبيعات

Territory	Sales y	Time with company X ₁	Market potential X ₂	Advertising X ₃	Market share X ₄	Market share change X ₅	Total no. of account X ₆	Work load X ₇	Rating by Sales – manager X ₈
1.	3669.88	43.10	74065.11	4582.88	2.51	0.34	74.86	15.05	4.9
2.	3473.95	108.13	58117.30	5539.78	5.51	0.15	107.32	19.97	5.1
3.	2295.10	13.82	21118.49	2950.38	10.91	-0.72	96.75	17.34	2.9
4.	4675.56	186.18	68521.27	2243.07	8.27	0.17	195.12	13.40	3.4
5.	6125.96	161.79	57805.11	7747.08	9.15	0.50	180.44	17.64	4.6
6.	2134.94	8.94	37826.94	402.44	5.51	0.15	104.88	16.22	4.5
7.	5031.66	365.04	50935.26	3140.62	8.54	0.55	256.10	18.80	4.6
8.	3367.45	220.32	35602.08	2086.16	7.07	-0.49	126.83	19.86	2.3
9.	6519.45	127.64	46176.77	8846.25	12.54	1.24	203.25	17.42	4.9
10.	4876.37	105.69	42053.24	5673.11	8.85	0.31	119.51	21.41	2.8
11.	2468.27	57.72	36329.71	2761.76	5.38	0.37	116.26	16.32	3.1
12.	2533.31	23.58	33612.67	1991.85	5.43	-0.65	142.28	14.51	4.2
13.	2408.11	13.82	21412.79	1971.52	8.48	0.64	89.43	19.35	4.3
14.	2337.38	13.82	20416.87	1737.38	7.80	1.01	84.55	20.02	4.2
15.	4586.95	86.99	36272.00	10694.20	10.34	0.11	119.51	15.26	5.5
16.	2729.24	165.85	23093.26	8618.61	5.15	0.04	80.49	15.87	3.6
17.	3289.40	116.26	26878.59	7747.89	6.64	0.68	136.58	7.81	3.4
18.	2800.78	42.28	39571.96	4565.81	5.45	0.66	78.86	16.00	4.2
19.	3264.20	52.84	51866.15	6022.70	6.31	-0.10	136.58	17.44	3.6
20.	3453.62	165.04	58749.82	3721.10	6.35	-0.03	138.21	17.98	3.1
21.	1741.45	10.57	23990.82	860.97	7.37	-1.63	75.61	20.99	1.6
22.	2035.75	13.82	25694.86	3571.51	8.39	-0.43	102.44	21.66	3.4
23.	1578.00	8.13	23736.35	2845.50	5.15	0.04	76.42	21.46	2.7
24.	4167.44	58.54	34314.29	5060.11	12.89	0.22	136.58	24.78	2.8
25.	2799.97	21.14	22809.53	3552.00	9.14	-0.74	88.62	27.96	3.9

Source: Applied statistics: Improving Business processes" by BRUCE L.Bowerman & R.T. O'connell, USA, 1997.

(*) اعماما للفائدة حول المعيار المقترح فقد نظمت حلقة نقاشية حوله بتاريخ 2011/3/31 حضرها عدد من اساتذة قسم الإحصاء في كلية الإدارة والاقتصاد في جامعتي بغداد والمستنصرية، فضلا عن ممثل رئيس هيئة التعليم التقني، وقد اقترح عدد من المحاضرين ان يطبق المعيار على أكثر من مثال واحد للتأكد من صلاحيته (كفأته) علما بان المثال الذي طبق عليه في حينه كان مثالا نموذجيا استخدمه العديد من الباحثين الا الجانب لهذا الغرض الا وهو مثال Hald حول الحرارة المنبعثة كدالة لكميات اربعة عناصر في مزيج مادة السمنت).



خطوات استخدام RUF للحصول على افضل نموذج (جزئي) للاحدار الخطي

أ. طريقة كل الانحدارات الممكنة واختيار النموذج الافضل All possible Regressions
لقد استخدم الباحثان Bowerman & Connell في كتابهما Applied statistics
[19] Improving Business processes حزمة برامج SAS في ايجاد كل نماذج الانحدار الممكنة
لبينات الجدول رقم (1) والبالغ عددها (255) نمودجا لكي يختارا من بينها النموذج الافضل على اساس
اصغر MSE واكبر R^2 فحصا على النماذج الثلاث الاتية باعتبارها الافضل من بين ذلك العدد الكبير من
النماذج.

اولا: النموذج بدلالة المتغيرات التفسيرية السبعة الاولى: ومعادلته هي:

$$\hat{y}_i = -1485.881 + 1.975X_1 + 0.037X_2 + 0.152X_3 + 198.308X_4 + 285.866X_5 \\ + 5.610X_6 + 19.899X_7; MSE = 189811.970; RMSE = 435.674; R^2 = .992$$

sig. .043 .287 .000 .003 .007 .090
sig. .234 .550 ; $i = 1, \dots, 25$

ثانيا: النموذج بدلالة المتغيرات التفسيرية الستة الاولى، ومعادلته هي:

$$\hat{y}_i = -1165.479 + 2.269X_1 + 0.038X_2 + 0.141X_3 + 221.605X_4 + 285.109X_5 + 4.378X_6$$

sig. .013 .198 .000 .002 .000 .093 .288
 $MSE = 183187.153; RMSE = 428.004; R^2 = .920$

ثالثا: النموذج بدلالة المتغيرات الخمسة الاولى، ومعادلته هي:

$$\hat{y}_i = -1113.788 + 3.612X_1 + 0.042X_2 + 0.129X_3 + 256.956X_4 + 324.533X_5$$

sig. .016 .006 .000 .003 .000 .053
 $MSE = 185099.475; RMSE = 430.232; R^2 = .915$

ثم استبعدا النموذج (الاول) لكون MSE له كبيرا قياسا بالنموذجين الاخرين على الرغم من ان
المعيار الثاني (R^2) له هو الاكبر، ولو كانا قد استخدمنا R_a^2 لاكتشفا بانه الاصغر مقارنة بالنموذجين الاخرين.
وعلى اية حال فقد انحصرت الافضلية وفق المعيارين المذكورين ما بين النموذج (ثانيا) والنموذج (ثالثا)،
الذي يتميز بان عدد متغيراته التفسيرية اقل. ولكي يحسما هذا الامر استخدمنا معيارا ثالثا هو طول المجال
التنبؤي Prediction Interval Length لاحدى قيم المتغير المعتمد (القيمة الاخيرة) في كلا النموذجين
وبمستوى ثقة 95% فحسلا على مجال ثقة تنبؤي [3426.1, 1489.3] للنموذج (ثانيا) وبطول مجال تنبؤي
مقداره (1936.8) بينما كان مجال الثقة التنبؤي للنموذج (ثالثا) [3456.6, 1520.8] وبطول مجال تنبؤي
مقداره (1935.8) فاستنتجا ان القدرة التنبؤية للنموذج (ثالثا) اكبر منها للنموذج (ثانيا) وعليه اختارا
النموذج (ثالثا) الذي دالته $\hat{y} = f(x_1, \dots, x_5)$ باعتباره هو النموذج الافضل.

ب. طريقة step-wise: وعندما استخدم الباحثان المذكوران طريقة step-wise حسلا على نموذج اخر
ليس من بين النماذج الثلاثة الفضلى المذكورة انفا وفقا لمعايير المفاضلة نفسها، فقد انتجت هذه الطريقة
النموذج الذي دالته $\hat{y} = f(X_2, X_3, X_4, X_6)$ ومعادلته هي:

$$\hat{y}_i = -1441.932 + 0.038X_2 + 0.175X_3 + 190.144X_4 + 9.214X_6$$

sig. .003 .000 .000 .001 .004
 $MSE = 205967.323; RMSE = 453.840; R^2 = .900$

Pred.conf.Interval=[1606.273, 3605.350] & pred. Interval length=1999.077

وهذا يعني ان طريقة step-wise قد فشلت في اختيار النموذج الذي يطابق البيانات ويفسر اكثر من
غيره الاختلافات في قيم متغير الاستجابة (المتغير المعتمد) ويتنبأ بها بصورة دقيقة.



خطوات استخدام RUF للحصول على أفضل نموذج (جزئي)

الانحدار الخطي

ج. طريقة RUF: ان طريقة RUF التي اقترحها الباحث تضمن تحقيق ليس معايير MSE, R^2 او R_a^2 وحسب في تقييم النماذج وانما تضيف اليها Cp و PRESS لكل قيم y وليس لقيمة مختارة منها، وكذلك تضمن تحقيق معاملات انحدار معنوية ومعقولة في آن واحد، كل ذلك في حدود النماذج التي تحقق الفرضيات القياسية لحد الخطأ كما سنلاحظ ذلك فيما يلي:

الدورة الاولى: انحدار y على جميع المتغيرات المستقلة وحساب قيم RUF من العلاقة رقم (1)، فنحصل على مصفوفة الارتباط ما بين المتغيرات، ومعادلة الانحدار، ومؤشراتها الاتية:

	y	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈
y	1								
X ₁	.623	1							
X ₂	.598	.454	1						
X ₃	.596	.249	.174	1					
X ₄	.484	.106	-.211	.264	1				
X ₅	.489	.251	.268	.377	.085	1			
X ₆	.754	.758	.479	.200	.403	.327	1		
X ₇	-.117	-.171	-.259	-.272	.349	-.288	-.199	1	
X ₈	.402	.101	.359	.411	-.024	.549	.229	-.277	1

ويلاحظ من خلال مصفوفة الارتباطات ان الارتباط الخطي بين المتغيرات المستقلة ومتغير الاستجابة هو عموماً متوسط القوة او اقل وكذلك الارتباط الخطي ما بين بعض المتغيرات المستقلة، وعليه فلا يصح اختيار مستوى معنوية دقيق مثل 0.01 لاختبار معنوية النموذج ومعاملات الانحدار فيه وانما يجب اختيار المستوى 0.05 او 0.10 ويفضل 0.10 لكي نسمح لتأثير عدم التحيز في نموذج الانحدار الخطي المتعدد من ان يظهر المعنوية (الحقيقية) لمعاملات انحدار المتغيرات المستقلة التي ارتباطها بمتغير الاستجابة قريب من 0.50 مما يبيحها في النموذج فنزداد مطابقته للبيانات وقدرته التنبؤية، اما معادلة الانحدار ومؤشراتها فهي:

$$\hat{y}_i = -1507.814 + 2.010X_1 + 0.037X_2 + 151X_3 + 199.024X_4 + 290.855X_5 + 5.551X_6 + 19.794X_7 + 8.189X_8$$

s.e (778.635) (1.931) (.008) (.047) (67.028) (186.782) (4.776) (33.677) (128.506)

sig.	.071	.313	.000	.006	.009	.139	.262	.565	.950
riy.	-----	.623	.598	.596	.484	.489	.754	-.117	.402
Beta	-----	.133	.445	.309	.373	.138	.193	+0.055	.006
RUF	-----	-.787	-.256	-.482	-.229	-.718	-.744	-1.480	-.985
VIF	-----	3.343	1.978	1.916	3.236	1.602	5.639	1.818	1.809

MSE=201624.042 ; RMSE=S=449.03 ; R_a^2 =.883 ; SSE=3225985 ; \overline{VIF} =2.667 ; D.W=1.872 ; Sig. Runs Test (eis)=.450 ; Sig. Normality Test (eis)=.503 (Shapiro & wilks) Range (eis)= [-1.224 , 1.757] ⇒

لا توجد نقطة خارج Box-Plot (لاحظ الرسم في الملحق) كذلك فان رسم البواقي مقابل القيم التنبؤية لا يشير الى أي خرق للفرضيات القياسية لحد الخطأ؟ فهل هذا هو النموذج الوافي؟ الجواب سيكون بعد استخدام RUF لحذف المتغير الأقل استقراراً في النموذج وهو هنا المتغير X₇ الذي تشوّهت علاقته بمتغير الاستجابة حيث RUF لمعامل انحداره هو الاكبر سالبية (الأقل قيمة) ثم اجراء الانحدار ثانية وملاحظة تغير قيمة MSE او R_a^2 او اختبار عشوائية البواقي للنموذج الجديد ومدى تحقيقه للفرضيات القياسية لحد الخطأ.



خطوات استخدام RUF للحصول على أفضل نموذج (جزئي)

الانحدار الخطي

الدورة الثانية: انحدار y على المتغيرات المستقلة عدا المتغير المحذوف (X_7) فيكون:

$$\hat{y}_i = -1199.780 + 2.318X_1 + 0.038X_2 + 0.139X_3 + 222.464X_4 + 277.917X_5 + 4.301X_6 + 11.890X_8$$

	<i>s.e</i>	(564.664)	(1.822)	(.008)	(.042)	(52.825)	(181.825)	(4.193)	(125.856)
sig.	.049	.220	.000	.004	.001	.145	.319	.926	
riy.	-----	.623	.598	.596	.484	.489	.754	.402	
Beta	-----	.153	.457	.285	.417	.131	.149	.009	
RUF	-----	-.754	-.236	-.522	-.138	-.732	-.802	-.978	
VIF	-----	3.096	1.902	1.572	2.090	1.579	4.521	1.804	

MSE=193861.097 ; RMSE=S=440.300 ; $R_a^2=.888$; $\overline{VIF}=2.366$; SSE=3295639 ; D.W=1.863 ; Sig. Runs Test (eis)=1 ; Sig. Normality Test (eis)=.282 (eis)=[-1.191 , 1.895]

لا توجد نقطة خارج Box-Plot ولا يوجد خرق للفرضيات القياسية لحد الخطأ، من رسم البواقي مقابل القيم التنبؤية. وحيث ان النموذج في هذه الدورة لم يخرق فرضيات حد الخطأ ولم تزداد قيمة MSE له ولا نقصت قيمة R_a^2 ولا معنوية عشوائية توزيع البواقي فيه، وانما على العكس، قلت قيمة MSE وازدادت قيمة R_a^2 وأصبحت معنوية العشوائية 100% لذلك لا يكون النموذج في الدورة السابقة (الاولى) هو النموذج الوافي ونستمر بحذف المتغير الاقل استقرارا وغير المعنوي في هذه الحالة وهو المتغير X_8 ثم نجري انحدار y على المتغيرات الباقية ونلاحظ تغير MSE او R_a^2 او العشوائية.

الدورة الثالثة: انحدار y على المتغيرات التفسيرية الستة الاولى (بعد حذف X_8, X_7) فيكون:

$$\hat{y}_i = -1165.479 + 2.269X_1 + 0.038X_2 + 0.141X_3 + 221.605X_4 + 285.109X_5 + 4.378X_6$$

	<i>s.e</i>	(420.373)	(1.699)	(.008)	(.038)	(50.583)	(160.560)	(3.999)
sig.	.013	.198	.000	.002	.000	.093	.288	
riy.	-----	.623	.598	.596	.484	.489	.754	
Beta	-----	.150	.458	.288	.415	.135	.152	
RUF	-----	-.759	-.234	-.517	-.143	-.724	-.798	
VIF	-----	2.849	1.843	1.396	2.028	1.303	4.352	

MSE=183187.153 ; RMSE=S=428.004 ; $R_a^2=.894$; $\overline{VIF}=2.295$; SSE=3297369 ; D.W=1.858 ; Sig. Runs Test (eis)=1 ; Sig. Normality Test (eis)=.296 (eis)=[-1.242 , 1.919]

لا توجد نقطة خارج Box-Plot ولا يوجد خرق لفرضيات حد الخطأ، اما قيمة MSE فلم تزداد وانما تناقصت وازداد R_a^2 وحافظت البواقي في النموذج على معنوية عشوائية حول وسطها الحسابي، ولذلك لا يكون النموذج في الدورة السابقة (الدورة الثانية) هو النموذج الوافي وبالنظر لوجود متغيرات تفسيرية معاملات انحدارها غير مستقرة وغير معنوية في النموذج، عليه نستمر باستخدام RUF في حذف المتغير غير المعنوي الاقل استقرارا وهو هنا المتغير X_2 ثم نجري انحدار y على المتغيرات الباقية ونلاحظ تغير MSE او R_a^2 او العشوائية.



خطوات استخدام RUF للحصول على أفضل نموذج (جزئي)

الانحدار الخطي

الدورة الرابعة: انحدار y على المتغيرات الخمسة الاولى (بعد حذف X_8, X_7, X_6) فيكون:

$$\hat{y}_i = -1113.788 + 3.612X_1 + 0.042X_2 + 0.129X_3 + 256.956X_4 + 324.533X_5$$

sig.	(419.887)	(1.182)	(.007)	(.037)	(39.136)	(157.283)
sig.	.016	.006	.000	.003	.000	.053
riy.	-----	.623	.598	.596	.484	.489
Beta	-----	.239	.504	.264	.481	.153
RUF	-----	-.616	-.157	-.557	-.006	-.687
VIF	-----	1.364	1.451	1.286	1.202	1.237

MSE=185099.475 ; RMSE=S=430.232 ; $R^2=.893$; $\overline{VIF}=1.308$; SSE=3516890 ;
D.W=1.762 ; $d_L=.95$, $du=1.890$; Sig. Runs Test (eis)=.650 ; Sig. Normality Test (eis)=.184

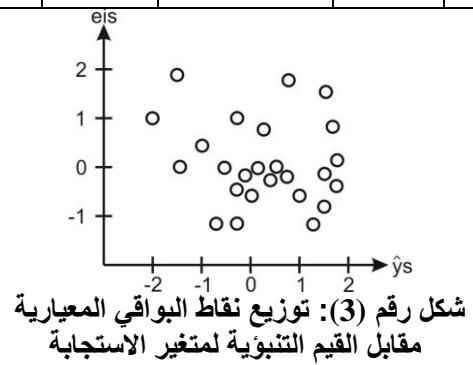
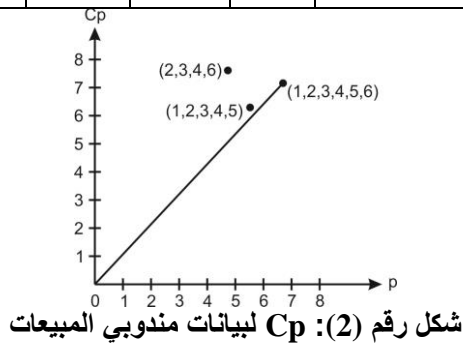
ولا توجد نقطة بيانات خارج Box-plot حيث $[-1.401, 1.703]$ (eis)، والنموذج مازال يحقق كافة فرضيات حد الخطأ. وحيث ان قيمة MSE لهذا النموذج قد ازدادت وقيمة R^2 وقد نقصت ولو قليلا كما تناقصت معنوية التوزيع العشوائي للبواقي حول متوسطها، ولذلك فالنموذج الوافي هو النموذج الذي حصلنا عليه في الدورة السابقة (الدورة الثالثة) لانه يضم (جميع) المتغيرات التفسيرية فقط وهي المتغيرات الستة الاولى وهو ما جعل MSE له هو الاقل مع تحقيقه لكافة الفرضيات القياسية لحد الخطأ، وبذلك يكون $[7=p=Cp]$ اما النموذج في هذه الدورة (الدورة الرابعة) فهو نموذج معنوي (باصغر) MSE (واكبر) R^2 واقل تحيز حيث Cp هي الاقرب الى P مقارنة بكل النماذج الجزئية الاخرى (غير النموذج الوافي) لاحظ الشكل رقم (2) وهو يحقق جميع فرضيات حد الخطأ ويتميز على النموذج الوافي بان المتغيرات التفسيرية فيه معنوية ومستقرة (تقريبا) لان تضخم تبايناتها هو الاصغر حيث بلغ متوسطه ($\overline{VIF}=1.308$) لاحظ الشكل رقم (3) كذلك فان هذا النموذج هو الافضل في التنبؤ بقيم متغير الاستجابة (المتغير المعتمد) حيث بلغ الجذر التربيعي لمجموع مربعات البواقي التنبؤية له (2328.640) وهي الاصغر بين كل نماذج الانحدار الجزئية الممكنة ان هذا كله يعني ان النموذج الافضل هو النموذج الذي حصلنا عليه مباشرة وببساطة باستخدام RUF بينما حصل عليه الباحثان المذكوران بعد ان قيما (255) نموذجا جزئيا للانحدار الخطي. ونلخص فيما يلي ميزات: النموذج الوافي، نموذج Stepwise، ونموذج RUF لغرض المقارنة في الجدول الاتي:



خطوات استخدام RUF للحصول على أفضل نموذج (جزئي)
للانحدار الخطي

جدول رقم (2): ميزات أفضل نماذج الانحدار الجزئية

النماذج	عدد المعلمات المقدر	فرضيات حد الخطأ	VIF	$\rho_{bibj} \geq .50$	معنوية واستقرار معاملات الانحدار	RMSE (S)	RPRESS	R_a^2	Cp
النموذج الوافي $f(X_1, \dots, X_6)$	7	متحققة	2.295	$\rho_{46} = -.638$ $\rho_{16} = -.722$ $\rho_{24} = .514$	بعضها غير معنوية وغير مستقرة	428.004	2459.002	0.894	7
نموذج step-wise $f(X_2, X_3, X_4, X_6)$	5	متحققة	1.678	$\rho_{62} = -.696$ $\rho_{64} = -.579$ $\rho_{42} = .534$	معنوية ومستقرة (نسبيا)	453.840	2409.240	.881	7.487
نموذج RUF النموذج الأفضل $f(X_1, \dots, X_5)$	6	متحققة	1.308	لا يوجد	معنوية ومستقرة	430.232	2328.640	.893	6.198





خطوات استخدام RUF للحصول على أفضل نموذج (جزئي)

الانحدار الخطي

المثال التطبيقي الثاني: وهو دراسة العوامل المؤثرة في مجمل الاستيرادات الفرنسية للفترة من (1959-1949) وكما وردت في الكتاب الموسوم والمترجم Regression Analysis By Examples [20] ونعرضها في الجدول الآتي:

جدول رقم (3) بيانات عن الاستيرادات الفرنسية والعوامل المؤثرة فيها للفترة من (1959-1949) بمليارات الفرنكات الفرنسية

Years	Y Imports	X ₁ Domestic-product	X ₂ Stock-formation	X ₃ Domestic-consumption
1949	15.9	149.3	4.2	108.1
1950	16.4	161.2	4.1	114.8
1951	19.0	171.5	3.1	123.2
1952	19.1	175.5	3.1	126.9
1953	18.8	180.8	1.1	132.1
1954	20.4	190.7	2.2	137.7
1955	22.7	202.1	2.1	146.0
1956	26.5	212.4	5.6	154.1
1957	28.1	226.1	5.0	162.3
1958	27.6	231.9	5.1	164.3
1959	26.3	239.0	0.7	167.6

أ. باستخدام RUF:

الدورة الأولى: انحدار y على جميع المتغيرات المستقلة (X₃, X₂, X₁) وحساب قيم RUF لمعاملات انحدارها من العلاقة رقم (1) المشار إليها انفا فنحصل على مصفوفة الارتباطات ما بين المتغيرات، ومعادلة الانحدار، ومؤشراتها الآتية:

$$\hat{y}_i = -10.128 - .051X_1 + .587X_2 + .287X_3$$

$$s.e. (1.212) (.070) (.095) (.102)$$

	Y	X ₁	X ₂	X ₃
sig.	.00	.488	.000	.026
0				
riy.j	-----	-.266	.920	.728
riy	-----	.965	.251	.972
Beta	-----	-.339	.213	1.303
RUF	-----	-1.351	-.151	.341
VIF	-----	185.99	1.019	186.110

7

MSE=.239 ; $R_a^2=.988$; F=285.610 ; $\sqrt{VIF}=124.375$; $\rho b_1b_3=-.997$; Sig. Runs Test (eis)=.777 ; Sig. Normality Test (eis)=.409 ; eis=[-1.071 , 1.602]



خطوات استخدام RUF للحصول على أفضل نموذج (جزئي)

الانحدار الخطي

ويلاحظ من مدى القيم المعيارية للبواقي ومن Box-Plot عدم وجود نقطة متطرفة ومؤثرة في تحليل الانحدار كما يلاحظ ان النموذج يحقق فرضيات حد الخطأ ولكنه يعاني وبشدة كما تشير مصفوفة الارتباطات من العلاقة الخطية التامة تقريبا ما بين X_1 , X_3 ($r_{13} = .997$) وكما يشير $[RUF > -1]$ الى عدم استقرار X_1 وتشوه علاقته بمتغير الاستجابة مما يقتضي حذفه من النموذج لزيادة قدرته التفسيرية (الصحيحة) وقدرته التنبؤية من جهة، وزيادة معنوية واستقرار معاملات الانحدار الاخرى فيه من جهة اخرى. كما ان معامل الارتباط ما بين معاملي انحدار X_1 , X_3 $[p \ b_1 b_3 = -.997]$ يشير الى ان الميل الحدي لمتغير الاستجابة على احدهما يرافقه الميل الحدي لمتغير الاستجابة على الاخر ولكن باتجاه معاكس يجعل محصلة الميل الحدي لمتغير الاستجابة على أي منهما عند وجودهما معا لا يعبر عن أي منهما لوحده وقد يساوي صفرا وبذلك يكون لا فائدة مطلقا من وجودهما معا في النموذج ماداما متغيرين (معتمدين) وليس (مستقلين) وهو ما يستوجب ايضا حذف احدهما حيث تحدد قيم RUF المتغير الذي يجب حذفه من بينهما وهو X_1 لانه المتغير الاكثر تأثيرا بهذه العلاقة الخطية أي الاكثر عدم استقرار كما يشير $RUF_{(1)}$ الذي يساوي (-1.351) . وبالنظر لعدم وجود علاقة خطية (تذكر) ما بين X_2 , X_3 ($r_{23} = .036$) ولقوة الارتباط ما بين y_2 ($r_{3y} = .972$) و X_3 , لذلك يمكننا ان نحدد مستوى معنوية دقيق مثل 0.01 لاختيارهما واختبار النموذج بدالتهما.

الدورة الثانية: اجراء انحدار y على X_2 , X_3 (بعد حذف X_1 من النموذج) فيكون:

$\hat{y}_i = -9.743 + .596X_2 + .212X_3$	MSE=.225	$R_a^2 = .989 \cong .99$	$VIF = 1.001$
s.e. (1.059) (.091) (.007)	$\rho \ b_2 b_3 = -.036$	$V(b_2) \cong .008$	$V(b_3) = .000$
sig. ----- .000 .000	Sig. Runs Test (eis) = .777 ;		
riy.j ----- .918 .995	Sig. Normality Test (eis) = .727 ;		
riy ----- .251 .972	eis-Range = [-1.420 , 1.558] ;		
Beta ----- .216 .964	F = 454.581 ; sig. model = .000		
RU ----- -.139 -.008			
F			
VIF ----- 1.001 1.001			

كما ان رسم البواقي المعيارية (eis) مقابل القيم التنبؤية لمتغير الاستجابة \hat{y} يشير الى تحقيق النموذج لكافة فرضيات حد الخطأ (لاحظ الشكل رقم (4)) وحيث ان قيمة MSE لم تزداد وانما تناقصت بحذف X_1 وكذلك ازدادت قيمة R_a^2 بينما حافظ توزيع البواقي على معنوية عشوائيته، وحافظ النموذج على تحقيقه للفرضيات القياسية لحد الخطأ، لذلك فان النموذج في الدورة السابقة (الدورة الاولى) بدلالة جميع المتغيرات المستقلة ليس هو النموذج الوافي. ولكون معامل انحدار كل من X_2 , X_3 في دالة النموذج $\hat{y} = f(X_2, X_3)$ في الدورة الثانية هو مستقر تماما ومعنوي جدا لذلك فالنموذج في هذه الدورة الذي حصلنا عليه باستخدام RUF هو النموذج الوافي حيث $[3 = \rho = Cp]$ وهو النموذج الافضل للانحدار الخطي لكونه نموذج معنوي بمعاملات معنوية ومستقرة تماما (تقريبا) وغير متحيزة في ان واحد. ب. باستخدام step-wise في حزمة برامج SPSS نحصل على نفس معادلة الانحدار التي حصلنا عليها باستخدام RUF. (لاحظ التفاصيل في الملحق).

وجدير بالذكر انه لمن يريد التأكد من ان RUF يحدد المتغير الذي يجب حذفه من النموذج من بين المتغيرات المستقلة التي ترتبط فيما بينها بعلاقات خطية قوية، لنجرب في مثالنا هذا حذف X_3 بدلا من X_1 وذلك خلافا لما يشير اليه RUF وعند ذلك نحصل على معادلة الانحدار بدلالة X_1 , X_2 الاتية:

$$\hat{y} = f(X_1, X_2) = -8.440 + .145X_1 + .622X_2$$

s.e. (1.435) (.007) (.128)

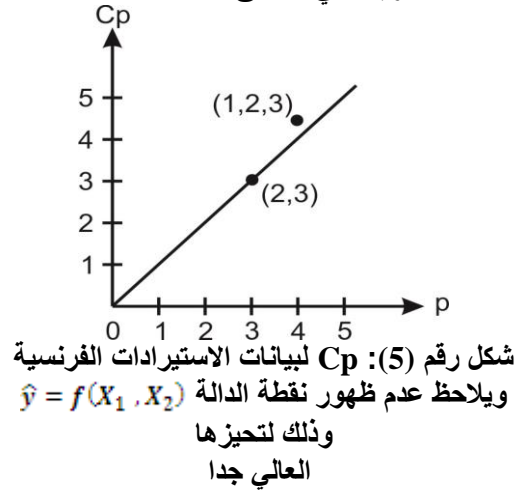
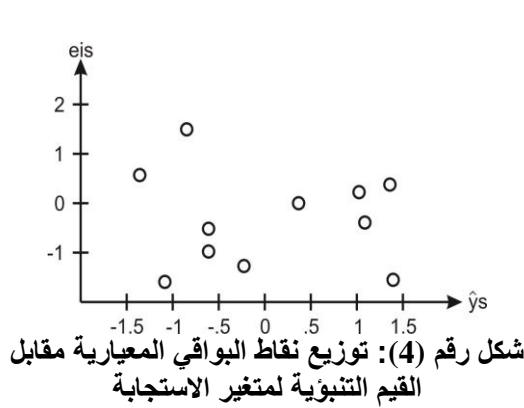
MSE = .444 ; $R_a^2 = .978$; $VIF = 1.001$; $Cp = 10.8$; $\rho = 3 \Rightarrow Bias = 7.8$
Sig. Normality Test = .154 ; sig. Runs Test = 1



خطوات استخدام RUF للحصول على أفضل نموذج (جزئي)

الانحدار الخطي

وبالرغم من ان النموذج يحقق فرضيات حد الخطأ ومعاملاته معنوية ومستقرة الا ان مطابقته للبيانات اصبحت اقل بنسبة 50% عما كانت عليه عند استخدام نموذج RUF كذلك فان تقدير معاملات الانحدار اصبحت متحيزا وبصورة كبيرة لا يمكن قبولها [Bias = 7.8] حسب احصاءة Mallows فضلا عن تناقص كبير في قدرة هذا النموذج التنبؤية حسب احصاءة Allen [PRESS]، لاحظ الشكل رقم (5) مع الجدول الاتي الذي يلخص ميزات النماذج الثلاثة، اما قيم البواقي التنبؤية prediction of residuals للنماذج الثلاثة فقد ادرجت في الملحق.



جدول رقم (4): ميزات افضل نماذج الانحدار الخطية (الجزئية)

النماذج	عدد المعلمات المقدره	فرضيات حد الخطأ	VIF	$ \rho_{bij} \geq .50$	معنوية واستقرار معاملات الانحدار	MSE (S)	RPRESS	R_a^2	Cp
$\hat{y} = f(X_1, X_2, X_3)$	4	متحققة	124.375	$\rho_{1b3} = -.997$	بعضها غير معنوية وغير مستقرة على الاطلاق	.239	4.790	.988	4.436
$\hat{y} = f(X_2, X_3)$	3	متحققة	1.001	لا يوجد	معنوية ومستقرة	.225	3.224	.989	3
$\hat{y} = f(X_1, X_2)$	3	متحققة	1.001	لا يوجد	معنوية ومستقرة	.444	7.361	.978	10.8

المثال التطبيقي الثالث: وهو مثال صارخ على المتغيرات المستقلة التي تعاني جميعها من العلاقات الخطية القوية جدا فيما بينها حتى تبدو وكأنها جميعا ليست الا متغير واحد فقط، فكل منها ولوحده يمكنه ان يفسر اكثر من 90% من التغيرات في الظاهرة موضوع البحث والتي هي تقدير الحاجة الفعلية للعمالة في (17) مستشفى اختيروا بصورة عشوائية، وذلك بدلالة عدد ساعات العمل المطلوبة شهريا التي تمثل متغير الاستجابة (المتغير المعتمد y) والتي تحدها خمسة عوامل (متغيرات مستقلة) هي الاتية:

- X_1 : معدل عبا العمل اليومي لكل مريض Average daily patient load
 X_2 : عدد الذين يحتاجون الى فحص باشعة اكس شهريا Monthly X-ray Exposures
 X_3 : معدل عدد الاسرة المشغولة Monthly occupied Bed Days
 X_4 : عدد السكان في المنطقة المؤهلين لاعمال المستشفى Eligible population
 X_5 : معدل طول مدة البقاء في المستشفى Average length of stay
والجدول الاتي يعرض بيانات المتغير المعتمد والمتغيرات المستقلة:



خطوات استخدام RUF للحصول على أفضل نموذج (جزئي)

للانحدار الخطي

جدول رقم (5) بيانات تقدير حجم العمالة المطلوبة في المستشفيات

Hospital	X ₁	X ₂	X ₃	X ₄	X ₅	y
1	15.57	2463	472.92	18.0	4.45	566.52
2	44.02	20.48	1339.75	9.5	6.92	696.82
3	20.42	3940	620.25	12.8	4.28	1033.15
4	18.74	65.5	568.33	36.7	3.90	1603.62
5	40.20	5723	1497.60	35.7	5.50	1611.37
6	44.92	11520	1365.83	24.0	4.60	1613.27
7	55.48	5779	1687.00	43.3	5.62	1854.17
8	59.28	5969	1639.92	46.7	5.15	2160.55
9	94.39	8461	2872.33	78.7	6.18	2305.58
10	128.02	20106	3655.08	180.5	6.15	3503.93
11	96.00	13313	2912.00	60.9	5.88	3571.89
12	131.42	10771	3921.00	103.7	4.88	3741.40
13	127.21	15543	3865.67	126.8	5.50	4026.52
14	252.90	36194	7684.10	157.7	7.00	10343.81
15	409.20	34703	12446.33	169.4	10.78	11732.17
16	463.70	39204	14098.40	331.4	7.05	15414.94
17	510.22	86533	15524.00	371.6	6.35	18854.45

Source: Procedures and analysis for staffing standards development: Regression Analysis handbook (San Diego, CA: Navy man power and material analysis-center, 1979).

أ- استخدام كل الانحدارات الممكنة (All possible Regressions):

لو اجرينا انحدار Y على جميع المتغيرات التفسيرية معا وبجميع نقاط بياناتها فاننا نكتشف وجود نقطة بيانات متطرفة هي النقطة (14) التي لها باقي معياري كبير قياسا ببقية البواقي المعيارية، وهي تؤثر في قوة النموذج التفسيرية والتنبؤية وفي التوزيع الطبيعي للبواقي المعيارية وقوة ارتباطها الذاتي مما يمنع الحصول على النموذج السوافي وفقا لمواصفاته المذكورة في الجانب النظري من هذا البحث (بحق الفرضيات القياسية لحد الخطأ وله اصغر MSE)، فضلا عن تشويه كافة المعايير والمؤشرات الاحصائية حول معنوية التقديرات واستقرارها ومقارنة النماذج. وقد قام الباحثان O'Connell & Bowerman في كتابهما المشار اليه انفا- في المثال التطبيقي الاول- باجراء كل الانحدارات الممكنة ومقارنة (31) نمودجا خطيا تمثل هذه البيانات (لاحظ التفاصيل في الملحق) فحصلنا على دالة النموذج الخطي الجزئي $\hat{Y} = f(X_2, X_3, X_5)$ باعتباره النموذج الافضل وفقا للمعايير المذكورة اما معادلته التي تمثل تقديرات معلمات النموذج ومؤشراتها فهي كما يلي:

$$\hat{y} = 1523.389 + 0.053X_2 + 0.978X_3 - 320.951X_5 \quad n=17$$

s.e	(786.898)	(.020)	(.105)	(153.192)	RMSE=614.779
sig.	.075	.021	.000	.056	RPRESS=4224.537
riy.j	-----	.590	.932	-.502	$R^2_a = .988$; $VIF = 7.166$
riy.	-----	.945	.986	+.579	$\rho_{23} = -.916$; $\rho_{35} = -.706$
Beta	-----	.203	.863	-.091	$\rho_{25} = .520$, $C_p = 2.92$
RUF	-----	-.785	-.125	-1.157	D.W=2.546
VIF	-----	7.737	11.269	2.493	

4-D.W = 1.454 ; d_L = .900 ; d_U = 1.710 ; sig. Run test (eis) = .605 sig. Normality test (eis)

shapiro & wiks

→ .023 ; skewness = 1.593 ; kurtosis = 3.739 ; eis-Range = [-1.118 , 2.652]

→ Data point (14) is outside Box-plot.



خطوات استخدام RUF للحصول على أفضل نموذج (جزئي)

الانحدار الخطي

فهل يمكننا القول مع هذين الباحثين- في ضوء هذه المؤشرات الاحصائية - بان هذا النموذج هو نموذج الانحدار الخطي (الجزئي) الافضل؟ الاجابة القاطعة والواضحة هي كلا، لسبب بسيط هو ان نموذج الانحدار الخطي الافضل يجب ان لا يخرق الفرضيات القياسية لحد الخطأ، وهذا النموذج كما هو واضح من المؤشرات الاحصائية ومن رسم البواقي المعيارية مقابل القيم التنبؤية لمتغير الاستجابة (لاحظ شكل رقم (6)) لا تتوزع بواقية توزيعا طبيعيا وهي تعاني من مشكلة الارتباط الذاتي، وبذلك تكون تقديرات الاخطاء المعيارية ومعنوية هذه المعاملات فضلا عن قيم معايير المقارنة جميعها مشكوكا فيها ولا يصح اعتمادها. هذا من جانب معنوية النموذج ومعنوية تقديرات معاملات الانحدار اما من جانب استقرار هذه التقديرات ومعقوليتها الذي يمثل الجانب الاخر من النموذج الافضل فان النموذج الذي اختاره الباحثان على اساس المعايير المذكورة [Cp , R_a^2 , RPRESS , RMSE] يعاني بوضوح من عدم استقرار مقدرات معاملات الانحدار فيه، ومؤشرات ذلك هي ما يلي:

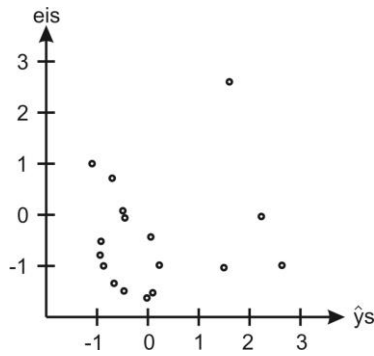
1. تشوه معامل انحدار y على X_5 فبدلا من العلاقة الطردية بينهما اصبحت العلاقة عكسية اذ تغيرت اشارة معاملي الانحدار المعياري للمتغير X_5 في نموذج الانحدار الخطي المتعدد ونموذج الانحدار الخطي البسيط [Beta = -.091 , $r_{5y} = +.579$] وبذلك اعطى معامل الانحدار الجزئي في النموذج تفسيراً غير معقول لهذه العلاقة، فزيادة مكوث المريض الواحد في المستشفى يوما واحدا يتطلب ان (تقل) ساعات العمل المطلوبة في المستشفى شهريا بمقدار (321) ساعة تقريبا أي يجب ان (تقل) بمعدل (11) ساعة تقريبا في اليوم!! وهذا تفسير غير معقول يقدمه النموذج المذكور.

2. ان قيمة

$$-1 > RUF(b_5) \Leftrightarrow \text{وبالتحديد } RUF(b_5) = -1.157$$

3. ان قيمة $VIF(b_3) = 11.269 < 10$ كذلك فان $\overline{VIF} = 7.166$

4. ان قوة الارتباط كبيرة ما بين معاملات الانحدار في النموذج: $\rho_{25} = .520$; $\rho_{35} = -.706$; $\rho_{23} = -.916$ ولذلك فمعاملات الانحدار جميعها (غير مستقرة)، واحدها (غير مستقر وغير معقول ايضا) وهذا يعني ان هذا النموذج وبالرغم من حصوله على افضل قيم معايير المقارنة الا انه نموذج غير مقبول لتمثيل البيانات لانه يخالف فرضيات حد الخطأ كما ان تقديرات معلماته غير مستقرة وغير معقولة.





خطوات استخدام RUF للحصول على أفضل نموذج (جزئي)

الانحدار الخطي

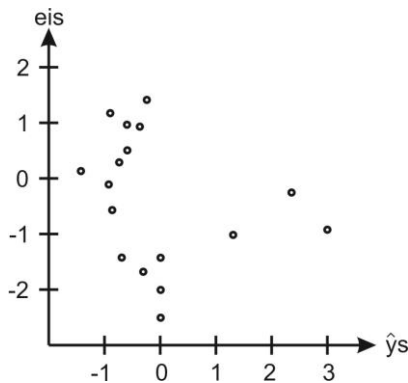
كما استخدم الباحثان المذكوران المعايير ذاتها لاختبار النموذج الجزئي الأفضل بطريقة كل الانحدارات الممكنة وطريقة Step-Wise أيضا بعد حذف نقطة البيانات (14) المتطرفة والمؤثرة، وقد توصلنا الى نفس النموذج الذي دالته $\hat{y} = F(X_2, X_3, X_5)$ وبذلك اقتنعنا ان هذا النموذج هو الأفضل، اما معادلته ومؤشرات تقديراته فهي كما يلي: (*)

شكل رقم (6): توزيع البواقي المعيارية مقابل القيم التنبؤية لمتغير الاستجابة

\hat{y}	$= 1946.802 + .039X_2 + 1.039X_3 - 413.758X_5$	$n=16$; $RMSE=387.160$
$s.e$	(504.182) (.013) (.068) (98.598)	$R^2=.996$; $R_a^2=.995$
sig.	.002 .012 .000 .001	$VIF=7.248$; $\rho_{23}=-.919$;
riy.	----- .943 .989 +.560	$\rho_{35}=-.718$; $\rho_{25}=.543$
Beta	----- .149 .933 -.120	$D.W=2.334$; $4-D.W.=1.666$
RUF	----- -.842 -.057 -1.214	$d_L=.860$; $d_U=1.730$
VIF	----- 7.828 11.396 2.520	

Sig, Runs Test (eis)=.844 sig. Normality test (eis) $\xrightarrow{\text{shapiro \& wiks}}$.806 ; skewness = -.372 ; kurtosis = -.666 ; eis-Range = [-1.749 , 1.337] ; $F=1028.131$; sig. of Model = .000

وبملاحظة رسم البواقي مقابل القيم التنبؤية لمتغير الاستجابة الشكل رقم (7) تبدو واضحة العلاقة بينهما، كذلك تبدو العلاقة واضحة عند رسم كل من هذه المتغيرات المستقلة في النموذج مقابل قيم البواقي المعيارية مما يدل صراحة على خرق ما لا يقل عن فرضيتين اساسيتين من الفرضيات القياسية لحد الخطأ، وهذا يعني ان حد الخطأ كان له دور في التفسير والتنبؤ بقيم متغير الاستجابة خلافا للفرضيات وهو ما اظهر قيمة $RMSE$ صغيرة كتقدير لتباين الخطأ الحقيقي وجعل قيمة R_a^2 له قريبة جدا من الواحد (.995) وقيمة



شكل رقم (7): توزيع نقاط البواقي المعيارية

مقابل القيم التنبؤية لمتغير الاستجابة

وانه لكل يوم اضافي يمكنه المريض في المستشفى (نقل) الحاجة الى ساعات العمل فيه بمقدار (414) ساعة شهريا أي بمعدل (14) ساعة يوميا!! وهذا تفسير غير معقول على الاطلاق، ولذلك فالنموذج غير مقبول لكونه يخرق فرضيات حد الخطأ من جهة ولعدم استقرار ومعقولية معاملات الانحدار فيه من جهة اخرى.

(*) لاحظ تفاصيل قيم المعايير لكافة نماذج الانحدار الممكنة في الملحق.



خطوات استخدام RUF للحصول على أفضل نموذج (جزئي)

الانحدار الخطي

ب- استخدام RUF:

حيث ان الدورة الاولى لانحدار y على جميع المتغيرات التفسيرية وبكافة نقاط بياناتها قد اظهرت النقطة (14) متطرفة ومؤثرة على نتائج التحليل لذلك سوف نستبعدا ونعتبر حجم العينة (16) مستثنى بدلا من (17) فيكون:

الدورة الاولى: انحدار y على جميع المتغيرات التفسيرية وحساب قيم RUF لمعاملات انحدارها وفق العلاقة (1) والحصول على مصفوفة الارتباط بين المتغيرات، ومعادلة الانحدار ومؤشراتها الاتية:

	X ₁	X ₂	X ₃	X ₄	X ₅	y
X ₁	1	.90738	.99990	.93569	.67120	.98565
X ₂	.90738	1	.90715	.91048	.44665	.94517
X ₃	.99990	.90715	1	.93317	.67111	.98599
X ₄	.93569	.91047	.93317	1	.46286	.94036
X ₅	.67120	.44665	.67111	.46286	1	.57858
y	.98565	.94517	.98599	.94036	.57858	1

بالنظر للارتباطات الخطية القوية جدا ما بين المتغيرات التفسيرية التي تؤدي الى تضخم تباينات تقديرات معاملات انحدارها بصورة كبيرة جدا تؤدي الى عدم استقرارها وعد معقوليتها وعدم معنويتها، لذلك لابد من استخدام مستوى المعنوية (0.10) للاختبار لكي نسمح بدخول متغيرين في الاقل في نموذج الانحدار الخطي (الجزئي) الافضل).

\hat{y}	$= 2270.415 + 9.297X_1 + 0.041X_2 + 1.413X_3 - 3.223X_4 - 467.861X_5$
s.e	(670.786) (60.810) (.014) (1.925) (4.474) (131.627)
sig.	.007 .882 .013 .480 .488 .005
riy.j	----- -.048 .689 .226 -.222 -.747
riy	----- +.989 .943 .989 +.947 +.560
Beta	----- -.274 .159 1.269 -.064 -.135
RUF	----- -1.277 -.831 .283 -1.068 -1.241
VIF	----- 9334.456 8.077 8684.208 23.005 4.213

RMSE = S = 399.712 ; SSE = 1597695 ; $R_a^2 = .995$; $\overline{VIF} = 3610.792$; $\rho_{13} = -.999$; $\rho_{41} = -.672$; $\rho_{43} = .643$; $\rho_{45} = .632$; D.W = 2.513 ; du = 2.150 ; Sig. Run Test (eis) = .796 ; sig. ormality Test (eis) $\xrightarrow{\text{shapiro \& wiks}}$.856 ; eis-Rang = [-1.682 , 1.210] ; F = 578.995 ; sig. of the model = .000

ان المؤشرات الاحصائية تشير الى التوزيع الطبيعي والعشوائي للبواقي الا ان رسم البواقي مقابل القيم التنبؤية لمتغير الاستجابة (لاحظ الشكل رقم (8)) يشير الى خرق لاحدى فرضيات حد الخطأ اما تقديرات معاملات الانحدار فالمؤشرات الاحصائية تشير بوضوح الى كونها جميعا (تقريبا) غير مستقرة وغير معقولة بسبب علاقتها الخطية القوية التي تصل الى حد الارتباط التام (تقريبا) ما بين X_3, X_1 واستنادا الى قيم RUF لمعاملات الانحدار فان $RUF(b_1)$ هو الاكبر سالبية (الاقبل قيمة) لذلك يكون X_1 هو المتغير الذي يجب حذفه من النموذج لانه الاقل استقرارية ومعقولة ولان وجوده يضعف من قوة النموذج التفسيرية والتنبؤية ويؤثر في تضخم تباينات معاملات انحدار المتغيرات المستقلة الاخرى.



خطوات استخدام RUF للحصول على أفضل نموذج (جزئي)

الانحدار الخطي

الدورة الثانية: انحدار y على المتغيرات المستقلة عدا المتغير المحذوف X_1 وحساب قيم RUF فيكون:
 $\hat{y} = f(X_2, X_3, X_4, X_5) = 2311.275 + .041X_2 + 1.119X_3 - 3.682X_4 - 477.169X_5$

s.e	(587.301)	(.013)	(.095)	(3.163)	(111.398)
sig.	.002	.009	.000	.269	.001
riy.j	-----	.689	.962	-.331	-.791
riy	-----	.943	.989	+.947	+.560
Beta	-----	.159	1.005	-.073	-.138
RUF	-----	-.831	.016	-1.077	-1.246
VIF	-----	8.067	23.367	12.624	3.311

RMSE = S = 381.555 ; SSE = 1601430 ; $R_a^2 = .995$; $\overline{VIF} = 11.842$ D.W = 2.547 ; 4-D.W = 1.453 ; du = 1.930 ; sig = Runs Test (eis) = .796 ; $\rho_{35} = -.788$; $\rho_{32} = -.509$; $\rho_{34} = -.716$; sig. Normality Test (eis) = .795

ما زال النموذج يخرق الفرضيات القياسية لحد الخطأ وما زالت تقديرات معاملات انحداره حسب المؤشرات الاحصائية - غير مستقرة وغير معقولة، لذلك نستمر بحذف المتغيرات وفقاً لقيمة RUF الأكبر سالبة ولذلك نحذف المتغير X_5 .

الدورة الثالثة: انحدار y على المتغيرات المستقلة عدا المتغيرين المحذوفين (X_5 , X_1)، وبحساب قيم RUF يكون:

$$\hat{y} = f(X_2, X_3, X_4) = -138.978 + .063X_2 + .797X_3 + 2.943X_4$$

s.e	(208.113)	(.019)	(.092)	(4.315)
sig.	.517	.006	.000	.508
riy.j	-----	.692	.929	.193
riy	-----	.943	.989	.947
Beta	-----	.241	.716	.059
RUF	-----	-.744	-.276	-938
VIF	-----	6.885	8.874	9.605

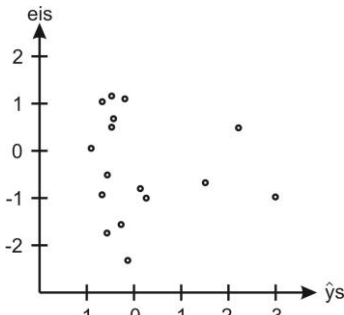
RMSE = S = 596.701 ; $R_a^2 = .988$; SSE = 4272625 ; F = 430.512 ; sig.Model = .000 ; $\overline{VIF} = 8.455$; $|P_{bi\ bj}| \geq .50 \Rightarrow \rho_{34} = -.615$; $\rho_{24} = -.446$; $\rho_{23} = -.364$; D.W = 2.091 ; 4-D.W = 1.909 ; du = 1.780 ; sig. Runs Test (eis) = 1 ; sig. Normality Test (eis) = shap. & wilks = .977) kolm. & semir. = 1) ; skewness = -.246 ; kurtosis = -.301 ; eis Range = [-1.777 , 1.484]

ويلاحظ من المؤشرات الاحصائية ان هذا النموذج يحقق جميع الفرضيات القياسية لحد الخطأ، فتوزيع البواقي طبيعي جدا ولا يوجد ارتباط ذاتي فيما بينها ولا ما بينها وبين المتغيرات المستقلة او القيم التنبؤية لحد الخطأ فمعنوية عشوائية توزيع البواقي حول متوسطها 100% (تساوي 1) والرسم يؤكد ذلك (لاحظ الشكل رقم (9)) فهل هذا هو النموذج الوافي؟ لكي نجيب عن هذا السؤال، وحيث ان لدينا متغيراً مستقلاً بمعامل انحدار غير معنوي وغير مستقر هو المتغير X_4 [RUF(b_4) = -.938] لذلك نحذفه من النموذج ونلاحظ تغير قيم RMSE او R_a^2 او عشوائية توزيع البواقي.

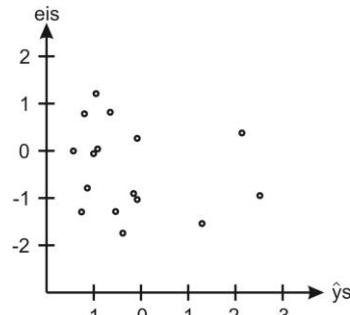


خطوات استخدام RUF للحصول على أفضل نموذج (جزئي)

الانحدار الخطي



شكل رقم (8): البواقي المعيارية مقابل القيم التنبؤية



شكل رقم (9): البواقي المعيارية مقابل القيم التنبؤية

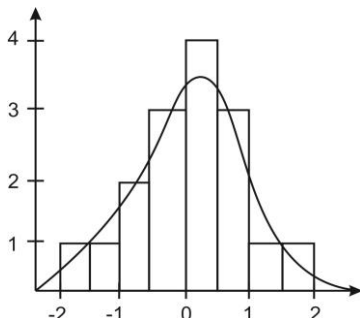
الدورة الرابعة: انحدار y على المتغير التفسيري (X₂, X₃) فقط، مع حساب قيم RUF نحصل على ما يلي:

$$\hat{y} = f(X_2, X_3) = -98.158 + .068X_2 + .836X_3 \quad n = 16 ; RMSE = S = 584.300 ;$$

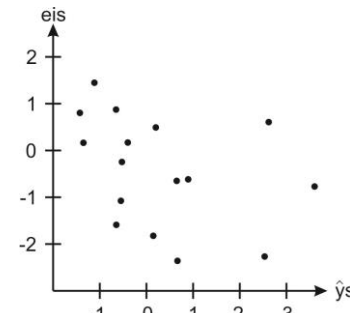
s.e	(195.180)	(.017)	(.071)	$R_a^2 = .989$; SSE = 4438288 ;
sig.	.623	.001	.000	$\overline{VIF} = 5.516$ $\rho b_2 b_3 = -.905$;
riy.j	-----	.754	.956	D.W = 2.323 ; 4-D.W = 1.677 ;
riy	-----	.943	.989	du = 1.540 \Rightarrow 4-D.W > du sig.
Beta	-----	.263	.751	Runs Test (eis) = 1 ;
RUF	-----	-.721	-.241	sig. Normality Test = .977 ;
VIF	-----	5.516	5.516	skewness = .057 ; kurtosis = -

.188 ; eis-Range = [-1.613 , 1.800]

ويلاحظ من المؤشرات الاحصائية ان هذا النموذج يحقق جميع فرضيات حد الخطأ وان معاملات الانحدار فيه جميعها معنوية ومستقرة تقريبا كما ان تقدير ثابت التقاطع (معقول) لانه (غير معنوي)، وان قيمة RMSE للنموذج اصغر مما كانت في النموذج السابق - الذي كان يحقق ايضا فرضيات حد الخطأ - كذلك فان قيمة R_a^2 اكبر اما معنوية عشوائية توزيع البواقي المعيارية حول متوسطها فلم تتغير فهي تساوي (1)، لذلك يمكننا القول ان هذا النموذج الوافي الذي نبحت عنه حيث فيه $[3 = \rho = Cp]$ ، وهو ايضا النموذج الجزئي الافضل لتمثيل البيانات لان تقديرات معاملات انحداره معنوية وغير متحيزة ومعقولة ومستقرة (تقريبا)، ورسم البواقي المعيارية مقابل تكراراتها ومقابل القيم التنبؤية لمتغير الاستجابة يؤكد ذلك (لاحظ الشكل رقم (10) والشكل رقم (11)).



شكل رقم (10): رسم المدرج والمنحني التكراري للبواقي المعيارية



شكل رقم (11): رسم توزيع البواقي المعيارية مقابل القيم التنبؤية لمتغير الاستجابة



خطوات استخدام RUF للحصول على أفضل نموذج (جزئي) للانحدار الخطي

- ان (معقولة) تقديرات معلمات هذا النموذج يمكن ايجازها بما يلي:
1. ان معنوية ثابت التقاطع (bo) -صفرية- أي لا رصيد له في المجتمع الاحصائي، فالقيمة المتوقعة لساعات العمل التي يحتاجها المستشفى تساوي (صفر) عند عدم وجود مرضى يراجعونه (X_2, X_3) وهذا تفسير منطقي ومعقول.
 2. ان زيادة عدد المرضى الذين تؤخذ لهم اشعة اكس، بمقدار مريض واحد (X_2) او زيادة مدة بقاء المريض الراقد في المستشفى يوميا اضافيا (X_3) يؤدي الى زيادة ساعات العمل المطلوبة فيه، وهذا ايضا امر بديهي ومعقول... ولذلك فهذا النموذج مقبول وفقا للمنطق والمؤشرات الاحصائية والرسوم البيانية. ونلخص في الجدول الاتي ميزاته مقارنة بميزات النموذج الذي حصل عليه الباحثان المذكوران انفا بطريقتي كل الانحدارات الممكنة، وطريقة step-wise.

جدول رقم (6) ميزات RUF ونموذج step-wise

النموذج	فرضيات حد الخطأ	معنوية واستقرار معاملات الانحدار	تقدير ثابت التقاطع	VIF	تحيز النموذج ^(*)	RMSE	RPRESS ^(*)	R _a ²
Stepwise & All possible Regressions $\hat{y} = f(X_2, X_3, X_5)$	غير متحققة	معنوية ولكن غير مستقرة وغير معقولة	معنوي ولكن غير معقول	7.248	متحيز Cp=-2.731 $\rho = 4$	387.160	2053.204	.995
RUF $\hat{y} = f(X_2, X_3)$	متحققة جميعها وبصورة مثالية	معنوية ومعقولة ومستقرة (تقريبا)	غير معنوي (صفر) ومعقول جدا	5.516	غير متحيز Cp = $\rho = 3$	584.3	3243.945	.989

(*) لاحظ التفاصيل في الملحق

(**) مقدار Cp او التحيز الوارد في الجداول (2, 4, 6) للامثلة التطبيقية قد حسب وفقا لمفهوم النموذج الوافي الذي عرضه الباحث في الجانب النظري، وليس على اساس انه مجرد النموذج الذي يضم جميع المتغيرات (المستقلة)، وذلك توخيا للدقة.



خطوات استخدام RUF للحصول على أفضل نموذج (جزئي)

الانحدار الخطي

8- الاستنتاجات والتوصيات Conclusions & suggestions

أ. الاستنتاجات Conclusions

1. ان استخدام RUF يجب ان يسبقه استبعاد النقطة (النقاط) التي يواقيها المعيارية كبيرة قياسا بعموم نقاط البيانات موضوع البحث، لان عدم استبعادها يمكن ان يؤدي الى عدم الحصول على النموذج الوافي، فضلا عن اضعاف القوة التفسيرية والتنبؤية لنموذج الانحدار.
2. ان استخدام RUF هو لغرض التخلص من التأثيرات السلبية للعلاقات الخطية القوية ما بين المتغيرات (المستقلة) من دون اضعاف للقوة التفسيرية والتنبؤية لنموذج الانحدار الخطي غير المتحيز وذلك لانه يستهدف حذف المتغيرات التي تؤدي وجودها في نموذج الانحدار الخطي المتعدد الى تشويه تفسيره للعلاقة (الدالية) ما بين متغير الاستجابة والمتغيرات (المستقلة) او اضعاف مغنوية هذه الدالة ومطابقتها للبيانات.
3. ان استخدام RUF في حالة وجود العلاقات الخطية القوية ما بين واحد او اكثر من المتغيرات (المستقلة) يؤدي بالنتيجة للوصول الى نموذج الانحدار الخطي (الجزئي) الذي يحقق جميع الفرضيات القياسية لحد الخطأ والذي تكون معاملات انحدار المتغيرات التفسيرية فيه (مغنوية) و (مستقرة) و (معقولة) و (غير متحيزة او قليلة التحيز) في ان واحد.
4. ان استخدام RUF يكون في مديين هما:
 - $[-1 < RUF \leq -0.5]$ ، وفي ذاك يحذف المتغير الذي معامل انحداره (غير مغنوي) والذي قيمة RUF لذلك المعامل هي الاصغر (الاكبر سالبية) من بين المتغيرات التي معاملات انحدارها (غير مغنوية) فقط.
 - $[RUF \leq -1]$ ، وفي ذلك يحذف المتغير الذي قيمة RUF لمعامل انحداره هي الاصغر (الاكبر سالبية) من بين (جميع) قيم RUF لمعاملات الانحدار.
5. ان الدورة الاولى لانحدار متغير الاستجابة على المتغيرات (المستقلة) تكتسب اهمية خاصة لانها تكشف في كل برامج الانحدار (الجاهزة) عن وجود النقطة (النقاط) التي يؤثر وجودها في التوصل لافضل نموذج انحدار خطي (جزئي)، فضلا عن العلاقات الخطية ما بين المتغيرات (المستقلة) وبينها وبين متغير الاستجابة ولو بصورة اولية وبما يسمح بتحديد جيد لمستوى المغنوية لاختبار النموذج ومعاملات الانحدار فيه.
6. ان تقييم أي نموذج انحدار خطي يجب ان يتضمن شقين هما:
 - تحقيقه لجميع الفرضيات القياسية لحد الخطأ.
 - مغنويته (قدرته التفسيرية والتنبؤية) العالية، ومغنوية واستقرار ومعقولية معاملات الانحدار فيه وبدون ذلك فان معايير تقييم النماذج (MSE , PRESS , R_a^2 , Cp) تكون مضللة misleading للباحث حتى وان استخدم طريقة كل الانحدارات الممكنة.
7. ان النموذج الوافي هو ليس (دانما) مجرد النموذج الذي يضم جميع المتغيرات (المستقلة) التي اختارها الباحث لتفسير الظاهرة موضوع البحث، لانه قد يكون من بينها متغير واحد او اكثر (غير تفسيري) في نموذج الانحدار الخطي المتعدد وقد يؤدي الى ان لا يحقق النموذج جميع الفرضيات القياسية لحد الخطأ وهو ما يجعل تقدير تباين الخطأ باستخدام MSE للنموذج بعيدا عن حقيقته فيضلل الباحث لانه يؤدي به الى حسابات خاطئة لقيمة احصاءه Mallows (Cp) ومقدار تحيز النموذج.



خطوات استخدام RUF للحصول على أفضل نموذج (جزئي)

للاحدار الخطي

ب- التوصيات suggestions

- يوصي الباحث باستخدام RUF للحصول على نموذج الانحدار الخطي الأفضل في حالة وجود علاقات خطية قوية ما بين المتغيرات (التفسيرية) في نموذج الانحدار الخطي المتعدد، وذلك وفق الخطوات الآتية:
1. اكتشاف واستبعاد النقطة (النقاط) المتطرفة والمؤثرة في النموذج (ان وجدت) باستخدام Box-plot للبواقي المعيارية، وتحديد مستوى المعنوية لاختبار النماذج وتقديرات معاملاتها، وحساب قيم RUF لجميع تقديرات معاملات الانحدار باستخدام العلاقة $RUF = \frac{B_i - r_{iy}}{r_{iy}}$ وذلك من خلال إجراء الدورة الأولى لانحدار متغير الاستجابة على جميع المتغيرات (التفسيرية) التي اختارها الباحث، وتعاد هذه الدورة في حالة وجود نقطة بيانات أو أكثر يتم استبعادها.
 2. يحذف المتغير (التفسيري) من نموذج الدورة اللاحقة للانحدار كما يلي:
 - إذا كانت قيمة RUF لواحد أو أكثر من معاملات الانحدار كلها، اصغر أو يساوي سالب واحد $RUF \leq -1$ ، فيحذف المتغير الذي يقابل اصغر RUF من بينها (الأكبر سالبية).
 - إذا كانت قيمة RUF لواحد أو أكثر من معاملات الانحدار (غير المعنوية فقط) تقع في المدى $[-1 < RUF \leq -5]$ فيحذف المتغير (غير المعنوي) المقابل لاصغر RUF من بينها (الأكبر سالبية).
 3. نستمر في حساب قيم RUF واستبعاد متغير واحد في كل دورة انحدار وفقاً لما تقدم حتى نحصل على نموذج بمعاملات انحدار (معدنية) و(مستقرة) تبعد فيها قيم RUF عن سالب واحد باتجاه (الصفر) حيث ان معاملات الانحدار تكون أكثر استقراراً كلما اقترب RUF لكل منها من الصفر.



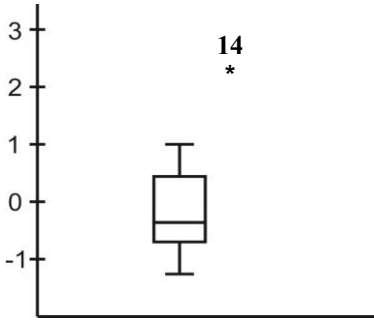
خطوات استخدام RUF للحصول على أفضل نموذج (جزئي)
للانحدار الخطي

المصادر Reference

1. د. عبد الحسين، صباح فرج، "طريقة ومعيان جديان للكشف عن مشكلة التعدد الخطي ومعالجتها، واختيار متغيرات الانحدار" مجلة الادارة والاقتصاد - تصدر عن كلية الادارة والاقتصاد/ الجامعة المستنصرية - العدد 86، بغداد، 2011.
2. Efroymson, M.A., "Multiple Regression Analysis in Mathematical Methods for Digital computers", Eds. By A. Ralsyon and H.S. wilf, wiley, New York, 1962.
3. Hocking, R.R. and Leslie, R.N., "Selection of the best subset in regression analysis", Technometrics, 9. 1967, pp. 531-540.
4. Allen, D.M., "The prediction sum of squares as a criterion for selecting predictor variables", Technical Report No.23, Department of statistics, University of Kentucky, 1971.
5. Hocking, R.R., "Criteria for selection of a subset regression: which one should be used?", Technometrics, 14, 1972, pp. 967-970.
6. Mckay, R.J. "The adequacy of variables subsets in multivariate regression", Technometrics, 21, 1979, pp. 475-480.
7. Recher, A. C. and F. C. Pun., "Inflation of R^2 in best subset Regression", Technometrics, 22, 1980, pp. 49-53.
8. Harold Hotelling, "Analysis of a complex of statistical variables in to Principall components", Journal of Eolucational psychology, 24, 1934, PP. 417-441 & 489-520.
9. Hoerl, A.E. & Rennard, R.W., "Ridge regression: biased estimation for non orthogonal problems", Technometrics, 12, 1970, PP. 55-67.
10. Bowerman, Bruce L. and O'connel, Richard T., "Applied statistics: Improving Business processes". Richard D. Irwin, a times mirror higher education group, Inc, company, 1997.
11. Draper, N.R. and Smith, H., "Applied Regression Analysis", 3rd Edition, John Wiley & sons, 1999.
12. Haan, C.T., "Statistical methods in Hydrology", 2nd Edition, Iowa state University press, Ames, Iowa, 2002.
13. Ostrom, C.W., Jr., "Time series Analysis, Regression techniques: Quantitative Applications in the social sciences" , 2nd Edition, V.07-009, Newbury park, Sage publications 1990.
14. دومينيك سالفاتور، "نظريات ومسائل في الاحصاء والاقتصاد القياسية" ترجمة د. سعيد حافظ منتصر، دار ماكجر وهيل للنشر، 1990.
15. ساميريت جاترجي وبيترام برايس، "تحليل الانحدار بالامتثلة"، ترجمة محمد مناجد الدليمي، مطابع التعليم العالي في الموصل، 1990.
16. Ronald D. Snee, "Validation of Regression Models: Methods and Examples", Technometrics, Vol, 19, No. 4, November, 1977.
17. Mallows, C.L., "Some comments on Cp", Technometrics, 15 1973, PP. 661-675.
18. Hocking, R.R., " The analysis and selection of variables in linear regression", Biometrics 32 (1), 1976, pp. 1-50.
19. Bowerman, Bruce L. and O'connel, Richard T., "Ibid" pp. 856-860.
20. سابريت جترجي وبيترام برايس، مصدر سابق، ص191.

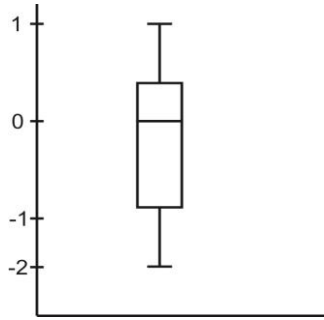


خطوات استخدام RUF للحصول على أفضل نموذج (جزئي)

الانحدار الخطي
الملاحق

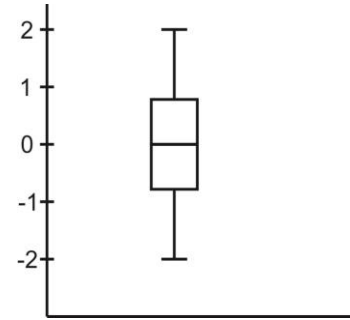
(1)

Box – plot للبواقي المعيارية للمثال الثالث والانحدار على كل نقاط البيانات وكل المتغيرات



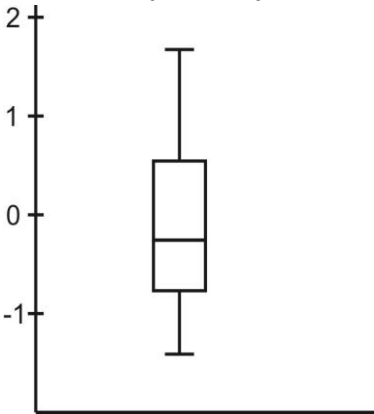
(2)

Box – plot للبواقي المعيارية للمثال الثالث بطريقة step - wise مع حذف النقطة (14)

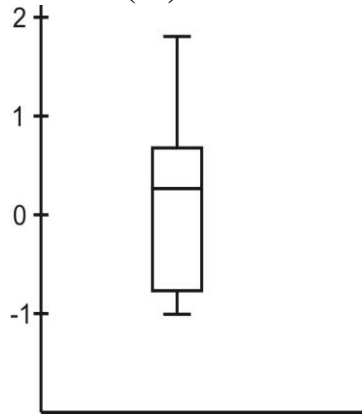


(3)

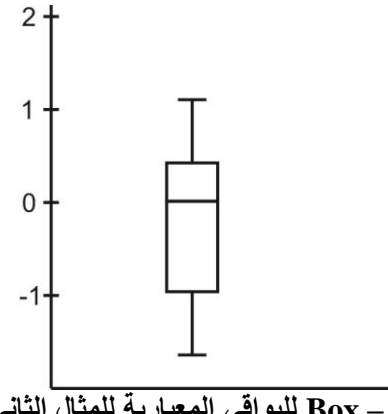
Box – plot للبواقي المعيارية للمثال الثالث بطريقة RUF مع حذف النقطة (14)



Box – plot للبواقي المعيارية للمثال الاول (نموذج الانحدار بطريقة RUF)



Box – plot للبواقي المعيارية للمثال الثاني (نموذج الانحدار على جميع المتغيرات المستقلة)



Box – plot للبواقي المعيارية للمثال الثاني - نموذج الانحدار بطريقة -RUF

قيم البواقي التنبؤية لنموذج RUF للمثال التطبيقي الاول		قيم البواقي التنبؤية للنموذج الوافي للمثال التطبيقي الاول		قيم البواقي التنبؤية لنموذج step wise للمثال التطبيقي الاول	
296.75158	-420.8568	555.56353	545.24524	618.83923	187.0886
-492.30493	-696.66910	-410.92517	-562.21263	-368.17068	-866.75301
-620.33653	+525.01465	-564.37559	-84.00250	-695.02137	-546.39706
-336.85903	-509.36133	-364.59231	-474.90424	-376.57986	21.02814
+850.22617	-43.53056	802.04207	407.36047	+722.95310	-721.68120
+182.66095	-552.98485	66.21523		55.20450	-79.68502
-217.36129	+368.54179	-495.13523		-12.66021	-409.71636
+359.72840		613.88731		618.09436	219.33287
+721.55403		606.98991		562.28606	
+779.62903		908.6751		1011.54488	
-39.08465		-117.15161		-82.52741	
+814.42272		812.75273		-1.65486	
-90.46964		-10.48103		+286.32409	
-2130984		93.44517		+496.16042	
-313.68674		-274.42757		-428.84827	
-332.78730		-170.48284		94.23432	
-95.88835		-272.42757		-240.01811	
-122.89965		-33.18151			
		-628.24032			
		-631.06644			
	PRESS= 5422547.291 RPRESS= 2328.64		PRESS= 6046691.528 RPRESS= 2459.002		PRESS= 5804449.646 RPRESS= 2409.240



خطوات استخدام RUF للحصول على افضل نموذج (جزئي)

الانحدار الخطي

قيم البواقي التنبؤية لنموذج RUF (المثال التطبيقي الثاني)	قيم البواقي التنبؤية للنموذج $f(X_1, X_2, X_3)$ (المثال التطبيقي الثاني)	قيم البواقي التنبؤية للنموذج $f(X_1, X_2)$ (المثال التطبيقي الثاني)
.29657 , -.91638 , .87564 , .06156 , -.21972 , -.46632 , .23017 , .28049 , .58671 , -.88065 , .09352	.36395 , -.93657 , .94489 , .03168 , -.60967 , -.55055 , .13174 , .10208 , .53195 , -.82849 , 1.10223	.04681 , -1.49807 , .69394 , .12423 , .39627 , -.27889 , .54763 , .86244 , .81972 , -1.29709 , -.99232
PRESS = 3.224	PRESS = 4.790	PRESS = 7.361

معايير تقييم نماذج الانحدار لبيانات المثال التطبيقي الثالث				قيم البواقي التنبؤية لنموذج step wise , RUF لبيانات المثال الثالث	
النموذج (n=16)	R_a^2	RMSE	Cp ^(*)	بواقي نموذج RUF	بواقي نموذج Step wise
f(X ₃)	.976	856.707	52.313	112.55769	-142.87735
f(X ₁)	.976	867.672	53.970	-513.86497	185.22905
f(X ₄)	.888	1857.938	290.480	381.64579	69.57763
f(X ₂)	.880	1924.143	312.421	880.46227	509.73985
f(X ₅)	.265	4769.166	1981	72.98571	178.36199
f(X ₃ , X ₅)	.992	489.126	4.467	-244.54028	-332.30332
f(X ₁ , X ₅)	.992	509.821	11.149	160.69014	280.43102
f(X ₂ , X ₃)	.989	564.300	17.779	523.03373	447.16886
f(X ₁ , X ₂)	.988	600.882	19.378	-621.85289	-434.11504
f(X ₃ , X ₄)	.980	794.085	41.308	-905377063	-546.78424
f(X ₁ , X ₄)	.978	821.206	44.872	351.19383	564.76499
f(X ₁ , X ₃)	.974	885.318	53.775	-188.77699	-823.88426
f(X ₂ , X ₄)	.923	1546.207	184.529	179.27792	-280.78615
f(X ₂ , X ₅)	.907	1693.459	223.345	1567.57285	-103.93230
f(X ₄ , X ₅)	.905	1710.298	228.009	2209.63481	1026.92551
f(X ₂ , X ₃ , X ₅)	.995	387.160	3.258	995.89573	919.78921
f(X ₁ , X ₂ , X ₅)	.994	415.472	4.965	PRESS=	PRESS=
f(X ₃ , X ₄ , X ₅)	.992	504.334	11.105	10523181.41	5215645.846
f(X ₁ , X ₃ , X ₅)	.992	504.441	11.110	RPRESS=	RPRESS=
f(X ₁ , X ₄ , X ₅)	.991	514.552	11.885	3243.945	2053.204
f(X ₁ , X ₂ , X ₃)	.989	594.082	18.508		
f(X ₂ , X ₃ , X ₄)	.988	596.701	18.742		
f(X ₁ , X ₂ , X ₄)	.988	619.237	20.801		
f(X ₁ , X ₃ , X ₄)	.982	746.845	33.894		
f(X ₂ , X ₄ , X ₅)	.942	1343.261	127.522		
f(X ₂ , X ₃ , X ₄ , X ₅)	.995	381.555	4.023		
f(X ₁ , X ₂ , X ₃ , X ₅)	.995	390.876	4.519		
f(X ₁ , X ₂ , X ₄ , X ₅)	.995	391.236	4.538		
f(X ₁ , X ₃ , X ₄ , X ₅)	.991	525.841	13.037		
f(X ₁ , X ₂ , X ₃ , X ₄)	.989	573.366	16.634		
f(X ₁ , X ₂ , X ₃ , X ₄ , X ₅)	.995	399.712	6.000		

(*) قيم Cp محسوبة على اساس ان النموذج غير المتحيز هو الذي يضم جميع المتغيرات (المستقلة) بغض النظر عن تحقيقه لفرضيات حد الخطأ وتمييزه بان له اصغر MSE من بين النماذج التي تحقق هذه الفرضيات، وهي طريقة (غير دقيقة) كما يرى الباحث.



خطوات استخدام RUF للحصول على أفضل نموذج (جزئي)
نتائج استخدام طريقة step – wise لاختبار نموذج الإنحدار الخطي الأفضل وتحليل بيانات المثال التطبيقي
الإنحدار الخطي
الثاني باستخدام SPSS

1-Variables Entered / Removed

Model	Variables entered	Variables removed	D.W-Statistic
1	X ₃		2.818
2	X ₂		

2- A NOVA – TABLE

Model	S.O.V.	S.S.	d.f.	M.S.	R _a ²	F	Sig.
1	Regression	194.997	1	194.997	.938	153.246	.000
	Residuals	11.452	9	1.272			
	Total	206.449	10				
2	Regression	204.648	2	102.324	.989	454.581	.000
	Residuals	1.801	8	.225			
	Total	206.449	10				