اختيار انموذج حصين في الانحدار الخطي

صباح حسيب حسن
كلية العلوم–جامعة كركوك

الخلاصة

تناول البحث اقتراح صيغة حصينة لمعيار تجميع خطأ التنبوء(APE) المستخدم في اختيار انموذج الانحدار الخطي . وتم تقييم اداء المعيار من خلال تجربة محاكاة حيث أظهر الصيغة المقترحة اداءاً جيداً مقارنة مع بعض المعايير الحصينة الاخرى، اذ كانت اكثر مقاومة للقيم الشاذة وخاصة عند الحجوم الصغيرة للعينة.

المقدمة

من المسائل المهمة المتعلقة بالنماذج الخطية، اختيار الانموذج المناسب الذي يوفق المعلومات المتضمنة في متغير الاستجابة بشكل كفء ومختصر. وتعد اختيار افضل مجموعة جزئية من مجموعة المتغيرات التوضيحية (Best Subset Selection) من احدى الاساليب الاحصائية المستخدمة في مسائل اختيار الانموذج المناسب، إذ يتم ذلك من خلال توفيق كل النماذج الممكنة باستخدام كل المجاميع الجزئية من مجموعة المتغيرات التوضيحية المتوفرة واختيار الانموذج الذي يحوي تلك المجموعة الجزئية من المتغيرات التوضيحية الاساسية في العلاقة للحصول على المعلومات نفسها التي نحصل عليها لا سيما لو شمل الانموذج جميع المتغيرات.تناول البحث اختيار الانموذج في النماذج الخطية في حالة عدم تحقق بعض فرضيات الانموذج الخطي مثل فرضية التوزيع الطبيعي للاخطاء العشوائية وحالة وجود القيم الشاذة في البيانات،اذ ان مثل هذه الحالات لها أثر سلبي على أساليب اختيار النماذج مما تؤدي الى أن تكون النتائج غير دقيقة ومضللة. ولذلك جاءت الاحصاءات الحصينة كأسلوب مقاوم للقيم الشاذة والخروقات الاخرى لفرضيات الانموذج الخطي لاسيما المعايير الحصينة المستخدمة في اختيار الانموذج (حسن، ٢٠٠٢).يهدف البحث الى ايجاد صيغة حصينة لاحد المعايير المستخدمة في اختيار الانموذج الخطي وهو معيار تجميع خطأ التنبؤ Accumulate Prediction Error (APE)

واستخدامه الى جانب معايير حصينة أخرى في اختيار انموذج الانحدار الخطي في حالة وجود القيم الشاذة في البيانات وذلك من خلال تجربة محاكاة بالاستناد الى برنامج حاسبة يقوم بتكوين انموذج تام واشتقاق كل النماذج الجزئية الممكنة منه وتقدير معلمات هذه النماذج بطريقة M الحصينة وحساب معايير اختيار الانموذج في حالتين وهي الحالة الطبيعية وحالة وجود قيم شاذة في البيانات . لغرض تحقيق اهداف البحث، فقد تم تقسيمه الى مبحثين ، المبحث الاول يمثل الجانب النظري وقد خصص للنماذج الخطية وتجزئته الى نماذج جزئية واستعراض لطريقة التقدير M الحصينة في تقدير معلمات الانموذج الخطي وبعد ذلك التعرف على فكرة تحوير صيغة معيار APE الى صيغة حصينة ، المبحث الثاني يمثل الجانب التجريبي للبحث ، اذ يتم فيه التطرق الى نماذج الانحدار المستخدمة في تجارب المحاكاة والحالات المراد دراستها واسلوب تنفيذ التجارب قيد البحث ، واخيراً تحليل نتائج المحاكاة والخروج ببعض الاستنتاجات حول اداء المعايير.

الجانب النظري

١- الانموذج الخطي

يعرف الانموذج الخطي بأنه علاقة رياضية تربط بين المتغيرات، ويتم التعبير عن هذه العلاقة بصيغة معادلة خطية تحتوي على متغير الاستجابة مع واحد او اكثر من المتغيرات التوضيحية.وبصيغة المصفوفات وبوجود (n) من المشاهدات و (k-1) من المتغيرات التوضيحية، يتم التعبير عن الانموذج الخطي كالاتي (1976,Hocking):

$$Y = XB + U \qquad ...(٢-١)$$

اذ ان Y متجه (١ ×n) لمشاهدات متغير الاستجابة ، X مصفوفة (n × k) لمشاهدات المتغيرات التوضيحية التي يتضمنها الانموذج ومن ضمنها (X_0) المرافق للحد الثابت ، B متجه(k × 1) للمعلمات المجهولة في الانموذج ومن ضمنها الحد الثابت ، U متجه (١ ×n) للاخطاء العشوائية بمتوسط صفر ومصفوفة تباين وتباين مشترك $\sigma^2 I$. يعرف الانموذج (٢-١) بالانموذج الخطي العام والذي قد يظهر بعدة حالات معتمدا على رتبة المصفوفة X ، الاِ ان هذه الدراسة ستركز على انموذج الانحدار الخطي كحالة خاصة من النماذج الخطية .

ولاجل تقدير معلمات الانموذج (٢-١) فانه بموجب طريقة M الحصينة يتم تحوير الاسلوب

لمتبع في طريقة المربعات الصغرى، والذي يتطلب تصغير مجموع مربعات الخطأ وبعبارة اخرى:

$$\min \sum_{i=1}^{n} (y_i - x_i' B)^2$$

الى الصيغة الاتية (1979,Launer & Wilkingson) :

$$\min \sum_{i=1}^{n} \rho\left(\frac{Y_i - x_i' B}{\sigma}\right) \qquad \ldots(٢-٢)$$

حيث ان x_i' يمثل صف (i) من المصفوفة X.

اذ تم استبدال مربعات القيم بدالة عمومية ρ . وباتباع الاسلوب المستخدم في طريقة المربعات الصغرى نفسه يتم اشتقاق دالة ρ بالنسبة الى المتجه B ومساواتها الى الصفر

$$\sum_{i=1}^{n} \psi\left(\frac{y_i - x_i' B}{\sigma}\right) x_{ij} = 0 \qquad j = 1,2,\ldots,k$$

لنحصل على المعادلة الاتية:

حيث ان $\psi = \rho'$

ويمكن اعادة كتابة المعادلة الاخيرة بالصيغة الاتية:

$$\sum_{i=1}^{n} w_i x_{ij} (y_i - x_i' B) / S = 0 \qquad j = 1,2,\ldots,k \qquad \ldots(٣-٢)$$

حيث ان الاوزان w_i تعطى بالصيغة الاتية:

$$w_i = \frac{\psi[(y_i - x_i' B_0)/S]}{(y_i - x_i' B_0)/S} \qquad i = 1,2,\ldots,n$$

وبالتالي يتم حل المعادلة (٣-٢) لايجاد قيمة الموجه b_M وسيكون وفق طريقة المربعات الصغرى الموزونة الاتية (1977,Holland &Welsch):

$$b_i = (X'WX)^{-1} X'WY \qquad i = 1,2,\ldots,M$$

حيث ان W مصفوفة قطرية عناصر القطر فيها قيم الاوزان w_i ، B_0 تمثل القيمة الاولية لموجه المعلمات، ويمكن ان تقدر بطريقة المربعات الصغرى الاعتيادية، وبالتعويض في صيغة w_i لايجاد الاوزان نحصل على المتجه b_1 في التكرار الاول ونحصل على المتجه b_2 في التكرار الثاني وهكذا تستمر التكرارات الى ان تتقارب القيم في التكرارات المتتابعة فيكون b_M هو المقدر النهائي. حيث ان معلمة القياس (σ) تقدر مرة واحدة باستخدام القيم الاولية للمعالم في الانموذج التام بموجب الصيغة الاتية:

$$S = 1.483 \; median \left| r_i - median(r_i) \right|$$

اذ ان $r_i = y_i - x_i' B_0$

ونظرا لكون حصانة المقدر الناتج عن هذه الطريقة تعتمد على دالة ρ ، فأنه تم اقتراح صيغ عديدة لهذه الدالة من قبل الباحثين في مجال الاحصاءات الحصينة هدفها الحصول على مقدر حصين ضد القيم الشاذة في البيانات، وأهم هذه الصيغ هي دالةHuber الاتية(Huber 1981):

$$\rho(r) = \begin{bmatrix} r^2/2 \\ C|r| - C^2/2 \end{bmatrix} \qquad \psi(r) = \begin{bmatrix} r & |r| \le C \\ C\,sign(r) & |r| > C \end{bmatrix}$$

حيث ان C=1.345

ان مجموعة المتغيرات التوضيحية المتوفرة التي تحتويها المصفوفة X قد يكون بعض منها غير اساسي في الانموذج ، وسنرمز الى عدد مثل هذه المتغيرات بالرمز (q). في حين هناك متغيرات توضيحية تؤدي دورا اساسيا في الانموذج وليكن عددها (p-1)، ويمكن عزل المتغيرات (q) عن المتغيرات (p-1) من خلال اعادة كتابة الانموذج (١–٢) بالصيغة الاتية (1976,Hocking):

$$Y = X_p B_p + X_q B_q + U \qquad \ldots(٢–٤)$$

اذ ان المصفوفة X تم تجزئتها الى المصفوفتين المجزئتين X_p و X_q ذات المرتبة $(n \times p)$ و $(n \times q)$ على التوالي. فيما يخص المصفوفة الجزئية X_p فأنها تضم المتغيرات التوضيحية الاساسية في الانموذج بضمنها المتغير X_0 المرافق للحد الثابت، وهذا يعني ان الحد الثابت موجود في النماذج الجزئية، بينما المصفوفة X_q تضم المتغيرات التوضيحية غير الضرورية المرشحة للاستبعاد من الانموذج. ونتيجة لتجزئة المصفوفة X كما في اعلاه فأن المتجه B الخاص بالمعلمات سوف يجزأ بدوره ايضا الى المتجهات الجزئية B_p و B_q ذات المرتبة $(p \times ١)$ و $(q \times ١)$ على التوالي ولذلك فأن الرمز p سيشير الى عدد المعالم بضمنها الحد الثابت في الانموذج الجزئي في حين الرمز q يشير الى عدد المتغيرات المستبعدة من الانموذج، حيث ان $k = p + q$. وبتغير اعمدة المصفوفتين X_p و X_q يمكن الحصول على جميع التراكيب الممكنة للمصفوفتين المذكورتين، اذ سيكون هناك 2^{k-1} من المجاميع الجزئية الممكنة للمتغيرات التوضيحية بضمنها المجموعة التامة. ومن هذه المجاميع

يمكن الحصول على 2^{k-1} من النماذج الجزئية بضمنها الانموذج التام الـذي يحـوي جميـع المتغيرات التوضيحية .

٢–١ معايير APE الحصين

ان الهدف من الاحصاءات الحصينة هو توفير اساليب احصائية جديدة قليلة الحساسية والتأثر تجاه القيم الشاذة وتعمل بشكل جيد تحت توزيعات مختلفة، وبالنظر لتطور طرائق تقدير معلمات النماذج الخطية وايجاد طرائق حصينة مختلفة، فقد دعت الحاجة الى تطوير المعايير التقليدية المستخدمة في اختيار النماذج وايجاد معايير جديدة تسمح اولا باختيار الانموذج المناسب الذي يمثل البيانات مع الاخذ بنظر الاعتبار وجود قيم شاذة وحالة توزيع الاخطاء العشوائية قد لا يكون توزيعا طبيعيا. وثانيا استخدام متوافق لهذه المعايير مع النماذج المقدرة وفق الطرائق الحصينة.وعليه فقد حظيت مسالة الاختيار الحصين للنماذج الخطية بالاهتمام المتزايد من قبل الباحثين في الاونة الاخيرة،وظهرت محاولات عديدة تهدف الى ايجاد صيغ جديدة حصينة لبعض المعايير المستخدمة في اختيـار النماذج الخطية نذكر منها معيـار RAIC المقترح من قبل(Ronchetti Ronchetti,1985)،معيارRSIC المقترح من قبل Machado(Machado 1993)، معيار RCV المقترح من قبل Ronchetti وBlanchard (Ronchetti and Blanchard 1997) ، معيار RF المقترح من قبل Huber (Huber 1981, Schrader and Mckean 1977) ، معيار RC_p المقترح من قبل Ronchetti و Staudt (Ronchetti and Robert 1994) ومعيار RSC المقترح من قبل Qian وKunsch (Qian and Kunsch:1998) .

وفي ادناه نستعرض معيار APE والصيغة الحصينة المقترحة لهذا المعيار (حسن، ٢٠٠٢) :

انطلاقا من مبدأ تجميع خطأ التنبؤ،فقد اقترح Rissanen (Speed &Yu 1993;Wei,1992;Lee&Lai,1997) معيارا يقوم على اساس تجميع مربع خطأ التنبؤ في مرحلة معينة لتكن (i) من المشاهدات المتوفرة. وقد جاء تسمية هذا المعيار (APE) نسبة الى تجميع مربعات خطأ التنبؤ(Accumulate Prediction Error) وكذلك يطلق على هذا المعيار تسمية PLS (Predictive Least Squares)، وعلى اية حال فأن هذا المعيار يعرف وفق الصيغة الاتية:

$$APE = \sum_{i=m}^{n} (y_i - b'_{i-1} x_i)^2$$

اذ ان: (y_i) : تمثل المشاهدة (i) في متغير الاستجابة.

(b_j) : تمثل تقدير المربعات الصغرى باستخدام المشاهدات $(x_i\,, y_i\ :\ i \le j)$.

(m) : يمثل اول قيمة من قيم j بحيث ان b_j يكون تقديرا وحيدا (Uniquely Defined).

وبموجب هذا المعيار يتم اختيار الانموذج الجزئي الذي ينتج عنـــه اقــل قيمـــة للمعيـــار الاتـــــي:

$$APE = \sum_{i=m}^{n} (y_i - b'_{p(i-1)} x_{p(i)})^2$$

ويلاحظ على هذا المعيار انه يقوم على اساس تصغير مجموع مربعات خطأ التنبؤ، ولاجل تطبيق اجراءات الحصانة على هذا المعيار فأنه بالامكان اعتماد المبدأ الذي تقوم عليه طريقـــة M الحصينة في حالة النماذج الخطية بعد تحوير مبدأ طريقة المربعات الصغرى، اذ بموجـــب المعادلة (٢-٢) يتم تغيير مجموع مربعات الخطأ الى صيغة عمومية (ρ) من الخطأ، بعبارة اخرى يتم تغيير الصيغة الاتية:

$$\min \sum_{i=1}^{n} (y_i - b'_i x_i)^2$$

الى الصيغة:

$$\min \sum_{i=1}^{n} \rho(y_i - b'_i x_i)$$

وعليه وكخطوة اولى يتم تقدير معالم الانموذج بطريقة M الحصينة من البيانات المتوفرة واستخدام المقدر الحصين الناتج (b_M) في حساب قيمة المعيار من الصيغة الحصينة المقترحة، اذ بموجب التحوير المشار اليه في اعلاه يقترح البحث تعويض مربعات خطأ التنبـــؤ المتجمـــع بالدالة العمومية ρ ، بعبارة اخرى يقترح البحث الصيغة الحصينة الاتية لهذا المعيار:

$$RAPE = \sum_{i=m}^{n} \rho(y_i - b'_{M(i-1)} x_i)$$

ولاجل جعل هذا المعيار يملك خاصية Scale Invariant وفق المبدأ المتبع فـــي طريقـــة التقدير M بموجب المعادلة (٢-٢)، يتم تقسيم خطأ التنبؤ على معلمة القياس S . وعليه بموجب المعطيات التي تم ذكرها في اعلاه فأن الصيغة الحصينة المتقرحة لمعيار APE تكون كالاتي:

$$RAPE = \sum_{i=m}^{n} \rho\left(\frac{y_i - b'_{M(i-1)} x_i}{S}\right)$$

حيث ان m يشير الى اول مرحلة من المشاهدات التي تجعل مقــدر M وحيــدا، و b_M مقدر M حصين لمعالم الانموذج. وفيما يخص معلمة القياس S فأنها تقدر من الانموذج التام

بموجب الصيغة المشار اليها في المعادلة (٤-٣). وفيما يخص دالة ρ فأن هنــاك صــيغا عديدة لها وقد تم استخدام الصيغة المشار اليها في هذا البحث.

٢: الجانب التجريبي

تناول هذا الجانب اهم الجوانب في البحث ممثلة بتجربة المعايير التي تناوله البحث ، اذ تــم استخدام المعايير في اختيار النماذج من خلال تجربة محاكاة في حالتين وهي الحالــة الطبيعيــة وحالة وجود قيم شاذة في البيانات وبأستخدام أنموجين مختلفين في عدد المتغيرات التوضــيحية وبحجوم مختلفة للعينة.

١-٢:نماذج المحاكاة

تم استخدام انموذجين انحدار خطي في البحث وايجاد جميع النماذج الجزئية الممكنة منــه (النماذج المنتخبة)،وقد تم ترميز كل انموذج برمز يشير الى المتغيرات التوضيحية المتضمنة فيه كما يأتي:

١. **الانموذج الاول:** يتضمن متغيرين X_1 و X_2 ، وبذلك فأن عدد النماذج المنتخبة سيكون $2^2 = 4$ وهي كالأتي:

النماذج المنتخبــــة	الرمــز
$Y = B_0 + B_1X_1 + B_2X_2$	(٠١٢)
$Y = B_0 + B_1X_1$	(٠١)
$Y = B_0 + B_2X_2$	(٠٢)
$Y = B_0$	(٠)

١. **الانموذج الثاني:** يتضمن ثلاثة متغيرات توضيحية X_1 و X_2 و X_3، وبذلك فــأن عــدد النماذج المنتخبة سيكون $2^3 = 8$ وهي كالاتي:

النماذج المنتخبــــة	الرمــز

(٠١٢٣) $Y = B_0 + B_1X_1 + B_2X_2 + B_3X_3$

(٠١٢) $Y = B_0 + B_1X_1 + B_2X_2$

(٠١٣) $Y = B_0 + B_1X_1 + B_3X_3$

(٠٢٣) $Y = B_0 + B_2X_2 + B_3X_3$

(٠١) $Y = B_0 + B_1X_1$

(٠٢) $Y = B_0 + B_2X_2$

(٠٣) $Y = B_0 + B_3X_3$

(٠) $Y = B_0$

بعد الانتهاء من استعراض نماذج الانحدار المستخدمة في تجربة المحاكاة لابد من ذكر اسلوب تنفيذ تجارب المحاكاة.بالنسبة للانموذج الاول، فقد تم ايجاد مشاهدات متغير الاستجابة (y) من خلال الانموذج المفترض الاتي:

$$y_i = 2 + 7x_{i2} + u_i \qquad i = 1,2,...,n$$

وعليه فأن الانموذج بالرمز (٠٢) يكون هو الانموذج المطلوب اختياره من بين النماذج الاربعة المنتخبة.وفيما يخص الانموذج الثاني المتضمن ثلاثة متغيرات، فقد تم ايجاد قيم المشاهدات (y) باستخدام الانموذج المفترض اعلاه نفسه، وعلى هذا الاساس فأن الانموذج بالرمز (٠٢) سيكون هو المطلوب اختياره من بين ثمانية نماذج منتخبة.

٢–٢: الحالات المدروسة

وتمت دراسة أداء المعايير في حالتين كالاتي:

١. **الحالة الطبيعية:** تشير الى عدم وجود قيم شاذة في البيانات والى توفر شرط التوزيع الطبيعي للاخطاء العشوائية.

٢. **حالة القيم الشاذة:** بغية توضيح التغير في اداء معايير اختيار الانموذج نتيجة وجود قيم شاذة في البيانات فأنه تم استخدام التوزيع الملوث من جانب واحد للأخطاء العشوائية وبنسبة تلوث(١٠%) وفق الصيغة الاتية: $0.90\ N(0,1) + 0.10\ N(30,1)$ واعادة حساب المعايير في هذه الحالة وملاحظة التغيرات التي يمكن ان تطرأ على اختيار المعايير للنماذج مقارنة بالحالة الطبيعية. ومن الجدير بالذكر هنا انه تم توليد مشاهدات المتغيرات التوضيحية بحيث تتبع التوزيع المنتظم U(0,1) في حين تم توليد مشاهدات الاخطاء العشوائية بحيث تتبع التوزيع الطبيعي في الحالة الطبيعية والتوزيع الملوث في حالة وجود قيم شاذة في البيانات وتجربة الحالتين المذكورتين في كل انموذج من النماذج باستخدام حجوم العينة (n) ١٠

و ٢٠ مشاهدة واستخدام طريقة M الحصينة لتقدير معلمات النماذج وحساب المعايير في كل تكرار، اذ تم تكرار كل تجربة ٢٠٠ مرة وحساب معدل قيم المعايير في كل التكرارات.

٣-٢ : تحليل نتائج المحاكاة

بعد تطبيق الافكار الواردة في المباحث السابقة ، فقد تم الحصول على النتائج المبنية في الجداول الواردة في الصفحات القادمة ولكل حالة من الحالتين المدروسة، وفي ضوء ذلك ولغرض اعطاء صورة واضحة عن هذه النتائج، فأنه سيتم تحليل ومناقشة هذه النتائج لكل حالة على انفراد وكما يأتي:

١. الحالة الطبيعية

تشير الجداول (١) و (٢) الى معدل القيم المحسوبة للمعايير لاختيار الانموذج المفترض (٠٢) من بين النماذج المنتخبة بالنسبة الى كل انموذج من النماذج التي تم استخدامها في تجارب المحاكاة في الحالة الطبيعية وبحجوم عينة ١٠ و٢٠.

تشير قيم المعايير المبينة في الجدولين الى اختيار الانموذج الصحيح (٠٢) من بين النماذج الجزئية في الانموذجين عند حجمي العينة ١٠و٢٠ وهذه النتيجة متوقعة لكون جميع الظروف متوفرة في هذه الحالة لتطبيق المعايير المدروسة والبيانات خالية من القيم الشاذة.

٢. حالة القيم الشاذة

تشير الجداول (٣) و (٤) الى معدل القيم المحسوبة للمعايير لاختيار الانموذج المفترض (٠٢) من بين النماذج المنتخبة وذلك بعد تلوث البيانات بقيم شاذة وبحجوم عينة ١٠ و ٢٠.يلاحظ من الجدول رقم (٣) انه في حالة حجم العينة ١٠ فان بعض المعايير كانت متأثرا" بالقيم الشاذة وهذه المعايير هي RSIC,RSC,RC ,RF أما المعايير RAIC ,RCV,RAPE فانها اختارت الانموذج المفترض (٠٢) بشكل صحيح. بينما في حالة حجم العينة ٢٠ فان جميع المعايير اختارت الانموذج (٠٢) بشكل صحيح.فيما يخص الجدول (٤) الخاص بالانموذج الثاني فانه في حالة حجم العينة ١٠ كانت بعض المعايير ايضا متأثرا بالقيم الشاذة بحيث لم تختار الانموذج المفترض عدا معياري RCV, RAPE فانها اختارت الانموذج المفترض. أما في حالة حجم العينة ٢٠ فان جميع المعايير ترفض الانموذج الذي يحتوي على المتغير X_3 والانموذج الثابت وتميل الى اختيار النماذج الحاوية على المتغير X_2 وان افضل المعايير كانت معياري RCV ,RAPE ايضا.وعليه فانه من خلال تحليل النتائج اعلاه نستنتج ان الصيغة الحصينة المقترحة في البحث كان اداءه جيدا مقارنة مع بعض المعايير الحصينة الاخرى ولذلك

فانه لاختيار انموذج انحدار خطي في حالة وجود وعدم وجود القيم الشاذة يفضل استخدام المعايير الحصينة ولاسيما معياري RAPE , RCV ،اذ اثبتت هذه المعايير مقاومة جيدة للقيم الشاذة في البيانات.

جدول رقم (١) : قيم المعايير الحصينة لاختيار الانموذج (٠٢) في الانموذج الاول

(الحالة الطبيعية)

المعايير	النمـــــاذج المنتخبـــــة			
	(٠١٢)	(٠١)	(٠٢)	(٠)
	n =10			
RAIC	٤,٣١٥	٢٠,٤٣١	٤,٥١٨	١٩,٩٧٨
RSIC	٤,٨١٠	٩,٦٥٥	٣,٩٢١	٩,٣٧١
RSC	٤,٩٥٨	٩,٠٥٢	٤,٤٣٤	١٠,٣٥٥
RC_P	–	٤,٨٣١	٠,٩٢٠	٢,٩٥١
RF	–	٣٧,٤٣٨	١,٣٥٤	٢١,٥٢٧
RCV	٠,٣٢٦	١,٠٦١	٠,٢٣٥	٠,٩٧١
RAPE	٨,٩٧٦	١٢,٤٢٠	٨,٧٦٧	١١,٨٣٠
n = 20				
RAIC	١٤،٨٠٨	٢٩,٢٩٠	١٤,٥٤٧	٢٨,٤٨١
RSIC	١٠,٦٦٢	١٥,٥٣٠	٩,٤٢٠	١٤,٥٩٦
RSC	١٠,١٦٠	١٤,٧٧٢	٩,٥١٧	١٥,٢٩٤
RC_P	–	٢,١٩٢	٠,٩٢١	١,٤١٧١
RF	–	٢٩,٤٦٨	١,١٥١	١٥,٩١٥
RCV	١,١٣٠	٠,٨٤٢	٠,٤١٦	٠,٨٠٣
RAPE	١١,٣٧١	١٦,٧٢١	١١,١٥٠	١٦,٠٢٢

جدول رقم (2) : قيم المعايير الحصينة لاختيار الانموذج (٠٢) في الانموذج الثاني

(الحالة الطبيعية)

المعايير	النمـــــاذج المنتخبـــــة							
	(٠)	(٠٣)	(٠٢)	(٠١)	(٠٢٣)	(٠١٣)	(٠١٢)	(٠١٢٣)
	n =10							
RAIC	١٨,٤٤	١٨,٨٣٠	٤,١٧٥	١٨,٨٧١	٤,٠٠٨	١٨,٣٤٩	٣,٩٥٤	٣,٦٣٠
RSIC	٨,٨٣٢	٩,١٨٤	٣,٨٠٣	٩,١٨٩	٤,٧١٨	٩,٥٠٢	٤,٧٠٠	٥,٦٢٠
RSC	٩,٨١٢	٨,٦١٠	٤,٣١١	٨,٦٥٦	٤,٨٣١	٨,٧٢٨	٤,٨١٧	٦,٤٨٥
RC_P	٥,٦٨١	٦,٨٢٠	١,٦٨٨	٦,٨١٢	٠,٧٤٩	٤,١٤١	٠,٧٢١	–
RF	١٦,٣٠	٢١,٩١٧	١,٤٢٠	٢١,٧١٥	١,٤٣٢	٣٧,٩٦٩	١,٤٠٣	–
RCV	٠,٩٧٧	١,٠٩٤	٠,٢٥٩	١,٠٩٠	٠,٢٩١	١,٢٢٣	٠,٢٨٤	٠,٢٨٧
RAPE	١١,٤٣	١١,٨٠٣	٨,٥٣٩	١٢,٠١٤	٨,٦٩٢	١٢,٦٦٩	٨,٧٤١	٩,٠٤٦
n = 20								
RAIC	٢٥,٤٧	٢٦,٥٢١	١٣,٥٣٤	٢٦,٦٤٢	١٣,٧٧٨	٢٦,٩٧٦	١٣,٨١٥	١٤,٠١٦
RSIC	١٣,٤٦	١٤,٣٦٠	٨,٩٥٥	١٤,٣٩٦	١٠,٢٠٧	١٥,٢٨٢	١٠,٢٢٠	١١,٤٦٣
RSC	١٤,١٣	١٣,٦٤٩	٩,٠١٧	١٣,٦٨٠	٩,٦٦٦	١٤,٠٢٤	٩,٦٧٦	١١,٣٧٠
RC_P	٣,٠٢٤	٣,٥٢٥	١,٧٢٣	٣,٥٨١	٠,٨٣٢	١,٩٧٠	٠,٨٤٦	–
RF	٩,٦٦٤	١٣,٣٤٨	١,١٢٤	١٨,٢٧٦	١,١١٠	٢٤,٣٠٣	١,١١١	–
RCV	٠,٧٣٦	٠,٧٦٦	٠,٣٩٣	٠,٧٧٣	٠,٤١٢	٠,٨١٠	٠,٤١٣	١,١١٢
RAPE	١٤,٤٥	١٥,٠٨٥	١٠,٠٩٨	١٥,٠٥٢	١٠,٣١٠	١٥,٧٣٦	١٠,٢٩٥	١٠,٦٢٤

جدول رقم (٣) : قيم المعايير الحصينة لاختيار الانموذج (٠٢) في الانموذج الاول

(حالة القيم الشاذة)

المعايير	النمـــــاذج المنتخبـــــة			
	(٠١٢)	(٠١)	(٠٢)	(٠)
n = 10				
RAIC	٧,١١٠	٧,٢٢٩	٦,٨٣٤	٦,٩١٢
RSIC	٦,١٥٢	٥,٣٣٣	٥,١٩٤	٤,٣٣٧
RSC	٥,٣٤٠	٤,٨٢٢	٤,٧٦٥	٣,٩٤٧
RCP	–	٠,٥٦٠	٠,٥٩٨	٠,٤٦٨
RF	–	١,٨٢٢	١,٠٨٢	١,٤١٣
RCV	٢,٦٣٨	٣,٢٩٣	٢,٦٣٢	٣,١٥٢
RAPE	٣,٥٣١	٣,٥٩٢	٣,٥٢٢	٣,٥٥٤
n = 20				
RAIC	١٦,٥٥٧	٢٠,٢٦٨	١٦,٤٢٧	٢٠,٠٧٠
RSIC	١٢,٣١٥	١٢,٥٠١	١٠,٩١٤	١١,٢٠٥
RSC	١١,٤١١	١١,٨٦٠	١٠,٦٨٣	١١,٤٣٣
RCP	–	٠,١٦٧	٠,٢٧٠	٠,٣١٠
RF	–	١٨,٤٧٦	١,١٤٢	١٠,٣١٣
RCV	١,٦٥٨	٢,٢٨٦	١,٥١١	٢,٢٤١
RAPE	٩,٦٤٨	١١,٠٩٢	٩,٥٧٩	١٠,٨٧٧

جدول رقم (٤) : قيم المعايير الحصينة لاختيار الانموذج (٠٢) في الانموذج الثاني

(حالة القيم الشاذة)

المعايير	النمـــــاذج المنتخبـــــة							
	(٠١٢٣)	(٠١٢)	(٠١٣)	(٠٢٣)	(٠١)	(٠٢)	(٠٣)	(٠)
	n = 10							
RAIC	٥,٨١٥	٥,٦٩٨	٥,٩٤٥	٥,٧٣٢	٥,٧١٤	٥,٤٦١	٥,٧٤٦	5.455
RSIC	٦,٤٠٨	٥,٤٨١	٥,٥٦٥	٥,٤٩٢	٤,٦٠٤	٤,٥٢٦	٤,٦١٩	٣,٦٢٣
RSC	٦,٤١١	٤,٦٦٣	٤,٧١٤	٤,٦٧٠	٤,١٢٥	٤,٠٧٧	٤,١٢٩	٣,٠٨٩
RCP	–	٠,٨٥٧	٠,٧٧٥	٠,٨٧٩	١,٣٨٨	1.357	١,٤٠٧	١,٢١٦
RF	–	١,١٩٩	١,٦١٠	١,٣٧١	١,٤١٠	1.253	١,٤٩٠	١,٢١٦
RCV	٢,٧١٨	٢,٩٥٠	٣,٩٠٨	٢,٩٦٢	٣,٥٥٣	2.767	٣,٥٦٩	٣,٣٩٦
RAPE	٢,٧٢٣	٢,٧٠٦	٢,٧٧٠	٢,٧٠٥	٢,٧٥٠	2.699	٢,٧٥٤	٢,٧٣٩
n = 20								
RAIC	١٤,٧٤١	١٤,٥٥٩	١٧,٧٤٣	١٤,٥٨١	١٧,٥٠٩	14.415	١٧,٥١٦	١٧,٢٦٣
RSIC	١٢,٧٦٢	١١,٣٤٢	١٢,٥٥٠	١١,٣٥٢	١١,٢٠٣	9.926	١١,٢٠٦	٩,٨٥٢
RSC	١٢,١٨٥	١٠,٣٧٦	١١,٢٤٥	١٠,٣٧٦	١٠,٥٣٤	٩,٦٣٩	١٠,٥٤٠	٩,٩٦٤
RCP	–	٠,٢٧٠	٠,١٣٣	٠,٢٦٩	٠,٦٤١	0.505	٠,٦٣٩	٠,٧٤٦
RF	–	٠,٨٧٥	١٣,٤٨٤	٠,٩٩١	٧,٥٦٣	0.935	٧,٥٥٣	٥,٥٧٦
RCV	١,٦١٩	١,٤٩٦	٢,٢٨٠	١,٤٩٨	٢,٢٢٧	1.476	٢,٢٢٤	٢,١٦٦
RAPE	٨,٣٧٦	٨,٣٠٦	٩,٦٤٧	٨,٢٩٣	٩,٤٢٨	8.238	٩,٤١٦	٩,٢٦٧

المصادر

- حسن،صباح حسيب،(٢٠٠٢) : المعايير الحصينة في أختيـــار النمـــاذج الخطيـــة المناسبة،اطروحة دكتوراه في الاحصاء ،كلية الادارة والاقتصاد ، جامعة بغداد.
- Hocking,R.R.,(1976): The Analysis and Selection of Variables in Linear Regression, Biometrics,Vol. 32,pp.1- 40.
- Holland,P.W.&Welsch,R.E.,(1977): Robust Regression Using Alternatively Reweighted Least Squares, Comm. Statist.,Vol. 6, pp. 813-827.
- Huber, P. J., (1981): Robust Statistics, New York, John Wiley & Sons.
- Launer, R.L., & Wilkingson, G. N., (1979): Robustness in Statistics, Acadimic Press, Inc., New York.
- Lee, C. D., & Lai, T. L., (1997): Information and Prediction Criteria for Model Selection in Stochastic Regression and ARMA Model, Statistic Sinica,Vol.7,pp. 285 – 309.
- Machado,J. A. F., (1993): Robust Model Selection and M - Estimation, Economatric Theory,Vol. 9,pp. 478 – 493.
- Qian, G. & Kunsch, H., (1998): Some Notes on Rissanen's Stochastic Complexity, IEEE, Trans. Inform. Theory, Vol.44, pp.782–786. (By Internet).
- Qian, G. & Kunsch, H., (1998): On Model Selection in Robust Linear Regression, J. Stat. Plan. & Inform. (By Internet).
- Ronchetti, E., (1985): Robust Model Selection in Regression, Statistics of Probability Letters, Vol.3, pp. 21- 23.
- Ronchetti, E.,& Robert, S., (1994): Robust Version of Mallows C_p, J. Amer. Statist. Assoc., Vol. 89,pp. 550 – 559.

- Ronchetti, E., Field, C., & Blanchard, W.,(1997): Robust Linear Model Selction by Cross - Validation, JASA, Vol.92,pp. 1017 – 1023.
- Schrader, R. M. & Mckean, J. W., (1977): Robust Analysis of Variance, Comma Statist., A6 (9),pp. 879 – 894.
- Speed, T. P. & Yu, B.,(1993): Model Selection and Prediction: Normal Regression, Ann. Inst. Statist. Math.Vol. 45, pp. 35 – 54.
- Wei, C. Z,(1992): On Predictive Least Squares Principles, The Annals of Statistics,Vol. 20,pp. 1– 42.

Robust Model Selection in Linear regression

Dr. Sabah Haseeb Hassan

College of Science\Kirkuk University

ABSTRACT:

The research deals with the proposing of robust formula for the accumulate prediction error (APE) criterion which is used in selecting regression model. The proposed formula evaluated with a simulation study. The formula performed very well against outliers than other robust criteria used in this area especially in the case of small samples.