المجلة العراقية للعلوم الاحصائية (16) 2009
ص ص [131 – 148]

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

صباح حسيب حسن*

الملخص

تناول البحث اختيار المتغيرات المناسبة في الانحدار الخطـي بالأسـاليب الحصينة بوصفه أسلوبا مقاوما للقيم الشاذة والخروق الأخرى لفرضـيات النمـوذج الخطي.ويهدف البحث إلى استخدام المعايير الحـصينة ($RAIC$, $RSIC$, RC_p, RVC, RSC, RF, $RAPE$) في تحديد المتغيرات المـستقلة المهمـة فـي نمـوذج الانحدار الخطي لتقدير السكر في الدم بوصفه متغيرا معتمـدا ومتغيـرات أخـرى مستقلة في حالة وجود القيم الشاذة في البيانات والمقارنة بين أداء هذه المعايير . وقد اظهرت المعايير نتائج جيدة في اختيار المتغيرات المهمة في النموذج لا سيما معيار RAPE مقارنة مع المعايير الاخرى .

The Use of Robust Criteria in Selecting Effective Variables in Linear Regression Model for Blood Sugar Analogy

Abstract:

The research deals with the topic of variable selection in linear regression using robust procedures as resistant to outliers and other failures of assumptions . The objective of the research is using robust criteria (RAIC , RSIC, RC_p, RVC, RSC, RF,

* استاذ مساعد/ كلية العلوم/جامعة كركوك

تاريخ التسلم :2009/3/16 ـــــــــــ تاريخ القبول : 29 /6 /2009

RAPE) in selection of most adequate independent variables in the regression model used to estimate blood sugar as dependent variable and other independent variables and comparing the performance of these criteria. The results shows that the RAPE criterion was the best in selecting the most important variable compared with the other criteria.

المقدمـــة

من المسائل المهمة المتعلقة بالنمـاذج الخطيـة، اختيـار النمـوذج المناسب الذي يوفق المعلومات المتضمنة فـي متغيـر الاسـتجابة بـشكل كـفء ومختصر. ويعد اختيار أفضل مجموعة جزئية من مجموعة المتغيرات التوضـيحية من أحدى الأساليب الإحصائية المستخدمة في مسائل اختيار النموذج، إذ يتم ذلك من خلال توفيق كل النماذج الممكنة باستخدام كل المجـاميع الجزئيـة مـن مجموعـة المتغيرات التوضيحية المتوفرة واختيار النموذج الذي يحوي تلك المجموعة الجزئية من المتغيرات التوضيحية الأساسية في العلاقة للحصول على المعلومات نفسها التي نحصل عليها لا سيما لو شمل النموذج جميع المتغيرات.

تناول البحث اختيار المتغيرات في النماذج الخطية في حالة عـدم تحقـق بعض فرضيات النموذج الخطي مثل فرضية التوزيع الطبيعي للأخطاء العـشوائية وحالة وجود القيم الشاذة في البيانات،إذ ان مثل هذه الحالات لها أثـر سـلبي فـي أساليب اختيار النماذج وتحديد المتغيرات المهمة مما تؤدي إلى أن تكون النتائج غير دقيقة ومضللة ,ولذلك جاءت الإحصاءات الحصينة بوصفه أسلوبا مقاوما للقيم الشاذة والخروق الأخرى لفرضيات النموذج الخطي لاسيما المعايير الحصينة المـستخدمة في اختيار النموذج [1].

يهدف البحث إلى استخدام المعايير الحصينة في تحديد المتغيرات المهمة في نموذج الانحدار الخطي في حالة وجود القيم الشاذة في البيانات وذلك من خلال تحليل بيانات السكر في الدم بوصفه متغيرا معتمدا ومتغيرات أخرى مستقلة وبالاستناد إلى برنامج حاسبة يقوم بتكوين نموذج تام واشتقاق كل النماذج الجزئية الممكنة منه وتقدير معلمات هذه النماذج بطريقة M الحصينة وحساب معايير حصينة تفيد في تحديد المتغيرات المهمة في النموذج والمقارنة بين أداء هذه المعايير

لغرض تحقيق أهداف البحث، فقد تم تقسيمه على مبحثين ، المبحث الأول يمثل الجانب النظري وقد خصص للنماذج الخطية وتجزئته إلى نماذج جزئية واستعراض لطريقة التقدير M الحصينة في تقدير معلمات النموذج الخطي وبعد ذلك استعراض للمعايير الحصينة المستخدمة في تحديد المتغيرات واختيار النموذج المناسب, أما المبحث الثاني فيمثل الجانب التطبيقي للبحث ، إذ يتم فيه التطرق إلى نموذج الانحدار المستخدم في تقدير قياسات السكر والنماذج الجزئية المشتقة من النموذج التام وحساب المعايير قيد البحث ,وأخيرا تحليل نتائج التطبيق العملي والخروج ببعض الاستنتاجات حول أداء المعايير.

أولا: الجانب النظري

1-1 النموذج الخطي

يعرف النموذج الخطي بأنه علاقة رياضية تربط بين المتغيرات، ويتم التعبير عن هذه العلاقة بصيغة معادلة خطية تحتوي على متغير الاستجابة مع واحد أو أكثر من المتغيرات التوضيحية.

وبصيغة المصفوفات وبوجود (n) من المشاهدات و (k-1) من المتغيرات التوضيحية، يتم التعبير عن النموذج الخطي كالأتي [2]:

$$Y = XB + U \qquad \ldots(1)$$

اذ ان Y متجه (n× 1) لمشاهدات متغير الاستجابة , X مـصفوفة (n × k) لمشاهدات المتغيرات التوضيحية التي يتضمنها النموذج ومن ضمنها (X_0) المرافق للحد الثابت , B متجه (k × 1) للمعالم المجهولة في النموذج ومن ضـمنها الحـد الثابت , U متجه (n× 1) للأخطاء العشوائية بمتوسط صفر ومصفوفة تباين وتباين مشترك $\sigma^2 I$.

يعرف النموذج (1) بالنموذج الخطي العام والذي قد يظهر بعدة حـالات معتمدا على رتبة المصفوفة X , إلا أن هذه الدراسة ستركز على نموذج الانحـدار الخطي بوصفه حالة خاصة من النماذج الخطية .

ولأجل تقدير معلمات النموذج (1) فانه بموجب طريقة M الحصينة يتم تحوير الأسلوب المتبع في طريقة المربعات الصغرى، والذي يتطلـب تـصغير مجمـوع مربعات الخطأ وبعبارة أخرى:

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

إلى الصيغة الآتية [5]:

$$\min \sum_{i=1}^{n} \rho\left(\frac{Y_i - x_i' B}{\sigma}\right) \qquad \ldots(2)$$

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

اذ تم استبدال مربعات القيم بدالة عامة ρ . وبإتباع الأسلوب المستخدم في طريقة المربعات الصغرى نفسه يتم اشتقاق الدالة ρ بالنـسبة الـى المتجـه B ومساواتها الى الصفر لنحصل على المعادلة الاتية:

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

حيـ ـث ان $\psi = \frac{\partial \rho}{\partial B}$

ويمكن إعادة كتابة المعادلة الأخيرة بالصيغة الاتية:

$$\sum_{i=1}^{n} w_i x_{ij}(y_i - x_i'B)/S = 0 \qquad j = 1,2,\dots,k \qquad \dots(3)$$

حيث ان الأوزان w_i تعطى بالصيغة الآتية:

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

ومن ثم يتم حل المعادلة (3) لإيجاد قيمة الموجـ ـه b_M وسـ ـيكون وفـ ـق طريقة المربعات الصغرى الموزونة الآتية [3]:

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

حيث ان W مصفوفة قطرية عناصر القطر فيها قـ ـيم الأوزان w_i ، B_0 تمثل القيمة الاولية لموجه المعالمات، ويمكن ان تقدر بطريقة المربعات الـ ـصغرى الاعتيادية، وبالتعويض في صيغة w_i لإيجاد الأوزان نحصل على المتجه b_1 في التكرار الأول ونحصل على المتجه b_2 في التكرار الثاني وهكذا تستمر التكرارات الى ان تتقارب القيم في التكرارات المتتابعة فيكون b_M هو المقدر النهـ ـائي. اذ ان

معلمة القياس (σ) تقدر مرة واحدة باستخدام القيم الاولية للمعالمات فـي النمـوذج 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 الإحـصاءات الحـصينة هدفها الحصول على مقدر حصين ضد القيم الشاذة في البيانات، وأهم هذه الـصيغ هي دالة Huber الاتية[4]:

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

حيث ان sign(r) يشير الى اشارة r وان C هـو ثابـت القطـع ويأخـذ القيمـة C=1.345

ان مجموعة المتغيرات التوضيحية المتوافرة التي تحتويها المـصفوفة X قد يكون بعض منها غير أساسي في النموذج , وسـنرمز الـى عـدد مثـل هـذه المتغيرات بالرمز (q). في حين هناك متغيرات توضيحية تؤدي دورا أساسيا فـي النموذج وليكن عددها (p-1)، ويمكن عزل المتغيـرات (q) عـن المتغيـرات (p-1) من خلال إعادة كتابة النموذج (1) بالصيغة الآتية [2]:

$$Y = X_p B_p + X_q B_q + U \qquad \ldots(4)$$

اذ ان المصفوفة X تم تجزئتها الى المصفوفتين المجـزأتين 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 الأساسية في النموذج 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 النمـوذج. ونتيجة لتجزئة المصفوفة X كما في أعلاه فأن المتجه B الخـاص بالمعالمـات سوف يجزأ بدوره أيضا الى المتجهين الجزأيين B_p و B_q ذوات المرتبتين $(1 \times p)$ و $(1 \times q)$ على التوالي ولذلك فأن الرمز p سيشير الـى عـدد المعلمـات بضمنها الحد الثابت في النموذج الجزئي في حين الرمـز q يـشير الـى عـدد المتغيرات المستبعدة من النموذج، حيث ان $k = p + q$.

وبتغير أعمدة المصفوفتين X_p و X_q يمكن الحصول على جميع التراكيـب الممكنة للمصفوفتين المذكورتين، اذ سيكون هناك 2^{k-1} مـن المجـاميع الجزئيـة الممكنة للمتغيرات التوضيحية بضمنها المجموعة التامة. ومن هذه المجاميع يمكـن الحصول على 2^{k-1} من النماذج الجزئية بضمنها النموذج التام الذي يحوي جميـع المتغيرات التوضيحية .

2-1 المعايير الحصينة في اختيار النموذج

ان الهدف من الاحصاءات الحصينة هو ايجاد أساليب إحصائية جديدة قليلة الحساسية والتأثر تجاه القيم الشاذة وتعمل بشكل جيد تحت توزيعات مختلفة، وبالنظر لتطور طرائق تقدير معلمات النماذج الخطية وايجاد طرائق حصينة مختلفـة، فقـد دعت الحاجة الى تطوير المعايير التقليدية المستخدمة في اختيار النمـاذج وايجـاد معايير جديدة تسمح اولا باختيار النموذج المناسب الذي يمثل البيانات مـع الاخـذ

بنظر الاعتبار وجود قيم شاذة وحالة توزيع الاخطاء العشوائية قد لا يكون توزيعـا طبيعيا. وثانيا استخدام متوافق لهذه المعايير مع النماذج المقـدرة وفـق الطرائـق الحصينة.

وعليه فقد حظيت مسالة الاختيار الحـصين للنمـاذج الخطيـة بالاهتمـام المتزايد من قبل الباحثين في الآونة الأخيرة، وظهرت محاولات عديدة تهدف الـى إيجاد صيغ جديدة حصينة لبعض المعايير المستخدمة في اختيار النمـاذج الخطيـة. وفيما يأتي استعراض للمعايير التي تمت إعادة صياغتها او تحويرها بحيث أصبحت حصينة تجاه القيم الشاذة في البيانات وللخروق الأخرى لفرضيات النموذج الخطي:

1. معيار AIC الحصين

معيار AIC الحصين بالصيغة الآتية [9]:

$$RAIC = 2\sum_{i=1}^{n}\rho(\mathrm{r_i}) + 2\frac{\mathrm{E}\psi^2}{\mathrm{E}\psi'}p \qquad \ldots(5)$$

ولاختيار النماذج، فأنه يتم اختيار النموذج المقدر بالطريقة الحصينة الـذي ينتج عنه اقل قيمة لمعيار RAIC.

2. معيار SIC الحصين

ان صيغة معيار SIC الحصين هي كالأتي [6]:

$$RSIC = \sum_{i=1}^{n}\rho(r_i) + \frac{1}{2}p\,Ln\,n \qquad \ldots(6)$$

ولاختيار النماذج، فأنه يتم اختيار النموذج المقدر بالطريقة الحصينة الـذي ينتج عنه اقل قيمة لمعيار RSIC.

3. معيـار C_p الحصيـن

الصيغة الحصينة لهذا المعيار كالأتي [10]:

$$RC_p = \frac{W_p}{\hat{\sigma}^2} - (U_p - V_p) \qquad \ldots(7)$$

ولأغراض تطبيق الصيغة الحصينة للمعيار نحتاج الى ايجاد صيغ حسابية للمقدارين U_p و V_p ، ولقد تم تعريف صيغ تقريبية لها وفق نظريات فـي هـذا الصدد وكانت كالاتي:

$$V_p \cong tr(RM^{-1}\ QM^{-1})$$

$$U_p - V_p \cong E\|\psi\|^2 - 2tr(NM^{-1}) + tr(LM^{-1}\ QM^{-1})$$

حيث ان L , N , Q , R , M مصفوفات معرفة كما يأتي:

$$M = E(\psi' XX'), R = E(w^2 XX'), Q = E(\psi^2 XX'), N = E(\psi^2 \psi' XX')$$

$$L = E\left[(\psi'^2 + 2\psi' w - 3w^2)\ XX'\right]$$

$$\|\psi\|^2 = \sum_{i=1}^{n} \psi_i^2$$

ولأغراض اختيار النماذج، فبموجب هذا المعيار يتم اختيار النموذج الذي تنتج عنه قيمة RC_p قريبا من قيمة V_p.

وفيما يخص تباين الخطأ σ^2 فأنه يتم تقديره من النموذج التام (full) بموجب الصيغة الاتية:

$$\hat{\sigma}^2 = W_{full} / U_{full}$$

4. معيـار CV الحصين

صيغة هذا المعيار كالاتي [11]:

$$RCV = \sum_{i=1}^{n} \rho(y_i - x_i' b_M^{(i)}) / \hat{\sigma} \qquad ...(8)$$

حيث ان $b_M^{(i)}$ يمثل مقـدر M المحـصل عليـه باسـتخدام جميـع المشاهدات عدا المشاهدة (i)، $\hat{\sigma}$ هو تقدير معلمة القياس .

ولذلك يتم حساب معيار RCV لكل النماذج الجزئيـة ولجميـع التراكيـب الممكنة لتجزئة العينة واختيار النموذج الذي ينتج عنه اقـل قيمـة لمعـدل معيـار RCV بالنسبة لعدد التراكيب.

5. معـيار SC الحصين

صيغة هذا المعيار كالاتي[8,7]:

$$RSC = \sum_{i=1}^{n} \rho(\frac{r_i}{\sigma}) + \frac{1}{2} Ln\, E\psi' + \frac{1}{2} Ln\left|X'X\right| + Ln \prod_{i=1}^{p} \frac{\left|b_i\right| + n^{-1/4}}{\sigma} \qquad ...(9)$$

ويتم احتساب المعيار RSC لكل نموذج جزئي، وسيكون النمـوذج المناسـب الذي ينتج عنه اقل قيمة لهذا المعيار.

6. احصاءة F الحصين

صيغة احصاءة F الحصين تقوم على أساس التحوير في فكرة المربعات الصغرى اذ يتم تقـدير تشتت البواقي في النموذج التـام والنموذج الجزئي وإيجاد الفرق بينهما كالأتي [12]:

$$RF = \frac{\sum_{i=1}^{n} \rho(r_{pi}) - \sum_{i=1}^{n} \rho(r_i)/q}{\hat{\lambda}_M} \qquad \ldots(10)$$

حيث ان

$$\hat{\lambda}_M = D^2 \frac{\sum_{i=1}^{n} \psi_i^2 /(n-\rho)}{[\sum_{i=1}^{n} \psi_i' / n]^2}$$

اذ ان D يمثل معامل التصحيح الذي اوجده Huber لكي يكون المقدر غير متحيز، وتكون صيغة هذا المعامل كالاتي:

$$D = 1 + \frac{p}{n} \frac{Var(\psi')}{n[E(\psi')]^2}$$

اذ ان:

$$E(\psi') = \sum_{i=1}^{n} \psi_i' / n$$

$$Var(\psi') = \sum_{i=1}^{n} [\psi_i' - E(\psi')]^2 / n$$

الاحصاءة RF تتبع تقريبا توزيع F المركزي (تحت الفرضية H_0) بدرجة حرية q في البسط ودرجة حرية (n - k) في المقام، ومن ثم يتم حساب قيمة RF لكل نموذج جزئي ومقارنتها مع قيمة F الجدولية لمستوى معنوية معينة، فإذا كانت قيمة الاحصاءة المحسوبة اكبر من القيمة الجدولية تستنتج معنوية مجموعة المتغيرات في النموذج الجزئي، اما العكس فيعني عدم صلاحية النموذج الجزئي.

7. معيار APE الحصين

الصيغة الحصينة لهذا المعيار هي كالأتي [1]:

$$RAPE = \sum_{i=m}^{n} \rho \left(\frac{y_i - b'_{M(i-1)} x_i}{S} \right) \quad ...(11)$$

حيث ان m يشير الى أول مرحلة من المشاهدات التي تجعـل مقـدر M وحيدا، و b_M مقدر M حصين لمعالمات النموذج. وفيما يخص معلمة القياس S فأنها تقدر من النموذج التام بموجب الصيغة المشار إليها في المبحـث الأول مـن الجانب النظري من البحث.

ولأغراض اختيار النماذج، فأنه يتم حساب المعيار RAPE لكل نمـوذج من النماذج الجزئية واختيار النموذج الذي ينتج عنه اقل قيمة لهذا المعيار.

ثانيا: الجانب التطبيقـي

إن مرض السكر من أكثر الأمراض المزمنة انتشارا في عصرنا. وبالرغم من خطورة هذا المرض فانه يمكن السيطرة عليه من خلال متابعة تطورات المرض والعوامل المهمة المؤثرة فيه . لهذا فان هذا البحث يمكن أن يـساعد فـي تحديـد المتغيرات الأساسية ذوات التأثير المهم في المرض وذلك مـن خـلال اسـتخدام المعايير الحصينة في تحديد المتغيرات التوضيحية الداخلة في نموذج خطي لقيـاس السكر في الدم ، فقد تم تهيئة بيانات حقيقية وذلك من خلال تسجيل معلومـات مـن سجلات المصابين بداء السكري المراجعين لمركز أمراض الغدد الصم والسكري في بغداد ، وقد شملت العينة (35) سجلا، وكانت المعلومات عبارة عن قياسات الـسكر في الدم بوصفه متغيرا معتمدا وعمر المريض ووزنه ومقدار الكولسترول في الـدم وأخيرا قياسات اليوريا في الدم بوصفها متغيرات توضيحية نعتقد أنهـا تـؤثر فـي

التغيرات في مقدار السكر في الدم. وعند استشارة الأطباء المتخصصين في المركز تبين انه ليس هناك أسباب واضحة ومحددة للإصابة بهذا المرض ولكن الآراء تشير إلى أن زيادة وزن الشخص تزيد من احتمالية الإصابة بالسكري بينمـا المتغيـرات الأخرى المسجلة لا تمثل متغيرات مهمة تؤثر في الإصابة، وعلى هذا الأساس فقـد تم ترميز المتغيرات كالأتي:

Y = قياسات السكر في الدم

X_1 = عمر الشخص

X_2 = وزن الشخص

X_3 = قياسات اليوريا في الدم

X_4 = قياسات الكولسترول في الدم

وبهدف معرفة أية من المتغيرات التوضيحية المذكورة أعلاه التـي تـؤدي دورا أساسيا في الإصابة، فقد تم استخدام المعايير الحصينة موضوع البحث لاختيار أحسن نموذج جزئي من بين جميع النماذج الجزئية الممكنة المشتقة من النموذج التام الذي يحتوي على جميع المتغيرات المشار إليها في أعلاه وكالاتي:

$$Y = B_0 + B_1X_1 + B_2X_2 + B_3X_3 + B_4X_4 + u$$

وبذلك تكون هناك نماذج جزئية بعدد 16 نموذجا ($2^4 = 16$)، بحيـث ان كلا من هذه النماذج يحتوي على الحد الثابت (B_0). والجدول الأتي يبـين النمـاذج المنتخبة التي تتم المفاضلة بينها للوصول إلى النموذج المناسب .

رموز النماذج المنتخبة	النماذج المنتخبــة
(01234)	$Y = B_0 + B_1X_1 + B_2X_2 + B_3X_3 + B_4X_4$
(0123)	$Y = B_0 + B_1X_1 + B_2X_2 + B_3X_3$
(0124)	$Y = B_0 + B_1X_1 + B_2X_2 + B_4X_4$
(0134)	$Y = B_0 + B_1X_1 + B_3X_3 + B_4X_4$
(0234)	$Y = B_0 + B_2X_2 + B_3X_3 + B_4X_4$
(012)	$Y = B_0 + B_1X_1 + B_2X_2$
(013)	$Y = B_0 + B_1X_1 + B_3X_3$
(014)	$Y = B_0 + B_1X_1 + B_4X_4$
(023)	$Y = B_0 + B_2X_2 + B_3X_3$
(024)	$Y = B_0 + B_2X_2 + B_4X_4$
(034)	$Y = B_0 + B_3X_3 + B_4X_4$
(01)	$Y = B_0 + B_1X_1$
(02)	$Y = B_0 + B_2X_2$
(03)	$Y = B_0 + B_3X_3$
(04)	$Y = B_0 + B_4X_4$
(0)	$Y = B_0$

ولغرض معرفة المشاكل التي يعاني منها النموذج أعلاه فقد تم تقـدير المعلمـات وحساب المؤشرات الخاصة به مـن خـلال البرنـامج الجـاهز (Statgraf - 4) باستخدام البيانات المبينة في الجدول رقم (1)، وقد أشارت النتائج الـى أن هنـاك ارتباطا معنويا بين المتغيرين X_1 و X_2 ،وكذلك أشارت قيم البواقي الى ان هنـاك قيمتين غير مألوفتين (كبيرتين) فيها،مما تشير الى احتمال وجود قيمتين شاذتين في

البيانات، ومن ثم وفي ضوء هذه المشاكل فقد تم تطبيق البيانات الخاصة بالجانب التطبيقي على برنامج الحاسوب المصمم بالبحث الحالي لغـرض تقـدير معلمـات النموذج بطريقة M الحصينة وكذلك حساب المعايير الحصينة وفـق الـصيغ فـي المعادلات (5), (6) , (7) , (8) , (9) ,(10) , (11) التي تم التطرق إليهـا فـي الجانب النظري من البحث.

الجدول (2) يشير إلى القيم المحسوبة للمعايير الحصينة لجميع النماذج الممكنة من هذه المتغيرات والقيم التي تحتها خط تشير إلى اصغر قيمة محسوبة لكل معيار ومن ثم فأن النموذج المقابل لهذه القيم يمثل انسب نموذج وفق المعيار الذي تعـود إليه تلك القيمة. يلاحظ أن المعايير RCV, RF, RC_P, RAIC تميل إلى اختيـار النماذج التي تحتوي على المتغير X_2 ، وهي نماذج صحيحة ولكنها تحتوي علـى متغيرات إضافية يمكن أن تستبعد من النموذج. أما معيار RAPE فأنـه يختـار النموذج (02) الذي يحتوي على المتغير X_2 فقط وبذلك فأن هذا المعيار يختـار انسب نموذج يوافق رأي المتخصصين في مجال الطب. فـي حـين ان المعـايير RSIC و RSC اختارت النموذجين (03) و (0) على التوالي وهما نموذجان غير مناسبين لتمثيل الظاهرة المدروسة.

ثالثا: الاستنتاجات

في ظل المشاكل التي تم توضيحها حول النموذج الخطـي المـستخدم لتمثيـل قياسات السكر في الدم نستنتج ما ياتي:

1- أفضل المعايير أداءً كان المعيار RAPE ، اذ ان اختيـار هـذا المعيـار للنموذج جاء متوافقا تماما مع رأي المتخصصين في مجال الطب، اذ أشار المعيار RAPE الى أن متغير الوزن هو المتغير الأساسي والوحيد من بين المتغيـرات التوضيحية الأخرى المسجلة في التطبيق العملي.

2- المعايير RCV, RF, RC_P, RAIC كان أداؤها اقل دقـة فـي اختيـار النموذج المناسب، اذ تضمن النموذج وفق هذه المعايير متغيرات اخرى اضـافية غير مهمة ويمكن أن يعزى ذلك إلى المشاكل التي يعاني منها النموذج المـدروس أو إلى مشاكل أخرى لم يتم تحديدها في النموذج المستخدم في البحث.

3- وفيما يخص المعيارين RSIC و RSC فان اداؤهما كانـا غيـر جيـدين بالنسبة الى النموذج المستخدم في البحث الحالي لانهما اديا الى اختيار متغيـرات غير مناسبة فيما يخص قياس السكر في الدم.

الجدول (1) : قياسات السكر، العمر، الوزن، اليوريا في الدم والكولسترول في الدم لمرضى السكري

No.	السكر	العمر	الوزن	اليوريا	الكولسترول
1	288	50	71	23	212
2	358	55	60	48	253
3	170	49	61	23	185
4	225	65	65	32	272
5	238	36	85	26	280
6	260	52	98	54	148
7	292	53	56	22	148
8	260	43	60	22	130
9	250	60	80	38	254
10	415	70	71	48	171
11	225	45	75	18	275
12	302	54	60	26	206
13	224	53	96	23	166
14	285	65	52	26	265
15	180	48	72	24	220
16	243	48	71	27	240
17	188	52	75	24	240
18	247	55	70	23	262
19	212	54	80	29	231
20	200	58	73	26	233
21	247	60	68	24	186
22	280	45	72	32	119
23	195	70	61	32	175
24	288	48	73	28	156
25	352	44	68	23	220
26	392	50	73	30	280

27	193	47	72	19	216
28	175	55	81	31	125
29	166	50	87	33	185
30	170	57	80	38	150
31	204	70	60	38	220
32	190	46	66	23	198
33	192	57	81	40	355
34	278	52	86	29	320
35	285	67	68	72	257

الجدول (2) : قيم المعايير الحصينة لاختيار النموذج لبيانات مرضى السكري

النماذج المنتخبة	**المعاييـ ـر الحصينة**						
	RAIC	**RSIC**	**RSC**	**RC_P**	**RF**	**RCV**	**RAPE**
(01234)	14.396	14.509	16.398	-	-	57.007	11.815
(0123)	**<u>13.860</u>**	12.767	12.983	0.691	**<u>0.200</u>**	0.257	11.232
(0124)	15.680	13.553	13.238	**<u>0.661</u>**	4.193	0.290	11.839
(0134)	15.655	13.485	13.299	0.668	3.806	0.288	12.313
(0234)	14.664	13.056	13.143	0.687	1.804	0.270	11.651
(012)	15.085	11.817	10.237	1.464	2.197	0.275	11.438
(013)	15.012	11.741	10.218	1.507	2.010	0.370	11.574
(014)	15.842	12.126	11.380	1.522	2.962	0.292	12.001
(023)	13.993	11.288	10.120	1.472	0.952	**<u>0.254</u>**	11.028
(024)	15.029	11.786	10.300	1.469	2.110	0.270	11.153
(034)	15.096	11.766	10.279	1.525	2.87	0.276	12.015
(01)	15.180	10.387	9.487	2.411	2.096	0.277	11.649
(02)	14.398	10.044	9.380	3.286	1.470	0.257	**<u>10.807</u>**
(03)	14.399	10.008	**<u>9.372</u>**	2.390	1.442	0.260	11.267
(04)	15.090	10.352	9.54	2.386	1.985	0.273	11.426
(0)	14.428	**<u>8.620</u>**	9.492	2.508	1.548	0.261	11.092

المصادر

1- صباح حسيب حسن,(2007)," اختيار نموذج حصين في الانحدار الخطي",
مجلة جامعة كركوك, الدراسات العلمية , المجلد 2 ,العدد 1,ص55-69.

2. Hocking, R. R., (1976), "**The Analysis and Selection of Variables in Linear Regression**", Biometrics, 32, 1- 40.

3. Holland, P. W. & Welsch, R. E., (1977), "**Robust Regression Using Alternatively Reweighted Least Squares**', Comm. Statist., A(6), 813-827.

4. Huber, P. J., (1981), "**Robust Statistics**", New York, John Wiley & Sons.

5. Launer, R.L., & Wilkingson, G. N., (1979), "**Robustness in Statistics**",
Acadimic Press, Inc., New York.

6. Machado, J. A. F., (1993), "**Robust Model Selection and M - Estimation**", Economatric Theory, 9, 478 – 493.

7. Qian, G. & Kunsch, H., (1998), "**Some Notes on Rissanen's Stochastic Complexity**", IEEE, Trans. Inform. Theory, 44, 782–786. (By Internet).

8. Qian, G. & Kunsch, H., (1998), "**On Model Selection in Robust Linear Regression**", J. Stat. Plan. & Inform. (By Internet).

9. Ronchetti, E., (1985), "**Robust Model Selection in Regression**", Statistics of Probability Letters, 3, 21- 23.

10. Ronchetti, E.,& Robert, S., (1994), "**Robust Version of Mallows C_p**", J. Amer. Statist. Assoc., 89, 550 – 559.

11. Ronchetti, E., Field, C., & Blanchard, W., (1997), "**Robust Linear Model Selction by Cross - Validation**", JASA, 92, 1017 – 1023.

12. Schrader, R. M. & Mckean, J. W., (1977), "**Robust Analysis of Variance**", Comma Statist., A6 (9), 879 – 894.