متى ولماذا تفشل طريقة stepwise في اختيار أفضل مجموعة جزئية لمتغيرات الانحدار؟

أ.م. د. صباح فرج عبد الحسين*

المستخلص:

لاشك أن طريقة stepwise هي أحدى أفضل طرائق الاختيار المتسلسل لمتغيرات الانحدار، إلا أنها -حسب تقديري- تعاني من عيب أساس هو أنها في غالب التطبيقات لا يمكنها التخلص من تأثير المتغير- الذي يتم أختياره في المرحلة الأولى منها- على اختيار المتغيرات المهمة الأخرى مما يؤدي في حالة وجود علاقات خطية قوية بينه وبين هذه المتغيرات إلى استبعادها (كلاً أو جزءاً) من المجموعة الجزئية النهائية المختارة مما يعني في النتيجة الفشل في أختيار أفضل مجموعة لمتغيرات الانحدار. وهذا يشير إلى أن ذلك المتغير يقف حاجزاً مانعاً أمام اختيار المتغيرات التي تتلوه في الأهمية الابتدائية لقوة علاقاته الخطية بها وبذلك تنتفي الفائدة من إعادة تقييمه في كل مرحلة من مراحل الطريقة ضمن المجموعات الجزئية التي يسمح بتكوينها إلا إذا كسر هذا الحاجز بطريقة ما، كرفع مستوى معنوية إدخال المتغيرات إلى المجموعات الجزئية بحيث لا يقل عن 0.10. وحسب الأهمية الابتدائية لمتغيرات الانحدار التي يراد من بينها اختيار أفضل مجموعة جزئية، وهو ما اقترحه الباحث لتفعيل التعديل الذي جاءت به طريقة stepwise على طريقة foreword، مع ملاحظة أن ذلك قد يؤدي إلى اختيار متغيرات إضافية ربما تكون ضعيفة المساهمة في التأثير في الظاهرة موضوع البحث في المجتمع.

Abstract:

No doubt that step wise procedure (SP) is one of the best procedures of serial choice of regression variables, but it suffers – as I see – from abasic defect that, in most applications it can't getrid of the effect of first selected variable on selecting the other ones in case of presence of linear correlations between them. The first selected variable stands as a screen before joining the other important variables "totally or partially" the finalsubset. Consequently, (SP) fails selecting the best subset regression variables (BSRV). So there will be no use of retesting it during the procedure unless one breaks the screen in such away like raising the significance level of testing the selected variables such that it won't be less than 0.10 and according to their importance. This is what the researcher has proposed to make the (SP) more efficient than foreword to select (BSRV). But one must

* أستاذ مساعد ورئيس قسم أنظمة الحاسوب في معهد الإدارة / الرصافة.

مقبول للنشر بتأريخ 2012/11/22

notice that this proposal may lead to select additional variables that have weakened contribution in explanation of social phenomenon.

I- المقدمة والهدف:

كثيراً ما يحتاج الباحث في دراسته لظاهرة عشوائية أن يحدد- من بين مجموعة العوامل العديدة المؤثرة فيها- مجموعة جزئية معينة هي الأفضل من بين المجموعات الجزئية الممكنة التي لها القدرة على تفسير تلك الظاهرة والتنبؤ بمستقبلها وذلك اعتماداً على معايير المفاضلة المعروفة (R_a^2, PRESS, CP, MSE). ولكن ايجاد وفحص جميع المجموعات الجزئية الممكنة هو اجراء غير عملي فهو مكلف في الجهد والوقت لأن الزيادة في عدد هذه المجموعات هي زيادة هندسية تتناسب وعدد العوامل (المتغيرات المستقلة أو التفسيرية) التي تؤثر فيها. ولذلك اقترحت طرائق الاختيار المتسلسل لتلك المتغيرات وصولاً إلى المجموعة الجزئية الأفضل. ومن أشهر هذه الطرائق هي طريقتان؛ أحداهما تسمى الاختيار الأمامي Fore word selection التي تعتمد على اختيار المتغيرات تباعاً (متغير في كل مرة) على أساس قوة مساهمته (المعنوية) في المجموعة الجزئية أو ما يفسره من الانحرافات في قيم الظاهرة موضوع البحث. أما الأخرى فهي الحذف التراجعي Backword Elemination وفيها يجري حذف المتغيرات من المجموعة الكلية تباعاً (متغير في كل مرة) على نفس الأساس السابق ولكن بصورة معكوسة أي يجري حذف المتغير من المجموعة الذي له أضعف مساهمة (مساهمة غير معنوية) في تفسير الانحرافات في قيم الظاهرة. وأما الإحصاءه المستخدمة في اختبار معنوية هذه المساهمات (أهمية المتغيرات) فهي إحصاءة T أو F- الجزئية وذلك بمستوى معنوية يحدد ابتداءً وعادة ما يكون 0.05 أو 0.10. وقد نشر الباحث الأمريكي Efroymson عام 1962 طريقته في الاختيار المتسلسل للمتغيرات التفسيرية[1].

وهي نفسها الطريقة للاختيار الأمامي مع تعديل بسيط ولكن مهم وهو أن يستمر اختبار كل متغير في المجموعات الجزئية المتكونة وهذا يعني إمكانية أن يكون احد المتغيرات الداخل في مجموعة جزئية، خارجاً من مجموعة جزئية لاحقة في مرحلة من مراحل هذه الطريقة على أساس أن معنوية أي متغير في المجموعات الجزئية ليست ثابتة وأنما هي تتأثر بعلاقته الخطية مع المتغيرات الأخرى في تلك المجموعات ومقدار التحيز في تقدير انحرافه المعياري، وبكلمة واحدة فأنها تتأثر بالاتجاه والقيمة المطلقة لمركبه انحداره غير المباشر[2]. وقد أسمى طريقته في تعديل الاختيار الأمامي للمتغيرات التفسيرية بطريقته stepwise لاعتقاده بأنه بذلك قد استكمل خطوات الاختيار الحكيم للمتغيرات التي تفسر كمجموعة جزئية أفضل من غيرها الاختلافات في قيم الظاهرة موضوع البحث. فهل كان الباحث المذكور محقاً في اعتقاده؟! وإذا لم يكن كذلك كما لوحظ في التطبيق على العديد من المسائل الاجتماعية والاقتصادية والحيوية، فلماذا فشلت طريقته في اختيار أفضل مجموعة جزئية لمتغيرات الانحدار في بعض التطبيقات؟؟ ومتى تكون طريقته ناجحة ومتى تكون فاشلة؟ إن الإجابة عن هذه الأسئلة هي محور هذا البحث وهدفه.

II- التحليل النظري لطريقة step wise في اختيار متغيرات الانحدار: تقوم طريقة step wise كما هي طريقة الاختيار الأمامي في مرحلتها الأولى على انتقاء المتغير الذي له أكبر قيمة معامل ارتباط بسيط بالمتغير المعتمد والذي تقدير معنوية معامل انحداره على وفق اختبار t أو f -الجزئية بمستوى معنوية معين أكبر من المعنوية الناجمة عن عشوائية العينة بنفس احتمال المعنوية، ثم يصار إلى اختيار المتغيرات الأخرى تباعاً في المراحل اللاحقة وهنا لدينا حالتان:

الحالة الأولى:

ضعف أو عدم وجود علاقات خطية ما بين المتغيرات التفسيرية، وبذلك لا تتأثر سلباً معنوية تقدير معامل انحدار المتغير الجديد بوجود المتغير أو المتغيرات التي تشكل المجموعة الجزئية التي يدخل إليها وخصوصاً المتغير الأهم الذي تم اختياره في المرحلة الأولى وإنما على العكس فأن تقدير معنوية معامل انحدار المتغير الجديد تزداد بدخوله على المجموعة الجزئية وذلك لتناقص التحيز في تقدير انحرافه المعياري وهذا يعني أن اختيار المتغيرات يكون تباعاً حسب أهميتها الابتدائية (قوة علاقتها الخطية بالمتغير المعتمد) وفي هذه الحالة لا يخرج أي متغير من المجموعة بعد اختياره فتتساوى طريقة الاختيار الأمامي مع طريقة step wise.

الحالة الثانية:

وجود علاقات خطية قوية ما بين متغيرات الانحدار المهمة وفي هذه الحالة يقف المتغير الذي يتم اختياره في المرحلة الأولى من عملية انتقاء المتغيرات حاجزاً مانعاً لانضمام المتغيرات إليه تباعاً في مجموعات جزئية حسب أهميتها الابتدائية التي تشير إليها قوة علاقتها الخطية بالمتغير المعتمد، فتكون محصلة مركبة الانحدار غير المباشر لكل متغير فيها هي التي تحدد تقدير معنويته وبالتالي استبعاده من المجموعة أو إدخاله إليها وفق نفس إحصاءة الاختبار (t أو f – الجزئية) وإذا كانت طريقة الاختيار الأمامي

fore ward لا تسمح بإعادة تقييم متغير تفسيري سبق إدخاله في مجموعة فأن طريقة stepwise وأن كانت تتضمن إعادة التقييم فأن ذلك لا يكون فعالاً لأنه يصطدم بالمتغير (الحاجز) الذي لا يسمح بدخول المتغير التالي له في الأهمية الابتدائية نتيجة قوة علاقته الخطية به، وبذلك لا يتحقق إعادة تقييم المتغير الأول (المتغير الحاجز) مع المتغير الجديد (المهم) وإنما يكون ذلك مع متغير آخر أقل أهمية وأضعف علاقة خطية بالمتغير الأول وتكون النتيجة المتوقعة للتقييم هي لصالح بقاء المتغير (الحاجز) في أغلب الأحيان وعلى حساب الأهمية الابتدائية لمتغيرات الانحدار الأخرى، وهذا ما يؤدي إلى فشل طريقة stepwise في اختيار أفضل مجموعة جزئية لمتغيرات الانحدار. وقد دفع هذا الفشل، البعض، إلى وصف هذه الطريقة بـ(step down) بدلاً من (step wise)[3]. بينما دفع البعض الآخر إلى البحث عن بدائل أخرى لتقدير معاملات انحدار المتغيرات التفسيرية مثل طريقة stage wise[4] التي يعاب على معاملات انحدار مجموعة المتغيرات التفسيرية التي تختارها كبر أخطائها المعيارية بحيث يصبح متوسط مربعات البواقي لمعادلة انحدارها كبيراً قياساً بطرائق الاختيار المعروفة التي تعتمد تقديرات المربعات الصغرى. وكذلك طريقة انحدار الـ(Ridge) التي يعاب عليها الافتراض المسبق لحجوم معالم نموذج الانحدار المراد تقديرها فضلاً عن الحصول على مقدرات متحيزة لتلك المعالم لا يمكن معالجة تحيزها[6,5].

III- مستوى المعنوية لدخول وخروج المتغير التفسيري من المجموعة الجزئية: لوحظ أن مستوى المعنوية لدخول وخروج المتغير التفسيري من المجموعة الجزئية عند استخدام طريقة step wise تؤثر في تكوين المجموعة الجزئية النهائية، وقد استقر الرأي عند أغلب الباحثين في مجال تطبيقات الانحدار الخطي أنه من الحكمة أن لا يزيد مستوى معنوية أدخال المتغير التفسيري عن مستوى معنوية إخراجه من المجموعة الجزئية لأن خلاف ذلك يعني التضحية بالمتغيرات المهمة التي يتم إدخالها[7]. ولذلك انقسم الباحثون إلى قسمين أحدهما يفضل أن يكون مستوى المعنوية واحداً لإدخال وإخراج المتغيرات التفسيرية لكي لا تحدث حالة الدوران في اختيار المتغيرات (cycling pattern) ومن أبرز الداعين إلى ذلك (Draper and Smith)، ونجد هذا الاتجاه مترجماً في بعض أنظمة الحاسوب للتحليل الإحصائي مثل نظام SAS (statistical analysis system).

أما القسم الثاني فهو الذي يدعو إلى أن يكون مستوى معنوية اخراج (إبقاء) المتغير التفسيري أكبر من مستوى معنوية ادخاله لضمان المحافظة على المتغيرات (المهمة) الداخلة ولتحقيق خاصية الاتساق (consistency) في اختيار المتغيرات التفسيرية ضمن المجموعة الجزئية النهائية ومن أبرز الداعين إلى ذلك BOWERMAN & O'CONNELL[8]. ونجد هذا الاتجاه مترجماً في برامج الانحدار ضمن أنظمة حاسوب أخرى مثل نظام SPSS :.(Statistical package for social sciences)[9] والحقيقة أن كلا الاتجاهين لا يفيد في أضفاء الفاعلية على اختبار (بقاء) المتغير (الحاجز) بل أن الاتجاه الثاني يعزز بقاءه وبالتالي يؤدي إلى تكوين مجموعة جزئية منسجمة معه (ضعيفة الارتباط الخطي به) حتى وأن ادى ذلك إلى استبعاد عدد من المتغيرات المهمة وهو ما ينتهي بعملية step wise إلى اختيار مجموعة جزئية هي ليست المجموعة الأفضل لتفسير الظاهرة والتنبؤ بمستقبلها. وهكذا فأن التعديل الذي جاءت به طريقة step wise لا يمكن تفعيله إلا بزيادة مستوى معنوية ادخال المتغير التفسيري بحيث لا يقل عن 0.10 وحسب الأهمية الابتدائية للمتغيرات التفسيرية لكي تسمح بكسر حاجز المتغير الذي تم انتقاؤه في المرحلة الأولى ودخول المتغيرات المهمة التي لا ترتبط فيما بينها بعلاقات خطية قوية وانما ترتبط مع ذلك المتغير فقط، وعند ذاك يمكن أن يفعل أختبار المتغير (الحاجز) كمتغير داخل على تلك المتغيرات المهمة فيبقى أو يخرج من المجموعة، كما سنلاحظ ذلك في التطبيق العملي، ولكن ذلك قد يسمح بادخال متغيرات إضافية ضعيفة المساهمة في تفسير الظاهرة موضوع البحث والتنبؤ بمستقبلها.

IV- التطبيق العملي وتحليل النتائج:

تم اختيار مسألتين شائعتي الاستخدام في مجال الانحدار الخطي وهما مسألة Hald[10]. حول الحرارة المنبعثة من مزج أربعة عناصر من مادة السمنت، ومسألة cravens حول أداء مندوبي المبيعات في عدة مناطق (أقاليم) والتي أعاد تحليل بياناتها BOWERMAN and O'CONNELL عام 1997[11].

المسألة الأولى: مسألة Hald(*): وبحساب معاملات الارتباط البسيط لبيانات كل زوج من المتغيرات نحصل على المصفوفة الآتية:

(*) لاحظ بيانات هذه المسألة في الملحق.

جدول رقم (1)
مصفوفة الارتباط للمتغيرات (المستقلة) في المسألة والمتغير المعتمد

	X_1	X_2	X_3	X_4	Y
X_1	1	229	-.829	-.245	.731
X_2		1	-.139	-.973	.816
X_3			1	.030	-.535
X_4				1	-.821
Y					1

من خلال مصفوفة الارتباط لأزواج المتغيرات وفي ضوء التحليل النظري لطريقة step wise ودور العامل الحاجز فيها، يمكن القراءة الأولية للمجموعات الجزئية التي ستتكون بهذه الطريقة وكما يلي:

1) **أن المتغير X_4 هو المتغير الأكثر أهمية ابتدائية لأنه الأكثر ارتباطاً بالمتغير المعتمد، ويليه المتغير X_2- المقارب له في الأهمية- ثم X_1 وفي الأخير X_3.**
2) **أن المتغير المستقل X_4 يرتبط بعلاقة خطية قوية جداً بالمتغير X_2 مما يعني أن دخوله (X_4) في المرحلة الأولى سيكون (حاجزاً) أمام دخول المتغير المقارب له في الأهمية (X_2) عند مستوى معنوية معين (0.05).**
3) **أن المتغير الثالث في الأهمية الابتدائية (X_1) يرتبط بعلاقة خطية ليست قوية (ضعيفة نسبياً) بالمتغير الأكثر أهمية (X_4) مما يسمح بدخوله معه في مجموعة جزئية واحدة ولكن مع أضعاف نسبي لقوته التفسيرية والتنبؤية.**
4) **أن المتغير الأخير في الأهمية (X_3) يرتبط بعلاقة خطية قوية جداً مع المتغير (X_1) تمنع دخوله إلى المجموعة الجزئية المتكونة في المرحلة الثانية من (X_1, X_4).**

وهذا يعني أن المجوعة الجزئية النهائية لمتغيرات الانحدار المتوقعة بطريقة step wise هي نفسها المجموعة الجزئية التي نحصل عليها باستخدام طريقة الاختيار الأمامي fore ward وذلك لأن اعادة اختبار المتغيرات في المجموعات الجزئية لهذه المسألة لا يؤدي إلى اخراج أي متغير منها، وبذلك تكون معادلة الانحدار الخطي وفق هذه الطريقة هي:

$$\hat{y} = f(X_1, X_4) = 103.097 + 1.440\ X_1 - 0.614\ X_4$$

sig. 0.000 0.000 0.000

MSE = 7.476, R_a^2= 0.964, C_p= 5.50, $\overline{VIF}$= 1.064, PRESS= 121

وللتأكد من صحة هذه القراءة لمسيرة خطوات طريقة step wise، نطبق هذه الخطوات عملياً ونلخصها في الجدول الأتي:

جدول رقم (2)
ملخص نتائج تطبيق خطوات طريقة step wise بمستوى معنوية 0.05

مراحل التطبيق	F- الجزئية للمتغيرات	FIN = Fout	المتغير الداخل	المتغير الخارج
الدورة(1): الانحدار على كل متغير تفسيري	$F_{(1)}$= 12.60, $F_{(2)}$ = 21.96 $F_{(3)}$= 4.40, $F_{(4)}$ = 22.80	F (0.05, 1, 11) = 4.48	X_4	-
الدورة(2):الانحدار على المتغيرين (Xj, X4) حيث j=1,2,3	$F_{(1)}$= 108.22,$F_{(2)}$ = 0.1725 $F_{(3)}$=40.29,$F_{(4/1)}$ = 159.930	F (0.05, 1, 10) = 4.96	X_1	-
الدورة(3): الانحدار على المتغيرات (x_1, x_4,x_j) j= 2,3	$F_{(2)}$ = 5.02, $F_{(3)}$= 4.23	$F_{(.05,1,9)}$= 5.12	-	-

وبذلك تكون المجموعة الجزئية النهائية لمتغيرات الانحدار المتحققة بطريقة step wise عند مستوى معنوية 0.05 هي المجموعة التي تم قراءتها مسبقاً وهي{X_4, X_1}.

وحيث أن X_4 وX_2 متقاربان جداً في الأهمية ومرتبطان خطياً بقوة لذلك فأن رفع مستوى معنوية إدخال المتغيرات من 0.05 إلى 0.10 قد يكون كافياً لكسر حاجز اختيار أي منهما لدخول معادلة الانحدار الخطي في المرحلة الأولى، ومن ثم تفعيل اختبار دخوله ثانية بوجود المتغير الآخر أي اختبار بقائه في مجموعة جزئية مع الآخر في مرحلة لاحقة. وللتأكد من ذلك نعيد ايجاد معادلة الانحدار الخطي بطريقة step wise ولكن بمستوى معنوية 0.10 لدخول وخروج المتغيرات.

جدول رقم (3)
ملخص نتائج تطبيق خطوات طريقة step wise بمستوى معنوية 0.10

مراحل التطبيق	F- الجزئية للمتغيرات	FIN = Fout	المتغير الداخل إلى النموذج	المتغير الخارج
الدورة الأولى	$F_{(1)}$= 12.60, $F_{(2)}$ = 21.96 $F_{(3)}$= 4.40, $F_{(4)}$ = 22.80	3.23	X_4	-
الدورة الثانية	$F_{(1)}$=108.22, $F_{(2)}$ = 0.1725 $F_{(3)}$= 40.29, $F_{(1/4)}$ = 159.930	3.29	X_1	-
الدورة الثالثة	$F_{(2)}$=5.02, $F_{(3)}$ = 4. 23 F(4/1,2) = 1.86 F (1/4,2) = 154.01	3.36	X_2	X_4
الدورة الرابعة	$F_{(3)}$= 1.83	3.36	-	-

∴ المعادلة الأخيرة لتفسير الظاهرة (Y) والتنبؤ بمستقبلها عند مستوى معنوية 0.10 هي بدلالة المجموعة الجزئية لمتغيرات الانحدار $\{X_2, X_1\}$:

$\hat{y}$= 52.5773 +1.4683 X_1 +.6623 X_2
sig. 0.000 0.000 0.000
MSE = 5.790, R_a^2= 0.975, C_p= 2.68, $\overline{VIF}$= 1.055, PRESS= 95

ويلاحظ أن معادلة الانحدار المتحققة بطريقة step wise. قد تحسنت بعد كسر حاجز المتغير الأول الذي له علاقة خطية قوية بالمتغيرات المهمة التي لا يوجد بينها علاقات خطية قوية، وهذا ما اوضحناه في تحليلنا النظري لهذه الطريقة.

المسألة الثانية: مسألة Cravens(*): وبحساب قيم معاملات الارتباط البسيط لبيانات كل زوج من المتغيرات نحصل على المصفوفة الأتية:

جدول رقم (4)
مصفوفة الارتباط للمتغيرات (المستقلة) في المسألة والمتغير المعتمد

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_7	Y
X_1	1	.454	.249	.106	.251	.758	-.171	.101	.623
X_2		1	.174	-.211	.268	.479	-.259	.359	.598
X_3			1	.264	.377	.200	-.272	.411	.596
X_4				1	.085	.403	.349	-.024	.484
X_5					1	.327	-.288	.549	.489
X_6						1	-.199	.229	.754
X_7							1	-.277	-.117
X_8								1	.402
Y									1

من مصفوفة الارتباط لكل زوج من المتغيرات وباتباع نفس التحليل النظري لطريقة step wise ودور المتغير الحاجز فيها يمكن القراءة الأولية للمجموعات الجزئية المتكونة من المتغيرات التفسيرية وكما يلي:

1) أن المتغير التفسيري الأكثر ارتباطاً بالمتغير المعتمد هو X_6، وبذلك فهو الأكثر أهمية في المرحلة الأولى لذلك يتم انتقاؤه لدخول معادلة الانحدار.
2) أن المتغير X_6 يرتبط بعلاقة خطية قوية بالمتغير X_1 الذي يليه في الأهمية الابتدائية مما يغني أن دخوله (X_6) في المرحلة الأولى سيكون حاجزاً أمام دخول المتغير التالي له في الأهمية (X_1) عند مستوى المعنوية 0.10 بل وحتى 0.15 لأن ارتباط المتغيرات التفسيرية بالمتغير المعتمد ليس عالياً.
3) أن المتغيرين X_2 وx_3 متقاربان في الأهمية الابتدائية من خلال قوة ارتباط كل منهما بالمتغير المعتمد ولكن X_3 يتميز على X_2 بأن علاقته الخطية بالمتغير الحاجز (X_6) هي الأضعف، لذلك فأن من المنطقي أن يسمح X_6 لـX_3 بالدخول في مجموعة جزئية في المرحلة الثانية حيث تكون معنويته هي الأكبر.
4) حيث أن X_3 لا يرتبط بعلاقة خطية قوية بـX_2 لذلك يمكن أن يكون X_2 هو المتغير الثالث في المجموعة الجزئية في المرحلة الثالثة إذا لم يكن X_2 دالة خطية لـX_3 وX_6 معاً.

(*) لاحظ بيانات هذه المسألة في الملحق.

أما المرحلتين الباقيتين فقد يكون فيهما X_4 أو X_5 وقد لا يكون فذلك يتوقف على قوة ارتباط كل منهما الجزئي بالمتغير المعتمد في ظل وجود المجموعة الجزئية المتكونة في المرحلة الثالثة $\{X_6, X_3, X_2\}$. ويلاحظ أن اعادة اختبار المتغيرات في المراحل اللاحقة للمرحلة الأولى لا يؤدي إلا إلى تعزيز بقائها وذلك لأنه يتم اختيارها بحيث لا يكون بينها وبين المتغير الحاجز علاقة خطية قوية، أضف إلى ذلك أن توسع المجموعات الجزئية بادخال متغيرات مهمة إليها يؤدي إلى تناقص تحيز تقدير انحرافاتها المعيارية وبذلك تزداد معنوية معاملاتها، وهكذا تنتفي الفائدة من اعادة الاختبار وتتساوى طريقة step wise مع طريقة fore ward.

مراحل التطبيق	F- الجزئية للمتغيرات	FIN = Fout	المتغير الداخل إلى النموذج	المتغير الخارج
الدورة الأولى	$F_{(1)}$= 14.583, $F_{(2)}$ = 21.791 $F_{(3)}$= 12.683, $F_{(4)}$ = 7. 018 $F_{(5)}$= 7.235, $F_{(6)}$ = 30.302 $F_{(7)}$= 0.320, $F_{(8)}$ = 4.430	2.94	X_6	-
الدورة الثانية	$F_{(1)}$= 0.323, $F_{(2)}$ = 4.466 $F_{(3)}$= 20.210, $F_{(4)}$ = 2.157 $F_{(5)}$= 3.956, $F_{(7)}$ = 0.057 $F_{(8)}$= 3.253, $F_{(613)}$ = 41.053	2. 95	X_3	-
الدورة الثالثة	$F_{(1)}$=0.009, $F_{(2)}$ =6. 414 $F_{(4)}$= 1. 090, $F_{(5)}$ = 1.197 $F_{(7)}$= 2.259, $F_{(8)}$ = 0.422 F(6/3,2) = 26.822 F (3/6,2) = 22.723	2.96	X_2	-
الدورة الرابعة	$F_{(1)}$=0.070, $F_{(4)}$ =14. 611 $F_{(5)}$= 0. 881, $F_{(7)}$ = 5.347 $F_{(8)}$= 0.601, $F_{(6/2,3,4)}$ = 10.341 F(3/2,4,6) = 22.481 F (2/3,4,6) = 22.954	2.97	X_4	-
الدورة الخامسة	$F_{(1)}$= 1.198, $F_{(5)}$ = 2.596 $F_{(7)}$= 0.46, $F_{(8)}$ = 0.17	2.99	-	-

يتضح من تطبيق step wise أن اختيار المتغير X_6 في الدورة الأولى كأن حاجزاً مانعاً للمتغير X_1 – التالي له في الأهمية الابتدائية- من الدخول معه في مجموعة جزئية تفسر الاختلافات في قيم المتغير المعتمد وذلك لأنه يهبط بأهميته إلى الصفر تقريباً في الدورات اللاحقة بينما يسمح للمتغير X_3 الرابع في الأهمية الابتدائية من مشاركته في المجموعة الجزئية في الدورة الثانية لكون علاقته الخطية به هي الأضعف ولارتفاع معنويته في المجموعة من 12.683 إلى 20.210 بسبب تناقص تحيز تقدير انحرافه المعياري. وقد أتيحت الفرصة في الدورة الثالثة لدخول المتغير الثالث في الأهمية الابتدائية X_2 إلى مجموعة الدورة الثانية (X_3, X_6). لزيادة أهمية X_2 في الدورة الثالثة لتناقص تحيز تقدير انحرافه المعياري وضعف علاقته بالمتغير المشارك للمتغير الحاجز ولكن بعد أن فقد نصف أهميته الابتدائية بسبب ذلك الحاجز. ثم دخل X_4 في الدورة الرابعة بعد ارتفاع معنويته (أهميته) عند انضمامه لمجموعة الدورة الثالثة. وبعدها توقف دخول المتغيرات الباقية لعدم تجاوز معنويتها حاجز العتبة FIN بمستوى معنوية 0.10. ويلاحظ أن المتغير X_1 وهو المتغير الثاني في الأهمية الابتدائية قد بقي إلى نهاية خطوات التطبيق يبدو غير مهم على الأطلاق بسبب المتغير الحاجز X_6، كما بقي X_6 وغيره من المتغيرات الداخلة لا تتأثر باعادة الاختبار لغرض البقاء أو عدم البقاء في المجموعة، وبذلك تساوت stepwise مع foreward وفقد التعديل الذي اقترحه Efroymson فائدته، وانتهت خطوات step wise إلى المجموعة الجزئية في الدورة الرابعة $\{X_6, X_4, X_3, X_2\}$ وإلى معادلة الانحدار الآتية:

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

sig. .003 .000 .000 .001 .004

MSE=205967.323, RMSE= 453.840, R_a^2=.881, C_p= 7.487, PRESS= 2409.240

ولكن لو رفعنا مستوى معنوية دخول المتغيرات إلى 0.20 فقد نتمكن من كسر حاجز المتغير X_6 والحصول على مجموعة جزئية جديدة لمتغيرات الانحدار قد تكون أفضل من غيرها في تفسير الظاهرة موضوع البحث والتنبؤ بمستقبلها. ولعدم توافر قيم جدولية لاختبار F بمستوى معنوية (مخاطرة) 0.20 لذلك سوف نعتمد على اختبار T ومستوى المعنوية (المخاطرة) لدخول أو بقاء كل متغير في المجموعة وذلك باستخدام برنامج الانحدار الخطي في حزمة برامج SPSS، ونواصل تطبيق الطريقة من المجموعة الجزئية التي توقفت عندها، أي من نهاية الدورة الرابعة ولكن بمستوى معنوية 0.20 وليس 0.10.

جدول رقم (6)

استكمال طريقة step wise في انتقاء متغيرات الانحدار بمستوى معنوية (مخاطرة) 0.20

المتغيرات في نهاية الدورة	قيمة t للمتغيرات الباقية	مستوى المخاطرة في إدخالها	المتغير الداخل	المتغير الخارج	مستوى المخاطرة في بقائه
X_2,X_3,X_4X_6	T_{x1}= 1.095 T_{x5}=1.611 T_{x7}= 0.681 T_{x8}= 0.414	0.287 0.124 0.504 0.683	X_5	لا يوجد	-
X_2,X_3,X_4,X_5,X_6	T_{x1}= 1.336 T_{x7}=0.935 T_{x8}=- 0.271	0.198 0.362 0.789	X_1	X_6	0.288
X_1,X_2,X_3,X_4,X_5	T_{x7}= 0.067 T_{x8}=0.298	0.947 0.769	لا يوجد	لا يوجد	-

وبذلك تتكون لدينا مجموعة جزئية جديدة بعد كسر حاجز المتغير X_6 بتفعيل اعادة اختباره مع المتغير التالي له في الأهمية X_1 وذلك برفع مستوى المعنوية إلى 0.20 وتكون معادلة الانحدار على هذه المجموعة الجديدة هي:

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

sig. .016 .006 .000 .003 .000 .053

MSE=185099.475,RMSE=430.232, R_a^2=.893, $\overline{VIF}$=1.308 C_p=6.198, PRESS= 2328.640

وبالمقارنة مع المجموعة الجزئية المتكونة بنفس الطريقة ولكن بمستوى معنوية 0.10 فان هذه المجموعة الجزئية الجديدة هي الأفضل لأنها تحقق مطابقة أكبر للبيانات ويمكنها التنبؤ بمستقبل الظاهرة موضوع البحث بصورة أدق إضافة إلى صغر تحيز معادلتها.

V- الاستنتاجات والتوصيات conclusions and suggestions

1- الاستنتاجات

أولاً- إن طريقة step wise تفشل في الوصول إلى أفضل مجموعة جزئية لمتغيرات الانحدار في حالة وجود علاقات خطية قوية بين متغيرات الانحدار ولاسيما عندما يكون مستوى المعنوية المستخدم في دخول وبقاء المتغيرات دقيقاً (0. 01 أو 0.05) وبذلك تتساوى في النتيجة النهائية مع طريقة foreward.

ثانياً- إن سبب فشل طريقة stepwise في اختيار المجموعة الجزئية الأفضل لمتغيرات الانحدار عند وجود المتغير الأهم (الحاجز) هو عدم تفعيل إعادة اختبار هذا المتغير والمتغيرات الأخرى في المراحل اللاحقة للمرحلة الأولى من هذه الطريقة، إلا برفع مستوى المعنوية لدخول وبقاء المتغيرات في المجموعة لكي يكون أكبر أو يساوي 0.10 وحسب أهمية المتغيرات التفسيرية وقوة ارتباطها الخطي بالمتغير الحاجز، على أن لا يزيد مستوى المعنوية عن 0.25 لكي لا تصبح المجموعة الجزئية المنتقاة كبيرة وتحتوي على متغيرات ضعيفة المساهمة في تفسير الظاهرة موضوع البحث والتنبؤ بمستقبلها.

2 - التوصيات

أولاً- نوصي باستخدام طريقة step wise عند عدم وجود علاقات خطية قوية ما بين متغيرات الانحدار أما عندما يرتبط المتغير الأهم بعلاقة خطية مع المتغيرات المهمة الأخرى فأنه يمكن استخدام هذه الطريقة ولكن على مرحلتين:

أ- بمستوى معنوية 0.10 لدخول وبقاء المتغير التفسيري في المجموعة.

ب- بمستوى معنوية أعلى (0.20 أو 25.).

ومن ثم المفاضلة ما بين المجموعتين الجزئيتين الناتجتين واختيار أحدهما وفقاً لمعايير المفاضلة المعروفة.

ثانياً- لا نوصي باستخدام طريقة step wise ولا الطرائق الأخرى المعروفة في الاختيار المتسلسل للمتغيرات في حالة وجود علاقات خطية قوية متعددة ما بين كل متغير مهم والمتغيرات المهمة الأخرى وإنما نوصي باستخدام الطريقة والمعيار الذي اقترحه الباحث والذي اسماه (RUF) والمنشور في

مجلة الإدارة والاقتصاد- الجامعة المستنصرية ومجلة العلوم الاقتصادية والإدارية- جامعة بغداد، لأنه يؤدي إلى الحصول على أفضل مجموعة جزئية لمتغيرات الانحدار تحقق الفرضيات القياسية لحد الخطأ في نموذجها الرياضي وتتمتع بتقديرات المربعات الصغرى لمعاملات انحدارها(*).

المصادر References

1- Efroymson, M.A, "Multiple Regression Analysis in mathematical methods for digital computers", Eds. By A. Ralsyon and H.S.WILF, wiley, new york, 1962.

2- عبد الحسين، صباح فرج، "طريقة ومعيار جديدان للكشف عن مشكلة التعدد الخطي ومعالجتها واختيار متغيرات الانحدار" مجلة الإدارة والاقتصاد- تصدر عن كلية الإدارة والاقتصاد / الجامعة المستنصرية، العدد 86، بغداد، 2011.

3- Mantel, N. "why step down procedures in variables selection", Technometrics, 12, 621-625, 1970.

4- Draper, N.R. and smith, H. "Applied Regression Anaysis", 3rd Edition, John wiley and sons, 1999.

5- نفس المصدر السابق ط2، ص320 .

6- Hoerl, A.E. and kennrad, R.W., "Ridge Regression: Biased Estimation for non orthogonal problems", Technometrics, 12, 1970, PP. 55-67.

7- Draper and Smith، مصدر سابق، ص310.

8- 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.

9- Normal Nie, Dale H. Bent and C. Hadlai Hull, "statistical package for the social sciences", new york, Mc Graw-Hill, 1970.

10- Hald A. "Statistical theory with Engineering Applications, wiley, new york, 1952.

11- Bowerman and O'connel، مصدر سابق، ص 855.

(*) لاحظ رجاءً مجلة الإدارة والاقتصاد، العدد 86، السنة 2011، الصفحة 274-243

وكذلك مجلة العلوم الاقتصادية والإدارية، المجلد 18، العدد 66، السنة 2012، الصفحة 386-357.

الملحق

جدول رقم (7)

بيانات مسالة Hald

ob.	Y	x_1	x_2	x_3	x_4
1	78.5	7	26	6	60
2	74.3	1	29	15	52
3	104.3	11	56	8	20
4	87.6	11	31	8	47
5	95.9	7	52	6	33
6	109.2	11	55	9	22
7	102.7	3	71	17	6
8	72.5	1	31	22	44
9	93.1	2	54	18	22
10	115.9	21	47	4	26
11	83.8	1	40	23	34
12	113.3	11	66	9	12
13	109.4	10	68	8	12

جدول رقم (8)

بيانات مسالة Cravens

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.	13.82	21412.7	1971.52	8.48	0.64	89.43	19.3	4.3

	11		9					5	
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.									

……
……
……..