مجلة تنمية الرافدين

تنمية الرافدين ٨٦ (٢٩) ٢٠٠٧
ص ص [١٧١-١٧٩]

مقارنة طريقتي .H.S.L والطريقة التكرارية لتقدير معلمة الحرف

محاسن صالح الطالب
مدرس - قسم الاحصاء
كلية علوم الحاسبات والرياضيات
جامعة الموصل
Mhasn64@yahoo.com

عزه مصطفى القصيمي
مدرس مساعد -قسم نظم معلومات
كلية الادارة والاقتصاد
جامعة الموصل

عبير اسامة الشبخون
م.رئيس ابحاث -وحدة الحاسبة
كلية الادارة والاقتصاد
جامعة الموصل

المستخلص

من المشاكل التي تواجه الباحثين مشكلة تعدد العلاقات الخطية، إذ قد يؤدي وجود هذه العلاقات بين المتغيرات قيد الدراسة الى أن تكون مقدرات معلمات إنموذج الانحدار باستخدام طريقة المربعات الصغرى وهي احدى الطرائق غير المتحيزة غير مقبولة، ولذلك يجب اللجوء الى إحدى الطرائق المتحيزة ومنها طريقة إنحدار الحرف.

في هذه الدراسة اعتمدنا طريقة إنحدار الحرف وتم اختيار طريقتين من طرائق تقدير معلمة الحرف (K)، وهما الطريقة التكرارية (Iteration method)، وطريقة (.H.S.L) (Hocking, Speed and Lynn)، وتمت المقارنة لاختيار أفضل طريقة باعتماد معيار أقل متوسط مربعات الخطأ الكلي.

Comparison Between Two Methods, H.S.L. and The Iteration Method to Estimate Parameter of Ridge

Mhasen Salh Al-talb
Assistant lecturer Coll. of comp. scie. & math Dept. of Statistic

Azza Mustfa Al-qusaymy
Lecturer- Coll. of Admi. & Econ. Dept. of Informatic system

Abeer Osama Al-shabkon
Assistant Dean of scientific Research- unit of computer- Coll. of Admi. & Econ.

Abstract

The multi - linearity is one of the problems that the researchers are avoided. In this problem, the least squares estimates will be inadmissible Therefore, we used the Ridge Regression which has a biased estimates and has a more efficient estimates than the least squares.

In this paper we used the Ridge Regression and we used two methods for the estimation of Ridge factor (K) , the first , is the iterative method , and the second , is Hocking , Speed and lynn (H. S. L.) .The comparison is made between the two methods by using the total mean square error criterion.

تاريخ التسلم ٢٠٠٦/٥/١٤ ________________ تاريخ قبول النشر ٢٠٠٦/٩/٢٤

المقدمة

يعد إنموذج الإنحدار من النماذج الاحصائية المهمة والتي يمكن أن تطبق على بعض الظواهر بعد دراستها ، اذيعتمد نجاح هذا الإنموذج أو فشله على الطريقة المستخدمة لتقدير معلماته ، فضلاً عن على صفات هذه المعلمات التي سوف يتم تقديرها وهناك عدة طرائق لتقدير معلمات إنموذج الانحدار، منها طريقة المربعات الصغرى التي تعد من الطرائق المهمة والسهلة الاستخدام ، وتعتمد هذه الطريقة على عدد من الفروض التي تتعلق بالمتغير العشوائي وطبيعة المتغيرات المستقلة، واذا سقط أحد فروضها الرئيسة و أهمها عدم وجود تداخل خطي بين المتغيرات المستقلة يؤدي ذلك الى إعطاء تقديرات غير دقيقة ، من هنا يتم اللجوء الى طرائق التقدير المتحيزة، فعلى الرغم من تحيزها إلا أنها تعطي تقديرات أفضل من تقديرات طريقة المربعات الصغرى، ومن هذه الطرائق طريقة إنحدار الحرف Ridge Regression method ، وهذا ما أكده (Gunst and Mason and 1975 ,4(3) Webster,) اذ نبهوا على خطورة محاولة التنبؤ في حالة وجود تداخل خطي باستخدام طريقة المربعات الصغرى، وأكدوا على وجوب استخدام الطرائق المتحيزة بديلاً لهذه الطريقة ، وركزوا على طريقة إنحدار الحرف بوصفها الطريقة المثلى.

تعتمد طريقة إنحدار الحرف على اضافة ثابت يسمى بمعلمة الحرف ، وهناك عدة طرائق لتقدير هذا الثابت منها الطريقة التكرارية (Iiteration method) التي تعد احدى الطرائق المهمة والمثلى (القصيمي، ٢٠٠٠، ٣٤) ،وطريقة Hocking, Speed and Lynn اللتين سيتم تطبيقهما في هذه الدراسة لمقارنة نتائجهما واستنتاج أيهما أفضل.

هدف الدراسة

تهدف الدراسة الى مقارنة طريقتين من طرائق تقدير معلمة الحرف (K) ومقارنة مقدرات هاتين الطريقتين باعتماد معيار متوسط مربع الخطأ الكلي ، ومن ثم اختيار الطريقة الأفضل التي تعطي أقل متوسط مربع خطأ كلي.

أولاً - الجانب النظري

يتمثل إنموذج الانحدار الخطي بالمعادلة

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + \ \dots + \hat{\beta}_p X_{ip} \qquad \dots(1-1)$$

اذ إن

$\hat{Y}_i$: المتغير المعتمد.

X_{ij} :المتغيرات المستقلة.

$\hat{\beta}_j$:معلمات الإنموذج .

i :عدد المشاهدات (حيث i =1,2,…,n)

j :عدد المتغيرات المستقلة (حيث $j = 1,2, \dots ,p$)

١. طريقة المربعات الصغرى

تعد طريقة المربعات الصغرى من الطرائق المهمة والأكثر شيوعاً لتقدير معلمات إنموذج الانحدار الخطي،وتتميز بسهولة حساب تقدير المعلمات ، اذ تمتلك أقل التباينات مقارنة بباقي الطرائق غير المتحيزة. والصيغة العامة لايجاد تقديرات معلمات إنموذج الانحدار الخطي بهذه الطريقة هي:

$$\hat{\underline{\beta}}_{ls} = \left(X'X\right)^{-1} X'Y \qquad \dots (1-2)$$

٢. طريقة إنحدار الحرف

وهي طريقة معدلة لطريقة المربعات الصغرى في حالة وجود تداخل خطي بين المتغيرات المستقلة ، وتعدمن الطرائق المتحيزة ، اذ تعتمد على إضافة المصفوفة (KI_p) الى المصفوفة ($X'X$) قبل اخذ المعكوس لها وكالآتي:

$$\hat{\underline{\beta}} = \left(X'X + K_{Ip}\right)^{-1} X'Y \qquad \dots (1-3)$$

اذ إن:

KI_p : مصفوفة قطرية ذات بعد (pxp).

١. معلمة الحرف

يمكن أن يعرف إنحدار الحرف على أنه دالة لمعلمة الحرف (K)، واختيار قيمة مناسبة لـ (K) سوف يؤدي الى إعطاء تقديرات اكثر دقة من تقديرات المربعات الصغرى، وذلك لان إضافة الثابت (K) يقود الى تقليل عناصر قطر المصفوفة $(X'X)^{-1}$ ومن ثم تقليل قيم المقدرات والتباينات المتضخمة ، كما يتم التوصل الى متوسط مربعات الخطأ الكلي بقيمة أقل من قيمته لطريقة المربعات الصغرى.

٢. طرائق تقدير معلمة الحرف

توجد عدة طرائق لتقدير معلمة الحرف (K) منها الطريقة التكرارية التي تعد من الطرائق الناجحة لاختيار معلمة الحرف كما ذكرنا سابقاً ، إذ انها تتمتع بـ صغر متوسط مربع الخطأ مقارنة بالطرائق الأخرى (القصيمي، ٢٠٠٠، ٣٤).

أ. الطريقة التكرارية Iteration method

يتم في هذه الطريقة إعتماد قيم لمعلمة الحرف (K) وتنحصر هذه القيم ضمن الفترة (0,1) وتعويضها من أجل الحصول على عدة قيم لمقدرات الحرف ، ومن ثم تطرح كل قيمة سابقة من مقدر الحرف من القيمة اللاحقة للمقدر ، وبإفتراض خطأ معين يمكن السماح به يتم اختيار أحد المقدرات، ومن ثم يعد هذا المقدر هو الافضل، وقيمة (K) التي استخدمت لايجاد هذا المقدر هي القيمة المثلى.

ب. طريقة (.H. S. L) (Hocking , Speed and Lynn, 1989, 131-148)

يرمز لهذه الطريقة بـ (.H.S.L)، اذ عرضت في سنة ١٩٧٦، ويتم حساب قيمة معلمة الحرف (K) باستخدام الصيغة الاتية:

$$K = \sigma^2 \frac{\sum_{j=1}^{p} \lambda_j^2 \hat{\beta}_j^2}{\sum_{j=1}^{p} \lambda_j^2 \hat{\beta}_j^4} \qquad \ldots(1-4)$$

اذ إن:

λ_j: تمثل القيم المميزة لمصفوفة الارتباط لقيم X .

$\hat{\beta}_j$: قيمة المعلمات المقدرة بطريقة المربعات الصغرى.

Σ^2: التباين والذي يحسب بالصيغة:

$$\sigma^2 = \frac{RSS}{n-p}$$

$$= \frac{\underline{Y}'\underline{Y} - \hat{\underline{\beta}}'\underline{X}'\underline{X}}{n-p} \qquad \ldots(1-5)$$

وإن

$\underline{Y}'\underline{Y} = 1$ وذلك لأن البيانات قياسية.

$\underline{\hat{\beta}}$: متجه المعلمات المقدرة بطريقة المربعات الصغرى.

١. متوسط مربعات الخطأ الكلي (TMSE) Total mean square error

يعد هذا المعيار من المعايير التي يتم عن طريقها مقارنة المقدرات التي يتم الحصول عليها بأكثر من طريقة، والذي يحتسب على وفق الصيغة :

$$TMSE = \sum_{j=1}^{p} MSE\left(\hat{\beta}_j(K)\right) \qquad \ldots(1-6)$$

اذ إن $\left(MSE\left(\hat{\beta}_j(K)\right)\right)$ يحتسب من الصيغة :

$$MSE\left(\hat{\beta}_j(K)\right) = \frac{1}{L}\sum_{j=1}^{p}\left(\hat{\beta}_{ji}(K) - \lambda_j\right)^2 \qquad \ldots(1-7)$$

$$i = 1,2,\ldots,L$$

وإن:

L: عدد العينات .

$\hat{\beta}_{ji}(K)$: مقدر المعلمة (j) للعينة (i) .

λ_j : القيمة المميزة لمصفوفة الارتباط $(X'X)$.

ثانياً - الجانب التطبيقي

١. توصيف الإنموذج ومحاكاته

أ. توصيف الإنموذج

تم تحديد عدد المتغيرات المستقلة بخمسة متغيرات هـــي $(X_1, X_2, X_3, X_4, X_5)$ وعليه فإن الإنموذج يمثل المعادلة:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + e_i \qquad \dots(2\text{-}1)$$

اذ إن :

e_i :الخطأ العشوائي ويكون ذا توزيع طبيعي بوسط (0) وتباين (σ^2).

X_i و e_i تولد باستخدام أحد أساليب المحاكاة، وهـــو أســـلوب Monte carlo simulation ($\beta_1 , \dots , \beta_5$) تحسب بالتطابق مع قيم الجـــذور المميـــزة لمـــصفوفة إرتباط متغيرات التنبؤ (المستقلة)، وβ_0 تساوي الصفر، ومن ثـــم تحـــول البيانـــات جميعها الى الصيغة القياسية ، وذلك باستخدام إحدى صيغ التحويل ، وهـــي صـــيغة الثابت بطول واحد unit length scaling والتي تتمثل بالمعادلة الاتية:

$$X^*_{ij} = \frac{X_{ij} - \bar{X}_j}{\sqrt{\sum_{i=1}^{n} (X_{ij} - \bar{X}_j)^2}} \qquad \dots(2\text{-}2)$$

وعليه يكون الإنموذج بالشكل الاتي:

$$\underline{Y} = X^* \underline{\beta} + \underline{e} \qquad \dots(2\text{-}3)$$

اذ إن

X^* : مصفوفة المتغيرات المستقلة المعيارية .

$\underline{\beta}$: متجه المعلمات .

$\underline{e}$: متجه الأخطاء .

$\underline{Y}$: متجه الاستجابة بالصيغة المعيارية .

كما استخدم التحويل $X^{**} = X^* B$ ، اذ تمثل B مـــصفوفة المتجهـــات المميـــزة لمصفوفة ارتباطات المتغيرات المستقلة ذات بعد (5 x 5) وكذلك $D = B' X^{*'} X^* B$ اذ إن D هي مصفوفة قطرية عناصر قطرها هي القيم المميزة لمـــصفوفة إرتبـــاط المتغيرات المستقلة ذات بعد (5 x 5)، كما يمكـــن حـــسابها ايـــضاً عـــن طريـــق $D = X^{**'} X^{**}$ ، وبالرجوع الى المعادلة (2-3) فان $\underline{\beta}$ يمثل متجه المعلمات ذات بعد (5 x 1)، و $\underline{e}$ متجه الاخطاء ذا بعد (100 x 1)، و $\underline{Y}$ متجه الاستجابة ذا بعد (100 x 1) أيضاً.

ب. المحاكاة

تم استخدام أسلوب المحاكاة المذكور لتوليد قيم X_{ij} والـــذي يتمثـــل بالمعادلـــة الاتية:

$$X_{ij} = (1-\alpha^2)^{(1/2)} Z_{ij} + \alpha\, Z_{i6} \qquad \dots(2\text{-}4)$$

$i = 1, 2, \dots, n \;,\; j = 1, 2, \dots, 5$

اذ إن :

α : قيمة الارتباط بين المتغيرات المستقلة .

$Z_{i1}, \dots, Z_{i6}$:متغيرات مستقلة عشوائية طبيعية قياسية $N(0,1)$.

وتحسب قيم e_i عن طريق المعادلة

$$e_i = Z_{i6} - \bar{Z} \qquad \dots(2\text{-}5)$$

$\bar{Z}$: يمثل متوسط العمود (Z_{i6}) .

٢. التطبيق العملي (المعاينة)

باستخدام البرمجية الجاهزة Minitab under windows الاصدار 13.20 تم توليد (٢٠) عينة لكل عينة (٦٠٠) مشاهدة، تتبع التوزيع الطبيعي القياسي $(Z_{ij} \quad N(0,1))$ موزعة على ستة أعمدة كل عمود (١٠٠) مشاهدة، ومن هذه المشاهدات تحسب قيم (X_{ij}) حسب الصيغة (2-4)وبارتباط قدره $(\alpha = 0.9)$. ومن ثم حساب الخطأ (e_i)من تطبيق الصيغة (2-5) وبالتالي ايجاد قيم (Y_i)، وبعد ذلك تم تحويل (X_{ij} , Y_i) الى الصيغة القياسية.

٣. تقدير معلمة الحرف (K)

أ. الطريقة التكرارية

تم الاعتماد على عدة قيم لـ (K) ضمن الفترة (0,1) وبذلك تم الحصول على عدد من القيم لمقدرات الحرف ، وبطرح كل قيمة سابقة لهذه المقدرات من القيمة اللاحقة للمقدر وبخطأ مسموح به قدر بـ (0.0001)،اخترنا أفضل المقدرات ، ومن ثم تم التوصل الى أفضل قيمة لـ (K) التي تعد القيمة المثلى، بعدها تم حساب معدل معلمة الحرف، وتم ذلك باستخدام البرمجية الجاهزة Statgraph، والجدول ١ يوضح هذه النتائج :

الجدول ١

قيم المعلمة (K) للطريقة التكرارية

رقم العينة	المعلمة K	رقم العينة	المعلمة K
1	0.75	11	0.90
2	0.90	12	0.85
3	0.90	13	0.90
4	0.85	14	0.90
5	0.80	15	0.85
6	0.90	16	0.75
7	0.80	17	0.90
8	0.75	18	0.85
9	0.75	19	0.85
10	0.95	20	0.90
معدل قيمة K	0.85		

المصدر: نتائج مستخرجة

ب. طريقة (.H. S. L)

تم حساب معلمة الحرف (K) من المعادلة (1-4) باستخدام برنامج احصائي ضمن إمكانيات فقرة الـ (Macro) ضمن البرمجية الجاهزة (Minitab under windows) والجدول ٢ يمثل نتائج تطبيق هذه الطريقة :

الجدول ٢

قيم المعلمة (K) لطريقة (.H. S. L)

رقم العينة	المعلمة K	رقم العينة	المعلمة K
1	0.000077	11	0.000064
2	0.000052	12	0.000077
3	0.000069	13	0.000066
4	0.000078	14	0.000056
5	0.000088	15	0.000066
6	0.000063	16	0.000072
7	0.000063	17	0.000053
8	0.000059	18	0.000072
9	0.000100	19	0.000061
10	0.000077	20	0.000057
معدل قيمة K	0.000069		

المصدر: نتائج مستخرجة

٤ . تقدير المعلمات

تم تقدير المعلمات بطريقة انحدار الحرف على وفق المعادلة الاتية:

$$\underline{\hat{\beta}}(K) = (D + KI_p)^{-1} \underline{X}^{*\prime} \underline{Y} \qquad \ldots(2\text{-}6)$$

اذ إن

D : مصفوفة قطرية تمثل عناصرها القيم المميزة لمصفوفة الارتباط بين المتغيرات المستقلة(مذكورة سابقاً) .

ولجميع قيم (K) لكل العينات ومن ثم أخذ المعدل لهذه المعلمات ولكلا الطريقتين، وكما موضح في الجدولين ٣ للطريقة التكرارية، وجدول ٤ لطريقة (.H. S. L):

الجدول ٣
معلمات الطريقة التكرارية

$\beta_5(K)$	$\beta_4(K)$	$\beta_3(K)$	$\beta_2(K)$	$\beta_1(K)$	العينة
-0.158062	0.597339	0.031786	-0.002088	-0.470172	1
-0.301060	0.082263	0.119961	0.551486	-0.459498	2
0.160196	-0.247716	0.534075	-0.026891	-0.457522	3
0.404752	-0.231881	0.048146	0.263898	-0.471478	4
0.057340	0.557467	-0.067739	-0.066549	-0.476757	5
0.582344	-0.004698	-0.269427	0.154692	-0.467948	6
-0.051634	-0.133051	0.044342	0.605089	-0.461527	7
0.452995	-0.042004	0.340497	-0.355301	-0.453811	8
0.023104	0.405461	0.210031	-0.321343	-0.453281	9
-0.089101	0.414232	-0.110384	0.182289	-0.457290	10
-0.169069	-0.154277	0.285917	0.453740	-0.460802	11
0.491124	0.356615	-0.145718	-0.264582	-0.464533	12
0.451820	0.346302	-0.040513	-0.292587	-0.467337	13
0.198701	0.055782	0.533375	-0.335323	-0.461387	14
0.030508	0.402135	0.270531	-0.303643	-0.455181	15
0.231947	0.080074	0.189186	0.017973	-0.476551	16
0.023875	0.175042	0.544470	-0.312656	-0.461367	17
-0.107911	0.500583	-0.004368	0.007833	-0.458601	18
0.305852	-0.255980	0.451298	-0.030953	-0.464487	19
0.047557	-0.310596	0.485343	0.212848	-0.461843	20
0.129260	0.129650	0.172540	0.0068966	-0.46307	المعدل

المصدر: نتائج مستخرجة

الجدول ٤
معلمات طريقة .H. S. L

$\beta_5(K)$	$\beta_4(K)$	$\beta_3(K)$	$\beta_2(K)$	$\beta_1(K)$	العينة
-0.163792	0.592840	0.027786	-0.005758	-0.470387	1
-0.353155	0.035715	0.086218	0.521828	-0.460909	2
-0.116366	-0.480127	0.354680	-0.170381	-0.465424	3
0.501057	-0.167576	0.104780	0.313064	-0.468809	4
0.074726	0.570830	-0.055411	-0.056145	-0.476019	5
0.545409	-0.032497	-0.291085	0.135154	-0.469007	6
-0.031046	-0.116638	0.057384	0.617402	-0.460931	7
0.131167	-0.289168	0.151317	-0.525313	-0.464065	8
0.283552	0.304643	-0.421587	-0.139976	-0.468772	9
-0.481446	0.105507	-0.378178	-0.041175	-0.470895	10
-0.429789	-0.348078	0.108559	0.305522	-0.470449	11
0.344423	0.229595	-0.249272	-0.354090	-0.470029	12
0.437727	0.333657	-0.052507	-0.299939	-0.467828	13
0.138121	0.005512	0.491491	-0.367490	-0.463249	14
-0.322325	0.149671	0.038137	-0.494370	-0.466352	15
0.424979	0.257509	0.353762	0.147683	-0.468978	16
-0.138542	0.043950	0.441044	-0.398896	-0.467278	17
-0.467909	0.232868	-0.242640	-0.214227	-0.471242	18
0.341935	-0.223648	0.477179	-0.007361	-0.463437	19
-0.153423	-0.462025	0.347083	0.119777	-0.467433	20
0.028265	0.037127	0.067437	-0.045735	-0.46757	المعدل

المصدر: نتائج مستخرجة

٥. مقارنة المقدرات

استخدم معيار متوسط مربعات الخطأ الـ كلي في مقارنة المقدرات التي حصلنا عليها من الطريقتين المستخدمتين لانحدار الحرف لتمييز الطريقة الافضل في التقدير وحسب المعادلة (6-1) والنتائج مبينة بالجدول ٥ .

الجدول ٥
متوسط مربعات الخطأ الكلي للطرائق المستخدمة

الطريقة	TMSE
التكرارية	22.4287
H. S. L.	22.6222

الاستنتاجات

من ملاحظة الجدول ٥ يتبين أن قيمة TMSE للطريقة التكرارية أقل من طريقة .H.S.L ، وهذا يدل على أن الطريقة التكرارية هي الأفضل.

المراجع

اولاً - المراجع باللغة العربية

١. عزه مصطفى القصيمي"استخدام اسلوب المحاكاة في مقارنة مقدرات انحدار الحرف"، رسالة ماجستير (غير منشورة)، مقدمة الى كلية الادارة والاقتصاد، جامعة الموصل،٢٠٠٠.

ثانياً - المراجع باللغة الاجنبية

1. R. F. Gunst , and R. L. Moson and J. F. Webster,"Regression Analysis and Problems of Multicollinearity", Communications in Statistics, 1975.
2. R. R. Hocking , F. M. Speed and M. J. Lynn,"On Identification of Transfer Function Models by Regression Methods", J.Statist. Simul. , 1989.