الحصانة في مقدرات معاملات الارتباط القويم

المدرس طاهر ريسان دخيل   المدرس علي جواد كاظم
جامعة القادسية/ كلية الإدارة والاقتصاد

الخلاصة

إن تحليل الارتباط القويم يدرس الارتباطات بين مجموعتين من المتغيرات العشوائية ، حيث يعد هذا التحليل من التحليلات التقليدية والتي تعمل بتحقق افتراضات معينة.إن خرق هذه الافتراضات يؤدي إلى إن يكون أداء الطرق التقليدية غير كفوء،ومن هذه الافتراضات عدم تحقق التوزيع الطبيعي وذلك بسبب وجود قيم شاذة في البيانات ،إن عدم تحقق هذا الشرط يدعونا للبحث عن طرق اقل حساسية للخروقات في تلك الافتراضات وهذا ما يدعى بالطرق الحصينة.

في هذا البحث تم استخدام طريقة الانحدار البديل الحصين **Robust Alternative Regression** بمقدرات **L** ومقدرات **M** للحصول على تقديرات الارتباط القويم ولقد تم استخدام المحاكاة لغرض إجراء المقارنة بين مقدرات الارتباط القويم المحسوبة باستخدام طريقة الانحدار البديل الحصين **Robust Alternative Regression** التي تكون فيها المعاملات القويمة مقدرة على وفق طريقة المربعات الصغرى مع تلك المعاملات المقدرة على وفق طريقة المربعات الموزونة باستخدام دالتي وزن **Huber** و **Hample** وذلك بهدف معرفة أفضلية المقدرات وذلك بالاعتماد على معيار متوسط مربعات الأخطاء **MSE** للحكم على المقدر الأفضل.

المقدمة

إن هدف تحليل الارتباط القويم هو التشخيص والقياس الكمي للعلاقات بين متغير عشوائي يتكون من **p** من الأبعاد هو **X** ومتغير يتكون من **q** من الأبعاد هو **Y** ، هدفنا هو الحصول عل تركيبات خطيه $a^T X$ ، $b^T Y$ من المتغيرات ألأصلية تمتلك أعظم تباين ، ويمكن التعبير عن ذلك بصيغة رياضية بالشكل التالي:

$$(\alpha,\beta) = MAX\left|corr(a^T X, b^T Y)\right| \quad ...(1)$$

إن المتغيرات الناتجة $U=a^TX$ ، $V=b^TY$ تدعى المتغيرات القويمة ونلاحظ إن المتجهات α و β محددة بثابت كما في المعادلة رقم (١) وحتى تأخذ كل من α و β قيمه وحيده لابد من إضافة قيد جعلها طبيعية **(Normalization)** لكل من المتغير **U,V** حتى يمتلك كل منهما تباين مقداره واحد أو بصوره أخرى حتى يمتلك α و β مستوى معياري مقداره واحد.

إن الارتباط القويم الأول ρ يعرف كقيمه مطلقه للارتباط بين مجموعتين من المتغيرات القويمة وكما في العلاقة رقم (1) ، إن المتغير القويم ذو الرتبة k ، $1 < k < Min(p,q)$ سيكون غير مرتبط مع كل المتغيرات القويمة ذات الرتب الدنيا .

وبفرض إن Σ تمثل مصفوفة التباين المشترك للمجتمع للمتغير العشوائي Z حيث إن $Z = (X^T, Y^T)^T$ ويمكن وضع Σ بالشكل التالي :

$$\Sigma = \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix}$$

إن المتجهات α و β تمثل متجهات مميزه مترافقة مع اكبر قيمه مميزه للمصفوفات :

$$..\Sigma_{XX}^{-1}\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\,,\Sigma_{YY}^{-1}\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\ \dots\ (2)$$

إن كلا المصفوفتين أعلاه تمتلك نفس القيم المميزة الموجبة واكبر تلك القيم يمثل الارتباط القويم الأول ، ولتقدير α و β و ρ لابد من حساب مصفوفة التباين المشترك للعينة $\hat{\Sigma}$ وذلك من خلال العينة $Z_1,Z_2,..,Z_n$ حيث إن $Z_i = (X_i^T, Y_i^T)^T \in IR^P * IR^q$.

إن حساب المتجهات المميزة والقيم المميزة المحسوبة في المصفوفات في العلاقة (2) يعطي تقديرات للمتغيرات القويمة والارتباطات ، إن التقدير التقليدي لمصفوفة التباين المشترك غير حصين بالنسبة للمشاهدات الشاذة ولذلك فان القيم المميزة والمتجهات المميزة والتي تعتمد على تقديرات التباين المشترك سوف تكون حساسة جدا لهذه المشاهدات الشاذة .

طريقة الانحدار البديل الحصين

تم اقتراح أسلوب الانحدار البديل Alternative Regression من قبل Wold سنة 1966 .إن هذه الأسلوب اخذ اهتماما كبيرا في السنوات الأخيرة . حيث يمكن استخدام هذا الأسلوب لتقدير المعاملات القويمة وكما يأتي :

$$\beta = MAX\left|corr(\alpha^T X, b^T Y)\right|$$

حيث ان β هي نسبة لمعامل الانحدار b في نموذج الانحدار التالي :

$$...\alpha^T X = b^T Y + \gamma_1 + \varepsilon_1 \dots\dots\dots\dots\ (3)$$

وبنفس الطريقة فان α تساوي المعلمة a في نموذج الانحدار التالي:

$$..b^T Y = a^T X + \gamma_2 + \varepsilon_2 \dots\dots\dots\ (4)$$

إن قيمة α_0 الأولية يمكن الحصول عليها باستخدام أسلوب تحليل المركبات الأساسية الحصينة لمصفوفة البيانات المكونة من $(x_1,x_2,..,x_n)$.

وعلى أساس المعادلة رقم(3) يمكن الحصول على قيمة β1 الأولى من خلال انحدار المتغير $\alpha_0^T X$ على متغيرات Y وبعد ذلك يتم استخدام المعادلة رقم (4) ، أما قيمة α_1 الجديدة فيتم الحصول عليها بانحدار $\beta_1^T Y$ على المتغير X ، نستمر بهذا الأسلوب التكراري حتى يتم الحصول على التقارب المطلوب ، وبعد ذلك يتم حساب الارتباط القويم ρ بين المتغيرات.

لقد تم الاعتماد في هذا البحث على أيجاد المقدرات في المعادلتين (٣) و(٤) باستخدام طريقتين للتقدير هما طريقة المربعات الصغرى الاعتيادية **O.L.S** وطريقة المربعات الصغرى الموزونة **W.L.S** عندما تكون هناك قيم شاذة وعند عدم وجود قيم شاذة علما إن الأوزان كانت تمثل دوال وزن من نوع مقدرات **M** وتم الاعتماد على دالة وزن **Huber** وحسب صيغتها المذكورة أدناه:

$$\psi(x) = \begin{cases} x;|x| < h \\ h * \operatorname{sgn}(x);|x| \geq h \end{cases}$$

حيث إن **h** تأخذ القيم التالية:**2 , 1.5**
وان **sgn** تمثل إشارة **x** .
ودالة وزن **Hample** وحسب صيغتها المذكورة في أدناه:

$$\psi(x) = -\psi(-x) = \begin{cases} x;0 \leq x \leq a \\ a;a \leq x < b \\ \dfrac{c-x}{c-b}a, b \leq x < c \\ 0, x \geq c \end{cases}$$

وتمت المقارنة بين النتائج على أساس معيار متوسط مربعات الأخطاء **MSE** لمقدرات الارتباط القويم الأول .

المحاكاة

لقد تم استخدام المحاكاة لغرض إجراء المقارنة بين مقدرات الارتباط القويم المحسوبة باستخدام طريقة الانحدار البديل التي تكون فيها المعاملات القويمة مقدرة وفق طريقة المربعات الصغرى مع تلك المعاملات المقدرة وفق طريقة المربعات الموزونة باستخدام دالتي وزن **Huber** و **Hample** وذلك بهدف معرفة أفضلية المقدرات وذلك بالاعتماد على معيار متوسط مربعات الأخطاء (**MSE**) للحكم على المقدر الأفضل حيث تم كتابة برنامج باستخدام لغة (**V.B**) خاص بالتجربة حيث تم تكرارها (**1000**) مرة لغرض الوصول آلى نتائج مُقنعة حيث يمكن أدراج خطوات ، أجراء المحاكاة وكالتالي :

١-نولد قيم لثلاث متغيرات **(x1,x2,x3)** لتكوين تركيبة خطية هي**Z** وبأحجام عينات **(n=20,50,80,100,200)** أي لكل متغير من متغيرات **x** نولد **n** من المشاهدات المذكورة وبتوزيع طبيعي ذو متوسط صفر وتباين واحد ،ونولد قيم للمتغير ε_1 مرة بالتوزيع الطبيعي بالمتوسط صفر وتباين واحد ومرة أخرى يتم تلويثه بمشاهدات من توزيعات اخرى وبنسب تلوث **0.1** و **0.2** .

٢-يتم توليد متغيرات**(y1,y2,y3)**لتكوين تركيبة خطية هي**W**وبأحجام عينات **(n=20,50,80 ,100,200)** أي لكل متغير من متغيرات **y** نولد **n** من المشاهدات المذكورة وبتوزيع طبيعي ذو متوسط صفر وتباين واحد ،ونولد قيم للمتغير ε_2 مرة بالتوزيع الطبيعي بالمتوسط صفر وتباين واحد ومرة أخرى يتم تلويثه بمشاهدات من توزيعات أخرى وكما مبينة في الجداول للنتائج المستحصلة وبنسب تلوث **0.1** و **0.2** .

٣- يتم حساب معامل الارتباط بين **S,T** والتي تمثل التراكيب الخطية **Z** و **W** بعد تقدير المعاملات **ai,bi** مرة وفق طريقة المربعات الصغرى ومرة أخرى وفق طريقة المربعات الصغرى الموزونة باستخدام دالة وزن **Huber** ودالة وزن **Hample** .

٤- إن معامل الارتباط المحسوب هو معامل الارتباط القويم .

تحليل النتائج :

بعد أجراء تجربة المحاكاة تم الحصول على نتائج موضحة في الجداول أدناه حيث نلاحظ من خلال النتائج في الجدول (١) والذي يمثل قيم **MSE** عند عدم وجود تلوث في البيانات بان طريقة المربعات الصغرى الاعتيادية **OLS** أعطت أفضل النتائج وذلك لان لها اقل **MSE** ويأتي بعدها طريقة المربعات الصغرى الموزونة **WLS** عند استخدام دالة **Huber** أما أسوء النتائج فقد أعطتها طريقة المربعات الصغرى الموزونة **WLS** عند استخدام دالة **Hample** .

الجدول (١):يمثل قيم MSE عند عدم وجود تلوث في البيانات

الطريقة		20	50	80	100	200
WLS	Huber	0.0925	0.0551	0.0612	0.0479	0.0412
	Hample	1.5119	0.6546	0.3682	0.1562	0.2775
OLS		0.09٠1	0.0525	0.0604	0.04٥0	0.04٠9

نلاحظ من خلال الجدول (٢) والذي يمثل قيم **MSE** عند وجود ثلوث في البيانات بنسبة ١٠% و ٢٠% ببيانات من التوزيع المنتظم المتقطع بان طريقة المربعات الصغرى الموزونة **WLS** عند استخدام دالة **Huber** هي الأفضل بامتلاكها اقل **MSE** وتأتي بعدها طريقة المربعات الصغرى الاعتيادية **OLS** وأسوء النتائج أعطتها طريقة المربعات الصغرى الموزونة **WLS** عند استخدام دالة **Hample** ولكن أدائها يتحسن بزيادة حجم العينة.

الجدول (2):يمثل قيم MSE بوجود تلوث في البيانات لكلتا المجموعتين بتوزيع من نوع منتظم متقطع

الطريقة		20		50		80		100		200	
نسبة التلوث		0.1	0.2	0.1	0.2	0.1	0.2	0.1	0.2	0.1	0.2
WLS	Huber	0.0972	0.1022	٠.٠٥٨٧	٠.٠٤٧٨	٠.٠٣٣٧	٠.٠٣٠٧	٠.٠٤٦	٠.٠٢٣٢	٠.٠١٨٥	٠.٠١٤٧
	Hample	١.٦٣٧٧٣	١.١١٩٨	٠.١١٩٨	٠.١١٧٥	٠.٠٥٤٢	٠.٠٦٨٢	٠.٠٣٢٥	٠.٠٥٠٨	٠.٠١٣٢	٠.٠١٧٦
OLS		٠.٠٩٥٩	٠.١١٠٤	٠.٠٦٠٧	٠.٠٤٩٥	٠.٠٣٥٦	٠.٠٣١	٠.٠٤٨٨	٠.٠٢٤٦	٠.٠١٩٩	٠.٠١٥١

ويتبين من خلال النتائج في الجدول (٣) والذي يبين قيم **MSE** عند وجود ثلوث في البيانات بنسبة ١٠% و ٢٠% ببيانات من التوزيع المنتظم المتقطع وتوزيع بواسون بان طريقة المربعات الصغرى الموزونة **WLS** عند استخدام دالة **Huber** تكون أفضل وخاصة عندما يزداد حجم العينة وتأتي بعدها طريقة المربعات الصغرى الاعتيادية **OLS** وبالمرتبة الأخيرة طريقة المربعات

الصغرى الموزونة WLS عند استخدام دالة **Hample** ولكن أدائها يتحسن بزيادة حجم العينة.

الجدول (٣):يمثل قيم MSE بوجود تلوث في البيانات للمجموعة الأولى بتوزيع من نوع بواسون والمجموعة الثانية بتوزيع من نوع منتظم متقطع

الطريقة		20		50		80		100		200	
نسبة التلوث		0.1	0.2	0.1	0.2	0.1	0.2	0.1	0.2	0.1	0.2
WLS	**Huber**	٠.٨١٨٥	٠.٢٥٦٩	٠.١٩٦	٠.٢٥٩٦	٠.٢٥٤٢	٠.٣٩٢٦	٠.٢٩٨	٠.٦١٠٥	٠.٦٢٣٣	٠.٦٠٦٩
	Hample	٢.٢٩٤	٢.٣٧٧٧	٠.٣١٠١	٠.٢٢٣٤	٠.١٨٤٥	٠.٢٥٥٤	٠.١٦٦٥	٠.٢٩٩٦	٠.٣٠٢٢	٠.٢٠٦
OLS		٠.٢٣٢١	٠.٢٥٧١	٠.٢١٥٣	٠.٢٧٠٧	٠.٢٨٣٦	٠.٣٩١٤	٠.٣١٢٨	٠.٦٣٥٩	٠.٧٤٦٦	٠.٦١٢٣

الاستنتاجات

١- نستنتج مما سبق بان طريقة المربعات الصغرى الموزونة WLS عند استخدام دالة **Huber** تعطي نتائج أفضل عندما لا يكون هناك تلوث في البيانات وتأتي طريقة المربعات الصغرى الاعتيادية **OLS** بالمرتبة الثانية بينما كانت طريقة المربعات الصغرى الموزونة WLS عند استخدام دالة **Hample** بالمرتبة الأخيرة.

٢- عندما يكون هناك تلوث في البيانات في المجموعتين بتوزيع منتظم متقطع فان طريقة المربعات الصغرى الموزونة WLS عند استخدام دالة **Huber** تعطي نتائج أفضل.

٣- بوجود تلوث في البيانات للمجموعة الأولى بتوزيع من نوع بواسون والمجموعة الثانية بتوزيع من نوع منتظم متقطع فان طريقة المربعات الصغرى الموزونة WLS عند استخدام دالة **Huber** تعطي نتائج أفضل.

وبالتالي نستنتج بان طريقة المربعات الصغرى الموزونة WLS عند استخدام دالة **Huber** تعطي نتائج أفضل في جميع الحالات.

المصادر

1. Dehon,C. Filzmoser, P. and Croux,C. ((2003)"Robust methods for canonical correlation analysis", Journal of multivariate analysis , vol. 84 ,p. 61-67.
2. Filzmoser, P. ((1999)"Robust principle components and factor analysis in the geostatistical treatment of environmental data, environmetrics ,10, 363-375.
3. Huber, P. ((1981)), "robust statistics", John Wiley and sons, New York.
4. Romanazzi,M.((1992)),"Influence in canonical correlation analysis", Psychometrika, 57, 237-259.
5. Staudte, R. G. ((1989)),"robust estimation &testing" ,Wiley inter science New York.
6. Wilcox ,R. R. ((1997)), "introduction to robust estimation &hypothesis testing" Academic Press , New York