

# مقارنة بعض الطرائق الحصينة في تحليل الارتباط القويم

## بإستخدام المحاكاة مع تطبيق عملي

الخبير  
طارق عزيز- وزارة الصناعة

أ.م.د. لقاء علي محمد  
جامعة بغداد- كلية الادارة والاقتصاد  
قسم الاحصاء

### المستخلص

تبرز أهمية الحاجة إلى طرائق تقدير كفاءة والتي تسمى بالطرائق الحصينة (Robust Methods) عندما تكون بيانات الظاهرة المدروسة ملوثة، أو وجود شواذ في المشاهدات، والتي ينتج بسببها مقدرات تؤدي إلى زيادة أو نقصان في متوسط مربعات الخطأ (MSE) مما يؤدي إلى استدلال أحصائي غير دقيق.

ومن هنا جاء هدف هذا البحث في الحصول على مقدرات حصينة لغرض تحليل الارتباط القويم (Canonical Correlation Analysis) من خلال دراسة بعض الطرائق الحصينة منها (مقدرات-M، مقدرات-MVE، مقدرات-MCD، ومقدرات-S)، حيث تم إجراء نوعين من الدراسات التطبيقية، الأولى استخدم فيها أسلوب المحاكاة للمقارنة بين طرائق التقدير المدروسة في تحليل الارتباط القويم والتعرف على أفضل مقدر بالاعتماد على المقياسين الأحصائيين متوسط التحيز ومتوسط مربعات الخطأ والذي يعطي أقل (MSE) لمعامل الارتباط القويم، والثانية استخدمت فيها البيانات الحقيقية للتحقق من أداء الطرائق الحصينة في الواقع العملي. وبشكل عام تم التوصل إلى أن مقدر (MCD) هو الأفضل في تقدير معامل الارتباط القويم مقارنة بطرائق التقدير المدروسة.

### ABSTRACT

The important of the necessity to my efficient estimation methods, which are called the robust methods, appears when the data of the studied phenomenon are contaminated, it means the observations contains outliers, which may produce estimators which result in increasing (decreasing) in the (MSE), That would leads to an inaccurate statistical inference.

From this point was the good behind this research in reaching robust estimators of canonical correlation analysis that can be achieved through the study of some robust methods such as (estimators – M, estimators – MVE, estimators-MCD and estimators-S) . which tow types of applied studies were made, The first one uses the simulation method to compare among the studied methods of estimation in canonical correlation analysis and detect the best of the estimator depending on the two statistical measurements bias mean and mean square error which renders the minimum MSE of canonical correlation analysis, the second study uses the truthful data to verity the performance in a practical actuality.

In general it is regarded that MCD estimator is the best in the estimation of canonical correlation analysis in comparison with the other studied methods of estimation.



## 1. المقدمة Introduction

أن وجود الشواذ في مجموعة البيانات أصبح في الوقت الحاضر أمراً مألوفاً ويجب التعامل معه بحذر وأن هذه الشواذ تؤدي إلى أخراج مصفوفة تباين مشترك غير صحيحة ضمن هذه المرحلة وبالتالي إلى مصفوفة ارتباط وسلسلة علاقات بين المتغيرات غير صحيحة لقد تناولت العديد من البحوث والدراسات موضوع تحليل الارتباط القويم والذي يعد من المواضيع المهمة في الأحصاء لما له من تطبيقات واسعة في معظم المجالات ومختلف العلوم.

إن تحليل الارتباط القويم هو أسلوب يدرس قوة العلاقة بين مجموعتين من المتغيرات العشوائية، وأن قياس المقارنة يعتمد على مصفوفة التباين المشترك للعينة.

لذا أصبح الاعتماد على الأساليب الحصينة في حساب مصفوفة التباين المشترك أو مصفوفة الارتباط أمراً ضرورياً للغاية إذ تقود هذه المصفوفات الحصينة إلى تحاليل حصينة ضمن تحليل متعدد المتغيرات.

## 2. هدف البحث Purpose of Search

إن هذا البحث يهدف إلى دراسة مقارنه المقدرات الحصينة في تقدير مصفوفة التباين المشترك لغرض تحليل الارتباط القويم والتعرف على خصائصها مع توضيح تأثير القيم الشاذة في أسلوب تحليل الارتباط القويم، ومن خلال هذه الدراسة يتم المقارنة بين الطرائق الحصينة المختلفة لتقدير معامل الارتباط القويم وهي (مقدرات-M، مقدرات-MVE، مقدرات-MCD، مقدرات-S) بغية الوصول إلى الهدف والتوصل إلى أفضل أسلوب للتقدير من بين الطرائق أعلاه باستخدام المحاكاة والتطبيق العملي.

## 3. الجانب النظري

يتضمن هذا المبحث دراسة بعض الطرائق الحصينة المستخدمة في تقدير مصفوفة التباين المشترك والتي من خلالها يتم تقدير معامل الارتباط القويم.

### 3.1. تحليل الارتباط القويم (Canonical Correlation Analysis)

نشأ هذا النوع من الارتباط من قبل الباحث (Hotelling) عام (1936)، إن تحليل الارتباط القويم هو أسلوب يبحث في تشخيص وتقدير قوة العلاقة بين مجموعتين من المتغيرات .

وهو يوضح الارتباط بين مجموعتين من المركبات الخطية للمتغيرات في أول مجموعة والمركبات الخطية للمتغيرات في المجموعة الأخرى، والفكرة هي تحديد أول زوج من المركبات الخطية التي تمتلك أكبر ارتباط وأن الزوج التالي في المركبات الخطية تمتلك ثاني أعلى ارتباط من بين كل الأزواج غير المرتبطة مع الأزواج السابقة وهكذا بالنسبة لبقية الارتباطات القويمية إلى أن نصل إلى أقل ارتباط وبهذه الطريقة تكون الارتباطات القويمية بين أزواج المتغيرات متناقصة بالتسلسل.

وبين الباحث (Hotelling) عام (1936) أن تحليل الارتباط المتعدد هو حالة خاصة من الارتباط القويم، فالارتباط المتعدد يحدد درجة العلاقة الخطية بين المتغير المنفرد (Y) والتركيب الخطي لمجموعة المتغيرات التوضيحية X's والتي هي (X<sub>1</sub>, X<sub>2</sub>, ..... X<sub>p</sub>) والتي لها أعلى ارتباط مع y.

ويتميز الارتباط القويم عن الارتباط البسيط في كونه يقيس قوة العديد من العلاقات البسيطة في الوقت نفسه وعن الارتباط المتعدد في كونه يقيس قوة العلاقة بين مجموعة من المتغيرات التوضيحية ومجموعة من المتغيرات المعتمدة، وأن تحليل الارتباط القويم هو الحالة الأكثر عمومية في حالات الأنموذج الخطي العام وتطبق طريقة تحليل الارتباط القويم في المجالات كافة ومختلف العلوم.

إن الارتباط القويم يشمل مجموعتين من المتغيرات العشوائية الأولى X's والتي هي (X<sub>1</sub>, X<sub>2</sub>, ..... X<sub>p</sub>) والثانية y's والتي هي (y<sub>1</sub>, y<sub>2</sub>, ..... y<sub>q</sub>) ثم إيجاد التراكيب الخطية للمجموعتين وهما a'x و b'y على التوالي، إذ يكون لها أكبر ارتباط ممكن مع المتغيرات والهدف هو إيجاد المركبات الخطية الآتية:

$$U = a_1x_1 + a_2x_2 + \dots + a_px_p = a'x \quad (1)$$

$$V = b_1y_1 + b_2y_2 + \dots + b_qy_q = b'y \quad (2)$$



إن هذه التراكيب الخطية تكون مساوية إلى أقل عدد من متغيرات المجموعتين أي أن:

$$U_i \text{ or } V_i = \min(p, q) \quad \dots\dots\dots (3)$$

إذ تحتوي المجموعة الأولى على (p) من المتغيرات والمجموعة الثانية على (q) من المتغيرات ولتكن  $q \leq p$  فإن عدد التراكيب الخطية ستكون مساوية إلى (q) وفق المعادلة (3). وعندما يكون التباين لمجموعتي المتغيرات القويمة يساوي واحد وبالشكل الذي يكون فيه الارتباط ما بين U, V, أعظم ما يمكن:

$$\text{Corr}(U, V) = \text{Corr}(\underline{a}'\underline{x}, \underline{b}'\underline{y}) = \frac{\text{Cov}(\underline{a}'\underline{x}, \underline{b}'\underline{y})}{\sqrt{\text{var}(\underline{a}'\underline{x})}\sqrt{\text{var}(\underline{b}'\underline{y})}}$$

وفقاً للشروط

$$\text{Var}(\underline{a}'\underline{x}) = a' \sum_{xx} a = 1 \quad \dots\dots\dots (4)$$

$$\text{Var}(\underline{b}'\underline{y}) = b' \sum_{yy} b = 1 \quad \dots\dots\dots (5)$$

$$\therefore a' \sum_{xx} a = b' \sum_{yy} b = 1$$

$$\text{Cov}(\underline{a}'\underline{x}, \underline{b}'\underline{y}) = a' \sum_{xy} b \quad \dots\dots\dots (6)$$

وعليه فإن  $(\rho_1^2, \rho_2^2, \dots, \rho_p^2)$  هي الجذور والمتجهات المميزة على التوالي للمعادلة المميزة:

$$\left( \sum_{xy} \sum_{yy}^{-1} \sum_{yx} - \rho^2 \sum_{xx} \right) a = 0 \quad \dots\dots\dots (7)$$

وتحل باخذ المحددة لها وجعلها مساوية للصفر أي ان:

$$\left| \sum_{xy} \sum_{yy}^{-1} \sum_{yx} - \rho^2 \sum_{xx} \right| = 0 \quad \dots\dots\dots (8)$$

وبحل هذه المعادلة نحصل على قيم  $(\rho^2)$  التي تسمى بالجذور المميزة (eigen values). فالجذور غير الصفريّة التي هي الارتباطات القويمة تكون قيمها تنازلية أي أن:

$$\rho_1 \geq \rho_2 \geq \dots > \rho_p$$

حيث أن  $(\rho_1)$  هي أكبر قيمة ارتباط إلى  $(\rho)$  الذي هو الجذر التربيعي إلى  $(\rho^2)$  ويأتي بعده  $(\rho_2)$  وهكذا لبقية الارتباطات.



فإن  $(\rho_1)$  يسمى الارتباط القويم بين المتغيرات القوية  $U_i = a_{ix}$  ,  $V_i = b_{iy}$  ومن خلال المعادلة المميزة التالية:

$$\left( \sum_{yx} \sum_{xx}^{-1} \sum_{xy} - \rho^2 \sum_{yy} \right) b = 0 \dots \dots \dots (9)$$

يتم احتساب قيم الاوزان  $(b_q)$  حيث أن  $(\rho_1^2, \rho_2^2, \dots, \rho_q^2, b_1, b_2, \dots, b_q)$  هي الجذور والمتجهات المميزة على التوالي.

### 3.2 تحليل الارتباط القويم الحصين

#### Robust Canonical Correlation Analysis. (RCCA)

إن الطريقة الواضحة لتحليل الارتباط القويم الخطي الحصين (RLCCA) هي تقدير مصفوفة التباين المشترك الحصينة  $(\Sigma)$  وحساب الجذور المميزة من خلال المصفوفتين الآتيتين:

$$\sum_{xx}^{-1} \sum_{xy} \sum_{yy}^{-1} \sum_{yx} \cdot \text{ and } \sum_{yy}^{-1} \sum_{yx} \sum_{xx}^{-1} \sum_{xy} \cdot$$

علماً أن الطريقتين أعلاه تعطي النتائج نفسها للجذور المميزة . إن تحليل الارتباط القويم (CCA) الاعتيادي يبحث في تشخيص وتقدير قوة العلاقة المشتركة بين مجموعتين من المتغيرات العشوائية وأن سائر النتائج تكون غير حصينة بوجود الشواذ في البيانات لأن مصفوفة الارتباط أو مصفوفة التباين المشترك للعينة سريعة التأثر جداً بوجود المشاهدات الشاذة. ولذلك أستعمل بعض الباحثين طرائق بديلة عن الطرائق التقليدية التي لا تتأثر أو تتحسس بوجود الشواذ وأطلق عليها أسم المقدرات الحصينة (Robust Estimators) من قبل الباحث Box عام (1953). ولأكتشاف الشواذ يبدو من الأفضل استعمال مسافات مهلنوبس التربيعية الحصينة الآتية:

$$RD_i^2 = (\underline{Xi} - \underline{\mu_n})' \sum_n^{-1} (\underline{Xi} - \underline{\mu_n}) \dots \dots \dots (11)$$

مع التقديرات الحصينة للموقع  $(\mu_n)$  والقياس  $(\sum_n)$ ، مع ملاحظة أن مسافات مهلنوبس التربيعية هي ليست ملائمة تماماً لاكتشاف الشواذ لأنها تعتمد على التقديرات العادية (غير الحصينة)، ومن خلال هذا أصبح ضرورياً أبدال مصفوفة التباين المشترك أو مصفوفة الارتباطات التقليدية بالمقدر الحصين لأجراء تحليل الارتباط القويم للبيانات ذات القيم الشاذة.

### 3.3 المقدرات الحصينة (Robust Estimators)

تطلق كلمة الحصانة (Robustness) على المقدرات التي لا تتأثر أو تتحسس بسهولة لوجود انتهاك في إحدى فرضيات التوزيع الطبيعي المتعدد المتغيرات أو وجود قيم شاذة (تلوث) قي البيانات الأحصائية ولهذه الأسباب تم إيجاد طرائق بديلة للتقدير تتعامل مع البيانات بأسلوب يختلف عن الطرائق التقليدية. والهدف الأساس من إيجاد طرائق تقدير حصينة هو لتقليل تأثير القيم الشاذة على المقدر.



### 3.4. المقدرات الحصينة لمصفوفة التباين المشترك

#### (Robust Estimators for Covariance Matrix)

لقد برزت الحاجة الى توظيف طرائق جديدة (بديلة) للتقدير والتي لا تتأثر أو تتحسس بسهولة لوجود المشاهدات الشاذة (Outliers) في العينة قيد الدراسة والتي تسمى بالمقدرات الحصينة (Robust Estimators) عند احتواء البيانات على بيانات شاذة. يتم تحليل البيانات لمتعدد المتغيرات بأسلوب الارتباط القويم من خلال حساب الجذور المميزة والمتجهات القوية لمصفوفة التباين المشترك وأيجاد معامل الارتباط القويم (p) المناظر لأكبر جذر مميز. ومن خلال هذا الأسلوب تبديل مصفوفة التباين المشترك العادية بالتقدير الحصين، فيتم إيجاد المقدر الحصين لمعلمة القياس لغرض تحليل الارتباط القويم لذا سوف نتطرق إلى بعض الطرائق الحصينة وهي:-

#### 3.4.1 مقدرات M – (M- Estimators)

لغرض حساب مقدرات M لمتعدد المتغيرات لـ  $\underline{\mu}$  و  $\underline{\Sigma}$  يتم أتباع الخطوات الآتية:-

- 1- حساب المقدرات التقليدية بطريقة دالة الأماكن الأعظم (Maximum Likelihood Method) وهي موجه الاوساط  $\underline{T}(x)$  ومصفوفة التباين المشترك  $S(x)$ ، ويتم عدّها كنقاط ابتدائية أولية (Initial Starting points) ( $z=0$  في بداية التكرارات).
- 2- حساب مسافات مهنوبس التربيعية العادية (Squared Mohalonobis distances) على وفق الصيغة التالية:-

$$di^2 = (\underline{Xi} - \underline{\bar{X}})' S^{-1} (\underline{Xi} - \underline{\bar{X}})$$

$$\underline{\bar{X}} = \underline{T}(x) = \frac{\sum_{i=1}^n \underline{Xi}}{n} \quad \text{وحساب المتوسط:}$$

$$S(x) = \frac{\sum (\underline{Xi} - \underline{\bar{X}})(\underline{Xi} - \underline{\bar{X}})'}{n-1} \quad \text{ومصفوفة التباين المشترك:}$$

- 3- يعاد حساب موجه الأوساط ومصفوفة التباين المشترك بأستعمال منظومة التكرارات الآتية:-

$$\underline{T}(x)^{j+1} = \frac{\sum_{i=1}^n W(di^{(j)}) \underline{Xi}}{\sum_{i=1}^n W(di^{(j)})} \quad \dots\dots\dots (33)$$

$$S(x)^{(j+1)} = \frac{\sum W(di^{(j)}) (\underline{Xi} - \underline{T}(x)^{(j+1)})(\underline{Xi} - \underline{T}(x)^{(j+1)})'}{\sum W(di^{(j)}) - 1} \dots\dots\dots (34)$$

إذ أن  $W(di^{(j)})$  هي دالة الوزن التي تعتمد على مسافات مهنوبس التربيعية المعتمدة بدورها على المقدرات التقليدية.

وفيما يلي مقدرات M – المعروفة بدلالة وزن Huber  $W(.)$  التالية في حالة متعدد المتغيرات (Huber, 1964).

$$W(di) = \begin{cases} 1 & 0 \leq di < c \\ \frac{c}{di} & di > c \end{cases} \quad \text{إذ أن } C^2 \text{ هي وسيط قيم مربع كاي سكوير الجدولية عند درجة حريه } P \text{ وبمستوى معنوية } 0.95$$

$$C = \sqrt{\chi^2_{(p,0.95)}} \quad \text{أي أن :}$$



تمثل هذه الخطوات التكرار الأول ثم يعاد على وفق الخطوتين الأخيرتين بصورة تكرارية إذ أن كل تكرار يعتمد على نتائج التكرار الذي يسبقه وتتوقف التكرارات عندما يكون الفرق بين نتائج تكرارين متعاقبين غير ملحوظ أو قليل.

وأن نقطة أنهيارها لا يتجاوز  $\frac{1}{p+1}$  هذا يعني بأن مقدرات (M) تصبح أكثر تحسس للملاحظات الشاذة بتزايد عدد المتغيرات أي أن:

$$\delta^* \leq \frac{1}{p+1}$$

إذ أن  $\delta^*$  : تمثل نقطة الأنهيار.

P : يمثل عدد المتغيرات.

ويتم بعدها إيجاد الجذور والمتجهات المميزة ومعامل الارتباط القويم المناظر لأكبر جذر مميز.

### 3.4.2 مقدرات أصغر حجم مجسم قطع ناقص

#### Minimum Volume Ellipsoid Estimators - MVE

لغرض حساب مقدرات MVE- يتم اتباع الخطوات الآتية:

- 1- يتم تحديد مصفوفة البيانات التي تحتوي على (n) من الصفوف (المشاهدات) و P من الأعمدة المتمثلة بالمتغيرات بعدها يتم تحديد العينات الجزئية بحجم (P+1) من المشاهدات، أي أننا نقوم بأختيار عدد من العينات الجزئية (المصفوفات) مساوي إلى  $C_{P+1}^n$  وهذا العدد من التوافق يزيد بزيادة عدد المشاهدات أو عدد الأبعاد أو كليهما.
- 2- لكل عينة جزئية (j) بحجم (P+1) المستخرجة وفق الخطوة (1) يتم استخراج متجة الأوساط ( $\underline{\mu}_j$ ) ومصفوفة التباين المشترك  $\sum_j$  وكالاتي:

$$\underline{\mu}_j = \frac{1}{P+1} \sum_{i \in J} \underline{X}_i$$

$$\sum_j = \frac{1}{p} \sum_{i \in J} (\underline{X}_i - \underline{\mu}_j)(\underline{X}_i - \underline{\mu}_j)'$$

- 3- لكل أختيار لـ ( $\underline{\mu}_j, \sum_j$ ) أعلاه يجري استخراج المسافات التربيعية وفق الصيغة الآتية:

$$D_i^2(j) = (\underline{x}_i - \underline{\mu}_j)' \sum_j^{-1} (\underline{x}_i - \underline{\mu}_j)$$

- 4- بعد استخراج  $D_i^2(j)$  للعينات كافة، يتم في هذه الخطوة ترتيب هذه القيم تصاعدياً وأختيار  $D_{(hp)}^2(j)$  حيث  $hp = \frac{n+p+1}{2}$  لكل عينة جزئية (j).

- 5- بعد ماتم سابقاً يجري هنا أختيار أفضل عينة جزئية وهي العينة الجزئية التي تمتلك أصغر قيمة لمحددة مصفوفة التباين المشترك من بين كل العينات الجزئية في:

$$\tilde{J} = \arg_j \min \det \sum_j D_{(hp)}^2(j)$$



6- بعد تحديد أفضل مجموعة جزئية وتحديد متجه الأوساط ومصفوفة القياس لها، يتم تحديد معامل تصحيح مصفوفة القياس والذي تكون قيمته مساوية إلى [2]:

$$C_{\alpha} = 1 + \frac{15}{n - p}$$

7- بعد احتساب مصفوفة القياس لأفضل عينة جزئية ومعامل التصحيح لها نقوم بحساب مصفوفة التباين المشترك المضخمة (Inflated Covariance Matrix) من خلال الصيغة الآتية:

$$S(x) = (\chi_{p,0.5}^2)^{-1} \cdot C_{\alpha} \cdot \sum_{(j)} D_{(hp)}^2(j) \dots\dots\dots (38)$$

8- وفق المصفوفة  $S(x)$  المستخرجة من الفقرة أعلاه يتم حساب المسافات الحصينة (مسافات مهنوبس التربيعية الحصينة) وفق:

$$Rdi^2(j) = (\underline{x}_i - \underline{\mu}_j)' S^{-1}(x) (\underline{x}_i - \underline{\mu}_j) \dots\dots\dots (39)$$

ولمشاهدات العينة الأصلية كافة (n من المشاهدات).

9- بعد احتساب المسافات الحصينة يبدأ الأسلوب التعاقبي في الحساب والذي يتم بزيادة مشاهدة واحدة إلى حجم العينة الجزئية ليصبح حجم العينة الجزئية في المرحلة الثانية (P+2) من المشاهدات. يتم اختيار (P+2) من المشاهدات في المرحلة الثانية كحجم للمجموعة الجزئية، وهي المشاهدات التي لها أقل قيم المسافات التربيعية الحصينة المستخرجة من الصيغة الآتية:

$$di^2(j) = (\underline{X}_i - \underline{\mu}_j)' S^{-1}(X) (\underline{X}_i - \underline{\mu}_j)$$

وعادة تكون (P+1) نفسها من المشاهدات في المجموعة الأولى فضلاً عن مشاهدة جديدة.

10- بعد اختيار المجموعة الجزئية ذات الحجم (P+2) يتم تكرار الخطوات (2) و (9) لنحصل على عينة جزئية بحجم (P+3) من المشاهدات.

11- تستمر سلسلة العمليات التعاقبية في الحساب بزيادة مشاهدة واحدة في كل مرحلة إلى أن نصل حد التوقف.

ومعيار التوقف هنا هو الوصول إلى حجم العينة الجزئية (hp) حيث  $hp = \frac{n + p + 1}{2}$  إذ تكون هذه العينة متجانسة ومتسقة مع بعضها وتنتج نحو المركز.

12- للعينة المستخرجة وفق الفقرة أعلاه ذات (hp) من المشاهدات يجري أستخراج متجه الأوساط لها  $\mu_{(hp)}$  مع مصفوفة القياس  $\sum_{(hp)}$  والتي يتم بموجبها أستخراج المسافات التربيعية الحصينة

$Rdi^2(hp)$  ليتم من خلالها تشخيص المشاهدات الشاذة (المشاهدات الشاذة هي التي تزيد قيمة

المسافة التربيعية الحصينة لها عن قيمة  $\chi_{(P,0.975)}^2$  الجدولية).

إن طريقة حساب مقدرات (MVE) هي طريقة تكرارية تقود إلى وقت وكلفة حسابات عالية إلا أنها تضمن شروط الحصانة للمقدرات المستخرجة.

ويتم بعدها إيجاد الجذور والمتجهات المميزة ومعامل الارتباط القويم المناظر لأكثر جذر مميز.



### 3.4.3 مقدرات أصغر محددة مصفوفة تباين مشترك

#### Minimum Covariance Determinant estimators -MCD

لغرض حساب مقدرات - MCD يتم اتباع الخطوات الآتية:

1- للينة المكونة من (n) من المشاهدات (الصفوف) و (P) من المتغيرات (الاعمدة) يتم أستخراج العينات الجزئية ذات الحجم (h)، حيث أن  $h = \frac{n+p+1}{2}$  أي أننا نختار  $C_h^n$  من العينات الجزئية.

2- لكل عينة جزئية (j) يجري أستخراج متجة الأوساط  $\bar{x}_j$  ومصفوفة التشتت  $\sum_j$ .

3- بعد حساب مصفوفات التشتت للعينات الجزئية كافة ذات الحجم (h) وتحديد قيمة المحددة كل مصفوفة يتم أختيار أفضل عينة جزئية وهي العينة التي لها أصغر محدد مصفوفة تباين مشترك.

$$\tilde{J} = \arg_j \min \left| \sum_j \right| \dots \dots \dots (41)$$

4- بعد أن تم تحديد في الفقرة (3) أعلاه أفضل مجموعة جزئية  $\tilde{J}$  لنحصل من خلالها على مقدر (MCD)

الحصين للمجموعة الجزئية  $\tilde{J}$  كمتوسط  $\hat{\mu}_{j,MCD}$  ومصفوفة تباين مشترك  $\sum_{j,MCD}$  إذ أن مقدر المتوسط معرف كالاتي:

$$\hat{\mu}_{j,MCD} = \frac{1}{h} \sum_{i=1}^h X_i \dots \dots \dots (42)$$

ومقدر التشتت:

$$\sum_{j,MCD} = \frac{1}{h-1} \sum (x_i - \hat{\mu}_{j,MCD})(x_i - \hat{\mu}_{j,MCD})' \dots \dots \dots (43)$$

5- بالأعتماد على التقديرات أعلاه  $\hat{\mu}_{j,MCD}$  و  $\sum_{j,MCD}$  نحسب مسافات مهلنوبس التربيعية الحصينة الآتية:

$$RD_i = \sqrt{(X_i - \hat{\mu}_{j,MCD})' S_{j,MCD}^{-1} (X_i - \hat{\mu}_{j,MCD})} \dots \dots \dots (44)$$

وتعد  $X_i$  قيمة شاذة إذا فقط إذا :

$$RD_i = \sqrt{\chi^2_{(p,0.975)}} \dots \dots \dots (45)$$

لتشخص بعد ذلك البيانات الشاذة عن غيرها.

ويتم بعدها إيجاد الجذور والمتجهات المميزة ومعامل الارتباط القويم المناظر لأكبر جذر مميز.





### 3.4.4 مقدرات S – (S – Estimators)

يمكن تلخيص خوارزمية مقدر (S) بالخطوات الآتية:-

1- إن مقدرات (S) معرفة كحلول للموقع والتشتت ( $t_n, C_n$ ) لمسألة تقليل محددة التباين المشترك C وفقاً الى :

$$\frac{1}{n} \sum_{i=1}^n \rho[d(\underline{X}_i, t, C)] = bo$$

2- للينة المكونة من (n) من المشاهدات (الصفوف) و (P) من المتغيرات (الأعمدة) يتم أستخراج العينات الجزئية ذات الحجم (P+1) أي نختار  $C_{P+1}^n$  من العينات الجزئية.

3- بعد أستخراج العينات الجزئية يجري أستخراج متجه الموقع (Location) ومصفوفة التشتت لكل عينة جزئية.

4- يكون أختيار أفضل عينة جزئية من العينات المستخرجة وفقاً لمحددة مصفوفة التباين المشترك إذ يتم أختيار العينة الجزئية التي تكون فيها محددة مصفوفة التباين المشترك لها أقل ما يمكن أي أن:

$$\tilde{J} = \arg_j \min \left| \sum_j \dots \dots \dots \right| \quad (49)$$

5- بعد أختيار أفضل عينة جزئية من خلال موجة الموقع ومصفوفة التباين المشترك يتم استخراج مسافات مهلنوبس التربيعية العادية بين نقاط المشاهدات  $X_i$  والموقع t بالاعتماد على مصفوفة التباين المشترك إذ:

$$d(\underline{x}_i, t, C) = \left[ (\underline{x}_i - t)' C^{-1} (\underline{x}_i - t) \right]^{\frac{1}{2}} \quad \dots \dots \dots (50)$$

إذ أن العينة  $X_1, \dots, X_n$  تعرف كثائي ( $t_n, C_n$ ) والتي تقلل C .  
6- يتم أستخراج  $\rho(u)$  حيث أن  $\rho$  هو دالة وزن (Biweight) لتحقيق المعادلة:  
بعد التعويض بقيمة ( $u=0.7$ ) في المعادلة :

$$\rho(u) = \min \left[ \frac{u^2}{2} - \frac{u^4}{2C^2} + \frac{u^6}{6C^4}, \frac{C^2}{6} \right] \quad \dots \dots \dots (47)$$

نحصل على  $\rho(u) = \frac{1}{6}$  وبتطبيق المعادلة  $\rho(u) = \frac{b}{\delta}$  يتم تحديد قيمة b وبالشكل التالي:

$$\rho(u) = \frac{b}{\delta}$$

$$\frac{1}{6} = \frac{b}{0.50}$$

$$\therefore b = 0.083$$

7- بعد أيجاد قيمة b أعلاه ندخل في عملية تكرار المعادلة  $\rho(u)$  ليتم فيها البحث عن قيمة u وذلك لتحقيق شرط المقدر.



8- باستخراج قيم الثوابت يحقق شرط المقدر ليتم بعدها اختيار مصفوفة التشتت التي تحقق الشرط مع متجة الموقع ليعدا مقدرات ( S ) الحصينة للموقع والتشتت والتي يتم بموجبها استخراج مسافات مهلنويس التربيعية الحصينة لتشخيص بعد ذلك نقاط البيانات الشاذة عن غيرها وتعد المشاهدة شاذة إذا وفقط إذا :

$$RDi > \sqrt{\chi^2_{(p,0.975)}}$$

ويتم بعدها إيجاد الجذور والمتجهات المميزة ومعامل الارتباط القويم المناظر لأكبر جذر مميز.

#### 4. الجانب التطبيقي

لغرض توضيح الطرائق المدروسة في البحث والوصول الى افضل النتائج فقد تم استخدام نوعين من التطبيقات الاولى استخدم فيها اسلوب المحاكاة والثانية تم استخدام البيانات الحقيقية لتطبيقها على الطرائق الحصينة التي تم عرضها في الجانب النظري وتم الحصول على كافة النتائج للدراستين من خلال برنامج (Matlab) حيث تم حساب الجذور المميزة ومنها معامل الارتباط القويم المناظر لأكبر جذر مميز ولجميع طرائق التقدير الحصينة المدروسة.

#### 4.1 الاجراء التطبيقي الاول (استخدام اسلوب المحاكاة):

##### 4.1.1 وصف مراحل تجربة المحاكاة:

لقد تضمنت تجارب المحاكاة المراحل الأتية لغرض حساب معامل الارتباط القويم وتطبيق الطرائق الحصينة لتقدير مصفوفة التباين المشترك، ومنها يتم تقدير معامل الارتباط القويم ( $\hat{\rho}$ ).

##### المرحلة الأولى:

وهي مرحلة اختيار القيم الافتراضية، إذ تعد من المراحل المهمة التي تعتمد المراحل الأخرى عليها، ومن أجل حساب معامل الارتباط القويم، تم افتراض المعالم  $\epsilon, n, P_2, P_1$  وتم دراسة ثلاث تجارب والجدول الآتي يوضح تشكيل الحالات التي درست في عملية المحاكاة وكالاتي:

جدول رقم (1)

Experiment	$P_1 = P_2$	$\epsilon = 0.0$	$\epsilon = 0.10$	$\epsilon = 0.20$
		n	n	n
I	4	20 50 100	20 50 100	20 50 100
II	5	20 50 100	20 50 100	20 50 100
III	6	20 50 100	20 50 100	20 50 100

حيث أن كل تجربة من التجارب الثلاث بحثت عند ثلاثة مستويات من التلوث وهي ( $\epsilon = 0.0, 0.10, 0.20$ ) علما ان نسبة التلوث في المجموعة الاولى هي نفسها في المجموعة الثانية، وبأحجام عينات مختلفة وهي (100, 50, 20) وكررت كل تجربة (100) مرة. وأن  $P_1$  يمثل المجموعة الاولى من المتغيرات و  $P_2$  : يمثل المجموعة الثانية من المتغير حيث يكون عدد المتغيرات في المجموعتين متساوي لضمان تحقق وجود علاقة خطية بين جميع المتغيرات وهي فكرة الارتباط القويم.



### المرحلة الثانية

توليد البيانات المتجانسة بالشكلين النظيف والملوث والمجموعة الثانية باستخدام طريقة (Box-Muller) والتي تتبع التوزيع الطبيعي متعدد المتغيرات. إذ يتم توليد بيانات تتبع التوزيع الطبيعي متعدد المتغيرات (Multivariate Normal) وعددها  $n(1-\epsilon)$  حيث أن  $\epsilon$  تمثل نسبة التلوث وعليه تم الحصول على  $X \sim N(\mu_x, \Sigma_x)$  وهو التوزيع الأصلي (النظيف) ثم لوثت هذه البيانات ببيانات توزيع آخر طبيعي بمتوسط (0) وتباين يختلف عن الأول أي تم توليد  $\epsilon$  من المشاهدات التي تتوزع وفق التوزيع  $Z \sim MvN(\mu_z, \Sigma_z)$  وهو التوزيع الملوث ولكون هذه الدراسة تحتاج إلى توزيع آخر غير التوزيع الأساس وهو التوزيع الملوث فقد تم توليد بيانات متعدد المتغيرات، والذي يتضمن مشاهدات شاذة لدراسة مدى تأثير القيم الشاذة على تقدير المعلمات. وبعد ذلك ينتج توزيع ثالث  $g \sim MvN(\mu_g, \Sigma_g)$  وهو التوزيع الخليط، إذ أن:

$$\underline{g} = \begin{bmatrix} \underline{X} \\ \underline{Z} \end{bmatrix}$$

and

$$g \sim (1-\epsilon)N(\mu_x, \Sigma_x) + \epsilon N(\mu_z, \Sigma_z)$$

وفي هذه المرحلة تم استخدام برنامج (MatLap) لحساب الجذور المميزة ومنها معامل الارتباط القويم ( $\rho_1$ ) المناظر لأكبر جذر مميز ثم حساب المتجهات القوية ( $a, b$ ) المناظرة لأكبر جذر مميز.

### المرحلة الثالثة

يتم تقدير معامل الارتباط القويم ( $\hat{\rho}$ ) على وفق صيغ الطرائق الحصينة والمبينة في الجانب النظري:-

- 1- مقدرات - M.
- 2- مقدرات - MVE.
- 3- مقدرات - MCD.
- 4- مقدرات - S.

### المرحلة الرابعة

المقارنة بين طرائق التقدير الحصينة لغرض الوقوف على مدى كفاءة ودقة أسلوب التقدير. إذ تم استعمال المقياسين الآتيين:-

- 1- متوسط التحيز: (Bias Mean)

$$E \text{ Bias}(\hat{\rho}) = \frac{\sum_{i=1}^L [\hat{\rho}_{(i)} - \rho]}{L} = E \hat{\rho} - \rho$$

إذ أن  $i = 1, 2, \dots, L$   
 وأن  $L$ : يمثل عدد التكرارات (Replications) لكل تجربة.  
 $\hat{\rho}_{(i)}$ : يمثل مقدر ( $\rho$ ) على وفق الأسلوب المستعمل في التقدير.

- 2- متوسط مربعات الخطأ: (Mean Square error)

$$E \text{ MSE}(\hat{\rho}) = \frac{\sum_{i=1}^L [\hat{\rho}_{(i)} - \rho]^2}{L}$$



- وبالتالي تحدد أفضل أسلوب من بين الأساليب المستعملة في التقدير والتي تمتلك أقل (MSE). في هذا المبحث سيتم عرض وتحليل نتائج تجارب المحاكاة لتقدير معامل الارتباط القويم حسب الطرائق المبينة في الجانب النظري.
- وقد تم الحصول على هذه النتائج بالاعتماد على الامكانية العالية لبرنامج (Matlab) فقد تبين بشكل عام ولعموم التجارب ما يلي والجداول (2) و(3) و(4) توضح ذلك.
1. أظهرت مقدرات-MCD الافضلية من حيث امتلاكها الى اقل (MSE) عند مقارنتها مع باقي طرائق التقدير المدروسة لتقدير معامل الارتباط القويم.
  2. اظهرت كل من مقدرات -MVE ومقدرات - S تقارباً كبيراً في النتائج من حيث اقتراب قيم (MSE) لمعامل الارتباط القويم لكلا المقدرين.
  3. تبين ان مقدرات-M اقل كفاءة في تقدير معامل الارتباط القويم مقارنة مع باقي طرائق التقدير المدروسة.









## 4.2 الاجراء التطبيقي الثاني:

### 4.2.1 استخدام البيانات الحقيقية الملوثة:

في هذا التطبيق تم اعتماد بيانات حقيقية ملوثة (تم التحقق من ان البيانات ملوثة من خلال استخدام مسافات مهنوبس التربيعية) اخذت من احدى المدارس الثانوية للبنات في محافظة بغداد لتحليل العلاقة بين اداء (25) طالبة في المرحلة المتوسطة ومرحلة الاعدادية (علما ان مجموعة الاعدادية هي نفس مجموعة المتوسطة) ومعرفة قوة العلاقة بين المجموعتين من المتغيرات ودراسة اثر القيم الشاذة على تقدير معامل الارتباط القويم .







## جدول رقم (7)

يبين قيم مسافات مهلنوبس التربيعية لكل مشاهدة من مجموعتي  
البيانات الحقيقية  $X_S'$  و  $Y_S'$

مشاهدات مجموعة المتغيرات $X_S'$	$d_i^2$	مشاهدات مجموعة المتغيرات $Y_S'$	$d_i^2$
1	8.8598	1	6.1999
2	9.2034	2	5.6362
3	7.6161	3	4.1488
4	1.1811	4	5.7760
5	7.9796	5	3.8029
6	8.1773	6	6.8925
7	2.9974	7	3.6767
8	12.1728	8	16.6953
9	8.2953	9	6.1673
10	5.3448	10	5.7783
11	5.7003	11	3.9437
12	10.2516	12	7.2910
13	7.1186	13	5.6692
14	17.4193	14	11.2198
15	4.6730	15	5.5755
16	18.8762	16	4.9898
17	8.1825	17	7.6281
18	7.2996	18	1.5688
19	4.0416	19	7.6836
20	16.6258	20	9.1105
21	1.7374	21	3.3609
22	1.6353	22	6.4349
23	3.7245	23	8.5097
24	10.3509	24	13.0380
25	2.5360	25	7.2027



سوف يتم تطبيق جميع الطرائق الحصينة عليها لغرض تقدير معامل الارتباط القويم، في البداية يتم اختبار البيانات من حيث توافقها مع التوزيع الطبيعي متعدد المتغيرات ام لا ولاختبار توفر شرط التوزيع الطبيعي المتعدد المتغيرات (Multivariate Normal Distribution) فقد اقترح الباحث Mardia في الاعوام (1970 و1980) اختباراً يعتمد على معامل التفلطح (Kurtosis) ومعامل الالتواء (Skewness) في حالة متعدد المتغيرات (Multivariate)، تم إجراء اختبار البيانات الحقيقية باستخدام صيغة معامل التفلطح للعينة متعددة المتغيرات ويتضمن الاختبار وضع الفرضية الآتية.

$$H_0 : B = 0$$

$$H1: B \neq 0$$

حيث أن:

B : تمثل معامل التفلطح الطبيعي متعدد المتغيرات.

علماً أن معامل التفلطح للعينة متعددة المتغيرات (b<sub>2</sub>, p) هو :

$$b_{2,p} = \frac{1}{n} \sum_{i=1}^n [(Xi - \bar{X})' S^{-1} (Xi - \bar{X})]^2$$

وتتم المقارنة بعد إيجاد معامل التفلطح الطبيعي متعدد المتغيرات (Normalized Kurtosis) الآتي:

$$B = \frac{b_{2,p} - \left[ \frac{p(p+2)(n-1)}{(n+1)} \right]}{\sqrt{\frac{8p(p+2)}{n}}} \sim ASYM N(0,1)$$

علماً أن أحصاءة الاختبار B تتوزع طبيعياً بمتوسط p(p+2) وتباين 8p(p+2)/n.

إذ أن : p = p<sub>1</sub>+p<sub>2</sub> تمثل عدد المتغيرات في المجموعتين.

فإذا كانت قيمة |B| أقل من قيمة Z الجدولية لمستوى معنوية (α = 0.05) والتي تساوي 1.96 فإن البيانات تتوزع طبيعياً لمتعدد المتغيرات.

بعدها يتم اختبار وجود المشاهدات الشاذة في بيانات متعددة المتغيرات فإن القياس هو مسافات مهلبوس التربيعية، فإذا كانت قيمة مسافات مهلبوس التربيعية للمشاهدة أكبر من قيمة القطع (cut of value)

$$\chi^2_{(p,0.975)}$$

إذ يتم حساب مسافات مهلبوس التربيعية لكل مشاهدة وفق الصيغة الآتية:

$$d_i^2 = (Xi - \bar{X})' S^{-1} (Xi - \bar{X})$$

حيث أن:

d<sub>i</sub><sup>2</sup> : تمثل مسافة مهلبوس التربيعية.

وعند حساب مسافة مهلبوس التربيعية لكل مشاهدة من مشاهدات العينة يتم مقارنتها مع (p,0.975)

χ<sup>2</sup> الجدولية وأظهرت النتائج وجود قيم شاذة في البيانات فتعد المشاهدة (16) في مجموعة البيانات X<sub>S</sub>'

شاذة لكون قيمة d<sub>i</sub><sup>2</sup> لها أكبر من قيمة χ<sup>2</sup><sub>(8,0.975)</sub> الجدولية والتي تساوي (17.53).

وتعد المشاهدة (8) في مجموعة البيانات Y<sub>S</sub>' شاذة لكون قيمة d<sub>i</sub><sup>2</sup> لها أكبر من قيمة

χ<sup>2</sup><sub>(7, 0.975)</sub> الجدولية والتي تساوي (16.01) وكما مبين في الجدول رقم(7). بالاعتماد على برنامج

خاص كتب بلغة (matlab) لغرض حساب مسافات مهلبوس التربيعية، ومن أجل الحصول على معامل الارتباط القويم واختبار معنويته كان لا بد من حساب موجه الجذور المميزة ومنه حصلنا على معامل الارتباط القويم (ρ<sub>1</sub>) المناظر لأكبر جذر مميز وهو (0.9271) إذ كان مساوياً إلى (0.9628)، ويتم بعد ذلك إيجاد معامل الارتباط القويم المناظر لأكبر جذر مميز لجميع طرائق التقدير المدروسة ومن خلال النتائج التي حصلنا عليها من خلال برنامج (Matlab) وبعد مقارنتها بالقيمة الحقيقية لمعامل الارتباط القويم



ولجميع الطرائق الحصينة اظهرت مقدرات-MCD الافضلية عن باقي الطرائق لاقترب قيمتها التقديرية من القيمة الحقيقية.

#### 4.2.2 أستعمال البيانات الحقيقية النظيفة (بدون تلوث):

أعتمد هذا التطبيق على البيانات الحقيقية النظيفة (بدون تلوث) أي بعد أستبعاد القيم الشاذة من البيانات. يتم اختبار توفير شرط التوزيع الطبيعي المتعدد المتغيرات لنفس الفرضية السابقة

$$H_0: B = 0$$

$$H_1: B \neq 0$$

وبعد التأكد من توفر شرط التوزيع الطبيعي للبيانات (أي ان البيانات لها توزيع طبيعي متعدد المتغيرات) يتم ايجاد معامل الارتباط القويم المناظر لأكبر جذر مميز لجميع طرائق التقدير المدروسة ومن خلال النتائج التي حصلنا عليها وبعد مقارنتها بالقيمة الحقيقية لمعامل الارتباط القويم ولجميع الطرائق الحصينة اظهرت مقدرات-MCD الافضلية عن باقي الطرائق لاقترب قيمتها التقديرية من القيمة الحقيقية يليها في الافضلية مقدرات-MVE ويليها مقدرات-S ثم مقدرات-MI والجدول أدناه يوضح ذلك.





نلاحظ من خلال التطبيقين الأول في حالة وجود التلوث في البيانات والثاني في حالة البيانات النظيفة، ظهر أن وجود القيم الشاذة في بيانات العينة أدت إلى زيادة قيمة معامل الارتباط القويم حيث ارتفعت قليلاً قياساً إلى تلك التي حسبت في حالة عدم وجود قيم شاذة (البيانات النظيفة) ولجميع الطرائق الحصينة. أن الهدف من دراسة البيانات الحقيقية هو إعطاء إدارة المدرسة صورة واضحة عن أداء الطالبات في المرحلتين (الأعدادية والمتوسطة) إذ تبين أن هناك علاقة طردية قوية بين أداء الطالبات في الصف السادس العلمي والصف الثالث المتوسطة وهذا يدل على أن أي تدني أو ارتفاع في مستوى أداء الطالبات في مرحلة المتوسطة يرافقه تدني أو ارتفاع في مستوى أداء الطالبات نفسها في مرحلة الأعدادية لذلك على إدارة المدرسة الأهتمام بمرحلة المتوسطة والتخطيط الجيد من خلال مآتراه مناسباً من توفير كادر تدريسي متخصص لتلك المرحلة فضلاً عن توفير المستلزمات الخاصة الضرورية بالدراسة من أجل الارتقاء والنهوض بالمستوى العلمي .

## 5. الأستنتاجات والتوصيات

### 5.1 الأستنتاجات: Conclusion

- 1- من خلال ماتم عرضه من دراسات والتي تمت بأستعمال أسلوب المحاكاة ونتائج البيانات الحقيقية تم التوصل إلى جملة من الأستنتاجات الآتية:
- 1- بشكل عام أثبتت مقدرات (MCD) كفاءتها في تقدير معامل الأرتباط القويم مقارنة مع باقي طرائق التقدير المدروسة بأستعمال المقياسين الأحصائيين متوسط التحيز ومتوسط مربعات الخطأ (MSE) وحجوم العينات جميعاً (n=20,50,100).
- 2- تبين أنه كلما زادت نسبة التلوث في البيانات فإن أسلوب مقدرات (MCD) يكون أنسب في تقدير معامل الأرتباط القويم ويعود سبب ذلك إلى نقطة أنهياره العالية والى حد 50%.
- 3- حافظ كل من أسلوب التقدير مقدرات (MVE) ومقدرات (S) على إعطاء قيم تقديرية متقاربة ومتطابقة لمعامل الأرتباط القويم ولعموم التجارب .
- 4- ظهر أن مقدرات (M) (متعدد المتغيرات) أقل كفاءة في تقدير معامل الأرتباط القويم مقارنة مع باقي طرائق التقدير المدروسة .
- 5- من خلال نتائج الجانب التطبيقي تبين أن قيم معامل الأرتباط القويم تزداد بزيادة نسبة التلوث في البيانات ولعموم التجارب .

### 5.2 التوصيات: Recommendations

- 1- مما تقدم في الجوانب السابقة للدراسة ونتائجها نوصي بما يلي :
- 1- اعتماد أسلوب مقدرات (MCD) الحصينة لتقدير معامل الأرتباط القويم.
- 2- من خلال دراسة مقدرات (MVE) و مقدرات (S) أوصي مستقبلاً بأستعمال إحداهما في تقدير معامل الأرتباط القويم لكونها تعطي نتائج متقاربة بسبب أن مقدرات (S) هي حالة خاصة من مقدرات (MVE) ويمكن أن تستعمل كنقاط أبتدائية لمقدرات (S).
- 3- دراسة الطرائق الحصينة لتقدير معامل الأرتباط القويم وذلك بزيادة عدد المتغيرات المستعملة وحجوم العينات المختلفة مع زيادة نسبة التلوث في البيانات تصل إلى (50%) وملاحظة مما يحدث من تغير في القيم الحصينة المستخرجة.
- 4- تطبيق طرائق تقدير حصينة أخرى غير المذكورة في البحث لمعرفة مدى كفاءتها في تقدير معامل الأرتباط القويم مثل مقدرات (SD) ومقدرات (EMVE) .



## المصادر

- 1- صالح، طارق عزيز، (2009)، "مقارنة بعض الطرائق الحصينة في تحليل الارتباط القويم الخطي باستخدام المحاكاة مع تطبيق عملي" رسالة ماجستير، كلية الإدارة والاقتصاد، جامعة بغداد.
- 2- فائز، نبيلة عبد الهادي، (1998)، "مقارنة بعض التقديرات الحصينة لمعلمة الموقع في متعدد المتغيرات مع تطبيق عملي"، رسالة ماجستير، كلية الإدارة والاقتصاد، جامعة بغداد.
- 3-Alexander Von Eye and G.Anne Bogat,(2004),"Testing the assumption of multivariate normality",psychology science,volume46,p.243-258.
- 4- Branco, J. A., Croux, C. Filzmoser, P. and Oliveira, M. R. (2005), "Robust Canonical Correlation : Acomparative Study" Computational statistic ,Vol 20(2), PP. 203-229.
- 5-Croux,C. and Dehon, C. (2001), "Robust Linear discriminate Analysis Using S-estimators " The Canadian Journal of statistics, Vol, 29, PP.473-492.
- 6-Croux, C. and Dehon, C, (2002), " Analyse Canonique basée sor des estimateurs robustes de La matrice de Covariance ", Revuede statistique Appliquee 2, PP. 5-26.
- 7-Croux, C, Broeck, G.H, Rousseeuw,P.J. (2002), "Location Adjustment for The Minimum Volume Ellipsoid Estimator", Volume 12, PP.191-200.
- 8- Danijel Skočaj and Aleš Leonardis. (2004)," Canonical Correlation Analysis for appearance-based orientation estimation and self- localization", University of Ljubljana.
- 9-David M.Rocke,(1996)," Robustness Properties of S-estimators of Multivariate Location and shap in high dimension", Ann. Statist, Volumen 24, Number 3 (1996),PP. 1327-1345.
- 10-David Weenink,(2003),"canonical correlation analysis",Instaitute of phonetic sciences,university of Amsterdam,proceeding 25.pp.81-99.
- 11-Dehon, C. , Filzmoser, P. and Croux, C. (2000)," Robust Methods for Canonical Correlation Analysis", In H. A. L. Kiers. J. P. Rasson, P. J. F. Geonen and M. Schrader, Eds., Data Analysis, Classification , and Related Methods Berlin, Springer-Verlage, PP. 321-326.
- 12-Mia Hubert,PeterJ.Rousseeuw and Stefan Van Aelst,(2008),"High- Breakdown Robust Multivariate Methods", statistic science,vol.23,no.1.pp.92-119.
- 13-Ricardo Antonio Maronna,(1976),"Robust M-Estimators of multivariate location and scatter",The Annals of statistics,vol.4,no.1.
- 14-Taskinen, S, Croux, C. Kankainen, A. Ollila, E. and Oja,H, (2006), " Canonical Analysis based on scatter Matrices", Journal of Multivariate Analysis, 97(2). PP.359-384.
- 15-Willis A. Jensen, Jeffrey B. Brich, and William H. Woodall, (2007), "High Breakdown Estimation Methods for Phase I Multivariate Control Charts" ,Department of Statistics, Virginia Polytechnic Institute and State University Blacksburg, VA 24061-0439,USA, Volume 23, Issue 5,PP. 615-629.
- 16-Wl poston, Ej wegman, Jl Solka, (1993), "Robust Estimation of Multivariate Location and Scatter based on the Effective Independence Distribution", Wendyl. Poston Naval,pp.1-18.
- 17-Yijun Zuo, (2005), " Robust Location and Scatter Estimators in multivariate Analysis", East Lansing ,ML 488824, USA,pp.1-23.