

## مقارنة المقدرات الحصينة في اسلوب التحليل العاملي

م.م. افراح كاظم جويد  
معهد الادارة / واسط

أ.م.د. لقاء علي محمد  
كلية الادارة والاقتصاد/جامعة بغداد

### الملخص

لتحليل عدد كبير من المتغيرات بهدف استخلاص عوامل جديدة عددها اقل من العوامل التي يعزى لها تباين تلك المتغيرات ، نستخدم اسلوب التحليل العاملي وهو اسلوب احصائي يمكن استخدامه على نطاق واسع من المجالات و الانشطة العلمية .

الدراسات السابقة اثبتت ان البيانات بدون اي شواذ هي نوعا ما حالة مستثناة ضمن الواقع العملي ، ويمكن تجنب هذه المشكلة بأستخدام بعض الطرائق الحصينة المناسبة للنموذج او التنظيف الاولي للبيانات من الشواذ ومن ثم تطبيق الطرائق الاحصائية العادية او الكلاسيكية عليها مثل التحليل العاملي .

ان تقدير الموقع و التشتت ضمن متعدد المتغيرات مع خاصية تساوي التباين فضلا عن الانهيار العالي يصبح صعب التنفيذ و يحتاج الى وقت اكبر للحسابات بزيادة الابعاد .

في هذا البحث تم دراسة عينة مؤلفة من (123) شخص مصاب بالامراض الباطنية و الصدرية مع (16) متغير و الاعتماد على المقدرات الحصينة .

- \*Minimum Volume ellipsoid estimator (MVE)
- \* Fast Minimum Covariance determinant estimator (Fast MCD)
- \*Scalable robust estimators with high breakdown Point R .

### Abstract

Robust Factor Analysis is used to analyze a large number of variables to extract new ones whose number is fewer than those factors to which the difference in these variables is attributed . Such a statistical analysis can be used in a broader scope of fields and scientific activities .

Related studies have proved that data without any outliers is but exception within scientific realm . Such a problem can be avoided through using certain suitable robust methods for a model or via primary clearance of data from outliers , and then applying classical statistical methods of factor analysis . The estimation of location and Scatter with multivariate together with affine equivariant and high break down .

Become hard to apply and need much time for calculation as dimensions increase .

In this paper we study asample with (123) person with conditions of the chest and medical diseases and (16) variable deals with three robust estimators .

- i-Minimum volume ellipsoid estimator (MVE) .
- ii-Fast Minimum covariance determinant estimator (Fast MCD) .
- iii-Scalable robust estimators with high break down point R . (R estimators )

## 1- المقدمة Introduction

التحليل العاملي Factor Analysis هو أسلوب من أساليب التحليل الإحصائي متعدد المتغيرات Multivariate Analysis يهدف إلى تقليص عدد المتغيرات المدروسة الأصلية (P) إلى عدد أقل من العوامل (K) يعزى لها تباين تلك المتغيرات .

بداية العمل بهذا التحليل تتم من خلال حساب القيم الذاتية Eigon Values لمصفوفة التباين المشترك Covariance Matrix او مصفوفة الارتباط Correlation Matrix مع العلم ان هذه المصفوفات تكون حساسة وتتأثر نتائجها كثيرا بوجود الشواذ Outliers في مجاميع البيانات , لهذا اصبح الاعتماد على مصفوفة التباين المشترك الحصينة ملزما في الاستخدام في حالة التعامل بالبيانات الملوثة او البيانات غير المتجانسة non homogeneous data (اي ان نقاط البيانات الموجودة في مجموعة معينة لاتتبع كلها نفس التوزيع المفترض) . اكثر من مقدر وجد للتعامل مع بيانات بهذه المواصفات مثل مقدر Minimum Covariqnce (MCD) determinant او مقدر Fast (MCD) او المقدرات rrcov,M,S,Mve,RMCD,GM اضافة الى مقدرات الامكان الاعظم (MLE).

يهدف هذا البحث الى دراسة اهم العوامل التي تؤثر على المرضى المصابين بالامراض الباطنية و الصدرية لمرضى مستشفى الزهراء العام في الكوت (مركز محافظة واسط) و ذلك من خلال استمارة الفحوص المختبرية للكيمياء السريرية لكل مريض .

## 2- الجانب النظري

يفسر التحليل العاملي لـ (P) في المتغيرات لعينة حجمها (n) على اساس دالة خطية لـ (K) من العوامل المشتركة Factors Common حيث (K<P) لاسيما نموذج التحليل العاملي المتعامد و كما يلي :

$$X - m = \wedge f + e \quad \dots\dots\dots(1)$$

حيث

$$X = (X_1, X_2, \dots, X_p)^t \quad \text{متجه المتغيرات}$$

$$m = (m_1, m_2, \dots, m_p)^t \quad \text{متجه الاوساط الحسابية}$$

$$A \in IR^{P \times K} \quad \text{مصفوفة تحميلات العوامل Factor loading وهي من درجة P x k}$$

$$f = (f_1, f_2, \dots, f_k)^t \quad \text{متجه لـ k من الابعاد للعوامل المشتركة}$$

$$e = (e_1, e_2, \dots, e_p)^t \quad \text{متجه الخطأ}$$

المصفوفة  $\wedge$  تكون فقط محددة بمستوى المضاعفات الصحيحة بواسطة المصفوفة المتعامدة Orthogonal التي تكون عواملها غير مترابطة ، اما عندما تكون العوامل مترابطة فتسمى مائلة Oblique .

بعد حساب مصفوفة الارتباط للبيانات و التي تكون من درجة (PxP) يتم حساب التشبعات  $\lambda$  بأستخدام احدى طرائق التحليل العاملي حيث قمنا في هذا البحث بالاعتماد على طريقة العامل الرئيسي (P.F.A) Method Principle Factor ( و التي تبدا اولاً بحساب المتجه الذاتي Eigen vector المقابل الى اكبر قيمة ذات ( $I_1$ ) الذي يمثل مقدار مساهمة العامل الاول  $F_1$  يؤكد نسبة من التباين الاول اي :

$$I_1 = \sum_{i=1}^n a_{i1}^2 = A_1 \dots\dots\dots(2)$$

هذا وقد طور العالم هوتلك طريقة ايجاد قيم و الموجهات الذاتية و ذلك بالاعتماد على اسلوب التالي Iterative Method في استخراجها .

التعامل بالبيانات النظيفة او البيانات المتجانسة يكون مجرد شيء افتراضي في حالات كثيرة و ذلك لان معظم البيانات تحتوي على قيم شاذة او منطرفة تقود الى مجموعة بيانات غير متجانسة يكون معها التعامل بالمقدر التقليدي شيء غير مجدي و يقود الى نتائج خاطئة في معظم الاحيان ،لذا اصبح التعامل مع المقدرات الحصينة Robust Estimators ملزماً في هكذا ظروف .

ولا يتم اكتشاف الشواذ من خلال حساب مسافات مهلونوبس التربيعية Squared Mahalanobis Distances كالاتي :

$$MDi = \{(x_i - t)' C^{-1} (x_i - t)\}^{1/2} \dots\dots\dots(3) \quad i = 1, \dots, n$$

حيث:

$$G \in PSD \{A \in R^{P \times P} | x' Ax > 0 \text{ for all } x \neq 0 \in R^P\} t \in R^P$$

حيث SPD(P) مجموعة كل المصفوفات المتماثلة الموجبة قطعاً Symmetric Positive Definite وان SDi هي المسافة التربيعية بين  $t, x$  مجموعة كل المصفوفات المتماثلة الموجبة قطعاً Symmetric Positive Definite .

بعد اكتشاف الشواذ في البيانات هناك طريقتان للتعامل مع البيانات الاولى تكمن في ازالة المشاهدة الشاذة (وهذا تشويه لطبيعة البيانات) والثانية ان يتم التعامل بالمقدرات الحصينة بدلا من المقدرات العادية وكما قلنا سابقا ان هناك اكثر من طريقة حصينة تكون المفاضلة بينها من خلال تمتعها بمجموعة من الخصائص اهمها نقطة الانهيار العالية اضافة الى خاصية تساوي التباين ،علما ان Croux قد عرف نقطة الانهيار عام (1999) على انها اصغر كسر للمشاهدات و الذي نحتاجه لابدال مواقع اعتباطية قبل تقدير التشتت (اكبر قيمة ذاتية ترسل الى المالانهاية) او (اصغر قيمة ذاتية ترسل الى الصفر) . اما المقدرات التي تتمتع بخاصية تساوي التباين فهي المقدرات التي لاتتغير بالتحويلات التي تجري على الموقع و القياس معا و بالنوع  $X \rightarrow AX + b$  وان ذلك يقود الى جسم قطع ناقص حجمه مساو الى الحجم الاصلي مضروباً في محدد المصفوفة A .

## 1-2 المقدرات الحصينة لمصفوفة التباين المشترك

مما تقدم نرى انه في اغلب حالات التعامل بالبيانات ذات الابعاد العالية تتم خلال اسلوب التحليل العاملي و المعتمد اساسا على اسلوب حساب مصفوفة التباين المشترك لمصفوفة البيانات الاصلية و التي تكون حساسة جدا لنسبة صغيرة من الشواذ في البيانات لذا يتم الاعتماد مع هذه الحالة على المقدرات الحصينة لمصفوفة التباين المشترك بدلا من المصفوفة التقليدية للتباين .

بين Croux عام (2000) ان العوامل الرئيسية الناتجة عن التحليل تكون غير مستقلة عن التشتت حيث حسبت من المتغيرات الاصلية ،لهذا فقد اوصى باستخدام مصفوفة الارتباط لاشتقاق العوامل الرئيسية وفي كلا الحالتين العادية و الحصينة .

### 1-1-2 مقدر اصغر قطع بيضوي Minimum Volume Ellipsoid Estimator

في التحليل الحصين مقدر اصغر قطع بيضوي (MVE) غالبا ما يستخدم لتقدير الموقع و التشتت في متعدد المتغيرات حيث يعرف مقدر (MVE) لمصفوفة التباين المشترك على انه اصغر قطع بيضوي يحوي نصف المشاهدات بينما تقدير الموقع في (MVE) فهو نقطة الوسط لذلك الشكل البيضوي . العمل بخوارزمية مقدر (MVE) يتبع الخطوات التالية :

1-اختيار عدد من العينات الجزئية Subset مساو الى  $C_{p+1}^n$  من مصفوفة البيانات الاصلية و التي تحتوي على (n) من الصفوف (المشاهدات) و (P) من الاعمدة(المتغيرات).

2-حساب قيمة الاوساط الحسابية  $\mu(j)$  و مصفوفة التباين  $\sum(j)$  لكل عينة جزئية (j) بحجم (p+1) .

3-لكل اختيار  $(\sum_{(j)}, m_{(j)})$  اعلاه يتم استخراج المسافات التربيعية وفق المعادلة :

$$D_i^2(J) = (x_i - m_j)^t \sum_{(j)}^{-1} (x_i - m_j)$$

4- ترتيب المسافات التربيعية المستخرجة في الخطوة السابقة  $D_i^2(j)$  ترتيبا تصاعديا و تحديد قيمة  $D_{(hp)(j)}^2$  حيث  $hp = \frac{n+p+1}{2}$  لكل عينة جزئية (j) .

5-اختيار افضل عينة جزئية (j) وهي العينة التي تمتلك اصغر دالة هدف  $\tilde{J} = \arg \min_j \det \sum_{(j)} D_{(hp)(j)}^2$

و تحديد  $\mu, \sum$  لهذه العينة الجزئية .

6-يحسب معامل تصحيح مصفوفة التشتت وهو  $C = 1 + \frac{15}{n-p}$  .

7-تحسب مصفوفة التباين المشترك المضخمة Inflated Covariance Matrix وفق المعادلة ادناه :

$$S(x) = (x_{p,0.5}^2)^{-1} * C_a * \sum_{(j)} D_{(h)(j)}^2 \dots\dots\dots(4)$$

8-تحسب المسافات الحصينة Robust Distance مرة ثانية وفقا للمصفوفة S(x) اعلاه وفق :

$$d_{i(j)}^2 = (x_i - m_{(j)})^t S^{-1}(x)(x_i - m_{(j)}) \dots\dots\dots(5)$$

- و لكافة مشاهدات العينة الاصلية (n) من المشاهدات .
- 9- يضاف مشاهدة واحدة لحجم المجموعة الجزئية ليصبح حجمها (P+2) وعادة ما تكون نفس (P+1) من المشاهدات في المجموعة الاولى مضاف لها مشاهدة جديدة .
- 10- بعد اختيار العينة الجزئية ذات الحجم (P+2) يتم تكرار العمليات 2-9 اعلاه لنحصل على عينة بحجم (P+3) من المشاهدات .
- 11-تستمر سلسلة العمليات التعاقبية في الحساب بزيادة مشاهدة واحدة لكل مرحلة الى ان نصل الى حد التوقف وهو حد الوصول الى حجم العينة الجزئية (np) حيث  $hp = \frac{n+p+1}{2}$  وتكون هذه العينة متجانسة ، منسقة مع بعضها و تتجه نحو المركز . اخيرا من الجدير بالذكر ان نقطة الانهيار لهذا المقدر مساوية الى  $h=n(1-\alpha)$  حيث  $\alpha$  هي نسبة التلوث في البيانات المدروسة .
- 12- للعينة المستخرجة اعلاه ذات (hp) من المستخرجات يجري استخراج متجه الاوساط  $\mu_{(hp)}$  و مصفوفة التشتت  $\sum_{(hp)}$  والتي من خلالها نستخرج المسافات التربيعية الحصينة  $d_{i(hp)}^2$  و تشخص من خلالها المشاهدات الشاذة (وهي المشاهدة التي تزيد قيمة المسافة الحصينة لها عن قيمة  $(c_{r,0.975}^2)$  .

## 2-1-2 مقدار اصغر محددة تباين مشترك

### Minimum Covariance Determent Estimator (MCD)

يمتاز هذا المقدر بحصانة عالية لموقع متعدد المتغيرات و مصفوفة التشتت و هدفه ايجاد المشاهدات h ،حيث  $(h \approx 3n/4)$  و التي مصفوفة التباين المشترك لها تمتلك اصغر محدد .

لقد تم تطوير خوارزمية جديدة لحساب مقدر (MCD) سميت بخوارزمية: Fast-Minimum covariance determintes (Fast MCD) و اساسها وضع الرتب الاحصائية و المحددات Order Statistics & determent لمجموعة المجموعات الجزئية من العينات الصغرى المستخرجة من مجموعة البيانات الاصلية و اساس عملها كما يلي :

- 1- تكوين مجموعات جزئية عشوائية بحجم (P+1) التي تكبر الى (h) من المشاهدات بأستخدام C-Step [1] كما موضح في الخطوة التالية، وهذه الحسابات يمكن اجراؤها في كل المجموعات الجزئية بحجم (p+1) الممكن استخدامها.
- 2- ضمن كل مجموعة جزئية (h) تنفذ C-Step مرتين ،وتشمل على حساب الوسط التقليدي  $(\hat{m}_0)$  و مصفوفة التباين المشترك التقليدية  $\sum_0$  في مشاهدات ال (h) .

تحسب هنا المسافات الحصينة لكل نقطة كما يلي :

$$RD_{\hat{m}_0, \hat{\Sigma}_0}(x_i) = \sqrt{(X_i - \hat{m}_0)' \sum_0^{-1} (X_i - \hat{m}_0)} \dots\dots\dots(6)$$

و بعد الترتيب نحصل على اصغر مسافة حصينة .

3- فيما يخص المجموعة الجزئية (h) العاشرة مع القيمة الافضل لدالة الهدف و استخدام C-Step الى ان نحصل على قيمة واحدة و في بعض الاحيان يمكن استخدام تقنية (1999) , Rousseeuw & Van Driessen .

4- لزيادة كفاءة المقدرات تضاف خطوة اعادة الوزن Reweighted حيث يعطى لكل مشاهدة وزن يعتمد على المسافة الحصينة لها ، فعندما تكون مسافة الحصانة تزيد على حد القطع Cut value وهو  $\sqrt{c_{p,0.975}^2}$  تعطى الوزن (0) والا تكون مساوية (1) .

5- تحسب العوامل الرئيسية (k) من المتجهات الذاتية لـ  $\sum_{MCD}$  وهذه العوامل الرئيسية هي عمود مخزون في مصفوفة التحويلات للعوامل  $P_{p \times k}$  وهذه الخطوة مشابهة الى الخطوة (2) حيث ان تقدير  $x_i$  في قياس المسافة K-dimensional يكون وفق المعادلة التالية :

$$\hat{X}_{i,k} = P_{p,k} P'_{k,p} (X_i - \hat{m}_{MCD}) + \hat{m}_{MCD} \dots\dots\dots(7)$$

اخيرا من الضروري الاشارة الى ان نقطة الانهيار لهذا المقدار مساوية الى  $a = \frac{n-p+1}{2} / n$

### 3-1-2 مقدرات القياس الحصينة مع نقاط الانهيار العالية R [7,8] Scalable Robust Estimators With High Break down Point R

لقد تم تصميم هذا النظام عام (2005) لدعم الحاسبات الاحصائية و ليكون مرجعا للاشخاص المهتمين بموضوع حصانة المقدرات ، سمي هذا المقدار بهذا الاسم نسبة الى الاسماء الاولى من اول حرفين للمؤلفين Ross Ihaka & Robert Gentle man  
الايجازات لبلوغ مقدرات R تكون بالشكل :

1- بعد الضغط على ايقونة النظام تظهر شاشة R2.1.1

2- من الواجهة الاساسية وبعد الضغط على (Pachages) و ظهور الخيارات الستة لها يتم الضغط على الخيار الاول (Load Pachage)

3- تظهر قائمة بالدوال الممكن اختيارها في الحساب متضمن (26) خيار فتختار الخيار (18) تحت اسم (rrcov) المراد تنفيذه

4- الضغط على زر (ok) لتظهر لنا المقدرات الحصينة لمتجه الموقع و مصفوفة التباين المشترك الحصينة و المسماة بمقدرات (rrcov) الحصينة

### 3- الجانب التطبيقي

لغرض تطبيق الهدف من هذا البحث في التحقق من اداء طرائق التقدير الحصينة فقد قمنا بدراسة بيانات حقيقية لعينة من المرضى المصابين بالامراض الباطنية و الصدرية اخذت بياناتهم من مستشفى الزهراء العام في الكوت (مركز محافظة واسط) من خلال استمارة الفحوص المختبرية للكيمياء السريرية لكل مريض ، وقد تضمنت المعلومات التالية :

- 1- الجنس ( ذكر ، انثى ) .
- 2- العمر .
- 3- المهنة ( طفل ، ربة بيت ، موظفة ، كاسب ، متقاعد، مزارع ، عاجز ، طالب ) .
- 4- الحالة الاجتماعية (مطلق، ارملة ، اعزب ، متزوج ) .
- 5- عنوان السكن (مركز المحافظة ، قضاء ، ناحية ، ريف ) .
- 6- التدخين (نعم ، كلا ) .
- 7- الامراض (تسارع ضربات القلب ، عجز القلب ، التهاب القصبات ، جلطة دماغية ، ذبحة صدرية ، ارتفاع ضغط الدم ، جلطة قلبية ، قصور في الدورة الدموية ، داء السكر ، داء الملوك (النقرس) ، عجز الكليتين ) .
- 8- Blood Sugar (F) فحص السكر بالدم ، تشخيص مرض السكري في حاله ارتفاعه عن النسبة الطبيعية (3.6-6.1 mmol/L) .
- 9- Blood Urea فحص اليوريا :تشخيص عجز الكليتين في حاله ارتفاعه عن النسبة الطبيعية (3.3-7.3 mmol/L) .
- 10- S.Creatinine فحص الكرياتينين :لتشخيص عجز الكليتين في حالة ارتفاع Creatinine عن النسبة الطبيعية (62-124 mmol/L) .
- 11- S.Calcium فحص الكالسيوم : تشخيص مرض الكساح لدى الاطفال في حاله انخفاض الكالسيوم عن القيمة الطبيعية (2.1-2.6 mmol/L) وكذلك امراض العظام .
- 12- S.Chlesterol فحص الكولسترول و ارتفاعه عن النسبة الطبيعية (3.9-6.5 mmol/L) الذي يؤدي الى امراض القلب و تصلب الشرايين .
- 13- S.Hdl فحص الدهون ذات الكثافة العالية ، وان انخفاض Hdl عن النسبة الطبيعية (3.9-6.5 mmol/L) يؤدي الى زيادة نسبة الكولستيرول ومن ثم امراض القلب و تصلب الشرايين .
- 14- S.Uric Acid فحص حامض اليورك و ارتفاعه عن النسبة الطبيعية (180-420 mmol/L) و يشخص مرض داء الملوك (النقرس) .
- 15- S.A.L.T(G.P.T) فحص GPT وفي حالة ارتفاعه عن (20U/L) يعني التهاب الكبد .
- 16- S.Iron فحص الحديد وله علاقة بفقر الدم حيث ان النسبة الطبيعية هي (13-32 mmol/L) .



حيث تم استخدام التحليل العاملي العادي (الكلاسيكي ، التقليدي ) ثم تطبيق الطرق الحصينة على مجموعة البيانات { (123) مشاهدة مع (6) متغير } حيث يتم تحديد الشواذ و استبعادها ثم استخدام طريقة العامل الرئيسي للبيانات النظيفة في التحليل العاملي .

علما ان البرامج المستخدمة في هذا البحث تمت وفق :

- 1- لغة Fortran Power Station 4.0 .
- 2- Mat lab ver (7) .
- 3- Spss ver (10.0) for windows .
- 4- Mini tab .
- 5- Coro IDRAW Ver (8.0) .
- 6- Microsoft Excel .
- 7- R.2.1.1 .

### 1-3 مناقشة نتائج التحليل

ان الهدف من هذا التطبيق هو تحليل البيانات الحقيقية لعينة سحبت من البيانات الخاصة بمرضى الامراض الصدرية و الباطنية ، وان مثل تلك البيانات تكون بعض قيمها او التسجيلات الخاصة بمتغيراتها متجاوزة النسبة الطبيعية فتعد شاذة في مثل هذه الحالة .

عند تطبيق التحليل العاملي على مجموعة البيانات بالابعاد (16\*123) تبين وجود سبعة عوامل مشبعة Factors loading و تطبيق مسافات مهلو نوبس Mahalanobis distance اتضح وجود (25) مشاهدة شاذة ضمن مجموعة البيانات هذه و بأستبعادها و تطبيق الاساليب الاحصائية ( MVE , rrcov , Fast MCD ) اصبح البيانات بعد الحذف غير ملوثة (نظيفة) و بالابعاد الموضحة في الجدول (1) ادناه :

#### جدول رقم (1)

احجام العينات المتعامل بها مع كل مقدار

عدد المتغيرات P	عدد المشاهدات N	الطرائق الحصينة
16	123	الطريقة العادية
16	98	MVE
16	98	Fast MCD
16	106	Rrcov



وعند تطبيق التحليل العاملي العادي اظهرت النتائج وجود سبعة عوامل لكل من الطريقة العادية و مقدر (MVE) بالاضافة لمقدر (Fast MCD) و ستة عوامل للمقدر (ITCOV). و فيما يأتي مناقشة نتائج التحليل العاملي لكل طريقة على حدة

جدول رقم (2)

عدد العوامل المحملة (Factors Loading)

MVE n=98 p=16							
Factors							
7	6	5	4	3	2	1	Var
-0.136	0.112	-0.181	7.51E.02	0.797	5.04E.02	-0.254	X <sub>1</sub>
-0.271	0.172	-0.619	3.46E.02	-0.432	-3.69E.02	-0.335	X <sub>2</sub>
-0.169	-4.90E.02	-1.43E.02	7.14E.02	-0.838	-3.28E.02	-0.132	X <sub>3</sub>
-0.311	-0.807	0.26	-2.83E.02	.488E.02	6.44E.02	-9.53E.02	X <sub>4</sub>
0.488	-7.57E.02	0.18	0.492	4.57E.02	6.75E.02	-0.244	X <sub>5</sub>
-3.51E.02	5.81E.02	0.789	4.25E.03	-8.16E.02	-4.78E.02	-0.233	X <sub>6</sub>
5.41E.02	4.71E.02	0.154	-3.29E.02	0.819	-9.36E.02	-9.69E.02	X <sub>7</sub>
4.57E.02	2.51E.02	-4.29E.02	2.96E.02-	-0.139	2.28E.02	0.948	X <sub>8</sub>
-0.116	-2.89E.03	-7.44E.02	0.868	-5.39E.02	-8.39E.03	0.147	X <sub>9</sub>
-0.16	-4.40E.02	2.07E.02	0.519	-6.13E.02	0.531	0.507	X <sub>10</sub>
2.90E.03	6.33E.02	5.66E.02	-0.223	1.04E.02	0.904	0.208	X <sub>11</sub>
8.31E.02	-6.42E.02	3.99E.02	4.70E.02	3.58E.02	0.705	-0.482	X <sub>12</sub>
-0.149	0.804	0.341	-0.103	4.55E.02	8.41E.02	-0.118	X <sub>13</sub>
0.108	-1.56E.02	-8.47E.02-	8.12E.02	-5.37E.02	-0.202	0.879	X <sub>14</sub>
9.28E.02	-5.41E.02	9.88E.02	0.175	-1.62E.02	0.882	-0.206	X <sub>15</sub>
0.813	9.86E.02	4.32E.02	-0.116	5.08E.02	0.11	0.14	X <sub>16</sub>

Classical n=123 p=16							
Factors							
7	6	5	4	3	2	1	Var
-0.113	4.29E.02	-0.197	7.82E.02	0.82	0.01E.02	-0.272	X <sub>1</sub>
-0.251	7.44E.02	-0.706	3.74E.02	-0.366	-2.44E.02	-0.294	X <sub>2</sub>
-0.128	-1.68E.02	-0.152	8.27E.02	-0.838	-5.81E.02	-0.77E.02	X <sub>3</sub>
-0.348	-0.745	0.288	-2.66E.02	-7.75E.02	0.163	-0.16	X <sub>4</sub>
0.474	-0.162	0.12	0.516	5.48E.02	1.03E.02	-0.245	X <sub>5</sub>
-4.18E.02	0.11	0.727	-2.39E.04	-2.95E.02	-3.40555E.02	-0.265	X <sub>6</sub>
5.41E.02	4.71E.02	0.154	-3.29E.02	0.819	-9.36E.02	-9.69E.02	X <sub>7</sub>
4.57E.02	2.51E.02	-4.29E.02	2.96E.02-	-0.139	2.28E.02	0.948	X <sub>8</sub>
-0.116	-2.89E.03	-7.44E.02	0.868	-5.39E.02	-8.39E.03	0.147	X <sub>9</sub>
-0.16	-4.40E.02	2.07E.02	0.519	-6.13E.02	0.531	0.507	X <sub>10</sub>
2.90E.03	6.33E.02	5.66E.02	-0.223	1.04E.02	0.904	0.208	X <sub>11</sub>
8.31E.03	-6.42E.02	3.99E.02	4.70E.02	3.58E.02	0.705	0.482	X <sub>12</sub>
-0.149	0.804	0.341	-0.103	4.55E.02	8.41E.02	-0.118	X <sub>13</sub>
0.108	-1.556E.02	-8.47E.02-	8.12E.02	-5.37E.02	-0.202	0.879	X <sub>14</sub>
9.28E.02	-6.41E.02	9.88E.02	0.175	-1.62E.02	0.882	-0.206	X <sub>15</sub>
0.813	9.86E.02	4.32E.02	-0.116	5.08E.02	0.11	0.14	X <sub>16</sub>

Fast MCD n=98 p=16							
Factors							
7	6	5	4	3	2	1	Var
-0.113	4.29E.02	-0.197	7.82E.02	0.82	5.01E.02	-0.272	X <sub>1</sub>
-0.251	7.44E.02	-0.706	3.74E.02	-0.366	-2.44E.02	-0.294	X <sub>2</sub>
-0.126	-1.68E.02	-0.152	8.27E.02	-0.838	-5.81E.02	-9.77E.02	X <sub>3</sub>

-0.348	-0.745	0.288	-2.66E.02	-7.75E.02	0.163	-0.15	X <sub>4</sub>
0.474	-0.152	0.12	0.516	5.48E.02	-1.83E.02	-0.245	X <sub>5</sub>
-4.18E.02	0.11	0.727	-2.39E.02	-2.95E.02	-3.45E.02	-0.265	X <sub>6</sub>
5.41E.02	4.71E.02	0.154	-3.29E.02	0.819	-9.36E.02	-9.69E.02	X <sub>7</sub>
4.57E.02	2.51E.02	-4.29E.02	2.96E.02-	-0.139	2.28E.02	0.948	X <sub>8</sub>
-0.116	-2.89E.03	-7.44E.02	0.868	-5.39E.02	-8.39E.03	0.147	X <sub>9</sub>
-0.16	-4.40E.02	2.07E.02	0.519	-6.13E.02	0.531	0.507	X <sub>10</sub>
2.90E.03	6.33E.02	5.66E.02	-0.223	1.04E.02	0.904	0.208	X <sub>11</sub>
8.31E.02	-6.42E.02	3.99E.02	4.70E.02	3.58E.02	0.705	0.482	X <sub>12</sub>
-0.149	0.804	0.341	-0.103	4.55E.02	8.41E.02	-0.118	X <sub>13</sub>
0.108	-1.56E.02	-8.47E.02-	8.12E.02	-5.37E.02	-0.202	0.879	X <sub>14</sub>
9.28E.02	-5.41E.02	-9.88E.02	0.175	-1.62E.02	0.882	-0.206	X <sub>15</sub>
0.813	9.86E.02	4.32E.02	-0.116	5.08E.02	0.11	0.14	X <sub>16</sub>

rrcov n=105 p=16						
Factors						
6	5	4	3	2	1	Var
9.15E.02	-0.17	5.99E.02	0.824	4.05E.02	-0.314	X <sub>1</sub>
0.138	-0.735	1.97E.02	-0.302	-6.87E.02	-0.316	X <sub>2</sub>
-6.34E.02	-0.262	3.62E.02	-0.822	-6.94E.02	-8.35E.02	X <sub>3</sub>
-0.867	7.41E.02	-9.15E.02	-7.18E.02	0.196	-0.126	X <sub>4</sub>
4.70E.02	0.274	0.613	1.75E.02	3.24E.02	-0.187	X <sub>5</sub>
-1.24E.02	0.604	-9.31E.02	-2.86E.02	-5.91E.02	-0.234	X <sub>6</sub>
8.73E.04	8.27E.02	-2.01E.02	0.87	-4.47E.02	-5.63E.02	X <sub>7</sub>
3.39E.02	2.47E.02	-4.74E.02	-0.13.	0.111	0.948	X <sub>8</sub>
-0.105	-0.182	0.772	-1.12E.02	3.67E.02	0.212	X <sub>9</sub>
-2.49E.02	-7.75E.02	0.333	-3.97E.02	0.736	0.533	X <sub>10</sub>
2.21E.02	6.83E.02	-0.31	6.26E.02	0.902	8.24E.02	X <sub>11</sub>
-0.141	0.186	6.97E.02	6.41E.02	0.584	-0.582	X <sub>12</sub>
0.631	-3.91E.03	-0.399	6.33E.02	0.122	-8.89E.02	X <sub>13</sub>
1.39E.02	-2.91E.02	0.107	-5.99E.02	-0.177	0.924	X <sub>14</sub>
-7.73E.02	-2.85E.02	0.213	-9.47E.02	0.846	-0.371	X <sub>15</sub>
0.316	0.529	0.201	3.52E02	8.54E.02	7.04E.02	X <sub>16</sub>

### 1-1-3 مناقشة نتائج التحليل العاملي العادي Classical

عند تطبيق التحليل العاملي العادي على مجموعة البيانات (16\*123) ظهر ان هناك سبعة عوامل رئيسية تفسر ما نسبته 78.645% من اجمالي التباين و التي تؤثر في الامراض الباطنية و الصدرية التي مثلتها القيم المميزة Eigen values التي تكون قيمتها اكبر من الواحد الصحيح و هي :

#### 1- العامل الاول

و يفسر (16.487%) من اجمالي التباين كما موضح في الجدول (3) ومن خلال متابعة متغيرات هذا العامل نلاحظ ظهور اربعة متغيرات معنوية (انظر جدول (2)) وهي المتغير الثامن ، الرابع عشر ، العاشر ، و المتغير الثاني عشر ومنه نلاحظ ان هذا العامل يهتم بتشخيص الامراض ذات العلاقة بالموضوع ، و منه تتوضح اهمية التشخيص المبكر للامراض و علاقته بمرحلة العلاج اللاحقه .

## 2- العامل الثاني

ويأتي في المرتبة الثانية من الأهمية حيث يفسر (15.489%) من إجمالي التباين أما متغيراته فهي المتغير الحادي عشر ، الخامس عشر ، الثاني عشر و المتغير العاشر وهي متغيرات امراض العظام و التهاب الكبد و امراض القلب و تصلب الشرايين و عجز الكليتين .

## 3- العامل الثالث

ويفسر (13.912%) من إجمالي التباين (جدول رقم (3)) ووجد ان هذا العامل له متغيرين فقط (جدول (4)) هما المتغيرين الاول و السابع ما يبين ان هناك علاقة بين الجنس و الاصابة بالامراض الصدرية و الباطنية .

## 4- العامل الرابع

يفسر العامل الرابع (8.867%) من إجمالي التباين و متغيراته هي التاسع ، العاشر ، و الخامس وهي متغيرات تخص امراض الكلية وما يرافقها من امراض اخرى .

## 5- العامل الخامس

و هو عامل التدخين و يفسر (8.478%) من إجمالي التباين وقد اظهر هذا العامل متغير واحد وهو المتغير السادس الذي يبين ان كان المريض مدخنا ام لا ، ومنه نرى الدور السيء الذي يلعبه التدخين في الاصابة بالامراض الباطنية و الصدرية .

## 6- العامل السادس

ويفسر (7.948%) من إجمالي التباين وقد اظهر متغيرا واحد فقط هو المتغير الثالث عشر الذي يشخص مرض السكري . العامل السادس بمتغيره يؤكد الاحصاءات و البيانات التي قامت بها وزارة الصحة من حيث انتشار هذا الداء ولاسيما بعد الحروب التي خاضها و تعرض له بلدنا العزيز وما وافقها من احداث .

## 7- العامل السابع

وهو العامل الاخير و يفسر (7.464%) من إجمالي التباين وقد اظهر متغيرين فقط هما السادس عشر و الخامس ، حيث يوضح هذا العمل ان الاصابة بمرض فقر الدم (الانيميا) تقود في معظم الاحيان الى الاصابة بالامراض الباطنية و الصدرية الاخرى .

جدول رقم (3)

التباين المفسر للبيانات الكلية (n=123 ,P=16)

Total Variance explained Rotation sums of Squared loadings			Extraction sums of Squared loadings			Initial Eigenvalues			Comp onent
Cumulativ e%	% of variance	Total	Cumulative %	% of variance	Total	Cumulative %	% of variance	Total	
16.487	16.487	2.638	18.701	18.701	2.992	18.701	18.701	2.992	1
31.977	15.489	2.478	34.481	15.781	2.525	34.481	15.781	2.525	2
45.889	13.912	2.226	48.310	13.828	2.212	48.310	13.828	2.212	3
54.755	8.867	1.419	57.593	9.284	1.485	57.593	9.284	1.485	4
63.233	8.478	1.357	65.120	7.526	1.204	65.120	7.526	1.204	5
71.182	7.948	1.272	72.176	7.056	1.129	72.176	7.056	1.129	6
78.645	7.464	1.194	78.645	6.470	1.035	78.645	6.470	1.035	7
						83.799	5.154	.825	8
						88.258	4.459	.713	9
						91.837	3.579	.573	10
						94.612	2.775	.444	11
						97.001	2.389	.382	12
						98.294	1.293	.207	13
						99.215	.921	.147	14
						99.686	.471	7.537E.02	15
						100.00	.314	5.031E.02	16

Extraction method: principal component Analysis.

**2-1-3 مناقشة نتائج التحليل العاملي لمقدر (Fast MCD) و مقدر (MVE)**

بعد عملية التحليل العادي التي تمت على مجموعة البيانات ، قمنا في هذه المرحلة بأجراء التحليل الحصين مع المقدر (MVE) ثم المقدر (Fast MCD) حيث اتفق هذان المتغيران في نتائجهما ، فبداية اظهر هذان المقدران وجود (25) مشاهدة شاذة ضمن مجموعة البيانات وقد تم استبعادها ليتم التعامل فقط مع (98) مشاهدة (انظر الشكل (1)) عند العمل ضمن خوارزميات هذين المقدرين اتضح وجود سبعة عوامل رئيسية تفسر ما نسبته (79.090%) من اجمالي التباين (انظر جدول (2)).

العامل الاول يفسر (16.431%) من اجمالي التباين وقد انصب غلى الامراض (السكري ، داء الملوك ، عجز الكلوتين ) (جدول (4)).

جدول رقم (4)

التباين المفسر للبيانات (n=98 , p=16) لمقدر (MVE & Fast MCD)

Total Variance explained Rotation sums of Squared loadings			Extraction sums of Squared loadings			Initial Eigenvalues			Comp onent
Cumulativ e%	% of variance	Total	Cumulative e%	% of variance	Total	Cumulative %	% of variance	Total	
16.431	16.431	2.629	18.005	18.005	2.881	18.005	18.005	2.881	1
31.097	14.666	2.347	33.476	15.471	2.475	33.476	15.471	2.475	2
45.140	14.044	2.247	47.864	14.388	2.302	47.864	14.388	2.302	3
54.543	9.403	1.504	57.567	9.703	1.552	57.567	9.703	1.552	4
63.165	8.622	1.380	65.778	8.211	1.314	65.778	8.211	1.314	5
71.388	8.223	1.316	72.736	6.959	1.113	72.736	6.959	1.113	6
79.090	7.701	1.232	79.090	6.353	1.017	79.090	6.353	1.017	7
						84.439	5.350	.856	8
						88.943	4.504	.721	9
						92.517	3.574	.572	10
						95.109	2.591	.415	11
						97.313	2.204	.353	12
						98.561	1.248	.200	13
						99.458	.897	.144	14
						99.816	.358	5.729E.02	15
						100.000	.184	2.938R.02	16

Extraction method: principal component Analysis.

اما العامل الثاني فيفسر (14.666%) من اجمالي التباين (جدول (4)) وقد مثل هذا العامل ثلاثة متغيرات مهمة هي (التهاب الكبد ، الكساح ، امراض القلب وتصلب الشرايين) . العامل الثالث فسر (14.044%) من اجمالي التباين وله متغيران رئيسيان هما الجنس و الامراض الباطنية و الصدرية .

(9.403%) من اجمالي التباين فسرها العامل الرابع وقد شخص امراض الكليتين اضافة الى منطقة سكن المريض كونها مركز المدينة اوريف ، و كما هو معلوم يعتمد الاشخاص في الريف على مياه السواقي و الانهر في اغلب الاحيان ما يؤدي الى الاصابة بأمراض الكلية .

العامل الخامس فسر (8.622%) من اجمالي التباين وهو عامل التدخين وهي نتيجة تدعم ما نعلمه من وجود علاقة قوية بين التدخين و الاصابة بالامراض الصدرية و امراض الرئة .

العامل السادس يفسر (8.223%) من اجمالي التباين وله متغيران فقط هما الحالة الاجتماعية للمريض اضافة الى فحص الدهون له .

في المرتبة الاخيرة جاء العامل السابع ليفسر (7.701%) من اجمالي التباين وليهتم بمتغيرين تؤثران على الاصابة بالامراض الصدرية و الباطنية هما متغير منطقة السكن اضافة الاصابة بفقر الدم .

### 3-1-3 مناقشة نتائج التحليل العاملي لمقدر rrcov

بعد ان اجري في المبحث السابق التحليل الحصين من خلال المقدرين (MVE) و (Fast MCD) فقد تم هنا اجراء التحليل الحصين لمقدر (rrcov) الذي اوضح بداية وجود (17) مشاهدة شاذة في مجموعة البيانات (شكل 1) من خلال اجراء التحليل تبين وجود ستة عوامل تفسر ما نسبته (74.508%) من اجمالي التباين ، كما موضح في الجدول (5) .

#### جدول (5)

التباين المفسر للبيانات (n=106 , p=16) لمقدر (rrcov)

Total Variance explained Rotation sums of Squared loadings			Extraction sums of Squared loadings			Initial Eigenvalues			Comp onent
Cumulativ e%	% of variance	Total	Cumulative %	% of variance	Total	Cumulative %	% of variance	Total	
18.089	18.089	2.894	19.993	19.993	3.199	19.993	19.993	3.199	1
33.939	15.805	2.536	35.711	15.718	2.515	35.711	15.718	2.515	2
47.972	14.033	2.245	50.295	14.583	2.333	50.295	14.583	2.333	3
57.124	9.152	1.464	59.899	9.604	1.537	59.899	9.604	1.537	4
66.237	9.113	1.458	67.467	7.568	1.211	67.467	7.568	1.211	5
74.508	8.271	1.323	74.508	7.041	1.127	74.508	7.041	1.127	6
						80.606	6.098	.976	7
						86.131	5.526	.884	8
						90.502	4.370	.699	9
						93.455	2.953	.473	10
						95.947	2.492	.399	11
						98.341	2.394	.383	12
						99.339	.998	.160	13
						99.752	.413	6.606E.02	14
						99.964	.211	3.381E.02	15
						100.000	3.630E.02	5.808E.03	16

Extraction method: principal component Analysis.

حيث فسر العامل الاول (18.089%) من اجمالي التباين وقد اهتم بثلاثة متغيرات [جدول(2)] هي الثامن و الرابع عشر و العاشر ومنه نرى ان مقدر rrcov قد اتفق مع المقدرات السابقة من حيث متغيرات العامل الاول و هي الاصابة بالامراض (السكري ، داء الملوك ، عجز الكلبيين ) وهي كما معروف بالفعل من الاهمية بحيث ترتب كمتغيرات للعامل الرئيسي الاول الخاص بالامراض الصدرية و الباطنية .  
فسر العامل الثاني (15.850%) من اجمالي التباين الكلي للمتغيرات وقد ادرج تحته اربع متغيرات هي الحادي عشر ، الخامس عشر ، العاشر و الثاني عشر وهي امراض العظام و التهاب الكبد و عجز الكلبيين و امراض القلب .

حل متغير الجنس المرتبة الاولى ضمن العامل الثالث الذي فسر (14.033%) من اجمالي التباين يليه متغير الامراض الصدرية و الباطنية و بعدها جاء العامل الرابع ليفسر (9.152%) من اجمالي التباين و الذي اعتمد على متغيري عجز الكليتين و منطقة سكن المريض ليفوق ذلك مع المقدرات الحصينة التي سبقته جدول (2) .

العامل الخامس فسر (9.113%) من اجمالي التباين وقد اظهر هذا العامل متغيرين رئيسيين هما السادس و السادس عشر و هما متغيري التدخين و الاصابة بمرض فقر الدم يليه اخيرا العامل السادس الذي يفسر (8.271%) من اجمالي التباين و متغيريه هما الرابع و الثالث عشر و هما الحالة الاجتماعية للمريض مع فحص الدهون و من ثم امراض القلب و تصلب الشرايين [ جدول (5) ] .

مما تقدم نرى ان الطرق الحصينة الثلاثة التي استخدمت كانت النتائج لها مقاربة فيما يخص عدد المشاهدات الشاذة المكتشفة و التي استبعدت في المرحلة الثانية ثم اعداد العوامل المؤثرة على العلاقة بين المتغيرات المدروسة ثم متغيرات هذه العوامل و هذا ان دل على شيء فأنما يدل على التوافق العالي بين المقدرات الحصينة المستخدمة (rrcov , Fast MCD , MVE) مع مجموعة البيانات التي اجري التحليل لها فكما نعلم ان لكل مجموعة من البيانات المدروسة يوجد مقدرات حصينة تصلح للتعامل معها و ذلك حسب نسبة الشواذ في مجموعة البيانات ام حجم البيانات بحد ذاته وصولا الى نسبة عدد المتغيرات الى عدد المشاهدات .

#### 4- النتائج :

من دراسة هذا البحث ممكن التوصل الى جملة من النتائج اهمها :

- 1- السلوك الجيد للمقدر الحصين في حالة التعامل بالبيانات الشاذة و التي تقود الى نتائج يمكن الوثوق بها في هكذا حالة .
- 2- اثبت المقدر rrcov سهولة مع سرعة في الانجاز ، كونه برنامجا واحدا مع مجموعة برامج حصينة ضمن ملف جاهز يمكن التعامل معه في اغلب الاحيان ، لاسيما في حالة التعامل بالبيانات ذات الابعاد العالية كذلك فأن مقدر (MCD) وفق خوارزميته المعروفة يكون كبرنامج متاح في المجاميع الاحصائية القياسية (Becker and Gather 2001)
- 3- ان جميع المقدرات الحصينة التي تم التعامل بها في هذا البحث (rrcov, Fast MCD , MVE) هي مقدرات ذات خصائص جيدة من حيث الانهيار العالي وصولا الى تساوي التباين .
- 4- ان نتائج التحليل العاملي بالطريقة العادية هي نفسها بطريقة مقدر (Fast MCD) على الرغم من كون عدد المشاهدات بالطريقة التقليدية (n=1232) و عدد المشاهدات بطريقة المقدر (Fast MCD) (n=98) حيث تم استبعاد خمس و عشرين مشاهدة ملوثة . كذلك نلاحظ ان هذا المقدر استبعد اكبر عدد من المشاهدات مقارنة بالمقدرات (MVE , rrcov) .



#### 5- التوصيات :

من النتائج التي تم الوصول إليها و من مطالعتنا للبحوث الاحصائية العديدة و المتعلقة بجانب الحصانة للمقدرات ، نوصي ان يتم اعتماد اكثر من متغير حصين واحد لكل مجموعة بيانات ليتسنا لنا المفاضلة بين تلك المقدرات سيما وانها ومن خلال نقطة الانهيار لها تكون معتمدة على نسبة التلويث في مجموعة البيانات اضافة الى نوعية التلويث فقد يكون تلويث البيانات بنفس التوزيع المتبع في توليد البيانات الاصلية ولكن بمعالم مختلفة حيث ان الاستقلالية بين المشاهدات تضمن لنا مثل هذه الحالة .

كذلك نوصي ان تكون هناك دراسات مستقبلية حول مقدرات حصينة اخرى كالمقدر (S) و مقدر (Average MVE) ذوات الكفاءة العالية اضافة الى خصائص الحصانة الجيدة .

في الجانب الطبي و من خلال ما توصلنا اليه في تأثير التدخين على الاصابة بالامراض الصدرية نوصي بالابتعاد قدر الامكان عن التدخين و المدخنين على حد سواء ، اضافة الى عدم الاعتماد على مياه السواقي و الانهر في الشرب وان كانت الحاجة ملحة الى ذلك فتكون بغلي الماء جيدا قبل شربه . و اخيرا وللاصحاء و المرضى نوصي بأجراء فحص دوري شامل للجسم على الاقل مرة واحدة في السنة لان التشخيص المبكر للامراض يقي المريض التفاعلات التي قد تصاحب المرض ان ظل بدون علاج كفانا الله و اياكم شر جميع الامراض انه سميع مجيب .

#### المصادر

- 1- ناسي ، نبيل جورج (2001) "طرق تقدير القيم الشاذة لنموذج الانحدار " اطروحة دكتوراه ، كلية الادارة و الاقتصاد ، جامعة بغداد
2. جويد، افراح كاظم (2007) " استخدام التحليل العاملي في المجال الطبي " اطروحة ماجستير، كلية الادارة والاقتصاد ، جامعة بغداد.
- 3-Groux , C .& Haesbroeck , G.(1999) , "Influence function and efficiency of the Minimum Covariance determinany Scatter Matrix estimator " , journal of multivariate analsis , 71 , 161 -190 .
- 4-Croux , C.& Haesbroeck , G.(2000) , "Principal Component Analysis Based on Robust estimators of Covariance or correlation Matrix : Influence Function and efficiencies " Biometrika ,87, 603 -618 .
- 5-Groux , c. & Dehon , C.(2001) " Robust Linear discriminant analysis Using S-estimation " the Canadion Journal of statistics , vol .29 , 473-492 .
- 6-Matias salibion–barrara ,(2004) "Afast algorithm for S-regression estimates , "University of British columbia and victor J.Yohai , Univer sided Buenos Aires october 14.
- 7-Mia Hubert , sanne Engelen (2004) . " Fast cross-vali dation of high-breakdown resampling methods for PCA " .
- 8-.Rousseeuw , P.J & Leroy , A.M. (1987) ." Robust regression and outlier detection " , wiley , P.79.table 2.
- 9-Ross Ihaka , Robert Gentlman (2005) . " Scalable robust estimators with high breakdown Point R " . ISBN 3 -900051-07-0 . http :// cran . r – project .org.