إستخدام بعض الطرائق التمييزية الحصينة

لتشخيص أمراض القلب في البصرة

م.م. / دريد حسين بدر
جامعة البصرة / كلية الإدارة والاقتصاد
قسم الاحصاء

الملخص

في هذا البحث تمت محاولة إيجاد مقدرات حصينة وكفوءة لموجّـه المتوسط (معلمة الموقع) ومصفوفة التباين والتباين المشترك (معلمة القياس) لدوال التمييز الخطية والتربيعية والى إيجاد دوال تمييز خطية وتربيعية حصينة في حالة وجود قيم شاذة باستخدام بعض طرائق التقدير الحصينة. ومن هذه الطرق طريقة مقدر أصغر محدد تباين مشترك وطريقة مقدر أصغر محدد تباين مشترك معاد الأوزان وطريقة مقدر أصغر حجم مجسم قطع بيضوي، ومقارنة دوال التمييز الحصينة مع دوال التمييز التقليدية للوصول إلى أفضل دالة تمييزية وذلك لتشخيص نوعين من أمراض القلب ، مرض الذبحة الصدرية ومرض الجلطة القلبية .

Abstract

In this research I've tired to find out to robust and efficient estimations mean vector (location parameter) and, variance covariance matrix (scull parameter) for linear and quadratic discriminant functions. And also attempting to provide the robust linear and quadratic discriminant functions in the case of existing outliers resents through using some of high quality robust methods estimation. of these methods are minimum covariance determinant estimator, reweighed minimum covariance determinant estimator and minimum volume ellipsoid estimator , moreover , comparing robust discriminant functions with classical discriminant functions to draw out the best discriminant function in order to diagnose two types of diagnosis diseases heart which are angina pectoris and clot hearty .

المقدمة

إن التحليل التمييزي هو تحليل إحصائي يتم بموجبه استخدام مجموعة من المتغّيرات للتمييز بين مجموعتين أو أكثر عن طريق دالة تمييزية محددة وهي توليفة خطـّية للمتغّيرات التوضيحية وطريقة إيجاد هذه الدالة هو بإيجاد المعاملات للدالة وفقاً للقياسات أو المعايير التي يتم الحصول عليها من المشاهدات.

وفي قواعد طرائق التمييز التقليدية غالباً ما يتم الاعتماد على موجـّه المتوسطات التجريبي ومصفوفة التباين والتباين المشترك للبيانات أو جزء من البيانات والنتيجة تكون مقدرات تتأثر بشكلٍ عالٍ بالمشاهدات الشاذة فمقدرات المربعات الصغرى تتأثر كثيراً بوجود عدد قليل من القيم الشاذة، والنموذج المقدر سوف يعكس خصائص غير طبيعية، ومع تطور الأساليب الحصينة التي تزودنا بطرقٍ جديدة للتعامل مع البيانات حيث تزودنا هذه الأساليب بمقدرات تمتاز بكونها قليلة التأثر بالمشاهدات الشاذة، لذا ظهرت الحاجـّة حول معالجة مشكلة اختراق البيانات لشروط التوزيع الطبيعي من خلال إيجاد عدد من الطرائق الحصينة إذ تتصف بأن تقديراتها تكون قريبة من تقديرات المربعات الصغرى عند عدم وجود الشواذ وتكون أفضل عند وجود القيم الشاذة .

مشكلة البحث :

تعد أمراض القلب من الأمراض الخطرة على المصابين ، وتسبب للمرضى عدم إمكانية القيام بالأعمال الإعتيادية، التي تكفل لهم العيش مما يتسبب بخسائرٍ مادية كبيرة للمرضى ،و للبلد خاصة،وبالتالي فان المجالات الطبية مليئة ببيانات تتصف بأن بعض قياساتها أو التسجيلات الخاصة بمتغّيراتها تكون متجاوزة لنسبها الطبيعية داخل جسم الإنسان، وهذه القياسات تعد حينئذ شاذةلأنها تحقق الغرض من هذه الدراسة، ونتيجةً لخطورة هذين المرضين (مرض الذبحة الصدرية ومرض الجلطة القلبية) فقد تم تناولهما في البحث باستخدام عدد من طرائق التقدير الحصينة عند وجود الشواذ.

هدف البحث

يهدف البحث إلى إيجاد مقدرات حصينة وكفوءة لموجـّه المتوسط (معلمة الموقع) ، ومصفوفة التشتت (معلمة القياس) ،لدوال التمييز الخطية والتربيعية، والى إيجاد دوال تمييز خطية وتربيعية حصينة في حالة وجود قيم شاذة (Outliers) باستخدام عدد من طرائق التقدير الحصينة ذات التقنية العالية لما لها من دورٍ كبيرٍ في التحصين ضد وجود الشواذ في البيانات. ومن هذه الطرق طريقة مقدر أصغر محدد تباين مشترك، وطريقة مقدر أصغر محدد تباين مشترك معاد الأوزان، وطريقة مقدر أصغر حجم مجسم قطع بيضوي، ومقارنة دوال التمييز الحصينة مع دوال التمييز التقليدية للوصول إلى أفضل دالة تمييزية وذلك لتشخيص أمراض القلب والحصول على دالةِ تمييزٍ بين نوعين من إمراض القلب، مرض الذبحة الصدرية (A.P) ومرض الجلطة القلبية (C.H) .

٢. الجانب النظري :

١.٢ طرائق التقدير الحصينة:

إن المفهوم الإحصائي للحصانة(Robustness) كما وصفه (Sarhan و Box) فالطرائق الحصينة ينبغي أن تكون ذات كفاية قريبة من طرائق التقدير التقليدية في حالة تحقق الافتراضات، وأفضل منها في حالة الانحراف عن تلك المشاهدات.ومن طرائق التقدير الحصينة لمعلمتي الموقع، والقياس التي ساتناولها في دراستي هذه هي طريقة مقدر أصغر محدد تباين مشترك (MCD) ، وطريقة مقدر أصغر محدد تباين مشترك معاد الأوزان(RMCD)، وطريقة مقدر أصغر حجم مجسم قطع ناقص (MVE). التي تستخدم لتقدير موجـّه المتوسط $\underline{\mu}$ ومصفوفة التباين والتباين المشترك Σ . (Hubert,Driessen,2004)

١.١.٢ طريقة مقدّر أصغر محدد تباين مشترك (MCD) :

وهي إحدى طرائق التقدير الحصينة لمعلمتي الموقع والقياس، وتستخدم لتقدير موجـّه المتوسط ($\underline{\mu}$) ومصفوفة التباين، والتباين المشترك (Σ)،في حالة

البيانات تتبع التوزيع الطبيعي متعدد المتغيّرات، وتقوم فكرة هذه الطريقة على اساس تقدير موجـّه المتوسط ، ومصفوفة التباين، والتباين المشترك لعينات جزئية بحجم (h) من المشاهدات الكلية، وحساب المحددات لتلك المصفوفات فالمصفوفة التي تمتلك اصغرمحدد تعتبر هي افضل مقدر من بين المقدرات الاخرى لمصفوقة التباين والتباين المشترك للمجموعة ويطلق عليها بمقدر أصغر محدد تباين مشترك الحصين (MCD). (Rousseeuw, 1985)

ويتم اختيار قيمة لـ(h) تساوي(P+1) لأنها تضمن الحصول على مقدرات حصينة تمتلك أعلى نقطة انهيار (Breakdown Point) تصل إلى (50%).

وإن عدد العينات الجزئية يكون مساوٍ إلى:

$$H = C^{n}_{P+1} \qquad \textbf{....(2-1)}$$

وان مقدرات أصغر محدد تباين مشترك (MCD) تتمتع بخاصية ثبات التغاير (Affine Equivariant) وتكون التقديرات كالآتي: (Hubert,Driessen,2004)

$$\underline{\overline{X}}_{MCD} = \frac{1}{h}\sum_{i=1}^{h}\underline{x}_i \qquad \text{... (2-2}$$

$$\hat{\Sigma}_{MCD} = dp \times \frac{1}{h}\sum_{i=1}^{h}(\underline{x}_i - \underline{\overline{X}}_{MCD})(\underline{x}_i - \underline{\overline{X}}_{MCD})' \qquad \text{... (٣ -٢}$$

إذ إن: dp :عامل الاتساق المحدد ويحسب كالآتي: (Pison, Van Aelst,2002)

$$[1+\frac{15}{(n-p)}]^2 \times \frac{1}{\chi^2_{p,\alpha}} \qquad \textbf{....(2-4)}$$

إذ أن:

P : عدد المتغيّرات.

α :مستوى المعنوية.

٢.١.٢ طريقة مقدّر أصغر محدد تباين مشترك معاد الأوزان (RMCD):

وتقوم هذه الطريقة بنفس فكرة طريقة مقدر أصغر محدد تباين مشترك (MCD)،حيث يتم حساب تقديرات ابتدائية لموجـّه المتوسط ($\underline{\mu}$) ومصفوفة التباين والتباين المشترك(Σ) للمجموعات الجزئية بحجم (h) واختيار مصفوفة التباين

المشترك التي تمتلك أصغر محدد، وتعتبر بمثابة مقدر أولي لمصفوفة التباين المشترك (Σ) ويعبر عنها بـ $(\hat{\Sigma}_{j,0})$ وكذلك يعبرعن موجـّه المتوسط $(\bar{x}_{j,0})$:
[8](Rousseeuw , Leroy, 1987)

إذ أن:

....(2-5)

$$\underline{\bar{x}}_{j,0} = \frac{1}{h}\sum_{i=1}^{h}\underline{x}_{ij}$$

$$\hat{\Sigma}_{j,0} = dp \times \frac{1}{h}\sum_{i=1}^{h}(\underline{x}_{ij} - \underline{\bar{x}}_{j,0})(\underline{x}_{ij} - \underline{\bar{x}}_{j,0})' \qquad \textbf{....(2-6)}$$

إذ أن: dp :عامل الاتساق المحدد وفق الصيغة (2-4) .

وفقا لهذه التقديرات الأولية يتم حساب وزن لكل مشاهدة وكالآتي:

$$w_i = \begin{cases} 1 & if \quad \sqrt{(\underline{x}_{ij} - \underline{\bar{x}}_{j,0})'\hat{\Sigma}_0^{-1}(\underline{x}_{ij} - \underline{\bar{x}}_{j,0})} \le \chi^2_{P,1-\alpha} \\ 0 & if \quad o.w \end{cases}$$

....(2-7)

ويتم بعد ذلك حساب مقدرات أصغر محدد تباين مشترك معاد الأوزان (RMCD) النهائية للمشاهدات التي أوزانها (1) حيث تكون التقديرات كالأتي:

$$\underline{\bar{X}}_{RMCD} = \frac{\sum_{i=1}^{h} w_{ij}\,\underline{x}_{ij}}{\sum_{i=1}^{h} w_{ij}} \qquad \textbf{....(2-8)}$$

$$\hat{\Sigma}_{RMCD} = dp \times \frac{\sum_{i=1}^{h} w_{ij} (\underline{x}_{ij} - \overline{X}_{RMCD})(\underline{x}_{ij} - \overline{X}_{RMCD})'}{\sum_{i=1}^{h} w_{ij}} \qquad \text{....(2-9)}$$

$$\text{....(2-9)} \qquad \sum_{i=1}^{h} w_{ij} \neq 0$$

هذا وتتصف مقدرات أصغر محدد تباين مشترك معاد الأوزان (RMCD) بخاصية ثبات التغاير (Affine Equivariant). والتقديرات الموزونة الجديدة تقود إلى كفاءة إحصائية أفضل. (Rousseeuw, Katrin, 1999) [9]

3.1.2 طريقة مقدّر أصغر حجم مجّسم قطع بيضوي (MVE):

وهي إحدى طرائق التقدير الحصينة لمعلمتي الموقع والقياس، مقدر (MVE) لمصفوفة التباين والتباين المشترك، يعرف بأنه أصغر شكل بيضوي يحوي نصف المشاهدات، في حين مقدر(MVE) لموجّـه المتوسط هو مركز ذلك الشكل البيضوي. ويمكن توضيح هذه الطريقة على النحو الآتي :

لتكن F دالة توزيع (Distribution Function) لـ (P) من الأبعاد معرفة $\sum_n(F)$ أصغر شكل بيضوي يحوي نصف كتلة التوزيع و $\mu_n(F)$ هو مركز ذلك القطع البيضوي كذلك فان :

$$\Sigma \in PSD_{(P)}\{A \in R^{P\times P} \mid x' A x > 0 \ \ for\, all\, x \neq 0 \in R^P\}$$

إذ إن:

$x\, and\, \mu \in R^P$،كذلك فان :

$$D^2_{\mu,\Sigma}(x) = (x-\mu)' \, \Sigma^{-1} \, (x-\mu) \qquad \text{....(2-10)}$$

هي المسافة التربيعية بين μ , x .

في أكثر الأوقات يستعمل مقدر(MVE) لتقدير معالم الإنموذج البيضوي ذي الشكل الآتي :

....(2-11)

$$p._{\sum,\mu}(x)=[\det(\Sigma)]^{-\frac{1}{2}}g([(x-\mu)'\Sigma^{-1}(x-\mu)]^{\frac{1}{2}})$$

وعندما تكون المعاينة في هذا الإنموذج نحن نحصل على:

$$\mu_{\alpha}(F_{\mu,\Sigma})=\mu \quad and \quad C_{\alpha}\sum_{\alpha}(F_{\mu,\Sigma})=\Sigma \qquad ....(2\text{-}12)$$

إذ ينتج إتساق فيشر(Fisher Consistency) عندما تضرب مصفوفة التباين والتباين المشترك في (MVE) بثابت مناسب (في الإنموذج الطبيعي متعدد المتغّيرات يكون مساوياً لـ ـ: $C_{\alpha}=1+\frac{15}{n-p}$). Rousseeuw , Leroy, (1987)

اختيار $(\alpha=\frac{1}{2})$ ينتج أصغر شكل بيضوي يتضمن نصف كتلة التوزيع، ويطلق على المقدرات الناتجة لهذه الطريقة بمقدرات (MVE) المضبوطة (EMVE) (Exact Minimum Volume Ellipsoid Estimator) لإعادة المعاينة في مقدر (MVE) سوف نعتبر هنا مقدر المجموعة الجزئية (P+1)، (لان هذا أصغر حجم عينة جزئية ممكن أن يؤخذ ،لإن هذا هو أصغر حجم يؤدي إلى مصفوفة التباين والتباين المشترك غير المفردة (Nonsingular) ويعظم فرصة الحصول على عينة جزئية خالية من الشواذ).

لكل مجموعة جزئية (J) تتكون من(P+1) من المشاهدات أي:

$$J=\{i_1,i_2,....,i_{p+1}\} \ of \ \{1,2,.....,n\}$$

نحسب الوسط الحسابي (μ_j)، ومصفوفة التباين والتباين المشترك (Σ_j) للمشاهدات (P+1) إذ إن $\{x_j / j\in J\}$.المسافة التربيعية (x_i) و(μ_j) في المصفوفة (Σ_j) تكون بالصيغة الآتية:

....(2-13)

$$D_i^2(J)=(x_i-\mu_j)'\Sigma_j^{-1}(x_i-\mu_j)$$

وبأخذ الإحصاءة المرتـّبة (Order Statistic) ذات الرتبة $(hp=\frac{n+p+1}{2})$ من $D_i^2(J)$ تنتج مسافة الوسيط (Median Distance) وهي $D_{(hp)}^2(J)$. إذا جعلنا الشكل البيضوي وفقا إلى $\sum_j$ مع العامل $D_{(hp)}^2(J)$ وبالتالي فان المجموعة الجزئية المثالية معرفة على النحو الآتي :

$$J = \arg\min_{i\in j} \det(\textstyle\sum_j D_{(hp)}^2(J))$$

....(2-14)

مع مقدرات المجموعة الجزئية (P+1) :

$$\hat{\mu}_n^p = \mu_j \quad \text{and} \quad \hat{\Sigma}_n^p = \Sigma_j D_{(hp)}^2(J) \qquad \textbf{....(2-15)}$$

ومن ثم فكرة حساب مقدر(MVE)لمقدرات المجموعة الجزئية ((Average (P+1) موضحة كالآتي: (Rousseeuw , 1985) [7]

يتم تحديد العينات الجزئية بحجم (P+1) من(المشاهدات) أي إننا نقوم باختيار عدد من العينات الجزئية (المصفوفات) مساوي إلى (C_{P+1}^n) .ثم لكل عينة جزئية (J) بحجم $J=\{i_1, i_2, ..., i_{p+1}\}$ of $\{1,2,.....,n\}$ من (P+1) (المشاهدات) يتم استخراج موجـّه المتوسط $\mu_{(j)}$ ومصفوفة التباين والتباين المشترك $\Sigma_{(j)}$ وكالاتي:

$$\overline{X}_J = \frac{1}{P+1}\sum_{i\in j} x_i \qquad \textbf{....(2-16)}$$

$$\Sigma_J = \frac{1}{P}\sum_{i\in j}(x_i - \overline{X}_J)(x_i - \overline{X}_J)' \qquad \textbf{....(2-17)}$$

ولكل اختيارلـ($\mu_{(j)}$ و $\Sigma_{(j)}$) أعلاه يجري استخراج المسافات التربيعية وفق الصيغة

$$D_i^2(J) = (x_i - \mu_j)' \Sigma_j^{-1} (x_i - \mu_j) \qquad \textbf{....(2-18)}$$

وبعد استخراج $D_i^2(J)$ لكافة مشاهدات العيّـنة، يتم بعد ذلك ترتيب هذه القيم تصاعديا واختيار $D_{(hp)}^2(J)$ حيث ($hp=\frac{n+p+1}{2}$) مسافة الوسيط (Median Distance) لكل عينة جزئية (J). وثم يتم اختيار أفضل مجموعة جزئية مثالية معرفة كما في الصيغة

$$J = \arg\min_{i \in j} \det(\Sigma_j D_{(hp)}^2(J)) \qquad ...(٢-١٩)$$

بعدها يتم حساب المقدرات الحصينة موجّـه المتوسط ($\hat{\mu}_i$) ومصفوفة التباين والتباين المشترك ($\hat{\Sigma}_i$) كما في الصيغة وكالآتي :

$$\hat{\mu}_i = \mu_j(i) \qquad ...(2\text{-}20)$$

....(2-21)

$$\hat{\Sigma}_i = \Sigma_j(i)\, D_{(hp)}^2(j_{(i)})$$

ثم يتم ترتيب تصاعدي للأزواج ($\hat{\Sigma}_i$ و $\hat{\mu}_i$) وفقاً لكمية $|\hat{\Sigma}_i|$. ليتم بعدها حساب المتوسطات إلى أصغر نصف من هذه الأزواج لتنتج:

....(2-22)

$$\overline{X}_n^c = \frac{1}{hp}\sum_{i=1}^{hp} \overline{X}(i)$$

$$\hat{\Sigma}_n^c = \frac{1}{hp}\sum_{i=1}^{hp} \hat{\Sigma}_{(i).j} \qquad .... (٢-٢٣)$$

هنا نأخذ فقط النصف الأصغر من المقدرات الحصينة في العينات النهائية والتي تبقى تنتج مقدرات تمتاز بخاصيةِ ثبات للتغاير (Affine Equevariant). (Croux,Haesbroeck, 2001)[3] ،ويتم بعد ذلك استخدام قواعد التمييز الخطية والتربيعية التقليدية والحصينة ولطرائق التقدير التقليدية والحصينة.

٢.٢ قواعد التمييز الحصينة:

١.٢.٢قواعد التمييز الخطّيّة الحصينة Rules:

1.1.2.2 طريقة مقدّر أصغر محدد تباين مشترك (MCD) :

على وفق هذه الطريقة يتم لكل مجموعة j إذ إن (j = 1,2,3,……,L) تقدير موجـّه متوسط ($\underline{\mu}_j$)، ومصفوفة التباين والتباين المشترك(Σ_j) وخطا التصنيف (p_j) باستعمال مقدرات طريقة (MCD) للمجموعة الجزئية (h) التي تمتلك أصغرمحدد من بين (H) من المجموعات الجزئية الكلية لكل مجموعة إذ تكون المقدرات على النحو الآتي :

$$\underline{\bar{x}}_{j,MCD} = \frac{\sum_{i=1}^{h} \underline{x}_{ij}}{h} \quad , \; j = 1,2,...,L \qquad ....(2\text{-}24)$$

$$\hat{\Sigma}_{j,MCD} = dp \times \frac{\sum_{i=1}^{h} (\underline{x}_{ij} - \underline{\bar{x}}_{j,MCD})(\underline{x}_{ij} - \underline{\bar{x}}_{j,MCD})'}{h} \qquad ....(2\text{-}25)$$

وان مقدار احتمال خطأ التصنيف الحصين ($\hat{P}_j^R$) يمكن أن يقدر. [4](Hubert, Driessen, 2004)

$$\hat{P}_j^R = \frac{\tilde{n}_j}{\tilde{n}} \qquad ....(2\text{-}26)$$

إذ إن:

$\tilde{n}_j$: عدد المشاهدات غير الشاذة المقدرة في المجموعة j .

$\tilde{n}$: عدد المشاهدات الكلية غير الشاذة المقدرة لكل المجتمعات.

وبعدها يتم حساب مصفوفة التباين والتباين المشترك المدمجة العامة (Pooled Var-Cov Matrix) على وفق الصيغة الآتية :

$$\hat{\Sigma}_{PMCD} = \frac{\sum_{j=1}^{L} n_j \hat{\Sigma}_{j,MCD}}{\sum_{j=1}^{L} n_j} \quad \text{....(2-27)}$$

إذ إن قاعدة التمييز الخطية الحصينة(RLDR) وفق المقدرات في الصيغ(2-24) ,(2-26)،(2-27) تصبح :

بتعيين x إلى $\prod_k$ وإذا كانت :

$$\hat{d}_k^{RL}(\underline{x}) > \hat{d}_j^{RL}(\underline{x}) \quad , \forall j=1,...,L \ , \ j \neq k \quad \text{....(2-2٨)}$$

إذ إن:

$$\hat{d}_j^{RL}(\underline{x}) = \hat{d}_j^{RL}(\underline{x}, \hat{\underline{\mu}}_{j,MCD}, \hat{\Sigma}_{PMCD}) = \\ \underline{\bar{x}}'_{j,MCD} \hat{\Sigma}_{PMCD}^{-1} \underline{x} - \frac{1}{2} \underline{\bar{x}}'_{j,MCD} \hat{\Sigma}_{PMCD}^{-1} \underline{\bar{x}}_{j,MCD} + Ln(\hat{p}_j^R) \quad \text{....(2-29)}$$

وعندما $\pi_1 = \pi_2$ وان (L=2) فإن الصيغة (29-2) تعرف بقاعدة التمييز الحصينة لفشر التي توصف على النحو الآتي:

$$\begin{cases} x \in \pi_1 & \text{if} \quad (\bar{x}_1 - \bar{x}_2)' \hat{\Sigma}_{PMCD}^{-1} (\underline{x} - (\bar{x}_1 + \bar{x}_2)/2) > 0 \\ x \in \pi_2 & \text{if} \quad \text{other wise} \end{cases}$$

فإن الصيغة (٣٠-٢) يتم من خلالها تحديد المسافة الحصينة . **(2-30)....**

2.1.2.2 طريقة مقدّر أصغر محدد تباين مشترك معاد الأوزان (RMCD):

وتقوم هذه الطريقة على فكرة طريقة مقدر أصغر محدد تباين مشترك(MCD) لدالة التمييز الخطية الحصينة (RLDR) نفسها. إذ يتم حساب تقديرات أولية لموجــّه المتوسط ($\underline{\bar{x}}_{j,o}$) ، ومصفوفة التباين المشترك ($\hat{\Sigma}_o$) بطريقة (RMCD) ووفقاً لهذه التقديرات الأولية يتم حساب وزن لكل مشاهدة في المجموعة j ، ويتم بعد ذلك حساب مقدرات أصغر محدد تباين مشترك معاد الأوزان

(RMCD) النهائية للمشاهدات التي أوزانها (1) حيث تكون التقديرات على النحو الآتي :

$$\bar{x}_{j,RMCD} = \sum_{i=1}^{nj} w_{ij} x_{ij} \Big/ \sum_{i=1}^{nj} w_{ij} \qquad ....(2\text{-}31)$$

....(2-32)

$$\hat{\Sigma}_{PRMCD} = \frac{\sum_{j=1}^{L}\sum_{i=1}^{nj} w_{ij}(x_{ij}-\bar{x}_{j,RMCD})(x_{ij}-\bar{x}_{j,RMCD})'}{\sum_{j=1}^{L}\sum_{i=1}^{nj} w_{ij}}$$

وان مقدار احتمال خطأ التصنيف الحصين ($\hat{P}_j^R$) يتم حسابه لكل مشاهدة x_{ij} بالمجموعة j .وبالتالي فان المسافة الحصينة النهائية تحسب على النحو الآتي:

$$RD_{ij} = \sqrt{(x_{ij} - \bar{x}_{j,MCD})\hat{\Sigma}_{j,MCD}^{-1}(x_{ij} - \bar{x}_{j,MCD})}$$

....(2-33)

ويعد x_{ij} قيمة شاذة إذا و فقط إذا:

$$RD_{ij} > \sqrt{\chi^2_{P,0.975}} \qquad ....(2\text{-}34)$$

و بفرض أن $\tilde{n}_j$ تشير إلى عدد غير الشواذ في المجموعة j وأن:

$$\tilde{n} = \sum_{j=1}^{g} \tilde{n}_j \qquad ....(2\text{-}35)$$

ومن ثم فان تقدير احتمال خطأ التصنيف الحصين تم تقديره في الصيغة (2-٢٦)،وللحصول على تقديرات مصفوفة التباين المشترك الأولية $\hat{\Sigma}_o$ في الصيغة (٢-7) يكون بدمج مصفوفة التباين والتباين المشترك لكل مجموعة بعد الحصول على تقديرات $\hat{\mu}_{j,MCD}$ و $\hat{\Sigma}_{j,MCD}$ لكل مجموعة، ومن ثم يتم حساب مصفوفة التباين والتباين المشترك المدمجة العامة (Pooled Var-Cov Matrix) وعلى وفق الصيغة (2-27) وان قاعدة التمييز الخطية الحصينة (RLDR)على وفق المقدرات (2-26) ،(2-31)، (2-32) تصبح : (Hubert, Driessen, 2004)[4]

بتعيين x إلى π_k وإذا كانت :

$$\hat{d}_k^{RL}(\underline{x}) > \hat{d}_j^{RL}(\underline{x}) \quad , \forall j=1,...,L \ , \ j \neq k$$

إذ إن:

....(2-36)

$$\hat{d}_j^{RL}(\underline{x}) = \hat{d}_j^{RL}(\underline{x}, \underline{\hat{\mu}}_{j,RMCD}, \hat{\Sigma}_{PRMCD}) =$$

$$\underline{\bar{x}}'_{j,RMCD}\ \hat{\Sigma}^{-1}_{PRMCD}\ \underline{x} - \frac{1}{2}\underline{\bar{x}}'_{j,RMCD}\ \hat{\Sigma}^{-1}_{PRMCD}\ \underline{\bar{x}}_{j,RMCD} + Ln(\hat{p}_j^R)$$

3.1.2.2 طريقة مقدّر أصغر حجم مجّسم قطع ناقص (MVE):

(Minimum Volume Ellipsoid Estimator)

على وفق هذه الطريقة يتم لكل مجموعة j .

إذ إن (j=1,2,3,……,L) تقدير موجـّـه متوسط ($\underline{\mu}_j$) ومصفوفة التباين والتباين المشترك (Σ_j) وخطا التصنيف (p_j) باستعمال مقدرات طريقة (MVE) للمجموعة الجزئية (h_p)التي تمتلك أصغرمحدد من بين (H) من المجموعات الجزئية الكلية لكل مجموعة من مسافة الوسيط (Median Distance) إذ تكون المقدرات على النحو الآتي:

$$\underline{\bar{x}}_{j,MVE} = \frac{\sum_{i=1}^{hp} \underline{x}_{ij}}{hp} \quad , j=1,2,...,L \qquad \text{....(2-37)}$$

....(2-38)

$$\hat{\Sigma}_{j,MVE} = \frac{1}{hp}\sum_{i=1}^{h_p}\hat{\Sigma}_{(i).j} \quad , j=1,2,....,L$$

وان مقدار احتمال خطأ التصنيف الحصين ($\hat{P}_j^R$) تم تقديره كما في الصيغة (2-٢٦).

وبعدها يتم حساب مصفوفة التباين والتباين المشترك المدمجة العامة (Pooled Var-Cov Matrix) على وفق الصيغة الآتية:

$$\hat{\Sigma}_{PMVE} = \frac{\sum_{j=1}^{L} n_j \, \hat{\Sigma}_{j,MVE}}{\sum_{j=1}^{L} n_j} \qquad \text{....(2-39)}$$

إذ إن قاعدة التمييز الخطية الحصينة (RLDR) وعلى وفق المقدرات في الصيغ (2-37)،(2-39)،(2-26) تصبح (Hubert,Driessen,2004)[٤] :

بتعيين x إلى Π_k وإذا كانت :

$$\hat{d}_k^{RL}(\underline{x}) > \hat{d}_j^{RL}(\underline{x}) \quad , \forall j = 1,...,L \ , \ j \neq k$$

إذ إن:

$$\hat{d}_j^{RL}(\underline{x}) = \hat{d}_j^{RL}(\underline{x}, \underline{\hat{\mu}}_{j,MVE}, \hat{\Sigma}_{PMVE}) = \underline{\bar{x}}'_{j,MVE}\, \hat{\Sigma}_{PMVE}^{-1}\, \underline{x} - \frac{1}{2} \underline{\bar{x}}'_{j,MVE}\, \hat{\Sigma}_{PMVE}^{-1}\, \underline{\bar{x}}_{j,MVE} + Ln(\hat{p}_j^R) \qquad \text{....(2-40)}$$

2.2.2 قواعد التمييز التربيعية الحصينة:

1.2.2.2 طريقة مقدّر أصغر محدد تباين مشترك (MCD) :

وتقوم هذه الطريقة على إيجاد المقدرات ($\underline{\bar{x}}_{j,MCD}$) و($\hat{\Sigma}_{j,MCD}$) و($\hat{P}_j^R$) لكل مجموعة j إذ ان $j = 1,2,3,...,L$ كما في الصيغ،(2-٢٤)،(2-25)، (2-26) . وبالتالي فان قاعدة التمييز التربيعية الحصينة (RQDR) وعلى وفق هذه الطريقة تصبح (Hubert,Driessen,2004)[٤] :

بتعيين x إلى Π_k وإذا كانت :

$$\hat{d}_k^{RQ}(\underline{x}) > \hat{d}_j^{RQ}(\underline{x}) \quad , \ \forall j = 1,2,...,L \ , \ j \neq k$$

إذ إن:

....(2-41)

$$\hat{d}_j^{RQ}(\underline{x})=\hat{d}_j^{RQ}(\underline{x},\underline{\hat{\mu}}_{j,MCD},\hat{\Sigma}_{j,MCD})$$
$$=-\frac{1}{2}Ln\left|\hat{\Sigma}_{j,MCD}\right|-\frac{1}{2}(\underline{x}-\underline{\bar{x}}_{j,MCD})'\hat{\Sigma}_{j,MCD}^{-1}(\underline{x}-\underline{\bar{x}}_{j,MCD})+Ln(\hat{p}_j^R)$$

2.2.2.2 طريقة مقدّر أصغر محدد تباين مشترك معاد الأوزان (RMCD):

وتقوم هذه الطريقة كذلك على إيجاد المقدرات($\underline{\bar{x}}_{j,RMCD}$) و($\hat{\Sigma}_{j,RMCD}$) و($\hat{P}_j^R$) في الصيغ (2-٣١)، (2-9)، (2-26) .

وبالتالي فان قاعدة التمييز التربيعية الحصينة (RQDR) على وفق هذه الطريقة تصبح (Hubert,Driessen,2004):

بتعيين x إلى $\prod_k$ وإذا كانت:

$$\hat{d}_k^{RQ}(\underline{x}) > \hat{d}_j^{RQ}(\underline{x}) \quad , \forall j=1,2,...,L \ , j\neq k$$

إذ إن:

....(٤٢ - ٢)

$$\hat{d}_j^{RQ}(\underline{x})=\hat{d}_j^{RQ}(\underline{x},\underline{\hat{\mu}}_{j,RMCD},\hat{\Sigma}_{j,RMCD})$$
$$=-\frac{1}{2}Ln\left|\hat{\Sigma}_{j,RMCD}\right|-\frac{1}{2}(\underline{x}-\underline{\bar{x}}_{j,RMCD})'\hat{\Sigma}_{j,RMCD}^{-1}(\underline{x}-\underline{\bar{x}}_{j,RMCD})+Ln(\hat{p}_j^R)$$

3.2.2.2 طريقة مقدّر أصغر حجم مجّسم قطع ناقص (MVE):

وتقوم هذه الطريقة كذلك على إيجاد ($\underline{\bar{x}}_{j,MVE}$) و($\hat{\Sigma}_{j,MVE}$) و($\hat{P}_j^R$) في الصيغ (2-٣٧)،(2-38)،(2-26) .

وبالتالي فان قاعدة التمييز التربيعية الحصينة (RQDR) على وفق المقدرات في الصيغ المذكورة تصبح (Hubert,Driessen,2004)[٤] :

بتعيين x إلى $\prod_k$ وإذا كانت :

$$\hat{d}_k^{RQ}(\underline{x}) > \hat{d}_j^{RQ}(\underline{x}) \quad , \forall j=1,2,...,L \ , j\neq k$$

إذ إن:

$$\hat{d}_j^{RQ}(\underline{x}) = \hat{d}_j^{RQ}(\underline{x}, \underline{\hat{\mu}}_{j,MVE}, \hat{\Sigma}_{j,PMVE})$$
$$= -\frac{1}{2} Ln\left|\hat{\Sigma}_{j,MVE}\right| - \frac{1}{2}(\underline{x} - \underline{\bar{x}}_{j,MVE})' \hat{\Sigma}_{j,MVE}^{-1} (\underline{x} - \underline{\bar{x}}_{j,MVE}) + Ln(\hat{p}_j^R) \quad \text{....(2-43)}$$

$$\hat{d}_j^{RL}(x) = \hat{d}_j^{RL}(\underline{x}, \underline{\hat{\mu}}_{j,MCD}, \hat{\Sigma}_{PMCD}) =$$
$$\underline{\bar{x}}'_{j,MCD}\, \hat{\Sigma}_{PMCD}^{-1}\, \underline{x} - \frac{1}{2} \underline{\bar{x}}'_{j,MCD}\, \hat{\Sigma}_{PMCD}^{-1}\, \underline{\bar{x}}_{j,MCD} + Ln(\hat{p}_j^R) \quad \text{...(2-44)}$$

حيث عندما $\pi_1 = \pi_2$ وان (L=2) فإن الصيغة (٢-٤٤) تعرف بقاعدة التمييز: الحصينة لفشر والتي توصف كالآتي:

$$\begin{cases} x \in \pi_1 & \text{if} \quad (\underline{\bar{x}}_1 - \underline{\bar{x}}_2)' \hat{\Sigma}_{\text{PMCD}}^{-1} (\underline{x} - (\underline{\bar{x}}_1 + \underline{\bar{x}}_2)/2) > 0 \\ x \in \pi_2 & \text{if} \quad \text{other wise} \end{cases} \quad \text{... (2- 45)}$$

[3](Croux, Haesbroeck, 2001)

٣.٢ اختبار دالة التمييز :

1.3.2 اختبار المتغيرات :

لاختيار المتغيرات ذات القوة التمييزية المعنوية و التي تتبع أقل خطأ تصنيف. هناك عدة أساليب منها طرق الاختيار المتدرج (Stepwise Selection) و فيها يكون اختيار دالة التمييز القياسي لمعنوية الفروق بين متوسطات المجاميع تعميم متعدد المتغيرات لمقياس Wilk's Lambda أحادي المتغير و بأخذ التباينات المشتركة بنظر الاعتبار حيث يوفر مقياس Wilk's Lambda متعدد المتغيرات و لـ P من المتغيرات: (Neil.H.Timm, 2002)[5].

$$\hat{P} = |W_{PP}| / |T_{PP}| \quad \text{....(2-46)}$$

إذأن :

W: مصفوفة مجموع مربعات الانحرافات داخل المجاميع.

T : مصفوفة مجموع مربعات الانحرافات الكلية.

و تكون إحصاءة F المناظرة المستندة على أساس هذا المقياس كالآتي:

$$F(\hat{P}) = \left[1 - \hat{P}\right] \left[n - P - 1\right] / \hat{P}P \quad ....(2\text{-}46)$$

والتي تتوزع $F_{1,n-P-1}$.

2.3.2 اختبار دالة التمييز بعد حذف المتغيرات غير التمييزية :

ومن هذه الاختبارات:

2 . 3 . 2 .١ اختبار معنوية الدالة المميزة (تساوي المجموعات) :

تم استخدام مقياس مربع كاي وتكون صيغته كالآتي : (Neil.H.Timm,2002).

$$\chi^2 = -\left[n - 1 - 1/2\,(P+g)\right] \log\hat{e} \quad ....(2\text{-}48)$$

إذ أن:

P : المتغّيرات المقدرة.

g :عدد المجاميع .

^ : مقياس ولكس ويكون بين (صفر) و(الواحد).

و الذي يتوزع تقريباً مربع كاي بدرجة حرية P(g-1).

2 . 3 . 2 .٢ اختبار تساوي مصفوفة التباين و التباين المشترك لجميع المجاميع :

وتكون احصاءة هذا الاختبار كما يلي:(Neil.H.Timm, 2002).

$$\mu = \left(\sum_{i=1}^{k} V_i\right) \text{Ln } |S| - \sum_{i=1}^{k} \left(V_i \text{ Ln } |S|\right) \quad ....(249)$$

وقد أثبت Box عام (1949) أنه إذا ضرب μ في ثابت C^{-1} والذي يساوي:

$$C^{-1} = 1 - \frac{2m^2 + 3m + 1}{6(m+1)(k-1)} \left[\sum_{i=1}^{k} \frac{1}{V_i} - \frac{1}{\sum_{i=1}^{k} V_i}\right] \quad ....(2\text{-}50)$$

سوف نحصل على مقياس Box'M يتوزع بالتقريب توزيع χ^2 وبدرجة حرية

.(m - 1)(k - 1)/2

إذ أن:

m : عدد المتغيّرات المدروسة في المصفوفة.

k : عدد المصفوفات المختارة .

٣. الجانب التطبيقي :

١.٣ أمراض القلب أعراضها السريرية والعوامل المسببة للمرض وطرق تشخيصها :

إن التصلب الذي يحدث لجدران الشرايين ينتهي بحدوث جلطة تؤدي الى أنسداد الشريان الضيق تماماً، فينقطع الدم عن جزء من عضلات القلب، إذ تقل تغذية القلب نتيجةً لانسداد الشرايين ،فيصبح أي عملٍ إضافيٍ مجهوداً للقلب، كعدم القدرة على المشي لمسافة كان سابقاً يستطيع مشيها، وعدم القدرة على صعود مرتفع، وعند تفاقم الحالة يكون صعود الدرج صعب، كذلك يعاني من ضيق التنفس، وعدم القدرة على النوم، إذ تبدأ تكون الآلام في الصدر، ثم تنتشر إلى الكتف الأيسر عادةً، وتكون مصحوبةً بتعرقٍ وشحوبٍ في الوجه، مع خفقان القلب، مما يؤدي إلى عدم تشخيصه كذبحة صدرية، أو جلطة قلبية، وإذا كان الانسداد كبيراً، فإن بطين القلب يبدأ بالارتجاف، ثم يتوقف عن العمل، ويموت المصاب، ولما كان ارتفاع نسبة الكولسترول في الدم من أهم الأعراض، إذ تلتصق الدهون بجدران الأوعية لتشكّل ترسبات تسبب ضيق في الأوعية الدموية، وتشكّل في بعض الأحيان انتفاخاً يشبه البصلة داخل تجويف الأوعية، وعند الجهد المفاجئ قد تنفجر هذه البصلة، و تتوزع محتوياتها على شكل كتل صغيرة داخل الوعاء الدموي، وقد تتسبب في حالات اغلاق شريان في القلب، إذ تسبب جلطة في القلب، أواغلاق شريان في المخ، وتسبب الجلطة الدماغية، وحالات أخرى تسبب مضاعفات خطيرة كالشلل، والسكتة القلبية. لذا يجب الاعتماد على الزيوت ذات المصدر النباتي، وكذلك ممارسة الرياضة، إذ تقوم بتخليص الدم من ترسبات الدهون العالقة، والابتعاد عن التدخين. ولتشخيص مثل هذه الأعراض يستند أولاً للأعراض السريرية بالإضافة للأنواع الأخرى من التشخيص، وهي تخطيط القلب، وتحليل الدم للتحقق من بعض الانزيمات التي تفرز بالدم. عند حدوث الجلطة يجب عمل الايكو، وعمل رنين للدماغ، وهنا يبرز أهمية التشخيص المبكر لأمراض القلب .
(تشارلز، و.أيرتج وآخرون ،1970)

2.3 وصف العيّـنة و المتغيّرات المستخدمة وأسلوب جمع البيانات:

تم الأعتماد على ملفات المرضى (طبلات المصابين) في المستشفى التعليمي والمستشفى الجمهوري، في محافظة البصرة للأعوام (2008 و2009 و2010) ، واشتملت بيانات البحث على (١٤٠) مريضاً، وقد تم سحب عينة عشوائية من المجتمع الأول (الذبحة الصدرية A.P) بحجم (n_1=70) ومن المجتمع الثاني (الجلطة القلبية C.H) بحجم (n_2=70)، والتي تـّمثل مجموعتين من التشخيص لأمراض القلب الذي يمثـّل المتغيّر المعتمد (متغيّر الاستجابة) والذي يتمثل بنوع المرض (مرض الذبحة الصدرية =1) و(مرض الجلطة القلبية =2) وعدد المتغيّرات الإيضاحية (المستقلة) التي تم الاتفاق عليها مع الأطباء أصحاب الاختصاص هي (٧) متغيّرات يمكن توضيحها بإيجاز على النحو الآتي:

1. العمر (age) :تتراوح الفئات العمرية لدى الجنسين بين (١٥-٨٥) سنة.

٢. الوزن (weight): تم حساب الزيادة المثلى عن الوزن المثالي .

٣.ضغط الدم (Blood Pressure): تم حساب الزيادة عن المقدار الطبيعي لضغط الدم الطبيعي عند الأنسان المتوسط العمر، والذي تتراوح نسبته بين (١٢٠/٨٠) .

٤. نسبة السكر في الدم (Sugar): تم حساب الزيادة عن المقدار الطبيعي للسكر في الدم والتي تتراوح نسبته بين (١٤٠- ٨٠) ملغم لكل مئة مليلتر من الدم .

5. نسبة اليوريا في الدم (Urea): تم حساب الزيادة عن المقدار الطبيعي لليوريا في الدم والتي تبلغ (% ٤٠).

6. نسبة الكولسترول في الدم (cholesterol): تم حساب الزيادة عن المقدار الطبيعي للكولسترول في الدم والتي تتراوح نسبته بين (٢٠٠- ١٨٠) ملغم لكل مئة مليلتر من الدم.

7. الجنس (sex): و يمثـل متغيّر ثنائي (ذكر = 1) ، (أنثى = 2).

وكان توزيع مفردات العيّـنة بين المجموعتين و نسبة كل مجموعة كالآتي:

Group Name	Prior	No. of Cases	Group
الذبحة الصدرية	٠0.5	70	1
الجلطة القلبية	٥٠0.	٧٠	2
		٤٠1	Total

٣.٣ نتائج الاختبارات :

١. اختزال عدد المتغّيرات:ولتقليص عدد المتغّيرات لكي يشتمل النموذج على المتغّيرات ذات الطابع التمييزي، فقط تم استخدام الاختيار المتدرج (Stepwise Discriminant Analysis) وباستخدام إحصاءتي Wilk's Rao's V&Lambda كمعيار لاختيار المتغّيرات ذات القيمة الأصغر لهذه الإحصاءة. وأظهرت نتائج التحليل دخول ستة متغّيرات واستبعاد متغّير الجنس (Sex) كما في الجدول رقم (١)،(٢)،(3).

جدول رقم (1)

نتائج اختيار المتغيّرات

Variable in the Analysis (اختيار المتغيّرات لتقدير الدوال المميزة)							
Step	**Variables**	**Tolerance**	**F to Remove**	**Wilk's Lambda**	**مستوى المعنوية**	**Rao's V**	**مستوى المعنوية**
1	Cholesterol	1.000	.565١2	.804	.000		
2	Cholesterol	.966	.429٧2	.804	.000	.126٥1	.000
	Weight	.966	02٩21.	.878	.000	.555٣2	.000
3	Cholesterol	.963	٧٧٧24.	.774	.000	.559٨3	.000
	Weight	.950	.008١2	.771	.000	6٥٥040.	.000
	Blood Pressure	.980	15.337	.733	.000	.161٧4	.000
4	Cholesterol	.920	27.956	.711	.000	.234٣4	.000
	Weight	.888	12.290	.627	.000	.908٥6	.000
	Blood Pressure	.977	16.332	.688	.000	.471٨5	.000
	Sugar	.876	9.176	.634	.000	.309٧٠	.000
5	Cholesterol	.900	31.930	.777	.000	.318١5	.000
	Weight	.879	13.588	.654	.000	.513٣7	.000
	Blood Pressure	.976	18.383	.666	.000	.21٨6	.000
	Sugar Urea	.878	7.555	.654	.000	.837٢2	.000
		.955	6.317	.587	.000	.474٤8	.000

.000	٨٨٨85.	.000	.614	10.432	.687	Cholesterol	6
.000	.006٧7	.000	.644	15.578	.850	Weight	
.000	.581٢8	.000	.678	13.739	.965	Blood Pressure	
.000	.473٧8	.000	.665	9.325	.876	Sugar	
.000	.327٢9	.000	.566	7.576	.976	Urea	
.000	.976٤9	.000	.509	5.991	.683	Age	

جدول رقم (٢)

نتائج إحصاءة Wilk's Lambda

Exact F				df3	df2	df1	Lambda	Number of Variables	Step
Sig.	df2	df1	Statistic						
.000	136.000	1	21.555	136	1	1	.758	1	1
.000	135.000	2	23.504	136	1	2	.638	2	2
.000	134.000	3	22.863	136	1	3	.562	3	3
.000	133.000	4	21.408	136	1	4	.420	4	4
.000	132.000	5	17.242	136	1	5	.491	5	5
.000	131.000	6	16.094	136	1	6	.370	6	6

جدول رقم (٣)

نتائج إحصاءة Rao’s V

v in Change		Rao’s V			Entered	Step
Sig.	Statistic	Approx. sig.	df	Statistic		
.000	25.655	.000	1	21.333	Cholesterol	1
.000	26.406	.000	2	48.222	Weight	2
.000	23.333	.000	3	68.321	Blood Pre	3
.000	15.322	.000	4	80.554	Sugar	4
.000	11.345	.000	5	90.564	Urea	5
.000	9.567	.000	6	105.666	Age	6

٢. تم اختبار البيانات لمعرفة هل إن المتغّيرات تتوزع توزيعاً طبيعياً (Normal) أم لا، من خلال إختبار كاي سكوي للبيانات في برنامج (SPSS.16) وحسب الاختبارات أظهرت النتائج أن المتغّيرات الوزن (Weight)، نسبة اليوريا في الدم (Urea)، ضغط الدم (Blood Pressure)، ونسبة السكرفي الدم (Sugar)،ونسبة الكولسترول في الدم (cholesterol)، لا تتوزع توزيعاً طبيعياً، و فقط متغّير العمر (Age) كان يتوزع توزيعاً طبيعياً.

٣. تم الكشف عن وجود القيم الشاذّة بطريقة (Box- and -Whisker Plot) حيث أظهرت النتائج أن المتغّيرات (الوزن، نسبة اليوريا في الدم ، ضغط الدم والعمر) تحتوي على قيم شاذة و بأعداد متفاوتة من طرف واحد أو طرفين ولم تظهر قيم شاذة في متغّيري(نسبة السكر في الدم) و(نسبة الكولسترول في الدم) والرسوم معطاة في الملحق (١).

٤. اختبار معنوية الدالة المميزة و بالاعتماد على المعادلة (2.48) كانت قيمة مربع كاي $\chi^2 = 66.841$ وبمستوى معنوية 0.01 و هذا يدل على إن الدالة لها إمكانية جيدة على التمييز.

٥. اختبار تساوي التباينات و بالاعتماد على المعادلة (2.49) و (2.50) فإن Box'sM=132.488 وهذا يدل على عدم التجانس بين المجموعات .

نتائج تحليل طرائق التمييز لبيانات أمراض القلب : 4.3

لغرض تطبيق طرائق التقدير الحصينة في الواقع العملي تم استخدام بيانات حقيقية عن أمراض القلب، وهي بيانات عن مرض الذبحة الصدرية، و مرض الجلطة القلبية، لأنها تحقق الغرض من هذه الدراسة،وذلك يعود إلى إن المجالات الطبية مليئة ً ببيانات تتصف بان بعض قياساتها، أو التسجيلات الخاصة بمتغّيراتها، تكون متجاوزة لنسبها الطبيعية داخل جسم الإنسان، وهذه القياسات تعد حينئذ شاذة، فقد قام الباحث بكتابة برنامج بلغة (Q.Basic) لغرض تطبيق طرائق التقدير الحصينة. (والبرنامج معطى في الملحق(٢)) ، وإن نتائج التقدير لجميع متغّيرات الدراسة للمجموعتين الأولى والثانية، باستخدام دالة التمييز الخطية والتربيعية التقليدية والحصينة، ولطرائق التقدير التقليدية والحصينة ،وهي موضحة في الجداول رقم **(٤)** **(5)**على التوالي:

جدول رقم (4)

نتائج تحليل طرائق التمييز لبيانات الذبحة الصدرية

المجموعة الأولى باستعمال طرائق التمييز التقليدية والحصينة

Variables	LDF	QDF	RLM CD	RQ MCD	RLR MCD	RQR MCD	RLM VE	RQ MVE
constant	1.3834	-160.08	4.2108	-242.35	-137.72	4.4319	114.49	-211.43
Age	80.56	79.25	2.29	2.01	0.96	.99	1.11	1.12
Weight	5.81	4.73	51.45	2.12	0.69	3.07	1.14	1.01
Urea	21.93	18.75	2.61	1.97	1.06	1.1	0.96	0.4
Blood Pressure	14.41	13.74	624.23	1.92	0.98	-7.30	0.68	0.79
Sugar	.02	.01	682.78	2.22	1.30	2.77	0.87	0.71
Cholesterol	5.85	4.93	419.34	1.93	1.30	1.39	1.12	0.77

جدول رقم (5)

نتائج تحليل طرائق التمييز لبيانات الجلطة القلبية

المجموعة الثانية باستعمال طرائق التمييز التقليدية والحصينة

Variables	LDF	QDF	RLM CD	RQM CD	RL RM CD	RQ RM CD	RL MVE	RQ MVE
constant	4.21	-180.09	4. 34	-303.52	-403.21	6.55	102.65	-301.65
Age	19.95	42.62	1.34	1.56	0.03	0.68	1.05	1.06
Weight	0.29	0.44	1.94	2.16	2.03	2.29	1.18	0.77
Urea	3.81	4.61	1.93	1.94	1.88	1.88	0.990	0.76
Blood. Pressure	2.56	3.01	1.77	2.56	2.26	4.14	0.78	-1.03
Sugar	0.01	0.01	1.81	1.88	1.84	2.90	1.28	1.39
Cholesterol	1.29	1.08	2.09	1.85	2.26	1.54	1.17	0.73

وان نتائج احتمال خطأ التصنيف لجميع متغيّرات الدراسة لبيانات (الذبحة الصدرية والجلطة القلبية) لأمراض الذبحة الصدرية والجلطة القلبية للمجموعتين الأولى والثانية موضحة في الجداول رقم **(6)(7)**.

جدول رقم (6)

احتمال خطأ التصنيف لبيانات الذبحة الصدرية المجموعة الأولى

RQMVE	RLMVE	RQRMCD	RLRMCD	RQMCD	RLMCD	QDF	LDF
0.30	0.36	0.09	0.10	0.29	0.32	0.42	0.53

جدول رقم (7)

احتمال خطأ التصنيف لبيانات الجلطة القلبية المجموعة الثانية

RQMVE	RLMVE	RQRMCD	RLRMCD	RQMCD	RLMCD	QDF	LDF
0.23	0.26	0.04	0.08	0.19	0.22	0.31	0.42

ويلاحظ من الجدولين **(٦،٧)**. إن مقدر RMCD الحصين، لبيانات(تصاب الشرايين والجلطة الدماغية) لأمراض القلب للمجموعتين الأولى والثانية تفوق بإعطائه اقل احتمال لخطأ التصنيف.

٤. الاستنتاجات:

في ظل تحليل نتائج الجانب التطبيقي تم التوصل إلى الاستنتاجات الآتية :

١. تم التوصل الى أن مقدر أصغر محدد تباين مشترك معاد الأوزان (RMCD) الحصين حقق نتائج كفوءة ودقيقة إذا أعطى أقل احتمال لخطأ التصنيف للدالة التمييزية التربيعية الحصينة (RQDF) فالدالة التمييزية الخطية الحصينة (RLDF) لأمراض القلب للمجموعتين الأولى والثانية ، يليه مقدر أصغر محدد تباين مشترك (MCD) الحصين ثم مقدر أصغر حجم مجسم قطع بيضوي (MVE) الحصين.

٢. إن مقدر أصغر محدد تباين مشترك معاد الأوزان (RMCD) الحصين أكثردقة من استخدام مقدري (MCD) الحصين ،(MVE) الحصين.

٥. التوصيات:

١. أوصي باستخدام مقدر أصغر محدد تباين مشترك معاد الأوزان (RMCD) الحصين وفي حالة تعدد المجتمعات.

٢. أوصي بتطبيق الأساليب الأخرى في الدوال التمييزية كالدالة التمييزية اللوجستية.

المصادر

١. تشارلز، وأيرتج وآخرون، (1990). "الموسوعة الطبية الحديثة "الطبعة الثانية ، مؤسسة سجل العرب ،القاهرة .

٢. رشيد, ظافر حسين, خمو، خلود يوسف، حميد، رند سليم،العدد (26) لسنة (2001)، "استخدام بعض الطرائق التمييزية الحصينة في تشخيص أمراض قرحة ألاثني عشر"، بحث منشور، مجلة العلوم الاقتصادية والإدارية، كلية الإدارة والاقتصاد، جامعة بغداد.

٣. النداوي ، سرى صباح، (٢٠٠٨). "مقارنة بعض المقدرات الحصينة في الدوال التمييزية مع تطبيق عملي ، رسالة ماجستير , كلية الإدارة والاقتصاد ، جامعة بغداد .

3. Croux, C.& Hasbrouck, G.,(2001), "A Note on finite–sample efficiencies of Estimators for the Minimum Volume Ellipsoid ", Submitted.

4. Hubert, M., Driessen, K.V., (2004),"Fast and Robust Discriminant Analysis", Computational statistics and Data Analysis, vol .45, 301-20.

5. Neil.H.Timm. (2002),Applied Multivariate analysis,John Wiley & sons.

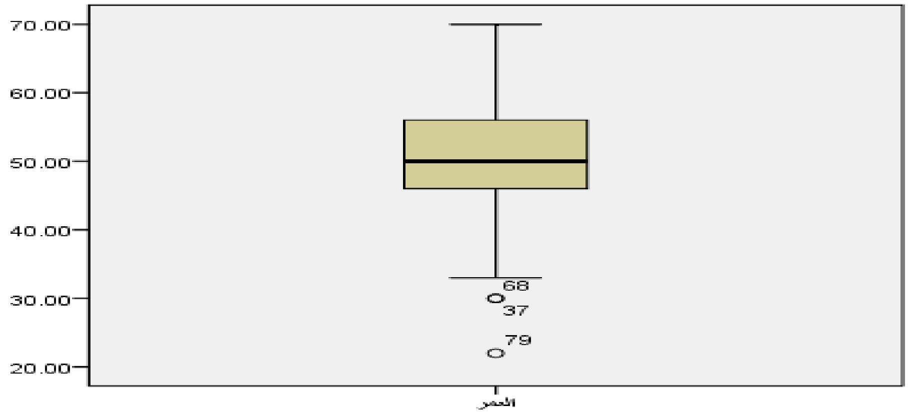
6. Pison, G., Van Aelst, S., and Willems, G. (2002b), “Small Sample Corrections for LTS and MCD,” Metrika, 55, 111-123.

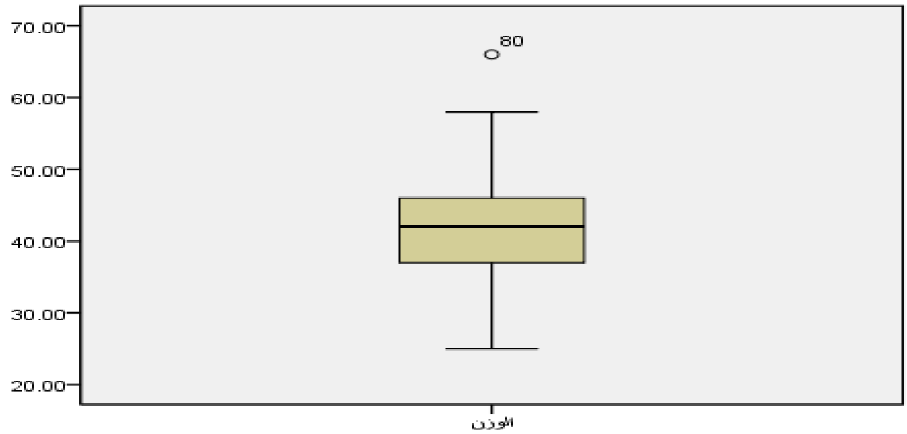
7. Rousseeuw, P.J., (1985) "Multivariate Estimation with High Breakdown Point", Mathematical Statistics and Application, B., pp. 283-397.

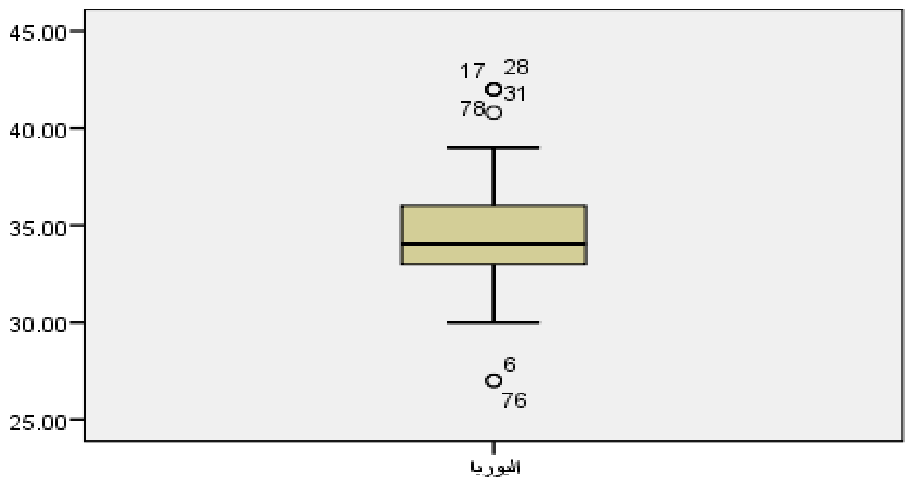
8. Rousseeuw, P.J.&Leroy,A.M.,(1987)"Robust Regression and Outlier Detection", John Wiley & Sons, New York.

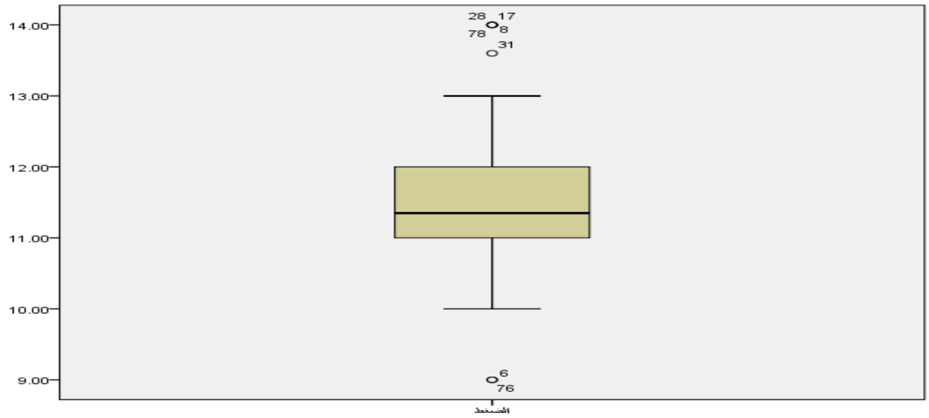
9. Rousseeuw, P.J., &KatrinV.D.,(1999)"A fast Algorithm for the Minimum Covariance Determinant Estimator", Techno metrics, 41, pp. 212-223.

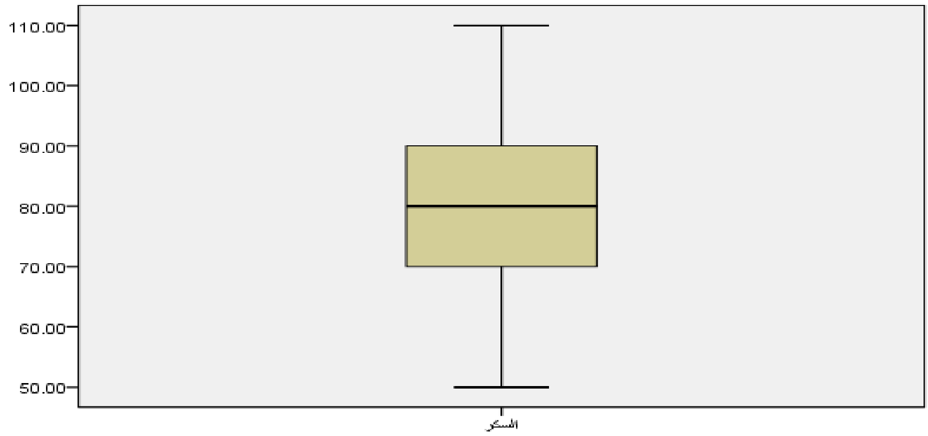
الملحق ١ الكشف عن وجود القيم الشاذة

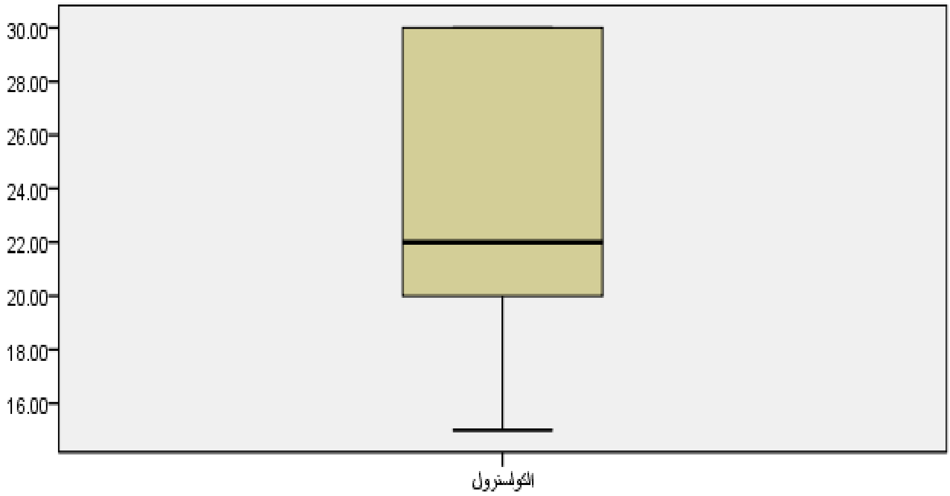







14.00
13.00
12.00
11.00
10.00
9.00
28
17
8
78
31
6
76
الضغط


110.00
100.00
90.00
80.00
70.00
60.00
50.00
السكر


30.00
28.00
26.00
24.00
22.00
20.00
18.00
16.00
الكولسترول