

إستخدام بعض طرائق التمييز الحصينة لتشخيص

أمراض سرطان الدم (اللوكميا)*

أ.م.د. حمزة اسماعيل شاهين*

م.م. دريد حسين بدر الياسين**

المستخلص

فكرة البحث تتضمن تشخيص بعض انواع سرطان الدم (اللوكميا) بأستخدام طرائق تمييز حصينة تعتبر مهمة لقلّة تناولها في البحوث في قطرنا، اضافة لكون مرض سرطان الدم (اللوكميا) يعتبر من اكثر امراض السرطان شيوعاً واوسعها انتشاراً في العراق حيث تكون معدلات الوفيات فيه عاليه جداً بين كبار السن والاطفال مقارنة بالامراض الاخرى.

وتم التركيز في هذا البحث على ايجاد مقدرات حصينة لمعلمتي الموقع M (موجه المتوسط) والقياس Σ (مصفوفة التباين والتباين المشترك) من خلال استخدام طرائق تقدير حصينه منها طريقة مقدر أصغر محدد تباين مشترك (MCD) ، وطريقة مقدر أصغر محدد تباين مشترك معاد الاوزان (RMCD) وطريقة مقدر أصغر حجم مجسم قطع بيضوي (MVE) وبالتالي استخدام دوال تمييز حصينه لما لها من دور كبير في التحصين ضد وجود الشواذ في البيانات ومقارنتها مع دوال التمييز التقليدية للوصول الى افضل دالة تمييز لتشخيص بعض انواع مرض سرطان الدم (اللوكميا) معتمدين على بعض انواع المرض والصفات والمتغيرات المصاحبه للمرض في بناء نموذج احتمالي للتمييز بين نوعين من سرطان الدم(اللوكميا) هما سرطان الدم النخاعي الحاد وسرطان الدم اللمفاوي الحاد .

الدم(اللوكميا) والمقدم الى كلية الادارة والاقتصاد الجامعة المستنصرية (2009) .

* أستاذ مساعد/ الجامعة المستنصرية /كلية الادارة والاقتصاد/قسم الإحصاء

** مدرس مساعد/ جامعة البصرة

مقبول للنشر بتاريخ 2010/5/9

Abstract

The idea of this research has been contained the diagnostic some of leukemia using robut methods discriminant that leukemia is very dangers and widely in Iraqi.

In this research an attempt has been made to find robust estimators namely :location parameter $\underline{\mu}$ (mean vector) and scall parameter Σ (var-cov matrix) depend upon using robust methods estimation .Of these methods are minimum covariance determinant (MCD) ,reweighted minimum covariance determinant (RMCD)and minimum volume ellipsoid (MVE) as they have vital role in robust against outliers ,Moreover comparing these methods with the classical discriminant function for diagnoses two types of leukemia which are acute myeloid leukemia and acute lymphoblastic leukemia based on some variables which it is belongs to this sickness.

1. المقدمة وصف البحث:

كما هو معروف أن التحليل المميز يستخدم بشكل رئيس في عملية تصنيف مفردة واحده أو اكثر الى أحد المجتمعات بأقل احتمال لخطأ التصنيف. وفي قواعد التمييز التقليدي غالباً ما يتم الاعتماد على المتوسط التجريبي كتقدير لمعلمة الموقع ($\underline{\mu}$) وعلى مصفوفة التباين والتباين المشترك للبيانات كتقدير لمعلمة القياس (Σ). ومن المعروف ان هذين التقديرين يتأثران بشدة بوجود عدد قليل من البيانات الشاذة وبالتالي فإن النموذج المقدر سوف يعكس خصائص غير طبيعية.

ومع تطور الاساليب الحصينة ظهرت طرائق تقدير حصينة لمعلمتي الموقع والقياس للتوزيع الطبيعي متعدد المتغيرات وبالتالي فإن هذه الطرائق سوف تزودنا بطرق جديدة بالتعامل مع البيانات حيث تعمل اساليب تحسين دوال التمييز على تقليل الشواذ على التقديرات وبالتالي تزودنا بتقديرات تمتاز بكونها قليلة التأثر بالملاحظات غير العادية .

لقد درست دوال التمييز في تشخيص بعض الامراض التي تشترك في اعراض عديده يصعب التمييز بينها من قبل العديد من الباحثين فعلى سبيل المثال في عام (2001) قام الباحثون (ظافر،خلود،رند)⁽²⁾ بأستخدام اربعة دوال تمييزيه حصينه وهي دالة التمييز الطبيعي الحصين

والتربيعي الحصين والتميز الطبيعي للرتب الحصين والتميز التربيعي للرتب الحصين لتشخيص مرضى قرحة الاثنى عشر وتوصلوا الى ان نموذج التميز التربيعي الحصين هو اكثر كفاءة من بقية النماذج في تشخيص أمراض قرحة الاثنى عشر.

يهدف بحثنا بالدرجة الاساس الى تشخيص بعض انواع سرطان الدم (اللوكيميا) بأستخدام طرائق تمييز حصينة تعتبر مهمة لقلّة تناولها في البحوث في قَطرنا، إضافة لكون مرض سرطان الدم (اللوكيميا) يعتبر من اكثر امراض السرطان شيوعاً ووسعها انتشاراً في العراق حيث تكون معدلات الوفيات فيه عاليه بين كبار السن والاطفال مقارنة بالامراض الاخرى.

كذلك يهدف البحث التركيز على استخدام طرائق تقدير حصينة لمعلمتي الموقع والقياس منها طريقة أصغر محدد تباين مشترك (MCD) وطريقة أصغر محدد تباين مشترك معاد الاوزان (RMCD) وطريقة مقدر اصغر حجم مجسم قطع بيضوي (MVE) وبالتالي استخدام دوال تمييز حصينه لما لها من دور كبير في التحصين ضد وجود الشواذ في البيانات ومقارنتها مع دوال التمييز التقليدية للوصول الى افضل دالة تمييز لتشخيص بعض انواع سرطان الدم (اللوكيميا) معتمدين على بعض انواع المرض والصفات والمتغيرات المصاحبه للمرض في بناء نموذج احتمالي للتمييز بين نوعين من سرطان الدم هما سرطان الدم النخاعي الحاد وسرطان الدم اللمفاوي الحاد.

2. طرائق التقدير الحصينة:

1.2 : طريقة مقدر أصغر محدد تباين مشترك (MCD) :

(Minimum Covariance Determinant Estimator)

طريقة مقدر أصغر محدد تباين مشترك الحصين (MCD) هي إحدى طرائق التقدير لحصينة لمعلمتي الموقع والقياس وتستخدم لتقدير موجه المتوسط (μ) ومصفوفة التباين والتباين المشترك (Σ) (في حالة البيانات تتبع التوزيع الطبيعي متعدد المتغيرات وتقوم فكرة هذه الطريقة على اساس تقدير موجه المتوسط ومصفوفة التباين والتباين المشترك لعينات جزئية بحجم (h) من المشاهدات الكلية وحساب المحددات لتلك المصفوفات فالمصفوفة التي تمتلك اصغر محدد تعتبر هي افضل مقدر من بين المقدرات الاخرى لمصفوفة التباين والتباين المشترك للمجموعة ويطلق عليها بمقدر أصغر محدد تباين مشترك الحصين (MCD, Rousseeuw, 1985)

(1985)⁽⁷⁾

ويفضل اختيار قيمة لـ (h) تساوي (P+1) لأنها تضمن الحصول على مقدرات حصينة تمتلك أعلى نقطة انهيار (Breakdown Point) تصل إلى (50%).
وأن أعداد العينات الجزئية التي يتم اختيارها يكون مساوياً إلى:

$$....(1) H = C_{P+1}^n$$

وان مقدرات أصغر محدد تباين مشترك (MCD) تتمتع بخاصية ثبات التغيرات (Affine Equivariant) وتكون المقدرات كالأتي: (Hubert, Driessen, 2004)⁽⁴⁾

$$\bar{X}_{MCD} = \frac{1}{h} \sum_{i=1}^h x_i \quad \dots(2)$$

$$\hat{\Sigma}_{MCD} = dp \times \frac{1}{h} \sum_{i=1}^h (x_i - \bar{X}_{MCD})(x_i - \bar{X}_{MCD})' \quad \dots(3)$$

إذ أن: dp: عامل الاتساق المحدد ويحسب كالأتي: (Pison, Van Aelst, 2002)⁽⁶⁾

$$....(4) \left[1 + \frac{15}{(n-p)} \right]^2 \times \frac{1}{\chi_{p, \alpha}^2}$$

إذ أن:

P: عدد المتغيرات.

α: مستوى المعنوية.

2.2: طريقة مقدر أصغر محدد تباين مشترك معاد الأوزان (RMCD):

(Reweighted Minimum Covariance Determinant Estimator)

وتقوم هذه الطريقة بنفس فكرة طريقة مقدر أصغر محدد تباين مشترك (MCD)، حيث

يتم حساب تقديرات ابتدائية لموجه المتوسط (μ) ومصفوفة التباين والتباين المشترك (Σ) للمجموعات الجزئية بحجم (h) واختيار مصفوفة تباين مشترك التي تمتلك أصغر محدد وتعتبر

بمماثلة مقدر أولي لمصفوفة التباين المشترك (Σ) ويعبر عنها بـ $(\hat{\Sigma}_{j,0})$ وكذلك يعبر عن موجه

$$\text{المتوسط} \quad (8) \text{ (Rousseeuw , Leroy, 1987): } (\bar{x}_{j,0})$$

إذ أن:

$$\dots(5) \bar{x}_{j,0} = \frac{1}{h} \sum_{i=1}^h x_{ij}$$

$$\hat{\Sigma}_{j,0} = dp \times \frac{1}{h} \sum_{i=1}^h (x_{ij} - \bar{x}_{j,0})(x_{ij} - \bar{x}_{j,0})' \dots(6)$$

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

وفقا لهذه التقديرات الأولية يتم حساب وزن لكل مشاهدة وكالاتي:

$$w_i = \begin{cases} 1 & \text{if } \sqrt{(x_{ij} - \bar{x}_{j,0})' \hat{\Sigma}_0^{-1} (x_{ij} - \bar{x}_{j,0})} \leq \chi_{P,1-\alpha}^2 \dots(7) \\ 0 & \text{if } o.w \end{cases}$$

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

$$\bar{X}_{RMCD} = \frac{\sum_{i=1}^h w_{ij} x_{ij}}{\sum_{i=1}^h w_{ij}} \dots(8)$$

$$\hat{\Sigma}_{RMCD} = dp \times \frac{\sum_{i=1}^h w_{ij} (x_{ij} - \bar{X}_{RMCD})(x_{ij} - \bar{X}_{RMCD})'}{\sum_{i=1}^h w_{ij}} \dots(9) \quad \sum_{i=1}^h w_{ij} \neq 0$$

هذا وتتصف مقدرات أصغر محدد تباين مشترك معاد الأوزان (RMCD) بخاصية ثبات التغير (Affine Equivariant). والتقديرات الموزونة الجديدة تقود إلى كفاءة إحصائية أفضل. (9)

. (Rousseeuw, Katrin, 1999)

3.2 : طريقة مقدر أصغر حجم مجسم قطع بيضوي (MVE):

(Minimum Volume Ellipsoid Estimator)

إن فكرة حساب مقدر (MVE) موضحة كالآتي: (Rousseeuw , 1985) ⁽⁷⁾ يتم تحديد العينات الجزئية بحجم (P+1) من (المشاهدات) أي أننا نقوم باختيار عدد من العينات الجزئية (المصفوفات) مساوي إلى (C_{P+1}^n) ثم لكل عينة جزئية (J) $J = \{i_1, i_2, \dots, i_{p+1}\}$ of $\{1, 2, \dots, n\}$ يتم استخراج موجه المتوسط $\mu_{(j)}$ ومصفوفة التباين والتباين المشترك $\Sigma_{(j)}$ وكالاتي:

$$\dots(10) \bar{X}_J = \frac{1}{P+1} \sum_{i \in j} x_i$$

$$\dots(11) \Sigma_J = \frac{1}{P} \sum_{i \in j} (x_i - \bar{X}_J)(x_i - \bar{X}_J)'$$

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

$$\dots(12) D_i^2(J) = (x_i - \mu_j)' \Sigma_j^{-1} (x_i - \mu_j)$$

وبعد استخراج $D_i^2(J)$ لكافة مشاهدات العينة،. وتم يتم اختيار أفضل مجموعة جزئية مثالية معرفة كما في الصيغة

$$\dots(13) J = \arg \min_{i \in j} \det(\Sigma_j D_{(hp)}^2(J))$$

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

$$\dots(14) \hat{\mu}_i = \mu_j(i)$$

$$\dots(15) \hat{\Sigma}_i = \Sigma_j(i) D_{(hp)}^2(j(i))$$

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

$$\bar{X}^c_n = \frac{1}{hp} \sum_{i=1}^{hp} \bar{X}(i) \quad ..(16)$$

$$\hat{\Sigma}^c_n = \frac{1}{hp} \sum_{i=1}^{hp} \hat{\Sigma}(i).j \quad ..(17)$$

هنا نأخذ فقط النصف الأصغر من المقدرات الحصينة في العينات النهائية
(Croux,Haesbroeck, 2001)⁽³⁾.

3. قواعد التمييز الحصينة: (Robust Discriminant Rules)

1.3: قواعد التمييز الخطية الحصينة:

(Robust Linear Discriminant Rules)

1-1-3: طريقة مقدر أصغر محدد تباين مشترك (MCD):

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

$$\bar{x}_{j,MCD} = \frac{\sum_{i=1}^h x_{ij}}{h}, \quad j=1,2,\dots,L \quad(18)$$

$$\hat{\Sigma}_{j,MCD} = dp \times \frac{\sum_{i=1}^h (x_{ij} - \bar{x}_{j,MCD})(x_{ij} - \bar{x}_{j,MCD})'}{h}$$

وان مقدار احتمال خطأ التصنيف الحصين (\hat{P}_j^R) ويمكن أن يقدر:

$$\dots(20) \hat{P}_j^R = \frac{\tilde{n}_j}{\tilde{n}}$$

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:

$$\dots(21) \hat{\Sigma}_{PMCD} = \frac{\sum_{j=1}^L n_j \hat{\Sigma}_{j,MCD}}{\sum_{j=1}^L n_j}$$

إذ إن قاعدة التمييز الخطية الحصينة (RLDR) وفق المقدرات في الصيغ (18)، (20)، (21)

تصبح^[4]:

بتعيين x إلى \prod_k وإذا كانت:

$$\hat{d}_k^{RL}(x) > \hat{d}_j^{RL}(x), \forall j=1, \dots, L, j \neq k \dots(22)$$

إذ إن:

$$\begin{aligned} \hat{d}_j^{RL}(x) &= \hat{d}_j^{RL}(x, \hat{\mu}_{j,MCD}, \hat{\Sigma}_{PMCD}) = \\ & \bar{x}'_{j,MCD} \hat{\Sigma}_{PMCD}^{-1} x - \frac{1}{2} \bar{x}'_{j,MCD} \hat{\Sigma}_{PMCD}^{-1} \bar{x}_{j,MCD} + Ln(\hat{p}_j^R) \end{aligned} \dots(23)$$

و عندما $\pi_1 = \pi_2$ وان $(L=2)$ فإن الصيغة (23) تعرف بقاعدة التمييز الحصينة لفشر التي

توصف على النحو الآتي:

$$\begin{cases} x \in \pi_1 & \text{if } (\bar{x}_1 - \bar{x}_2)' \hat{\Sigma}_{\text{PMCD}}^{-1} (x - (\bar{x}_1 + \bar{x}_2)/2) > 0 \\ x \in \pi_2 & \text{if other wise} \end{cases} \dots(24)$$

3-1-2: طريقة مقدر 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 لموجه المتوسط $(\bar{x}_{j,0})$ ، و مصفوفة التباين المشترك $(\hat{\Sigma}_0)$ بطريقة (RMCD) ووفقا لهذه التقديرات الأولية يتم حساب وزن ولكل مشاهدة في المجموعة j ، ويتم بعد ذلك حساب مقدرات أصغر محدد تباين مشترك معاد الأوزان (RMCD) النهائية للملاحظات التي أوزانها (1) حيث تكون التقديرات على النحو الآتي :

$$\bar{x}_{j,\text{RMCD}} = \frac{\sum_{i=1}^{n_j} w_{ij} x_{ij}}{\sum_{i=1}^{n_j} w_{ij}} \dots(25)$$

$$\hat{\Sigma}_{\text{PRMCD}} = \frac{\sum_{j=1}^L \sum_{i=1}^{n_j} w_{ij} (x_{ij} - \bar{x}_{j,\text{RMCD}})(x_{ij} - \bar{x}_{j,\text{RMCD}})'}{\sum_{j=1}^L \sum_{i=1}^{n_j} w_{ij}} \dots(26)$$

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:

$$RD_{ij} = \sqrt{(x_{ij} - \bar{x}_{j,\text{MCD}})' \hat{\Sigma}_{j,\text{MCD}}^{-1} (x_{ij} - \bar{x}_{j,\text{MCD}})} \dots(27)$$

و يعد x_{ij} قيمة شاذة إذا و فقط إذا:

$$RD_{ij} > \sqrt{\chi_{P,0.975}^2} \dots(28)$$

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

$$\tilde{n} = \sum_{j=1}^g \tilde{n}_j \quad \dots(29)$$

ومن ثم فإن تقدير احتمال خطأ التصنيف الحصين تم تقديره في الصيغة (20) ، وللحصول على تقديرات مصفوفة التباين المشترك الأولية $\hat{\Sigma}_0$ في الصيغة (7) يكون بدمج مصفوفة التباين والتباين المشترك لكل مجموعة بعد الحصول على تقديرات $\hat{\mu}_{j,MCD}$ و $\hat{\Sigma}_{j,MCD}$ ولكل مجموعة (Pooled Var-Cov Matrix) وعلى وفق الصيغة (21) وان قاعدة التمييز الخطية الحصينة (RLDR) على وفق المقدرات (20) ، (25) ، (26) تصبح [4].

بتعيين x إلى π_k وإذا كانت :

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

إذ إن :

$$\begin{aligned} \hat{d}_j^{RL}(x) &= \hat{d}_j^{RL}(x, \hat{\mu}_{j, RMCD}, \hat{\Sigma}_{PRMCD}) = \\ & \bar{x}'_{j, RMCD} \hat{\Sigma}_{PRMCD}^{-1} x - \frac{1}{2} \bar{x}'_{j, RMCD} \hat{\Sigma}_{PRMCD}^{-1} \bar{x}_{j, RMCD} + Ln(\hat{p}_j^R) \end{aligned} \quad \dots(30)$$

3-1-3: طريقة مقدر أصغر حجم مجسم قطع ناقص (MVE):

على وفق هذه الطريقة يتم لكل مجموعة j إذ إن ($j=1,2,3,\dots,L$) تقدير موجه متوسط ($\underline{\mu}_j$) ومصفوفة التباين والتباين المشترك ($\hat{\Sigma}_j$) وخطا التصنيف (P_j) باستعمال مقدرات طريقة (MVE) للمجموعة الجزئية (h_p) التي تمتلك أصغر محدد من بين (H) من المجموعات الجزئية الكلية لكل مجموعة من مسافة الوسيط (Median Distance) إذ تكون المقدرات على النحو الآتي:

$$\dots(31) \bar{x}_{j, MVE} = \frac{\sum_{i=1}^{hp} x_{ij}}{hp}, \quad j=1,2,\dots,L$$

$$\dots(32) \hat{\Sigma}_{j, MVE} = \frac{1}{hp} \sum_{i=1}^{hp} \hat{\Sigma}_{(i).j}, \quad j=1,2,\dots,L$$

وان مقدار احتمال خطأ التصنيف الحصين (\hat{P}_j^R) تم تقديره كما في الصيغة (20) .
وبعدها يتم حساب مصفوفة التباين والتباين المشترك المدمجة العامة
(Pooled Var-Cov Matrix) على وفق الصيغة الآتية:

$$\dots(33) \hat{\Sigma}_{PMVE} = \frac{\sum_{j=1}^L n_j \hat{\Sigma}_{j, MVE}}{\sum_{j=1}^L n_j}$$

إذ إن قاعدة التمييز الخطية الحصينة (RLDR) وعلى وفق المقدرات في الصيغ
(20)، (31)، (33) .

بتعيين x إلى \prod_k وإذا كانت :

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

إذ إن :

$$\dots(34) \hat{d}_j^{RL}(x) = \hat{d}_j^{RL}(x, \hat{\mu}_{j, MVE}, \hat{\Sigma}_{PMVE}) =$$

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

2.3: قواعد التمييز التربيعية الحصينة:

(Robust Quadratic Discriminant Rules)

1-2-3: طريقة مقدر أصغر محدد تباين مشترك (MCD):

قاعدة التمييز التربيعية الحصينة (RQDR) وعلى وفق هذه الطريقة تصبح .

بتعيين x إلى \prod_k وإذا كانت :

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

إذ إن :

$$\begin{aligned} \hat{d}_j^{RQ}(\underline{x}) &= \hat{d}_j^{RQ}(\underline{x}, \hat{\mu}_{j,MCD}, \hat{\Sigma}_{j,MCD}) \\ &= -\frac{1}{2}Ln \left| \hat{\Sigma}_{j,MCD} \right| - \frac{1}{2}(\underline{x} - \bar{x}_{j,MCD})' \hat{\Sigma}_{j,MCD}^{-1} (\underline{x} - \bar{x}_{j,MCD}) + Ln(\hat{p}_j^R) \end{aligned} \quad \dots(35)$$

2-2-3: طريقة مقدر أصغر محدد تباين مشترك معاد الأوزان (RMCD):

قاعدة التمييز التربيعية الحصينة (RQDR) على وفق هذه الطريقة تصبح .

بتعيين x إلى \prod_k وإذا كانت :

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

إذ إن :

$$\begin{aligned} \hat{d}_j^{RQ}(\underline{x}) &= \hat{d}_j^{RQ}(\underline{x}, \hat{\mu}_{j,RMCD}, \hat{\Sigma}_{j,RMCD}) \\ &= -\frac{1}{2}Ln \left| \hat{\Sigma}_{j,RMCD} \right| - \frac{1}{2}(\underline{x} - \bar{x}_{j,RMCD})' \hat{\Sigma}_{j,RMCD}^{-1} (\underline{x} - \bar{x}_{j,RMCD}) + Ln(\hat{p}_j^R) \end{aligned} \quad \dots(36)$$

3-2-3 : طريقة مقدر أصغر حجم مجسم قطع ناقص (MVE):

قاعدة التمييز التربيعية الحصينة (RQDR) على وفق هذه الطريقة تصبح .

بتعيين x إلى \prod_k وإذا كانت :

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

إذ إن:

$$\begin{aligned} \hat{d}_j^{RQ}(x) &= \hat{d}_j^{RQ}(x, \hat{\mu}_{j,MVE}, \hat{\Sigma}_{j,PMVE}) \\ \dots(37) \quad &= -\frac{1}{2} \text{Ln} \left| \hat{\Sigma}_{j,MVE} \right| - \frac{1}{2} (x - \bar{x}_{j,MVE})' \hat{\Sigma}_{j,MVE}^{-1} (x - \bar{x}_{j,MVE}) + \text{Ln}(\hat{p}_j^R) \end{aligned}$$

4. اختبار صالة التمييز:

ومن هذه الاختبارات:

1-4 : اختبار معنوية الصالة المميزة (تساوي المجموعات):

تم استخدام مقياس مربع كاي وتكون صيغته كالتالي: (Neil.H.Timm , 2002)⁵.

$$\chi^2 = -[n - 1 - 1/2(P + g)] \log \hat{e} \quad \dots(38)$$

إذ أن:

P : المتغيرات المقدره.

g : عدد المجاميع .

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

2.4 : اختبار تساوي مصفوفة التباين والتباين المشترك لجميع المجاميع: وتكون

احصاءة هذا الاختبار كما يلي: (Neil.H.Timm, 2002)⁽⁵⁾.

$$\mu = \left(\sum_{i=1}^k V_i \right) \text{Ln} |S| - \sum_{i=1}^k (V_i \text{Ln} |S_i|) \quad \dots(39)$$

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

$$C^{-1} = 1 - \frac{2m^2 + 3m - 1}{6(m+1)(k-1)} \left[\sum_{i=1}^k \frac{1}{V_i} - \frac{1}{\sum_{i=1}^k V_i} \right] \dots(40)$$

$$\frac{1}{2} (k-1) m (m+1)$$

سوف نحصل على مقياس يتوزع بالتقريب توزيع χ^2 وبدرجة حرية إذ أن:

m: عدد المتغيرات المدروسة في المصفوفة.

k: عدد المصفوفات المختارة .

5- الجانب التطبيقي

1.5 : فكرة موجزة عن مرض سرطان الدم (Leukemia's):

اللوكيميا هي النمو الشاذ لخلايا الدم البيضاء، وهي الخلايا الحديثة الولادة غير الناضجة، وعند حدوث اللوكيميا تبقى هذه الخلايا غير ناضجة، ويتوقف نموها عند حد معين بأحد أطوار نموها، ثم تزيد زيادة مفرطة في الدورة الدموية (أي زيادة هائلة في أعدادها بطور نمو غير بالغ)، وهذا ما يُطلق عليه بسرطان الدم "ابيضاض الدم" لطغيان كرات الدم البيضاء فيه على سائر الخلايا الأخرى المكوّنة للدم.. ويعد مرض السرطان من أكثر الأمراض شيوعاً وأوسعها انتشاراً في العراق، والعالم والأسباب الكامنة وراء الإصابة بسرطان الدم غير معلومة حتى الآن، لكن العلماء والباحثين يشكون باحتمالية وجود عوامل متعددة: فيروسية، جينية، بيئية، ومناعية التي لها دخل في الإصابة بهذا المرض. (تشارلز، و.أيرتج وآخرون 1970) (1).

2.5 : وصف العينة والمتغيرات المستخدمة وأساليب جمع البيانات:

تم اعتماد ملفات المرضى (طبلات المصابين) في المستشفى التعليمي ومستشفى الجمهوري في محافظة البصرة للأعوام (2005 و 2006 و 2007)، واشتملت بيانات البحث على (164) مريضاً، وقد تم سحب عينة عشوائية من المجتمع الأول (سرطان الدم النخاعي الحاد AML) بحجم (n1=80) ومن المجتمع الثاني (سرطان الدم اللمفاوي الحاد ALL) بحجم (n1=84)، والتي تمثل

مجموعتين من التشخيص لأمراض سرطان الدم (اللوكيميا) الذي يمثل المتغير المعتمد (متغير الاستجابة) والذي يتمثل بنوع المرض السرطاني (سرطان الدم النخاعي الحاد=1) و(سرطان الدم اللمفاوي الحاد=2) وعدد المتغيرات الإيضاحية (المستقلة) التي تم الاتفاق عليها مع الأطباء أصحاب الاختصاص هي (7) متغيرات يمكن توضيحها بإيجاز على النحو الآتي:

1. العمر (age).
2. الوزن (weight).
3. نسبة الكريات الحمراء في الدم % Packed Cell Volume (P.C.V).
4. هيموجلوبين الدم (H.b).
5. معدل الكريات البيضاء في الدم (W.B.C) White Blood Cell.
6. نسبة سرعة ترسب كريات الدم الحمراء (E.S.R) Erythrocyte Sedimentation .Percentage
7. الصفائح الدموية (Platelets Count).

3.5 : نتائج الاختبارات

1. اختبارات تحقق شرط التوزيع الطبيعي:

وقد تم اختبار البيانات من خلال جودة توفيق البيانات في البرنامج الإحصائي الجاهز (SPSS.16) وحسب الاختبارات أظهرت النتائج قيم اختبار كولموكروف سمر نوف إن جميع متغيرات الدراسة لا تتوزع توزيعاً طبيعياً.

2. اختبار وجود أو انعدام القيم الشاذة بطريقة (Box- and -Whisker Plot):

فيما يخص عينة البحث فقد أظهرت نتائج طريقة (Box-and-Whisker Plot) باستخدام البرنامج الإحصائي الجاهز (SPSS.16) وجود القيم الشاذة لخمس متغيرات وبأعداد متفاوتة من طرف واحد أو طرفين للبيانات العمر (Age) والجنس (sex) والوزن (weight) والهيموغلوبين (H.b) ونسبة الكريات الحمراء في الدم % (P.C.V) وبقيّة المتغيرات لم تظهر فيها قيم شاذة.

3. اختبار معنوية الدالة المميزة:

تم اختبار معنوية الدالة المميزة وبالاعتماد على المعادلة (38) كانت قيمة

$\chi^2_{cal} =$ (53.106) وبمستوى معنوية 0.05 وهذا يدل على إن متوسطات المجموعتين غير متساوية، إي يوجد فرق معنوي بين المجموعتين.

4- اختبار تجانس التباينات

تم اختبار تساوي التباينات و بالاعتماد على المعادلتين (39) و (40) فإن

Box's M=128.073 وهذا يدل على عدم التجانس بين المجموعات .

4-5: نتائج تحليل طرائق التمييز لبيانات سرطان الدم (اللوكيميا)

لغرض تطبيق طرائق التقدير الحصينة المذكورة في الجانب النظري من هذه الدراسة، ففقد أثرنا اللجوء إلى عينة سحبت من بيانات خاصة بالجانب الطبي لأنها تحقق الغرض من هذه الدراسة، وذلك يعود إلى أن المجالات الطبية مليئة ببيانات تتصف بان بعض قياساتها أو التسجيلات الخاصة بمتغيراتها تكون متجاوزة لنسبها الطبيعية داخل جسم الإنسان، وهذه القياسات تعد حينئذ شاذة، ونتيجة لعدم توفر برامج جاهزة بهذا الخصوص، فقد قام الباحثان بكتابة برنامج بلغة (Q.Basic) لغرض تطبيق طرائق التقدير الحصينة.

وإن نتائج التقدير لجميع متغيرات الدراسة لنوعي بيانات سرطان الدم النخاعي الحاد والمفاوي الحاد باستخدام دالة التمييز الخطية والتربيعية التقليدية والحصينة ولطرائق التقدير التقليدية والحصينة هي موضحة في الجداول رقم (3) و(4) على التوالي:

جدول رقم (3)

نتائج تحليل طرائق التمييز لبيانات سرطان الدم (اللويميا)
النخاعي الحاد بأستعمال طرائق التمييز التقليدية والحصينة

التمييز التربيعي الحصين بطريقة MVE	التمييز الخطي الحصين بطريقة MVE	التمييز التربيعي الحصين بطريقة RMCD	التمييز الخطي الحصين بطريقة RMCD	التمييز التربيعي الحصين بطريقة MCD	التمييز الخطي الحصين بطريقة MCD	التمييز التربيعي	التمييز الخطي	Variables
-211.43	114.49	4.4319	-137.72	-242.35	4.2108	-160.08	1.3834	constant
1.12	1.11	.99	0.96	2.01	2.29	79.25	80.56	Age
1.01	1.14	3.07	0.69	2.12	51.45	4.73	5.81	Weight
0.4	0.96	1.1	1.06	1.97	2.61	18.75	21.93	H.b
0.79	0.68	-7.30	0.98	1.92	624.23	13.74	14.41	(P.C.V)%
0.71	0.87	2.77	1.30	2.22	682.78	.01	.02	W.B.C
0.77	1.12	1.39	1.30	1.93	419.34	4.93	5.85	E.S.R
0.81	0.98	1.18	1.05	1.97	2.32	8.05E-	1.9E-5	Platelets

جدول رقم (4)

نتائج تحليل طرائق التمييز لبيانات سرطان الدم (اللويميا) للمفاوي الحاد
بأستعمال طرائق التمييز التقليدية والحصينة

التمييز التربيعي الحصين بطريقة MVE	التمييز الخطي الحصين بطريقة MVE	التمييز التربيعي الحصين بطريقة RMCD	التمييز الخطي الحصين بطريقة RMCD	التمييز التربيعي الحصين بطريقة MCD	التمييز الخطي الحصين بطريقة MCD	التمييز التربيعي	التمييز الخطي	Variables
-301.65	102.65	6.55	-403.21	-303.52	4.34	-180.09	4.21	constant
1.06	1.05	0.68	0.03	1.56	1.34	42.62	19.95	Age
0.77	1.18	2.29	2.03	2.16	1.94	0.44	0.29	Weight
0.76	0.990	1.88	1.88	1.94	1.93	4.61	3.81	H.b
-1.03	0.78	4.14	2.26	2.56	1.77	3.01	2.56	(P.C.V)%
1.39	1.28	2.90	1.84	1.88	1.81	0.01	0.01	W.B.C
0.73	1.17	1.54	2.26	1.85	2.09	1.08	1.29	E.S.R
0.85	0.96	1.81	1.84	1.81	2.04	-7.7E-6	1.3E-5	Platelets

من نتائج الجدولين (3و4) يتضح من قيم تقديرات المعالم الاهمية التمييزية لمتغير معدل الكريات البيضاء في الدم (W.B.C) في تشخيص مرض سرطان الدم في جميع دوال التمييز الحصينة يليه متغير نسبة الكريات الحمراء في الدم (P.C.V%) ثم متغير نسبة سرعة ترسب

كريات الدم الحمراء (E.S.R) ، وان نتائج احتمال خطأ التصنيف لجميع متغيرات الدراسة لنوعي بيانات سرطان الدم النخاعي الحاد واللمفاوي الحاد موضحة في الجدول رقم (5) .

جدول رقم (5)

احتمال خطأ التصنيف للدوال المميزة

نوع المرض	التمييز الخطي	التمييز التريبي	التمييز الخطي الحصين بطريقة MCD	التمييز التريبي الحصين بطريقة MCD	التمييز الخطي الحصين بطريقة RMCD	التمييز التريبي الحصين بطريقة RMCD	التمييز الخطي MVE	التمييز التريبي الحصين بطريقة MVE
سرطان الدم النخاعي الحاد	0.53	0.42	0.32	0.29	0.10	0.09	0.36	0.30
سرطان الدم النخاعي اللمفاوي	0.42	0.31	0.22	0.19	0.08	0.04	0.26	0.23

ويلاحظ من الجدولين اعلاه ان نموذج التمييز التريبي الحصين بطريقة (RMCD) لبيانات سرطان الدم (اللويميا) النخاعي الحاد واللمفاوي الحاد اعطت اقل احتمال لخطأ التصنيف مقارنة بدوال التمييز الاخرى .

6. الاستنتاجات

- 1- من خلال قيم تقديرات المعالم يلاحظ ان متغير معدل الكريات البيضاء في الدم (W.B.C) هو الاكثر اهمية في تشخيص امراض سرطان الدم النخاعي الحاد واللمفاوي الحاد يليه متغير نسبة الكريات الحمراء في الدم % (P.C.V) ثم متغير نسبة سرعة ترسب كريات الدم الحمراء.
- 2- ان تقديرات نموذج التمييز التريبي الحصين بطريقة (RMCD) اعطت تفوقاً واضحاً في النسب المئوية للتصنيف الصحيح وهذا يتلائم مع اختبار عدم تساوي التباينات اضافة الى وجود بعض القيم الشاذة مما يعطيها الاهمية في اعتمادها كأسلوب مساعد في عملة الشخيص لنوعي سرطان الدم النخاعي الحاد واللمفاوي الحاد.
- 3- ان دالتي التمييز الخطي التقليدية اعطت اعلى نسبة لاحتمال خطأ التصنيف مما يستدعي عدم استخدامهما في عملية التشخيص خصوصاً في المجال الطبي والذي يتطلب الدقة في التشخيص .
- 4- فيما يتعلق بالتقديرات الحصينة نجد ان نسب احتمال خطأ التصنيف قد تحسنت في معظم الدوال المميزة.

7. التوصيات

1. نوصي باستخدام دالة التمييز التربيعي الحصين بطريقة (RMCD) وفي حالة تعدد المجتمعات.
2. نوصي بتطبيق الأساليب الأخرى في الدوال التمييزية كالدالة التمييزية اللوجستية.

8. المصادر

1. تشارلز، و.أيرتج وآخرون (1970) "الموسوعة الطبية الحديثة" الطبعة الثانية، مؤسسة سجل العرب، القاهرة .
2. رشيد، ظافر حسين، خمو، خلود يوسف، حميد، رند سليم، العدد (26) لسنة (2001)، "استخدام بعض الطرائق التمييزية الحصينة في تشخيص أمراض قرحة الأثني عشر"، بحث منشور، مجلة العلوم الاقتصادية والإدارية، كلية الإدارة والاقتصاد، جامعة بغداد.
- 3.Croux,C.& Haesbroeck, 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 Regressionand Outlier Detection", John Wiley & Sons, New York.
- 9.Rousseeuw,P.J.,&KatrinV.D.,(1999)"Afast Algorithm for the Minimum Covariance Determinant Estimator", *Techno metrics*, 41, pp. 212-223.