إستخدام التحليل المميز لتشخيص بعض أمراض العيون

عائدة هادي صالح*

المستخلص :

تم إستخدام التحليل المميز لتصنيف بعض أمراض العيون إلى ثلاث مجموعات إعتماداً على متغيرات ذات صفات تمييزية . تضمنت الدراسة سحب عينة عشوائية لكل مجموعة على حدة ، وباستخدام الدالة المميزة الخطية تم تصنيف بعض أمراض العيون على أساس مجموعة من المتغيرات المصاحبة للمرض .

Abstract

The discriminant ananlysis has been used to classify some of the eye diseases into three groups based on changes of discrimination features . The study has been employed a random sample for each group by using the linear discriminant function , some diseases of eye have been classified on the base of the disease accompanied changes .

هدف البحث :

يهدف البحث إلى تشخيص العوامل المؤثرة في أمراض العيون ومعرفة تأثير كل عامل من هذه العوامل من خلال استخدام الدالة التمييزية الخطية .

* مدرس مساعد / وزارة التعليم العالي والبحث العلمي

مقبول للنشر بتاريخ 2007/11/13

المقدمة :

يعد التحليل المميز Discriminant Analysis أحد طرق تحليل متعدد المتغيرات Multivariate Analysis ، يهتم بدراسة تأثير مجموعة من العوامل في مفردات مجموعات مختلفة متميزة عن بعضها البعض . ويستخدم للتمييز بين مجموعتين أو أكثر تتشابه في الكثير من الصفات على أساس وجود عدد من المتغيرات .

يستفاد من التحليل المميز في تصنيف مفردة واحدة أو أكثر إلى مجتمع أو مجموعة من المجتمعات بالاعتماد على متغيرات لها صفات معينة ، وكذلك يستفاد منه في التعرف على المتغيرات التي تساهم في عملية التصنيف .

يستخدم التحليل المميز في مجالات مختلفة وكمثال على ذلك يستخدم في حالة تصنيف المرضى أو المصابين بمرض معين بالاعتماد على مجموعة من المعلومات عن حالتهم الصحية .

الجانب الطبي :

العين البشرية هي جوهرة وهي أداة النظر التي تمكننا من رؤية الأشياء من حولنا ، إنها هبة من الله لايمكن تقديرها بثمن . إن العوامل البيئية وأساليب وأنماط الحياة والسلوك والممارسات اليومية للفرد قد تسبب أضرار للعين ولكن الخطر الأعظم هو الإلتهاب فأخذ الحيطة والحذر واجب .

الأمراض التي تصاب بها العين (الإلتهابات) هي التراخوما ، الجدري ، الجذام ، التهاب المنظمة الصديدي ، القرنية ، التهاب العصب البصري ، الساد (الماء الأبيض) Cataract ، الزرقاء (الماء الأزرق) Glaucoma ، أمراض الجهاز التناسلي كالسيلان والسفلس . أما التغيرات والشكوى المصاحبة لأمراض العيون هي الألم ، الورم ، الإحمرار ، التغير المفاجئ بالنظر كالرؤيا المزدوجة والتشوش بالرؤيا ، الصداع ، والحساسية الشديدة للضوء . أما أسباب فقدان البصر الجزئي هو الإصابات ، أمراض وراثية ، أمراض خلقية ، أمراض سوء التغذية ، الإلتهابات ، الأورام ، داء الزرقاء ، قصر البصر ، ضعف البصر .

الساد هو عتامة العدسة الشفافة الموجودة داخل العين ، أما الكلوكوما (داء الزرقاء) أو الماء الأزرق هو مرض يصيب العصب البصري نتيجة ارتفاع الضغط بالعين فيحصل نتيجة ذلك تلف في أنسجة العصب البصري ، مما يؤدي الى تكوين بقع عمياء داخل العين (فقد أجزاء من المجال البصري للرؤيا) ، وإذا لم يعالج المرض يحدث تلفاً كلياً في العصب البصري وبذلك تفقد العين قدرتها على الابصار .

يعرف الكلوكوما عند عامة الناس بالماء الأزرق ، وفي الحقيقة فإن تسميته بهذا الأسم خطأ شائع إذ لاتوجد مياه زرقاء بداخل العين ولكن أتت هذه التسمية من مفهوم كلمة الكلوكوما عند الإغريق والتي تعني شلالات زرقاء ، لأن المريض أحياناً يشاهد هالات زرقاء حول مصدر الضوء فيعطي انطباع أن بداخل العين مياه زرقاء . الكلوكوما هي السبب الرئيس للعمى في الأشخاص كبار السن ويمكن منع الإصابة بالعمى لو بدأ العلاج مبكراً .أما انفصال الشبكية **Retinal detachment** فهو عبارة عن انفصال الطبقة الطلائية الصبغية عن الطبقة العصبية الداخلية للشبكية ، يحدث الانفصال نتيجة الى 1-تعويم الشبكية للداخل بسبب وجود سوائل أو نزيف أو مخلفات التهابية تعوم عليها الشبكية . 2- شد الشبكية للداخل نتيجة تلف بالجسم الزجاجي أو الغشاء الهدبي . 3- دفع الشبكية للداخل نتيجة ورم بالشبكية .

المتغيرات التي أستخدمت في الدراسة هي :

1-الجنس (Sex)

2-العمر (Age)

3-ضغط العين (Intra Ocular Pressure) ، يتراوح الضغط الطبيعي الفسيولوجي داخل العين بين (12-19) ملم زئبق ، إن إزدياد الضغط الداخلي للعين عن 19 ملم زئبق يعد بداية الحالة المرضية التي يجب معالجتها وإعادة الفحص بشكل دوري كل شهر لمدة 3 أشهر أو 6 أشهر حسب الحالة المرضية ، وإذا انخفض ضغط العين عن معدله الطبيعي فهذا يؤدي إلى ضمور في العين ، وهذه الحالة نادراً ما تحدث .

4-السكر (Sugar) هذا المتغير مهم جداً ، وله تأثير كبير على الإصابة بأمراض العيون ، المعدل الطبيعي للسكر عند الإنسان يتراوح مابين (4.5-6.5) ملي مول لكل لتر ، فإذا انخفض السكر عن مستواه الطبيعي يسبب الإغماء اما إذا أرتفع فيسبب نزف داخل العين ويسبب العمى . وحيث أن اكتمال الماء الأبيض عند مرضى السكر يكون أسرع مما هو في الأشخاص غير المصابين بالسكر .

5-فقر الدم (Packed Cell Volume) أو هو نسبة الهيموغلوبين في الدم ، النسبة الطبيعية تتراوح مابين (40%-45%) ، يحتاج المريض الذي يتعرض إلى تخدير عام الى نسبة دم طبيعية .

6-اليوريا (Urea) نسبة اليوريا الطبيعية في الجسم هي مابين (5-7) ملي غرام لكل ملي لتر mg/mL ، إذا ارتفعت هذه النسبة تسبب عجز في الكلى .

الجانب النظري :

التحليل المميز هو إسلوب إحصائي لتحليل البيانات متعددة المتغيرات ، يصنف مفردات العينة إلى مجموعتين أو أكثر . الدالة المميزة هي عبارة عن تركيب خطي للمتغيرات المستقلة ، والدالة المميزة هي التي تقوم بعملية التمييز . وعملية التصنيف تأتي بعد عملية تكوين الدالة المميزة ، حيث يعتمد عليها في تصنيف المفردات الجديدة لأحدى المجموعات قيد الدرس بأقل خطأ تصنيف ممكن .

التمييز الخطي هو إحدى حالات التمييز التي يتطلب فيها توفر شرط تساوي التباينات للمجموعات المبحوثة ، فهناك تمييز خطي في حالة مجموعتين ، وتمييز خطي في حالة أكثر من مجموعتين . أما التمييز غير الخطي فيستخدم في حالة عدم تساوي التباينات .

تستخدم دالة التمييز الخطية عندما تكون المجتمعات المبحوثة ذات توزيع طبيعي متعدد المتغيرات بمتجهات متوسطات مختلفة ومصفوفات تباين وتباين مشترك متساوية .

أولاً : الدالة المميزة الخطية في حالة مجموعتين

The Linear Discriminant Function – Two Groups

إن دالة التمييز هي نموذج يمكن صياغته إعتماداً على مؤشرات العينة التي أختيرت مفرداتها بشكل عشوائي ووضعت في مجموعتين مختلفتين ، بواسطة هذه الدالة نستطيع أن نختبر المفردة ونحدد عائديتها إلى أي مجموعة .

فلو فرضنا أن مجال العينة هو W سوف يقسم إلى قسمين :

(R) يعود إلى المجموعة الأولى ، و (W-R) يعود إلى المجموعة الثانية ، أما الحد الفاصل بين المجموعتين فيمكن أن يعود إلى أية مجموعة من هاتين المجموعتين وفي هذه الحالة فإن :

$$f_1(\underline{X}) = f_2(\underline{X})$$

$$Logf_1 = Logf_2$$

وبافتراض أن مفردات المجموعتين تتوزع توزيعاً طبيعياً متعدد المتغيرات بمتوسطي μ_1 ، μ_2 ومصفوفة تباين وتباين مشترك $\sum$.

$$(\underline{X} - \underline{\mu}_1)' \sum^{-1} (\underline{X} - \underline{\mu}_1) = (\underline{X} - \underline{\mu}_2)' \sum^{-1} (\underline{X} - \underline{\mu}_2)$$

$$(\underline{X} - \underline{\mu}_1)' \sum^{-1} (X - \underline{\mu}_1) - (\underline{X} - \underline{\mu}_2)' \sum^{-1} (\underline{X} - \underline{\mu}_2) = 0$$

$$(\underline{X}' - \underline{\mu}_1')'\sum^{-1}(\underline{X} - \underline{\mu}_1) - (\underline{X}' - \underline{\mu}_2')'\sum^{-1}(\underline{X} - \underline{\mu}_2) = 0$$

$$\underline{X}'\sum^{-1}(\underline{X} - \underline{\mu}_1) - \underline{\mu}_1'\sum^{-1}(\underline{X} - \underline{\mu}_1) - X'\sum^{-1}(\underline{X} - \underline{\mu}_2) + \underline{\mu}_2'\sum^{-1}(\underline{X} - \underline{\mu}_2) = 0$$

$$\underline{X}'\sum^{-1}\underline{X} - \underline{X}'\sum^{-1}\underline{\mu}_1 - \underline{\mu}_1'\sum^{-1}\underline{x} + \underline{\mu}_1'\sum^{-1}\underline{\mu}_1 - \underline{X}'\sum^{-1}\underline{X} + \underline{X}'\sum^{-1}\underline{\mu}_2 + \underline{\mu}_2'\sum^{-1}\underline{X} + \underline{\mu}_2'\sum^{-1}\underline{\mu}_2 = 0$$

$$\underline{X}'\sum^{-1}(\underline{\mu}_1 - \underline{\mu}_2) - (\underline{\mu}_1 - \underline{\mu}_2)'\sum^{-1}\underline{X} + (\underline{\mu}_1 + \underline{\mu}_2)'\sum^{-1}(\underline{\mu}_1 - \underline{\mu}_2) = 0$$

$$\underline{X}'\sum^{-1}(\underline{\mu}_1 - \underline{\mu}_2) = (\underline{\mu}_1 - \underline{\mu}_2)\sum^{-1}\underline{X}'$$

$$2\underline{X}'\sum^{-1}(\underline{\mu}_1 - \underline{\mu}_2) + (\underline{\mu}_1 + \underline{\mu}_2)'\sum^{-1}(\underline{\mu}_1 - \underline{\mu}_2) = 0$$

$$\underline{X}'\sum^{-1}(\underline{\mu}_1 - \underline{\mu}_2) - \frac{1}{2}(\underline{\mu}_1 + \underline{\mu}_2)'\sum^{-1}(\underline{\mu}_1 - \underline{\mu}_2) = 0$$

وهي تمثل قيم دالة التمييز عند الحد الفاصل بين المجموعتين عندما تكون معالم المجتمع معلومة ، أما عندما تكون المعالم مجهولة فتقدر اعتماداً على قيمة العينتين n1 و n2 وبأوساط حسابية $\bar{x}_1$ و $\bar{x}_2$.

وبتقدير مصفوفة التباين والتباين المشرك فإن دالة التمييز تكون :

$$\underline{X}'S^{-1}(\underline{\bar{X}}_1 - \underline{\bar{X}}_2)' - \frac{1}{2}(\underline{\bar{X}}_1 + \underline{\bar{X}}_2)'S^{-1}(\underline{\bar{X}}_1 - \underline{\bar{X}}_2) = 0 \quad .....(1)$$

حيث أن :

$$S = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{n_1 + n_2 - 2}$$

$$S = \frac{1}{n_1 + n_2 - 2}(X_1'X_1 - X_2'X_2)$$

S_1 ، S_2 تمثل تقدير مصفوفة التباين المشترك اعتماداً على العينة الأولى والثانية على التوالي .

والمعادلة (1) تمثل الدالة المميزة الخطية المقدرة عند الحد الفاصل بين المجموعتين ، نعتمد على هذه المعادلة في تصنيف أية مفردة جديدة ، فالمفردة أما تعود إلى المجموعة الأولى إذا كـــان

الجانب الأيسر أكبر من صفر ، أو تعود إلى المجموعة الثانية إذا كان أقل من صفر ، وتصنف عشوائياً فيما عدا ذلك .

إذا كان لدينا المتغيرات $(X_1, X_2, ..., X_p)$ فالدالة الخطية تأخذ الصيغة التالية :

$$Y = b_1X_1 + b_2X_2 + ... + b_pX_p$$

حيث Y تمثل تركيبة خطية للمتغيرات التوضيحية تدعى بدالة التمييز وتكتب بالصيغة التالية وعلى شكل مصفوفة :

$$Y = \underline{b}'\underline{X}$$

حيث أن المعاملات $(b_1, b_2, ..., b_p)$ تقدر بحيث تجعل الدالة تعطي أفضل تميز بين المجموعتين وذلك بجعل مربع الفرق بين متوسطي المجموعتين إلى التباين (مجموع المربعات داخل المجموعات) أكبر ما يمكن .

$$Q = \frac{Between \quad groups}{Within \quad groups}$$

$$Q = \frac{\left[\overline{Y}_1 - \overline{Y}_2\right]^2}{\sum_{i=1}^{2}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_i)^2} = \frac{\left[\underline{b}'\mu_1 - \underline{b}'\mu_2\right]^2}{\sum_{i=1}^{2}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_i)^2}$$

حيث إن :

$$\overline{Y}_i = b_1\overline{X}_{i1} + b_2\overline{X}_{i2} + ... + b_p\overline{X}_{ip}$$

وعندما تكون المعلمات غير معلومة نستخدم تقديرات العينة $\overline{X}_i$ بدلاً من μ_i و S بدلاً من $\sum$

حيث إن :

$$\overline{X}_i = (\overline{X}_{i1}, \overline{X}_{i2}, ..., \overline{X}_{ip}) \qquad i = 1,2$$

$$S = \frac{1}{n_1 + n_2 - 2}(X_1{}'X_1 - X_2{}'X_2)$$

$$\therefore \quad Q = \frac{\left[\underline{b}'(\overline{X}_1 - \overline{X}_2)\right]^2}{\sum_{i=2}^{2}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_i)^2}$$

نفرض أن $\underline{\overline{X}_1} - \underline{\overline{X}_2} = \underline{\alpha}$

$$d = \bar{y}_1 - \bar{y}_2$$

$$= \underline{b}'(\overline{X}_1 - \overline{X}_2)$$
$$D = \overline{Y}_1 - \overline{Y}_2 = \underline{b}'\underline{\alpha}$$
$$d^2 = (\underline{b}'\underline{\alpha})^2$$

ومجموع مربعات الفرق داخل المجاميع هو :

$$(Y_{ij} - \bar{y}_i) = b_1(X_{1ij} - \overline{X}_{1i}) + b_2(X_{2ij} - \overline{X}_{2i}) + ... + b_p(X_{pij} - \overline{X}_{pi})$$

وبأخذ المجموع وتربيع طرفي المعادلة نحصل على :

$$\sum_{i=1}^{2}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_i)^2 = \underline{b}'S\underline{b}$$
$$Q = \frac{(\underline{b}'\underline{\alpha})}{\underline{b}'S\underline{b}}$$
$$\frac{\partial Q}{\partial \underline{b}} = \frac{\underline{b}'S\underline{b} * 2(\underline{b}'\underline{\alpha})\alpha - (\underline{b}'\underline{\alpha})^2 * 2S\underline{b}'}{(\underline{b}'S\underline{b})^2} = 0$$
$$\underline{b}'S\underline{b} * \underline{b}'\underline{\alpha}\underline{\alpha} - (\underline{b}'\underline{\alpha})^2 * S\underline{b} = 0$$

وبقسمة الطرفين على $(\underline{b}'\underline{\alpha})^2$

$$\frac{\underline{b}'S\underline{b}}{\underline{b}'\underline{\alpha}}\underline{\alpha} - S\underline{b} = 0$$

أو إن :

$$\frac{\underline{b}'S\underline{b}}{\underline{b}'\underline{\alpha}}\underline{\alpha} = S\underline{b}$$

$$\underline{b}' = S^{-1}\frac{\underline{b}'S\underline{b}}{\underline{b}'\underline{\alpha}}\underline{\alpha}$$

ويمكن اعتبار المقدار $\frac{\underline{b}'S\underline{b}}{\underline{b}'\underline{\alpha}}$ **ثابتاً ومساوياً إلى الواحد فإن الحل سيكون كالآتي :**

$$\underline{b}' = S^{-1}\underline{\alpha}$$
$$= S^{-1}(\overline{X}_1 - \overline{X}_2)'$$

وهي توليفة خطية ناتجة عن تعظيم نسبة التباين بين المجاميع إلى داخل المجاميع .
في حالة وجود مجموعتين يكون لدينا دالة تمييز واحدة فقط ، وفي حالة وجود ثلاث مجاميع يكون لدينا دالتين تمييزيتين .

وبعد استخراج المعاملات $\underline{b}'$ ، تصنف المشاهدة إلى إحدى المجموعتين بالاعتماد على نقطة وسط المجموعتين (L) التي تجعل احتمال التصنيف الخاطئ أقل ما يمكن .
حيث إن :

$$L = \frac{\overline{Y}_1 + \overline{Y}_2}{2}$$

تصنف المشاهدة إلى المجموعة الأولى إذا كانت $\hat{Y} > L$.
تصنف المشاهدة إلى المجموعة الثانية إذا كانت $\hat{Y} < L$.
تصنف المشاهدة عشوائياً إلى المجموعة الأولى او الثانية إذا كانت $\hat{Y} = L$.
حيث إن :

$$\hat{Y} = (\overline{X}_1 - \overline{X}_2)' S^{-1} X = \hat{\alpha} X$$
$$Y_1 = (\overline{X}_1 - \overline{X}_2)' S^{-1} X_1$$
$$\overline{Y}_1 = \hat{\alpha}' \overline{X}_1$$
$$\overline{Y}_2 = \hat{\alpha}' \overline{X}_2$$

حيث إن : $\hat{\alpha}' = (\overline{X}_1 - \overline{X}_2) S^{-1}$

$\overline{Y}_2, \overline{Y}_1$ تمثل متوسطي المجموعتين الأولى والثانية .

ثانياً : الدالة المميزة الخطية في حالة أكثر من مجموعتين

The Linear Discriminant Function – More Than Two Groups

نفرض لدينا K من المجموعات ، وكل مجموعة لها n من المشاهدات ، وكل مشاهدة تتضمن P من المتغيرات .

n_i هو حجم العينة المسحوبة من المجموعة i .

$$n = \sum_{i=1}^{k} n_i$$

نفرض إن T تمثل مصفوفة التباين والتباين المشترك الكلي .

$$T = \sum_{i=1}^{K} \sum_{j=1}^{n} (X_{ij} - \overline{X})(X_{ij} - \overline{X})'$$

وإن W_i تمثل مصفوفة التباين والتباين المشترك للمجموعة i .

$$W_i = \sum_{i=1}^{n_i} (X_{ij} - \overline{X}_i)(X_{ij} - \overline{X}_i)'$$

وإن مصفوفة التباين والتباين المشترك داخل المجاميع تساوي W :

$$W = W_1 + W_2 + ... + W_K$$

وإن مصفوفة التباين والتباين المشرك بين المجاميع هي :

$$B = T - W$$

$$T = \begin{bmatrix} S_{11T} & S_{12T} & S_{1pT} \\ S_{21T} & S_{22T} & S_{2pT} \\ \\ S_{p1T} & S_{p2T} & S_{ppT} \end{bmatrix} \qquad W = \begin{bmatrix} S_{11W} & S_{12W} & S_{1pW} \\ S_{21W} & S_{22W} & S_{2pW} \\ \\ S_{p1W} & S_{p2W} & S_{ppW} \end{bmatrix}$$

$$B = \begin{bmatrix} S_{11B} & S_{12B} & S_{1pB} \\ S_{21B} & S_{22B} & S_{2pB} \\ \\ S_{p1B} & S_{p2B} & S_{ppB} \end{bmatrix}$$

هدفنا هو إيجاد مجموعة من التراكيب الخطية والتي هي :

$$Y = [Y_1, Y_2, ..., Y_r]$$

التي تعظم مقياس التمييز عن طريق تعظيم λ بالنسبة لكل b .

$$\lambda = \frac{Between \quad groups}{Within \quad groups}$$

$$\lambda = \frac{\underline{b}' B \underline{b}}{\underline{b}' W \underline{b}}$$

ولجعل λ أعظم مايمكن نأخذ المشتقة الجزئية بالنسبة لـ b .

$$\frac{\partial \lambda}{\partial \underline{b}} = \frac{\left[2(\underline{b}' W \underline{b} * B\underline{b} - \underline{b}' B \underline{b} * W \underline{b}) \right]}{(\underline{b}' W \underline{b})^2}$$

$$\frac{\partial \lambda}{\partial \underline{b}} = 0$$

$$(\underline{b}'W\underline{b})B\underline{b} - (\underline{b}'B\underline{b})W\underline{b} = 0$$

بقسمة الطرفين على $\underline{b}'W\underline{b}$ وبالتعويض عن λ بما يساويها نحصل على :

$$B\underline{b} - \lambda W\underline{b} = 0$$
$$(B - \lambda W)\underline{b} = 0$$
$$(W^{-1}B - \lambda I)\underline{b} = 0$$

نجد قيم λ ، أكبر قيمة إلى λ هي أكبر جذر مميز لمصفوفة $W^{-1}B$ والذي يقابل أكبر متجه مميز b_1 .

$$b_1 = (b_{11}, b_{12}, ..., b_{1p})$$

b_1 يمثل مقياس تمييز للدالة الأولى Y_1 والتي تساوي :

$$Y_1 = b_{11}X_1 + b_{22}X_2 + ... + b_{1p}X_p$$

وثاني أكبر جذر مميز لمصفوفة $W^{-1}B$ هو λ_2 والذي يقابل ثاني أكبر متجه مميز b_2 الذي يمثل مقياس التمييز للدالة الثانية والتي تساوي :

$$Y_2 = b_{21}X_1 + b_{22}X_2 + ... + b_{2p}X_p$$

من الضروري أن تكون Y_1 غير مرتبطة مع Y_2 .

Y_3 تمتلك ثالث أكبر متجه مميز .

$$Y_3 = b_{31}X_1 + b_{32}X_2 + ... + B_{3P}X_P$$

Y_3 غير مرتبطة مع Y_1 و Y_2 .

وهكذا نستمر إلى Y_r والتي تكون غير مرتبطة مع $Y_1, Y_2, ..., Y_{r-1}$ ويطلق على الدوال $(Y_1, Y_2, ..., Y_r)$ الدوال الخطية المميزة والتي يمكن التعبير عنها بشكل مصفوفة .

$$\underline{Y} = \underline{b}\underline{x}$$

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \\ Y_r \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} & b_{1P} \\ b_{21} & b_{22} & b_{2P} \\ \\ b_{r1} & b_{r2} & b_{rP} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \\ X_P \end{bmatrix}$$

يحدد r بعدد الدوال المميزة بالإعتماد على رتبة المصفوفة المركبة $W^{-1}B$.

إن رتبة المصفوفة W_{PXP} يساوي P وإن رتبة W^{-1} = رتبة W ، وإن رتبة مصفوفة B يكون أصغر $(P, K-1)$ وعادة يكون $(K-1)$ أصغر من (P) ، وبهذا تكون رتبة المصفوفة $W^{-1}B$:

$$rank(W^{-1}B) = \min(K-1, P)$$

أي يكون عدد الدوال المميزة لـ K من المجموعات و P من المتغيرات هو :

No.of discriminant function =min(P,K-1)

وعملية التصنيف تكون عن طريق تعويض قيم المتغيرات الخاصة بأي مشاهدة يراد تصنيفها في جميع الدوال المميزة ، وتصنف المشاهدة إلى الدالة المقابلة لأكبر مقدار .

تستخدم في التحليل المميز الاختبارات التالية :

1-اختبار معنوية الدالة المميزة الخطية .

2-اختبار تساوي مصفوفات التباين والتباين المشترك .

اختبار معنوية الدالة المميزة الخطية :

عندما يراد التمييز بين مجموعتين ، فإنه يمكننا أن نختبر الفرضية التي تنص على تساوي متوسطات المجموعات .

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

وعندما يراد التمييز بين أكثر من مجموعتين فإن الفرضية تكون .

$$H_0 : \mu_1 = \mu_2 = ... = \mu_K$$

H_1: على الأقل اثنان منهما غير متساويان

احصاءة الاختبار المستخدمة في حالة التمييز بين مجموعتين هي **Hotelling T^2** وصيغته الرياضية بالشكل التالي :

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} * D^2$$

حيث أن :

$$D^2 = (\overline{X}_1 - \overline{X}_2)' S^{-1} (\overline{X}_1 - \overline{X}_2)$$

وباستخدام اختبار F والتي تكون صيغته بالشكل التالي :

$$F = \frac{n_1 + n_2 - p - 1}{(n_1 + n_2 - 2)p} * T^2$$

بدرجات حرية $(p, n_1 + n_2 - p - 1)$

ترفض H_0 بمستوى معنوية F_α إذا كانت :

$$F_{Cal} \rangle F_{\alpha(p, n_1 + n_2 - p - 1)}$$

ونقبل H_1 وهذا يدل على أن متوسطات المجموعتين غير متساوية ، أي يوجد فرق معنوي بين المجموعتين ، ومعناه أن الدالة المميزة الخطية قابلة للتمييز بدرجة عالية .

وفي حالة التمييز بين أكثر من مجموعتين تكون احصاءات الاختبار كما يلي :

1- مقياس ولكس Wilks-criteria

يحسب هذا المقياس وفق الصيغة التالية :

$$\Lambda = \frac{|W|}{|T|}$$

حيث تمثل كل من :

T : مصفوفة التباين والتباين المشترك الكلي للمجموعات

W : مصفوفة التباين والتباين المشترك داخل المجموعات

تتراوح قيمة (Λ) بين الصفر والواحد ، فإذا كانت قريبة أو مساوية للواحد فإن ذلك يشير إلى ان متوسطات المجموعات متساوية وبذلك فلا يوجد تميز بين المجموعات ، وهذا يعني ان دالة التمييز المحسوبة فاشلة . أما إذا كانت قيمتها قريبة من الصفر فإن ذلك يدل على قوة التمييز .

2-مقياس χ^2

يعد هذا المقياس أكثر دقة من مقياس (Λ) وتكون صيغته الرياضية كالاتي :

$$\chi^2 = -N Log(\Lambda)$$

ويكون توزيعه مقارباً إلى χ^2 بدرجة حرية $P(K-1)$ ، وقد طورت صيغته من قبل **Barttlete** إلى الشكل التالي :

$$\chi^2 = -[N-1-1/2(P+K)]Log(\Lambda)$$

بدرجة حرية $P(K-1)$

3-مقياس F

وبالنظر لصعوبة الحصول على القيمة الجدولية لعدم توفر الجداول الخاصة بمعيار ولكس ، وضع معياراً بديلاً آخراً من قبل **Rao** ، وهذا المعيار هو مقياس F وصيغته الرياضية هي :

$$F = \frac{1-\Lambda^{1/5}}{\Lambda^{1/5}} * \frac{ms-2\lambda}{P(K-1)}$$

بدرجات حرية $df_1 = P(K-1)$

$$df_2 = ms-2\lambda$$

$$m = N - 1/2[P+K]$$

$$S = \left[\frac{P^2(1-K)^2-4}{(1-K)^2+P^2-5}\right]^{1/2}$$

$$\lambda = \frac{P(K-1)-2}{4}$$

اختبارتساوي مصفوفات التباين والتباين المشترك لجميع المجموعات :

يستخدم هذا الاختبار لمعرفة النوع الملائم من النماذج لتمثيل دالة التمييز بين المجموعات.

الفرضية تكون :

$$H_0 : \sum_1 = \sum_2 = ... = \sum_k$$

H_1 : على الأقل اثنان منهم غير متساويان

S_i هو تقدير غير متحيز لـ $\sum_i$

عندما تكون H_0 صحيحة فأن :

$$S = \frac{\sum_{i=1}^{K} n_i S_i}{\sum_{i=1}^{K} n_i}$$

S هو تقدير لمصفوفة التباين والتباين المشترك

احصاءة الاختبار تكون :

$$M = Ln|S|\sum_{i=1}^{K} n_i - \sum_{i=1}^{K} n_i Ln|S_i|$$

وقد أثبت **Box** أنه إذا ضرب M في ثابت C^{-1} والذي يساوي :

$$C^{-1} = 1 - \frac{2P^2 + 3(P-1)}{6(P+1)(K-1)}\left[\sum_{i=1}^{K}\frac{1}{n_i} - \frac{1}{\sum n_i}\right]$$

وبذلك نتوصل الى مقياس يتوزع χ^2 بدرجة حرية $\frac{1}{2}(K-1)(P-1)$ عندما تكون n_i كبيرة . لذا فأن :

$$Box's\ \ M = MC^{-1} \sim \chi^2_{\frac{1}{2}(K-1)(P-1)}$$

احتمال خطأ التصنيف :

هناك نوعان من احتمال خطأ التصنيف هما :

1-احتمال خطأ التصنيف P_{12} وهو احتمال تصنيف المفردة الى المجموعة الثانية وهي أصلاً تعود الى المجموعة الأولى .

$$P_{12} = \phi(-\frac{\delta}{2})$$

حيث أن ϕ تمثل دالة التوزيع الطبيعي القياسي

2-احتمال خطأ التصنيف P_{21} وهو احتمال تصنيف المفردة الى المجموعة الأولى وهي أصلاً تعود إلى المجموعة الثانية .

$$P_{21} = \phi(-\frac{\delta}{2})$$

δ^2 يمثل مربع مسافة وكذلك D^2 يمثل مربع مسافة مقياس مهالونوبيس

إذن :

$$\delta^2 = D^2 = (\overline{X}_1 - \overline{X}_2)' S^{-1} (\overline{X}_1 - \overline{X}_2)$$

وبذلك سوف يكون تقدير احتمال التصنيف :

$$P_{21} = P_{12} = \phi(-\frac{D}{2})$$

حيث أن D هو جذر مقياس مهالونوبيس .

الجانب التطبيقي :

جمعت البيانات من خلال عينة أخذت من طبلات المرضى الراقدين في مستشفى ابن الهيثم التعليمي لأمراض العيون وكذلك من المرضى المراجعين لهذا المستشفى لغرض العلاج .

عينة البحث هي عينة عشوائية بسيطة تتضمن نوعين من المتغيرات الثنائية والمتصلة لثلاث مجموعات مختلفة من امراض العيون هي (الساد ، الزرقاء ، انفصال الشبكية) والتي تمثل المتغير المعتمد ، حيث سحبت من ثلاثة مجتمعات ، كل مجموعة سحبت من المجتمع الذي تنتمي اليه .

تتكون عينة البحث موضوع الدراسة من 300 مريض ، تم دراسة المتغيرات التالية لكل مجموعة .

1-الجنس 2-العمر 3-ضغط العين اليمنى 4-ضغط العين اليسرى

5-السكر 6-فقر الدم 7-اليوريا

كان توزيع مفردات العينة حسب المجموعات ونسبة كل مجموعة على النحو التالي :

جدول (1) يبين توزيع مفردات العينة حسب المجموعات

المجموعات	أسماء المجموعات	عدد الحالات	الاحتمال
1	ساد	150	0.500
2	زرقاء	62	0.207
3	انفصال الشبكية	88	0.293
		300	1.000

المتغيرات التي أستخدمت في الدراسة :

أعتمدنا في تكوين دالة التمييز الخطية على عدة متغيرات جمعت عن كل مفردة (مريض) من مفردات العينة ، وحيث أن جزء من المتغيرات هي متغيرات متقطعة والجزء الآخر هي متغيرات متصلة .

أ-المتغير المعتمد والذي يمثل نوع المرض

(ساد يمثل 1) ، (زرقاء يمثل 2) ، (انفصال الشبكية 3)

ب–المتغيرات المستقلة فتمثلت بما يلي :

1–الجنس (Sex) (X_1)

ذكر (0) أنثى (1)

2– العمر (Age) (X_2)

3– ضغط العين (I.O.P) ضغط العين اليمنى (X_3) ضغط العين اليسرى (X_4)

4– السكر (Sugar) (X_5)

5– فقر الدم (P.C.V) (X_6)

6– اليوريا (Urea) (X_7)

نحتاج إلى احتساب المتوسطات الحسابية للمتغيرات الداخلة في المجموعات وهي كما في الجدول التالي :

جدول (2)

يبين المتوسطات الحسابية للمتغيرات الداخلة في عينة البحث للمجموعات الثلاثة

متوسطات المتغيرات / المجموعات	$\overline{X}_1$	$\overline{X}_2$	$\overline{X}_3$	$\overline{X}_4$	$\overline{X}_5$	X_6	$\overline{X}_7$
ساد	0.4933	62.9667	14.7267	14.5533	6.5407	0.4289	5.2927
زرقاء	0.5000	55.5484	19.0806	18.1129	6.1903	0.4279	5.5194
انفصال الشبكية	0.4205	44.9318	13.3636	13.3068	6.2534	0.4203	5.1102
المتوسط العام	0.4733	56.1433	15.2267	14.9233	6.3840	0.4262	5.2860

أ–اختبار معنوية الدالة المميزة الخطية

أولاً: عندما يراد التمييز بين ثلاث مجموعات فإنه يمكننا أن نختبر الفرضية التي تنص على تساوي متوسطات المجموعات .

$$H_0 : \mu_1 = \mu_2 = \mu_3$$
$$H_1 : \mu_1 \neq \mu_2 \neq \mu_3$$

كانت نتائج المقاييس المستخدمة في التحليل الإحصائي كما يلي :

Test of Function(s)	Wilks' Lambda	Chi-square	d.f	Sig.
1 through 2	0.638	131.953	14	.000
2	0.864	43.079	6	.000

d.f=P(K-1)
d.f=7(3-1)=14

حيث أن P : عدد المتغيرات = 7 ، K : عدد المجموعات = 3

$$\chi^2 = -\left[300 - 1 - \frac{1}{2}(7+3)\right] Log_e(0.638)$$

=57.33

حيث أن القيمة الجدولية لـــ (χ^2) تحت مستوى معنوية (0.0001) ودرجة حرية (14) هي :

$$\chi^2_{(14)} = 38.11$$

القيمة المحسوبة أكبر من القيمة الجدولية ، نرفض فرضية العدم ونقبل الفرضية البديلة، وهذا يدل على وجود فرق معنوي أي هناك تميز بين المجموعات .

أما بالنسبة لإحصاءة الاختبار مقياس F فكانت النتيجة كالآتي :

$$m = 300 - \frac{1}{2}(7+3)$$

$$m = 250$$

$$S = \left[\frac{(7)^2(1-3)^2 - 4}{(1-3)^2 + (7)^2 - 5}\right]^{1/2}$$

$$S = 2$$

$$\lambda = \frac{7(3-1) - 2}{4}$$

$$\lambda = 3$$

علماً بأن : $\Lambda = 0.638$

$$\therefore \quad F = \frac{1-(0.638)^{1/5}}{(0.638)^{1/5}} * \frac{(250)(2)-2(3)}{7(3-1)}$$

$$F = 3.320$$

بدرجات حرية $df_1 = 7(3-1)$

$$= 14$$

$$df_2 = (250)(2)-2(3)$$

$$= 494$$

ملاحظة : بما أنه لاتوجد في الجداول الاحصائية درجة حرية (14) لذلك أخذنا القيمة الجدولية عند درجة حرية (15) .

$F_{(0.05,15,494)} = 1.67$

لقد ظهر أن القيمة المحسوبة أكبر من القيمة الجدولية ، وهذا يدل على أن الدالة المميزة الخطية لها القابلية على التمييز اعتماداً على قيم المتغيرات ($\underline{X}$) .

ثانياً : نختبر الفرضية التالية

$$H_0 : \sum_1 = \sum_2 = \sum_3$$
$$H_1 : \sum_1 \neq \sum_2 \neq \sum_3$$

وباستخدام احصاءة الاختبار M واستخراج قيمة الثابت C^{-1} والتي تساوي :

$$C^{-1} = 1 - \frac{2(7)^2 + 3(7-1)}{6(7+1)(3-1)}\left[\frac{1}{150} + \frac{1}{62} + \frac{1}{88} - \frac{1}{300}\right]$$

$C^{-1} = 0.9628$

$M = 5.97$

$MC^{-1} = 5.75$

$MC^{-1} \sim \chi^2_{(\alpha,df)}$

$Box'M = MC^{-1} = 5.75$

حسبت القيمة أعلاه عند مستوى معنوية (0.000)

القيمة الجدولية لـ χ^2 بمستوى معنوية (0.001) ودرجة حرية $\frac{1}{2}(P-1)(K-1)$

$$df = \frac{(7-1)(3-1)}{2} = 6$$

$$\chi^2_{(6)} = 24.10$$

فإذا كانت القيمة الجدولية أكبر من القيمة المحسوبة فهذا يشير إلى قبول فرضية العدم ورفض الفرضية البديلة ، وهذا معناه تساوي التباينات فبذلك يكون النموذج الملائم لتمثيل دالة التمييز هو النموذج الخطي .

اختبار معنوية المتغيرات في الدالة المميزة

يتم اختبار معنوية جميع المتغيرات وذلك لمعرفة أهمية كل متغير في الدالة ، ومدى تأثيره في تحليل النتائج باستخدام تحليل التباين الأحادي ، وكانت النتائج بالشكل التالي :

جدول (3) يبين اختبار F لكل متغير في الدالة المميزة الخطية
Tests of Equality of Group Means

Variables	F	df_1	df_2	Sig.
Sex	.355	2	297	.701
Age	41.385	2	297	.000
R.iop	18.438	2	297	.000
L.iop	12.399	2	297	.000
Sugar	.600	2	297	.549
Pcv	1.147	2	297	.319
Urea	2.364	2	297	.096

نلاحظ من خلال الجدول أن المتغير X_2 (العمر) يمتاز بمعنوية عالية ، وله تأثير كبير في عملية التصنيف ، يليه المتغير X_3 (ضغط العين اليمنى) ، والمتغير X_4 (ضغط العين اليسرى)

لهما تأثير معنوي ولكن بشكل أقل من المتغير الأول ، ثم المتغيرين X_7 (اليوريا) و X_6 (فقر الدم) فكان تأثيرهما ضعيفاً جداً لأن تأثيرهما قليل جداً في عملية التصنيف ، ثم المتغير X_5 (السكر) ، أما المتغير X_1 (الجنس) فليس له تأثير معنوي .

تحتسب الدالة المميزة للمجموعات باستخدام نموذج Fisher للدالة المميزة وفق الصيغة التالية :

$$Z = A + B_1X_1 + B_2X_2 + B_3X_3 + \ldots\ldots\ldots\ldots + B_PX_P$$

ويتم تقدير معلمات الدالة المميزة β وفق الصيغة التالية :

$$\beta = S^{-1}(\overline{X}_{Pi} - \overline{X}_{qi}) \qquad P = q$$

بعد تقدير معاملات الدالة المميزة كانت أقيامها كما في الجدول الآتي :

جدول (4) يبين معاملات الدوال المميزة الخطية

Variables	Function (1)	Function(2)
Sex (X_1)	0.184	0.043
Age (X_2)	0.896	-0.463
R.iop (X_3)	0.245	0.619
L.iop (X_4)	0.368	0.452
Sugar (X_5)	-0.123	-0.113
Pcv (X_6)	0.324	0.017
Urea (X_7)	0.055	0.274

بما أن قيمة معيار (Wilks' Lambda) للدالة الأولى هي (0.638) وللدالة الثانية (0.864) وهذا يدل على القوة التمييزية للدالة الأولى . لذلك سوف نعتمد على الدالة الأولى في تفسير أهمية المتغيرات الداخلة في الدالة وتأثير هذه المتغيرات في تحليل النتائج .

نلاحظ أن المتغير X_2 (العمر) ذو تأثير عال قياساً بالمتغيرات الأخرى ، يليه المتغيران X_4 (ضغط العين اليسرى) ، و X_5 (فقر الدم) ، ثم المتغير X_3 (ضغط العين اليمنى) ، ثم المتغير X_1 (الجنس) ، ثم المتغير X_5 (السكر) ، وأخيراً المتغير X_7 (اليوريا) .

حساب احتمال التصنيف الصحيح :

أما بخصوص النسب المئوية للتصنيف الصحيح فقد جاءت بالشكل الآتي :

جدول (5) يبين نتائج التصنيف التنبؤي

Group		Predicted Group Membership			Total
		1	2	3	
Original Count	1	123	13	14	150
	2	26	29	7	62
	3	29	7	52	88
%	1	82.0	8.7	9.3	100.0
	2	41.9	46.8	11.3	100.0
	3	33.0	8.0	59.1	100.0

نسبة التصنيف الصحيح في دالة التمييز الخطية كانت (68 %) ، حيث أن (204) مفردة من أصل (300) مفردة صنفت بشكل صحيح الى المجاميع التي تنتمي اليها ، بينما أن (96) مفردة صنفت بشكل خاطئ ، وسوء التصنيف هذا حدث بسبب التشابه الكبير بين المتغيرات الموجودة في الدالة .

الإستنتاجات :

1- لقد أظهرت النتائج التي توصلنا إليها بأن هناك تأثيراً معنوياً عالياً قياساً بالمتغيرات الأخرى ، للمتغير (X_2) الذي يمثل العمر ، في تمييز المرض من حيث كونه ساد أو زرقاء أو انفصال شبكية .هذا المتغير يؤثر بشكل كبير جداً بالنسبة لمرضى الساد والزرقاء ، أما انفصال الشبكية قد يحدث في جميع الأعمار . الساد والزرقاء ممكن أن يحدثا في الأعمار الصغيرة جداً (منذ الولادة ولحد 16 سنة) ، قد يحدث الساد في أي عمر نتيجة شدة خارجية على العين ، فالساد والزرقاء الثانوي يحدث نتيجة استعمال الأدوية التي تحتوي على الكورتزونات أو نتيجة أمراض تصيب العين ، والساد يحدث للكبار في العمر فوق سن الخمسين نتيجة تغيرات فسلجية في عدسة العين ، أما الزرقاء يصيب الكبار في عمر أكثر من 40 سنة .

2- كذلك أظهرت نتائج التحليل إن المتغيرين (X_3) الذي يمثل ضغط العين اليمنى و (X_4) الذي يمثل ضغط العين اليسرى لهما تأثيراً معنوياً ولكن بشكل أقل من تأثير متغير العمر .

3- ظهر من نتيجة التحليل أن المتغير (X_1) والذي يمثل الجنس ، ليس له تأثير معنوي من حيث الأمراض التي تصيب العيون ، أي ليس هناك فرق بين الذكور والأناث في معدل الإصابة .

4- بالنسبة للمتغيرين (X_5) السكر و (X_7) اليوريا ظهرا في التحليل الاحصائي عكس ما موجود فعلاً في الحياة ، حيث بعد الاستفسار عن نتيجة التحليل من الطبيب الإخصائي ثبت مايلي : إن المتغيرين السكر واليوريا لهما تأثير كبير ومعنوي ، لأن الإصابة بالإلتهابات وأمراض العيون تكون أعلى عند المصابين بمرض السكر ، لكن طبلات المرضى تحتوي آخر معلومات تدون عن حالة المريض قبل إجراء العملية الجراحية بحيث يتم تخفيض نسبة السكر المرتفع الى المستوى الطبيعي للمريض الذي يتعرض الى تخدير عام لغرض إجراء العملية ، والتئام الجروح يحتاج الى نسبة سكر طبيعية في الجسم ، وكذلك بالنسبة لليوريا لأن المريض الذي هو بحاجة الى تخدير عام يحتاج الى نسبة يوريا طبيعية في الجسم ، لأن تخليص الجسم من أدوية التخدير العام يعتمد على عمل الكلى .

5-المتغير (X_6) الذي يمثل نسبة الهيموغلوبين في الدم ، ليس له تأثير معنوي في حالات التصنيف المدروسة.

التوصيات :

1-ضرورة تسجيل المعلومات والبيانات الكاملة لكل مريض في الكارت الخاص به (طبلة المريض) طبقاً لنوع المرض ، مع ضرورة الإحتفاظ بنتائج التحليلات السابقة للإستفادة منها في مقارنة حالة المريض الصحية وكذلك للإستفادة منها في إجراء الدراسات . حيث لاتتوفر مثل هذه المعلومات لدى قسم الإحصاء في المستشفى .

2-تجنب تعاطي الحبوب والأدوية بالنسبة للأم الحامل في الأسابيع الأولى من الحمل .

3-بالنسبة لمرضى السكر يجب المحافظة على بقاء نسبة السكر في الدم طبيعية قدر الإمكان .

4-على كل فرد تعدى عمره الأربعين سنة أن يفحص عينيه سنوياً .

5-إجراء العملية الجراحية في وقت مبكر .

6-عدم التزاوج بين الأقارب حتى لا تتزايد احتمالات إصابة الأطفال بالمرض ، لأن الأسر التي لها تاريخ وراثي بالمرض ، يمكن توارث هذا المرض فإذا كانت الأم مصابة ، فإن احتمال إصابة مولودها به ترتفع ست أو سبع مرات عن الأشخاص الذين ليس لديهم تاريخ وراثي للمرض .

7-التشخيص والعلاج المبكر للمرض هما العاملان الرئيسان للوقاية من الإصابة بالعمى .

8-الإهتمام بمراجعة استشاري العيون وإجراء الفحص الدوري للعينين كل ثلاثة أشهر أو كل ستة أشهر .

9-تجنب التعرض للأشعة تحت الحمراء كما هو الحال عند عمال أفران الزجاج ، وكذلك تجنب التعرض للإشعاعات المضرة بالعين كأشعة جهاز لحيم المعادن ، وضرورة ارتداء الأقنعة الواقية لحماية العينين .

10-الإهتمام بنظافة العينين واستعمال القطرات والمراهم العلاجية التي يصفها الطبيب بصورة صحيحة ، والمحافظة على العينين من أي شدة خارجية أو أجسام غريبة تصيبها .

المصادر :

1-القصاب ، موفق محمد ، *"استخدام الاسلوب التمييزي في تصنيف الحوامل طبقاً لدرجات الخطورة"* ، مجلة تنمية الرافدين ، عدد 63 ، مجلد 23 ، سنة 2001 .

2-حمة أمين ، چرو علي ، *"استخدام التحليل المميز في دراسة العوامل المؤثرة في أمراض الجهاز العصبي لدى الأطفال في ظل الحصار"* ، رسالة ماجستير ، قسم الاحصاء ، كلية الادارة والاقتصاد ، الجامعة المستنصرية ، 1999 .

3-داود ، عربية عبد الرحمن ، *"استخدام الدالة المميزة لبيان تأثير الحمية في مرضى السمنة"* ، مجلة تنمية الرافدين ، عدد 61 ، مجلد 22 ، سنة 2000 .

4-محمد ، سميرة محمد صالح ، *"استخدام التحليل المميز (التصنيفي) لتحديد أهم العوامل المؤثرة في تسرب ورسوب الطلبة في جميع المراحل الدراسية"* ، رسالة ماجستير ، قسم الاحصاء ، كلية الادارة والاقتصاد ، جامعة السليمانية ، 2003 .

5-Anderson, T.W., *"An Introduction to Multivariate Statistical Analysis"* , 2nd Edition , John Wiley and Sons , New York , USA , 1984 .

.....

.....