بناء الدالة التمييزية لبيانات القياسات المكررة بافتراض بنية التباين المشترك المتماثلة الكروية (CS) وتوظيفها في المجال الصحي

أ.د. ظافر حسين رشيد /كلية الادارة والاقتصاد/ جامعة بغداد
الباحث/ مؤمن عباس موسى/كلية الادارة والاقتصاد/ جامعة بغداد

المستخلص

يعرف التحليل التمييزي بأنهُ تَقْنِيَة تستعمل لتمييز أو لتصنيف فرد ما إلى مجموعة من بين عدد من المجموعات استنادا إلى تركيبة خطية لمجموعة من المتغيرات ذات العلاقة تعرف بالدالة التمييزية. في هذا البحث أستعمل التحليل التمييزي لتحليل بيانات تصاميم القياسات المكررة، حيث تركزت أهميته في مشكلة التمييز والتصنيف ثنائي المجموعة بافتراض تركيبة التباين المشترك المتماثلة الكروية (Compound Symmetric, CS) وبوجود حالة التوزيع الطبيعي للمجموعات لبيانات القياسات المكررة المستقاة طوليا عبر الزمن على متغير الاستجابة نَفسُهُ ولكل وحدة تجريبية.

أهمية البحث هذا تمثلت بإيجاد أفضل أنموذج لتصنيف مجموعة من مرضى داء السكري وذلك لغرض دراسة تأثيرات كل من عدد الارتباطات، و عدد التباينات، وعدد القياسات المكررة على أداء قواعد التصنيف لهذا النوع من البيانات، حيث تم اعتماد القياسات الشهرية لنسبة بروتين الهيموغلوبين الغليكوزيلاتي (HbA1c) في الدم والمأخوذة في ثلاث مراحل، المرحلة الأولى كانت في بداية التجربة، المرحلة الثانية كانت بعد ثلاث أشهر، أما المرحلة الثالثة فكانت بعد ست أشهر ولمجموعتين من المرضى تضمنت المجموعة الأولى بعدد (38) مريض يعاني من داء السكري من النمط الأول (I)، في حين كانت المجموعة الثانية تمثل منهم (33) مريض يعاني من داء السكري من النمط الثاني (II).

ومن خلال ما تم عرضه في البحث هذا والتي تمت باستعمال نتائج البيانات الحقيقية تم التوصل إلى أن انه كلما زادت عدد المعلمات المطلوب تقديرها لبناء الأنموذج التمييزي فأن نسبة التصنيف الخطأ الظاهرة الكلية (APER) تبدأ بالزيادة وهذا ما يقلل من كفاءة قواعد التصنيف لهذا النوع من البيانات، واعتماداً على ما توصل إليه البحث فأن ما يوصى به وعند التركيز على أقل عدد من المعلمات لبناء قاعدة التصنيف فانه من الممكن لإتباع أسلوب التحليل التمييزي الخطي، وأسلوب التحليل التمييزي ألتربيعي والمتمثل بتساوي معلمة التباين واختلاف معلمة الارتباط بوجود تركيبة التباين المشترك المتماثلة الكروية (CS) لتصنيف المرضى.

المصطلحات الرئيسية للبحث/التحليل التمييزي- بيانات القياسات المكرره – معامل الارتباط- مركبة التباين – تركيبة التباين المشترك المتماثله الكروية (cs)

مجلة العلوم الاقتصادية والإدارية
المجلد ٢١ العدد ٨٥
الصفحات ٣٣٦-٣٥٩

*البحث مستل من رسالة ماجستير

1. المقدمة وهدف البحث (Introduction and Objective)

يستفاد عادة من التحليل التمييزي في تصنيف مفردة واحدة أو أكثر إلى إحدى المجتمعات بالاعتماد على متغيرات ذات صفات تمييزية وهذا ما يعرف بالتحليل التمييزي التنبؤي والذي يركز بدوره على وضع قواعد تصنيف كفوءة، ويمكن الاستفادة منها أيضا لوصف الأهمية النسبية للمتغيرات للتمييز بين المجموعات وهذا ما يعرف بالتحليل التمييزي الوصفي والذي يركز على تحديد الأهمية النسبية لعدد من المتغيرات في قدرتها على التمييز بين المجموعات.

وفي الأعوام القليلة الماضية كان هناك اهتمام متزايد لاستعمال التحليل التمييزي لتحليل بيانات تصاميم القياسات المكررة والتي تنشأ عندما يتم جمع القياسات عند مناسبتين أو أكثر ولنفس الوحدة التجريبية أما لمتغير وحيد (بيانات قياسات مكررة أحادية المتغير)، أو عدة متغيرات (بيانات قياسات مكررة متعددة المتغيرات).

وبما أن أسلوب التحليل التمييزي الخطي التقليدي لديه بعض الافتراضات الخاصة والتي يجب توفرها عند استعمالها لتحليل هذا النوع من البيانات ومنها أن تكون عدد المشاهدات للوحدة التجريبية كاملة، صعوبة تضمين تأثيرات المتغيرات المشاركة، وكذلك فأنها لا يمكن تطبيقها عندما تكون عدد الوحدات التجريبية الكلي أقل من عدد مشاهدات القياسات المكررة، وفضلا عن ذلك فأنها تعتمد على تقدير عدد كبير من المعلمات المطلوب تقديرها لبناء الأنموذج التمييزي والتي يزداد عددها بزيادة عدد القياسات المكررة.

ألا انه غالبا ما يكون هناك فجوة بين الأسس النظرية لهذه الأساليب وبين توظيفها في التطبيقات العملية فنتائج هذه الأساليب تتأثر بشدة بوجود عدد كبير من المعلمات غير المعلومة للأنموذج المقدر والذي سوف يعكس خصائص من شأنها أن تقلل من دقة قواعد التصنيف لهذا النوع من البيانات.

ومن ثم ومع تطور أساليب التحليل التمييزي للقياسات المكررة والتي تزودنا بطرائق حديثة للتعامل مع هذا النوع من البيانات حيث تعمل وبوجود تراكيب مختلفة للتباين للمشترك (Covariance Structures) على تقليص عدد المعلمات المطلوبة لبناء الأنموذج التمييزي والتصنيفي والتي من شأنها أن تزيد من دقة قواعد التصنيف لبيانات القياسات المكررة.

أن فكرة البحث هذا والمتضمنة تشخيص مرضى داء السكري وباستعمال طرائق تمييز للقياسات المكررة تعد مهمة لكونها وسيلة مساعدة تشخيصية للطبيب المختص، فضلا عن كون مرض داء السكري من الإمراض الخطيرة والذي يؤدي إلى مضاعفات خطيرة وحتى إلى الوفاة المبكرة مع تطور المرض إلى الحالات المتقدمة منه، وأن لمرض داء السكري بأنواعه العديد من الأسباب ولعل من أهمها نقص هرمون الأنسولين أو عدم استجابة الجسم له والذي تفرزه غدة البنكرياس.

وأن التشخيص المبكر للمرض عامل مهم في التقليل من حدة المرض فضلا عن أنه يساعد في إنقاذ حياة المريض ويجنبه الحالات المتطورة للمرض، ولان التشخيص مسألة ذات جوانب معقدة لان عدد من الإمراض تشترك في عدة إعراض يصعب التمييز فيما بينها كالشك بين الإصابة بمرض داء السكري من النمط الأول I والمرض السكري من النمط الثاني II ، لذلك فأن الأساليب الإحصائية ومنها التحليل التمييزي تساعد الطبيب المختص في التشخيص المبكر لإمراض أنماط السكري.

حيث أُتسِم هدف البحث هذا ببناء أكثر من أنموذج تمييزي وتصنيفي بوجود تركيبة التباين المشترك المتماثلة الكروية (Compound Symmetric, CS) لما لها من دور فعال في الحد من زيادة عدد المعلمات المطلوب تقديرها لبناء الأنموذج التصنيفي، والمقارنة بينها عبر مجموعة من شروط التجانس وعدم التجانس لمصفوفة التباين المشترك والوصول إلى أفضل قاعدة تصنيفية لتشخيص بعض أنماط داء السكري وبالاستناد إلى القياسات الشهرية لنسبة بروتين الهيموغلوبين الغليكوزيلاتي (HbA1c) في الدم والمأخوذة في ثلاث مراحل مختلفة ولمجموعتين من المرضى في بناء نموذج احتمالي للتمييز بين نوعيين من أمراض السكري.

ومن الجدير بالذكر فأن دوال التمييز للقياسات المكررة درست من قبل بعض الباحثين ففي عام (2005) قدم كل من (Roy & *Khattree*) عملا وصفيا لأسلوبين مختلفين بوجود تراكيب (Structures) مختلفة ولحالتين مختلفتين من شروط عدم التجانس لمصفوفة التباين والتباين المشترك، وذلك للتقليص من عدد المعلمات غير المعلومة والمطلوب تقديرها لبناء قاعدة التميز. وفي عام (2010) قام كل من (Pluta , Madry وأخرون) بمقارنة الطرائق التقليدية مع الطرائق الحديثة لأساليب التحليل التمييزي من خلال مجموعة بيانات حقيقية لتشخيص بعض أنواع النباتات، وتوصلوا هؤلاء إلى أن الطرائق الحديثة للتحليل التمييزي بوجود تراكيب التباين المشترك هي ذات كفاءة أعلى في التصنيف من الطرائق التقليدية من حيث نسبة التصنيف الخطأ الظاهرة الكلية [11],[8].

2. الجانب النظري

1-2 القياسات المكررة بوجود تركيبة التباين المشترك المتماثلة الكروية [2],[3],[4],[7],[8],[9]، (Repeated Measures with Compound Symmetric Covariance Structure) [10]

لتَفسِير مشكلة أسلوب التحليل التمييزي و التصنيفي لبيانات القياسات المكررة المستقاة طوليا عبر الزمن على متغير الاستجابة نفسه ولكل وحدة تجريبية وفي حالة التصنيف الثنائي المجموعة وبوجود حالة التوزيع الطبيعي للمجموعات. لنفترض أن y هو متجه عمودي أبعادهُ $(p \times 1)$ يمثل القياسات المكررة على الوحدة التجريبية ولمتغير الاستجابة نَفسُهُ المأخوذة عبر p من النقاط الزمنية. وبافتراض أن y_{jit} يمثل القياس على الفرد ith في ألمجموعه jth عند النقطة الزمنية tth، حيث أن $(t = 1, 2, 3, \dots, p)$، $(j = 1, 2)$، و $(i = 1, 2, 3, \dots, n_j)$.

ومن ثم فأن $y_{ji} = (y_{ji1}, y_{ji2}, \dots, y_{jip})'$ هو متجه أبعادهُ $(p \times 1)$ للقياسات المكررة المشاهدة للفرد ith المشارك في الدراسة في المجموعة jth. حيث أن لكلتا المجموعتين j ، فأن y_{ji} يتبع التوزيع الطبيعي مع متوسط مساوي إلى μ_j ومصفوفة التباين والتباين المشترك أبعادها $(p \times p)$ المساوية إلى Ω_j أي أن:

$$y_{ji} \sim N_p(M_j, \Omega_j) \quad , \quad j = 1, 2$$

وبوجود تركيبة التباينات المتماثلة الكروية CS لمصفوفة التباين والتباين المشترك Ω_j ومن ثم فأن Ω_j تعطى بالصيغة الآتية :

$$\Omega_j = {\sigma_j}^2[(1 - \rho_j)I_p + \rho_j 1_p 1_p'] \quad \dots\dots(1)$$

حيث أن ρ هو مقدار الارتباط بين القياسات المكررة، σ^2 هو تباين القياسات المكررة ويفترض هنا أن يكون ثابت عبر مناسبات القياس، I_p تمثل مصفوفة الوحدة، وأن 1_p متجه عمودي ذو أبعاد $(p \times 1)$ جميع عناصره مساوية إلى الواحد الصحيح، و أن قيمة الارتباط ρ_j تتراوح ما بين $(0 < \rho < 1)$.

وأن قيمة محدد ومعكوس المصفوفة Ω_j يعطى بالصيغتين الآتيتين :

$$|\Omega_j| = ({\sigma_j}^2)^p[1 + (p - 1)\rho_j](1 - \rho_j)^{p-1} \quad \dots\dots(2)$$

$${\Omega_j}^{-1} = \frac{1}{{\sigma_j}^2(1 - \rho_j)}\left[I_p - \frac{\rho_j}{1 + (p - 1)\rho_j} 1_p 1_p'\right] \quad \dots\dots(3)$$

وأن المصفوفة Ω_j في ظل هذا الافتراض هي دالة من معلمتين غير معلومة، أحداهما التباين ${\sigma_j}^2$ للمشاهدات والأخرى هي الارتباط ρ_j بين أي زوج للقياسات المكررة من نفس الوحدة التجريبية.

2-2 دوال التمييز والتصنيف (Discrimination and Classification Functions)

يتناول البحث هذا بعض دوال التمييز والتصنيف الخطية و التربيعية بوجود تركيبة (CS)

للقياسات المكررة عبر مجموعة من شروط التجانس وعدم التجانس لمصفوفة التباين المشترك وكالاتي :

1-2-2 دالة التمييز والتصنيف الخطية للحالة الأولى

تتمثل الحالة الأولى بتجانس مصفوفة التباين المشترك (CS) مع تساوي كل من مركبتي التباين والارتباط لكلا المجموعتين أي [4], [7],[8],[9] :

$$\text{Case 1}: \Omega_1 = \Omega_2 \ (\sigma_1{}^2 = \sigma_2{}^2 = \sigma^2,\ \rho_1 = \rho_2 = \rho)$$

يعبر عن الدالة التمييزية (discriminant function) في هذه الحالة (CS) كتركيبة خطية من القياسات المكررة p والتي تعطى بالصيغة الآتية :

$$D = \hat{a}'y_{ji} \ , \quad \hat{a} = \hat{\Omega}^{-1}(\hat{\mu}_1 - \hat{\mu}_2) \qquad \dots\dots(4)$$

حيث أن $\hat{a}$ هو متجة تقديرات معاملات الدالة التمييزية (Discriminant function coefficients) وبوجود متجهات متوسط تركيبية منتظمة للمجموعات بدون تأثير عامل الوقت فأن :

$$\hat{\mu}_j = c_j 1_p \ , \ j = 1,2 \qquad \dots\dots(5)$$

حيث c_j هو متوسط مناسبات القياسات المكررة للمجموعة jth .

وأيضا في المعادلة (4)، فأن المصفوفة Ω هي مصفوفة التباين المشترك المتماثلة الكروية والمعطاة في المعادلة رقم (1). وطبقا للحالة هذه فأن قاعِدَةُ التمييز والتصنيف الخطية (linear classification rule) هي: أنهُ يتم تخَصّيَصَ المفردة ith مع المشاهدات $(y_{ji1}, y_{ji2}, \dots, y_{jip})$ للمجموعة الأولى إذا كان :

$$\lambda(y_i) = \sum_{t=1}^{p} y_{it} \geq \left(\frac{c_1 + c_2}{2}\right) p \qquad \dots\dots(6)$$

وعدا ذلك يتم تخَصّيَصَه إلى المجموعة الثانية. وواضح أن قاعِدَةُ التمييز والتصنيف للحالة هذه مستقلة عن Ω ، حيث أن y_{it} يمَثّلَ المشاهدة للفرد ith المُشَارِك أو الجديد في الدراسة عند القياس المكرر أو النقطة الزمنية t^{th} والمراد تمييزية أو تصنيفه على التوالي، و p يدل على عدد النقاط الزمنية.

إما مقدرات الإمكان الأعظم لكل من c_1 و c_2 يمكن الحصول عليها كما يأتي [4],[5],[6],[7],[8],[9].

بوجود تركيبة منتظمة لمتوسطات المجموعة، ولـيكن θ متجه من معلمات الأنموذج، حيث أن العناصر الأولى والثانية يدلان على المتوسط μ_1 و μ_2 والعنصران الأخيران يمثلان التباين σ^2 والارتباط ρ .

وأن y_{ij} ليكون متجه أبعادهُ $(1 \times p)$ من القياسات المكررة على الفرد ith المشارك في الدراسة حيث أن $(i = 1, 2, \dots, n_j;\ N = n_1 + n_2)$ في المجموعة jth وأن $Y_1 = (y_{11}, y_{12}, \dots, y_{1n_1})$ تمثل المشاهدات من المجتمع الأول $N_p(\mu_1, \Omega)$ و $Y_2 = (y_{21}, y_{22}, \dots, y_{2n_2})$ تدل على المشاهدات من المجتمع الثاني $N_p(\mu_2, \Omega)$ بالتالي، فأن لوغاريتم دالة الإمكان الأعظم المشتركة (log joint likelihood function) يعطى بالصيغة التالية :

$$\log L(\theta, \Omega ; Y_1, Y_2, \dots, Y_N) = -\frac{Np}{2}\log 2\pi - \frac{N}{2}\log|\Omega| - \frac{1}{2}\mathrm{tr}\left(\sum_{j=1}^{2}\sum_{i=1}^{n_j}(y_{ij} - \mu_j)'\Omega^{-1}(y_{ij} - \mu_j)\right) \quad \dots\dots(7)$$

بوجود تركيبة متوسط منتظمة لمتوسطات المجموعة فأن :

$$\mu_j = c_j 1_p \quad \dots(8)$$

وبتعويض (8) في (7) نجد أن :

$$\log L(\theta, \Omega ; Y_1, Y_2) = -\frac{Np}{2}\log 2\pi - \frac{N}{2}\log|\Omega| - \frac{1}{2}\mathrm{tr}\,(\Omega^{-1}W) - \frac{1}{2}\left(\sum_{j=1}^{2} n_j(\bar{y}_j - c_j 1_p)'\,\Omega^{-1}(\bar{y}_j - c_j 1_p)\right) \quad \dots(9)$$

وبتبسيط المقدار الأخير، ومن ثم يمكن إعادة كتابة المعادلة (9) بالشكل الأتي :

$$\log L(\theta, \Omega ; Y_1, Y_2) = -\frac{Np}{2}\log 2\pi - \frac{N}{2}\log|\Omega| - \frac{1}{2}\mathrm{tr}\,(\Omega^{-1}W) - \frac{1}{2}\left[n_1(\bar{y}_1 - c_1 1_p)'\Omega^{-1}(\bar{y}_1 - c_1 1_p) + n_2(\bar{y}_2 - c_2 1_p)'\Omega^{-1}(\bar{y}_2 - c_2 1_p)\right] \quad \dots(10)$$

حيث أن $\bar{y}_1$: متجه متوسط العينة الأولى للمجتمع الأول.

$\bar{y}_2$: متجه متوسط العينة الثانية للمجتمع الثاني.

أما W و W_j تعطى بالصيغ الآتية :

$$W = W_1 + W_2 \quad \dots\dots(11)$$

$$W_j = \sum_{i=1}^{n_j}(y_{ij} - \bar{y}_j)(y_{ij} - \bar{y}_j)', \quad j = 1, 2. \quad \dots\dots(12)$$

وأن :

$$\bar{y}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} y_{ji} \quad , \quad j = 1,2 \qquad \dots\dots (13)$$

وللحصول على مقدرات الإمكان الأعظم (MLEs) لكل من c_j نشتق المعادلة (10) ونضعها مساوية للصفر ثم بعد عملية التبسيط فأن كل من c_1 و c_2 تساوي :

$$\hat{c}_j = \frac{1_p{}' \Omega^{-1} \bar{y}_j}{1_p{}' \Omega^{-1} 1_p} \quad , \quad j = 1,2$$

أما مصفوفة التباين والتباين المشترك Ω^{-1} هي متماثلة كروية (CS)، ومن ثم فأن المعادلة المذكورة أنفاً تختزَل إلى :

$$\hat{c}_j = \frac{1_p{}' \bar{y}_j}{1_p{}' 1_p} = \frac{1_p{}' \bar{y}_j}{p} = \frac{1}{p}\sum_{t=1}^{p} \hat{\mu}_{jt} \ , \quad j = 1,2 \qquad \dots\dots (14)$$

ولإيجاد تقديرات الإمكان الأعظم لكل من معامل الارتباط ρ ومركبة التباين σ^2 للحالة الأولى والمطلوب تقديرها لغرض تقدير مصفوفة التباين المشترك (CS) لبناء الدالة التمييزية، نقوم بتعويض كل من محددة ومعكوس المصفوفة Ω_j والمعطاة في المعادلتين (2) و (3) على التوالي في معادلة رقم (9) ومن ثم تصبح بالصيغة الآتية :

$$\log L(\theta, \Omega ; Y_1, Y_2) = -\frac{Np}{2}\log 2\pi - \frac{N}{2}\log\sigma^2 + \frac{N}{2}\log(1+(p-1)\rho) + \frac{N(p-1)}{2}\log(1-\rho) - \frac{1}{2\sigma^2(1-\rho)}\operatorname{tr}(W) - \frac{\rho}{1+(p-1)\rho}\operatorname{tr}(JW), \quad \dots (15)$$

حيث أن W تم تعريفها في معادلة (11)، إما J فتمثل مصفوفة إبعادها $(p \times p)$ جميع عناصرها مساوية للواحد الصحيح، وباشتقاق المعادلة (14) المذكورة أنفاً مره واحده بالنسبة إلى σ^2 و ρ نجد المعادلتين (16) و (17) الآتيتين وبحلها نجد تقديرات الإمكان الأعظم لكل من σ^2 و ρ [7],[8],[9].

$$-Np\hat{\sigma}^2(1-\hat{\rho})(1+(p-1)\hat{\rho}) + (a_1+a_2)(1+(p-1)\hat{\rho}) - (b_1+b_2)\hat{\rho} = 0, \qquad \dots\dots (16)$$

و

$$-Np\hat{\rho}\hat{\sigma}^2(p-1)(1-\hat{\rho})(1+(p-1)\hat{\rho}) - (a_1+a_2)(1+(p-1)\hat{\rho})^2 + (b_1+b_2)(1+(p-1)\hat{\rho}^2) \qquad \dots\dots (17)$$

في المعادلتين المذكورتين أنفاً، القيم التقديرية لكل من a_1 ، b_1 ، a_2 و b_2 يتم الحصول عليها بتطبيق كل من المعادلات الآتية :

$$\hat{a}_1 = \operatorname{tr}(W_1), \quad \hat{b}_1 = \operatorname{tr}(J_p W_1), \quad \hat{a}_2 = \operatorname{tr}(W_2), \quad \hat{b}_2 = \operatorname{tr}(J_p W_2) \qquad \dots (18)$$

وأن القيمة التقديرية للمصفوفة W_j يتم إيجادها بتطبيق الصيغة (12) .

2-2-2 قاعدة التمييز والتصنيف التربيعية للحالة الثانية [4],[7],[8]

للحالة الثانية والتي تتمثل بعدم تجانس مصفوفتي التباين والتباين المشترك المتماثلة الكروية مع اختلاف كل من مركبة التباين ومعامل الارتباط لكلتا المجموعتين أي :

$$\text{Case 2}: \Omega_1 \neq \Omega_2 \ (\sigma_1{}^2 \neq \sigma_2{}^2 \ , \ \rho_1 \neq \rho_2)$$

فأن قاعِدَةُ التمييز والتصنيف التربيعية $(\text{Quadratic classification rule}, \text{QCR})$ هي : أنهُ يتم تخَصّيَصَ الفرد ith مع المشاهدات $(y_{ji1}, y_{ji2}, \dots, y_{jip})$ إلى ألمجموعه الأولى إذا كان :

$$\begin{aligned}
\lambda(y_i) = &-\frac{1}{2}\left(\frac{1}{\sigma_1{}^2(1-\rho_1)} - \frac{1}{\sigma_2{}^2(1-\rho_2)}\right)\sum_{t=1}^{p} y_{it}{}^2 \\
&+\frac{1}{2}\left(\frac{\rho_1}{\sigma_1{}^2(1-\rho_1)(1+(p-1)\rho_1)}\right. \\
&\left.-\frac{\rho_2}{\sigma_2{}^2(1-\rho_2)(1+(p-1)\rho_2)}\right)\left(\sum_{t=1}^{p} y_{it}\right)^2 \\
&+\left(\frac{c_1}{\sigma_1{}^2(1+(p-1)\rho_1)} - \frac{c_2}{\sigma_2{}^2(1+(p-1)\rho_2)}\right)\sum_{t=1}^{p} y_{it} \\
&\geq \frac{p}{2}\left(\frac{c_1{}^2}{\sigma_1{}^2(1+(p-1)\rho_1)} - \frac{c_2{}^2}{\sigma_2{}^2(1+(p-1)\rho_2)}\right) - \frac{p}{2}\ln\frac{\sigma_2{}^2}{\sigma_1{}^2} \\
&-\frac{p-1}{2}\ln\left(\frac{1-\rho_2}{1-\rho_1}\right) - \frac{1}{2}\ln\left(\frac{1+(p-1)\rho_2}{1+(p-1)\rho_1}\right), \qquad \dots (19)
\end{aligned}$$

عدا ذلك يتم تخَصّيَصَ الفرد ith مع الاستجابات $(y_{ji1}, y_{ji2}, \dots, y_{jip})$ إلى ألمجموعه الثانية. في قاعِدَةُ التصنيف المذكورة أنفاً فأن تقديرات الإمكان الأعظم للمتوسطات c_j معطاة في معادلة رقم (14). إما تقديرات الإمكان الأعظم لكل من ρ_1 و ρ_2 ، $\sigma_1{}^2$ و $\sigma_2{}^2$ وبوجود تركيبة (CS) للتباين المشترك فأن لوغاريتم دالة الإمكان الأعظم يعطى بالصيغة التالية :

$$\begin{aligned}
&\log L(c_1, c_2, \Omega_1, \Omega_2; Y_1, Y_2) \\
&= -\frac{Np}{2}\log 2\pi - \frac{n_1}{2}\log|\Omega_1| - \frac{n_2}{2}\log|\Omega_2| \\
&-\frac{1}{2}\text{tr}\Omega_1{}^{-1}\left[W_1 + n_1(\bar{y}_1 - c_1 1_p)(\bar{y}_1 - c_1 1_p)'\right] \\
&-\frac{1}{2}\text{tr}\Omega_2{}^{-1}\left[W_2 + n_2(\bar{y}_2 - c_2 1_p)(\bar{y}_2 - c_2 1_p)'\right] \qquad \dots (20)
\end{aligned}$$

نجد أن تقدير الإمكان الأعظم للمعلمة c_j مطابق للمعادلة (14). وبافتراض إن $\Lambda_j = \Omega_j^{-1}$ وباشتقاق المعادلة رقم (20) بالنسبة إلى Λ_j ولكل من $j = 1, 2$ فأن :

$$\frac{\partial \log L(c_1, c_2, \Omega_1, \Omega_2; Y_1, Y_2)}{\partial \Lambda_j} = \frac{n_j}{2}\Lambda_j^{-1} - \frac{1}{2}\Lambda_j \quad j = 1, 2. \ldots\ldots (21)$$

حيث أن :

$$\Lambda_j = W_j + n_j(\bar{y}_j - c_j 1_p)(\bar{y}_j - c_j 1_p)' \quad j = 1, 2.$$

وبمساواة المشتقة أعلاه للصفر وبعد عملية التبسيط ولكل $j = 1$ و $j = 2$ نجد منظومة المعادلات أدناه [8],[7]:

$$-n_1 p(1 - \rho_1)\{1 + (p-1)\rho_1\}\sigma_1^2 + a_1\{1 + (p-1)\rho_1\} - b_1\rho_1 = 0, \ldots (22)$$
$$-n_2 p(1 - \rho_2)\{1 + (p-1)\rho_2\}\sigma_2^2 + a_2\{1 + (p-1)\rho_2\} - b_2\rho_2 = 0, \ldots (23)$$
$$n_1 p\sigma_1^2\rho_1(p-1)(1 - \rho_1)\{1 + (p-1)\rho_1\} - a_1\{1 + (p-1)\rho_1\}^2 + b_1\{1 + (p-1)\rho_1^2\} = 0, \quad \ldots (24)$$

و

$$n_2 p\sigma_2^2\rho_2(p-1)(1 - \rho_2)\{1 + (p-1)\rho_2\} - a_2\{1 + (p-1)\rho_2\}^2 + b_2\{1 + (p-1)\rho_2^2\} = 0, \quad \ldots (25)$$

كل من a_1، b_1، a_2 و b_2 يمكن إيجادها بتطبيق الصيغة :

$$\hat{a}_1 = \mathrm{tr}(W_{11}), \hat{b}_1 = \mathrm{tr}(J_p W_{11}), \quad \hat{a}_2 = \mathrm{tr}(W_{21}), \hat{b}_2 = \mathrm{tr}(J_p W_{21}), \quad \ldots (26)$$

وبتطبيق الصيغة الآتية نجد القيمة التقديرية للمصفوفة W_{j1} :

$$\hat{W}_{j1} = W_j + n_j(\bar{y}_j - m_{j1}1_p)(\bar{y}_j - m_{j1}1_p)', \quad j = 1, 2. \quad \ldots\ldots (27)$$

حيث أن كل من J_p ، 1_p ، $\bar{y}_j$ و W_j سبق ذكرهما، وأن القيمة التقديرية m_{j1} يتم الحصول عليها بتطبيق الصيغة:

$$\hat{m}_{j1} = 1_p'\bar{y}_j/p, \quad j = 1, 2. \quad \ldots\ldots (28)$$

3-2-2 قاعدة التمييز والتصنيف التربيعية للحالة الثالثة [9],[8],[7]

للحالة الثالثة والموضحة فيما يأتي :

$$\text{Case 3} : \Omega_1 \neq \Omega_2 \ (\sigma_1^2 = \sigma_2^2 = \sigma^2, \ \rho_1 \neq \rho_2)$$

فأن قاعِدَةُ التمييز والتصنيف التربيعية (QCR)هي : أنهُ يتم تخَصّيَصَ الفرد ith مع المشاهدات $(y_{ji1}, y_{ji2}, \ldots, y_{jip})$ إلى ألمجموعه الأولى إذا كان:

$$\lambda(y_i) = -\frac{1}{2\sigma^2}\left[\left(\frac{1}{1-\rho_1}-\frac{1}{1-\rho_2}\right)\sum_{t=1}^{p} {y_{it}}^2 - \left(\frac{\rho_1}{(1-\rho_1)(1+(p-1)\rho_1)}\right.\right.$$
$$\left.\left. - \frac{\rho_2}{(1-\rho_2)(1+(p-1)\rho_2)}\right)\left(\sum_{t=1}^{p} y_{it}\right)^2\right]$$
$$+\frac{1}{\sigma^2}\left(\frac{c_1}{1+(p-1)\rho_1}-\frac{c_2}{1+(p-1)\rho_2}\right)\sum_{t=1}^{p} y_{it}$$
$$\geq \frac{p}{2\sigma^2}\left(\frac{{c_1}^2}{1+(p-1)\rho_1}-\frac{{c_2}^2}{1+(p-1)\rho_2}\right)-\frac{p-1}{2}\ln\left(\frac{1-\rho_2}{1-\rho_1}\right)$$
$$-\frac{1}{2}\ln\left(\frac{1+(p-1)\rho_2}{1+(p-1)\rho_1}\right), \qquad \dots\dots\dots(29)$$

عدا ذلك يتم تخَصّيَص الفرد ith مع الاستجابات $(y_{ji1}, y_{ji2}, \dots, y_{jip})$ إلى المجموعة الثانية. في قاعِدَةُ التصنيف المذكورة أنفاً فأن تقديرات الإمكان الأعظم لكل من c_1 و c_2 هي نفسها المعطاة في معادلة رقم (14). إما تقديرات الإمكان الأعظم لكل من ρ_1 ، ρ_2 و σ^2 يتم الحصول عليها بحل المعادلات الثلاثة أنيا :

$$n_1 p\hat{\sigma}^2\,\hat{\rho}_1(p-1)(1-\hat{\rho}_1)\{1+(p-1)\hat{\rho}_1\} - a_1\{1+(p-1)\hat{\rho}_1\}^2 + b_1\{1+(p-1){\hat{\rho}_1}^2\} = 0, \qquad \dots\dots(30)$$

$$n_2 p\hat{\sigma}^2\,\hat{\rho}_2(p-1)(1-\hat{\rho}_2)\{1+(p-1)\hat{\rho}_2\} - a_2\{1+(p-1)\hat{\rho}_2\}^2 + b_2\{1+(p-1){\hat{\rho}_2}^2\} = 0, \qquad \dots\dots(31)$$

$$-Np\hat{\sigma}^2(1-\hat{\rho}_1)\{1+(p-1)\hat{\rho}_1\}(1-\hat{\rho}_2)\{1+(p-1)\hat{\rho}_2\}$$
$$+ a_1\{1+(p-1)\hat{\rho}_1\}(1-\hat{\rho}_2)\{1+(p-1)\hat{\rho}_2\}$$
$$- b_1\,\hat{\rho}_1(1-\hat{\rho}_2)\{1+(p-1)\hat{\rho}_2\}$$
$$+ a_2(1-\hat{\rho}_1)\{1+(p-1)\hat{\rho}_1\}\{1+(p-1)\hat{\rho}_2\}$$
$$- b_2\,\hat{\rho}_2(1-\hat{\rho}_1)\{1+(p-1)\hat{\rho}_1\} = 0\,, \dots(32)$$

أما القيم التقديرية لكل من a_1 ، b_1 ، a_2 و b_2 تم تعريفها في الحالة الثانية في (26).

4-2-2 قاعدة التمييز والتصنيف التربيعية للحالة الرابعة [8],[7],[4]

للحالة الرابعة :

$$\text{Case 4}: \Omega_1 \neq \Omega_2 \ ({\sigma_1}^2 \neq {\sigma_2}^2\,,\ \rho_1 = \rho_2 = \rho)$$

فأن قاعِدَةُ التمييز والتصنيف التربيعية (QCR) هي : أنهُ يتم تخَصّيَص الفرد ith مع المشاهدات $(y_{ji1}, y_{ji2}, \dots, y_{jip})$ إلى ألمجموعه الأولى إذا كان:

$$\lambda(y_i) = -\frac{1}{2(1-\rho)}\left(\frac{1}{{\sigma_1}^2}-\frac{1}{{\sigma_2}^2}\right)\left[\sum_{t=1}^{p}{y_{it}}^2-\frac{\rho}{1+(p-1)\rho}\left(\sum_{t=1}^{p}y_{it}\right)^2\right]$$
$$+\frac{1}{1+(p-1)\rho}\left(\frac{c_1}{{\sigma_1}^2}-\frac{c_2}{{\sigma_2}^2}\right)\sum_{t=1}^{p}y_{it}$$
$$\geq \frac{\rho}{2(1+(p-1)\rho)}\left(\frac{{c_1}^2}{{\sigma_1}^2}-\frac{{c_2}^2}{{\sigma_2}^2}\right)-\frac{p}{2}\ln\frac{{\sigma_2}^2}{{\sigma_1}^2}, \quad \dots\dots(33)$$

عدا ذلك يتم تخصّيص الفرد ith مع الاستجابات $(y_{ji1}, y_{ji2}, \dots, y_{jip})$ إلى المجموعة الثانية. في قاعِدَةُ التصنيف (31) المذكورة أنفاً فأن تقديرات الإمكان الأعظم للمتوسطات c_j معطاة في معادلة رقم (7). إما تقديرات الإمكان الأعظم لكل من ρ ، ${\sigma_1}^2$ و ${\sigma_2}^2$ يتم الحصول عيها بواسطة حل المعادلات الثلاث أنيا :

$$-n_1p(1-\hat{\rho})\{1+(p-1)\hat{\rho}\}{\hat{\sigma}_1}^2+a_1\{1+(p-1)\hat{\rho}\}-b_1\hat{\rho}=0, \quad \dots\dots(34)$$
$$-n_2p(1-\hat{\rho})\{1+(p-1)\hat{\rho}\}{\hat{\sigma}_2}^2+a_2\{1+(p-1)\hat{\rho}\}-b_2\hat{\rho}=0, \quad \dots\dots(35)$$
$$Np\hat{\rho}{\hat{\sigma}_1}^2{\hat{\sigma}_2}^2(p-1)(1-\hat{\rho})\{1+(p-1)\hat{\rho}\}-(a_1{\hat{\sigma}_2}^2+a_2{\hat{\sigma}_1}^2)\{1+(p-1)\hat{\rho}\}^2$$
$$+(b_1{\hat{\sigma}_2}^2+b_2{\hat{\sigma}_1}^2)\{1+(p-1)\hat{\rho}^2\}=0, \quad \dots\dots(36)$$

كل من a_1، b_1، a_2 و b_2 تم تعريفها مسبقا في معادلة (25) في الحالة الثانية.

3-2 تقدير نسب التصنيف الخاطئ
Estimating Misclassification Rates

لتقيم فعالية وقدرة أساليب التصنيف للتنبؤ بعضوية المجموعة يجب أن تؤخذ بالحسبان كمية أو قدر سوء التصنيف، وهناك نوعان من الأخطاء في التصنيف وهما [4],[1] :

1-3-2 نسبة التصنيف الخطأ الظاهرة (Apparent Error Rate, APER): وهي عبارة عن نسبة عدد الإفراد اللذين تغير تصنيفها إلى عدد الإفراد الكلي في المجموعتين ، وتعطى بالصيغة الآتية :

$$APER=\frac{n_{12}+n_{21}}{n_1+n_2} \quad \dots\dots(37)$$

2-3-2 نسبة التصنيف الصحيح الظاهرة (Apparent Correct rate, APCR) :

وهي عبارة عن نسبة عدد الإفراد التي تثبت تصنيفها إلى عدد الإفراد الكلي في المجموعتين ، وتعطى بالصيغة الآتية :

$$APCR=\frac{n_{11}+n_{22}}{n_1+n_2}. \quad \dots\dots(38)$$

3 . الجانب التطبيقي

3-1 مرض داء السكري، إعراضه السريرية، أنواعه، العوامل المسببة للمرض وطرائق تشخيصها :

داء السكري (Diabetes mellitus) هو متلازمة تتصف باضطراب الأستقلاب وارتفاع شاذ في تركيز سكر الدم عن المعدل الطبيعي الذي يتراوح بين (80 – 100ملیجرام/دیسیلتر) الناجم عن عوز هرمون الأنسولين الذي تفرزه غدة البنكرياس أو انخفاض حساسية الأنسجة للأنسولين، أو كلا الأمرين.

إن من أهم الأسباب للإصابة بهذا المرض هو نقص الأنسولين الذي يعد الهرمون الأساسي الذي ينظم نقل الغلوكوز من الدم إلى معظم خلايا الجسم ، ولا سيما الخلايا العضلية والخلايا الدهنية، ولكن لا ينقله إلى خلايا الجهاز العصبي المركزي. ولذلك يؤدي نقص الأنسولين أو عدم استجابة الجسم له إلى أي نمط من أنماط السكري وهي :

- النمط الأول : يتميز النمط الأول (I) من السكري بخسارة الخلايا بيتا المنتجة للأنسولين في خلايا لانغرهانس بالبنكرياس مما يؤدي إلى نقص الأنسولين والسبب الرئيس لهذه الخسارة هو مناعة ذاتية تتميز بهجوم الخلايا تاء المناعية على خلايا بيتا المنتجة للأنسولين، أي أن مرضى هذا النوع من السكري معتمدون في علاجهم على الأنسولين الذي يعطى بشكل حقن تحت الجلد.
- النمط الثاني : يتميز النمط الثاني من السكري باختلافه عن النمط الأول من حيث وجود مقاومة مضادة لمفعول الأنسولين فضلاً عن قلة إفراز الأنسولين ولا تستجيب مستقبلات الأنسولين الموجودة في الأغلفة الخلوية لمختلف أنسجة الجسم بصورة صحيحة للأنسولين، أي أن مرضى هذا النوع من السكري هم مرضى سكري معتمدون في علاجهم على الحبوب.

وهناك عدة أعراض توحي بالإصابة بهذا المرض ومنها زيادة عدد مرات التبول (بسبب ارتفاع الضغط التناضحي)، زيادة الإحساس بالعطش (تنتج عنها زيادة تناول السوائل لمحاولة تعويض زيادة التبول)، التعب الشديد والعام، فقدان الوزن رغم تناول الطعام بانتظام (شهية أكبر للطعام)، تباطؤ شفاء الجروح، وتغيّم الرؤية.

أما تشخيص النمط الأول والعديد من حالات النمط الثاني من السكري فيتم بناءا على الأعراض الأولية التي تظهر في بداية المرض مثل كثرة التبول والعطش الزائد وقد يصاحبها فقد للوزن، ويتم عادة تشخيص بقية أنماط السكري بطرائق أخرى مثل اكتشاف ارتفاع مستوى غلوكوز الدم أثناء أجراء أحد التحاليل؛ ويتم عادة اكتشاف المرض عندما يعاني المريض من مشكلها يسببها السكري بكثرة مثل السكتات القلبية، اعتلال الكلى، بطئ التئام الجروح أو تقيح القدم، مشكلة معينة في العين.

ويمكن تشخيص المرض من خلال نسبة الهيموغلوبين الغليكوزيلاتي (Glycosylated hemoglobin (HbA1c)) وهو ما يسمى بالتحليل التراكمي لسكر الدم، حيث يعتبر بروتين (HbA1c) تحليل تشخيصي لمرض السكري في حال إذا كانت نسبته في الدم أكثر من (6.5%) حيث أن النسبة الطبيعية المعتمدة هي أقل من (5.9%) ويزداد في مرضى السكري في حالة عدم الانتظام في العلاج، ويذكر أن هذا التحليل يتميز بثبات نسبته وعدم تأثره بصيام المراجع أو بعوامل أخرى يمكن أن تؤثر عند سحب عينات دم لتحليلها، وبذلك فيمكن عمله في أي وقت ولا حاجة لأن يكون الشخص صائما قبل سحب العينة.

ومن خلال ذلك تتضح أهمية التشخيص المبكر للمرض لتجنب الحالات المعقدة والتي تؤدي إلى مضاعفات خطيرة وحتى إلى الوفاة المبكرة مع تطور المرض إلى الحالات المتقدمة، وأن استخدام الأساليب الإحصائية ومنها الدال التمييزية هي لمساعدة ذوي الاختصاص لتشخيص مجاميع المرض وفي حالاتها المبكرة.

٤-٢ وصف بيانات الدراسة

شملت عينة البحث موضوع الدراسة بيانات لمجموعتين تضمنت المرضى المصابين بداء السكري، إذ تم الحصول على البيانات من البطاقات الخاصة بالمرضى في المركز الوطني للسكري/الجامعة المستنصرية لفحص تأثير عقارين هما عقار حبوب ريباجلينيد (Repaglinide) وعقار الأنسولين المخلوط (Novo Mix) على استجابة مرضى السكري والسيطرة على المرض ولمدة زمنية قدرها ست أشهر، إذ تمثل هذه البيانات القياسات الشهرية لنسبة بروتين الهيموغلوبين الغليكوزيلاتي (HbA1c) في الدم والمأخوذة في ثلاث مراحل، المرحلة الأولى كانت في بداية التجربة (month 0)، المرحلة الثانية كانت بعد ثلاث أشهر (month 3) ، أما المرحلة الثالثة فكانت بعد ست أشهر (month 6). ولمجموعتين من المرضى، تمثلت المجموعة الأولى بعدد (38) مريضاً يعانون من داء السكري من النمط الأول (I) (مرضى سكري معتمدون في علاجهم على الأنسولين)، في حين كانت المجموعة الثانية تمثل منهم (33) مريضاً يعانون من داء السكري من النمط الثاني (II) (مرضى سكري ضعيفي السيطرة السكرية معتمدون في علاجهم على الحبوب).

حيث قمنا بتشخيص نوعين من أمراض السكر بالاعتماد على قواعد التمييز والتصنيف بوجود تركيبة التباين المشترك المتماثلة الكروية (CS) لما لها من دور فعال في الحد من عدد المعلمات غير المعلومة والمطلوب تقديرها لبناء الأنموذج التصنيفي، والمقارنة بينها عبر مجموعة من شروط التجانس وعدم التجانس لمصفوفة التباين المشترك والوصول إلى أفضل قاعدة تصنيفية لتشخيص المرضى.

وعادة ما تكون هناك أخطاء في عملية التشخيص حالها حال التشخيص الطبي الذي قد يصاحبه بعض الأخطاء، لذلك تمت مقارنة النسب المئوية للتصنيف الخاطئ واعتماد قاعدة التصنيف التي تعطي أقل نسبة خطأ تصنيف ظاهرة كأسلوب مساعد في عملية التشخيص لمرضى المجموعتين.

3-3 التحليل التمييزي والتصنيفي لبيانات القياسات المكررة في حالة تركيبة التباين المشترك المتماثلة الكروية (Compound Symmetric)

فيما يأتي الحالات الأربعة للدالة التمييزية والتصنيفية بوجود تركيبة التباين المشترك المتماثلة الكروية $(\mathbf{CS})$:

3-3-1 الحالة الأولى

بافتراض وجود الحالة الأولى :

$$\mathbf{Case\ 1}: \Omega_1 = \Omega_2\ (\sigma_1{}^2 = \sigma_2{}^2 = \sigma^2,\ \rho_1 = \rho_2 = \rho)$$

للحصول على نتائج الدالة التمييزية في هذه الحالة ينبغي أولا تقدير كل من متوسط مناسبات القياسات المكررة للمجموعة الأولى $\hat{c}_1$ والثانية $\hat{c}_2$ والموضحة في معادلة (13) وذلك لغرض تَخْمِين القيم التقديرية لكل من متجه متوسطات القياسات المكررة المنتظمة بدون تأثير عامل الوقت للمجموعة الأولى $\hat{\mu}_1$ والثانية $\hat{\mu}_2$ وكما يأتي:

نجد $\bar{y}_1$ و $\bar{y}_2$ للمجموعتين بتطبيق (13) والمقدرين بالمتجهين أدناه :

$$\bar{y}_1 = [8.342 \quad 9.474 \quad 10.939]'$$
$$\bar{y}_2 = [5.712 \quad 6.091 \quad 6.412]'$$

ومن ثم فأن القيمة التقديرية لكل من $\hat{c}_1$ و $\hat{c}_2$ من خلال (14) هي :

$$\hat{c}_1 = 9.585\ , \qquad \hat{c}_2 = 6.072$$

ومنهما نجد تقدير $\hat{\mu}_1$ و $\hat{\mu}_2$ باعتماد (5) :

$$\hat{\mu}_1 = \hat{c}_1 1_3 = [9.585 \quad 9.585 \quad 9.585]$$
$$\hat{\mu}_2 = [6.072 \quad 6.072 \quad 6.072]$$

وبعدها نجد تقدير مصفوفة التباين والتباين المشترك المتماثلة الكروية $(\mathbf{CS})$، ولغرض تقديرها في هذه الحالة فيبغي أولا إيجاد القيم التقديرية لكل من $\hat{a}_1$ ، $\hat{b}_1$ و $\hat{a}_2$ و $\hat{b}_2$ بالاستناد إلى صيغة كل منهم والتي تم إيضاحها في (18) وكما يأتي :

لإيجاد قيمة a_1 و b_1 نجد W_1 وبتطبيق الصيغة (12) فأن :

$$W_1 = \begin{bmatrix} 58.830 & 49.247 & 34.188 \\ & 88.652 & 62.937 \\ & & 131.757 \end{bmatrix}$$

ومن ثم فان تقدير قيمة a_1 و b_1 تكون مساوي إلى :

$\hat{a}_1 = 279.239, \quad \hat{b}_1 = 571.983$

ولإيجاد قيمة a_2 و b_2 نجد W_2 وبتطبيق الصيغة (12) فأن :

$$W_2 = \begin{bmatrix} 43.296 & 34.880 & 34.272 \\ & 38.880 & 34.592 \\ & & 53.504 \end{bmatrix}$$

ومنها فأن تقدير قيمتي a_2 و b_2 مساوي إلى :

$\hat{a}_2 = 135.68, \quad \hat{b}_2 = 343.168$

وبعد إيجاد القيم التقديرية $\hat{a}_1$ ، $\hat{b}_1$ ، $\hat{a}_2$ و $\hat{b}_2$ يتم تعويض كل منهما في (16) و (17) لغرض الحصول على تقديرات الإمكان الأعظم $\hat{\rho}$ و $\hat{\sigma}^2$ المطلوب إيجادها لبناء الدالة التمييزية للحالة الأولى للتحليل التمييزي بوجود تركيبة (CS).وباستعمال برنامج (MATLAB, V 7.6.0) وبتَوظِيف الطرائق التكرارية (Iterative methods) لحل أنظمة المعادلات الآتية تم الحصول على النتائج الآتية :

$$\hat{\theta} = \begin{bmatrix} \hat{\rho} \\ \hat{\sigma}^2 \end{bmatrix} = \begin{bmatrix} 0.1647 \\ 2.9696 \end{bmatrix}$$

ومن ثم فأن مصفوفة التباين والتباين المشترك المتماثلة الكروية في هذه الحالة تكون مساوية إلى:

$$\hat{\Omega}_{cs} = \begin{bmatrix} 2.9696 & 0.4891 & 0.4891 \\ & 2.9696 & 0.4891 \\ & & 2.9696 \end{bmatrix}$$

ومن ثم بعد إيجاد القيم التقديرية لكل من $\hat{\mu}_1$ ، $\hat{\mu}_2$ و $\hat{\Omega}_{cs}$ فانه من الممكن لإيجاد متجه تقديرات معاملات الدالة التمييزية $\hat{a}$ في حالة تركيبة التباين المشترك المتماثلة الكروية والمعطى في المعادلة رقم (4) وكالاتي :

$$\hat{a} = \begin{bmatrix} 2.9696 & 0.4891 & 0.4891 \\ & 2.9696 & 0.4891 \\ & & 2.9696 \end{bmatrix}^{-1} \begin{bmatrix} 3.513 \\ 3.513 \\ 3.513 \end{bmatrix} = [0.8899 \quad 0.8899 \quad 0.8899]'$$

وبهذا فأن تقدير الدالة التمييزية في حالة تركيبة التباين المشترك هذه والمعطاة في المعادلة رقم (4) تكون مساوية إلى :

$$D = 0.8899\, y_{ji1} + 0.8899\, y_{ji2} + 0.8899\, y_{ji3}$$

ألان وبعد تقدير الدالة التمييزية لهذه الحالة وبعد تعويض كل من $\hat{c}_1$ و $\hat{c}_2$ في قاعدة التمييز والتصنيف $\lambda(y_i)$ المعطاة في معادلة (6) فأنه يتم تخَصّيَص الفرد ith مع المشاهدات $(y_{ji1}, y_{ji2}, \cdots, y_{jip})$ إلى ألمجموعه الأولى فيما إذا كان $\lambda(y_i)$ بالصيغة الآتية :

$$\lambda(y_i) = \sum_{t=1}^{p} y_{it} \geq 23.4855$$

وعدا ذلك يتم تخَصّيَصَهُ إلى المجموعة الثانية .وبعد بناء قاعدة التمييز والتصنيف ، فأن الجدول الأتي يعطي كِلَ من عدد المشاهدات في المجموعتين الأولى والثانية مع الإعداد الصحيحة والخاطئة التي تم التنبؤ بها إلى كلتا المجموعتين :

الجدول (1): نتائج التحليل التمييزي لبيانات القياسات المكررة بوجود تركيبة التباين المشترك (CS) (جدول النسب المئوية للتصنيف في كل مجموعة للحالة الأولى)

From Group	I	II	Total
I	36 94.74	2 5.26	38 100.00
II	1 3.03	32 96.97	33 100.00
Total	37 52.11	34 47.89	71 100.00

في الجدول المذكورة أنفاً فأن الرقم (94.74 %) يمثل نسبة الإفراد الذين تم تصنيفهم بشكل صحيح إلى المجموعة الأولى، وهكذا $\left({}^{36}/_{38} = 0.9474\right)$ بالنسبة لبقية النسب، وبتطبيق الصيغة رقم (37) نجد أن نسبة التصنيف الخطأ الظاهرة الكلية مساوية إلى :

$$APER = \frac{2+1}{38+33} = 0.04225$$

أي أن العدد الكلي للإفراد المصنفين بشكل غير صائب هو (3) أي ما يعادل (4.225%) لحجم العينة الكلي من 71 فرداً مشاركاً في الدراسة.

وكذلك بتطبيق صيغة (38) فأن نسبة التصنيف الصحيح الظاهرة الكلية أيضا :

$$APCR = 0.9577$$

أي أن العدد الكلي للإفراد المصنفين بشكل صحيح هو (68) أي ما يعادل (95.77%) لحجم العينة الكلي من 71 فرداً مشاركاً في الدراسة.

2-3-3 الحالة الثانية

بافتراض وجود الحالة الثانية :

$$\text{Case 2} : \Omega_1 \neq \Omega_2 \left(\sigma_1{}^2 \neq \sigma_2{}^2 \, , \, \rho_1 \neq \rho_2\right)$$

لبناء قاعدة التمييز والتصنيف التربيعية للفرد المشارك أو الجديد في الدراسة والمراد تمييزه أو تصنيفه فيبغي أولا إيجاد القيم التقديرية لكل من $\hat{a}_1$ ، $\hat{b}_1$ و $\hat{a}_2$ ، $\hat{b}_2$ بالاستناد إلى صيغة كل منهم والتي تم إيضاحها في في (26) وكما ياتي :

ولغرض إيجاد $\hat{a}_1$ و $\hat{b}_1$ يجب تقدير المصفوفة W_{11} فينبغي إيجاد القيم التقديرية لكل من W_1 (سبق تقديرها في الحالة الأولى) و m_{11} ومن خلال تطبيق الصيغة (28) نجد :

$$\hat{m}_{11} = 9.585$$

وبتعويضها في معادلة (27) نجد :

$$\hat{W}_{11} = \begin{bmatrix} 117.542 & 54.490 & -29.767 \\ & 89.120 & 57.226 \\ & & 201.423 \end{bmatrix}$$

ومن ثم فان تقدير قيمتي a_1 و b_1 وبتطبيق (26) يكون مساوياً إلى :

$$\hat{a}_1 = 408.085, \quad \hat{b}_1 = 571.983$$

وكذلك بالنسبة إلى $\hat{a}_2$ و $\hat{b}_2$ نجد تقدير المصفوفة W_{21} فينبغي إيجاد القيم التقديرية لكل من W_2 (سبق تقديرها في الحالة الأولى) و m_{21} من خلال تطبيق (28) فتساوي :

$$\hat{m}_{21} = 6.072$$

ومن ثم فأن W_{21} يكون مساوياً إلى :

$$\hat{W}_{21} = \begin{bmatrix} 47.443 & 34.661 & 30.355 \\ & 38.892 & 34.799 \\ & & 57.203 \end{bmatrix}$$

ومنها فأن تقدير قيمتي a_2 و b_2 يكون مساوي إلى :

$$\hat{a}_2 = 143.538, \quad \hat{b}_2 = 343.168$$

وبعدها نقوم بتعويض هذه القيم في المعادلات (22)، (23)، (24) و (25) والمعطاة في الحالة الثانية لغرض الحصول على تقديرات الإمكان الأعظم للمعالم $\hat{\rho}_1$ و $\hat{\sigma}_1^2$ ، $\hat{\rho}_2$ و $\hat{\sigma}_2^2$ المطلوب إيجادها لبناء قاعدة التمييز والتصنيف التربيعية للحالة هذه، وبحل تلك المعادلات فأن المتجة الأتي يعطي تقديرات الإمكان الأعظم لهذه الحالة :

$$\hat{\theta} = \begin{bmatrix} \hat{\rho}_1 \\ \hat{\sigma}_1^2 \\ \hat{\rho}_2 \\ \hat{\sigma}_2^2 \end{bmatrix} = \begin{bmatrix} 0.1954 \\ 3.5127 \\ 0.1671 \\ 3.2328 \end{bmatrix}$$

حيث أن متوسط مناسبات القياسات المكررة للمجموعة الأولى والثانية قد سبق تقديرهما في الحالة الأولى وهما :

$$\hat{c}_1 = 9.585, \quad \hat{c}_2 = 6.072.$$

وبعد تعويض كل من $\hat{\rho}$ ، $\hat{\sigma}_1^2$ ، $\hat{\sigma}_2^2$ ، $\hat{c}_1$ و $\hat{c}_2$ في قاعدة التمييز والتصنيف $\lambda(y_i)$ المعطاة في معادلة (19) فأنه يتم تخصّيص الفرد ith مع المشاهدات $(y_{ji1}, y_{ji2}, \dots, y_{jip})$ إلى ألمجموعه الأولى فيما إذا كان $\lambda(y_i)$ بالصيغة الآتية :

$$\lambda(y_i) \geq 15.4133$$

وعدا ذلك يتم تخصيصه إلى المجموعة الثانية. وبالتالي وبالاستناد إلى قاعدة التصنيف هذه فأن الجدول التالي يعطي كِلَ من عدد المشاهدات في المجموعتين الأولى والثانية مع الإعداد الصحيحة والخاطئة التي تم التنبؤ بها إلى كلتا المجموعتين:

الجدول (2): نتائج التحليل التمييزي لبيانات القياسات المكررة بافتراض تركيبة التباين المشترك (CS)
(جدول النسب المئوية للتصنيف في كل مجموعة للحالة الثانية)

From Group	I	II	Total
I	32 84.21	6 15.79	38 100.00
II	4 12.12	29 87.88	33 100.00
Total	36 50.70	35 49.30	71 100.00

ومن ثم وبالاستناد إلى قاعدة التصنيف هذه فأن نسبة التصنيف الخطأ الظاهرة الكلية، ونسبة التصنيف الصحيح الظاهرة الكلية أيضا وبتطبيق الصيغتين (37) و (38) على التوالي :

$$APER = 0.1408$$
$$APCR = 0.8592$$

واضح من خلال APER و APCR أن العدد الكلي للإفراد المصنفين بشكل غير صائب هو (10) أشخاص أي ما يعادل (14.08%) أما الذين تم تصنيفهم بشكل صائب فكان عددهم (32 + 29 = 61) فرد إي ما يعادل (85.92%) لحجم العينة الكلي من 71 فرداً مشاركاً في الدراسة.

3-3-3 الحالة الثالثة

وبافتراض وجود الحالة الآتية :

$$\text{Case 3}: \Omega_1 \neq \Omega_2 \ (\sigma_1{}^2 = \sigma_2{}^2 = \sigma^2 \ , \ \rho_1 \neq \rho_2)$$

ولغرض بناء قاعدة التمييز والتصنيف للفرد المشارك أو الجديد في الدراسة والمراد تمييزه أو تصنيفه نقوم بتعويض القيم التقديرية والتي تم تقديرها في الحالة السابقة لكل من :

$$\hat{a}_1 = 408.085, \hat{b}_1 = 571.983, \quad \hat{a}_2 = 143.538, \hat{b}_2 = 343.165$$

في المعادلات (30)، (31) و (32) لغرض الحصول على تقديرات الإمكان الأعظم ρ_1 ، ρ_2 و σ^2 المطلوب إيجادها لبناء قاعدة التمييز والتصنيف التربيعية للحالة الثالثة للتحليل التمييزي بوجود تركيبة (CS)، ويمثل المتجه الأتي تقديرات الإمكان الأعظم التي تم التوصل إليها بحل تلك المعادلات :

$$\hat{\theta} = \begin{bmatrix} \hat{\rho}_1 \\ \hat{\rho}_2 \\ \hat{\sigma}^2 \end{bmatrix} = \begin{bmatrix} 0.1714 \\ 0.8477 \\ 2.9963 \end{bmatrix}$$

حيث أن متوسط مناسبات القياسات المكررة للمجموعة الأولى والثانية قد سبق تقديرهما في الحالة الأولى وهما :

$$\hat{c}_1 = 9.585 \,, \quad \hat{c}_2 = 6.072$$

وبعد تعويض كل من $\hat{\rho}_1$ ، $\hat{\rho}_2$ ، $\hat{\sigma}^2$ ، $\hat{c}_1$ و $\hat{c}_2$ في قاعدة التمييز والتصنيف $\lambda(y_i)$ المعطاة في معادلة (29) فأنه يتم تخصّيص الفرد ith مع المشاهدات $(y_{ji1}, y_{ji2}, \dots, y_{jip})$ إلى ألمجموعة الأولى فيما إذا كان $\lambda(y_i)$ بالصيغة الآتية :

$$\lambda(y_i) \geq 28.749$$

وعدا ذلك يتم تخصّيصه إلى المجموعة الثانية. وبعد بناء قاعدة التمييز والتصنيف ذات تركيبة التباينات المتماثلة الكروية وللحالة الثانية، فأن الجدول الأتي يعطي كِلَ من عدد المشاهدات في المجموعتين الأولى والثانية مع الإعداد الصحيحة والخاطئة التي تم التنبؤ بها إلى كلتا المجموعتين :

الجدول (3): نتائج التحليل التمييزي لبيانات القياسات المكررة بوجود تركيبة التباين المشترك (CS) (جدول النسب المئوية للتصنيف في كل مجموعة للحالة الثالثة)

From Group	I	II	Total
I	35 92.11	3 7.89	38 100.00
II	3 9.09	30 90.91	33 100.00
Total	38 53.52	33 46.48	71 100.00

ومن ثم وبالاستناد إلى قاعدة التصنيف هذه فأن نسبة التصنيف الخطأ الظاهرة الكلية، ونسبة التصنيف الصحيح الظاهرة الكلية أيضا قدرت إلى :

$$APER = 0.0845$$

$$APCR = 0.9155$$

واضح من خلال APER و APCR أن العدد الكلي للإفراد المصنفين بشكل غير صائب هو (6) أشخاص أي ما يعادل (8.45%) أما الذين تم تصنيفهم بشكل صائب فكان عددهم (65) فرداً إي ما يعادل (91.55%) لحجم العينة الكلي من 71 فرداً مشاركاً في الدراسة.

3-4-3 الحالة الرابعة

وبافتراض وجود الحالة الأخيرة :

$$\text{Case 4} : \Omega_1 \neq \Omega_2 \ (\sigma_1{}^2 \neq \sigma_2{}^2 \ , \ \rho_1 = \rho_2 = \rho)$$

لغرض بناء قاعدة التمييز والتصنيف للفرد المشارك أو الجديد في الدراسة والمراد تمييزه أو تصنيفه نقوم بتعويض القيم التقديرية والتي تم تقديرها في الحالة الثانية لكل من :

$$\hat{a}_1 = 408.085, \hat{b}_1 = 571.983, \quad \hat{a}_2 = 143.538 \ , \hat{b}_2 = 343.165$$

في المعادلات (34)، (35) و (36) والمعطاة في الحالة الرابعة لغرض الحصول على تقديرات الإمكان الأعظم للمعالم ρ و $\sigma_1{}^2$، $\sigma_2{}^2$ المطلوب إيجادها لبناء قاعدة التمييز والتصنيف التربيعية للحالة هذه ويمثل المتجه الأتي تقديرات الإمكان الأعظم التي تم التوصل إليها بحل تلك المعادلات :

$$\hat{\theta} = \begin{bmatrix} \hat{\rho} \\ \hat{\sigma}_1{}^2 \\ \hat{\sigma}_2{}^2 \end{bmatrix} = \begin{bmatrix} 0.4882 \\ 4.5724 \\ 3.1599 \end{bmatrix}$$

حيث أن متوسط مناسبات القياسات المكررة للمجموعة الأولى والثانية قد سبق تقديرهما في الحالة الأولى وهما :

$$\hat{c}_1 = 9.585 \ , \quad \hat{c}_2 = 6.072 \ .$$

وبعد تعويض كل من $\hat{\rho}$ ، $\hat{\sigma}_1{}^2$ ، $\hat{\sigma}_2{}^2$ ، $\hat{c}_1$ و $\hat{c}_2$ في قاعدة التمييز والتصنيف $\lambda(y_i)$ للحالة الرابعة المعطاة في معادلة (33) فأنه يتم تخَصّيَصَ الفرد ith مع المشاهدات $(y_{ji1}, y_{ji2}, \cdots, y_{jip})$ إلى المجموعة الأولى فيما إذا كان $\lambda(y_i)$ بالصيغة الآتية :

$$\lambda(y_i) \geq 5.8399$$

وعدا ذلك يتم تخَصّيَصَه إلى المجموعة الثانية. وبعد بناء قاعدة التمييز والتصنيف ذات تركيبة التباينات المتماثلة الكروية لهذه الحالة، فأن الجدول الأتي يعطي كِلَ من عدد المشاهدات في المجموعتين الأولى والثانية مع الإعداد الصحيحة والخاطئة التي تم التنبؤ بها إلى كلتا المجموعتين :

الجدول (4): نتائج التحليل التمييزي لبيانات القياسات المكررة بافتراض تركيبة التباين المشترك (CS)
(جدول النسب المئوية للتصنيف في كل مجموعة للحالة الرابعة)

From Group	I	II	Total
I	37 97.37	1 2.63	38 100.00
II	6 18.18	27 81.82	33 100.00
Total	43 60.56	28 39.44	71 100.00

ومن ثم وبالاستناد إلى قاعدة التصنيف هذه فأن نسبة التصنيف الخطأ الظاهرة الكلية، ونسبة التصنيف الصحيح الظاهرة الكلية أيضا قدرت إلى :

$$APER = 0.0986$$
$$APCR = 0.9014$$

واضح من خلال APER و APCR أن العدد الكلي للإفراد المصنفين بشكل غير صائب هو (7) أشخاص أي ما يعادل (9.86%) أما الذين تم تصنيفهم بشكل صائب فكان عددهم (64) فرداً إي ما يعادل (90.14%) لحجم العينة الكلي من 71 فرداً مشاركاً في الدراسة.

4 . الاستنتاجات والتوصيات

4-1 الاستنتاجات (Conclusions)

من خلال ما تم عرضه في بحثنا هذا لمجموعة بيانات حقيقية تم التوصل إلى الاستنتاجات الآتية:

١- ظهر إن أنموذج التحليل التمييزي الخطي (LDA) بوجود تركيبة التباين المشترك المتماثلة الكروية (CS) كان الأكثر ملاءمة في تصنيف مرضى داء السكري ويعود السبب في ذلك إلى امتلاكه اقل نسبة من نسب التصنيف الخطأ الظاهرة الكلية والتي بلغت فيه (APER = 4.23%).

٢- تبين أن أسلوب التحليل التمييزي التربيعي (QDA) وبوجود تركيبة (CS) للحالة الثالثة فأن قواعد التصنيف كانت أنسب في تصنيف الأشخاص ويعود سبب ذلك إلى امتلاكها أقل نسبة من نسب التصنيف الخطأ الظاهرة الكلية (APER = 8.45%).

٣- أظهرت قواعد التصنيف انه كلما زادت عدد المعلمات المطلوب تقديرها لبناء قاعدة التمييز والتصنيف للمرضى فأن نسبة التصنيف الخطأ الظاهرة الكلية (APER) تبدأ بالزيادة وهذا ما يقلل من كفاءة قواعد التصنيف لهذا النوع من البيانات.

2-4 التوصيات (Recommendations)

اعتماداً على ما توصل إليه هذا البحث من استنتاجات نوصي بما يأتي :

١- عند التركيز على أقل عدد من المعلمات لبناء قاعدة التمييز والتصنيف فانه من الممكن لإتباع أنموذج التحليل التمييزي الخطي (LDA) بوجود تركيبة التباين المشترك المتماثلة الكروية (CS) لتصنيف مرضى داء السكري.

٢- اظهر أنموذج التحليل التمييزي ألتربيعي (QDA) للحالة الثالثة كفاءته في تصنيف الإفراد لذا يوصى باستعماله مقارنة مع الحالات الأخرى بوجود تركيبة التباين المشترك المتماثلة الكروية (CS) .

٣- كذلك نوصي بتوظيف تراكيب أخرى من نماذج نمط التغاير لمصفوفة التباين والتباين المشترك ومنها، نموذج تركيبة الأوساط المتحركة (Moving average structure)، وأنموذج التركيبة المعممة أو الدائرية (Circular structure) لأساليب التحليل التمييزي للقياسات المكررة.

المصادر References

أولا : المصادر العربية

١- الراوي، عمر فوزي و دبدوب، مروان عبد العزيز (2007).استخدام السيطرة النوعية والدالة التمييزية في الدراسات التطبيقية.مجلة التربية والعلم، مجلد (19)، العدد الأول، كلية علوم الحاسبات والرياضيات، جامعة الموصل.ص (220 – 203).

ثانيا : المصادر الأجنبية

2- Choi, S.C. (1972). Classification of Multiply Observed Data ,Biometrical Journal, 14(1),8–11.

3- Fitzmaurice , G. M., Laird, N. M., and Ware, J. H. (2004). Applied Longitudinal Analysis. New Jersey: John Wiley & Sons, Inc., Hoboken.

4- Lix, L. M., and Sajobi, T. T. (2010). Discriminant Analysis for repeated Measures Data: a review. Frontiers in Quantitative Psychology and Measurement, 1, 1 – 9.

5- McLachlan, G. J. (1992). Discriminant Analysis and Statistical Pattern Recognition. New York :Wiley.

6- Rencher, A.C. (2002).Methods of multivariate analysis, Second Edition. New York :John Wiley & Sons, Inc.

7- Roy A, Khattree R .(2008).Classification rules for repeated measures data from biomedical research. In :Khattree, R and Naik, D (Eds) Computational methods in biomedical research. Chapman and Hall/CRC Biostatistics Series ,London, pp 323–370.

8- Roy, A., & Khattree, R. (2005a). Discrimination and classification with repeated measures data under different covariance structures. Communications in Statistics Simulation and Computation, 34, 167 – 178.

9- Roy, A., and Khattree, R. (2005b). On discrimination and classification with multivariate repeated measures data. Journal of Statistical Planning and Inference, 134, 462 – 485

10- Yilliam,L. Li,L. Lix,L and Sajobi,T.(2011).Discriminant Analysis for Repeated Measures Data: Effects of Mean and Covariance Misspecification on Bias and Error in Discriminant Function Coefficients.Journal of Modern Applied Statistical Methods,10 (2), 571-582.

11- Wolynski,W , Krzysko, M., Madry, W., Pluta, S., and Skorzybut, M.,.(2010). Analysis of multivariate repeated measures data. Colloquium Biometricum,40,117-133.

Building discriminant function for repeated measurements data under compound symmetry (CS) covariance structure and applied in the health field

Abstract

Discriminant analysis is a technique used to distinguish and classification an individual to a group among a number of groups based on a linear combination of a set of relevant variables know discriminant function. In this research discriminant analysis used to analysis data from repeated measurements design. We will deal with the problem of discrimination and classification in the case of two groups by assuming the Compound Symmetry covariance structure under the assumption of normality for univariate repeated measures data.

The importance of this research represented to find the best model to classify a group of patients who suffer from diabetes. For the purpose of studying the effects of the number of correlations, variances, and umber of repeated measurements on the performance of classification rules for this type of data based on monthly measurements of glycosylated hemoglobin (HbA1C) in the blood was taken in three stages, which is the beginning of the experiment, and after three months, and then six months for two groups of patients, the first group consists of (38) patients was suffered from diabetes type (I) and the second group includes (33) patients suffered from diabetes type (II).

And through this research, concluded that when the number of parameters began to increase. Thus, the apparent error rate begin to increasing, and this is what reduces the efficiency of classification rules for this type of data. And we recommend by using the linear discriminant function when you focus on the least number of parameters to build the classification rule. And quadratic discriminant procedure Represented by equal the variance and different correlation parameters under compound symmetry covariance structures.

Key words/ discriminant analysis – repeated measures date- correlation coefficient- variances- com pound symmetry –covariance structure(cs)