

Building Discriminant Model For Repeated Measurements Data Under Autoregressive (AR-1) Covariance Structure For patients with diabetes

بناء أنموذج تمييزي لبيانات القياسات المكررة بوجود بنية أنموذج الانحدار الذاتي ذو المرتبة الأولى (AR-1) للتباين المشترك لمرضى داء السكري

أ.د. ظافر حسين رشيد مؤمن عباس موسى
قسم الإحصاء/ كلية الإدارة والاقتصاد/ جامعة بغداد

بحث مستقل

المستخلص

يعرف التحليل التمييزي بأنه أسلوب إحصائي يعتمد على عينه من المفردات المأخوذة من مجتمعات معلومة، وذلك لبناء قاعدة يمكنها المساعدة مستقبلاً في تعيين المجتمع الذي تنتمي إليه المفردات الجديدة، هذا البحث أستعمل التحليل التمييزي لتحليل بيانات تصاميم القياسات المكررة، حيث تركزت أهميته في مشكلة التمييز ثنائي المجموعة بوجود حالة التوزيع الطبيعي للمجموعات لبيانات القياسات المكررة المستقاة طولياً عبر الزمن على متغير الاستجابة نفسه ولكل وحدة تجريبية.

من الباحثين الذين تناولوا هذه المشكلة (Roy & Khattree, 2005) فقد قدما عملاً وصفاً 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 التمييز. ويختلف عمل هذين الباحثين عن عمل الباحثين (Kshirsaga & Albert, 1993) حيث قدما أسلوبين للتحليل التمييزي الوصفي وذلك لوصف أو لتقييم الأهمية النسبية لمناسبات القياسات المكررة للتمييز بين المجموعات، الأسلوب الأول مستند إلى تحليل (MANOVA) أما الأسلوب الثاني اعتمد على نموذج (Growth curve) في حين أن هذه الأساليب لن تضع أية قيود أو تراكيب على مصفوفة التباينات، وأن أهمية البحث هذا تمثلت بإيجاد أفضل أنموذج لتصنيف مجموعة من مرضى داء السكري وذلك لغرض دراسة تأثيرات كل من عدد الارتباطات، 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)،

حيث تمت نمذجة هذه البيانات بوجود تركيبة التباين المشترك لأنموذج الانحدار الذاتي من المرتبة الأولى (AR – 1) وذلك للحد من عدد المعلمات غير المعلومة لبناء الأنموذج التمييزي عبر مجموعة من شروط التجانس وعدم التجانس لمصفوفة التباين المشترك، وبالإضافة إلى وجود تركيبة لمتجهات المتوسط حيث سنعتبر هذه المشكلة مع متجهات متوسط منتظمة أي بدون تأثير عامل الوقت، والتي من شأنها أن تزيد من دقة التصنيف لهذا النوع من البيانات، وسنوضح بعض العمليات الحسابية لتقديرات الإمكان الأعظم لمعلمات المجتمع غير المعلومة لأساليب التحليل التمييزي لهذا النوع من البيانات.

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 فإن نسبة التصنيف الخاطئ الظاهرة الكلية (APER) تبدأ بالزيادة وهذا ما يقلل من كفاءة قواعد التصنيف لهذا النوع من البيانات، واعتماداً على ما توصل إليه البحث فإن ما يوصى به وعند التركيز على أقل عدد من المعلمات لبناء قاعدة التصنيف فانه من الممكن لإتباع أسلوب التحليل التمييزي الخطي بوجود تركيبة التباين المشترك (AR – 1) لتصنيف المرضى.

الكلمات الرئيسية : تركيبة التباين المشترك، الانحدار الذاتي من الدرجة الأولى (AR-1)، قانون التصنيف، بيانات القياسات المكررة، مقدرات الإمكان الأعظم، تراكيب حول المتوسطات.

Abstract

discriminant analysis is a statistical technique Based on a sample of individuals Taken from communities known in advance, In order to build a model that could help to assigned the group that belong to the new individual. 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 Under The Assumption of Multivariate Normality For Univariate Repeated Measures Data .

Researchers who studied this problem (Roy & Khattree, 2005), where he presented a descriptive study

of the two methods under different structures of the covariance matrix To reduce the number of parameters is required to build a classification rule, While researchers (Kshirsagar & Albert, 1993) studied two methods Growth curve and ANCOVA models for descriptive discriminant analysis To describe the relative importance of the occasions repeated measurements to distinguish between groups.

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 Number 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,

which has modeled by assuming the Autoregressive (AR-1) covariance structure To reduce the number of parameters is required to build a classification rule Across a Range of Conditions of Homogeneity and Heterogeneity For The Covariance Matrix. In Addition to Assuming Covariance Structures we Will Assume The Structured Mean Vectors Without Time Effect on each Individual. And Some of Computational Schemes For Maximum Likelihood Estimates of Required Population Parameters are Given.

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 under (AR-1) Covariance Structures, When you focus on the least number of parameters to build the Classification rule.

Keywords : Covariance Structures; Autoregressive (AR-1); Classification rule; Maximum likelihood estimates; Repeated measures data; Structure on mean.

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 أو عدة متغيرات.

وأن لأساليب التحليل التمييزي الخطي التقليدي (LDA classical) بعض الافتراضات الخاصة والتي يجب توفرها عند استعمالها لتحليل هذا النوع من البيانات؛فإنها تفترض أن تكون عدد المشاهدات للوحدة التجريبية كاملة، صعوبة تضمين تأثيرات المتغيرات المشاركة، وكذلك فإنها لا يمكن تطبيقها عندما تكون عدد الوحدات التجريبية الكلي أقل من عدد مناسبات القياسات المكررة، وبالإضافة إلى ذلك فإنها تعتمد على تقدير عدد كبير من المعلمات المطلوب تقديرها لبناء النموذج التمييزي والتي يزداد عددها بزيادة عدد القياسات المكررة، وعند عدم تساوي مصفوفة التباينات للمجموعات فإن استعمال أسلوب التحليل التمييزي التريبيعي يكون ضروريا.

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 للنموذج المقدر والذي سوف يعكس خصائص من شأنها أن تقلل من دقة قواعد التصنيف لهذا النوع من البيانات.

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

دوال التمييز للقياسات المكررة قد درست لأول مره في عام (1972) من قبل العالم تشوي Choi حيث قام بتطوير نموذج التأثيرات المختلط لتصنيف بيانات قياسات مكررة أحادية المتغير بافتراض أن مشاهدات متعددة تؤخذ لنفس المتغير ولنفس الوحدة التجريبية في التجربة حيث أن هذه القياسات تحتفظ بنفس المتوسط على مر الزمن [2].

وفي السنوات الأخيرة كان هناك اهتمام متزايد في أساليب التحليل التمييزي لبيانات القياسات المكررة، ففي عام 2010 قام كل من (Pluta, Madry وأخرون) بمقارنة الطرق التقليدية مع الطرق الحديثة لأساليب التحليل التمييزي من خلال مجموعة بيانات حقيقية لتشخيص بعض أنواع النباتات، وفي العام نفسه قام كل من (Lix & Sajobi) بنفس العمل من خلال مجموعة بيانات حقيقية لتشخيص عدد من المرضى إلى مجموعتين مختلفتين من مرضى الصم الذين قاموا بزراعة قوقعة الإذن

(Cochlear Implant)، وتوصلوا هؤلاء إلى أن الطرق الحديثة للتحليل التمييزي بوجود تراكيب التباين المشترك هي ذات كفاءة أعلى في التصنيف من الطرق التقليدية من حيث نسبة التصنيف الخطأ الظاهرة الكلية [4],[9].

أن فكرة البحث هذا والمتضمنة تشخيص مرضى داء السكري وباستعمال طرائق تمييز للقياسات المكررة تعتبر مهمة لكونها وسيلة مساعدة تشخيصية للطبيب المختص، إضافة لكون مرض داء السكري من الأمراض 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 إضافة إلى أنه يساعد في إنقاذ حياة المريض ويجنبه الحالات المتطورة للمرض، ولكون أن التشخيص مسألة ذات جوانب معقدة لكون عدد من الأمراض تشترك في عدة أعراض يصعب التمييز فيما بينها كالشك بين الإصابة بمرض داء السكري من النمط الأول I والمرض السكري من النمط الثاني II ، لذلك فإن الأساليب الإحصائية ومنها التحليل التمييزي تساعد الطبيب المختص في التشخيص المبكر لأمراض أنماط السكري.

حيث أتسّم هدف البحث هذا ببناء أكثر من نموذج تمييزي وتصنيفي بوجود تركيبة التباين المشترك لأنموذج (AR – 1) 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 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 نوعين من أمراض السكري.

2. الجانب النظري

1.2 القياسات المكررة بوجود تركيبة التباين المشترك لأنموذج الانحدار الذاتي من الدرجة الأولى

Repeated Measures with AR-1 Covariance Structure

لتفسير مشكلة أسلوب التحليل التمييزي والتصنيفي لبيانات القياسات المكررة المستقاة طولياً 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 أبعاده $(p \times 1)$ يمثل القياسات المكررة على الوحدة التجريبية ولمتغير الاستجابة نفسه المأخوذة عبر p من النقاط الزمنية. وبافتراض أن y_{jit} يمثل القياس على الفرد it في المجموعه j th عند النقطة الزمنية t th، حيث أن $t = (j = 1, 2, 3, \dots, p)$ ، و $(i = 1, 2, 3, \dots, n_j)$.

بالتالي فإن $y_{ji} = (y_{ji1}, y_{ji2}, \dots, y_{jip})'$ هو متجه أبعاده $(p \times 1)$ للقياسات المكررة المشاهدة للفرد it المشارك في الدراسة في المجموعة j th. حيث أن لكلا المجموعتين j ، فإن y_{ji} يتبع التوزيع الطبيعي مع متوسط مساوي إلى μ_j ومصفوفة التباين والتباين المشترك أبعاده $(p \times p)$ المساوية إلى Ω_j أي أن:

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

وبوجود تركيبة التباينات لأنموذج (AR – 1) لمصفوفة التباين المشترك Ω_j ، وبالتالي لكل من $j = 1, 2$ فإن Ω_j تعطى بالصيغة الآتية [3],[6],[8]:

$$\Omega_j = \sigma_j^2 \begin{bmatrix} 1 & \rho_j & \rho_j^2 & \dots & \rho_j^{p-1} \\ \rho_j & 1 & \rho_j & \dots & \rho_j^{p-2} \\ \rho_j^2 & \rho_j & 1 & \dots & \rho_j^{p-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_j^{p-1} & \rho_j^{p-2} & \rho_j^{p-3} & \dots & 1 \end{bmatrix} \dots \dots (1)$$

وأن مقدار مُحدّدة (Determinant) المصفوفة Ω_j يعطى بالصيغة التالية:

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

وكذلك فإن معكوس (Inverse) المصفوفة Ω_j هو مصفوفة ثلاثية الأقطار (tridiagonal matrix) ويعطى بالصيغة التالية:

$$\Omega_j^{-1} = \frac{1}{\sigma_j^2 (1 - \rho_j)} \Omega_{0j} \dots \dots (3)$$

حيث أن Ω_{0j} مساوية إلى:

$$\Omega_{0j} = \begin{bmatrix} 1 & -\rho_j & 0 & \dots & 0 & 0 \\ -\rho_j & 1 + \rho_j^2 & -\rho_j & \dots & 0 & 0 \\ 0 & -\rho_j & 1 + \rho_j^2 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 + \rho_j^2 & -\rho_j \\ 0 & 0 & 0 & \dots & -\rho_j & 1 \end{bmatrix} \dots \dots (4)$$

وأن المصفوفة Ω_j في ظل هذا الافتراض هي دالة من معلمتين غير معلومة، أحدهما التباين σ_j^2 للملاحظات 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 (AR - 1) 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 (AR - 1) مع تساوي كل من مركبتي التباين والارتباط لكلا المجموعتين أي [9],[8],[7].

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

يعبر عن الدالة التمييزية (discriminant function) في هذه حالة (AR - 1) كتركيبة خطية من القياسات المكررة p والتي تعطى بالصيغة الآتية [4],[6],[7],[8].

$$D = \hat{a}' y_{ji} \dots (5)$$

حيث أن المتجه y_{ji} هو متجه أبعاده (1 × p) للقياسات المشاهدة للفرد i th المشارك في الدراسة في المجموعة j th ويعطى بالشكل الآتي :

$$y_{ji} = (y_{ji1}, y_{ji2}, \dots, y_{jip})'$$

أما \hat{a} هو متجه تقديرات معاملات الدالة التمييزية (Discriminant function coefficients) والذي يمكن تقديره بواسطة المعادلة التالية :

$$\hat{a} = \hat{\Omega}^{-1}(\hat{\mu}_1 - \hat{\mu}_2) \dots (6)$$

وبافتراض متجهات متوسط تركيبية منتظمة بدون تأثير عامل الوقت أي أن متوسطات القياسات المكررة المختلفة تبقى ثابتة على مر الزمن لذلك فإن :

$$\hat{\mu}_j = c_j \mathbf{1}_p, \quad j = 1, 2 \dots (7)$$

حيث c_j هو متوسط مناسبات القياسات المكررة للمجموعة j th والذي يتم تعويضه بمقدر الإمكان الأعظم (MLE) للحالة الأولى والذي سيأتي توضيحه. إما المصفوفة Ω في المعادلة (6) هي مصفوفة التباين المشترك لتركيبية نموذج الانحدار الذاتي من الدرجة الأولى (AR - 1) والمعطاة في المعادلة رقم (1).

وطبقاً للحالة هذه فإن قاعدة التمييز والتصنيف الخطية (linear classification rule) هي: أنه يتم تحصيل الفرد i th مع المشاهدات $(y_{ji1}, y_{ji2}, \dots, y_{jip})$ إلى المجموعه الأولى فيما إذا كان [4],[6],[7],[8].

$$\lambda(y_i) = [p\bar{y}_{1,p} - \hat{\rho}(p-2)\bar{y}_{2,p-1}] \geq \frac{1}{2}[p - \hat{\rho}(p-2)](\hat{c}_1 + \hat{c}_2) \dots (8)$$

وعدا ذلك فإنه يتم تحصيل الفرد i th مع الاستجابات $(y_{ji1}, y_{ji2}, \dots, y_{jip})$ إلى المجموعه الثانية. وواضح أن قاعدة التصنيف لهذه الحالة مستقلة عن مركبة التباين σ^2 ، حيث أن $\bar{y}_{l,m}$ يقدر بتطبيق الصيغة التالية :

$$\hat{y}_{l,m} = \frac{y_l + y_{l+1} + \dots + y_m}{m - l + 1}, \quad l, m = 1, 2, \dots, p \dots (9)$$

حيث أن y_l تمثل مقدار الاستجابة للفرد i th المشارك أو الجديد في الدراسة عند القياس المكرر أو النقطة الزمنية t th والمراد تمييزية أو تصنيفه على التوالي. و p هي عدد القياسات المكررة.

أما تقديرات الإمكان الأعظم لكل من c_1, c_2, ρ و σ^2 يتم الحصول عليها كما يل [6],[7].

بوجود تركيبية منتظمة لمتوسطات المجموعة، وليكن θ متجه من معلمات النموذج، 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 أبعاده (1 × p) من القياسات المكررة على الفرد i th المشارك في الدراسة حيث أن $(i = 1, 2, \dots, n_j; N = n_1 + n_2)$ في المجموعة j th 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، فإن لوغاريتم دالة الإمكان الأعظم المشتركة (log joint likelihood function) يعطى بالصيغة التالية :

$$\log L(\theta, \Omega ; Y_1, Y_2, \dots, Y_N)$$

$$= -\frac{Np}{2} \log 2\pi - \frac{N}{2} \log |\Omega| - \frac{1}{2} \text{tr} \left(\sum_{j=1}^2 \sum_{i=1}^{n_j} (y_{ij} - \mu_j)' \Omega^{-1} (y_{ij} - \mu_j) \right) \dots (10)$$

وبوجود تركيبة متوسط منتظمة لمتوسطات المجموعة فإن :

$$\mu_j = c_j \mathbf{1}_p \dots (11)$$

حيثما c_j هو المتوسط لمناسبات (حوادث) القياس (measurement occasions) للمجموعة j th و $\mathbf{1}_p$ متجه عمودي ذو أبعاد $(p \times 1)$ جميع عناصره مساوية إلى الواحد الصحيح ، وبالتالي يمكن إعادة كتابة المعادلة (10) بالشكل الآتي :

$$\log L(\theta, \Omega ; Y_1, Y_2)$$

$$= -\frac{Np}{2} \log 2\pi - \frac{N}{2} \log |\Omega| - \frac{1}{2} \text{tr} (\Omega^{-1} W) - \frac{1}{2} \left(\sum_{j=1}^2 n_j (\bar{y}_j - c_j \mathbf{1}_p)' \Omega^{-1} (\bar{y}_j - c_j \mathbf{1}_p) \right) \dots (12)$$

وبوجود تركيبة التباين لأنموذج $(AR - 1)$ على المصفوفة Ω_j نقوم بتعويض قيم كل من محددة (2) ومعكوس (3) المصفوفة Ω_j في معادلة (12) فإنه يعطي :

$$\log L(c_1, c_2, \Omega ; Y_1, Y_2)$$

$$= -\frac{Np}{2} \log 2\pi - \frac{Np}{2} \log \sigma^2 - \frac{N(p-1)}{2} \log(1 - \rho^2) - \frac{1}{2} \text{tr} \frac{1}{\sigma^2(1 - \rho^2)} \Omega_0 \times [W_0 - n_1 c_1 W_5 - n_2 c_2 W_6 + (n_1 c_1^2 + n_2 c_2^2) J_p] \dots (13)$$

حيث أن قيمة $(W_j, j = 0, 3, 4, 5, 6)$ موضحة في معادلة (22) .

الآن، وباستقاف المعادلة رقم (13) مرة واحدة بالنسبة لكل من c_1 و c_2 على التوالي ونضعها مساوية إلى الصفر وبعد عملية تبسيطها نحصل على المعادلتين أدناه :

$$(p-2)\hat{\rho}\hat{c}_1 - p\hat{c}_1 + pm_{11} - (p-2)\hat{\rho}m_{12} = 0, \dots (14)$$

$$(p-2)\hat{\rho}\hat{c}_2 - p\hat{c}_2 + pm_{21} - (p-2)\hat{\rho}m_{22} = 0, \dots (15)$$

في المعادلات أعلاه فإن كل من m_{21} و m_{11} تقدر من قبل المعادلتين الآتيتين :

$$\hat{m}_{11} = p^{-1}(1_p' \bar{y}_1) , \hat{m}_{21} = p^{-1}(1_p' \bar{y}_2) \dots (16)$$

في حين كل من m_{12} و m_{22} تعطى في المعادلتين الآتيتين :

$$\hat{m}_{12} = (p-2)^{-1}(1_p' \bar{y}_1 - \bar{y}_{11} - \bar{y}_{1p}), \hat{m}_{22} = (p-2)^{-1}(1_p' \bar{y}_2 - \bar{y}_{21} - \bar{y}_{2p}) \dots (17)$$

ويمثل \bar{y}_{1p} و \bar{y}_{11} العنصر الأول والأخير من متجه متوسطات المجموعة الأولى \bar{y}_1 . وبالمثل فإن \bar{y}_{2p} و \bar{y}_{21} هما العنصر الأول والأخير من متجه متوسطات المجموعة الثانية \bar{y}_2 .

وباقتاف المعادلة (13) مرة واحدة بالنسبة إلى σ^2 ومساواتها إلى الصفر وتبسيطها نجد أن :

$$\sigma^2 = \frac{1}{np(1 - \rho^2)} [\text{tr} \Omega_0 W_0 - n_1 c_1 \text{tr} \Omega_0 W_5 - n_2 c_2 \text{tr} \Omega_0 W_6 + (n_1 c_1^2 + n_2 c_2^2) \text{tr}(\Omega_0 J)] \dots (18)$$

$$\text{tr} \Omega_0 W_0 = \beta_1 \rho^2 - 2\gamma_1 \rho + \alpha_1 , \text{tr} \Omega_0 W_5 = \beta_2 \rho^2 - 2\gamma_2 \rho + \alpha_2 ,$$

$$\text{tr} \Omega_0 W_6 = \beta_3 \rho^2 - 2\gamma_3 \rho + \alpha_3$$

حيث أن القيم التقديرية لكل من $\gamma_1, \beta_1, \alpha_1$ و $\gamma_2, \beta_2, \alpha_2$ و $\gamma_3, \beta_3, \alpha_3$ يمكن إيجادها بتطبيق صيغة كل منها :

$$\hat{\alpha}_1 = \text{tr} W_0 , \hat{\beta}_1 = \text{tr} W_0 - w_{0,11} - w_{0,pp} , \hat{\gamma}_1 = \sum_{t=2}^p w_{0,t-1t} \quad t = 1, 2, \dots, p \dots (19)$$

$$\hat{\alpha}_2 = \text{tr}W_5, \hat{\beta}_2 = \text{tr}W_5 - w_{5,11} - w_{5,pp}, \quad \hat{\gamma}_2 = \sum_{t=2}^p w_{5,t-1t} \dots \dots (20)$$

$$\hat{\alpha}_3 = \text{tr}W_6, \hat{\beta}_3 = \text{tr}W_6 - w_{6,11} - w_{6,pp}, \quad \hat{\gamma}_3 = \sum_{t=2}^p W_{6,t-1t} \dots (21)$$

حيث أن الكمية $w_{ij,l}$ هنا تمثل العنصر $(i,j)^{th}$ من المصفوفة W_l ($l = 0, \dots, 6$) والتي يتم تقديرها بتطبيق الصيغ التالية :
 $W_0 = W_1 + W_2 + n_1 W_3 + n_2 W_4$, $W_3 = \bar{y}_1 \bar{y}_1'$, $W_4 = \bar{y}_2 \bar{y}_2'$, $W_5 = (1_p \bar{y}_1' + \bar{y}_1 1_p')$, $W_6 = (1_p \bar{y}_2' + \bar{y}_2 1_p')$ (22)

وفيها يمكن تقدير كل من المصفوفتين W_2 و W_1 بتطبيق :

$$\hat{W}_j = \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)(y_{ij} - \bar{y}_j)', \quad j = 1, 2 \dots (23)$$

وبالتالي وبتعويض القيم أعلاه وبعد عملية التبسيط فإن المعادلة (18) تختصر إلى :

$$N\hat{\sigma}^2 p(1 - \hat{\rho}^2) - (\beta_1 \hat{\rho}^2 - 2\gamma_1 \hat{\rho} + \alpha_1) + n_1 \hat{c}_1 (\beta_2 \hat{\rho}^2 - 2\gamma_2 \hat{\rho} + \alpha_2) + n_2 \hat{c}_2 (\beta_3 \hat{\rho}^2 - 2\gamma_3 \hat{\rho} + \alpha_3) - (n_1 \hat{c}_1^2 + n_2 \hat{c}_2^2) \{ (p-2)\hat{\rho}^2 - 2(p-1)\hat{\rho} + p \} = 0, \dots (24)$$

وباشتقاق المعادلة (13) مره أخرى بالنسبة إلى ρ وبمساواتها إلى الصفر نحصل على المعادلة التالية :

$$N\hat{\sigma}^2 \hat{\rho}(p-1) - N\hat{\sigma}^2 \hat{\rho}^3(p-1) - \{ \hat{\rho}(\alpha_1 + \beta_1) - \gamma_1 \hat{\rho}^2 - \gamma_1 \} + n_1 \hat{c}_1 \{ \hat{\rho}(\alpha_2 + \beta_2) - \gamma_2 \hat{\rho}^2 - \gamma_2 \} + n_2 \hat{c}_2 \{ \hat{\rho}(\alpha_3 + \beta_3) - \gamma_3 \hat{\rho}^2 - \gamma_3 \} - (n_1 \hat{c}_1^2 + n_2 \hat{c}_2^2) \{ \hat{\rho}(2p-1) - (p-1)\hat{\rho}^2 - (p-1) \} = 0, \dots (25)$$

بالتالي، ومن منظومة المعادلات (14)، (15)، (24) و (25) أعلاه وبحلها أنيا نجد تقديرات الإمكان الأعظم لكل من c_1 ، c_2 و ρ ، σ^2 .

2-2-2 قاعدة التمييز والتصنيف التربيعية للحالة الثانية

للحالة الثانية والتي تتمثل بعدم تجانس مصفوفتي التباين والتباين المشترك لتركيبية أنموذج (AR - 1) بتساوي مركبة التباين واختلاف معامل الارتباط لكلا المجموعتين أي :

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

فإن قاعدة التمييز والتصنيف التربيعية (Quadratic classification rule) هي : أنه يتم تخصيص الفرد i^{th} مع المشاهدات $(y_{ji1}, y_{ji2}, \dots, y_{jip})$ إلى المجموعه الأولى إذا كان [6] :

$$\begin{aligned} \lambda(y_i) = & -\frac{1}{2\sigma^2} \left(\frac{1}{1 - \rho_1^2} - \frac{1}{1 - \rho_2^2} \right) \sum_{t=1}^p y_{it}^2 - \frac{1}{2\sigma^2} \left(\frac{\rho_1^2}{1 - \rho_1^2} - \frac{\rho_2^2}{1 - \rho_2^2} \right) \sum_{t=2}^{p-1} y_{it}^2 \\ & + \frac{1}{\sigma^2} \left(\frac{\rho_1}{1 - \rho_1^2} - \frac{\rho_2}{1 - \rho_2^2} \right) \sum_{t=1}^{p-1} y_{it} y_{it+1} + \frac{p}{\sigma^2} \left(\frac{c_1}{1 + \rho_1} - \frac{c_2}{1 + \rho_2} \right) \bar{y}_{1,p} \\ & - \frac{p-1}{\sigma^2} \left(\frac{\rho_1 c_1}{1 + \rho_1} - \frac{\rho_2 c_2}{1 + \rho_2} \right) \bar{y}_{2,p-1} \\ & \geq \frac{1}{2\sigma^2} \left[\frac{c_1^2}{1 + \rho_1} (p - (p-2)\rho_1) - \frac{c_2^2}{1 + \rho_2} (p - (p-2)\rho_2) \right] \\ & - \frac{p-1}{2} \ln \left(\frac{1 - \rho_2^2}{1 - \rho_1^2} \right) \dots \dots (26) \end{aligned}$$

وعدا ذلك فأنه يتم تخصيص الفرد i^{th} مع الاستجابات $(y_{ji1}, y_{ji2}, \dots, y_{jip})$ إلى المجموعه الثانية. في قاعدة التصنيف أعلاه فأن تقديرات الإمكان الأعظم لكل من c_1 ، c_2 ، ρ_1 ، ρ_2 و σ^2 يتم الحصول عليها بواسطة الحل وبشكل إني للمعادلات الآتية [6].

$$(p-2)\hat{\rho}_1 \hat{c}_1 - p\hat{c}_1 + pm_{11} - (p-2)\hat{\rho}_1 m_{12} = 0, \dots (27)$$

$$(p-2)\hat{\rho}_2 \hat{c}_2 - p\hat{c}_2 + pm_{21} - (p-2)\hat{\rho}_2 m_{22} = 0, \dots (28)$$

$$\begin{aligned} & \hat{\sigma}^2 N^2 p^2 (1 - \hat{\rho}_1^2)(1 - \hat{\rho}_2^2) \\ & - Np(1 - \hat{\rho}_2^2)[(\beta_8 \hat{\rho}_1^2 - 2\gamma_8 \hat{\rho}_1 + \alpha_8) - n_1 \hat{c}_1 (\beta_2 \hat{\rho}_1^2 - 2\gamma_2 \hat{\rho}_1 + \alpha_2) \\ & + n_1 \hat{c}_1^2 ((p-2) \hat{\rho}_1^2 - 2(p-1) \hat{\rho}_1 + p)] \\ & - Np(1 - \hat{\rho}_1^2)[(\beta_9 \hat{\rho}_2^2 - 2\gamma_9 \hat{\rho}_2 + \alpha_9) - n_2 \hat{c}_2 (\beta_3 \hat{\rho}_2^2 - 2\gamma_3 \hat{\rho}_2 + \alpha_3) \\ & + n_2 \hat{c}_2^2 ((p-2) \hat{\rho}_2^2 - 2(p-1) \hat{\rho}_2 + p)] = 0, \quad \dots (29) \end{aligned}$$

$$\begin{aligned} & n_1 \hat{\sigma}^2 \hat{\rho}_1 (p-1) - n_1 \hat{\sigma}^2 \hat{\rho}_1^3 (p-1) - \{(\alpha_8 + \beta_8) \hat{\rho}_1 - \gamma_8 \hat{\rho}_1^2 - \gamma_8\} \\ & + n_1 \hat{c}_1 \{(\alpha_2 + \beta_2) \hat{\rho}_1 - \gamma_2 \hat{\rho}_1^2 - \gamma_2\} \\ & - n_1 \hat{c}_1^2 \{ \hat{\rho}_1 (2p-1) - (p-1) \hat{\rho}_1^2 - (p-1) \} = 0, \quad \dots (30) \end{aligned}$$

$$\begin{aligned} & n_2 \hat{\sigma}^2 \hat{\rho}_2 (p-1) - n_2 \hat{\sigma}^2 \hat{\rho}_2^3 (p-1) - \{(\alpha_9 + \beta_9) \hat{\rho}_2 - \gamma_9 \hat{\rho}_2^2 - \gamma_9\} \\ & + n_2 \hat{c}_2 \{(\alpha_3 + \beta_3) \hat{\rho}_2 - \gamma_3 \hat{\rho}_2^2 - \gamma_3\} \\ & - n_2 \hat{c}_2^2 \{ \hat{\rho}_2 (2p-1) - (p-1) \hat{\rho}_2^2 - (p-1) \} = 0, \quad \dots (31) \end{aligned}$$

وبقية القيم التقديرية يتم الحصول عليها بتطبيق الصيغ التالية :

$$\hat{\alpha}_8 = \text{tr}W_8, \quad \hat{\beta}_8 = \text{tr}W_8 - w_{8,11} - w_{8,pp}, \quad \hat{\gamma}_8 = \sum_{t=2}^p w_{8,t-1t} \quad \dots (32)$$

$$\hat{\alpha}_9 = \text{tr}W_9, \quad \hat{\beta}_9 = \text{tr}W_9 - w_{9,11} - w_{9,pp}, \quad \hat{\gamma}_9 = \sum_{t=2}^p w_{9t-1,t} \quad \dots (33)$$

وفيهما يمكن تقدير كل من المصفوفتين W_8 و W_9 بتطبيق الصيغة أدناه :

$$W_8 = W_1 + n_1 W_3, \quad W_9 = W_2 + n_2 W_4 \quad \dots \dots (34)$$

إما $\alpha_2, \beta_2, \gamma_2$ و $\alpha_3, \beta_3, \gamma_3$ وكل من المصفوفتين W_3, W_4 و W_1, W_2 فان صيغة كل منها تم تعريفها في الحالة السابقة.

3-2- قاعدة التمييز والتصنيف التربيعية للحالة الثالثة

للحالة الثالثة والموضحة أدناه :

Case 3 : $\Omega_1 \neq \Omega_2$ ($\sigma_1^2 \neq \sigma_2^2$, $\rho_1 = \rho_2 = \rho$)

فإن قاعدة التمييز والتصنيف التربيعية (Quadratic classification rule) هي : أنه يتم تخصيصة الفرد i th مع المشاهدات $(y_{ji1}, y_{ji2}, \dots, y_{jip})$ إلى المجموعه الأولى إذا كان [6]: [7]

$$\begin{aligned} \lambda(y_i) &= -\frac{1}{2(1-\rho^2)} \left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2} \right) \left[\sum_{t=1}^p y_{it}^2 + \rho^2 \sum_{t=2}^{p-1} y_{it}^2 - 2\rho \sum_{t=1}^{p-1} y_{it} y_{it+1} \right] \\ &+ \frac{1}{(1+\rho)} \left(\frac{c_1}{\sigma_1^2} - \frac{c_2}{\sigma_2^2} \right) [p\bar{y}_{1,p} - \rho(p-2)\bar{y}_{2,p-1}] \\ &\geq \frac{1}{2(1+\rho)} \left(\frac{c_1^2}{\sigma_1^2} - \frac{c_2^2}{\sigma_2^2} \right) (p - \rho(p-2)) - \frac{p}{2} \ln \frac{\sigma_2^2}{\sigma_1^2} \quad \dots (35) \end{aligned}$$

وعدا ذلك فأنه يتم تخصيصة الفرد i th مع الاستجابات $(y_{ji1}, y_{ji2}, \dots, y_{jip})$ إلى المجموعه الثانية. في قاعدة التصنيف أعلاه فأن تقديرات الإمكان الأعظم لكل من $\sigma_1^2, \rho, c_2, c_1$ يتم الحصول عليها بواسطة حل للمعادلات الآتية وبشكل إنبي [6] :

$$(p-2)\hat{\rho}\hat{c}_1 - p\hat{c}_1 + pm_{11} - (p-2)\hat{\rho}m_{12} = 0, \quad \dots \dots (36)$$

$$(p-2)\hat{\rho}\hat{c}_2 - p\hat{c}_2 + pm_{21} - (p-2)\hat{\rho}m_{22} = 0, \quad \dots \dots (37)$$

$$\begin{aligned} & n_1 \hat{\sigma}_1^2 p(1 - \hat{\rho}^2) - (\beta_8 \hat{\rho}^2 - 2\gamma_8 \hat{\rho} + \alpha_8) + n_1 \hat{c}_1 (\beta_2 \hat{\rho}^2 - 2\gamma_2 \hat{\rho} + \alpha_2) \\ & - n_1 \hat{c}_1^2 ((p-2) \hat{\rho}^2 - 2(p-1) \hat{\rho} + p) = 0, \quad \dots \dots (38) \end{aligned}$$

$$\begin{aligned} & n_2 \hat{\sigma}_2^2 p(1 - \hat{\rho}^2) - (\beta_9 \hat{\rho}^2 - 2\gamma_9 \hat{\rho} + \alpha_9) + n_2 \hat{c}_2 (\beta_3 \hat{\rho}^2 - 2\gamma_3 \hat{\rho} + \alpha_3) \\ & - n_2 \hat{c}_2^2 ((p-2) \hat{\rho}^2 - 2(p-1) \hat{\rho} + p) = 0, \quad \dots \dots (39) \end{aligned}$$

$$\begin{aligned}
 & N(p-1)(\hat{\rho} - \hat{\rho}^3)\hat{\sigma}_1^2\hat{\sigma}_2^2 - \hat{\sigma}_2^2\{(\alpha_8 + \beta_8)\hat{\rho} - \gamma_8\hat{\rho}^2 - \gamma_8\} \\
 & + n_1\hat{c}_1\hat{\sigma}_2^2\{(\alpha_2 + \beta_2)\hat{\rho} - \gamma_2\hat{\rho}^2 - \gamma_2\} \\
 & - n_1\hat{c}_1^2\hat{\sigma}_2^2\{(2p-2)\hat{\rho} - (p-1)\hat{\rho}^2 - (p-1)\} - \hat{\sigma}_1^2\{(\alpha_9 + \beta_9)\hat{\rho} - \gamma_9\hat{\rho}^2 - \gamma_9\} \\
 & + n_2\hat{c}_2\hat{\sigma}_1^2\{(\alpha_3 + \beta_3)\hat{\rho} - \gamma_3\hat{\rho}^2 - \gamma_3\} \\
 & - n_2\hat{c}_2^2\hat{\sigma}_1^2\{(2p-2)\hat{\rho} - (p-1)\hat{\rho}^2 - (p-1)\} = 0. \dots\dots (40)
 \end{aligned}$$

القيم التقديرية المطلوب إيجادها لهذه المعادلات وهي $\alpha_2, \beta_2, \gamma_2, \alpha_3, \beta_3, \gamma_3, \alpha_8, \beta_8, \gamma_8, \alpha_9, \beta_9, \gamma_9$ معطاة في الحالتين السابقتين.

4-2-2 قاعدة التمييز والتصنيف التربيعية للحالة الرابعة للحالة الرابعة :

Case 4 : $\Omega_1 \neq \Omega_2 (\sigma_1^2 \neq \sigma_2^2, \rho_1 \neq \rho_2)$

فإن قاعدة التمييز والتصنيف التربيعية (Quadratic classification rule) هي : أنه يتم تخصّيص الفرد i th مع المشاهدات $(y_{ji1}, y_{ji2}, \dots, y_{jip})$ إلى المجموعه الأولى إذا كان [7],[6]:

$$\begin{aligned}
 \lambda(y_i) &= -\frac{1}{2} \left(\frac{1}{\sigma_1^2(1-\rho_1^2)} - \frac{1}{\sigma_2^2(1-\rho_2^2)} \right) \sum_{t=1}^p y_{it}^2 \\
 & - \frac{1}{2} \left(\frac{\rho_1^2}{\sigma_1^2(1-\rho_1^2)} - \frac{\rho_2^2}{\sigma_2^2(1-\rho_2^2)} \right) \sum_{t=2}^{p-1} y_{it}^2 \\
 & + \left(\frac{\rho_1}{\sigma_1^2(1-\rho_1^2)} - \frac{\rho_2}{\sigma_2^2(1-\rho_2^2)} \right) \sum_{t=1}^{p-1} y_{it} y_{it+1} \\
 & + \left(\frac{c_1}{\sigma_1^2(1+\rho_1)} - \frac{c_2}{\sigma_2^2(1+\rho_2)} \right) p \bar{y}_{1,p} \\
 & - (p-2) \left(\frac{\rho_1 c_1}{\sigma_1^2(1+\rho_1)} - \frac{\rho_2 c_2}{\sigma_2^2(1+\rho_2)} \right) \bar{y}_{2,p-1} \\
 & \geq \frac{1}{2} \left(\frac{c_1^2(p-(p-2)\rho_1)}{\sigma_1^2(1+\rho_1)} - \frac{c_2^2(p-(p-2)\rho_2)}{\sigma_2^2(1+\rho_2)} \right) - \frac{p}{2} \ln \frac{\sigma_2^2}{\sigma_1^2} \\
 & - \frac{p-1}{2} \ln \left(\frac{1-\rho_2^2}{1-\rho_1^2} \right) \dots\dots (41)
 \end{aligned}$$

وعدا ذلك فأنه يتم تخصّيص الفرد i th مع الاستجابات $(y_{ji1}, y_{ji2}, \dots, y_{jip})$ إلى المجموعه الثانية. في قاعدة التصنيف أعلاه فإن تقديرات الإمكان الأعظم لكل من σ_1^2, ρ_1, c_1 و σ_2^2, ρ_2, c_2 يتم الحصول عليها بواسطة حل للمعادلات الآتية أنيا [7],[6]:

$$(p-2)\hat{\rho}_1\hat{c}_1 - p\hat{c}_1 + pm_{11} - (p-2)\hat{\rho}_1m_{12} = 0, \dots\dots (42)$$

$$(p-2)\hat{\rho}_2\hat{c}_2 - p\hat{c}_2 + pm_{21} - (p-2)\hat{\rho}_2m_{22} = 0, \dots\dots (43)$$

$$\begin{aligned}
 & n_1\hat{\sigma}_1^2p(1-\hat{\rho}_1^2) - (\beta_8\hat{\rho}_1^2 - 2\gamma_8\hat{\rho}_1 + \alpha_8) + n_1\hat{c}_1(\beta_2\hat{\rho}_1^2 - 2\gamma_2\hat{\rho}_1 + \alpha_2) \\
 & - n_1\hat{c}_1^2((p-2)\hat{\rho}_1^2 - 2(p-1)\hat{\rho}_1 + p) = 0, \dots\dots (44)
 \end{aligned}$$

$$\begin{aligned}
 & n_2\hat{\sigma}_2^2p(1-\hat{\rho}_2^2) - (\beta_9\hat{\rho}_2^2 - 2\gamma_9\hat{\rho}_2 + \alpha_9) + n_2\hat{c}_2(\beta_3\hat{\rho}_2^2 - 2\gamma_3\hat{\rho}_2 + \alpha_3) \\
 & - n_2\hat{c}_2^2((p-2)\hat{\rho}_2^2 - 2(p-1)\hat{\rho}_2 + p) = 0, \dots\dots (45)
 \end{aligned}$$

$$\begin{aligned}
 & n_1(p-1)(\hat{\rho}_1 - \hat{\rho}_1^3)\hat{\sigma}_1^2 - \{(\alpha_8 + \beta_8)\hat{\rho}_1 - \gamma_8\hat{\rho}_1^2 - \gamma_8\} \\
 & + n_1\hat{c}_1\{(\alpha_2 + \beta_2)\hat{\rho}_1 - \gamma_2\hat{\rho}_1^2 - \gamma_2\} \\
 & - n_1\hat{c}_1^2\{(2p-2)\hat{\rho}_1 - (p-1)\hat{\rho}_1^2 - (p-1)\} = 0, \dots\dots (46)
 \end{aligned}$$

$$\begin{aligned}
 & n_2(p-1)(\hat{\rho}_2 - \hat{\rho}_2^3)\hat{\sigma}_2^2 - \{(\alpha_9 + \beta_9)\hat{\rho}_2 - \gamma_9\hat{\rho}_2^2 - \gamma_8\} \\
 & + n_2\hat{c}_2\{(\alpha_3 + \beta_3)\hat{\rho}_2 - \gamma_3\hat{\rho}_2^2 - \gamma_3\} \\
 & - n_2\hat{c}_2^2\{(2p-2)\hat{\rho}_2 - (p-1)\hat{\rho}_2^2 - (p-1)\} = 0. \dots\dots (47)
 \end{aligned}$$

القيم التقديرية المطلوب إيجادها لحل هذه المعادلات وهي $\alpha_2, \beta_2, \gamma_2, \alpha_3, \beta_3, \gamma_3, \alpha_8, \beta_8, \gamma_8, \alpha_9, \beta_9, \gamma_9$ هي نفسها المعطاة في الحالتين الأولى والثانية.

3. تقدير نسب التصنيف الخاطئ Estimating Misclassification Rates

لتقييم فعالية وقدرة أساليب التصنيف للتنبؤ بعضوية المجموعة يجب أن تؤخذ بعين الاعتبار كمية أو قدر سوء التصنيف، وهناك نوعان من الأخطاء في التصنيف وهما [1]. [4]:

1.3 نسبة التصنيف الخطأ الظاهرة (Apparent Error Rate, APER): وهي عبارة عن نسبة عدد الأفراد اللذين تغير تصنيفها إلى عدد الأفراد الكلي في المجموعتين وتعطى بالصيغة الآتية:

$$APER = \frac{n_{12} + n_{21}}{n_1 + n_2} \quad \dots \dots (48)$$

2.3 نسبة التصنيف الصحيح الظاهرة (Apparent Correct rate, APCR): وهي عبارة عن نسبة عدد الأفراد التي ثبتت تصنيفها إلى عددهم الكلي في المجموعتين وتعطى بالصيغة الآتية:

$$APCR = \frac{n_{11} + n_{22}}{n_1 + n_2} \quad \dots \dots (49)$$

4. الجانب التطبيقي

1.4 مرض داء السكري، إعراضه السريرية، أنواعه، العوامل المسببة للمرض وطرق تشخيصها:

داء السكري (Diabetes mellitus) هو متلازمة تتصف باضطراب الأستقلاب وارتفاع شاذ في تركيز سكر الدم عن المعدل الطبيعي الذي يتراوح بين (80 – 100مليجرام/ديسليتر) 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 والذي يعتبر الهرمون الأساسي الذي ينظم نقل الغلوكوز من الدم إلى معظم خلايا الجسم، خصوصاً 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 الرئيسي لهذه الخسارة هو مناعة ذاتية تتميز بهجوم الخلايا تاء المناعية على خلايا بيتا المنتجة للأنسولين، أي أن مرضى هذا النوع من السكري هم مرضى سكري 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 لمفعول الأنسولين بالإضافة إلى 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 (ضعيفي السيطرة السكرية).

وهناك عدة أعراض توحى 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، وتغييم الرؤية.

أما تشخيص النمط الأول والعديد من حالات النمط الثاني من السكري فيتم بناء على الأعراض الأولية التي تظهر في بداية المرض مثل كثرة التبول والعطش الزائد وقد يصاحبها فقد للوزن، ويتم عادة تشخيص بقية أنماط السكري بطرق أخرى مثل اكتشاف ارتفاع مستوى غلوكوز الدم أثناء إجراء 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 صائماً 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.

2.4 وصف العينة ومتغير 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) مريض يعاني من داء السكري من النمط الأول (I) (مرضى سكري معتمدون في علاجهم على الأنسولين)، في حين كانت المجموعة الثانية تمثل منهم (33) مريض يعاني من داء السكري من النمط الثاني (II) (مرضى سكري ضعيفي السيطرة السكرية معتمدون في علاجهم على الحبوب).

حيث قمنا بتشخيص نوعيين من أمراض السكر بالاعتماد على قواعد التمييز والتصنيف بوجود تركيبة التباين المشترك ($AR - 1$) 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 الخاطي واعتماد قاعدة التصنيف التي تعطي أقل نسبة خطأ تصنيف ظاهرة كأسلوب مساعد في عملية التشخيص لمرضى المجموعتين.

3.4 التحليل التمييزي والتصنيفي لبيانات القياسات المكررة في حالة تركيبة التباين المشترك لأنموذج ($AR - 1$)
توضح الأقسام التالية الدالة التمييزية والتصنيفية ونتائجها بوجود تركيبة التباين المشترك لأنموذج ($AR - 1$) ولكل حالة من الحالات الأربعة التي تم إيضاحها مسبقاً :

1.3.4 الحالة الأولى

بافتراض وجود الحالة الأولى :

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 أولاً تقدير كل من \hat{c}_1 و \hat{c}_2 وذلك لغرض تخمين القيم التقديرية $\hat{\mu}_1$ و $\hat{\mu}_2$ للمجموعتين الأولى والثانية، وكذلك تقدير كل من $\hat{\rho}$ و $\hat{\sigma}^2$ وبالتالي تقدير مصفوفة التباينات ($AR - 1$) والمطلوب تقديرها لبناء الدالة التمييزية لهذه الحالة، ولغرض الحصول على هذه التقديرات نجد بالترتيب القيم التقديرية الآتية والتي تم تفسيرها في الحالة الأولى لكل من m_{11} و m_{21} من (16) :

$$\hat{m}_{11} = 9.585, \quad \hat{m}_{21} = 6.072.$$

ثم نجد أيضاً تقديري كل من m_{12} و m_{22} باعتماد (17) :

$$\hat{m}_{12} = 9.474, \quad \hat{m}_{22} = 6.091.$$

وأيضاً نجد تقدير كل من $\hat{\alpha}_1$ ، $\hat{\beta}_1$ ، $\hat{\gamma}_1$ بالاستناد إلى صيغة كل منهم من (19)، وينبغي أولاً إيجاد كل من W_1 و W_2 ، W_3 و W_4 وبتطبيق صيغة رقم (23) تكون مساوية إلى :

$$W_1 = \begin{bmatrix} 58.830 & 49.247 & 34.188 \\ & 88.652 & 62.937 \\ & & 131.757 \end{bmatrix}, \quad W_2 = \begin{bmatrix} 43.296 & 34.880 & 34.272 \\ & 38.880 & 34.592 \\ & & 53.504 \end{bmatrix}$$

وبتطبيق معادلة (22) نجد :

$$W_3 = \begin{bmatrix} 69.5890 & 79.0321 & 91.2531 \\ & 89.7567 & 103.6361 \\ & & 119.6617 \end{bmatrix}, \quad W_4 = \begin{bmatrix} 32.6269 & 34.7918 & 36.6253 \\ & 37.1003 & 39.0555 \\ & & 41.1137 \end{bmatrix}$$

وبتعويض كل من W_1 ، W_2 و W_3 ، W_4 في W_0 ومن (22) نجد تقدير المصفوفة W_0 مساوية إلى :

$$W_0 = \begin{bmatrix} 3838.4557 & 4235.4762 & 4744.7127 \\ & 4762.5965 & 5324.5323 \\ & & 4755.5193 \end{bmatrix}$$

بالتالي فإن تقدير $\hat{\alpha}_1$ ، $\hat{\beta}_1$ و $\hat{\gamma}_1$ يكون مساوي إلى :

$$\hat{\alpha}_1 = 13356.5715, \quad \hat{\beta}_1 = 4762.5965, \quad \hat{\gamma}_1 = 9560.0085$$

وبتطبيق الصيغة (22) نجد تقدير المصفوفة W_5 لغرض تقدير كل من $\hat{\alpha}_2$ ، $\hat{\beta}_2$ ، $\hat{\gamma}_2$ بالاستناد إلى صيغة كل منهم في (20) :

$$W_5 = \begin{bmatrix} 16.684 & 17.816 & 19.281 \\ & 18.948 & 20.413 \\ & & 21.878 \end{bmatrix}$$

وبالتالي ومن التقدير أعلاه نجد $\hat{\alpha}_2$ ، $\hat{\beta}_2$ و $\hat{\gamma}_2$:

$$\hat{\alpha}_2 = 57.51, \quad \hat{\beta}_2 = 18.948, \quad \hat{\gamma}_2 = 38.229$$

ومن صيغة (22) نجد تقدير المصفوفة W_6 لغرض تقدير كل من $\hat{\alpha}_3$ ، $\hat{\beta}_3$ ، $\hat{\gamma}_3$ بالاستناد إلى صيغة كل منهم في (21):

$$W_6 = \begin{bmatrix} 11.424 & 11.803 & 12.124 \\ & 12.182 & 12.503 \\ & & 12.824 \end{bmatrix}$$

وبالتالي ومن التقدير أعلاه نجد أن:

$$\hat{\alpha}_3 = 36.43, \quad \hat{\beta}_3 = 12.182, \quad \hat{\gamma}_3 = 24.306$$

وبعد إيجاد القيم التقديرية α_1 ، β_1 ، γ_1 ، α_2 ، β_2 ، γ_2 و α_3 ، β_3 ، γ_3 يتم تعويض كل منهما في (14)، (15)، (24) و (25) لغرض الحصول على تقديرات الإمكان الأعظم \hat{c}_1 ، \hat{c}_2 ، $\hat{\rho}$ و $\hat{\sigma}^2$ المطلوب إيجادها لبناء الدالة التمييزية للحالة الأولى للتحليل التمييزي بوجود تركيبة (AR - 1). وباستعمال برنامج (MATLAB, V 7.6.0) لحل أنظمة المعادلات هذه، فإن المتجه أدناه يمثل تقديرات الإمكان الأعظم التي تم التوصل إليها بحل تلك المعادلات:

$$\hat{\theta} = \begin{bmatrix} \hat{c}_1 \\ \hat{c}_2 \\ \hat{\rho} \\ \hat{\sigma}^2 \end{bmatrix} = \begin{bmatrix} 8.4323 \\ 5.7443 \\ 0.8201 \\ 9.9383 \end{bmatrix}$$

ومن خلال \hat{c}_1 و \hat{c}_2 نجد تقدير متجه متوسطات القياسات المكررة المنتظمة للمجموعة الأولى $\hat{\mu}_1$ والثانية بتطبيق الصيغة (7) وكما يلي:

$$\hat{\mu}_1 = 8.4323 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 8.4323 \\ 8.4323 \\ 8.4323 \end{bmatrix}, \quad \hat{\mu}_2 = \begin{bmatrix} 5.7443 \\ 5.7443 \\ 5.7443 \end{bmatrix}$$

وبعدها نجد تقدير مصفوفة التباين والتباين المشترك لأنموذج (AR - 1) والمعطاة في معادلة رقم (1) من خلال تعويض قيمة كل من $\hat{\sigma}^2$ و $\hat{\rho}$ في الصيغة الآتية [3]:

$$\sigma_{tt'} = 9.9383 \times 0.8201^{|t-t'|}$$

وبالتالي فإن تقدير مصفوفة التباين المشترك (AR - 1) لهذه الحالة تكون مساوية إلى:

$$\hat{\Omega}_{AR-1} = \begin{bmatrix} 9.9383 & 8.1504 & 6.6841 \\ 8.1504 & 9.9383 & 8.1504 \\ 6.6841 & 8.1504 & 9.9383 \end{bmatrix}$$

وبالتالي بعد إيجاد القيم التقديرية لكل من $\hat{\mu}_1$ ، $\hat{\mu}_2$ و $\hat{\Omega}_{AR-1}$ فإنه من الممكن لإيجاد متجه تقديرات معاملات الدالة التمييزية \hat{a} في حالة تركيبة التباين المشترك لأنموذج (AR - 1) والمعطى في معادلة (6) وكالاتي:

$$\hat{a} = \begin{bmatrix} 9.9383 & 8.1504 & 6.6841 \\ 8.1504 & 9.9383 & 8.1504 \\ 6.6841 & 8.1504 & 9.9383 \end{bmatrix}^{-1} \begin{bmatrix} 2.688 \\ 2.688 \\ 2.688 \end{bmatrix} = [0.1486 \quad 0.0267 \quad 0.1486]'$$

وبهذا فإن تقدير الدالة التمييزية في حالة تركيبة التباين المشترك هذه والمعطاة في المعادلة رقم (5) تكون مساوية إلى:

$$D = 0.1486 y_{ji1} + 0.0267 y_{ji2} + 0.1486 y_{ji3}$$

الآن وبعد تقدير الدالة التمييزية ، وبعد تعويض كل من \hat{c}_1 ، \hat{c}_2 و $\hat{\rho}$ في قاعدة التمييز والتصنيف $\lambda(y_i)$ المعطاة في معادلة (8) فإنه يتم تَحْصِيص الفرد i th مع المشاهدات ($y_{ji1}, y_{ji2}, \dots, y_{ji3}$) إلى المجموعه الأولى فيما إذا كان بالصيغة الآتية:

$$\lambda(y_i) = 3\bar{y}_{1,3} - 0.8201\bar{y}_{2,2} \geq 15.4518$$

وعدا ذلك يتم تَحْصِيصَه إلى المجموعة الثانية.

وبعد بناء قاعدة التمييز والتصنيف الخطية ذات تركيبة التباينات (AR - 1) لهذه الحالة، فإن الجدول التالي يعطي كِل من عدد المشاهدات في المجموعتين الأولى والثانية مع الإعداد الصحيحة والخاطئة التي تم التنبؤ بها إلى كلا المجموعتين:

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

From Group	I	II	Total
I	38 100.00	0 00.00	38 100.00
II	4 12.12	29 87.88	33 100.00
Total	42 59.15	29 40.85	71 100.00

ومن خلال الجدول أعلاه نلاحظ أن في حالة التحليل التمييزي الخطي لبيانات القياسات المكررة مع بنية التباين المشترك (1 - AR)، وبالنسبة إلى المجموعة الأولى فإن جميع الأفراد والذين عددهم 38 أي بنسبة (100.00%) صنفوا بشكل صحيح إلى مجموعتهم. وبالمثل؛ بالنسبة للمجموعة الثانية فإنه من أصل (33) من الأفراد هناك 29 فرد أي بنسبة (87.88%) صنفوا بشكل صحيح إلى مجموعتهم. إما نسبة التصنيف الخطأ الظاهرة الكلية يمكن إيجادها وباعتماد معادلة (48)، وكما يلي:

$$APER = \frac{4 + 0}{33 + 38} = 0.0563$$

أي أن العدد الكلي للأفراد المصنفين بشكل غير صائب هو (4) أي ما يعادل (5.63%) لحجم العينة الكلي من 71 فرد مشارك في الدراسة. وأن نسبة التصنيف الصحيح الظاهرة الكلية وبتطبيق (49) أيضا:

$$APCR = \frac{38 + 29}{33 + 38} = 0.9437$$

أي أن العدد الكلي للأفراد المصنفين بشكل صحيح هو (67) أي ما يعادل (94.37%) لحجم العينة الكلي من 71 فرد مشارك في الدراسة.

2.3.4 الحالة الثانية

بافتراض وجود الحالة الثانية:

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

ولغرض بناء قاعدة التمييز والتصنيف للفرد المشارك أو الجديد في الدراسة والمراد تمييزه أو تصنيفه، فيبغى أولاً إيجاد القيم التقديرية لكل من $\hat{\alpha}_8$ ، $\hat{\beta}_8$ ، $\hat{\gamma}_8$ بالاستناد إلى صيغة (32)، وينبغي أولاً تقدير للمصفوفة W_8 وبتطبيق صيغتها (34) تكون مساوية إلى:

$$W_8 = \begin{bmatrix} 2703.212 & 3052.4668 & 3501.8058 \\ & 3499.4066 & 4001.1088 \\ & & 4678.9016 \end{bmatrix}$$

حيث أن كل من W_1 و W_3 قد سبق تقديرهما، وبالتالي فإن:

$$\hat{\alpha}_8 = 10881.5202, \quad \hat{\beta}_8 = 3499.4066, \quad \hat{\gamma}_8 = 7053.5756$$

ونجد أيضا القيم التقديرية لكل من $\hat{\alpha}_9$ ، $\hat{\beta}_9$ ، $\hat{\gamma}_9$ بالاستناد إلى صيغة (33) وكما يلي، حيث نجد أولاً القيمة التقديرية للمصفوفة W_9 وبتطبيق صيغتها (34) تكون مساوية إلى:

$$W_9 = \begin{bmatrix} 1119.9837 & 1183.0094 & 1242.9069 \\ & 1263.1899 & 1323.4235 \\ & & 1392.2561 \end{bmatrix}$$

حيث أن كل من W_2 و W_4 قد سبق تقديرهما، وبالتالي فإن:

$$\hat{\alpha}_9 = 3775.4297, \quad \hat{\beta}_9 = 1263.1899, \quad \hat{\gamma}_9 = 2506.4329$$

وبعد إيجاد القيم التقديرية α_8 ، β_8 ، γ_8 و α_9 ، β_9 ، γ_9 يتم تعويض كل منهما في (27)، (28)، (29)، (30) و (31) لغرض الحصول على تقديرات الإمكان الأعظم \hat{c}_1 ، \hat{c}_2 ، \hat{c}_1 ، \hat{c}_2 ، \hat{c}_1 و $\hat{\sigma}^2$ المطلوب إيجادها لبناء قاعدة التمييز والتصنيف للحالة الثانية للتحليل التمييزي بوجود تركيبة (1 - AR)، ويمثل المتجه أدناه تقديرات الإمكان الأعظم للحالة الثانية التي تم التوصل إليها بحل تلك المعادلات:

$$\hat{\theta} = \begin{bmatrix} \hat{c}_1 \\ \hat{c}_2 \\ \hat{\rho}_1 \\ \hat{\rho}_2 \\ \hat{\sigma}^2 \end{bmatrix} = \begin{bmatrix} 17.7370 \\ 15.5067 \\ 0.0508 \\ 0.0621 \\ 9.5352 \end{bmatrix}$$

وبعد تعويض كل من \hat{c}_1 ، \hat{c}_2 ، $\hat{\rho}_1$ ، $\hat{\rho}_2$ و $\hat{\sigma}^2$ في قاعدة التمييز والتصنيف $\lambda(y_i)$ المعطاة في معادلة (26) فإنه يتم تخصّيص الفرد i th مع المشاهدات $(y_{ji1}, y_{ji2}, \dots, y_{jip})$ إلى المجموعه الأولى فيما إذا كان $\lambda(y_i)$ بالصيغة الآتية:

$$\lambda(y_i) \geq 11.4238$$

وبعد ذلك يتم تخصّيصه إلى المجموعة الثانية. وبالتالي وبالاستناد إلى قاعدة التمييز والتصنيف ذات تركيبة التباينات (AR – 1) للحالة الثانية فأن الجدول التالي يعطي كل من عدد المشاهدات في المجموعتين الأولى والثانية مع الإعداد الصحيحة والخاطئة التي تم التنبؤ بها إلى كلا المجموعتين :

الجدول (2): نتائج التحليل التمييزي لبيانات القياسات المكررة بوجود تركيبة التباين المشترك (AR – 1) (جدول النسب المئوية للتصنيف في كل مجموعة للحالة الثانية)

From Group	I	II	Total
I	36 94.74	2 5.26	38 100.00
II	5 15.15	28 84.85	33 100.00
Total	41 57.75	30 42.25	71 100.00

وبالتالي وبالاستناد إلى قاعدة التصنيف هذه فأن نسبة التصنيف الخطأ الظاهرة الكلية، ونسبة التصنيف الصحيح الظاهرة الكلية أيضا وبتطبيق الصيغتين (48) و (49) على التوالي قدرت إلى :

$$APER = 0.0986$$

$$APCR = 0.9014$$

واضح من خلال $APER$ و $APCR$ على التوالي، أن العدد الكلي للأفراد المصنفين بشكل غير صائب هو (7) أي ما يعادل (9.86%) أما الذين تم تصنيفهم بشكل صائب فكان عددهم (64) فرد إي ما يعادل (90.14%) لحجم العينة الكلي من 71 فرد مشارك في الدراسة.

3.3.4 الحالة الثالثة

بافتراض وجود الحالة الثالثة :

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

للحصول على نتائج الدالة التمييزية في هذه الحالة نقوم بتعويض القيم التقديرية والتي تم تقديرها في الحالات السابقة لكل من:

$$\hat{\alpha}_8 = 10881.5202, \quad \hat{\beta}_8 = 3499.4066, \quad \hat{\gamma}_8 = 7053.5756,$$

$$\hat{\alpha}_2 = 57.51, \quad \hat{\beta}_2 = 18.948, \quad \hat{\gamma}_2 = 38.229$$

$$\hat{\alpha}_9 = 3775.4297, \quad \hat{\beta}_9 = 1263.1899, \quad \hat{\gamma}_9 = 2506.4329,$$

$$\hat{\alpha}_3 = 36.43, \quad \hat{\beta}_3 = 12.182, \quad \hat{\gamma}_3 = 24.306$$

في المعادلات (36)، (37)، (38)، (39) و (40) والمعطاة في الحالة الثالثة لغرض الحصول على تقديرات الإمكان الأعظم للمعالم \hat{c}_1 ، \hat{c}_2 ، $\hat{\sigma}_1^2$ ، $\hat{\sigma}_2^2$ و $\hat{\rho}$ المطلوب إيجادها لبناء قاعدة التمييز والتصنيف التريبيعية للحالة الثالثة للتحليل التمييزي بوجود تركيبة (AR – 1)، وبحل تلك المعادلات فأن المتجه أدناه يعطي تقديرات الإمكان الأعظم للحالة الثالثة :

$$\hat{\theta} = \begin{bmatrix} \hat{c}_1 \\ \hat{c}_2 \\ \hat{\rho} \\ \hat{\sigma}_1^2 \\ \hat{\sigma}_2^2 \end{bmatrix} = \begin{bmatrix} 10.6033 \\ 7.0264 \\ 0.0712 \\ 3.4224 \\ 2.1991 \end{bmatrix}$$

وبعد تعويض كل من \hat{c}_1 ، \hat{c}_2 ، $\hat{\sigma}_1^2$ ، $\hat{\sigma}_2^2$ ، $\hat{\rho}$ في قاعدة التمييز والتصنيف $\lambda(y_i)$ للحالة الثالثة المعطاة في معادلة (35) فإنه يتم تخصّيص الفرد i th مع المشاهدات $(y_{ji1}, y_{ji2}, \dots, y_{jip})$ إلى المجموعه الأولى فيما إذا كان $\lambda(y_i)$ بالصيغة الآتية :

$$\lambda(y_i) \geq 14.8823$$

وعدا ذلك يتم تخصّيصه إلى المجموعة الثانية. وبالتالي وبالاستناد إلى قاعدة التمييز والتصنيف ذات تركيبة التباينات $(AR - 1)$ للحالة الثالثة فإن الجدول الآتي يعطي كل من عدد المشاهدات في المجموعتين الأولى والثانية مع الإعداد الصحيحة والخاطئة التي تم التنبؤ بها إلى كلا المجموعتين :

الجدول (3): نتائج التحليل التمييزي لبيانات القياسات المكررة بوجود تركيبة التباين المشترك $(AR - 1)$ (جدول النسب المئوية للتصنيف في كل مجموعة للحالة الثالثة)

From Group	I	II	Total
I	33 86.84	5 13.16	38 100.00
II	4 12.12	29 87.88	33 100.00
Total	37 52.11	34 47.89	71 100.00

وبالتالي وبالاستناد إلى قاعدة التصنيف هذه فإن نسبة التصنيف الخطأ الظاهرة الكلية، ونسبة التصنيف الصحيح الظاهرة الكلية أيضا وبتطبيق الصيغتين (48) و (49) على التوالي قدرت إلى :

$$APER = 0.1268$$

$$APCR = 0.8732$$

واضح من خلال $APER$ و $APCR$ أن العدد الكلي للإفراد المصنفين بشكل غير صائب هو (9) أشخاص أي ما يعادل (12.68%) أما الذين تم تصنيفهم بشكل صائب فكان عددهم (62) فرد أي ما يعادل (87.32%) لحجم العينة الكلي من 71 فرد مشارك في الدراسة.

4.3.4 الحالة الرابعة

وبافتراض وجود الحالة الرابعة :

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

ولبناء الدالة التمييزية في هذه الحالة ، نقوم بتعويض القيم التقديرية والتي تم تقديرها في الحالتين الأولى والثانية لكل من:

$$\hat{\alpha}_8 = 10881.5202, \hat{\beta}_8 = 3499.4066, \hat{\gamma}_8 = 7053.5756, \hat{\alpha}_2 = 57.51$$

$$\hat{\beta}_2 = 18.948, \hat{\gamma}_2 = 38.229$$

$$\hat{\alpha}_9 = 3775.4297, \hat{\beta}_9 = 1263.1899, \hat{\gamma}_9 = 2506.4329, \hat{\alpha}_3 = 36.43$$

$$\hat{\beta}_3 = 12.182, \hat{\gamma}_3 = 24.306$$

في المعادلات (42)،(43)،(44)،(45)،(46) و (47) المعطاة في الحالة الرابعة لغرض الحصول على تقديرات الإمكان الأعظم للمعالم \hat{c}_1 ، $\hat{\rho}_1$ ، $\hat{\sigma}_1^2$ و \hat{c}_2 ، $\hat{\rho}_2$ ، $\hat{\sigma}_2^2$ المطلوب إيجادها لبناء قاعدة التمييز والتصنيف التربيعية للحالة الأخيرة للتحليل التمييزي بوجود تركيبة $(AR - 1)$ ، ويمثل المتجه أدناه تقديرات الإمكان الأعظم للحالة الرابعة التي تم التوصل إليها بحل تلك المعادلات:

$$\hat{\theta} = \begin{bmatrix} \hat{c}_1 \\ \hat{\rho}_1 \\ \hat{\sigma}_1^2 \\ \hat{c}_2 \\ \hat{\rho}_2 \\ \hat{\sigma}_2^2 \end{bmatrix} = \begin{bmatrix} 17.2512 \\ 0.1640 \\ 6.5306 \\ 19.9956 \\ 0.7227 \\ 7.4854 \end{bmatrix}$$

وبعد تعويض كل من \hat{c}_1 ، $\hat{\rho}_1$ ، $\hat{\sigma}_1^2$ و \hat{c}_2 ، $\hat{\rho}_2$ ، $\hat{\sigma}_2^2$ في قاعدة التمييز والتصنيف $\lambda(y_i)$ للحالة الرابعة المعطاة في معادلة (41) فإنه يتم تخصّيص الفرد i th مع المشاهدات $(y_{ji1}, y_{ji2}, \dots, y_{jip})$ إلى المجموعه الأولى فيما إذا كان $\lambda(y_i)$ بالصيغة الآتية :

$$\lambda(y_i) \geq 20.7168$$

وبعد ذلك يتم تخصّيصه إلى المجموعة الثانية. وبالتالي وبالاستناد إلى قاعدة التمييز والتصنيف ذات تركيبة التباينات (AR – 1) للحالة هذه فإن الجدول الآتي يعطي كل من عدد المشاهدات في المجموعتين الأولى والثانية مع الإعداد الصحيحة والخاطئة التي تم التنبؤ بها إلى كلا المجموعتين :

الجدول (4): نتائج التحليل التمييزي لبيانات القياسات المكررة بوجود تركيبة التباين المشترك (AR – 1) (جدول النسب المئوية للتصنيف في كل مجموعة للحالة الرابعة)

From Group	I	II	Total
I	35 92.11	3 7.89	38 100.00
II	6 18.18	27 81.82	33 100.00
Total	41 57.75	30 42.25	71 100.00

وبالتالي وبالاستناد إلى قاعدة التصنيف هذه فإن نسبة التصنيف الخاطئ الظاهرة الكلية، ونسبة التصنيف الصحيح الظاهرة الكلية أيضا وبتطبيق الصيغتين (48) و (49) على التوالي قدرت إلى :

$$APER = 0.1268, \quad APCR = 0.8732$$

واضح من خلال APCR و APER أن العدد الكلي للإفراد المصنفين بشكل غير صائب هو (9) أشخاص أي ما يعادل (12.68%) أما الذين تم تصنيفهم بشكل صائب فكان عددهم (62) فرد أي ما يعادل (87.32%) لحجم العينة الكلي من 71 فرد مشارك في الدراسة.

5-1 الاستنتاجات Conclusions

- من خلال ما تم عرضه في بحثنا هذا ولمجموعة بيانات حقيقية تم التوصل إلى الاستنتاجات الآتية :
- 1- ظهر إن أسلوب التحليل التمييزي الخطي (LDA) بوجود تركيبة التباين المشترك لأنموذج الانحدار الذاتي من المرتبة الأولى (AR – 1) بأنه الأكثر ملائمة في تصنيف مرضى داء السكري ويعود السبب في ذلك إلى امتلاكه اقل نسبة من نسب التصنيف الخاطئ الظاهرة الكلية والتي بلغت فيه (APER = 5.63%).
 - 2- تبين كذلك كلما زادت عدد المعلمات المطلوب تقديرها لبناء قاعدة التمييز والتصنيف فإن نسبة التصنيف الخاطئ الظاهرة الكلية (APER) تبدأ بالزيادة وهذا ما يقلل من كفاءة قواعد التصنيف لهذا النوع من البيانات.

المصادر
References

أولاً : المصادر العربية

- 1- الراوي، عمر فوزي و دبذوب، 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 - 203 (220).

ثانياً : 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- Rencher, A.C. (2002). **Methods of multivariate analysis**, Second Edition. New York :John Wiley & Sons, Inc.
- 6- 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.
- 7- 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.
- 8- 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.
- 9- Wolynski,W , Krzysko, M., Madry, W., Pluta, S., and Skorzybut, M.,(2010). **Analysis of multivariate repeated measures data**. Colloquium Biometricum,40,117-133.