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 مستل

المستخلص

يعرف التحليل التمييزي بأنه أسلوب إحصائي يعتمد على عينه من المفردات المأخوذة من مجتمعات معلومة, وذلك لبناء قاعدة يمكنها المساعدة مستقبلا 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 طوليا عبر الزمن على متغير الاستجابة نَفسُهُ ولكل وحدة تجريبية.

من الباحثين الذين تناولوا هذه المشكلة (**Roy** & **Khattree**, **2005**) فقد قدما 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. 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 ($\mathbf{AR-1}$) 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 سنعْتَبَرَ 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 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 بوجود تركيبة التباين المشترك ($\mathbf{AR-1}$) 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. المقدمـــة 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;فأنها تفترض 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 مناسبات القياسات المكررة, وبالإضافة إلى ذلك فأنها تعتمد على تقدير عدد كبير من المعلمات المطلوب تقديرها لبناء الأنموذج التمييزي والتي يزداد عددها بزيادة عدد القياسات المكررة, وعند عدم تساوي مصفوفة التباينات للمجموعات فأن استعمال أسلوب التحليل التمييزي ألتربيعي 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 للأنموذج 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 للمشترك (**Covariance Structures**) على تقليص عدد المعلمات المطلوبة لبناء الأنموذج 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 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 إضافة إلى أنه يساعد في إنقاذ حياة المريض ويجنبه الحالات المتطورة للمرض, ولكون أن التشخيص مسألة ذات جوانب معقدة لكون 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 أُتسِمّ **هدف** البحث هذا ببناء أكثر من أنموذج تمييزي وتصنيفي بوجود تركيبة التباين المشترك لأنموذج $(\mathbf{AR-1})$ 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 نوعيين 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

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 $\boldsymbol{y}$ هو متجه عمودي أبعادهُ $(\mathbf{p\times 1})$ يمثل القياسات المكررة على الوحدة التجريبية ولمتغير الاستجابة نَفسُهُ المأخوذة عبر $\mathbf{p}$ من النقاط الزمنية. وبافتراض أن $\mathbf{y_{jit}}$ يمثل القياس على الفرد $\boldsymbol{i}\mathbf{th}$ في ألمجموعه $\boldsymbol{j}\mathbf{th}$ عند النقطة الزمنية $\boldsymbol{t}\mathbf{th}$, حيث أن $(\mathbf{t} = \mathbf{1,2,3,\ldots,p})$,$(\mathbf{j=1,2})$, و $(\mathbf{i=1,2,3,\ldots,n_j})$.

بالتالي فأن $\mathbf{y_{ji}} = \left(\mathbf{y_{ji1}, y_{ji2}, \ldots, y_{jip}}\right)'$ هو متجه أبعادهُ $(\mathbf{p\times 1})$ للقياسات المكررة المشاهدة للفرد $\boldsymbol{i}\mathbf{th}$ المشارك في الدراسة في المجموعة $\boldsymbol{j}\mathbf{th}$. حيث أن لكلا المجموعتين $\mathbf{j}$, فأن $\mathbf{y_{ji}}$ يتبع التوزيع الطبيعي مع متوسط مساوي إلى $\boldsymbol{\mu}_\mathbf{j}$ ومصفوفة التباين والتباين المشترك أبعادها $(\mathbf{p\times p})$ المساوية إلى $\boldsymbol{\Omega}_\mathbf{j}$ أي أن :

$$\mathbf{y_{ji}} \sim \mathbf{N_p}\left(\mathbf{M_j}, \boldsymbol{\Omega}_\mathbf{j}\right) \quad , \quad \mathbf{j=1,2}$$

وبوجود تركيبة التباينات لأنموذج $(\mathbf{AR-1})$ لمصفوفة التباين المشترك $\boldsymbol{\Omega}_\mathbf{j}$, وبالتالي لكل من $\mathbf{j=1,2}$ فأن $\boldsymbol{\Omega}_\mathbf{j}$ تعطى بالصيغة الآتية [3],[6],[8] :

$$\boldsymbol{\Omega}_\mathbf{j} = \boldsymbol{\sigma}_\mathbf{j}^{\mathbf{2}} \begin{bmatrix} \mathbf{1} & \boldsymbol{\rho}_\mathbf{j} & \boldsymbol{\rho}_\mathbf{j}^{\mathbf{2}} & \cdots & \boldsymbol{\rho}_\mathbf{j}^{\mathbf{p-1}} \\ \boldsymbol{\rho}_\mathbf{j} & \mathbf{1} & \boldsymbol{\rho}_\mathbf{j} & \cdots & \boldsymbol{\rho}_\mathbf{j}^{\mathbf{p-2}} \\ \boldsymbol{\rho}_\mathbf{j}^{\mathbf{2}} & \boldsymbol{\rho}_\mathbf{j} & \mathbf{1} & \cdots & \boldsymbol{\rho}_\mathbf{j}^{\mathbf{p-3}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\rho}_\mathbf{j}^{\mathbf{p-1}} & \boldsymbol{\rho}_\mathbf{j}^{\mathbf{p-2}} & \boldsymbol{\rho}_\mathbf{j}^{\mathbf{p-3}} & \cdots & \mathbf{1} \end{bmatrix} \qquad \ldots\ldots(\mathbf{1})$$

وأن مقدار مُحَدِّدَة (**Determinant**) المصفوفة $\boldsymbol{\Omega}_\mathbf{j}$ يعطى بالصيغة التالية :

$$\left|\boldsymbol{\Omega}_\mathbf{j}\right| = \left(\boldsymbol{\sigma}_\mathbf{j}^{\mathbf{2}}\right)^{\mathbf{p}}\left(\mathbf{1}-\boldsymbol{\rho}_\mathbf{j}^{\mathbf{2}}\right)^{\mathbf{p-1}} \qquad \ldots\ldots(\mathbf{2})$$

وكذلك فأن معكوس (**Inverse**) المصفوفة $\boldsymbol{\Omega}_\mathbf{j}$ هو مصفوفة ثلاثية الأقطار(**tridiagonal matrix**) ويعطى بالصيغة التالية :

$$\boldsymbol{\Omega}_\mathbf{j}^{\mathbf{-1}} = \frac{\mathbf{1}}{\boldsymbol{\sigma}_\mathbf{j}^{\mathbf{2}}\left(\mathbf{1}-\boldsymbol{\rho}_\mathbf{j}\right)}\boldsymbol{\Omega}_{\mathbf{0j}} \qquad \ldots\ldots(\mathbf{3})$$

حيث أن $\boldsymbol{\Omega}_{\mathbf{0j}}$ مساوية إلى :

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

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 $(AR-1)$ 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 $(\text{discriminant function})$ في هذه حالة $(AR-1)$ كتركيبة خطيه من القياسات المكررة p والتي تعطى بالصيغة الآتية [8],[7],[6],[4]:

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

حيث أن المتجه y_{ji} هو متجه أبعادهُ $(p \times 1)$ للقياسات المشاهدة للفرد ith المشارك في الدراسة في المجموعة jth ويعطى بالشكل الأتي :

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

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

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

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

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

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

وطبقا للحالة هذه فأن قاعِدَةُ التمييز والتصنيف الخطية $(\text{linear classification rule})$ هي: أنهُ يتم تخَصّيَصَ الفرد ith مع المشاهدات $(y_{ji1}, y_{ji2}, \dots, y_{jip})$ إلى ألمجموعه الأولى فيما إذا كان [8],[7],[6],[4]:

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

وعدا ذلِك فأنهُ يتم تخَصّيَصَ الفرد ith مع الاستجابات $(y_{ji1}, y_{ji2}, \dots, y_{jip})$ إلى ألمجموعه الثانية. وواضح أن قاعِدَةُ التصنيف لهذه الحالة مستقلة عن مركبة التباين σ^2, حيث أن $\bar{y}_{l,m}$ يقدر بتطبيق الصيغة التالية :

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

حيث أن y_l تمَثّلَ مقدار الاستجابة للفرد ith المُشَارِك أو الجديد في الدراسة عند القياس المكرر أو النقطة الزمنية tth 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 عيها كما يل [7],[6]

بوجود تركيبة منتظمة لمتوسطات المجموعة, وليكن 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

$\mathbf{Y_1} = (\mathbf{y_{11}}, \mathbf{y_{12}}, \dots, \mathbf{y_{1n_1}})$ تمثل المشاهدات من المجتمع الأول $\mathbf{N_p}(\boldsymbol{\mu}_1, \boldsymbol{\Omega})$ و $\mathbf{Y_2} = (\mathbf{y_{21}}, \mathbf{y_{22}}, \dots, \mathbf{y_{2n_2}})$ تدل على المشاهدات من المجتمع الثاني $\mathbf{N_p}(\boldsymbol{\mu}_2, \boldsymbol{\Omega})$. بالتالي, فأن لوغاريتم دالة الإمكان الأعظم المشتركة $(\textbf{log joint likelihood function})$ يعطى بالصيغة التالية :

$$\log L(\theta, \Omega\,; Y_1, Y_2, \dots, Y_N) = -\frac{Np}{2}\log 2\pi - \frac{N}{2}\log|\Omega| - \frac{1}{2}\mathrm{tr}\left(\sum_{j=1}^{2}\sum_{i=1}^{n_j}(y_{ij} - \mu_j)'\Omega^{-1}(y_{ij} - \mu_j)\right) \quad \dots (10)$$

وبوجود تركيبة متوسط منتظمة لمتوسطات المجموعة فأن :

$$\mu_j = c_j 1_p \qquad \dots\ (11)$$

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

$$\log L(\theta, \Omega\,; Y_1, Y_2) = -\frac{Np}{2}\log 2\pi - \frac{N}{2}\log|\Omega| - \frac{1}{2}\mathrm{tr}\,(\Omega^{-1}W) - \frac{1}{2}\left(\sum_{j=1}^{2} n_j(\bar{y}_j - c_j 1_p)'\,\Omega^{-1}(\bar{y}_j - c_j 1_p)\right) \quad \dots (12)$$

وبوجود تركيبة التباين لأنموذج $(\mathbf{AR-1})$ على المصفوفة $\boldsymbol{\Omega}_\mathbf{j}$ نقوم بتعويض قيم كل من محددة $(\mathbf{2})$ ومعكوس $(\mathbf{3})$ المصفوفة $\boldsymbol{\Omega}_\mathbf{j}$ في معادلة $(\mathbf{12})$ فأنه يعطي :

$$\log L(c_1, c_2, \Omega\,; Y_1, Y_2) = -\frac{Np}{2}\log 2\pi - \frac{Np}{2}\log\sigma^2 - \frac{N(p-1)}{2}\log(1-\rho^2) - \frac{1}{2}\mathrm{tr}\frac{1}{\sigma^2(1-\rho^2)}\Omega_0 \times \left[W_0 - n_1c_1W_5 - n_2c_2W_6 + (n_1{c_1}^2 + n_2{c_2}^2)J_p\right] \quad \dots\dots (13)$$

حيث أن قيمة $\left(\mathbf{W_j}, \mathbf{j} = \mathbf{0}, \mathbf{3}, \mathbf{4}, \mathbf{5}, \mathbf{6}\right)$ موضحة في معادلة $(\mathbf{22})$.

ألان, وباشتقاق المعادلة رقم $(\mathbf{13})$ مرة واحدة بالنسبة لكل من $\mathbf{c_1}$ و $\mathbf{c_2}$ على التوالي وَنضْعها مساوية إلى الصفر وبعد عملية تبسيطها نحصل على المعادلتين أدناه :

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

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

في المعادلات أعلاه فأن كل من $\mathbf{m_{11}}$ و $\mathbf{m_{21}}$ تقدر من قبل المعادلتين الآتيتين :

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

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

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

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

وباشتقاق المعادلة $(\mathbf{13})$ مرة واحدة بالنسبة إلى $\boldsymbol{\sigma}^\mathbf{2}$ ومساواتها إلى الصفر وتبسيطها نجد أن :

$$\sigma^2 = \frac{1}{np(1-\rho^2)}\left[\mathrm{tr}\Omega_0W_0 - n_1c_1\mathrm{tr}\Omega_0W_5 - n_2c_2\mathrm{tr}\Omega_0W_6 + (n_1{c_1}^2 + n_2{c_2}^2)\mathrm{tr}(\Omega_0 J)\right] \quad \dots (18)$$

$$\mathrm{tr}\Omega_0W_0 = \beta_1\rho^2 - 2\gamma_1\rho + \alpha_1\ , \quad \mathrm{tr}\Omega_0W_5 = \beta_2\rho^2 - 2\gamma_2\rho + \alpha_2\ ,$$

$$\mathrm{tr}\Omega_0W_6 = \beta_3\rho^2 - 2\gamma_3\rho + \alpha_3$$

حيث أن القيم التقديرية لكل من $\boldsymbol{\gamma}_\mathbf{1}, \boldsymbol{\beta}_\mathbf{1}, \boldsymbol{\alpha}_\mathbf{1}$ و $\boldsymbol{\gamma}_\mathbf{3}, \boldsymbol{\beta}_\mathbf{2}, \boldsymbol{\alpha}_\mathbf{2}$ و $\boldsymbol{\gamma}_\mathbf{3}, \boldsymbol{\beta}_\mathbf{3}, \boldsymbol{\alpha}_\mathbf{3}$ يمكن إيجادها بتطبيق صيغة كل منها :

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

$$\widehat{\alpha}_2 = \mathrm{tr}W_5 \ , \widehat{\beta}_2 = \mathrm{tr}W_5 - w_{5,11} - w_{5,pp} \ , \qquad \widehat{\gamma}_2 = \sum_{t=2}^{p} w_{5,t-1t} \ \ldots\ldots(20)$$

$$\widehat{\alpha}_3 = \mathrm{tr}W_6 \ , \widehat{\beta}_3 = \mathrm{tr}W_6 - w_{6,11} - w_{6,pp} \ , \qquad \widehat{\gamma}_3 = \sum_{t=2}^{p} W_{6,t-1t} \ \ldots(21)$$

حيث أن الكمية $w_{ij,l}$ هنا تمثل العنصر $(i,j)^{th}$ من المصفوفة W_l $(l = 0, \ldots, 6)$ والتي يتم تقديرها بتطبيق الصيغ التالية :

$$W_0 = W_1 + W_2 + n_1W_3 + n_2W_4 \ , \ \ W_3 = \bar{y}_1\bar{y}_1' \ \ , W_4 = \bar{y}_2\bar{y}_2', \ \ W_5 = \left(1_p\bar{y}_1' + \bar{y}_1 1_p'\right), \ \ W_6 = \left(1_p\bar{y}_2' + \bar{y}_2 1_p'\right) \ \ldots\ldots(22)$$

وفيها يمكن تقدير كل من المصفوفتين W_1 و W_2 بتطبيق :

$$\widehat{W}_j = \sum_{i=1}^{n_j}\left(y_{ij} - \bar{y}_j\right)\left(y_{ij} - \bar{y}_j\right)', \qquad j = 1,2 \qquad \ldots(23)$$

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

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

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

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

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

2-2-2 قاعدة التمييز والتصنيف التربيعية للحالة الثانية

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

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

فأن قاعِدَةً التمييز والتصنيف التربيعية (**Quadratic classification rule**) هي : أنهُ يتم تَخَصّيَصَ الفرد ith مع المشاهدات $\left(y_{ji1}, y_{ji2}, \ldots, y_{jip}\right)$ إلى ألمجموعه الأولى إذا كان [6] :

$$\lambda(y_i) = -\frac{1}{2\sigma^2}\left(\frac{1}{1-{\rho_1}^2} - \frac{1}{1-{\rho_2}^2}\right)\sum_{t=1}^{p}{y_{it}}^2 - \frac{1}{2\sigma^2}\left(\frac{{\rho_1}^2}{1-{\rho_1}^2} - \frac{{\rho_2}^2}{1-{\rho_2}^2}\right)\sum_{t=2}^{p-1}{y_{it}}^2$$
$$+\frac{1}{\sigma^2}\left(\frac{\rho_1}{1-{\rho_1}^2} - \frac{\rho_2}{1-{\rho_2}^2}\right)\sum_{t=1}^{p-1}y_{it}\,y_{it+1} + \frac{p}{\sigma^2}\left(\frac{c_1}{1+\rho_1} - \frac{c_2}{1+\rho_2}\right)\bar{y}_{1,p}$$
$$-\frac{p-1}{\sigma^2}\left(\frac{\rho_1c_1}{1+\rho_1} - \frac{\rho_2c_2}{1+\rho_2}\right)\bar{y}_{2,p-1}$$
$$\geq \frac{1}{2\sigma^2}\left[\frac{{c_1}^2}{1+\rho_1}(p-(p-2)\,\rho_1) - \frac{{c_2}^2}{1+\rho_2}(p-(p-2)\,\rho_2)\right]$$
$$-\frac{p-1}{2}\ln\left(\frac{1-{\rho_2}^2}{1-{\rho_1}^2}\right) \qquad \ldots\ldots(26)$$

وعدا ذلِك فأنهُ يتم تخَصّيَصَ الفرد ith مع الاستجابات $\left(y_{ji1}, y_{ji2}, \ldots, y_{jip}\right)$ إلى ألمجموعه الثانية. في قاعِدَةً التصنيف أعلاه فأن تقديرات الإمكان الأعظم لكل من c_1 , c_2 , ρ_1 , ρ_2 و σ^2 يتم الحصول عيها بواسطة الحل وبشكل إني للمعادلات الآتية [6]:

$$(p-2)\,\widehat{\rho}_1\hat{c}_1 - p\hat{c}_1 + pm_{11} - (p-2)\,\widehat{\rho}_1m_{12} = 0 \ , \qquad \ldots(27)$$
$$(p-2)\,\widehat{\rho}_2\hat{c}_2 - p\hat{c}_2 + pm_{21} - (p-2)\,\widehat{\rho}_2m_{22} = 0 \ , \qquad \ldots(28)$$

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

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

و

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

وبقية القيم التقديرية يتم الحصول عليها بتطبيق الصيغ التالية :

$$\hat{\alpha}_8=trW_8,\quad \hat{\beta}_8=trW_8-w_{8,11}-w_{8,pp}\,,\quad \hat{\gamma}_8=\sum_{t=2}^{p}w_{8,t-1t} \quad \dots(32)$$

$$\hat{\alpha}_9=trW_9,\quad \hat{\beta}_9=trW_9-w_{9,11}-w_{9,pp}\,,\quad \hat{\gamma}_9=\sum_{t=2}^{p}w_{9t-1,t} \quad \dots(33)$$

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

$$W_8=W_1+n_1W_3\,,\quad W_9=W_2+n_2W_4 \quad \dots\dots(34)$$

إما α_2, β_2, γ_2 و α_3, β_3, γ_3 وكل من المصفوفتين W_3,W_1 و W_4,W_2 فان صيغة كل منها تم تعريفها في الحالة السابقة.

3-2-2 قاعدة التمييز والتصنيف التربيعية للحالة الثالثة

للحالة الثالثة والموضحة أدناه :

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

فأن قاعِدَةُ التمييز والتصنيف التربيعية (**Quadratic classification rule**) هي : أنهُ يتم تَخَصّيَصَ الفرد ith مع المشاهدات $(y_{ji1},y_{ji2},\dots,y_{jip})$ إلى ألمجموعه الأولى إذا كان [7],[6] :

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

وعدا ذلِك فأنهُ يتم تَخَصّيَصَ الفرد ith مع الاستجابات $(y_{ji1},y_{ji2},\dots,y_{jip})$ إلى ألمجموعه الثانية. في قاعِدَةُ التصنيف أعلاه فأن تقديرات الإمكان الأعظم لكل من c_1, c_2, ρ, $\sigma_1^{\,2}$ و $\sigma_2^{\,2}$ يتم الحصول عليها بواسطة حل للمعادلات الآتية وبشكل إني [6] :

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

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

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

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

و

$$\mathbf{N(p-1)(\hat\rho-\hat\rho^3)\hat\sigma_1{}^2\hat\sigma_2{}^2-\hat\sigma_2{}^2\{(\alpha_8+\beta_8)\hat\rho-\gamma_8\hat\rho^2-\gamma_8\}}$$
$$\mathbf{+\,n_1\hat c_1\hat\sigma_2{}^2\{(\alpha_2+\beta_2)\hat\rho-\gamma_2\hat\rho^2-\gamma_2\}}$$
$$\mathbf{-\,n_1\hat c_1{}^2\hat\sigma_2{}^2\{(2p-2)\hat\rho-(p-1)\hat\rho^2-(p-1)\}-\hat\sigma_1{}^2\{(\alpha_9+\beta_9)\hat\rho-\gamma_9\hat\rho^2-\gamma_9\}}$$
$$\mathbf{+\,n_2\hat c_2\hat\sigma_1{}^2\{(\alpha_3+\beta_3)\hat\rho-\gamma_3\hat\rho^2-\gamma_3\}}$$
$$\mathbf{-\,n_2\hat c_2{}^2\hat\sigma_1{}^2\{(2p-2)\hat\rho-(p-1)\hat\rho^2-(p-1)\}=0\,.}\quad \dots\dots(40)$$

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

4-2-2 قاعدة التمييز والتصنيف التربيعية للحالة الرابعة

للحالة الرابعة :

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

فأن قاعِدَةٌ التمييز والتصنيف التربيعية (**Quadratic classification rule**) هي : أنهُ يتم تخَصّيَصَ الفرد $\boldsymbol{ith}$ مع المشاهدات $\left(y_{ji1},y_{ji2},\dots,y_{jip}\right)$ إلى ألمجموعه الأولى إذا كان [7],[6]:

$$\boldsymbol{\lambda(y_i)=-\frac{1}{2}\left(\frac{1}{\sigma_1{}^2(1-\rho_1{}^2)}-\frac{1}{\sigma_2{}^2(1-\rho_2{}^2)}\right)\sum_{t=1}^{p}y_{it}{}^2}$$
$$\boldsymbol{-\frac{1}{2}\left(\frac{\rho_1{}^2}{\sigma_1{}^2(1-\rho_1{}^2)}-\frac{\rho_2{}^2}{\sigma_2{}^2(1-\rho_2{}^2)}\right)\sum_{t=2}^{p-1}y_{it}{}^2}$$
$$\boldsymbol{+\left(\frac{\rho_1}{\sigma_1{}^2(1-\rho_1{}^2)}-\frac{\rho_2}{\sigma_2{}^2(1-\rho_2{}^2)}\right)\sum_{t=1}^{p-1}y_{it}\,y_{it+1}}$$
$$\boldsymbol{+\left(\frac{c_1}{\sigma_1{}^2(1+\rho_1)}-\frac{c_2}{\sigma_2{}^2(1+\rho_2)}\right)p\bar y_{1,p}}$$
$$\boldsymbol{-(p-2)\left(\frac{\rho_1c_1}{\sigma_1{}^2(1+\rho_1)}-\frac{\rho_2c_2}{\sigma_2{}^2(1+\rho_2)}\right)\bar y_{2,p-1}}$$
$$\boldsymbol{\geq\frac{1}{2}\left(\frac{c_1{}^2(p-(p-2)\rho_1)}{\sigma_1{}^2(1+\rho_1)}-\frac{c_2{}^2\,(p-(p-2)\rho_2)}{\sigma_2{}^2(1+\rho_2)}\right)-\frac{p}{2}\ln\frac{\sigma_2{}^2}{\sigma_1{}^2}}$$
$$\boldsymbol{-\frac{p-1}{2}\ln\left(\frac{1-\rho_2{}^2}{1-\rho_1{}^2}\right)}\quad \dots\dots(41)$$

وعدا ذلِك فأنهُ يتم تخَصّيَصَ الفرد $\boldsymbol{ith}$ مع الاستجابات $\left(\mathbf{y_{ji1},y_{ji2},\dots,y_{jip}}\right)$ إلى ألمجموعه الثانية. في قاعِدَةٌ التصنيف أعلاه فأن تقديرات الإمكان الأعظم لكل من $\mathbf{\sigma_1{}^2,\rho_1,c_1}$ و $\mathbf{\sigma_2{}^2,\rho_2,c_2}$ يتم الحصول عليها بواسطة حل للمعادلات الآتية أنيا [7],[6] :

$$\mathbf{(p-2)\,\hat\rho_1\hat c_1-p\hat c_1+pm_{11}-(p-2)\,\hat\rho_1m_{12}=0\,,}\quad \dots\dots(42)$$
$$\mathbf{(p-2)\,\hat\rho_2\hat c_2-p\hat c_2+pm_{21}-(p-2)\,\hat\rho_2m_{22}=0\,,}\quad \dots\dots(43)$$
$$\mathbf{n_1\hat\sigma_1{}^2p\left(1-\hat\rho_1{}^2\right)-\left(\beta_8\,\hat\rho_1{}^2-2\gamma_8\,\hat\rho_1+\alpha_8\right)+n_1\hat c_1\left(\beta_2\,\hat\rho_1{}^2-2\gamma_2\,\hat\rho_1+\alpha_2\right)}$$
$$\mathbf{-\,n_1\hat c_1{}^2((p-2)\,\hat\rho_1{}^2-2(p-1)\,\hat\rho_1+p)=0,}\quad \dots\dots(44)$$
$$\mathbf{n_2\hat\sigma_2{}^2p\left(1-\hat\rho_2{}^2\right)-\left(\beta_9\,\hat\rho_2{}^2-2\gamma_9\,\hat\rho_2+\alpha_9\right)+n_2\hat c_2\left(\beta_3\,\hat\rho_2{}^2-2\gamma_3\,\hat\rho_2+\alpha_3\right)}$$
$$\mathbf{-\,n_2\hat c_2{}^2((p-2)\,\hat\rho_2{}^2-2(p-1)\,\hat\rho_2+p)=0,}\quad \dots\dots(45)$$
$$\mathbf{n_1(p-1)(\,\hat\rho_1-\hat\rho_1{}^3)\hat\sigma_1{}^2-\left\{(\alpha_8+\beta_8)\,\hat\rho_1-\gamma_8\,\hat\rho_1{}^2-\gamma_8\right\}}$$
$$\mathbf{+\,n_1\hat c_1\left\{(\alpha_2+\beta_2)\,\hat\rho_1-\gamma_2\,\hat\rho_1{}^2-\gamma_2\right\}}$$
$$\mathbf{-\,n_1\hat c_1{}^2\left\{(2p-2)\,\hat\rho_1-(p-1)\,\hat\rho_1{}^2-(p-1)\right\}=0\,,}\quad \dots\dots(46)$$

و

$$\mathbf{n_2(p-1)(\,\hat\rho_2-\hat\rho_2{}^3)\hat\sigma_2{}^2-\left\{(\alpha_9+\beta_9)\,\hat\rho_2-\gamma_9\,\hat\rho_2{}^2-\gamma_8\right\}}$$
$$\mathbf{+\,n_2\hat c_2\left\{(\alpha_3+\beta_3)\,\hat\rho_2-\gamma_3\,\hat\rho_2{}^2-\gamma_3\right\}}$$
$$\mathbf{-\,n_2\hat c_2{}^2\left\{(2p-2)\,\hat\rho_2-(p-1)\,\hat\rho_2{}^2-(p-1)\right\}=0\,.}\quad \dots\dots(47)$$

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

3 . تقدير نسب التصنيف الخاطئ Estimating Misclassification Rates

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

1.3 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 :

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

2.3 نسبة التصـــنيف الصـحيح الظاهرة (**Apparent Correct rate, APCR**): وهي عبارة عن نسبة عدد الإفراد التي ثَبتَ تصنيفها إلى عددهم الكلي في المجموعتين وتعطى بالصيغة الآتية :

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

4. الجانب التطبيقي

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

داء السـكري (**Diabetes mellitus**) هـو متلازمـة تتصـف باضـطراب الأسـتقلاب وارتفـاع شـاذ فـي تركيـز سـكر الـدم عـن المعـدل الطبيعـي الـذي يتـراوح بـين (**80 – 100**مليجرام/ديسيلتر) النـاجم عـن عـوز هرمـون الأنسـولين الـذي تفرزه غدة البنكرياس أو انخفاض حساسية الأنسجة للأنسولين، أو كلا الأمرين.

إن مـن أهـم الأسـباب للإصـابة بهـذا المـرض هـو نقـص الأنسـولين والـذي يعتبـر الهرمـون الأساسـي الـذي يـنظم نقـل الغلوكـوز مـن الـدم إلـى معظـم خلايـا الجسـم ، خصوصـاً الخلايـا العضـلية والخلايـا الدهنيـة، ولكـن لا ينقلـه إلـى خلايـا الجهـاز العصبي المركزي. ولذلك يؤدي نقص الأنسولين أو عدم استجابة الجسم له إلى أي نمط من أنماط السكري وهي :

- الـــــنمط الأول : يتميـــــز الـــــنمط الأول (**I**) مـــــن الســـــكري بخســـــارة الخلايـــــا بيتـــــا المنتجـــــة للأنســـــولين فـــي خلايـــا لانغرهـــانس بالبنكريـــاس ممـــا يـــؤدي إلـــى نقـــص الأنســـولين والســـبب الرئيســـي لهـــذه الخســـارة هـــو مناعـــة ذاتيـــة تتميـــز بهجـــوم الخلايـــا تـــاء المناعيـــة علـــى خلايـــا بيتـــا المنتجـــة للأنســـولين, أي أن مرضـــى هـــذا النـــوع مـــن الســـكري هـــم مرضى سكري معتمدون في علاجهم على الأنسولين الذي يعطى بشكل حقن تحت الجلد.
- الـنمط الثـاني : يتميـز الـنمط الثـاني مـن السـكري باختلافـه عـن الـنمط الأول مـن حيـث وجـود مقاومـة مضـادة لمفعـول الأنسـولين بالإضـافة إلـى قلـة إفـراز الأنسـولين ولا تسـتجيب مسـتقبلات الأنسـولين الموجـودة فـي الأغلفـة الخلويـة لمختلـف أنسـجة الجسـم بصـورة صـحيحة للأنسـولين, أي أن مرضـى هـذا النـوع مـن السـكري هـم مرضـى سـكري معتمدون في علاجهم على الحبوب (ضعيفي السيطرة السكرية).

وهنـاك عـدة أعـراض تـوحي بالإصـابة بهـذا المـرض ومنهـا زيـادة عـدد مـرات التبـول (بسـبب ارتفـاع الضـغط التناضـحي), زيـادة الإحسـاس بـالعطش (تنـتج عنهـا زيـادة تنـاول السـوائل لمحاولـة تعـويض زيـادة التبـول), التعـب الشـديد والعـام, فقدان الوزن رغم تناول الطعام بانتظام (شهية أكبر للطعام), تباطؤ شفاء الجروح, وتغيّم الرؤية.

أمـا تشـخيص الـنمط الأول والعديـد مـن حـالات الـنمط الثـاني مـن السـكري فيـتم بنـاءا علـى الأعـراض الأوليـة التـي تظهـر فـي بدايـة المـرض مثـل كثـرة التبـول والعطـش الزائـد وقـد يصـاحبها فقـد للـوزن، ويـتم عـادة تشـخيص بقيـة أنمـاط السـكري بطـرق أخـرى مثـل اكتشـاف ارتفـاع مسـتوى غلوكـوز الـدم أثنـاء أجـراء أحـد التحاليـل؛ ويـتم عـادة اكتشـاف المـرض عنـدما يعـاني المـريض مـن مشـكلها يسـببها السـكري بكثـرة مثـل السـكتات القلبيـة، اعـتلال الكلـى، بطـئ التئـام الجـروح أو تقـيح القـدم، مشـكلة معينـــة فـــي العـــين, ويمكـــن تشـــخيص المـــرض مـــن خـــلال نســـبة الهيموغلـــوبين الغليكـــوزيلاتي (**Glycosylated hemoglobin** (**HbA1c**)) وهـــو مـــا يســـمى بالتحليـــل التراكمـــي لســـكر الـــدم, حيـــث يعتبـــر بـروتين (**HbA1c**) تحليـل تشخيصـي لمـرض السـكري فـي حـال إذا كانـت نسـبته فـي الـدم أكثـر مـن (**6.5%**) حيـث أن النسـبة الطبيعيـة المعتمـدة هـي أقـل مـن (**5.9%**) ويـزداد فـي مرضـى السـكري فـي حالـة عـدم الانتظـام فـي العـلاج, ويـذكر أن هـذا التحليـل يتميـز بثبـات نسـبته وعـدم تـأثره بصـيام المراجـع أو بعوامـل أخـرى يمكـن أن تـؤثر عنـد سـحب عينات دم لتحليلها، وبذلك فيمكن عمله في أي وقت ولا حاجة لأن يكون الشخص صائما قبل سحب العينة.

ومـن خـلال ذلـك تتضـح أهميـة التشـخيص المبكـر للمـرض لتجنـب الحـالات المعقـدة والتـي تـؤدي إلـى مضـاعفات خطيـرة وحتـى إلـى الوفـاة المبكـرة مـع تطـور المـرض إلـى الحـالات المتقدمـة, وأن اسـتخدام الأسـاليب الإحصـائية ومنهـا الـدال التمييزية هي لمساعدة ذوي الاختصاص لتشخيص مجاميع المرض وفي حالاتها المبكرة.

2.4 وصف العينة ومتغير الدراسة

شـملت عينـــة البحـــث موضـــوع الدراســـة بيانـــات لمجمـــوعتين تضـــمنت المرضـــى المصـــابين بـــداء الســـكري, إذ تـــم الحصـــول علـــى البيانـــات مـــن البطاقـــات الخاصـــة بالمرضـــى فـــي المركـــز الـــوطني للســـكري/الجامعة المستنصـــرية لفحـــص تـــأثير عقـــارين همـــا عقـــار حبـــوب ريباجلينيـــد (**Repaglinide**) وعقـــار الأنســـولين المخلـــوط (**Novo Mix**)علـــى اســـتجابة مرضـــى الســـكري والســـيطرة علـــى المـــرض ولمـــدة زمنيـــة قـــدرها ســـت أشـــهر, إذ تمثـــل هـــذه البيانـــات القياســـات الشـــهرية لنســـبة بـــروتين الهيموغلـــوبين الغليكـــوزيلاتي (**HbA1c**) فـــي الـــدم والمـــأخوذة فـــي ثـــلاث مراحـــل, المرحلـــة الأولـــى

كانت في بداية التجربة ($\mathbf{month_0}$), المرحلة الثانية كانت بعد ثلاث أشهر ($\mathbf{month_3}$) , أما المرحلة الثالثة فكانت بعد ست أشهر ($\mathbf{month_6}$). ولمجموعتين من المرضى, تمثلت المجموعة الأولى بعدد ($\mathbf{38}$) مريض يعاني من داء السكري من النمط الأول ($\mathbf{I}$) (مرضى سكري معتمدون في علاجهم على الأنسولين), في حين كانت المجموعة الثانية تمثل منهم ($\mathbf{33}$) مريض يعاني من داء السكري من النمط الثاني ($\mathbf{II}$) (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 الخاطئ 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 $(\mathbf{AR-1})$ ولكل حالة من الحالات الأربعة التي تم إيضاحها مسبقا :

1.3.4 الحالة الأولى

بافتراض وجود الحالة الأولى :

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

للحصول على نتائج الدالة التمييزية في هذه الحالة ينبغي أولا تقدير كل من $\hat{\mathbf{c}}_1$ و $\hat{\mathbf{c}}_2$ وذلك لغرض تَخْمِين القيم التقديرية $\hat{\boldsymbol{\mu}}_1$ و $\hat{\boldsymbol{\mu}}_2$ للمجموعتين الأولى والثانية , وكذلك تقدير كل من $\hat{\boldsymbol{\rho}}$ و $\hat{\boldsymbol{\sigma}}^2$ وبالتالي تقدير مصفوفة التباينات $(\mathbf{AR-1})$ 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 $\mathbf{m_{11}}$ و $\mathbf{m_{21}}$ من $(\mathbf{16})$:

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

ثم نجد أيضا تقديري كل من $\mathbf{m_{12}}$ و $\mathbf{m_{22}}$ باعتماد $(\mathbf{17})$:

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

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

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

وبتطبيق معادلة $(\mathbf{22})$ نجد :

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

وبتعويض كل من $\mathbf{W_1}$, $\mathbf{W_2}$ و $\mathbf{W_3}$, $\mathbf{W_4}$ في $\mathbf{W_0}$ ومن $(\mathbf{22})$ نجد تقدير المصفوفة $\mathbf{W_0}$ مساوية إلى :

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

بالتالي فان تقدير $\hat{\boldsymbol{\alpha}}_1$, $\hat{\boldsymbol{\beta}}_1$ و $\hat{\boldsymbol{\gamma}}_1$ يكون مساوي إلى :

$$\hat{\boldsymbol{\alpha}}_1 = \mathbf{13356.5715}, \qquad \hat{\boldsymbol{\beta}}_1 = \mathbf{4762.5965}, \qquad \hat{\boldsymbol{\gamma}}_1 = \mathbf{9560.0085}$$

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

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

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 $\gamma_1, \beta_1, \alpha_1$ و $\gamma_2, \beta_2, \alpha_2$ و $\gamma_3, \beta_3, \alpha_3$ 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, فأن 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 فان 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 $\hat{\Omega}_{AR-1}$ فانه من المكن 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 فأن تقدير الدالة التمييزية في حالة تركيبة التباين المشترك هذه والمعطاة في المعادلة رقم (5) تكون مساوية إلى :

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

ألان 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) فأنه يتم تخَصّيَصَ الفرد ith مع المشاهدات $(y_{ji1}, y_{ji2}, \dots, y_{jip})$ إلى ألمجموعه الأولى فيما إذا كان $\lambda(y_i)$ بالصيغة الآتية :

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

وعدا ذلك يتم تخَصّيَصَه إلى المجموعة الثانية.

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

الجدول ($\mathbf{1}$): نتائج التحليل التمييزي لبيانات القياسات المكررة بوجود تركيبة التباين المشترك ($\mathbf{AR-1}$) (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 ($\mathbf{AR-1}$), وبالنسبة إلى المجموعة الأولى فأن جميع الإفراد والذين عددهم **38** أي بنسبة ($\mathbf{100.00\%}$) صنفوا بشكل صحيح إلى مجموعتهم. وبالمثل; بالنسبة للمجموعة الثانية فانه من أصل ($\mathbf{33}$) من الإفراد هناك **29** فرد أي بنسبة ($\mathbf{87.88\%}$) صنفوا بشكل صحيح إلى مجموعتهم.

إما نسبة التصنيف الخطأ الظاهرة الكلية يمكن إيجادها وباعتماد معادلة ($\mathbf{48}$), وكما يلي :

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

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

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

أي أن العدد الكلي للإفراد المصنفين بشكل صحيح هو ($\mathbf{67}$) أي ما يعادل ($\mathbf{94.37\%}$) لحجم العينة الكلي من **71** فرد مشارك في الدراسة.

2.3.4 الحالة الثانية

بافتراض وجود الحالة الثانية :

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

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

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

حيث أن كل من $\mathbf{W_1}$ و $\mathbf{W_3}$ قد سبق تقديرهما, وبالتالي فأن :

$$\boldsymbol{\hat{\alpha}_8} = \mathbf{10881.5202}, \qquad \boldsymbol{\hat{\beta}_8} = \mathbf{3499.4066}, \qquad \boldsymbol{\hat{\gamma}_8} = \mathbf{7053.5756}$$

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

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

حيث أن كل من $\mathbf{W_2}$ و $\mathbf{W_4}$ قد سبق تقديرهما, وبالتالي فأن :

$$\boldsymbol{\hat{\alpha}_9} = \mathbf{3775.4297}, \qquad \boldsymbol{\hat{\beta}_9} = \mathbf{1263.1899}, \qquad \boldsymbol{\hat{\gamma}_9} = \mathbf{2506.4329}$$

وبعد إيجاد القيم التقديرية $\boldsymbol{\alpha_8}$ $\boldsymbol{\beta_8}$, $\boldsymbol{\gamma_8}$ و $\boldsymbol{\alpha_9}$, $\boldsymbol{\beta_9}$, $\boldsymbol{\gamma_9}$ يتم تعويض كل منهما في ($\mathbf{27}$), ($\mathbf{28}$), ($\mathbf{29}$), ($\mathbf{30}$) و ($\mathbf{31}$) لغرض الحصول على تقديرات الإمكان الأعظم $\boldsymbol{\hat{c}_1}$, $\boldsymbol{\hat{c}_2}$, $\boldsymbol{\hat{\rho}_1}$, $\boldsymbol{\hat{\rho}_2}$ و $\boldsymbol{\hat{\sigma}^2}$ المطلوب إيجادها لبناء قاعدة التمييز والتصنيف للحالة الثانية للتحليل التمييزي بوجود تركيبة ($\mathbf{AR-1}$), ويمثل المتجه أدناه تقديرات الإمكان الأعظم للحالة الثانية التي تم التوصل إليها بحل تلك المعادلات :

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

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

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

وعدا ذلك يتم تخَصّيَصَه إلى المجموعة الثانية.
وبالتالي وبالاستناد إلى قاعدة التمييز والتصنيف ذات تركيبة التباينات $(AR-1)$ للحالة الثانية فأن الجدول التالي يعطي كِلَ 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 للإفراد 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:

$$\widehat{\alpha}_8 = 10881.5202, \quad \widehat{\beta}_8 = 3499.4066, \quad \widehat{\gamma}_8 = 7053.5756,$$
$$\widehat{\alpha}_2 = 57.51, \quad \widehat{\beta}_2 = 18.948, \quad \widehat{\gamma}_2 = 38.229$$

$$\widehat{\alpha}_9 = 3775.4297, \quad \widehat{\beta}_9 = 1263.1899, \quad \widehat{\gamma}_9 = 2506.4329,$$
$$\widehat{\alpha}_3 = 36.43\,, \quad \widehat{\beta}_3 = 12.182, \quad \widehat{\gamma}_3 = 24.306$$

في المعادلات (36), (37), (38), (39) و(40) والمعطاة في الحالة الثالثة لغرض الحصول على تقديرات الإمكان الأعظم للمعالم $\hat{c}_1$, $\hat{c}_2$, $\widehat{\sigma}_1{}^2$, $\widehat{\sigma}_2{}^2$ و $\widehat{\rho}$ المطلوب إيجادها لبناء قاعدة التمييز والتصنيف التربيعية 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 $\hat{c}_1, \hat{c}_2, \hat{\sigma}_1^2, \hat{\sigma}_2^2$ و $\hat{\rho}$ في قاعدة التمييز والتصنيف $\lambda(y_i)$ للحالة الثالثة المعطاة في معادلة (35) فأنه يتم تخَصّيَصَ الفرد ith مع المشاهدات $(y_{ji1}, y_{ji2}, \dots, y_{jip})$ إلى ألمجموعه الأولى فيما إذا كان $\lambda(y_i)$ بالصيغة الآتية :

$\lambda(y_i) \geq 14.8823$

وعدا ذلك يتم تخَصّيَصَه إلى المجموعة الثانية. وبالتالي وبالاستناد إلى قاعدة التمييز والتصنيف ذات تركيبة التباينات $(AR-1)$ للحالة الثالثة فأن الجدول الأتي يعطي كِلَ 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 فأن 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 إي ما يعادل (87.32%) لحجم العينة الكلي من 71 فرد مشارك في الدراسة.

4.3.4 الحالة الرابعة

وبافتراض وجود الحالة الرابعة :

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

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 $\hat{\sigma}_1^2, \hat{\rho}_1, \hat{c}_1$ و $\hat{\sigma}_2^2, \hat{\rho}_2, \hat{c}_2$ في قاعدة التمييز والتصنيف $\lambda(y_i)$ للحالة الرابعة المعطاة في معادلة (41) فأنه يتم تخَصّيَصَ الفرد ith مع المشاهدات $(y_{ji1}, y_{ji2}, \dots, y_{jip})$ إلى ألمجموعه الأولى فيما إذا كان $\lambda(y_i)$ بالصيغة الآتية :

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

وعدا ذلك يتم تخَصّيَصَه إلى المجموعة الثانية.

وبالتالي وبالاستناد إلى قاعدة التمييز والتصنيف ذات تركيبة التباينات $(AR-1)$ للحالة هذه فأن الجدول الأتي يعطي كِلَ 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 فأن نسبة التصنيف الخطأ 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 $APER$ 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 إي 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 الخطأ الظاهرة الكلية والتي بلغت فيه $(APER = 5.63\%)$.

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

References المصادر

أولا : المصادر العربية

1- 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 (**203 – 220**).

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- 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.