AL KUT JOURNAL OF ECONOMIC AND ADMINISTRATIVE SCIENCES
Publisher: College of Economics and Management - Wasit University

التحليل التميزي بوجود مشكلة تعدّد خطّي (تطبيق عملي)

م. وهاب سالم محمد

Wahabsta@uodiyala.edu.iq

جامعة ديالى / كلية الإدارة والاقتصاد/ قسم الاحصاء

مستخلص

تُعدّ طريقة الدالّة المميزة الخطّية Linear discriminant functionمن أهّم الطرائق الإحصائية المستعملة في التصنيف، في حالة وجود مشكلة التعدّد الخطّي Multicollinearity بين البيانات (إنّ البيانات يوجد فيها ارتباطات عالية بين المتغيرات) أصبح عدم إمكان استعمال الانحدار اللوجيستي والدالّة المميزة الخطية، ولحل هذه المشكلة نلجأ إلى طريقة انحدار المربعات الصغرى الجزئية Partiial least squeare لتقدير معلمات الانموذج لمشكلة التعدّد الخطي،

يُعدّ تشخيص أمراض القلب باستعمال نماذج تصنيفية وتمييزية إحصائية أمرا مهماً ، لأنّ أمراض القلب من الأمراض الخطيرة والتي تفتك بحياة المصاب لذلك يجب التشخيص المبكر وذلك من خلال تصنيف وتمييز المرضى إلى مصابين وغير مصابين استنادا إلى العوامل المؤثّرة على الإصابة بالمرض باستعمال تقنيات مثل التحليل التمييزي ، ويهدف البحث لإيجاد بعض العوامل التي لها دور في تمييز أمراض القلب المصابين من غير المصابين ، ويوصي البحث بتطوير النموذج التمييزي وذلك بإضافة عوامل أخرى لها تأثير في تمييز أمراض القلب المصابين من غير المصابين ، واستعمال الدالّة التمييزية مع توسيع الدراسة وذلك بزيادة المتغيّرات التي لم تدخل في البحث ولها تأثير كبير على الإصابة بأمراض القلب.

الكلمات المفتاحية : (المميزة ، الجزئية ، الهرمي ، المتدرّج ، التوضيحية ، المرتدة)

Abstract

The method of the linear discriminant function is one of the most important statistical methods used in classification, in the event that there is a problem of multicollinearity between the data (that the data has high correlations between variables), it has become impossible to use the logistic regression and the linear characteristic function, and to solve this problem we resort to

Partial least square regression method to solve the multilinearity problem, diagnosing heart diseases using taxonomic and statistical discriminatory models is an important matter because heart disease is a serious disease that kills the patient's life. Therefore, early diagnosis must be made by classifying and distinguishing patients into injured and uninfected based on the Factors affecting the incidence of the disease using techniques such as discriminatory analysis, and the research aims to identify the most important factors that have a role in distinguishing heart disease, infected from those who are not, and the research recommends developing a discriminatory model by adding other factors that have an impact in distinguishing heart disease infected from those without, and the use of The discriminatory function with the expansion of the study by increasing the variables that were not included in the research and It has a significant impact on heart disease

Keywords= (discriminant ، Partiial ، Hierarchical ، Stepwise ، explanatory ، Lagged)

1 - المقدّمة

في العالم توجد هناك الكثير من الأمراض والعراق وأحد من البلدان الموجودة في العالم التي طالما تعاني الكثير من الأمراض المتنوعة التي تصيب الإنسان سواء أكانت مزمنة أم مكتسبه أم متوارثة عن طريق الأجيال وتُعدّ فكرة البحث والمتضمنة تشخيص أمراض القلب باستعمال نماذج تصنيفية وتمييزية إحصائية متقدمة أمراً مهماً لكون أمراض القلب من الأمراض الخطيرة ، والتي تفتك بحياة الشخص المصاب ويساعد التشخيص المبكر للمرض في إنقاذ حياة الشخص المصاب ويمنعه الحالات المتطورة للمرض. ولأنّ التشخيص مسألة صعبة من الناحية الطبية ، لوجود أمراض عديدة تشترك في أعراض عدّة يصعب التمييز بينها لذلك فإنّ الأساليب الإحصائية المتقدمة تساعد ذوي الاختصاص في تشخيص بعض أمراض القلب بصورة صحيحة وفي مرحلة مبكرة ،وذلك باستعمال أسلوب التحليل-التمييزي الذي يُعدّ أحد أساليب تحليل متعدّد المتغيّرات (Multivariate Analysis) واستعملت لتحليل العلاقات المتداخلة بين المتغيّرات في الانموذج بطرائق مترابطة بين المتغيّرات كما أنّه يسعى إلى تكوين أنموذج إحصائي يصور العلاقة المتبادلة بين المتغيّرات المختلفة ،وتعود أهمّيته بصفة أساس إلى فعاليته في التميز بين مشاهدات العديد من المتغيّرات ، وذلك بإيجاد تركيبة خطّية لمجموعة المتغيّرات.

و التحليلِ التمييزي هو أسلوب إحصائي ذات أهمّية كبيرة في التحليل التي يمكن الاعتماد عليها في بناء قاعدة لإعادة توزيع المتغيّرات أو تصنيف المجتمعات التي تشترك في خصائص أو صفات معينّة، أيْ أن التحليل التمييزي يمثّل تقنية تستعمل لتصنيف الأفراد إلى وأحد من مجتمعين أو أكثر بالاعتماد على مقاييس معينة وكذلك ، على خصائص المشاهدة التي لابُدّ أن تتوافق بدرجة ما مع خصائص المجتمع الذي ستنتسب اليه، وذلك من خلال تكوين دالّة تسمّى الدالّة التمييزية.

إنّ تحليلَ متعدّد المتغيّرات بأساليبه المختلفة والذي يعتمد على وصف وتحليل الظواهر ذات الأبعاد والمتغيّرات المتعددة فإذا كانت المتغيّرات (X1,X2,X3,…….Xn) تشترك فيما بينها بمجموعة من الصفات والخصائص بدرجات متفاوتة فإنّ تحليلَ متعدّد المتغيّرات يتناول دراسة بيانات تلك المشاهدات والتعبير عنها من خلال اكثر المتغيّرات تأثيراً في الظاهرة محل الدراسة . وباستعمال طريقة مربعات الصغرى الجزئية للكشف عن هذه المشكلة فضلاً عن استعمال البرنامج الإحصائي (SPSS) في تحليل البيانات للكشف عن مشكلة تعدّد خطيٍ

2- مشكلة البحث

لغرض توظيف بيانات مرض القلب والتي تعاني من وجود مشكلة التعدّد الخطّي من أجل تكوين نموذج احتمالي للمريض بالقلب والذي يَتَمّ على أساس تصنيف أو تحديد المريض بالقلب ثم استعمال أسلوب من أساليب التصنيف لغرض إيجاد أفضل نموذج احتمالي بأقلّ خطأ تصنيف ممكن.

3- أهداف البحث

استعمال التحليل التمييزي متعدّد المتغيّرات وذلك لتحديد العوامل المؤثّرة في الإصابة بمرض القلب ويَتَمّ ذلك بإنشاء دالّة تمييزية تصنف الاشخاص إلى مجموعتين (مصاب وغير مصاب) وتحدد الأهمّية النسبية للعوامل المؤثّرة في الإصابة بمرض القلب ومدى مساهمة كُلّ عامل في التمييز والتصنيف وما يعانيه من مشكلة التعدّد الخطّي والتأثيرات المترتبة من وجود هذه المشكلة.

4- أهمّية البحث

هو تحديد وتوضيح طرائق الدالّة المميزة في مختلف المجالات بأسلوبها الإحصائي وكيفية تطبيقهاً وماهي العوامل المؤثّرة على مستوى الإصابة بأمراض القلب عن طريق الدالّة التمييزية. وبناء أنموذج إحصائي يقدّر احتمال الإصابة ممّا يؤدّي إلى تقليل قيمة الاحتمال بالإصابة للمريض بشكُلّ مناسب قادر على التنبُؤ فيما إذا كان الشخص يصاب أو لا يصاب بمرض القلب ،وتقدير احتمال الإصابة بالمرض

5- التحليل التمييزي [3, 7 , 8]

يلعب التحليل الإحصائي دورا مهماً في تحليل وتفسير النتائج للظواهر الطبيعة و الاجتماعية في المجتمع أحد طرائق البحث الذي يستخدم في دراسة المشاكُلّ الاجتماعية والاقتصادية وغيرها . وتَمّ تفسيره إلى تحليل أحادي وثنائي المتغيّرات ويبنى على حزمة من المتغيّرات أو العوامل الإحصائية لتحليل تعدّد التغيرات وكيفية التمييز بين مجموعتين أو أكثر من الأفراد أو الاشياء وتصنيف المفردات الجديدة على المجموعات ،ويعتمد أسلوب التمايز على الوصول إلى الدالّة تسمّى دالّة التمايز تعمل على زيادة الفروق بين متوسّط المجموعات كلما كان التمييز كفوء ومِنْ ثَمّ يقل خطأ التصنيف ويُعدّ التحليل التمييزي استكشافياً بطبيعته إذْ يكتشف أسباب الاختلاف المشاهدة عندما لا تستطيع فهم العلاقات المسببة بدرجة كافية الدقة.

إذْ يبيّن مفهوم التحليلِ التمييزيٍ وبشكُلّ عام نفرض أنّ باحثا يبحث مجتمعين A ,B مشتركين بدرجات متفاوتة في بعض الخواص التي يمكن قياسها عدديا ، إذا كان لدى الباحث عينة عشوائيا يعلم أنها من أحد هذين المجتمعين لكنه لا يعلم ما إذا كانت من المجتمع الأول أو المجتمع الثاني ،وإذا كان الباحث قد قام من تلك الخواص وحصل على القيم العددية :K بقياس هذه العينة من إذْ (X1,X2,X3,……..Xp) فكيف تستعمل هذه القيم لتحديد المجتمع الذي تنتمي اليه العينة؟ أن مثل هذه المشكلات تحل عن طريق إيجاد دالّة التمييز.

وتُعدّ أهمّية التحليل التمييزي كأحد طرائق تحليل المتغيّرات المقدرة للتمييز بين مجموعتين أو أكثر ويَتَمّ ذلك بإنشاء دوال التمايز "Discriminate Function" تعمل على تعظيم الاختلاف أو الفروق بين المجموعات بأقلّ خطأ للتصنيف

إذْ هناك أهداف عدّة للتحليل التمييزي أهمها :

1_ إنشاء دوال تمييزية للفصل أو التمييز بين فئات المتغيّر التابع.

2_ تعمل هذه الدوال على تعظيم الفروق بين المجموعات (فئات المتغير- التابع).

3_ترتيب المتغيّرات التي تسهم بقدر كبير في التمييز أو توضيح الاختلافات بين المجموعات (للمتغيّر -التابع).

4_تصنيف المشاهدات الجديدة وتوزيعها على مجموعات (للمتغيّر التابع).

5_ الوصول إلى أقلّ نسبة خطأ للتوصيف _ تقييم دقة التصنيف كنسبة مئوية.

وبعد إنشاء أو إيجاد الدوال المميزة يجب تديدها من خلال معرفة هل هي تابعة إلى إحدى أنواع التحليل التمييزي الآتية :

1_التحليل التميزي المباشر Direct Discriminate analysis : إذْ تدخل المتغيّرات إلى التحليل دفعة واحدة دون إعطاء أيْ أهمّية لأيْ متغيّر .

2_ التحليل التمييزي الهرمي Hierarchical Discriminate analysis : ويَتَمّ فيها إدخال المتغيّرات حسب رؤية الباحث.

3_التحليل التمييزي المتدرّج Stepwise Discriminate analysis:يَتَمّ إدخال المتغيّرات للتحليل حسب معيار إحصائي يحدد اولوية ادخال المتغيّرات إلى النموذج إذْ يَتَمّ إضافة المتغيّرات إلى الدوال التمييزية وأحد تلو الاخر حتى نجد أن إضافة متغيرات لا يعطي تمييزا افضل .

ويَتَمّ إيجاد الدالّة التمييزية على تصنيف البيانات والفئات إلى مجموعتين أساس وذلك بالاعتماد على مجموعة من المتغيّرات أو العوامل وتعمل الدالّة على زيادة درجة التجانس بين مفردات المجموعة الواحدة وتقليل درجة التجانس بين المجموعتين ومِنْ ثَمّ تسهيل إمكانية تصنيف أيْ مشاهدة جديدة إلى إحدى المجموعتين بأقلّ خطأ تصنيف كما تعمل الدالّة على استبعاد المتغيّرات التي ليس لها تأثير معنوي في التمييز والفصل بين المجموعتين.

وتكون جاهزة للتطبيق عند أخذ عينة عشوائية مجموعها N1 من المجتمع الأول وN2 من المجتمع الثاني ، نحدد K من الخواص التي نرى أنّها ذات تأثير في التمييز بين مجتمعين (k) ثم نقيس كُلّ وحدة من الوحدات العينتين من إذْ الخواص فتكون القياسات الناتجة وعددها (N1+N2)K هي الأساس التي تعتمد عليه دالّة التمييز .

إنّ جدول البيانات لهذا التحليل سيكون بالشكّل الآتي ، علما بأنّ Xi هي متغيرات مستقلّة وهذه المتغيّرات هي $(X_1, X_2, X_3, \dots \dots . . X_k)$.

الجدول رقم (1) يمثّل تحليل المتغيّرات للمجموعتين

المجموعة (2)				**المجموعة (1)**			
X_k	…	X_2	X_1	X_k	…	X_2	X_1
X_{1k}	…	X_{12}	X_{11}	X_{1k}	…	X_{12}	X_{11}
X_{2k}	…	X_{22}	X_{21}	X_{2k}	…	X_{22}	X_{21}
:	:	:	:	:	:	:	:
X_{nk}	…	X_{n2}	X_{n1}	X_{nk}	…	X_{n2}	X_{n1}

دالّة التمييز عبارة عن معادلة خطّية تحتوي على المتغيّرات المستقلّة الموجودة في المسألة ويمكن توضيح خطوات حساب الدالّة كما يأتي

*- نحسب متوسّطات المتغيّرات في كُلّ مجموعة ثم نجد الفرق بينهما كما يأتي :

المجموعة الاولى: [7,1]

$$\bar{X}_1^{(1)} = \frac{\sum_{i=1}^{n_1} x_{1i}}{n_1} \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots . (1)$$

$$\bar{X}_2^{(1)} = \frac{\sum_{i=1}^{n_1} x_{2i}}{n_1} \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots . (2)$$

$$\bar{X}_k^{(1)} = \frac{\sum_{i=1}^{n_1} x_{ki}}{n_1} \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots . (3)$$

$$\bar{X}_i^{(1)} \rightarrow i = 1, 2, 3, \dots \dots .., k$$

المجموعة الثانية:

$$\bar{X}_1^{(2)} = \frac{\sum_{i=1}^{n_2} x_{1i}}{n_2} \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots . (4)$$

$$\bar{X}_2^{(2)} = \frac{\sum_{i=1}^{n_2} x_{2i}}{n_2} \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots . (5)$$

:

$$\bar{X}_k^{(2)} = \frac{\sum_{i=1}^{n_2} x_{ki}}{n_2} \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots . (6)$$

$$\bar{X}_i^{(2)} \rightarrow i = 1, 2, 3, \dots \dots .., k$$

إذْ:

$\bar{X}_i^{(2)}$ تمثّل متوسّط المتغيّر i في المجموعة الثانية.

الفرق بين متوسّطي كُلّ متغيّر يحسب كالاتي:[8,1]

$$d_1 = \bar{X}_1^{(1)} - \bar{X}_1^{(2)} \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots . (7)$$

$$d_2 = \bar{X}_2^{(1)} - \bar{X}_2^{(2)} \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots . (8)$$

$$d_k = \bar{X}_k^{(1)} - \bar{X}_k^{(2)} \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots . (9)$$

بشكُلّ عام فإنّ:

$$d_j = \bar{X}_j^{(1)} - \bar{X}_j^{(2)} \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots . (10)$$

إذْ :

$$j = 1, 2, 3, \dots \dots .., k$$

في المجموعتين تَمّ توضيح هذه الفروق في شكُلّ متجه عمودي يمثّل الفرق بين متوسطي المتغيّر ويرمز له بالرمز d

$$d = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix}$$

ثانياً:

نحسب مجموع مربعات كُلّ متغيّر في كُلّ مجموعة وكذلك مجموع حاصل ضرب كُلّ متغيرين في كُلّ مجموعة كالاتي:

$$SS_{X_1}^{(1)} = (S_{x1\,x2}) = S_{11}^{(1)} = \sum_{i=1}^{n_1} x_{1i}^{n_1} - \frac{\left[\sum_{i=1}^{n_1} x_{1i}\right]^2}{n_1} = \sum_{i=1}^{n_1} (x_{1i} - \bar{x})^2 \dots\dots\dots (11)$$

عبارة عن مجموع مربعات المتغيّر الأول في المجموعة الأولى وبصورة عامة فأن:

$$S_{kk}^{(1)} = \sum_{i=1}^{n_1} x_{k1}^2 - \frac{\left[\sum_{i=1}^{n_1} x_{ki}\right]^2}{n_1} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (12)$$

وهذا يمثّل مجموعة مربعات المتغيّر الأخير في المجموعة الأولى، وبصورة عامة فأن:

$$S_{jj}^{(1)} = \sum_{i=1}^{n_1} x_{j1}^2 - \frac{\left[\sum_{i=1}^{n_1} x_{ji}\right]^2}{n_1} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (13)$$

إذْ:

$$j = 1, 2, 3, \dots\dots\dots, k$$

وبنفس الأسلوب يمكن كتابة الصيغ الرياضية المشابهة للصيغ السابقة للمجموعة الثانية مع تغير الرمز (1) إلى الرمز (2) و n_1 الى n_2

أمّا مجموعة حاصل ضرب كُلّ متغيرين في المجموعة الأولى يحسب كالآتي:[7,3,1]

$$(S_{x1\,x2})^1 = S_{12}^{(1)} = \sum_{i=1}^{n_1} (x_{1i} - \bar{x}_1)\,(x_{2i} - \bar{x}_1) \dots\dots\dots\dots\dots\dots\dots\dots (.14)$$

$$(S_{x1\,x2})^1 = S_{12}^{(1)} = \sum_{i=1}^{n_1} x_{1i}\,x_{2i} - \frac{\sum_{i=1}^{n_1} x_{1i}\ \sum_{i=1}^{n_1} x_{1i}}{n_1} \dots\dots\dots\dots\dots\dots\dots\dots (15)$$

وهذا يمثّل مجموع حاصل ضرب المتغيرين الأول والثاني في المجموعة الأولى.

وبصورة عامة:

$$S_{ij}^{(1)} = \sum_{i=1}^{n_1} x_{ji}\, x_{ij} - \frac{\sum_{i=1}^{n_1} x_{ij}\ \sum_{i=2}^{n_1} x_{ji}}{n_1} \dots\dots\dots\dots\dots\dots\dots\dots. (16)$$

وبنفس الطريقة يمكن كتابة الصيغ الرياضية للمجموعة الثانية مع تغيير الرمز (1) إلى الرمز (2) و n_2 الى n_1

علماً بأَنّ مجاميع حاصل ضرب كُلّ متغيرين مختلفين في المجموعة هو: [8,6]

$$\boldsymbol{C_2^k = \frac{k!}{2!\,(k-2)!}} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \boldsymbol{(17)}$$

ثالثاً:

حسب التباينات المجمعة أو المدمجة لكُلّ متغيّر بين المجموعتين، وكذلك حساب التغايرات المشتركة المجمعة بين كُلّ متغيرين وبين المجموعتين كالآتي:

$$\boldsymbol{V_{11} = \frac{s_{11}^{(1)} + s_{11}^{(2)}}{n_1 + n_2 - 2}} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \boldsymbol{(18)}$$

يمثّل تباين المجمع للمتغيّر الأول بين المجموعتين وبصورة عامة فإنّ:

$$\boldsymbol{V_{11} = \frac{s_{jj}^{(1)} + s_{jj}^{(2)}}{n_1 + n_2 - 2}} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \boldsymbol{(19)}$$

إذْ:

$$J = 1, 2, 3, \dots\dots.., k$$

يمثّل التباين المجمع للمتغير

أمّا التغايرات المشتركة المجمعة فتحسب كما في الصيغة العامّة

$$\boldsymbol{V_{12} = \frac{s_{12}^{(1)} + s_{12}^{(2)}}{n_1 + n_2 - 2}} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \boldsymbol{(20)}$$

يمثّل التغاير المشترك المجمع للمتغيرين الأول والثاني بين المجموعتين.

$$\boldsymbol{V_{jj} = \frac{s_{jj}^{(1)} + s_{jj}^{(2)}}{n_1 + n_2 - 2}} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \boldsymbol{(21)}$$

إذْ $\boldsymbol{j \neq j}$

يمثّل التغايير المشترك المجمع للمتغيرين j و j بين المجموعتين

توضّح هذه النتائج في مصفوفة يرمز لها بالرمز V لتمثّل مصفوفة التباينات والتغايرات المشتركة وتأخذ الشكْل الآتي: [8,6]

$$V = \begin{bmatrix} v_{11} & v_{12} & v_{1k} \\ v_{21} & v_{22} & v_{2k} \\ v_{k1} & v_{k2} & v_{kk} \end{bmatrix}$$

6- خصائص مصفوفة التباين والتغاير المشترك[13,9,1]

1_مصفوفة مربعة عدد صفوفها يساوي عدد أعمدتها.

2_مصفوفة متماثلة بمعنى أنّ العنصر في الصف رقم 1 والعمود رقم 1 يساوي العنصر في الصف رقم 1 والعمود رقم 1.

3_ التباينات عبارة عن عناصر القطر الرئيس للمتغيّرات على التوالي ولذلك يجب أن تكون موجبة وباقي العناصر تمثّل التغايرات المشتركة.

4_مصفوفة غير شاذة أيْ محدّدها لا يساوي صفر.

$$\hat{a} = v^{-1}d$$

$$\hat{a} = \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \end{bmatrix}$$

نحصل على معادلة خطّية تمثّل دالّة التمييز وتأخذ الشكْل الاتي:

$$\hat{L}_i = \hat{a}_1 x_{11} + \hat{a}_2 x_{21} + \ \dots\dots\dots\dots + \ \hat{a}_k x_{ki} \ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (22)$$

إذْ :

$$i = 1, 2, 3, \dots\dots\dots, n$$

تمثّل هذه المعادلة بيانات المجموعتين يعني أنّه عند التعويض بالقيم المستغلّة من للمتغيرات لغرض إيجاد القيم المميزة نحصل على القيم التمييزية لتلك المفردة. [6,1]

المشترك في المتجه d وكذلك كالاتي:

لإمكانية التحليل الإحصائي سوف نضع الافتراضات الثلاثة الآتية :

وبعد ذلك نضرب معكوس مصفوفة التباين والتغاير المشترك في المتجه d وذلك كالآتي:

أولاً: دالّة التمييز خطّية على الصورة

$$\hat{L}_i = \hat{a}_1 x_{11} + \hat{a}_2 x_{21} + \dots\dots\dots\dots + \hat{a}_k x_{ki} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (23)$$

إذْ إنّ $\hat{a}_n$ ، ... ... ، $\hat{a}_2$ ، $\hat{a}_1$ معلمات مجهولة مطلوب تقديرها من العينتين بشرط أن نحصل على هذه المتغيّرات على أقلّ احتمال خطأ التقسيم.

ثانياً: التغيرات هي متغيرات عشوائية توزيعها المشترك هو توزيع معتدل (ذو k من المتغيرات)سواء في المجتمع A أو المجتمع B . من هذا الافتراض ينتج أنّ دالّة التمييز تكون ذات توزيع معتدل لأنّها تتضمن اعتدالية كُلّ متغيّر على حدا.

ثالثاً: العينات التي تأخذ من كُلّ مجتمع هي عينات عشوائية . [6,1]

7- تعريف مشكلة التعدّد الخطي:-[14,2]

مشكلة التعدّد الخطّي تظهر فقط عندما توجد علاقة خطّية بين بعض أو كُلّ المتغيّرات وان الارتباطات بين هذه المتغيّرات يعرف بالتعدّد الخطي، إنّ الرتبة من أهّم شروط انموذج الانحدار هو:

$$RANK\ (X) = P \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (24)$$

إذْ إنّ:

X: مصفوفة ابعادها (n*p) للمشاهدات التوضيحية.

عندما تكون المتغيّرات التوضيحية (explanatory variables) مستقلّة خطياً (linearly in dependent) فأنه يمكن إيجاد معكوس مصفوفة المعلومات (X`X) ومِنْ ثَمّ يمكن إيجاد مقدرات المربعات الصغرى الاعتيادية (β) (Ordinary least squares) (OLS). وعند وجود علاقة خطّية تامة بين اثنين أو اكثر من المتغيّرات التوضيحية فإنّ ذلك سيؤدّي إلى انتهاك شرط الرتبة، أيْ إنّ: [14,2]

$$RANK\ (X) < P \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (25)$$

عليه فإنّه لا يمكن إيجاد معكوس مصفوفة المعلومات (X'X) ومِنْ ثَمّ لا توجد امكانية لتقدير معلمات النموذج (β)باستعمال طريقة (OLS) ، وهذه الحالة تسمّى بالتعدّد الخطّي التام (perfect multicollinearity) ، ولغرض معالجة هذه المشكلة هناك عدة طرق احداها حذف متغيرات التوضيحية المسببة للتعدّد الخطّي ومن ثم تقدير الانموذج.

أمّا إذا كان محدد مصفوفة المعلومات لا يساوي صفر وانما قريب منه (أيْ إنّ 0≠ |X'X |) ، وتظهر هذه الحالة عندما تميل المتغيّرات للتحرك سوية بالزيادة أو الانخفاض أو في حالة استعمال المتغيّرات المرتدة زمنيا (Lagged variables) ، ففي هذا الحالة يمكن تقدير معلمات الانموذج ولكن هذه التقديرات سوف تكون غير دقيقة وغير ممثلة لواقع المشكلة المدروسة بسبب أنّ تباينات المعلمات المقدرة ستكون كبيرة جدا ومِنْ ثَمّ سيظهر اختبار T عدم معنوية المعلمات بسبب ارتباط المتغيّرات ببعضها ببعض

8- كشف مشكلة التعدّد الخطّي [11,2]

كشف مشكلة التعدّد الخطّي تعني كشف درجة التعدّد الخطّي وليس الغرض منه هو وجود أو عدم وجود التعدّد الخطّي ، ولغرض الكشف عن التعدّد الخطّي في النموذج يحتوي على اكثر من متغيرين توضيحين لابُدّ من الاخذ بالحسبان النقاط الآتية :

1- ارتفاع قيمة R^2 وعدم معنوية المعلمات المقدرة للمتغيرات التوضيحية[11,2]

2 -ارتفاع معاملات الارتباط بين المتغيّرات التوضيحية بما أنّه لا يمكن أنْ يعول ع معاملات الارتباط بين أزواج المتغيّرات في تشخيص مشكلة التعدّد الخطّي لأنّ العلاقة المتبادلة بين 3 متغيرات أو أكثر قد يؤديان إلى درجة عالية من التعدّد الخطّي ع الرقم من أنّه الارتباطات بين أزواج المتغيّرات منخفضه. لذا فأنّ الاجراء الأفضل في قياس درجة التعدّد الخطّي هو احتساب الجذور المميزة لمصفوفة الارتباط والدليل الشرطي المقابلة لها فضلا عن عوامل تضخّم التباين ونسب تحلّل التباين .

3_ احتساب معامل تضخّم التباين (VIF) (variance inflation factor) لكُلّ من المتغيّرات التوضيحية ، إذْ يستفاد منه في قياس مدى ارتباط كُلّ متغيّر توضيحي مع المتغيّرات الأخرى في الانموذج ، فإذا كانت قيمة $VIF_j>10$ فأنه يدل على وجود مشكلة التعدّد الخطّي للمتغيرات التوضيحية ، وتستعمل الصيغة الاتية في إيجاد قيمة VIF_j

$$\frac{\mathbf{1}}{\left(\mathbf{1}-\boldsymbol{R}_{\boldsymbol{J}}^{\mathbf{2}}\right)} \boldsymbol{VIF_J} = \cdots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots (\mathbf{26})$$

$$J = 1, 2, 3, \ldots \ldots .., P$$

إذْ إنّ :

P: تمثّل عدد المتغيّرات.

R2j: تمثّل معامل التحديد للمتغيّر التوضيحي Xj المستخرج من انحدار Xj على بقية المتغيّرات التوضيحية

4_ إيجاد الدليل الشرطي (Cl) ونسب التباين (proportion variance) ، إذْ يؤخذان سوية ويستفاد منهما في بيان درجة التعدّد الخطّي والمتغيّرات المرتبطة مع بعضها البعض. فإذا كانت قيمة الدليل الشرطي بحدود 5_10 وان اثنين أو أكثر من

نسب التباين أقلّ من 0.50 فأنه يدل على ان الارتباط ضعيف. أمّا إذا كانت بين 10_30 فهذا يعني أن هناك تعدّداً خطّياً من المعتدل إلى العالي ، أمّا إذا تجاوزت 30 فهذا يدل على ان التعدّد الخطّي بدرجة أكبر .

ولغرض إيجاد الدليل الشرطي لابُدّ اولا من احتساب الجذور المميزة (إذْ توضع كميه الاختلافات الكلية بين المتغيرات)، فعندما تكون الجذور المميزة مساويه للصفر فانه يدل ع التعدّد الخطّي التام، أمّا إذا كانت قريبة من الصفر فهذا مؤشر ع وجود تعدّد خطّي عالي، أمّا إذا كانت مساوية إلى وأحد فتُعدّ الحالة الأمثل في عدم وجود التعدّد الخطي. ومِنْ ثَمّ يمكن إيجاد العدد الشرطي (k) بالصيغة الآتية:[14,2]

$$k = \frac{Maximum\ eigenvalue}{Minimum\ eigenvalue} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \ (27)$$

ومن ثم فإنّ الدليل الشرطي (CI) يعرف بالصيغة الآتية:

$$CI = \sqrt{\frac{Maximum\ eigenvalue}{Minimum\ eigenvalue}} = \sqrt{k} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \ (28)$$

لتوضيح طريقة تحلل نسب التباين (proportion variance-decomposition method) ليكن لدينا أنموذج الانحدار الآتي:

$$y = X\beta + \varepsilon \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots .. \ (29)$$

إذْ إنّ :

موجة ذو (n*1) لمشاهدات متغيّر معتمد(y)

موجة ذو (p*1) للمعلمات المطلوب تقديرها، علما بأنّ العنصر الأول منه يمثّل الحدّ الثابت للأنموذج قيد الدرس.(β)

موجة ذو (n*1) الأخطاء العشوائية (Σ)

مصفوفة ذات أبعاد (n*p)لمشاهدات المتغيّرات التوضيحية علما أن العمود الأول من هذا المصفوفة يمثّل الحد الثابت(X) .

كما يمكن كتابة هذه المصفوفة بالشكُلّ الاتي

$$\boldsymbol{X = \ UMV'} \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \ (\boldsymbol{30})$$

إذْ إنّ

مصفوفة تحتوي على الجذور المميزة , (V) كما يمكن كتابة , vij=V إذْ تمثّل تباين العنصر (i) من المعلمة. b وإنّ

$$U'U = V'V = Ip \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.. (31)$$

وإنّ m مصفوفة قطرية (diagonal matrix) قطرها الجذور المميزة (بمعنى ان القطر هو (λ_p .... $\lambda_1\ \lambda_2$)) ، أي

$$M = diag(\lambda_1, \ldots\ldots\ldots\ldots, \lambda_p) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (32)$$

كذلك يمكن القول إنّ

$$X'X = VM^2\, V' \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.. (33)$$

ولغرض تحليل التباين فمن المعلوم أن تباين المعلمات المقدر هو

$$Var\,(\hat{b}) = \hat{\sigma}^2(X'X)^{-1} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.. (34)$$

إذْ إنّ $\hat{\sigma}^2$ تباين الخطأ العشوائي المقدر.

أذن

$$Var\,(b) = \hat{\sigma}^2 V M^{-1} V' \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.. (35)$$

وعليه فإنّ موجه معلمات الانحدار يمكن أنْ يحلل إلى مجموع P من المكوّنات وكُلّ منها مرتبط بجذر ميز وأحد وكالآتي:

$$Var\,(\hat{b}_k) = \hat{\sigma}^2 \sum_{G=1}^{P} \frac{V_{IJ}^2}{\lambda_J} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.. (36)$$

بما أنّ الجذور المميزة تظهر في المقام، فإنّ مكوّنات التباين ستكون معتمدة علية، وإذا ما كانت الجذور المميزة صغيرة فان مكوّنات التباين ستكون أكبر نسبيا بالمقارنة مع المكوّنات الأخرى. وهكذا فأن النسبة العالية لاثنين أو اكثر من معلمات التي ارتبطت بالجذر المميز الصغير نفسه تعطي دليلا على وجود مشكلة تعدّد خطّي بسبب الاعتماد المتبادل للمتغيرات. [12,10]

لتكن

$$\phi_{kj} = \frac{v_{ki}^2}{\lambda_j} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.. (37)$$

وإنّ

$$\phi_k = \sum_{j=1}^{p} \phi_{kj} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots .. (38)$$

إذن

$$Var\ (b_k) = \hat{\sigma}^2 \phi_k \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots .\dots .. (39)$$

وإنّ نسب تحلّل التباين هي

$$\pi_{jk} = \frac{\phi_{kj}}{\phi_k} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots .. (40)$$

الصيغة الأخيرة هي التي ستظهر أيْ المتغيّرات تسبّب في مشكلة تعدّد خطّي والتي ممكن ترتيبها بالجدول الآتي:

Associated singular value eigen values	*Variance proportion*			
	$Var\ (b_1)$	$Var\ (b_2)$	……………..	$Var\ (b_k)$
λ_1	π_{11}	π_{12}	……………..	π_{1p}
λ_2	π_{21}	π_{22}	……………..	π_{2p}
:	:	:	:	:
:	:	:	:	:
λ_p	π_{p1}	π_{p2}	……………..	π_{pp}

يفضل للحكم على التعدّد الخطّي بأنْ يَتَمّ تحويل المتغيّرات إلى الصيغة المعيارية (*Standarized*) بمعنى $Z_i = \frac{(X_i - \bar{X})}{S_{X_1}}$ [12,10].

9- طريقة انحدار المربعات الصغرى الجزئية partial least square Regression (PLSR

إنّ طريقة انحدار المربعات الصغرى الجزئية (partial least square) والتي يرمز لها بالرمز PLS هي تقنيه عُمّمَتْ خصائصها من خلال تحليل المركبات الرئيسية(principal components Analysis)والتي يرمز لها بالرمز PCAوالانحدار المتعدّد (Multiple Regression)هذا إذا كان لدينا متغيّر الاستجابة وأحد ، كذلك إذا كان لدينا عدة متغيرات استجابة فإنّ طريقة انحدار المربعات الصغرى الجزئية مفيدة عندما نحتاج إلى التنبُؤ لمجموعة من المتغيّرات وإنّ طريقة المربعات الصغرى الجزئية تعتمد على خطوتين اساسيتين الأولى هي إيجاد المتغيّرات الكامنة (latent variable) بين Y و X من خلال تعظيم مصفوفة التباين والتباين المشترك والخطوة الثانية هي انحدار Y على المركبات t ، توجد عدة خوارزميات تتعلق بالمربعات الصغرى الجزئية ومن الخوارزميات التي تَمّ الاعتماد عليها هي (PI.S2 , PI.S1) NIPALS،إذْ أنّ PLS1 تستعمل في حالة متغيّر الاستجابة وأحد أمّا PLS2 تستعمل في حالة وجود عدة متغيرات استجابة وهي من أول الخوارزميات ثم تليها خوارزمية Kernel إذْ الاثنان تعطيان النتائج نفسها لكن الفرق هو في كيفية حساب المركبات ففي الأول يَتَمّ بصوره تكرارية أمّا الخوارزمية الأخرى يَتَمّ عن طريق إيجاد المتجهات الذاتية .

نفرض لدينا المصفوفة xn.p والمتجه Yn.1 فطريقة المربعات الصغرى الجزئية تعتمد على النموذج الثاني بينY و X وكالآتي :

$$X=TP'+E$$

$$Y=Uq'+f$$

إذْ

T=مصفوفة_ مصفوفة درجات x-score ذات رتبة nxr إذْ ان T هي مصفوفة متماثلة symmetric

U=_مصفوفه درجات Y-score ذات رتبة nxr

P'=_مصفوفة تحميلات x-loading ذات رتبة pxr

q'=_متجه تحميلات Y-loading 1 بعدxr

E=_مصفوفة البواقي x-residual ذات رتبة nxp

F= متجه البواقي _ Y-residual بعد nx1

المصفوفة p' والمتجة q' له r من الاعمدة وهو محدد بما ياتي :

$$(r < \min (p,n))$$

إذْ إنّ :

P: عدد المتغيّرات

n: عدد المشاهدات

r: عدد المركبات

والعلاقة الداخلية التي تربط بين scores تعطى كالآتي [10]

$$\boldsymbol{U} = \boldsymbol{TD} + \boldsymbol{H} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \ (\mathbf{41})$$

إذْ Dمصفوفة قطرية ذات رتبةrxr

Hمصفوفة البواقي ذات رتبهnxr

الفكرة الاساسية في المربعات الصغرى الجزئية هو في كيفية إيجاد المتجه w من مجال X والمتجه ci من مجال Y بحيث إن

Max COV (Xw, Yc) With

$$\mathbf{t\ _Xw_1\ and\ u\ _Yc\ \ _\ 1} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \ (\mathbf{42})$$

إذْ إنّ Max COVهو تقدير التباين المشترك

وان t,u هي أعمدة في المصفوفتين T,U ويَتّمّ تنفيذ التكرارات بطريقة متسلسلة وهذا يعني ان المتجهات scores يَتّمّ احتسابها الواحد بعد الاخر حتى يَتّمّ استخراج كافة المتجهات إلى rتحت القيد عدم الارتباط بين المتجهات وتوجد طرائق عدة لحل المعادلة (41) منها خوارزمية NIPALS وغيرها من الخوارزميات وفي هذا البحث تَمّ الاعتماد على الخوارزمية التكرار غير الخطّي للمربعات الصغرى الجزئية[14,11]

فإذا كان لدينا الثوابت $C_1 = C_2 = \cdots = C_P$ تحقق المعادلة المذكورة انفاً في حين إذا كانت $C_1 = 0$ فيمكن إيجاد أيْ متغيّر توضيحي أو متجه عبارة عن تركيبة خطّية من المتجهات الأخرى وهذا يدل على التعدّد الخطّي أيْ بمعنى متجه بدلالة المتجهات الأخرى وفق المعادلة الاتية [10]

$$X_j = \frac{1}{C_1}\left[C_1X_1 + C_2X_2 + \cdots + C_PX_P\right] \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \ (43)$$

يقسم التعدّد الخطّي إلى :

أولا: التعدّد الخطّي التام: يحدث بين المتغيّرات التفسيرية إذا توفر أحد الشروط الاتية

1- الثوابت في المعادلة رقم 2 المتمثلة في $C_1 = C_2 = \cdots = C_P$ على الأقلّ أحدهما لا يساوي صفرا

2- المصفوفة الخاصة بالمعلومات $\grave{X}X$ لها محدد يساوي إلى الصفر $\left|\grave{X}X\right| = 0$ ومِنْ ثَمّ لا يمكن إيجاد مقدرات المعلمات لأنّه لا يمكن إيجاد معكوس المصفوفة.

ثانياً: التعدّد الخطّي شبه التام هذا النوع من التعدّد يحدث إذا توفر أحد الشروط الاتية [6,2]

1- إنّ المتغيّرات التفسيرية دالّة على متغيرات تفسيرية أخرى بشكُلّ تركيب خطّية كما في المعادلة رقم 2 وبوجود قيم عشوائية ε_i تحقق المعادلة الآتية:

$$a_1X_1 + a_2X_2 + \cdots + a_PX_P + \varepsilon_i = 0 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (44)$$

عندما a_1 , a_P ثوابت

2- محدد مصفوفة المعلومات $\grave{X}X$ صغيرا جدا أو قريب من الصفر $\left|\grave{X}X\right| \approx 0$ لذا يمكن إيجاد مقدرات المعلمات التي تكون كبيرة جدا وغير دقيقة لعدم دقة معكوس المصفوفة.

10- اختبارات التحليل التمييزي Discriminate analysis tests [11,1]

عندما يراد التمييز بين مجموعتين ،فإنّه يمكننا أنْ نختبر الفرضية التي تنصّ على تساوي متوسّطات المجموعتين .

$$H_0 = \mu_1 = \mu_2$$

$$H_1 = \mu_1 \neq \mu_2$$

وأنْ نختبر فرضية تساوي مصفوفة التباين والتباين المشترك للمجموعتين

$$H_0 = \sum\nolimits_1 = \sum\nolimits_2$$

$$H_1 = \sum\nolimits_1 \neq \sum\nolimits_2$$

بعد تقدير النموذج التمييزي، لابُدّ من اختبار معنويته، واختبار المتغيّرات الداخلة في النموذج والتي يجب أن تبقى في النموذج وأيْ منها غير معنوية ويجب أن تحذف.

وهذه الاختبارات هي:

1- اختبار F_Test F .

2- اختبار ويلكس لامبدا Wilkes's Lambda Test .

3- اختبار هوتلين Hoteling _Lawley Test (T2).

4- اختبار بيللأيْ بارتليت Pillai_Bartlett(V).

1-10 اختبار F_Test F [13,3]

source	Ss	Df	Ms	F
بين المجموعات Between x s	SSB	k-1	M_{SB}	M_{SB}
الخطأ Within x s	SSE	n-k	M_{SE}	M_{SE}
الكلي Total	SST	n-1		

إذ إنّ: -

1- مجموع مربعات الأخطاء يحسب كالآتي : -

$$SSE = D^2 = \hat{\alpha}_1 \quad d_1 + \hat{\alpha}_2 \quad d_2 + \cdots \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots + \hat{\alpha}_k \quad d_k$$

2- مجموع مربعات الكلّي: -

$$SSB = \frac{n_1 n_2}{(n_1 + n_2)(n_1 + n_2 - 2)} * (D^2)^2$$

3- مجموع مربعات الكلي: -

$$SST = SSB + SSE$$

وذلك لاختبار قدرة الدالّة على التمييز وعن طريق الفرضية التي تنص على أن الدالّة ليس لديها القدرة على التمييز (H_0) ضد الدالّة لديها قدرة على التمييز (H_1) ويعتمد هذا الاختبار على قياس الاختلافات بين المجموعات وداخل المجموعات بين المفردات ويَتَمّ ذلك من خلال تكوين جدول تحليل التباين الآتي :- ويَتَمّ الاختبار كالآتي :-

1- صياغة الفروض:

H_0 : الدالّة ليس لها قدرة على التمييز

H_1 : الدالّة لها قدرة على التمييز

2- القيمة المحسوبة: $F = \frac{M_{SB}}{M_{SE}}$

3- القرار:

إذا كانت F المحسوبة اكبر من F الجدولية نرفض الفرض العدمي ونقبل بالفرض البديل ويكون للدالّة قدرة عالية على التمييز والعكس صحيح.

2_10 اختبار ويلكس لامبدا Wilkes's Lambda Test [13,3]

الدالّة ليس لها قدرة على التمييز :

$$H_0 = \mu_1 = \mu_2$$

الدالّة لها قدرة على التمييز :

$$H_1 = \mu_1 \neq \mu_2$$

القيمة المحسوبة: $\boldsymbol{\Lambda} = \prod_{i=1}^{k} \frac{1}{1+\lambda_i}$

λ_i الجذر الكائن (eigenvalues) لكُلّ المتغيرات.

K عدد المتغيرات

القرار : تنحصر قيمة $0 \leq \boldsymbol{\Lambda} \leq 1$

$\boldsymbol{\Lambda} = 1$ معناها تساوي متوسّطات المجموعتين ومِنْ ثَمّ عدم مقدرة الدالّة على التمييز والفصل.

$\boldsymbol{\Lambda} = 0$ معناها عدم تساوي متوسّطات المجموعتين والدالّة لها القدرة عالية على التمييز.

تأخذ الفروض الشكْلَ الآتي :-

إذا اقتربت قيمة A من الواحد دليل على عدم مقدرة الدالّة على التمييز، وإذا اقتربت من الصفر دليل على قدرة الدالّة على التمييز.

وتستعمل إحصائية " وليكس لمدا " لاختبار معنوية متغيرات الانموذج إذْ الإبقاء على المتغيّرات لها ادنى قيمة لإحصائية WILK,S Lambda واعلى قيمة لF

$$F = \frac{P, n_1 + n_2 - P - 1}{(n_1 + n_2 - 2)\ P} = T^2 \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \ (45)$$

بدرجات حرية $(p, n_1 + n_2 - p - 1)$ ، إذْ إنّ p عدد المتغيرات وإنّ n_1, n_2 حجما المجموعتين .

ترفض H_0 بمستوى معنوية H_a كانت:

$$F_{CAL} > Fa(p, n_1 + n_2 - p - 1)$$

ونقبل H_1 وهذا يدل على وجود فرق معنوي ، ومعناه أنّ النموذج التمييزي قابل للتمييز بدرجة عالية.

هو مجموع القيم الخاصة لكُلّ متغيّر للتابع المميز للبيانات، ويكافئ النسبة F لتحليل تباين من ناحية المبدأ، أيْ أنّه مجموع النسب $\frac{SS_R}{SS_T}$ لكُلّ من المتغيّرات ولذلك يمكن مقارنته بالنسبة F في تحليل التباين. ويأخذ الصيغة الآتية :

$$T^2 = \sum_{i=1}^{S} \lambda_1 \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots .. (46)$$

إذْ λ تمثّل القيم الخاصة (الشعاع، الجذر الكامن) لكُلّ من المتغيّرات المميزة و S عدد المتغيرات.

3_10 اختبار هوتلين (Hoteling_ Lawley Test (T2 [14,3]

إحصاء هوتلنج تأخذ الشكُلّ الآتي:-

$$\boldsymbol{T^2} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{s}} \lambda_i$$

إذْ إنّ

λ_i : الجذور المميزة للمتغيرات eigenvalues

S: عدد المتغيرات

وتعادل إحصائية هوتلنج قيمة F من جدول تحليل التباين ويمكن تحويله إلى قيمة لها توزيع F تقريبي صيغته كالآتي :

$$\boldsymbol{F} = \frac{\boldsymbol{n_1 + n_2 - k - 1}}{\boldsymbol{(n_1 + n_2 - 2)k}} * \boldsymbol{T^2} \dots \dots \dots \dots \dots \dots (47)$$

والقيمة الجدولية: $F \propto (k - 1, n_1 + n_2 - k - 1)$

إذا كانت F المحسوبة اكبر من F الجدولية نرفض الفرض العدمي وقبول البديل بأن للدالّة قدرة عالية على التمييز

4_10 اختبار بيللأيْ بارتليت (V)Pillai_Bartlett [14,3]

هو مجموع نسب التباين المبرر في التوابع المميزة ويأخذ الصيغة الآتية:

$$V = \sum_{i=1}^{s} \frac{\lambda_i}{1+\lambda_i} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (48)$$

إذْ λ تمثّل القيم الخاصة (الشعاع) لكُلّ من المتغيّرات المميزة و s عدد المتغيرات. وقد طورت صيغته من قبل Bartlett وتحويله لتوزيع x^2 تقريبي بدرجة حرية p(k-1) إلى الشكْل التالي:

$$x^2 = -\left[N - 1 - \frac{1}{2}(p-k)\right]\log(\mathbf{\Lambda}) \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (49)$$

إذْ إنّ

N : حجم العينة

P : عدد المتغيرات

k : عدد المجموعات

$\mathbf{\Lambda}$: قيمة اختبار ويلكس لمبدا وتحسب من الصيغة (49)

ويستعمل لاختبار تساوي متوسّطات المجموعات، ويمكن الاعتماد عليه في حالة خرق فرضية عدم تساوي مصفوفات التباين والتباين المشترك.

11- معامل التحديد للنموذج التمييزي Coefficient of determination R2 [13,12,4]

يمكن استعمال إحصائية (R2) لاختبار قوة النموذج التمييزي ،أيْ نسبة مساهمة العوامل المؤثّرة التي يتضمنها النموذج المقدر للمتغيّر التابع (الإصابة) من خلال قيمة الجذر الكامن Eigen values وهي نسبة التباين المفسر بين المجموعات والتي تعود إلى الفروق بينها في النموذج التمييزي ، ومعامل الارتباط القانوني Canonical Correlation الذي يقيس الارتباط بين النموذج التمييزي والعوامل المؤثّرة التي تمثّل تمييز الإصابة ، وبتربيع معامل الارتباط القانوني نحصل على قيمة معامل التحديد R2 التي نحصل عليها عندما نجري تحليل التباين الاحادي ، ومِنْ ثَمّ تعود قيمة R2 من التباين إلى الفروق بين المجموعتين في نموذج التمييز ، أيْ هي نسبة ما يبرره النموذج إلى المقدار الذي كان تبريره مطلوبا في البداية.

12- متغيرات الدراسة

أخذت البيانات من مستشفى عام بعقوبة في محافظة ديالى مكونة من عينة (167) شخص ممن يعانون من أمراض القلب واستعمال إحدى البرامج الإحصائية SPSS وتمثلت متغيرات الدراسة بالعوامل المؤثّرة على القلب

1. حالة المريض المتمثّل بالمتغير(Y) : إذْ يكون متحسن (1) أو عدا ذلك (0).
2. نوع المرض المتمثّل بالمتغيّر (X1):إذْ يكون حسب نوع المرض .
   - احتشاء عضلة القلب (جلطة قلبية) (1).
   - ارتفاع ضغط الدم (2).
   - ذبحة صدرية (3).
   - قصور الشرايين التاجية (4).
   - تباطؤ أو تسارع دقات القلب (5).
   - عجز القلب (6).
3. فترة الرقود في المستشفى المتمثّل بالمتغيّر (X2).
4. العمر تمثّل بالمتغيّر (X3).
5. الجنس تمثّل بالمتغيّر (X4) :إذْ يكون ذكر(1) أو انثى (0).
6. الوزن (السمنة) تمثّل بالمتغيّر (X5).
7. التدخين تمثّل بالمتغير(X6) : إذْ يكون نعم (1) أو كلا (0) .
8. السكري تمثّل بالمتغيّر (X7) : إذْ يكون نعم (1) أو كلا (0).
9. الضغط تمثّل بالمتغيّر (X8): إذْ يكون نعم (1) أو كلا (0).

13-التحليل الإحصائي البيانات

بعد إجراء التحليل الأولي للبيانات في برنامج ال SPSS وجد الآتي :

1. عدد جنس المرضى هم (167) بنسبة 100.0% ، وعدد نوع المرض هم (0) بنسبة 0.0% وقت جمع البيانات ما بين الملفات لمرضى رقدوا سابقا في المستشفى (طبلات المرضى) وكما مبين في الجدول رقم (1)

الجدول رقم (1) يبيّن عدد ونسبة جنس المرضى ونوع المرض

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
المريض جنس * المرض نوع	167	100.0%	0	0.0%	167	100.0%

2. أمّا عدد الذكور والإناث في العينة مبين في الجدول رقم (2) الاتي :

الجدول رقم (2) يبيّن عدد الذكور والإناث للعينة

نوع المرض	ذكور	إناث	المجموع
I21	32	24	56
I10	6	6	12
I20	20	10	30
I24	14	6	20
I47	1	0	1
I50	22	26	48
المجموع	95	72	167

1) عدد الذكور (32) وعدد الإناث (24) في المرض I21 (احتشاء عضلة القلب جلطة قلبية)
2) عدد الذكور (6) وعدد الإناث (6) في المرض I10 (ارتفاع ضغط الدم)
3) عدد الذكور (20) وعدد الإناث (10) في المرض I20 (ذبحة صدرية)
4) عدد الذكور (14) وعدد الإناث (6) في المرض I24 (قصور الشرايين التاجية)
5) عدد الذكور (1) وعدد الإناث (0) في المرض I47 (تباطؤ أو تسارع دقات القلب)
6) عدد الذكور (22) وعدد الإناث (26) في المرض I50 (عجز القلب)

14- طريقة المربعات الصغرى الجزئية (partial least square method)

التحليل الإحصائي باستعمال طريقة المربعات الصغرى الجزئية وتكون كالأتي:

1) نقوم بالتأكد من وجود مشكلة التعدّد الخطّي من خلال اختبار VIF كما مبين في الجدول (3)

الجدول رقم (3) يبيّن مشكلة التعدّد الخطي

Coefficients[a]

Model		Collinearity Statistics	
		Tolerance	VIF
1	نوع المرض	0.924	1.082
	فترة الرقود في المستشفى	0.937	1.067
	عمر المريض	0.784	1.275
	جنس المريض	0.969	1.032
	وزن المريض	0.725	1.379
	التدخين	0.618	1.618
	السكري	0.252	5.967
	الضغط	0.220	6.545

a. Dependent Variable: حالة المريض بعد الخروج

من جدول رقم (3) نلاحظ بأنّ قيمة ال VIF > 5 للمتغيرات (السكري _ الضغط) وهذا يبيّن وجود مشكلة تعدّد خطّي فلذلك سوف نلجأ إلى حل هذه المشكلة باستعمال طريقة المربعات الصغرى الجزئية (partial least square method) ولمعالجة مشكلة التعدّد الخطّي نطبق طريقة المربعات الصغرى الجزئية.

الجدول رقم (4) يمثّل نتائج مشكلة التعدّد الخطي

Source	DF	SS	MS	F	P
Regression	2	1.7616	0.880797	4.07	0.019
Residual Error	164	35.4600	0.216219		
Total	166	37.2216			

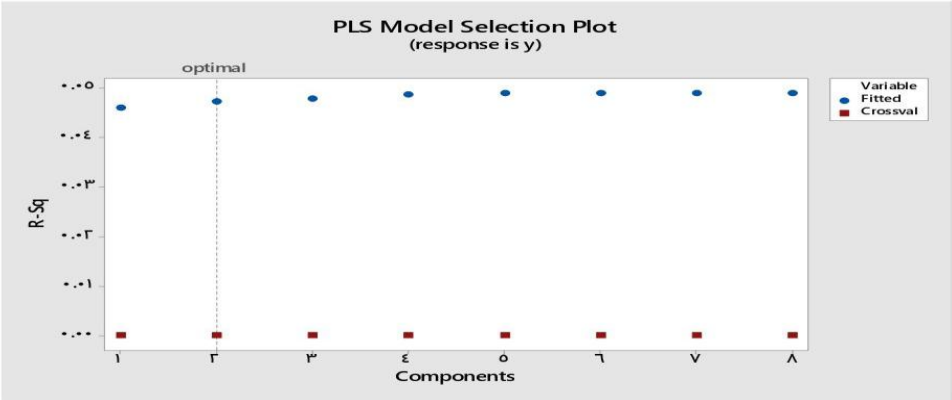
يلاحظ من جدول رقم (4) أنّ متغيّر الاستجابة لتحليل التباين للأنموذج الذي يحتوي على مركبين إذْ وجدت قيمة متغيّر الاستجابة ل P_value المحسوبة هي (0.019) أقلّ من قيمة $\alpha=0.05$ يثبت ذلك ان المركبين في الانموذج معنوية.

2) تَمّ اختبار الانموذج بطريقة cross_ validation لمركبين ويمتلك معامل التحديد R_Sq =0.047 ومعامل تنبؤي 0 وان X_variance تشير إلى مقدار التباين وقد لمركبين 0.361 من تباين المركبات المستقلّة وكما في الشكُلّ رقم (5)

الجدول رقم (5) يبيّن معاملات التحديد ومعاملات التحديد التنبؤية

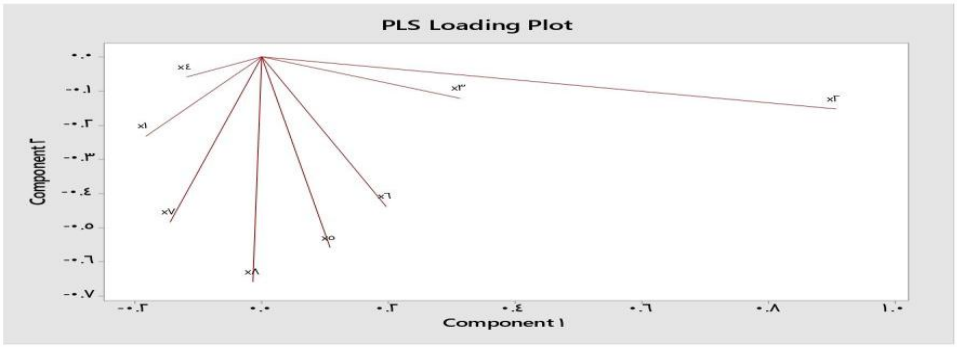
Components	X Variance	Error	R-Sq	PRESS	R-Sq (pred)
1	0.128797	35.5041	0.0461408	39.4377	0
2	0.361277	35.4600	0.0473273	38.9785	0
3		35.4350	0.0479989	39.0977	0
4		35.4044	0.0488213	39.9839	0
5		35.3967	0.0490280	39.6084	0
6		35.3950	0.0490738	39.4733	0
7		35.3949	0.0490741	39.4490	0
8		35.3949	0.0490741	39.4411	0

3) إنّ أفضل نموذج لطريقة المربعات الصغرى الجزئية حسب طريقة العبور الشرعي هل لمركبين إذْ إنّ المحور العمودي يشير إلى قسم معامل التحديد التنبؤي أمّا المحور الأفقي فيشر إلى المركبات إذْ أن القيم المركبات تتزايد حتى المركبة السابقة وبعدها تتناقص وكما في الشكُلّ رقم (1) الآتي :



الشكُلّ رقم (1) يبيّن عدد المركبات المختارة من المتغيّرات المستقلة

4) إنّ الشكُلّ ألبياني لمركبات التحميل (Pls Loading) إذْ إنّ X4 المتمثّل (بالجنس) وX1 (حسب نوع المرض) يمتلك خط قصير جدا وهذا يدل ع امتلاكها X_ Loading منخفضة وليست ذات علاقة مع متغيّر الاستجابة (مرض القلب) وان المتغيّرات X2 هي (فترة الرقود في المستشفى) و X3 تمثّل (العمر) وX5 تمثّل (الوزن) وX6 تمثّل (التدخين) وX7 تمثّل (السكري) و X8 تمثّل (الضغط) تمتلك خطوط طويلة أيْ إنّ لها تحملات عالية أيْ ترتبط بعلاقة معنوية مع متغيّر الاستجابة وكما مبين في الشكُلّ رقم (2) الاتي:

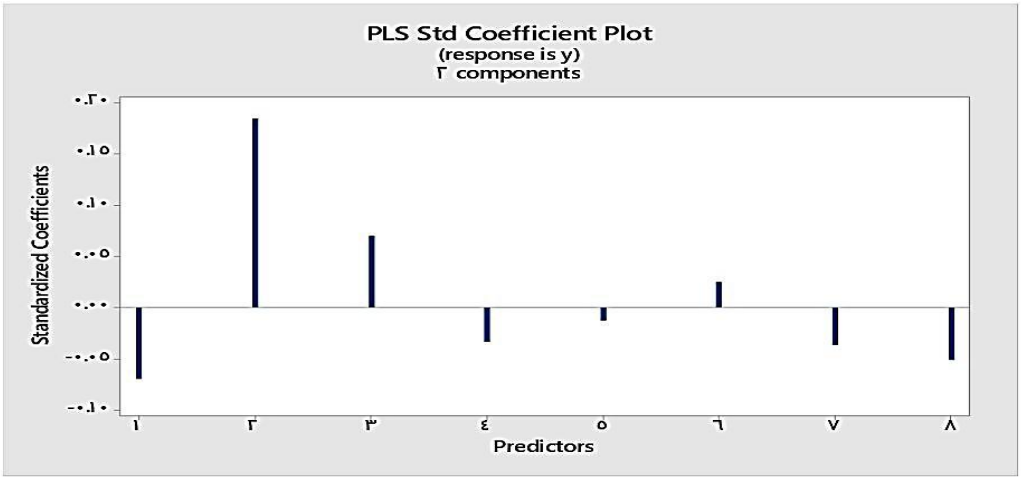


الشكُلّ رقم (2) يبيّن علاقة المتغيّرات المستقلّة بالمتغيّر التوضيحي

5) الشكُلّ رقم (3) يوضّح القيم القياسية المعالم (Standardized Coefficients) وللمتغيرات التوضيحية إذْ نجد أن X2 تمثّل (فترة الرقود في المستشفى) و X1 تمثّل (نوع المرض) و X3 تمثّل (العمر) و X8 تمثّل (الضغط) و X7 تمثّل (السكري

) و X4 تمثّل (الجنس) وتمتلك أكبر معاملات قياسية إذْ إنّ X3 وX2 وX1 وX6 ترتبط بصورة ايجابية مع متغيّر الاستجابة حالة المريض (Y) والمتغيّرات X5 وX4 وX7 و X8

ترتبط بصورة سلبية مع متغيّر الاستجابة.



شكّلّ رقم (3) يبيّن القيم القياسية للمعالم

7_ يكتب الانموذج الرياضي باستعمال طريقة المربعات الصغرى الجزئية كالآتي:

$$Y=0.578 \_ 0.016x1 + 0.047x2 + 0.002x3 \_ 0.031x4 \_ 0.0004x5 +0.026x6 \_0.037x7 \_0.049x8$$

15- الدالّة المميزة الخطّية باستعمال المربعات ألصغرى الجزئية

تَمّ استعمال المربعات الصغرى الجزئية للتخلص من مشكلة التعدّد الخطّي وتَمّ استعمال طريقة العبور الشرعي لتحديد عدد المركبات وقد بلغ عدد المركبات الداخلة في نموذج المربعات الصغرى الجزئية (2) وسيَتَمّ تطبيق الاختبارات الإحصائية الخاصة بالدالّة المميزة ع هذه المركبات وتكون الاختبارات كالآتي

Tests of Equality of Group Means

	Wilks' Lambda	F	df1	df2	Sig.
pc1	0.966	5.840	1	165	0.017
pc2	0.999	0.158	1	165	0.692

الجدول رقم (6) يبيّن معنوية المركبات

في جدول رقم (6) تبين أنّه من خلال اختبار Will's نجد أن المركبة الأولى والثاني لها تأثير معنوي ومهم في تكوين الدالّة المميزة الخطّية ولها تأثير كبير في التفرقة بين المجموعتين .

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	0.039	100.0	100.0	0.193

الجدول رقم (7) يبيّن معنوية الدالّة المميزة الخطية

من جدول رقم(7) تبين انه تَمّ اختبار معنوية الدالّة المميزة الخطّية ووجود قوة التميز من خلال التباين بين المجموعتين والتي قد فسرت 100% من التباين

الجدول رقم (8) يبيّن اختبار Will's Lambda

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	0.963	6.229	2	0.044

من خلال جدول رقم (8) تبين انه تَمّ اختبار وجود علاقة بين المتغيّرات خطّية إذْ إنّ قيمة Will's Lambda = 0.963 ونجد معنوية Chi_square = 6.229 وحيث بلغ قيمة sig.= 0.044 وهي اصغر من قيمة α=0.05 أيْ توجد علاقة خطّية بين المجموعتين في المركبات.

4_ من خلال اختبار معادلة الانحدار الخطّية نجد أنّ معاملات الدالّة المميزة الخطّية المركبات وهي كالآتي:

$$Y=0.997pc1 \_ 0.296pc2$$

	Function
	1
pc1	0.997
pc2	-0.296

الجدول رقم (9) يبيّن معاملات الدالّة المميزة الخطية

_ومن جدول رقم (10) ادناه نجد أن :

$$\hat{M} = 1/2\ (-0.275+0.139).$$

فإذا كانت المفردة أكبر من $\hat{M}$ تصنف على أنها تعود للمجموعة الأولى (وفاة) وإذا كانت المفردة الجديدة أقلّ من $\hat{M}$ تصنف على أنها تعود للمجموعة الثانية

	Function
حالة المريض بعد الخروج	1
وفاة	-0.275
متحسن	0.139

الجدول رقم (10) تمثّل حالة المريض بعد الخروج

			Predicted Group for Analysis		
			وفاة	متحسن	Total
حالة المريض بعد الخروج	وفاة	Count	40	16	56
	متحسن	Count	51	60	111
Total		Count	91	76	167
			54.5%	45.5%	100.0%

الجدول رقم (11) يبيّن نسب التصنيف للدالّة المميز

من جدول رقم (11) نجد أن 40 صنفت بشكُلّ صحيح من 56 أيْ إنّ 54.5% صنفت بشكُلّ صحيح من حالة المريض (وفاة) و60 صنفت بشكُلّ صحيح من 111 أيْ إنّ 45.5% من حالة المريض (متحسن) ، وإنّ نسبة التصنيف الصحيحة 91% وهذا يعني أن احتمال خطأ التصنيف 9% .

16- الاستنتاجات

نستنتج من هذا البحث الآتي

1. نلاحظ أن معامل تضخّم التباين للبيانات قد قل بعد استعمال انحدار المربعات الصغرى الجزئية ولجميع المتغيّرات إذْ أصبحت جميعها أقلّ من (5) وهذا يعني أنه تَمّ التخلص من مشكلة التعدّد الخطي.
2. إنّ النموذج التمييزي ملائم لنمذجة العوامل المؤثّرة على الإصابة بأمراض القلب ، أيْ يمكن استعماله في التمييز وتصنيف المفردات الجديدة إلى مصاب أو غير مصاب وفقا لقيم المتغيّرات المستقلّة للمفردات الجديدة.
3. أن النموذج التمييزي المقدر يمكن أن يستعمل للتنبؤ والتمييز في المجتمع وفقا لاختبار ويلكس لامبدا لجودة الملائمة.
4. دالّة التمييز التي استعملت لها القدرة على التمييز بين مجموعتين (المصابين وغير المصابين).

17- التوصيات

بناءه على ما أستنتج في البحث نوصي بالآتي:

1. استعمال التحليل التمييزي بشكُلّ اوسع ولإيجاد أو تحليل تعدّد المتغيّرات
2. اعتماد الدالّة التمييزية كأداة دقيقة للتصنيف وعمل حزم متخصصة فيها .
3. نوصي الي استعمال طرائق مربعات الصغرى الجزئية للتخلص من مشكلة تعدّد خطّي لأنها عالجت الارتباط بين المتغيّرات التوضيحية.

18 -المصادر

1. البكري ،رباب عبد الرضا (2015) ، مقارنة بعض الطرائق الخطّية لمعالجة مشكلة التعدّد الخطّي في النماذج مع تطبيق عملي ، رسالة دكتوراه ، كلية الإدارة والاقتصاد ، جامعة بغداد.
2. رباب عبد الرضا ، محمد شاكر محمود العزي ،(2017) ، التحليل التمييزي والانحدار اللوجيستي بوجود مشكلة التعدّد الخطّي (دراسة تطبيقية على مرض فقر الدم).
3. عبد الرحيم عوض عبد الخالق بسيوني، استعمال التحليل التمييزي في التصنيف والتنبُؤ (دراسة تطبيقية).
4. حيدر جميل الله محمد ابو دومه، (2019)، استعمال اسلوبي تحليل الانحدار اللوجيستي والتحليل التمييزي للعوامل المؤثّرة على الإصابة بأمراض القلب (دراسة مقارنة مركز جراحة القلب وزراعة الكلى في مستشفى احمد قاسم بالخرطوم البحري 2017 م)، رسالة دكتوراه.
5. دريد حسين بدر، استعمال بعض الطرائق التمييزية الحصينة لتشخيص أمراض القلب في البصرة، جامعة البصرة، كلية الإدارة والاقتصاد، قسم الإحصاء.
6. فريال النويري محمد النويري، (2013)، استعمال الدالّة التمييزية الخطّية لتمييز مرضى السكري المصابين من غير المصابين بالفشل الكلوي بمستشفى الجزيرة لأمراض وجراحة الكلى، ولاية الجزيرة، السودان، كلية الاقتصاد والتنمية الريفية، قسم الإحصاء التطبيقي والديموغرافيا، رسالة ماجستير.
7. جواني حسينة، دراسة مقارنة بين الانحدار اللوجيستي والتحليل التمييزي للتنبؤ بفشل المؤسسات_ دراسة عينة من المؤسسات الاقتصادية لولاية ام البواقي، رسالة ماجستير، جامعة العربي بن مهيدي _ ام البواقي_ كلية العلوم الاقتصادية والتجارية وعلوم التسيير _ قسم العلوم الاقتصادية.

8. كنان احمد علي، (2015) ، فاعلية استعمال التحليل العنقودي والتحليل التمييزي في التحقق من الدلالة التمييزية لاختبارات الذكاء والشخصية (دراسة مبدئية مقارنة في محافظة دمشق) ، رسالة ماجستير ، جامعة دمشق _ كلية التربية_ قسم القياس والتقويم التربوي والنفسي.
9. شمس الدين احمد علي احمد (2013) ، استعمال الدالّة التمييزية الخطّية لدراسة مستوى الإصابة بسرطان الغدة الدرقية ، رسالة ماجستير، جامعة الجزيرة_ كلية العلوم الرياضية والحاسوب _ قسم الإحصاء.
10. Erik Brorson , Asterios Geroukis ,2014 , "A Comparison between discriminant analysis and Logistic regression using principal components " , Department of statistics , Uppsala University , Uppsala University.
11. Ahmed, M. S. ,(1998) " A companion of the Discriminant and Logistic Regression Approach " ph .p thesis,ISSR ,Cairo University.
12. Anderson, T. W., (1984) .An introduction to Multivariate statistical analysis. and Edition, John and Sons , New York, USA .
13. Chandran, R.k., (2009) . The Effectiveness of stepwise Discriminant Analysis. Antimicrob Agents Chemother .2009 July; 53 (7) : 2887 _ 289.
14. Geofry J. M. , (1992) , Discriminant Analysis and statistical pattern Rocognitien the university of Queen slan, Wiley.