



استعمال خوارزمية تجزئة القيم المفردة لمعالجة مشكلة التعدد الخطي (عالية الابعاد) لتحديد وتميز أهم العوامل المؤثرة على امراض القلب

م. م. اياد حبيب شمال/ كلية الإدارة والاقتصاد/ جامعة ديالى/ العراق<sup>(1)</sup>

م. م. احمد رزاق عبد/ كلية الإدارة والاقتصاد/ جامعة ديالى/ العراق<sup>(2)</sup>

<sup>(1)</sup> [ayad@ecomang.uodiyala.edu.iq](mailto:ayad@ecomang.uodiyala.edu.iq)

<sup>(2)</sup> [ahmedrazzaq@uowasit.edu.iq](mailto:ahmedrazzaq@uowasit.edu.iq)

**المقدمة:**

في هذا البحث تم دراسة أحد المشاكل التي تعاني منها البيانات الإحصائية في حالة المتغيرات المستقلة المتعددة وهي مشكلة التعدد الخطي إذ تم معالجة مشكلة التعدد الخطي باستعمال خوارزمية تجزئة القيمة المفردة التي تقوم بتغيير هيكلية البيانات لتخلص من المشكلة مع الحفاظ على طبيعة البيانات من حيث التأثير على المتغير التابع.

تم استعمال التحليل التمييزي والتحليل اللوجستي بعد تطبيق خوارزمية تجزئة القيمة المفردة على بيانات طبيه تمثل حالة البقاء على قيد الحياة وحالة الوفاة لمرض النوبة القلبية ( $y=0$  حالة وفاة ،  $y=1$  حالة يقاء على قيد الحياة ) والعوامل التي تؤثر على مرض النوبة القلبية ( متغيرات مستقلة تعاني مشكلة التعدد الخطي ) وبعد اجراء التحليل التمييزي وجد ان اهم العوامل ذات التأثير العالي على مرض النوبة القلبية هي ( الانفعال ، امراض القلب ، التدخين ، العمر ) وفي حالة التحليل اللوجستي وجد ( الانفعال ، امراض القلب ، والتدخين ، والضغط ، والسكر ، والعمر ) هي التي تؤثر على مرض النوبة القلبية

الكلمات المفتاحية: مشكلة التعدد الخطي، خوارزمية القيمة المفردة، التحليل التمييزي، التحليل اللوجستي، مرض النوبة القلبية.

**Introduction:**

In this paper, one of the problems of statistical data was examined in the case of multicollinearity variables. This is the problem of linear multicollinearity. The problem multicollinearity was solved using a Singular Value Decomposition that changes the structure of the data to eliminate the problem while preserving the nature of the data in terms of the effect on the variable Affiliate.

Comparative analysis and logistic analysis were used to compare the two after the application of a single value fragmentation algorithm to medical data representing recovery status and death of heart attack ( $y = 0$  deaths,  $y = 1$  healing) and factors affecting heart attack After the differential analysis, the most important factors with a high effect on heart attack were (emotion, heart disease, smoking, age). In the case of logistic analysis (emotion, heart disease, smoking, pressure, sugar and age) Affect the heart attack disease.

Keywords: multicollinearity, Singular Value Decomposition, Discriminant Analysis, logistic analysis, heart attack disease

**المبحث الاول الجانب النظري**

**1-1 المقدمة**

ان الاستعمالات العديدة والمتنوعة لعلم الإحصاء في مجالات الحيات المختلفة ومنها المجالات الطبية جعل منه الركيزة الأساسية لحل المشاكل التي تعاني منها البيانات وخاصة في حالة كون البيانات عالية الابعاد (كثيرة المتغيرات المستقلة وكثيرة المشاهدات) تظهر مشكلة التعدد الخطي بين المتغيرات. [ جيريل، (2014) ].

قبل اجراء التحليل الاحصائي لبيانات عالية الابعاد سواء كان التحليل التمييزي Discriminatory analysis او التحليل اللوجستي Logistics Analysis اللذان يتعاملان مع البيانات ثنائية الاستجابة Binary data للمتغير التابع، يجب معالجة مشكلة التعدد الخطي وأحدي الطرق التي تستعمل لحل هذه المشكلة خوارزمية تجزئة القيم المفردة Singular Value Decomposition (SVD) التي تعتمد بالدرجة الأساسية على تغير هيكلية البيانات. [ Baker, (2005) ]

وفي هذا البحث نحاول ان نبين فكره واضحه عن خوارزمية تجزئة القيم المفقودة (SVD) التي تعالج مشكلة التعدد الخطي عالي الابعاد من خلال تطبيق عملي لدراسة عدد الاشخاص الباقين على قيد الحياة من مرض النوبة القلبية كمتغير تابع (متغير استجابة) والعوامل الأخرى التي تؤثر على مرض النوبة القلبية ( المتغيرات المستقلة ) ومن ثم اجراء المقارنة بين التحليل اللوجستي والتحليل التمييزي للبيانات التي جمعها من الأشخاص المصابين بمرض النوبة القلبية في مستشفى الزهراء العام في الكوت باستعمال البرنامج الاحصائي SPSS ولغة برمجة احصائية هي لغة MATLAB ولتحقيق هدف البحث قسم الى اربعة مباحث هي:

- **المبحث الاول:** تضمن منهجية البحث (المستخلص، المقدمة، مشكلة البحث، هدف البحث)
- **المبحث الثاني:** تضمن الجانب النظري الذي تم التطرق فيه حل مشكلة التعدد الخطي عالية الابعاد من خلال خوارزمية تجزئة القيم المفردة والتحليل التمييزي والتحليل اللوجستي.
- **المبحث الثالث:** تضمن الجانب التطبيقي تطبيق بيانات مرض النوبة القلبية على التحليل التمييزي والتحليل اللوجستي واجراء المقارنة بينها.
- **المبحث الرابع:** تضمن الاستنتاجات والتوصيات.

### 1-2مشكلة البحث: research problem

ان التطور العلمي يؤدي بطبيعة الحال الى زيادة الدقة والمعرفة فعند دراسة مرض النوبة القلبية يكون المتغير التابع ثنائي الاستجابة ( 0 = الموت، 1 = البقاء على قيد الحياة) وهناك الكثير من العوامل التي تؤثر عليها تمثل المتغيرات المستقلة فتعاني هذه المتغيرات من مشكلة التعدد الخطي عالية الابعاد التي تؤثر على دقة النتائج في التحليل التمييزي والتحليل اللوجستي عند المقارنة بينهما.

### 3-1 هدف البحث Object of research

يهدف هذا البحث لمعالجة مشكلة التعدد الخطي عالية الابعاد باستعمال خوارزمية تجزئة القيم المفردة (SDV) لبيانات النوبة القلبية في المتغيرات المستقلة، ولتحسين الأساليب الإحصائية المستعملة التحليل التمييزي والتحليل اللوجستي عند المقارنة بينهما لعطاء أفضل وأدق النتائج التي يتم الاعتماد عليها في اتخاذ القرار.

### المبحث الثاني الجانب النظري

#### 1-2 مشكلة التعدد الخطي عالي الابعاد High-dimensional The problem of Multicollinearity

يمكن تعريف مشكلة التعدد الخطي على انها العلاقة بين اثنين او أكثر من المتغيرات التفسيرية على شكل علاقة خطية قوية، وكذلك يمكن وصف مشكلة التعدد الخطي من خلال التعامد عندما تكون المتغيرات متعامدة تكون غير مرتبطة مع بعضها البعض. [ Bager، & Roman & Meshal، (2017) ].

ان احد الشروط التي يجب توفرها في نماذج الانحدار عند تقدير المعلمات الرتب Rank Condition اذ يقال المصفوفة كاملة الرتب اذا كان عدد المتغيرات التفسيرية ( التوضيحية ) للمصفوفة يساوي رتبة المصفوفة بمعنى اخر المصفوفة  $X_{n \times p}$  كاملة الرتب اذا  $\text{rank}(X)=p$ .

ويمكن التعبير عن مشكلة التعدد الخطي عالي الابعاد رياضيا فاذا كانت المتغيرات التفسيرية  $X_1 X_2 \dots X_p$  وجد  $P$  من الثوابت  $C_1 C_2 \dots C_p$  ليست جميعها اصفار وتحقق المعادلة الآتية: [ Bager، & Roman & Meshal، (2017) ].

$$\sum_{j=1}^p C_j X_j = 0 \quad \dots \dots \dots (1)$$

يدل ذلك على ان المتغيرات غير مستقلة خطياً.

فاذا كان لدينا الثوابت  $C_1 = C_2 = \dots = C_p$  تحقق المعادلة أعلاه بينما اذا كانت  $C_j = 0$  فيمكن إيجاد أي متغير توضيحي او متجه عبارة عن تركيبة خطية من المتجهات الأخرى وهذا يدل على التعدد الخطي أي بمعنى متجه بدلالة المتجهات الأخرى وفق المعادلة الآتية<sup>[4]</sup>.

$$X_j = \frac{1}{C_1} [C_1 X_1 + C_2 X_2 + \dots + C_p X_p] \dots \dots \dots (2)$$

يقسم التعدد الخطي الى:

أولاً: التعدد الخطي التام: هذا النوع من التعدد يحدث بين المتغيرات التفسيرية اذا توفر احد الشروط الآتية: [جبريل، (2014)].

- 1- الثوابت في المعادلة رقم 2 المتمثلة في  $C_1 = C_2 = \dots = C_p$  على الأقل احدهما لا يساوي اصفراً
- 2- المصفوفة الخاصة بالمعلومات  $XX$  لها محدد يساوي الى الصفر  $|XX| = 0$  وبالتالي لا يمكن إيجاد مقدرات المعلمات لأنه لا يمكن إيجاد معكوس المصفوفة.

ثانياً: التعدد الخطي شبه التام هذا النوع من التعدد يحدث اذا توفر احد الشروط الآتية: [جبريل، (2014)].

- 1- اذا كانت المتغيرات التفسيرية دالة على المتغيرات تفسيرية أخرى بشكل تركيب خطية كما في المعادلة رقم 2 وبوجود قيم عشوائية  $\varepsilon_i$  تحقق المعادلة الآتية:

$$\alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_p X_p + \varepsilon_i = 0 \dots \dots (3)$$

عندما  $\alpha_1 \dots \alpha_p$  ثوابت

- 2- محدد مصفوفة المعلومات  $XX$  صغير جداً او قريب من الصفر  $|XX| \approx 0$  لذا يمكن إيجاد مقدرات المعلمات التي تكون كبيرة جدا وغير دقيقة لعدم دقة معكوس المصفوفة.

## 2-2 خوارزمية تجزئة القيم الفردية (SVD) Singular Value Decomposition

تستعمل خوارزمية تجزئة القيم الفردية SVD للتخلص من مشكلة التعدد الخطي بين المتغيرات التوضيحية ويمكن النظر الى خوارزمية القيمة الفردية SVD من خلال ثلاث نقاط متوافقة مع بعضها البعض وهي. [ Baker, (2005) ]

- 1- تعتبر SVD طريقة لتحويل المتغيرات المترابطة الى مجموعة من المتغيرات الغير مترابطة والتي تبين العلاقة المختلفة بين عناصر البيانات الاصلية بشكل أفضل.
- 2- خوارزمية القيمة الفردية SDV وسيلة لتحديد وترتيب المتغيرات التي تظهر بها نقاط البيانات الأكثر تبايناً وبالتالي العثور على أفضل تقريب لنقاط البيانات.
- 3- يمكن النظر الى خوارزمية القيمة الفردية SDV كوسيلة لاختزال البيانات مع الحفاظ على طبيعة البيانات.

ولتوضيح هذه الأفكار أعلاه عند رسم خط الانحدار بين نقاط البيانات ثنائية الابعاد للمتغير التابع والمتغيرات التوضيحية فان خط الانحدار يمر خلال أفضل تقريب للبيانات الاصلية ومن هنا يمكن تعريف خوارزمية القيمة الفردية SVD انه الخط الذي يقلل المسافة بين كل نقطة اصلية وخط الانحدار، أي اخذ مجموعة من النقاط (مشاهدات) البيانات عالية الابعاد ومتغيرة للغاية وتقليصها الى مساحة ابعاد اقل تعرض البنية الأساسية للبيانات الاصلية بشكل أكثر وضوحاً وترتيباً من معظم الاختلافات الى الأقل [Mevik & Wehrens (2007)]

يعتمد تحليل القيمة الفردية SDV على نظرية الجبر الخطي التي تنص على ان المصفوفة المستطيلة A (مصفوفة المتغيرات الخالية من مشكلة التعدد الخطي) يمكن الحصول عليها من ثلاث مصفوفات وبشكل الآتي:

$$A_{mn} = U_{mn} * S_{mn} * V_{nn} \dots \dots \dots (3)$$

عندما

$U_{mn}$  عبارة عن مصفوفة متعامدة من المتجهات المميزة لمصفوفة المعلومات  $XX$ .

$V_{nn}$  عبارة عن مصفوفة متعامدة من المتجهات المميزة للمصفوفة  $XX$ .

$S_{mn}$  عبارة عن مصفوفة قطرية تحتوي على الجذور المربعة للقيم المميزة للمصفوفة  $U_{mn}$  او  $V_{nn}$  وبالترتيب التنازلي.

## 3-2 التحليل التمييزي Discriminant Analysis

ان التحليل التمييزي هو أحد أساليب الإحصاء في تحليل المتغيرات المتعددة (Multivariate analysis) فان اول من استعمل مصطلح التمييز هو العالم فشر Fisher عند معالجة مشكلة الفصل separation والتمييز Discriminant اذ أطلق عليها باسم دالة فشر الخطية التمييزية (Fisher Linear Discriminant Function). [الطائي، (2011)]

اذ يستعمل التحليل التمييزي لتصنيف البيانات الى مجموعات كأن تكون مجموعتين او أكثر في توليف خطي للمتغيرات المستقلة أي ان التحليل التمييزي يدرس العلاقة السببية بين المتغيرات فهو يشبه الى حد كبير تحليل الانحدار الخطي (Regression Analysis) الذي يكون فيه المتغير المعتمد (Dependent variable) هو متغير مستمر (continuous variable) لكن في حالة كون المتغير المعتمد (Nominal variable) يتم استعمال التحليل التمييزي او التحليل اللوجستي. [Ramayah, (2010)]

ان المعادلة الخطية للتحليل التمييزي تقوم بتحليل المتغيرات المتعددة (المتغيرات المستقلة) بأفضل تمييز بين مجموعتين معرفتين او أكثر من خلال توظيف خطيا للأوزان Weightings او الحالات لهذه المتغيرات اذ ان معادلة التحليل التمييزي اشبه بمعادلة الانحدار وتكون كالآتي: [Ramayah & others (2010)]

$$Z = a + W_1X_1 + W_2X_2 + \dots + W_kX_k \dots \dots \dots (4)$$

Z: الدالة التمييزية

( $W_1W_2 \dots W_p$ ): معاملات التمييزي للمتغيرات المستقلة ( $X_1X_2 \dots X_p$ )

a: ثابت يمثل المقطع الطولي لمحور الصادات .

k: عدد المتغيرات المستقلة الخطية.

#### 4-2 الدالة التمييزية الخطية في حالة مجموعتين

ان عدد الدوال التمييزية يحدد من خلال العدد الأقل لـ [ (عدد المجموعات-1) او (عدد المتغيرات الكلية) ] فاذا كانت لدينا اربع مجاميع وثمان متغيرات مستقلة فان عدد الدوال التمييزية يساوي العدد الأقل [ (1-4) او (8) ] = 3 .

فان الدالة التمييزية في حالة مجموعتين يشترط ان يكون لدينا مصفوفتي تباينات متساوية ( $\Sigma_1 = \Sigma_2$ ) للمجموعتين ومتجهات المتوسطات الحسابية مختلفة ( $\mu_1, \mu_2$ ) لذلك فان الدالة التمييزية الخطية هي تراكيب خطية يضم p من المتغيرات التي تعظم الفروق بين المتوسطات وتكتب الدالة التمييزية في حالة مجموعتين كالآتي: [الطائي، (2011)]

$$Z = \underline{a} \underline{X} = (\bar{X}_1 - \bar{X}_2) S_p^{-1} \underline{X} \dots \dots \dots (5)$$

عندما

$$\underline{a} = S_p^{-1} (\bar{X}_1 - \bar{X}_2) \dots \dots \dots (6)$$

قيم المتوسطات للمتغيرات لكل مجموعة تستعمل في مقياس mahalanobis الذي يعتمد على قياس المسافات القصيرة بين قيم المتغيرات للملاحظات الجديدة اذ ان يمكن كتابته بالشكل الآتي:

$$D_i^2 = (\bar{X}_1 - \bar{X}_i) S_p^{-1} (\bar{X}_1 - \bar{X}_i) \dots \dots \dots (7)$$

$\bar{X}_i$ : متجه متوسطات المتغيرات ضمن المجموعة i.

$S_p^{-1}$ : معكوس مصفوفة التباينات المقدرة داخل العينتين.

ان من اهم الشروط التي يجب توفرها في التحليل التمييزي ان تكون المجاميع ذات توزيع طبيعي ويشترط ان تكون مصفوفة التباينات متساوية في الدالة المميزة الخطية وحجم العينة كبير وعدم وجود ارتباط بين المتغيرات المستقلة. [الطائي، (2011)]

#### 5-2 التحليل اللوجستي logistic Analysis

هو أسلوب احصائي لفحص العلاقة بين المتغير التابع والمتغيرات التفسيرية (المستقلة) اذ يكون المتغير التابع من نوع المتقطع Discrete والذي يطلق عليه متغير الاستجابة والمتغيرات التوضيحية (المستقلة) من مستويات القياس ويكون التحليل اللوجستي على نوعين التحليل اللوجستي متعدد الاستجابة Analysis of multiple logistic regression والتحليل اللوجستي ثنائي الاستجابة Analysis of Binary logistic regression. [بايطن، (2008)]

## 6-2 الانحدار اللوجستي ثنائي الاستجابة Binary logistic regression

يعرف انموذج الانحدار اللوجستي على أساس المتغير التابع  $y$  متغير الاستجابة يتبع توزيع برنولي ويأخذ القيمة (1) باحتمال  $p$  حدوث استجابة والقيمة (0) باحتمال  $1-p$  عدم حدوث استجابة وكما هو معروف فان نموذج الانحدار اللوجستي يكون بالصيغة الآتية:

$$E(y|x) = p(y = 1) = p = \frac{e^{B_0+B_1X_1+\dots+B_KX_K}}{1 + e^{B_0+B_1X_1+\dots+B_KX_K}} \dots \dots (8)$$

ومن اهم الصفات التي يمتلكها هذا الانموذج هي ان قيمة الطرف الأيمن للنموذج محصورة بين الصفر والواحد الصحيح (0,1) على خلاف انموذج الانحدار الخطي الآتي. [Kleinbaum, (2005)]

$$E(y|x) = B_0+B_1X_1 + \dots + B_KX_K \dots \dots \dots (9)$$

الذي يكون فيه الطرف الأيمن بين  $(-\infty, \infty)$ .

ان نموذج الانحدار اللوجستي غير قابل للتطبيق بسبب قيمته المحصورة بين الصفر والواحد  $0 < E(y|x) < 1$  ولحل هذا المشكلة يتم ادخال تحويل رياضي مناسب يسمى تحويل دالة اللوجت  $\frac{p}{q}$  او  $\frac{p}{1-p}$  والذي تكون قيمته محصورة بين  $(0, \infty)$  وعند اخذ اللوغاريتم الطبيعي لدالة اللوجت  $\log \frac{p}{q}$  تصبح قيمته محصورة بين  $(-\infty, \infty)$  ويكون بالشكل الآتي: [بايطن، (2008)]

$$\log \frac{p}{1-p} = B_0+B_1X_1 + \dots + B_KX_K \dots \dots \dots (10)$$

ان من اهم الشروط الواجب توفرها في الانحدار اللوجستي ان المتغير التابع ثنائي الاستجابة وعد وجود ارتباط بين المتغيرات المستقلة وان يكون حجم العينة الكبيرة (أكبر من عدد المتغيرات) بسبب استخدام الطرق تكرارية في تقدير المعلمات. [بايطن، (2008)]

## 7-2 الاختبارات الإحصائية

سوف يتم عرض اهم الاختبارات الإحصائية التي تستعمل في التحليل التمييزي والتحليل اللوجستي:

### lowely-hotelling: اختبار 1-7-2

وهو من اهم الاختبارات التي تستعمل لا اختبار معنوية الدالة التمييزية في حالة التمييز بين مجموعتين أي ان المجموعتين تعود لنفس المجتمع او لمجتمعات مختلفة في حالة تساوي مصفوفة التباين والتباين المشترك، فيتم قياس المسافة بين المجموعات وتعرف بمسافة مهلنوبس ((mahlanobis distance) كما تم ذكرها في المعادلة رقم 7 أعلاه كلما زادت مسافة مهل ونيس يعني زيادة انفصال المجموعتين وبالتالي فان اختبار  $T^2$  hotelling يكون بالصيغة الآتية: [الطائي، (2010)]

$$T^2 = \frac{n_1 n_2}{(n_1 + n_2)} D_{12}^2 \dots \dots (11)$$

وبالتالي فان اختبار F

$$F = \frac{(n_1 + n_2 - k - 1)}{k(n_1 + n_2 - 2)} \sim F(k, n_1 + n_2 - k - 1) \dots \dots (12)$$

1- فاذا كانت F المحتسبة كبر من الجدولية يدل على ان الدالة التمييزية لها القابلية على التمييز بين المجموعتين. [Suleiman Usman (2014)]

### Wald test: ولد: 2-7-2

يعد من اهم الاختبارات الإحصائية التي يتم استعمالها في انموذج الانحدار اللوجستي لاختبار الدلالة الإحصائية لكل المعلمات  $\hat{B}$  اختبار ولد يحسب إحصاءات Z كالاتي: [Suleiman Usman (2014)]

$$Wald = \left( \frac{\hat{B}}{SE(\hat{B})} \right)^2 \dots \dots \dots (13)$$

اذان  $\hat{B}$  : معلمة المتغير.

$SE(\widehat{B})$ : الخطأ القياسي لمعلمة المتغير.

اختبار Wald يشبه توزيع Chi-square حيث تقارن قيمة الاختبار مع القيمة الجدولية بمستوى معنوية ( $\alpha$ )، فإذا كانت قيمة الاختبار اكبر من القيمة الجدولية ترفض فرضية العدم وتقبل البديلة وهذا يدل على معنوية انموذج الانحدار اللوجستي. [Tygret، ward، (2013)]

### 3-7-2 اختبار Wilks' Lambda

من اهم الاختبارات الإحصائية التي تستعمل لاختبار وجود العلاقة الخطية بين مجموعتين من المتغيرات ويعرف بالصيغة الآتية: [Suleiman Usman (2014)]

$$\lambda = \frac{|\omega|}{|H + \omega|} \dots \dots \dots (14)$$

اذ ان  $\omega$ : مصفوفة التباين والتباين المشترك بين المجموعات.

$H + \omega$ : مصفوفة التباين والتباين المشترك الكلية.

تقارن قيمة  $\lambda$  مع قيمة توزيع chi-square الجدولية بدرجة حرية  $k(n-1)$  وبمستوى معنوية ( $\alpha$ )، فإذا كانت قيمة الاختبار اكبر من القيمة الجدولية ترفض فرضية العدم وتقبل البديلة وهذا يدل على عدم وجود علاقة خطية بين المجموعات. [Tygret، ward، (2013)]

### المبحث الثالث الجانب التطبيقي

#### 1-3 وصف البيانات

بعد الاستعانة بمجموعة من الأطباء اخذت عينة من المرضى المصابين بالنوبة القلبية قوامها 62 شخص من كلا الجنسين لمستشفى الزهراء التعليمي في محافظة واسط للعام 2018 ومن خلال دراسة اهم العوامل التي تؤثر على المرضى والتي يتم تعريفها من خلال المتغيرات الآتية:

- حالة المريض والتي اعتبرت كمتغير استجابة ( $y$ ) اذ ان في حالة الوفاة ( $y=0$ ) وفي حالة البقاء على قيد الحياة ( $y=1$ ).
- جنس المريض وهو اول المتغيرات التوضيحية ( $X1$ ) ففي حالة كون المريض ذكر فان ( $X1=1$ ) وإذا كانت انثى ( $X1=2$ ).
- عمر المريض يمثل ثاني المتغيرات التوضيحية ( $X2$ ) ومقاس بالسنة.
- وزن المريض ويمثل ثالث المتغيرات التوضيحية ( $X3$ ) ومقاس كغم.
- التغذية وهي المتغير التوضيحي الرابع ( $X4$ ) اذ ان في حالة التغذية الصحية فان ( $X4=0$ ) وفي حالة التغذية غير الصحية (تحتوي على نسب عالية من الدهون والكربوهيدرات) فان ( $X4=1$ ).
- ممارسة الأنشطة الرياضية وهي المتغير التوضيحية الخامس ( $X5$ ) ففي حالة كون المريض يمارس الأنشطة الرياضية تكون ( $X5=0$ ) اما في حالة كون المريض لا يمارس أي نشاط رياضي فان ( $X5=1$ ).
- الحالة الانفعالية للمريض وهي المتغير التوضيحي السادس ( $X6$ ) فتكون ( $X6=1$ ) إذا كان المريض يتمتع بهدوء الاعصاب دائما وتكون ( $X6=2$ ) إذا كان المريض منفعل دائما.
- ضغط الدم ويمثل ( $X7$ ) ويقاس بوحدة ملم زئبق تكون النسب الطبيعية لضغط الدم العالي للشخص البالغ (120-130) وتكون النسب الطبيعية لضغط الدم الواطئ بين (70 - 80).
- السكر في الدم ويمثل ( $X8$ ) ويقاس بوحدة mmol/L تكون النسب الطبيعية للسكر من (85-126) قبل الاكل صيام لمدة ثمان ساعات ومن (165-200) بعد الاكل.
- الارتجاج الأذيني وهو المتغير التوضيحي التاسع ( $X9$ ) ففي حال كون المريض لا يعاني منه فان ( $X9=1$ ) وفي حالة كون المريض يعاني منه ( $X9=2$ ).
- ادوية تخثر الدم تمثل المتغير التوضيحي العاشر ( $X10$ ) إذا كان المريض يأخذ ادوية تخثر الدم فان ( $X10=2$ ) وفي حالة كون المريض لا يأخذ ادوية تخثر الدم فان ( $X10=1$ ).
- أمراض القلب تمثل المتغير التوضيحي الاحد عشر ( $X11$ ) فإذا كان المريض يعاني من امراض القلب فان ( $X11=2$ ) وفي حالة كون المريض لا يعاني من امراض القلب فان ( $X11=1$ ).
- نوبات قلبية متكررة تمثل المتغير التوضيحي الاثنا عشر ( $X12$ ) فإذا كان المريض مصاب بنوبة قلبية سابقا فان ( $X12=2$ ) وفي حالة كون المريض غير مصاب بنوبة قلبية فان ( $X12=1$ ).
- التدخين يمثل المتغير التوضيحي الثالث عشر ( $X13$ ) فإذا كان المريض لا يدخن فان ( $X13=0$ ) وفي حالة كون المريض يدخن فان ( $X13=1$ ).

### 3-3 التحليل الإحصائي:

في هذا البحث تم إجراء التحليل الإحصائي باستخدام البرنامج الإحصائي R والبرنامج حزمة الإحصاء SPSS 22 ومن خلال الجداول الآتية:

#### جدول (1) يبين مشكلة التعدد الخطي

| variable | Eigenvalue | Condition index | Tolerance | VIF    |
|----------|------------|-----------------|-----------|--------|
| X1       | 11.129     | 1               | 0.05      | 19.953 |
| X2       | 0.792      | 3.749           | 0.575     | 1.74   |
| X3       | 0.623      | 4.227           | 0.724     | 1.381  |
| X4       | 0.493      | 4.752           | 0.761     | 1.314  |
| X5       | 0.442      | 5.017           | 0.734     | 1.363  |
| X6       | 0.128      | 9.316           | 0.825     | 1.212  |
| X7       | 0.098      | 10.654          | 0.024     | 13.213 |
| X8       | 0.088      | 11.26           | 0.08      | 14.136 |
| X9       | 0.065      | 13.122          | 0.804     | 1.244  |
| X10      | 0.054      | 14.404          | 0.69      | 1.448  |
| X11      | 0.049      | 15.076          | 0.075     | 12.105 |
| X12      | 0.029      | 19.58           | 0.082     | 11.466 |
| X13      | 0.01       | 34.065          | 0.051     | 19.791 |

يتضح من الجدول أعلاه أن قيمة VIF كانت لبعض المتغيرات أكبر من 10 وهي (x1)، (x7)، (x8)، (x11)، (x12)، (x13) وهذا يدل على وجود مشكلة التعدد الخطي، وأن قيم Tolerance أقل من 0.5 فإن معنى ذلك أنه لا يوجد مشكلة ارتباط ذاتي بين المتغيرات المستقلة وكذلك يمكن مشاهدة ذلك من خلال العدد الشرطي الذي كانت قيمته أكبر من 15 .

جدول رقم (2) يبين قيم المتغيرات بعد تطبيق SVD

| x13  | x12  | x11  | x10  | x9   | x8   | x7   | x6   | x5   | x4   | x3     | x2     | x1   |
|------|------|------|------|------|------|------|------|------|------|--------|--------|------|
| 0.32 | 1.16 | 1.56 | 2.00 | 1.73 | 0.46 | 1.43 | 2.21 | 0.63 | 1.27 | 102.12 | 73.73  | 1.74 |
| 0.14 | 1.11 | 1.43 | 2.54 | 0.95 | 1.13 | 1.53 | 1.39 | 0.48 | 1.13 | 55.47  | 64.54  | 1.78 |
| 0.73 | 1.31 | 0.24 | 1.46 | 1.32 | 0.66 | 0.97 | 1.91 | 0.34 | 0.29 | 91.34  | 70.73  | 1.30 |
| 1.35 | 0.81 | 1.77 | 0.94 | 1.20 | 0.87 | 1.70 | 0.94 | 0.54 | 0.20 | 96.64  | 66.63  | 0.77 |
| 0.95 | 1.17 | 1.20 | 1.29 | 1.74 | 0.93 | 0.84 | 0.75 | 0.06 | 0.78 | 77.06  | 60.29  | 1.17 |
| 0.22 | 1.14 | 1.50 | 2.16 | 1.10 | 0.43 | 2.31 | 1.80 | 0.89 | 1.23 | 63.34  | 52.81  | 2.06 |
| 0.89 | 1.45 | 1.70 | 1.47 | 2.14 | 0.27 | 1.59 | 1.81 | 0.98 | 0.95 | 67.23  | 62.23  | 0.94 |
| 1.33 | 1.00 | 1.01 | 0.92 | 0.93 | 0.20 | 2.09 | 1.79 | 0.21 | 0.67 | 86.83  | 68.57  | 0.79 |
| 1.10 | 1.08 | 1.71 | 2.04 | 1.56 | 0.88 | 1.12 | 1.65 | 0.99 | 0.23 | 98.24  | 54.07  | 0.80 |
| 0.73 | 1.70 | 1.41 | 1.18 | 2.74 | 0.48 | 1.30 | 1.70 | 0.70 | 0.38 | 88.59  | 67.20  | 1.27 |
| 0.01 | 0.86 | 0.91 | 1.42 | 1.73 | 0.30 | 1.28 | 2.16 | 0.61 | 0.41 | 107.87 | 56.29  | 2.18 |
| 0.34 | 0.36 | 1.18 | 1.33 | 0.77 | 0.56 | 2.17 | 0.81 | 0.49 | 0.17 | 51.29  | 79.62  | 1.42 |
| 0.76 | 1.13 | 2.00 | 1.27 | 2.10 | 0.22 | 1.01 | 1.12 | 0.93 | 0.15 | 67.83  | 60.06  | 1.07 |
| 0.11 | 2.15 | 0.81 | 1.42 | 1.23 | 0.62 | 1.88 | 1.85 | 0.58 | 0.03 | 100.42 | 49.59  | 2.35 |
| 0.34 | 0.64 | 1.55 | 1.19 | 1.41 | 0.75 | 1.33 | 2.47 | 0.17 | 1.62 | 56.41  | 79.67  | 1.67 |
| 0.24 | 2.31 | 1.47 | 1.42 | 1.69 | 0.40 | 2.33 | 1.38 | 0.35 | 1.35 | 84.17  | 39.54  | 2.10 |
| 0.36 | 0.90 | 1.86 | 2.84 | 1.18 | 0.71 | 1.94 | 1.69 | 0.10 | 0.44 | 56.66  | 55.12  | 1.33 |
| 0.03 | 2.10 | 1.63 | 1.71 | 1.46 | 0.58 | 1.42 | 1.88 | 0.18 | 0.63 | 92.98  | 42.89  | 2.17 |
| 0.65 | 2.07 | 2.51 | 1.26 | 1.63 | 0.65 | 2.05 | 1.65 | 0.38 | 0.42 | 68.20  | 62.03  | 1.58 |
| 0.46 | 0.57 | 2.03 | 0.75 | 2.14 | 0.18 | 1.41 | 1.91 | 0.33 | 0.18 | 91.98  | 53.29  | 1.82 |
| 0.49 | 2.42 | 2.40 | 1.87 | 1.94 | 0.36 | 2.01 | 1.41 | 0.02 | 0.40 | 94.93  | 42.51  | 1.81 |
| 0.38 | 1.33 | 1.43 | 1.95 | 1.12 | 0.65 | 1.53 | 1.45 | 0.01 | 0.23 | 94.48  | 66.04  | 2.44 |
| 1.15 | 1.62 | 1.55 | 1.75 | 2.00 | 0.14 | 1.80 | 1.60 | 0.82 | 0.77 | 47.37  | 75.30  | 1.46 |
| 0.30 | 0.55 | 1.24 | 1.34 | 2.36 | 0.46 | 1.76 | 1.81 | 0.79 | 0.03 | 87.07  | 54.25  | 1.53 |
| 0.98 | 1.17 | 1.78 | 0.67 | 1.36 | 0.08 | 1.98 | 1.72 | 0.86 | 1.31 | 73.90  | 65.01  | 1.14 |
| 0.05 | 0.24 | 0.74 | 1.00 | 1.77 | 0.36 | 1.39 | 1.31 | 0.68 | 0.74 | 55.28  | 63.57  | 1.70 |
| 0.69 | 1.96 | 1.94 | 1.22 | 1.10 | 0.36 | 2.98 | 1.47 | 0.71 | 0.61 | 92.87  | 73.50  | 1.40 |
| 0.17 | 0.95 | 1.50 | 1.70 | 1.21 | 0.03 | 1.97 | 0.65 | 0.24 | 0.27 | 77.40  | 72.46  | 1.96 |
| 0.45 | 2.08 | 2.30 | 1.49 | 0.67 | 1.10 | 1.86 | 1.82 | 0.04 | 0.59 | 87.01  | 69.55  | 1.87 |
| 1.23 | 1.16 | 1.28 | 1.76 | 0.87 | 0.42 | 1.24 | 1.01 | 0.09 | 0.35 | 87.97  | 79.54  | 0.70 |
| 1.16 | 1.70 | 2.44 | 1.02 | 1.40 | 0.44 | 1.47 | 1.46 | 0.06 | 0.68 | 109.72 | 86.49  | 1.19 |
| 0.43 | 0.62 | 1.29 | 1.67 | 2.03 | 0.88 | 1.50 | 1.97 | 0.32 | 0.60 | 54.43  | 80.03  | 1.15 |
| 0.09 | 1.21 | 1.65 | 1.18 | 2.38 | 1.02 | 2.42 | 1.77 | 0.12 | 0.21 | 60.11  | 77.90  | 1.77 |
| 0.55 | 1.23 | 1.42 | 1.44 | 1.98 | 0.53 | 1.25 | 2.15 | 1.56 | 0.34 | 75.31  | 51.47  | 1.28 |
| 0.72 | 1.77 | 1.14 | 1.56 | 1.59 | 1.72 | 2.14 | 2.21 | 0.07 | 0.00 | 72.48  | 73.41  | 1.21 |
| 1.59 | 1.07 | 2.05 | 1.36 | 0.92 | 0.44 | 1.81 | 1.19 | 0.26 | 0.59 | 97.38  | 80.75  | 0.51 |
| 0.09 | 1.26 | 2.18 | 1.88 | 0.82 | 1.10 | 1.70 | 1.75 | 0.38 | 0.69 | 60.44  | 100.26 | 1.81 |
| 0.88 | 1.45 | 2.49 | 1.17 | 1.49 | 0.73 | 1.01 | 2.54 | 0.18 | 0.75 | 77.47  | 52.06  | 1.45 |
| 1.10 | 1.34 | 2.80 | 1.44 | 1.21 | 0.40 | 1.58 | 2.40 | 0.92 | 0.20 | 103.53 | 55.08  | 1.11 |
| 1.47 | 1.48 | 1.61 | 1.14 | 2.10 | 0.61 | 2.01 | 2.25 | 0.08 | 0.89 | 92.36  | 55.25  | 0.82 |
| 0.91 | 1.04 | 0.80 | 1.68 | 1.30 | 0.92 | 1.45 | 1.61 | 1.33 | 0.57 | 92.35  | 60.34  | 0.98 |
| 0.14 | 0.98 | 2.01 | 2.34 | 1.57 | 0.85 | 1.63 | 1.46 | 0.69 | 0.28 | 59.34  | 73.96  | 1.67 |
| 0.03 | 1.12 | 1.51 | 1.88 | 2.32 | 0.42 | 2.44 | 1.29 | 0.76 | 0.08 | 63.66  | 70.06  | 1.58 |
| 0.01 | 1.96 | 1.58 | 1.69 | 1.15 | 1.28 | 1.18 | 1.17 | 0.42 | 0.32 | 70.35  | 67.73  | 2.02 |
| 0.00 | 2.04 | 1.25 | 1.72 | 0.86 | 0.69 | 1.14 | 1.31 | 0.11 | 0.41 | 83.83  | 84.44  | 1.97 |
| 0.39 | 1.46 | 0.86 | 1.87 | 1.41 | 0.60 | 1.97 | 1.84 | 0.29 | 0.05 | 96.82  | 72.73  | 2.36 |
| 0.55 | 0.15 | 0.93 | 2.13 | 0.87 | 0.38 | 0.83 | 1.31 | 0.30 | 0.20 | 76.94  | 95.98  | 1.01 |
| 1.28 | 1.20 | 0.87 | 0.89 | 1.74 | 1.17 | 1.47 | 1.35 | 0.03 | 0.51 | 76.51  | 109.32 | 0.59 |
| 1.55 | 1.40 | 1.98 | 2.26 | 1.70 | 0.09 | 1.40 | 1.24 | 0.36 | 0.00 | 93.02  | 89.74  | 0.18 |
| 0.63 | 0.73 | 2.39 | 1.06 | 1.60 | 0.42 | 1.50 | 1.54 | 0.24 | 1.25 | 83.70  | 68.18  | 1.68 |
| 0.51 | 1.23 | 1.00 | 1.42 | 1.67 | 0.05 | 1.07 | 2.26 | 1.18 | 0.29 | 76.22  | 71.66  | 1.19 |
| 0.38 | 1.94 | 1.74 | 1.74 | 0.74 | 0.44 | 1.96 | 0.88 | 0.23 | 0.10 | 82.49  | 72.47  | 1.59 |
| 0.91 | 1.03 | 0.59 | 1.72 | 1.53 | 0.25 | 1.26 | 2.18 | 0.86 | 0.30 | 81.74  | 63.43  | 0.83 |
| 0.04 | 1.10 | 0.71 | 2.10 | 1.55 | 1.50 | 2.08 | 1.58 | 0.22 | 0.00 | 76.30  | 51.28  | 2.01 |
| 0.82 | 1.89 | 1.50 | 1.35 | 1.95 | 0.12 | 1.50 | 1.56 | 0.56 | 1.28 | 98.78  | 72.34  | 1.37 |
| 0.22 | 2.43 | 1.46 | 1.98 | 1.23 | 0.77 | 2.02 | 1.91 | 0.61 | 0.91 | 94.28  | 65.07  | 1.93 |
| 0.33 | 2.23 | 1.81 | 1.62 | 1.62 | 0.90 | 1.71 | 0.58 | 0.26 | 0.03 | 98.54  | 81.54  | 1.74 |
| 0.25 | 2.38 | 1.10 | 1.47 | 1.38 | 0.19 | 2.22 | 1.50 | 0.00 | 0.61 | 92.73  | 62.32  | 1.99 |
| 0.03 | 1.68 | 1.42 | 1.92 | 2.14 | 0.30 | 2.41 | 2.00 | 0.31 | 0.36 | 84.12  | 54.84  | 2.08 |
| 0.98 | 2.05 | 1.22 | 1.41 | 1.52 | 0.31 | 2.71 | 2.15 | 0.67 | 0.01 | 90.75  | 62.69  | 1.24 |
| 1.27 | 1.03 | 1.20 | 1.41 | 1.03 | 0.08 | 1.78 | 2.82 | 0.39 | 0.18 | 81.55  | 62.46  | 0.72 |
| 0.33 | 0.38 | 1.89 | 1.45 | 1.79 | 0.27 | 1.27 | 0.75 | 0.82 | 0.60 | 79.39  | 66.98  | 1.52 |

ان الجدول أعلاه يبين قيم المتغيرات بعد تطبيق خوارزمية تجزئة القيمة المفردة عليها التي تقوم بحل مشكلة التعدد بين المتغيرات من خلال اختزال البيانات مع الحفاظ على طبيعتها فهي وسيلة لتحديد وترتيب المتغيرات التي تظهر بها نقاط البيانات الأكثر تبايناً وبالتالي العثور على أفضل تقريب لنقاط البيانات.

**الجانب التطبيقي للتحليل التمييزي**

قبل اجراء التحليل التمييزي يجب التأكد من ان المجاميع تتبع التوزيع الطبيعي وتم ذلك من خلال اختبار Kolmogorov-Smirnov Z فوجد ان مستوى المعنوية لجميع المتغيرات في المجاميع أكبر من 0.05 مما يدل على انها تتوزع طبيعي، وعدم وجود مشكلة التعدد الخطي بين المتغيرات إذا تم معالجة مشكلة التعدد الخطي من خلال خوارزمية القيمة المفردة، وعدم وجود قيم شاذة بين المتغيرات ويتم ذلك من خلال نتائج الجداول الاتية:



جدول (3) مسافة مهل نوبس

| Mahlanobis Distance |       |       |       |       |       |       |
|---------------------|-------|-------|-------|-------|-------|-------|
| 16.30               | 12.15 | 12.84 | 10.67 | 7.79  | 8.50  | 11.23 |
| 12.60               | 14.94 | 16.92 | 8.30  | 12.13 | 8.11  | 13.23 |
| 9.68                | 11.80 | 33.93 | 14.47 | 11.11 | 14.78 |       |
| 13.51               | 10.84 | 9.42  | 12.19 | 11.90 | 10.32 |       |
| 16.13               | 16.18 | 11.08 | 16.50 | 18.45 | 9.01  |       |
| 13.51               | 14.24 | 12.21 | 9.38  | 14.54 | 11.68 |       |
| 7.07                | 18.97 | 14.68 | 15.56 | 18.73 | 10.07 |       |
| 11.13               | 9.05  | 10.25 | 15.33 | 19.30 | 8.66  |       |
| 12.29               | 8.46  | 9.44  | 16.13 | 9.99  | 10.76 |       |
| 9.62                | 14.54 | 8.42  | 12.51 | 12.58 | 16.32 |       |

من الجدول أعلاه بعد مراجعة كافة قيم **Mahlanobis Distance** فإننا نجد أن كل من هذه القيم أقل من القيمة الجدولية لمربع كأي  $X^2$  عند درجة حرية (62-1) \* (2-1) ومستوى معنوية 0.001 والتي تساوي 43.19 هذا يدل على عدم وجود قيم شاذة بين كل البيانات المتعلقة بكافة المتغيرات المستقلة والبالغة 62 مشاهدة.

لمعرفة معنوية الدالة المميزة الخطية تم تطبيق اختبار **hotelling  $T^2$**  الذي تم ذكره وفق المعادلة رقم (12) فوجد ان قيمته تساوي **hotelling  $T^2 = 134.23$**  ومن خلال المعادلة رقم (13) وجد ان قيمة المحتسبة  $F=11.543$  وعند مقارنة قيمة  $F$  المحتسبة مع  $F(k, n_1 + n_2 - k - 1)$  الجدولية بمستوى معنوية 0.001 نجد ان  $f(13, 48) = 3.64$  لذا ترفض فرضية العدم وتقبل البديلة هذا يدل على ان الدالة التمييزية جيدة في التمييز بين المجموع.

جدول رقم (4) بين اختبار F واختبار Wilks' Lambda في الدالة المميزة الخطية

|                   | Wilks' Lambda | F     | df1 | df2 | Sig. |
|-------------------|---------------|-------|-----|-----|------|
| الجنس             | .956          | 2.772 | 1   | 60  | .101 |
| العمر             | .989          | 1.669 | 1   | 60  | .000 |
| وزن               | .961          | 2.448 | 1   | 60  | .123 |
| الغذاء            | .942          | 3.716 | 1   | 60  | .059 |
| الرياضة           | .990          | 1.609 | 1   | 60  | .438 |
| الانفعال          | .954          | 2.916 | 1   | 60  | .000 |
| ضغط               | .969          | 1.947 | 1   | 60  | .168 |
| السكر             | .973          | 1.660 | 1   | 60  | .000 |
| ارتجاج اذيني      | .996          | 2.239 | 1   | 60  | .027 |
| تخثر ادوية        | .992          | .2475 | 1   | 60  | .093 |
| امراض القلب       | 1.000         | 2.003 | 1   | 60  | .000 |
| قلب نوبات التدخين | 1.000         | 2.001 | 1   | 60  | .074 |
| التدخين           | .987          | 1.810 | 1   | 60  | .000 |

من خلال الجدول رقم 4 لاختبار F واختبار **Wilks' Lambda** نجد ان المتغيرات (العمر، والانفعال، والسكر، وامراض القلب، والتدخين) كان لها تأثير أكبر من بقية المتغيرات من خلال مستوى المعنوية للتمييز بين حالة البقاء على قيد الحياة المريض وحالة الوفاة.

جدول رقم (5) يبين Eigenvalues

| Function | Eigenvalue        | % of Variance | Cumulative % | Canonical Correlation |
|----------|-------------------|---------------|--------------|-----------------------|
| 1        | .361 <sup>a</sup> | 100.0         | 100.0        | .415                  |

a. First 1 canonical discriminant functions were used in the analysis.

من خلال الجدول رقم 5 تم اختبار معنوية الدالة التمييزية الخطية اذ وجد قوة التمييز من خلال التباين بين المجموعتين ولكل المتغيرات قد فسرت 100% من التباين وان معامل الارتباط وجد 0.415.

جدول رقم (6) اختبار Wilks' Lambda Chi-square

| Test of Function(s) | Wilks' Lambda | Chi-square | df | Sig.  |
|---------------------|---------------|------------|----|-------|
| 1                   | .735          | 76.507     | 13 | .0002 |

من خلال الجدول رقم 6 نجد ان قيمة اختبار **Wilks' Lambda** تساوي 0.735 وهي اقل من واحد مما يدل على عدم وجود علاقة خطية بين المتغيرات المستقلة وان قيمة اختبار Chi-square تساوي 76.507 هذا يدل على ان الدالة التمييزية الخطية لها القابلية على التمييز بين المجموعتين من خلال المتغيرات المستقلة كما هو مشاهد من خلال مستوى معنوية الاختبار اذ كانت اقل من مستوى معنوية 0.001.

جدول رقم (7) تقدير المعلمات

|            | Function |
|------------|----------|
|            | 1        |
| الجنس      | .884     |
| العمر      | .265     |
| وزن        | .526     |
| الغذاء     | .439     |
| الرياضة    | .246     |
| الانفعال   | .398     |
| ضغط        | -.419-   |
| السكر      | -.234-   |
| ارتجاج     | .083     |
| اذيني      |          |
| تخثر ادوية | -.056-   |
| مراض       | .046     |
| القلب      |          |
| قلب نوبات  | -.214-   |
| التدخين    | .336     |

ان الجدول أعلاه يمثل تقدير معلمات المتغيرات المستقلة.

جدول رقم (8) تصنيف مشاهدات المتغير التابع

|          |       | حالة المريض | Predicted Group Membership |         | Total |
|----------|-------|-------------|----------------------------|---------|-------|
|          |       |             | .00000                     | 1.00000 |       |
| Original | Count | .00000      | 13                         | 5       | 18    |
|          |       | 1.00000     | 11                         | 33      | 44    |
|          | %     | .00000      | 72.2                       | 27.8    | 100.0 |
|          |       | 1.00000     | 25.0                       | 75.0    | 100.0 |

ان الجدول أعلاه يبين التصنيف الصحيح الى المجموعات اذ نجد  $46=33+13$  من المشاهدات صنفنا بشكل صحيح وان احتمال التصنيف الصحيح لها هو 0.74 وان  $16=5+11$  من المشاهدات صنفنا بشكل خاطئ وباحتمال تصنيف 0.26.

## الجانب التطبيقي للتحليل اللوجستي

يتم عرض نتائج التحليل اللوجستي من خلال الجدول الآتي:

جدول رقم (9) يبين الحد الثابت

| Iteration | -2 Log likelihood | Coefficients |
|-----------|-------------------|--------------|
|           |                   | Constant     |
| Step 0 1  | 74.742            | .839         |
| 2         | 74.703            | .893         |
| 3         | 74.703            | .894         |

من الجدول أعلاه تم استعمال طريقة الإمكان الأعظم بالخطوات التكرارية ومن خلال الحصول على أقل قيمة سالبة لضعف لوغاريتم دالة الإمكان لقد حصلنا في التكرار الثالث على أقل قيمة سالبة لضعف لوغاريتم دالة الإمكان للحد الثابت وكانت تساوي 74.703 وبمعامل الحد الثابت 0.894.

جدول رقم (10) يمثل النسب المئوية لتصنيف المتغير التابع

| Observed           | Predicted | المريض حالة |         | Percentage Correct |
|--------------------|-----------|-------------|---------|--------------------|
|                    |           | .00000      | 1.00000 |                    |
|                    |           | Step 0 حالة | .00000  |                    |
| المريض             | 1.00000   | 0           | 44      | 100.0              |
| Overall Percentage |           |             |         | 71.0               |

من الجدول أعلاه ان التصنيف الصحيح الى مشاهدات المتغير التابع بلغت نسبته 71.0% والتصنيف الخاطئ نسبته 29%.

جدول رقم (11) يبين اختبار Wald

|                 | B    | S.E. | Wald   | df | Sig.  | Exp(B) |
|-----------------|------|------|--------|----|-------|--------|
| Step 0 Constant | .894 | .280 | 10.205 | 1  | .0001 | 2.444  |

من الجدول أعلاه عند مقارنة قيمة wald مع القيمة الجدولية ل chi-square بمستوى معنوية 0.001 ودرجة حرية K-1=12 والتي تساوي 32.91 لذا يتم رفض فرضية العدم وقبول البديلة وهذا يدل على ان المتغيرات المستقلة لها تأثير على المتغير التابع وكذلك يمكن مشاهدة ذلك من خلال sig.

جدول رقم (12) يبين اختبار Cox & Snell R Square

| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
|------|-------------------|----------------------|---------------------|
| 1    | 55.897            | .762                 | .794                |

من الجدول أعلاه نجد ان قيمة اختبار كوس المناظر الى معامل التحديد  $R^2$  بلغت نسبته 0.762 هذا يعني ان المتغيرات المستقلة فقد فسرت ما هو نسبته 76% من نموذج التحليل اللوجستي اما الباقي والذي نسبت 24% يدرج ضمن حد الخطأ.

جدول رقم (13) يبين تقدير المعلمات

|                     |       | B      | S.E.  | Wald  | df   | Sig.      | Exp(B) |
|---------------------|-------|--------|-------|-------|------|-----------|--------|
| Step 1 <sup>a</sup> | X1    | 2.366  | 1.628 | 2.113 | 1    | .002      | .094   |
|                     | X2    | 1.923  | .032  | .497  | 1    | .000      | .978   |
|                     | X3    | -.043- | .030  | 2.023 | 1    | .155      | .958   |
|                     | X4    | 1.561  | .955  | 2.673 | 1    | .102      | .210   |
|                     | X5    | 1.146  | 1.127 | 1.035 | 1    | .309      | .318   |
|                     | X6    | 2.938  | .798  | 1.381 | 1    | .000      | .391   |
|                     | X7    | 1.425  | .929  | 2.355 | 1    | .125      | 4.159  |
|                     | X8    | 1.156  | 1.315 | .773  | 1    | .379      | 3.177  |
|                     | X9    | .173   | .891  | .038  | 1    | .846      | .841   |
|                     | X10   | .099   | 1.041 | .090  | 1    | .000      | 1.104  |
|                     | X11   | .081   | .641  | .016  | 1    | .900      | .923   |
|                     | X12   | .284   | .730  | .151  | 1    | .697      | 1.328  |
|                     | X13   | 3.496  | 1.819 | .074  | 1    | .000      | .609   |
| Constant            | 9.729 | 6.419  | 2.297 | 1     | .130 | 16800.641 |        |

a. Variable(s) entered on step 1: X1، X2، X3، X4، X5، X6، X7، X8، X9، X10، X11، X12، X13.

من الجدول أعلاه نجد ان المتغيرات التي كانت لها دور كبير من باقي المتغيرات في تفسير نموذج التحليل اللوجستي من خلال تقدير معالمها هي (التدخين وامراض القلب والعمر والانفعال والسكر والضغط) أي ان هذه المتغيرات هي التي تؤثر على المريض في تحديد حالة البقاء على قيد الحياة او حالة الوفاة ويمكن ملاحظة ذلك من خلال مستوى المعنوية لاختبار ولد لتلك المتغيرات.

#### المبحث الرابع الاستنتاجات والتوصيات:

تم عرض اهم الاستنتاجات والتوصيات التي تم التوصل اليها من خلال هذا البحث وهي كالآتي:

#### 1-4 الاستنتاجات

- 1- تعد خوارزمية تجزئة القيمة المفردة طريقة جيدة لحل مشكلة التعدد الخطي مع الحفاظ على طبيعة البيانات بعد اختزالها.
- 2- ان جميع بيانات المتغيرات التي تم اخذها والتي تؤثر على النوبة القلبية لا تحتوي على القيم الشاذة، كما بينت مسافات مهلبوس.
- 3- ان المتغيرات المستقلة (العمر، الانفعال، امراض القلب، التدخين) كان لها اثر كبير على مرض النوبة القلبية للتمييز بين حالات البقاء على قيد الحياة وحالات الوفاة من خلال اختبار F واختبار Wilks' Lambda واختبار chi-square للدالة التمييزية.
- 4- في التحليل التمييزي ان احتمال التصنيف الصحيح لبيانات المتغير التابع هو 0.74.
- 5- في التحليل اللوجستي تم استعمال مقدرات الإمكان الأعظم التكرارية وتم التوقف في التكرار الرابع والتصنيف الصحيح الى بيانات المتغير التابع بلغت نسبته 71%.
- 6- في التحليل اللوجستي قيمة اختبار ولد كانت 10.205 هذا يدل على ان المتغيرات المستقلة كان لها تأثير على المتغير التابع (مرض النوبة القلبية).
- 7- ان قيمة اختبار كوكس كانت 0.762 أي ان المتغيرات المستقلة فسرت 76% من نموذج التحليل اللوجستي.
- 8- في التحليل اللوجستي المتغيرات المستقلة (التدخين، العمر، الانفعال، امراض القلب، السكر، والضغط) كان لها اثر كبير على مرض النوبة القلبية لبيان حالة البقاء على قيد الحياة وحالة الوفاة من خلال تقدير المعلمات ومستوى المعنوية لتلك المعلمات.

#### 2-4 التوصيات:

- 1- يوصي الباحثان باستعمال التحليل اللوجستي في دراسة مشاكل الحالات الصحية والاجتماعية ذات المتغيرات المستقلة التي تعاني من مشكلة التعدد الخطي.

- 2- يوصي الباحثان تطوير أساليب جمع البيانات في وزارات الدولة والقطاع الصحي للحصول على بيانات ذات أهمية كبرى وحل المشاكل عن طريق اتخاذ القرار الاحصائي.
- 3- يوصي الباحث استعمال أسلوب احصائي اخر للمقارنة مع التحليل التمييزي او التحليل اللوجستي او الثلاث معاً
- 4- يوصي الباحثان عدم التدخين والانفعال وذلك لظهور هذان المتغيران في التحليل التمييزي والتحليل اللوجستي لأثرهما البالغ على مرض النوبة القلبية.

#### المصادر

- 2- الطائي، خالد ضاري عباس، وعتاب، محمد رمضان (2011)، " استخدام الدالة المميزة التربيعية في تمييز أنماط الأرقام العربية"، مجلة جامعة النهريين، المجلد 14 والعدد 1 ص 58-78.
- 3- بايطن، عادل احمد (2008)، "الانحدار اللوجستي وكيفية استخدامه في بناء نماذج التنبؤات ذات المتغيرات التابعة ثنائية القيمة"، أطروحة دكتوراه كلية التربية، جامعة ام القرى.
- 4- جبريل، محمد سليمان محمد، (2014)، " التعدد الخطي أسبابها تأثيراتها والمعالجة بانحدار الحافة وانحدار المركبات الرئيسية مع التطبيق على بيانات افتراضية.
- 5- Bager، Ali & Roman، Monica & Algedih، Meshal، (2017)، " Addressing multicollinearity in regression models: a ridge regression application"، MPRA Paper No. 81390، posted 16 September 2017 09:04 UTC.
- 6- Baker، Kirk (2005)، " Singular Value Decomposition Tutorial"، book، ROUGH DRAFT - USE AT OWN RISK: suggestions kbaker@ling.osu.edu
- 7- Kleinbaum، D. G. and Klein، M. (2005)، "Logistic Regression"، book، Springer، Third Edition.
- 8- Mevik، Bj\_rn-Helge & Wehrens، Ron (2007)، "Principal Component and Partial Least Squares Regression in R"، Journal of Statistical Software January 2007، Volume 18، Issue 2.
- 9- Ramayah، T. & others (2010)، " Discriminant analysis: An illustrated example" ، African Journal of Business Management Vol. 4(9)، pp. 1654-1667، 4 August.
- 10- Suleiman1، S. & Suleman، Issa & U.Usman3، U.، (2014)، "Predicting an Applicant Status Using Principal Component، Discriminant and Logistic Regression Analysis"، International Journal Of Mathematics And Statistics Invention (IJMSI)، E-ISSN: 2321 – 4767 P-ISSN: 2321 – 4759.
- 11- Tygret، Mark & ward، Rachel (2013) "Testing goodness-of-fit for logistic regression"، June 21، 2013، [arXiv:1306.0959v2](https://arxiv.org/abs/1306.0959v2) [stat.ME] for this version).