



إستخدام طرق التحليل المتعدد لدراسة أهم العوامل المؤثرة التي تميز

بين الجنسين بمرض السكري

Use multiple analysis methods to study the most important influencing factors
Between the sexes and diabetes

المدرس المساعد زينب عبدالله محمد

قسم الاحصاء / كلية الادارة والاقتصاد/ جامعة صلاح الدين- أربيل

zainab.Muhammad@su.edu.krd

المدرس المساعد إسراء مؤيد علي

قسم الاحصاء / كلية الادارة والاقتصاد / جامعة صلاح الدين- أربيل

Israa.Ali@su.edu.krd

الملخص

معلومات البحث

إن التحليل الإحصائي متعدد المتغيرات يتناول دراسة بيانات تلك المشاهدات والتعبير عنها من خلال أكثر المتغيرات تأثيراً في الظاهرة محل الدراسة، اعتمد البحث على الربط بين تحليل متعدد المتغيرات باستخدام طريقة تحليل المكونات الرئيسية (Principle Component method) والتحليل التمييزي (Discriminate Analysis)، لذا يهدف هذا البحث الى تحديد أهم العوامل المؤثرة على المرضى من الجنسين المصابين بمرض

تاريخ البحث:

الإستلام: 2018/5/11

القبول: 2018/6/10

النشر: خريف 2018

Doi:

10.25212/lfu.qzj.3.4.21

الكلمات المفتاحية:

السكري من خلال تحديد عدد العوامل المستخلصة باستخدام التحليل العاملي وكذلك باستخدام التحليل المميز في عملية التمييز والتصنيف ومن ثم تقدير أفضل معادلة تصنيفية بالأعتماد على القياسات الخاصة والمستخلصة بواسطة المكونات الرئيسة للتمييز بين الجنسين والمقارنة بين النتائج على أساس نسبة خطأ التصنيف وبيان تأثيرها على عملية تصنيف، وقسم البحث الى الجانبين الجانب النظري تم التطرق فيها الى أسلوب التحليل العاملي والتحليل التمييزي، أما الجانب التطبيقي فقد تضمن بيانات من الجنسين للمرضى المصابين بمرض السكري، تم الحصول عليها من مستشفى (ليلى قاسم) في اربيل، وتم التوصل الى أن المتغيرات (H.D.L) نسبة كورسترول الجيد (الدهون الثلاثية)(X_3) و (S.CR) مستوى الكرياتين (X_2) و (HBA1C) مستوى هيموكلوبين (X_4)، لهم التأثير الأكبر في الدالة التمييز بين المجموعتين.

Variables, phenomenon, analysis of major components, differential analysis, diabetes, hospital, gender.

1. المقدمة (2) (4) (7):

يعد تحليل المركبات الرئيسية احد اقدم وأوسع الأساليب المستخدمة في تحليل البيانات متعددة المتغيرات، حيث يستخدم في تحليل البيانات ذات العدد الكبير للمتغيرات التوضيحية نسبة الى عدد المشاهدات، كما نلجاء الى تحليل المركبات الرئيسية عندما تكون المتغيرات التوضيحية ذات ارتباط عالي (اي وجود مشكلة التعدد الخطي) بالإضافة الى استخدام تحليل المكونات الرئيسة لتحديد الاسقاط الملائم للبيانات والقيم الشاذة للبيانات و تم استخدام التحليل المميز في عملية التمييز والتصنيف ومن ثم تقدير أفضل معادلة تصنيفية بالأعتماد على القياسات الخاصة والمستخلصة بواسطة المكونات الرئيسية للتمييز بين النوعين.

الجانب النظري

(1-2) التحليل العاملي (8) (9) (12):

يعد التحليل العاملي أحد الأساليب الإحصائية المهمة، ويهدف إلى دراسة الظواهر المعقدة من اجل استخراج أهم العوامل (Extraction of Factors) التي أثرت فيها من خلال تحليل الارتباط بين متغيرات الظاهرة المدروسة.

لقد توسع استخدام أسلوب التحليل العاملي في مجالات الحياة المختلفة نتيجة للانتشار الواسع للمتغيرات الظواهر الطبيعية والتي ترتبط فيما بينها بعلاقات متداخلة، بوصفه نموذجاً رياضياً لتحليل العلاقات بين عدد كبير من تلك المتغيرات وتفسيرها في عدد قليل من العوامل.

(2-2) طريقة المركبات الرئيسية (principle component method) (5) (2) (10):

تعد طريقة المركبات الرئيسية من اكثر طرق التحليل العاملي انتشاراً، وهي طريقة استكشافية يمكن الاستفادة منها للتوصل الى تفسير أو فهم العلاقات المتداخلة بين المتغيرات، وتعمل على تحويل المتغيرات المرتبطة الى متغيرات جديدة مستقلة فيها بينها، وتدعى هذه المتغيرات بالمركبات الرئيسية وقد تم استخدام هذه الطريقة في هذه الدراسة.

الفكرة الأساسية لهذه الطريقة هي إيجاد المركبات الرئيسية، والتي عددها مساو لعدد المتغيرات المدروسة $(X_j) (j = 1, 2, \dots, p)$ والتي تتبع التوزيع الطبيعي المتعدد المتعدد المتغيرات بمتوسط مجتمع (μ) ومصفوفة التباين - التباين المشترك (Σ) أي ان $X \sim N(\mu, \Sigma)$ ومن ثم اختيار أول مركب من المركبات الرئيسية حيث يتم تعديلها لإيجاد النموذج الآتي:

$$\begin{aligned} x_1 &= L_{11}F_1 + L_{12}F_2 + \dots + L_{1m}F_m + U_1 \\ &\vdots \\ x_p &= L_{p1}F_1 + L_{p2}F_2 + \dots + L_{pm}F_m + U_p \end{aligned} \quad \dots(1-2)$$

حيث ان:

F_1, \dots, F_m : تمثل العوامل العامة غير المرتبطة التي تم استخلاصها من (P) من المتغيرات.

L_{ji} : تمثل تحميل العامل (i) للمتغير (j) .

U_1, \dots, U_p : تمثل العوامل الوحيدة.

والسبب في اختيار أول مركب من المركبات الرئيسية هو لتفسيره أكبر نسبة من التباين الكلي للمتغيرات، حيث يعد توليفة خطية للمتغيرات الأصلية التي تسهم بأعلى قدر من التباين الكلي، أما المركبة الرئيسية الثانية فهي تسهم بأعلى ثاني قدر للتباين المتبقي، وهكذا حتى نصل الى مجموع كل التباينات، ولهذا السبب فان المركبات الرئيسية تترتب ترتيباً تنازلياً حسب تبايناتها.

وان كل مركبة رئيسية لها تباين مساوٍ لقيمة الجذر المميز (Eigen Root) للمصفوفة المستخدمة فإذا كان.

$$PC_i = \sum_{j=1}^p a_{ji} X_j \quad \dots (2-2) \quad j = 1, 2, \dots, p$$

حيث ان:

PC_i : تمثل المركبة الرئيسية (i) .

X_j : يمثل المتغير (j) .

a_{ji} : يمثل معامل المتغير (j) للمركبة (i) والتي هي قيم المتجهات المميزة (a_i) (Eigen Vectors)

المرافقة للجذور المميزة (Eigen Roots) للمصفوفة المستخدمة، فان:

$$Var(PC_i) = \lambda_i$$

حيث ان:

$Var(PC_i)$: يمثل تباين المركبة الرئيسية (i).

λ_i : يمثل الجذر المميز للمركبة (i).

ويمكن الحصول على العوامل F_1, F_2, \dots, F_m من خلال قسمة كل مكون رئيسي على الانحراف المعياري لمكون رئيسي ($\sqrt{\lambda_i}$) كما يأتي:

$$F_i = \frac{PC_i}{(Var(PC_i))^{1/2}} = \frac{PC_i}{(\lambda_i)^{1/2}} \quad \dots(3-2)$$

حيث ان:

F_i : يمثل العامل (i).

PC_i : تمثل المركبة الرئيسية (i).

λ_i : يمثل تباين المركبة الرئيسية (i).

ومن نموذج المركبات الرئيسية (2-2) نلاحظ ان (PC_i) هي دالة خطية لـ (X_j) ولجعل (X_j) دالة خطية لـ (PC_i) فأن:

$$X_1 = a_{11}PC_1 + a_{12}PC_2 + \dots + a_{1m}PC_m \quad \dots(4-2)$$

$$X_p = a_{p1}PC_1 + a_{p2}PC_2 + \dots + a_{pm}PC_m$$

ومن معادلة (4-2) يمكن التعبير عن (PC_i) كما يأتي:

$$PC_i = F_i(Var(PC_i))^{1/2} \quad \dots(5-2)$$

ويمكن التعبير عن المعادلة (5-2) ل (zth) من المتغيرات بالشكل الآتي:

$$X_j = a_{1j}F_1(Var(PC_1))^{1/2} + \dots + a_{mj}F_m(Var(PC_m))^{1/2} \quad \dots(6-2)$$

أي ان:

$$j = 1, 2, \dots, p \quad \dots(7-2) \quad L_{ji} = a_{ij}(Var(PC_i))^{1/2} \quad \dots i = 1, 2, \dots, m$$

ويمكن التعبير عن كل متغير (X_j) كما يأتي:

$$X_j = L_{j1}F_1 + L_{j2}F_2 + \dots + L_{jm}F_m + e_j \quad \dots(8-2)$$

وان تباين المتغير (X_j) هو:

$$Var(X_j) = h_j^2 + U_j^2 \quad \dots(9-2)$$

$$h_j^2 = L_{j1}^2 + L_{j2}^2 + \dots + L_{jm}^2 \quad \dots(10-2)$$

وبما ان المتغيرات بصيغتها القياسية فإن:

$$Var(X_j) = 1 = h_j^2 + U_j^2$$

$$U_j^2 = 1 - h_j^2 \quad \text{وهذا يعني:}$$

(3-2) المكونة الرئيسية (Principle Component) (11)(3):

هي تركيبة خطية للمتغيرات العشوائية والميزة المهمة للمكونة الجديدة هي أنها تفسر أكبر نسبة تباين للمتغيرات المدروسة وتساوي مجموع المتغيرات المدروسة بعد ضربها بالمتجهات المميزة.

- 1- المتجهات المميزة (Eigen Vectors): وهي معاملات التركيبة الخطية للمكونات الرئيسية أو يسمى بمعاملات المتغيرات الأصلية للمكونات الرئيسية ويرمز لها بـ (a) .
- 2- الجذور المميزة (Eigen Values): وهي تباين المكونة الرئيسة ويرمز لها بـ (λ_j) .
- 3- التحميل (Loading): التحميل (jj') يمثل معامل الارتباط البسيط بين قيم المكون (j) وقيم المتغير (j') وفق الصيغة الآتية:

$$L_{jj'} = (a_j) \sqrt{\lambda_j} \quad \dots (11-2)$$

(4-2) طرق اختبار عدد المكونات الرئيسية (4) (9) (1)

هناك معايير عدة لاختبار عدد المكونات الرئيسية المهمة (المعنوية) نذكر منها :

1. معيار (Kaiser): الطريقة التي توصل اليها (Kaiser, 1960) وهي تقوم على أساس اختيار عدد المكونات الرئيسية المساوية لعدد الجذور المميزة التي تزيد عن الواحد الصحيح (وتجدر الإشارة الى أن هذا المعيار يستخدم فيما إذا كانت المتغيرات المدروسة تمتلك وحدات قياس مختلفة (أي استخدام مصفوفة الارتباط لحساب المكونات الرئيسية)).
2. معيار العرض البياني: أفرح (Cattle, 1966) طريقة بيانية لتحديد عدد المكونات، حيث يقوم على أساس رسم المنحنى الذي يمثل علاقة عدد المكونات والجذور المميزة ويكون قاعدة القراري أن تستبقي المكونات الرئيسية الى الحد الذي يكون فيه المنحنى تقوسا وترفض المكونات التي يصبح عنها المنحنى بشكل مستقيم.
3. معيار نسبة التباين المفسر: هذا المعيار يعتمد على النسبة المفسرة من التباين الكلي، حيث يتم اختيار عدد المكونات على أساس النسبة التجميعية للتباين المفسرة من قبل المكونات، كما أشار (Green, 1976) الى ان هذه النسبة يجب ان تكون أكبر أو مساوي لـ (80%) بينما أشار كل من (Draper & Smith, 1981) الى ان تفسير (75%) من التباين الكلي يكون كافياً في حين أشار الباحثان (Afifi & Clark, 1984) الى أنه يمكن تحديد عدد المكونات باعتماد رأي الباحث في تحديد العدد حسب قناعته بكمية المعلومات، يستخدم هذا المعيار حينما تكون المتغيرات المدروسة ذات

وحدات قياس مماثلة (أي عند استخدام مصفوفة التباين والتباين المشترك لحساب المكونات الرئيسية).

(5-2) تدوير المحاور (Rotation of Axes) (6) (4) (12):

يهدف التحليل العائلي الى إعطاء صورة واضحة عن طبيعة العلاقات المتداخلة بين المتغيرات من خلال إبراز العوامل الكامنة وراء هذه العلاقات ووصفها وتفسيرها على ضوء البيانات. يتناول تفسير العوامل البحث في طبيعتها واصلها ومدى استقلالها أو ارتباطها، ويتطلب ذلك تحديد الخصائص التي تشترك فيها كل مجموعة من المتغيرات ذات التشعبات العالية بأحد العوامل. ان الهدف الأساس من تدوير المحاور هو الحصول على عوامل تكون معاملاتها سهلة التفسير و ذات دلالة معنوية (لاتتغير من تحليل لآخر) ، ويتوقف أسلوب التدوير على مدى ارتباط العوامل أو استقلالها، وعلى الباحث أن يختار الأسلوب المناسب للتدوير.

من هنا تأتي أهمية تدوير المحاور لتحويل مصفوفة العوامل غير المدورة الى مصفوفة يطلق عليها التركيب البسيط Simple Structure لمصفوفة العوامل المستخلصة، ويتم الحصول عليها بتدوير هذه العوامل بأحد الأسلوبين:

1. أسلوب التدوير المتعامد (Orthogonal Rotation).

2. أسلوب التدوير المائل (Oblique Rotation).

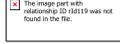
(6-2) اختبار معنوية العوامل (4) (9) (3):

1. اختبار (Burt and Banks): من اكثر الاختبارات استخداماً،الاختبار الذي قدمه كل من (Burt and Banks) لاختبار معنوية تحميلات (تشعبات) العوامل مستخدمين الخطا المعياري للتحميلات (Standard Error of the Loading) لتحديد معنوية التحميلات لكل عامل وحسب الصيغة التالية:

$$... (12-2) \quad (a_{jp}) = s(rx_j, y_j) \sqrt{\left(\frac{k}{k+1-p}\right)^4} \dots (33-2)$$

حيث ان:

هى القيمة الجدولية لمعنوية معامل الارتباط



K: عدد المتغيرات المدروسة

P: عبارة عن ترتيب العامل فى عمليات الاستخلاص (Extraction Process) (ترتيب استخلاصه) ويكون التحميل معنويا للعامل عندما تزيد او تساوى قيمته قيمة الخطأ المعياري لتحميلات ذلك العامل.

2. الاختبار التجريبي: فى هذا الاختبار تكون التحميلات معنوية إذا كانت قيمها أكبر من (± 0.3) بشرط أن تحتوي العينة أكثر من (50) مشاهدة.

3. معيار المقارنة: وهى عبارة عن عملية المقارنة بين تحميلات المكونات لمتغير ما بدءاً من المكون الأول الى آخر مكون مستخلص من عملية التحليل ، بحيث يتم اختيار أكبر تحميل من التحميلات الموجودة فى المكون الخاص بالمتغير المعنى ، ثم يتم مقارنته مع التحميلات بقية المكونات للمتغير نفسه ، فإذا كان أكبر من تحميلات بقية المكونات والمتغير نفسه يكون التحميل معنوياً ، أما إذا كان غير ذلك يعد غير معنوي.

(7-2) التحليل التمييزي (8)(11)(1):

إن التحليل التمييزي (Discriminate Analysis) هو أسلوب إحصائي يتم بموجبه استخدام مجموعة من المتغيرات للتمييز بين مجموعتين أو أكثر عن طريق دالة تمييزية محددة وهى توليفة خطية للمتغيرات التوضيحية وطريقة إيجاد هذه الدالة هو بإيجاد المعاملات للدالة على وفق القياسات أو المعايير التي يتم الحصول عليها من المفردات ، ويعد التحليل التمييزي من الأدوات المهمة فى التحليل الإحصائي لمتعدد المتغيرات (Multivariate Statistical Analysis) والذي يهتم بكيفية التمييز بين مجموعتين أو أكثر من المجتمعات ، إن عملية التصنيف وهى العملية اللاحقة لعملية تكوين الدالة التمييزية إذ يتم الاعتماد على



هذه الدالة بالتنبؤ وتصنيف المفردة الجديدة لإحدى المجموعات قيد الدراسة بأقل خطأ تصنيف ممكن، بمعنى آخر إن التصنيف عرف بأنه جمع للمتشابهات في خصائص الأشياء أو العلاقات بينها في فئات معينة، أما الهدف الرئيس من التحليل التمييزي هو نسبة المشاهدة أو مجموعة من المشاهدات إلى مجاميعها التصنيفية وبأقل خطأ ممكن.

ان التحليل التمييزي يختلف عن تحليل الانحدار في المتغير المعتمد، في التحليل التمييزي هو متغير اسمي (Nominal variable) وهو من المتغيرات النوعية (Qualitative variable) بينما المتغير المعتمد في تحليل الانحدار هو على الاكثر متغير مستمر (Continuous variable)، وهو من المتغيرات الكمية (Quantitative variable)، ولكن التحليلين متشابهان فكلاهما يحاول وصف العلاقة بين المتغير المتعمد والمتغيرات المستقلة عن طريق نموذج خطي.

يعد التمييز والذي يسمى بدالة فيشر (Fisher) طريقة فعالة فيما لو تحققت شروطها الخاصة وهي توزيع المتغيرات التوضيحية توزيعاً طبيعياً وتساوي مصفوفات التباين والتباين المشترك، ويستخدم التحليل التمييزي في عملية التوقع حيث يأتي الباحث بعدة متغيرات يتوقع ان تميز بين المجتمعين المراد دراستهما لنحصل على دالة تمييزية تستخدم في تصنيف المشاهدات بين المجتمعين في المستقبل، ولكي نحصل على اعلى تمييز بين المجموعات ان يكون نسبة التباين بين المجموعات الى التباين داخل المجموعات كبيراً.

(8-2) الفكرة الأساسية (Basic Idea) (5) (3) (12).

بفرض أن فرداً يمكن أن يعود إلى أحد المجتمعين، سنبدأ بالبحث عن طريقة يمكننا من خلالها تصنيف فرد إلى مجتمع من مجتمعين على أساس عدد من المتغيرات (مقاييس الشكل والتوزيع) المقاسة، وبفرض أننا قمنا بسحب عينة من كل مجتمع، وهذا يساعدنا في إيجاد توزيع X ومتوسطاتها وهذه التوزيعات.



(9-2) الدالة التمييزية الخطية للمجموعتين⁽⁴⁾⁽⁷⁾⁽¹¹⁾:

لفرض صياغة مشكلة التحليل التمييزي، نفترض بأن لدينا مجتمعين، المجتمع الأول يُخصص له الرقم (1)، والمجتمع الثاني يُخصص له الرقم (2)، كذلك لدينا n_1, n_2 التي تم اختيارها من كل مجتمع على التوالي.

نفترض أن لدينا قيم مشاهدة لـ m من المتغيرات العشوائية التي يمكن الاعتماد عليها بالتصنيف وهي: X_1, X_2, \dots, X_m لتكون الدالة التصنيفية هي:

$$Z = \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_m X_m \quad \dots(13-2)$$

إذ أن: α : معاملات الأنموذج وتستخدم في عملية التصنيف. m : عدد المتغيرات

إن هذه الدالة تمكننا من تصنيف أية مفردة وتحديد المجتمع الذي تنتمي إليه.

يعد هذا النوع من التمييز من ابسط حالات التمييز تتطلب توفر الشروط الآتية:

- 1- يفترض أن تتوزع المتغيرات التوضيحية توزيعاً طبيعياً متعدد المتغيرات.
- 2- تساوي التباينات لكل المجاميع المدروسة (مصفوفات التباين والتباين المشترك)، أي قبول فرضية العدم عند اختبار الفرضية:

$$H_0 : \Sigma_1 = \Sigma_2 = \dots = \Sigma_k \quad \dots(14-2)$$

H_A : at least two of them are not equal

إذ أن: Σ : مصفوفة التباين والتباين المشترك k : عدد المجاميع

- 3- تصنيف المشاهدات الموجودة إلى مجاميع n_1, n_2, \dots بشكل دقيق.

إن بناء دالة التمييز يتم باختيار المعاملات (α 's) بحيث تعطي أعلى تمييز بين المجموعتين، أي أن الاختلافات بين قيم المجموعتين (Between-group variation) أكبر من الاختلافات بين القيم داخل المجموعتين (Within-group variation).

فإذا رمزنا لنسبة الاختلافات بين المجموعتين إلى الاختلافات داخل المجموعتين بالرمز Q:

$$Q = \frac{[\bar{Z}_{(1)} - \bar{Z}_{(2)}]^2}{\sum_{i=1}^{n_1} [Z_i - \bar{Z}_{(1)}]^2 + \sum_{j=1}^{n_2} [Z_j - \bar{Z}_{(2)}]^2} \dots (15-2)$$

$$Q = \frac{\text{Between - group variation}}{\text{Within - group variation}}$$

اذ إن: $\bar{Z}_{(1)}$: متوسط قيم مشاهدات المجموعة الأولى. $\bar{Z}_{(2)}$: متوسط قيم مشاهدات المجموعة الثانية.
 $\bar{Z}_{(i)}$: قيمة المشاهدة i في المجموعة الأولى. $\bar{Z}_{(j)}$: قيمة المشاهدة j في المجموعة الثانية.

أي اننا نختار (α 's) بحيث تكون (Q) أكبر ما يمكن.

إن الدالة التمييزية الخطية تأخذ الشكل التالي:

$$Z = \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_m X_m \dots (16-2)$$

ويمكن كتابة الدالة بصيغة المصفوفات:

$$Z = \underline{X}' \underline{S}^{-1} [\underline{\bar{X}}_{(1)} - \underline{\bar{X}}_{(2)}] \dots (17-2)$$

$$S\alpha = d \quad \text{علماً أن:}$$

ويمكن إيجاد قيمة (α) كالآتي:

$$\underline{\alpha} = \underline{S}^{-1}[\underline{\bar{X}}_{(1)} - \underline{\bar{X}}_{(2)}] \quad \dots(18-2)$$

إذ إن $\underline{\alpha}$: متجه معامل المتغيرات \underline{X} : متجه المتغيرات.

ويمكن تلخيص خطوات إيجاد الدالة التمييزية في النقاط الآتية:

1. إيجاد متوسط كل متغير في كل مجموعة ثم إيجاد الفرق بين متوسطي كل متغير في المجموعتين:

$$d_i = \bar{X}_{i(1)} - \bar{X}_{i(2)}$$

2. إيجاد مجموع المربعات لكل متغير في كل مجموعة وكذلك مجموع حاصل ضرب كل متغيرين داخل

كل مجموعة: S_{ii}, S_{ij} على التوالي.

3. إيجاد التباين والتباين المشترك المدمج (داخل المجاميع (Within Group).

4. ومن ثم إيجاد الدالة التمييزية: $Z = \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_m X_m$

(10-2) اختبار تساوي مصفوفات التباين والتباين المشترك (اختبار Bartlett) (6) (9) (3):

يعد هذا الاختبار من أهم الاختبارات والتي يجب أن تتحقق لكي نتمكن من استخدام التمييز الخطي،
ويستخدم هذا الاختبار لمعرفة النوع الملائم من النماذج لتمثيل دالة التمييز بين المجموعات وكذلك
النتائج المستخرجة عند الدقة أو مقدار خطأ التصنيف، ولكي يكون تطبيق تحليل التباين صحيحاً لابد
من التأكد من تجانس تباينات المجموعات، حيث أن فرضيتنا الأختبار هي:

$$H_0 : \Sigma_1 = \Sigma_2 = \dots = \Sigma_k$$

H_A : at least two of them are not equal

وأختبار بارتليت (Bartlett Test) وأقترح هذا الاختبار من قبل (Bartlett) في عام (1937) والذي يمتاز
بكفاءة عالية وهو الأكثر شيوعاً، وتكون صيغة هذا الاختبار كما يلي:

$$\mu = \left(\sum_{i=1}^k V_i \right) \ln |S| - \sum_{i=1}^k (V_i \ln |S|) \quad \dots(19-2)$$

$$S = \frac{1}{\sum_{i=1}^k V_i} \sum_{i=1}^k V_i S_i$$

إذ أن: S: مصفوفة التباين والتباين المشترك: S_i : تباين العينة i. $V_i = n_i - 1$: درجة الحرية للعينة i. K: عدد المجموع المختبرة

وقد اثبت Box عام (1949) أنه إذا ضرب (μ) في ثابت C^{-1} والذي يساوى.

$$C^{-1} = 1 - \frac{2m^2 + 3m - 1}{6(m+1)(k-1)} \left[\sum_{i=1}^k \frac{1}{V_i} - \frac{1}{\sum_{i=1}^k V_i} \right] \quad \dots(20-2)$$

سوف نحصل على مقياس يتوزع بالتقريب توزيع (χ^2) وبدرجة حرية $\frac{1}{2}(k-1)m(m+1)$ إذ أن:

m: عدد المتغيرات المدروسة في المصفوفة.

$$\mu = C^{-1} \sim \chi^2 \left[\frac{1}{2}(k-1)m(m+1) \right] \quad \text{أي أن:}$$

ومن ثم نقارن قيمة المقدار (μC^{-1}) المحسوبة مع قيمة (χ^2) الجدولية، فإذا كان المقدار (μC^{-1}) اكبر من (χ^2) الجدولية فإننا نرفض فرضية تجانس التباين وعدا ذلك نقبل الفرضية عند مستوى معنوية (α) .
وستستخدم هذا الاختيار عندما يكون حجم العينة كبيراً اي ان ($N \geq 30$) .

(11-2) نقطة القطع (Cut point) (4) (9) (10):

لجعل دالة التمييز الخطية في أداة لتصنيف الأفراد فإننا نحتاج إلى نقطة فاصلة تفصل بين المجموعتين، ونستخدم المعادلة التالية لإيجاد الدالة التمييزية بعد أن يتم تعيين المتغيرات المهمة

الواجب إدخالها في الدالة، إذ تم إيجاد نقطة القطع (الفصل) بتطبيق دالة الكثافة الاحتمالية والمبينة ادناه:

$$Z = (\bar{X}_1 - \bar{X}_2)' S^{-1} X \quad \dots(21-2)$$

$$Z = \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_m X_m$$

وبافتراض (كمرحلة أولية) أن مفردات المجموعتين تتوزع توزيعاً طبيعياً متعدداً وبمتوسط μ_1

$$X_{j2} \sim N(\mu_2, \Sigma_2) \quad X_{j1} \sim N(\mu_1, \Sigma_1) \quad \Sigma$$

لو أخذنا \bar{X} بدلاً من X نحصل:

$$Z = (\bar{X}_1 - \bar{X}_2)' S_1^{-1} \bar{X}_1 \Rightarrow \bar{y}_1 \bar{Z}_1 \quad \dots(22-2)$$

$$Z = (\bar{X}_1 - \bar{X}_2)' S_1^{-1} \bar{X}_2 \Rightarrow \bar{y}_2 \bar{Z}_2$$

كذلك لو أخذنا $(\bar{X}_1 - \bar{X}_2)$ بدلاً من X ف سنحصل على إحصاء مربع المسافة مهالونوبيس D^2

$$\text{Cut Point (C)} = \frac{\bar{Z}_1 + \bar{Z}_2}{2} \text{ يمكن أن تساوي: Cut Point (CP)}$$

ولتسهيل التصنيف يمكن طرح نقطة الفصل من القيمة التي تعطيها الدالة التمييزية ليكون بذلك نقطة القطع هي الصفر وكما يلي:

$$\therefore W = Z - CP \quad \dots(23-2)$$

$$W = \begin{cases} > 0 \text{ Group I} \\ < 0 \text{ Group II} \\ 0 \text{ Can not classify} \end{cases}$$



والأخيرة تمثل قيمة الدالة التمييزية عند الحد الفاصل بين المجموعتين، فإذا كانت $W > 0$ فإن المفردة تعود إلى المجتمع الأول أما إذا كانت $W < 0$ فإن المفردة تعود إلى المجتمع الثاني، إذ إن (W) تدعى بإحصائية التصنيف التي وضعها (Anderson) عام [1984].

(12-2) خطأ التصنيف (Error Rate) (5) (4) (7):

يمكن تعريف خطأ التصنيف بأنه احتمال تصنيف مفردة (مشاهدة) معينة إلى المجموعة (i) بينما هي في الحقيقة تعود للمجموعة (i') وبالعكس، وهذا التصنيف غير صحيح يحدث عند تحديد نقطة القطع (Cut off Point) بين المجموعتين.

إن خطأ التصنيف هو عامل مهم لإثبات كفاءة الدالة التمييزية، بمعنى آخر أن الدالة التمييزية التي تعطي أقل خطأ تصنيف هي الدالة الأكثر كفاءة وتكون الأفضل من بين دوال التمييز، وهناك طرق عدة لاحتساب خطأ التصنيف ومن هذه الطرق هي:

أولاً: الخطأ الحقيقي:

من الأمور المتفق عليها إحصائياً أنه كلما كبر حجم العينة للمفردات فإنها تتوزع بالتقريب التوزيع الطبيعي وذلك بتطبيق نظرية الغاية المركزية (Central Limit Theorem)، وهذا يجب أن يؤخذ بنظر الاعتبار عند حساب احتمال خطأ التصنيف إذ يعتمد هذا الاحتمال على كون العينة تتوزع توزيعاً طبيعياً أو يقترب من الطبيعي.

$$P_{12} = P(\text{Classify from (1) | } X \text{ from (2)}) \\ = \Phi\left(-\frac{1}{2}\sqrt{D^2}\right) \quad \dots(24-2)$$

$$P_{21} = P(\text{Classify from (2) | } X \text{ from (1)}) \\ = \Phi\left(-\frac{1}{2}\sqrt{D^2}\right) \quad \dots(25-2)$$

إذ إن: D^2 : هي إحصاءة مهالونوبيس

Φ : هي دالة التوزيع الطبيعي ويمكن إيجاد قيمة هذا الاحتمال وهو P_{12} او P_{21} من جداول التوزيع الطبيعي، وبذلك فإن تقدير احتمال خطأ التصنيف يكون:

$$\hat{P}_1 = \hat{P}_2 = \Phi\left(-\frac{1}{2}\sqrt{D^2}\right) \dots (26-2)$$

ويسمى أيضاً الاحتمال عند تقدير المعلمات، ومن الجدير بالذكر أن هذه الطريقة تكون مثالية عندما يكون حجم العينة كبيراً وبعبارة أخرى فإنها تكون متحيزة.

ثانياً: نسبة الخطأ الظاهري:

عند ادخال متغيرات الى الدالة التمييزية قد يحدث خطأ في التصنيف وهذا التصنيف يشير إلى الخلل الموجود في الدالة.

3. الجانب التطبيقي:

إن الجانب التطبيقي هو أهم جزء من البحث وإن الهدف من هذا البحث هو وصف العلاقات التصنيفية بين المتغيرات المدروسة، والحصول على دالة تصنيفية تقديرية يعتمد على بعض المقاييس الإحصائية بحيث يمكن بواسطتها التمييز بين الجنسين المصابين بمرض السكري ، لذا استخدمنا طريقة تحليل المكونات الرئيسية (Principle Component method) لاستخلاص أهم المتغيرات المؤثرة في عملية التصنيف ومن ثم طبقنا التحليل التمييزي (Discriminate Analysis) على المتغيرات المستخلصة، ويتم الاعتماد على مصفوفة التحميلات (Loading Matrix) في تحليل المكونات الرئيسية (Principle Component method) كل مرة لاختيار المتغيرات الداخلة في التحليل لبيان تأثير (Principle Component method) في عملية التصنيف، وذلك بالاعتماد على نسبة خطأ التصنيف ، وقد تم استخراج نتائج التحليل باستخدام الحاسبة الآلية وذلك من خلال اعتماد على البرمجيات الإحصائية الجاهزة (SPSS-V:22) (Statistical Package for Social Sciences).

(1-3) وصف البيانات:

لقد تم جمع البيانات من مستشفى ليلي قاسم في أربيل بالاعتماد على مجموعة من الفحوصات الطبية حيث سحبت عينة عشوائية مكونة من (100) مريضاً من كلا الجنسين المصابين بمرض السكري , وعدد المتغيرات الإضافية وهي (8) متغيرات يمكن توضيحها بالشكل الآتي:

1. الجنس (X_1).
2. (S.CR) مستوى الكرياتين في الدم (X_2).
3. (H.D.L) نسبة كورسترول الجيد أو (الدهن الثلاثية) (X_3).
4. (HBA1C) مستوى هيموكلوبين A1C (نسبة التراكمي) (X_4).
5. (B.Su) نسبة السكر في الدم (X_5).
6. (S.Cho) نسبة الكورسترول في الدم (X_6).
7. (B.Ur) نسبة اليوريا في الدم (X_7).
8. (T.G) نسبة التراكليسيراييد (X_8).

(2-3) اختبار التوزيع الطبيعي وتجانس التباين للمتغيرات المدروسة:

أن كثيراً من التحليلات والاختبارات الاحصائية تفترض بأن مشاهدات العينة يجب أن تكون مختارة من مجتمع يتبع توزيعاً طبيعياً، وتم استخدام احصاءة (Kolmogorov-Smirnov) لمعرفة إذا كانت البيانات تتبع التوزيع الطبيعي أم لا ؟ ويجب وضع الفرضيتين الاتيتين لأختبار ذلك:

H_0 : البيانات تتوزع طبيعياً.

H_1 : البيانات لا تتوزع طبيعياً.

ولاتخاذ القرار الاحصائي نعتمد على القيمة الاحتمالية ($\text{Sig.}=\text{P-value}$) فإذا كانت أقل من مستوى

المعنوية ($\alpha = 0.05$) فإننا نرفض فرضية العدم.

الجدول (3-1) اختبار(Kolmogorov-Smirnov) للتوزيع الطبيعي للبيانات

Tests of Normality						
المتغيرات	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
B.Su (نسبة السكر في الدم)	0.085	100	0.073	0.928	100	0.000
S.Cho (نسبة الكورسترول (في الدم)	0.058	100	0.200	0.975	100	0.055
S.CR (مستوى الكرياتين)	0.094	100	0.031	0.968	100	0.017
H.D.L (نسبة كورسترول (الجيد)	0.085	100	0.072	0.926	100	0.000
HBA1C (مستوى هيموكلوبين)	0.071	100	0.200	0.977	100	0.074
B.Ur (نسبة اليوريا)	0.078	100	0.136	0.970	100	0.021
T.G (نسبة التراكليسيريد)	0.088	100	0.054	0.910	100	0.000

نلاحظ في الجدول (3-1) بأن المتغيرات التي قيمتها الاحتمالية ($Sig. \geq 0.05$) هي متغيرات تتوزع طبيعياً و تم تظليلها ماعدا متغير واحد وهي (S.CR مستوى الكرياتين) ($Sig. < 0.05$) وهي قريب من توزيع طبيعي.

(3-3) نتائج المركبات الرئيسية:

ولإجراء أسلوب التحليل العاملي فقد تم الاعتماد على طريقة المركبات الرئيسية لتحليل مصفوفة الارتباط الخاصة بالمتغيرات (أنظر مصفوفة الارتباط في الملحق رقم (1))، وذلك بسبب أختلاف وحدات قياس المتغيرات الداخلة في التحليل. ومعيار(Kaiser) لأستخلاص المكونات الرئيسية المساوية لعدد الجذور المميزة التي تزيد عن الواحد الصحيح ($\lambda > 1$). ويعتبر المتغير ذي دلالة إحصائية إذا ساوت أو

زادت القيمة المطلقة للتحميل عن (0.5) وإظهار تسلسل المتغيرات من حيث أهميتها وتأثيرها في الظاهرة المدروسة ومدى مساهمتها في تفسير التباين والحصول على مكونات يمكن تفسيرها بسهولة وكانت النتائج كما في الجداول المبينة أدناه.

الجدول (3-2) اختبار مدى كفاية العينة و معنوية مصفوفة الارتباطات

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.497
Bartlett's Test of Sphericity	Approx. Chi-Square	57.269
	Df	21
	Sig.	0.000

نلاحظ في الجدول (3-2) أنه تم إجراء اختبارين قبل البدء باستخدام طريقة تحليل المكونات الرئيسية (PCA) وهي:

1. اختبار (KMO) للمعانة:

يستخدم مقياس (KMO) لتحديد مدى كفاية العينة بحيث يقارن مقادير معاملات الارتباط للمشاهدة بمقادير معاملات الارتباط الجزئي، وتقع قيمة المقياس ضمن المدى $(0 \leq KMO \leq 1)$ ، هنا نلاحظ أن قيمة المقياس تساوي (0.497) أي انه يمكن لنا استخدام تحليل (PCA).

2. اختبار بارتلليت الدائري (Bartlett's Test):

يتم تطبيق اختبار بارتلليت لاختبار المعنوية الكلية لكل الارتباطات داخل المصفوفة، فنلاحظ هنا أن قيمة (χ^2) كبيرة جداً وتساوي (57.269) وبقية احتمالية معنوية (0.000) أي أننا نرفض الفرض (إن كل المتغيرات غير مرتبطة).

مجلة قهلاى زانست العلمية

مجلة علمية دورية محكمة تصدر عن الجامعة اللبنانية الفرنسية - اربيل، كوردستان،
العراق

المجلد (٣) - العدد (٤) ، خريف ٢٠١٨

رقم التصنيف الدولي: ISSN 2518-6558 (Online) - ISSN 2518-6566 (Print)



الجدول (3-3) التحليل العاملي باستخدام طريقة المركبات الرئيسية للبيانات

Total Variance Explained						
Component العوامل	Initial Eigenvalues			Rotation Sums of Squared Loadings		
	Total تباين العامل القيمة المميزة	Variance % of نسبة تباين العامل	Cumulative % النسبة التراكمية للتباين	Total تباين العامل القيمة المميزة	% of Variance نسبة تباين العامل	Cumulative % النسبة التراكمية للتباين
1	1.593	22.753	22.753	1.552	22.172	22.172
2	1.485	21.208	43.961	1.522	21.744	43.916
3	1.141	16.296	60.257	1.144	16.341	60.257
4	0.921	13.160	73.417			
5	0.813	11.612	85.029			
6	0.570	8.150	93.179			
7	0.477	6.821	100.000			

يلاحظ من الجدول المذكور نتيجة لتطبيق التحليل العاملي على مصفوفة الارتباط فقد تم اختيار ثلاثة عوامل معنوية، والتي تمثل عدد العوامل التي تزيد قيمتها المميزة عن الواحد الصحيح فسرت (60.257%) من التباين الكلي. وهذه العوامل الثلاثة تفسر كل منها على التوالي (22.753%، 21.208%، 16.296%) من التباين الكلي.

الجدول رقم (3-4) مصفوفة التحميلات العوامل المعنوية بعد تدوير (Varimax) لعوامل المستخلصة

Rotated Component Matrix ^a				
variables المتغيرات	Component			Communality كمية الشبوع
	1	2	3	
B.Su (نسبة السكر في الدم)	.826	.099	-.096	.701
S.Cho (نسبة الكورسترول في الدم)	.278	.210	.769	.713
S.CR (مستوى الكرياتين)	-.080	.586	.291	.434
H.D.L (نسبة كورسترول الجيد)	-.056	.694	-.013	.485
HBA1C (مستوى هيموكلوبين)	.833	-.177	.110	.737
B.Ur (نسبة اليوريا)	.101	.740	-.227	.609
T.G (نسبة التراكليسرايد)	-.281	-.254	.629	.539

يشير الجدول المذكور اعلاه الى التحميلات المعنوية (المتغيرات المهمة) لكل العامل. ويلاحظ من الجدول إن العامل الاول فسرَ نسبة قدرها 22.753 % من التباين الكلي، وقد ضم هذا العامل متغيرين ذات دلالة إحصائية وحسب التسلسل وهي مستوى هيموكلوبين (X₄)، ونسبة السكر في الدم (X₅) وفسر العامل الثاني نسبة قدرها 21.208 % من التباين الكلي، وضم هذا العامل ثلاث متغيرات ذات دلالة إحصائية وحسب التسلسل وهي: نسبة اليوريا في الدم (X₇) ، ونسبة الكورسترول الجيد (X₃) ، ومستوى الكرياتين في الدم (X₂)، وفسر العامل الثالث نسبة 16.296% من التباين الكلي، وضم هذا العامل متغيرين ذات دلالة إحصائية وحسب التسلسل وهي: نسبة الكورسترول في الدم (X₆) ونسبة التراكليسرايد (X₈).

(4-3) تطبيق التحليل التمييزي لتصنيف المرضى المصابين بمرض السكري:

تم تطبيق التحليل التمييزي باستخدام المتغيرات على مرحلتين بحيث تم:

1. استخدام المتغيرات المستخلصة من المكون الرئيس الأول والثاني (PC₁+PC₂).

2. إستخدام المتغيرات المستخلصة من المكون الرئيس الأول والثاني والثالث ($PC_1+PC_2+PC_3$). وذلك بهدف الوصول إلى أفضل أنموذج يمكن أن نعتمد عليه في عملية التصنيف والتمييز بين الجنسين المصابين بمرض السكري وبالاعتماد على أعلى درجة تصنيف وأقل نسبة للخطأ، والمقارنة بين درجات قوة التصنيف بين الحالات المختلفة وكانت النتائج كما يأتي:

أ. إختبار (Wilks' Lambda):

ويمكن أن نختبر مدى اهمية الدالة التمييزية في التمييز بين المجموعتين باستخدام الأحصاء (Wilks' Lambda) وبالاعتماد على إختبار (Chi-square goodness of fit) كما هو موضح في الجدول (5):

الجدول (3-5) إختبار (Wilks' Lambda) لأهمية الدالة التمييزية ($PC_1 + PC_2$)

Wilks' Lambda				
Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	0.836	17.061	5	0.04

الجدول (3-6) إختبار (Wilks' Lambda) لأهمية الدالة التمييزية ($PC_1 + PC_2 + PC_3$)

Wilks' Lambda				
Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	0.827	17.996	7	0.012

إذ نلاحظ أن القيمة الاحتمالية ($Sig. \leq 0.05$) أي يمكن الاستنتاج بأن هناك فروقاً معنوية بين المجموعتين تعود إلى المتغيرات المستقلة.

ب. إيجاد القيم التقديرية لمعاملات الدالة الخطية التمييزية

يمكن حساب الدالة التمييزية الخطية حيث يتم استخراج قيم معاملات الدالة (α) اعتماداً على الصيغة

$$\alpha = S^{-1} (\bar{X}_1 - \bar{X}_2)$$

تم استخدام المتغيرات المعنوية (الأكثر تحميلاً في (PC1+PC2) في الدالة التمييزية وكانت النتائج كما هو موضح في الجدول (6) وكما يلي:

الجدول (3-7) يبين حساب الدالة التمييزية الخطية لجميع متغيرات (PC1 + PC2)

Var.	Function(Z)
(مستوى الكرياتين) S.CR	-0.627
نسبة كورسترول (H.D.L الجيد)	0.790
مستوى (HBA1C هيموكلوبين)	0.474
نسبة السكر في (B.Su الدم)	0.110
(نسبة اليوريا) B.Ur	-0.213

$$z=0.627z_1 + 0.790 z_2 + 0.474 z_3 + 0.110 z_4 + 0.213 z_5$$

نلاحظ في الجدول أعلاه أن المتغيرات (H.D.L) نسبة كورسترول الجيد (X_3) و (S.CR) مستوى الكرياتين (X_2) و (HBA1C) مستوى هيموكلوبين (X_4) , لهم التأثير الأكبر في الدالة التمييزية لمجموعة الأولى من المتغيرات الأخرى.

مجلة قهلاى زانست العلمية

مجلة علمية دورية محكمة تصدر عن الجامعة اللبنانية الفرنسية - اربيل، كوردستان،
العراق

المجلد (٣) - العدد (٤) ، خريف ٢٠١٨

رقم التصنيف الدولي: ISSN 2518-6558 (Online) - ISSN 2518-6566 (Print)



الجدول (3-8) يبين حساب الدالة التمييزية الخطية لجميع متغيرات (PC1 + PC2+PC3)

Var.	Function(Z)
B.Su (نسبة السكر في الدم)	0.131
S.Cho (نسبة الكورسترول (في الدم)	0.108
S.CR (مستوى الكرياتين)	-0.629
H.D.L (نسبة كورسترول (الجيد)	0.766
HBA1C (مستوى هيموكلوبين)	0.452
B.Ur (نسبة اليوريا)	-0.169
T.G (نسبة التراكليراييد)	0.233

$$z=0.131 z_1 + 0.108 z_2 + 0.629 z_3 + 0.766 z_4 + 0.452 z_5 + 0.169 z_6 + 0.233 z_7$$

نلاحظ في الجدول أعلاه أن المتغيرات (H.D.L) نسبة كورسترول الجيد (X_3) و (S.CR) مستوى الكرياتين (X_2) و (HBA1C) مستوى هيموكلوبين (X_4) , لهم التأثير الأكبر في الدالة التمييزية لمجموعة الغائية من المتغيرات الأخرى.

ت. الدالة التمييزية بحسب متوسطات المجموعات (نقطة القطع):

ويمكن ملاحظة الدالة الخطية التمييزية مقيمة بحسب متوسطات المجموعات كما هو موضح في الجدول (3-9) والجدول (3-10):

الجدول (3-9) الدالة التمييزية بحسب متوسطات (PC1 +PC2)

Functions at Group Centroids	
FACTOR	Function(1)-Cut point
male	-0.447
female	0.429

الجدول (3-10) الدالة التمييزية بحسب متوسطات (PC1 +PC2+PC3)

Functions at Group Centroids	
FACTOR	Function(1)-Cut point
male	-0.463
female	0.444

إذ نلاحظ أن هناك متوسطين الأول (الذكور) والثاني (الإناث) قيمة المتوسطين غير متساويتين بسبب (عدم تساوي أفراد العينتين) ويمكن تفسير الإشارة الموجبة (+) بإرتفاع درجات المتغيرات الداخلة في التحليل تؤدي إلى إرتفاع احتمالات الانضمام إلى المجموعة والثاني (الإناث) وكذلك الإشارة السالبة (-) تعني أن ارتفاع درجات المتغيرات الداخلة في التحليل يؤدي إلى إرتفاع احتمالات الانضمام إلى المجموعة الأول (الذكور)، أي أن نقطة التقاطع تساوي CP=0.444 لمجموعة (1) ونقطة التقاطع تساوي CP=0.429 لمجموعة (2)

وللحصول على معادلة التصنيف (W) يمكن دمج نقطة القطع (CP) مع الدالة التمييزية (Z) وكما يلي:

$$\therefore W = Z - CP$$

مجلة قهلاى زانست العلمية

مجلة علمية دورية محكمة تصدر عن الجامعة اللبنانية الفرنسية - اربيل، كوردستان،

العراق

المجلد (٣) - العدد (٤)، خريف ٢٠١٨

رقم التصنيف الدولي: ISSN 2518-6558 (Online) - ISSN 2518-6566 (Print)



ولتصنيف أي مفردة بالاعتماد على الدالة (W) فإننا نعوض قيم المتغيرات التابعة لهذه المفردة في المعادلة فإذا كانت $W > 0$ فإنها تعود إلى المجموعة الأولى أما إذا كانت $W < 0$ فإن المفردة تعود إلى المجموعة الثانية.

(5-3) أخطاء التصنيف:

إن خطأ التصنيف هو احتمال تصنيف مفردة معينة إلى المجموعة الأولى بينما هي في الحقيقة تعود للمجموعة الثانية وبالعكس وهناك نوعان من الخطأ الأول: نسبة الخطأ الظاهر والثاني هو نسبة الخطأ الحقيقي.

مجلة قهلاى زانست العلمية

مجلة علمية دورية محكمة تصدر عن الجامعة اللبنانية الفرنسية - اربيل، كوردستان،
العراق

المجلد (٣) - العدد (٤) ، خريف ٢٠١٨

رقم التصنيف الدولي: ISSN 2518-6558 (Print) - ISSN 2518-6566 (Online)



Classification Results					
		gendar	Predicted Group Membership		Total
			غير المصابين	المصابين	
Original	Count	male	36	13	49
		female	19	32	51
	%	male	73.5	26.5	100.0
		female	37.3	62.7	100.0
68.0% of original grouped cases correctly classified.					

جدول (3-11) تصنيف الكتل باستخدام متغيرات المكون الرئيس (PC1+PC2) المعنوية

Classification Results					
		gendar	Predicted Group Membership		Total
			غير المصابين	المصابين	
Original	Count	male	38	11	49
		female	16	35	51
	%	male	77.6	22.4	100
		female	31.4	68.6	100
73.0% of original grouped cases correctly classified.					

جدول (3-12) تصنيف الكتل باستخدام متغيرات المكون الرئيس (PC1+PC2 +PC3) المعنوية



وهكذا تم تطبيق التحليل التمييزي على المتغيرات المعنوية (الأكثر تحميلاً) للمكونات الرئيسية المستخلصة والبالغة (3) مكونات رئيسة وذلك بشكل تجميعي (Cumulative)، أي استخدام المتغيرات الأكثر تحميلاً.

4. الأستنتاجات:

1. نلاحظ في الجدول (1) بأن المتغيرات التي قيمتها الاحتمالية ($Sig. \geq 0.05$) هي متغيرات تتوزع طبيعياً و تم تحليلها ماعدا واحد وهي (S.CR مستوى الكرياتين) ($Sig. < 0.05$) وهي قريب من توزيع طبيعي.
2. عند استخدام أختبار (KMO) للمعينة نلاحظ أن قيمة المقياس تساوي (0.497) أي انه يمكن لنا إستخدام تحليل (PCA) , وقد تم تطبيق اختبار بارتليت لاختبار المعنوية الكلية لكل الارتباطات داخل المصفوفة، نلاحظ هنا أن قيمة (χ^2) كبيرة جداً وتساوي (57.269) وبقيمة إحتتمالية معنوية (0.000) أي أننا نرفض الفرض (إن كل المتغيرات غير مرتبطة).
3. نلاحظنا عند تطبيق (Principle Component method) لأستخلاص أهم المتغيرات والقياسات الخاصة إستخلاص (3) مكونات رئيسة تفسر نسبة (60.257) من التباين الكلي وهي : إن العامل الاول فسر نسبة قدرها 22.753% من التباين الكلي، وقد ضم هذا العامل متغيرين ذات دلالة إحصائية وحسب التسلسل وهي مستوى هيموكلوبين (X_4)، ونسبة السكر في الدم (X_5) وفسر العامل الثاني نسبة قدرها 21.208% من التباين الكلي، وضم هذا العامل ثلاث متغيرات ذات دلالة إحصائية وحسب التسلسل وهي: نسبة اليوريا في الدم (X_7) ، ونسبة الكورسترول الجيد (X_3) ، ومستوى الكرياتين في الدم (X_2)، وفسر العامل الثالث نسبة 16.296% من التباين الكلي، وضم هذا العامل متغيرين ذات دلالة إحصائية وحسب التسلسل وهي: نسبة الكورسترول في الدم (X_6) ونسبة التراكليسرايد (X_8).
4. من خلال اختبار (Wilks' Lambda) لأهمية الدالة التمييزية في التمييز بين المجموعتين , نلاحظ أن القيمة الاحتمالية ($Sig. \leq 0.05$) أي يمكن الاستنتاج بأن هناك فروقاً معنوية بين المجموعتين تعود إلى المتغيرات المستقلة.

5. نلاحظ في الجدول (6) و(7) أن المتغيرات (H.D.L) نسبة كورسترول الجيد (X_3) و (S.CR) مستوى الكرياتين (X_2) و (HBA1C) مستوى هيموكلوبين (X_4) , لهم التأثير الأكبر في الدالة التمييزية لمجموعتين من المتغيرات الأخرى.

المصادر:

1. اثناسيوس، زكريا زكي، والبياتي، عبد الجبار توفيق (1977): "المدخل الى التحليل العاملي"، مطبعة مؤسسة الثقافة العالمية، بغداد-العراق.
2. أحمد، رزطار مغديد: (2005) " المكونات الرئيسية وتحميلاتها مع التطبيق على المستوى التعليم المهني"، رسالة ماجستير، قسم الاحصاء، كلية الادارة والاقتصاد، جامعة صلاح الدين، أربيل- إقليم كردستان، العراق.
3. الجبوري، شلال حبيب، وعبد، صلاح حمزة (2000): " تحليل متعدد المتغيرات". دار الكتب للطباعة والنشر، جامعة بغداد- العراق.
4. حمد، بخشان احمد (2003): " استخدام التحليل العاملي لتحديد مؤشرات الرضا الوظيفي لدى المعلمين في مدينة اربيل". رسالة ماجستير - كلية الادارة والاقتصاد - جامعة صلاح الدين
5. داود، الا عبد الستار(2005): "استخدام الدالة التمييزية وطرق تحديد متغيراتها". رسالة ماجستير - كلية علوم والحاسبات والرياضيات - جامعة الموصل.
6. الهاشمي، ندى محمد جواد (1999): " استخدام التحليل العاملي لتحديد العوامل المؤثرة ذات العلاقة بظهور حالات الاصابة باورام الدماغ". رسالة ماجستير - كلية الادارة والاقتصاد - جامعة المستنصرية، بغداد , العراق.
7. Anderson, T. W. (1984): "An Introduction to Multivariate Statistical Analysis". 2nd- Edition, John Wiley and Sons, New York-U.S.A.
8. Fazzo, L. et al (2008), "Cluster analysis of mortality and malformations in the Provinces of Naples and Caserta (Campania Region)", Ann Ist Super Sanità , Rome, Italy, Vol. (44), No. (1), PP(99-111).
9. Hardle, W.& Simar, L (2003), "Applied Multivariate Statistical Analysis", Berlin and Louvain-la-Neuve, Germany.



10. Jolliffe,(1986):"Principle Component Analysis", Spring-Verlag New York, Inc. USA.
11. Rencher, A. C (2002) ,"Methods of Multivariate Analysis" , Second Edition, John Wiley & sons, New York , USA,..
12. Timm, N.H (2002) ,"Applied Multivariate Analysis" Springer-Verlag ,New York, Inc. USA.

ثوخته:

شيكردنقوى ناماري فرط طوراو باس لة ليكولينقوى داتاكانى نغو بينراوانة و طوزارشت لبيان دكات لة ميانقوى زورترين نغو طوراوانقوى كاريطريان لقسر دياردكة هقوة كة لة ذير ليكولينقوى، ليكولينقوى كشت دقبسنتيت بة يك بستن لة نيوان شيكردنقوى فرط طوراوكان بة بكارهيناني شيكردنقوى نيكهاتة سقرقوةكان (Principle Component method) وشيكردنقوى جياكرقوىي (Discriminate Analysis)، لقبقر نمة نامانجي نغم ليكولينقوى دسنتيشانكردي طرنطرين فاكترط كاريطرقكان لة سقر هقردوو جورى رطقوى تووشبوو بة نقوشى شقكرة لة ميانقوى دسنتيشانكردي نامارقي فاكترط كورنقراوكان بة بكارهيناني شيكردنقوى فاكترقي و هقروها بة بكارهيناني شيكردنقوى جياكرقوة لة كرداري جياكردنقوى و ثولينكردي، ناشان مقردنقردني باشترين هاوكيشقوى ثولينقراو بة كشت بستن بة ثيوقة تايبقكان و كورنقراو بة هوي نيكهاتة سقرقوةكان بو جياكردنقوى لة نيوان هقردوو جورى رطقوى تووشبوو و بقراوركردي لة نيوان نجانماكان لة سقر بنضيقوى ريذقوى هلقوى ثولينقراو و بلاوكرديقوى كاريطقري لة سقر كرداري ثولين، و بؤ جيقجيكردني نامانجي تويدنقوىقوة تويدنقوىقوة دابقش كرا بو لايقنى نيوري كة باسي شيقوازي شيكردنقوى فاكترقي و شيكردنقوى جياكرقوىي دكات، بلام لايقنى ثراكتيكي برينيق لة داتاي نقوشة تووشبووكان بة نقوشى شقكرة لة هقردوو جورى رطقوىقوة دقت كقوت لة نقوشانقوى (ليلى قاسم) لة هقولير و نجانماكانمان دقت كقوت كة بة نيزانيني نيمطرنقطة ودقبت نييني بكري لة لايقن لايقنقوة تندروسنيقكان.

Abstract:

The multivariate statistical study deals with those views and expressed through the most influential variables in the phenomena under study, research on the link between the adopted multivariate analysis using analysis of the main components of (Principle Component method) and discriminate analysis of (Discriminate Analysis), so this research aims to determine the most important factors affecting patients infected of both Sexes with diabetes determining the

number of factors derived using the analysis of the factor as well as using the distinctive analysis in the process of discrimination and classification and then estimating the best equation based on measurements and derived by the main components of discrimination between the two sexes and the comparison of results based on the ratio of error classification and show its effect on the classification process, and to achieve the search of objectives the research partition on the theoretical side was addressed to the analysis of the functioning and discriminatory analysis, the applied side included data for patients with infected of both Sexes with diabetes, obtained from (Layla Qasim Hospital) in Erbil have been reached to conclusions that we believe are important and should be noted by health providers.

الملحق رقم (1) مصفوفة الارتباطات

	Correlation Matrix						
	S.CR	H.D.L	HBA1C	B.Su	S.Cho	B.Ur	T.G
S.CR	1.000	.093	-.041	-.165	-.102	.146	.101
H.D.L	.093	1.000	-.094	.185	.251	.228	.006
HBA1C	-.041	-.094	1.000	.491	.154	.077	.296
B.Su	-.165	.185	.491	1.000	.175	.081	.169
S.Cho	-.102	.251	.154	.175	1.000	-.035	.433
B.Ur	.146	.228	.077	.081	-.035	1.000	-.092
T.G	.101	.006	.296	.169	.433	-.092	1.000