

تطبيق المركبات الرئيسية والتحليل المميز لتمييز بين المرضى المصابين وغير المصابين بالاورام الخبيثة

(PP 41 - 57)

<http://dx.doi.org/10.21271/zjhs.22.1.4>**م. قطنية محمد حمدامين****م. هدى قرداغ يلددا**

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

qutnia.hamad@su.edu.krd

huda.yalda@yahoo.com

الاستلام: 2017/05/29**القبول: 2017/08/27****النشر: 2018/03/26****ملخص**

يعتمد أسلوب التحليل الإحصائي متعدد المتغيرات بأساليبه المختلفة على وصف وتحليل الظواهر ذات الأبعاد والمتغيرات المتعددة. فإذا كانت المشاهدات تشترك فيما بينها بمجموعة من الخصائص والصفات بدرجات متفاوتة، فإن التحليل الإحصائي متعدد المتغيرات يتناول دراسة بيانات تلك المشاهدات والتعبير عنها من خلال أكثر المتغيرات تأثيراً في الظاهرة تحت الدراسة، اعتمد البحث على الربط بين تحليل متعدد المتغيرات باستخدام تحليل المكونات الرئيسية (PCA) والتحليل التمييزي (DA)، لذا يهدف هذا البحث الى تحديد أهم العوامل المؤثرة على المرضى المصابين وغير المصابين بالاورام الخبيثة 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 الرئيسية للتمييز بين النوعين والمقارنة بين النتائج على أساس نسبة خطأ التصنيف وبيان تأثيرها على عملية تصنيف. ولتحقيق اهداف البحث قسم البحث الى الجانب النظري تم التطرق فيها الى أسلوب التحليل العاملي والتحليل التمييزي، اما الجانب التطبيقي فقد تضمن بيانات للمرضى المصابين وغير المصابين بالاورام الخبيثة تم الحصول عليها من مستشفى رزكري ومستشفى نانة كلي في اربيل تم التوصل الى استنتاجات نعتقد انها مهمة وتجدر ملاحظتها من قبل الجهات الصحية.

1- المقدمة

يستخدم لتحويل مجموعة من المتغيرات المرتبطة الى عدد اقل من المتغيرات بالمركبات الرئيسية، كما نلجأ الى تحليل المركبات الرئيسية عندما تكون المتغيرات التوضيحية ذات ارتباط عالي (اي وجود مشكلة التعدد الخطي) بالإضافة الى استخدام تحليل المكونات الرئيسية لتحديد الاسقاط الملائم للبيانات والقيم الشادة للبيانات.

2- الجانب النظري**2-1 تحليل المكونات الرئيسية**

يعد تحليل المكونات الرئيسية (Principal Component Analysis-PCA) أحد الأساليب المهمة في دراسة مجموعة كبيرة من المتغيرات وخاصة المتغيرات المرتبطة فيما بينها بعلاقات متداخلة والتي تسمى بتعدد العلاقة الخطية (Multi collinearity) ويمكن تعريف تحليل المكونات الرئيسية بأنه (طريقة رياضية تلتخص في تحويل مجموعة كبيرة من المتغيرات المترابطة إلى مجموعة صغيرة من المتغيرات غير المترابطة والمتمثلة بتركيبية خطية لمجموعة من المتغيرات الأصلية، وهذه التركيبية تحتوي على نسبة كبيرة من المعلومات الأصلية) وتتلخص طريقة المكونات الرئيسية بتحويل المتغيرات التوضيحية التي تعاني من ارتباط خطي إلى تراكيب خطية تسمى المكونات الرئيسية بحيث تفسر معظم التباين الكلي للمتغيرات الأصلية ويتم بناء المكونات الرئيسية بحيث تستوفي الشرطين الآتيين:-

1- المكونات الرئيسية غير مرتبطة (متعامدة Orthogonal).

2- المكون الرئيسي الأول يفسر أكبر نسبة من تباينات المتغيرات التوضيحية، أما المكون الرئيسي الثاني فيفسر ثاني أكبر نسبة من التباينات، وهكذا بالنسبة لبقية المكونات الرئيسية على المتغيرات التوضيحية الأخرى.

اولا- المتجه المميز والقيم المميزة (Eigen Vector & Eigen Value)

الجدور المميزة لمصفوفة مثل A ذات بعد (p,p) هي حل للمعادلة المميزة: $|A - \lambda I| = 0$ بحيث إن: λ : مصفوفة أحادية ذات رتبة (p,p): λ : معلمة المعادلة (الجدور المميزة).الحل للمعادلة أعلاه هو متعدد الحدود (Polynomial) بدرجة p للمعلمة λ ، و لهذا فإن A لها p من الجدور المميزة أي:



$$\lambda^p + C_{n-1}\lambda^{p-1} + \dots + C_1\lambda + C_0 = 0$$

وعند حل هذه المعادلة نحصل على p من الجذور المميزة $\lambda_1, \lambda_2, \dots, \lambda_p$ ، و إن لكل جذر مميز متجه مميز (V_i) ببعده $(p,1)$ مرافق له، ويتم إيجاده بعد التعويض بقيمة λ_i في العلاقة: $(A - \lambda_i I)X = 0$ بمعنى يوجد متجه مثل X بحيث إذا ضرب بالمصفوفة المربعة A لنتج المتجه X نفسه مضروباً بمقدار عددي λ أي $AX = \lambda X$

ثانياً- نموذج المكونات الرئيسية:

نقصد بالمكونات الرئيسية تلك المتغيرات التي تم الحصول عليها من تحويل المتغيرات التفسيرية (الأصلية) X_i ($i=1,2,\dots,p$) المرتبطة خطياً إلى تراكيب خطية (PC_j (Linear Combinations) ($j=1, 2, \dots, k$)، قليلة ($p > k$)، مشتقة من المتغيرات الأصلية (X) لتحل محلها بحيث تكون مؤهلة لتفسير معظم التباين الكلي للقيم الأصلية، وتكون هذه المكونات متعامدة - أي لا يوجد ارتباط بينها.

$$Y = X\beta + \varepsilon$$

نفرض النموذج الخطي التالي:

Y : متجه متغير الاستجابة، ببعده $(n,1)$

X : مصفوفة المتغيرات التفسيرية ذات بعد p من المتغيرات n من المشاهدات، (n,p)

β : متجه معاملات الانحدار، ببعده $\varepsilon(p,1)$: متجه الخطأ العشوائي، ببعده $(n,1)$

ويمكن كتابة المكونات الرئيسية (PC) كما يلي:

$$PC_j = a_{1j}x_1 + a_{2j}x_2 + \dots + a_{pj}x_p, \quad j = 1, 2, \dots, p$$

$$PC_j = Xa_j$$

وبشكل عام يمكن كتابة المكون الرئيسي z بالتركيب الخطي الآتي:

$$PC_j = \sum_{k=1}^p a_{kj} X_k, \quad j, k = 1, 2, \dots, p$$

إذ إن: PC_j : المكون الرئيسي j .

a_{kj} : معامل المتغير k في المكون z وهي قيم المتجهات المميزة a_j (Characteristic Vectors) وقد تسمى (Eigen Vectors) المرافقة للجذور المميزة (λ_j) (Characteristic Roots) وقد تسمى (Eigen Values).

$$(PC = XA) \quad \text{أو بلغة المصفوفات}$$

إذ إن: X : متجه المتغيرات المدروسة ذات بعد $A: 1 \times p$: مصفوفة معاملات المتغيرات المدروسة ذات بعد $p \times p$.

PC : متجه المكونات الرئيسية ذات بعد $1 \times p$

ثالثاً - كيفية احتساب المكونات الرئيسية

تعتمد فكرة احتساب المكونات الرئيسية على خصائص الجذور المميزة و ما يرافقها من المتجهات المميزة، إن وحدات قياس المتغيرات التفسيرية التي توضحها المصفوفة X لها دور كبير في احتساب المكونات الرئيسية من حيث كونها وحدات قياس متشابهة أو مختلفة، عندما تكون كل المتغيرات التوضيحية (X 's) لها نفس وحدات القياس فإنه يتم استخراج المكونات الرئيسية اعتماداً على مصفوفة التباين و التباين المشترك للمتغيرات (X 's)، لتكون معاملات المكونات الرئيسية (a_{ij}) قيم المتجهات المميزة لمصفوفة التباين و التباين المشترك (V).

عندما تكون للمتغيرات X 's وحدات قياس مختلفة يستحسن في هذه الحالة تحويل تلك المتغيرات إلى متغيرات قياسية Z 's

$$Z_i = \frac{(X_i - \bar{X})}{\sqrt{V_{ii}}}, \quad i = 1, 2, \dots, p$$

و في هذه الحالة فإن معاملات المكونات الرئيسية (a_{ij}) هي قيم المتجهات المميزة لمصفوفة الارتباط R للمتغيرات X 's بصيغتها القياسية.



رابعا- اختبار كفاية العينة (KMO) لإستخدام تحليل (PCA)

مقياس (Kaiser-Meyer-Olkin measure) (KMO) لتحديد مدى كفاية العينة يقارن مقادير معاملات الارتباط المشاهدة (بمقادير

$$KMO = \frac{\sum_{i \neq j} \sum_j r_{kj}^2}{\sum_{i \neq j} \sum_j r_{kj}^2 + \sum_{i \neq j} \sum_j r_{k,j-1,2,\dots,k,j,\dots,p}^2}$$

بحيث أن: r_{kj} : معامل الارتباط البسيط بين المتغيرين k, j

معامل الارتباط الجزئي فإن قيمة مقياس KMO تكون $0 \leq KMO \leq 1$

خامسا:- تدوير المحاور Rotation of Axes

ان الهدف الاساس من تدوير المحاور هو الحصول على عوامل تكون معاملاتها سهلة التفسير وذات دلالة معنوية، ويتوقف أسلوب التدوير على مدى ارتباط العوامل أو استقلالها، حيث يعمل التدوير على رفع التجميعات الكبيرة وخفض التجميعات الصغيرة. وتم استخدام في هذا البحث أسلوب التدوير المتعامد باستخدام طريقة (Kaiser - Varimax) التي تطبق قانون التباين على مربعات تجميعات العامل (i).

$$Var(F_k) = \frac{1}{p} \left[\sum_{j=1}^p (a_{jk}^2)^2 - \frac{1}{p} \left(\sum_{j=1}^p a_{jk}^2 \right)^2 \right], \quad k = 1, \dots, m$$

حيث ان: $Var(F_k)$: يمثل تباين العامل (k) ، a_{jk} : تمثل قيمة تشعب المتغير (j) بالعامل (k) ، p : عدد المتغيرات.

سادسا:- عدد العوامل المعنوية Number of significant Factors

هنالك عدة معايير لتحديد العوامل المعنوية ومن اهمها طريقة (Kaiser) حيث توصل العالم Kaiser الى اساس علمي لاجاد عدد العوامل المعنوية الضرورية لتوضيح العلاقة بين المتغيرات، حيث بين ان عدد العوامل المشاعة (Common Factors) مساوي الى عدد القيم المميزة (Eigen Values) والتي تزيد قيمتها عن الواحد الصحيح ($\lambda > 1$) (Afifi and Azen 1979). وقد اشارة الى اهمية راي الباحث في اختيار عدد العوامل مع الاخذ بنظر الاعتبار ارتفاع نسبة التباين المفسر من التباين الكلي. أما لاختبار معنوية التجميعات العوامل، فقد تم في هذا البحث بالاعتبار تحميل (تشعب) معنوي إذا كان أكبر أو يساوي $(0.5\bar{F})$.

2- للتحليل التمييزي

يعد التحليل التمييزي (Discriminate Analysis) من الأدوات المهمة في التحليل الإحصائي لمتعدد المتغيرات (Multivariate Statistical Analysis) والذي يهتم بكيفية التمييز بين مجموعتين أو أكثر من المجتمعات، بمعنى آخر إن التحليل التمييزي هو أسلوب إحصائي يتم بموجبه استخدام مجموعة من المتغيرات للتمييز بين مجموعتين أو أكثر عن طريق دالة تمييزية محددة وهي توليفة خطية للمتغيرات التوضيحية وطريقة إيجاد هذه الدالة هو بإيجاد المعاملات للدالة على وفق القياسات أو المعايير التي يتم الحصول عليها من المفردات، إن عملية التصنيف وهي 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.

اولا- الفكرة الأساسية Basic Idea

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.

ثانيا- الدالة التمييزية الخطية للمجموعتين

لغرض 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

$$Z = \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_m X_m \quad \dots \text{ الدالة التصنيفية هي:}$$

إذ أن: α : معاملات النموذج 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 أبسط حالات التمييز تتطلب توفر الشروط الآتية:

1- يفترض أن توزع المتغيرات التوضيحية توزيعاً طبيعياً متعدد المتغيرات.

2- تساوي التباينات لكل المجاميع المدروسة (مصفوفات التباين والتباين المشترك)، أي قبول فرضية العدم عند اختبار

الفرضية:

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

$$H_A : \text{at least two of them are not equal}$$

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).

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

$$Q = \frac{\text{Between - group variation}}{\text{Within - group variation}} = Q = \frac{[\bar{Z}_{(1)} - \bar{Z}_{(2)}]^2}{\sum_{i=1}^{n_1} [Z_i - \bar{Z}_{(1)}]^2 + \sum_{j=1}^{n_2} [Z_j - \bar{Z}_{(2)}]^2}$$

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

ويمكن كتابة الدالة بصيغة المصفوفات: $Z = \underline{X}' \underline{S}^{-1} [\bar{X}_{(1)} - \bar{X}_{(2)}]$

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

ويمكن إيجاد قيمة (α) كالآتي: $\alpha = \underline{S}^{-1} [\bar{X}_{(1)} - \bar{X}_{(2)}]$

إذ إن: α : متجه معامل المتغيرات 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

$$d_i = \bar{X}_{i(1)} - \bar{X}_{i(2)} \quad \text{المجموعتين:}$$

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

$$S_{ii}, S_{ij} \quad \text{على التوالي.}$$

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

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



ثالثا- اختبار تساوي مصفوفات التباين والتباين المشترك (Bartlett Test)

يعد هذا الاختبار من أهم الاختبارات والتي يجب أن تتحقق لكي تتمكن 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:

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

$$H_A : \text{at least two of them are not equal}$$

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

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

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

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

رابعا- نقطة القطع Cutoff point

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 :

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

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

وبافتراض (كمرحلة أولية) أن مفردات المجموعتين تتوزع توزيعاً طبيعياً متعددًا وبمتوسط μ_1 و μ_2 وبمصفوفة تباين

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

لو أخذنا \bar{X} بدلاً من X نحصل:

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

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

كذلك لو أخذنا $(\bar{X}_1 - \bar{X}_2)$ بدلاً من X ف سنحصل على إحصاءة مربع المسافة مهالونويس D²

$$Cutoff Point (CP) = \frac{\bar{Z}_1 + \bar{Z}_2}{2}$$

إذاً نقطة القطع (C) Cutoff Point يمكن أن تساوي:

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

$$\therefore W = Z - CP$$

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

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.



خامسا- خطأ التصنيف Error Rate

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 Cutoff 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:

أ- 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.

$$P_{12} = P(\text{Classify from (1)} | X \text{ from (2)})$$

$$= \Phi\left(-\frac{1}{2}\sqrt{D^2}\right)$$

$$P_{21} = P(\text{Classify from (2)} | X \text{ from (1)})$$

$$= \Phi\left(-\frac{1}{2}\sqrt{D^2}\right)$$

إذ إن D^2 : هي إحصاءة مهالونويس

Φ : هي دالة التوزيع الطبيعي ويمكن إيجاد قيمة هذا الاحتمال وهو P_{12} أو P_{21} من جداول التوزيع الطبيعي، وبذلك فإن تقدير

$$\hat{P}_1 = \hat{P}_2 = \Phi\left(-\frac{1}{2}\sqrt{D^2}\right) \quad \text{احتمال خطأ التصنيف يكون:}$$

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

ب - 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 الهدف من هذا البحث هو وصف العلاقات التصنيفية بين المتغيرات المدروسة، والحصول على دالة تصنيفية تقديرية يعتمد على بعض المقاييس الأحصائية بحيث يمكن بواسطتها التمييز بين المرضى المصابين وغير المصابين بالاورام والخبيثة، لذا استخدمنا (PCA) لاستخلاص أهم المتغيرات المؤثرة في عملية التصنيف ومن ثم طبقنا التحليل التمييزي (LDA) على المتغيرات المستخلصة، ويتم الاعتماد على مصفوفة التحميلات (Loading Matrix) في (PCA) كل مرة لاختيار المتغيرات الداخلة في التحليل لبيان تأثير (PCA) في عملية التصنيف، وذلك بالاعتماد على نسبة خطأ التصنيف.

اولا - وصف البيانات

تم جمع البيانات من كل مستشفى زركاري ومستشفى نانة كلي في اربيل بالاعتماد على مجموعة من الفحوصات الطبية حيث سحبت عينتين عشوائية مكونة من (75) و (60) مريضاً يمثلون مجموعتين من المرضى المصابين وغير المصابين بالاورام الخبيثة، حيث تمثل المجموعة الاولى فحوصات المرضى غير المصابين و المجموعة الثانية تمثل فحوصات المرضى المصابين بالاورام الخبيثة واعتبرت الفحوصات كمتغير استجابة فقد اعطيت القيمة (1) للمجموعة الاولى والقيمة (2) للمجموعة الثانية اماعد المتغيرات الاضافية وهي (7) متغيرات يمكن توضيحها بالشكل الاتي :

1-(كلوكوز) نسبة السكر في الدم: X1 2- يوريا: X2 3- مستوى الكرياتينين: X3 4- نسبة الكولسترول في الدم: X4 5- الدهون الثلاثية: X5 6- الدهون العالية: X6 7- الدهون الواطئة: X7



تم تحليل النتائج عن طريق ادخال البيانات في البرنامج الاحصائي الجاهز (SPSS (Version 20).

ثانيا - إختبار التوزيع الطبيعي وتجانس التباين للمتغيرات المدروسة

مما لا شك فيه 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 (Sig.=P-value) فإذا كانت أقل من مستوى المعنوية ($\alpha=0.05$) فإننا نرفض فرضية العدم، بالنسبة لتجانس التباين هي أيضاً إحدى الفروض الضرورية والمهمة التي يجب توافرها قبل لإجراء التحليل الاحصائي للبيانات (تحليل المكونات الرئيسية والتحليل التمييزي) وهنالك اختبار الفرضيتين الأتيتين:

H_0 : التباينات في المجموعة (1) متجانسة مع تباينات المجموعة (2)

H_1 : التباينات في المجموعة (1) غير متجانسة مع تباينات المجموعة (2)

وتم استخدام احصاءة (Levene Test) لأختبار تجانس التباين للمتغيرات المدروسة عى أساس النوعين من الكتل (مجتمعين) ويوضح الجدولان (1)(2) نتائج الاختبارين:

جدول (1) اختبار (Kolmogorov-Smirnov) للتوزيع الطبيعي للبيانات

Var.	Normal Parameters(a,b)		Most Extreme Differences			Kolmogorov-Smirnov Z	Asym. Sig.
	Mean	SD	Abs.	Posi.	Nega.		
كلوكوز	33.572	9.748	0.79	0.79	-0.045	.079	0.037
يوريا	0.8456	0.2586	0.068	0.068	-0.029	.068	0.200
مستوى الكرياتينين	25.574	7.8596	0.144	0.144	-0.080	.144	0.000
كولسترول	27.880	10.984	0.120	0.120	-0.073	.120	0.000
الدهون الثلاثية	0.7844	.3344	0.086	0.086	-0.055	0.086	0.017
الدهون العالية	134.80	60.532	0.172	0.172	-0.157	0.172	0.000
الدهون الواطئة	183.83	34.074	0.069	0.069	-0.015	0.069	0.200

نلاحظ في الجدول (1) بأن المتغيرات التي قيمتها الاحتمالية ($Sig. \leq 0.05$) هي متغيرات لا تتوزع طبيعياً و تم تظليلها.

جدول(2) أختبار(Levene Test) لتجانس التباينات للمتغيرات المدروسة

Var.	F	df1	df2	Sig.
كلوكوز	6.391	1	133	.013
يوريا	1.389	1	133	.241
مستوى الكرياتينين	3.425	1	133	.066
كولسترول	2.167	1	133	.143
الدهون الثلاثية	1.676	1	133	.198
الدهون العالية	.001	1	133	.974
الدهون الواطئة	.191	1	133	.663

نلاحظ من الجدول(2) بأن مجموعة من المتغيرات التي قيمتها الاحتمالية لها ($Sig. \geq 0.05$) وهذا يعني المتغيرات متجانسة التباين اي لا يمكن رفض فرضية العدم ولمعالجة مشكلة عدم توزيع الطبيعي للمتغيرات المدروسة تم إجراء التحويل اللوغاريتمي ذات الأساس 10 (Log_{10} Transformation) الذي أدى بدوره الى معالجة المشكلة بدرجة كبيرة كما هو موضح في الجدول ادناه.



جدول (3) اختبار (Kolmogorov-Smirnov) للتوزيع الطبيعي للبيانات بعد تحويل (Log) للبيانات

Var.	Normal Parameters(a,b)		Most Extreme Differences			Kolmogorov-Smirnov Z	Asym. Sig.
	Mean	SD	Abs.	Posi.	Nega.		
كلوكوز	1.5076	.12815	0.79	0.47	-0.079	.079	0.057
يوربا	0.8456	0.2586	0.068	0.068	-0.029	.068	0.200
مستوى الكرياتينين	1.3898	.12351	0.088	0.088	-0.059	.088	0.012
كولسترول	1.4150	.16133	0.051	0.051	-0.047	.051	0.200
الدهون الثلاثية	-1.1707	.20447	0.074	0.041	-0.074	0.074	0.065
الدهون العالية	2.0983	.15609	0.111	0.111	-0.089	0.111	0.000
الدهون الواطئة	183.837	34.0744	0.069	0.069	-0.015	0.069	0.200

نلاحظ في الجدول (3) بان مجموعة من المتغيرات التي قيمتها الاحتمالية لها ($Sig. \geq 0.05$) وهذا يعني ان البيانات تتوزع التوزيع الطبيعي.

ثالثا - 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 اختلاف 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.

جدول (4) اختبار مدى كفاية العينة و معنوية مصفوفة الارتباطات

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.589
Bartlett's Test of Sphericity	Approx. Chi-Square	101.340
	Df	21
	Sig.	.000

نلاحظ في الجدول (4) أنه تم إجراء اختبارين قبل البدء باستخدام تحليل (PCA) وهي:

1- اختبار (KMO) للمعانية:

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 (0.589) أي انه يمكن لنا إستخدام تحليل (PCA).

2- اختبار بارنليت الدائري (Bartlett's Test) :

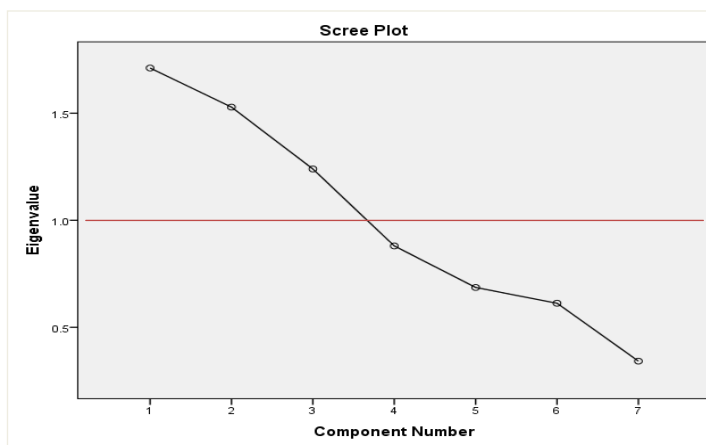
يتم تطبيق اختبار بارنليت لاختبار المعنوية الكلية لكل الارتباطات داخل المصفوفة، فنلاحظ هنا أن قيمة (χ^2) كبيرة جداً وتساوي (101.340) وبقيمة احتمالية معنوية (0.000) أي أننا نرفض الفرض (إن كل المتغيرات غير مرتبطة).



جدول (5) التحليل العاملي باستخدام طريقة المركبات الرئيسية للبيانات

رقم العامل	قيمة المميزة	نسبة تباين العامل %	النسبة التراكمية للتباين
1	1.688	24.121	24.121
2	1.439	20.554	44.675
3	1.237	17.677	62.352
4	.896	12.805	75.158
5	.743	10.615	85.773
6	.631	9.017	94.790
7	.365	5.210	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 (62.352 %) من التباين الكلي. وهذه العوامل الثلاثة تفسر كل منها على التوالي (24.121، %20.554، %17.677)، من التباين الكلي.



شكل (1) عوامل معنوية وغير معنوية

يوضح الشكل أعلاه عدد عوامل معنوية وغير معنوية، حيث تمثل المربعات الواقعة فوق الخط القياسي (المساوية للواحد) والتي تزيد قيمتها المميزة عن الواحد الصحيح عدد عوامل معنوية، أما المربعات التي تقع أسفل الخط تمثل عدد عوامل غير معنوية.

جدول رقم (6) مصفوفة التحميلات العوامل المعنوية بعد تدوير Varimax لعوامل المستخلصة

المتغير	العامل 1	العامل 2	العامل 3
كلوكوز		.701	
يوريا		.762	
مستوى الكرياتينين	.883		
كولسترول	.871		
الدهون الثلاثية		.558	
الدهون العالية			.753
الدهون الواطئة			.799
نسبة تباين العامل %	،%24.121	%20.554	%17.677



يشير الجدول المذكور اعلاه الى التحميلات المعنوية (المتغيرات المهمة) لكل العامل. ويلاحظ من الجدول ان العامل الاول فسّر نسبة قدرها 24.121% من التباين الكلي، وقد ضم هذا العامل متغيرين ذات دلالة إحصائية وحسب التسلسل وهي مستوى الكرياتينين (X_3)، ونسبة الكلوسترول في الدم (X_4) وفسر العامل الثاني نسبة قدرها 20.555% من التباين الكلي، وضم هذا العامل ثلاثة متغيرات ذات دلالة إحصائية وحسب التسلسل وهي: كلوكوز (نسبة السكر في الدم) (X_1)، يوريا (X_2)، الدهون الثلاثية (X_5). أما العامل الثالث فقد بلغت نسبة تفسيره 17.677% من التباين الكلي، وضم هذا العامل متغيرين ذات دلالة إحصائية وحسب التسلسل وهي الدهون العالية X_6 ، الدهون الواطئة X_7 .

رابعا - تطبيق التحليل التمييزي لتصنيف المرضى المصابين وغير المصابين بالاورام الخبيثة

تم تطبيق التحليل التمييزي باستخدام المتغيرات على مرحلتين بحيث تم:

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

2- استخدام المتغيرات المستخلصة من المكون الرئيس الأول والثاني والثالث ($PC1+PC2+PC3$).

وذلك بهدف الوصول إلى أفضل نموذج يمكن أن نعتمد عليه في عملية التصنيف والتمييز بين المرضى المصابين وغير المصابين بالاورام الخبيثة وبالاعتماد على أعلى درجة تصنيف وأقل نسبة للخطأ، والمقارنة بين درجات قوة التصنيف بين الحالات المختلفة وكانت النتائج كما يأتي:

أ- إختبار (Box's-M Test) لتجانس مصفوفة التباين المشترك

إن أحد الشروط التي يجب أن تتوافر لكي نستطيع تطبيق التحليل التمييزي الخطي هو تساوي مصفوفات التباين والتباين المشترك للمجموعتين كما هو موضح في الجدول (7)، واختبار (Box's-M Test) الذي يعتمد أساساً على اختبار (F) وهو أحد الاختبارات التي طبقت للتأكد من شرط تجانس التباينات، إذ يختبر الفرضية الخاصة بهذا الاختبار وهي:
 H_0 : التباينات في المجموعة (1) متجانسة مع تباينات المجموعة (2).
 H_1 : التباينات في المجموعة (1) غير متجانسة مع تباينات المجموعة (2).

جدول (7) إختبار (Box's-M Test) لتجانس مصفوفة التباين المشترك

Log Determinants		
FACTOR	Rank	Log Determinant
غير المصابين	5	-19.015
المصابين	5	-18.657
Pooled within-groups	5	-18.553
Test Results		
Box's M		40.289
F	Approx.	2.576
	df1	15
	df2	64091.368
	Sig.	0.001

نلاحظ في الجدول (7) قيمة الاختبار (Box's-M) تساوي (40.289) وقيمة (F) المحسوبة تساوي (2.576) وبقيمة احتمالية (Sig.=.001) وهي أكبر من مستوى المعنوية ($\alpha=0.05$) وهذا يشير بدوره إلى عدم رفض فرضية العدم وهذا يعني تساوي تباين المجموعتين، وبهذا تم تحقيق شرط تساوي مصفوفات التباين والتباين المشترك للمجموعتين، وأيضاً نلاحظ قيمة (Log Determinants) للمجموعتين متقاربة جداً من بعضها أي يفترض تجانس مصفوفات التباين-التباين المشترك.



2- 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 (8):

جدول (8) إختبار (Wilks' Lambda) لأهمية الدالة التمييزية في التمييز بين المجموعتين

Wilks' Lambda				
Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	.878	16.908	7	0.005

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

3- 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 (9) وكما يلي:

جدول (9) يبين حساب الدالة التمييزية الخطية لجميع متغيرات PC1 + PC2

Var.	Function(Z)
كلوكوز	.114
يوريا	.518
مستوى الكرياتينين	.161
كولسترول	.345
الدهون الثلاثية	.646

$$Z = .114Z_1 + .518Z_2 + .161Z_3 + .345Z_4 + .646Z_5$$

نلاحظ في الجدول أعلاه أن المتغيرات، يوريا (X_2)، ونسبة الكولسترول في الدم (X_4) والدهون الثلاثية (X_5). لهم التأثير الأكبر في الدالة التمييزية من المتغيرات الأخرى.

وتم استخدام طريقة الأنحدار المتدرج (Stepwise ANOVA DA) وكانت النتائج كما هي موضحة في الجدول (10):

جدول (10) يبين حساب الدالة التمييزية الخطية بطريقة (Stepwise) لمتغيرات PC1+PC2

Var.	Function(Z)
الدهون الثلاثية	5.065
Constant	0.865

نلاحظ في الجدول (10) أن المتغير الدهون الثلاثية (X_5). بقيا في التحليل.

4- الدالة التمييزية بحسب متوسطات المجموعات (نقطة القطع)

ويمكن ملاحظة الدالة الخطية التمييزية مقيمة بحسب متوسطات المجموعات كما هو موضح في الجدول (11):

جدول (11) الدالة التمييزية بحسب متوسطات المجموعتين

Functions at Group Centroids	
Factor	Function(1)-Cutoff point
غير المصابين	.330
المصابين	-.413



إذ نلاحظ أن هناك متوسطين الأول (المرضى غير المصابين) والثاني (المرضى المصابين) قيمة المتوسطين غير متساويتين بسبب (عدم تساوي أفراد العينتين) ويمكن تفسير الإشارة الموجبة (+) بارتفاع درجات المتغيرات الداخلة في التحليل تؤدي إلى إرتفاع احتمالات الأنضمام إلى المجموعة غير المصابين) وكذلك الإشارة السالبة (-) تعني أن ارتفاع درجات المتغيرات الداخلة في التحليل يؤدي إلى إرتفاع احتمالات الأنضمام إلى المجموعة المصابين، أي ان نقطة القطع تساوي $CP=0.330$.

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

4-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.

1- نسبة الخطأ الظاهر

وهي تعويض جميع قيم المصنوفة X في المعادلة المستخرجة (W الجدول 12) يمثل اخطاء التصنيف بلاحظ بأنه يمثل مدى صحة تصنيف الكتلة إلى (غير المصلبين) أم (المصابين)، علماً أننا استخدمنا عملية التصنيف على متغيرات المكون الرئيس الأول جميعها (الأكثر تحميلاً)، ونرى أن النسبة المئوية لصحة التصنيف هي (63.6%)، ويمكن إيجاد نسبة الخطأ الحقيقي للمجموعتين كما يلي

$$P_{12} = 38/73 = 0.373$$

أي ان

$$P_{12} : \text{احتمال تنسيب الكتلة الى المجموعة الاولى في حين انهم يعودون للمجموعة الثانية} = 22/60 = 0.367$$

أي ان P_{21} : احتمال تنسيب الكتلة الى المجموعة الثانية في حين انهم يعودون للمجموعة الاولى

$$\text{ام نسبة الخطا الكلي هي : } TER = \hat{P} = \frac{28 + 22}{135} = 0.370$$

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



Classification Results ^(b,c)					
		Factor	Predicted Group Membership		Total
			غير المصابين	المصابين	
Original	Count	غير المصابين	47	28	75
		المصابين	22	38	60
	%	غير المصابين	2	0	2
		المصابين	62.7	37.3	100.0
Cross-validated ^(a)	Count	غير المصابين	36.7	63.3	100.0
		المصابين	100.0	.0	100.0
	%	غير المصابين	45	30	75
		المصابين	23	37	60
63.0% of original grouped cases correctly classified.					
60.7% of cross-validated grouped cases correctly classified.					

جدول (13) تصنيف الكتل باستخدام متغيرات المكون الرئيسي (PC1+PC2 +PC3) المعنوية

Classification Results ^{a,c}					
		Factor	Predicted Group Membership		Total
			غير المصابين	المصابين	
Original	Count	غير المصابين	47	28	75
		المصابين	20	40	60
	%	غير المصابين	62.7	37.3	100.0
		المصابين	33.3	66.7	100.0
Cross-validated ^b	Count	غير المصابين	42	33	75
		المصابين	22	38	60
	%	غير المصابين	56.0	44.0	100.0
		المصابين	36.7	63.3	100.0
a. 64.4% of original grouped cases correctly classified.					
c. 59.3% of cross-validated grouped cases correctly classified.					



وهكذا تم تطبيق التحليل التمييزي على المتغيرات المعنوية (الأكثر تحملاً) للمكونات الرئيسية المستخلصة والبالغة (3) مكونات رئيسية وذلك بشكل تجميعي (Cumulative)، أي استخدام المتغيرات الأكثر تحملاً في و (PC2+PC1) و (PC3+PC2+PC1) ثم والمقارنة بين المتغيرات الداخلة في الدالة و نسبة درجة التصنيف ونسبة الخطأ كما هو موضح في جدول المقارنات.

رابعا - جدول المقارنات بين النماذج

من خلال ما تقدم يمكن تلخيص النتائج التي تم التوصل اليها والتي تضم المتغيرات المختارة في الدالة و نسبة درجة التصنيف ونسبة الخطأ الكلي (TER) كما هو موضح في الجدول (15) إذ نلاحظ بأن المتغيرات المعنوية داخل المكونات الرئيسية المستخلصة لها تأثير في زيادة دقة التصنيف، فنلاحظ بأنه باستخدام المتغيرات التابعة للمكون الأول بطريقة المتغيرات جميعها داخلة في التحليل (Enter) فإن نسبة درجة التصنيف بلغت (63.0%) ونسبة خطأ كلي (0.370). وبطريقة (Stepwies) فان درجة دقة التصنيف بلغت (60.7%) ونسبة خطأ الكلي (0.277) وهكذا لبقيّة المكونات الأخرى، ويمكن الملاحظة بان أفضل أنموذج توصل اليه البحث هو النموذج الأخير، إذ بلغت درجة دقة التصنيف (64.4%) ونسبة خطأ كلي (355)، ويمكن ملاحظة أن المتغيرات المعنوية في المكون الرئيس الثاني لم يكن له تأثير في درجة دقة التصنيف ونسبة الخطأ، ويمكن ملاحظتها في جدول مقارنة النتائج.

جدول (14) مقارنة النتائج

التحليل التمييزي (للتصنيف بين الكتلة الخبيثة والحميدة)				المكونات
Stepwise ANOVA DA		Enter all Sig. Variables ANOVA DA		
Var.	Function(Z)	Var.	Function 1	PC1+PC2
		كلوكوز	.114	
		يوريا	.518	
		مستوى الكرياتينين	.161	
		كولسترول	.345	
		الدهون الثلاثية	.646	
(60.7%) (0.277)		(63.0%) (0.370)		دقة التصنيف ونسبة الخطأ الكلي (TER)
Var.	Function(Z)	Var.	Function(Z)	PC1+PC2+PC3
		كلوكوز	.122	
		يوريا	.561	
		مستوى الكرياتينين	.126	
		كولسترول	.377	
		الدهون الثلاثية	.609	
		الدهون العالية	-.279-	
		الدهون الواطئة	.148	
		Constant	0.865	
		الدهون الثلاثية	5.065	



(59.3%) (0.407)	(64.4%) (0.355)	دقة التصنيف ونسبة الخطأ الکلي (TER)
--------------------	---------------------	---

4 / الأستنتاجات والتوصيات

اولا- الاستنتاجات

- 1- عند تطبيق (PCA) لأستخلاص أهم المتغيرات والقياسات الخاصة إستخلاص (3) مكونات رئيسة تفسر مانسبته (62.352) من التباين الكلي وهي : إن العامل الاول فسر نسبة قدرها 24.121% من التباين الكلي، وقد ضم هذا العامل متغيرين ذات دلالة إحصائية وحسب التسلسل وهي مستوى الكرياتينين (X_3)، ونسبة الكلوسترول في الدم (X_4) وفسر العامل الثاني نسبة قدرها 20.555% من التباين الكلي، وضم هذا العامل ثلاثة متغيرات ذات دلالة إحصائية وحسب التسلسل وهي: كلوكوز (نبيبة السكر في الدم) (X_1) ، يوريا (X_2)، والدهون الثلاثية (X_5). أما العامل الثالث فقد بلغت نسبة تفسيره 17.677% من التباين الكلي، وضم هذا العامل متغيرين ذات دلالة إحصائية وحسب التسلسل وهي الدهون العالية X_6 ، والدهون الواطئة X_7 .
- 2- من الأستنتاجات التي يمكن بيانها هي أن المتغير والدهون الثلاثية (X_3). بقيا في التحليل ظهرا في المعادلات التصنيفية في النماذج كافة وهذا يعني أنه من المتغيرات المؤثرة في معرفة المرضى المصابين وغير المصابين بالاورام وتمييزها.
- 3- وأفضل النماذج المقترحة للتصنيف في البحث تصل دقة تصنيفها تساوي (64,4%) و (63,0%) على التوالي.
- 4- إذ نلاحظ بأن المتغيرات المعنوية داخل المكونات الرئيسية المستخلصة لها تأثير في زيادة دقة التصنيف، فنلاحظ بأنه باستخدام المتغيرات التابعة للمكون الأول بطريقة المتغيرات جميعها داخله في التحليل (Enter) فإن نسبة درجة التصنيف بلغت (63.0%) ونسبة خطأ كلي (0.370) وبطريقة (Stepwies) فان درجة دقة التصنيف بلغت (60.7%) ونسبة خطأ الكلي (0.277)

ثانيا - التوصيات

نوصي بإعتماد الأطباء المختصين على بعض التحليلات والقياسات الاحصائية للأورام ومتابعة التغييرات الحاصلة خلال فترات العلاج لزيادة معلوماتهم في إتخاذ قراراتهم إستخدام طرق اخرى لتحليل متعدد المتغيرات في معالجة المرضى المصابين وغير المصابين بالاورام مثل (Cluster Analysis, Principle component Analysis) والمقارنة بينهما

5 - المصادر

- 1- اثناسيوس، زكريا زكي، والبياتي، عبد الجبار توفيق (1977): " المدخل الى التحليل العاملي". مطبعة مؤسسة الثقافة العمالية، بغداد، العراق.
- 2- 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.
- 3- حمد، بخشان احمد (2003): " استخدام التحليل العاملي لتحديد مؤشرات الرضا الوظيفي لدى المعلمين في مدينة اربيل". رسالة ماجستير، كلية الادارة والاقتصاد، جامعة صلاح الدين . اربيل، العراق.
- 4- داود، الا عبد الستار (2005): "استخدام الدالة التمييزية وطرق تحديد متغيراتها". رسالة ماجستير، كلية علوم الحاسبات والرياضيات، جامعة الموصل، العراق.
- 5- Afifi, A. and Clark, V (1984): "Computer Aided Multivariate Analysis". Life time Learning Publications, USA.
- 6- Anderson, T. W. (1984): "An Introduction to Multivariate Statistical Analysis". 2nd - Edition, John Wiley and Sons, New York-USA.
- 7- Jolliffe, (1986): "Principle Component Analysis", Spring-Verlag New York, Inc.



پوخته

شیتوازی شیکردنهوی ئاماری فره گۆراو پشت ده به سیتت به شیتوازه جیاوازه کان بۆ وه سف کردن و شیکردنهوی دیارده کان که فره ره هه ند و فره گۆراو وی هه یه، ته گهر بینراوه کان به شدار بن له ئیوانیان به کۆمه لیک له تایه تمه ندیه کان و سیفه ته کان به پله جیاوازه کان، تهوا شیکردنهوی ئاماری فره گۆراو باس له لیکۆلینهوی داناکانی تهو بینراوانه و گوزارشت لیان ده کات له میانهی زۆرتین تهو گۆراوانه ی کاریگه ریان له سه ر دیارده که هه یه که له ژیر لیکۆلینهوی هه یه، لیکۆلینهوی که پشت ده به سیتت به یه ک به ستن له ئیوان شیکردنهوی فره گۆراوه کان به به کاره یئانی شیکردنهوی یتکهاته سه ره کیه کان (PCA) وه شیکردنهوی جیاکه ره وه یی (DA)، له بهر ته مه ئامانجی ته م لیکۆلینهوی ده ستینشانکردنی گرنگترین فاکته ره کاریگه ره کان له سه ر نه خۆشه تووشبوه کان وتووشنه بوه کان به شپیره نه چه پیسه کان له میانهی ده ستینشانکردنی ژماره ی فاکته ره کورتکراوه کان به به کاره یئانی شیکردنهوی فاکته ری وه هه ره وه به به کاره یئانی شیکردنهوی جیاکه ره وه له کرداری جیاکه ردنه وه و پۆلینکردن، پاشان مه زه نده کردنی باشته رین هاوکتیشه ی پۆلینکراو به پشت به ستن به پتوه ره تایه ته کان وه کورتکراوه به هۆی یتکهاته سه ره کیه کان بۆ جیاکه ردنه وه له ئیوان دوو چۆر به راوردکردن له ئیوان ته نجامه کان له سه ر بنچینه ی ریژه ی هه له ی پۆلینکراو وه بلاوکردنهوی کاریگه ری له سه ر کرداری پۆلین، وه بۆ جیه جیکردنی ئامانجی توژیینه وه که دابهش کرا بۆ لایه نی تیوری که باسی شیتوازی شیکردنهوی فاکته ری و شیکردنهوی جیاکه ره وه یی ده کات، به لام لایه نی پراکتیکی بریتیه له داتای نه خۆشه تووشبوه کان وتووشنه بوه کان به (شپیره نه چه پیسه کان) که ده ست کهوت له نه خۆشخانه ی رزگاری ونه خۆشخانه ی نانه که لی له هه ولێر وه ته نجامه کانمان ده ست کهوت که به پیزانیی ئیمه گرنگه و ده بیته تیبنی بکریته له لایه ن لایه نه ته ندروستیه کان.

Abstract

Statistical analysis of the style of multi -dependent variables of its different ways to describe and analyze the phenomena of multi - dimensional variables. If they share views with each set of characteristics and traits with varying degrees, 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 (PCA) and discriminate analysis of (DA), so this research aims to determine the most important factors affecting patients infected and not infected with malignant tumours by 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 types 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 and non-tumours infected, obtained from (Rizgarry Hospital) and (Nana Kali 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								
		كلوكوز	يوریا	مستوى الكرياتينين	كولوسترول	الدهون الثلاثية	الدهون العالية	الدهون الواطئة
Correlation	كلوكوز	1.000	.327	-.171-	-.242-	.101	.166	-.117-
	يوریا	.327	1.000	-.147-	-.265-	.253	.189	-.052-
	مستوى الكرياتينين	-.171-	-.147-	1.000	.752	-.031-	.062	.087
	كولوسترول	-.242-	-.265-	.752	1.000	-.061-	.071	.087
	الدهون الثلاثية	.101	.253	-.031-	-.061-	1.000	.620	.088
	الدهون العالية	.166	.189	.062	.071	.620	1.000	.106
	الدهون الواطئة	-.117-	-.052-	.087	.087	.088	.106	1.000
Sig. (1-tailed)	كلوكوز		.005	.095	.031	.221	.103	.187
	يوریا	.005		.131	.020	.026	.074	.347
	مستوى الكرياتينين	.095	.131		.000	.406	.320	.254
	كولوسترول	.031	.020	.000		.322	.296	.254
	الدهون الثلاثية	.221	.026	.406	.322		.000	.252
	الدهون العالية	.103	.074	.320	.296	.000		.210
	الدهون الواطئة	.187	.347	.254	.254	.252	.210	