

مقارنه بين التحليل التمييزي الخطي واحتمال الاستجابة في تصنيف البيانات*

باحثه ازهار كاظم جبارة

أ.م.د حمزة اسماعيل شاهين

الجامعة المستنصرية // كلية الادارة والاقتصاد // قسم الاحصاء

المستخلص

من اوسع الاساليب الاحصائية استخداماً في مجال تحليل البيانات هو اسلوب التحليل التمييزي الخطي. ان استخدام التحليل التمييزي الخطي يشترط توفر عدد من الافتراضات من اهمها ان تكون بيانات المتغيرات التوضيحية ذات توزيع طبيعي متعدد المتغيرات.

يهدف هذا البحث بالدرجة الاساس الى استخدام دالة التمييز الخطي كدالة تصنيف خطية وتصنيف البيانات بصيغتين ،الصيغه الاولى تتمثل بدالة التمييز الخطي ،والصيغه الثانيه تمثل دالة احتمال الاستجابة التي تم اشتقاقها لدالة التمييز الخطي .وتم اجراء المقارنات بين هذه الصيغتين وفق معيار احتمال خطأ التصنيف (Misclassification) .

اثبتت النتائج التصنيف حسب الصيغه الاحتماليه لدالة التمييز الخطي تعطي اقل احتمال خطأ التصنيف من خلال التطبيق على بيانات تخص ثلاثه انواع من امراض العيون.

Abstract

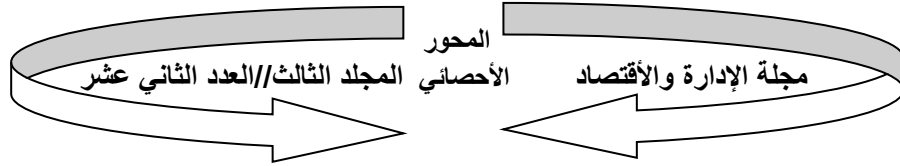
Linear discriminant analysis is the most widely used in statistical analysis linear discriminant analysis makes assumption on the distribution of explanatory data.It has been developed for normally distributed explanatory variables.

The objective of this paper is using the linear discriminant function as linear classification models and classified the data with respect two forms ,The first is the linear discriminant analysis and the second the form of the response probability Which it derivative for the linear disccriminant function .

The Comparison between these methods is based on the measure of probability is misclassification.

We show that the results of probability response has the minimum probability misclassification through the application the data of three types of eyes disease.

* البحث مستل من رسالة الماجستير بعنوان مقارنة بين طرائق التحليل التمييزي والتحليل اللوجستي متعدد الاستجابة ولم تناقش



١- المقدمة وهدف البحث

ان تصنيف مفردة او مشاهد جديدة لاحدى المجموعات قيد الدراسة يمكن ان تتم وفق اسلوب التحليل التمييزي (Discriminant Analysis) (٢٦) ، ان التحليل التمييزي يعتبر احد الاساليب المهمة في التحليل الاحصائي متعدد المتغيرات ، حيث يتم بموجبه استعمال مجموعة من المتغيرات للفصل (التمييز) بين مجموعتين او اكثر عن طريق دوال تمييزية (Discriminant functions) قد تكون خطية أو تربيعية وهذه عبارة عن توليفه خطية للمتغيرات التوضيحية اذ يتم الاعتماد على هذه الدالة في التنبؤ ، وتحديد المتغيرات التوضيحية (المستقلة) التي تساهم بشكل مؤثر في التمييز ما بين مجموعتين فأكثر وفي تصنيف مفردة جديدة لاحدى المجموعات قيد الدراسة بأقل خطأ تصنيف ممكن (١٦) .

أن استخدام التحليل التمييزي الخطي يشترط توفر عدد من الافتراضات منها ان تكون بيانات المتغيرات التوضيحية ذات توزيع طبيعي متعدد المتغيرات وان تكون التباينات (مصفوفة التباين والتباين المشترك) متساوية لكل المجموع واخيرا ان تكون موجهاً المتوسطات مختلفة في كل مجموعة من المجموع (٢٧) . ويهدف البحث بالدرجة الاساس الى تصنيف البيانات وفق طريقتين ، طريقة دالة التمييز الخطي وطريقة احتمالات الاستجابة لدالة التمييز الخطي ، ويهدف البحث ايضا الوصول الى افضل نموذج خطي وذلك لتشخيص بعض انواع امراض العيون والحصول على دالة تميز بين ثلاث انواع من امراض العيون ، الساد (الماء الابيض)، ونزف السائل الزجاجي، وداء الزرقاء (الماء الازرق) .

وتتم المقارنه بين الطريقتين على اساس معيار احتمال خطأ التصنيف (Misclassification)

الجانب النظري

(٢) التحليل التمييزي: Discriminant Analysis (8)، (22)، (23)

يعتبر التحليل التمييزي (Discriminant Analysis) من أساليب تحليل متعدد المتغيرات التي تهتم بفصل مجموعات مختلفة من المفردات (أو المشاهدات) ويتوزع المفردات (أو المشاهدات) الجديدة على مجموعات سبق تعريفها. أما عملية التصنيف فهي العملية اللاحقة لعملية تكوين الدالة المميزة حيث يتم الاعتماد على هذه الدالة لتصنيف المفردة الجديدة لإحدى المجموعات قيد الدرس بأقل خطأ تصنيف ممكن.

وفيما يأتي بعض انواع نماذج التحليل التمييزي وسوف نركز على التمييز الخطي في هذه الدراسة

(1-2) الدالة التمييزية الخطية لمجموعتين: (8)، (9)، (27)

The Linear Discriminant Function – Two Groups

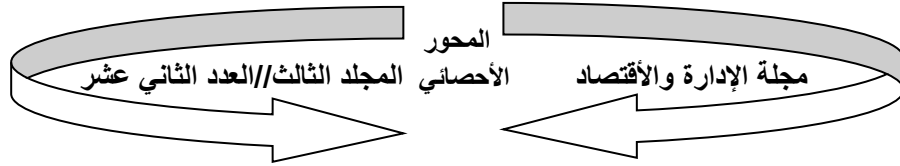
وهي الدالة المعروفة لفشر (Fisher Function) حيث لايفترض فيها توزيع المشاهدات توزيعاً طبيعياً

وتعتبر هذه الدالة من الدوال اللامعلمية. لو فرضنا انه لدينا مجتمعين يراد المقارنه بينهما تحت حالة افتراض ان المجتمعين لهما نفس مصفوفة التباين والتباين المشترك ($\Sigma_0 = \Sigma_1$) وبموجه متوسطات (μ_0, μ_1) على التوالي وتم

سحب عينتين عشوائيتين ($x_{01}, x_{02}, x_{03}, \dots, x_{0n_0}$) و ($x_{11}, x_{12}, x_{13}, \dots, x_{1n_1}$)

لكل من المجتمعين، وأن كل موجه (x_{ij}) يضم (p) من المتغيرات (الصفات) .

وبالتالي فإن دالة التمييز هي عبارة عن تركيب خطي يضم (p) من المتغيرات التي تعمل على تعظيم الفرق بين متوسطي المجموعتين أن هذه الدالة تكون كالآتي :



$$z_{0i} = \underline{\hat{a}} \underline{x}_{0i} = a_1 x_{0i_1} + a_2 x_{0i_2} + \dots + a_p x_{0i_p} \quad (1)$$

حيث ان $i = 1, 2, \dots, n_0$

$$z_{1i} = \underline{\hat{a}} \underline{x}_{1i} = a_1 x_{1i_1} + a_2 x_{1i_2} + \dots + a_p x_{1i_p} \quad (2)$$

حيث ان $i = 1, 2, \dots, n_1$

وبالتالي يجب إيجاد موجه (\underline{a}) الذي يعظم الفرق بين متوسطي المجموعتين فإذا رمزنا لنسبة الاختلافات بين المجموعتين الى الاختلافات داخل المجموعتين بالرمز (Q) إذ أن:

$$Q = \frac{\text{Between Groups}}{\text{Within Groups}} = \frac{[\bar{z}_0 - \bar{z}_1]^2}{S_z^2}$$

ومن خلال الطرق الجبرية نتوصل الى ان

الدالة التمييزية الخطية لفشر تأخذ النحو الآتي :

$$Z = \underline{a}' \underline{x} = (\bar{x}_0 - \bar{x}_1)' S_p^{-1} \underline{x} \quad (3)$$

وتجدد الإشارة الى أن مسافة (مقياس) مها لونيوس تعتمد اساساً في قياسات اقل مسافة ممكنة بين قيم المتغيرات للملاحظات الجديدة وقيم متوسطات المتغيرات لكل مجموعة حيث أن :

$$D_i^2 = (\underline{x} - \bar{x}_i)' S_p^{-1} (\underline{x} - \bar{x}_i) \quad (4)$$

حيث أن :

\bar{x}_i : متجه متوسطات المتغيرات لقيمة المجموعة (i)

S_p^{-1} : هو معكوس مصفوفة التباين والتباين المشترك المقدرة داخل العينتين معاً

حيث ان

$$S_p = \frac{(n_0 - 1) S_0 + (n_1 - 1) S_1}{n_0 + n_1 - 2} \quad (5)$$

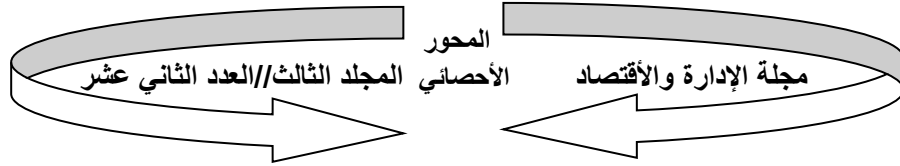
(3-2) دالة التمييزية الخطية في حالة دالة الكثافة الاحتمالية معلومه (16)·(28)

لنفترض لدينا مجتمعين (G_1, G_0) ولكل منهما داله كثافه احتماليه $f(\underline{x}|G_1)$ ، $f(\underline{x}|G_0)$ على الترتيب ولنفرض ان (π_1, π_0) هي احتمالات اوليه (ors probability) ان المشاهده \underline{x} تاتي من (G_1, G_0) على الترتيب وعلى افتراض ان دوال الكثافه تتبع توزيعاً طبيعياً متعدد المتغيرات تحت حاله تساوي مصفوفتا التباين والتباين المشترك ($\Sigma_0 = \Sigma_1$) إذ أن :-

$$\begin{aligned} f(\underline{x}|G_0) &\sim Np(\underline{\mu}_0, \Sigma) \\ f(\underline{x}|G_1) &\sim Np(\underline{\mu}_1, \Sigma) \end{aligned} \quad (6)$$

النسبه لدالتي الكثافه هي

$$h(\underline{x}) = \frac{\pi_0 f(\underline{x}|G_0)}{\pi_1 f(\underline{x}|G_1)}$$



$$h(\underline{x}) = \frac{\frac{\pi_0}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp \left[-\frac{1}{2} (\underline{x} - \underline{\mu}_0)' \Sigma^{-1} (\underline{x} - \underline{\mu}_0) \right]}{\frac{\pi_1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp \left[-\frac{1}{2} (\underline{x} - \underline{\mu}_1)' \Sigma^{-1} (\underline{x} - \underline{\mu}_1) \right]} \quad (7)$$

وبالطرق الجبرية نتوصل الى ان

$$h(\underline{x}) = \exp \left[(\underline{\mu}_0 - \underline{\mu}_1)' \Sigma^{-1} \underline{x} - \frac{1}{2} (\underline{\mu}_0 + \underline{\mu}_1)' \Sigma^{-1} (\underline{\mu}_0 - \underline{\mu}_1) \right] * \frac{\pi_0}{\pi_1} \quad (8)$$

وحيث ان لوغاريتم الدالة اعلاه يعطي نفس نتيجة الدالة نفسها فللبساطة تكتب المعادلة (8) كالآتي:-

$$\ln h(\underline{x}) = \left[(\underline{\mu}_0 - \underline{\mu}_1)' \Sigma^{-1} \underline{x} - \frac{1}{2} (\underline{\mu}_0 + \underline{\mu}_1)' \Sigma^{-1} (\underline{\mu}_0 - \underline{\mu}_1) + \ln \frac{\pi_0}{\pi_1} \right] \quad (9)$$

يمثل الحد الأول من المعادلة (9) داله التمييز الخطي لفشر التي مر ذكرها سابقاً، إما الحد الثاني فيمثل نقطة الفصل بين المجموعتين G_1, G_0 .

ان المعالم المجهوله $\Sigma, \underline{\mu}_1, \underline{\mu}_0$ في المعادلة (9) يتم تقديرها من بيانات العينة بطريقة الامكان الاعظم (Maximum Likelihood) حيث افضل تقدير لـ $\underline{\mu}_0$ هو \bar{x}_0 ولـ $\underline{\mu}_1$ هو \bar{x}_1 اذ:-

$$\bar{x}_0 = \frac{\sum_{i=1}^{n_0} x_{0i}}{n_0} \quad (10)$$

$$\bar{x}_1 = \frac{\sum_{i=1}^{n_1} x_{1i}}{n_1} \quad (11)$$

اما مصفوفه التباين والتباين المشترك المدمجه Σ فيكون تقديرها بمصفوفه التباين والتباين المشترك المدمجه للعينه S_p وحسب الصيغه (4).

اما π_1, π_0 فيتم تقديرها باسلوبين اما بافتراضهما متساويين اي ($\pi_0 = \pi_1$) للمجتمعين او انها يقدران بحسب الآتي:-

$$\begin{aligned} \hat{\pi}_0 &= \frac{n_0}{n} \\ \hat{\pi}_1 &= \frac{n_1}{n} \end{aligned} \quad (12)$$

اذ ان

n_0 : حجم العينه المسحويه من المجتمع G_0

n_1 : حجم العينه المسحويه من المجتمع G_1

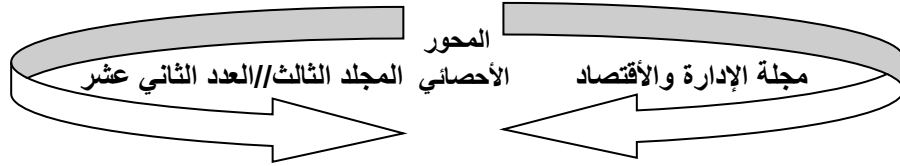
وان $n = n_0 + n_1$

وبالتعويض عن المعالم المجهوله في المعادلة (9) نحصل على الاحصاء الآتية:-

$$W = (\bar{x}_0 - \bar{x}_1)' S_p^{-1} \underline{x} - \frac{1}{2} (\bar{x}_0 + \bar{x}_1)' S_p^{-1} (\bar{x}_0 - \bar{x}_1) + \ln \frac{\hat{\pi}_0}{\hat{\pi}_1} \quad (13)$$

وفي حالة ($\pi_0 = \pi_1$) فان المعادلة (9) تصبح

$$W^* = (\bar{x}_0 - \bar{x}_1)' S_p^{-1} \underline{x} - \frac{1}{2} (\bar{x}_0 + \bar{x}_1)' S_p^{-1} (\bar{x}_0 - \bar{x}_1) \quad (14)$$



(2-4) دالة التمييز الخطية في حالة اكثر من مجموعتين: (8)، (22)، (27)

The Linear Discriminant Function More than two group

لنفرض لدينا (k) من المجتمعات بمتجهات متوسطات μ_1, \dots, μ_k ومصفوفه تباين مشترك لكل المجتمعات Σ .

ولنفترض ان T تمثل مصفوفة التباين والتباين المشترك التقديرية لـ Σ .
اذ أن:

$$T = \frac{1}{n-k} \sum_{i=1}^k \sum_{j=1}^n (x_{ij} - \bar{x})(x_{ij} - \bar{x})' \quad (15)$$

و w_i تمثل مصفوفة التباين والتباين المشترك للمجموعة i

$$w_i = \frac{1}{n_i-1} \sum_{j=1}^n (x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)' \quad (16)$$

وان مصفوفة التباين والتباين المشترك بين المجاميع هي:-

$$B = T - W$$

الهدف هو إيجاد مجموعة من الدوال (التراكيب) الخطية

$$Y = \{y_1, y_2, \dots, y_r\}$$

ويطلق على الدوال الخطية (y_1, y_2, \dots, y_r) الدوال المميزة الخطية ويمكن كتابتها بشكل مصفوفة:

$$\underline{y} = b \underline{x}$$

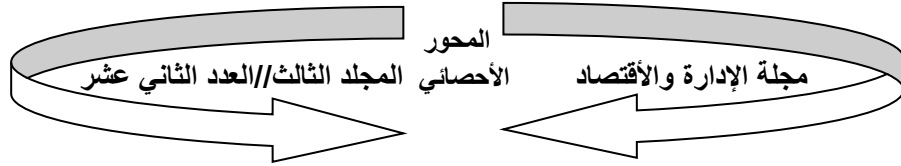
$$\begin{pmatrix} y_1 \\ \vdots \\ y_p \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} & \dots & \dots & b_{1p} \\ \vdots & \vdots & \dots & \dots & \vdots \\ b_{p1} & b_{p2} & \dots & \dots & b_{pp} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix} \quad (17)$$

وتحديد (r) اي عدد الدوال المميزة بالاعتماد على رتبة المصفوفة المركبة $(B w^{-1})$ ، أن رتبة المصفوفة $(w_{p \times p})$ تساوي (p) (وان رتبة $w^{-1} =$ رتبة w) رتبة المصفوفة (B) تكون اصغر لـ (k-1, p) وعادة تكون (k-1) اصغر من (p) لانه نادراً ما يستخدم عدد المتغيرات اقل من عدد المجاميع لهذا تكون رتبة المصفوفة $(B w^{-1})$ كالاتي :

$$\text{rank}(w^{-1} B) = \text{Min}(k-1, p) \quad (18)$$

اي ان تكون عدد الدوال المميزة لـ (k) من المجاميع و (p) من المتغيرات هو

$$r = \text{Min}(k-1, p)$$



(5-2) التصنيف وفق الدالة التمييزية الخطية لفشر (8)، (16)، (28)

Classification according to discriminant function for linear fisher

ان تصنيف المشاهدات التي تعود لمجتمعات غير معروفه وفق دالة التمييز الخطية لفشر يتم وفق الاتي:-

1- احتساب قيم متوسطات دالة التمييز الخطية للعينتين وفق الصيغه (3) أن

$$\bar{Z}_i = (\bar{x}_0 - \bar{x}_1)' S_P^{-1} \underline{x}, \quad i = 0, 1$$

2- نحسب متوسط متوسطات دالة التمييز الخطية بالشكل

$$\frac{\bar{Z}_1 + \bar{Z}_2}{2} = \frac{1}{2} (\bar{x}_0 - \bar{x}_1)' S_P^{-1} (\bar{x}_0 + \bar{x}_1)$$

3- تكون قاعده التصنيف بان تصنف المشاهده \underline{x} الى المجموعه G_0 اذا كان :

$$Z = (\bar{x}_0 - \bar{x}_1)' S_P^{-1} \underline{x} > \frac{1}{2} (\bar{x}_0 - \bar{x}_1)' S_P^{-1} (\bar{x}_0 + \bar{x}_1)$$

والى المجموعه G_1 اذا كان

$$Z = (\bar{x}_0 - \bar{x}_1)' S_P^{-1} \underline{x} \leq \frac{1}{2} (\bar{x}_0 + \bar{x}_1)' S_P^{-1} (\bar{x}_0 - \bar{x}_1)$$

(2-7) التصنيف وفق صيغة دالة الكثافة الاحتماليه (16)

قاعده التصنيف وفق صيغه دالة الكثافه الاحتماليه تستند على تطبيق نظرية بيز (Bayes Theorem) والتي تعطي افضل تصنيف بأقل احتمال لخطأ التصنيف حيث تصنف المشاهده \underline{x} الى المجموعه G_0 وفق الصيغه (14) اذا كان :-

$$(\bar{x}_0 - \bar{x}_1)' S_P^{-1} \underline{x} > \frac{1}{2} (\bar{x}_0 + \bar{x}_1)' S_P^{-1} (\bar{x}_0 - \bar{x}_1) + Ln \frac{\hat{\pi}_1}{\hat{\pi}_0}$$

وفي حاله $\pi_0 = \pi_1$ فان تصنيف المشاهدات يماثل صيغه فشر المعطاة في الفقره (5-2).

وفي حاله لدينا k من المجتمعات يمكن استخدام مسافه (مقياس) مهالونوبيس لتصنيف المشاهده \underline{x} الى المجموعه i حسب الصيغه (4) اذا كان

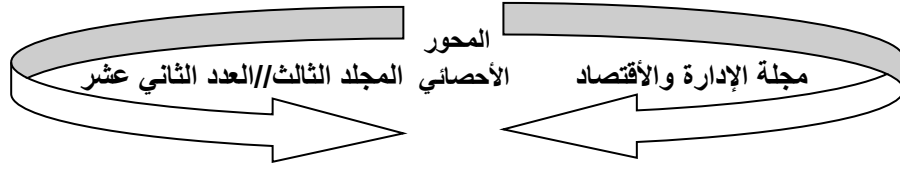
$$D_i^2 < D_j^2, \quad i \neq j, \quad i = 1, 2, \dots, k$$

وبالامكان ايجاد دوال تمييزيه خطيه بقدر عدد المجاميع من خلال التوسع في استخدام الصيغه (4) (16) اذ

$$\begin{aligned} D_i(\underline{x}) &= \underline{x}' S_P^{-1} \underline{x} - \underline{x}' S_P^{-1} \bar{x}_i - \bar{x}_i' S_P^{-1} \underline{x} + \bar{x}_i' S_P^{-1} \bar{x}_i \\ &= \underline{x}' S_P^{-1} \underline{x} - 2\bar{x}_i' S_P^{-1} \underline{x} + \bar{x}_i' S_P^{-1} \bar{x}_i \end{aligned} \quad (19)$$

الحد الاول من الطرف الايمن للصيغه (19) يكون ثابت لكل المجاميع ، والحد الثاني عباره عن دالة خطيه في الموجه \underline{x} ، اما الحد الاخير فانه لا يعتمد على الموجه \underline{x} وبالتالي يمكن اهمال الحد الاول من الصيغه (19) والحصول على دالة تصنيف خطيه ويعبر عنها $d_i(\underline{x})$ ولكي تكون هذه الصيغه متوافقه مع داله التمييز الخطي في الفقره (3-2) يتم ضرب هذه الداله بالمقدار $(-\frac{1}{2})$ فتكون داله التمييز الخطي كالاتي :-

$$d_i(\underline{x}) = \underline{x}' S_P^{-1} \underline{x} - \frac{1}{2} \bar{x}_i' S_P^{-1} \bar{x}_i \dots \dots \dots (20)$$



وتكون قاعده التصنيف بان تصنف المشاهده \underline{x} الى المجموعه i اذا كانت قيمه الدالة $d_i(\underline{x})$ اكبر ما يمكن من بقية المجاميع الاخرى .

وفي حاله افتراض دوال الكثافه الاحتماليه لكل مجموعه تتبع التوزيع الطبيعي متعدد المتغيرات تحت حاله تساوي مصفوفات التباين والتباين المشترك وباحتمالات اوليه $\pi_1, \pi_2, \dots, \pi_k$, اذ

$$f(\underline{x}|G_i) = N_p(\underline{\mu}_i, \Sigma) \quad (21)$$

$$i = 1, \dots, k$$

فان داله التمييز الخطيه في الصيغه (20) تصيح

$$d_i^*(\underline{x}) = \ln \hat{\pi}_i + \bar{\underline{x}}_i' S_P^{-1} \underline{x} - \frac{1}{2} \bar{\underline{x}}_i' S_P^{-1} \bar{\underline{x}}_i \quad (22)$$

وتصنف المشاهده \underline{x} الى المجموعه i اذا كانت قيمه الدالة $d_i^*(\underline{x})$ اكبر ما يمكن من بقية المجاميع الاخرى علماً انه في حاله $\pi_1 = \pi_2 = \dots = \pi_k$ فان الصيغه (22) تختزل الى الصيغه (20) -7

(2) التصنيف وفق الصيغه الاحتماليه ⁽¹⁶⁾ (27) (49)

-1 : حالة مجموعتين

لنفرض لدينا مجموعتين G_1, G_0 ولكل منها دالة كثافة احتمالية (p.d.f) هي $f(\underline{x}|G_1)$ ، $f(\underline{x}|G_0)$ ولنفترض ان (π_1, π_0) هي احتمالات اوليه ان متجه المشاهده \underline{x} تأتي من المجتمع (G_1, G_0) على الترتيب فإن التوزيع اللاحق ان \underline{x} تأتي من المجموعه G_1 يكون .

$$P_r(G_1|\underline{x}) = \frac{\pi_1 f(\underline{x}|G_1)}{\pi_0 f(\underline{x}|G_0) + \pi_1 f(\underline{x}|G_1)} \quad (23)$$

والذي يمكن كتابته كالآتي:

$$= \frac{1}{1 + \frac{\pi_0}{\pi_1} \frac{f(\underline{x}|G_0)}{f(\underline{x}|G_1)}}$$

بافتراض إن دوال الكثافة الاحتمالية $f(\underline{x}|G_1)$ ، $f(\underline{x}|G_0)$ دوال تتبع التوزيع الطبيعي متعدد المتغيرات تحت حالة افتراض تساوي مصفوفتا التباين والتباين المشترك $(\Sigma_1 = \Sigma_0)$ وبالتحديد

$$f(\underline{x}|G_0) \sim Np(\underline{\mu}_0, \Sigma)$$

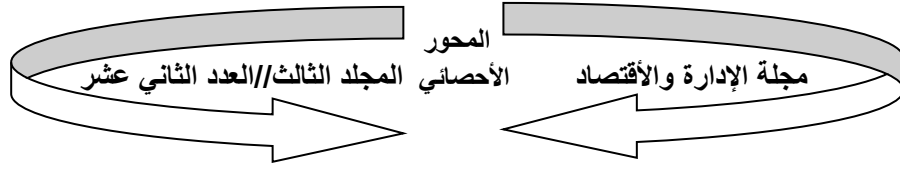
$$f(\underline{x}|G_1) \sim Np(\underline{\mu}_1, \Sigma)$$

(24)

عند نذ يمكن كتابة المعادله (23) حلاي :

$$P_r(G_1|\underline{x}) = \frac{1}{1 + \frac{\pi_0}{\pi_1} * \frac{\exp\left(-\frac{1}{2} (\underline{x} - \underline{\mu}_0)' \Sigma^{-1} (\underline{x} - \underline{\mu}_0)\right)}{\exp\left(-\frac{1}{2} (\underline{x} - \underline{\mu}_1)' \Sigma^{-1} (\underline{x} - \underline{\mu}_1)\right)}}$$

$$= \frac{1}{1 + \frac{\pi_0}{\pi_1} * \exp\left[-\frac{1}{2} (\underline{x} - \underline{\mu}_0)' \Sigma^{-1} (\underline{x} - \underline{\mu}_0) + \frac{1}{2} (\underline{x} - \underline{\mu}_1)' \Sigma^{-1} (\underline{x} - \underline{\mu}_1)\right]}$$



$$\begin{aligned}
 &= \frac{1}{1 + \frac{\pi_0}{\pi_1} * \exp \left[\frac{1}{2} (\underline{x} - \underline{\mu}_1)' \Sigma^{-1} (\underline{x} - \underline{\mu}_1) - (\underline{x} - \underline{\mu}_0)' \Sigma^{-1} (\underline{x} - \underline{\mu}_0) \right]} \\
 &= \frac{1}{1 + \frac{\pi_0}{\pi_1} \exp \left\{ \frac{1}{2} (\underline{x}' \Sigma^{-1} \underline{x} - \underline{x}' \Sigma^{-1} \underline{\mu}_1 - \underline{\mu}_1' \Sigma^{-1} \underline{x} + \underline{\mu}_1' \Sigma^{-1} \underline{\mu}_1 - \underline{x}' \Sigma^{-1} \underline{x} + \underline{x}' \Sigma^{-1} \underline{\mu}_0 \right. \\
 &\quad \left. + \underline{\mu}_0' \Sigma^{-1} \underline{x} - \underline{\mu}_0' \Sigma^{-1} \underline{\mu}_0) \right\}} \\
 &= \frac{1}{1 + \frac{\pi_0}{\pi_1} \exp \left\{ \frac{1}{2} (\underline{x}' \Sigma^{-1} \underline{x} - \underline{x}' \Sigma^{-1} \underline{\mu}_1 - \underline{\mu}_1' \Sigma^{-1} \underline{x} + \underline{\mu}_1' \Sigma^{-1} \underline{\mu}_1 - \underline{x}' \Sigma^{-1} \underline{x} + \underline{x}' \Sigma^{-1} \underline{\mu}_0 \right. \\
 &\quad \left. + \underline{\mu}_0' \Sigma^{-1} \underline{x} - \underline{\mu}_0' \Sigma^{-1} \underline{\mu}_0) \right\}} \\
 &= \frac{1}{1 + \frac{\pi_0}{\pi_1} \exp \left\{ \frac{1}{2} (-2 \underline{\mu}_1' \Sigma^{-1} \underline{x} + 2 \underline{\mu}_0' \Sigma^{-1} \underline{x} + \underline{\mu}_1' \Sigma^{-1} \underline{\mu}_1 - \underline{\mu}_0' \Sigma^{-1} \underline{\mu}_0) \right\}} \\
 &= \frac{1}{1 + \frac{\pi_0}{\pi_1} \exp \left\{ \frac{1}{2} (-2 (\underline{\mu}_1' \Sigma^{-1} \underline{x} - \underline{\mu}_0' \Sigma^{-1} \underline{x}) + (\underline{\mu}_1 + \underline{\mu}_0)' \Sigma^{-1} (\underline{\mu}_1 - \underline{\mu}_0)) \right\}}
 \end{aligned}$$

وبالطرق الجبرية نتوصل الى ان:

$$P_r(G_1 | \underline{x}) = \frac{1}{1 + \left\{ e^{\text{Log} \frac{\pi_0}{\pi_1} - \frac{1}{2} (\underline{\mu}_1 + \underline{\mu}_0)' \Sigma^{-1} (\underline{\mu}_1 - \underline{\mu}_0) + (\underline{\mu}_1 - \underline{\mu}_0)' \Sigma^{-1} \underline{x}} \right\}^{-1}} \quad (25)$$

والمعادلة (25) يمكن تكتب بالشكل الاتي

$$P_r(G_1 | \underline{x}) = \frac{1}{1 + \left\{ e^{\alpha + \beta \underline{x}} \right\}^{-1}} \quad (26)$$

إذ أن:

$$\begin{aligned}
 \alpha &= \text{Log} \frac{\pi_0}{\pi_1} - \frac{1}{2} (\underline{\mu}_1 + \underline{\mu}_0)' \Sigma^{-1} (\underline{\mu}_1 - \underline{\mu}_0) \\
 \beta &= (\underline{\mu}_1 - \underline{\mu}_0)' \Sigma^{-1}
 \end{aligned}$$

حيث ان المعالم $(\pi_1, \pi_0, \underline{\mu}_1, \underline{\mu}_0, \Sigma)$ تكون غير معلومة ويجب تقديرها من بيانات العينة وفق الصيغ (5)، (10)، (11)، (12) على الترتيب.

وأن احتمال المشاهدات \underline{x} تنتمي للمجموعة G_1 هو:

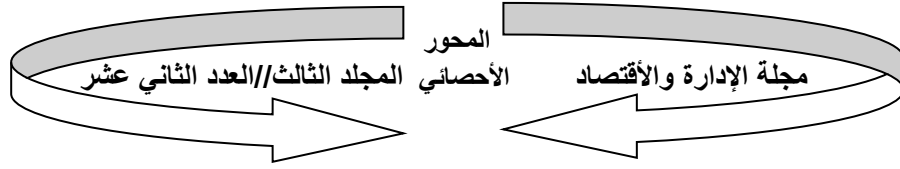
$$\bar{P}_r(G_1 | \underline{x}) = \frac{1}{1 + \left\{ e^{\hat{\alpha} + \hat{\beta} \underline{x}} \right\}^{-1}} \quad (27)$$

وا احتمال المشاهدات \underline{x} تنتمي للمجموعة G_0 هو:

$$\bar{P}_r(G_0 | \underline{x}) = 1 - \bar{P}_r(G_1 | \underline{x}) \quad (28)$$

وان قاعدة التصنيف وفق صيغة الدالة الاحتمالية

ان المشاهدة \underline{x} تعود للمجموعة G_1 اذا كان



$$\bar{P}_r(G_1|\underline{x}) > \bar{P}_r(G_0|\underline{x})$$

وتعود للمجموعة G_0 إذا كان

$$\bar{P}_r(G_1|\underline{x}) < \bar{P}_r(G_0|\underline{x})$$

2 : حالة اكثر من مجموعتين^{(٢٥)،(٢٩)}

لنفترض لدينا ثلاث مجاميع G_2, G_1, G_0 ولكل منها دالة كثافة احتمالية (p.d.f) هي $f(\underline{x}|G_0)$ هي $f(\underline{x}|G_2), f(\underline{x}|G_1)$ ولنفترض ان π_2, π_1, π_0 هي احتمالات اولية ان متجه المشاهدة \underline{x} تأتي من المجتمع G_2, G_1, G_0 على الترتيب فان التوزيع اللاحق ان \underline{x} تأتي من المجموعة G_1 هو:

$$P_r(G_1|\underline{x}) = \frac{\pi_1 f(\underline{x}|G_1)}{\sum_{i=0}^2 \pi_i f(\underline{x}|G_i)} \quad \forall i = 0, 1, 2 \quad (29)$$

وبنفس الطرق الجبريه في الفقرة السابقه فان:

$$P_r(G_1|\underline{x}) = \frac{1}{1 + \left\{ e^{\text{Log} \frac{\pi_1}{\pi_0} - \frac{1}{2}(\underline{\mu}_1 + \underline{\mu}_0)' \Sigma^{-1}(\underline{\mu}_1 - \underline{\mu}_0) + (\underline{\mu}_1 - \underline{\mu}_0)' \Sigma^{-1} \underline{x}} \right\}^{-1} + \left\{ e^{\text{Log} \frac{\pi_1}{\pi_2} - \frac{1}{2}(\underline{\mu}_1 + \underline{\mu}_2)' \Sigma^{-1}(\underline{\mu}_1 - \underline{\mu}_2) + (\underline{\mu}_1 - \underline{\mu}_2)' \Sigma^{-1} \underline{x}} \right\}^{-1}} \quad (30)$$

وأن

$$P_r(G_2|\underline{x}) = \frac{1}{1 + \left\{ e^{\text{Log} \frac{\pi_2}{\pi_0} - \frac{1}{2}(\underline{\mu}_2 + \underline{\mu}_0)' \Sigma^{-1}(\underline{\mu}_2 - \underline{\mu}_0) + (\underline{\mu}_2 - \underline{\mu}_0)' \Sigma^{-1} \underline{x}} \right\}^{-1} + \left\{ e^{\text{Log} \frac{\pi_2}{\pi_1} - \frac{1}{2}(\underline{\mu}_1 + \underline{\mu}_2)' \Sigma^{-1}(\underline{\mu}_1 - \underline{\mu}_2) + (\underline{\mu}_1 - \underline{\mu}_2)' \Sigma^{-1} \underline{x}} \right\}^{-1}} \quad (31)$$

من المعادلتين (30)،(31) نحصل على ان:

$$P_r(G_0|\underline{x}) = 1 - P_r(G_1|\underline{x}) - P_r(G_2|\underline{x}) \quad (32)$$

ويمكن اعادة كتابه المعادلات (30)،(31) بالصيغ الآتية :

$$P_r(G_1|\underline{x}) = \frac{1}{1 + \{e^{\alpha_1 + \beta_1 \underline{x}}\}^{-1} + \{e^{\alpha_2 + \beta_2 \underline{x}}\}^{-1}} \quad (33)$$

$$P_r(G_2|\underline{x}) = \frac{1}{1 + \{e^{\alpha_3 + \beta_3 \underline{x}}\}^{-1} + \{e^{\alpha_4 + \beta_4 \underline{x}}\}^{-1}} \quad (34)$$

حيث إن

$$\alpha_1 = \text{Log} \frac{\pi_1}{\pi_0} - \frac{1}{2}(\underline{\mu}_1 + \underline{\mu}_0)' \Sigma^{-1}(\underline{\mu}_1 - \underline{\mu}_0)$$

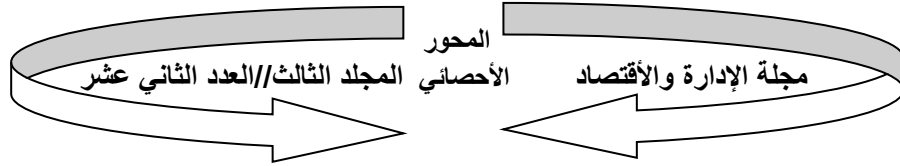
$$\beta_1 = (\underline{\mu}_1 - \underline{\mu}_0)' \Sigma^{-1}$$

وكذلك

$$\alpha_2 = \text{Log} \frac{\pi_1}{\pi_2} - \frac{1}{2}(\underline{\mu}_1 + \underline{\mu}_2)' \Sigma^{-1}(\underline{\mu}_1 - \underline{\mu}_2)$$

$$\beta_2 = (\underline{\mu}_1 - \underline{\mu}_2)' \Sigma^{-1}$$

وان



$$\alpha_3 = \text{Log} \frac{\pi_2}{\pi_0} - \frac{1}{2} (\underline{\mu}_2 + \underline{\mu}_0)' \Sigma^{-1} (\underline{\mu}_2 - \underline{\mu}_0)$$

$$\underline{\beta}_3 = (\underline{\mu}_2 - \underline{\mu}_0)' \Sigma^{-1}$$

وكذلك

$$\alpha_4 = \text{Log} \frac{\pi_2}{\pi_1} - \frac{1}{2} (\underline{\mu}_2 + \underline{\mu}_1)' \Sigma^{-1} (\underline{\mu}_2 - \underline{\mu}_1)$$

$$\underline{\beta}_4 = (\underline{\mu}_2 - \underline{\mu}_1)' \Sigma^{-1}$$

حيث ان المعامل $(\pi_1, \pi_0, \underline{\mu}_2, \underline{\mu}_1, \underline{\mu}_0, \Sigma)$ تكون غير معلومة ويجب تقديرها من بيانات العينة وفق الصيغ المشار اليها سابقاً .

هذا وأن احتمال المشاهدة \underline{x} تنتمي الى المجموعة G_1 هو

$$\bar{P}_r(G_1|\underline{x}) = \frac{1}{1 + \{e^{\hat{\alpha}_1 + \hat{\beta}_1 \underline{x}}\}^{-1} + \{e^{\hat{\alpha}_2 + \hat{\beta}_2 \underline{x}}\}^{-1}} \quad (35)$$

وا احتمال المشاهدة \underline{x} تنتمي الى المجموعة G_2 هو

$$\bar{P}_r(G_2|\underline{x}) = \frac{1}{1 + \{e^{\hat{\alpha}_3 + \hat{\beta}_3 \underline{x}}\}^{-1} + \{e^{\hat{\alpha}_4 + \hat{\beta}_4 \underline{x}}\}^{-1}} \quad (36)$$

وا احتمال المشاهدة \underline{x} تنتمي الى المجموعة G_0 هو

$$\bar{P}_r(G_0|\underline{x}) = 1 - \bar{P}_r(G_2|\underline{x}) - \bar{P}_r(G_1|\underline{x}) \quad (37)$$

وان قاعدة التصنيف وفق صيغة الدالة الاحتمالية لحاله ثلاث مجاميع فأكثر يكون :-

أن المشاهدات \underline{x} تعود للمجموعه G_1 اذا كان:

$$\bar{P}_r(G_1|\underline{x}) > \bar{P}_r(G_2|\underline{x})$$

$$\bar{P}_r(G_1|\underline{x}) > \bar{P}_r(G_0|\underline{x})$$

وهكذا بالنسبه لبقية المجاميع

٣- الجانب التطبيقي

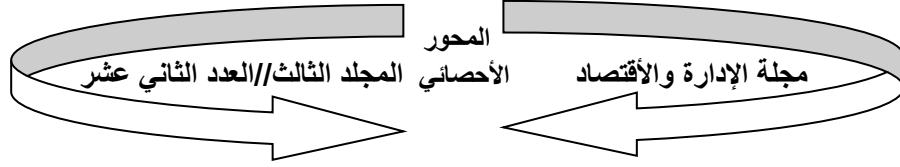
٣-١ وصف عينه البحث

تم جمع البيانات من مستشفى ابن الهيثم للعيون في محافظه بغداد بالاعتماد على الملفات الخاصه بالمرضى (طبيلات المصابين) والموجودة في قسم الاحصاء للفترة من 1/3/2012 الى 1/3/2013 .

حيث سحبت ثلاث عينات عشوائيه مكونه من (133) ، (116) ، (51) مريضاً يمثلون ثلاث مجاميع من المرضى المصابين بالتهاب العيون ،حيث تمثل المجموعه الاولى المرضى المصابين بمرض الساد (الماء الابيض) والمجموعه الثانيه تمثل المرضى المصابين بنزف السائل الزجاجي والمجموعه الثالثه تمثل المرضى المصابين بمرض داء الزرقاء.

واعتبرت الاصابه بالامراض الثلاثه .كمتغير استجابيه (Response variable). فقد اعطيت قيمه (٠)

للمريض بالساد والقيمه (١) لمريض نزف السائل الزجاجي والقيمه (٢) لمرض داء الزرقاء.



أما عدد المتغيرات التوضيحية التي تم الاتفاق عليها مع الاطباء اصحاب الاختصاص هي (٧) متغيرات يمكن توضيحها بإيجاز على النحو الآتي .

X1 : الجنس مصنف الى (ذكر (١) ، انثى (٢))

X2 : العمر (تتراوح اعمار المشمولين بالعينه بين (١-٨٨)سنة والذي قسم الى ثلاث فئات .

أ- (اطفال) بالرمز (١)

ب- (شباب) بالرمز (٢)

ج- (كبار السن) بالرمز (٣)

X3 : ضغط العين(تتراوح بين ٤-٦٠)ملم زئبق

وقسمت الى ثلاث فئات

أ- (طبيعي) بالرمز (١)

ب- (واطيء) بالرمز (٢)

ج- (عالي) بالرمز (٣)

X4 : مستوى السكر بالدم

يشعر المريض بالسكر بتغيير في حده النظر خلال اليوم الواحد حيث يتراوح بين (4.0-6.4)ملي مول لكل لتر.وقد قسمت الى ثلاث فئات

أ- (طبيعي) بالرمز (١)

ب- (متوسط) بالرمز (٢)

ج- (عالي) بالرمز (٣)

X5 : نسبة اليوريا في الدم (تتراوح بين 2.5-7.5)ملي غرام لكل ملي متر

أ- (طبيعي) بالرمز (١)

ب- (متوسط) بالرمز (٢)

ج- (عالي) بالرمز (٣)

٦-مستوى الرؤيا(تتراوح بين (6/60-6/6)

وصنفت الى ثلاث فئات

أ- (اعتيادي) بالرمز (١)

ب- (متوسط) بالرمز (٢)

ج- (ضعيف) بالرمز (٣)

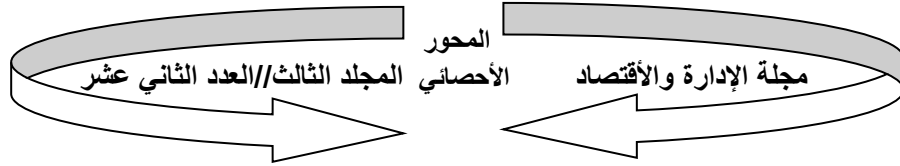
٧- وضع العصب البصري

وصنف الى ثلاث فئات

أ- (طبيعي) بالرمز (١)

ب- (متوسط) بالرمز (٢)

ج- (ضعيف) بالرمز (٣)



(٣-٢) التحليل الاحصائي للدالة التمييزية

قبل تحديد الدوال التمييزية للمجاميع الثلاث لأمراض العيون لابد من إجراء بعض الاختبارات التي تخص تحقق شرط التوزيع الطبيعي ومعنوية الدالة التمييزية وتجانس التباينات بين المجاميع الثلاث.

١- اختبار تحقق شرط التوزيع الطبيعي

تم اختبار البيانات لمعرفة فيما إذا كانت المتغيرات التوضيحية للمجموعات الثلاث لأمراض العيون تتوزع توزيعاً طبيعياً أم لا باستخدام اختبار (كولمكروف - سميرنوف) وبموجب البرنامج الإحصائي الجاهز (SPSS V20) والجدول (١) يبين نتائج الاختبار .

جدول (١) نتائج اختبار البيانات للتوزيع الطبيعي

Kolmogorov –Smirnov

Variables	Kolmogorov –Smirnov		
	Statistic	Df	Sig.
الجنس (x_1)	٦,٤٤٢	٣٠٠	0.000
العمر (x_2)	٤,٦١	٣٠٠	0.000
ضغط العين (x_3)	٥,٢٥	٣٠٠	0.000
مستوى السكر (x_4)	٤,٧٢	٣٠٠	0.000
نسبه اليوريا (x_5)	٤,٤٣	٣٠٠	0.000
مستوى الرؤيا (x_6)	٤,٥٥	٣٠٠	0.000
وضع العصب البصري (x_7)	٤,٦٥	٣٠٠	0.000

أظهرت نتائج قيم اختبار كولمكروف - سميرنوف أن متغيرات الدراسة لا تتوزع توزيعاً طبيعياً وبالنظر لكون حجم البيانات تجاوزت ٣٠ مشاهد يمكن اعتبار البيانات تتوزع بالتقريب التوزيع الطبيعي وذلك حسب نظرية الغاية المركزية^(٩)، لاغراض انجاز الجانب التطبيقي الخاص بهذه الفقرة .

٢- اختبار معنوية الدالة التمييزية الخطية

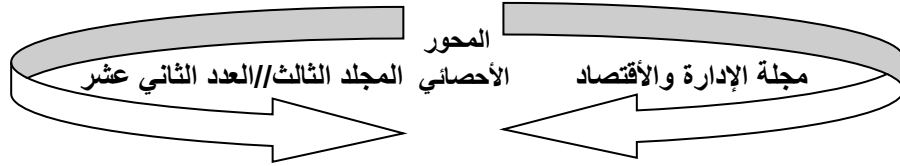
عندما يراد اختبار التمييز بين ثلاثة مجاميع فأكثر وتكوين دوال تمييزية مقبولة احصائيه بمستوى معنوية فانه لابد من اختبار معنوية الفروق بين متوسطات المجاميع قيد الدراسة وذلك بالاعتماد على الفرضية الاتية^(١٠) :-

$$H_0: \underline{\mu}_0 = \underline{\mu}_1 = \underline{\mu}_2$$

$$H_1: \underline{\mu}_0 \neq \underline{\mu}_1 \neq \underline{\mu}_2$$

وهناك عدة مقاييس لاختبار الفرضية اعلاه منها مقياس (Wilks') وتكون صيغته كالآتي^(١٤)

$$\Lambda = \frac{|W|}{|W + B|} = \frac{|W|}{|T|} \dots \dots \dots (1)$$



اذ ان

W : مصفوفة التباين والتباين المشترك داخل المجاميع

T : مصفوفة التباين والتباين المشترك الكلي للمجاميع

B : مصفوفة التباين والتباين المشترك بين المجاميع

والصيغه (1) تتوزع تقريباً مربع كاي (χ^2) بدرجة حريه $p(k-1)$ ومستوى معنويه (α) وتم اختبار معنويه الفروق بين متوسطات المجاميع الثلاث لامراض العيون بموجب الصيغه (1) وكانت نتائج هذا الاختبار كما مبينه في الجدول (2) .

جدول (٢) اختبار معنويه الداله التمييزيه

Teast fuctions	Wilks' Lamda	Chi-Squar	Df	Sig.
١ throug(٢)	0.867	41.806	١٤	0.000
٢	0.967	9.745	٦	0.013

اظهرت النتائج من خلال الجدول (٢) وجود فروق معنويه بين المتوسطات الثلاث وذلك دليل على ان هنالك فروق معنويه بين متوسطات المرضى المصابين بمرض الساد والمرضى المصابين بمرض نزف السائل الزجاجي والمصابين بمرض داء الزرقاء وهذا يعني ان الدوال التمييزيه لها القدره على التمييز اي يمكن الاعتماد عليها لتصنيف ايه مفرده الى احدى المجموعات الثلاثه .

٣- اختبار تجانس التباينات بين المجاميع الثلاث

ان احد الشروط التي يجب ان تتوفر لكي نستطيع تطبيق دالة التمييز الخطي هو تساوي مصفوفات التباين والتباين المشترك ، وأختبار بارتلنت هو أحد الاختبارات التي تطبق للتأكد من شرط تجانس التباين حيث يختبر الفرضيه الخاصه بهذا الاختبار هي :-

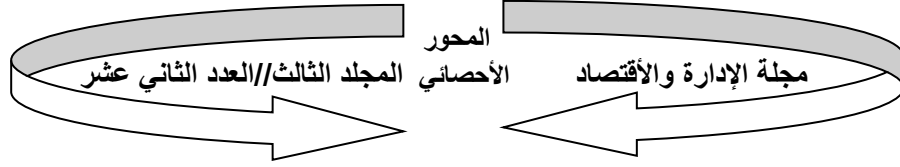
$$H_0 : \Sigma_0 = \Sigma_1 = \Sigma_2$$

$$H_1 : \Sigma_0 \neq \Sigma_1 \neq \Sigma_2$$

وتكون صيغه احصاءة الاختبار كما يأتي (12)

$$\mu = \left(\sum_{i=1}^k V_i \right) \text{Ln} |S| - \sum_{i=1}^k (V_i \text{Ln} |S_i|) \quad (2)$$

$$S = \frac{1}{\sum_{i=1}^k V_i} \sum_{i=1}^k V_i S_i \quad (3)$$



حيث ان

S_p : مصفوفة التباين والتباين المشترك التقديرية (pooled covariance).

S_i : تباين العينة i اذ $(i=1, \dots, k)$

درجة الحرية للعينة i : $V_i = n_i - 1$.

k : عدد المجموع.

وقد أثبت Box عام (1949) أنه إذا ضرب μ في ثابت C^{-1} يساوي:

$$C^{-1} = 1 - \frac{2p^2 + 3p - 1}{6(p+1)(k-1)} \left[\sum_{i=1}^k \frac{1}{V_i} - \frac{1}{\sum_{i=1}^k V_i} \right] \quad (4)$$

اذ ان

p : عدد المتغيرات التوضيحية

فسوف نحصل على مقياس يتوزع بالتقريب توزيع χ^2 وبدرجه حريه $\frac{1}{2}(k-1)p(p+1)$

حيث

$$\mu C^{-1} \sim \chi^2 \left[\frac{1}{2}(k-1)p(p+1) \right] \quad (5)$$

وفقاً لهذا الاختبار تم التوصل الى النتائج في الجدول (3)

جدول (3) اختبار تجانس التباينات بين المجموع الثلاث

Box's M	76.460
F Approx.	1.309
Df1	56
Sig	.060

وفقاً لجدول (3) اظهرت النتائج ان قيمه $P.Value > 0.05$ هذا يشير الى قبول فرضيه العدم ورفض

الفرضيه البديله وهذا يدل على تجانس التباينات بين المجموع الثلاث. وهذا يعني تحقق شرط استخدام الدالة

التمييزيه الخطيه

(3-3) اختبار معنويه المتغيرات في الداله التمييزيه

تم اختبار معنويه المتغيرات لمعرفة اهميه كل متغير بشكل منفرد ومدى تأثيره في بناء الداله التمييزيه الخطيه

وكانت النتائج كما في الجدول (4)

جدول (٤) اختبار معنوية متغيرات الداله التمييزيه

Variables	Wilks' Lamda	F	df1	df2	Sig.
الجنس (x_1)	1.000	0.016	2	297	.0284
العمر (x_2)	0.944	8.870	2	297	0.000
ضغط العين (x_3)	0.985	2.201	2	297	0.0011
سكر (x_4)	0.995	0.741	2	297	0.04
اليوريا (x_5)	0.996	0.564	2	297	.016
مستوى الرؤيا (x_6)	0.962	5.799	2	297	.003
وضع العصب البصري (x_7)	0.982	2.653	2	297	.002

من خلال جدول (٤) يتضح ان جميع متغيرات الدراسة لها تاثير واهميه في تكوين وبناء الداله التمييزيه الخطيه .

(٣-٤) تقدير الدوال التمييزية الخطية

تم تقدير الدوال التمييزيه الخطيه للمجاميع الثلاث لمرضى العيون وفق الصيغه (22) وذلك بعد تقدير موجهاً متوسطات العينات الثلاث ومصنوفه التباين والتباين المشترك المدمجه باستخدام مقدرات طريقه الامكان الاعظم وكما في الجدول (5) .

جدول (٥) الدوال التمييزيه الخطيه التقديرية

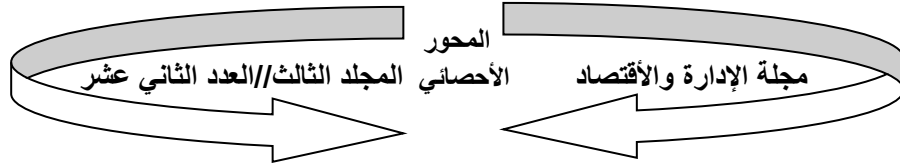
Variables	الداله الاولى	الداله الثانية	الداله الثالثة
Constant	-19.127	-19.461	-21.005
الجنس (x_1)	5.421	5.585	5.467
العمر (x_2)	3.070	2.329	3.023
ضغط العين (x_3)	2.766	2.559	3.059
مستوى السكر (x_4)	1.381	1.563	1.646
نسبه اليوريا (x_5)	2.706	2.654	2.346
مستوى الرؤيا (x_6)	2.097	2.603	2.471
وضع العصب البصري (x_7)	3.371	3.838	3.851

ووفقاً لجدول (٥) يمكن كتابه الدوال التمييزيه الخطيه كما يلي

$$d_0^*(x) = 5.42X_{01} + 3.07X_{02} + 2.76X_{03} + 1.38X_{04} + 2.70X_{05} + 2.09X_{06} + 3.37X_{07} - 19.12.....(6)$$

$$d_1^*(x) = 5.58X_{11} + 2.32X_{12} + 2.55X_{13} + 1.56X_{14} + 2.65X_{15} + 2.60X_{16} + 3.83X_{17} - 19.46.....(7)$$

$$d_2^*(x) = 5.46X_{21} + 3.02X_{22} + 3.05X_{23} + 1.64X_{24} + 2.34X_{25} + 2.47X_{26} + 3.85X_{27} - 21.005.....(8)$$



ويتضح من خلال قيم معاملات المتغيرات التوضيحية ووفق دوال التمييز الخطي الثلاثة الأهمية التمييزية لمتغير الجنس ، يليه متغير وضع العصب البصري ، ثم متغير نسبة اليوريا و متغير ضغط العين ، و متغير نسبة الرؤيا ، و متغير نسبة السكر و أخيراً متغير العمر في تشخيص الأمراض الثلاثة من أمراض العيون في جميع دوال التمييز الخطية .

(3-5) التصنيف حسب دالة التمييز الخطي

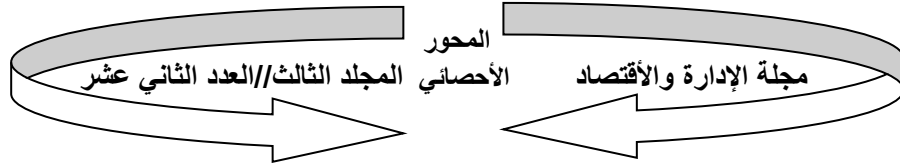
لتصنيف أي مفردة بالاعتماد على الدوال التمييزية التقديرية بالصيغ (6)، (7)، (8) فإننا نعوض قيم المتغيرات التابعة لهذه المفردة في المعادلات المذكورة فإذا كانت $d_0^*(x) > d_1^*(x), d_2^*(x)$ فإن المفردة تصنف للمجموعة الأولى (0) وإذا كانت $d_1^*(x) > d_0^*(x), d_2^*(x)$ فإنها تصنف للمجموعة الثانية (1) وإذا $d_2^*(x) > d_0^*(x), d_1^*(x)$ فتصنف للمجموعة الثالثة (2) .

إن عملية التصنيف قد تؤدي إلى الوقوع فيما يعرف بخطأ التصنيف (misclassification) وهو احتمال تصنيف مفردة معينة إلى المجموعة الأولى بينما هي في الحقيقة تعود للمجموعة الثانية أو الثالثة وبالعكس وهناك نوعان من الخطأ الأول يسمى بنسبه الخطأ الظاهر والثاني بنسبه الخطأ الحقيقي. وفي هذه الدراسة تم التركيز على احتساب النوع الأول من الخطأ لجميع مشاهدات العينات الثلاث والجدول (6) يمثل التصنيف حسب دوال التمييز الخطية الثلاث .

جدول (6) تصنيف المشاهدات حسب دوال التمييز الخطي الثلاثة

التصنيف				الحالة
نسبه التصنيف الصحيح	اصبح المريض عائد الى المجموعة الثالثة (2)	اصبح المريض عائد الى المجموعة الثانية (1)	اصبح المريض عائد الى المجموعة الاولى (0)	
49.6%	34	33	66	المريض عائد الى المجموعة الاولى (0)
50.8%	23	59	34	المريض عائد الى المجموعة الثانية (1)
47.0%	24	14	13	المريض عائد الى المجموعة الثالثة (2)
49.6%	27.0%	35.4%	37.6%	نسبه التصنيف الكلي

أظهرت النتائج من الجدول (6) ان احتمال التصنيف الصحيح لمريض يعود إلى المجموعة الأولى (0) كانت 49.6%، بينما احتمال خطأ التصنيف 50.4%. كذلك أشارت النتائج بان احتمال التصنيف الصحيح لمريض يعود إلى المجموعة الثانية (1) كان 50.8% واحتمال خطأ التصنيف لتلك المجموعة 49.2% في حين كان احتمال التصنيف الصحيح لمريض يعود إلى المجموعة الثالثة (2) 47% واحتمال خطأ التصنيف 53% وقد بلغت نسبة التصنيف الصحيح الكليه 49.6% في حين نسبة التصنيف الخاطيء الكليه 51.4%.

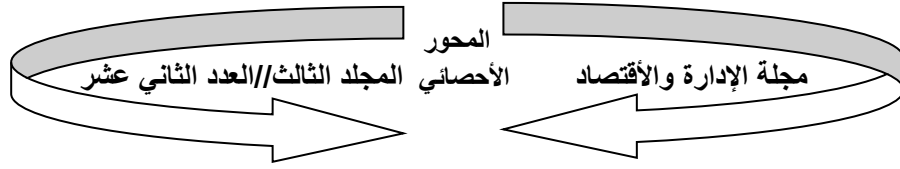


(3-6) /التصنيف حسب الصيغة الاحتمالية

لتصنيف اي مفردة من مفردات المجموعات الثلاثة قيد الدراسة باستخدام الصيغة الاحتمالية في الفقره (6-1) ولحالة ثلاث مجموعات فقد تم تقدير احتمال المشاهده \bar{X} تنتمي لاحدى المجاميع الثلاث حسب الصيغ (35) ، (36) ، (37) . والجدول رقم (٧) يمثل التصنيف حسب الصيغة الاحتمالية
جدول (٧) تصنيف المشاهدات حسب الصيغة الاحتمالية لدالة التمييز الخطيه

التصنيف				الحالة
نسبه التصنيف الصحيح	اصبح المريض عائد الى المجموعه الثالثه (٢)	اصبح المريض عائد الى المجموعه الثانيه (١)	اصبح المريض عائد الى المجموعه الاولى (٠)	
66.1%	2	43	88	المريض عائد الى المجموعه الاولى (٠)
53.4%	1	63	52	المريض عائد الى المجموعه الثانيه (١)
60.78%	31	19	1	المريض عائد الى المجموعه الثالثه (٢)
60.67%	11%	42%	47%	نسبه التصنيف الكلي

يتبين من نتائج جدول (٧) ان احتمال التصنيف الصحيح لمفرده تعود الى المجموعه الاولى (٠) كانت 66.1% بينما احتمال خطأ التصنيف 39.9% وكذلك احتمال التصنيف الصحيح لمفرده تعود الى المجموعه الثانيه (١) كان 53.4% واحتمال خطأ التصنيف 46.6% وكان احتمال التصنيف لمفرده تعود الى المجموعه الثالثه (٢) هو 60.78% واحتمال خطأ التصنيف 39.2% وقد بلغت نسبة التصنيف الصحيح الكليه 60.67% في حين التصنيف الخاطيء الكليه 39.33% .
ان نتائج احتمال خطأ التصنيف الكلي للجدولين (7,6) تشير إلى ان تصنيف المشاهدات باستخدام الصيغة الاحتمالية اعطى تفوقاً واضحاً على التصنيف باستخدام دوال التمييز الخطي باعطائه اقل احتمال لخطأ التصنيف .



(4) الاستنتاجات والتوصيات

(4-1) الاستنتاجات

١- تبين من خلال تحليل البيانات ان الإصابة بأمراض العيون تتركز في الفئة العمرية (32-63) حيث بلغت النسبة (46.3%) وللذكور اكثر من الاناث في الإصابة بتلك الامراض حيث بلغت نسبة الذكور (50%) وتبين من النتائج كذلك انه في ضغط العين عند المستوى (4-24) هم اكثر اصاباه بأمراض العيون حيث كانت النسبة عند هذا المستوى (88%) اما بالنسبة لمستوى السكر وعند المستوى (1.70-7.7) هم اكثر عرضه للاصابة بتلك الامراض وبلغت النسبة (87.3%) وبالنسبة لليوريا تكون النسبة الاكثر اصابة في المستوى (3.6-9.6) وكانت نسبتها (97.7%) اما بالنسبة الى مستوى الرؤيا تكون النسبة الاكثر اصابة هي عند المستوى (0.1-0.4) حيث بلغت نسبتها (44.7%) وكذلك بالنسبة لوضع العصب البصري بين (24-44) ونسبتها (48%) وفق نتائج جدول (١٠).

٢- ظهر لنا من خلال نتائج (الدالة التمييزية الخطية) ان جميع متغيرات الدراسة لها تأثير واهمية في تكوين وبناء الدالة التمييزية الخطية وتبين ان لمتغير الجنس الاهمية الاكبر ويليه متغير وضع العصب البصري ثم متغير نسبة اليوريا ومتغير ضغط العين ومتغير مستوى الرؤيا ويليه متغير نسبة السكر واخيراً متغير العمر.

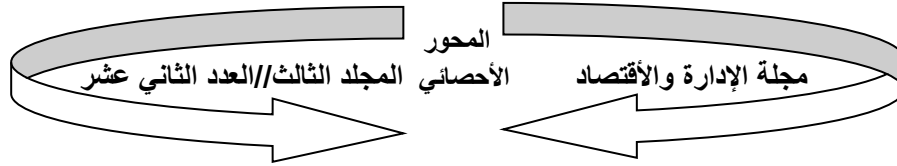
٣- تبين من خلال تصنيف البيانات وفق صيغه الدالة التمييزية الخطية وصيغه احتمال الاستجابة لدالة التمييز الخطي وتم التوصل الى ان طريقة التصنيف حسب الصيغه الاحتمالية لدالة التمييز الخطي تعطي نتائج اقل احتمال خطأ التصنيف حيث بلغت نسبة التصنيف الصحيح الكلي لها 60.67% في حين نسبة التصنيف الخاطيء بلغت 39.33% .

(4-2) التوصيات

- ١- نوصي باستخدام دوال التمييز الاخرى (التربيعية- اللوجستي الخطي -اللوجستي التربيعي) في تصنيف البيانات ذات المتعدده الاستجابة .
- ٢- نوصي ان تكون مقارنه احتمال خطأ التصنيف وفق احتمالات الاستجابة .
- ٣- نوصي بتطوير مراكز الاحصاء في المستشفيات وازافه معلومات جديدة للمرضى المصابين.
- ٤- نوصي بتشخيص المرض واجراء العمليات الجراحية اللازمة في وقت مبكر .

أولاً: المصادر العربية

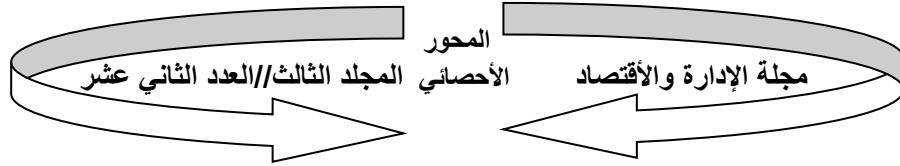
١. الجاعوني، فريد وغانم(٢٠٠٧) "التحليل الاحصائي متعدد المتغيرات (التحليل التمييزي) (في توصيف وتوزيع الاسر داخل الهيكل الاقتصادي الاجتماعي في المجتمع" بحث منشور ،مجلة كلية الاقتصاد، جامعه دمشق.
٢. الجبوري، د. شلال حبيب وعبد. د. صلاح حمزه (٢٠٠٠) "تحليل متعدد المتغيرات ،دار الكتب للطباعة والنشر .بغداد ،العراق.
٣. الراوي، خاشع محمود، (1987)، "المدخل إلى تحليل الانحدار"، طبع مؤسسة دار الكتب للطباعة والنشر، جامعة الموصل.
٤. العباسي. عبد الحميد محمد(٢٠١١) "تطبيقات في العلوم الاجتماعية باستخدام SPSS قسم الاحصاء الحيوي والسكاني جامعه القاهرة.



٥. الياسين .دريد حسين بدر " استخدام بعض طرائق التمييز الحصينه لتشخيص امراض سرطان الدم(اللوكيميا)رساله ماجستير احصاء.كلية الادارة والاقتصاد ،الجامعة المستنصرية
٦. جونسون ،ريتشارد و شرن "التحليل الاحصائي للمتغيرات المتعددة من الوجهه التطبيقية "طبع مؤسسه دار المريخ للطباعه والنشر ، جامعه الملك سعود.
٧. حمودات،الاء عبد الستار (٢٠٠٥) "الدالة التمييزيه وطرق تحديد متغيراتها" رساله ماجستير في علوم رياضيات .جامعه الموصل.
٨. حميد ،رند سليم (١٩٩١) " استخدام الداله المميزه في تشخيص بعض الاورام السرطانيه "رساله ماجستير في الاحصاء مقدمه الى كلية الادارة والاقتصاد ،جامعه بغداد .
٩. شومان ،عبد اللطيف حسن(1977) و" التحليل المميز وتطبيقه في تصنيف طلاب المدرسة الثانوية " ، رساله ماجستير ، جامعه الموصل
١٠. محمد . سميره محمد صالح (٢٠٠٣)"استخدام التحليل المميز (التصنيفي) لتحديد اهم العوامل المؤثره في تسرب ورسوب الطلبة في جميع المراحل الدراسيه "،رساله ماجستير في الاحصاء ،كلية الادارة والاقتصاد ،جامعه السليمانيه.

ثانياً :- المصادر الأجنبية

- 11- Abbas, F. M. Azhar, M. E (2008) " Comparing Discriminant Analysis and Logistic Regression Model"as a Statistical Assessment Tools of Arsenic and Heavy Metal Contents in Cockles"School of Industrial Technology, Environmental Technology Division Universiti Sains Malaysia, 11800 Penang, Malaysia.
- 12- Agresti,A.(1990)" Categorical Data Analysis "، John Wiley Sons , Inc , New York
- 13- Al-Thabhawe ,G .D (2012) "A Comparison between Discriminant Analysis and Logistic Regression on the Classification of Cancer Patients" for the Degree of Master of Science in Mathematics (Mathematical Statistic) University of Kufa.
- 14- Alvin C.Rencher"Methods of Multivariate Analysis" John Widy and Sons,puplication.
- 15- Anderson,T.W.(1918)"An introduction to multivariate statistical analysis"by John Wiley &Sons.Inc
- 16- Davidson ,R .and James ,G. Mackinon .(2003)," Econometric Theory and Method " ، Answer to Starred Exercise .
- 17- Dillon,W.R and Goldistein,M(1984)"Multivariate Analysis Methods and Application".John Wily &Sons,Inc,New York,USA.
- 18- Feighner, J. P., & Sverdlov, L. (2002) " The use of discriminant analysis to separate a study population by treatment subgroups in a clinical trial with a new pentapeptide antidepressant". *Journal of Applied Research*, 2; 17 – 18.
- 19- George C.J Fernandez"Discriminant Analysis ,Apowerful Classification Technique in Data mining".
- 20- Harris,R.J(1975)"Aprimer of Multivariate Statistics.Academic press,INC,New York,USA.



- 21- James.b,Chew.c&bron.a(2002)"Ophthalmology book ninth edition"Blackweiy puplishing.
- 22- Kashler,Sagar,M.A.&Keysers(2003),"Multivariate Analysis, New York."
- 23- Kenneth ,T .(2003) ” Discrete Choice Model with Simulation “,Cambridge University Press . <http://elsa.berkeley.edu>
- 24- Krieng kitbumrungrat (2012)"Comparison logistic regreesion and Discriminant analysis in classification groups for breast cancer"Faculty of Information Technology, Rangsit University, Thailand
- 25- Lebart,L..Morineau,A.andWarwick,K.M(1984):Multivariate Descriptive Statistical Analysis.John Widy and Sons,NewYork,Chester,Bisbance,Toruta,Singapore.
- 26- Lei, P., & Koehly, L. M. (2003). Linear discriminant analysis versus logistic regression: a comparison of classification errors in the two–group case. *Journal of Experimental Education*, 72; 25 – 49.