Using the Discriminate function to classify the death and survivors cases of premature infants

استخدام الدالة المميزة لتصنيف حالات الوفاة والبقاء عند الاطفال الخدج

م. د. شروق عبد الرضا سعيد السباح
جامعة كربلاء / كلية الطب

الملخص//

من المعلوم ان الدالة المميزة تستخدم في تصنيف المفردات الى واحد او أكثر من المجاميع بالاعتماد على متغيرات ذات صفات تمييزيه، وتمتاز هذه المجاميع بالتجانس بين مفرداتها ولكنها مختلفة بين مجموعة وأخرى. ولغرض فحص مدى وجود فروق ذات دلالة معنوية بين مجموعتين (وفاة وبقاء الطفل الخديج) بالنسبة للمتغيرات المستقلة وتحديد المتغيرات المستقلة التي تساهم بأكبر قدر من الاختلاف بين فئات المتغير التابع ولتقسيم الحالات بين فئات المتغير التابع بناء على قيم المتغيرات المستقلة وتقيم دقة التقسيم (كنسبة مئوية) أي اختبار الجانب النظري على ضوء التصنيف ان كان حسب المتوقع ام المتوقع غير صحيح نستخدم الدالة المميزة.

اذ كانت البيانات التي تمثل وفاة او عدم وفاة الطفل الخديج هو متغير الاستجابة [والجنس (ذكر ، أنثى) والمستشفى (كربلاء ، نسائية) ونوع الولادة (طبيعية ، قيصرية ، اخرى) وسبب الرقود(نقص الوزن ، قصر مدة الحمل ، ذات الرئة الولادي ، متلازمة شفط السوائل الرحمية ، يرقان ولادي ، تشوهات خلقية ، متلازمة ضيق التنفس (عسر التنفس) ، اختناق ولادي ، تسمم الدم الجرثومي ، اضطرابات في دم الجنين ، إصابة أثناء الولادة ، أخرى) وعمر الأم ، وزن الطفل ، مدة حمل الطفل] كمتغيرات مستقلة تؤثر في حياة الطفل.

وعليه فقد توصل البحث الى ان المتغيرين الأفضلين في التمييز كانا مدة الحمل وسبب الرقود فقد كان معامل الارتباط بين الدالة المميزة ومدة الحمل 0.72 ومعامل الارتباط الدالة المميزة وسبب الرقود كان 0.511- ،وتبين ان اعلى معامل ارتباط ثنائي كان بين مدة الحمل ووزن الطفل وهو 0.716 يليه معامل الارتباط الثنائي بين نوع الولادة ووزن الطفل وهو0.327، والى وجود فروق ذات دلالة معنوية بين مجموعتي الوفيات والباقين على قيد الحياة بالنسبة للمتغيرات المستقلة فقد كان هذا واضحا من ملاحظة معنوية تباين الجذور المميزة للدالة المميزة وان المتغير المستقل مدة الحمل كان له الوزن الأكبر المؤثر في زيادة قوة التمييز بين المجموعتين لذا تم تسمية الدالة بأسمه وهي (دالة مدة الحمل)

Summary

It is known that the discriminate function is used to classify items into one or more groups depending on discriminate variables. Each group has characteristics that are available in its items and differs from the others. In order to investigate whether there are significant differences between two groups (the death of premature infant) in relation to the independent variables , to specify these independent variables that play a significant role of differing the categories of the followed variable, to divide the cases among the followed variable depending on the value of independent variables, and to evaluating the accuracy of division (as a percentage) i.e testing the theoretical side in the light of the classification whether the expected is true or not we use the discriminate function.

If the data that represent the death or survivors of the premature infant is the response variable, Sex (male, female), Hospital (kerbala, women), The way of delivery (normal, suzerain, other), The cause of admission (Least of weight, Shortage of pregnancy period, congenital pneumonias ,Meconium Aspiration syndrome ,Respiratory distress syndrome , Neonatal jaundice ,Hematological disorder of fetus , Septicemia , congenital anomalies ,Birth asphyxia ,Birth trauma ,other disorders) The age of the mother , the weight of the infant, the period of pregnancy as independent variables that affect on the baby's staying alive.

Through out the research, we found that the best variables of distinction were the period

of pregnancy and the cause of admission. The relation factor between the discriminate function and the period of pregnancy was 0.72., the relation factor between the discriminate function and the cause of admission was 0.511. It is also found that the highest binary relation factor was between the period of pregnancy and the weight of the infant which is 0.716, followed by the binary relation factor between the type of delivery and the weight of the baby which is 0.327. It is also found that there are differences of significant reference between the group of dead and alive babies in relation to the independent variables. That was clear through noting the difference between the origins of the discriminate function. The independent variable (period of pregnancy) had the biggest effect in increasing the distinction between the two groups. This is why the function called with its name (period of pregnancy function)

مقدمة ومشكلة البحث

يستخدم التحليل المميز في التفريق بين مجتمعات متداخلة ذات خصائص مشتركة بالشكل الذي يجعلنا غير واثقين او متأكدين من انتماء بعض المفردات الى أي من المجاميع ، ويستخدم في تصنيف المفردات الى واحد او أكثر من المجاميع بالاعتماد على متغيرات ذات صفات تميزيه، وتمتاز هذه المجاميع بالتجانس بين مفرداتها ولكنها مختلفة بين مجموعة وأخرى ، ويمكن وصف المجتمعات باستخدام التحليل المميز من خلال الكشف عن الاختلافات الموجودة بين مفردات مجتمع معين ويستفاد منه في التعرف على المتغيرات التي تساهم في عملية التصنيف

هدف البحث

تتحدد اهداف البحث بما يلي :

1-تصميم وظائف التمييز او التوليفات الخطية للمتغيرات المستقلة الأفضل في التمييز والتي تتمثل بالجنس (ذكر ، أنثى) ونوع المستشفى (كربلاء للأطفال . نسائية) ونوع الولادة (طبيعية . قيصرية . أخرى) وسبب الرقود (نقص الوزن ، قصر مدة الحمل ، ذات الرئة الولادي ، متلازمة شفط السوائل الرحمية ، يرقان ولادي ، تشوهات خلقية ، متلازمة ضيق التنفس (عسر التنفس) ، اختناق ولادي ، تسمم الدم الجرثومي ، اضطرابات في دم الجنين ، إصابة أثناء الولادة ، أخرى) وعمر الأم ، وزن الطفل ، مدة حمل الطفل] بين فئات المتغير التابع حياة الطفل الخديج .

2-فحص مدى وجود فروق ذات دلالة معنوية بين مجموعتين (وفاة وبقاء الطفل الخديج) بالنسبة للمتغيرات المستقلة

3-تحديد المتغيرات المستقلة التي تساهم بأكبر قدر من الاختلاف بين فئات المتغير التابع

4-تقسيم الحالات بين فئات المتغير التابع بناء على قيم المتغيرات المستقلة

5-تقيم دقة التقسيم (كنسبة مئوية) أي اختبار الجانب النظري على ضوء التصنيف ان كان حسب المتوقع ام المتوقع غير صحيح

وإجمالا فإنه يمكن القول بأن التحليل التميزي يعمل على ايجاد دالة للتمييز وذلك من خلال احتساب قيم المتغيرات الكمية اذ تقوم دالة التمييز بالتنبؤ برقم المجموعة التي ينتمي اليها كل طفل فالدالة التميزية هي عبارة عن توليفه من المتغيرات المستقلة التي يمكن استخدامها في عملية التنبؤ بانتماء الحالة الى إحدى المجموعتين او أكثر وبالتالي تتم عملية التصنيف على أساس دالة تميزية واحدة عندما يكون هناك مجموعتين (وفاة وبقاء)وكقاعدة عامه فإن عدد الدوال التميزية مساوي الى عدد المجموعات مطروح منه واحد .

فرضية ومنهجية البحث

ان فرضية البحث تتجسد في وفاة وبقاء الطفل الخديج في عديد من المتغيرات توثر على كل طفل اذ تمثلت بالمعلومات الاساسية المتوفرة في طبلة كل طفل عند دخوله الى ردهة الخدج وهدفنا تحديد أي المتغيرات اهم من غيرها بالدرجة الأولى تؤثر في بقاء الطفل الخديج ومن خلال استخدام الدالة المميزة وعليه تم تقسيم البحث الى جزئين الأول يتعلق بالجانب النظري للدالة المميزة وكيفية اشتقاقهما والوصول الى الصيغة النهائية لها فيما يتعلق الثاني بالجانب العملي حيث اخذت بيانات من كل مستشفيات المحافظة وجمعت المعلومات عن كل طفل لتمثل لدينا متغيرات مستقلة(المستشفى (كربلاء . نسائية) ,الجنس(ذكر . أنثى) ,مدة الحمل ,الوزن .نوع الولادة (قيصرية, طبيعية ,اخرى),سبب الرقود(نقص الوزن ، قصر مدة الحمل ، ذات الرئة الولادي ، متلازمة شفط السوائل الرحمية ، يرقان ولادي ، تشوهات خلقية ، متلازمة ضيق التنفس (عسر التنفس) ، اختناق ولادي ، تسمم الدم الجرثومي ، اضطرابات في دم الجنين ، إصابة أثناء الولادة ، أخرى) وعمر الأم ، وزن الطفل ، مدة حمل الطفل)توثر في حياة الطفل او وفاته من خلال تطبيق الدالة المميزة تبين ان بعضها لها دور أساسي في بقاء الطفل الخديج على قيد الحياة دون البعض الاخر.

الدراسات السابقة

في عام 1921 بدأ parson بإيجاد مقياس التشابه للأشياء في المجتمعات المختلفة [4].

في عام 1925 أوجد Mahalanobios مقياس المسافات والوصول الى الدالة المميزة الخطية لمجموعتين بأفتراض أخذ المتغيرات قيم مستمرة ، وصيغة مقياس مهاللنوبس [6]

$$D^2=(Xi-Xj)\ ´S^{-1}(Xi-Xj) \quad \text{---------- (1)}$$

$$S=\frac{(n1-1)s1+(n2-1)s2}{n\square+n2-2} \quad \text{----------(2)}$$

وفي عام 1931 استخدم hotelling مقياس Mahalanobis في اختبار t لعدة متغيرات وأوجد الصيغة الآتية[3]

$$T^2=\frac{n_1n_2}{n_1+n_2}D^2 \quad \text{---------- (3)}$$

في عام 1936 قام Fisher بتطوير مقياس مهالنوبس وذلك بإيجاد الدالة المميزة الخطية لتصنيف المفردات الى إحدى مجموعتين معينتين وبنى الدالة على أساس اختيار الدالة التي تعظم النسبة بين المجموعات الى داخل المجموعات والدالة مبنية على أساس تساوي تباين المجموعات والصيغة الرياضية هي[7]

$$\underline{X}\,'\sum^{-1}(\underline{\mu_1}-\underline{\mu_2}) \quad \text{----------(4)}$$

وفي حالة الاعتماد على مفردات العينة

$$\underline{X}\,'\,S^{-1}\,(\bar{\mathbf{X}}_1-\bar{\mathbf{X}}_2) \quad \text{---------- (5)}$$

وقاعدة التصنيف هي ان نضيف المفردة الجديدة x الى المجموعة الأولى إذا كانت قيم الدالة (بعد ان يتم التعويض بقيم x للمفردة المراد ان تصنف) اكبر من ثابت معين وبعكسه يتم تصنيفها الى المجموعة الثانية .

في عام 1939 لاحظ Welch ان اختبار نسبة الإمكان الأعظم للتوزيعات الطبيعية متعددة المتغيرات تعطي نتائج تكون ضرورية لمعرفة معالم التوزيع وقد اوجد Welch(1944) Anderson (1951) Sitgreares (1952) توزيع المعادلة.[5]

$$\underline{X}\,'\,S^{-1}\,(\bar{\mathbf{X}}_1-\bar{\mathbf{X}}_2)\text{-}1\backslash 2(\bar{\mathbf{X}}_1-\bar{\mathbf{X}}_2)\,'\,S^{-1}(\bar{\mathbf{X}}_1+\bar{\mathbf{X}}_2) \quad \text{---------- (6)}$$

وفي عام 1940 استخدم Guttman طريقة اخرى بسيطة للتعامل مع البيانات التي تكون على شكل اختبارات متعددة أي التعامل مع البيانات الوصفية (النوعية) حيث أن أوزان الاستجابة لكل متغير في الدراسة سيحدد بالشكل الذي يجعل نسبة التباين بين الاختبارات الى التباين الكلي أقل ما يمكن . وتتلخص هذه الطريقة بتحديد أوزان الاستجابة ثم تقدير المتغيرات بألاعتماد على متوسط الاستجابة الموزون حيث أن أوزان الاختبارات سوف تعطي بوصفها نسبة لمتوسط هذه الاختبارات .

في عام 1944 اشتق Vonmises العلاقة التي تعظم الاحتمال الأصغر للتصنيف الصحيح .[9]

في عام 1947 عالج Perirose مشكلة التصنيف لبيانات تتبع التوزيع الطبيعي وبمصفوفتي تباين وتباين مشترك مختلفة .

وفي سلسلة دراسات Rao في الأعوام 1946, 1947, 1948 , 1949, 1950, إقترح الطرق المختلفة لتصنيف مجموعتين أو أكثر ومن هذه الطرق مقياس للمسافة بين المجاميع مع إفتراض أن التوزيعات معلومة [8].

وفي عام 1951 حاول J.B.Bryan تعميم الدالة المميزة ل(Fisher) إلى أكثر من مجموعتين.

وفي عام 1966 أعطى Hills خلاصة لمشكلة تحديد إحتمال التصنيف الخاطئ للوحدات الجديدة غير معروفة التصنيف باستخدام دالة التمييز المعتمدة على العينة السابقة . بتقدير الوسط الحسابي والخطأ المعياري إلى D الذي يمكن تقديره

$$P1=\Phi\left(\frac{-D}{2}\right) \quad \text{---------- (7)}$$

$$D^2=(\bar{\mathbf{X}}_1-\bar{\mathbf{X}}_2)'\,S^{-1}(\bar{\mathbf{X}}_1-\bar{\mathbf{X}}_2) \quad \text{----------(8)}$$

وهي إحصائية Mahalanobias لقياس المسافة او البعد بين عينتين وهي طريقة تعتمد على إفتراض أن المتغيرات تتوزع طبيعيا [2].

اما البحث الحالي فيستخدم الدالة المميزة لتشخيص حالات الوفاة والبقاء عند الأطفال الخدج

الدالة المميزة الخطية في حالة مجموعتين

تعتبر هذه الطريقة من الطرق المعلمية ويستخدم هذا النوع من الدوال عندما يكون المجتمع المدروس ذا توزيع طبيعي متعدد المتغيرات وبمتوسطات μ_1 , μ_2 مختلفة وبمصفوفة تباين وتباين مشترك متساوي $\sum$.

فأن مجال العينة R يقسم الى قسمين R_1 يعود الى المجموعة الأولى و R_2 يعود الى المجموعة الثانية والحد الفاصل بين المجموعتين R_1,R_2 يمكن ان يعود الى المجموعة الأولى أو الثانية وفي مثل هذه الحالة سيكون

$$f_1=f_2 \quad \text{----------(9)}$$

حيث أن : f_2,f_1 دوال احتمالية تتوزع توزيعا طبيعيا متعدد المتغيرات وللمجموعتين الأولى والثانية على التوالي

$$\log f_1 = \log f_2 \quad \text{----------(10)}$$

وعلى فرض تساوي التباينات وبمتوسطات مختلفة μ_1 , μ_2 .

$$(X-\mu_1)' \sum^{-1} (X-\mu_1) = (X-\mu_2)' \sum^{-1} (X-\mu_2) \text{----------(11)}$$

$$X' \sum^{-1} (\mu_1 - \mu_2) - \tfrac{1}{2}(\mu_1 + \mu_2)' \sum^{-1} (\mu_1 - \mu_2) = 0 \text{----------(12)}$$

وهي دالة التمييز عند الحد الفاصل بين المجموعتين عندما تكون معالم المجتمع معلومه وفي حالة المعالم مجهوله فتقدر بالاعتماد على قيم العينتين بحجم n_1 , n_2 وتصبح الدالة كما يأتي.

$$DL(\square) = X' S^{-1}(\bar{X}_1 - \bar{X}_2) - \frac{1}{2} (\bar{X}_1 + \bar{X}_2)' S^{-1}(\bar{X}_1 - \bar{X}_2) \text{----------(13)}$$

$$S = \frac{(n1-1)s1 + (n2-1)s2}{n\square + n2 - 2}$$

$\bar{X}_1$, $\bar{X}_2$:تمثل تقدير الأمكان الأعظم ل μ_1 ,μ_2

S : يمثل تقدير الأمكان الأعظم لمصفوفة التباين والتباين المشترك $\sum$.

S_1,S_2 :يمثل تقدير مصفوفة التباين والتباين المشترك للعينة الأولى والثانية على التوالي .

وأن $DL(\square)$ هي الحد الفاصل المقدر للدالة الخطية المميزة بين مجموعتين , وسوف تصنف المشاهدات الجديدة على اساس انها تعود الى المجموعة الاولى اذا كانت اكبر من الصفر وإلا فأنها تصنف الى المجموعة الثانية إذا كانت لأم أصغرمن الصفر عدا ذلك فانها تصنف عشوائيا الى أي من المجتمعين .

لو فرضنا ان Z هي دالة خطية حيث ان :

$$Z = C' X$$

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

المعادلة سوف تصبح

$$DL(\square) = Z - \frac{1}{2} C(\bar{X}_1 + \bar{X}_2) > \leq 0 \quad \text{----------(14)}$$

سوف يكون الحد الاول هي الدالة المميزة التي وضعها فيشر وهي دالة خطية لها اعظم تباين ممكن بين المجموعات بالنسبة الى التباين داخل المجموعات .

والكمية المراد تعظيمها هي :

$$q = \frac{[C'[\mu 1 - \mu 2]]^2}{C' \square C} \quad \text{----------(15)}$$

حيث ان C هي متجه المعالم للدالة المميزة الخطية والتي تقدر بحيث تجعل الدالة افضل تمييز وذلك بتعظيم النسبة (q) اما العملية التي تلي عملية التمييز فهي تصنيف المفردات الجديدة الى المجموعة الاولى اذا كانت Z >H والى المجموعة الثانية اذا كانت Z < H وعشوائيا عدا ذلك حيث ان :

$$H = \frac{\overline{Z1} + \overline{Z2}}{2} \quad \text{----------(16)}$$

و H كمية ثابتة تجعل احتمال التصنيف الخاطيء اقل ما يمكن .

احتمال خطأ التصنيف

1-الاحتمال عندما المعالم يمكن تقديرها

يستخدم هذا الاسلوب عندما تكون المفردات المدروسة تتوزع توزيعا طبيعيا

$$X_1 \sim N_p(\mu_1, \sum)$$

$$X_2 \sim N_p(\mu_2, \sum)$$

لذلك فإنه عندما المفردة (x) تعود الى المجموعة الاولى فإن :

$$\alpha'(X - \mu) \sim N_p(\alpha'(\mu_1 - \mu_2), \alpha'\Sigma\alpha) \quad \text{----------(17)}$$

حيث ان $\alpha'(X - \mu)$ هي الدالة المميزة الخطية وترمز لها ب h(x)

$h(x) \sim N_p((\underline{\alpha}'(\underline{\mu}_1 - \underline{\mu}_2), -(\alpha')\Sigma\alpha)$ ---------- (18)
$=\sum^{-1}(\underline{\mu}_1 - \underline{\mu}_2)\underline{\alpha}$

وان المفردة (x) تعود الى المجموعة الاولى في حالة

$h(x) \sim N[1/2\ \Delta^2, \Delta^2]$
$\Delta^2 = (\underline{\mu}_1 - \underline{\mu}_2)'\sum^{-1}(\underline{\mu}_1 - \underline{\mu}_2)$

حيث ان Δ^2 هو مربع المسافة ل Mahalanobis بين المجموعتين وان المفردة (x) تعود الى المجموعة الثانية في حالة

$h(x) \sim N[-1/2\ \Delta^2, \Delta^2]$

وعليه فإن احتمال خطا التصنيف P12 (احتمال تصنيف المفردة الى المجموعة الثانية وهي تعود الى المجموعة الاولى)

$P12 = p(h(x) > 0 \backslash H_2)$ ----------(19)

$\Phi \quad (-1/2)\Delta)\Phi\ [\ \frac{-E(h)}{\sqrt{V(h)}}\ ] =$

و (Φ) هي دالة التوزيع الطبيعي القياسي و P21 (احتمال تصنيف المفردة الى المجموعة الاولى وهي تعود الى المجموعة الثانية)

$\Phi\ P21 = ((-1/2)\Delta)$ ----------(20)

وتقدير احتمال التصنيف الخاطئ عندما يتم تقدير Δ والتي تقدر من مفردات العينة بالاعتماد على طريقة الإمكان الأعظم حيث ان :

$\widehat{\boldsymbol{\Delta^2}} = D^2 = (\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2)' S^{-1}(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2)$ ----------(21)

إذا سيكون تقدير احتمال التصنيف الخاطئ

$(-1/2\Delta) = P12 = P21\Phi$ ----------(22)

وهذه الطريقة تستخدم وتكون مثالية عندما حجم العينة كبير وبعكسه فإنها تكون متحيزة .

2- طريقة التعويض

أسلوب هذه الطريقة يكون بالاعتماد على مصفوفة X حيث ان(nj) عدد المشاهدات التي تعود الى المجموعة J , حيث J هو عدد المجاميع تحت البحث وان (nij) هو عدد المشاهدات في المجموعة (Hj) وصنفت على انها تعود الى المجموعة (Hi) وعليه فإن تقدير احتمال خطا التصنيف لأي مجموعة مساويا الى :

$\hat{P}\mathbf{i}\square = \frac{nij}{nj}$ ---------- (23)

حيث ان $\hat{P}\mathbf{i}\square$ هو القيمة التقديرية

وبذلك يكون إحتمال خطأ التصنيف الكلي للدالة المميزة هو :

$\hat{P} = \frac{n12 + n21}{n1 + n2}$ ---------- (24)

اما متوسط إحتمال التصنيف الخاطئ فهو :

$\hat{P} = \frac{p12 + p21}{2}$

اختبار كوهين كابا Cohen's kappa

هو مقياس للموثوقية فيما بين مقدرين كل منهما يصنف الى n من الوحدات , فهو يأخذ عامل الصدفة بنظر الاعتبار ويقوم بتصحيحه ويحسب من خلال الصيغة التالية :

$\kappa = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)},$ ---------- (25)

Pr(a): القيمة الاحتمالية النسبية المشاهدة .
Pr(e):لقيمة الاحتمالية الفرضية لتوافق الصدفة.
فذا كانت قيمة احصاءة اختبار كابا معنوية احصائيا فعلينا رفض فرضية العدم القائلة إن معامل كابا =0 وقبول الفرضية البديلة .
علما ان قيم كابا الاقل من (0.40) تشير الى تنبؤ منخفض الدقة ومن (0.40-0.75) تنبؤ متوسط الدقة واعلى من (0.75) تنبؤ عالي الدقة

بعض الاختبارات المتعلقة بالتحليل المميز

توجد عدة اختبارات إحصائية مهمه تستخدم في التحليل المميز ومن هذه الاختبارات هي :

1- اختبار معنوية دالة التمييز

H_0: $\mu 1 = \mu 2 = \cdots = \mu \square$

2- اختبار تساوي مصفوفة التباين والتباين المشترك

H_0: $\sum \square = \sum \square = \cdots = \sum k$

3- طريقة اختبار المتغيرات التمييزية

1-اختبار معنوية دالة التمييز

لغرض اختبار معنوية الدالة أي اختبار امكانية التمييز بين ا لمجاميع بإستخدام عدة متغيرات مشتركة بينها وتكوين دوال التمييز بمستوى معنوية مقبول إحصائيا فإنه تختبر الفرضية الاتية

H_0: $\mu 1 = \mu 2$

أي اختبار ان المجموعتين تعودان لنفس المجتمع أم لمجتمعات مختلفة بإفتراض تساوي مصفوفتين التباين والتباين المشترك للمجموعتين $\sum_2 , \sum_1$ والمقياس المستخدم لإختبار هذه الفرضية هو مقياس (T^2 Hotelling) والذي يساوي :

$$T^2=\frac{n1n2}{n1+n2}\ (\bar{X}_1 - \bar{X}_2)' S^{-1}(\bar{X}_1 - \bar{X}_2) \quad \text{---------- (26)}$$

$$= \quad \text{---------- (27)}$$

$$(n_1-n_2-2)S= (n_1- 1)S_1 +(n_2- 1)S_2 \quad \text{---------- (28)}$$

بدرجة حرية (P , $n1+n2-p-1$) فإذا كانت F الحسابية أكبر من F الجدولية دل على ان الدالة المميزة لها القابلية على التمييز , أما إذا كانت الحسابية أقل من F الجدولية دل ذلك أنه لا يمكن تكوين الدالة المميزة وذلك لعدم وجود فروق معنوية بيت متجهات المتوسطات للمجموعتين .

2-اختبار تساوي مصفوفة التباين والتباين المشترك لجميع المجاميع

ويستخدم هذا الاختبار للتأكد من معنوية الدالة المميزة سواء كانت باستخدام البيانات الأصلية او الرتب وذلك من خلال اختبار الفرضية الخاصة بعدم تجانس تباينات المجموعة .

H_0: $\sum_1 \square = \sum_2 \square$

H_1: $\sum_1 \square \neq \sum_2 \square$

وإحصاءه الاختبار المستخدم هي

$$M=\sum_{i=1}^{k} ni\ln|s| - \sum_{i=1}^{k} n1\ln\left|s\square\right| \square \quad \text{---------- (29)}$$

Or

$$M= \ln\frac{|s|^{n1+n2}}{|s1|^{n1}\left|s\square\right|^{n2}}$$

$$c^{-1} = 1 - \frac{2p^2 + 3p - 1}{6(p+1)(k-1)} - \left[\sum_{i=1}^{k} \frac{1}{ni} - \frac{\square}{\sum_{i=1}^{k} ni}\right]$$

وبضرب M مع الثابت $c^{-\square}$ فإننا نحصل على مقياس يقترب توزيعه من توزيع مربع كاي وبدرجة حرية $\frac{1}{2}(k-1)(p-1)$ عندما ni كبيرة .

اختبار المتغيرات التمييزية

لغرض الحصول على انسب أنموذج تمييزي بأقل خطأ تصنيف ممكن نلجأ الى استخدام اختبارات إحصائية تحدد عدد المتغيرات التمييزية المعنوية والتي تستخدم في دالة التمييز ومن هذه الاختبارات الإحصائية التي تستخدم لتحديد المتغيرات التمييزية المعنوية لمجموعتين تستخدم طريقة Forward stepwise Selection Method وتعتمد هذه الطريقة على إيجاد مقياس مهالنوبس والذي يعتمد على حساب المسافة بين المتوسطات لمتغيرات المجموعتين ويرمز له D_p^2 ويعرف كالآتي :

$$D_p^2 = (\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2)' S^{-1} (\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2) \quad \text{---------- (30)}$$

ويستخدم هذا المقياس لاختبار مدى اعتماد الدالة على بعض المتغيرات والتي تتكون منها للقيام بعملية التمييز .

ولنفرض أن´ $X_k = (X_{j1}, X_{j2}, \ldots, X_{jn-1})$ حيث أن $Z \le P$ من المتغيرات فان إبعاد مهالنوبس الى Xk تكون كالآتي:

$$D_k^2 = (\bar{\mathbf{X}}_{1k} - \bar{\mathbf{X}}_{2k})' S^{-1}_{xk} (\bar{\mathbf{X}}_{1k} - \bar{\mathbf{X}}_{2k}) \quad \text{---------- (31)}$$

ان طريقة Stepwise تعتمد على اعادة ترتيب X_p الى متجه جديد ´$(X_{j1}, X_{j2}, \ldots, X_{jn-1})$ بحيث اذا مددنا المتجه $(X_{j1}, X_{j2}, \ldots, X_{jn-1}) = X_{k-1}$ مع X_{jk} نحصل على المتجه´ $X_k = (X_{j1}, X_{j2}, \ldots, X_{jn-1})$ بحيث يحقق أكبر D_k^2 عند تحديد X_{k-1} ولجميع قيم K حيث K= 1, 2,…p وتستمر الى ان نحصل على أفضل دالة K من المتغيرات باستخدام احصاءة الاختبار :

$$F = \frac{(n1 + n2 - p - 1)}{((p-k))} \; \frac{(n1\, n2(dp^{\intercal}2 - dk^{\intercal}2)}{((n1 - n2)(\square 1 + n2 - 2) + n1\, n2\, dk^{\intercal}(2\ ))} \quad \text{---------- (32)}$$

فإذا كانت $F_k < F_{1-\alpha}(p-k, n_1-n_2-p-1)$ دل ذلك على ان المتغيرات P-K ليست تمييزية إذ أننا نحصل على بعض قوة التمييز باستخدام K بدلا من P من المتغيرات واذا كانت F المحسوبة أكبر من الجدولية دل ذلك على ان المتغيرات المتبقية (p-k) متغيرات تمييزية [1].

الجانب العملي:

تضمن الجانب العملي تحليل إحصائي باستخدام spss إذ سحبت عينه من (80) طفل {(40) طفل متوقي,(40) باقي على قيد الحياة} تم أيجاد الدالة للتمييز من خلال احتساب قيم المتغيرات لبيانات حقيقية وذلك للتحقق من أدائها في الواقع العملي . وبهدف دراسة البيانات والعلاقة بين المتغيرات اعتمدت بيانات من سجلات دائرة صحة كربلاء لمرضى الأطفال الخدج (حديثي الولادة) جمعت من مستشفيات المحافظة وكانت البيانات التي تمثل وفاة او عدم وفاة الطفل الخديج هو متغير الاستجابة [والجنس (44 ذكور ، 36 أناث) والمستشفى (36 طفل في م كربلاء ، 44 طفل في م نسائية) ونوع الولادة (60 ولادة طبيعية ،20 ولادة قيصرية ،) وسبب الرقود(نقص الوزن ، قصر مدة الحمل ، ذات الرئة الولادي ، متلازمة شفط السوائل الرحمية ، يرقان ولادي ، تشوهات خلقية ، متلازمة ضيق التنفس (عسر التنفس) ، اختناق ولادي ، تسمم الدم الجرثومي ، اضطرابات في دم الجنين ، إصابة أثناء الولادة ، أخرى) وعمر الأم ، وزن الطفل ، مدة حمل الطفل] كمتغيرات مستقلة تؤثر في حياة الطفل

بعض الاختبارات المهمة :

قبل البدء بتطبيق التحليل لابد من إجراء بعض الاختبارات للتحقق من شروط تطبيق التحليل التمييزي وكالاتي :

أولا: البيانات تتوزع توزيعا طبيعيا لكل مجتمع

في هذا الشرط يتم اختبار البيانات لمعرفتها ان كانت تتوزع طبيعي أم لا مع ملاحظة حجم العينة فاذا كان حجم العينه كبيرا فهذا يؤدي الى افتراض اقترابها من التوزيع الطبيعي لذلك يمكننا القول ان البيانات تقترب من التوزيع الطبيعي أجمالا.

ثانيا :التأكد من عدم وجود ارتباط متعدد بين المتغيرات المستقلة

لغرض اختبار وجود مشكلة الارتباط المتعدد بين المتغيرات المستقلة يتم حساب معامل Tolerance كل من المتغيرات المستقلة حيث ان :

$$\text{Tolerance} = 1 - R^2_{xi\,,\,others}$$

$R^2_{xi\,,\,others}$: يمثل مربع معامل الارتباط المتعدد بين المتغير المستقل وبقية المتغيرات المستقلة .

ثم يستخرج معامل VIF كل متغير مستقل والذي يعني (Variance Inflation Factor) حيث ان:

VIF=1\ Tolerance

من ملاحظة النتائج كانت جميع قيم المعامل VIF اقل من (5) هذا يشير الى ان الأنموذج لا يعاني من مشكلة التعدد الخطي . وكما في الجدول

جدول رقم (1)

Collineaty Statistics		
	Tolerance	VIF
X1	.432	2.315
X2	.937	1.067
X3	.360	2.779
X4	.389	2.570
X5	.708	1.412
X6	.946	1.058
x7	.505	1.982

ثالثا: اختبار شرط تجانس التباين

لمعرفة مدى تجانس أفراد المجموعتين وحسب الفرضية الأتية :

$H_0: \sum_1 \square = \sum_2 \square$

$H_1: \sum_1 \square \neq \sum_2 \square$

ومن خلال تطبيق الصيغة (29) المذكورة في الجانب النظري نلاحظ النتائج التالية:

جدول رقم (2)

Log Determinants

Y	Rank	Log Determinant
.00	2	4.118
1.00	2	4.319
Pooled within-groups	2	4.231

جدول رقم (3)

Test Results

Box's M		.987
F	Approx.	.320
	df1	3
	df2	1095120.000
	Sig.	.811

يمكن الاستعانة باختبار BOX s' M للتجانس حيث أشارت النتائج في الجدول الاول الى ان قيم log Determinant تقريبا متساوية للمجموعتين مما يدل الى تجانس افراد المجموعتين ومن ملاحظة المعنوية في الجدول الثاني

(BOX s'M) نجد ان مستوى الدلالة المستخرج هو 0.811 والذي هو أكبرمن مستوى المعنوية مما يوجب قبول فرضية العدم القائلة تجانس بين مصفوفات التباينات المشتركة (تجانس افراد المجموعتين) .

أجراء التحليل التمييزي

لغرض احتساب الدالة المميزة نعتمد نتائج طريقة stepwise وكما في الجداول ادناه:

جدول رقم (4)

Group Statistics

Y		Mean	Std. Deviation	Valid N (listwise) Unweighted	Weighted
.00	X1	.4750	.5057	40	40.000
	X2	1.4000	.4961	40	40.000
	X3	34.4250	3.3504	40	40.000
	X4	2.0363	.5476	40	40.000
	X5	1.3250	.4743	40	40.000
	X6	6.3250	2.4744	40	40.000
	x7	33.6500	8.5681	40	40.000
1.00	X1	.4250	.5006	40	40.000
	X2	1.5000	.5064	40	40.000
	X3	31.0500	3.4785	40	40.000
	X4	1.4925	.6694	40	40.000
	X5	1.1750	.3848	40	40.000
	X6	8.0750	2.5155	40	40.000
	x7	32.2000	8.2094	40	40.000
Total	X1	.4500	.5006	80	80.000
	X2	1.4500	.5006	80	80.000
	X3	32.7375	3.7945	80	80.000
	X4	1.7644	.6664	80	80.000
	X5	1.2500	.4357	80	80.000
	X6	7.2000	2.6309	80	80.000
	x7	32.9250	8.3693	80	80.000.

يشير الجدول الى الاوساط الحسابية والانحرافات المعيارية للمتغيرات المستقلة المنبئة وذلك لكل مجموعة من المجموعتين ولإجمالي المجموعتين علما ان (0) يشير الى الباقين على قيد الحياة و(1) يشير الى حالات الوفاة و[x1:المستشفى (كربلاء ، نسائية).x2 :الجنس (ذكر ، أنثى) .x3 : مدة حمل الطفل . x4: وزن الطفل ، x5: نوع الولادة (طبيعية ، قيصرية ، اخرى)x6 :سبب الرقود(نقص الوزن ، قصر مدة الحمل ، ذات الرئة الولادي ، متلازمة شفط السوائل الرحمية ، يرقان ولادي ، تشوهات خلقية ، متلازمة ضيق التنفس (عسر التنفس) ، اختناق ولادي ، تسمم الدم الجرثومي ، اضطرابات في دم الجنين ، إصابة أثناء الولادة ، أخرى) x7:عمر الأم]

جدول رقم (5)

Pooled Within-Groups Matrices

		X1	X2	X3	X4	X5	X6	X7
Correlation	X1	1.000	-.056	.150	.052	-.304	-.066	-.682
	X2	-.056	1.000	.151	.072	.136	-.048	-.084
	X3	.150	.151	1.000	.716	.305	.229	-.075
	X4	.052	.072	.716	1.000	.327	.114	-.070
	X5	-.304	.136	.305	.327	1.000	-.057	.128
	X6	-.066	-.048	.229	.114	-.057	1.000	.075
	X7	-.682	-.084	-.075	-.070	.128	.075	1.000

يشير الجدول الى معاملات الارتباط الثنائي بين المتغيرات المستقلة السبعة فمثلا معامل الارتباط بين x3 مدة الحمل وx4 وزن الطفل بلغ 0.716 وبين x5نوع الولادة و x4 وزن الطفل كان 0.327 وهكذا .

جدول رقم (6)

Variables Entered/Removed

Min. F

Step	Entered	Statistic	df1	df2	Sig.	Between Groups
1	X3	19.534	1	78.000	0.00	0.00 and 1.00
2	X6	18.613	2	77.000	0.00	0.00 and 1.00

يشير الجدول الى الحد الأدنى لقيمة F المحسوبة وفق الصيغة (26) في الجانب النظري , ويبين الخطوتين التي تم في كل منها ادخال المتغير الذي يضاعف نسبة Fالصغرى بين أزواج المجموعتين إذ نطبق القاعدة الأساسية القائلة بأن الحد الأدنى لقيمة F الجزئية لإدخال أي متغير في التحليل يجب ان لا تقل عن 3.84 وان الحد الأعلى لقيمة F الجزئية لإخراج أي متغير من التحليل هو 2.71

جدول رقم (7)

Variables in the Analysis

Step		Tolerance	F to Remove	Min. F	Between Groups
1	X3	1.000	19.534		
2	X3	.947	24.431	9.839	.00 and 1.00
	X6	.947	14.349	19.534	.00 and 1.00

يشير الجدول الى المتغيرات الداخلة في التحليل وبالذات الى الخطوتين التي اتبعت لتحديد المتغيرات الداخلة في التحليل حيث بدأت الخطوة الأولى باستخراج قيمة F to Remove لمتغير واحد وفي الخطوة الثانية كانت لمتغيرين المفترض إدخالهم في التحليل وكانت قيمة F لكل منهما أكبر من 3.84 وفقا للقاعدة .

جدول رقم (8)

Variables Not in the Analysis

Step		Tolerance	Min. Tolerance	F to Enter	Min. F	Bet. Groups
0	X1	1.000	1.000	.197	.197	.00 and 1.00
	X2	1.000	1.000	.796	.796	.00 and 1.00
	X3	1.000	1.000	19.534	19.534	.00 and 1.00
	X4	1.000	1.000	15.811	15.811	.00 and 1.00
	X5	1.000	1.000	2.412	2.412	.00 and 1.00
	X6	1.000	1.000	9.839	9.839	.00 and 1.00
	X9	1.000	1.000	.597	.597	.00 and 1.00

1	X1	.978	.978	.038	9.666	.00 and 1.00
	X2	.977	.977	1.968	10.872	.00 and 1.00
	X4	.488	.488	1.070	10.310	.00 and 1.00
	X5	.907	.90	.036	9.664	.00 and 1.00
	X6	.947	.947	14.349	18.613	.00 and 1.00
	X9	.994	.994	.968	10.247	00 and 1.00
2	X1	.967	.920	.293	12.392	00 and 1.00
	X2	.970	. 921	2.506	13.487	00 and 1.00
	X4	.485	.466	.476	12.483	00 and 1.00
	X5	.890	.846	.091	12.292	.00 and 1.00
	X9	.985	.939	1.512	12.995	00 and 1.00

يشير الجدول أعلاه الى المتغيرات المحذوفة من التحليل والى الخطوات الثلاث التي اتبعت لتحديد المتغيرات الخارجة من التحليل إذ بدأت الخطوة ما قبل الأولى باستخراج قيمة F to Remove المذكورة في الصيغة (28) للمتغيرات السبعة وانتهت الخطوة الأخيرة باستخراج قيمة F to Remove للمتغيرات الخمسة المفترض إخراجها من التحليل كون قيمة F لهذه المتغيرات اقل من 2.71 وفقا للقاعدة.

جدول رقم (9)

Wilks' Lambda

Step	Number of Var.	Lambda	df1	df2	df3	Exact F Statistic	df1	df2	Sig.
1	1	.800	1	1	78	19.534	1	78.00	0.00
2	2	.674	2	1	78	18.613	2	77.00	0.00

تحتسب قيمة Wilks' Lambda في كل خطوة من الخطوتين إذ تم ادخال متغير واحد إضافي في كل خطوة ففي الخطوة الأولى بلغت قيمة Lambda للمتغير الأول في التحليل 0.8 بينما في الخطوة الثانية بلغت للمتغيرين الأول والثاني الداخلين في التحليل 0.674 أي نلاحظ ان قيمتها انخفضت فهي تقل كلما اضفنا متغير مؤثرا في التحليل وهذا يدل على وجود فروق بين المجموعتين فقد كانت قيمة F في كلا الخطوتين أكبر من قيمتها الجدولية ومما يؤكد
ذلك ان مستوى الدلالة الإحصائية في كل خطوة منها كان صفرا

جدول رقم (10)

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	.483	100.0	100.0	.571

الجدول رقم (10) يبين الارتباط القانوني وحجم التباينات المفسرة من قبل المتغيرات المستقلة في المتغير المعتمد ويفيد في تقيم اهمية الدوال عندما تكون أكثر من واحدة فترتب حسب اهميتها تنازليا علما ان عدد الدوال يحدد من عدد الطبقات في المتغير المعتمد -1 والعمود الثاني يمثل نسبة التباين المفسر بواسطة الدالة اما العمود الثالث يمثل نسبة التباين المتراكم المفسر والعمود الاخير يمثل نسبة التشتت التي ميزت بواسطة المتغيرات المستعملة وبعنى اخر ان 33% من التغير في عضوية المجموعة يعود الى التغير في المتغيرات المنبئه,وبهدف بيان اهمية الدالة المميزة في التمييز بين المجموعتين استخدمنا احصاءة 'Wilks Lambda التي لها توزيع مربم كاي وتحسب بقسمة مجموع مربعات التباينات داخل المجموعات على مجموع مربعات التباينات الكلي

جدول رقم (11)

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	.674	30.367	2	.000

من الجدول (11) نلاحظ معنوية تباين الجذور المميزة للدالة اي ان هناك فروق معنوي بين المجموعتين لان قيمة مربع كاي البالغة 30.367 اكبر من قيمتها الجدولية .

جدول رقم (12)

Standardized Canonical Discriminant Function Coefficients

	Function
	1
X3	.883
X6	-.713

الجدول يبين الارتباط التجميعي بين الدالة التمييزية وكل متغير من المتغيرين المستقلين التي تم ادخالها في التحليل التمييزي بوحدات القياس المعيارية اي يحدد الأهمية النسبية للمتغيرات المستقلة في تقدير المتغير المعتمد ونلاحظ ان مدة حمل الطفل لها الوزن الاكبر المؤثر في زيادة قوة التمييز بين المجموعتين وهو 0.883 يليه متغير سبب الرقود وعليه يتم تسمية هذه الدالة باسم المتغير المؤثر زيادة قوة التمييز بين المجموعتين وهي مدة الحمل

جدول رقم (13)

Structure Matrix

	Function
	1
X3	.720
X4[a]	.551
X6	-.511
X5[a]	.310
X1[a]	.179
X2[a]	.168
X9[a]	-.119

يبين جدول رقم (13) الارتباط داخل المجموعات بين كل متغير من المتغيرات المنبئه الداخلة في التحليل وقيمة الدالة التمييزية وقد كان معامل الارتباط مع مدة الحمل اقواها اذ بلغ 0.720 ثم الارتباط مع سبب الرقود الذي بلغ 0.511- اما بقية المتغيرات فقد تم استبعادها من التحليل ووضع رمز مشير اليها . وتعني الاشارة السالبة وجود علاقة عكسية بين المتغير المعتمد وفاة وبقاء الطفل وبين المتغير المستقل سبب الرقود

جدول رقم (14)

Canonical Discriminant Function Coefficients

	Function 1
X3	.259
X6	-.286
(Constant)	-6.408

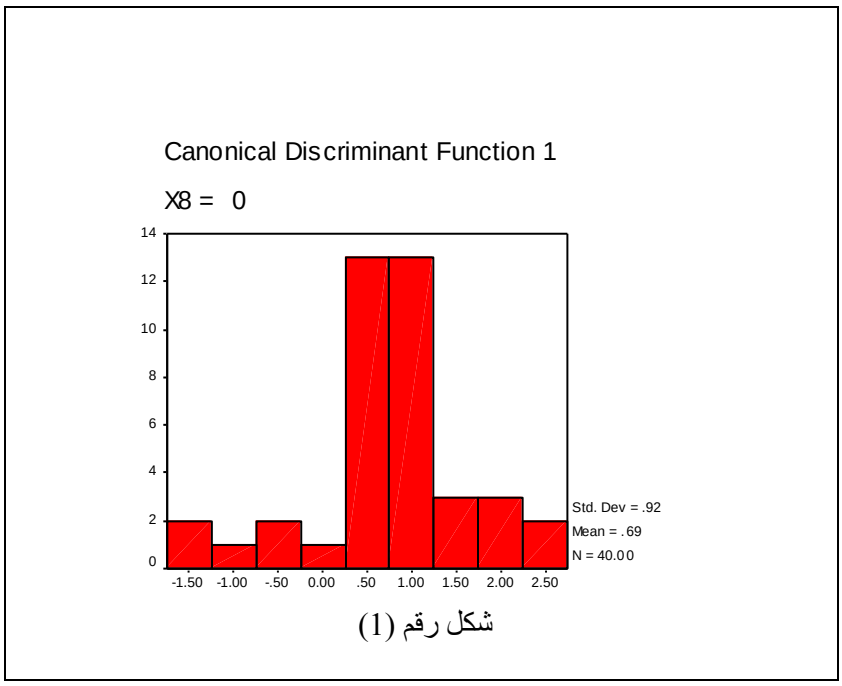
يوضح الجدول اعلاه المعاملات التمييزية غير المعيارية للارتباط بين كل متغير من المتغيرات المنبئه الداخلة في التحليل وبين الدالة التمييزية وتحسب الدرجة التمييزية من خلال ضرب المعاملات التمييزية غير المعيارية في قيم المتغيرات المدخلة ثم جمع الناتج واضافته الى القيمة الثابتة 6.408 -

جدول رقم (15)

Functions at Group Centroids

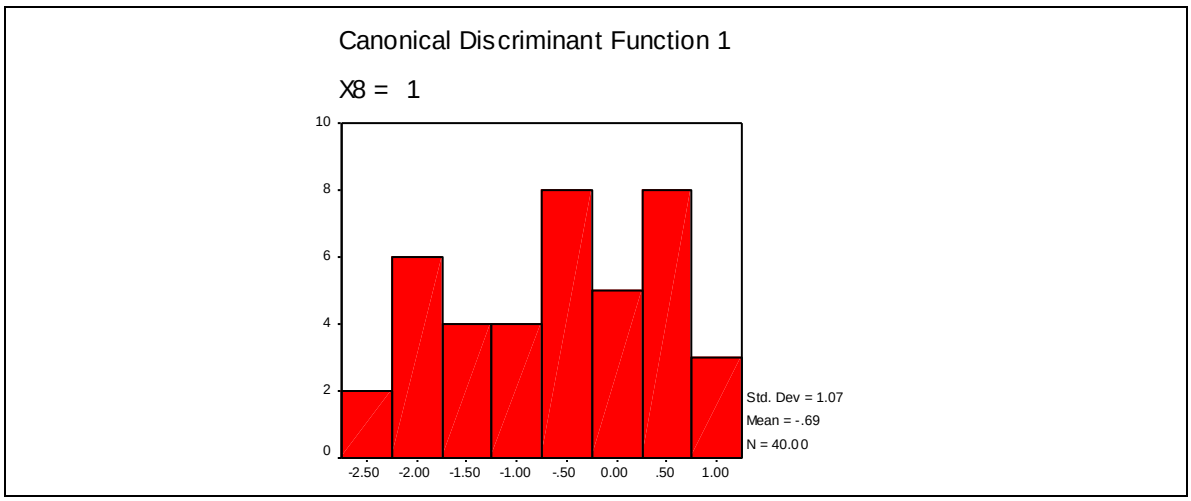
X8	Function 1
.00	.687
1.00	-.687

يبين الجدول رقم (15) متوسطين الاول يمثل المجموعة الاولى الباقين على قيد الحياة وقيمته (0.687) والثاني يمثل متوسط المجموعة الثانية للمتوفين ومن الجدير بالذكر ان متوسطي المجموعتين يكونا نفس القيمة اذا كان عدد افراد المجموعتين متساويا



شكل رقم (1)

يظهر الشكل ان الوسط الحسابي للدالة التمييزية للباقين على قيد الحياة قد بلغ 0.69 وبانحراف معياري 0.92



شكل رقم (2)

يظهر الشكل ان الوسط الحسابي للدالة التمييزية للمتوفين قد بلغ 0.69- وبانحراف معياري 1.07 أي أكبر من المجموعة السابقة

جدول رقم (15)

Classification Results

		X8	Predicted Group Membership .00	1.00	Total
Original	Count	.00	35	5	40
		1.00	12	28	40
	%	.00	87.5	12.5	100.0
		1.00	30.0	70.0	100.0

a78.8% of original grouped cases correctly classified.

يشير الجدول الى مدى دقة النتائج النهائية للتصنيف اذ يتبين ان (35) حالة من المجموعة الأولى وبنسبة 87.5 % قد تم تصنيفها بشكل صحيح وبنائا عليه فان باقي حالات المجموعة الأولى والبالغ عددها (5) حالات وبنسبة 12.5% قد تم تصنيفها بشكل خاطئ . وفي نفس الوقت يتبين ان (28) حالة في المجموعة الثانية وبنسبة 70.0% قد تم تصنيفها بشكل صحيح وبنائا عليه فان باقي حالات المجموعة الثانية والبالغة (12) حالة وبنسبة 30.0% قد تم تصنيفها بشكل خاطئ وكنتيجة عامة فقد دلت النتائج على ان ما نسبته 78.8% من الحالات في كلا المجموعتين قد تم تصنيفها بشكل صحيح وهذا يدل على جودة نتائج التصنيف.

الا انه على الرغم من ان الحالات المصنفة تصنيفا صحيحا كانت (63) حالة من الحالات في العينة البالغة (80) اي ما نسبته 78.8% الا ان النسبة المذكورة قد تتاثر بما يسمى عامل الصدفة ولحل هذه المشكلة استخدم اختيار كابا (kappa) وفق الصيغة (25) والذي يعد مقياسا جيدا للموثوقية لانه ياخذ عامل الصدفة بعين الاعتبار وتبين معنوية هذا الاختبار عند التحليل .

ومن ملاحظة الجدول أعلاه نجد ان احتمال التنبؤ بالوفاة علما ان الطفل متوفي اصلا يكون مساوي الى 74.47 وان احتمال البقاء على قيد الحياة علما ان الطفل قد عاش اصلا هو 84.85.

اما تقدير احتمال خطا التصنيف فهو من خلال تطبيق الصيغة (23) المذكورة في الجانب النظري.

$$\widehat{Pij} = \frac{nij}{nj}$$

P_{12} فقد كان مساويا الى 0.125

P_{21} فقد كان مساويا الى 0.3

تقدير احتمال خطا التصنيف الكلي مساويا الى 0.21

الاستنتاجات:

1- لوحظ ان المتغيرين الأفضلين في التمييز كانا مدة الحمل وسبب الرقود فقد كان معامل الارتباط بين الدالة المميزة ومدة الحمل 0.72 ومعامل الارتباط الدالة المميزة وسبب الرقود كان 0.511-
2- تبين ان اعلى معامل ارتباط ثنائي كان بين مدة الحمل ووزن الطفل وهو 0.716 يليه معامل الارتباط الثنائي بين نوع الولادة ووزن الطفل وهو0.327
3- وجود فروق ذات دلالة معنوية بين مجموعتي الوفيات والباقين على قيد الحياة بالنسبة للمتغيرات المستقلة فقد كان هذا واضحا من ملاحظة معنوية تباين الجذور المميزة للدالة المميزة
4- تبين ان المتغير المستقل مدة الحمل كان له الوزن الأكبر المؤثر في زيادة قوة التمييز بين المجموعتين لذا تم تسمية الدالة بأسمه وهي (دالة مدة الحمل)
5- لقد كان تقسيم الحالات بين فئات المتغير التابع بناء على قيم المتغير المستقل (35) حالة من المجموعة الأولى وبنسبة 87.5 % قد تم تصنيفها بشكل صحيح وبنائا عليه فان باقي حالات المجموعة الأولى والبالغ عددها (5) حالات وبنسبة 12.5% قد تم تصنيفها بشكل خاطئ . وفي نفس الوقت يتبين ان (28) حالة في المجموعة الثانية وبنسبة 70.0% قد تم تصنيفها بشكل صحيح وبنائا عليه فان باقي حالات المجموعة الثانية والبالغة (12) حالة وبنسبة 30.0% قد تم تصنيفها بشكل خاطئ وكنتيجة عامة فقد دلت النتائج على ان ما نسبته 78.8% من الحالات في كلا المجموعتين قد تم تصنيفها بشكل صحيح وهذا يدل على جودة نتائج التصنيف.اذ ان الحالات المصنفة تصنيفا صحيحا كانت (63) حالة من عدد الحالات الكلية وهي (80) اي ما نسبته 78.8%.

المصادر:

1-البكري، رباب عبد الرضا صالح (1997) :مقارنة بعض طرق التقدير للدالة المميزة لمجموعتين مع تطبيق عملي ، رسالة ماجستير في الاحصاء كلية الادارة والاقتصاد جامعة بغداد.

2-Broffit J.andRandles R.and Hogg R.(1976):Distribution-Free pratial Discriminant Analysis .JASA.Vol.17, No:356pp.934-939

3-Donald F.Morrison,(1976) Multivariate statistical Method second Edition,Tokyo,London.Mexico,New Delhi.

4-Gunandesikan . R,(1977) Method for statistical data Analysis for Multivariate Observation,jhon wiley and son New York, London.

5-Knke,JD,(1982): Discriminant Analysis with discret and continuous variable Biometrics Vol.38.

6-Mahalanbios p.c.(1949):SANKHYA

7-Mike James.(1985): Classification Algorthem John Wiley and Son.New york chichester Brisbane Torontes.

8-Rao C.Radhakrishan(1952)Advance statistical Method in Biometric Research,Jhon Wiley New york London.

9-Von Mises; Discriminant and Classification using Both Binaryand continuous variable , JASA.Vol. 70, No:352