معالجة القصور في معدلات التصنيف المقدرة بدالة fisher

دراسة وتطبيق

د.عبد الحكيم المنصوب*

المستخلص

تستخدم دالة fisher الخطية للتميز في الفصل بين المجتمعات الاحصائية المتداخلة . وغالبا ما يتم استخدام هذه الدالة في صياغة نموذج لتصنيف المفردات حسب مجتمعاتها . الا ان هذا النموذج قد يفرز معدلات مدنية للتصنيف الصحيح بالرغم من المعنوية العالية لدالة التميز . وهذه المشكلة ناقشها الباحث في دراسة سابقة له ، حيث تم شرح واستخدام اهم الطرق الخاصة بمعالجة قصور معدلات التصنيف الناتجة عن النموذج Anderson . وفي هذه الدراسة يقدم الباحث اسلوبا اخرا لمعالجة هذا القصور ، ولكن في المعدلات الخاصة بدالة fisher التي تعد الاساس لنماذج التصنيف .

Abstract :

fisher"s linear discriminant function is used to separate interbetween statistical populations . often , it is used for formulating a model which used to classify the observations on their populations . sometimes , the model leads to low classification rates , despite of the high significant discriminant function . this broblem was dicussed in a previous study . some important procedures of defects treatment in classification rates of anderson's model were discussed . in this study the researcher present a new process based on the classification rates estimated by fisher's function , the base of the classification model .

* استاذ مساعد/جامعة ال /كلية التجارة والعلوم الادارية

مقدمة :

اذا كان المتغير العشوائي متغيرا اسميا nominal ثنائي أو متعدد الصفات (التقسيمات) بحيث يعبر عن كل صفة برقم معين (....0,1,2,) واذا كان هذا المتغير يعتمد خطيا على مجموعة من المتغيرات المستقلة independent ، ويراد التنبؤ باحدى تقسيماته بمعلومية المتغيرات المستقلة ، فان تحليل التميز Discriminant Analysis هو الاسلوب الاحصائي المناسب اذا توفرت افتراضيات Assumptions هذا التحليل . { press & wilson , 1978 } ،ويتم التعبير عن القوة التنبؤية للنموذج المتواصل الية بمعدلات التصنيف Classification rates الصحيحة ، التي تشير الى نسبة المفردات المصنفة تصنيفا صحيحا في كل تقسيم من تقسيمات المتغير العشوائي الاسمي (التابع) .

فالمعدل p (k / g) يشير الى نسبة المفردات التي صنفت – باستخدام النموذج – في التقسيم k وهي تنتمي اصلا الى التقسيم g . ولكننا قد نحصل على دالة تمييز ذات معنوية مرتفعة ، في حين ان استخدامها ، في تكوين نموذج التصنيف ، يفرز قيما متدنية جدا لمعدلات التصنيف الصحيح ، الامر الذي يدفع الباحث الى فحص توفر افتراضات تحليل التمييز اولا ، ثم مناقشة حساباته واشتراطاتها ، وصولا الى افضل الطرق التي تخفف من حدة هذا التناقض . وذلك بالتطبيق على بيانات تنظيم الاسرة كما وردت في المسحين اليمنيين حول صحة الام والطفل 1991 و 1997 ، وبغرض المقارنة مع نتائج الباحث التي توصل اليها في دراسة سابقة له (المنصوب ، 2004) .

وتتمثل اهمية استخدام دالة الفصل بين المجتمعات الاحصائية ، المتداخلة ، في ان لا معنى لها اذا لم تستخدم في تحديد المجتمع الذي تنتمي الية مفردات جديدة ، أو اذا لم تستخدم على الاقل في اعادة توزيع مفردات الدراسة على المجتمعات التي تنتمي اليها ، لان التمييز والتصنيف معا يساعد على معرفة القوة التمييزية للمتغيرات المستقلة المستخدمة ، وعلى معرفة افضل الطرق التي تؤدي الى تخفيض اخطاء التصنيف ، والى تخفيض تكلفة هذه الاخطاء (اذا تم مراعاتها) . ويتفق على هذا الرأي العديد من علماء الاحصاء مثل Agresti (1996) و jackson (1983) و kleinbum (1998) .

وعلى الرغم من امكانية الحصول على دالة تمييز Discriminant function معنوية، يتم عند تقديرها مراعاة احجام مجموعات المتغير التابع ، الا ان معدلات التصنيف ، باعتبارها اداة لتقيم دالة التمييز ، قد تكون اقل من 60 % كحد ادنى مقبول لها { randles et al , 1978} وهذا ما واجهه الباحث في دراسته السابقة (المنصوب ، 2004) عندما تم تطبيق نموذج

Anderson على بيانات تنظيم الاسرة اليمنية ، فقسمت المعالجات الاحصائية الخاصة بمواجهة هذا التناقض الى نوعين من المعالجات :

النوع الاول :

يتمثل في الحلول والبدائل اللازمة لمواجهة انتهاك واحد أو اكثر من افتراضيات تحليل التمييز .

النوع الثاني :

يتمثل في توفير اشتراطات حسابات مكونات النموذج . وهي معالجة اشمل لانها لاتتعلق بنموذج Anderson فقط ولكنها تتعلق بدالة fisher الخطية للتمييز ، التي تستخدم في بناء نموذج Anderson اصلا. حيث يتضمن الجزء الاول من هذه الدراسة اشارة سريعة الى اهم الافتراضات الخاصة بدالة fisher الخطية للتميزونموذج Anderson المشتق منها ويتضمن الجزء الثاني عرضا مختصرا لمعالجة القصور في معدلات التصنيف في دراسة السابقة للباحث . اما الجزء الثالث فيتضمن المعالجة المقترحة في هذه الدراسة ، ومقارنة نتائجها بنتائج الدراسة السابقة للباحث .

1.اهم افتراضات تحليل التمييز :

اذا كان لدينا عدد قدرة (p) من المتغيرات المستقلة (المميزة) تنتمي الى المجتمعين أو المجموعتين $\prod 1, \prod 2$ ويراد الفصل بينها طبقا للخاصية (أو المتغير) Y ، فان اهم الافتراضات والاجراءات اللازمة لذلك تتمثل في الاتي:

الافتراض الاول :

ان مجتمعات الدراسة المتداخلة والقابلة للتحديد ، هي مجتمعات طبيعية NORMAL وكل مجتمع له توزيع معتدل مختلف عن الاخر . فاذا كانت مجموعة المتغيرات المستقلة أو المميزة X's تتمثل بالمتجهة

$X' = \{X1, X2, X3, \ldots\ldots, XP\}$ فان :

$$X \approx Np[\mu, \Sigma]$$

$$\mu = \begin{bmatrix} \mu 1 \\ \mu 2 \\ . \\ . \\ \mu P \end{bmatrix} \quad , \quad \sum = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ \sigma & \sigma_2^2 & \ldots & 0 \\ \vdots & \vdots & \vdots & 0 \\ 0 & 0 & \cdots & \sigma_p^2 \end{bmatrix}$$

الافتراض الثاني :

ان مجتمعات الدراسة لها مجتمعات اوساط مختلفة الا انها نفس مصفوفة التباين – التغـاير وتحت هذا الافتراض يمكن تحويل المشاهدة متعددة المتغيرات X الى مشاهدة وحيدة المتغيـر Y باستخدام دالة فيشر الخطية { JOHNSON & WWICHERN , 1992}

$$Y=(\mu_1 - \mu_2)'\Sigma^{-1}X \qquad (1)$$

حيث : μ_k هو متجه متوسطات متغيرات المستقلة (X'S) في المجتمع K .

وتكون قاعدة التصنيف :

$$\text{Put x in if y} \prod 1 \geq (\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 + \mu_2) \qquad (2)$$

PUT X IN Π_2 OTHERWISE

حيث : Y معرفة في (1)

ولان معالم المجتمعات غالبا ما تكون مجهولة ، يمكن استخدام احصاءات العينة لنحصل على :

$$Y = ((\overline{X}_1 - \overline{X}_2)'S_{pooled}X \qquad (3)$$

$$S_{pooled} = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{n_1 + n_2 - 2}$$

حيث :

$\overline{X}_k$: متجهة متوسطات المتغيرات المستقلة X'S في العينة المسحوبة من المجتمع K .

N_K : حجم العينة المسحوبة من المجتمع K .

S_K : مصفوفة التباين – التغاير في العينة المسحوبة من المجتمع K .

ومن ثم تكون قاعدة التصنيف :

$$\text{PUT X IN } \Pi_1 \text{ if } Y \geq \frac{1}{2}(\overline{X}_1 - \overline{X}_2)' S^{-1}_{pooled}(\overline{X}_1 + \overline{X}_2) \qquad (4)$$

PUT X in Π_2 OTHERWISE

حيث Y معرفة في (3)

ولان مجموعات الدراسة (مجتمعات أو عينات) غالبا ما تكون مختلفة الاحجام ، فقد اقتـرح ANDERSON (1972) قاعدة ادنى تكلفة متوقعة للتصنيف الخاطى باستخدام بيانات العينـة ، وذلك بمراعاة كل من الاحتمالات القبلية للعينات ، وتكلفة التصنيف الخاطى . فاذا كـان الاحتمـال القبلي لظهور العينة المسحوبة من المجتمع k هو P_K والذي يقدر من العلاقة :

$$P_K = \frac{n_k}{n} \qquad (5)$$

حيث N هو حجم العينتين معا .

واذا كان (C (K/G هو تكلفة وضع مشاهدة في المجتمع K في حين انها تنتمي الى المجتمـع G اصلا ، فان قاعدة التصنيف تصبح :

$$\text{PUT X IN } \Pi_1 \text{ if } Y \geq \frac{1}{2}(\overline{X}_1 - \overline{X}_2)' S^{-1}_{pooled}(\overline{X}_1 + \overline{X}_2) \geq Lin\frac{c(1/2)P_1}{c(2/1)P_2} \qquad (6)$$

Put X in Π_2 OTHERWISE

حيث Yمعرفة في (3) .

واذا كانت التكلفة غير معلومة ، أو غير ضرورية ، أو متساوية ، فيمكن اهمالها ليصبح

الطرف الايمـن (فـي العلاقـة الاخيـرة) هـو

$$lin \frac{p_1}{p_2}$$

الافتراض الثالث :

عدم وجود ارتباط بين المتغيرات المميزة . فكلما زادت قوة الازدواج الخطي multicollinearity كلما زادت صعوبة تفسير نتائج تحليل التمييز ، بما في ذلك صعوبة تحديد المساهمة النسبية لكل متغير على حدة في القوة الكلية للتميز { Lacnenbruch , 1975} .

2. معالجة قصور معدلات التصنيف في دراسة سابقة للباحث :

في دراسة سابقة للباحث (المنصوب ، 2004) تم فيها استخدام بيانات المسح اليمني حول صحة الام والطفل 1991 ونظيرة الخاص بعام 1997 ، وذلك لمعرفة العوامل المؤثرة على موقف السيدة اليمنية من استخدام وسائل تنظيم الاسرة . اذا مثل Y الخاصية أو المتغير الذي بموجبه تم التمييز بين مجموعة السيدات المستخدمات للوسائل والسيدات غير المستخدمات .

حيث :

Y = [1 FOR NOT USER
[2 FOR USER

اما المتغيرات المميزة فقد وصل عددها الى 26 متغيرا ، وهي المتغيرات الواردة في جدول رقم (1) . ورغم الحصول على دالة تمييز معنوية حسب الاختبار :

$$-\left[\frac{n-1-(p+m)}{2}\right] Lin\Lambda \approx \chi^2_{p(m-1)} \qquad (7)$$

$$\Lambda = \frac{|W|}{|B+W|}$$

حيث :

M : عدد مجموعات الدراسة .

W : مصفوفة التباين داخل WITHIN المجوعات .

B: مصفوفة التباين بين between المجموعات .

ورغم مراعاة حجمي المجموعتين (4168 سيدة غير مستخدمة مقابل 607 سيدة مستخدمة في مسح 1991 ، زدن الى 6747 و 1838 سيدة على التوالي في مسح 1997) الا ان اعادة توزيع السيدات (في كل مسح) باستخدام نموذج ANDERSON (علاقة رقم 6 بعد اهمال عنصر التكلفة) اسفر عن معدلات تصنيف متناقضة للغاية ، حتى مع اعادة تقديرها بطريقة Lachenbruch (1975) التي تناسب العينات كبيرة الحجم . فلم يتجاوز معدل التصنيف

الصحيح للسيدات المستخدمات (2/2) p ال 0.46 في نموذج مسح 1991 ، مقابل حوالي 0.97 لمعدل التصنيف الصحيح للسيدات غير المستخدمات (1/1) p . وانخفض هذان المعدلان في نموذج مسح 1997 – على التوالي – الى 0.44 و 0.95 . واذا كانت نسبة التطابق Hit-Ratio ضمن قيمها المقبولة (اكثر من 60 %) في كل النماذج الموفقة ، فأن ذلك يفسر بالتصنيف الصحيح بمعدل اعلى للسيدات غير المستخدمات الاتي يمثلن الاكثرية في المسيحين ، اذ تشير نسبة التطابق الى نسبة التصنيف الصحيح لعينة المسح الاجمالية . أي ان :

$$\text{Hit–Ratio} = \frac{n_{11} + n_{22}}{n} \qquad (8)$$

حيث : n_{kk} تشير الى عدد المفردات المصنفة تصنيفا صحيحا في المجموعة k .

والجدول رقم (2) يلخص مؤشرات التصنيف الصحيح الخاصة بستة نماذج ، يمكن تصنيفها في مجموعتين رئيسيتين ، المجموعة الاولى تضم النماذج الممكن استخدامها عند الاخلال بافتراض اكثر من افتراضات تحليل التميز . والمجموعة الثانية تضم النموذج المقترح في الدراسة السابقة للباحث ، وهو النموذج الذي قدر بتحقيق احد اشتراطات حسابات نموذج ANDERSON .وفيما يلي عرض سريع لهذه النماذج واساليب توفيقها :

اولا)) نماذج المجموعة الاولى :

1.نموذج ANDERSON باتباع التدريج في ادخال المتغيرات المستقلة {kleinbaum et al , 1998 } .

2.نموذج ANDERSON مع ادخال متغيرات التفاعل بين كل متغيرين مستقلين مرتبطين بقوة { NETER & WASSERMAN , 1996 } .

3. نموذج ANDERSON باستخدام المكونات الرئيسية PRINCPAL COMPONENTS وهذه النماذج الثلاثة تم استخدامها بغرض تخفيف اثر الازدواج الخطي MULTICOLLINEARITY بين المتغيرات المستقلة ، حيث وجد ارتباط قوي ومعنوي بين الكثير منها . ورغم ذلك لم تتحسن النتائج كثيرا فيما يخص P(2/2) ، وفي نموذجي المسحين ، بحيث لم تتجاوز 0.43 في احسن تقدير .

4.نموذج الانحدار اللوجستي التدريجي Stepwise logistic regression .

وتم استخدامه للاسباب التالية :

أ . تفادي انتهاك الافتراض الخاص بالتوزيع الطبيعي للمتغيرات المستقلة مع صعوبة استخدام التحويلات transformation لتقريب توزيعها الى الطبيعة { dillon & goldstein , 1984 } ، فاغلب هذه المتغيرات هي وصفية ، كما انه لايمكن اختيار تحويلة واحدة لتكون هي المناسبة لكل المتغيرات .

ب. تخفيف اثر الازدواج الخطي ، ورغم ذلك لم تتجاوز p(2/2) ال 0.66 في نموذج مسح 1991 وانخفض الى حوالى 0.42 في نموذج مسح 1997 .

5.الدالة التربيعية Quadratic function :

وهي الاسلوب المناسب لمواجهة عدم تساوي مصفوفتي التباين – التغاير في مجموعتي الدراسة { hair et al , 1987} الا انها قد تقود الى النتائج غريبة لايمكن تفسيرها أو قبولها { johnson & wichern , 1992 } . ورغم ان استخدام هذه الدالة ادى الى زيادة ملحوظة ومقبولة في p(2/2) ، الا ان النتائج رفضت بسبب ما افرزته من حجم عينة اكبر من الحجم المستخدم في التحليل .

ووفقا لنتائج السابقة ، فقد انتهى الباحث في دراسته السابقة الى ان تدني معدل التصنيف الصحيح للسيدات المستخدمات p(2/2) لا يرجع في اغلبه الى انتهاك واحد أو اكثر من افتراضات تحليل التمييز .

ثانيا)) نموذج المجموعة الثانية :

وهو نموذج ANDERSON بعد تعديل احتمالات القبلية لمجموعتي السيدات ، في نموذجي المسحين . فقد تم استخدام p_1=0.1 و p_2 = 0.9 في نموذج مسح 1991 ، وتم استخدام p_1 =0.6 و p_2 = 0.4 في نموذج مسح 1997 . وهذه الاحتمالات هي التي حققت افضل معدلات تصنيف ، بحيث كان الحصول على النموذجين التاليين :

1.نموذج مسح 1991 :

يمكن وضع السيدة اليمنية في مجتمع غير المستخدمات اذا كان :

$$31.5594-7.69208\,(x15) - 1.10455\,(x20) - 1.01351\,(x19) - .92267\,(x1) - .6825\,(x18) - .66992\,(x13) - .46212\,(x14) - .33901\,(x6) + .31880\,(x3) - .13906\,(x10) + .13229\,(x5) + .12248\,(x11) + .01774\,(x2) \geq 2.1972 \qquad (9)$$

ويمكن وضعها في مجتمع المستخدمات في غير ذلك .

حيث X ' S معرفة في جدول رقم (1) وفي هذا النموذج كان P(1/1) = 0.90 مقابل ل P (2/2) وبنسبة تطابق 89% تقريبا .

2. نموذج مسح 1997 :

يمكن وضع السيدة اليمنية في مجتمع غير المستخدمات اذا كان :

9.84215 – 1.83306 (x19) + 1.25878 (x18) – 1.20184 (x16) + .78724 (x13) - .68078 (x1)
-.66398 (x20) - .44740 (x14) - .38862 (x3) - .33384 (x4) + 23522 (x11) - .16291 (x 10)
-.08354 (x21) -.02886 (x 26) -.07488 (x15) + .07053 (x12) + .02674 (x2) ≥ -.4057
(10)

ويمكن وضعها في مجتمع المستخدمات في غير ذلك .

حيث X ' S معرفة في جدول رقم (1)

وفي النموذج كان P (1/1) = 0.82 مقابل 0.72 ل P (2/2) وبنسبة تطابق 80 5% تقريبا.

ولاختيار احد هذين النموذجين ليكون معبرا عن العوامل المحددة لموقف السيدة اليمنية من تنظيم الاسرة،كانت المفاضلة بينهما من حيث معدلات التصنيف التي يفرزها كل نموذج عند تطبيقه على :

أ . بيانات عينة من السيدات في نفس المسح .

ب. بيانات السيدات قس مسح الاخر .

وكانت نتيجة المفاضلة لصالح نموذج مسح 1991 (علاقة رقم 9) خاصة وانه يحتوي على عدد اقل من المتغيرات المميزة .

3. مقترح الدراسة :

سبقت الدراسة الى ان نتائج تحليل التمييز يتم الحصول عليها بالترجيح بحجمي مجموعتي الدراسة (مجتمعات أو عينات) . وبسبب عدم تساوب هذين الحجمين في اغلب الحالات التطبيقية ، فان الفرق بين هذين الحجمين لا يقتصر تأثيره على الاحتمالات القبلية فقط ، وانما يمتد ليطال تاثيره مصفوفة التباين – التغاير المشتركة . ففي العلاقة رقم (3) نجد ان تقدير العناصر المصفوفة S_{Pooled} يتجه نحو قيم عناصر مصفوفة الخاصة بالعينة الاكبر حجما ، وكلما زاد الفرق بين حجمي العينتين كلما زاد هذا الاتجاه . لذلك ، فان المعالجة الممكن اقتراحها في هذا البحث تتمثل في عدم استخدام العينة الاكبر حجما(وهي هنا مجموعة السيدات غير المستخدمات للوسائل) بل عينة جزئية منها يكون حجمها مساويا لحجم العينة الاصغر (وهي هنا مجموعة السيدات المستخدمات) مع اشتراط ان تظل هذه العينة الجديدة ممثلة للمجتمع الاصلي تمثيلا جيدا .

ان مثل هذا الاجراء يحقق فائدتين :

الاولى :

تتمثل في التقدير المتوسط غير المرجح لعناصر المصفوفة S_{Pooled} . فلا تميل قيمها نحو قيم عناصر المصفوفة الخاصة بالمجموعة الاكبر حجما .

الثانية :

تتمثل في التخلص من شرط الاحتمالات القبلية ، التي يتطلب تقديرها معرفة حجم مجتمع كل عينة ، وهو الامر الذي لا يتوفر في كثير من الاحيان . واذا اضفنا الى ذلك اتباع التدريج في ادخال المتغيرات الى النموذج (باستخدام 8 المعبر عنها رقم 7) فاننا نخفف من حدة اثر الازدواج الخطي ، ونبسط النموذج بتقليل عدد متغيراته ، خاصة وان لدينا 26 متغيرا مميزا .

اولا : نموذج مسح 1991 :

ان صفة استخدام وسائل تنظيم الاسرة يعبر عنها بنسبة عدد النساء المستخدمات الى اجمالي عدد السيدات المعرضات للحمل { Bongaarts & Rodriguez , 1991 } ولتعظيم حجم العينة العشوائية البسيطة (n) المراد سحبها من مجتمع حجمه (N) مفردة ، وذلك لتقدير نسبة ما ، يفترض ان هذه النسبة = 0.5 (ابو يوسف ، 1989) ومن ثم يقدر هذا الحجم وفقا للعلاقة :

$$n = \frac{N}{(N-1)\,B^2 + 1} \qquad (11)$$

حيث B هو حد خطا التقدير .

وقد سبقت الاشارة الى ان عدد السيدات غير المستخدمات للوسائل في مسح 1991 هو 4168 سيدة مقابل 607 سيدة مستخدمة ، وبذلك تم سحب بيانات 607 سيدة غير مستخدمة عشوائيا . أي اننا وضعنا n=607 و N=4168 في العلاقة رقم 11 . وهذا يجعل حد خطا التقدير في حدود 0.037 فقط .

وشرط هذه المعالجة ، المتمثل في ان العينة التي تم سحبها يجب ان تظل معبرة عن مجتمعها الاصلي ، قد تم مراعاته ، وذلك كالتالي :

1. اذا كان عدد سكان اليمن في عام 1991 لا يزيد على 12 مليون نسمة { CENTRAL STATISTICAL ORGANIZATION , 1994 } فان هذا الحجم (607 سيدة) وطبقا للعلاقة رقم 11 ، يمكن ان يمثل 6 مليون سيدة عند حد خطا لايتجاوز 0.041 .

2. تم سحب بيانات ال 607 سيدة (غير مستخدمة) من كل محافظة بحسب نسبة مساهمة كل محافظة في المجموعة الاصلية (4168 سيدة) .

وبذلك، وفقا للنتائج الواردة في ملحق رقم (1) الخاص بخلاصة التمييز الخطي التدريجي ، فان :

1. النموذج تضمن 17 متغيرا مميزا .
2. دالة التمييز الخطية ، المعبر عنها بالعلاقة رقم 3 ، هي:

$$Y = -.076373 (x1) + 0.58686 (x2) + 0.65346 (x3) - 0.79408 (x5) - 0.72474 (x6) - 1.31303 (x7) - 0.09576 (x8) - 0.13463 (x10) - 2.56505 (x13) - 1.49576 (x14) - 3.88883 (x15) - 0.99203 (x16) - 1.84747 (x18) - 3.89692 (x19) - 3.83914 (x20) - 1.35795 (x22) + 5.75429 (x23) \quad (12)$$

حيث X ' S معرفة في جدول رقم (1) .

3. قاعدة التصنيف ، المعبر عنها بالعلاقة رقم 4 ، هي :

ضع السيدة اليمنية في مجتمع غير المستخدمات اذا كان :

$$Y \geq 30.35534 \quad (13)$$

وضعها في مجتمع المستخدمات في غير ذلك .

حيث Y معرفة في (12) .

وهذه القاعدة افرزت معدلات تصنيف افضل من المعدلات الواردة في جميع النماذج السابقة الخاصة بمسح 1991 ، حيث تم استخدامها في :

أ. اعادة توزيع النساء اللاتي تم استخدام بياناتهن في اشتقاق دالة التمييز ومن ثم قاعدة التصنيف ، وتم الحصول على :

P(1/1) = 0.977 و P (2/2) = 0.941 ونسبة تطابق تزيد على 95 % (وذلك كما هو وارد في ملحق رقم 1) .

ب. اعادة توزيع سيدات عينة المسح (4168 سيدة غير مستخدمة مقابل 607 سيدة مستخدمة) وتم الحصول على :

P (1/1) = 0.89 و P (2/2) = 0.83 ونسبة تطابق تزيد على 74 % .

ثانيا : في مسح 1997 :

بتباع نفس الخطوات المذكورة في اولا ، تم الحصول على نتائج الواردة في ملحق رقم 2 والتي منها :

1. تضمن النموذج 15 متغيرا مميزا

2. دالة التميز الخطية تكون :

Y = -0.32321 (x1) + 0.05692 (x2) -0.27403 (x3) -0.33462 (x4) -.058353 (x8) -0.25958 (x10) + 0.31090 (x11) -0.75815 (x13) -0.34023 (x14) -2.98405 (x16) -2.86773 (x18) -3.47906 (x19) -2.71955 (x20) -0.10224 (x21) -0.33226 (x23) (14)

حيث X ' S معرفة في جدول رقم (1) .

3. تكون قاعدة التصنيف :

ضع السيدة اليمنية في مجتمع السيدات المستخدمات اذا كانت :

$Y \geq 45.69521$ (14)

وضعها في مجتمع السيدات المستخدمات في غير ذلك .

حيث : Y معرفة في جدول (14) .

وهذه القاعدة افرزت معدلات تصنيف افضل من المعدلات المتحصل عليها من جميع النماذج الخاصة بمسح 1997 . حيث تم استخدامها في :

أ. اعادة توزيع النساء اللاتي تم استخدام بياناتهن في اشتقاق دالة التمييز ومن ثم قاعدة التصنيف ، وتم الحصول على :

P (1/1) = 0.753 و P (2/2) = 0.839 ونسبة تطابق 80 % تقريبا (وذلك كما هو وارد في ملحق رقم 2) .

ب. اعادة توزيع سيدات عينة المسح (6747 سيدة غير مستخدمة مقابل 1838 سيدة مستخدمة) وتم الحصول على :

P (1/1) = 0.74 و P (2/2) = 0.80 ونسبة تطابق 75.3 % .

ج. اعادة توزيع سيدات مسح 1991 (4168 سيدة غير مستخدمة مقابل 607 سيدة مستخدمة) وتم الحصول على :

P (1/1) = 0.65 و P (2/2) = 0.67 ونسبة تطابق تزيد على 65 % .

وبمقارنة هذه المعدلات بنظيراتها الخاصة بنموذج الدراسة السابقة للباحث (علاقة 9) يمكن القول بان :

1. نموذج الدراسة السابقة تم تقديره بتعديل الاحتمالات القبلية لمجموعتي الدراسة ، بينما تم تقدير نموذج هذه الدراسة بمساواد حجمي المجموعتين ، وهذا الاجراء اكثر حيدة . مع ملاحظة ان نموذج الدراسة السابقة يبقى اكثر قبولا اذا لم يتحقق شرط مساواة حجمي المجموعتين ،

المتمثل في ان العينة التي يتم سحبها من المجموعة الاكبر حجما ، يجب ان تظل معبرة عن المجتمع الاصلي لهذه المجموعة .

2. النموذج السابق احتوى على 14 متغير مستقل ، والنموذج الخاص بهذه الدراسة احتوى على 17 متغير مستقل . ورغم ذلك ، فان هذا الاخير اكثر قبولا بسبب ما افرزته من معدلات تصنيف صحيح اعلى .

3. النموذجا اشتركا في 13 متغير من المتغيرات المستقلة . كما اشتركا النموذجان في سيادة تاثير المتغيرات ذات الطابع الثقافي على موقف السيدة اليمنية من تنظيم الاسرة ، هذه المتغيرات هي : معلومات السيدة عن الوسائل ، معرفة السيدة لمصادر الوسائل ، المعتقد الديني تجاه تنظيم الاسرة ، السماع عن الوسائل ، الاثار الجانبية للوسائل .

التوصيات :

ورد في المقارنة السابقة ان نموذج كل من هذه الدراسة والدراسة السابقة للباحث قد اشتركا في 13 متغير من المتغيرات المستقلة . كما اشتركا النموذجان في سيادة تاثير المتغيرات ذات الطابع الثقافي على موقف السيدة اليمنية من تنظيم الاسرة ، هذه المتغيرات هي : معلومات السيدة عن الوسائل ، معرفة السيدة لمصادر الوسائل ، المعتقد الديني تجاه تنظيم الاسرة ، السماع عن الوسائل ، الاثار الجانبية للوسائل .ولعل هذا الامر ساهم في تفسير الثبات التقريبي لمعدل الخصوبة الكلي Total Fertility Rate بين المسحين . فبالرغم من زيادة معدل استخدام الوسائل من 9.7 % في مسح 1991 { Central Statistical Organization , 1994 } الى 20.8 % في مسح 1997 { Central Statistical Organization , 1998 } . ولذلك لانجد مانعا من التاكيد على نفس التوصيات الواردة في الدراسة السابقة ، والمتمثلة في تنفيذ الحملات الاعلامية والندوات الدورية على مستوى المديرات ، والتي يكون من اهدافها :

أ. التعريف بالوسائل وانواعها واماكن الحصول عليها وعلى خدمات تنظيم الاسرة .

ب. التوعية بامكانية تخفيف الاثار الجانبية للوسائل عن طريق اختيار الوسيلة المناسبة واستخدامها بطريقة صحيحة .

ج.التوعية بعدم تعارض تنظيم الاسرة مع الاسلام .

4. الجداول

جدول رقم (1)

متغيرات الدراسة

المتغيرات المستقلة	الترميز
محل الاقامة	X 1 : 2 للريف ، 3 للحضر
عمر الزوجة	X 2
الحالة التعليمية للزوجة	X3 :2 امية ، 3 تقرا وتكتب ، 4 ابتدائيـــة ، 5 اعلى من الابتدائية
قراءة الزوجة لصحيفة واحدة على الاقل في الاسبوع	X4 : 2 لا ، 3 نعم
مشاهدة التلفزيون المحلي	X5 : 2 لا ، 3 نعم
الاستماع الى الاذاعة المحلية	X6: 2 لا ، 3 نعم
اشتغال الزوجة باجر نقدي	X7 : 2 لا ، 3 نعم
عمر السيدة عند الزواج بالسنوات	X8
مدة الزواج بالسنوات	X9
عدد المواليد السابق انجابهم	X10
عدد وفيات الاطفال	X11
عدد حالات الاجهاض	X12
موافقة الزوج على ممارسة تنظيم الاسرة	X13: 2 لا ، 3 نعم ، 4 لا تعرف
سماع الزوجة عن الوسائل	X14: 2 نعم ، 3 لا
معرفة الزوجة لمصادر الوسائل	X15 : 2 لا ، 3 نعم
معرفة الزوجة في مزيد من الاطفال	X16 : 2 لا ، 3 نعم
رغبة الزوجة في طفل ذكر اخر	X17: 2 لا ، 3 نعم
الزوجة تعارض تنظيم الاسرة دينيا	X18 : 2 لا ، 3 نعم
الزوجة ترفض الوسائل لاثارها الجانبية	X19 : 2 لا ، 3 نعم
الزوجة ترفض الوسائل لقصور معلوماتها	X20 : 2 لا ، 3 نعم
الحالة التعليمية للزوج	X21 : 2 امي ، 3 يقرا ويكتب ، 4 ابتدائية ، 5 اعدادية ، 6 اعلى من الاعدادية ، 7 الزوجـــة لا تعرف
الحالة العملية للزوج	X22 :2 لا تعمل ، 3 تعمل
اشتغال الزوج بالزراعة	X23: 2 نعم ، 3 لا
عمر الزوج	X24
ملكية وحدة السكن	X25 :3 نعم ، 2 لا
عدد السنع المعمرة	X26

جدول رقم (2)

مؤشرات التصنيف الصحيحة حسب طريقة توفيق النموذج وحسب المسح *

Hit – Ratio		P (2/2)		P (1/1)		المؤشر والمسح
1997	1991	1997	1991	1997	1991	الطريقة
% 82	% 93	0.39	0.42	0.94	1.00	المجموعة الاولى : -نموذج Anderson باستخدام التدريج
% 82	% 93	0.42	0.43	0.93	1.00	-نموذج Anderson مع متغيرات التفاعل
% 79	% 87	0.00	0.00	1.00	1.00	-نموذج Anderson باستخدام المكونات الرئيسية
% 82	% 94	0.42	0.66	0.92	0.98	-نموذج الانحدار اللوجستي التدريجي
% 72	% 80	0.64	0.66	0.74	0.82	- الدالة التربيعية
% 80	% 89	0.72	0.81	0.82	0.90	-نموذج Anderson بعد تعديل الاحتمالات القبلية

* في الدراسة السابقة للباحث .

5. المراجع

المراجع العربية :

1. ابو يوسف ، محمد (1989) " الاحصاء في البحوث العملية " المكتبة الاكاديمية ، القاهرة .
2. المنصوب ، عبد الحكيك عبد الرحمن (2004) " معالجة القصور في معدلات التصنيف المقدرة بنموذج Anderson – دراسة وتطبيق " مجلة الباحث الجامعي " ، العدد السابع ، جامعة إب . الصفحات 239 –262 .

المراجع الاجنبية :

1.Agresti ; A. (1996) '' An Introduction to Categorical Data Anlaysis '' john wily &sons , new york .

2.Anderson ; T.W. (1972) '' An Introduction to Multivariate staistical analysis ' ' john wily &sons , new york.

3.Bongaarts ; J.& Rodriguez ;G . (1991) "A new method for estimating contraceptive failure rates " depart of international economics and social affairs , U.N , NEW YORK .

4.CENTRAL STATISTICAL ORGANIZATION (1994) " DEMOGRAPHIC AND MATERNAL AND CHILD HEALTH SURVEY 1991 / 1992 "SANA'A .

5.DEMOGRAPHIC AND MATERNAL AND CHILD HEALTH SURVEY (1997) SANA'A " (1998)

6.DILLON ; W.&Goldstein ; M. (1984) Multivariate Analysis – Methods and Applications "" john wily &sons , new york.
7.Hair ; J.F, Anderson , R.E. & Tatham ;R.L. (1987) '' Multivariate data analysis with readings "second edition , macmillan publishing company , new york .
8.Jackson , B.B. (1983) '' Multivariate data analysis –An introduction "richard D.Irwin , inc , Georgetown .
9.JOHNSON ; R.A.& Wichern ; D.W. (1992) '' Applied Multivariate Statistical Analysis''third edition , prentice – hall international , inc , new jersey .
10.Kleinbaum ; D.J., Kupper ; L.L. , Muler ; K.E., & Nizam ; A. (1998) '' Applied Regression Analysis and other Multivariable methods ''thired edition , duxbury press , new york .
11.Lachenbruch ; peter A. (1975) ''Discrminant Anlysis '' Hanfer press , new york .
12.Neter ; J. & Wasserman ; W. (1997) ''Applied Linear Statical Mothods ; regression , ANALYSIS OF Variance and Experimental Designs '' third Edition , MCGRAW – HILL PUBLISHING COMPANY , NEW YORK .
13.Press ,J.& Wilson , S. (1978)CHOOSING Between logistic regreession and discriminant analysis "Journal of the american statical association , VOL. 73 , NO. 364 , PP. 699-705 .
14.randles ; R.H, Broffitt ; J.D.,Ramberg ; S.R.& HOGG ;R.V.(1978) '' Discriminant Analysis Based on rank '' JOURNAL OF THE AMERICAN Statistical Association , VOL.73 , NO.362, PP 379-384 .

الملاحق :

ملحق رقم (1) :

خلاصة التمييز الخطي عند مساواة حجمي مجموعتي الدراسة – مسح 1991

---------- DISCRIMINAT ANALYSIS -------

ON GROUPS DEFINED BY Y CURRENTLY USING A FP METHOD (91) ?

PRIOR PROBABILITIES

GROUP	PRIOR	LABEL
1	.50000	NO
2	.50000	YES
TOTAL	1.00000	

-------- VARIBLES IN THE ANLYSIS AFTER STEP 17 -------

X1	.6581989	33.6304	.2603369
X10	.6121448	45.9594	.2630125
X13	.9158381	172.0815	2903828
X14	.8370180	19.1633	.2571973
X15	.1946653	27.3120	.2589657
X16	.8161876	11.9842	.2556394
X18	.9174356	10.6502	.2553499
X19	.9237465	52.2979	2643880
X2	.6439846	649.8353	.3940621
X20	.8588787	76.5778	.2696571
X22	.8879941	4.7900	. 2540781
X23	.8786891	87.9024	.2721147
X3	.1623242	4.7471	.2540688
X5	.6403851	6.2230	.2543891
X6	.8547374	7.6034	.2546887
X7	.7926002	7.8483	.2547418
X8	.7822559	9.4442	.2550881

(fisher's linear discriminant functions)

Classification function coefficients

Y =	1 NO	2 YES
X1	-9.8564027	-8.0908654
X10	1.3435015	1.6581345
X13	9.4317698	11.9968154
X14	13.6053460	15.1011140
X15	58.3141078	62.2029420
X16	1.7350071	2.7270401
X18	84.7238124	86.5712760
X19	80 .8477749	84.7446898
X2	.5120851	-.0747695
X20	64.942661	68.7817955
X22	74.2097850	75.5677142
X23	72.748009	66.9937092
X3	-18.9524517	-19.6059056
X5	5.8259579	6.6200430
X6	1.4021805	31.2614638
X7	29.9484303	1.4690872
X8	-678.5254043	-708.8807351
(CONSTANT)	-678.5254043	-708.8807351

CANONICAL DISCRIMINANT FUNCTIONS

FCN	EIGENVALUE	PCT OF VARIANCE	CUM PCT	CANONICAL CORR	AFTER FUN	WILKS' LAMBDA	CHI-SQUARE	DF	SIG
:	0.253039	1612.639	17	.0000					
1*	2.9520	100.00	100.00	.8643	:				

* MARKS THE 1CANONICAL DISCRIMINANT FUNCTIONS ERMAING IN THE ANALYSIS .

CLASSIFICTION ERSULTS –

ACTUAL GROUP	NO. OF CASES	PREDICTED GROUP MEMBERSHIP 1	2
GROUP 1 NO	607	593 97.7 %	14 2.3 %
GROUP 2 YES	607	36 5.9 %	571 94.1 %

PERCENT OF "GROUPED" CASES CORRECTLY CLASSIFIED : 95.88 %

ملحق رقم 2

خلاصة التمييز الخطي التدريجي عند مساواة حجمي مجموعتي الدراسة – مسح 1997

---------- DISCRIMINANT ANALYSIS ----------

ON GROUPS DEFINED BY Y CURENTLY USE OF FT (97)

Prior probabilities

Group	prior	label
1	.50000	no
2	.50000	yes
total	1.00000	

--------- VARIABBLES IN THE ANALYSIS AFTER STEP 15 -------

VARIBLE	TOLERANCE	F TO REMOVE	WILK'S LAMBDA
X1	.5862948	7.6860	.5754460
X10	.2907005	119.2642	.5937925
X11	.6300620	40.7639	.580849
X13	.839700	107.4986	.5918580
X14	.8026182	11.5668	.5760841
X16	.8764339	260.5761	.6170282
X18	.9571554	113.8571	.5929035
X19	.9346059	339.8840	.6300686
X2	.4385279	48.8085	.5822077
X20	.9446969	102.5608	.5910460
X21	.6336697	8.4588	.5755730
X23	.8438935	9.0715	.5756738
X3	.5218354	12.6148	.5762564
X4	.6654716	7.6911	.5754468
X8	.8091240	669.2774	.6842300

Classification function coefficients

(fisher's linear discriminant functions)

Y =	1	2
	NO	YES
X1	6.2481403	6.5713506
X10	-.0660680	.1935118
X11	1.4683803	1.1574768
X13	14.3862040	13.6280459
X14	10.5023172	10.8425519
X16	77.6586874	80.6427443
X18	136.3989485	139.2666757
X19	81.2406910	84.7197452
X2	.5800428	.5231224
X20	134.5616219	137.2811650
X21	1.1198958	1.2221384
X23	8.6987658	9.0310331
X3	-.3663906	-.0923555
X4	11.7077023	12.0423183
X8	5.1735168	5.7570514
(CONSTANT)	-736.4402919	-782.1355029

CANONICAL DISCRIMINANT FUNCTIONS

FCN	EIGENVALUE	PCT OF VARIANCE	CUM PCT	CANONICAL CORR	AFTER FUN	WILKS' LAMBDA	CHI-SQUARE	DF	SIG
					: 0	.574182	1940.998	15	.0000
1*	.7416	100.00	100.00	.6525	:				

* MARKS THE 1 CANONICAL DISCRIMINANT FUNCTIONS ERMAING IN THE ANALYSIS.

CLASSIFICTION ERSULTS –

ACTUAL GROUP	NO. OF CASES	PREDICTED GROUP MEMBERSHIP 1	2
GROUP 1 NO	1838	1384 75.3 %	454 24.7 %
GROUP 2 YES	1838	269 16.1 %	15421 83.9 %

PERCENT OF "GROUPED" CASES CORRECTLY CLASSIFIED : 79.60 %

.....

.....

.....