

استخدام الطرائق التمييزية الاحصائية لتشخيص بعض امراض القلب

أنوار ضياء عبدالكريم

كلية العلوم - جامعة كركوك

الخلاصة

تناول البحث التحليل المميز والذي يعد من الطرائق الاحصائية المهمة في تصنیف مفردة واحدة او اکثر الى احد المجتمعات بالاعتماد على متغيرات ذات صفات تمیزیة ، والطرق الى بعض الانواع من الدوال التمييزية المتقدمة في هذا المجال وتطبیقها على نوعین من امراض القلب لبناء نموذج احتمالي للتمیز بینها بالاعتماد على بعض الاعراض والصفات والمتغيرات المصابة للمرض ، وقد كانت نموذج اللوجستك ذو تفوق على بقیة النماذج المستخدمة في البحث من حيث قلة نسبة التصنیف الخاطئ بالمقارنة مع بقیة النماذج .

المقدمة

يستفاد عادة من التحليل المميز في تصنیف مفردة واحدة او اکثر الى احد المجتمعات بالاعتماد على متغيرات ذات صفات تمیزیة وعلى اساس تشكیلة من المقایيس والصفات تكون المجامیع مختلفة فيما بینها والمفردة الواحدة تتبع الى مجموعة واحدة فقط ويمكن الاستفاده من التحليل المميز في التعرف على المتغيرات التي تساهم في عملية التصنیف اضافة الى التنبو الذي يزوّدنا بقدر شامل لکفاءة قواعد التصنیف.

ان دالة التمييز الخطیة المسندة على تركیب خطی للمتغيرات لكي تكون مثلى يجب أن تنتج اصغر احتمال لخطأ التصنیف علما بان هناك افتراضات يجب توفرها حول البيانات المستخدمة في التحليل وهي ان يكون موجه المتغيرات المسئللة ذو توزیع طبیعی متعدد المتغيرات ومصفوفات تباينات مشترکة متساوية وموجهات متosteatas مختلفة في كل مجموعة من المجامیع ولكن غالبا ما نواجه اختراق بعض الفرضیات وبالتالي فان تقديرات دالة التمييز الخطیة تفقد خواصها المثلی. مثلا عند عدم تساوی مصفوفة التباين والتباين المشترک فان استخدام دالة التمييز التربیعیة يكون ضروري كذلك عندما تكون بعض او جميع المتغيرات الاضافیة ذات طبیعة ثنائیة او مصنفة ستفقد دالة التمييز الخطیة خواصها المثلی ويكون النموذج انحدار اللوجستك بديلا

فعلاً للتمييز الطبيعي والذي يكون أكثر حصانة من الطبيعي إلا أنه أقل كفاءة عند تحقيق الافتراض الطبيعي وعند زيادة المسافة بين المجتمعات كما أن النموذج اللوجستيك قابل للتطبيق لفئة واسعة من التوزيعات متعددة المتغيرات متضمنة متغيرات متصلة و متقطعة وخليط من الاثنين .

ان دوال التمييز اللوجستيمية قد تناولت من قبل العديد من الباحثين للمساعدة في التشخيص الطبي ففي عام ١٩٩٦ قامت الباحثة تسنيم حسن البلداوي (البلداوي، ١٩٩٦) بمقارنة تحليلية لنموذجي التمييز الخطي واللوجستيك للحالة متعددة المجاميع لامراض الكبد المزمن اضافة الى التحري عن اداء الاسلوبين في حالة التعامل مع البيانات الملوثة باستخدام نموذجين حصينين فقط هما نموذج هوبر(Huber) الحصين ونموذج خطأ التصنيف ، وتوصلت الى ان تقديرات نموذج اللوجستيك المعين اكثراً تفوقاً في التمييز بين مجاميع المرض .اما في عام ٢٠٠١ قام ظافر حسين رشيد وأخرون (رشيد، ٢٠٠١) ببناء افضل نموذج احتمالي لتشخيص بعض امراض قرحة المعدة ولثلاثة مجامييع باستخدام دالة التمييز اللوجستيكية وعدة دوال تميزية معينة تعرض لأول مرة اضافة لدوال غير معينة وتوصلوا الى ان نموذج اللوجستيك المعين عموماً اعطى تفوقاً واضح في النسب المئوية للتصنيف الصحيح مما يعطيه الاهمية في اعتماده كاسلوب مساعد من ذوي الاختصاص في عملية تشخيص بعض امراض قرحة المعدة .

ان فكرة البحث والمتضمنة تشخيص بعض امراض القلب باستخدام دوال تميزية احصائية متقدمة يعتبر مهم لكون امراض القلب من الامراض الخطيرة والتي قد تفتك بحياة المصاب كما ان التشخيص المبكر للمرض يساعد في انقاذ حياة المصاب ويعنده الحالات المتطرفة للمرض . ولكون التشخيص مسألة صعبة من الناحية الطبية لكون امراض عديدة تشتراك في عدة اعراض يصعب التمييز بينها لذلك فان الاساليب الاحصائية المتقدمة تساعد ذوي الاختصاص في تشخيص بعض امراض القلب بصورة صحيحة وفي مرحلة مبكرة ، لذا سناحول في هذا البحث مقارنة اسلوبين الاول اسلوب يستند على تحليل متعدد المتغيرات ويكون ملائماً عند تحقق الفرضية الطبيعية والثاني يستند على تحليل انحدار غير خطي يكون ملائماً في تحليل البيانات مصنفة الاستجابة ذات متغيرات اضافية متصلة ومصنفة للتمييز بين مجموعتين من امراض القلب لبيانات واقعية .

الجانب النظري

دوال التمييز:

يتناول البحث بایجاز بعض دوال التمييز الطبيعية واللوجستكية وكالاتي (1996,Alvin)

١- دالة التمييز الطبيعية

ليكن الموجه العشوائي x ذو p من المكونات تتنمي إلى واحدة من g من المجتمعات

الطبيعية متعددة المتغيرات π_1, \dots, π_g .

وان $x \sim N_p(\mu_j, \Sigma)$ حيث $j=1,2,\dots,g$ وقد تختلف المجتمعات الطبيعية في متوسطاتها لكنها تتشابه من حيث تركيبة تبايناتها ، ولنفرض ان لدينا N من المشاهدات العشوائية المستقلة ذات الخصائص والصفات X وان كل مشاهدة تتمثل بالزوج (y_i, x_i) حيث تشير y_i إلى المجموعة التي تتنمي إليها المشاهدة i وان x_i موجه المقاييس او الصفات لذاك المشاهدة .ان قاعدة بيز للتصنيف تنص على تعين x إلى المجتمع π_j اذا كانت قيمة الاحتمال اللاحق اعظم ما يمكن. ففي حالة كون المعالم μ_j, Σ مجهولة فيمكن تقديرها من العينة.فإذا كان s_{j,x_j,n_j} تمثل حجم العينة لموجة المتوسطات ومصفوفة تباينات المجموعة j على التوالي ($j=1,\dots,g$) حيث يتم التعويض عن μ_j بمتوسط العينة \bar{x}_j وعن Σ بمصفوفة التباينات المشتركة للعينة S (*Pooled cov-matrix*) فان دالة التمييز الخطية سوف تكون :

$$Z_j(x) = \hat{B}_{oj} + \hat{B}_j x \quad (j = 1, \dots, g-1)$$

وان

$$\hat{B}_{oj} = \log \frac{\pi_j}{\pi_g} - \frac{1}{2} B_j (\bar{x}_j + \bar{x}_g) \quad \dots(1)$$

$$\dots(2) \hat{B}_j = S^{-1} (\bar{x}_j + \bar{x}_g)$$

٢- دالة التمييز التربيعية

وتشتمل في حالة عدم تساوي مصفوفة التباين والتباين المشترك للمجموعات وان معاملات

دالة التمييز التربيعية تكون:

$$\hat{B}_{oj} = \frac{1}{2} \log \frac{|S_g|}{|S_j|} - \frac{1}{2} (\bar{x}_j S_j^{-1} \bar{x}_j - \bar{x}_g S_g^{-1} \bar{x}_g) \quad \dots(3)$$

والتي تمثل الحد الثابت ، اما المعاملات x_i, x_j

$$\hat{B}_j = (S_j^{-1} \bar{x}_j + S_g^{-1} \bar{x}_g) \quad j=1, \dots, g-1 \quad \dots (4)$$

٣- تقديرات معلم انموذج اللوجستك متعددة الاستجابة (تقديرات الارجحية العظمى) (Albert, 1986) :
لنفرض ان لدينا N من المتجهات المستقلة ذات توزيع متعدد الحدود كل موجه ينتمي الى واحد من g من مجاميع الاستجابة ترمز للمشاهدات بـ (y_1, \dots, y_N) حيث ان كل مشاهدة y_i محددة بوحدة من g من القيم الممكنة $y_i = (y_{il}, \dots, y_{ig})$ وان $\sum_{ij} y_{ij}$ يكون محدداً لكل j تكون دالة لوغاريم دالة الارجحية بالشكل:

$$\dots (5) \ell = \sum_i^N \sum_j^S y_{ij} \log p_j(x_i)$$

نشتق دالة لوغاريم الارجحية بالنسبة للمعلمة B_{jk} وعند مساواة المشتقة بالصفر نحصل على المعادلات الطبيعية التالية

$$\sum_i^N [y_{ij} - n_i p_j(x_i)] x_{ik} = 0$$

ان جذور المعادلات تمثل المعلم التقديرية الناتجة عن عملية التعظيم ولكن هذه المعادلات غير خطية في المعلم نلجم إلى استخدام طريقة نيوتن - رافسون التكرارية والتي تنتج تقديرات مناسبة بعد بعض دورات تعاقبية وباستخدام طريقة المصفوفات في الحل تكون المعادلات الطبيعية بالشكل التالي $x_r^T = 0$ وان مصفوفة x الكاملة تكون ذات ابعاد $(p+I)(g-I)$ وان كل r_i عبارة عن موجه البوافي (r_1, \dots, r_N) ذو بعد I وان كل r_i عبارة عن موجه بعده (r_{1i}, \dots, r_{gi}) وكل عنصر يكون بالشكل التالي :

$$r_{ji} = y_{ji} - n_i p_i(x_i)$$

اما مصفوفة المشتقات الثانية فهي عبارة عن سالب مصفوفة فشر للمعلومات وتكون:

$$-\sum_{i=1}^N n_i p_j(x_i) p_i(x_i) x_{ik} x_{ik}^T = x_r^T x$$

حيث ان $v = diag(v_1, \dots, v_N)$ وان كل v_i عبارة عن مصفوفة مربعة تكتب بالشكل:

$$\dots (6) v_i = n_i [p_\delta(x_i)(\delta_{st} - p_i(x_i))\delta_t]$$

وان δ_{st} دلتا كرونكر ويعرف كالاتي :

$$\delta_{st} = \begin{cases} 1 & \delta = t \\ 0 & \delta \neq t \end{cases}$$

وتكون تقديرات المعاليم بطريقة نيوتن - رافسون وفق الصيغة التالية:

$$\dots (7) \beta(t-1) = \beta(t) + [x^T v(t)x]^{-1} xv(t)$$

والقيم الابتدائية تؤخذ عادة بمساواة موجه المعالم الابتدائي بالصفر أو تؤخذ من تقديرات المربعات الصغرى ونستمر في تقديرات متsequة للمعلم ونتوقف عندما يكون الفرق بين القيم المستخرجة في الدورات المتsequة مقدار صغير جداً حسب مستوى الثقة.

الطرق اللا معلمية باستخدام الرتب (1978,Rondle) :

ان طريقة الرتبة Rank تعتبر من الطرق اللامعلمية والتي تستخدم بغض النظر عن نوع البيانات سواء كانت تتوزع توزيعاً طبيعياً أم لا لذلك نلجم إليها في حالة عدم وفر الفرضيات الخطية بالدالة المميزة وفيها يتم استبدال البيانات الأصلية برتبها حيث تدمج المجتمعات وتضع لها الرتب $R(x_j)$ أي رتبة المشاهدة j وكل متغير y_1, y_2, \dots, y_g وكل متغير x_1, x_2, \dots, x_n وذلك بحسب التوزيع الطبيعي للرتب y يأخذ القيمة صفر للمشاهدة في المجتمع الأول وواحد للمشاهدة في المجتمع الثاني وهكذا دالة $(y_j R(x))$ مرتبطة فيما بينها بارتباط موجب وتضع الرتب التمييزية بشكل تصاعدي وتحدد متطلبات الرتب لكل متغير.

١- دالة التمييز الطبيعية للرتب

ان دالة التمييز الطبيعية للرتب للمجموعة j تكون :

$$Z_j^\wedge(x) = \beta_{oj}^\wedge + \beta_{jx}^\wedge \quad j=1,2,3,\dots,g-1 \quad \dots (8)$$

حيث ان:

$$\dots \beta_{oj}^\wedge = \log \frac{\pi_1^\wedge}{\pi_g^\wedge} - \frac{1}{2} \beta_j^{\wedge''} (\bar{x}_j^* - \bar{X}_j^*) \quad (9)$$

$$\dots (10) \beta_j^\wedge = S^{-1*} (\bar{X}_j^* - \bar{X}_g^*)$$

حيث

\bar{X}_j^*, \bar{X}_g^* متوسطات المجاميع g, j بعد اخذ الرتب للمشاهدات.

S^* مصفوفة التباينات المشتركة بعد اخذ الرتب للمشاهدات .

٢- دالة التمييز التربوية للرتب

بعد استخدام تحويل الرتب للبيانات يعاد تطبيق طريقة الدالة المميزة التربوية والمنكرة سابقاً .

احتمال خطأ التصنيف (1986,Lachenbrach) :

هناك العديد من الاساليب المتوفرة في تقدير احتمالات خطأ التصنيف من بيانات العينة والطريقة الشائعة الاستخدام هي اسلوب Jackknife حيث تزال كل مشاهدة تباعاً من العينة

وتحسب دالة التمييز للعينة بدون المشاهدة Z للمجموعة Z مثلاً ، ثم تصنف المشاهدة المحفوظة الى احدى المجتمعات وفق قاعدة التصنيف المحسوبة . ان تقدير احتمالات خطأ التصنيف في حالة التمييز متعدد المجاميع يكون مربكاً بعض الشيء لذا يفضل العمل مع بديله أو مكملة احتمال التصنيف الصحيح فعندما تكون معاليم المجتمع معلومة وعند تساوي نسب المجتمعات فان اعظم احتمال للتصنيف الصحيح هو $PC = \phi(8/2)$.اما في حالة كون المعاليم مجهولة فان احتمال التصنيف يكون بابداً مسافة مهلوبيس Mahalanobis للمجتمع بتلك المقدرة من العينة فيكون احتمال التصنيف الصحيح هو $PC = \phi(D/2)$.

اختيار المتغيرات:

ان تقليل عدد المتغيرات في دالة التمييز يفيد في قياس المتغيرات ذات العلاقة المعنوية وذات الصلة الاكبر بالمسألة موضوع الدراسة في الدراسات المقبلة ولتقليل من صعوبة وتعقيد المسألة وللتعرف على المتغيرات ذات القوة التمييزية المعنوية والتي تعطي اقل خطأ تصنيف هناك عدة اساليب منها (James, 1985):

١. مقياس مسافة مهلوبيس وعند استخدامه كمعيار لاختيار المتغيرات تحسب اولاً مسافات مهلوبيس بين كل زوج من المجاميع ثم تقام مسافة مهلوبيس بين اقرب المجاميع عند ادخال كل متغير ويكون له اكبر مسافة بين المجموعتين الاقرب له قابلية تمييز افضل

$$D_y^2 = d' w d$$

حيث ان d موجه الفروقات بين المتوسطات أي ان $d = (\bar{x}_i - \bar{x}_j)$: وان j, i هي المجتمعات الاقرب ، w مصفوفة مجاميع المربعات داخل المجتمع وتكون احصاءة F المكافئة بدرجات حرية $(q, n-q)$ كالاتي :

$$F_{ij} = [(n-g-q-1)n_i n_j q(n_i + n_j)][D_g^2 / n - g]$$

٢. احصاءة F الاكبر بين المجتمع ، حيث لاختبار فرضية تساوي متواسطي مجموعتين يمكن استخدام مسافة مهلوبيس او احصاءة F ، حيث :

$$\dots (11) F = \frac{(n-1-q)n_1 n_2}{q(n-2)(n_1 + n_2)} D_{ij}^2$$

بدرجات حرية $(q, n-q-1)$ ، ويمكن استخدام احصاءة F هذه في اختيار المتغيرات، في كل خطوة يدخل المتغير الذي له اكبر قيمة الى F . وبما ان مسافة مهلوبيس

موزونة بحجوم العينات عندما تستخدم F بين المجاميع كمعيار لاختيار المتغيرات فإن نتائج الطرق قد تختلف.

اختبار دالة التمييز:

١- اختبار تساوى المجموعات :

هناك عدة صيغ للاختبار منها مقاييس مربع كاي حيث:

$$\dots(12)\chi^2 = -[N-1-\frac{1}{2}(p+g)]\log e^\wedge$$

. والذى يتوزع تقريباً مربع كاي بدرجة حرية $(g-1)$.

٢- اختبار تساوي مصفوفة التباين والتباين المشترك لجميع المجاميع. ان احصاء الاختبار هي:

$$\dots(13)M = \sum_{i=1}^g n_i Ln|S| - \sum_{j=1}^g n_j Ln|S_j|$$

وقد أثبتت Box انه اذا ضرب M في ثابت C^1 والذى يساوى:

$$C^{-1} = 1 - \frac{2P^2 + 3P + 1}{6(P+1)(g-1)} \left[\sum_{i=1}^g \frac{1}{n_i} - \frac{1}{\sum_{i=1}^g n_i} \right]$$

فإننا نحصل على مقياس $Box's M$ والذي يتوزع تقريباً مربع كاي بدرجة حرية $(g-1)(p-1)/2$

الجانب التطبيقي

عينة البحث:

ان عينة البحث موضوع الدراسة هي عينة طبقية عشوائية *Stratified Random Sampling* وتشمل نوعين من المتغيرات الثانية والمتصلة لمجموعتين مختلفتين من امراض القلب هي تصلب الشرايين والجلطة القلبية والذي يمثل المتغير المعتمد حيث اشتملت بيانات العينة على عينة مؤلفة من ٢٠٦ مريض جمعت من كل من مستشفى اليرموك التعليمي ومستشفى الشهيد عدنان ومستشفى بعقوبة واستنادا الى طرق التشخيص السابقة عدد المتغيرات الاضافية هي (٧) متغيرات يمكن توضيحها بشكل موجز كالاتي:

- ١- العمر ٢- الوزن بالكيلوغرام ٣- الطول بالسنتيمتر

٤- ضغط الدم الوراثي ٥- ضغط الدم العالي ٦- نسبة الكوليسترول في الدم

٧- الجنس وهو متغير شرطي (ذكر = ١) ، (أنثى = ٢)

وقد كانت توزيع مفردات العينة حسب المجاميع ونسبة كل مجموعة كما في الجدول الآتي:

جدول يبين توزيع مفردات العينة حسب المجاميع ونسبة كل مجموعة

Groups	Group Name	Prior	No. of Cases
1	تصلب الشرايين	0.47	96
2	جلطة قلبية	0.53	110

وقد تم القيام بتشخيص نوعين من امراض القلب بالاعتماد على طرائق تمييزية مختلفة للتوصل الى افضل طريقة تمييز بين المجاميع. ولعدم توفر برامج فيما يتعلق بالتمييز التربيري وتمييز اللوجستك تم الاعتماد على برامج خاصة كتبت بلغة *Qbasic* ولكن هناك اخطاء تصاحب عملية التشخيص وكما في حالة التشخيص الطبي فقد يصاحبها بعض الاخطاء من قبل الاطباء، لذا تم مقارنة النسب المئوية للتصنيف الصحيح واعتماد الدوال التي تعطي اقل نسبة لخطأ التصنيف كاسلوب مساعد في عملية التشخيص لامراض القلب.

نتائج بعض الاختبارات:

- ١- تم اختبار البيانات لمعرفة فيما اذا كانت تتوزع توزيعا طبيعيا وذلك من خلال اختبار جودة توفيق البيانات ظهر ان المتغيرات التي لا تتبع التوزيع الطبيعي هي (الجنس ، العمر ، نسبة الكوليسترول) على الرغم من ان كبر حجم العينة يؤدي الى افتراض اقربها من التوزيع الطبيعي .
- ٢- اختبار معنوية الدالة المميزة بالاعتماد على المعادلة (١٢) . كانت قيمة مربع كای $\chi^2 = 85.252$ وبمستوى معنوية ٠،٠٠٠ هذا يدل على ان الدالة لها امكانية جيدة للتمييز وكذلك قيمة احصاء $Wilk's\ Lambda = 0.654$ وبمستوى معنوية ٠،٠٠٠
- ٣- اختزال عدد المتغيرات لكي يشمل الانموذج على المتغيرات ذات الطابع التميزي فقط تم استخدام اسلوب الاختبار المتدرج Stepwise Discreminate Analysis ، حيث ان استخدام احصاء *Wilks Lambda* كانت النتائج ان جميع المتغيرات معنوية عدا متغير الوزن وضغط الدم الواطئ والمعطاة في الجدول (١) ، (٢) ، وبعد استبعاد المتغيرات غير المعنوية تم احتساب معاملات الدالة المميزة الطبيعية والتربيري و الدالة اللوجستيكية كما في الجدول (٥).
- ويتبين من قيم تقديرات المعامل الاهمية التمييزية لمتغير نسبة الكوليسترول في الدم وضغط الدم العالي والعمر مما يدل على اهمية هذه المتغيرات في تشخيص امراض القلب .
- ٤- اختبار تساوي التباينات وجد ان قيمة احصاء $Box's M = 143.718$ وبمستوى معنوية ٠،٠٠٠ هكذا يدل على عدم تجانس بين المجموعات .

جدول (١) إحصاءات للمتغيرات المختارة stepwise

Step	Entered	Wilks Lambda						Sig.
		Statistic	df ₁	df ₂	df ₃	Exact F		
						Statistic	df ₁	df ₃
1	Var7	0.765	1	1	204.000	62.541	1	204.000
2	Var6	0.732	2	1	204.000	37.228	2	203.000
3	Var5	0.700	3	1	204.000	28.797	3	202.000
4	Var1	0.682	4	1	204.000	23.288	4	201.000
5	Var3	0.663	5	1	204.000	20.312	5	200.000

At each step the variable that minimizes the overall Wilk's Lambda is entered.

- a. Maximum number of steps is 14.
- b. Minimum partial F to enter is 3.84.
- c. Maximum partial F to remove is 2.71.
- d. F level, Tolerance, or V/N in sufficient for further computation.

جدول (٢) المتغيرات الباقية في التحليل

Step	Tolerance	F to remove	Wilks Lambda
1	Var 7	1.00	62.541
2	Var 7	0.933	39.463
	Var 6	0.933	09.354
3	Var 7	0.913	43.528
	Var 6	0.912	11.741
	Var 5	0.941	8.999
4	Var 7	0.912	40.962
	Var 6	0.892	8.822
	Var 5	0.941	8.794
	Var 1	0.973	5.316
5	Var 7	0.888	32.234
	Var 6	0.882	10.016
	Var 5	0.941	8.377
	Var 1	0.969	5.833
	Var 3	0.964	5.783

نتائج الطرق اللامعلمية :

تم اعادة اختبار معنوية الدالة المميزة لكون الطرق اللامعلمية تعتمد على تحويل البيانات الى رتبها حيث كانت قيمة مربع كاي $\chi^2 = 98.612$ وبمستوى معنوية ٠٠٠٠٠، هذا يدل على ان الدالة لها امكانية جيدة على التمييز والاختبار تساوي التباينات فان قيمة $Box'sM = 94.242$ وبمستوى معنوية ٠٠٠٠٠، هذا يدل على عدم التجانس بين المجموعات، ولتحديد المتغيرات المعنوية بهدف الاكتفاء بالمتغيرات ذات العلاقة المعنوية واستبعاد المتغيرات غير المعنوية فقد اظهرت نتائج اختبار المتغيرات باستخدام احصاء *Wilks Lambda* ان جميع المتغيرات معنوية عدا متغير

ضغط الدم العالي كما في الجدولين (٣)، (٤). وقد حسبت معاملات الدالة المميزة الطبيعية للرتب والتربيعية للرتب بعد استبعاد المتغيرات غير التمييزية كما في الجدول (٥).

جدول (٣) إحصاء Stepwise للمتغيرات المختارة لحالة الرتب Variables entered/removed a,b,c,d

Step	Entered	Wilks Lambda							fSig.	
		Statistic	df ₁	df ₂	df ₃	Exact F				
						Statistic	df ₁	df ₃		
1	Var.7	0.765	1	1	204.000	62.541	1	204.000	0.000	
2	Var.3	0.712	2	1	204.000	40.986	2	203.000	0.000	
3	Var.6	0.682	3	1	204.000	31.391	3	202.000	0.000	
4	Var.2	0.647	4	1	204.000	27.418	4	201.000	0.000	
5	Var.4	0.628	5	1	204.000	23.665	5	200.000	0.000	
6	Var.1	0.605	6	1	204.000	20.806	6	199.000	0.000	

At each step the variable that minimizes the over all Wilk's Lambda is entered

- a. Maximum number of steps is 14.
- b. Minimum partial F to enter is 3.84.
- c. Maximum partial F to remove is 2.71.
- d. F level, Tolerance, or V/N in sufficient for further computation.

جدول (٤) المتغيرات الباقية في التحليل لحالة الرتب

Step		Tolerance	F to remove	Wilks Lambda
1	Var 7	1.00	62.541	
2	Var 7	0.980	44.782	0.869
	Var 3	0.980	15.107	0.765
3	Var 7	0.899	26.240	0.771
	Var 3	0.979	15.065	0.733
	Var 6	0.917	08.978	0.712
4	Var 7	0.897	26.066	0.731
	Var 3	0.977	15.462	0.697
	Var 6	0.557	19.677	0.710
	Var 2	0.578	10.888	0.682
5	Var 7	0.894	26.879	0.713
	Var 3	0.972	13.404	0.620
	Var 6	0.524	23.251	0.704
	Var 2	0.576	11.403	0.664
	Var 4	0.902	05.954	0.647
6	Var 7	0.873	25.251	0.693
	Var 3	0.964	14.450	0.659
	Var 6	0.515	20.032	0.676
	Var 2	0.576	11.023	0.649
	Var 4	0.901	05.448	0.631
	Var 1	0.964	04.460	0.628

جدول رقم (٥) نتائج تحليل الطرق التمييزية

المعاملات	تمييز الطبيعي	تمييز التربيعي	تمييز اللوجستك	المعاملات	تمييز الرتب الطبيعي	تمييز الرتب التربيعي
١ الحد الثابت	-245.03 -234.11	-105.32 -133.65	-167.57 -144.32	١ الحد الثابت	-11.746 -13.358	-5.41 -3.22
Var.1	١ 0.42 ٢ 0.45	٢.٥٤ ١.٣٦	١.٣٨ ١.٧٩	Var . 7 ١ ٠.٠٧٢ ٢ ٠.٠٩٢	٠.٧١ ٠.٨٨	
Var.3	١ ٢.١٥ ٢ ٢.١٠	٢.٤٩ ٢.٣٢	٢.٣٣ ٤.٥٥	Var. 3 ١ ٠.٠٤٧ ٢ ٠.٠٣٥	٠.٢٦ ٠.٧٨	
Var.5	١ ٢.٧٦ ٢ ٢.٩٨	١.٦٥ ١.٧٧	١.٣٦ ١.٧٩	Var. 6 ١ ٠.٠٤٦ ٢ ٠.٠٢٧	٠.٧١ ٠.٤٣	
Var.6	١ ٠.٢٢ ٢ ٠.٢٠	٠.٥٣ ٠.٢٩	١.٢٢ ١.٦٥	Var . 2 ١ ٠.٠١١ ٢ ٠.٠٢٤	٠.٠٤١ ٠.٠٦٣	
Var.7	١ ٢٠.٢٨ ٢ ٢٢.٥٧	١١.٥٦ ١٠.٣٢	١٥.٤٣ ١٣.٢٢	Var. 4 ١ ٠.٠٢٥ ٢ ٠.٠٣٣	٠.٠٣٣ ٠.٠٦٤	
				Var . 1 ١ ٠.٠٢٨ ٢ ٠.٠٣٥	٠.٠١٢ ٠.٠١٥	

جدول رقم (٦) نسب التصنيف الخاطيء للدوال المميزة

Group	تمييز الطبيعي	تمييز الطبيعي للرتب	تمييز التربيعي	تمييز التربيعي للرتب	تمييز اللوجستك
1	٠,٣٤	٠,٢٩	٠,٣١	٠,٢٠	٠,١٥
2	٠,٤٧	٠,٣٥	٠,٣٦	٠,١٨	٠,٠٩
AMR	٠,٤١	٠,٣٢	٠,٣٤	٠,١٩	٠,١٢

AMR يشير الى معدل نسبة التصنيف الخاطئ

الاستنتاجات

نتيجة لما سبق يمكن ادراج ما يلي :

- فيما يتعلق بمتغير نسبة الكوليسترول في الدم فقد بلغ أعلى متوسط له في المجموعة الثانية اذ بلغ ٢,٥٤ وفي المجموعة الاولى ٢,٣٠ .

- ٢- كان متغير نسبة الكولستروول في الدم معنويا في كل الدوال التمييزية والاشارة الموجبة للمعلومة تدل على زيادة التعرض لامراض المحاميع بزيادة نسبة الكولستروول في الدم.
- ٣- ان تقدير نموذج اللوجستك اعطى تفوق واضح في المجموعتين وهذا واضح من النسبة المئوية للتصنيف الخاطئ والمعطاة في جدول (٦) وهذا يتلائم مع طبيعة البيانات ذات طبيعة مصنفة ومتصلة.
- ٤- كانت تقديرات النموذج التربيعي جيدة وهذا يتلائم مع اختبار عدم تساوي التباينات اضافة الى عدم ابتعاد بعض المتغيرات عن النموذج الطبيعي .
- ٥- ان دالة التمييز الطبيعية اعطى اعلى نسبة للتصنيف الخاطئ والمعطاة في جدول (٦) ، مما يستدعي عدم اعتمادها في عملية التشخيص في المجال الطبي والذي يتطلب الثقة في التشخيص.
- ٦- اعطى النموذج التربيعي للرتب نتائج جيدة. تلي انموذج اللوجستك وهذا يتلائم مع فرضية عدم تجانس التباين.

References

- Albert , A. & lesaffre , E.,(1986): Multiple groups logistic Discrimination ,Computer Math.Vol.12, pp. 30-52
- Alvin , C.R.,(1996): Methods of Multivariate Analysis , John Wiley & sons.,pp. 20-70
- James ,M.,(1985): Classification Algorithm , London.,pp. 58-80
- Lachenbrach , P.A. & Mickey , M.R.,(1986) : Estimation of Error Rate in Discriminate Analysis , Technometrics ,Vol.10,pp. 20-30
- Rondle ,H . Borffitt,JD. Rawberry , J.S. & Hogg , R .V, (1978): Diseriminate Analysis based on Ranks JASA, Vol .73,pp. 110-117

المصادر

- البلداوي ، تسنيم حسن (١٩٩٦) : مقارنة تحليلية بين نموذج اللوجستك ونماذج الدوال التمييزية ، اطروحة دكتوراه فلسفة في الاحصاء، كلية الادارة والاقتصاد ، جامعة بغداد.
- رشيد، ظافر حسين ، خمو، خلود يوسف ، حميد ، رند سليم (٢٠٠١) .

؛ تشخيص أمراض قرحة المعدة باستخدام دالة التمييز اللوجستية، مجلة كلية الإدراة والاقتصاد، جامعة بغداد، المجلد ٨، عدد ٤٥، ٢٥٠٦ ص.

مجلة جامعة Kirkuk - الدراسات العلمية المجلد (١) - العدد (٢) - ٢٠٠٦

Using of Statistical Discriminate methods for the Diagnosis of Some Heart Diseases^{٢٧}

Anwaar Dhia'a Abdul-Kareem
College of science – Kirkuk University

Abstract

The research use discriminate analysis which considered one of the important statistical methods in the classification of single variable or more for the populations under study depending on specific variable taking in consideration some types of advance discriminate functions in this field and apply them on two types of heart diseases for the build of probability model to distinguish two differences between them depending on signs, symptoms and some factors that may lead to cause the disease. This type of analysis show that the logistic model is superior to other models in this research regarding the ratio of wrong classification in comparison with the other models.