

تتفليس العوامل المسببة للإصابة بالتشوهات الخلقية بأستخدام الطوال التمييزية

م . انتصار مجيد جاسم *

المستخلص :

تم استخدام التحليل المميز لتصنيف درجة التشوهات الخلقية وقد صنفت الى ثلاث مجموعات اعتمادا على متغيرات ذات صفات تمييزية . تضمنت الدراسة سحب عينة عشوائية لكل مجموعة وباستخدام الدالة المميزة الخطية تم تصنيف التشوهات الخلقية على اساس مجموعة من المتغيرات المصاحبة للتشوه .

Abstract

The discriminatory Analysis has been used to classify some of the congenital deformities into three groups based on changes of discrimination features . The study has employed a random sample for each group by using the linear discriminant function some deformities have been classified on the base of the deformities accompanied changes .

المقدمة

يعد التحليل التمييزي او التصنيفي من الادوات المهمة في التحليل الاحصائي المتعدد المتغيرات والذي يهتم لمسألة التمييز بين اثنين او اكثر من المجتمعات المبنية على قياسات متعددة المتغيرات . فالمسألة المطروحة للتحليل في هذا الصدد ستكون متركزة على ايجاد عدد من الصيغ او القواعد المعينة التي تمكننا من تصنيف مشاهدة جديدة الى مجتمعها الصحيح وباقل خطأ تصنيف ممكن . ان حدوث التشوهات الخلقية في ازدياد من حيث نسبة الحدوث من جهة ومن حيث تنوع تلك التشوهات من جهة أخرى ، وكثرة حالات التشوه والولادات غير الطبيعية ، لذلك فمن الضروري دراستها والتعرف على اسبابها حيث ان التشوهات الخلقية هي سببا مهما في وفيات الاطفال فضلا عما تحدثه هذه التشوهات من تأثير على المدى البعيد على الطفل المصاب والعائلة والمجتمع . وعلى الرغم من ان اسباب التشوهات الخلقية هي وراثية ، كذلك الالتهابات والعوامل البيئية تؤدي دورا مهما في الإصابة ، وعادة مايصعب تحديد العامل المسبب .

هدف البحث

يهدف البحث الى تحديد عوامل الإصابة بالتشوهات الخلقية وتشخيص أهم المسببات المؤثرة في حدوثها ، اي تحديد العوامل الأكثر تمييز والتي تساعد في تصنيف الافراد الجدد على احدى المجموعات الثلاث المتمثلة بـ (التشوهات الخلقية) باستخدام الدوال التمييزية .

الجانب الطبي :

التشوه الخلقي Congenital deformity هو تخلف غير طبيعي في احد اعضاء الجسم او الانسجة في مرحلة تخلق الجنين، ووفقا لمنظمة الصحة العالمية فان التشوهات الخلقية هي شذوذات بنيوية او

* الجامعة المستنصرية / كلية الادارة والاقتصاد .

تأريخ استلام البحث 2016/4/17

تأريخ قبول النشر 2016/6/13

وظيفية موجودة منذ الولادة سواء تم التعرف عليها عند الولادة او في وقت لاحق ويمكن ان تكون اسبابها وراثية او بيئية او مجموعة عوامل متعددة .
يمكن تصنيف التشوهات الخلقية بالاستعانة بالمختصين في الجانب الطبي على ثلاثة انواع [5].

- 1- التشوهات البسيطة (Deformities Simple) وتشمل :
الشفة المشقوقة (الارنبية) (Cleft Lip) ومن اشكالها شق الشفة وشق الحنك .
والتشوهات الخلقية للاطراف (العلوية او السفلية) (Congenital Limb Defects)
- 2- التشوهات المتوسطة : (Deformities medium) وتشمل :
الشرج المسدود (Imperforate anus) وقيلة سحائية (Meningeal) وهي نوع من شلل الحبل الشوكي .
- 3- التشوهات الخطيرة (المميتة) : (Serious deformities) وتشمل :
انعدام الدماغ (الدماغية) (Anencephaly)
ومتلازمة ادواردز (Edward Syndrome) وهي اضطراب وراثي ناجم عن زيادة في العدد الاجمالي للكروموسومات .

الجانب النظري

التحليل التمييزي أو التصنيفي :

يستخدم التحليل التمييزي أو التصنيفي أما لتصنيف مفردات جديدة أو لمعرفة المتغيرات التمييزية ، وهو أحد أساليب متعدد المتغيرات الذي يؤكد على تحليل المتغيرات انيا وليس بشكل منفصل .واساس هذا التحليل هو ايجاد مركبات خطية للمتغيرات المستقلة كأساس في تصنيف حالات جديدة الى مجاميعها . ان المشكلة الاحصائية تكمن في كيفية ايجاد دالة تمييزية وفقاً للمعايير أو القياسات التي يمكن الحصول عليها من الافراد والتي بواسطتها يمكن تصنيف الافراد الجدد (المجهولي الانتماء) الى المجموعة الصحيحة .

دوال التمييز [4]

الدالة التمييزية الخطية لـ (Fisher)

ان الدالة التمييزية الخطية المستندة الى تركيب خطي للمتغيرات لكي تكون مثلى يجب ان تنتج احتمال خطأ التصنيف ويجب ان تحقق الافتراضات الاتية :

- 1- ان يكون موجه المتغيرات التوضيحية (المستقلة) ذا توزيع طبيعي متعدد المتغيرات .
- 2- تساوي التباينات لكل المجاميع المدروسة (مصفوفات التباين والتباين المشترك) .
- 3- ان تكون موجات المتوسطات مختلفة في كل مجموعة من المجاميع .

الدالة التمييزية لمجموعتين او اكثر :

هي تركيبة خطية تبني بالاعتماد على عدة مؤشرات من عينة اختيرت مفرداتها من مجتمعين او اكثر، اذ ان هذه الدالة تمكننا من تصنيف المفردة على المجتمع الاصلي وباقل خطأ تصنيف ممكن .
فلو افترضنا ان لدينا مجتمعين يراد المقارنة بينهما باستخدام دالة التمييز لهما نفس مصفوفة التباين والتباين المشترك وان متوسطي هذين المجتمعين μ_1 و μ_2 مختلفين ، تم سحب عينة

عشوائية حجمها n_1 تضم $(x_{11}, x_{12}, \dots, x_{1n_1})$ من المجتمع الاول ، وسحبت عينة عشوائية ثانية حجمها n_2 تضم $(x_{21}, x_{22}, \dots, x_{2n_2})$ من المجتمع الثاني ، وان كل موجه x_i يضم (P) من المتغيرات (الصفات) وبالتالي فان دالة التمييز عبارة عن تركيب خطي يضم (p) من المتغيرات التي تعمل على تعظيم الفرق بين متوسطي المجموعتين وهذه الدالة تكون :

$$Z = \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_p x_p \quad \dots (1)$$

$$Z = \underline{\alpha}' \times$$

حيث أن :

$\underline{\alpha} : (\alpha_1, \alpha_2, \dots, \alpha_p)$ متجه معلمات النموذج وتستهمل في عملية التصنيف .

$\underline{x} : (x_1, x_2, \dots, x_p)$ متجه المتغيرات .

Z : دالة التمييز

حيث نختار قيم α بحيث تعطي أعلى تمييز بين المجموعتين اي بشرط توفر خاصية تعظيم التباين والتباين المشترك بين المجموعات الى التباين داخل المجموعات .
والتعظيم يتم عندما :

$$\underline{\alpha} = s^{-1}(z_1 - z_2) \dots (2)$$

والمهم هنا هو ايجاد المتجه $\underline{\alpha}$ ($\alpha_1, \alpha_2, \dots, \alpha_p$) الذي يجعل قيمة Q اكبر مايمكن .

$$Q = \frac{\text{Between group variation}}{\text{With in group variation}}$$

$$Q = \frac{[\bar{z}_1 - \bar{z}_2]^2}{s_2^2} = \frac{[\bar{z}_1 - \bar{z}_2]^2}{\sum_{i=1}^{n_1} [z_{i1} - \bar{z}_1]^2 + [z_{i2} - \bar{z}_2]^2} \dots (3)$$

حيث ان :

Q : مسافة مهالونوبيس التربيعية .
 $\bar{z}_1$: متوسط قيم مشاهدات المجموعة الاولى .
 $\bar{z}_2$: متوسط قيم مشاهدات المجموعة الثانية .
 z_{i1} : قيمة المشاهدة i في المجموعة الاولى .
 z_{i2} : قيمة المشاهدة i في المجموعة الثانية .
 وبذلك يمكن ايجاد $\bar{z}_2, \bar{z}_1$ على النحو التالي :

$$\bar{z}_1 = \underline{\alpha}' \bar{x}_1 = \alpha_1 \bar{x}_{11} + \alpha_2 \bar{x}_{21} + \dots + \alpha_p \bar{x}_{p1}$$

$$\bar{z}_2 = \underline{\alpha}' \bar{x}_2 = \alpha_1 \bar{x}_{12} + \alpha_2 \bar{x}_{22} + \dots + \alpha_p \bar{x}_{p2}$$

وأن :

$$[\bar{z}_1 - \bar{z}_2] = [\underline{\alpha}' (\bar{x}_1 - \bar{x}_2)]^2 = \underline{\alpha}' (\bar{x}_1 - \bar{x}_2) (\bar{x}_1 - \bar{x}_2) \underline{\alpha}$$

$$\sum_{i=1}^{n_1} (z_{i1} - \bar{z}_1)^2 = \underline{\alpha}' \left[\sum_{i=1}^{n_1} (z_{i1} - \bar{z}_1)' (z_{i1} - \bar{z}_1) \right] \underline{\alpha}$$

$$= \underline{\alpha}' (n_1 - 1) s_1 \underline{\alpha}$$

$$\sum_{i=1}^{n_2} (z_{i2} - \bar{z}_2)^2 = \underline{\alpha}' \left[\sum_{i=1}^{n_2} (z_{i2} - \bar{z}_2)' (z_{i2} - \bar{z}_2) \right] \underline{\alpha}$$

$$= \underline{\alpha}' (n_2 - 1) s_2 \underline{\alpha}$$

اذ أن :

s_2, s_1 هما مصفوفتا التباين والتباين المشترك التقديرية داخل كل من المجموعتين الاولى والثانية على التوالي :
 ونعوض لاجداد الاختلافات داخل المجموعات على النحو الآتي :

$$\sum_{i=1}^{n_1} (z_{i1} - \bar{z}_1) + \sum_{i=1}^{n_2} (z_{i2} - \bar{z}_2) = \underline{\alpha}' (n_1 - 1) s_1 \underline{\alpha} + \underline{\alpha}' (n_2 - 1) s_2 \underline{\alpha}$$

$$= (n_1 + n_2 - 2) \underline{\alpha}' s \underline{\alpha} \dots (4)$$

وبفرض ان s هي مصفوفة التباين التجميعي (pooled) والتباين المشترك داخل العينتين .

$$S = \frac{(n_1 - 1) s_1 + (n_2 - 1) s_2}{(n_1 + n_2 - 2)}$$

$$\therefore (n_1 + n_2 - 2) S = (n_1 - 1) s_1 + (n_2 - 1) s_2$$

ولاجداد قيمة Q نعوض في المعادلة (3) وعلى النحو الآتي :

$$Q = \left[\frac{\alpha'(\bar{x}_1 - \bar{x}_2)(\bar{x}_1 - \bar{x}_2)\alpha}{(n_1 + n_2 - 2)(\alpha' s \alpha)} \right]$$

$$= \frac{1}{(n_1 + n_2 - 2)} \left[\frac{\alpha'(\bar{x}_1 - \bar{x}_2)(\bar{x}_1 - \bar{x}_2)}{(\alpha' s \alpha)} \right]$$

$$= \frac{1}{n_1 + n_2 - 2} Q^* \quad \dots (5)$$

حيث Q هي النسبة المراد تعظيمها .

فذلك يمكن وضع الشرط ($\alpha' s \alpha = 1$) ، أي ان التباين داخل المجموعات ($\alpha' s \alpha = 1$) ثابت (constant) ليستعمل من خلال مضروب لاكرانج في الدالة فتصبح على النحو الآتي :

$$L(\alpha, \lambda) = \alpha' (\bar{x}_1 - \bar{x}_2)' (\bar{x}_1 - \bar{x}_2) \alpha - \lambda (\alpha' s \alpha - 1)$$

$$\frac{\partial L}{\partial \alpha} = 2\alpha (\bar{x}_1 - \bar{x}_2)' (\bar{x}_1 - \bar{x}_2) - 2\lambda s \alpha$$

$$\frac{\partial L}{\partial \alpha} = 0$$

فتصبح ($\frac{\partial L}{\partial \alpha}$) على النحو الآتي :

$$\frac{\partial L}{\partial \alpha} = \alpha (\bar{x}_1 - \bar{x}_2)' (\bar{x}_1 - \bar{x}_2) = \lambda s \alpha$$

وبضرب طرفي المعادلة في أعلاه بـ α' ليجاد قيمة λ ينتج :

$$\lambda = \frac{\alpha (\bar{x}_1 - \bar{x}_2)' (\bar{x}_1 - \bar{x}_2) \alpha'}{\alpha s \alpha'}$$

$$\lambda = (\bar{x}_1 - \bar{x}_2)' s^{-1} (\bar{x}_1 - \bar{x}_2) = D^2$$

اذ ان D^2 مسافة مهالونوبيس وهي مقياس للمسافات بين مراكز المجموعات .
حيث :

$$\alpha (\bar{x}_1 - \bar{x}_2)' (\bar{x}_1 - \bar{x}_2) = (\bar{x}_1 - \bar{x}_2) D^2$$

$$\therefore \alpha (\bar{x}_1 - \bar{x}_2)' (\bar{x}_1 - \bar{x}_2) = \lambda s \alpha$$

$$\therefore \lambda s \alpha = (\bar{x}_1 - \bar{x}_2) D^2$$

$$\alpha = s^{-1} (\bar{x}_1 - \bar{x}_2) \quad \dots (6)$$

وبالتالي فإن الدالة التمييزية الخطية لـ (fisher) تأخذ النحو الآتي :

$$Z = \alpha' \times = (\bar{x}_1 - \bar{x}_2) s^{-1} \times$$

ثم بعد ذلك يتم اختيار معنوية الدالة المميزة الخطية بعد استخراج المعاملات وينصب الاهتمام حول قوة الدالة المميزة الخطية ومدى صلاحيتها للتمييز ويمكن معرفة ذلك بواسطة تحليل التباين (Analysis of variance) حيث تكون فيه مجموع المربعات بين المجموعات (sum square between groups) يساوي :

$$S.S \text{ between groups} = \left(\frac{n_1 n_2}{n_1 + n_2} \right) [D^2]^2$$

وبدرجة حرية k .

ومجموع المربعات داخل المجموعات (sum square within groups) يساوي :

$$s.s \text{ within groups} = D^2$$

وبدرجة حرية ($n_1 + n_2 - k - 1$)

فيكون متوسط المربعات بين المجموعات (Mean square between groups) يساوي :

$$Mean s.s. = \frac{\left(\frac{n_1 n_2}{n_1 + n_2}\right) [D^2] / k}{D^2 / (n_1 + n_2 - k - 1)} \dots (7)$$

اذ ان النسبة الاخيرة تتبع توزيع F بدرجة حرية k و $(n_1 + n_2 - k - 1)$ فعندما تكون هذه النسبة معنوية هذا يدل على ان الدالة صالحة للتمييز وان التغيرات المستخدمة تعطي احسن تمييز بين المجتمعين .
الدالة التمييزية في حالة التوزيع الطبيعي متعدد المتغيرات :

بافتراض ان $f_j(x)$ هي مجموعة الكثافة الاحتمالية الشرطية لـ x في المجموعة $(G=j)$ وان (π_j) يمثل الاحتمال المسبق (prior probability) للمجموعة (i) اذ ان $(\sum_{j=1}^l \pi_j = 1)$ فانه بتطبيق نظرية بيز (Bayes Theorem) يمكن الحصول على المجموعة اللاحقة (Group posterior) $P(G/x)$ والتي تعطي افضل تصنيف وبأقل معدل خطأ بين تقنيات التصنيف الاخرى .
اذ ان :

$$P\left(G = j / \underline{X} = \underline{x}\right) = \frac{f_j(\underline{x}) \pi_j}{\sum_{i=1}^l f_i(\underline{x}) \pi_i} \dots (8)$$

وعندما تتبع البيانات التوزيع الطبيعي متعدد المتغيرات فان :

$$f_j(\underline{x}) = \frac{1}{(2\pi)^{p/2} |\Sigma_j|^{1/2}} e^{-\frac{1}{2}(\underline{x} - \underline{\mu}_j)' \Sigma_j^{-1} (\underline{x} - \underline{\mu}_j)} \dots (9)$$

وعلى افتراض انها تساوي مصفوفات التباين والتباين المشترك للمجموعات أي ان :

$$\sum_j = \sum, \quad v_j = 1, 2, \dots, l$$

عندئذ قاعدة التمييز الخطية أو حدود القرار في حالة مجموعتين يمكن اشتقاقها للنسبة بين دالتي الكثافة الاحتمالية اللاحقة للمجموعتين وبحسب الصيغة الاتية :

$$\frac{P(G = 1 / \underline{X} = \underline{x})}{P(G = 2 / \underline{X} = \underline{x})} = \frac{f_1(\underline{x}) \pi_1 / \sum_{i=1}^l f_i(\underline{x}) \pi_i}{f_2(\underline{x}) \pi_2 / \sum_{i=1}^l f_i(\underline{x}) \pi_i} = \frac{f_1(\underline{x}) \pi_1}{f_2(\underline{x}) \pi_2} \dots (10)$$

ويمكن كتابة المعادلة (10) على النحو الاتي :

$$\begin{aligned} \frac{P(G = 1 / \underline{X} = \underline{x})}{P(G = 2 / \underline{X} = \underline{x})} &= \frac{\frac{\pi_1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\underline{x} - \underline{\mu}_1)' \Sigma^{-1} (\underline{x} - \underline{\mu}_1)\right]}{\frac{\pi_2}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\underline{x} - \underline{\mu}_2)' \Sigma^{-1} (\underline{x} - \underline{\mu}_2)\right]} \\ &= \exp\left[-\frac{1}{2}\left(\underline{x}' \sum \underline{x} - 2 \underline{x}' \sum \underline{\mu}_1 + \underline{\mu}_1' \sum \underline{\mu}_1 - \underline{x}' \sum \underline{x} + 2 \underline{x}' \sum \underline{\mu}_2 - \underline{\mu}_2' \sum \underline{\mu}_2\right)\right] * \frac{\pi_1}{\pi_2} \\ &= \exp\left[-\frac{1}{2}\left(-\underline{x}' \sum \underline{\mu}_1 + 2 \underline{x}' \sum \underline{\mu}_2 + \underline{\mu}_1' \sum \underline{\mu}_1 - \underline{\mu}_2' \sum \underline{\mu}_2\right)\right] * \frac{\pi_1}{\pi_2} \dots (11) \end{aligned}$$

وبإضافة وطرح $\underline{\mu}_2' \Sigma^{-1} \underline{\mu}_2$ للطرف الايمن من المعادلة يكون :

$$\frac{p(G=1/\underline{X}=\underline{x})}{p(G=2/\underline{X}=\underline{x})} = \exp \left(\sum_{j=1}^l (\underline{\mu}_1 - \underline{\mu}_2) - \frac{1}{2} (\underline{\mu}_1 + \underline{\mu}_2)' \sum_{j=1}^l (\underline{\mu}_1 - \underline{\mu}_2) \right) * \frac{\pi_1}{\pi_2} \quad \dots (12)$$

وبأخذ اللوغارتم الطبيعي لطرفي المعادلة (12) ينتج :

$$d_j^l(\underline{x}) = \left[\sum_{j=1}^l (\underline{\mu}_1 - \underline{\mu}_2) - \frac{1}{2} (\underline{\mu}_1 + \underline{\mu}_2)' \sum_{j=1}^l (\underline{\mu}_1 - \underline{\mu}_2) \right] + \ln \frac{\pi_1}{\pi_2} \quad \dots (13)$$

اذ ان :

$$d_j^l = \log \frac{p(G=1/\underline{X}=\underline{x})}{p(G=2/\underline{X}=\underline{x})}$$

ويمثل الحد الاول هنا دالة التمييز الخطي لـ (fisher) ، أما الحد الثاني فيمثل نقطة الفصل بين المجموعتين .

وبافتراض $p = \frac{\pi_1}{\pi_2}$ فإن المعادلة (13) لحالة j من المجاميع منها على نقاط دالة التمييز الخطية لـ $\underline{x}$ (Linear Discriminante Scores) التي تحقق قرار الخطية بين المجموعتين على النحو الآتي :

$$d_j^l(\underline{x}) = M_j \sum_{j=1}^l \underline{x} - \frac{1}{2} M_j' \sum_{j=1}^l M_j + \ln(p_j) \quad \forall j = 1, 2, \dots, l \quad \dots (14)$$

اذ ان $p_j, \sum, M_j$ غير معروفة فيجب تقديرها من بيانات العينة ويستعمل متوسط المجموعة $\bar{x}_j$ بوصفه تقدير لـ M_j ومصفوفة التباين المشترك العامة S بوصفها تقدير لـ $\sum$ وان تقدير احتمال خطأ التصنيف (p_j) يمكن ايجاده بأسلوبين :
أما بافتراض (p_j) تكون ثابتة لكل المجتمعات وتنتج :

$$p_j^{\hat{}} = \frac{1}{l} \quad , \quad \forall l \quad \dots (15)$$

أو أنها تقدير لكل مجتمع بحسب الآتي :

$$p_1^{\hat{}} = \frac{n_j}{n} \quad \dots (16)$$

اذ ان :

n_j : عدد المشاهدات التي تعود الى المجموعة j .
 n : عدد المشاهدات الكلية لكل المجموعات .

قواعد التصنيف [3]

قاعدة التصنيف وفقا للدالة التمييزية لـ (fisher) :

لاستخراج دالة التمييز الخطية في تصنيف المشاهدات التي تعود الى المجتمعات غير المعروفة فأننا :
نبدأ باحتساب قيم متوسطات دالة التمييز للعينتين على النحو الآتي :

$$\bar{Z}_i = (\bar{x}_1 - \bar{x}_2) s^{-1} \bar{x}_i \quad , \quad i = 1, 2$$

نحسب متوسط متوسطات دالة التمييز الخطية على النحو الآتي :

$$\begin{aligned} \bar{Z}_1 &= (\bar{x}_1 - \bar{x}_2) s^{-1} \bar{x}_1 \\ \bar{Z}_2 &= (\bar{x}_1 - \bar{x}_2) s^{-1} \bar{x}_2 \\ C &= \frac{\bar{Z}_1 + \bar{Z}_2}{2} = \frac{(\bar{x}_1 - \bar{x}_2)' s^{-1} (\bar{x}_1 - \bar{x}_2)}{2} \quad \dots (17) \end{aligned}$$

اذ ان C هي نقطة تقسيم وهي نقطة فاصلة وقاطعة بين المجتمعين (cut point) .
تكون قاعدة تصنيف المشاهدة $\times$ للمجتمع الاول اذا كان :

$$Z_1 = (\bar{x}_1 - \bar{x}_2)' s^{-1} \times > \frac{(\bar{x}_1 - \bar{x}_2)' s^{-1} (\bar{x}_1 + \bar{x}_2)}{2} \quad \dots (18)$$

الى المجتمع الثاني اذا كان :

$$Z_2 = (\bar{x}_1 - \bar{x}_2)' s^{-1} \times \leq \frac{(\bar{x}_1 - \bar{x}_2)' s^{-1} (\bar{x}_1 + \bar{x}_2)}{2} \quad \dots (19)$$

وتجدر الاشارة الى ان هناك دالة تمييز أخرى لـ (Rao) تستخدم في حالة وجود مجموعتين فأكثر اذ أن دالة (fisher) تختص لمجموعتين فقط ، ودالة (Rao) تستخرج لكل مجموعة دالة تمييزية بالشكل :

$$Z_i = A_i + \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_p x_p \quad \dots (20)$$

اختبار معنوية المتغيرات [3]

لأختيار المتغيرات ذات التمييز المعنوي والتي تنتج أقل خطأ تصنيف هناك عدة أساليب منها طريقة ادخال المتغيرات المستقلة سوية وطريقة الاختيار المتدرج (stepwise statistics) ومنها يكون اختيار دالة التمييز القياسي لمعنوية الفروق بين متوسطات المجموعات ويعرف بمقياس Wilk's lambda متعدد المتغيرات الى p من المتغيرات .

$$\Lambda = \frac{|W|}{|T|}$$

حيث أن :

W: مصفوفة التباين والتباين المشترك داخل المجموعات .

T: مصفوفة التباين والتباين المشترك الكلي .

أما اختبار F وضعت صيغته الرياضية كالآتي :

$$F = \frac{1 - \Lambda^{\frac{1}{5}}}{\Lambda^{\frac{1}{5}}} \frac{ms - 2\lambda}{p(k - 1)} \quad \dots (21)$$

بدرجات حرية $df_1 = p(k - 1)$

$df_2 = ms - 2\lambda$

$$m = N - \frac{1}{2}(p + k)$$

$$s = \left[\frac{p^2(1 - k)^2 - 4}{(1 - k)^2 + p^2 - 5} \right]^{\frac{1}{2}}$$

$$\lambda = \frac{p(k - 1) - 2}{4}$$

اختبار معنوية دالة التمييز [3]

1- اختبار تساوي المجموعات ويستخدم مقياس مربع كاي حيث ان :

$$\chi^2 = - \left[N - 1 - \frac{1}{2}(p + k) \right] \log(\Lambda) \quad \dots (22)$$

والذي يتوزع تقريبا مربع كاي بدرجة حرية $p(k-1)$.

2- اختبار تساوي مصفوفة التباين والتباين المشترك لجميع المجموعات :

أحصاءة الاختبار المستخدمة هي :

$$M = [n_i \ln|s| - [n_i \ln|s_i|]] \quad \dots (23)$$

وقد أثبت Box انه اذا ضرب M في ثابت C^{-1} والذي يساوي :

$$C^{-1} = 1 - \frac{2p^2 + 3(p-1)}{6(p+1)(k-1)} \left[\sum_{i=1}^k \frac{1}{n_i} - \frac{1}{\sum n_i} \right] \dots (24)$$

وبذلك نتوصل الى مقيس $Box's M$ يتوزع مربع كاي بدرجة حرية $\frac{(k-1)(p-1)}{2}$

$$Box's M = MC^{-1}$$

الجانب التطبيقي

جمعت البيانات لعينة عشوائية بسيطة تتكون من (150) مشاهدة من عدد من المستشفيات الحكومية لعام 2013 والتمثلة (مستشفى بغداد العام - مستشفى اليرموك التعليمي - مستشفى النعمان التعليمي - مستشفى الكاظمية التعليمي). تتضمن البيانات نوعين من المتغيرات الثنائية والمتصلة لثلاثة أنواع من المجموعات من التشوهات الخلقية .

وصف المتغيرات المستخدمة في البحث :

Y: (المتغير المعتمد) متغير الاستجابة ويمثل درجة التشوه ويقسم على :

تشوه بسيط ويرمز له 1

تشوه متوسط ويرمز له 2

تشوه خطير ويرمز له 3

المتغيرات التوضيحية تضم :

- عمر الام ويقاس بالسنوات ويرمز له (x_1)

- نوع فصيلة الدم (RH) للأُم (x_2) : وتم ترميزه Positive=0, Negative=1

- صلة القرابة بين الزوجين (x_3) وتم ترميزه : عند وجود صلة قرابة = 1 , عدم وجود صلة قرابة = 0

- عدد الاسقاطات السابقة ويرمز له (x_4)

- عدد الولادات الحية السابقة ويرمز له (x_5)

- عدد الولادات الميتة السابقة ويرمز له (x_6)

- التدخين (x_7) وتم ترميزه يدخن = 1 , لا يدخن = 0

كان توزيع مفردات العينة بين المجموعات ونسبة مساهمة كل مجموعة كما في جدول رقم (1)

جدول (1)

يبين توزيع مفردات العينة بين المجموعات

المجموعات	أسماء المجموعات	عدد الحالات	نسبة المساهمة
Groups	Group Name	No. of cases	Prior
1	تشوه بسيط	12	0.08
2	تشوه متوسط	75	0.50
3	تشوه خطير	63	0.42
المجموع		150	1.00

وباستخدام برنامج التطبيقات الاحصائية على sps حصلنا على مايلي .

وكانت المتوسطات الحسابية للمتغيرات التوضيحية موضحة في جدول رقم (2) موزعة حسب المجموعات .

جدول (2)

المتوسطات الحسابية الداخلة في عينة البحث للمجموعات الثلاث

متوسطات المتغيرات	المجموعات	$\bar{X}_1$	$\bar{X}_2$	$\bar{X}_3$	$\bar{X}_4$	$\bar{X}_5$	$\bar{X}_6$	$\bar{X}_7$
تشوه بسيط (1)	24.3333	0.2500	0.4167	1.9167	2.0000	0.5000	1.0000	
تشوه متوسط (2)	25.9733	0.4800	0.4933	1.9600	2.2267	0.5600	0.2267	
تشوه خطير (3)	25.06825	0.3968	0.4921	1.8413	2.2063	0.5397	0.1429	
	25.7200	0.4868	0.4867	1.9067	2.2000	0.5467	0.2533	

تم اختبار معنوية الدالة المميزة بالاعتماد على المعادلة (22) ومن خلال جدول رقم (3) كانت قيمة مربع كاي $\chi^2 = 53.006$ وحيث أن قيمة مربع كاي الجدولية بدرجة حرية (14) $(\chi^2_{(0.05,14)} = 23.69)$ وهذا يدل على أن الدالة لها امكانية جيدة على التمييز .

جدول (3)

يبين نتائج المقاييس المستخدمة في التحليل

Test of Function (s)	Wilks Lambda	Chi – Square	df	Sig
1 through 2	0.692	53.006	14	0.000
2	0.958	6.131	6	0.409

ولاختبار تساوي مصفوفة التباين والتباين المشترك للمجموعات وبالاعتماد على المعادلة (23) والمعادلة (24) فإن قيمة $Box's M = 50.645$.

حسبت القيمة اعلاه عند مستوى معنوية (0.011) ، القيمة الجدولية لمربع كاي بمستوى معنوية (0.011) ودرجة حرية $d.f = \frac{(7-1)(3-1)}{2} = 6$ هي : $\chi^2_{(6)} = 24.10$ وهذا يدل على عدم تساوي التباينات بين المجموعات ، وهذا يدل على عدم التجانس بين المجموعات .

ولاختبار معنوية المتغيرات :

ظهرت نتيجة اختبار المتغيرات كما في جدول رقم (4) .

جدول (4)

يبين اختبار F لكل متغير في الدالة المميزة الخطية

Tests of Equality of Group Mean

	Wilks Lambda	F	df ₁	df ₂	Sig
X1	0.995	0.353	2	147	0.703
X2	0.982	1.313	2	147	0.272
X3	0.998	0.126	2	147	0.882
X4	0.989	0.843	2	147	0.433
X5	0.991	0.131	2	147	0.878
X6	0.999	0.059	2	147	0.943
X7	0.735	26.469	2	147	0.000

نلاحظ من خلال الجدول رقم (4) ان المتغير x_7 (التدخين) يمتاز بمعنوية عالية وله تأثير كبير في عملية التصنيف، يليه المتغير x_2 (فصيلة الدم (RH) والمتغير x_4 (عدد الاسقاطات) ثم المتغير x_1 (عمر الام) والمتغير x_5 (عدد الولادات الحية) اما المتغير x_3 (صلة القرابة بين الزوجين) و x_6 (عدد الولادات الميتة) فليس لهما تأثير معنوي فكان تأثيرهما ضعيف جدا لان تأثيرهما قليل جدا في عملية التصنيف .

وتحتسب الدالة المميزة للمجموعات باستخدام نموذج Fisher للدالة المميزة وفق الصيغة الآتية .

$$Z = \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_p x_p$$

ويتم تقدير معاملات الدالة المميزة α وفق الصيغة التالية .

$$\alpha = s^{-1}(\bar{x}_{pi} - \bar{x}_{qi})$$

وبعد تقدير معاملات الدالة المميزة كانت أقيامها كما في جدول رقم (5) .

جدول (5)

يبين معاملات الدوال المميزة الخطية

Variables	Function (1)	Function (2)
X7	0.968	0.148
X3	0.066	0.045
X4	-0.034	0.675
X2	0.151	0.611
X1	0.100	0.201
X6	-0.033	0.125
X5	0.063	0.081

بما ان قيمة معيار (wilks lambda) للدالة الاولى هو (0.692) وللدالة الثانية (0.958)

يدل على القوة التمييزية للدالة الاولى لذلك سوف نعتمد على الدالة الاولى في تفسير أهمية المتغيرات الداخلة في الدالة وتأثير هذه المتغيرات في تحليل النتائج. نلاحظ ان المتغير x_7 (التدخين) ذو تأثير عال قياسا بالمتغيرات الاخرى، يليه المتغيرات x_2 (فصيلة الدم RH) و x_1 (عمر الام) ثم المتغير x_3 (صلة القرابة) ثم المتغير x_5 (الولادات الحية) ثم المتغير x_4 (عدد الاسقاطات) واخيرا المتغير x_6 (الولادات الحية) .

حساب احتمال التصنيف الصحيح

النسب المئوية للتصنيف الصحيح فهي موضحة في جدول رقم (6) .

جدول (6)

يبين التصنيف التنبؤي

Group	Predicted Group Membership			Total
	1	2	3	
Original Count				
1	10	2	0	12
2	7	46	22	75
3	4	33	26	63
%				
1	83.3	16.7	0.0	100.0
2	9.3	61.3	29.3	100.0
3	6.3	52.4	41.3	100.0
cross-validated count				
1				
2	6	6	0	12
3	9	42	24	75
	6	33	20	63
%				
1	50.0	50.0	0	100.0
2	12.0	56.0	32.0	100.0
3	9.5	52.4	38.1	100.0

نسبة التصنيف الصحيح في دالة التمييز الخطية كانت (54.7 %) حيث ان (82) مفردة من اصل (150) مفردة صنف بشكل صحيح الى المجاميع التي تنتمي اليها بينما (68) مفردة صنف بشكل خاطئ. وسوء التصنيف هذا حدث بسبب التشابه الكبير بين المتغيرات الموجودة في الدالة .

الاستنتاجات

- 1- من خلال النتائج التي توصلنا اليها فان هناك تأثير معنوي عالي للمتغير (x_7) الذي يمثل التدخين قياسا بالمتغيرات الاخرى، اي ان الام الحامل اذا كانت مدخنة فانه يؤثر بشكل كبير على صحة الجنين وتكوينه، حيث ان التدخين يؤدي الى ولادة اطفال مصابين بتشوهات خلقية مثل خلل بالمخ بنسبة تصل الى (25 %) واذا كانت الام مفرطة بالتدخين فتصل النسبة الى (65 %) وهي نسبة كبيرة جدا . ومن تأثيرات التدخين ولادة اطفال اصغر حجما من المعتاد يصل وزن الطفل الى (200) غم ومصابين بتشوهات بالرأس (صغر حجم الدماغ) وعيوب خلقية اخرى في سقف الحلق ومنطقة الكلام .
- 2- كذلك اظهرت نتائج التحليل ان المتغيرين (x_2) الذي يمثل فصيلة الدم RH و (x_4) الذي يمثل عدد الاسقاطات لها تأثير معنوي ولكن بشكل اقل من تأثير متغير التدخين .
- 3- المتغير (x_1) الذي يمثل عمر الام ايضا له تأثير معنوي ولكن بشكل اقل من تأثير متغير التدخين وفصيلة الدم RH وعدد الاسقاطات. حيث ان كبر عمر الام الحامل يضاعف من مخاطر اصابة جنينها بالتشوهات الخلقية ومرض التوحد ، فان الحمل بعد سن الاربعين عاما معرضا اكثر لزيادة نسبة التشوهات الخلقية مثل (متلازمة داون) (المنغوليين) وذلك لاضطراب عدد الكروموسومات حيث ان نسبة حدوث التشوهات في سن الاربعين (1) من كل (100) حمل اما عند سن الثلاثين (1) من كل (1000) حمل .
- 4- المتغير (x_6) الذي يمثل عدد الولادات الميتة والمتغير (x_3) الذي يمثل صله القرابة بين الزوجين ليس لهما تأثير معنوي في حالات التصنيف المدروسة .

النوصيات

- 1- ضرورة تسجيل المعلومات والبيانات الكاملة للام الحامل قبل الولادة وبعدها وضرورة الاحتفاظ بالبيانات السابقة لكل ولادة وبنائج التحليلات السابقة للاستفادة منها في مقارنة حالات الولادة وكذلك للاستفادة منها في اجراء دراسات مستقبلية ، حيث نلاحظ ان مثل هذه البيانات تكون غير كاملة المعلومات او غير موجودة في المستشفيات .
- 2- تجنب تعاطي الادوية للام الحامل خلال الاشهر الاولى من الحمل .
- 3- ننصح الام الحامل بعدم التدخين حيث ان التدخين يؤدي لولادة اطفال مصابين بتشوهات خلقية .
- 4- ننصح الام بتجنب الحمل بعد سن الاربعين عاما لان الام الحامل في هذا العمر تكون اكثر عرضة لزيادة نسبة التشوهات الخلقية .
- 5- الاهتمام بمراجعة العيادات الاستشارية للنسائية والتوليد ورعاية الحوامل واجراء الفحص الدوري والتحليل .

المصادر :

- 1- القصاب، موفق محمد ، 2001، استخدام الاسلوب التمييزي في تصنيف الحوامل طبقا لدرجة الخطورة ، مجلة تنمية الرافدين ، عدد 63، مجلة 23.
- 2- داود، عربية عبد الرحمن ، 2000 ، استخدام الداله المميزة لبيان تاثير الحمية في مرض السمنة ، مجلة تنمية الرافدين ، عدد 61، مجلد 22.
- 3- محمد ، سميرة محمد صالح ، 2003، استخدام التحليل المميز (التصنيفي) لتحديد اهم العوامل المؤثرة في تسرب ورسوب الطلبة في جميع المراحل الدراسية ، رسالة ماجستير ، قسم الاحصاء، كلية الادارة والاقتصاد ، جامعة السليمانية .
- 4- T.W. Anderson ,2003, "An Introduction to Multivariate statistical Analysis " ,third edition ,John Wileg of sons, Inc.
- 5- 5 – Campell N.Mcintosh,2002,For far and Arneils text book of pediatrics, vol.1.Appleton & Lange;ch.2,p.645.