

استخدام التحليل التمييزي وتطبيقه على ولادة الطفل الخديج في مستشفى السبعين العام للأمومة والطفولة أمانة العاصمة صنعاء

أمة اللطيف محمد علي الحمزي^{1*} و أماني محمد علي الحمزي²

¹ قسم الرياضيات، كلية التربية – صنعاء، جامعة صنعاء، اليمن
² قسم الأطفال، كلية الطب، جامعة صنعاء، اليمن

* الباحث الممثل: أمة اللطيف محمد علي الحمزي؛ البريد الإلكتروني: alhmzi2018@gmail.com

استلم في: 29 أكتوبر 2022 / قبل في: 31 يناير 2023 / نشر في 31 مارس 2023

المُلخَص

هدفت الدراسة الحالية إلى توظيف استخدام تحليل دالة التمييز لتصنيف بيانات الأطفال الخديج الجدد إلى أطفال خديج احياء أو اموات. استخدمت الدراسة أسلوب المنهج الوصفي التحليلي، تكونت الدراسة من عينة عشوائية من بيانات الأطفال الخديج بمرکز الإحصاء بمستشفى السبعين العام للأمومة والطفولة للأعوام 2018-2020م، و بلغ حجمها (130) طفل خديج، منهم (98) عاش و (32) توفي، واستخدم أسلوب تحليل الدالة التمييزية على البيانات بهدف الوصول إلى تصنيف الأطفال الخديج الجدد على احدى المجموعتين، و توصلت الدراسة لتحديد المتغيرات المستقلة الأكثر أهمية في تصنيف مجموعة الأطفال الخديج إلى احدى المجموعتين، وأولها أهمية متغير تدخين الأم الحامل و يليه في الأهمية متغير العمر الرحمي للطفل الخديج بالأسابيع ثم متغير أخذ المكملات الغذائية اثناء الحمل ثم بقية المتغيرات الستة المعتمدة في بناء دالة التمييز. كما اشارت الدراسة إلى قوة العلاقة بين المتغيرات الداخلة في التحليل حيث بلغ نسبة الارتباط القانوني لدالة التمييزية (0.85). نسبة التباين المفسر بواسطة دالة التمييز من خلال قيمة مربع آيتا بلغت (0.72) ويقابلها قيمة ذاتية تساوي (2.58)، كما اشارت النتائج إلى ان قيمة ولكس لمدا قد بلغت (0.280)، وقيمة كاي تربيع (159.33) وهي دالة احصائياً عند مستوى دلالة (0.05). وهذا يؤكد قدرة الدالة التمييزية على التمييز بين مجموعتي الأطفال الخديج، كما بلغت نسبة دقة التصنيف الصحيح (95.4%)، وهي نسبة عالية تؤكد على دقة التصنيف.

الكلمات المفتاحية: التحليل التمييزي، وبناء دالة التمييز، ودالة التصنيف.

1. المقدمة:

(Premature)، وتدل التسمية على عدم نضوج المولود ولذا فهناك عمليتين هامتين تتحكمان في النضوج عند الولادة هما: مدة الحمل ومعدل النمو (الوزن الطبيعي إذ يجب يكون 2.5 كجم أو أكثر) داخل الرحم (كرامر، 1999). وأسلوب التحليل التمييزي يستخدم للتصنيف والتنبؤ بالظواهر التي بها المتغير التابع ثنائي نوعي (الطفل الخديج حي_ميت). ولأن الطفل هو أساس بناء المجتمع الحاضر والمستقبل، فيكون لهذه الدراسة أهمية حيث تم ترشيح مجموعة من المتغيرات المستقلة (العوامل المؤثرة) للدارسة وذلك بعد استشارة أطباء أطفال استشاريين. وتوصلت الدراسة لتحديد المتغيرات المستقلة ذات الدلالة الإحصائية والبالغة تسعة متغيرات من بين خمسة عشر متغيراً مستقلاً، رشحت للدارسة، كما تم تحديد المتغيرات الأكثر أهمية في تصنيف مجموعة الأطفال الخديج إلى احدى المجموعتين، وبلغ عددها فقط ستة متغيرات من بين المتغيرات المستقلة التسعة الدالة احصائياً، والتي اعتمدت لبناء دالة التمييز. وقد اشارت نتائج الدراسة إلى قوة العلاقة بين المتغيرات الداخلة في التحليل حيث بلغ نسبة الارتباط القانوني لدالة التمييزية (0.85). كما ان قيمة ولكس لمدا بلغت (0.280)، وقيمة كاي تربيع بلغت (159.33) دالة احصائياً عند مستوى دلالة (0.05). وهذا يؤكد قدرة الدالة التمييزية على التمييز بين مجموعتي الأطفال الخديج، كما أظهرت النتائج دقة عالية في التصنيف فقد بلغت نسبة دقة التصنيف الصحيح (95.4%).

2.1 مشكلة الدراسة:

أن استخدام متغيرات مختلفة تخضع لرأي وظنون وخبرات الباحث تؤدي إلى نتائج مختلفة ومغايرة لنفس الفئة المستهدفة، وهنا محاولة لوضع معايير لتمييز مجموعتي الأطفال الخديج احياء والاموات، من خلال تحديد

يُعد البحث عن أسباب الظواهر وتفسيرها، من الاهداف التي يسعى لها العلم ويقاس تطوره بتطور تقنياته وأساليبه، ونتيجة لتعدد الظواهر و متغيراتها التي في اغلبها لا تخضع للقياس المباشر ولصعوبة التعبير عنها بصورة كمية، فيصبح من الصعب الوصول إلى نتائج وتفسيرها بطريقة مناسبة وفق شروط تتصف بالدقة والموضوعية، لذا ظهرت الحاجة لاستخدام أساليب إحصائية أكثر كفاية في معالجة البيانات وتحليل المتغيرات المتعددة بمختلف أنواعها، واستخلاص نتائج تصف الظاهرة المدروسة بدرجة كبيرة من الدقة، منها: تحليل التباين، والتحليل العاملي، وتحليل الانحدار والتحليل التمييزي الخطي وغير الخطي، ويعتبر التحليل التمييزي الخطي من الأساليب الإحصائية الأوسع استخداماً، وفكرته تعود لاقتراح العالم الإحصائي الإنجليزي (Karl Pearson) في عام 1920م، الذي اقترح مؤشر المسافة بين المجموعات، كما تم دراسة مؤشر المسافة من قبل العالم الهندي (C. Mahalanob) في عام 1930م، كما ظهرت أفكار التحليل التمييزي في مقال (Fisher) في عام 1936م من خلال ترجمة المسافة بين المجموعات إلى صورة خطية مركبة يستفاد منها في عملية التمييز. فالتحليل متعدد المتغيرات يُعد أحد الأساليب الإحصائية الهامة في المجالات الطبية والاقتصادية والاجتماعية وخصوصاً عندما يتعلق الأمر بعدد كبير من المتغيرات، ومن اهم الأساليب المستخدمة في المجال الطبي أسلوب التحليل التمييزي، [1]. وقد استخدم التحليل التمييزي في هذه الدراسة بغية التوصل إلى النموذج الذي يتكون من مجموعة متغيرات مستقلة (العوامل المؤثرة) صحية متنبئة (الطفل الخديج (Premature Baby) في مدينة صنعاء (الأمانة)، حيث عرقت منظمة الصحة العالمية الطفل الخديج بأنه المولود الحي لذي يولد قبل 37 أسبوع رحمي ويسمى (خديج –

د) الأدوات والحزم الإحصائية المستخدمة في التحليل:

استخدم نموذج الدالة التمييزية والدالة التصنيفية، وذلك باستخدام البرامج الإحصائية SPSS إصدار 18 وبرنامج Mathcad إصدار 2013.

الدراسات السابقة:

أجريت [2] دراسة أشارت إلى أن مشكلة تصنيف المجموعتين لها جوانب سلبية في عدد من المجالات مثل الاقتصاد والتمويل والطب، حيث استخدمت الدراسة ثلاثة نماذج للتصنيف، وقد استخدمت عينات محاكاة لمختلف التوزيعات وأحجام العينات وبينت نتائج الدراسة أن تحليل الدالة التمييزية ذو أفضلية في تصنيف البيانات. واستخدام [3] الدالة التمييزية الخطية لتمييز مرضى السكري المصابين من غير المصابين بالفشل الكلوي. تمت الدراسة في السودان، وهدفت إلى تحديد المتغيرات (العوامل) التي تؤدي للإصابة بالفشل الكلوي وتحديد أهمها في تمييز مرضى السكري المصابين بالفشل الكلوي من غير المصابين، وذلك باستخدام التحليل التمييزي واعتمد في ذلك على عدة متغيرات منها؛ (نسبة السكر في الدم، العمر، اليوريا، الكرياتينين) لمرضى السكري، واخذت عينه بأسلوب عشوائي، بحجم (200) مريض سكري منهم (100) مصابين بالفشل الكلوي والباقي غير مصاب بالفشل الكلوي، ونتجت الدراسة عن أن هناك فروق معنوية بين المتغيرات المستقلة في المجموعتين باستخدام اختبار F والذي يعني قدرة التحليل التمييزي على التصنيف، وأن هناك متغيرين لهما الأثر الأكبر في تمييز مرضى السكري المصابين بالفشل الكلوي وهما اليوريا الكرياتينين، وقد تم تصنيف (91%) من مفردات العينة مما يعني نسبة الدقة عالية في التصنيف ونسبة الخطأ صغيرة جداً (9%). ودراسة [4] هدفت إلى تحديد القدرة التمييزية لمقياس السلوك التكيفي للرابطة الأمريكية للتخلف العقلي لاستجابات عينة من التلاميذ العاديين والمتخلفين عقلياً بلغت 127 تلميذاً منهم 73 من الذكور و54 من الإناث كما بلغ عدد التلاميذ العاديين 264 تلميذاً منهم 138 من الذكور و126. واستخدم الصورة العربية السورية لمقياس السلوك التكيفي الأمريكية للتخلف العقلي، وظهرت النتائج أن هناك فروق دالة إحصائية بين التلاميذ العاديين والمتخلفين لصالح العاديين، بلغت دقة التصنيف في العاديين والمتخلفين عقلياً 100%، كما بلغت في عينة الذكور والإناث العاديين 60.2%، بينما بلغت في عينة الذكور والإناث المتخلفين عقلياً 60.6%. كما قدم [5] دراسة قامت بالمقارنة بين أسلوب تحليل الدالة التمييزية وتحليل الانحدار اللوغاريتمي الثنائي في الكشف عن قدرة بعض المتغيرات بالتنبؤ خلال فترة الحمل بعدم تعاطي الكحول والتي تمثلت في معارف وثقافة الحامل حول استمرار تعاطي الكحول والعمر للحامل أثناء الحمل، واخذت عينة عشوائية من مجموعة من الحوامل، واستخدم أسلوب تحليل الدالة التمييزية وتحليل انحدار الترجيح اللوغاريتمي لتحليل البيانات، وتوصلت الدراسة إلى أن أسلوب التحليل التمييزي كان أكثر دقة في تصنيف متعاطي الكحول وغير المتعاطي مقارنة مع أسلوب انحدار الترجيح اللوغاريتمي. كما تحدث [6] عن التحليل الإحصائي متعدد المتغيرات (التحليل التمييزي) في توزيع وتوصيف الأسر داخل الهيكل الاقتصادي الاجتماعي في المجتمع اليمني، و استخدم أسلوب التحليل التمييزي والذي يعد من الأساليب الإحصائية المتقدمة، واستخدمه الباحث لدراسة توزيع الأسر داخل الهيكل الاقتصادي والاجتماعي بهدف المساعدة في رسم الخطط الاقتصادية والاجتماعية التي ترسمها الدولة للوقوف على أنسب الطرق والأساليب لتحديد العبء الضريبي وعدالة توزيع الدخل والإعانات الحكومية للأسر في المجتمعات بصورة أكثر واقعية، وطبق الأسلوب من خلال عدة متغيرات حددها الباحث، لتحقيق أهداف البحث ومنها؛ توزيع الأسر داخل الهيكل الاقتصادي للمجتمع، تحديد المتغيرات التي لها دور في ذلك، تطبيق الأساليب الإحصائية الحديثة في الدراسات الاقتصادية. استخدم [7] الدالة التمييزية في تمييز الإصابة بمشكلات في القدمين لدى مرضى السكري، أجريت الدراسة في جامعة السودان للعلوم والتكنولوجيا، وتوصلت الدراسة باستخدام الدالة التمييزية إلى نموذج رياضي من خلاله يتمكن المصاب بالسكري من تحديد كونه مصاباً بمشاكل في القدمين أم لا، و طبق

المتغيرات المستقلة المؤثرة التي تتحكم بتصنيف الأطفال الخدج إلى إحدى المجموعتين السابقتين وذلك من خلال الاستفادة من أساليب التحليل الإحصائي متعدد المتغيرات ممثلاً بالتحليل التمييزي، لتوفير الجهد والوقت والمال.

3.1 تساؤلات الدراسة:

- هل البيانات والمعلومات المتوفرة عن الظاهرة محل الدراسة (متغيرات الأطفال الخدج الأحياء والأموات) ملائمة لأسلوب التحليل المستخدم؟
- ما أهم العوامل (المتغيرات) المؤثرة في تمييز الأطفال الخدج الأحياء والأموات؟
- ما الدالة التصنيفية التي تساعد في تصنيف الأطفال الخدج الجدد إلى أطفال خدج أحياء أو أموات؟

4.1 فروض الدراسة:

1. يوجد فروق دالة إحصائية بين متوسطات مجموعتي الأطفال الخدج الأحياء والأموات.
2. نموذج دالة التصنيف الذي يتم التوصل إليه باستخدام التحليل التمييزي يميز بدقة بين الأطفال الخدج الأحياء والأموات.

5.1 الهدف من الدراسة:

1. التعرف على معنوية الفروقات بين المتوسطات باستخدام اختبار (T^2 Hotelling) وعلى معنوية المعالم المستخدمة باستخدام اختبار (F).
2. تحديد المتغيرات المستقلة (العوامل) ذات الأهمية التصنيفية للدالة التمييزية التي لها دور في تمييز الأطفال الخدج الأحياء عن الأطفال الخدج الأموات.
3. استخدام أساليب التحليل الإحصائي متعدد المتغيرات ممثلاً في التحليل التمييزي والدالة التمييزية والتصنيفية لتتبع مجموعتي الأطفال الخدج الأحياء والأموات، وصولاً إلى تصنيف الأطفال الخدج الجدد على إحدى المجموعتين من خلال الدالة التصنيفية.

6.1 منهجية الدراسة:

أ) منهجية التحليل: بهدف تحقيق أهداف الدراسة استخدم الأسلوب الوصفي بواسطة جداول وملخصات رقمية تبين خصائص البيانات وتساعد في توضيح مدلول الدراسة، كما استخدم الأسلوب التحليلي الاستنتاجي الذي يعتمد على تطبيق التحليل التمييزي وتحليل التباين الوصول للدالة التمييزية التي تستخدم في عملية التصنيف والتنبؤ، وذلك بهدف تحقيق الأهداف المتوخاة لهذه الدراسة.

ب) مصادر البيانات: عينة من سجلات بيانات الأطفال الخدج بمركز الإحصاء بمستشفى السبعين العام للأمومة والطفولة للفترة من 2018-2020م.

ج) حدود الدراسة:

- الحدود البشرية: الأطفال الخدج.
- الحدود الزمانية: سجلات مركز الإحصاء بمستشفى السبعين العام للفترة من 2018-2020م.
- الحدود المكانية: مستشفى السبعين العام للأمومة والطفولة بأمانة العاصمة صنعاء.

النوع من الدوال يُعد من أبسط حالات التمييز، والتي تتطلب توفر الفروض الآتية:

1. توزيع المتغيرات المستقلة (التوضيحية) توزيعاً طبيعياً [10]

2. تساوي التباينات لكل المجموع (مصروفات التباين والتباين المشترك) أي قبول فرضية العدم عند اختبار الفرضية:

$$H_0 : \Sigma_1 = \Sigma_2 = \dots = \Sigma_k \quad VS \quad H_1 : \Sigma_1 \neq \Sigma_2 \neq \dots \neq \Sigma_k$$

حيث أن:

Σ : مصفوفة التباين والتباين المشترك، K: عدد المجموع. وتختبر باستخدام Box's M.

3. اختيار العينات بطريقة عشوائية وتصنيف مشاهداتها إلى مجموع n_1, n_2, \dots بشكل دقيق.

4. دالة التمييز خطية وستعرف لاحقاً.

1.3.2. صياغة التحليل التمييزي:

نفترض لدينا مجتمعين الأول نحدد له الرقم (1) والثاني الرقم (2)، تم اختبار مجموعتين (عينتين) حجمهما n_1, n_2 . من كلا المجتمعين السابقين وعلى الترتيب، وعلى افتراض أن لدينا قيم المشاهدات (المفردات) m من المتغيرات X_1, X_2, \dots, X_m والتي يُعتمد عليها بالتصنيف لتكوين دالة التمييز الخطية الآتية:

$$Y = \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m \quad (1)$$

حيث أن:

β : متجه معالم النموذج (معاملات التمييزية المتغيرات التفسيرية) تستخدم لعملية التصنيف، والمطلوب تقديرها من العينتين بشرط ان نحصل من هذه التقديرات على أقل احتمال لخطأ التصنيف [11].

X : متجه المتغيرات العشوائية.

m : عدد المتغيرات المستقلة.

2.3.2. إيجاد دالة التمييز [12]:

هناك عدة طرق لإيجاد دالة التمييز الخطية منها دالة الكثافة الاحتمالية لكل مجتمع أو باستخدام مضروب لاجرانج الذي استخدم لاستنتاج وبناء دالة التمييز الخطية، وذلك من خلال إيجاد قيم (تقديرات) β_i والتي تجعل التباين أقل ما يمكن، ويتم اختيار قيم المعاملات (β_i) لتعطي أعلى تمييز بين المجموعتين، أي أن الاختلاف بين قيم المجموعتين (Between group variation) أكبر من اختلافات القيم داخل المجموعتين (Within group variation)، فاذا رمزنا لنسبة الاختلافات بين المجموعتين إلى الاختلافات داخل المجموعتين بالرمز q [13].

$$q = \frac{\text{Between - group variation}}{\text{Within - group variation}} = \frac{(\bar{Y}^{(1)} - \bar{Y}^{(2)})^2}{\sum_i (Y_i^{(1)} - \bar{Y}^{(1)})^2 + \sum_j (Y_j^{(2)} - \bar{Y}^{(2)})^2} \quad (2)$$

أن اختيار قيم (β_i) يساعد على تكبير قيمة q [14].

حيث أن:

$\bar{Y}^{(1)}, \bar{Y}^{(2)}$: متوسط قيم المفردات للمجموعة الأولى والثانية على الترتيب.

$Y_j^{(2)}, Y_i^{(1)}$: قيم المشاهدة (i) و (j) في المجموعة الأولى والثانية، على الترتيب.

الأسلوب من خلال استخدام الباحث لعدة متغيرات منها؛ مدة الإصابة، مستوى الجلوكوز في الدم، العمر، الوزن، النوع، الضغط، ومعرفة مدى تأثير هذه المتغيرات على الإصابة بمشاكل القلب، وكان حجم العينة العشوائية التي اخذت للدراسة 154 مصابا بالسكري منهم 115 غير مصابين بمشاكل في القدمين 39 مصابين. وتوصلت الدراسة إلى ان هناك تأثير معنوي مباشر من قبل المتغيرات على كون مريض السكري مصاب بمشاكل في القدمين ام لا، وكان عامل مدة الإصابة أهم عامل يليه في الأهمية ضغط الدم، وأقلها أهمية متغير الوزن للمريض. وتناولت [8] بحثاً استخدام الدالة التمييزية الخطية في تصنيف المواليد من حيث الإصابة بالتشوه الخلقي.

2. الجانب النظري للتحليل التمييزي ودالة التمييز:

1.2 التحليل التمييزي Discriminate Analysis

التحليل التمييزي يُعد أحد العناصر والاساليب الإحصائية المهمة في التحليل المتعدد المتغيرات التي يمكن الاعتماد عليها في بناء قاعدة لإعادة توزيع المتغيرات (المشاهدات) وتصنيف المجموعات ضمن مجموعة من عدة مجموعات متداخلة تشترك في خصائص أو صفات معينة، لها الخصائص أو الصفات الفردية لهذه المشاهدة والتي يتم تحديدها مسبقاً، وتستخدم نوعين من المتغيرات في التحليل التمييزي: المتغيرات التابعة وهي نوعية مقاس بمقياس اسمي ذو طبيعة ثنائية أو متعددة، المتغيرات المستقلة (المؤثرة)، وتمثل الخصائص المميزة لكل مجموعة من المجموعتين الداخلتين في التحليل والتي من خلالها تنجح الدالة في التمييز بينهما، لتستخدم فيما بعد للتنبؤ وتصنيف الحالات المستقلة من العينة التي استخدمت لإيجاد الدالة. ولذا فإن عملية التصنيف تعتبر العملية اللاحقة لعملية تكوين الدالة التمييزية، إذ يُعتمد عليها في التنبؤ وتصنيف المشاهدة الجديدة لإحدى المجموعات باحتمال أقل لخطأ التصنيف. وطريقة إيجاد دالة التمييز تتم بإيجاد المعاملات للدالة وفقاً للقياسات أو المعايير التي يتم الحصول عليها من المشاهدات، وكلما كان حجم العينة كبير أعطت الدالة نتائج أفضل للتمييز. وللتحليل التمييزي ثلاثة أنواع: تحليل التمييزي المباشر يتم إدخال جميع المتغيرات مرة واحدة للتحليل، التحليل التمييزي الهرمي يتم إدخال المتغيرات طبقاً لجدول ينظمه الباحث، والتحليل التمييزي المتدرج وبه معيار احصائي يحدد من خلاله أولوية إدخال المتغيرات إلى النموذج، ومجالات استخدامه متعددة منها: العلوم البيولوجية والطبية والتربوية وعلم النفس والتمويل والتسويق، وفي حالة المجتمعات المتجانسة وغير المتجانسة. وهناك أنواع لدوال التمييز منها الخطية (وتسمى دالة فيشر نسبة إلى العالم الذي قام باشتقاقها) وتستخدم عندما العلاقة بين المتغيرات خطية (واستخدمت في هذه الدراسة)، والدالة الغير خطية تستخدم في حال كون المتغيرات غير خطية (تربيعية أو ذات درجة أعلى).

2.2 أهداف التحليل التمييزي

يهدف إلى تصميم تركيبة خطية للمتغيرات الأفضل بين مجموعات المتغير التابع موضوع الدراسة، بالإضافة إلى تحديد المتغيرات التي تسهم بأكثر قدر من التمييز ما بين مجموعات المتغير التابع، وكذلك التحقق من مدى وجود فروق احصائية بين المجموعات فيما يتعلق بالمتغيرات المستقلة (التوضيحية)، بالإضافة إلى تقييم دقة التصنيف كنسبة مئوية، كذا تقسيم المشاهدات بين مجموعة المتغير التابع بناءً على قيم المتغيرات المستقلة [9]، والهدف النهائي للتحليل التمييزي هو استخدام قاعدة التصنيف الكمي في تصنيف مشاهدة جديدة (مجهولة الانتماء) لأحد المجموعات المحددة والمعروفة مقدماً وذلك في ضوء قيم خصائصها، وبأقل احتمال خطأ تصنيف.

3.2 الجانب النظري الرياضي لدالة التمييز الخطية لمجموعتين

تُعد دالة التمييز الخطية نموذجاً رياضياً يصاغ من خلال مؤشرات عينة اختبرت مفرداتها بأسلوب عشوائي من مجتمعين مختلفين، حيث هذه الدالة تمكنا من تصنيف واختيار أية مفردة وتحديد المجتمع الذي تنتمي إليه. وهذا

$$\underline{S} = \frac{(n_1 - 1)\underline{S}^{(1)} + (n_2 - 1)\underline{S}^{(2)}}{n_1 + n_2 - 2} \quad (5)$$

والشكل الأخير لمصفوفة التباين والتباين المشترك يكون مصفوفة مربعة متماثلة، قطرها الرئيسي التباين وعناصرها الأخرى تمثل التباين المشترك كما بالمصفوفة الآتية:

$$S = \begin{bmatrix} s_{11} & s_{21} & \dots & s_{1k} \\ s_{21} & s_{22} & \dots & s_{2k} \\ \cdot & \cdot & \cdot & \cdot \\ s_{k1} & s_{k2} & \dots & s_{kk} \end{bmatrix} \quad (6)$$

وباستخدام المصفوفات تكون الصيغة الآتية: [15]

$$Y = X'S^{-1}(\bar{X}^{(1)} - \bar{X}^{(2)}) = X'\beta \quad (7)$$

وبذلك يمكن إيجاد $\bar{Y}^{(1)}, \bar{Y}^{(2)}$ كما يأتي:

$$\begin{aligned} \bar{Y}^{(1)} &= \bar{X}^{(1)'}\beta = \beta_1\bar{X}_1^{(1)} + \beta_2\bar{X}_2^{(1)} + \dots + \beta_m\bar{X}_m^{(1)} \\ \bar{Y}^{(2)} &= \bar{X}^{(2)'}\beta = \beta_1\bar{X}_1^{(2)} + \beta_2\bar{X}_2^{(2)} + \dots + \beta_m\bar{X}_m^{(2)} \end{aligned} \quad (8)$$

والاهتمام هنا بإيجاد $\beta_1, \beta_2, \dots, \beta_m$ التي تجعل قيمة q أكبر ما يمكن، وبالتعويض في قيمتها نحصل على:

$$\begin{aligned} \therefore (\bar{Y}^{(1)} - \bar{Y}^{(2)})^2 &= (\bar{X}^{(1)'}\beta - \bar{X}^{(2)'}\beta)^2 = (\beta'(\bar{X}^{(1)} - \bar{X}^{(2)}))'(\bar{X}^{(1)} - \bar{X}^{(2)})\beta \\ \therefore \sum_{i=1}^{n_1} (Y_i^{(1)} - \bar{Y}^{(1)})^2 + \sum_{j=1}^{n_2} (Y_j^{(2)} - \bar{Y}^{(2)})^2 &= \beta'(n_1 - 1)\underline{S}^{(1)}\beta + \beta'(n_2 - 1)\underline{S}^{(2)}\beta \end{aligned} \quad (9)$$

حيث أن $\underline{S}^{(1)}, \underline{S}^{(2)}$: مصفوفتا التباين داخل كلا المجموعتين. وبفرض أن \underline{S} تمثل التباين المشترك للعينتين معاً والمعرفة بالمعادلة (5):

$$\begin{aligned} \therefore (n_1 + n_2 - 2)\underline{S} &= (n_1 - 1)\underline{S}^{(1)} + (n_2 - 1)\underline{S}^{(2)} \\ \therefore \sum_{i=1}^{n_1} (Y_i^{(1)} - \bar{Y}^{(1)})^2 + \sum_{j=1}^{n_2} (Y_j^{(2)} - \bar{Y}^{(2)})^2 &= (n_1 + n_2 - 2)\beta'\underline{S}\beta \end{aligned} \quad (10)$$

وبالتعويض في المعادلة (2)

$$q = \frac{\beta'(\bar{X}^{(1)} - \bar{X}^{(2)})'(\bar{X}^{(1)} - \bar{X}^{(2)})\beta}{(n_1 + n_2 - 2)\beta'\underline{S}\beta} = \frac{1}{(n_1 + n_2 - 2)} q^* \quad (11)$$

حيث q^* هي النسبة المراد تكبيرها.

وعندما نفترض التباين داخل المجموعات مقدار ثابت، أي $\beta'\underline{S}\beta = 1$ وعملنا على تكبير التباين بين المجموعات، في هذه الحالة سوف نستخدم معامل مضروب لاجرنج فيصبح بالشكل الآتي:

$$\begin{aligned} L(\beta, a) &= \beta'(\bar{X}^{(1)} - \bar{X}^{(2)})'(\bar{X}^{(1)} - \bar{X}^{(2)})\beta - a(\beta'\underline{S}\beta - 1) \\ \therefore \frac{\partial L}{\partial \beta} &= 2(\bar{X}^{(1)} - \bar{X}^{(2)})'(\bar{X}^{(1)} - \bar{X}^{(2)})\beta - 2a\underline{S}\beta = 0 \\ 2(\bar{X}^{(1)} - \bar{X}^{(2)})'(\bar{X}^{(1)} - \bar{X}^{(2)})\beta - 2a\underline{S}\beta &= 0 \\ (\bar{X}^{(1)} - \bar{X}^{(2)})'(\bar{X}^{(1)} - \bar{X}^{(2)})\beta &= (\bar{X}^{(1)} - \bar{X}^{(2)})'D^2 \end{aligned}$$

حيث أن:

D^2 : مقياس (scalar) للمسافات بين مراكز المجموعات، وتدعى مسافة مهالونوبيس (Mahalanobis distance)، نسبة إلى العالم الهندي Mahalanobis (1930)، فكلما كبر قيمته دل ذلك على وجود اختلاف بين المجموعتين [16].

q : يمثل مقياس للبعد (المسافة) بين المجتمعين، فكلما كبرت قيمة q دل ذلك على وجود اختلاف بين المجتمعين. ودالة لتمييز الخطية تأخذ الصيغة (1).

3.3.2 تقدير معالم الدالة التمييزية: Estimation of the parameters discriminate function

ولتقدير معالم الدالة التمييزية نتبع الخطوات الآتية:

أولاً: نحسب متوسطات المتغيرات في كلاً من المجموعتين الأولى والثانية على الترتيب كما يأتي:

$$\bar{X}_1^{(1)} = \frac{\sum_{i=1}^{n_1} x_{1i}}{n_1}, \bar{X}_2^{(1)} = \frac{\sum_{i=1}^{n_1} x_{2i}}{n_1}, \dots, \bar{X}_k^{(1)} = \frac{\sum_{i=1}^{n_1} x_{ki}}{n_1}; \bar{X}_i^{(1)} \rightarrow i=1,2,\dots,k$$

$$\bar{X}_1^{(2)} = \frac{\sum_{i=1}^{n_2} x_{1i}}{n_2}, \bar{X}_2^{(2)} = \frac{\sum_{i=1}^{n_2} x_{2i}}{n_2}, \dots, \bar{X}_k^{(2)} = \frac{\sum_{i=1}^{n_2} x_{ki}}{n_2}; \bar{X}_i^{(2)} \rightarrow i=1,2,\dots,k$$

$\bar{X}_i^{(1)}$: تمثل متوسط المتغير (i) في المجموعة الأولى و $\bar{X}_i^{(2)}$: تمثل متوسط المتغير (i) في المجموعة الثانية.

ثانياً: نحسب الفرق بين متوسطي كل متغير كما يأتي:

$$d = \begin{bmatrix} d_1 \\ d_2 \\ \cdot \\ d_k \end{bmatrix} = \begin{bmatrix} \bar{X}_1^{(1)} - \bar{X}_1^{(2)} \\ \bar{X}_2^{(1)} - \bar{X}_2^{(2)} \\ \cdot \\ \bar{X}_k^{(1)} - \bar{X}_k^{(2)} \end{bmatrix} \text{ و } d_i = \bar{X}_i^{(1)} - \bar{X}_i^{(2)} \quad i=1,2,\dots,k \quad (3)$$

ثالثاً: نحسب مجموع مربعات كل متغير في كل مجموعة ومجموع حاصل ضرب كل متغيرين في كل مجموعة كما يأتي:

$$\begin{aligned} SS_{11}^1 &= \sum_{i=1}^{n_1} x_{1i}^2 - \frac{(\sum_{i=1}^{n_1} x_{1i})^2}{n_1}, \dots, SS_{kk}^1 = \sum_{i=1}^{n_1} x_{ki}^2 - \frac{(\sum_{i=1}^{n_1} x_{ki})^2}{n_1} \\ SS_{11}^2 &= \sum_{i=1}^{n_2} x_{1i}^2 - \frac{(\sum_{i=1}^{n_2} x_{1i})^2}{n_2}, \dots, SS_{kk}^2 = \sum_{i=1}^{n_2} x_{ki}^2 - \frac{(\sum_{i=1}^{n_2} x_{ki})^2}{n_2} \end{aligned}$$

حيث المعادلتان السابقتان تمثلان مجموع مربعات الأولى والأخيرة على الترتيب في المجموعة الأولى والثانية.

وبشكل عام

$$S_{ii}^1 = \sum_{i=1}^{n_1} x_{ii}^2 - \frac{(\sum_{i=1}^{n_1} x_{ii})^2}{n_1}, SS_{ii}^2 = \sum_{i=1}^{n_2} x_{ii}^2 - \frac{(\sum_{i=1}^{n_2} x_{ii})^2}{n_2}; i=1,2,\dots,n_1$$

والتي تمثل مجموع مربعات المتغير (i) في المجموعة الأولى والثانية.

ولحساب حاصل ضرب المتغيرين نستخدم الآتي:

$$S_{ij} = \sum x_{ij} - \frac{(\sum x_i)(\sum x_j)}{n}$$

رابعاً: إيجاد مصفوفة التباين والتباين (التغاير) المشترك للمجموعتين على الترتيب كما يأتي:

$$S_{ii} = \frac{S_i^{(1)} + S_i^{(2)}}{n_1 + n_2 - 2}, S_{ij} = \frac{S_{ij}^{(1)} + S_{ij}^{(2)}}{n_1 + n_2 - 2} \quad (4)$$

ويمكن إيجاد مصفوفة التباين والتباين المشترك داخل المجموعتين كما يأتي:

الإحصائية المتعلقة باتخاذ قرار حول المجموعتين الذي توزيع كل منهما طبيعياً متعدد المتغيرات ولهما مصفوفة التباين المشترك نفسها، ويعتبر المؤشر تطوير لاختبار (t) من مجموعة (عينة) أحادية إلى متعدد المتغيرات، حيث نستخدم اختبار (t) عند المقارنة بين متوسطي مجتمعين طبيعيين والذي يأخذ متغيراً واحداً (y)، ففي حالة مجموعتين والمطلوب التمييز بينهما بافتراض تساوي تباين المجموعتين المذكورتين سابقاً [17]. لاختبار الفرضية المعرفة بالمعادلة (16)، أي لمعرفة ما إذا كانا المجتمعان مختلفين فعلاً نستخدم المؤشر الإحصائي (T^2) قبل تطبيق التحليل التمييزي على مجتمعين طبيعيين متعدد المتغيرات لاختبار الفرضية في (18). ودالة الاختبار لذلك تُعرف بالمعادلة الآتية:

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} D^2 \quad (17)$$

نلاحظ مما سبق أن الإحصائية (T^2) Hotellin متناسبة مع الإحصائية (D^2) Mahalanobis وتعتمد عليها [12]. ونظراً لعدم توفر جداول لاستخراج القيمة الجدولية خاصة بهذا الغرض يمكن الحصول على قيمة (F) الجدولية مباشرة والذي وضعه (Rao) عام 1952، حيث يرتبط اختبار (F) بعلاقة مباشرة مع (T^2)، أي ان إحصائية (T^2) تعادل قيمة (F) من جدول تحليل التباين ويمكن تحويلها إلى قيمة لها توزيع (F) تقريبي، يأخذ الصيغة الآتية:

$$F_{cal} = \frac{MSB}{MSE} = \frac{n_1 + n_2 - m - 1}{(n_1 + n_2 - 2)m} * (T^2)$$

$$F_{tab}(\alpha, m - 1, n_1 + n_2 - m - 1)$$

والقيمة الجدولية:

القرار: في حال كانت قيمة (F_{cal}) أكبر من (F_{tab}) ترفض H_0 وتقبل H_a ، أي الدالة لها القدرة على التمييز.

2. تحليل التباين لاختبار معنوية الدالة التمييزية الخطية:

لاختبار الدالة التمييزية ومدى صلاحيتها وقدرتها على التمييز والتصنيف، يستخدم تحليل التباين والفرضية الخاصة بذلك [18]:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_m = 0 \quad VS \quad H_1: \beta_1 \neq \beta_2 \neq \dots \neq \beta_m \quad (18)$$

وباستخدام جدول تحليل التباين يتم قياس الاختلافات بين المجموعات وداخلها:

جدول تحليل التباين للتحليل التمييزي				
S	مجموع المربعات S.S	d.f	M.S	F
Between X,s بين المجموعات	$SSB = \frac{(n_1 n_2)(D^2)^2}{(n_1 + n_2)(n_1 + n_2 - 2)}$	$m - 1$	MSB	$F_{cal} = \frac{MSB}{MSE}$
Within X,s الخطأ	$SSE = D^2$ $= \hat{\beta}_1 d_1 + \hat{\beta}_2 d_2 + \dots + \hat{\beta}_m d_m$	$n_1 + n_2 - m$	MSE	$F_{tab}(\alpha, m - 1,$ $n_1 + n_2 - m)$
Total	$SST = SSB + SSE$	$n_1 + n_2 - 1$		

القرار: في حال كانت قيمة (F_{cal}) أكبر من (F_{tab})، ترفض H_0 وتقبل H_1 ، عند مستوى معنوية α ، وهذا يعني أن الدالة التمييزية لها القدرة على التمييز والمعاملات المستخدمة في التصنيف يمكن الاعتماد عليها في التصنيف والعكس صحيح.

$$D^2 = (\bar{X}^{(1)} - \bar{X}^{(2)})' \beta \quad (12)$$

$$\therefore (\bar{X}^{(1)} - \bar{X}^{(2)}) D^2 = a \underline{S} \beta$$

وبذلك يمكن إيجاد قيم تتناسب مع $\underline{S}^{-1}(\bar{X}^{(1)} - \bar{X}^{(2)})$ والتي تكبر التباين بين المجموعتين على شرط أن يكون التباين داخل المجموعتين ثابتاً، أي أن:

$$\hat{\beta} = \underline{S}^{-1}(\bar{X}^{(1)} - \bar{X}^{(2)}) = \underline{S}^{-1} d \quad (13)$$

حيث ان:

$(\bar{X}^{(1)} - \bar{X}^{(2)})$: متجه الفرق بين متوسطي (العينتين) المجموعتين الأولى والثانية.

\underline{S}^{-1} : معكوس مصفوفة التباين والتباين المشترك المقدر.

$\underline{S}^{-1}(\bar{X}^{(1)} - \bar{X}^{(2)})$: يمثل المعاملات المقدر $\hat{\beta}$ للدالة التمييزية. وبذلك فدالة التمييز الخطية المقدر تُعرف:

$$\hat{Y}_i = \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + \dots + \hat{\beta}_m X_{im} \quad ; i = 1, 2, \dots, n \quad (14)$$

4.3.2 اختبار معنوية الدالة التمييزية الخطية (قدرة الدالة على التمييز)

لمعرفة هل هناك فروق معنوية بين متوسطات المجاميع، نستخدم اختبار Hotelling (T^2) [17]، وللتعرف على معنوية المعالم المستخدمة باستخدام اختبار تحليل التباين (اختبار F). وذلك بهدف التمييز بين مجموعتين أو أكثر وتكوين دوال التمييز بمستوى معنوية مقبول إحصائياً.

5.3.2 اختبار معنوية الفروقات بين المتوسطات

يوجد ثلاثة أنواع من الاختبارات لاختبار الفرضية الآتية:

الدالة ليس لها القدرة على عملية التمييز: H_0 مقابل الدالة لها القدرة على عملية التمييز: H_1

$$H_0: \mu_1 = \mu_2 = \dots = \mu_m = 0 \quad VS \quad H_1: \mu_1 \neq \mu_2 \neq \dots \neq \mu_m \neq 0 \quad (16)$$

1. إحصائية (T^2) Hotelling

في عام 1931 قام Harold Hotelling بإيجاد المؤشر الإحصائي الذي سمي باسمه ورمز له بالرمز (T^2)، واستخدم المؤشر لمعالجة المشكلة

وهكذا إلى ان نصل إلى اقل المتغيرات المستقلة أهمية في عملية التمييز هو ذلك المتغير المقابل لأصغر معامل. [20].

9.3.2 النقطة الفاصلة (Cut Point=CP)

يمكن كتابة النقطة الفاصلة بين المجموعتين بالصورة الآتية:

$$CP = \frac{1}{2} (\bar{Y}^{(1)} + \bar{Y}^{(2)}) \quad (23)$$

للحصول على دالة التصنيف Y^* فإنه يمكن دمج نقطة الفصل CP مع الدالة التمييزية كالآتي:

$$Y^* = CP + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m \quad (24)$$

10.3.2 قاعدة التصنيف Classification Role

بالاعتماد على الدالة التصنيفية Y^* يتم تصنيف المشاهدات للمجموعات بحسب العلاقة الآتية:

$$Y^* \begin{cases} > 0 & \text{Group I} \\ < 0 & \text{Group II} \\ = 0 & \text{Can not classify} \end{cases} \quad (25)$$

11.3.2 احتمال خطأ التصنيف

عند استعمال دالة التمييز \hat{Y} واستخدام نقطة الفصل بين المجموعتين فقد ينتج تصنيف غير صحيح كتصنيف مشاهدة إلى مجموعة (n_1) بينما هي في الحقيقة تعود إلى المجموعة (n_2) وبالعكس، ولذا يترتب على ذلك نوعين من الأخطاء للتصنيف:

1. نسبة الخطأ الظاهري

لحساب نسبة الخطأ الظاهري الناتج من تصنيف مشاهدة هي في الأصل للمجموعة الأولى وتم تمييزها ضمن المجموعة الثانية يرمز له n_{21} ، والعكس n_{12} ، فيتم حساب النسب لكل منهما كالآتي [21]:

$$L_{21} = \frac{n_{21}}{n_2}, L_{12} = \frac{n_{12}}{n_1} \quad (26)$$

2. الخطأ الحقيقي وبحسب من الاحتمال الآتي:

$$L_{12} = L_{21} = \Phi\left(-\frac{\sqrt{D^2}}{2}\right) \quad (27)$$

حيث ان:

D^2 : إحصائية مهالونوبيس

Φ : دالة التوزيع الطبيعي ويمكن إيجاد قيمة هذا الاحتمال (L_{21} أو L_{12}) من جداول التوزيع الطبيعي. والدالة التمييزية التي تعطي أقل خطأ تصنيف تعتبر دالة ذات كفاءة، وهذا يدل على أهمية عامل خطأ التصنيف.

12.3.2 المتغيرات الصورية Dummy Variables

ان تحليل النماذج الخطية كدالة التمييز الخطية يتطلب ان تكون المتغيرات المستقلة (التفسيرية) كمية، لكن في واقع الامر قد نجد متغيرات نوعية تسهم في تفسير المتغير التابع، في دراستنا الحالية هناك متغيرات نوعية تؤثر على نوعية الطفل الخديج وفصله من حيث كونه خديج حي أو ميت، مثل نوع الولادة، اخذ المكملات الغذائية، تدخين الام وغيرها، ولإدخال تلك المتغيرات في نموذج التحليل يجب تحويلها الى متغيرات صورية تأخذ القيمة (1) إذا توفرت الصفة التي يعبر عنها المتغير النوعي والقيمة (0) إذا لم تتوفر تلك الصفة، وبذلك تتحول المتغيرات النوعية الى متغيرات كمية ذات فئات متساوية يمكن إدخالها في التحليل [15].

3. اختبار ويلكس لمدا Wilk's Lambda (Λ) ويستخدم لاختبار الفرضية في (17)، حيث القيمة المحسوبة:

$$\Lambda = \prod_{i=1}^m \frac{1}{1+\lambda_i} \quad (19)$$

و λ_i تمثل الجذر الكامن (eigenvalues) لكل المتغيرات، m عدد المتغيرات.

وقيمة لمدا تنحصر في الفترة $0 \leq \Lambda \leq 1$ ، فعندما $\Lambda = 1$ أو اقتربت منه دل ذلك على تساوي المتوسطات بين المجموعتين، أي عدم قدرة الدالة على التمييز، بنما في حال $\Lambda = 0$ أو اقتربت منه دل على عدم تساوي المتوسطات بين المجموعتين، أي ان الدالة لها القدرة على التمييز، وتستخدم إحصائية لمدا لاختبار معنوية المتغيرات المستقلة الداخلة في النموذج يتم إبقاء المتغيرات التي لها أدنى قيمة للمداء وأعلى قيمة ل-F.

6.3.2 اختبار معنوية كل متغير داخل في التحليل التمييزي [19]

إن العديد من المتغيرات التوضيحية التي تدخل في التحليل بأعداد كبيرة جداً، قد تكون غير معنوية فيتم استبعادها من التحليل مما يؤدي إلى نتائج جيدة في عملية التمييز. وذلك باختبار معنوية كل متغير توضيحي داخل في التحليل. وباستخدام اختبار (t) للمقارنة بين الأوساط الحسابية يمكن استخدامه لاختبار هذه المتغيرات، والفرضية الخاصة باختبار (t):

$$H_0: \mu_1 = \mu_2 \quad VS \quad H_a: \mu_1 \neq \mu_2 \quad (20)$$

ودالة الاختبار لذلك:

$$T = \frac{\bar{X}^{(1)} - \bar{X}^{(2)}}{S_p \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad (21)$$

وبمقارنة قيمة دالة الاختبار T في المعادلة (21) مع القيمة الجدولية $t_{(\alpha, n_1+n_2-2)}$ يتحدد كون المتغير معنوي أو غير ذلك، بمستوى معنوية (α) .

7.3.2 طريقة الانحدار المتدرج لاختيار المتغيرات Stepwise Regression Procedure

هناك عدة طرق تستخدم للتعرف على المتغيرات ذات القوة التمييزية المعنوية والتي تعطي أقل خطأ تصنيف، والطرق الإحصائية التي تساعد على الإدخال والحذف والإبقاء على المتغير المستقل المتميز ذو القيمة الأقل لإحصائية ويلكس لمدا والقيمة الأكبر ل-F، متعددة منها طريقة الانحدار المتدرج (Stepwise Method)، تستخدم في تحليل الانحدار لتحديد المتغيرات الأفضل والأكثر أهمية وتأثيراً في دالة الانحدار، وبالإمكان استخدامها في التحليل التمييزي باعتباره انحدار بمتغير تابع ثنائي نوعي وقد وضعت هذه الطريقة بواسطة (Efrogmson 1960).

8.3.2 الأهمية النسبية للمتغيرات المستقلة

الأهمية النسبية تساعد في معرفة وتحديد المتغيرات المستقلة التي تسهم وتؤثر في عملية التمييز والفصل بين المجموعات وترتيبها، ولترتيب كل متغير مستقل بحسب أهميته تطبيق المعادلة الآتية:

$$\beta'_i = \hat{\beta}_i \sqrt{S_{ii}} \quad (22)$$

حيث أن:

S_{ii} : تمثل التباين المستخرج من قطر مصفوفة التباين والتباين المشترك المتجمع S_p^2 .

وبإهمال الإشارة السالبة نتائج هذا المعامل يتم ترتيب القيم تصاعدياً أو تنازلياً بحيث أكبر المعاملات تعني أن المتغير المستقل المقابل له يعتبر أهم متغير في عملية التمييز يليه المتغير المستقل المقابل لثاني أكبر معامل

3. الجانب التطبيقي

الخصائص (المتغيرات) تخص طفل الخدج ماء، فإن أسلوب التحليل التمييزي سيحدد المعاملات التمييزية لكل من هذه الخصائص، وعند تطبيق المعاملات على المتغيرات الحقيقية يصبح لدينا أساس لتصنيف الأطفال الخدج الجدد على احدى المجموعتين من خلال دالة التصنيف.

المتغيرات المستقلة (المؤثرة) التي تم ترشيحها للدراسة من قبل أطباء الأطفال الاستشاريين موضحة بالجدول (1):

بعد جمع البيانات الخاص بالمتغيرات السابقة بأسلوب عشوائي من سجلات قسم الأطفال الخدج من وحدة الإحصاء بمستشفى السبعين العام بصنعاء للأعوام من 2018-2020م، تم تطبيق خطوات التحليل التمييزي الخطي لإيجاد أفضل مجموعة خطية من هذه المتغيرات (الخصائص) للتوصل إلى المعادلة التمييزية والتصنيفية والتي ستميز بين مجموعتين. وفي حال هذه

جدول (1): رمز وأسماء المتغيرات المستقلة المرشحة للدراسة

اسم المتغير	رمز المتغير	اسم المتغير	رمز المتغير
اخذ المكملات الغذائية أثناء الحمل	x_9	عمر الأم	x_1
نوع الولادة	x_{10}	العمر عند الزواج	x_2
وزن الطفل بالكيلوجرام	x_{11}	تدخين الام	x_3
عمر الطفل عند ادخاله الحضانه بالساعات	x_{12}	عدد مرات الحمل	x_4
فترة بقاء الطفل بالحضانه	x_{13}	عدد مرات الولادة	x_5
جنس الطفل	x_{14}	عدد مرات الإجهاض	x_6
أمراض الطفل الخديج	x_{15}	العمر الرحمي للطفل الخديج	x_7
		الفرق بين الحمل الحالي السابق بالأعوام	x_8

الامراض التي تصيب الطفل الخدج

يتعرض الطفل الخديج للعديد من الأمراض التي تصيبه باعتلال وزنه أو قد تؤدي بحياته ومنها ما موضح بالجدول (2):

جدول (2): الامراض التي قد تصيب الأطفال الخدج

اسم المرض	اسم المرض
Respiratory distress Tachypnea (RD I)	تسارع التنفس
Respiratory distress accessory muscle (RD II)	استخدام العضلات
Respiratory distress grunting (RD III)	اصدار صوت عند التنفس
Respiratory distress cyanosis (RD IV)	ازرقاق
Birth aphasics BA	الاختناق الولادي
Low birth weight L.B.W.	وزن الطفل ناقص
Very low birth weight V.L.B.W	وزن الطفل ناقص بشدة
Neonatal Sepsis	مولود مع التهاب الدم
Premature Rupture of membrane PRM	انفجار الغشاء المحيط بالطفل قبل موعد الولادة
Neonatal jaundice	صفار المواليد
Early neonatal sepsis ENS	التهاب وليدي مبكر
Past partum hemorrhage PP.	نزيف خلال الولادة

المصدر: أطباء الأطفال الاستشاريين

$$H_0 : \Sigma_0 = \Sigma_1 \quad VS \quad H_1 : \Sigma_0 \neq \Sigma_1$$

حيث:

Σ_0 تمثل تباين المتغيرات للأطفال الخدج الاحياء. و Σ_1 تمثل تباين

المتغيرات للأطفال الخدج الأموات

جدول (3): Log Determinants

حياة الطفل الخديج	Rank	Log Determinant
عاش	6	-.645
مات	6	-.927
Pooled within-groups	6	.284

المصدر: برنامج SPSS اصدار: 18

النتائج والمناقشة

1.3 التأكد من تحقق شروط تطبيق التحليل التمييزي

أولاً: اختبار التوزيع الطبيعي للعينة

نظراً لأن حجم العينة يزيد عن 30 مشاهدة وطبقاً لنظرية النهاية المركزية فإن البيانات تتبع التوزيع الطبيعي.

ثانياً: اختبار تساوي التباينات

باستخدام اختبار **Box's M** لتساوي مصفوفة التباين، نصيغ الفرضيات الاتية لاختبار تساوي التباينات للأطفال الخدج الاحياء والأطفال الخدج الأموات:

الجدول (5) والخاص باختبار ويلكس لامدا (ملخص لجدول تحليل التباين) بالنسبة لجميع المتغيرات المستقلة، فيوضح ما إذا كان هناك فروق معنوية بين متوسطات المتغيرات المستقلة كل على حده في مجموعتي المتغير التابع ((الخدج الاحياء والخدج الاموات) متغير التصنيف، فيبين ان متوسطات المجموعتين لجميع المتغيرات المستقلة ($X_2, X_3, X_6, X_7, X_8, X_9, X_{11}, X_{12}, X_{14}$) ذات دلالة (معنوية) احصائية (أقل من 0.05))، أي تحقق الفرضية H_1 ، أي ان هذه المتغيرات تؤثر في حالة الطفل الخدج، وعلى العكس بقية المتغيرات.

2.2.3 تحديد المتغيرات ذات الدلالة الإحصائية

باستخدام اختبار ويلكس لمدا F يتم تحديد المتغيرات المستقلة التي تمتلك دلالة (معنوية) إحصائية من بين (14) متغير.

جدول (6): المتغيرات المستقلة ذات الدلالة الإحصائية

رمز المتغير	ويلكس لمدا	F	Sig مستوى الدلالة
x_2	0.952	6.471	210.0
x_3	0.508	124.131	0.000
x_6	0.849	22.699	0.000
x_7	0.618	79.188	0.000
x_8	0.907	13.057	0.000
x_9	0.697	55.644	0.000
x_{11}	0.633	74.265	0.000
x_{12}	0.962	5.120	0.025
x_{14}	0.951	6.656	0.010

المصدر: تحليل الباحثة

يلاحظ من الجداول (6) أن المتغيرات المستقلة المعنوية التي يمكن استخدامها في التمييز بين المجتمعين تسعة متغيرات. لكن اختبار ويلكس لمدا و (F) لم يوضح أي المتغيرات المستقلة ذات التأثير المهم والإسهام الكبير في التمييز، وباستخدام التحليل التمييزي والأسلوب التدريجي (Stepwise Method) لاختيار المتغيرات الأفضل والأكثر تأثيراً وتمييزاً من بين متغيرات الأطفال الخدج تم الحصول على النتائج كما بالجدول (7):

الجدول (7): المتغيرات المستقلة والتي تمتلك أهمية وتمييز من بين المتغيرات التسعة المرشحة

رقم المتغير	رمز المتغير	اسم المتغير	إحصائية ويلكس لمدا	إحصائية F	درجات الحرية		مستوى الأهمية
					df1	df2	
1	x_3	تدخين الام	0.508	124.13	1	128	0.000
2	x_7	العمر الرحمي للطفل الخدج	0.398	96.18	1	128	0.000
3	x_{11}	وزن الطفل بالكيلوجرام	0.343	80.33	1	128	0.000
4	x_9	اخذ المكملات الغذائية اثناء الحمل	0.307	70.63	1	128	0.000
5	x_{12}	عمر الطفل عند ادخاله الحضانة بالساعات	0.290	60.72	1	128	0.000
6	x_6	عدد مرات الإجهاض	0.280	52.83	1	128	0.000

المصدر: تحليل الباحثة

يلاحظ من الجدول (3) ان قيم Log Determinant إلى حد ما متقاربة وهذا يعني تجانس متغيرات المجموعتين، كما يمثل عمود الرتبة عدد المتغيرات التي رشحت لتمثيل الدالة للتمييز والتنبيؤ وهي ستة متغيرات مستقلة من بين تسعة متغيرات دالة احصائياً.

جدول (4): نتائج اختبار Box's M لاختبار تجانس التباين

.Box's M	F		
	df2	df1	Approx.
31.576	12672.502	21	1.390

المصدر: تحليل الباحثة

يلاحظ من الجدول (4) أن قيمة (Sig = 0.110) أكبر من (0.05) وبالتالي قبول H_0 بتساوي مصفوفة التباين والتغاير للمجموعتين، أي تحقق افتراض تجانس التباين للمجموعتين.

2.3 بناء نموذج (دالة) التحليل التمييزي

1.2.3 اختبار تساوي متوسطات المتغيرات المستقلة للمجموعتين

لاختبار تساوي المتوسطات للأطفال الخدج الاحياء والأموات نختبر الفرضية (16):

جدول (5): الفروقات المعنوية بين متوسط المجموعات

رمز المتغير	الاحصائيات		درجات الحرية		مستوى الدلالة
	ويلكس لمدا	F	df11	df12	
x_1	0.986	1.828	1	128	0.179
x_2	0.952	6.471	1	128	0.012
x_3	0.508	124.13	1	128	0.000
x_4	1.000	0.063	1	128	0.803
x_5	0.994	0.724	1	128	0.397
x_6	0.849	22.699	1	128	0.000
x_7	0.618	79.188	1	128	0.000
x_8	0.907	13.057	1	128	0.000
x_9	0.697	55.644	1	128	0.000
x_{10}	0.981	2.460	1	128	0.119
x_{11}	0.633	74.265	1	128	0.000
x_{12}	0.962	5.120	1	128	0.025
x_{13}	1.000	0.005	1	128	0.946
x_{14}	0.951	6.656	1	128	0.011
x_{15}	0.991	1.109	1	128	0.294

المصدر: تحليل الباحثة

جدول (8): ترتيب المتغيرات المستقلة بحسب الأهمية النسبية

التسلسل	رمز المتغير	اسم المتغير	قيمة $\hat{\beta}$	قيمة معالم الدالة التمييزية بحسب الأهمية النسبية β'_i
1	X_3	تدخين الأم	-4.0293	-1.842
2	X_7	العمر الرحمي للطفل الخديج بالأسابيع	1.603	1.714
3	X_9	أخذ المكملات الغذائية أثناء الحمل	3.872	7851.
4	X_{11}	وزن الطفل عند الولادة بالكيلوجرام	6.816	1.554
5	X_{12}	عمر الطفل عند إدخاله الحضانة بالساعات	0.026	0.975
6	X_6	عدد مرات الإجهاض	-1.095	-0.833

المصدر: تحليل الباحثة

يلاحظ من الجدول (8) ان متغير (X_3) تدخين الام يعتبر المتغير الأول والاكثر أهمية في الدالة التمييزية في تحديد نوع الطفل الخديج من حيث كونه حي أو ميت بينما المتغير الذي يليه في الأهمية (X_7) العمر الرحمي للطفل الخديج بالأسابيع ثم (X_9) متغير اخذ المكملات الغذائية أثناء الحمل، ويلاحظ ان أقل المتغيرات أهمية في التمييز هو متغير عدد مرات الإجهاض.

4.2.3 اختبار معنوية الدالة التمييزية الخطية هناك عدة اختبارات:

1. اختبار ويلكس لمدا

جدول (9): اختبار دقة الدالة والمتغيرات التمييزية

الدالة	إحصائية ويلكس لمدا	كاي تربيع	درجة الحرية	مستوى الدالة (Sig)
1	0.280	159.33	6	0.000

المصدر: برنامج SPSS اصدار: 18

يتضح من الجدول (9) ان قيمة كاي تربيع دالة إحصائياً في المتغيرات المستقلة (المتنبئة)، حيث مستوى الدلالة ($00 = \text{Sig}$) أقل مع مستوى الدلالة المعتمد ($\alpha = 0.05$)، وقيمة إحصائية ويلكس لمدا وبالقيمة (0.280) التي تقترب من الصفر، كلاهما دليل على وجود فرق بين المتوسطات، وهذا يعني أن الدالة التمييزية لها القدرة على التمييز وتصنيف المشاهدات إلى مجتمعها الأصلي، وكذا التنبؤ بالأطفال الخديج الجدد الاحياء او الأموات.

جدول (10): الجذور الكامنة والارتباط الكلي القانوني

الدالة	الجذور الكامنة	نسبة التباين %	التباين النسبي المتجمع	الارتباط القانوني	مربع اينا معامل (التحديد)
1	2.578	100	100	0.849	0.721

المصدر: تحليل الباحثة

يلاحظ من الجدول (10) ان قيمة الجذر الكامن (2.58) وتشير إلى أن نسبة التباين المفسر بين مجموعتي الأطفال الخديج الاحياء والاموات، والتي تعود إلى الفروق بينها في دالة التمييز الوحيدة بينما بلغت قوة الارتباط القانوني الكلي (0.849) وهو ارتباط قوي موجب، ويعني وجود ملائمة بين المتغيرات المستقلة والمتغير التابع في تكوين هذه المجموعة، كما ان معامل التحديد بلغ (0.72) أي أن 72% من التباين يعود إلى الفرق بين المجموعتين في دالة التمييز.

يلاحظ من الجدول (7) أنه من بين المتغيرات المرشحة تم استبعاد ثلاثة متغيرات والإبقاء على ستة متغيرات الأكثر دلالة إحصائية ولها قدرة أعلى في التمييز والفصل بين مجموعتي الأطفال الخديج الاحياء والاموات، أي تمتلك أعلى قيمة (F) وأقل قيمة L ويلكس لمدا وتم الاختيار بناءً على الاختيار التدريجي على 6 مراحل.

3.2.3 إيجاد القيم التقديرية لمعامل الدالة التمييزية الخطية

تتبع الخطوات الآتية:

- إيجاد المسافة (d_i) بين المتغيرين وذلك بإيجاد الفرق بين متوسطي كل متغير بالمجموعتين كما في المعادلة (3):

$$d = \begin{pmatrix} -0.974 \\ -0.738 \\ 1.938 \\ -0.619 \\ 0.399 \\ 17.173 \end{pmatrix}$$

- إيجاد مصفوفة التباين والتباين المشترك للمجموعتين كما في المعادلة (6):

$$S = \begin{pmatrix} 0.184 & -0.023 & -0.009 & -0.021 & 0.460 & 0.015 \\ -0.023 & 1.144 & 0.015 & -0.006 & -1.950 & 0.021 \\ -0.009 & 0.015 & 0.052 & -0.003 & -0.930 & -0.016 \\ -0.021 & -0.006 & -0.003 & 0.166 & -2.404 & 0.019 \\ 0.460 & -1.950 & -0.930 & -2.404 & 1389.56 & -1.449 \\ 0.015 & 0.021 & -0.016 & 0.019 & -1.449 & 0.579 \end{pmatrix}$$

- إيجاد القيم التقديرية لمعامل الدالة التمييزية الخطية باستخدام المعادلة (13):

$$\hat{\beta} = \begin{pmatrix} -4.293 \\ 1.603 \\ 6.816 \\ 3.872 \\ 0.026 \\ -1.095 \end{pmatrix}$$

إن الدالة التمييزية التقديرية المعرفة بالمعادلة (14) تكتب بالصورة الآتية:

$$\hat{Y} = -4.293X_3 + 1.603X_7 + 6.816X_{11} + 3.872X_9 + 0.026X_{12} - 1.095X_6$$

ترتيب المتغيرات بحسب الأهمية وذلك باستخدام المعادلة (22).

$$\beta'_i = \begin{pmatrix} -1.841 \\ 1.714 \\ 1.554 \\ 1.578 \\ 0.975 \\ -0.833 \end{pmatrix}$$

جدول (12): نتائج تصنيف التحليل التمييزي

خط التصنيف الظاهري	ميت	عايش	المتغير التابع حالة الطفل الخديج	العدد
98	4	94	عايش	
32	30	2	ميت	
100	4.1	95.9	النسبة % عايش	
100.0	93.8	6.3	النسبة % ميت	

المصدر: تحليل الباحثة

يدل الجدول (12) على جودة التنبؤ بحالة الطفل الخديج من حيث كونه عايش ام ميت، فيلاحظ أنه 95.9% من الأطفال الخدج الاحياء قد صنفوا تصنيفاً صحيحاً، و93.8% من الأطفال الخدج الأموات الذين صنفوا تصنيفاً صحيحاً، بينما بلغت نسبة الخطأ الظاهري لتصنيف الأطفال الخدج الاحياء (4/98 = 4.1%) والأطفال الخدج الأموات (2/32 = 6.3%) ، كما ان المجموع الكلي المصنف تصنيفاً صحيحاً بلغ (124) من اصل (130) أي ما نسبته (95.4%)، أي أن دقة دالة التمييزي وقدرتها على التصنيف بشكل عام بلغ (95.4%)، وهي نسبة عالية جداً.

2. الخطأ الحقيقي

بتطبيق المعادلة (27) نحسب الخطأ الحقيقي ويساوي:

$$L_{12} = L_{21} = 0.467$$

8.2.3 الاستنتاجات والمقترحات

1. الاستنتاجات

نستنتج من خلال الدراسة أن:

1. هناك فروق بين المتوسطات، أي أن الدالة التمييزية لها القدرة على التمييز وتصنيف المشاهدات إلى مجتمعها الأصلي، ولها القدرة على التنبؤ بالأطفال الخدج الجدد الاحياء أو الأموات.
2. مجموعة المتغيرات المستقلة (المؤثرة) ذات الأهمية ولأولية في عملية التصنيف للأطفال الخدج إلى احياء أو اموات بلغ ستة متغيرات كما حددتها الدراسة وعلى راسها متغير تدخين الام ثم يليه في الأهمية العمر الرحمي للطفل الخديج ويليه متغير أخذ المكملات الغذائية اثناء الحمل.
3. دالة التصنيف يمكن الاعتماد عليها في تصنيف مجموعتي الأطفال الخدج الجدد إلى الاحياء والاموات حيث بلغت دقة التصنيف (95.4%)، وهي نسبة عالية جداً.

2. المقترحات

1. استخدام التحليل التمييزي في الدراسات الطبية بوصفه أسلوب احصائي متقدم.
2. استخدام التحليل التمييزي لتحديد العوامل المؤثرة على حياة الطفل الخديج.
3. إمكانية استخدام الدالة التمييزية التصنيفية التي استنتجتها في المستشفيات التي بها وحدات الأطفال الخدج لغرض تصنيف الأطفال الخدج الجدد إلى احياء أو اموات.
4. ابراز الدور الهام للتحليل الإحصائي في الجانب الطبي والاهتمام به.

2. إحصائية (Hotelling T^2)

يستخدم الإحصائية المعرفة في المعادلة (17) لاختبار الفرضية في (16):

$$D^2 = 13.659 \quad \therefore T^2 = 329.492$$

وباستخدام العلاقة المباشرة بين (T^2) و(F) والتي تأخذ الصيغة الآتية:

$$F_{cal} = \frac{n_1 + n_2 - m - 1}{(n_1 + n_2 - 2)m} * (T^2) = \frac{98 + 32 - 6 - 1}{(98 + 32 - 2)(6)} * (329.492) = 52.77$$

وباستخدام معنوية ($\alpha = 0.05$)، فإن القيمة الجدولية لتوزيع (F):

$$F_{tab}(\alpha, m - 1, n_1 + n_2 - m - 1) = F_{tab}(0.05, 5, 123) = 2.21$$

بما أن قيمة (F) المحسوبة أكبر من قيمة (F) الجدولية إذن نقبل الفرضية البديلة، والذي يدل على أن الدالة التمييزية الخطية ذات كفاءة في التمييز وأن الاختلافات معنوية بين المجموعتين.

3. تحليل التباين لاختبار معنوية معالم دالة التحليل التمييزي

يستخدم لاختبار الفرضية المعرفة في (20):

$$\therefore SSB = 35.16, \quad SSE = D^2 = 13.569$$

جدول (11): تحليل التباين للتحليل التمييزي

Source	S.S	d.f	M.S	F
Between X,s (بين المجموعات)	35.16	5	7.032	$F_{cal} = \frac{MSB}{MSE} = 64.51$ $F_{tab}(0.05, 5, 123) = 2.21$
Within X,s (الخطأ)	13.57	124	0.109	
Total	48.73	129		

المصدر: تحليل الباحثة

يلاحظ من الجدول (11) ان قيمة (F_{cal}) أكبر من قيمة (F_{tab})، وهذا يعني رفض H_0 ، أي أن دالة التمييز لها القدرة على التمييز وبذلك يمكن الاعتماد عليها لتصنيف أي مشاهدة جديدة إلى إحدى المجموعات.

5.2.3 نقطة الفصل

تحدد قيمتها باستخدام المعادلة (23):

$$CP = \frac{1}{2}(\bar{Y}^{(1)} + \bar{Y}^{(2)}) = \frac{21.1639}{2} = 10.582$$

6.2.3 دالة التصنيف

وبدمج قيمة نقطة الفصل بمعادلة دالة التمييز المقدره نحصل على دالة التصنيف المعرفة بالمعادلة (24):

$$Y^* = 10.582 - 4.293X_3 + 1.603X_7 + 6.816X_{11} + 3.872X_9 + 0.026X_{12} - 1.095X_{14}$$

بالاعتماد على الدالة التصنيفية Y^* يتم تصنيف المشاهدات للمجموعات بحسب المعادلة (25).

7.2.3 احتمال خطأ التصنيف

1. الخطأ الظاهري

باستخدام المعادلة (26) نحصل على أخطاء التصنيف الظاهري موضحاً بالجدول (12):

المصادر العربية:

- [15] علي بشر، المولى سليمان، "المقارنة بين التحليل التمييزي والنموذج اللوجستي الثنائي ونماذج الشبكات العصبية في تصنيف المشاهدات"، أطروحة دكتوراه، جامعة السودان للعلوم والتكنولوجيا، السودان، 2015.
- [16] R.D. Bock, "Multivariate in Statistical Methods in Behavioral Research", McGraw, Hill, Inc., USA, 1975.
- [17] D. F. Morrison "Multivariate Statistical Methods", 3rd Edition, McGraw-Hill Book Company, New York, 1976.
- [18] G. David, & L. Kupper, "Applied Regression Analysis and other multivariate methods", The University of North Carolina and Chapel Hill. 1978.
- [19] T.W. Anderson, "An Interdiction to Multivariate Statistical Analysis", 2nd Edition, JOHN WILEY & Sons, New York.1984.
- [20] آسيا محمد آدم، محمد، "استخدام الدالة التمييزية الخطية لتمييز الأطفال المصابين من غير المصابين بمرض السكري"، رسالة ماجستير، جامعة الجزيرة ود مدني، السودان، 2013. عبد الرضى حامد وبن جيلالي بو علام، عزام، "التحليل الاحصائي للمتغيرات المتعددة من الواجهة التطبيقية"، دار المريخ، الرياض، السعودية، 1997.
- [21] محمد أسامة، الكاتب، "الدالة المميزة الخطية في حالة أكثر من مجموعتين. مجلة التربية والعلم، 231 (4)، 157-169، 2010.
- [1] نورة، زبيري، "فعالية استخدام أسلوب التحليل التمييزي في تقدير مخاطر الائتمان-دراسة مجموعة من البنوك التجارية الجزائرية"، أطروحة دكتوراه، جامعة محمد بوضياف المسيلة. الجزائر، 2018.
- [2] F. H. Olmus, "A Comparison of Two Group Classification Approaches to Fat tailed and Skewed Data", Communication in Statistics- Simulation and Computation, no. 45, pp.17-32, 2016.
- [3] فريال، النويري، " استخدام الدالة التمييزية الخطية لتمييز مرضى السكري المصابين من غير المصابين بالفشل الكلوي"، رسالة ماجستير، جامعة السودان، 2013.
- [4] عبد الله، عبيد، " التحليل التمييزي لاستجابات عينة من التلاميذ العاديين والمتخلفين عقلياً على مقياس السلوك التكيفي للرابطة الامريكية للتخلف العقلي"، الجزء الأول. مجلة اتحاد الجامعات العربية للتربية وعلم النفس، بحث غير منشور، مجلة اتحاد الجامعات العربية، دمشق، سوريا، 2013.
- [5] Li, H. Sun, "Empirical Research of Hybridizing Principal Component Analysis with Multivariate Discriminant Analysis and Logistic Regression for Business Failure Prediction Expert Systems with Application", no.38, pp. 6244-56253 www.elsevier.com/locate/eswa, (2011).
- [6] فريد غانم، الجاعوني، "التحليل الاحصائي متعدد المتغيرات (التحليل التمييزي) لتوزيع وتوصيف الأسر داخل الهيكل الاقتصادي الاجتماعي في المجتمع"، مجلة جامعة دمشق للعلوم الاقتصادية والقانونية، 23، (2)، 2007.
- [7] هالة هاشم، فضل، " استخدام الدالة التمييزية وفقى تمييز الإصابة بمشكلات القدمين لدى مرضى السكري"، رسالة ماجستير، جامعة السودان، السودان، 2007.
- [8] عبير علي حسن، الناجي، "استخدام الدالة التمييزية الخطية في تصنيف المواليد من حيث الإصابة بالتشوهات الولادية"، رسالة ماجستير، الجامعة المستنصرية، العراق، 1999.
- [9] محفوظ، جودة، "التحليل الاحصائي باستخدام SPSS"، دار وائل للطباعة، الطبعة الأولى، عمان، الأردن، 2008.
- [10] A.A. Afifi & V. Clark, "Computer Aided Multivariate Analysis", Lifetime learning Publications, Belmont, California, USA. 1984.
- [11] أحمد علي، كنان، " فاعلية استخدام التحليل العنقودي والتحليل التمييزي في التحقق من الدلالة التمييزية لاختبارات الذكاء والشخصية"، جامعة دمشق، سوريا، 2015.
- [12] S.F. Arnold, "The Theory of Linear Models and Multivariate Analysis", John Wiley and Sons, New York, 1981.
- [13] خاشع محمود، الراوي، "المدخل إلى تحليل الانحدار"، مؤسسة دار الكتب للطباعة والنشر، جامعة الموصل، العراق، 1987.
- [14] T. Raykov, & G. A. Marcoulides, "An Introduction to Applied Multivariate Analysis", Taylor & Francis Group, LLC, New York, 2008.

RESEARCH ARTICLE

USING DISCRIMINATE ANALYSIS WITH APPLIED ON BIRTHS OF
PREMATURE BABY IN AL-SABEAN GENERAL HOSPITAL FOR
MATERNITY AND CHILDHOOD IN THE CAPITAL SANA'AAmtalteef M.A. Al-Hamzi^{1,*} & Amani M.A. Al-Hamzi²¹ Dept. of Mathematical, Faculty of Sana'a Education, Sana'a University, Yemen² Dept. of Pediatric, Faculty of Medicine, Sana'a University, Yemen*Corresponding author: Amtalteef M.A. Al-Hamzi; E-mail: alhmzi2018@gmail.com

Received: 29 October 2022 / Accepted: 31 January 2023 / Published online: 31 March 2023

Abstract

The current study aimed to employ the use of discriminant function analysis in classifying the new premature babies into living premature babies or dead premature babies. The study used the descriptive analytical method, the study consisted of a random sample of data on premature babies at the Statistics Center at Al-Sabeen General Hospital for Maternity and Childhood for the years 2018-2020, its size reached (130) premature babies, of whom (98) lived and (32) died. The method of analyzing the discriminatory function used on the data in order to reach the classification of the new preterm babies into one of the two groups. The study reached to identify the most important independent variables in classifying the group of premature babies into one of the two groups. The first in importance is the smoking variable of the pregnant mother, followed by the variable of uterine age of the premature baby in weeks, then the variable of taking nutritional supplements during pregnancy, then the rest of the six approved variables in constructing the function of discrimination. The study also indicated the strength of the relationship between the variables included in the analysis, where the percentage of legal correlation of the discriminatory function was (0.85). The percentage of variance explained by the discriminant function through the value of the eta-square was (0.72), and it corresponds to an eigenvalue equal to (2.58). The results also indicated that the value of Wilks' Lambda amounted to (0.280), and the value of chi-squared amounted to (159.33), which is statistically significant at the level of significance (0.05). This confirms the ability of the discriminatory function to distinguish between the two groups of premature babies. Amounted classification accuracy between the two groups of premature babies (95.4%), which is a high percentage that confirms the accuracy of the classification.

Keywords: Discriminant analysis, Building discriminative functions.

كيفية الاقتباس من هذا البحث:

أ.م. ع. الحمزي و أ.م. ع. الحمزي، "استخدام التحليل التمييزي وتطبيقه على ولادة الطفل الخديج في مستشفى السبعين العام للأمومة والطفولة أمانة العاصمة صنعاء"، مجلة جامعة عدن للإلكترونية للعلوم الأساسية والتطبيقية، المجلد 4، العدد 1، ص 1-12، مارس 2023. DOI: <https://doi.org/10.47372/ejua-ba.2023.1.214>

حقوق النشر © 2023 من قبل المؤلفين. المرخص لها EJUA، عدن، اليمن. هذه المقالة عبارة عن مقال مفتوح الوصول يتم توزيعه بموجب شروط وأحكام ترخيص (CC BY-NC 4.0) Creative Commons Attribution (CC BY-NC 4.0)

