دراسة مقارنة بين النموذج اللوجستي الثنائي وأسلوب التحليل التمييزي لتصنيف المرضى النساء المصابات باعتلال وظيفة الغدة الدرقية (دراسة ميدانية على مدينة بنغازي)

د. فائزة الخفيفى [1] ، أ. د عبد الغفار المنفي [2] ، صفاء بن غشير [3]
[1] جامعة بنغازي، بنغازي، ليبيا
[2] جامعة بنغازي ، بنغازي ، ليبيا
[3] وزارة التربية والتعليم ،بنغازي ،ليبيا
faizafarag2014@gmail.com[1]، abdstat@yahoo.com[2]، Safaaben725@gmail.com[3]

المستخلص

تتضمن هذه االدراسة المقارنة بين اسلوبين من الأساليب الاحصائية المستخدمة لتقدير معالم النماذج ذات المتغيرات التابعة ثنائية التصنيف فقد استخدمنا في هذه الدراسة أسلوب الانحدار اللوجستي الثنائي المتعدد وأسلوب التحليل التمييزي المتعدد بغرض المقارنة بين نتائجها. قد تم التطبيق العملي على بيانات تم الحصول عليها من اختيار عينة عشوائية عددها 285 مريض منهم 80 مريض لنوع نشاط الغدة و205 مريض لنوع كسل الغدة من النساء المصابات بالمرض بمدينة بنغازي وتم استخدام البرنامج الإحصائي (SPSS) لتحليل البيانات. توصلت الدراسة الى عدة نقاط أهمها ان النموذجين معنويين ويمكن استخدام اى منهما لتصنيف اى مريضة جديدة . كانت النتائج متقاربة في النموذجين حيث كانت العوامل المؤثرة في التصنيف هي الجرعة والهرمون T4 والهرمون TSH بالنسبة لنموذج الانحدار اللوجستي وكانت هذه المتغيرات أيضا من ضمن المتغيرات المسؤولة عن التميز في الدالة التميزية وكانت النتائج متقاربة في الأسلوبين غير أن أسلوب التحليل التمييزي ضم متغير واحد بالإضافة للمتغيرات الواردة في التحليل اللوجستي. بلغت نسبة التصنيف الصحيح للنموذج اللوجستي 98.2 %بينما بلغت نسبة التصنيف الصحيح للدالة التميزية 98.6 % .

مفتاح الكلمات : التحليل اللوجستي – الدالة التميزية – جدول التصنيف

Abstract

This comparison includes two statistical methods used to estimate the parameters of models with dependent variables of two classifications. We used in this study the logistic regression technique and the method of multiple discriminant analysis for the purpose of comparing their results. The practical application has been applied to data obtained from the selection of a random sample of 285 patients, including 80 patients for the type of gland activity and 205 patients for the type of laziness of the gland from women with the disease in the city of Benghazi, and the statistical program (SPSS) was used to analyze the data. The study reached several points, the most important of which is that

the two models are significant and either of them can be used to classify any new patient. The results were close to the two models, as the factors affecting the classification were the dose, the hormone T4 and the hormone TSH in relation to the logistical regression model, and these variables were also among the variables responsible for the discriminatory function. The results were close to the two methods, but the discriminant analysis method included one variable in addition to the variables contained in the Logistics analysis. The correct classification rate of the logistical model was 98.2 %, while the correct classification rate of the discriminatory function reached 98.6.%

Key Words: Logistics Analysis– Discriminatory Function– Classification rate

المقدمة

عند تحليل البيانات النوعية ثنائية التصنيف للمتغير التابع و ليكن Y فان هذا المتغير يأخذ القيمة واحد اذا توفرت صفة الظاهرة المراد دراستها والقيمة 0 في حالة عدم توفرها ومن اشهر الاساليب الاحصائية في هذه الحالة نموذج اللوجستي الثنائي (Binary Logistic Regression) و التحليل التمييزي (Discriminant Analysis) لتصنيف المفردة اعتمادا على قيم المتغيرات المستقلة المؤثرة في الظاهرة المدروسة (فضل2006). فى دراستنا هذه تم تطبيق الاسلوبيين للمفاضلة بينهما وابراز اهم العوامل المؤثرة في مرض اعتلال وظيفة الغدة الدرقية حيث يعتبر مرض اعتلال وظيفة الغدة الدرقية من الامراض الشائعة التي تصيب الانسان وبمجرد إفراز هرمونات الغدة الدرقية إلى الدم يبدأ الجسم بتنشيط التفاعلات الحيوية في الجسم وتقوم الغدة النخامية بالتحكم بإفرازات الغدة الدرقية من الهرمون ثيروكسين (T4-Thyroxine) ،ثلاثي يودوثيرونين (T3-Triiodothyonine) حيث إن الهرمون المسؤول عن تنظيم ذلك يسمى بهرمون TSH اختصاراً لـ(Thyroid stimulating hormone) ,والعنصر اليود أحد العناصر المعدنية الضرورية لجسم الإنسان كونه يدخل في تكوين هرمونات الغدة الدرقية (الثيروكسين) المسؤولة عن عمليات التمثيل الغذائي في الجسم[1] .وهناك بعض الحالات لا يشكو المريض من أي من هذه الأعراض ولكن عند التحليل تكون نسبة هرمونات الغدة الدرقية متأثرة أو ليست في المعدلات الهرمونية الطبيعية وخصوصاً يكون هناك ارتفاع في معدل الهرمون المحفز للغدة الدرقية والذي يسمى الهرمون TSH (الأغا 2010). فإذا كان هذا الهرمون مرتفعاً عن المعدل الطبيعي ولا يوجد هناك أعراض بالرغم من ذلك في بعض الحالات يفضل علاج هؤلاء المرضى بالهرمون الثيروكسين لان ارتفاع الهرمون TSH يدل على قصور في وظيفة الغدة الدرقية بصورة دقيقة ومع شرط وجود معدلات منخفضة او طبيعية من الهرمون T4 في نفس الوقت.

مشكلة الدراسةت

تمثل مشكلة هذه الدراسة في تحديد الاسلوب الأمثل لتحليل البيانات ذات المتغيرات التابعة النوعية

كمتغير تابع وفى دراستنا هنا المغير التابع هو نوع المرض (نشاط الغدة – كسل الغدة) وبعض العوامل المؤثرة في هذا المتغير مثل جرعة الدواء وبعض الهرمونات التي لها تأثير على المرض. مما سبق يكمن تحديد مشكلة الدراسة في التساؤل الآتي:

هل يعطى اسلوبي نموذج اللوجستي الثنائي (Analysis Binary Logistic Regression) والتحليل التمييزي (Discriminant Analysis) لتحليل المتغيرات التابعة النوعية ذات النتائج وبذات العدد من المتغيرات.

اهداف الدراسة

تكمن اهداف هذه الدراسة في

1- التعرف على ماهية أسلوب الانحدار اللوجستي وأسلوب التحليل التمييزي.
2- التطبيق العملي لهذه النماذج والمقارنة بين نتائجها باستخدام عينة الدراسة.
3- التعرف على اهم المتغيرات المؤثرة على مرض اعتلال وظيفة الغدة الدرقية وفق الاسلوبين.

فرضيات الدراسة

1- أسلوبي الانحدار اللوجستي و التحليل التمييزي يظهرا ذات النتائج .
2- لا يوجد اختلاف في المتغيرات المؤثرة في مرض اعتلال وظيفة الغدة وفقا للاسلوبين الإحصائيين المستخدمين.

مجتمع وعينة الدراسة:

تم تحديد مجتمع الدراسة بالنساء المصابات باعتلال وظيفة الغدة الدرقية في مدينة بنغازي، ونظرا لصعوبة اخذ جميع المرضى في كامل المدينة تم اختيار عينة عشوائية عددها 285 مريض حيث قسمت 80 مريض لنوع نشاط الغدة و205 مريض لنوع كسل الغدة من النساء المصابات بالمرض بالمدينة ،حيث سحبت عينة عشوائية منتظمة من ملفات المرضى المسجلين ، وتم أخذ العينة من اثنين من المستشفيات بالمدينة وهما: مستشفى الهواري العام وعيادة الكيش المجمعة.

متغيرات الدراسة: اولا: المتغير المعتمد (**المتغير التابع**) والذي يمثل نوع المرض متغير ثنائي الاستجابة يأخذ القيمة 1فى حالة نشاط الغدة والقيمة 0 في حالة كسل الغدة.

ثانيا: المتغيرات المستقلة: الوزن **(X_1)**، الجرعة المعطاة **(X_2)** ،الهرمون T4 **(X_3)** ، الهرمون TSH **(X_4)** ، نبضات القلب **(X_5)** ، قياس ضغط الدم (انقباضي) X_6 ، قياس ضغط الدم (انبساطي) X_7 ، الحالة الاجتماعية (X_8) و العمر(X_9).

الدراسات السابقة

كثرت الدراسات المستخدمة لاسلوب الانحدار اللوجستي الثنائي(Binary Logistic Regression) و التحليل التمييزي (Discriminant Analysis) كلا على حدا بينما قلت الدراسات التي تحاول المفاضلة بينهما. فيما يلي بعض من هذه الدراسات:

1. دراسة (على ابشر واخرون 2016) استخدمت الانحدار اللوجستي الثنائي و التحليل التمييزي لدراسة اهم العوامل المؤثرة في كفاية دخل الاسرة بولاية الخرطوم وتوصلت ان النموذج اللوجستي افضل من الدالة التميزية في تحليل بيانات الدراسة.
2. دراسة (علي، شمس الدين،2014) بعنوان (استخدام الدالة التمييزية الخطية لدراسة مستوى الإصابة بسرطان الغدة الدرقية) وهدفت الى استخدام الدالة التمييزية الخطية للتوصل إلى نموذج إحصائي يمكن من خلاله تصنيف الأشخاص المصابين وغير المصابين بمرض سرطان الغدة الدرقية، وذلك اعتمادا على المتغيرات (الهرمون TSH ،الهرمون T3، الهرمون T4، العمر والوزن) ومعرفة مدى تأثير هذه المتغيرات على الإصابة بسرطان الغدة الدرقية وبعد تطبيق الدالة التمييزية توصلت الدراسة إلى حقيقة أن هنالك تأثير معنوي من قبل المتغيرات المستقلة (الهرمون TSH،الهرمون T3 ،الهرمون T4 ،العمر والوزن) ووجد أن الهرمون T4 هو أكثر المتغيرات تأثيراً على نسبة الإصابة بسرطان الغدة الدرقية ويليه الهرمون T3 أما أقل المتغيرات تأثيراً هو الوزن وتم استخدام اختبار T للعينتين المستقلتين ووجد أنه توجد فروق ذات دلالة إحصائية بين المجموعتين، أي أن الدالة التمييزية التي توصلت إليها الدراسة لها القدرة على تصنيف الأفراد إلى مصابين وغير مصابين بسرطان الغدة الدرقية.
3. دارسة (لفاطمة أحمد حسن؛ مناهل سيد احمد, 2014) لمرض فقر الدم لدى الأطفال, تم تطبيق نموذج الانحدار اللوجستي , و تم التوصل إلى النتائج وملاحظة أن كلما قل الهيموجلوبين عن المعدل الطبيعي فهذا مؤشر يدل على احتمال وجود المرض أو عرضه للإصابة به وكذلك بالنسبة لكريات الدم البيضاء كلما زادت أو قلت عن المعدل الطبيعي لها دل ذلك على احتمال الإصابة بالمرض وينطبق نفس الشيء

بالنسبة للصفائح الدموية إي كلما قلت عن المعدل الطبيعي لها دل ذلك على احتمال الإصابة أو العرضة للإصابة بالمرض .

منهجية البحث

اولا الانحدار اللوجستي الثنائي(Binary Logistic Regression)

يعد الانحدار اللوجستي من الأساليب الإحصائية المهمة التي تقوم بتحليل العلاقة بين مجموعة من المتغيرات المستقلة ومتغير تابع يكون تصنيفيا ، أو رتبيا من خلال تقدير احتمال وقوع حدث من عدمه . في الانحدار اللوجستي الثنائي المتغير العشوائي التابع وليكن Y يكون متغير ثنائي حيث يأخذ القيمة واحد باحتمال p أذا تحقق حدث معين والقيمة صفر إذا لم يتحقق ذلك الحدث باحتمال 1-p .وهذا متغير عشوائي يسمى متغير برنولي (Kleinbaum et al 2000) و(1998chap) و له التوزيع:

$$\mathrm{p_r}(Y = y, p) = p^{y}\ (1 - p)^{1-y}\ ; y = 0,1 \qquad (1)$$

حيث:

$$p_i = \mathrm{p}(\mathrm{Y_i} = 1) and \quad 1 - p_i = p(Y_i = 0) \qquad (2)$$

بافتراض وجود عينة حجمها n .. و k من المتغيرات المستقلة $x_1, x_2, x_3 \ldots, x_k$ قد تكون كمية أو وصفية أو خليط من الاثنين.

تحدد العلاقة بين المتغير التابع الثنائي والمتغير المستقل كالأتي :

$$p_{\mathrm{i}} = \frac{e^{xi\beta}}{1 + e^{xi\beta}}$$

$$and \qquad 1 - p_{\mathrm{i}} = \frac{1}{1 + e^{xi\beta}} \qquad (3)$$

حيث : (1) $(k + 1)$ β متجه المعالم و $X_{\mathrm{n(k+1)}}$ متجه المتغيرات المستقلة النسبة بين p_{i} و 1_p_i تسمى نسبة الترجيح (*studenmund* 2001) وتعطى بالصيغة التالية :

$$odds = \frac{p_i}{1 - p_i} = e^{xi\beta} \qquad (4)$$

وتسمى المعادلة السابقة أعلاه بالدالة اللوجستية الثنائية (إبراهيم,2011) وتعرف بأنها احتمال وقوع حدث معين على احتمال وقوع الحدث المكمل له.

ومن خصائص الدالة اللوجستية أنه يمكن تحويلها لدالة خطية وبأخذ اللوغاريتم الطبيعي لنسبة الترجيح (اللوغاريتم الطبيعي لدالة الأرجحية) نحصل على :

(5)

$$\ln odds = \ln\left\{\frac{p_i}{1 - p_i}\right\} = e^{xi\beta}$$

وبهذا نحصل على علاقة خطية تربط بين المتغيرات المستقلة و لوغاريتم معامل الترجيح للمتغير التابع وليس في المتغير التابع نفسه كما هو الحال في تحليل الانحدار الخطي (Garson,2006).

تقدير واختبار معالم نموذج الانحدار اللوجستي الثنائي:

الطريقة المستخدمة في تقدير معالم النموذج الانحدار اللوجستي هي طريقة الإمكان الأعظم Maximum (Likelihood Method) وهي الطريقة الأكثر شيوعا, تعتمد على تكرار العمليات الحسابية عدة مرات حتى يتم الوصول إلى أفضل تقدير.(Sharma,1996) (EL.Khaffiai2002). يستخدم إحصائي الاختبار Wald Statisticلاختبار معنوية معالم المتغيرات المستقلة ومدي تأثيرها على النموذج .

ويتم حساب إحصاءه Wald كالتالي :

$$W_I = \frac{\beta_I^2}{SE_{\beta_I}{}^2} \qquad (6)$$

و حيث تتبع إحصاءه Wald توزيع مربع كاي χ^2 بفرض H_0:β=0

إذا كانت إحصاءه Wald ذات دلالة إحصائية، فإن ذلك يعني رفض الفرضية الصفرية القائلة بأن تأثير معامل الانحدار اللوجستي المرتبط بالمتغير المستقل يساوي صفر, وبذلك يكون المتغير المستقل X له تأثير في التنبؤ

بقيمة المتغير التابع Y أما إذا كانت إحصاءه Waldغير دال إحصائيا، فإن المتغير المستقل X ليس له تأثير في التنبؤ بقيمة المتغيرً التابع Y ،مما يعني إمكانية حذفه من النموذج لان ليس له أي دلالة إحصائية (2002,Menard).

جودة التوافق للنموذج :

هناك عدة طرق لتقييم معنوية النموذج تم استخدام

1- اختبار ($\chi 2$ model)

أن هذا الاختبار هو أحد مقاييس جودة المطابقة Goodness of- fit ومن الطرق المستخدمة في اختبار ملائمة النموذج يستخدم هذا الاختبار لجودة المطابقة لمعرفة ما أذا كان النموذج يمثل البيانات تمثيلا جدا أم لا.

2- *اختبار هوزمر-ليمشو Hosmer-lemeshow test*

يستخدم هذا الاختبار ايضا لمعرفة مدى مطابقة ما أذا كان النموذج يمثل البيانات تمثيلا جدا أم لا, وذلك باختبار الفرضية الآتية: H_0 :أذا كان النموذج مطبق للبيانات البحث و H_1 :إذا كان النموذج غير مطابق للبيانات البحث وتقارن مع القيمة الجد ولية لمربع كاي مع درجات الحرية المتساوية, فإذا كانت القيمة المحسوبة أقل من أو تساوي الجدولية هذا يعني أننا نقبل فرض العدم أي أن النموذج مطابق , أما إذا كانت اكبر فهذا يعنى رفض الفرضية, لتحديد تأثير متغيرات مستقلة ومقارنة هذا نموذج مع أي نموذج أخر.(Hosmer ,D.W., Lemeshow 2000).

3- طريقة أخرى لجودة التوافيق هي جداول التصنيف 2*2 هو جدول يوضح عدد احتمالات مصنفة في مجموعتين, حالات الظاهرة المشاهدة والتي لا تمتلك صفة وعدد حالات المشاهدة التي تمتلك الصفة مقابل عدد الحالات المتوقعة التي تملك صفة وعدد الحالات المتوقعة التي لا تملك صفة, ويوضح الجدول عدد الحالات التي تم تصنفيها بشكل الصحيح والحالات المصنفة بشكل غير صحيح, وتعتمد فكرة استخدامه لو أن معدل التصنيف كانت نسبته من 50% ما فوق هذا يعني بان النموذج مثل البيانات بشكل جيد.(Ferrer & Wang, 1999).

الانحدار التدرجي (الادخال الامامي):

عندما يكون عدد المتغيرات المستقلة كبيرا في هذه الحالة ,نريد تحديد مجموعة من المتغيرات التي له تأثير على المتغير التابع والانحدار اللوجستي له عدة طرق متاحة لبناء النموذج ,نستخدم الانحدار التدريجي فيمكن أن يكون الإدخال للأمام (forward) أو للخلف ,وهنا في دراستنا اختارنا التدرج الأمامي لبناء النموذج ,ونستخدم التدرج الأمامي بناءا على الإحصائية likelihood-ratio(LR) والتي تعتبر أفضل من إحصاء (Wald) لتحديد المتغيرات التي يجب حذفها وإزالتها من النموذج (Norusis,1993). حيث التغير في الإحصاءات (-2ln LR) يستخدم لاختبار الفرضية الصفرية , يبدأ التدرج الأمامي بأهم المتغيرات المستقلة في النموذج, ثم يضيف المتغيرات في كل خطوة ,حيث ندخل أولا المتغير الثابت b_0 ثم المتغير الذي له أعلى قيمة (maximum ln likelihood) بعد تحديد المتغير الثابت لإدخالها في النموذج., وفي كل خطوة يتم إضافة المتغير الذي لديه ارتباط مع المتغير التابع و والمتغير الذي له أعلى (ln likelihood) و هكذا حتى نصل إلى الحد الذي لا تأتي عنه أي زيادة في معامل الارتباط .

ثانيا التحليل التمييزي (Discriminant Analysis)

هو أحد الأساليب الإحصائية التي تستخدم لتحليل المتعددة التي تهتم بالتصنيف وتمييز بين مجموعتين أو أكثر بناء على المفردات المجتمعات المدروسة لإعداد دالة التمييز، ويعتمد التحليل التمييز على أن يكون المتغير التابع مقاسه بمقياس وصفي، كما يستخدم في التصنيف أو التنبؤ وذلك من خلال التعرف على المتغيرات المستقلة التي تساهم في عملية التصنيف وهيا العملية اللاحقة لعملية تكوين الدالة التمييزية إذا يتم الاعتماد على هذه الدالة بالتنبؤ والتصنيف المفردة الجديدة لإحدى المجموعات قيد الدراسة بأقل خطأ تصنيف ممكن(يعقوب 2017). ويهدف التحليل التمييزي إلى بناء نموذج رياضي لتوصل إلى تصنيف عدة مفردات إلى مجتمع معين من عدة مجتمعات وبناء قاعدة يمكنها المساعدة في تحديد المجتمع الذي تنتمي إليه مفردات جديدة مستقبلاً (حسينة 2018) فمثلا يمكن تصنيف المرضى النساء المصابات باعتلال وظيفة الغدة الدرقية إلى (نشاط الغدة وكسل الغدة) ،حسب بعض المتغيرات الطبية والاجتماعية –كالعمر ،الوزن ،الجرعة المعطاة، الهرمون T4 ،الهرمون TSH ،الهرمون T3،ضغط الدم (الانقباضي ،الانبساطي) ،نبضات القلب ،الحالة الاجتماعية ،عدد الأطفال ،عمر أصغر الطفل -. وينشى التحليل التمييزي دالة لتصنيف المرضى في إحدى المجموعتين ويمكن بعدها تطبيق تلك الدالة على

مشاهدات جديدة، أي هل تنتمى هذه المفردة الجديدة إلى مجموعة نشاط الغدة أو كسل الغدة. ومن الجدير بالذكر أن لدينا مجموعتين كنوع الغدة كمتغير تابع ،فإن التحليل التمييزي سينشئ دالة تمييزية واحدة، وإذا كان لدينا أكثر من مجموعتين في المتغير التابع فأنه يكون لدينا أكثر من دالة التمييزية بمعنى : (عدد الدوال التمييز= عدد المتغيرات التابعة _1) ، وتكون صيغة الدالة التمييزية والتي هي عبارة عن تركيبة خطية من المتغيرات المستقلة وتكتب كالآتي:

$$L = \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \ldots + \alpha_k X_k \qquad (7)$$

حيث أن:

L: الدالة التمييزية.

α: المعاملات التمييزية.

X: علامات الاستجابة للمتغير أي هي متغيرات توضيحية للدالة التمييزية وعددها k.

k: عدد المتغيرات المستقلة.

التحليل التدريجي التصنيفي Stepwise Discriminant Analysis (SDA)

وهو النوع الأكثر استخداماً و من أكثر الطرق شيوعاً ،ويشمل كل المتغيرات المستقلة في التحليل ويحسب قدرتها على التمييز بين المجموعات إذا يكون فيه ترتيب إضافة المتغيرات المستقلة إلى التحليل واستبعاده منها تبعاً لمعايير إحصائية بمعنى يتم تحديد معيار إحصائي يحدد أولوية إدخال المتغيرات إلى النموذج ولكن المشكلة في هذا التحليل هو التحيز الذي تسببه هذه الطريقة في الاختبارات المعنوية من حيث عدد المتغيرات المستقلة وتعتبر هذه الطريقة مفيدة في حالة البيانات التي لا تتبع التوزيع الطبيعي ومصفوفات التباين الغير متساوية، وأكثر هذه الاختبارات استخداماً هو إحصاءة((Wilks Lambda(Λ)) حيث تستخدم توزيع F في الحكم على معنوية التغير في قيمة الإحصاءة (Λ) الناتجة من إضافة أو حذف أي من المتغيرات المستقلة إلى التحليل وفي كل خطوة من الخطوات يضاف إلى التحليل ذلك المتغير الذي يتمتع بأكبر قيمة لدالة الاختبار F (F TO ENTER) ويتم تكرار عملية إضافة المتغيرات إلى التحليل لحين عدم بقاء أي متغير يتمتع بقيمة لدالة الاختبار F أكبر من القيمة الحرجة لهذا الاختبار عند مستوى معنوية محددة مسبقاً ،وفي نفس الوقت يتم فحص كل المتغيرات المستقلة التي تمت إضافتها في الخطوات السابقة وتحذف تلك المتغيرات التي لم تعد تساهم في تعظيم الدقة فيتمييز الفئات الصحيحة التي تنتمي إليها مفردات البحث نظراً لأن المتغيرات التي أضيفت للتحليل مؤخراً قد سلبتها هذا الدور، وهذا يتم عندما تنخفض قيمة دالة الاختبار F (F TO REMOVE) إلى مستوى أقل من القيمة الحرجة لهذا الاختبار عند مستوى المعنوية المحددة.

وفي أخر الأمر تنتهي عملية إضافة وحذف المتغيرات إلى ومن التحليل ويتم إدراج جدول يلخص النتائج ويبين أي المتغيرات التي أضيفت إلى وأيها حذفت من التحليل في كل خطوة من خطواته، والمتغيرات التي تبقت في التحليل هي التي تستخدم في بناء (دالة) التمييز ثم يبين الجدول الذي يليه مباشرة في قائمة النتائج أي الدوال يمكن الاعتماد عليها إحصائياً ،بما أن المتغير التابع في الدراسة ذو فئتين لذلك تكون لدينا دالة واحدة فقط للتنبؤ بالفئات الصحيحة التي تنتمي إليها المفردات الجديدة مجهولة التصنيف (سليمان وأخرون 2016) وإضافة إلى ذلك اهتم الكثير من الباحثين بهذه الطريقة لتكون المعينة لهم في تصنيف وتحديد المشاهدات الجديدة ومجتمعاتها وتتميز بعدة أهداف منها (جودة 2008) :

- تصنيف المشاهدات ضمن مجموعات مختلفة.
- التحقق من الفرق داخل المجموعات وبين المجموعات.
- تحديد أبسط طريقة للتمييز بين المجتمعات.
- إيجاد نسبة التباين في المتغيرات المستقلة في تصنيف المجموعات.
- إبعاد المتغيرات التي ليس لها تأثير في تصنيف المجموعات.

الجانب التطبيقي : تحليل ونتائج الدراسة

النتائج الخاصة بالنموذج اللوجستي الثنائي

يحتوى النموذج على المتغير التابع Y المتمثل في حالة الغدة الدرقية لدى النساء المصابات باعتلال الغدة الدرقية حيث يأخذ المتغير القيمة واحد اذا كانت حالة الغدة الدرقية نشط والقيمة صفر اذا كانت حالة الغدة الدرقية كسل, و المتغيرات المستقلة المتمثلة في الوزن **(X_1)**، الجرعة المعطاة **(X_2)** ،الهرمون T4 **(X_3)** ، الهرمون TSH **(X_4)** ، نبضات القلب (X_5) ، قياس ضغط الدم (انقباضي) X_6 قياس ضغط الدم (انبساطي) X_7 ، الحالة الاجتماعية (X_8)و العمر(X_9) تم استخدام الحزمة الاحصائية SPSS لتحليل البيانات وتقدير نموذج الانحدار

اللوجستي لعينة الدراسة وذلك بعد ترميز البيانات واستخدام الانحدار اللوجستي التدريجي(Stepwise Logistic Regression) تم الحصول على النتائج الموضحة بالجدول رقم (1) الاتي:

الجدول (1)نتائج تحليل الانحدار اللوجستي الثنائي باستخدام طريقة التدرج الامامية

المتغيرات الداخلة في النموذج في كل خطوة Variables in the Equation							
		B	S.E.	Wald	Df	Sig.	Exp (B)
Step 1[a]	الهرمونTSH	-2.369	.372	40.609	1	.000	.094
	Constant	4.019	.633	40.263	1	.000	55.665
Step 2[b]	الهرمون T_4 X_3	.106	.032	10.957	1	.001	1.112
	الهرمون X4 TSH	-1.676	.361	21.559	1	.000	.187
	Constant	-11.728	4.480	6.853	1	.009	.000
Step 3[c]	الجرعة X_2	-0.049	.022	4.923	1	.027	.953
	الهرمون T_4 X_3	0.195	.076	6.614	1	.010	1.215
	الهرمونTSH X4	-2.233	.803	7.724	1	.005	.107
	Constant	-18.972	7.552	6.311	1	.012	.000
a. Variable(s) entered on step 1: X4. الهرمون TSH							
b. Variable(s) entered on step 2: X3.الهرمون T4							
c. Variable(s) entered on step 3: X2.الجرعة							

يتضح من الجدول بعد استخدام طريقة التدرج الامامي في تحليل البيانات، لمعرفة أهم المتغيرات المعنوية التي لها تأثير على المتغير التابع تم الوصول بعد ثلاث خطوات إلى نموذج يحتوي على المتغيرات المعنوية وهي الجرعة (X_2) و الهرمون T_4 (X_3) و الهرمون TSH (X_4) حيث تم اختبار فرضية العدم (H_0 : bi= 0) التي تنص على أن تأثير المعامل لا يساوي الصفر في المجتمع الذي سحبت منه العينة وهو معنوي عند مستوى معنوية (0.05) أي أن المتغير المستقل له تأثير في التنبؤ بقيمة المتغير التابع ضد الفرضية البديلة (H_1 : bi≠ 0) باستخدام إحصاءه (Wald) التي تتبع توزيع مربع كاي (χ^2).

وتكون صيغة النموذج على النحو التالي:

$$\log\left(\frac{\hat{p}}{1-\hat{p}}\right) = -18.972 - 0.049x_2 + 0.195x_3 - 2.233x_4$$

تم ادخال المتغير الأول الهرمون TSH (X_4) أولا حيث كان معامل سالب (-2.233) فمن إشارة المعامل تبين لنا انه كلما زادت نسبة الهرمونTHS في الجسم كلما كانت حالة الغدة كسل عند المريضة بقيمة احتمالية (0.005 = sig) ومن ثم تم أدخال متغير الهرمون T_4 (X_3) حيث كانت إشارة معامله موجبة (0.1950) وهذا يعني انه كلما زادت نسبة الهرمون T_4 في الجسم كلما كانت حالة الغدة نشط عند المريضة بقيمة احتمالية (0.010 sig =) و أخيرا تم ادخال متغير الجرعة (X_2) حيث كانت إشارة المعامل سالب (-0.490) وهذا يعني كلما زادت نسبة الجرعة التي يتلقاها المريض كلما كانت حالة الغدة كسل عند المريضة بقيمة احتمالية (sig = 0.027). الجدير بالذكر انه تم طريقة التدرج الخلفية واعطت نتائج مماثلة لطريقة التدرج الامامية حيث تم الحصول في الخطوة الأخيرة على نموذج يختوي على المتغيرات الثلاثة السابقة.

لاختبار كفاية و جودة النموذج المتحصل عليه (Goodness Of Fit) تم استخدام اختبار مربع كاي (χ^2 model) لاختبار معنوية النموذج: H_0 النموذج غير معنوي ضد H_1: النموذج معنوي

فكانت قيمة احصائي الاختبار **(Model Chi- square) مربع كاي تساوي 322.525 مع مستوى دلالة** (sig = 0.000)

وبالتالي نرفض فرضية العدم القائلة بان النموذج غير معنوي.

لاختبار جود توفيق النموذج استخدامنا أيضا اختبار هوزمر كما موضح في الجدول رقم 2

الجدول (2) اختبار جودة توفيق النموذج باستخدام اختبار هوزمر.

اختبار هوزمر وليشمورز			
	Chi-square اختبار مربع كاي	D.f درجة الحرية	Sig. المعنوية
النموذج	0.284	8	1.00

من خلال جدول رقم 2 تتضح أن قيمة مربع كاي **0.284** ومستوى معنوية (=sig.0.05<1.00) وعليه نقبل فرض العدم ونستنتج أن النموذج مطابق لبيانات الدراسة أي أن النموذج يمثل البيانات تمثيلاً جيداً.

النسب المئوية لجدول التصنيف:يعتبر جدول التصنيف أحد طرق فحص جودة المطابقة النموذج للبيانات وكانت النتائج كما هي موضحة في الجدول التالي:

جدول (3) تصنيف البيانات لحالات الغدة الدرقية

جدول التصنيف					
المشاهدات		المتوقعة		النسبة	
		حالة الغدة الدرقية			
حالة الغدة الدرقية		كسل	نشط		
	كسل	203	2	99.00	
	النشط	3	77	96.3	
النسبة		%98.2			

يتضح من خلال الجدول **(3)** النسبة المئوية الإجمالية التي تم توقعها بشكل صحيح بواسطة النموذج زادت هذه النسبة حيث كانت 71.9% للنموذج الفارغ والتي أصبحت **98.2**% للنموذج كامل حيث صنفت الحالات المصابة بكسل بالغدة الدرقية بشكل صحيح بنسبة 99% إما الحالات المصابة بنشاط الغدة الدرقية صنفت بشكل صحيح بنسبة **96.3%.**

نتائج أسلوب التحليل التمييزي Discriminant Analysis

جدول (4) تحليل التباين للمتغيرات المستقلة

Tests of Equality of Group Means					
	Wilks' Lambda	F	df1	df2	Sig.
(X1) الوزن	0.833	56.609	1	283	0.000
(X2) الجرعة المعطاه	0.562	220.642	1	283	0.000
(X3)الهرمون T_4	0.413	401.644	1	283	0.000
(X4) الهرمون TSH	0.221	995.637	1	283	0.000
(X5) نبضات القلب	0.799	71.224	1	283	0.000

(X6) انقباضي	0.981	5.617	1	283	0.018
(X7) انبساطي	0.999	0.386	1	283	0.535
(X8) الحالة الإجتماعية	0.981	5.535	1	283	0.019
(X9) العمر	0.997	0.921	1	283	0.338

يبين الجدول(4) نتائج تحليل التباين للمتغيرات المستقلة وذلك لتحديد ما إذا كان هناك فروق ذات دلالة إحصائية بين متوسطات المتغيرات المستقلة لكل فئة من فئات المتغير التابع، ويتضح من الجدول (4) أن أغلب الفروق بين المتوسطات لها دلالة إحصائية أي فروق معنوية لأن القيمة الاحتمال (Sig) أقل من مستوى المعنوية (0.05) للمتغيرات المستقلة فيما عدا المتغير الانبساطي والمتغير العمر.

جدول رقم (5) : خطوات التي يمر بها البرنامج

Variables Entered/Removeda,b,c,d					
		Wilks' Lambda			
Step	Entered	Statistic	df1	df2	df3
1	TSHالهرمون	0.219	1	1	282.000
2	T4الهرمون	0.173	2	1	282.000
3	الجرعة المعطاه	0.153	3	1	282.000
4	نبضات القلب	0.147	4	1	282.000
At each step, the variable that minimizes the overall Wilks' Lambda is entered.					
b. Minimum partial F to enter is 3.84.					
c. Maximum partial F to remove is 2.71.					
d. F level, tolerance, or VIN insufficient for further computation.					

يبين الجدول رقم 5 الخطوات التي مر بها البرنامج من إدخال واستبعاد المتغيرات المستقلة من التحليل التمييزي التدريجي كما يبين الجدول ترتيب المتغيرات أو استبعادها من التحليل، حيث نلاحظ استبعاد متغيرات (الوزن ، انقباضي، انبساطي ، حالة اجتماعية ، العمر) ، بالإضافة إلى ذلك يبين الجدول قيمة لامدا والقيمة المعنوية (P-Value) المصاحبة لكل حالة أو مرحلة ، الحد الأدنى لقيمة دالة المتغير التي تحدد بقائه بالتحليل 3.84 ، بالإضافة إلى الحد الأعلى للقيمة التي يحذف عندها المتغير من التحليل 2.71 .

جدول رقم (6) : المتغيرات الداخلة للتحليل ب أربع خطوات

Variables in the Analysis				
Step		Tolerance	F to Remove	Wilks' Lambda
1	TSHالهرمون	1.000	1007.430	
2	TSHالهرمون	0.997	387.119	0.412
	T4الهرمون	0.997	73.405	0.219

3	الهرمونTSH	0.994	263.901	0.297
	الهرمونT4	0.993	70.603	0.196
	الجرعة المعطاة	0.994	37.469	0.173
4	الهرمونTSH	0.991	226.281	0.266
	الهرمونT4	0.987	72.405	0.185
	الجرعة المعطاه	0.992	37.569	0.167
	نبضات القلب	0.990	11.401	0.153

يبين الجدول رقم 6 المتغيرات المستقلة التي بقيت في التحليل وذلك لكل خطوة من خطوات التحليل التمييزي التدريجي ، كما يبين الجدول قيم دوال الاختبار F to Remove ، ونلاحظ جميع دوال الاختبار أكبر من الحد الأعلى لقيمة للدالة (2.71) التي يتم عندها استبعاد المتغير ، لذلك لم يتم استبعادها من التحليل. ، في الخطوة (1) تم استبعاد جميع المتغيرات فيما عدا المتغير (الهرمون TSH) . أما في الخطوة (2) تم استبعاد جميع المتغيرات فيما عدا المتغير (الهرمون T4) من تحليل . وفي الخطوة (3) تم استبعاد جميع المتغيرات فيما عدا المتغير (الجرعة المعطاه) من التحليل, أما في الخطوة (4) والأخيرة تم استبعاد جميع المتغيرات فيما عدا المتغير (نبضات القلب) من التحليل. وذلك لأن قيمة الدالة لذلك المتغير أصغر من الحد الأدنى للدالة المحدد مسبقاً (3.84) ، الذي يحدد بقاء أي متغير بالتحليل.

جدول رقم (7) الإحصاءات المتعلقة بدالة التمييز

Wilks' Lambda					Eigenvalues		
Test of Function(s)	Wilks' Lambda	Chi-square	Df	Sig.	Eigenvalue	% of Variance	Cumulative %
1	0.147	533.794	9	0.000	5.798	100.0	100.0

يبين الجدول رقم (7) الإحصاء المتعلقة بدوال التمييز، كما يبين النسبة المئوية من التباين التي تعزى إلى دوال التمييز وبما أن لدينا دالة واحدة للتمييز نجد أن النسبة المئوية للتباين 100% و نلاحظ من القيمة الاحتمالية (Sig) الخاصة بالإحصاءة لامدا **0.000** أن الدالة معنوية ولها القدرة على التمييز ، وبالتالي يمكن الاعتماد عليها في تمييز حالات جديدة من حيث نوع الإصابة بنشاط أو كسل الغدة.

جدول رقم (8) يبين معاملات الدالة التمييزية

Canonical Discriminant Function Coefficients	
Function	
0.009	الجرعة المعطاه
0.028-	الهرمون T4
0.713	الهرمون TSH
0.024-	نبضات القلب
2.338	(Constant)

في هذا الجدول يوضح المعاملات التمييزية للارتباط بين كل متغير من المتغيرات المستقلة المنبئة الداخلة في التحليل وبين الدالة التمييزية. ثم نحصل على الدالة التمييزية التالية:

$$L = 0.009\ X_2 - 0.028\ X_3 + 0.713\ X_4 - 0.024\ X_5$$

جدول (9) يوضح الدالة التمييزية ومتوسطات المجموعتين (نقطة الفصل)

Functions at Group Centroids	
نوع الغدة	**Function**
نشاط الغدة	**-3.834**
كسل الغدة	**1.504**

يبين هذا الجدول الدالة التمييزية التجميعية غير المعيارية مقيمة حسب متوسطات المجموعتين، فكانت القيمة وقيمة المتوسط الدالة للنوع النشاط الغدة حسب التحليل التمييزي المتدرج 3.834- المتوسط الدالة للنوع الكسل الغدة حسب التحليل التمييزي المتدرج 1.504 ، وهذا أن الاشارة السالبة تعني أن ارتفاع درجات المتغيرات الداخلة في التحليل تؤدي إلى ارتفاع احتمالات الانضمام إلى المجموعة المشار إليها ،أما الإشارة الموجبة تعني ارتفاع درجات المتغيرات الداخلة في التحليل تؤدي إلى ارتفاع احتمالات الانضمام إلى المجموعة .

جدول رقم (10) نتائج تصنيف بيانات العينة باستخدام دالة التصنيف

Classification Resultsa					
			حالة الغدة المتوقعة		
		حالة الغدة	نشاط الغدة	كسل الغدة	المجموع
الأصلية	**العدد**	**نشاط الغدة**	**79**	**1**	**80**
		كسل الغدة	**3**	**201**	**204**
	%	**نشاط الغدة**	**98.8**	**1.3**	**100.0**
		كسل الغدة	**1.5**	**98.5**	**100.0**
98.6%					

تشير النتائج في الجدول السابق أن النوع النشاط فقد تم تصنيف 79 حالة صحيحة من اجمالي 80 حالة من نوع النشاط أي بنسبة 98.8% و يوجد حالة واحدة تم تصنيفها خطأ في النوع الكسل بنسبة 1.3% . أما بالنسبة النوع الكسل فقد تم تصنيف 201 حالة صحيحة من اجمالي 204 حالة من نوع الكسل وذلك بنسبة 98.5% بينما كانت ثلاث حالات تصنيفها خاطئ تم تصنيفها بالنوع النشاط وذلك بنسبة 1.5%.أما في وقد بلغت النسبة الكلية للتصنيف الصحيح 99.6% أي أن الدالة التمييزية المشتقة لها القدرة على التمييز.

النتائـــج:

بعد تطبيق النموذج اللوجستي الثنائي على بيانات الدراسة تحصلنا على النتائج الاتية

1- النموذج مثل البيانات تمثيلا جيدا

2- المتغيرات او العوامل التي تساعد في تصنيف المرضى بناء على النموذج اللوجستي كانت الجرعة المعطاة والهرمون T4 والهرمون TSH .

3- بلغت نسبة التصنيف الصحيح للنموذج اللوجستي 98.2%

بعد تطبيق أسلوب الدالة التمييزية الخطية توصلنا إلى الاستنتاجات التالية:

1- وجود فروق ذات دلالة إحصائية بين المتوسطات في المجموعتين.

2- عند تطبيق التحليل التمييزي تم الاعتماد على 7 متغيرات فقط وهي الأكثر تأثيراً في التحليل – الوزن و الجرعة المعطاة والهرمون T4 والهرمون Tsh ونبضات القلب وانقباضي وحالة اجتماعية– من أصل 9 متغيرات من المتغيرات المستقلة التي تم استبعادها وذلك لخروج قيم إحصاءة F لها عن مجال القبول.

3- وجود فروق معنوية بين المتوسطين في المجموعتين أي أن الدالة التمييزية التي توصلنا إليها لها القدرة علي التمييز بين المجموعتين (النشاط الغدة والكسل الغدة) أي يمكن الاعتماد عليها لتصنيف.

4- تعتبر الدوال التمييزية التي تحصلنا عليه في هذه الدراسة ممتازة حيث تمثل التصنيف الصحيح 98.6% للقدرة على التمييز.

5- اعتمدت الدالة التمييزية على 4 متغيرات وهي – الجرعة المعطاة والهرمون T4 والهرمون Tsh ونبضات القلب– .

المراجع

اولا : المراجع العربية

1. احمد، شمس الدين احمد علي (2012–2013) ،استخدام الدالة التمييزية الخطية لدراسة مستوى الإصابة بسرطان الغدة الدرقية، جامعة شندي، سودان.
2. إبراهيم، عمر احمد خالد، (2011)، استخدام الانحدار اللوجستي في تحديد العوامل الإصابة بسرطان البروستاتا، جامعة السودان، رسالة ماجستير.
3. الأغا، عبد المعين : (2010): أمراض الغدة الدرقية لدى الأطفال: جامعة الملك عبدالعزيز بجدة، المملكة العربية السعودية ص(17).

4- جودة، صلاح : (2008) : التحليل الإحصائي المتقدم باستخدام SPSS ،دار وائل للنشر والتوزيع، طبعة الأولى ، عمان ، الأردن ، ص 118،117.

5- حسينة، جواني : (2018) : دراسة مقارنة بين الانحدار اللوجستي وتحليل التمييزي للتنبؤ بفشل المؤسسات – دراسة عينة من المؤسسات الاقتصادية لولاية أم البواقي ، كلية العلوم الاقتصادية والتجارية وعلوم التسيير ، جامعة العربي بن مهيدي ، الجزائر.

6- سليمان، على أبشر ، وأخرون : (2016) : (مقارنة بين النموذج اللوجستي الثنائي والدالة التمييزية في تصنيف بالتطبيق على أهم العوامل المؤثرة في كفاية دخل الأسرة) ،جامعة السلام كلية الاقتصاد وتنمية المجتمع وجامعة النيلين كلية الدراسات الاقتصادية والاجتماعية ، مجلة العلوم والاقتصادية، مجلد 17 ،العدد 2 ،ص 118 .

7- فاطمة احمد، حسن، (2014)، مرض فقر الدم الأطفال وخطورته، جامعة الخرطوم، رسالة ماجستير.

8- فضل ، هالة هاشم عثمان (2006) ،استخدام الدالة التمييزية في تمييز الإصابة بمشكلات في القدمين لدى مرض السكري، كلية العلوم ، جامعة السودان للعلوم وتكنولوجيا ، السودان.

9- على سليمان ، عبدالمنعم الشيخ و سعد الكرم (2016) مقارنة بين النموذج اللوجستي الثنائي والدالة التميزية في التصنيف. مجلة العلوم الاقتصادية.

10- يعقوب ، أسماء أيوب : (2017) : التحليل العنقودي والتمييزي في دراسة تطبيقية على بعض المصارف العراقية ،مجلة الاقتصاد الخليجي ،العدد 31،العراق،ص (97).

ثانيا : المراجع الاجنبية :

1- Chapt, LE (1998). Applied categorical Data Analysis. New York: John Wiley &So. ns,Inc

2 Faiza faraj EL-khafifi(2002),Demographic and socio-Economic factors influencing Disability phenoomeon .university of garounis.

3- Ferrer, J. and Wang, L. (1999)." Comparing the Classification Accuracy among Nonparametric, Parametric Discriminant Analysis and Logistic

4- Kleinbaum, D. and Klein, M. (2002)." Logistic Regression a Self-learning Text ". USA: Spri

5- Hosmer, David.W., & Lemesow, S., (2000) "Applied logistic regression",2nd ed., New York, NY: John Wiley &Sons Inc.

6-Garson, D. (2006)."logistic Regression " Available http://www2.class.ncsu at: Edu/Garson/pa765/logistic.

7-Menard, S. (2002)." Applied Logistic Regression Analysis" 2nd edition. Sage University Paper series on Quantitative Applications in the Social sciencies No. (07-106). Beverly Halls, CA: Sag.

8-Sharma, s (1996). Applied Multivariate Techniques. New York: Addison Wesley Longman.

9 - Studeenmund, A.H (2001). Using Econometrics. New York: Addison Wesley Longman