تصنيف مرضى التهاب الكبد باستعمال الدالة التمييزية

مآب نوري حسن
جامعة بغداد / كلية الادارة والاقتصاد / قسم الأحصاء

ابتسام كريم عبد الله
جامعة بغداد / كلية الادارة والاقتصاد / قسم الأحصاء

تاريخ استلام البحث: 2022/07/13

تاريخ قبول البحوث: 2022/08/15

نشر البحث في العدد عشرون: سبتمبر / ايلول 2023

رمز التصنيف ديوي / النسخة الالكترونية (Online): 2522-64X/512:616.3

رمز التصنيف ديوي / النسخة الورقية (Print): 2519-948X/512:616.3

تصنيف مرضى التهاب الكبد باستعمال الدالة التمييزية

مآب نوري حسن
جامعة بغداد / كلية الادارة والاقتصاد / قسم الأحصاء

ابتسام كريم عبد الله
جامعة بغداد / كلية الادارة والاقتصاد / قسم الأحصاء

مستخلص البحث

يهدف البحث الى إنشاء دالة تميزية تقوم بتحديد العوامل المؤثرة في الأصابة بأختلال انزيمات الكبد، من خلال التمييز والفصل بين المصابين إلى مجموعتين (الذكور، الأناث) ودراسة عدد من العوامل المؤثرة(Age ,weight , ALP, AST ,ALT ,GGT) عن طريق أخذ عينة من 100 شخص وبعد التأكد من ملائمة التحليل التمييزي للبيانات وتكوين دالة تمييزية للفصل والتمييز .

Abstract

The research aims to create a discriminatory function that determines the factors affecting the incidence of liver enzyme abnormalities, by distinguishing and separating the sufferers into two groups (males, females) and studying a number of influencing factors (Age, weight, ALP, AST, ALT, GGT) by taking A sample of 100 people and after ensuring the appropriateness of the discriminatory analysis of the data and the formation of a discriminatory function for segregation and discrimination.

مقدمة (Introduction):

أدى تطور الأبحاث ووجود العلوم الحديثة المختلفة الى وجود تعدد في المتغيرات سواء أكانت نوعية أم كمية والتي ادت الى ظهور بيانات متعددة, وان وجود هذه المتغيرات يتطلب من الباحثين استعمال التحليل الأحصائي المتعدد المتغيرات والذي يقدم للباحثين طرائق احصائية لتحليل واختبار تلك البيانات والتنبؤ بها ويعد التحليل التمييزي نوع من انواع التحليل الأحصائي متعدد المتغيرات [12]. التي تستعمل في معالجة البيانات الوصفية ويعتمد على بناء دالة تسمى دالة التمييز الخطية وهي عبارة عن توليفة خطية لمجموعة من المتغيرات المستقلة وهذه الدالة تعمل على تقليل التشابه في أخطاء التصنيف ويهدف التحليل التمييزي إلى تصنيف المشاهدات إلى مجموعاتها الصحيحة بأقل خطأ تصنيف ممكن، و يختلف ألأسلوب التمييزي مع تحليل التباين وتحليل الانحدار و المتغير التابع نوعي، بينما في الاسلوبين الاخيرين يكون تغير كمي, كما يتشابه التحليل التمييزي مع الانحدار اللوجستي إذ يفسر كل منهما متغير وصفي الا ان الانحدار اللوجستي لا يطلب ان تكون المتغيرات المستقلة تتبع التوزيع الطبيعي.[2] ويستعمل التحليل التمييزي في مجالات متعددة فعلى سبيل المثال يوضف التحليل التمييزي في مجالات الطب والزراعة والتعليم وعلم الاجتماع وغيرها من المجالات التطبيقية. ومن الناحية الطبية فإن التطرق الى أنزيمات الكبد هو من الدراسة المهمه و أن الكبد من الاعضاء الاساسية في جسم الانسان وان اي زيادة او نقصان في انزيمات الكبد تشير إلى وجود التهاب أو

تلف في خلايا الكبد، و تسبب خلاياه الملتهبة أو المصابة أعلى من الكميات الطبيعية من بعض المواد الكيميائية إلى مجرى الدم، مما قد يؤدي إلى ارتفاع إنزيمات الكبد في اختبارات الدم و يؤدي ارتفاع احد الأنزيمات الى اختلال في عمل وظائف الكبد ويتم في هذا البحث دراسة هذة الانزيمات ومدى تأثيرها من خلال اصابة الكبد بالألتهاب أو تلف خلاياه.

الدراسات السابقه

- في العام 2015م قامت الباحثة الاء عبد الستار حمودات[6] بدراسة الدالة التمييزية وطرق تحديد متغيراتها ولغرض التوصل إلى معلمات الانموذج المستعمل لوصف المجتمعين هي (T^2 Hotelling)وجدول تحليل التباين وهذا يدل على أن الدالة التمييزية ذات كفاءة في التمييز وكذلك في معرفة معنوية المتغيرات الداخلة واستعمال طريقة روي بوس لحدود الثقه وأختبارt.
- في عام 2008م قامت الباحثة عائدة[131]بكتابة مقال عن استعمال التحليل المميز لتشخيص بعض أمراض العيون و تم استعمال التحليل المميز لتصنيف بعض أمراض العيون إلى ثلاث مجموعات إعتماداً على متغيرات ذات صفات تمييزية . و تم استعمال المقاييس التالية مقياس مربع كاي ومقياس F ومقياس ولكس, وباستعمال الدالة المميزة الخطية تم تصنيف بعض أمراض العيون على أساس مجموعة من المتغيرات المصاحبة للمرض.
- في عام 2009م قام الباحث بدر [14]بدراسة عن استعمال بعض طرائق التمييز الحصينة لتشخيص مرضى سرطان الدم اللوكيميا عن طريق ايجاد مقدرات حصينة للمتوسطات ودالة التباين والتباين المشترك للدوال الخطية والتربيعية , وكذلك ايجاد دوال خطية وتربيعية حصينة في حالة وجود قيم شاذه باستعمال طرق تحصين ضد وجود الشواذ في البيانات . ومن هذه الطرق طريقة اصغر محدد تباين مشترك وطريقة اصغر قطع بيضوي .
- في عام 2018م قامت الباحثة شهد نجم عبد اللة[15] بدراسة حول أستعمال التحليل التمييزي في تصنيف درجة التشوه للولادات , وأن الدراسة شملت ثلاث مجموعات هي التشوه البسيط والتشوه المتوسط والخطير . وقد اظهرت المتغيرات ان هنالك تأثير معنوي كبير لمتغير التدخين والمتغير R و عدد الاجهاضات .
- في عام 2014م قامت الباحثة بسمة [2] بدراسة بعنوان تمييز الكادر الطبي حسب معرفتهم للتصنيف الدولي (ICD-10) باستعمال الدالة التمييزية , وطريقة الرتب لأستخراج الدالة التمييزية لمعرفة الاطباء الذين لديهم المام بالتصنيف الدولي للامراض وتمت المقارنة بين الدالتين بالأعتماد على خطأ التصنيف.

مراحل أختيار الدالة التمييزية:

أ. مرحلة الاشتقاق Derivation

تتضمن هذه المرحلة عدة خطوات أبتداء من اختيار المتغيرات وتقسيم العينة ثم اجراء الاختبارات الاحصائية. بالنسبة إلى خطوة اختيار المتغيرات، تتضمن تحديد ما هي المتغيرات التفسيرية و من هو المتغير التابع، الخطوة الثانية هي تقسيم العينة. في حين تتضمن الخطوة الثالثة أسلوبين إجراء التحليل التمييزي وهو ألأسلوب المباشر وألأسلوب الثاني هو أسلوب الخطوات (stepwise)، وسيتم الأعتماد في هذا البحث على هذا الأسلوب و يكون لدينا عدد كبير من المتغيرات التفسيرية للوصول الى أكثر المتغيرات معنوية للتمييز بين المشاهدات[6]

ب. مرحلة التحقيق Validation

في هذه المرحلة يتم الاعتماد على أكثر من مؤشر لتقييم مدى دقة المعادلة التمييزية، فإذا كانت معنوية ينصب الهدف على تحديد نقطة القطع (cut off value)، و يتم اعتماد هذه الدرجة كمقياس لمعرفة ما إذا كانت كل مفردة تعود إلى إحدى المجموعات.

ت. مرحلة التفسير

إذا كانت الدالة التمييزية معنوية إحصائيا والتصنيف مقبولً يمكن الاستمرار في هذه المرحلة، و أنها تتضمن خطوتين مهمتين، الخطوة الاولى تكون تتضمن الدوال المميزة لتحديد الاهمية النسبية لكل متغير مستقل في التمييز بين المجاميع .أما الخطوة الثانية تكون اختبار متوسطات المشاهدات في تحليل التباين اﻷحادي لكل متغير توضيحي ذي معنوية إحصائية في الدالة المميزة .

الانحدار التدريجي (Stepwise-regression)

وسيتم استعمال الأنحدار التدريجي في هذا البحث اذ تكون أهمية الأنحدار التدريجي في فحص الاهمية الأحصائية بصورة متكرره لكل المتغيرات و بشكل مستقل

خطوات الانحدار التدريجي

يتم استعمال المتغيرات التفسرية أثناء إنشاء دالة التحليل التمييزي بالترتيب و أن ترتيب المتغيرات فيه يكون وفق معيار احصائي و يتم البدء بأكثر المتغيرات تميزاً وهو من اشهر الاساليب الاحصائية[7]

أ- نأخذ المتغير الأكثر معنوية إلى الانموذج، وذلك بعد أن نختبر فيما إذا كان غير معنوياً، فإذا كان غيرمعنوي يتم التوقف.

ب-تختبر جميع المتغيرات في الانموذج فإذا كان منها متغيرات غير معنوية تحذف من الانموذج .

ج- أعادة الخطوات أ و ب إلى أن يتم التوقف ويلاحظ في الخطوة (أ) اَن اختيار المتغير الأكثر معنوية يكون آنياً وليس متفرداً [9].

تقدير حدود الثقة لروي بوس (Roy-Bose Confidence Interval Estimation):

يتم استعمال حدود الثقة لروي- بوس لتعيين المتغيرات المهمة التي ستدخل في الدالة، ويمكن استعمال هذه الطريقة بالخطوات التالية:

1- نجد قيمة F الجدولية بالرجوع إلى جداول خاصة وبدرجة حرية:

$$Tab'F = F(\alpha/2, m, n1 + n2 - 1)$$

2- ايجاد قيمة T الجدولية:

$$TabT = \left[\frac{(n_1+n_2-2)\mathrm{m}}{n_1+n_2-m-1} tab \cdot F\right]^2$$

3- إيجاد المتجه (Vector Selection) ويرمز له بالرمز a وهو متجه غير صفري ويكون عدد هذه المتجهات بعدد المتغيرات التوضيحية

$$a_{1=}[1 \quad 0 \quad 0 \ldots.]$$
$$a_{2=}[0 \quad 1 \quad 0 \ 0 \ldots]$$
$$\vdots$$
$$\ldots\ldots\ldots a_{m=}[0 \quad 0 \quad \ldots \ldots 1]$$

4- ثم بعد ذلك نجد حدود الثقة (Roy-Bose) وتكتب بالصيغة التالية:

$$\underline{\mathbf{a}}\boldsymbol{k}(\underline{\bar{X}}_1 - \underline{\bar{X}}_2) - \sqrt{\underline{\mathbf{a}}\mathbf{k}\sum \mathbf{a_k}\left[\frac{\mathbf{n_1}+\mathbf{n_2}}{\mathbf{n_1}\cdot\mathbf{n_2}}\right]\boldsymbol{T}} < \underline{\mathbf{a}}\mathbf{k} \quad < \underline{\mathbf{a}}\mathbf{k}(\underline{\bar{X}}_1 - \underline{\bar{X}}_2) + \sqrt{\underline{\mathbf{a}}\mathbf{k}\underline{\mathbf{a}}\mathbf{a_k}\left[\frac{\mathbf{n_1}+\mathbf{n_2}}{\mathbf{n_1}\cdot\mathbf{n_2}}\right]}$$

5- ثم نكتب المتباينة

$$\mathrm{C.I(L)} < a'm < C.I)$$

والقرار هو إذا احتوت المتباينة على الصفر يعني عدم وجود أهمية لذلك المتغير أي أن المتغير يختلف وسطه الحسابي في كلا المجموعتين ، إما إذا لم تحتوي المتباينة على الصفر يعني وجود أهمية لذلك المتغير.

نقطة القطع (Cut Off Value)

بعد تكوين تصنيف الدالة التمييزيّة واختبار قدرتها على التمييز والفصل بين المجموعتين يبدأ الاستعمال الثاني لها وهو كيفية تصنيف المشاهدة الجديدة إلى أي المجموعتين تنتمي. اي ان لتصنيف الأفراد فإننا نحتاج إلى نقطة تفصل بين المجموعتين و إذا قلت (مجموعة نقاط الأفراد) عن تلك النقطة فإن الفرد عندئذ يتبع إلى مجموعة معينة وإذا زادت مجموع نقاطه عن تلك النقطة الفاصلة أصبح تابعاً للمجموعة الأخرى. ويتم ذلك من خلال الخطوات الاتية[11]:

أ. تحديد نقطة التصنيف وهي تمثل متوسط المتوسطين:

$$\bar{\bar{y}} = \frac{\bar{y}_1 + \bar{y}_2}{2}$$

و ان :

$\bar{\bar{y}}$: نقطة التصنيف.

$\bar{y}_1$: متوسط القيم التميزية للمجموعة الأولى.

$\bar{y}_2$: متوسط القيم التميزية للمجموعة الثانية.

- **اختبار كفاءة دالة التمييزالخطي Testing of Linear Discriminate Function**

عندما يراد التمييز بين مجموعتين فأنه يمكننا إختبار الفرضية التي تنص على تساوي متوسط المجموعات.

$H_0: b_1 = b_2 = \cdots = b_k$

$H_1: b_1 \neq b_2 \neq \cdots \neq b_k$

وبذلك فأن اختبار الدالة المميزة تعتمد على رفض أو قبول فرضية العدم H_0.

وان إحصاءة الاختبار المستعملة في حالة التميز بين مجموعتين هي إحصاءة الاختبار (F) والتي تعتمد على إحصاءة الاختبار هوتلنك $(Hotelling)$ وصيغته الرياضية كالآتي [7]:

$$F = \frac{n_1 + n_2 - p - 1}{(n_1 + n_2 - 2)p}\ T^2$$

$$T^2 = \frac{n_1\, n_2}{n_1 + n_2}\ \ D_i^2$$

و ان :-

$$D_i^2 = (\bar{X}_{i1} - \bar{X}_{i2})' S^{-1} (\bar{X}_{i1} - \bar{X}_{i2})$$

D_i^2:- يمثل إحصاءه $(Mahalanobis)$ (مهلنوبيس)

p: عدد المتغيرات

وبهذا نرفض فرضية العدم H_0، عندما تكون F (المحسوبة) اكبر من F (الجد ولية) ونقبل الفرضية البديلة H_1 ، بمستوى معنوية α وبدرجة حرية $(n_1 + n_2 - p - 1)$.

$$F > F_{\alpha(p,n_1+n_2-p-1)}$$

وهذا يدل على إن متوسطات المجموعتين غير متساوية إي يوجد فروق معنوية بين المجموعتين بمعنى إن الدالة المميزة الخطية قابلة للتميز بدرجة عالية.

اختبار تساوي مصفوفة التباين والتباين المشترك لجميع المجاميع:

يستعمل هذا الاختبار لمعرفة النوع الملائم من النماذج لتمثيل دالـة التميز بين المجاميع. وتكون فرضيتي العدم والبديلة بالشكل التالي[13]:

$$H_0: \sum_1 = \sum_2 = \dots \sum_k \dots$$

يوجد على الاقل اثنان منهم غير متساويان :H_1

فعندما تكون فرضية العدم H_0 صحيحة فأن[20] :

$$S = \frac{\sum_{i=1}^{k} n_i s_i}{\sum_{i=1}^{k} n_i}$$

و أن :

S هو تقدير غير متحيز لـ $\sum$ ، وبذلك فأن S تمثل تقدير مصفوفة التباين والتباين المشترك وتكون إحصاءه الاختبار المستعملة في هذا كما يلي:

$$M = (\sum_{i=1}^{k} n_i)\, Ln\,|S| - \sum_{i=1}^{k}(n_i\, Ln\,|S|)$$

وقد اثبت Box عام (1949)م انهُ إذا ضرب M في ثابت C-1 والذي يساوي [8]:

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

نتوصل بذلك إلى مقياس يتوزع بالتقريب توزيع (χ^2) وخاصة عندما يكون حجم العينة كبير وبدرجة حرية $(k-1)(p-1)$

$$MC^{-1} \sim \chi^2\ with\ (k-1)(p-1)$$

و إن :

p : عدد المتغيرات المدروسة في المصفوفة .

k : عدد المجاميع.

أخطاء التصنيف (Classification Error)

تعرف أخطاء التصنيف بأنه احتمال تصنيف مفردة معينة إلى المجموعة الأولى وهي في الحقيقة تعود للمجموعة الثانية وبالعكس وعند حساب خطأ التصنيف نأخذ بعين الأعتبار أنه كلما كان حجم العينة كبيراً فأن توزيع المفردات يقترب من التوزيع الطبيعي وهذا الأمر مهم في حساب احتمال اخطأء التصنيف.[5]

ويقسم خطأ التصنيف الى نوعين:

أ- خطأ التصنيف الظاهري:

ويتم ايجاد الخطأ الظاهري كما يلي:

$$p_{12} = \frac{n_{12}}{n_1}$$

p_{12} : نسبة المفردات التي تنتمي للمجموعة الأولى وصنفت خطأ للثانية.

$$p_{21} = \frac{n_{21}}{n_2}$$

p_{21} : نسبة المفردات التي تنتمي للمجموعة الثانية وصنفت خطأ في الأولى،
و يتم الحساب عن طريق الجدول التالي :

جدول (1) خطأ التصنيف الظاهري

المجموعة	مشاهدات المجموعة الأولى	مشاهدات المجموعة الثانيه	المجموع
الأولى	n_{11}	n_{12}	n_1
الثانية	n_{21}	n_{22}	n_2

n_{11} : عدد المفردات من المجموعة الأولى والتي تم تصنيفها في نفس المجموعة وبالتالي هي صنفت بطريقة صحيحة.

n_{12} :عدد المفردات من المجموعة الأولى والتي تم تصنيفها خطأ في المجموعة الثانية .

n_{21}:عدد المفردات التي تنتمي بالاصل إلى المجموعة الثانية وتم تصنيفها خطأ في المجموعة الأولى. n_{22} :عدد المفردات في المجموعة الثانية التي تم تصنيفها في نفس المجموعة و هي صنفت بطريقة صحيحة.

ويمكن حساب معدل الخطأ الظاهري باستعمال المعادلة ادناه :

$$\frac{n_{12} + n_{21}}{n_1 + n_2}$$

الخطأ الحقيقي

يمثل نسبة التصنيف الخاطئ في المجتمع [4]

$$P_{12} = P_{21} = L\left[\frac{-\sqrt{D^2}}{2}\right]$$

و (L) دالة التوزيع الطبيعي المعياري، D أحصائية $(Mahalnobis)$
تحسب القيمة بين القوسين ويحسب الأحتمال المقابل لها من جدول التوزيع الطبيعي المعياري وكلما اقترب الأحتمال من الصفر دل على نوع وانخفاض في خطأ التصنيف وبالتالي قدرة الدالة على التصنيف والتمييز اما إذا كان الأحتمال قريب من الواحد يدل على ارتفاع خطأ التصنيف وانخفاض قدرة الدالة على التصنيف والتمييز ويعتبر خطأ التصنيف عامل مهم عند الحكم على كفاءة الدالة التمييزيّة.[2]

الدالة التمييزيّة بمعاملات غير معيارية[3] - :

تأخذ الشكل التالي:

$$y = b_o + b_1x_1 + b_2x_2 \dots\dots\dots\dots\dots\dots b_kx_k$$

(Y) الدالة التمييزيّة غير المعيارية.

$(b_\circ)$ ثابت التمييز.

$(b_n's)$ معاملات التمييز غير المعيارية.

$(x_n's)$ قيم المتغيرات غير المعيارية.
وللحكم على جودة الانموذج التمييزي من خلال معامل الارتباط القانوني Canonical correlation وان القيم المرتفعة لمعامل الارتباط القانوني تكون مؤشر للجودة العالية للأنموذج التمييزي وبتربيع قيمة عامل الارتباط القانوني تحصل على قيمة معامل التحديد R^2 الذي يحدد نسبة مساهمة المتغيرات المستقلة في التصنيف والتمييز[18].

أختبار قدرة الدالة على التمييز والفصل بين المجموعات:

تستعمل الأختبارات الآتية :
اختبار F $(\boldsymbol{F-TEST})$
وذلك لآختبار قدرة الدالة على التمييز وعن طريق الفرضية التي تنص على ان الدالة ليس لديها القدرة على التمييز H_0 ضد دالة لديها القدرة على التمييز (H_1) ويعتمد هذا الآختبار على قياس الآختلآفات بين المجموعات وداخل المجموعات بين المفردات ويتم ذلك من خلال جدول تحليل التباين (ANOVA Table) الأتي[7]:

جدول(2)ANOVA Table

Source of Variation	$S\,S$	$d\,f$	MSS	F
بين المجموعات Between x's	SSB	$k-1$	M_{SB}	$\frac{M_{SB}}{M_{SE}}$
الخطأ Within x's	SSE	$n-k$	M_{SE}	
الكلي Total	SST	$n-1$		

و ان :-

1. مجموع مربعات الاخطاء يحسب كالتالي [10]

$$\text{SSE} = D^2 = \hat{\alpha}_1 d_1 + \hat{\alpha}_2 d_2 + \cdots\cdots\cdots\cdots\cdots + \hat{\alpha}_k d_k$$

2. مجموع مربعات بين المتغيرات يحسب كالتالي:

$$SSB = \frac{n_1 n_2}{(n_1+n_2)(n_1+n_2-2)} \times (D^2)^2$$

3. مجموع المربعات الكلي يحسب كالتالي:

$$SST = SSB + SSE$$

ويتم الأختبار كالأتي:-

1- صياغة الفروض:
الدالة ليس لها القدرة على التمييز: H_0
الدالة لها القدرة على التمييز.: H_1

2- القيمة المحسوبة:

$$F = \frac{M_{SB}}{M_{SE}}$$

3- القيمة الجدولية:

$$F_{\alpha(k-1,\ \ n-k)}$$

4- القرار:

إذا كانت F المحسوبة أكبر من F الجدولية نرفض الفرض العدمي ونقبل بالفرض البديل ويكون للدالة قدرة عالية على التمييز والعكس صحيح.

جمع البيانات:

جمعت البيانات الخاصة بموضوع البحث من مختبرات التعليمية لمدينة الأمامين الكاظمين الطبية واحدى مختبرات المستشفيات الاهلية، وقد تم الاستعانة بمجموعة من المختصين بامراض الكبد للوقوف على اهم العوامل المؤثرة اذا تم جمع البيانات الخاصة لسنة 2021م وبعينة حجمها 100 مراجع، ولغرض تحليل البيانات فقد تم اعتبار متغير الاستجابة (y) هو (y=1)ذكور (y=2) أناث وكان عدد الاهلي 50 والحكومي 50 بالاعتماد على المتغيرات التفسيرية.

- (Age) متغير يمثل العمر
- (Weight) متغير يمثل الوزن
- (ALT)متغير يمثل انزيم ناقل امين ألانين (Alanine transaminase) يتركز هذا الانزيم بشكل رئيسي في الكبد وهو يحول البروتينات الى طاقة لذلك فأن اي خلل في النسب الطبيعية لهذا الأنزيم يدل على اصابة الكبد.
- (AST) متغير يمثل أنزيم ناقل أمين السباتات (Aspartate transaminase)
- وهو انزيم يوجد في العديد من اعضاء الجسم ويساعد في استقلاب الاحماض الأمينية.
- (ALP)متغير يمثل أنزيم الفوسفاتيز القلوي (Alkaline phosphatase) يوجد هذا الأنزيم بصورة اساسية في الكبد والذي يعمل على تحلل البروتين .
- (GGT)متغير يمثل أنزيم ناقلة الببتيد (Gamma-glutamy ltranspeptidase) وهو احد الأنزيمات التي يتم تصنيعها في الكبد وتتركز بشكل رئيسي في الدم.

وتدل الزيادة او النقصان في هذه الأنزيمات الى العديد من المشكلات مثل ألتهاب الكبد الحاد او المزمن او تليف الكبد او أمراض الغدة الدرقية.

تحليل البيانات:

تتم عملية تحليل البيانات باستعمال البرنامج الاحصائي (SPSS) وتمر مرحلة التحليل بثلاث

جدول(3): اختبارKolmogorov-Smirnov

One Kolmogorov-Smirnov Test

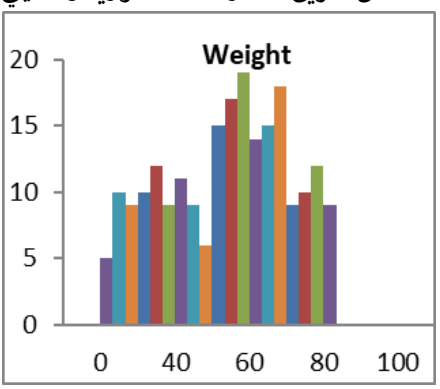
				Age	Weight	ALP	AST	ALT	GGT
Most Extreme Differences	Absolute			0.060	0.067	0.062	0.089	0.0	0.203
	Positive			0.053	0.067	0.062	0.040	0.043	0.192
	Negative			-0.060	-0.064	-0.054	-0.089	-0.045	-0.203
Test Statistic				0.060	0.067	0.062	0.089	0.045	0.203
Asymp. Sig. (2-tailed)[c]				0.200[d]	0.200[d]	0.200[d]	0.048	0.200[d]	0
Monte Carlo Sig. (2-tailed)[e]	Sig.			0.501	0.330	0.437	0.052	0.889	0
	99% Confidence Interval	Lower Bound	0.488		0.318	0.424	0.046	0.881	0
		Upper Bound	0.514		0.342	0.450	0.058	0.897	0

ألتأكد من توفر شروط التحليل التمييزي عن طريق:

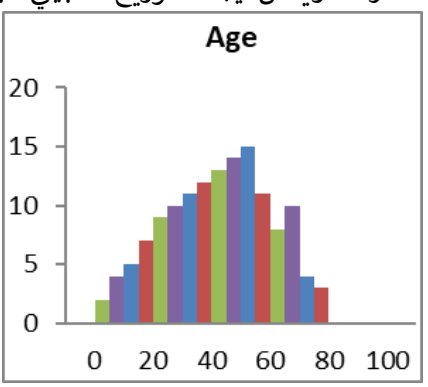
ايجاد التوزيع الطبيعي يتم ايجاد هل أن البيانات تتبع التوزيع الطبيعي عن طريق اجراء اختبار Kolmogorov-Smirnov

One Kolmogorov-Smirnov Test								
				Age	Weight	ALP	AST	ALT	GGT
Most Extreme Differences	Absolute			0.060	0.067	0.062	0.089	0.0	0.203
	Positive			0.053	0.067	0.062	0.040	0.043	0.192
	Negative			-0.060	-0.064	-0.054	-0.089	-0.045	-0.203
Test Statistic				0.060	0.067	0.062	0.089	0.045	0.203
Asymp. Sig. (2-tailed)[c]				0.200[d]	0.200[d]	0.200[d]	0.048	0.200[d]	0
Monte Carlo Sig. (2-tailed)[e]	Sig.			0.501	0.330	0.437	0.052	0.889	0
	99% Confidence Interval	Lower Bound	0.488		0.318	0.424	0.046	0.881	0
		Upper Bound	0.514		0.342	0.450	0.058	0.897	0

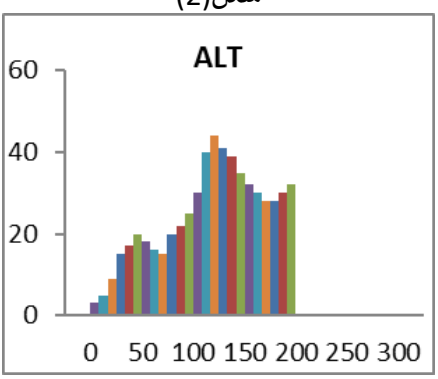
وتوضح نتائج الجدول (3) من المخرجات اعلاه ان البيانات المجمعة في كل من المتغيرات الستة تتبع التوزيع الطبيعي وان مستوى الدلالة لكل فيها أكبر من 0.05 المستوى المعتمدة لهذه الدراسة ويمكن ايجاد التوزيع الطبيعي للبيانات عن طريق المدرجات التكرارية وكما يلي:



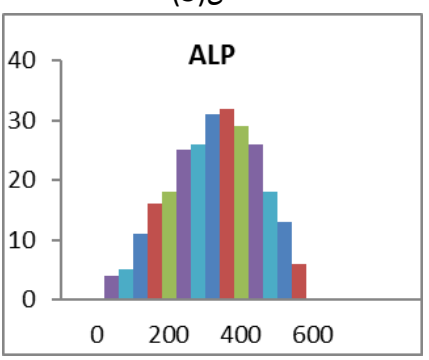
شكل(2)



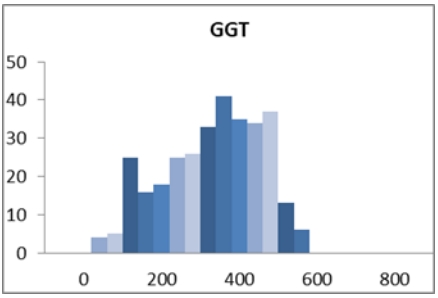
شكل(3)



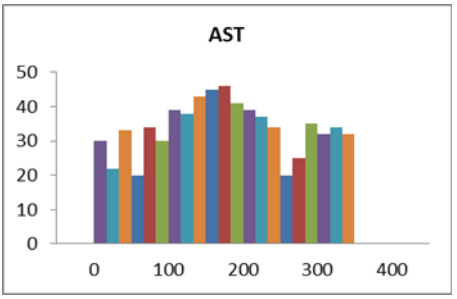
شكل(4)



شكل(5)



شكل(6)



شكل(7)

وللتأكد من ان البيانات لاتحتوي على قيم شاذة في كافة المتغيرات المستقلة نجري أختبار (Mahalanbis) ونلاحظ بوجود عمود جديد يحتوي على متغير جديد بأسم (MAH.1)، وبمراجعة هذه القيم الموجودة تحت عمود (MAH.1) نجد ان هذه القيم أقل من القيم الجدولة لمربع كاي والتي كانت تساوي 98 وبدرجة حرية 5 وذلك بعدم وجود قيم شاذه بكافة المتغيرات المستقلة .

التأكد من عدم وجود ارتباط ذاتي بين المتغيرات المستقلة يتضح من جدول رقم (4) أن قيم(VIF) لكل من المتغيرات التفسيرية الستة كانت 1.477 , 1.473 , 1.190 , 1.241 , 1.335 , 1.255, على التوالي وبما أن النتائج كانت أقل من 5 , يمكن الأستنتاج بأنة لايوجد ارتباط ,

وايضا يمكن الاستنتاج من خلال قيم (Tolerance) وبما أن القيم أكبر من 0.57 ($1\text{-}R^2$) فأن ذلك يعني عدم وجود ارتباط بين المتغيرات التفسيرية .

جدول(4): يوضح الارتباط بين المتغيرات

Classification Function Coefficients		
	genval	
	1.00	2.00
Age	-.034	.005
Weight	.408	.322
ALP	.007	.013
AST	.024	.017
(Constant)	-17.207	-13.282

التأكد من تجانس المجتمع Box’s M

الجدول (5) يوضح نتائج اختبارBox’s M

Test Results		
Box’s M		36.823
F	Approx.	12.003
	df1	3
	df2	1728720.000
	Sig.	<0.23

ونلاحظ من الجدول (5) أن قيمة مستوى الدلالة المستخرج تساوي 0.23 وهي أكبر من مستوى المعنوية المستعملة في الدراسة هذا يدل على التجانس بين افراد المجموعتين مما يستوجب بقبول الفرضية الصفرية .

جدول (6): نتائج القيمة النسبية

Log Determinants		
Type	Rank	Log Determinants
1	2	10.602
2	2	10.786
Pooled Within-groups	2	10.070

كما يلاحظ من الجدول رقم (6) أن قيم **(Log Determinants)** تقريبا متساوية بين المجموعتين وهذا يعني ان مصفوفات التباين المشتركة متساوية.

أجراء التحليل التمييزي:

الجدول (7): الأوساط الحسابية والانحرافات المعيارية لكل مجموعة من المجموعات.

Group statistics			
	Type	Mean	Std. Deviation
الذكور	Age	64.040	7.5376
	Weight	76.580	9.7940
	ALP	152.920	117.3491
	AST	42.370	52.0797
	ALT	24.772	27.9268
	GGT	130.100	151.3071
الاناث	Age	34.560	14.9805
	Weight	68.460	18.8487
	ALP	139.140	83.2422
	AST	63.974	87.2415
	ALT	38.078	46.2523
	GGT	88.840	69.3649
Total	Age	49.300	18.9382
	Weight	72.520	15.4909
	ALP	146.030	101.4566
	AST	53.172	72.3009
	ALT	31.425	38.5948

جدول(8): جدول مصفوفة الأرتباط داخل المتغيرات التفسيرية

Pooled Within-Groups Matrices

		Age	Weight	ALP	AST	ALT	GGT
Correlation	Age	1	0.505	-.047	0.014	-0.055	-0.040
	Weight	0.505	1	-0.042	-0.054	0.065	-0.216
	ALP	-0.047	-0.042	1	0.050	0.165	0.341
	AST	0.014	-0.054	0.050	1	0.413	-0.001
	ALT	-0.055	0.065	0.165	0.413	1	-0.088
	GGT	-0.040	-0.216	0.341	-0.001	-0.088	1

يشير جدول رقم (8) الى معاملات الارتباط الثنائي بين المتغيرات التفسيرية الستة, فمعامل الارتباط مثلاً بين (Age, Weight) كانت قيمتهُ 0,505 وبين(AST,ALP) كان يساوي 0.050 وهكذا بين باقي المتغيرات

جدول رقم(9) جدول التحليل التمييزي المتدرج

Variables Not in the Analysis						
Step		**Tolerance**	**Min. Tolerance**	**F to Enter**	**Min. F**	**Between Groups**
0	**Age**	1	1	154.513	154.513	الحكومي والاهلي
	Weight	1	1	7.307	7.307	الحكومي والاهلي
	ALP	1	1	0.459	0.459	الحكومي والاهلي
	AST	1	1	2.261	2.261	الحكومي والاهلي
	ALT	1	1	3.033	3.033	الحكومي والاهلي
	GGT	1	1	3.072	3.072	الحكومي والاهلي
1	**Weight**	0.745	0.745	6.607	84.981	الحكومي والاهلي
	ALP	.998	0.998	0.612	77.256	الحكومي والاهلي
	AST	1	1	1.089	77.871	الحكومي والاهلي
	ALT	0.997	0.997	0.428	77.020	الحكومي والاهلي
	GGT	0.998	0.998	1.940	78.967	الحكومي والاهلي
2	**ALP**	0.997	0.744	0.493	56.522	الحكومي والاهلي
	AST	0.995	0.741	1.397	57.351	الحكومي والاهلي
	ALT	0.985	0.736	0.133	56.191	الحكومي والاهلي
	GGT	0.947	0.706	0.641	56.658	الحكومي والاهلي

يوضح الجدول (9) خطوات التحليل التمييزي المتدرج (Stepwise) و توضح الخطوة رقم(0) عدم استبعاد اي متغير وتوضح الخطوة رقم(1) تم تحديد متغير العمر Age كأول متغير يدخل في الدالة التمييزية لأنه يتمتع باكبر قيمة لدالة الاختبار F و ان قيمتها(154.513) وتوضح الخطوة رقم(2) استبعاد متغير الوزن (Weight) لأن قيمتها اكبر من قيمة F ، في الخطوة (2) لم يتم ادخال اي من المتغيرات الاربعة المتبقية(GGT,ALT,ALP, AST) الى الدالة اطلاقاً وذلك لأن قيم الأختبار لهذه المتغيرات اصغر من الحد الأدنى الموضوع مسبقاً والذي يساوي (3.84).

جدول (10): يشير الى نتائج التحليل التمييزي التدريجي

Walk's' Lambda									
Step	Number of Variables	Lambda	df1	df2	df3	Exact F			
						Statistic	df1	df2	Sig.
1	1	0.388	1	1	98	154.513	1	98.000	<.001
2	2	0.363	2	1	98	84.981	2	97.000	<.001

يبن جدول (10) النتائج المعتمدة للتحليل التمييزي المتدرج يوضح ان المتغيرين (1,2) هي التي ستعتمد عليها دالة التمييز للمصابين وان اول هذه المتغيرات هو العمر (Age) لان له اكبر قيمة لنتيجة أختبار F واكبر قيمة ل(Walk's' Lambda) ويليه متغير التفسير الثاني الوزن (Weight).

جدول(11): يشير الى معاملات التمييز المعيلرية وغير المعيارية

Standardized Canonical Discriminant Function Coefficients

	Function
	1
Age	-.623
Weight	1.051
ALP	-.492
AST	.404

يوضح الجدول (11) معاملات التمييز المعيارية وغير المعيارية ويمكن كتابة نموذج الدالة التميزية كما في المعادلة التالية

$$y = -0.623x_1 + 1.051x_2 - 0.493x_3 + 0.404x_4$$

التصنيف (Classification):

جدول(12): نتائج التصنيف

Classification Results					
		Type	Predicted Group Membership		Total
			الحكومي	الاهلي	
Original	Count	الحكومي	50	0	50
		الاهلي	5	45	50
	%	الحكومي	100.0	.0	100.0
		الاهلي	10.0	90.0	100.0
A 95.0% of original grouped cases correctly classified.					

يشير جدول رقم (12) الى التصنيف الصحيح لمفردات العينة بالأعتماد على الدالة التمييزية ويبين الجدول ان معدل التصنيف الناجح هو 95% اي ان تم تصنيف مرضى الحكومي والأهلي تصنيفاً صحيحاً بدقة 95% وهذا يعني ان نسبة خطأ التصنيف قليلة وهي 5% .

الأستنتاجات والتوصيات

1. ملائمة اسلوب التحليل التمييزي لأية بيانات يحتاجها الباحث في تمييز وتصنيف المفردات الجديدة.
2. تم التوصل الى نموذج يوضح اهمية المتغيرات الداخلة في التمييز
3. تطبيق الدوال التمييزيّة في تحليل نتائج البيانات ولاسيما الطبية منها

المصادر:

1) بسيوني، عبدالرحيم عوض عبدالخالق (2021)، "استخدام التحليل التمييزي في التنبؤ". المجلة العلمية / التجارة والتمويل، المجلد 41، العدد 3.
2) الحمداني، بسمة رشيد، (2014)، "تمييز الكادر الطبي حسب معرفتهم للتصنيف الدولي بأستخدام الدالة المميزة، بحث دبلوم عالي جامعة بغداد، كلية الادارة والاقتصاد
3) بغرش، سعيدة، (2020) "استخدام التحليل التمييزي في تقدير خطر عدم تسديد القرض من طرق الوكالة الوطنية لتسيير القرض المصغر" مجلة نماء للاقتصاد والتجارة، المجلد 4 العدد1، الجزائر.
4) الجاعوني، فريد عدنان غانم (2007)، "التحليل الإحصائي متعدد المتغيرات (التحليل التمييزي) في توصيف وتوزيع الأسر داخل الهيكل الاقتصادي الاجتماعي في المجتمع"، مجلة جامعة دمشق للعلوم الاقتصادية والقانونية– المجلد 23 -العدد الثاني.
5) الحمداني، بسمة رشيد، (2014)، "تمييز الكادر الطبي حسب معرفتهم للتصنيف الدولي بأستخدام الدالة المميزة، بحث دبلوم عالي جامعة بغداد، كلية الادارة والاقتصاد
6) حمودات، الاء عبد الستار، (2015)، "الدالة التمييزية وطرق تحديد متغيراتها"، بحث ماجستير، جامعة الموصل، كلية علوم الحاسبات والرياضيات قسم الرياضيات.
7) خلف عماد محمد وعقيل محمد رشيد (2021) "مديونية القطاع المالي وآثرها على الاقتصاد المالي في العراق" مجلة الدراسات النقدية والمالية العدد التاسع.
8) خياط، شذى بنت شاكر وجواهر بنت محمد المزيد، (2018)، "القوة التمييزية للذكاءات المتعددة بين طالبات الكليات الأنسانية والعلمية بجامعة الملك سعود"، بحث بكالوريوس، كلية التربية- جامعة جدة، كلية التربية جامعة الملك سعود.
9) الراوي، خاشع، (1987)، "المدخل الى تحليل الانحدار"، جامعة الموصل، كلية الزراعة والغابات.
10) رجب، وليد خالد وحسين جميل محمود، (2012)، "التحليل التمييزي لبعض المتغيرات المهارية لدى حراس المرمى لكرة القدم"، مجلة الرافدين للعلوم الرياضية (نصف سنوية)، المجلد (18)، العدد (58)
11) سكينة، شامل جاسم (2020)، "دراسة مقارنة بين اسلوب التحليل التمييزي الخطي واسلوب التحليل التمييزي اللبي"، المجلة العراقية للعلوم الادارية.
12) سليمان، علي أبشر فضل المولى، (2015)، "المقارنة بين التحليل التمييزي والنموذج اللوجستي الثنائي ونماذج الشبكات العصبية في تصنيف المشاهدات.
13) هادي، عائدة صالح، (2008)، "إستخدام التحليل المميز لتشخيص بعض أمراض العيون"، مجلة الأدارة والأقتصاد، العدد السابع والستون.
14) الياسين، دريد حسين بدر، (2009)، "أستخدام بعض طرائق التميز الحصينة لتشخيص أمراض سرطان الدم"، بحث ماجستير، الجامعة المستنصرية، كلية الأدارة والأقتصاد.
15) عبدالله، شهد نجم، (2018)، "استعمال التحليل التمييزي في تصنيف درجة التشوه للولادات"، بحث دبلوم عالي، الجامعة المستنصرية، قسم الاحصاء