توظيف أسلوبي التحليل العنقودي والتحليل التمييزي في تصنيف البيانات وبناء الدوال التمييزية

إعداد

أ.د/ محمد موسى الشمراني

قسم علم النفس، كلية التربية، جامعة الطائف، المملكة العربية السعودية

توظيف أسلوبي التحليل العنقودي والتحليل التمييزي في تصنيف البيانات وبناء الدوال التمييزية

محمد موسى الشمراني

قسم علم النفس، كلية التربية، جامعة الطائف، المملكة العربية السعودية.

البريد الالكتروني: m.m.m.a1313@gmail.com

الملخص:

هدفت الدراسة الحالية الى توظيف استخدام اسلوب التحليل العنقودي وتحليل الدالة التمييزية في تصنيف بيانات الطلبة في الأداء الأكاديمي المرتفع والمنخفض ودقة تصنيف العناقيد. استخدمت الدراسة المنهج الوصفي، تكونت عينة الدراسة من عينة عشوائية لبيانات (62) طالباً وطالبة من جامعة أم القرى للعام الجامعي 1435/1434هـ. تم تطبيق اسلوبي التحليل العنقودي والتحليل التمييزي على البيانات وقد توصلت النتائج عند تطبيق أسلوب التحليل العنقودي للمتوسطات (K-Means) وجود مجموعتين من العناقيد، في العنقود الأول كانت المسافة بين الحالات ومركز العنقود تتراوح من (2.968) الى (19.775)، وتضم (28) حالة. بينما في العنقود الثاني تراوحت المسافة بين الحالات ومركز العنقود من (1.919) إلى (12.084) وضمت (34) حالة. كما أظهرت نتائج التحليل العنقودي أهمية المتغيرات (x2) و (x3) في تصنيف الحالات على العناقيد حيث كانت دالة إحصائياً عند مستوى دلالة (0.01). كما أظهرت النتائج تأكيد نتائج التحليل التمييزي في أهمية المتغيرات المستقلة (X2) و (X3) في تصنيف الحالات. كما اشارت النتائج ان الدالة التمييزية لها ارتباط قانوني بلغت نسبته (0.770) والتي تشير التي قوة العلاقة بين المتغيرات الداخلة في التحليل، يقابلها قيمة ذاتية (Eigenvalues) تساوي (1.453) وقد فسرت الدالة (100%) من التباين. كما اشارت النتائج للتحليل التمييزي ان قيمة ولكس لمدا قد بلغت (0.408)، وقيمة كاي تربيع والتي بلغت (52.042) وهي دالة احصائياً عند مستوى دلالة (0.01)، وهذا يشير الى قدرة الدالة التمييزية على التمييز بين المجموعتين. كما أظهرت نتائج التحقق من تصنيف الحالات عند مقارنة التصنيف للحالات الذي تم وفق التحليل العنقودي فقد وجد ان التصنيف كان صحيحاً بنسبة وصلت الى (98.4%) وهي نسبة عالية جداً تؤكد دقة التصنيف. كما قدمت الدراسة مجموعة من التوصيات والمقترحات.

الكلمات المفتاحية : التحليل العنقودي، التحليل التمييزي، القيمة الذاتية، بناء الدوال التمييزية.

Employing the two methods of Cluster Analysis and Discriminant Analysis in Data Classification and building discriminative functions.

Mohammad Muosa Al-Shumrani,

Taif University, KSA.

Email: m.m.m.a1313@gmail.com

ABSTRACT:

The current study aimed at employing the use of the cluster analysis method and the discriminant function analysis in classifying student data in high and low academic performance and accuracy of classification of clusters. The study used the descriptive approach. The study sample consisted of a random sample of data for (62) male and female students from Umm Al-Qura University for the academic year 1434/1435 AH. The two methods of cluster analysis and discriminant analysis were applied to the data. The results were obtained when applying the method of cluster analysis of the mean (K-Means), the presence of two groups of clusters, in the first cluster the distance between the cases and the center of the cluster ranged from (2.968) to (19.775), and includes (28) status. Whereas in the second cluster the distance between the cases and the center of the cluster ranged from (1.919) to (12.084) and included (34) cases. The results of the cluster analysis showed the importance of the variables (x2) and (x3) in classifying the cases on the clusters where they were statistically significant at the level of significance (0.01).The results also showed the confirmation of the results of the discriminant analysis on the importance of independent variables (X2) and (X3) in the classification of cases. The results also indicated that the discriminant function has a Canonical Correlation of (0.770) which indicates that the strength of the relationship between the variables included in the analysis is offset by an intrinsic value (Eigenvalues) equal to (1.453) and the function has been explained (100%) of the variance. The results also indicated for the discriminant analysis that the value of Wilks' Lambda for a period reached (0.408), and the value of Chi-square, which amounted to (52.042) which is statistically significant at the level of significance (0.01), and this indicates the ability of the discriminant function to distinguish between the two groups. The results of checking the classification of cases also showed that when comparing the classification of cases that were done according to the cluster analysis, it was found that the classification was correct at a rate of (98.4%), which is a very high percentage confirming the accuracy of the classification. The study also presented a set of recommendations and proposals.

Keywords: Cluster Analysis, Discriminant Analysis, Eigenvalue, building discriminative functions.

مقدمة:

يحتاج الباحثون في الكثير من الدراسات والأبحاث الى معرفة تصنيف الحالات الموجودة في مجموعات متشابهة ومختلفة، وأشهر هذه الأهداف البحثية هي بناء التصنيفات، وهو احد الأسباب التي تجعل التحليل العنقودي في غاية الأهمية حيث أن الباحثين في جميع المجالات بحاجة الى اجراء التصنيفات ومراجعتها باستمرار Romesburg (2004).

ويؤكد Ramdeen and Yim(2015) أننا في الحياة اليومية، نحاول تصنيف العناصر المتشابهة معًا وتصنيفها إلى مجموعات مختلفة، وهي طريقة طبيعية وأساسية لخلق نظام تصنيفي، و من الضروري أيضًا الكشف عن أوجه التشابه داخل البيانات لبناء مجموعات ذات معنى. ويمكن استخدام تقنيات التحليل الإحصائي متعدد المتغيرات ومنها استخدام التحليل العنقودي والذي يعمل على اكتشاف نظام لتنظيم الملاحظات حيث يشترك أعضاء المجموعة في خصائص معينة مشتركة. ويعد من التقنيات الإحصائية التي تصنف الحالات إلى مجموعات متجانسة نسبيًا داخلها وغير متجانسة نسبيًا فيما بينها.

ويعتبر التحليل العنقودي والتحليل التمييزي من الأساليب الاحصائية متعددة المتغيرات واسعة الاستخدام عند تصنيف الحالات او المتغيرات، فكلاهما مناسب للتصنيف ويختلف الأسلوبان في فكرتهما الأساسية، ولكنهما يتوصلان الى نتائج لتصنيف المشاهدات.

وفي هذا الصدد يشير Hardel and Simer (2003) ان التحليل التمييزي يركز على الحالات التي تكون فيها المجموعات المختلفة معروفة مسبقاً، حيث يتم توفير قواعد القرار في تصنيف الحالات بناءً على المجموعات المعروفة. كذلك Ramdeen and Yim(2015) تؤكد أن التحليل التمييزي يقوم بتصنف الحالات الجديدة إلى مجموعات تم تحديدها مسبقًا بناءً على معايير محددة.

وعلى الرغم من وجود العديد من الدراسات السابقة التي اهتمت باستخدام كلا الأسلوبين إلا أن توظيفهما في دراسات مع بعضهم البعض مازال بحاجة الى المزيد من اجراء الدراسات والأبحاث وعلى وجه الخصوص في المجال التعليمي حيث يشير Wilson & Hardgrave (1995) إلى أن تصنيف اداء الطلاب في برامج الدراسات الجامعية، وتوقع النجاح الأكاديمي او التعثر من المشكلات المهمة التي تتطلب استخدام أساليب إحصائية متقدمة مثل التحليل العنقودي والتحليل التمييزي. وتؤكد Ramdeen

and Yim(2015) على الرغم من أن هذه التقنيات قد تم استخدامها في العلوم الاجتماعية والصحية، إلا أنها لم تكتسب نفس الشعبية واسعة الانتشار كما هو الحال في العلوم الطبيعية، ومع ذلك، لا يزال استخدام التحليل العنقودي والتحليل التمييزي في مجال علم النفس والتربية نادرًا.

لذلك تهدف الدراسة لتناول هذه المشكلة وتوضيح كيفية استخدام كل من هذين الأسلوبين في تحليل البيانات وتصنيف المشاهدات، وتحديد مدى مساهمة المتغيرات المستقلة في عملية التصنيف وتكوين الدوال التمييزية التي يمكن استخدامها في التنبؤ بتصنيف الحالات في العناقيد التي تم تحديدها.

مشكلة الدراسة وتساؤلاتها:

تكمن مشكلة الدراسة الحالية في الإجابة على التساؤلات التالية:

1. كيف يمكن تصنيف الحالات في الأداء الأكاديمي المرتفع والمنخفض باستخدام التحليل العنقودي؟
2. كيف يمكن استخدام التحليل التمييزي في تكوين دالة تمييزية تقوم بتصنيف الحالات في العناقيد؟

أهداف الدراسة:

تهدف الدراسة الحالية الى توظيف استخدام اسلوبي التحليل العنقودي وتحليل الدالة التمييزية في تصنيف بيانات لعينة من الطلاب والطالبات إلى مجموعات متجانسة وفقاً للأداء الأكاديمي المرتفع والمنخفض وتحديد مدى الاختلاف والتقارب بين هذه الحالات ودقة تصنيف المجموعتين.

أهمية الدراسة:

تنبع أهمية الدراسة الحالية من تناولها لتوظيف اسلوبي التحليل العنقودي وتحليل الدالة التمييزية في تصنيف الحالات وبناء نماذج تمييزية، كما يمكن أن تفيد نتائج هذه الدراسة في المجالات التالية:

1. الأهمية النظرية: من خلال توضيح أسلوبي التحليل العنقودي وتحليل الدالة التمييزية.

2. الأهمية التطبيقية: تقديم المساعدة للباحثين والعاملين في المجالات العلمية المختلفة في تحديد الاسلوب الامثل للتصنيف.

مصطلحات الدراسة :

التحليل العنقودي Cluster Analysis: أسلوب احصائي يستهدف تصنيف مجموعة من الحالات او المتغيرات بطرق معينة وترتيبها داخل عناقيد بحيث تكون الحالات المصنفة داخل عنقود متجانسة فيما يتعلق بخصائص محددة (الشافعي، 2014).

الدالة التمييزية Discriminant Function: هي دوال إحصائية عددها يساوي عدد مستويات المتغير التابع مطروحاً منه الواحد الصحيح او عدد المتغيرات المستقلة، ايهما اقل (Hair et al, 2006)

حدود الدراسة:

اقتصر البحث الحالي على دراسة توظيف اسلوبي التحليل العنقودي وتحليل الدالة التمييزية في تحليل بيانات عينة من طلبة الجامعية.

الحدود المكانية: طلاب وطالبات المرحلة الجامعية في جامعة ام القرى بمكة المكرمة.

الحدود الزمانية: العام الدراسي 2014/2015.

أدبيات الدراسة:

اولا: الإطار النظري:

التحليل العنقودي:

عبارة عن مجموعة من الأدوات لانشاء مجموعات (عناقيد) من بيانات متعددة المتغيرات والهدف من ذلك هو تكوين مجموعات ذات خصائص متجانسة من أصل مجموعات كبيرة غير متجانسة. ويشير Romesburg (2004) ان التحليل العنقودي اسم عام لمجموعة متنوعة من الطرق الرياضية التي يمكن استخدامها أي الحالات في مجموعة متشابهة.

يقوم التحليل العنقودي بتطوير أساليب وأدوات في تصنيف مجموعة من الافراد ويتم ذلك عن طريق تجميع الافراد " المتشابهين" وفقاً لبعض المعايير المناسبة. ويتم تطبيق التحليل العنقودي في الكثير من المجالات مثل العلوم الطبية والاقتصادية والتسويق والعلوم الطبيعية وما الى ذلك (Hardel & Simer,2003).

يشير King (2015) الى ان الاهتمام بالتحليل العنقودي قد بدأ في الستينات 1960 م في تخصصات علم الأحياء وعلم البيئة وقد أدى حدثان الى انفجار الاهتمام بالتحليل العنقودي، فقد أتاح توفر وانتشار أجهزة الكمبيوتر الكبيرة عالية السرعة إمكانيات جديدة للباحثين، بالإضافة الى ذلك، نشر مبادئ التصنيف العددي من قبل سوكال وسنث Sokal and Sneath والتي غطت ثلاث مجالات، الأول: عددا من تقنيات التحليل العنقودي المختلفة، والثاني: استخدام أجهزة الكمبيوتر في البحوث التصنيفية، والثالث: المنهج التجريبي في التصنيف في علوم الحياة والحاجة الى التحليل العنقودي في الكثير من مجالات الدراسة.

ويعد التحليل العنقودي نوع من أنواع تقنيات تقليل البياناتReduction Technique . والتي تشمل أيضًا التحليل العاملي والتحليل التمييزي، وتعمل على تخفض البيانات بشكل أساسي، على سبيل المثال، التحليل العاملي يقلل من المتغيرات الداخل في النموذج ويحولها الى عوامل والتحليل التمييزي يصنف الحالات الجديدة إلى مجموعات تم تحديدها مسبقًا بناءً على معايير محددة. يعتبر التحليل العنقودي فريدًا من بين هذه التقنيات لأن هدفه هو تقليل عدد الحالات أو المشاهدات من خلال تصنيفها إلى مجموعات متجانسة، وتحديد المجموعات دون ضرورة معرفة عضوية المجموعة أو عدد المجموعات المحتملة مسبقًا. كما يسمح التحليل العنقودي أيضًا بالعديد من الخيارات فيما يتعلق بالخوارزمية لدمج المجموعات، مع كل خيار ينتج عنه بنية تجميع مختلفة. لذلك، يمكن أن يكون التحليل العنقودي أداة إحصائية ملائمة لاستكشاف الهياكل الأساسية في أنواع مختلفة من مجموعات البيانات (Ramdeen & Yim,2015).

مقاييس التشابه او الاختلاف Measures of Similarity or Dissimilarity:

يستخدم التحليل العنقودي العديد من مؤشرات التشابه او الاختلاف بين كل زوج من المشاهدات، ويعتبر مقياس المسافة بين مشاهدتين هو المقياس المناسب لتحديد درجة التقارب بينهما، وتعتبر دالة المسافة الشائعة هي المسافة الاقليدية (Rencher,2002) والتي يمكن حسابها وفقا للصيغة التالية:

$$d(x,y)=\sqrt{\sum_{j=1}^{p}(x_j-y_j)^2} \qquad (1)$$

X_j : تمثل قيم المجموعة X

Y_j : تمثل قيم المجموعة Y

أنواع التحليل العنقودي:

1) التحليل العنقودي الهرمي Hierarchical Clustering

تمثل الأساليب العنقودية الهرمية وغيرها من خوارزميات التجميع محاولة لتحديد المجموعات الجيدة في البيانات باستخدام تقنية فعالة حسابياً (Rencher,2002). يتكون التصنيف من سلسلة من الأقسام والتي قد تمتد من مجموعة واحدة تحتوي على جميع الافراد الى مجموعات n تحتوي كل منها على فرد واحد (Everitt et al.,2011). ولا يتطلب التحليل العنقودي الهرمي المعرفة المسبقة بعدد العناقيد .

$$N(n,g)=\frac{1}{g}\sum_{k=1}^{g}\binom{g}{k}(-1)^{g-k}k^{n} \qquad (2)$$

وهناك مجموعة من الطرق يمكن استخدامها:

1- طريقة الربط المنفرد او الجوار الأقرب Single Linkage (Nearest Neighbor)

يتم على أساسها إيجاد اصغر مسافة لكل زوج من المشاهدات، وتقوم على أساس اعتبار العنصرين الأكثر تشابهاً بين العناصر يشكل نواة العنقود، ثم تضاف بقية العناصر الأكثر تشابهاً ثم الأقل (Romesburg, 2004). وتتم عملية الربط بالاعتماد على اقرب مسافة بين ازواج الحالات وربطها معا بحسب الصيغة التالية (Rencher,2002):

$$D(A,B)=\min\left\{d(y_i,y_j),y_i\in A,y_j\in B\right\} \qquad (3)$$

2- طريقة الربط الشامل او الجوار الابعد Complete Linkage (Farthest Neighbor):

تقوم هذه الطريقة باعتبار العنصرين الأقل تشابهاً بين الحالات تشكل نواة العنقود ويتم إيجاد اكبر مسافة بين الحالات وربطها معا باستخدام الصيغة التالية:

$$D(A,B)=\max\left\{d(y_i,y_j),y_i\in A,y_j\in B\right\} \qquad (4)$$

3- طريقة الربط المتوسط Average Linkage

تستخدم بالاعتماد على متوسط المسافة بين نقطة من العنقود الأول ونقطة من العنقود الثاني باستخدام الصيغة التالية:

$$D(A,B)=\frac{1}{n_A n_B}\sum_{i=1}^{n_A}\sum_{j=1}^{n_B}d(y_i,y_j) \qquad (5)$$

4- الطريقة المركزية Centroid Method

الطريقة المركزية في هذه الطريقة يكون التشابه بين مجموعتين مساوياً للتشابه بين النقطتين الوسطى، حيث تكون النقطة الوسطى للمجموعة مركز عنقود، وعلى الرغم من سهولتها الا انها لا تستخدم كثيراً في الممارسة العملية (Romesburg, 2004)

$$D(A,B)=d(\bar{y}_A,\bar{y}_B) \qquad (6)$$

5- طريقة وادز Wad's Method

طريقة تجميع التباين الدنيا وهي من الطرق الأكثر استخداماً، تتبع هذه الطريقة سلسلة من خطوات التجميع التي تبدا بالعناقيد، يحتوي كل منها على عنصر واحد وتنتهي بمجموعة واحدة تحتوي جميع العناصر، تستخدم هذا الطريقة مؤشر مجموع المربعات او التباين (Romesburg, 2004) .

$$SSE_A=\sum_{i=1}^{n_A}(y_i-\bar{y}_A)'(y_i-\bar{y}_A) \qquad (7)$$

$$SSE_B=\sum_{i=1}^{n_B}(y_i-\bar{y}_B)'(y_i-\bar{y}_B) \qquad (8)$$

$$SSE_{AB}=\sum_{i=1}^{n_{AB}}(y_i-\bar{y}_{AB})'(y_i-\bar{y}_{AB}) \qquad (9)$$

$$I_{AB}=SSE_{AB}-(SSE_A+SSE_B) \qquad (10)$$

2) التحليل العنقودي غير الهرمي Nonhierarchical Methods

طريقة المتوسطات K-Means :

من اشهر الطرق استخدامً وتتطلب هذه الطريقة تحديد عدد العناقيد مسبقاً، وتقوم هذه الطريقة على أساس تصنيف الحالات في مجموعات متجانسة من حيث الخصائص من خلال استخدام خوارزميات يمكنها معالجة عدد كبير من الحالات (احمد، 2015).

$$d(x,y)=\sqrt{\sum_{j=1}^{p}(x_j-y_j)^2} \qquad (11)$$

صدق التحليل العنقودي Cluster Validity

يمكن استخدام طريقة التحقق المتبادل للتحقق من صحة او اثبات نتيجة التجميع. يتم تقسيم البيانات بشكل عشوائي الى مجموعتين فرعيتين A, B مثلاً ويتم اجراء التحليل العنقودي بشكل منفصل على كل من A و B، ويجب ان تكون النتائج متشابهة اذا كانت المجموعات صالحة (Rencher,2002).

افتراضات التحليل العنقودي:

هناك عدداً من الافتراضات او الاشتراطات يشير شيراز (2015) الى التأكد من تحققها وهي:

1- تتطلب المتغيرات قياس معياري موحد، عندما تختلف مستويات القياس من متغير الى آخر.

2- مراعاة ان البيانات الشاذة (Outliers) تؤثر على قدرة تصنيف البيانات.

3- ينبغي عدم اهمال المتغيرات الضرورية التي قد يكون ادراجها مؤثر في التحليل العنقودي.

تحليل الدالة التمييزية Discriminant Function :

هو تركيبه خطيه من متغير أو أكثر من المتغيرات المستقلة التي من شأنها أن تميز بين مجموعتين من المشاهدات او الموضوعات (افراد، شركات، ..) في مجموعات معرفة مسبقاً. يستخدم تحليل الدالة التمييزية لتقييم مدى مطابقة تصنيف ما، وذلك بمعرفة عضوية عناصر عينة الدراسة في كل مجموعة، وكذلك التنبؤ بالموقع التصنيفي لحالات جديدة لم تصنف بعد. (Huberty,1994). ويؤكد Hardel and Simer,(2003) ان التحليل التمييزي يركز على الحالات التي تكون فيها المجموعات المختلفة معروفة مسبقاً، حيث يتم توفير قواعد القرار في التصنيف الحالات بناءً على المجموعات المعروفة.

كما يشير احمد (2015) ان التحليل التمييزي يهدف الى تصنيف الحالات في مجموعتين او اكثر بالاعتماد على مجموعة من المتغيرات وغرض التحليل هو اكتشاف نمط معين ينظم الحالات التي غالبا ما تكون افراد او أشياء، ويقسمها الى مجموعات تتمتع عناصرها بخواص مشتركة.

ويتم التحقق من التمييز كما يشير(2006).Hair et al بحساب الأوزان لكل متغير مستقل للتباين بين المجموعات بالنسبة الى التباين داخل المجموعة وتحليل الدالة التمييزية يعرف بدالة التمييز Discrimininat function . وقد اقترحه فيشر في عام 1936 م، ويستخدم للفصل بين مجموعتين واستخلاص نموذج تصنيف لإنتماء مجموعة من المشاهدات الجديدة، ويأخذ الصيغة الآتية:

$$z_{ik} = a + w_1 x_{1k} + w_2 x_{2k} + \ldots\ldots + w_i x_{ik} \qquad (12)$$

حيث z_{ik} الدرجة التمييزية، a مقدار ثابت، w_i الوزن التمييزي، x_{ik} المتغير المستقل iفي k. و يحدد عدد الدوال التمييزية بالعلاقة $(k-1)$ حيث k عدد المجموعات أو مستويات المتغير التابع، فإذا كانت البيانات تتضمن متغيراً تابعاً له ثلاثة مستويات ($k=3$) فان تحليل الدالة التمييزية سيعطي دالتين تمييزيتين يمكن استخدامهما في تصنيف الحالات الجديدة (Hair,et al,2006).

ويذكر El-Hanjouri and Hamad (2015) أن التحليل التمييزي (DA) تقنية إحصائية متعددة المتغيرات تستخدم لبناء نموذج تنبؤي تصنيفي للتمييز وتصنيف كل ملاحظة في إحدى المجموعات، والتي تسمح للفرد بفهم اختلافات الأشياء بين اثنين أو ثلاثة أو أكثر من المجموعات فيما يتعلق بالعديد من المتغيرات في وقت واحد. وتعد أول طريقة تصنيف إحصائي متعدد المتغيرات تستخدم لعقود من قبل الباحثين والممارسين في تطوير نماذج التصنيف، في التحليل التمييزي يتم استخدام السمات الكمية المتعددة للتمييز بين متغير التصنيف، يختلف عن التحليل العنقودي في ضرورة المعرفة المسبقة بالتصنيف، ويعتبر الغرض الأساسي من التحليل التمييزي هو تقدير العلاقة بين متغير تابع تصنيفي ومجموعة من المتغيرات المستقلة الكمية.

فروض تحليل الدالة التمييزية: يشير Klecka(1980) و kardiyn & Olmus(2016) أن التحليل التمييزي يتطلب تحقيق بعض الافتراضات، منها وجود مجموعتين أو أكثر، كذلك وجود حالتين على الأقل لكل مجموعة . أن تكون المتغيرات المستقلة متغيرات كمية تتبع توزيعاً طبيعياً و تحقق تجانس مصفوفة التباين و التغاير لكل المجموعات.

الدراسات السابقة:

في دراسة أجرها Al-temimi , Al-saffar and Shebib (2018) هدفت الدراسة الى استخدام التحليل العنقودي والتحليل التمييزي في تصنيف النازحين بناءً على عوامل الأوبئة، اعتمدت الدراسة على بيانات موثوقة للحكومة والمنظمات الدولية تم جمعها من إدارة المعلومات والبحوث. أظهرت النتائج ان التحليل العنقودي صنف المحافظات التي تم تهجيرها الى مجموعات متجانسة وفقاً لأنواع المخيمات، كما أظهرت نتائج التحليل التمييزي اسهام المتغيرات في تمييز المحافظات.

وفي دراسة أجرها Huang, Li, Zhan, Liu, Tiah and Fan(2017) هدفت الى استخدام التحليل العنقودي والتحليل التمييزي في دراسة العلاقة بين المركبات المتطايرة والسمات لعجائن الدقيق الصيني المخمر. تكونت العينة من سبعة أنواع من عجائن الدقيق بنكهات مختلفة، أظهرت نتائج التحليل العنقودي عن ثلاث تصنيفات، وقد تم تحديد المركبات التمييزية المساهمة في التصنيف.

كما أجرى يعقوب (2017) دراسة استخدمت بعض الأساليب الإحصائية متعددة المتغيرات والمتمثلة في التحليل العنقودي والتمييزي في دراسة تطبيقية على بعض المصارف العراقية لتصنيف وتمييز مدى التجانس بين المصارف. طبقت على عينة من (20) مصرفاً، من خلال تطبيق التحليل العنقودي تم التوصل الى الى تصنيف المصارف الى مجموعتين متجانستين في طبيعة العمل والأداء، كما تم تطبيق التحليل التمييزي لتحديد اهم المتغيرات التي عدم التجانس بين المجموعات المستخلصة من التحليل العنقودي، وقد توصلت الدراسة الى ان مجموعة من المؤشرات كانت اكثر مساهمة في التمييز بين المصارف، كما تم صياغة الدالة التمييزية التي تساعد في معرفة مدى تاثير كل متغير وكذلك التنبؤ بمعرفة انتماء أي مصرف الى المجموعة الملائمة، كما اكدت النتائج من خلال الدالة التمييزية جرى التوصل الى ان التصنيف للعشرين مصرفاً كان صحيحاً.

وفي دراسة أجراها Rose et al.(2016) هدفت الى استخدام التحليل العنقودي وتحليل الدالة التمييزية في التعرف على العوامل النفسية العصبية المنفصلة وعلاقتها بالأعراض السريرية على مجموعة من الأطفال والمراهقين. تكونت العينة من (253) طفلة ومراهقة مصابة بفقدان الشهية، توصلت النتائج ان استخدام التحليل العنقودي تم لتحديد العدد الأمثل من المجموعات، وحدد التحليل التمييزي المتغيرات التفسية والعصبية التي تميز بشكل أفضل بين المحموعات.

كما أجرى Panagopoulos, Angelopoulou, Tzirtzilakis and Giannoulopouulos (2016) دراسة لاستخدام التحليل العنقودي والتحليل التمييزي في تصنيف أنظمة طبقات المياه الجوفية المعقدة، وتتطلب الطريقة المقترحة اجراء تحليلات كيميائية لعينات المياه للايونات الرئيسية من 57 عينة مأخوذة من الابار في جزيرة ليسفوس في اليونان، توصلت نتائج التحليل العنقودي الى تصنيف العينات في اربع مجموعات متميزة، وقد اكدت نتائج التحليل التمييزي جدوى التحليل العنقودي حيث كانت نسبة تصنيف العينات (94.7%) على النحو الصحيح .

أجرى Kardiyen and Olmus (2016) دراسة أشارت الى أن مشكلة تصنيف المجموعتين لها تداعيات في عدد من المجالات , مثل الطب، والاقتصاد، والتمويل، وقد استخدمت الدراسة ثلاث نماذج للتصنيف. وتم استخدام عينات محاكاة لمختلف التوزيعات وأحجام العينات و أظهرت نتائج الدراسة أن تحليل الدالة التمييزية كان ذو أفضلية في تصنيف البيانات.

وفي دراسة أجرتها الرعيض والشاوش وحنيش (2016) هدفت الى استخدام بعض طرق التحليل العنقودي في التصنيف، وقد تم اجراء الدراسة على بعض مصادر المياه في مصراته وتصنيفها بناءً على بعض المعادن، وقد توصلت الدراسة الى تصنيف مصادر المياه الى ثلاث مجموعات، كما أظهرت نتائج الدراسة ان طريقة الربط المنفرد كانت الأفضل في التصنيف مقارنة بطريقة الربط الشامل والربط الهرمية.

وفي دراسة قام بها Tanos, Kovacs, Anda and Hatvani (2015) هدفت الى استخدام التحليل العنقودي والتحليل التمييزي في تعظيم الاستفادة من شبكة مراقبة جودة المياه على نهر تيزاس في هنغاريا مع مراعاة الموسمية، ويعتبر ثاني اكبر نهر في اوربا الوسطى، تم تحديد 15 معلمة لنوعية المياه تم قياسها في 14 موقعاً لأخذ العينات خلال الفترة الزمنية من عام 1975 الى 2005، أظهرت نتائج التحليل العنقودي أربعة تصنيفات، كما أظهرت نتائج التحليل التمييزي على التحسن المكاني الزماني لشبكة المراقبة باستخدام التحليل العنقودي والتحليل التمييزي.

وفي دراسة أجراها El-Hanjouri and Hamad (2015) قامت هذه الدراسة بتطبيق أساليب التحليل الإحصائي متعدد المتغيرات، التحليل العنقودي (CA) من أجل التعرف على التفاوت في مستويات المعيشة للأسرة بين المناطق الفلسطينية. وخلصت نتائج الدراسة إلى وجود تقارب في مستويات المعيشة للأسرة في المجموعة الثلاث الأولى لمستويات المعيشة المرتفعة، و المجموعة الثانية من مستويات المعيشة المتوسطة، كما تشكلت المجموعة الثالثة من مستويات المعيشة المنخفضة. تم تطبيق طريقة التحليل

التمييزي (DA) لتمييز المتغيرات التي تساهم بشكل كبير في هذا التفاوت بين العائلات داخل المناطق الفلسطينية، وأشارت النتائج أن متغيرات الدخل الشهري، المساعدة، الأراضي الزراعية، الحيازات الحيوانية، إجمالي الإنفاق، الإيجار المحسوب، ساهمت بشكل كبير في التصنيف.

قامت علي (2015) بإجراء دراسة هدفت الى فاعلية استخدام التحليل العنقودي والتحليل التمييزي في التحقق من الدلالة التمييزية لاختبارات الذكاء والشخصية. تكونت عينة الدراسة من (610) طالباً وطالبة من المرحلة الثانوية بمحافظة دمشق، واستخدمت الدراسة اختبار المصفوفات المتتابعة لرافن واختبار ايزنك. أظهرت نتائج الدراسة فاعلية التحليل العنقودي في الكشف عن المجموعات المتجانسة بناءً على المسافات لكلا من الاختباراين، كما أظهرت الدراسة فاعلية التحليل التمييزي في التمييز بين المجموعات وكذلك من خلال التنبؤ بعضوية الأفراد في المجموعات التي انتجها التحليل العنقودي.

كما أجرى Jaiswara, Nandi and Balakrishnan (2013) دراسة فعالية الدالة التمييزية والتحليل العنقودي في تحديد الأنواع من الكريكيت، تم استخدام أسلوب تحليل الدالة التمييزية والتحليل العنقودي في تحديد وتصنيف يعتمد على الإشارات الصوتية لأنواع الكريكيت، أظهرت النتائج التصنيف باستخدام التحليل التمييزي بدقة 95-100% ولم تتأثر بعدد الأصناف المستخدمة. كما أظهرت النتائج دقة التصنيف باستخدام التحليل العنقودي والتي لا تتطلب معرفة مسبقة، كما تشير النتائج الى فعالية التحليل العنقودي في تصنيف الأنواع وتحديدها. كما اشارت الدراسة ان التحليل التمييزي يعتبر الطريقة الأقوى والأكثر دقة لكنها تتطلب تصنيفاً مسبقاً، ويمكن استخدامها فقط مع التصنيفات المعروفة مبدئياً، اما التحليل العنقودي فيعتبر اقل قوة ودقته مرتبطة بدرجة اكبر بعدد المتغيرات التي يتم فحصها معاً، ولكن يمكن استخدامه في الحالات التي لا تكون فيها التصنيفات معروفة من قبل، ويمكن استخدام كلتا الطريقتين لتطوير أدوات كمية والية لتحديد التصنيفات المختلفة.

وفي دراسة أجرها رشيد ومهدي (2011) هدفت الى استخدام طرائق التحليل العنقودي الهرمية والغير هرمية ومعرفة أفضل الطرائق الهرمية في تحليل واقع التربية والتعليم في العراق، وشملت الدراسة (25) متغيراً في هذا المجال لمعرفة التشابه في تقديم الخدمات ومعرفة أفضل الطرائق المستخدمة في التحليل، وقد توصلت الدراسة الى ان الطريقة الهرمية افضل الطرائق المستخدمة لامتلاكها اقل قيمة للمقياس وفق مقياس العلاقة المقلص.

في دراسة قام بها Lidi-Anne & Gaston (2011) للمقارنة بين بين اسلوب تحليل الدالة التمييزية وتحليل الانحدار اللوغاريتمي الثنائي في الكشف عن قدرة بعض العوامل بالتنبؤ بعدم تعاطي الكحول خلال فترة الحمل والتي تمثلت في معتقدات الحامل نحو اضرار تعاطي الكحول والعمر للحامل أثناء فترة الحمل وقد تم أخذ بيانات لعينة عشوائية من الحوامل، تم تحليل البيانات باستخدام اسلوب تحليل الدالة التمييزية وتحليل انحدار الترجيح اللوغاريتمي وقد توصلت الدراسة الى أن اسلوب تحليل الدالة التمييزية كان أكثر دقة في تصنيف متعاطي الكحول وغير المتعاطين مقارنة باسلوب انحدار الترجيح اللوغاريتمي.

كما أجرى Maroco et al. (2011) دراسة تنبؤية استخدمت عدة طرق تنبؤية ومنها تحليل الدالة التمييزية . وقد استخدمت الدراسة بيانات حقيقية تكونت العينة من سجلات (92) حالة مريض لتصنيف الحالات الى ضعف ادراكي معرفي وغير ذلك، واستخدمت محكات الدقة والحساسية للمقارنة و توصلت إلى أن تحليل الدالة التمييزية يتمتع بدرجة عالية من الدقة والحساسية وقوة التمييز مقارنة بالطرق الاخرى.

في دراسة اجراها مصطفى (2007) وهدفت الى مقارنة طرق التحليل العنقودي على عينة من البيانات تمثلت في مجموعة من دول المغرب العربي والتي تجمعها مجموعة من الصفات المشتركة، وقد توصلت الدراسة الى تصنيف لهذه الدول، كما توصلت النتائج الى ان طرق التحليل العنقودي لا تختلف نتائجها كثيراً عن بعضها البعض.

وفي دراسة أخرى أجراها بها Hwang (2001) تشير هذه الدراسة إلى أن دقة تصنيف المجموعة بالتركيز على تحليل الدالة التمييزية، إضافة إلى ذلك فان الهدف والغرض الأساسي لهذه الدراسة هو تقديم نظرة عامة لكيفية عمل تحليل الدالة التمييزية وكيف يمكن إن يساعد في الإجابة على العديد من أسئلة الباحثين المتنوعة، إضافة إلى ما سبق توضح هذه الدراسة وتشرح ما هو تحليل الدالة التمييزية، ولماذا هو على قدر من الأهمية، وكذلك تقدم الدراسة بيانات توضيحية لإستخدام تحليل الدالة التمييزية. كما تؤكد الدراسة على أفضلية تحليل الدالة التمييزية في تصنيف المشاهدات.

في دراسة أجرها الجاعوني وغانم (2001) هدفت الى استخدام أحد أساليب التحليل الاحصائي متعدد المتغيرات التحليل العنقودي والتحليل التمييزي في دراسة لتحديد مستويات الهيكل الاقتصادي والاجتماعي لأسر المجتمع. وقد تم تطبيق هذا الأسلوب على ستة متغيرات اقتصادية واجتماعية، وقد توصلت الدراسة الى تحديد المستويات للهيكل

الاقتصادي والاجتماعي باستخدام التحليل العنقودي والفصل بين المستويات المحددة باستخدام التحليل العنقودي والتمييز بين المجموعتين من خلال التحليل التمييزي.

التعليق على الدراسات السابقة:

من خلال عرض الدراسات السابقة يمكن ملاحظة أن اغلب الدراسات السابقة ركزت على توظيف هذه الأساليب على الظواهر الاجتماعية مثل Al-temimi et al.(2018) في تصنيف النازحين ودراسة يعقوب(2017) في تصنيف المصارف، و الجاعوني وغانم (2001) في التنصيف الاقتصادي والاجتماعي والظواهر العلمية مثل دراسة Jaiswara et al. (2013) في تحديد أنواع الكراكيت و Tanos et al. (2015) في مراقبة جودة المياه و الرعيض وآخرون (2016) في تصنيف المياه بناءً على بعض المعادن ودراسة Huang et al.(2017) في تصنيف المركبات المتطايرة، والقليل من الدراسات التي قامت على حد علم الباحث بتوظيف اسلوبي التحليل العنقودي والتحليل التمييزي على متغيرات تعليمية، كما تتفق الدراسة الحالية مع الدراسات السابقة والإجراءات وتختلف في طبيعة البيانات المستخدمة، ولعل هذه الدراسة تتميز عن الدراسات السابقة في كونها تناولت مفهوماً جديداً تمثل في توظيف اسلوبين معاً من الأساليب الإحصائية متعددة المتغيرات لتصنيف الطلاب في الأداء الأكاديمي لم يتم دراسته في الأدب التربوي في حدود علم الباحث.

الطريقة والإجراءات:

منهج الدراسة: استخدمت الدراسة المنهج الوصفي .

عينة الدراسة: تكونت عينة الدراسة من عينة عشوائية لبيانات (62) طالباً وطالبة من (1722) طالباً وطالبة بجامعة أم القرى للعام الجامعي 2014/2015.

أدوات الدراسة: تم الاعتماد على بيانات عينة الطلاب والطالبات.

متغيرات الدراسة:

المتغيرات المستقلة: X_1 متغير درجات الثانوية، X_2 درجات اختبار القدرات، X_3 درجات اختبار التحصيل، X_4 : معدل السنة التحضيرية.

المتغير التابع: Y متغير تصنيفي للمعدل التراكمي، يأخذ القيمة (1) إذا كان المعدل التراكمي مرتفعاً يساوي 1.75 أو أعلى، والقيمة (2) إذا كان المعدل التراكمي منخفضاً اقل من 1.75).

المعالجة الإحصائية: تم استخدام حزمة البرامج الإحصائية لمعالجة البيانات SPSS ، لإجراء التحليل العنقودي وتحليل الدالة التمييزية للبيانات للإجابة على تساؤلات الدراسة.

نتائج الدراسة :

تم تطبيق أسلوب التحليل العنقودي للمتوسطات (K-Means) وذلك لكون هذه الطريقة تتطلب المعرفة السابقة بعدد العناقيد التي تم تصنيف الحالات على أساسها.

جدول رقم(1): توزيع الحالات على العناقيد وبعد كل حالة عن مركز العنقود

رقم الحالة	العنقود	المسافة	رقم الحالة	العنقود	المسافة	رقم الحالة	العنقود	المسافة
1	1	12.368	23	1	6.689	45	2	3.286
2	2	1.919	24	2	7.091	46	1	3.174
3	2	8.953	25	2	4.210	47	1	5.494
4	1	10.648	26	1	2.968	48	2	6.741
5	1	14.824	27	2	7.112	49	2	4.431
6	1	5.170	28	2	5.638	50	1	4.423
7	2	6.288	29	2	2.520	51	2	9.830
8	1	7.215	30	1	6.781	52	1	10.298
9	1	3.185	31	1	10.452	53	1	3.274
10	1	2.076	32	2	6.164	54	2	4.966
11	1	5.097	33	2	6.483	55	1	4.952

رقم الحالة	العنقود	المسافة	رقم الحالة	العنقود	المسافة	رقم الحالة	العنقود	المسافة
12	2	7.663	34	1	8.086	56	2	10.155
13	1	5.955	35	1	6.242	57	2	10.797
14	2	5.710	36	2	10.886	58	2	5.755
15	1	6.137	37	2	12.084	59	2	11.727
16	2	6.449	38	2	6.568	60	2	5.761
17	1	19.775	39	2	2.179	61	1	6.309
18	1	4.361	40	2	4.789	62	2	4.348
19	2	4.347	41	1	5.218			
20	2	6.508	42	1	5.243			
21	1	4.320	43	2	2.745			
22	2	15.135	44	2	4.226			

يبين الجدول رقم (1) قائمة بجميع الحالات والعناقيد التي تنتمي اليها كل حالة، حيث أظهرت نتائج التحليل وجود مجموعتين من العناقيد كما هو واضح في العمود الثاني(Cluster) والخامس والثامن،، كما يبين العمود الثالث والسادس والتاسع المسافة (Distance) او البعد بين الحالة ومركز المجموعة، فمثلاً الحالة رقم (17) تنتمي الى المجموعة الأولى وهي ابعد حالة عن مركز هذه المجموعة حيث المسافة تساوي (19.775) بينما الحالة (26) هي اقرب حالة الى مركز المجموعة او العنقود الأول، حيث تساوي المسافة (2.968)، أي ان المسافة بين الحالات ومركز العنقود الأول تتراوح من (2.968) الى (19.775)، وتضم (28) حالة موضحة في جدول رقم (1). بينما في المجموعة الثانية الحالة (2) هي الأقرب الى العنقود الثاني والمسافة تساوي (1.919) بينما الحالة رقم (37) هي الابعد عن مركز المجموعة حيث تساوي المسافة (12.084)،

أي ان المسافة تتراوح بين الحالات ومركز العنقود بين (1.919) و (12.084) وتضم (34) حالة .

جدول رقم (2): لكل متغير من المتغيرات باستخدام مجموعتي العناقيد.

المتغير	العنقود		الخطأ		F	Sig.
	متوسط المربعات	df	متوسط المربعات	df		
X1	12.964	1	11.572	60	1.120	.294
X2	1202.287	1	20.308	60	59.202	.000
X3	900.875	1	25.460	60	35.384	.000
X4	.147	1	.988	60	.149	.701

يبين الجدول رقم (2) تحليل التباين لكل متغير من المتغيرات المستقلة باستخدام مجموعتي العناقيد حيث يعطي متوسط المربعات بين المجموعتين في عمود المجموعة (Cluster)، بينما يعطي متوسط المربعات داخل المجموعات في عمود الخطأ، وبالنظر الى مستوى الدلالة للمتغيرات (x2) و (x3) نجدها دالة عند مستوى دلالة (0.01) وهذا يعني انها تلعب دوراً هاماً في تصنيف الحالات على العناقيد، بينما نجد ان المتغيرات (x1) و (x4) لم تكن دالة احصائياً، مما يعني ان هذه المتغيرات لم تسهم في تصنيف الحالات على العناقيد.

كما هو ملاحظ من نتائج التحليل العنقودي للمتوسطات (K-Means) تم تأكيد الحصول على تصنيف الحالات الى مجموعتين من العناقيد تحتوي كل مجموعة عنقود على عدداً من الحالات المتقاربة.

ولايجاد الدالة التمييزية تم الاعتماد على التصنيف المستخلص من نتائج الحليل العنقودي، وبذلك تكون المتغيرات الداخلة في التحليل التمييزي، متغير تابع ثنائي التصنيف حيث يشير (1) الى الحالات في العنقود الأول، و (2) يشير الى تصنيف الحالات في العنقود الثاني، ومتغيرات مستقلة وهي المتغيرات التي تم من خلالها التصنيف

الى عنقودين، والتي تم استخدامها لإيجاد الدالة التمييزية لتصنيف الحالات بحسب مجموعة المتغير التابع.

جدول رقم (3): جدول تحليل التباين وقيمة ولكس لمدا لمجموعة المتوسطات

المتغير	Wilks' Lambda	F	df1	df2	Sig.
X1	.982	1.120	1	60	.294
X2	.503	59.202	1	60	.000
X3	.629	35.384	1	60	.000
X4	.998	.149	1	60	.701

يلاحظ بعد إجراء التحليل التمييزي ومن نتائج جدول رقم (3) ان المتغيرات المستقلة (X2) و (X3) كانت ذات دلالة إحصائية عند مستوى (0.01)، هذا يعني ان الفروق بين متوسطات المجموعتين كانت ذات دلالة إحصائية، اما المتغيرات المستقلة (X1) و (X2) فلم تكن دالة احصائياً، وبالنظر كذلك الى قيمة ولكس لمدا (Wilks' Lambda) لهذه المتغيرات نجد انها اعلى للمتغيرات المستقلة غير الدالة وقريبة من الواحد (0.982) (0.998) , وهذا يدل على عدم وجود دلالة إحصائية.

جدول رقم (4): يوضح القيم الذاتية للدالة التمييزية

الدالة	القيمة الذاتية	%التباين	التراكمي%	الارتباط القانوني
1	1.453(a)	100.0	100.0	.770

يتضح من جدول (4) ان الدالة التمييزية يقابلها قيمة ذاتية (Eigenvalues) تساوي (1.453) بنسبة ارتباط قانوني بلغت (0.770) والتي تشير التي قوة العلاقة بين المتغيرات الداخلة في التحليل، وقد فسرت الدالة (100%) من التباين.

جدول رقم (5): يوضح قيم ولكس لمدا للدالة التمييزية

الدالة التمييزية	Wilks' Lambda	مربع كاي	df	Sig.
1	.408	52.042	4	.000

يلاحظ من الجدول (5) ان قيمة ولكس لمدا قد بلغت (0.408)، وقيمة كاي تربيع والتي بلغت (52.042) وهي ذات دلالة إحصائية عند مستوى دلالة (0.01)، وهذا يشير الى قدرة الدالة التمييزية على التمييز بين المجموعتين.

جدول رقم (6): معاملات الدالة التمييزية المعيارية

المتغيرات	معاملات الدالة التمييزية المعيارية
	1
X1	-.243
X2	.733
X3	.566
X4	-.200

يمكن من جدول (6) صياغة نموذج الدالة التمييزية المعيارية من خلال تقدير المعاملات المميزة المعيارية والتي يمكن من خلالها معرفة مدى تأثير المتغيرات على النموذج، فكلما كانت قيمة المعامل كبيرة للمتغير سواءً كانت موجبة او سالبة دل ذلك على المساهمة العالية للمتغير في دالة التمييز. وبالتالي يمكن صياغة دالة التمييز المعيارية بالشكل التالي:

$$y = -0.243x_1 + 0.733x_2 + 0.566x_3 - 0.20x_4$$

جدول رقم (7): نتائج دقة التصنيف

		رقم العنقود	عضوية المجموعة المتنبئة		المجموع
			1	2	
الأصلي	العدد	1	28	0	28
		2	1	33	34
	%	1	100.0	.0	100.0
		2	2.9	97.1	100.0
الصدق العرضي	العدد	1	28	0	28
		2	2	32	34
	%	1	100.0	.0	100.0
		2	5.9	94.1	100.0
98.4 ٪ من الحالات المجمعة الأصلية المصنفة بشكل صحيح.					

بعد إيجاد دالة التمييزية يتم التحقق من تصنيف الحالات فيما اذا كانت بالفعل تقع ضمن العنقود الذي صنفت فيه، بمعنى هل كان التصنيف صحيحاً ؟، تشير النتائج من الجدول (7) عند مقارنة التصنيف للحالات الذي تم وفق التحليل العنقودي فقد وجد ان التصنيف كان صحيحاً بنسبة وصلت الى (98.4) وهي نسبة مرتفعة جداً تؤكد دقة التصنيف.

مناقشة النتائج:

تم تطبيق أسلوب التحليل العنقودي للمتوسطات (K-Means) وذلك لكون هذه الطريقة تتطلب المعرفة السابقة بعدد العناقيد والتي أشار اليها أحمد(54 ,2015) من استخدام هذه الطريقة في حالة تحديد عدد العناقيد. أعتمدت هذه الطريقة على المسافة الاقليدية في تحديد مدى قرب الحالة من العنقود، حيث أظهر الجدول رقم (1) قائمة بجميع الحالات والعناقيد التي تنتمي اليها كل حالة، حيث أظهرت نتائج التحليل وجود مجموعتين من العناقيد، في العنقود الأول كانت المسافة بين الحالات ومركز العنقود تتراوح من (2.968) الى (19.775)، وتضم (28) حالة. بينما في العنقود الثاني تراوحت المسافة بين الحالات ومركز العنقود من (1.919) إلى (12.084) وتضم (34) حالة، وهذا يتفق مع ما تم في دراسة (Al-temimi Et al.,2018) ودراسة (Huang, et al. 2017) ودراسة يعقوب (2017) ودراسة (Rose et al., 2016

من استخدام للتحليل العنقودي في تحديد الحالات وفق المسافة من حيث القرب والبعد عن كل عنقود. كما اتضح من الجدول رقم (2) مدى أهمية المتغيرات المستقلة باستخدام مجموعتي العناقيد حيث نلاحظ أن مستوى الدلالة للمتغيرات (x2) و (x3) كانت دالة عند مستوى دلالة (0.01) وهذا يعني انها تلعب دوراً هاماً في تصنيف الحالات على العناقيد، بينما نجد ان المتغيرات (x1) و (x4) لم تكن دالة احصائياً، مما يعني ان هذه المتغيرات لم تسهم في تصنيف الحالات على العناقيد.

تم الاعتماد على التصنيف المستخلص من نتائج الحليل العنقودي لايجاد الدوال التمييزية والتي تحدد مدى مساهمة كل متغير في عملية التصنيف، تكونت المتغيرات الداخلة في التحليل التمييزي من متغير تابع ثنائي التصنيف حيث يشير (1) الى الحالات المصنفة في العنقود الأول، و (2) الى تصنيف الحالات في العنقود الثاني، و مجموعة من المتغيرات المستقلة وهي المتغيرات التي تم من خلالها التصنيف الى عنقودين، والتي تم استخدامها لايجاد الدالة التمييزية لتصنيف الحالات بحسب مجموعة المتغير التابع. يكما اتضح بعد إجراء التحليل التمييزي ومن نتائج جدول رقم (3) ان المتغيرات المستقلة (X2) و (X3) كانت ذات دلالة إحصائية عند مستوى (0.01)، اما المتغيرات المستقلة (X1) و (X2) فلم تكن دالة احصائياً، كذلك كانت قيمة ولكس لمدا (Wilks' Lambda) لهذه المتغيرات اعلى للمتغيرات المستقلة غير الدالة وقريبة من الواحد (0.998) , (0.982) وهذا يدل على عدم وجود دلالة إحصائية. وهي تتفق مع دراسة (Jaiswara, et al. 2013) التي اكدت على ضرورة استخدام كلتا الطريقتين لتطوير أدوات كمية والية لتحديد التصنيفات المختلفة. ودراسة (الجاعوني وغانم 2001) في استخدام التحليل العنقودي للتصنيف والتمييز بين المجموعتين من خلال التحليل التمييزي.

كما اتضح من جدول (4) ان الدالة التمييزية لها ارتباط قانوني بلغت نسبته (0.770) والتي تشير التي قوة العلاقة بين المتغيرات الداخلة في التحليل، يقابلها قيمة ذاتية (Eigenvalues) تساوي (1.453) وقد فسرت الدالة (100%) من التباين. و يلاحظ كذلك من الجدول (5) ان قيمة ولكس لمدا قد بلغت (0.408)، وقيمة كاي تربيع والتي بلغت (52.042) وهي ذات دلالة إحصائية عند مستوى دلالة (0.01)، وهذا يشير الى قدرة الدالة التمييزية على التمييز بين المجموعتين. تم من جدول (6) صياغة الدالة التمييزية المعيارية من خلال تقدير المعاملات المميزة المعيارية والتي يمكن من خلالها معرفة مدى تاثير المتغيرات على النموذج. وبالتالي يمكن صياغة دالة التمييز المعيارية بالشكل التالي:

$$y = -0.243x_1 + 0.733x_2 + 0.566x_3 - 0.20x_4$$

كما قامت الدراسة بعد إيجاد دالة التمييزية بالتحقق من تصنيف الحالات فيما اذا كانت بالفعل تقع ضمن العنقود الذي صنفت فيه، حيث كانت تشير النتائج من الجدول (7) عند مقارنة التصنيف للحالات الذي تم وفق التحليل العنقودي فقد وجد ان التصنيف كان صحيحاً بنسبة وصلت الى (98.4) وهي نسبة مرتفعة جداً تؤكد دقة التصنيف وهي تتفق مع ما توصلت اليه دراسة (Panagopoulos, et al., 2016) والتي اكدت نتائج التحليل التمييزي جدوى التحليل العنقودي حيث كانت نسبة تصنيف العينات (94.7%) على النحو الصحيح . وهذه النتائج تتفق مع ما قامت به دراسة ,Huang (et al. ,2017) , (Al-temimi , 2018) حيث أظهرت نتائج التحليل العنقودي التصنيفات للحالات، واستخدام التحليل التمييزي لتحديد المتغيرات التمييزية المساهمة في التصنيف . ودراسة يعقوب (2017) من خلال تطبيق التحليل العنقودي تم التوصل الى تصنيف الحالات الى مجموعتين متجانستين في طبيعة العمل والأداء، كما تم تطبيق التحليل التمييزي لتحديد اهم المتغيرات التي أدت لعدم التجانس بين المجموعات المستخلصة من التحليل العنقودي. كما تتفق مع دراسة (Rose et al.,2016) التي توصلت ان استخدام التحليل العنقودي لتحديد عدد العناقيد، ومن ثم التحليل التمييزي لتحديد المتغيرات التي تميز بشكل افضل بين العناقيد.

التوصيات والمقترحات:

1- استخدام التحليل العنقودي في تصنيف الحالات الى عناقيد متجانسة او متقاربة في كل عنقود.

2- استخدام التحليل التمييزي في تحديد المتغيرات التي ساهمت في التصنيف وتحديد الدوال التمييزية.

3- استخدام الأساليب الإحصائية المتقدمة في تصنيف الحالات الى عناقيد.

4- اجراء دراسات لمقارنة الطرق المستخدمة في حساب المسافات في التحليل العنقودي ومقارنتها.

المراجع:

أحمد، طالب (2015). تصنيف المحافظات السورية حسب الاستهلاك للاسرة باستخدام التحليل العنقودي. *مجلة جامعة تشرين للبحوث والدراسات العلمية*، (2)37.

الجاعوني، فريد وغانم، عدنان (2001). التحليل الإحصائي متعدد المتغيرات (التحليل التجميعي) في دراسة تحديد مستويات الهيكل الاقتصادي الاجتماعي لأسر المجتمع. *مجلة جامعة دمشق*، (2)17 .

رشيد، أسيل ومهدي، نبأ (2011) . تحليل واقع التربية والتعليم في العراق باستخدام طرائق التحليل العنقودي (دراسة مقارنة). *مجلة القادسية للعلوم الإدارية والاقتصادية*، (2)13.

الشافعي، محمد منصور (2014). *الإحصاء التقليدي والمتقدم في البحوث العلمية والإنسانية (الكتاب الثاني)*. الرياض: مكتبة الرشد.

شيراز، محمد صالح (2015). *التحليل الاحصائي للبيانات SPSS*. جدة: خوارزم العلمية.

علي، كنان أحمد (2015). *فاعلية استخدام التحليل العنقودي والتحليل التمييزي في التحقق من الدلالة التمييزية لاختبارات الذكاء والشخصية*. (دارسة ميدانية مقارنة في محافظة دمشق). رسالة ماجستير غير منشورة، كلية التربية، جامعة دمشق.

مصطفى، نزار (2007). استخدام بعض طرق التحليل العنقودي في التصنيف مع تطبيق عملي. *مجلة التقني*، (2)20.

يعقوب، أسماء أيوب (2017). التحليل العنقودي والتمييزي في دراسة تطبيقية على بعض المصارف العراقية. *مجلة الاقتصاد الخليجي*، ع 31.

Aldenderfer,M. & Blashfield, R.(1984). *Cluster Analysis*. London: SAGA Publications.

AL-Temimi, S., AL-Saffar, R. & Shebib, H.(2018). Classification and Identification of IDP Camps After Mosul Events Based on Epidemics and Other Factors Using Cluster Analysis and Discriminant Analysis. *International*

***Journal of Pharmaceutical Research &Allied Sciences*, 2018, 7(3): 63-73**

El-Hanjouri,M. & Hamad, B.(2015). Using Cluster Analysis and Discriminant Analysis Methods in Classification with Application on Standard of Living Family in Palestinian Areas. International Journal of Statistics and Applications 2015, 5(5): 213-222 DOI: 10.5923/j.statistics.20150505.05.

Everitt, B. , Landau, S. , Leese, M. & Stahi, D. (2011).*Cluster Analysis (5th ed)*.Londen: Wily Series.

Giniyatullin, K., Valeeva,A. & Smirnova,E. (2017). Application of Cluster and Discriminant Analyses to Diagnose Lithological Heterogeneity of the Parent Material according to Its Particle-Size Distribution. *Eurasian Soil Science*, 50(8), pp. 917–924.

Hair, J. , Black, W. , Babin, B., Anderson, R. & Tatham, R.(2006). *Multivariate Data Analysis (6th ed)*. New Jersey: Pearson Education.

Hardle, W., Simar, L. (2003). *Applied Multivariate Statistical Analysis*. Berline: Springer.

Hawng, D.(2001).Issues in Predictive Discriminant Analysis: Using and Interpreting the Leave-One-Out Jackknife Method and the Improvement-Over-Change " I" Index Effect Size. Paper Presented at the Annual Meeting of the Southwest Educational Research Association.

Huang, M.,Li, Y. ,Zhan,P., Liu,P.,Tian,H. &Fan,J. (2017). Correlation of Volatile Compounds and Sensory Attributes of Chinese Traditional Sweet Fermented Flour Pastes Using Hierarchical Cluster Analysis and Partial Least Squares-Discriminant Analysis. *Journal of Chemistry* , ID 3213492, https://doi.org/10.1155/2017/3213492.

Huberty, C. (1994). *Applied Discriminant Analysis*. New York: John Willy & Sons.

Huberty, C. ; Olejnik, S. (2006).*Applied MANOVA and Discriminant Analysis (2nd ed)*. New Jersey: John Wiley & Sons.

Inoue, Y.(2001). Educational Research and Statistics: Examples of Questions and Answers. ERIC; ED 451242.

Jaiswara,R., Nandi,D. & Balakrishnan,R. (2013). Examining the Effectiveness of Discriminant Function Analysis and Cluster Analysis in Species Identification of Male Field Crickets Based on Their Calling Songs. PLoS ONE 8(9): e75930. doi:10.1371/journal.pone.0075930

Kardiyen,F. ; Olmus, H. (2016). A Comparison of Two Group Classification Approaches to Fat tailed and Skewed Data. Communication in Statistics-Simulation and Computation, 45, pp 17-32.Taylor&Francis, ISSN:0361-0918.

King, R. (2015). Cluster Analysis and Data Mining. Dulles: Mercury Learning and Information.

Klecka, W. (1980). *Discriminant Analysis*. California: Sage Publications.

Kutner, M. ; Neter, J. and Li, W.; Nachtsheim, C. (2005).*Applied Linear Statistical Models.(5nd ed)*.Irwin, McGraw-Hill, New York, U.S.A.

Li, H. ; Sun, j. (2011). Empirical Research of Hybridizing Principal Component Analysis with Multivariate Discriminant Analysis and Logistic Regression for Business Failure Prediction. Expert Systems with Application, (38).6244-6253. www.elsevier.com/locate/eswa.

Manly, B. (1994). Multivariate Statistical Methods A Primer. 2nd, Chapman & Hall.

Maroco, J. ; Silva,D.; Rodrigues, A.; Guerreiro,M ; Santana,L. ; Mendonca, A. (2011).Data Mining Methods in the Prediction of Dementia : A Real – Data Comparison of the Accuracy , Sensitivity and Specificity of Linear Discriminant Analysis, Logistic Regression , Neural Networks, Support vector machines, Classification Trees and Random forests. BMC Research Notes, 4 ; 299. Portugal.

Moor, J. (1996). Stepwise Methods are as in Discriminant Analysis as They are Anywhere Else. Paper Presented at the Annual Meeting of the Southwest Educational Research Association.

Nath, R. ; Jackson, W.(1992).A Comparison of the Classical and the Linear Programming approaches to the Classification

Problem in Discriminate. Journal of Statistical Computation and Simulation, 41, pp73-93.

Panagopoulos, G. , Angelopoulou, D., Tzirtzilakis, E. & Giannoulopoulos, P. (2016). The contribution of cluster and discriminant analysis to the classification of complex aquifer systems. Environ Monit Assess, (188) 591, DOI 10.1007/s10661-016-5590.

Pluker, A. (1995). Application of Discriminant Analysis in Research with Gifted Students. Paper Presented at the Annual Meeting of the American Educational Research Association.

Pohar, M. ; Blas, M. ; Turk, S. (2004). Comparison of Logistic Regression and linear Discriminant Analysis : A simulation study. Metodoloski , 1(1), pp143-161.

Ramdeen,K. and Yim, O. (2015). Hierarchical Cluster Analysis: Comparison of Three Linkage Measures and Application to Psychological Data. The Quantitative Methods for Psychology, 11(1). https://www.researchgate.net/publication/308015073

Rencher, A. (2002). Methods of Multivariate Analysis (2nd ed).Canada: A Wiley Interscience.

Richardson,A. & Abu Alhaija,E. (2003). Growth prediction in Class III patients using cluster and discriminant function analysis. European Journal of Orthodontics, 25, 599–608.

Romesburg, H. (2004). *Cluster Analysis for Researchers*. North Carolina: Lulu Press.

Rose,M., Stedal,K., Reville,M.,Noort,B. Kappel,V., Frampton,L., Watkins,B.& Lask ,B.(2016). Similarities and Differences of Neuropsychological Profiles in Children and Adolescents with Anorexia Nervosa and Healthy Controls Using Cluster and Discriminant Function Analyses. Archives of Clinical Neuropsychology ,31, 877–895.

Ruiz, S. (1991). Asymptotic Efficiency of Logistic Regression Relative to Linear Discriminant Analysis. Biometrika , V (78),2,pp.235-243.

Santos, F.; Guyomarch, P.; Bruzek, J.(2014). Statistical Sex Determination from Craniometrics: Comparison of Linear Discriminant Analysis, Logistic Regression , and Support Vector Machines. *Forensic Science International* ,245(205) .www.elsevier.com/locate/forscint.

Tabachnick, B. & Fidell, L. (2013).*Using Multivariate Statistics.(6th ed)*. Pearson Education,U.S.A.

Tanos,P. ,Kovács, J., Kovács, S., Anda, A. & Hatvani,I.(2015). Optimization of the monitoring network on the River Tisza (Central Europe, Hungary) using combined cluster and discriminant analysis, taking seasonality into account. *Environ Monit Assess*, (187) , DOI 10.1007/s10661-015-4777.

Thompson, B.(1994). Why Multivariate Methods are Usually Vital in Research: Some Basic Concepts. Paper Presented at the Annual Meeting of the Southwest Educational Research Association (Dallas, TX, Jun 27-29)

Wilson, L. & Hardgrave, C. (1995). Predicting Graduate Student Success in an MBA Program: Regression Versus Classification. *Educational and Psychological Measurement,* 55(2) , April 1995 , pp186-195.