

استخدام الدالة التمييزية الخطية واللاخطية مع التحليل العنقودي لدراسة مستوى التعليم
للمراحل المنتهية (الحكومية- الأهلية) في محافظة نينوى

د. صفوان ناظم راشد **

Safwan75nathem@uomosul.edu.iq

زينب عادل شهاب*

Zainab.csp159@student.uomosul.edu.iq

المستخلص:

الفكرة الأساسية من هذا البحث هي دراسة المتغيرات المؤثرة والتي أدت الى وجود فروقات بين اداء المدارس الحكومية والأهلية ، ومن خلال استخدام بعض الاساليب الاحصائية الخطية والمتمثلة بدالة التحليل التمييزي والاساليب اللاخطية المتمثلة بدالة الانحدار اللوجستي والتحليل العنقودي ، حيث تمت المقارنة بين مجموعتي المدارس الحكومية والأهلية ومعرفة مدى ملائمة كل من الاساليب المستخدمة مع البيانات التطبيقية لمجتمع الدراسة والذي يضم بيانات المرحلة المنتهية للعام الدراسي(2018-2019) لكل من المدارس (الابتدائية - المتوسطة - الاعدادية(الاحيائي)) في مديرية تربية محافظة نينوى، لتحديد المجموعة والمرحلة الافضل اضافة الى تحديد المتغيرات الأكثر تأثيرا على الاداء وذلك للوصول الى تحسين وتوحيد المسار التعليمي. الكلمات المفتاحية: الدالة التمييزية الخطية ، الانحدار اللوجستي الثنائي ، التحليل العنقودي الهرمي.

Using the linear and non-linear discriminant function with cluster analysis to study the level of education for the completed stages (governmental – private) In Nineveh Governorate

Extracted:

The basic idea of this research is to study the influential variables that led to the differences between the performance of public and private schools, and through the use of some linear statistical methods, which are the function of discriminatory analysis and nonlinear methods represented by the function of logistic regression and cluster analysis, where the comparison between the two groups of public and community schools and know the appropriateness of each of the methods used with the applied data of the study community, which includes data The finished stage for the school year(2018-2019) of each of the schools (primary- middle- preparatory(Bio)) in the Nineveh Governorate Education Directorate, to determine the group and the best stage in addition to identifying the variables that affect the performance in order to reach improvement and standardization of the educational path.

Keywords: linear discriminatory function, binary logistic regression, cluster or pyramidal analysis.

طالبة دراسات عليا/دبلوم عالي/ كلية علوم الحاسوب والرياضيات/قسم الاحصاء والمعلوماتية/ جامعة الموصل
**مدرس/ كلية علوم الحاسوب والرياضيات/قسم الاحصاء والمعلوماتية/ جامعة الموصل

تاريخ النشر: 2021/6/1

تاريخ القبول: 2021/2/10

تاريخ استلام البحث: 2021/1/29

المقدمة:

عند تحليل البيانات الثنائية "Binary Data" والتي يأخذ المتغير التابع y فيها احدى القيمتين (0,1) مثلا مصاب أو غير مصاب ، فانه يفضل استخدام اسلوب دالة التحليل التمييزي Discriminant Analysis أو استخدام اسلوب تحليل الانحدار اللوجستي Logistic Regression والذان يعتبران من الاساليب الاحصائية المهمة في تصنيف مفردة أو أكثر الى احدى المجتمعات بالاعتماد على قيم المتغيرات المستقلة لتلك المفردة أو المشاهدة، ومن ثم تم توظيف الاسلوبين مع اسلوب التحليل العنقودي لمعرفة تأثير كل متغير داخل كل مجموعة على حدى، حيث أن هناك اختلاف في الفكرة الاساسية للأساليب المستخدمة لكن في النهاية فانهم يتوصلون الى نتائج لتصنيف المشاهدات. على وجه الخصوص فان Wilson & Hardgrave(1995) يشير الى أن تصنيف اداء الطلاب في المدارس وتوقع النجاح أو التعثر من المشكلات المهمة التي تتطلب استخدام اساليب احصائية متقدمة ، وقد اقتصر البحث توظيف الاساليب المستخدمة في تحليل بيانات عينة من طلبة مدارس محافظة نينوى ومقارنة ادائهم في المدارس والمرحلة العائدين لها من خلال اخذ مجموعة من طلاب وطالبات المراحل المنتهية للمدارس الحكومية والأهلية في المحافظة للعام الدراسي 2018-2019. مبررات اختيار الموضوع:

1. تناسب الموضوع مع وضع وحالة التعليم في الوقت الحاضر .
2. الرغبة في معرفة أفضلية الاداء في كل مجموعة من مجموعات الدراسة.
3. بناء نموذج رياضي يُمكن التوصل الى تحديد الاسباب جود الفرق بين المجموعات لمعالجتها.

1- مفهوم التحليل التمييزي: Discriminant Analysis

أن اول من استخدم التحليل التمييزي هو Karl Pearson عام(1921) عندما اقترح اسلوب احصائي أطلق عليه معاملات التشابه للأشياء المتماثلة Coefficient of racial likeness ، أما Fisher فانه وفي عام (1931) أوجد دالة خطية لتصنيف المفردة الى احدى المجموعتين مع تساوي التباينات وأطلق عليها الدالة التمييزية الخطية Linear Discriminant Function ومنها أصبحت عامة للمجتمعات متعددة المتغيرات(الشكرجي وآخرون، 2008). الفكرة الاساسية من التمييز هي الفصل بين المجتمعات المتداخلة ولها نفس الخصائص والصفات ، حيث أن مهمة التحليل التمييزي هو ايجاد دالة يمكن من خلالها تصنيف المفردات الجديدة الى احدى مجتمعات الدراسة(دخيل وآخرون، (2014). التمييز والذي يسمى بدالة Fisher هو طريقة صحيحة للتصنيف لو توفرت شروط تطبيقها من حيث التوزيع الطبيعي للمتغيرات وتساوي مصفوفات التباين والتباين المشترك ويستخدم في عملية التوقع أن الباحث ينتبأ بعودة مفردات أو مشاهدات ويتوقع تمييزها بين مجموعتين في المستقبل (سليمان، 2012).

2-الدالة المميزة الخطية بين مجموعتين : Linear Discriminant Function Between 2Groups

دالة التمييز هي نموذج يمكن صياغته اعتماداً على مؤشرات العينة التي تم اختيارها ووضعت في مجموعتين مختلفتين ، وبواسطة هذه الدالة نستطيع أن نختبر المفردة الجديدة الى اي المجموعتين عائدة. بالتالي فان التركيب الخطي بشكل عام يكون حسب الصيغة الاتية(دخيل، 2009):

$$L = a_1x_1 + a_2x_2 + \dots + a_kx_k \quad \dots \dots \dots (1)$$

وأن (a_1, a_2, \dots, a_k) يمثل عدد معاملات الدالة التمييزية ، كما أن k تمثل عدد المتغيرات داخل كل مجموعة، حيث يتم تقدير معاملات الدالة باتباع الصيغة والخطوات الاتية:

$$a_j = v^{-1}d \quad ; j = 1, 2, \dots, k \quad \dots \dots \dots (2)$$

وحيث أن :

v^{-1} : معكوس مصفوفة التباين-التباين المشترك. d : الفروق بين متوسطات المشاهدات للمتغيرات.

1- ايجاد متوسط كل متغير في كل مجموعة وكالاتي :

$$\bar{x}_{k(p)} = \frac{\sum_{i=1}^{n_p} x_{n(p)}}{n_p} \quad \dots \dots \dots (3)$$

$P =$ عدد المجموعات قيد الدراسة (1,2) ، $k =$ عدد المتغيرات داخل المجموعات (1,2,3,.....)

2- ايجاد الفرق (المسافة) بين متوسطي كل متغيرين في كل من المجموعتين:

$$d_j = \bar{x}_{k(1)} - \bar{x}_{k(2)} \quad \dots \dots \dots (4); \quad i=1,2,3, \dots, n$$

3- ايجاد مصفوفة التباين والتباين المشترك (Variance-Covariance) وهي عبارة عن التباينات المجمعة بين كل متغير في مجموعة ونفس المتغير في المجموعة الثانية:

$$V = \begin{bmatrix} v_{11} & \dots & v_{1k} \\ \vdots & \ddots & \vdots \\ v_{n1} & \dots & v_{nk} \end{bmatrix} ; v_{ii} = \frac{S_i^2(1) + S_i^2(2)}{n(1) + n(2) - 2} \quad \dots \dots \dots (5)$$

$$S_i^2 = \sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n} \quad \dots \dots \dots (6)$$

أما بالنسبة للتغايرات المشتركة فيتم حسابها كالاتي:

$$v_{ij} = \frac{S_{ij}^2(1) + S_{ij}^2(2)}{n(1) + n(2) - 2} \quad \dots \dots \dots (7)$$

$$S_{ij} = \sum_{i=1}^n x_i x_j - \frac{\sum_{i=1}^n x_i \sum_{j=1}^n x_j}{n} \quad \dots \dots \dots (8)$$

3- نقطة الفصل : Cut Of Point

لغرض الفصل بين المجموعتين قيد البحث لابد من وجود نقطة للفصل بهدف تصنيف مفردة جديدة لإحدى المجموعتين ويعبر عنها بالصيغة الاتية(Copper and Jweaks, 1984):

$$l = \frac{\bar{l}_{(1)} + \bar{l}_{(2)}}{2} \quad \dots \dots \dots (9)$$

حيث أن $\bar{l}_{(1)}$ و $\bar{l}_{(2)}$ هي متوسط القيم التمييزية للمجموعة الاولى والمجموعة الثانية وتحسب قيم كل منها حسب الصيغ الآتية:

$$\bar{l}_{(1)} = \bar{x}_1 v^{-1} (\bar{x}_1 - \bar{x}_2) \quad \dots \dots \dots (10)$$

$$\bar{l}_{(2)} = \bar{x}_2 v^{-1} (\bar{x}_1 - \bar{x}_2) \quad \dots \dots \dots (11)$$

فاذا كانت قيمة نقطة الفصل أقل أو تساوي قيمة الدالة التمييزية للمشاهدة أو المفردة فإنها تُصنف ضمن المجموعة الاولى والا فإنها تُصنف ضمن المجموعة الثانية.

4- اختبار كفاءة الدالة التمييزية الخطية: The Efficiency Of The Linear Discriminant Function

تتعدد اساليب التحليل التمييزي وذلك في ضوء تحقق الافتراضات اللازم توفرها عند استخدام كل منها، حيث تعرف بالدالة التمييزية اللاخطية عند عدم تساوي مصفوفتي التباينات المشتركة ومنها ما تم استخدامها في هذا البحث والمتمثلة بدالة الانحدار اللوجستي Logistic Regression ، كما ويتم اعتماد الطريقة المناسبة عندما يكون خطأ التصنيف أقل ما يمكن وتعد الدالة التمييزية الخطية من الاساليب المستخدمة في هذا الصدد (Fisher, 1936) والتي قُدمت من قبل Fisher:

1- اختبار معنوية الفروق بين المجتمعات: test of differences between communities

لغرض اختبار معنوية الفروق بين متوسطات المجتمعات يتم اختبار الفرضية القائلة بان هناك تساوي بين المتوسطات ضد الفرض البديل وكالاتي وبحسب الصيغة الموضحة:

$$H_0: \mu_1 = \mu_2 = \dots = \mu_k \quad \& \quad H_1: \mu_1 \neq \mu_2 \neq \dots \neq \mu_k$$

$$F = \left(\frac{n_1 n_2}{n_1 + n_2} \right) \left(\frac{n_1 + n_2 - k - 1}{n_1 + n_2 - 2} \right) (D^2) \quad \dots \dots \dots (12); \text{ d. f} = (k, n_1 + n_2 - k - 1)$$

$$D^2 = (\bar{x}_{(1)} - \bar{x}_{(2)}) v^{-1} (\bar{x}_{(1)} - \bar{x}_{(2)}) \quad \dots \dots \dots (13)$$

وبمقارنة القيمة المحسوبة للمختبر الاحصائي مع القيمة الجدولية وعند مستوى معنوية معين يتم اتخاذ القرار حول رفض أو قبول فرضية العدم، فإذا كانت القيمة المحسوبة للمختبر أكبر من القيمة الجدولية عندها ترفض فرضية العدم ويقبل الفرض البديل بمعنى أن هناك فروق معنوية بين متوسطات المجموعات، والا فيتم قبول الفرض العدم.

2- اختبار تجانس مصفوفات التباين والتباين المشترك: Matrix Covariance Test

تعد أهم الفرضيات لكي يتم اعتماد الدالة التمييزية الخطية تُحقق تجانس مصفوفتي التباين وهي:

$$H_0: \Sigma_1 = \Sigma_2 = \dots = \Sigma_K \quad \& \quad H_1: \Sigma_1 \neq \Sigma_2 \neq \dots \neq \Sigma_K$$

ولتطبيق هذا الاختبار نستخدم اختبار (Bartlett test) والمقترح من قبل العالم (Bartlett) عام (1937) والذي يمتاز بكفاءة عالية وهو الأكثر شيوعاً، وصيغته المختبرية كالتالي :

$$M = \sum_{i=1}^p n_i \ln|S| - \sum_{i=1}^p n_i \ln|S_i| \dots\dots\dots(14)$$

وقد أثبت (Box) عام (1949) أنه إذا تم ضرب قيمة الصيغة اعلاه والمتمثلة بـ M في ثابت والذي يُرمز له بـ C^{-1} ويأخذ الصيغة الآتية :

$$C^{-1} = 1 - \frac{2k^2+3k-1}{6(k+1)(p-1)} \left[\sum_{i=1}^p \frac{1}{n_i} - \frac{1}{\sum_{i=1}^k n_i} \right] \dots\dots\dots(15)$$

وبالنتيجة فإننا نحصل على مقياس يتوزع بالتقريب توزيع (Chi-Square) بمستوى معنوية معين وعند درجة حرية $(k-1)(p-1)$ حيث p تمثل عدد المجموعات و k تمثل عدد المتغيرات، ومن ثم نُقارن القيمة المحسوبة للمقدار (MC^{-1}) مع قيمة المختبر Chi، فإذا كانت القيمة المحسوبة أقل من القيمة الجدولية فإننا نقبل فرضية العدم ، ونرفض الفرضية البديلة التي تُؤكد تجانس التباين.

5-تقدير نسبة الخطأ: Estimate Of Error Rate

بعد ايجاد نقطة الفصل وقيمة الدالة التمييزية للمشاهدة، لابد من ايجاد نسبة الخطأ لتصنيف اي مشاهدة، حيث أن هناك نوعين من اخطاء التصنيف(ابراهيم،2010):

1-الخطأ الظاهري: يمثل عدد المفردات التي صُنفت خطأ ظاهرياً وله حالتين:

الاولى: تصنيف مفردات للمجموعة الثانية في الاصل كانت تنتمي للمجموعة الاولى ويُرمز لعدد المفردات التي صُنفت خطأ بالرمز (n_{12}) وأن النسبة لها $p_{12} = \frac{n_{12}}{n_1}$ حيث يمثل نسبة المفردات التي تنتمي للمجموعة الاولى وصُنفت للمجموعة الثانية. والثانية: تصنيف مفردات للمجموعة الاولى وفي الاصل هي تنتمي للمجموعة الثانية ويرمز لعدد المفردات عندئذ بالرمز (n_{21}) وأن النسبة $p_{21} = \frac{n_{21}}{n_2}$ حيث أنها تمثل نسبة الخطأ للمفردات التي تنتمي للمجموعة الاولى وصُنفت للثانية.

2-الخطأ الحقيقي: نسبة التصنيف الحقيقي الخاطئ في المجتمع، يتم حسابه وفق المعادلة الآتية:

$$p_{12} = p_{21} = f\left(-\frac{\sqrt{D^2}}{2}\right) \dots\dots\dots(16)$$

بعد حساب قيمة داخل القوس يتم استخراج الاحتمال المقابل لها من جدول التوزيع الطبيعي المعياري، فكلما كان الاحتمال صغيراً وقريباً من الصفر دل ذلك على أن خطأ التصنيف كان ضئيلاً مما يدل على قدرة الدالة على التمييز، أما عندما تقترب قيمة الاحتمال من الواحد دل على أن احتمال خطأ التصنيف الحقيقي كبيراً مما يدل على ضعف الدالة على التصنيف والتمييز.

6- مفهوم نموذج الانحدار اللوجستي الثنائي: Binary Logistic Regression Model

تكمن أهمية الانحدار اللوجستي عند مقارنته بالأساليب الاحصائية الاخرى مثل (الانحدار الخطي والتحليل التمييزي)، في أنه أكثر قوة لأنه يقدم اختبار المعنوية للمعاملات، كما أنه يعطي الباحث فكرة عن مقدار تأثير المتغير المستقل في المتغير المعتمد النوعي ثنائي القيمة بالإضافة الى أنه يسمح للباحث بالاستنتاج بان متغيراً ما يعتبر اقوى تأثيراً من المتغير الاخر في فهم ظهور النتيجة المطلوبة (Lea, 1997). الانحدار اللوجستي الثنائي هو نوع من انواع الانحدار يكون فيه المتغير المعتمد متغير نوعي قد يأخذ قيمتين (الانحدار اللوجستي الثنائي) وقد يأخذ أكثر من قيمتين (الانحدار اللوجستي المتعدد)، الا أننا في الانحدار اللوجستي الثنائي يكون هدفنا ليس تفسير التغيير في قيم المتغير المعتمد وإنما تفسير احتمال حدوث وعدم حدوث ظاهرة معينة، وأن من مزايا الانحدار اللوجستي الثنائي أنه أقل حساسية تجاه الانحرافات عن التوزيع الطبيعي لمتغيرات الدراسة وذلك مقارنة مع اساليب احصائية اخرى، بالإضافة الى أنه يستطيع أن يتجاوز العديد من الافتراضات المقيدة لاستخدام طريقة المربعات الصغرى OLS في الانحدار الخطي (الفهود، 2014).

يقوم النموذج اللوجستي الثنائي على فرض اساسي وهو أن متغير معتمد يمثل ب (y) يأخذ القيمة (1) اذا حدث معين باحتمال (p)، ويأخذ القيمة (0) اذا لم يحدث ذلك الحدث وباحتمال (1-p)، اي حدوث الاستجابة من عدمها، وبما أن المتغير المعتمد في النموذج المقدر هو متغير برنولي يأخذ احدي القيمتين (0,1)، فان الدالة اللوجستية تأخذ الصيغة الاتية (عدلي وأبوشكان، 2014):

$$P_{x_i} = E(y_i | x_i) = \frac{e^{(B_0 + B_1 x_1 + \dots + B_k x_k)}}{1 + e^{(B_0 + B_1 x_1 + \dots + B_k x_k)}} \quad \dots \dots \dots (17)$$

حيث أن: P_{x_i} تمثل احتمال حدوث الاستجابة (i) من المتغير المعتمد.

من خلال المعادلة يتضح أن العلاقة بين المتغير المعتمد والمتغيرات المستقلة (المواد الدراسية لكل مرحلة من المراحل الثلاثة) علاقة غير خطية، لذلك هناك عدة تحويلات يمكن القيام بها لجعل العلاقة خطية ومن أشهر هذه التحويلات هو تحويل (Logit) (عباس، 2012)، فتصبح المعادلة بالصيغة الاتية:

$$L = \ln \left(\frac{p(x_i)}{1 - p(x_i)} \right) = B_0 + B_1 x_1 + \dots + B_k x_k \quad \dots \dots \dots (18)$$

من المعروف في الانحدار أن الطرف الايمن لهذه النماذج يأخذ قيماً من $(-\infty)$ الى $(+\infty)$ ولكن عندما يكون لدينا متغيران احدهما ثنائي (y) فان الانحدار الخطي البسيط لا يكون ملائم لان: $E(y|x) = p(y=1) = p$ وبالتالي يكون الطرف الايمن محصوراً بين (0,1) وبالتالي يكون النموذج غير قابل للتطبيق من وجهة نظر الانحدار (الجاعوني وغانم، 2011).

7- تقدير معلمات النموذج اللوجستي الثنائي: Estimate The Parameters Of The Binary Logistic Model

أشار *Lea* عام (1997) انه في ظل توافر الشروط للانحدار الخطي فان طريقة المربعات الصغرى الاعتيادية تعطي مقدرات تتصف بانها مقدرات خطية غير متحيزة ولها أقل تباين، حيث تبحث طريقة المربعات الصغرى عن أقل مجموع لمربعات انحرافات البيانات المشاهدة عن خط الانحدار، ولكن يتم استبعاد استخدام طريقة المربعات الصغرى لتقدير نموذج الانحدار اللوجستي، لذا نستخدم طريقة الامكان الأعظم لحساب المعاملات اللوجستية وتهدف هذه الطريقة الى تعظيم لوغارتم الاحتمال الذي يعكس مدى امكانية احتمال أن تكون تلك القيم المشاهدة للمتغير المعتمد في الامكان توقعها أو التنبؤ بها من خلال المتغيرات المستقلة (عابد، 2015)، حيث أن Logit هو اللوغارتم الطبيعي لمعاملات الاحتمال ويعبر عنه بدلالة الاحتمالات بالصيغة الآتية:

$$\text{Logit} = \ln\left(\frac{p}{1-p}\right) \dots \dots \dots (19)$$

أما بالنسبة لتفسير معاملات الانحدار اللوجستي الثنائي بدلالة الاحتمالات حيث أن الزيادة في المتغير المستقل بمقدار وحدة واحدة ستزيد $\ln\left(\frac{p}{1-p}\right)$ بان يكون المتغير المعتمد $(y=1)$.

8- اختبار معنوية معلمات النموذج اللوجستي الثنائي: Test Of Binary Logistic Model Parameters

عند بناء النموذج اللوجستي الثنائي نجد أن من المفيد الاجابة على السؤال "هل يمكن حذف اي من المتغيرات في نموذج الانحدار؟" وللإجابة على السؤال لابد من اجراء اختبار ما اذا كانت معلمات الانحدار B_j تساوي الصفر أم لا، لذا يتم استخدام اختبار Wald لاختبار معنوية معالم النموذج حيث أنه يختبر فرض العدم ضد الفرض البديل $H_0: B_j = B_{j0} \& H_1: B_j \neq B_{j0}$ ، والذي يستند على الاحصائية الآتية:

$$w = \frac{\hat{B}_j}{S.E(\hat{B}_j)} \dots \dots \dots (20)$$

حيث أن:

$S.E(\hat{B}_j)$ هي مقدر الخطأ المعياري للمعاملات \hat{B}_j والتي تتبع التوزيع الطبيعي، لهذا تُقارن قيمتها المشاهدة بالقيم الحرجة (p-value).

9- مفهوم التحليل العنقودي: Cluster Analysis

عبارة عن اجراءات تصنف مجموعة حالات أو متغيرات بطرق معينة وترتيبها داخل عناقيد بحيث تكون الحالات المصنفة داخل العنقود الواحد متجانسة فيما بينها في حين أنها تختلف عن الحالات والمتغيرات الموجودة داخل العناقيد الأخرى، عليه يمكن الاستنتاج بان اساس التحليل العنقودي هو ترتيب المتغيرات بشكل عناقيد بحيث يعمل على تصغير التباين داخل العنقود الواحد وتعظيم التباين بين العناقيد

المختلفة (سليمان وصالح، 2012)، كما أن التحليل العنقودي يمثل نوعاً من الأساليب المستخدمة في تجميع المفردات وتقليص عددها ويطلق أيضاً على عملية التحليل العنقودي بعملية التحليل التصنيفي أو عملية التصنيف الرقمي (Michael, 2001).

10- أنواع التحليل العنقودي : Type Of Cluster Analysis

يوجد نوعان من التحليل العنقودي هما الهرمية (Hierarchal Cluster Analysis) وغير الهرمية (طريقة المتوسطات) (K-means) المستخدمتان في تصنيف البيانات.

في هذا البحث تم التطرق إلى استخدام طريقة التحليل العنقودي الهرمي والذي يعد من الأساليب المفضلة في التحليل العنقودي وفيه يتم عنقدة (k) من المتغيرات وبشكل متسلسل في (m) من العناقيد إذ يكون (c₁) أضعف العناقيد و (c_m) أكثر قوة.

11- أساليب العنقدة الهرمية : Hierarchical Cluster Methods

هناك طرق عدة ضمن طريقة العنقدة الهرمية حيث أنه بعد تجميع المتغيرات على شكل عناقيد يجب حساب التماثل ما بين العناقيد، وهناك طرق عديدة تستخدم لهذا الغرض وهي (طريقة الربط المنفرد، طريقة الربط التام، طريقة الربط المتوسط، طريقة النقطة الوسطى، طريقة وارد) ، في هذا البحث تم استخدام طريقة الربط المنفرد في الجانب التطبيقي لذلك سيتم توضيح هذه الطريقة بشيء من التفصيل.

طريقة الربط المنفرد : Single Linkage

تسمى طريقة الجوار الأقرب، تعتمد هذه الطريقة على اعتبار أن العنصرين الأكثر تشابهاً بين العناصر بشكل نواة العنقود، ثم تُضاف باقي الوحدات إلى هذه النواة بالتسلسل وحسب درجة الشبه مع عناصر نواة العنقود، إذ تُضاف الأكثر شبهاً ثم الأقل وبالتدرج، وفي حالة ربط مجموعة من العناقيد مع بعضها فإن ذلك يتم بالاستناد على أقرب المسافات أو معاملات التماثل بين أزواج العناصر (المخلوفاً، 2012).

تعرف المسافة بين العنقودين (A,B) بأنها أقل مسافة بين نقطة في A ونقطة أخرى في B:

$$D(A, B) = \min\{d(y_i, y_j) \mid y_i \in A, y_j \in B\} \quad \dots \dots \dots (21)$$

حيث أن:

(i, j) تمثل العناصر في العناقيد A, B وأن (d(y_i, y_j)) هي المسافة المحسوبة :

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad \dots \dots \dots (22)$$

وأن العناصر اللذان يتم احتساب المسافة بينهما (رشيد وآخرون، 2011).

الجانب التجريبي:

في دراسة تطبيقية لتصنيف عينة مختارة من طلبة المدارس الحكومية والأهلية للمراحل المنتهية في محافظة نينوى والتي تم الحصول عليها من مديرية التربية في المحافظة ولمعرفة افضلية الأداء لكل مجموعة من هذه الجامعات المختارة تم تطبيق أسلوب الدالة التمييزية الخطية ودالة الانحدار اللوجستي فضلاً عن استخدام أسلوب التحليل العنقودي الهرمي حيث انه وبعد التأكد من أن البيانات التي تم الحصول عليها تتبع التوزيع الطبيعي من خلال اختبار البيانات المسحوبة للدراسة وفق فرضية العدم التي تشير الى توزيع البيانات توزيعاً طبيعياً ضد الفرضية البديلة لها وباستخدام البرنامج الاحصائي (Easy Fit) تم اختبار توزيع هذه البيانات بالاعتماد على المخبتر الاحصائي Chi – Square والمخبتر Kolmogorov – Smirnov فضلاً عن استخدام برنامج (Minitab17) في اختبار البيانات عن طريق الرسم والتي بينت توزيع البيانات توزيعاً طبيعياً.

تم حساب متجهات متوسطات المتغيرات لكل مجموعة قيد الدراسة فضلاً عن تقدير مصفوفتي التباين – التباين المشترك لغرض الاستفادة منه في ايجاد الدالة التمييزية، علماً أن متغيرات الدراسة كما في الجدول الآتي:

المرحلة الاعدادية(حكومية-أهلية)				المرحلة المتوسطة(حكومية-أهلية)				المرحلة الابتدائية(حكومية-أهلية)			
x_j	المتغيرات	Mean	Std.D	x_j	المتغيرات	Mean	Std.D	x_j	المتغيرات	Mean	Std.D
x_1	التربية الاسلامية	93.894	3.86038	x_1	التربية الاسلامية	77.327	6.64307	x_1	التربية الاسلامية	76.399	9.0821
x_2	اللغة العربية	75.811	8.42704	x_2	اللغة العربية	64.276	5.92501	x_2	اللغة العربية	67.551	8.4939
x_3	اللغة الانكليزية	83.758	8.49465	x_3	اللغة الانكليزية	60.742	8.04147	x_3	اللغة الانكليزية	56.121	12.1714
x_4	الاحياء	76.7305	9.99318	x_4	الرياضيات	55.012	9.98316	x_4	الرياضيات	60.936	14.1714
x_5	الرياضيات	75.866	11.6200	x_5	الاجتماعيات	56.740	6.61434	x_5	العلوم	62.937	15.5839
x_6	الكيمياء	77.695	10.4474	x_6	الاحياء	59.782	8.88504	x_6	الاجتماعيات	66.543	11.3480
x_7	الفيزياء	84.778	9.59449	x_7	الكيمياء	55.173	8.99914	x_1	التربية الاسلامية	94.286	4.20672
x_1	التربية الاسلامية	92.343	5.84179	x_8	الفيزياء	56.109	9.21239	x_2	اللغة العربية	90.818	6.34159
x_2	اللغة العربية	71.827	11.0659	x_1	التربية الاسلامية	68.936	9.27589	x_3	اللغة الانكليزية	84.836	7.95194
x_3	اللغة الانكليزية	83.124	12.3397	x_2	اللغة العربية	57.693	8.63978	x_4	الرياضيات	86.573	8.35671
x_4	الاحياء	73.457	15.0606	x_3	اللغة الانكليزية	58.601	16.01869	x_5	العلوم	88.305	5.36015
x_5	الرياضيات	71.151	16.4046	x_4	الرياضيات	50.078	14.84988	x_6	الاجتماعيات	89.253	5.21862
x_6	الكيمياء	77.335	13.9615	x_5	الاجتماعيات	57.273	11.46984				
x_7	الفيزياء	80.857	13.9215	x_6	الاحياء	59.494	13.13332				
				x_7	الكيمياء	53.581	11.60640				
				x_8	الفيزياء	47.836	14.38863				

12- دالة التمييز الخطي: يجب تقدير قيم معاملات الدالة a_i من خلال الصيغة (2) و (3) لكل

مجموعة من المجموعات ومن خلال النتائج السابقة تم الحصول على النتائج في الجدول الآتي الذي يوضح قيم المعاملات للدوال التمييزية لكل مرحلة من مراحل التصنيف :

دالة التميز للمراحل الثلاث (الحكومية - الأهلية)

المرحلة الابتدائية - للمدارس (الحكومية - الأهلية)

$$L = 0.059592x_1 + 0.269318x_2 + 0.158343x_3 - 0.09153x_4 - 0.05459x_5 + 0.089916x_6$$

المرحلة المتوسطة - للمدارس (الحكومية - الأهلية)

$$L = -0.1227x_1 - 0.17065x_2 + 0.048883x_3 + 0.015961x_4 + 0.062092x_5 + 0.119545x_6 - 0.02064x_6 - 0.08201x_7$$

المرحلة الاعدادية - للمدارس (الحكومية - الأهلية)

$$L = -0.01171x_1 - 0.12418x_2 + 0.082229x_3 + 0.006107x_4 - 0.04023x_5 + 0.136445x_6 - 0.08262x_7$$

من خلال ملاحظة معادلة الدالة التمييزية الخطية للمرحلة الابتدائية أن المتغيرات المتمثلة بالمواد (التربية الإسلامية واللغة العربية واللغة الانكليزية والاجتماعيات) عند زيادة قيمتها بمقدار درجة واحدة يؤدي الى زيادة قيمة الدالة التمييزية بمعدل (0.0595, 0.089916, 0.158343, 0.269318) لكل مادة على التوالي، أما المتغيرات التي تمثل مادتي (الرياضيات والعلوم) فان اشارة السالب فيها تؤدي الى ان الزيادة في قيمتها بمقدار درجة واحدة يؤدي الى انقاص قيمة الدالة بمعدل (0.09153, 0.05459) على التوالي، كذلك التفسير بالنسبة للمراحل المتوسطة والاعدادية مع اختلاف تأثير المواد على قيمة الدالة.

بعد أن تم تكوين الدالة التمييزية لكل مجموعة من مجاميع الدراسة، فان العملية التالية هي ايجاد نقطة الفصل لغرض التنبؤ وتصنيف المشاهدة الجديدة المجهول انتمائها لأي مجموعة من المجموعات قيد الدراسة بعد أن تتم مقارنتها مع قيمة الدالة التي يتم الحصول عليها عند تعويض قيم المشاهدة، ولهذا الغرض يتم الاعتماد على الصيغة الموضحة في المعادلة (8) و (9) و (10) لإيجاد كل من $\bar{L}_{(1)}$, $\bar{L}_{(2)}$ تم الحصول على قيم نقاط الفصل الموضحة وكالاتي:

الجدول (1) يبين قيم نقاط الفصل لكل مرحلة (حكومية - أهلية)

المرحلة	حكومية - أهلية		نقطة الفصل = L
	$\bar{L}_{(1)}$	$\bar{L}_{(2)}$	
الابتدائية	38.79186	28.60244	33.69715
المتوسطة	-9.00067	-11.6803	-10.340485
الاعدادية	-1.707799	-2.613261	-2.16053

وللوقوف عند كفاءة الدالة التمييزية الخطية لابد من اجراء الاختبارين التاليين لمعرفة مدى تحقق شروط استخدام التمييز الخطي وحسب المعادلة (12) و (13) وكالاتي:

الجدول (2) يبين القيم المحسوبة والجدولية للمختبر الاحصائي المستخدم لاختبار معنوية الفروق بين المجتمعات

	D^2	F_{cal}	d.f	F_{tab}
الابتدائية (أهلية- حكومية)	10.18941	322.54094**	5,141	3.02
المتوسطة (أهلية- حكومية)	2.67962	78.81408**	7,135	2.64
الاعدادية (أهلية- حكومية)	0.90546	36.10115**	6,162	2.80

بحسب النتائج الموضحة فانه تم التأكد من تحقق الشرط الاول لاستخدام الدالة التمييزية الخطية والمتمثل بوجود الفروق المعنوية للمتوسطات بمعنى أنه يتم قبول الفرض البديل الذي ينص على وجود الفروق .

الجدول (3) يبين القيم المحسوبة والجدولية لاختبار تجانس التباين

	M	C^{-1}	MC^{-1}	d.f	Chi-Square
الابتدائية(حكومية- أهلية)	127.768	0.9210317195	117.67838**	5	170.189
المتوسطة(حكومية- أهلية)	172.413	0.8901959409	153.48135	7	63.457
الاعدادية(حكومية- أهلية)	71.878	0.9260831527	66.56500	6	33.638

وباستخدام المعادلة (14) و(15) لاختبار تجانس مصفوفتي التباينات فانه تقبل فرضية العدم للمدارس الابتدائية والمتوسطة(حكومية - أهلية) كون القيمة المحسوبة أقل من القيمة الجدولية للمختبر Chi-Square وتُرفض عند المرحلة الاعدادية اي أن الشرط الثاني لم يتحقق في بيانات المرحلة الاعدادية مما يدعو لاستخدام الدالة التمييزية اللاخطية .

نظرا للنتائج التي تم التوصل اليها في السابق كان لا بد من معرفة اخطاء التصنيف التي تؤثر ويشكل على المعنوية والتجانس للمراحل الثلاثة، حيث تبين في الجدول الاتي :

الجدول (4) يبين اخطاء التصنيف لمفردات المتغيرات للمراحل الثلاث (الحكومية - الأهلية)

		المرحلة الابتدائية			المرحلة المتوسطة			المرحلة الاعدادية		
		أهلية	حكومية	Total	أهلية	حكومية	Total	أهلية	حكومية	Total
Count	أهلية	47	2	49	30	15	45	34	37	71
	حكومية	7	92	99	5	88	93	21	78	99
%	أهلية	95.9	4.1	100.0	66.7	33.3	100.0	47.9	52.1	100.0
	حكومية	7.1	92.9	100.0	5.4	94.6	100.0	21.2	78.8	100.0

اخطاء التصنيف كانت بنسبة كبيرة في مدارس الأهلية للمرحلة الاعدادية حيث بلغت قيمته (52.1) أما الحكومية بلغت قيمة الخطأ (21.2) تليها المدارس الأهلية للمرحلة المتوسطة والتي بلغت قيمة الخطأ فيها (33.3) تقابلها نسبة خطأ قليلة في المدارس الحكومية لنفس المرحلة حيث بلغت (5.4)، أما المرحلة الابتدائية كانت نسبة الخطأ ضئيلة جداً في المدارس الأهلية (4.1) والحكومية (7.1) بمقارنة بقية المراحل، مما يدل على وجود مشاهدات صنفت بطريقة خاطئة في كلا المجموعتين ولجميع المراحل.

13- دالة الانحدار اللوجستي الثنائي: تم الحصول على النتائج الآتية والمتمثلة بقيم الدالة ومعلماتها:

معاملات الانحدار اللوجستي الثنائي المقدر للمراحل الثلاثة(الحكومية - الأهلية)

المرحلة الابتدائية - للمدارس(الحكومية - الأهلية)

$$L = 5.122 + 0.102x_1 + 0.143x_2 - 0.262x_3 + 0.028x_4 - 0.054x_5 - 0.067x_6$$

المرحلة المتوسطة - للمدارس(الحكومية - الأهلية)

$$L = -10.721 + 0.155x_1 + 0.182x_2 - 0.063x_3 - 0.020x_4 - 0.095x_5 - 0.120x_6 + 0.023x_7 + 0.94x_8$$

المرحلة الاعدادية - للمدارس(الحكومية - الأهلية)

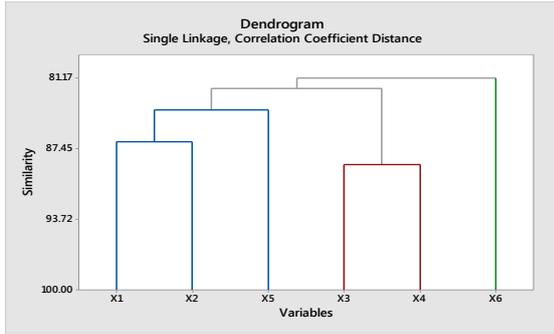
$$L = -4.142 + 0.044x_1 + 0.136x_2 - 0.091x_3 - 0.006x_4 + 0.046x_5 - 0.175x_6 + 0.103x_7$$

الجدول (5) يبين قيم اختبار Wald وقيمة الدالة الأسية للمراحل الثلاثة (الحكومية - الأهلية)

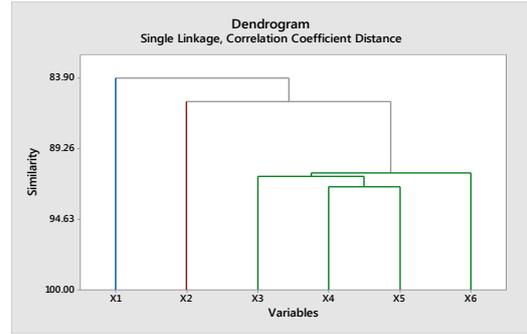
المرحلة	Variables	β	S.E	Wald	d.f	P-value	Exp(β)	
الابتدائية	الحكومية - الأهلية	التربية الإسلامية	0.102	0.095	1.134	1	0.287	1.107
		اللغة العربية	0.143	0.064	5.065	1	0.024*	1.154
		اللغة الانكليزية	-0.262	0.087	9.167	1	0.002**	0.769
		الرياضيات	0.028	0.057	0.234	1	0.629	1.028
		العلوم	-0.054	0.058	0.865	1	0.352	0.947
		الاجتماعيات	-0.067	0.067	0.994	1	0.319	0.935
		Constant	5.122	6.152	0.693	1	0.405	167.700
المتوسطة	الحكومية - الأهلية	التربية الإسلامية	0.155	0.049	10.132	1	0.001**	1.168
		اللغة العربية	0.182	0.057	10.047	1	0.002**	1.199
		اللغة الانكليزية	-0.063	0.036	3.086	1	0.079	0.939
		الرياضيات	-0.020	0.033	0.371	1	0.543	0.980
		الاجتماعيات	-0.095	0.045	4.346	1	0.037*	0.910
		الاحياء	-0.120	0.044	7.423	1	0.006**	0.887
		الكيمياء	0.023	0.044	0.285	1	0.593	1.024
		الفيزياء	0.94	0.044	4.451	1	0.035*	1.098
Constant	-10.72	3.150	11.583	1	0.001	0.000		
الاعدادية	الحكومية - الأهلية	التربية الإسلامية	0.044	0.046	0.898	1	0.343	1.045
		اللغة العربية	0.136	0.045	8.980	1	0.003**	1.145
		اللغة الانكليزية	-0.091	0.030	9.167	1	0.002**	0.913
		الاحياء	-0.006	0.029	0.038	1	0.845	0.994
		الرياضيات	0.046	0.027	2.867	1	0.090	1.048
		الكيمياء	-0.175	0.042	17.172	1	0.000**	0.840
		الفيزياء	0.103	0.040	6.835	1	0.009**	1.109
		Constant	-4.142	3.644	1.292	1	0.256	0.016

يوضح العمود الاول من الجدول اعلاه معاملات النموذج المقدره بوحدهات $\log\left(\frac{p}{1-p}\right)$ ، أما العمود الثالث يوضح احصاء اختبار wald لمعنوية المعلمات المقدره ومن خلال القيم الاحتمالية يتضح أن في العمود P-value في حالة (الأهلية - الحكومية) للمرحلة الابتدائية أن اللغة العربية والانكليزية لها تأثير معنوي على المتغير المعتمد Y حيث القيمة الاحتمالية كانت أقل من 0.01 و 0.05، أما بالنسبة لبقية المتغيرات فان تأثيرها غير معنوي حيث القيمة الاحتمالية كانت أكبر من 0.01 و 0.05، بالنسبة لمعنوية معاملات المرحلة المتوسطة فان التربية الإسلامية والعربية والاحياء كان تأثيرها معنوي عند مستوى المعنوية 0.01 أما مادة الاجتماعيات والفيزياء لها تأثير معنوي على قيمة المتغير المعتمد عند المستوى 0.05 للحالة (الحكومية - الأهلية)، أما ما يتعلق بمعنوية المعلمات بالنسبة للمرحلة الاعدادية فانه قد وجد أن مادة الفيزياء وبالدرجة الاولى تليها اللغة الانكليزية والعربية ثم الفيزياء تأثيرها المعنوي عند مستوى المعنوية 0.01 لحالة (الحكومية - الأهلية) وما يتعلق ببقية المتغيرات فان تأثيرها غير معنوي على المتغير المعتمد.

14- التحليل العنقودي: وبتطبيق Cluster analysis على درجات الطلبة تم تصنيف للحالات داخل العناقيد باستخدام طريقة الربط المنفرد حيث يحتوي كل عنقود على عددا من المتغيرات المتقاربة والمتجانسة فيما بينها، وباستخدام برنامج Minitab للحصول على المخطط العنقودي للمتغيرات، تم الحصول على النتائج الاتية:

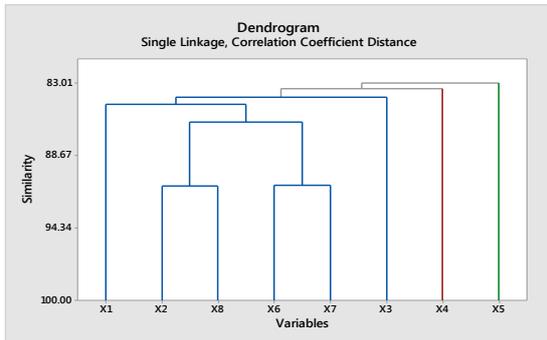


الشكل(2): المخطط الشجري للمدارس الابتدائية الأهلية

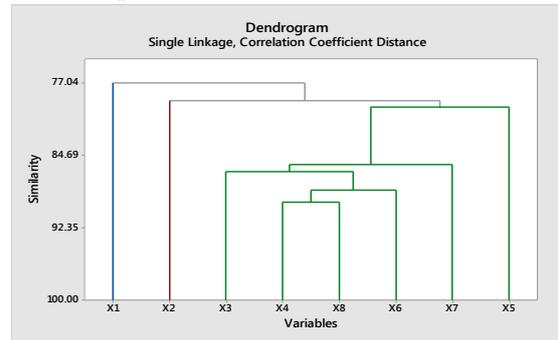


الشكل(1): المخطط الشجري للمدارس الابتدائية الحكومي

من المخطط الشجري الأكثر وضوحاً وبالنسبة للمدارس الابتدائية الحكومية وفي الشكل (1) فقد جُمعت كل من مادة (الرياضيات والعلوم واللغة الانكليزية والاجتماعيات) في عنقود واحد مما يدل على أن هناك تجانس فيما بينها أكثر من باقي المواد الاخرى مع ملاحظة أن مادتي التربية الاسلامية واللغة العربية كانت كل منها منفردة في عنقود واحد حيث اعطت فروق معنوية أكبر عن العنقود الاول، أما المدارس الأهلية في الشكل (2) كانت كل من مادة (التربية الاسلامية واللغة العربية والعلوم) متجانسة فيما بينها ومادتي (اللغة الانكليزية والرياضيات) متجانسة فيما بينها أما مادة الاجتماعيات فقد كانت منفردة في عنقود واحد مما أكد اعطائها فرق معنوي أكبر عن باقي المواد ، وهذا يظهر وجود فرق في أهمية المواد بين المدارس الحكومية والأهلية في الاداء.

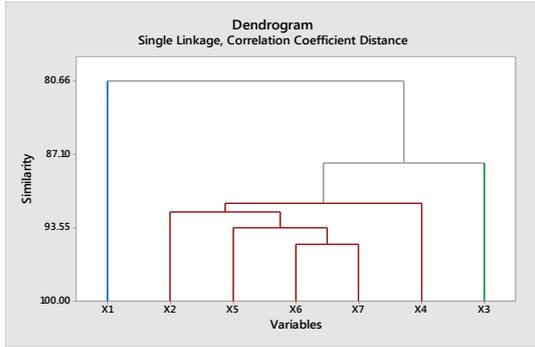


الشكل(4): المخطط الشجري للمدارس المتوسطة الأهلية

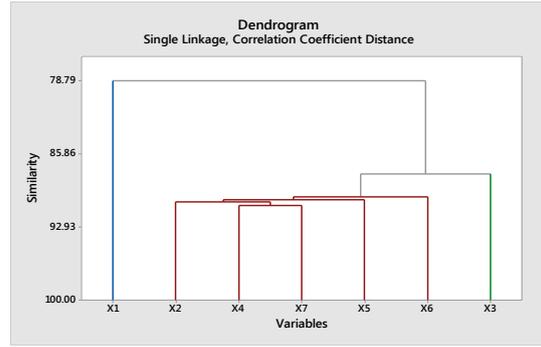


الشكل(3): المخطط الشجري للمدارس المتوسطة الحكومية

من المخطط الشجري للمدارس الحكومية وفي الشكل (3) تبين أن هناك تجانس بين مادة (الرياضيات والفيزياء واللغة الانكليزية والاحياء والكيمياء والاجتماعيات) مما جعلها في عنقود واحد بينما كانت مادتي اللغة العربية والتربية الاسلامية منفردة كل منها في عنقود واحد بشكل متميز عن باقي المواد لامتلاكها فروق معنوية كبيرة عن المواد الاخرى ، اما في المدارس الأهلية في الشكل (4) كان لمادتي الرياضيات والاجتماعيات تميز عن باقي المواد حيث كانت كل منها في عنقود منفرد فكانت الفروق لكل منهما أكبر أما بقية المواد جُمعت في عنقود واحد مما يدل على تجانسها مع بعضها، ونلاحظ اعطاء أهمية أكبر في الاداء لمادتي التربية الاسلامية واللغة العربية بشكل عام للمدارس الحكومية والأهلية.



الشكل (6): المخطط الشجري للمدارس الاعدادية الأهلية



الشكل (5): المخطط الشجري للمدارس الاعدادية الحكومية

ونلاحظ في المدارس الحكومية والأهلية في الشكلين (5) و(6) وجود المواد العلمية المتمثلة بـ(اللغة العربية والاحياء والفيزياء والرياضيات والكيمياء) في عنقود واحد مما يدل على تجانسها بينما تميزت مادتي اللغة الانكليزية والتربية الاسلامية كل منها في عنقود منفرد لوجود فروق معنوية أكبر جعلها تتميز عن باقي المواد.

وقد تبين ان مادتي (التربية الاسلامية واللغة الانكليزية) أعطت اهمية في الاداء الدراسي بشكل عام للمدارس الحكومية والأهلية ولجميع المراحل في عملية التحليل الاحصائي التي تم دراستها عن باقي المواد.

الاستنتاجات:

1. من خلال معادلة الدالة التمييزية الخطية تم ملاحظة في المرحلة الابتدائية ظهور تأثير واضح لمادتي التربية الاسلامية واللغة العربية على قيمة الدالة ويكون هذا التأثير بمقدار وحدة واحدة عن باقي المواد على النقيض من تأثيرهما في المرحلتين (المتوسطة والاعدادية) مع ملاحظة التأثير السلبي (المتناقض) لمرحلتي المتوسطة والاعدادية في مادة الفيزياء على قيمة الدالة، علماً أن انخفاض قيمة الدالة التمييزية يؤدي الى زيادة في نسبة الخطأ للتصنيف وهذا ما تم توضيحه وتأكيدُه في الجدول (4).
2. نستنتج من الجدول (2) و(3) تحقق شرطي كفاءة الدالة التمييزية الخطية للمرحلة الابتدائية (الحكومية- الأهلية) في ظهور الفروقات المعنوية اثناء الاختبار فضلاً عن تجانس مصفوفة التباين والتباين المشترك لهم على النقيض من ذلك للمرحلة المتوسطة والاعدادية (الحكومية- الأهلية) حيث اثبتوا عدم كفاءتهم، لذا ستكون الدالة اللاخطية لهما افضل.
3. في اختبار Wald التابع لدراسة الدالة اللوجستية الثنائية في المرحلة الابتدائية (الحكومية- الأهلية) تأثيرها المعنوي لمادتي اللغة العربية والانكليزية فضلاً عن ظهور التأثير المعنوي في المرحلة المتوسطة لمادة التربية الاسلامية واللغة العربية والاحياء فضلاً عن الاجتماعيات والفيزياء والمرحلة الاعدادية تبين أن مواد اللغة العربية والانكليزية والكيمياء والفيزياء كان لها تأثيراً معنوياً أما باقي المواد لم تظهر معنويتها ولجميع المراحل والمدارس (الحكومية- الأهلية).

4. من خلال النتائج التي تم التوصل إليها للمرحلة الإحصائية (الحكومية-الأهلية) فقد أثبتت الدالة التمييزية الخطية عدم كفاءتها وملائمتها لبيانات المرحلة لعدم تحققها شرط تجانس مصفوفتي التباين والتباين المشترك والضروري لتطبيق الدالة ، بعدما تم اللجوء الى استخدام دالة الانحدار اللوجستي الثنائي تبين عدم ملائمتها ايضاً وبحسب النتائج الموضحة في الجدول (2-25) لذا سيتم الاعتماد على نتائج التحليل العنقودي لبيانات المرحلة.

5. نتائج التحليل العنقودي فقد تبين أن هناك بعض المواد تتقارب في مستواها وذلك في ضوء تقاربها والتحامها في عملية تكوين العنقدة، بالنسبة للمرحلة الابتدائية الحكومية تبين أن هناك تقارب كبير بين مادة(اللغة الانكليزية والرياضيات والعلوم والاجتماعيات) بينما للمدارس الأهلية كانت مادة(التربية الاسلامية واللغة العربية والعلوم) متقاربة، أما المدارس المتوسطة الحكومية فكانت مادة(اللغة الانكليزية وانتهاءً بمادة الفيزياء) متقاربة فيما بينها كذلك في المدارس الأهلية بإضافة تقارب مادة التربية الاسلامية وفي ما يتعلق بالمرحلة الإحصائية الحكومية والأهلية فكان تقارب مستوى المواد العلمية واضح .

6. هناك تشابه كبير في المستوى التعليمي للمدارس الحكومية والأهلية كونت تشابه كبير في المخطط الهرمي، كما أنه يوجد تمييز لمادة التربية الاسلامية واللغة العربية واللغة الانكليزية من خلال الفروق المعنوية الكبيرة التي أعطتها المواد لجميع المراحل مما يعكس خصوصية لهذه المواد عن باقي المواد.

المصادر:

1. سيد أحمد، ابراهيم محمد ابراهيم، (2010)، "مقارنة بين نموذج الانحدار الخطي المتعدد ونموذج الدالة التمييزية" ، بحث تكميلي لنيل شهادة الماجستير في الاحصاء التطبيقي، جامعة السودان للعلوم والتكنولوجيا ، كلية الدراسات العليا، السودان.
2. غانم، عدنان والجاعوني، فريد خليل، (2011)، "استخدام تقنية الانحدار اللوجستي ثنائي الاستجابة في دراسة أهم المحددات الاقتصادية والاجتماعية لكفاية دخل الاسرة" ، مجلة جامعة دمشق للعلوم الاقتصادية والقانونية، قسم الاحصاء ، كلية الاقتصاد، المجلد(27)، العدد(1)، دمشق، ص119.
3. الشكرجي، ذنون يونس والنعمي، أسوان محمد طيب، (2008)، "بناء دالة التمييز بالاعتماد على متغيرات تحليل الانحدار" ، مجلة تكريت للعلوم الادارية والاقتصادية، العدد(8).
4. المخولفي، فؤاد عبده اسماعيل، (2012)، "تصنيف وتمييز المحافظات اليمنية بحسب مصادر الدخل الفردي باستخدام اسلوبي التحليل العنقودي والتحليل التمييزي" ، مكتب البحوث والنشر، جامعة الناصر، صنعاء، اليمن، ص11.
5. دخيل، طاهر ريسان وداؤود، زينب يوسف، (2014) ، "استخدام التحليل المميز لتحديد أهمية العوامل المؤثرة على المصابين بمرض اليرقان للأطفال حديثي الولادة" ، جامعة القادسية، العراق، ص5.

6. دخيل، طاهر ريسان وكاظم، هند جواد، (2009)، "دراسة مقارنة بين المدارس الأساسية والمدارس غير الأساسية في محافظة"، مجلة القادسية للعلوم الادارية والاقتصادية، العراق، المجلد(11)، العدد(1)، ص175-174.
7. رشيد، أسيل عبد الرزاق ومهدي، نبأ نعيم، (2011)، "تحليل ولقع التربية والتعليم باستخدام طرائق التحليل العنقودي"، مجلة القادسية للعلوم الادارية، المجلد(14)، العدد(2)، ص197.
8. الفرهود، سهيلة حمود عبدالله، (2014)، "استخدام الانحدار اللوجستي لدراسة العوامل المؤثرة على اداء الاسهم"، مجلة الازهر، كلية الدراسات التجارية، قسم الاحصاء، الكويت، ص54.
9. عابد، ياسر عبدالله، (2015)، "دراسة مقارنة بين الاساليب الاحصائية لدراسة العوامل المؤثرة على تعدد الزوجات في الاراضي الفلسطينية"، رسالة لاستكمال متطلبات الحصول على درجة الماجستير في الاحصاء، جامعة الازهر، غزة، ص22.
10. عباس، علي خضير،(2012)،"استخدام الانحدار اللوجستي في التنبؤ بالدوال ذات المتغيرات الاقتصادية التابعة النوعية"، مجلة كركوك للعلوم الادارية والاقتصادية، المجلد(2)، العدد(2)، ص240.
11. عدلي، ابراهيم وأبوشكان، محمد، (2014)، "استخدام الانحدار اللوجستي الثنائي في تفسير المتغيرات التابعة ثنائية القيمة في ميدان الأنشطة البدنية والرياضية"، مجلة علوم وممارسات الأنشطة البدنية الرياضية والفنية، رقم 06، جامعة الجزائر، اكتوبر، ص2-3.
12. سليمان، مثنى صبحي وصالح، رياض محمود، (2012)، "توظيف التحليل العنقودي وطريقة الجار الاقرب في التعرف على الانماط مع تطبيق على نوعية المياه الجوفية في محافظة نينوى"، المجلة العراقية للعلوم الاحصائية، العدد(21).
13. سليمان، هاشم أحمد وشفيق، عائد محمد، (2012)، "دالة التمايز البدنية لطلاب الدراسة المتوسطة في مدينة الموصل"، مجلة الرافدين للعلوم الرياضية، العدد(58).
14. Copper ,R. and A. Jweaks ,(1984),”Data Models and Statistic Analysis”, New York: Philip Allan Publisher Limited.
15. Fisher ,R,A,(1936) ”the use of multiple measurements in taxonomic problems”, Annals of Eugenics 7, pp179-188.
16. Lea, Stephen(1997), “Multivariate Analysis II: Manifest Variables Analysis”. Topic 4:Logistic Regression and Discriminant Analysis. University Of Exeter Department of Psychology. Revised 11th
17. Michael, W., Berry,(2001) “Survey of Text Mining :Clustering, Classification and Retrieval”, John Wiley & Sons ,UK.
18. Wilson, L. & Hardgrave, C.(1995), “Predicting Graduate Student Success in an MBA Program: Regression Versus Classification. Education and Psychological Measurement, 55(2),April ,pp186-195.