استخدام بعض طرق التحليل العنقودي في التصنيف مع تطبيق عملي[+]

USING SOME METHODS OF ACINIFORM ANALYSIS IN THE CLASSIFICATION WITH PRACTICAL ANALYSIS

نزار مصطفى[*]

المستخلص

من الطرق الاحصائية المهمة والمستخدمة في التصنيف هي اسلوب التحليل العنقودي والتي تعتمد بدورها على تحليل متغيرات محددة تعتمد على نقاط التشابه والاختلاف بين البيانات وقد استخدم هذا الاسلوب منذ قديم الزمان وقد استخدم في هذا البحث مجموعة من طرق التحليل العنقودي هو اسلوب الربط المنفرد، واسلوب الربط الشامل والطريقة الهرمية Ward. وتم تطبيق هذه الطرق على بيانات مجموعة من دول المغرب العربي والواقعة في قارة افريقيا بغية الوصول الى مجاميع الدول التي تمتاز فيما بينها والتي تجمعها مجموعة من الصفات المشتركة لغرض الوصول الى طريق التكامل الاقتصادي بين هذه الدول وتذليل الصعوبات فيما بينها وهذا هو واحد من اهداف علم الاحصاء الذي يسعى الى استخدام البيانات والمعلومات بشكل امثل لغرض تحقيق اهداف معينة يستفاد منها مستقبلاً.

Abstract

All activities for extraction of information from "raw data" involves some type of grouping or clustering of input items. The clustering of items of any kind is probably one of the first mental activities of men. A lot of special techniques has been developed in the past fifty years to simulate the thinking processes, but non of then can really claim to hit the essence of the recognition or prediction processes of human mind. From the theoretical point of view, the clustering procedures are probably too simple to attract a lot of attention, and from the practical aspect, all the methods can be applied to seventeen objects. In this study, a method for clustering (single linkage method, complete linkage methods and Ward's methods) are described and discussed. One of the important questions of research is investigating whether any of the clustering methods is evidently more effective than the others. The final conclusion of the research is that there isn't one particular method which is the best in a given situation. The choice of the method should depend mainly on the pattern of data and the aim of the study

المقدمة:

[+] تاريخ استلام البحث ٢٠٠٤/١٢/١ ، تاريخ قبول النشر ٢٠٠٥/٥/٣٠

[*] مدرس / الكلية التقنية الادارية

من المعلوم لدينا ان الدراسات التي تهتم بتصنيف البيانات لها الاثر الكبير في معرفة بعض الصفات المشتركة وأوجه الشبه والاختلاف للبيانات مما يسهل دراستها.

لذا سنحاول في هذا البحث استخدام اسلوب احصائي يساعد في تصنيف البيانات، وهذا الاسلوب هو التحليل العنقودي الذي يستخدم لتصنيف ودراسة تجمعات البيانات أو المشاهدات أو العناصر في مجاميع متجانسة فيما بينها ومختلفة عن باقي المجاميع بالاعتماد على مجموعة من الصفات أو المتغيرات.

ولقد تم استخدام التصنيف منذ العصور القديمة لغرض تصنيف البشر بالاعتماد على صفات اللون والجنس وغيرها ومن اهم البحوث التي تم استخدام مقاييس التصنيف فيها هي في عام (١٩٦٠) تم ايجاد مقياس للتشابه من قبل الباحثين (Bary and Curtis) وتم تطبيقه في مجال دراسات التربية وفي عام (١٩٦١) استطاع الباحث Solcal ايجاد مقياس اطلق عليه (Euclidian distance) والذي يقيس المسافة بين العنصرين المقارنين.

وفي عام (١٩٧١) استخدم الباحثان (Norris and Dule) طريقة التعنقد المعتمدة على نظرية المعلومات واجمالاً يمكن القول ان استخدام اساليب التصنيف العددي اصبح واسعاً فقط في الفترة الاخيرة وخاصة بعد ان تم استخدام الحاسوب في الحصول على نتائج سريعة ودقيقة.

<u>هدف البحث</u>:

يهدف البحث الى تحقيق دراسة احصائية تقوم بتقسيم بيانات بعض الأقطار العربية الى مجاميع وكل مجموعة تكون ذات تقارب (تجانس) كبير فيما بينها. لان تقدم البلدان يعتمد على التطور في القطاعات الحيوية في أي بلد كان، لذا لابد من ان تقوم بالاتفاق على جميع القطاعات الحيوية الخاصة بكل بلد. وان اهمال أي قطاع يؤدي الى تخلف هذه القطاعات وبالتالي تخلف ذلك البلد.

الجانب النظري:

من الاساليب المهمة في تحليل البيانات هو اسلوب التحليل العنقودي اذ يستخدم لغرض دراسة تجمعات البيانات حسب اسس معينة بغية الوصول الى وصف دقيق ذات متغيرات متعددة وان مايحدد نوع العلاقة بين العناصر المطلوب تصنيفها هي المتغيرات او الصفات التي تتمتع بها لذا لابد من معرفة كيفية التعامل مع جميع انواع المتغيرات.

ويستخدم هذا التحليل لتجميع مفردات (العناصر) بشكل عناقيد بالاعتماد على مقدار التشابه بينها، وتتم العنقدة بأشكال متعددة منها الشكل المتسلسل او الاسلوب غير المتسلسل وذلك بالاعتماد على مصفوفة المسافة او مصفوفة الارتباط. [1]

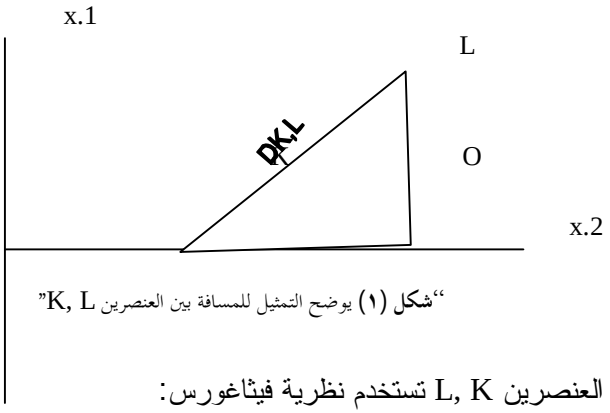
ويعتبر الاسلوب المتسلسل Hierarchical من الاساليب المفضلة في هذا التحليل اذ يتم فيه عنقدة (n) من المفردات وبشكل متسلسل في (M) من العناقيد المرتبة تصاعديا اذ يكون c_1 أضعف هذه العناقيد وc_m أقواها.

وهناك مستوى التحام Threshold distance a_i مرافق لكل عنقود(c_i; I=1, ...m) اذ تكون قيمها موجبة ومرتبة تصاعدياً.

كذلك في التحليل العنقودي المتسلسل هناك طرق مختلفة وبالاعتماد على طرق اختيار مسافة a_i بين عنقودين أو بين عنصر وعنقود، وباستخدام مجموع مربعات الانحرافات الكلي لأي نقطة عن متوسط العنقود، ومن الطرق المفضلة هي طريقة الربط المفرد، الربط الشامل وطريقة الربط الهرمية Ward. وتعتمد هذه الطرق على مقياس التشابه المستخدم بين العناصر تحديد مستوى الالتحام وهذا المقياس هو مقياس المسافة المسمى (Euclidian distance).

١ . مقياس المسافة المسمى Euclidian distance

وهو من المقاييس الاكثر شيوعاً في الاستخدام ولتوضيح فكرة هذا المقياس، نفرض لدينا كل من العنصرين K, L الممثلين بالشكل رقم (١) وكلاهما متصفان بالخاصيتين x.1 ,x.2. [1]



"**شكل (١)** يوضح التمثيل للمسافة بين العنصرين K, L"

لايجاد المسافة بين العنصرين L, K تستخدم نظرية فيثاغورس:

$$D^2_{K,L} = (o\overline{K})^2 + (o\overline{L})^L$$

$$= (X_{K1} - X_{L1})^2 + (X_{K2} - X_{L2})^2$$

$$D_{K,L} = \sqrt{\sum_{j=1}^{2} (X_{Kkj} - X_{Lj})^2}$$

وبصورة عامة

$$D_{K,L} = \sqrt{\sum_{j=1}^{n} (X_{Kkj} - X_{Lj})^2} \quad \text{.....(1)}$$

وعندما يكون عدد المشاهدات غير متساوٍ ولحساب جميع المسافات بين أزواج العناصر (حين توجد بيانات مفقودة) فمن الضروري حساب معدل المسافة، وحسب الصيغة:

$$d_{K,L} = \sqrt{\frac{1}{r} \sum_{j=1}^{n} (X_{Kkj} - X_{Lj})^2} \quad \text{.....(2)}$$

اذ ان P تمثل عدد العناصر (P < n) التي تؤخذ في الحسابات. وبالرجوع الى المعادلة (١) من الممكن الحصول على المسافات بشكل موجهات (Vector natation) وكما يأتي:- [2]

$\underline{X}K. = (X_{K1}, \ldots X_{Kj} \ldots X_{Kn})$

$\underline{X}L. = (X_{L1}, \ldots X_{Lj} \ldots X_{Ln})$

وعليه فان

$(\underline{X}K.- \underline{X}L.) = (X_{K1}- X_{L1} \ldots , X_{Kn} - X_{Ln})$

وهذه تمثل موجهات على شكل صفوف (row vectors) ومبدلتها هي:

$$(\underline{X}K. - XL.)' = \begin{bmatrix} X_{K1} - X_{L1} \\ X_{Kj} - X_{Lj} \\ X_{Kn} - X_{Ln} \end{bmatrix}$$

$$DK_1L = \sum_{j=\bullet}^{\wedge} (\underline{X}_{Kj}- X_{Lj})^2 = (\underline{X}_{K.} - \underline{X}_{L.})(\underline{X}_{K.}-\underline{X}_{L.})' \ldots\ldots\ldots\ldots\ldots(3)$$

٢ - طرق التعنقد *Clustering Methods*:

ان عملية التعنقد تعني وضع العناصر في مجاميع، أي انها تمثل الحصول على مجموعة واحدة او اكثر من المجاميع التي تضم عدداً من العناصر على درجة كبيرة من التشابه فيما بينها، وعملية الحصول على هذه المجاميع يتم عن طريق تجزءة العدد الكلي للعناصر الى مجاميع او كل مجموعة تضم عدد من العناصر بالاعتماد على مصفوفة التشابه.

وهناك العديد من طرق التعنقد التي تم تطويرها لتصف شكل العلاقات بين العناصر، وهذه الطرق تستند بالاساس الى مصفوفة المسافة أو مصفوفة الارتباط. [*3*]

اولاً: طريقة الربط المنفرد *Single Linkage Clustering*:

تعتبر هذه الطريقة من ابسط طرق التعنقد وتسمى ايضاً بطريقة الجوار الاقرب *(The nearest neighbor)* وأول من وضع هذه الطريقة هو العالم *(Sneath, 1957)* تعتمد هذه الطريقة بالاساس على اعتبار العنصرين الاكثر تشابهاً بين العناصر بشكل نواة العنقود. ثم تضاف باقي الوحدات الى هذه النواة بالتسلسل وحسب درجة الشبه مع عناصر نواة العنقود اذ يضاف الاكثر شبهاً ثم الاقل وبالتدريج.

وفي حالة ربط مجموعة من العناقيد مع بعضها فان ذلك يتم بالاستناد الى أقرب المسافات بين عناصر العناقيد وحسب الصيغة التالية:

$$d_{ij}^{*} = \min(d_{1j}, d_{2j})$$

حيث i, j تمثل العناصر في العناقيد I, J على التوالي.

ثانياً: طريقة الربط الشامل Complete Linkage Clustering:

وفي هذه الطريقة يتشكل العنقود بطريقة معاكسة للطريقة الاولى اذ انها تعتمد على الاقل تشابهاً بين العناصر. وبتعبير آخر ان العنصر المرشح للدخول الى العنقود يدخل فقط اذا كانت المسافة بينه وبين أي من

عناصر العنقود هي اكبر مسافة بينه وبين عنصر من عناصر العنقود. اما في حالة ربط مجموعة عناقيد مع بعضها فنلاحظ ابعد المسافات بين عناصر العناقيد وحسب الصيغة التالية:- [1]

$d(I,J) = Max\ (d_{ij})$ $\quad i \in I\ ,\ j \in J$ .....(4)

حيث ان i,j تمثل العناصر في العناقيد J, I على التوالي.

ثالثاً: الطريقة الهرمية Ward Method:

هذه الطريقة تعتمد على اساس اقل فقدان من المعلومات Loss Information لعمل العنقـــدة. لقـــد اعتمد مقياس المصاحبة لعمل العنقدة وهو مقياس مشابه لتباين العينة ذات البعد الواحد. اذا كانت قيمة المتغير (j) للعنصر (i) في العنقود (k) يقال لها (x_{ijk}) فان مركز العنقود (k) سيكون: [4]

$[\overline{X}.1K,\quad \overline{X}.2K..\quad \overline{X}.nK]$

اذ أن

$$\overline{X}.jK = \frac{1}{M_K}\sum_{j=1}^{MK} X_{ijK}$$

M_k: تمثل عدد العناصر في العنقود (k)

وعليه فان مجموع مربعات الخطأ الخاص بالعنقود (k) سيكون

$$E_K = \sum_{i=1}^{\wedge}\sum_{j=1}^{MK}(X_{ijK} - \overline{X}_{.jK})^2$$

ومجموع التشتت الكلي

$$E_{total} = \sum_{K=1}^{K} E_K$$

حيث أن

E_{total} (E_t): تمثل قياس لمقدار المعلومات المفقودة في حالة الاستعاضة عن العناصر بمراكز العناقيد عند اجراء عملية الربط.

في بداية عملية التعنقد يكون مجموع مربعات الخطأ يساوي صفراً ولكن عن ربط عنقودين ســـنلاحظ ان مجموع مربعات الخطأ سيزداد.

الجانب التطبيقي:

تتفاوت الامكانيات الاقتصادية المادية منها والبشرية بين اقطار الوطن العربي ويتسع هذا التفاوت بمرور الوقت مما يشكل تهديداً للتضامن الاقتصادي العربي الذي يشكل اولى خطوات التكامل الاقتصادي كذلك تتفاوت في عدم التوازن في الاتفاق على القطاعات المختلفة وتطويرها مما يؤدي الى تطور قطاعات على حساب قطاعات اخرى قد تكون اكثر فائدة للمجتمع، وان هدف البحث هو تكوين مجاميع من الأقطار العربية في قارة افريقيا والعراق ذات تجانس كبير فيما بينها باستخدام اسلوب التحليل العنقودي وقد تم تهيئة مصفوفة البيانات كما هو موضح في الجدول رقم (١) علماً أن هذه البيانات هي بيانات معيارية لـــ(١٠)

اقطار عربية تقع في قارة افريقيا مع العراق هي (جيبوتي والصومال والسودان ومصر وليبيا وتونس والجزائر والمغرب وموريتانيا والعراق).

وان المتغيرات (المعايير) المستخدمة في قياس التشابه أو الاختلاف هي:-

١- عدد الافراد الذين يحصلون على مياه مأمونة % للعام ١٩٩٧ Y1

٢- عدد الافراد الذين يحصلون على خدمات صحية % للعام ١٩٩٧ Y2

٣- عدد الافراد الذين يحصلون على مرافق صحية % للعام ١٩٩٧ Y3

٤- الانفاق العام على التعليم كنسبة مئوية من الناتج المحلي (١٩٩٧) Y4

٥- الانفاق العام على الصحة كنسبة مئوية من الناتج المحلي (١٩٩٧) Y5

٦- الانفاق العام على التجهيزات العسكرية من الناتج الاجمالي (١٩٩٧) Y6

٧- استيراد الاسلحة العسكرية مليون دولار (١٩٩٧) Y7

٨- استهلاك الكهرباء بالالف (١٩٩٧) Y8

٩- تعليم النساء البالغات لعام (١٩٩٧) Y9

أما الدول العربية فيرمز لها كالآتي:

١- المغرب x_1.

٢- الصومال x_2

٣- السودان x_3.

٤- مصر x_4.

٥- ليبيا x_5

٦- جيبوتي x6

٧- الجزائر x7

٨- تونس x8

٩- موريتانيا x9

١٠- العراق x10

جدول رق (١) مصفوفة البيانات المعيارية

	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9
x_1	**-0.1002**	**0.3327**	**0.0445**	**-0.4975**	**-1.4587**	**2.3060**	**-0.2795**	**-0.5030**	**-0.0889**
x_2	**0.5559**	**0.6596**	**0.9485**	**0.5880**	**0.5851**	**1.5369**	**0.1387**	**3.0432**	**0.8032**
x_3	**0.6106**	**0.6596**	**.9485**	**-.0505**	**1.0467**	**-01867**	**-.589**	**1.279**	**.9596**
x_4	**.6106**	**.6596**	**.4485**	**-.8807**	**1.1126**	**-.4519**	**-.0898**	**1.8423**	**.9596**
x_5	**-7746**	**.6129**	**.7430**	**-.1782**	**-.4038**	**-.484**	**.8647**	**1.2720**	**-7736**
x_6	**.6653**	**.5195**	**-4965**	**-8434**	**.3214**	**1.8817**	**3.742**	**.5657**	**.1689**
x_7	**-1731**	**-6129**	**-.2842**	**-.3698**	**-1.4587**	**.1580**	**-.2795**	**-.5644**	**-.8842**
x_8	**.8293**	**.5195**	**.1266**	**1.6097**	**.7170**	**.7149**	**-.4994**	**-.6054**	**.9850**
x_9	**.2278**	**.6129**	**.5789**	**.1410**	**-.6016**	**-.4784**	**.9726**	**-.6566**	**-.7613**

x_{10}	-1.1939	-2.2357	-2.0511	.8434	-1..0631	-.5845	-.5191	-.7840	-1.5657

ان حساب التشابه بين العناصر يمكن ان يقاس معرفة مقدار الاختلاف بين العناصر وذلك حسب العلاقة التالية
Similarity = 1- dissimilarity

ويشير جدول رقم (٢) الى المسافة المسماة Euclidian distance أو مصفوفة المسافة وقد تم استخراج النتائج باستخدام البرنامج الاحصائي الجاهز MATLAB بواسطة الحاسوب ولجميع نتائج البحث أما جدول رقم (٣) فيوضح طرق التعنقد المشار لها في الجانب النظري، والجدول رقم (٤) يشير الى مصفوفة المسافة المسماة Euclidian distance القياسية وللبيانات نفسها والجدول رقم (٥) يمثل نتائج طرق التعنقد باستخدام المسافة المسماة Euclidian distance القياسية.

جدول رقم (٢) مصفوفة المسافة بين مجموعة الاقطار العربية باستخدام المسافة المسماة Euclidian distance

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
x_1	**0**	**4.5648**	**4.2953**	**4.2358**	**3.9019**	**4.7981**	**2.2115**	**3.7183**	**3.376**	**4.871**
x_2		**0**	**2.0935**	**2.8124**	**3.0244**	**4.4902**	**4.7571**	**4.0316**	**4.7692**	**6.9634**
x_3			**0**	**1.1536**	**2.0950**	**5.0545**	**3.5714**	**2.8252**	**3.4927**	**6.0565**
x_4				**0**	**2.0312**	**5.1192**	**3.9843**	**3.824**	**3.8214**	**6.4746**
x_5					**0**	**4.014**	**2.8878**	**3.4060**	**2.5659**	**5.7944**
x_6						**0**	**5.0643**	**4.6886**	**4.1101**	**6.144**
x_7							**0**	**3.2813**	**2.04119**	**4.2005**
x_8								**0**	**3.3792**	**5.3347**
x_9									**0**	**4.547**
x_{10}										**0**

جدول رقم (3)

نتائج تطبيق الربط (المنفرد، الشامل، Ward)

الربط المنفرد		الربط الشامل		Ward	
مستوى الالتحام	العناصر الملتحمة	مستوى الالتحام	العناصر الملتحمة	مستوى الالتحام	العناصر الملتحمة
1.1536	(3,4), 1,2,5,6,7,8,9,10	1.1536	(3,4),1,2,5,6,7,8,9,10	0.8157	(3,4),1,2,5,6,7,8,9,10
2.0317	(3,4,5), 1,2,6,7,8,9,10	2.0419	(3,4),(7,9), 1,2,5,6,8,10	1.4439	(3,4),(7,9), 1,2,5,6,8,10
2.0419	(3,4,5),(7,9),1,2,6,8,10	2.0950	(3,4,5),(7,9) 1,2,6,8,10	1.6178	(3,4,5),(7,9) 1,2,6,8,10
2.2125	(3,4,5,7,9,1),2,6,8,10	3.0244	(3,4,5,2),(7,9),1,6,8,10	2.1762	(3,4,5),(7,9,1),2,6,8,10
2.5659	(3,4,5,7,9,1),2,6,8,10	3.3337	(3,4,5,2),(7,9,8),1,6,10	2.2930	(3,4,5,2),(7,9,1),6,8,10
2.6935	(3,5,4,7,9,1,2),6,8,10	3.7183	(3,4,5,2),(7,9,8,1),6,10	2.6872	(3,4,5,2),(7,9 1.8),6,10
2.8252	(3,5,4,7,9,1,2,8),6,10	4.7692	(3,4,5,2,7,9,8),6,10	3.8341	(3,4,5,2),(7,9,1,8,6),10
4.0121	(3,5,4,7,9,1,2,8,6),10	5.1192	(3,4,5,2,7,9,8,1,6),10	4.3046	(3,4,5,2),(7,9,1,8,6,10)
4.2005	(3,5,4,7,9,1,2,8,6,10)	6.9634	(3,4,5,2,7,9,8,1,6,10)	5.0098	(3,4,5,2),(7,9,1,8,6,10)

جدول رقم (4)

مصفوفة المسافة بين الأقطار العربية باستخدام المسافة المسماة Euclidian distance

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
X_1	0	4.1590	4.1898	4.5054	3.7226	4.2812	2.0796	4.3083	3.1931	5.4891
X_2		0	2.3502	2.8717	2.7353	3.5056	3.2560	3.3598	4.1326	7.2396
X_3			0	1.2624	1.8883	4.2166	3.5004	2.9313	3.2913	6.8180
X_4				0	2.0083	4.5826	3.8022	4.0738	3.6731	7.2621
X_5					0	3.5386	2.7142	3.4001	2.5224	6.5805
X_6						0	4.4366	3.8176	3.6414	7.0163
X_7							0	3.8412	1.9884	5.0228
X_8								0	3.6451	6.385
X_9									0	5.2168
X_{10}										0

جدول رقم (5)

نتائج تطبيق الربط (المنفرد، الشامل، Ward) باستخدام المسافة المسماة Euclidian distance

الربط المنفرد		الربط الشامل		Ward	
مستوى الالتحام	العناصر الملتحمة	مستوى الالتحام	العناصر الملتحمة	مستوى الالتحام	العناصر الملتحمة
1.2624	(3,4), 1,2,5,6,7,8,9,10	1.2624	(3,4),1,2,5,6,7,8,9,10	0.8926	(3,4),1,2,5,6,7,8,9,10
1.8883	(3,4,5), 1,2,6,7,8,9,10	1.9884	(3,4),(7,9), 1,2,5,6,8,10	1.4060	(3,4),(7,9), 1,2,5,6,8,10
1.9884	(3,4,5),(7,9),1,2,6,8,10	2.0083	(3,4,5),(7,9) 1,2,6,8,10	1.5058	(3,4,5),(7,9) 1,2,6,8,10
2.0796	(3,4,5),(7,9,1),2,6,8,10	2.2717	(3,4,5,2),(7,9),1,6,8,10	2.0448	(3,4,5),(7,9,1),2,6,8,10
2.3502	(3,4,5,2),(7,9,1),6,8,10	3.1931	(3,4,5,2),(7,9,1),6,8,10	2.1324	(3,4,5,2),(7,9,1),6,8,10
2.5224	(3,4,5,2),(7,9,1),6,8,10	3.2176	(3,4,5,2),(7,9,1),(6,8),10	2.6994	(3,4,5,2),(7,9,1),6,8,10
2.9313	(3,4,5,2,7,9,1,8),6,10	4.4366	(3,4,5,2),(7,9,1,6,8),10	3.3481	(3,4,5,2,6,8),(7,9,1),10
3.5056	(3,4,5,2,7,9,1,8,6),10	4.5826	(3,4,5,2,7,9,1,6,8),10	4.1040	(3,4,5,2,6,8,7,9,1),10
5.0228	(3,4,5,2,7,9,1,8,6,10)	7.2621	(3,4,5,2,7,9,1,6,8,10)	5.6472	(3,4,5,2,6,8,7,9,1,10)

الاستنتاجات

نلاحظ من نتائج التعنقد الموضحة في الجدولين (٣)، (٥) وللطرق الثلاث وباستخدام المسافة المسماة Euclidian distance القياسية ان نتائجها كانت مطابقة في اظهار ان الدولتين قطر والبحرين كانت الاقرب والاكثر تجانساً في اتجاهات الانفاق للدخل القومي.

كذلك اظهرت لنا النتائج ان المغرب كانت ذات تقارب وتجانس مع الدولتين الجزائر وموريتانيا وبجمع طرق التعنقد اضافة الى تونس التي اظهرت لنا طريقة الربط الشامل باستخدام مصفوفة المسافة المسماة Euclidian distance وكانت الصومال ذات تقارب وتجانس مع دولة السودان وليبيا مع اختلافات في مستويات التعنقد للطرق الثلاثة، وكانت دولة العراق من الدول العربية الاقل تقارباً مع باقي الدول التي تقع في قارة افريقيا.

ونلاحظ تطابق تقريبي للنتائج في طرق العنتقد مع اختلافات بسيطة في مستويات التعنقد.

المصادر

١. رشيد، ظافر حسين، وباقر، لميعة، " إستخدام التحليل العنقودي للتحري عن مصادر المياه الجوفية المغذية لعيون جبل سنجار في شمال العراق "، *مجلة العلوم الاقتصادية والإدارية*، العدد الأول، ص ص ٥٧-٦٥، ١٩٩٤

٢. الجبوري، شلال، وحمزة، صلاح، *تحليل متعدد المتغيرات*، دار الكتب لجامعة بغداد، بغداد، ٢٠٠٠

3. Everett, Brain, *Cluster Analysis*, Gox & Wyman, London, 1977
4. Anderson, T.W., *An Introduction to Multivariate Statistical Analysis*, John Wiley, New York, 1974