المحوران الخامس والسادس

التحليل العنقودي + التحليل العنقودي الهرمي

1- التحليل العنقودي (K − Means Cluster Analysis)

1.1- المفهوم

أول من استخدم مصطلح التحليل العنقودي هو العالم تريون $(Tryon)$ في عام 1939م، وهو تحليل يسعى إلى تصنيف المشاهدات في مجموعات ويعتمد على معامل التشابه $(Similarity)$ وعدم التشابه $(Dissimilarity)$، وذلك من خلال حساب المسافات بين المشاهدات.

ووفقا لعملية حساب المسافات تتكون المجموعات بطريقتين:

1.1.1-التجميع Agglomeration: وهي طريقة تنطلق من الجزء إلى الكل، بتعيين كل مشاهدة أو حالة في مجموعة مستقلة، ومن ثم إدماج الحالات المتشابهة تدريجيا حتى تتكون في النهاية مجموعة شاملة لكافة الحالات.

2.1.1-التقسيم $(Division)$: طريق تنطلق من الكل إلى الجزء، بتعيين كافة الحالات في مجموعة كلية واحدة، ومن ثم تبدأ عملية التقسيم حتى تكتمل في النهاية بتعيين مجموعات فرعية تضم كل منها حالة أو عدة حالات حسب طبيعة التشابه بينها.

بعبارة أخرى، يمكن القول بأن الطريقة الأولى تبدأ من أسفل إلى أعلى، بينما تبدأ الثانية من أعلى إلى أسفل، وأن معظم البرمجيات تعتمد على الطريقة التجميعية كما سنرى لاحقا في التحليل العنقودي الهرمي. وبغض النظر عن الطريقة المستخدمة، فإن التحليل العنقودي يتصف بخاصيتين هما:

3.1.1-التماسك الداخلي $(Internal\ Cohesion)$: وهو يشير إلى أن الحالات أو المشاهدات في كل مجموعة متشابهة.

4.1.1-العزل الخارجي $(External\ Isolation)$: وهذا يشير إلى أن الحالات في إحدى المجموعات تكون بمعزل عن بقية الحالات في المجموعات الأخرى تبعا للمسافات التي بينها.

2.1-الافتراضات

1) المعرفة المسبقة بعدد المجموعات المستهدفة $(Clusters)$.
2) أن يكون مستوى القياس في المتغيرات فتريا أو نسبيا.
3) أن يكون الهدف منصبا على تصنيف الحالات $(Cases)$ فقط.
4) تجانس وحدة القياس في المتغيرات.
5) أن يتضمن النموذج كافة المتغيرات ذات الصلة بالظاهرة محل التصنيف.

3.1-الآلية:

يعتمد التحليل العنقودي (K − Means)على حساب المسافة الإقليدية $(Euclidean\ Distance)$، وهي المسافة الهندسية بين متغيرين أو أكثر، وتعد المقياس الأكثر شيوعا في تقدير التشابه وعدم التشابه. وفي بعض الأحوال يمكن استخدام معامل الارتباط، ولاكن يعاب عليه إغفاله للفروق الحقيقية بين القيم، وهذا الأمر ما استدركته المسافة الإقليدية.

من خلال حساب هذه المسافة يتم إنشاء مصفوفة التقارب $(Proximity\ Matrix)$، فعلى سبيل المثال: إذا كانت لدينا البيانات التالية:

الطالب	درجة التحصيل
زياد	15
خالد	9
أسامه	13
حاتم	10

فستكون مصفوفة التقارب على النحو التالي:

	زياد	خالد	أسامة	حاتم
زياد	0	6	2	5
خالد		0	4	1
أسامة			0	3
حاتم				0

من خلال المصفوفة يتضح أنها تعتمد على الفروق بين القيم لكافة الحالات، ويمكن أن تصنف مبدئيا في مجموعتين:

- الأولى تضم خالدا، وحاتما على اعتبار أن المسافة بينهما هي الأقل (=1)؛
- الثانية تضم زيادا، وأسامة حيث بلغت المسافة بينهما (2).

تعد الصورة السابقة هي الصور البسيطة لحساب المسافات على اعتبار أن البيانات انطوت على متغير وحيد، في حين أن الآلية ستختلف في حالة وجود متغيرين أو أكثر، وذلك على النحو التالي:

الطالب	المتغير الأول	المتغير الثاني
زياد	15	18
خالد	9	12
أسامة	13	14
حاتم	10	16

في هذه الحال ينبغي حساب المسافة الإقليدية وفقا للعلاقة التالية:

$$d_{ij} = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2} \quad \dots (*)$$

وبتطبيق العلاقة $(*)$ أعلاه سنحصل على مصفوفة التقارب الآتية:

مصفوف التقارب Proximity Matrix

Case	Euclidean Distance			
	زياد: 1	خالد: 2	أسامة: 3	حاتم: 4
زياد: 1	0			
خالد: 2	8.485	0		
أسامة: 3	4.472	4.472	0	
حاتم: 4	5.385	4.123	3.606	0

This is a dissimilarity matrix

مثلا: المسافة بين زياد وخالد هي كالتالي: $d = \sqrt{(15-9)^2 + (18-12)^2} = 8.485$

وعندما تتضمن البيانات ثلاثة متغيرات أو أكثر فإن العلاق السابقة تمتد لتشملها بالطريقة نفسها.

تطبيق (TP)

يرغب مدير النشاط الطلابي في تصنيف خمس عشرة مدرسة إلى ثلاث مجموعات وفقا للدرجة التي حصلت عليها كل مدرسة في النشاط الطلابي بفروعه الستة وهي: النشاط الثقافي، والرياضي، والمسرحي، والبيئي، والفني، والكشفي، والدرجة الكلي لكل نشاط تساوي 100، وذلك قصد تكريم إدارات المدارس وفقا لإنجازاتها، والبيانات هي كالآتي:

النشاط / المدرسة	الثقافي	الرياضي	المسرحي	البيئي	الفني	الكشفي
1	78	88	75	85	81	91
2	79	66	72	75	68	66
3	84	67	68	58	62	63
4	45	68	67	54	67	62
5	75	74	70	69	63	65
6	90	75	72	61	69	57
7	78	61	68	47	71	58
8	48	58	64	48	71	54
9	49	57	58	52	44	55
10	58	59	57	48	74	71
11	88	65	51	47	81	72
12	91	62	60	46	80	75
13	78	70	71	89	79	67
14	71	63	59	84	74	65
15	74	49	62	66	75	69

تحليل مخرجات SPSS

ستظهر نتائج هذا التحليل على النحو التالي:

مراكز المجموعات الأولية Initial Cluster Centers

	Cluster		
	1	2	3
الثقافي	78	88	49
الرياضي	88	65	57
المسرحي	75	51	58
البيئي	85	47	52
الفني	81	81	44
الكشفي	91	72	55

يوضح الجدول أعلاه التقدير الأولي لمراكز المجموعات الثلاث، ويتم حسابها وفقا لقيم كل متغير في المجموعة التي ينتمي إليها.

خطوات التكرار Iteration History

Iteration	Change in Cluster Centers		
	1	2	3
1	26.819	20.340	18.055

2	4.146	4.794	5.802
3	0.000	0.000	0.000

يصف الجدول أعلاه عمليات التكرار في كل خطوة تتم فيها إعادة تعيين الحالات (المدارس) في مجموعة أخرى، وتبعا لذلك يتغير مركز المجموعة، ويعكس كل رقم في المجموعة الثانية المسافة التي ابتعد فيها مركز المجموعة عما كان عليه في المرحلة الأولى، وتستمر عملية التكرار حتى يكون التغير صفريا أو طفيفا جدا، وهذا ما نلحظه في الرتبة لثالثة حيث تشير القيم الصفرية إلى أن مراكز المجموعات اتسمت بالثبات والاستقرار الناتج عن عدم وجود انتقال أو إعادة تعيين للحالات بين المجموعات.

مراكز المجموعات النهائية Final Cluster Centers

	Cluster		
	1	**2**	**3**
الثقافي	76	84	50
الرياضي	72	63	61
المسرحي	67	64	62
البيئي	80	54	51
الفني	73	73	64
الكشفي	71	66	61

يوضح الجدول أعلاه التقدير النهائي لمراكز المجموعات الثلاث، وهو يفيد في تحديد مستوى أداء كل مجموعة، فيظهر جليا أن المجموعة الثالثة هي الأقل في درجات الأنشطة، بينما المجموعة الأولى هي الأفضل. ويشير الجدول التالي إلى ما خلصت إليه عملية التصنيف.

عدد الحالات في كل مجموعه Number of Cases in each Cluster

Cluster 1	5.000
2	6.000
3	4.000
Valid	15.000
Missing	0.000

نلاحظ أن الحالات تتوزع توزيعا منطقيا بين المجموعات، فالمجموعة الأولى تضم 5 مدارس وهي المجموعة المميزة في الأنشطة، بينما ضمت المجموعة الثانية 6 مدارس، في حين ضمت المجموعة الثالثة 4 مدارس.

وتجدر الإشارة أن ظهور مجموعة تستأثر بالعدد الأكبر من الحالات (المدارس) يعد مؤشرا على عدم فاعلية التصنيف، ويعزى ذلك إلى أن تحديد عدد المجموعات لم يكن سليما أو أن التحليل لم يشمل كافة المتغيرات المؤثرة.

بعد هذه الخطوة يتم الانتقال جدول الاقتران بحيث يدرج في الصفوف: رقم المدرسة، وفي الأعمدة: المتغير الجديد الذي ظهر باسم (Cluster Number of Cases) وستظهر النتيجة على النحو الذي يمكننا من معرفة المجموعة التي تنتمي إليها كل مدرسة.

Cluster Number of Cases Cross tabulation

		Cluster Number of Cases			Total
		1	2	3	
المدرسة	1	1			1
	10			1	1
	11		1		1
	12		1		1
	13	1			1
	14	1			1
	15		1		1
	2	1			1
	3		1		1
	4			1	1
	5	1			1
	6		1		1
	7		1		1
	8			1	1
	9			1	1
	Total	5	6	4	15

2- التحليل العنقودي الهرمي ($Hierarchical\ Cluster\ Analysis$)

1.2- المفهوم

من المناسب عقد مقارنة بين التحليل العنقودي والتحليل العنقودي الهرمي على النحو التالي:

	K-Means	Hierarchical
عدد المجموعات	يطلب تحديدها قبل إجراء التصنيف	لا يطلب ذلك
مستوى القياس	فتري أو نسبي	يتمتع بمرونة فائقة في التعامل مع مستويات قياس فترية وعددية وثنائية
الآلية	المساف الإقليدية	يستخدم المسافة الإقليدية وآليات أخرى عديدة
التصنيف	يصنف الحالات (Cases) فقط	يصنف الحالات والمتغيرات
العينة	يتناسب مع العينات كبير الحجم	يتناسب مع العينات صغيرة الحجم

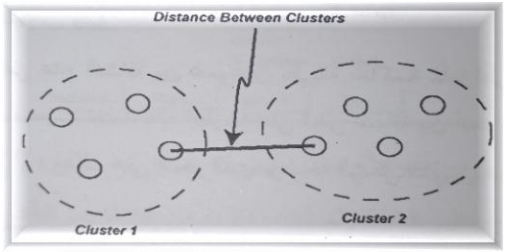
2.2-الآلية

يعتمد التحليل العنقودي الهرمي على حساب المسافات بعدة طرائق وعدة مقاييس، والطرائق هي كالتالي:

1.2.2-الترابط الأحادي ($Single\ Linkage$): ويطلق عليها في برنامج SPSS الجار القريب Nearest Neighbor وخطواتها كالآتي:

1) اعتبار كل حالة (Case) مجموعة مستقلة (Cluster).
2) البحث عن حالتين (أ) و(ب) تتسمان بأعلى قدر من التشابه ومن ثم إدماجهما معا في مجموعة واحدة.
3) تندمج حالة جديدة (ج) مع المجموعة السابقة وفقا لأعلى درجة من التشابه بينها وبين أي من الحالتين (أ) و(ب).
4) تستمر هذه العملية حتى تشمل كافة الحالات.

ويتم حساب عدم التشابه بين المجموعات في هذه الطريق من خلال قياس المسافة بين أقرب مشاهدتين في كل مجموعة، على النحو الذي يوضحه الشكل التالي:



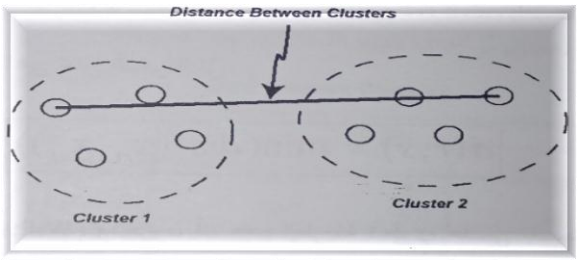
وعلاقة هذه الطريقة هي بالصيغة التالية: $d(r,s) = min\left(dist(x_{ri}, x_{sj})\right)$

ويرى ويلكس (Wilks) وآخرون أن هذه الطريقة تنطوي على بعض جوانب القصور، فهي غير قادرة على إنشاء مجموعات ممثلة على اعتبار أن المسافة التي تقيسها قصيرة، وبالتالي تميل هذه الطريقة إلى إدماج المجموعات معا، وهو ما يعرف بخطر السلسلة (Chain)، لذا تعد هذه الطريقة غير شائعة الاستخدام.

2.2.2-الترابط التام (Complete Linkage): ويطلق عليها في البرنامج اسم الجار البعيد (Furthest Neighbor)، وتتشابه في نقطة انطلاقها مع الطريقة السابقة، ولكن تختلف عنها في أمرين هما:

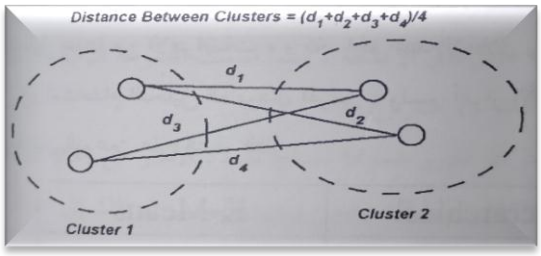
1) تدمج الحالة الجديدة (ج) بناء على تشابههما مع كلتا الحالتين (أ) و(ب) وليس بناء على تشابههما مع إحداهما، وتستمر العملية على هذا المنوال، وتعد هذه الطريقة الأكثر شيوعا والأفضل عند مقارنتها بطريقة الترابط الأحادي، على اعتبار أنها تعمل على تقليل عدم التشابه داخل المجموعة من خلال اعتمادها على التشابه ككل داخل المجموعة وليس تشابه حالة جديدة مع إحدى الحالات فقط.

2) يتم قياس عدم التشابه بين مجموعتين بطريق معاكسة تماما لطريق الترابط الأحادي، حيث تعتمد هذه الطريقة على قياس المسافة بين أبعد حالتين في كل مجموعة، وذلك على النحو الذي يوضحه الشكل التالي:



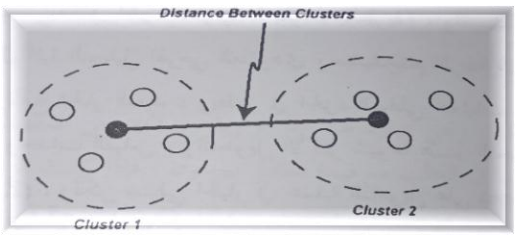
وعلاقة هذه الطريقة هي بالصيغة التالية: $\boldsymbol{d(r,s) = max\left(dist(x_{ri}, x_{sj})\right)}$

3.2.2-الترابط المتوسط أو المعتدل $\boldsymbol{(Avirage\ Likage)}$: يطلق عليها في البرنامج اسم (Between-Groups Linkage)، وهي طريقة وسيطة بين الطريقتين السابقتين، حيث تدمج الحالة الجديدة (ج) وفقا لأعلى درجة من تشابهها مع متوسط الحالتين (أ) و (ب)، ويتم قياس المسافة بين المجموعتين بناء على متوسط المسافة بين كافة الأزواج الممكنة، على النحو الذي يوضحه الشكل التالي:



وعلاقة هذه الطريقة هي بالصيغة التالية: $\boldsymbol{d(r,s) = \frac{1}{n_r n_s}\sum_{i=1}^{n_r}\sum_{j=1}^{n_s}\left(dist(x_{ri}, x_{sj})\right)}$

4.2.2-الترابط المتمركز $\boldsymbol{(Centroid\ Clustering)}$: **وهي** طريقة وسيطة أخرى تعمل وفقا لآلية الترابط المتوسط، ولكن تختلف عنها في أن قياس المسافة بين المجموعتين يتم وفقا للحالة الوسطى في كل مجموعة على النحو الذي يوضحه الشكل التالي:



وعلاقة هذه الطريقة هي بالصيغة التالية: $\boldsymbol{d(r,s) = d(x_r, x_s)}$

تطبيق (TP)

سعى مشرف تربوي إلى تصنيف المعلمين حسب مدى توافر الكفايات المهني اللازم لديهم (كفايات تعليمية، وعلمية، وشخصية)، وكانت البيانات على النحو التالي:

المعلم	تعليمية	علمية	شخصية
1	91	89	84
2	77	84	76
3	62	58	63
4	77	81	58
5	72	83	80
6	65	72	61
7	71	78	75
8	59	64	62
9	45	52	49
10	72	69	67
11	54	63	58
12	60	71	76
13	78	75	80
14	95	83	85
	70	71	74

وعلى اعتبار أن المثال يتضمن بيانات مستوى القياس فيها فتريا فسوف نتناول المقاييس الفترية المتوفرة في البرنامج وهي على النحو التالي:

- **المساف الإقليدية البسيط:** وهي الجذر التربيعي لمجموع مربعات الفروق بين القيم، ويتم حسابها وفقا للعلاقة التالية:

$$d_{ij} = \sqrt{\left(x_{i1} - x_{j1}\right)^2 + \left(x_{i2} - x_{j2}\right)^2}$$

ويمكن أن تأخذ الصيغة التالية: $d_{ij} = \sqrt{\left(x_{i1} - x_{j1}\right)^2} = \left|x_{i1} - x_{j1}\right|$

- **المسافة الإقليدية المربعة (Squared Euclidean):** وهي مجموع مربعات الفروق بين القيم، وهي الطريقة الأكثر شيوعا، والتي نستخدمها في مثالنا هذا، ويتم حسابها وفقا للعلاقة: $\boldsymbol{d(x,y) = \sum(x_i - y_i)^2}$.
- **معامل ارتباط بيرسون:** وهو مقياس الارتباط المعتاد.
- **جيب التمام (Cos):** وهي جيب تمام الزاوية بين اثنين من أشعة القيم، وتتشابه قيمتها في الغالب مع قيمة معامل بيرسون.
  - ✓ فإذا كانت الزاوية حادة فهي مؤشر على ارتباط موجب بين المتغيرين،
  - ✓ وإذا كانت الزاوية منفرجة فهي مؤشر على ارتباط سالب،
  - ✓ أما إذا كانت الزاوية قائمة فلا علاقة بين المتغيرين.

- **علاقة شيباشيف (Chebychev):** وهي مقياس مسافة يعتمد على تعظيم مجموع الفروق المطلقة بين القيم، وعلاقته تعطى بالصيغة التالية: $\boldsymbol{d(x,y) = Max|x_i - y_i|}$

- **طريقة الحاجز (Block):** وهي مقياس مسافة يعتمد على مجموع الفروق المطلق بين القيم، ويسمى أحيانا بمقياس **منهاتن** (Manhatten) على اعتبار عدم إمكانية الوصول بين نقطتين بشكل مباشر في تلك المنطقة التي تتسم بوجود عدد كبير من المباني الشاهقة، وصيغتها تعطى بالعلاقة التالية: $\boldsymbol{d(x,y) = \sum|x_i - y_i|}$

تحليل نتائج مخرجات SPSS

في البداية ستظهر مصفوفة التقارب على النحو التالي:

Proximity Matrix

	Squared Euclidean Distance														
Case	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1		285.000	2243.000	936.000	376.000	1494.000	602.000	2133.000	4710.000	1050.000	2721.000	1349.000	381.000	53.000	865.000
2	285.000		1070.000	333.000	33.000	513.000	73.000	920.000	2777.000	331.000	1294.000	458.000	98.000	406.000	222.000
3	2243.000	1070.000		779.000	1035.000	209.000	625.000	46.000	521.000	237.000	114.000	342.000	834.000	2198.000	354.000
4	936.000	333.000	779.000		504.000	234.000	334.000	629.000	1946.000	250.000	853.000	713.000	521.000	1057.000	405.000
5	376.000	33.000	1035.000	504.000		546.000	54.000	881.000	2706.000	366.000	1245.000	329.000	89.000	509.000	189.000
6	1494.000	513.000	209.000	234.000	546.000		268.000	101.000	944.000	94.000	211.000	251.000	539.000	1597.000	195.000
7	602.000	73.000	625.000	334.000	54.000	268.000		509.000	2028.000	146.000	803.000	171.000	83.000	701.000	51.000
8	2133.000	920.000	46.000	629.000	881.000	101.000	509.000		509.000	219.000	42.000	246.000	806.000	2186.000	314.000
9	4710.000	2777.000	521.000	1946.000	2706.000	944.000	2028.000	509.000		1342.000	283.000	1315.000	2579.000	4757.000	1611.000
10	1050.000	331.000	237.000	250.000	366.000	94.000	146.000	219.000	1342.000		441.000	229.000	241.000	1049.000	57.000
11	2721.000	1294.000	114.000	853.000	1245.000	211.000	803.000	42.000	283.000	441.000		424.000	1204.000	2810.000	576.000
12	1349.000	458.000	342.000	713.000	329.000	251.000	171.000	246.000	1315.000	229.000	424.000		356.000	1450.000	104.000
13	381.000	98.000	834.000	521.000	89.000	539.000	83.000	806.000	2579.000	241.000	1204.000	356.000		378.000	118.000
14	53.000	406.000	2198.000	1057.000	509.000	1597.000	701.000	2186.000	4757.000	1049.000	2810.000	1450.000	378.000		890.000
15	865.000	222.000	354.000	405.000	189.000	195.000	51.000	314.000	1611.000	57.000	576.000	104.000	118.000	890.000	

This is a dissimilarity matrix

جدول مصفوفة التقارب

وهي مصفوف تساعدنا في التعرف على درجه التقارب بين الحالات، فكلما كانت القيم صغيره فهي مؤشر على التشابه بين الحالات، والعكس صحيح، ومن خلال المصفوف يتبين أن أول حالتين ستتضمان في مجموعه واحده هي الحال الثانية والحالة الخامسة حيث بلغت المسافة الإقليدية المربعة بينهما 33، والتي يمكن حسابها من خلال البيانات الخام –المدرج في جدول بيانات المثال–على النحو التالي: $\boldsymbol{(77-73)^2 + (84-83)^2 + (76-80)^2 = 33}$.

ويوضح الجدول التالي عمليه اندماج الحالات في مجموعات، حيث اندمجت الحالة الثانية مع الحالة الخامسة في أول خطوه، على اعتبار قصر المساقة بينهما، ويشير العمود الأخير (Next Stage) إلى المرحلة اللاحقة، وحيث ظهرت المرحل السادسة فإن هذا يدل على أن إحدى الحالتين (2 أو 5) ستندمج مع حال جديه في المرحلة السادسة (اندمجت 2 مع 13)، ونلاحظ من الجدول أيضا أن معامل المساقة – العمود الرابع–يزيد كلما اندمجت الحالات، ويعزى لك إلى ميل الطريقة التجميعية في النهاية إلى وضع كافه الحالات في مجموعه كليه واحه.

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	3	5	33.000	0	0	6
2	8	11	42.000	0	0	7
3	7	15	51.000	0	0	8
4	1	14	53.000	0	0	13

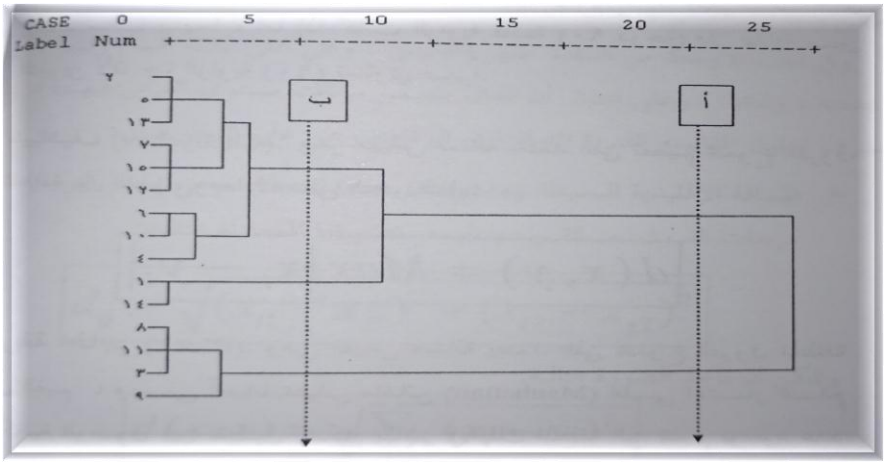
5	6	10	94.000	0	0	9
6	2	13	98.000	1	0	10
7	3	8	114.000	0	2	11
8	7	12	171.000	3	0	10
9	4	6	250.000	0	5	12
10	2	7	458.000	6	8	12
11	3	9	521.000	7	0	14
12	2	4	**713**.000	10	9	13
13	1	2	**1597**.000	4	12	14
14	1	3	4757.000	13	11	0

وتأسيسا على أننا اقترحنا مبدئيا أن يكون عدد المجموعات من (2–4) فسيظهر لنا جدول يوضح موقع كل حاله في كل خيار من الخيارات الثلاثة وذلك على النحو التالي:

Cluster membership

Case	4 Cluster	3 Cluster	2 Cluster
1	**1**	**1**	**1**
2	**2**	**2**	**1**
3	**3**	**3**	**2**
4	**4**	**2**	**1**
5	**2**	**2**	**1**
6	**4**	**2**	**1**
7	**2**	**2**	**1**
8	**3**	**3**	**2**
9	**3**	**3**	**2**
10	**4**	**2**	**1**
11	**3**	**3**	**2**
12	**2**	**2**	**1**
13	**2**	**2**	**1**
14	**1**	**1**	**1**
15	**2**	**2**	**1**

في الجزء الأخير من النتائج سيظهر رسم الشجرة وهو ما يعرف بالمصطلح (Dendrogram) على النحو التالي:



ويمكننا تقسيم الحالات إلى مجموعتين إذا رسمنا الخط العمودي (أ)، كما يمكن تقسيم الحالات إلى ثلاث مجموعات برسم الخط العمودي (ب)، والجدير بالذكر أن الخطوط العمودية المتقطعة لا تظهر في النتائج بل يرسمها المحلل.

وهنا يبرز تساؤل عن الآلية التي تمكننا من التحديد الأمثل لعدد المجموعات؟

للإجابة عن هذا السؤال ينبغي الرجوع إلى جدول عملية التجميع (Agglomeration) وننظر إلى الفقرة الاستثنائية في معامل المسافة، عندها يمكن حساب تلك المرحلة وما تبقى من مراحل على اعتبار أنها عدد المجموعات.

وبالنظر إلى الجدول يتبين أن الفقرة بدأت في المرحلة الثانية عشرة ما بين القيمة (**713**) والقيمة (**1597**)، وهي القيم التي تم إبرازها في الجدول بنمط مختلف، وتبعا لذلك فإن عدد المجموعات الأمثل هو (3) على اعتبار أن عدد المراحل ثلاث (12، 13، 14)، وبالتالي فإن الخط العمودي (ب) هو خط التقسيم الأمثل.

وبمقارنة توزيع الحالات –كما ظهرت في رسم الشجرة–بما ورد في جدول (Cluster Membership) يتبين أن النتيجة واحدة، فعندما تقرر أن خط التقسيم (ب) هو خط التقسيم الأمثل فسيكون لدينا ثلاث مجموعات تتوزع فيها الحالات بالطريقة نفسها الواردة في الجدول تحت عمود (Cluster 3)، وتأسيسا على ما سبق يمكن القول أن المعلمين يصنفون وفقا لدرجة توافر الكفايات المهنية إلى ثلاث مجموعات على النحو التالي:

	درجة توافر الكفايات المهنية		
	عالية	**متوسطة**	**منخفضة**
رقم المعلم	1، 14	3، 8، 9، 11	2، 4، 5، 6، 7، 10، 12، 13، 15