جامعة محمد خيضر (بسكرة)
كلية العلوم الاقتصادية والتجارية والتسيير
قسم العلوم التجارية

السنة الجامعية: 2026/2025
المستوى: أولى ماستر تسويق
مقياس: الأساليب الكمية في التسويق 2

أعمال تطبيقية رقم (01) حول التحليل العنقودي

أولا. المراجع المستخدمة:

محفوظ جودة (2008)، التحليل الإحصائي المتقدم باستخدام SPSS، دار وائل، عمان، الأردن.
نور الهدى بوهنتالة (2019-2018)، محاضرات في بحوث التسويق، مطبوعة موجهة لطلبة سنة ثالثة علوم تجارية، شعبة تسويق، جامعة باتنة.

ثانيا. ملخص نظري:

1. تعريف التحليل العنقودي Cluster Analysis

هو إجراءات تهدف إلى تصنيف مجموعة حالات Cases (أو متغيرات Variables) بطرق معينة وترتيبها داخل عناقيد Clusters بحيث تكون الحالات المصنفة داخل عنقود معين متجانسة فيما يتعلق بخصائص محددة وتختلف عن حالات أخرى موجودة في عنقود آخر.

2. التحليل العنقودي الهرمي Hierarchical Clustering:

لا يتطلب المعرفة المسبقة بعدد العناقيد التي سيتم تصنيف الحالات على أساسها، كما أنه يناسب العينات الصغيرة نسبيا، لكنه يتطلب الإلمام بكيفية تحديد التشابه أو المسافات بين الحالات، وكيفية تجميع الحالات في عناقيد، حيث أن هناك طرق عديدة لتحديد القرب Proximities بين الحالات، والمسافات distances بين الحالات أو المتغيرات، لكن أكثرها شيوعا طريقة المسافة الإقليدية Euclidean Distances، والتي تعتبر من أهم مقاييس التباعد في هذا المجال.

هناك نوعين من التحليل لتكوين العناقيد :

أ- التحليل العنقودي التقسيمي Divisive: يبدأ هذا النوع من التحليل العنقودي باعتبار أن جميع الحالات تتجمع في عنقود واحد ، وبعد ذلك يتم تصنيف الحالات تدريجيا في عناقيد أصغر فأصغر.

ب- التحليل العنقودي التجميعي Agglomerative: يبدأ التحليل التجميعي بعنقود واحد لكل حالة، ثم يتم تجميع العناقيد المتشابهة تدريجيا حتى نصل في النهاية إلى العدد المطلوب من العناقيد.

3. التحليل العنقودي غير الهرمي Non-Hierarchical Clustering:

هو أسلوب إحصائي لتصنيف البيانات إلى عدد محدد مسبقًا من المجموعات (العناقيد) (K) بناءً على تشابه الخصائص، حيث يتم تجميع الحالات حول مراكز (Centroids) لتقليل التباين داخل العنقود، وزيادة التباين بين المجموعات. أشهر خوارزمياته هي طريقة ك-متوسطات ($K-means$)، التي تتميز بالسرعة والفعالية في التعامل مع مجموعات البيانات الكبيرة.

تقوم طريقة ك- متوسطات (K-Means) على أساس تصنيف الحالات في مجموعات متجانسة من حيث خصائص أو صفات معينة، وذلك باستخدام برمجيات يمكن أن تعالج عدد كبير من الحالات، وتتم وفق الخطوات التالية:

1. تحويل البيانات الموجودة في المتغيرات إلى قيم معيارية، إذا كانت المتغيرات مختلطة (مقاسة بوحدات متنوعة)، من خلال اختيار القائمة الرئيسة Analyze، ثم القائمة الفرعية Descriptive Statistics، ثم الضغط على Descriptives، والتأشير على المربع الصغير أمام Save Standardized values as variables داخل صندوق الحوار Descriptives؛
2. تحديد عدد العناقيد المطلوب أن يجرى التصنيف على أساسها؛
3. تحديد متوسط قيم العناقيد (Centroids) بشكل مبدئي؛
4. حساب المسافات بين كل مفردة ومراكز المتوسطات؛
5. نقل المفردة إلى العنقود الذي يكون متوسطه أقرب إليها (استخدام مقياس Euclidean Distances)؛
6. إعادة حساب متوسطات قيم العناقيد الجديدة؛
7. تكرار نفس الخطوات (من 3 إلى 6)، حتى تتوصل إلى أنه ليس هنالك أية نقطة يمكن تحريكها إلى متوسط قيم العناقيد بشكل أكثر قرباً من الوضع الحالي.

مثال تطبيقي (1):

أجرى باحث تسويق في إحدى الشركات دراسة تتعلق بالعادات الاستهلاكية في 15 منطقة بيعية، حيث قام بتوزيع استبانات على عملاء الشركة نلخص نتائجها فيما يلي:

المنطقة	السكن	الطعام	الهواتف	السيارات	أخرى
A	200	100	60	110	80
B	150	90	40	10	60
C	400	250	30	200	30
D	420	210	40	180	80
E	80	40	10	20	10
F	390	220	30	190	170
G	140	170	80	60	20
H	250	200	80	120	110
I	270	180	50	60	90
J	350	230	30	180	50
K	110	70	40	90	70
L	140	90	70	20	30
M	240	200	60	120	40
N	240	120	30	120	200
O	320	210	20	210	70

المطلوب: إجراء التحليل العنقودي الهرمي، وتصنيف المناطق البيعية إلى ثلاثة عناقيد، وذلك لمساعدة الإدارة في تجزئة السوق ورسم الاستراتيجية التسويقية المناسبة.

حل المثال التطبيقي:

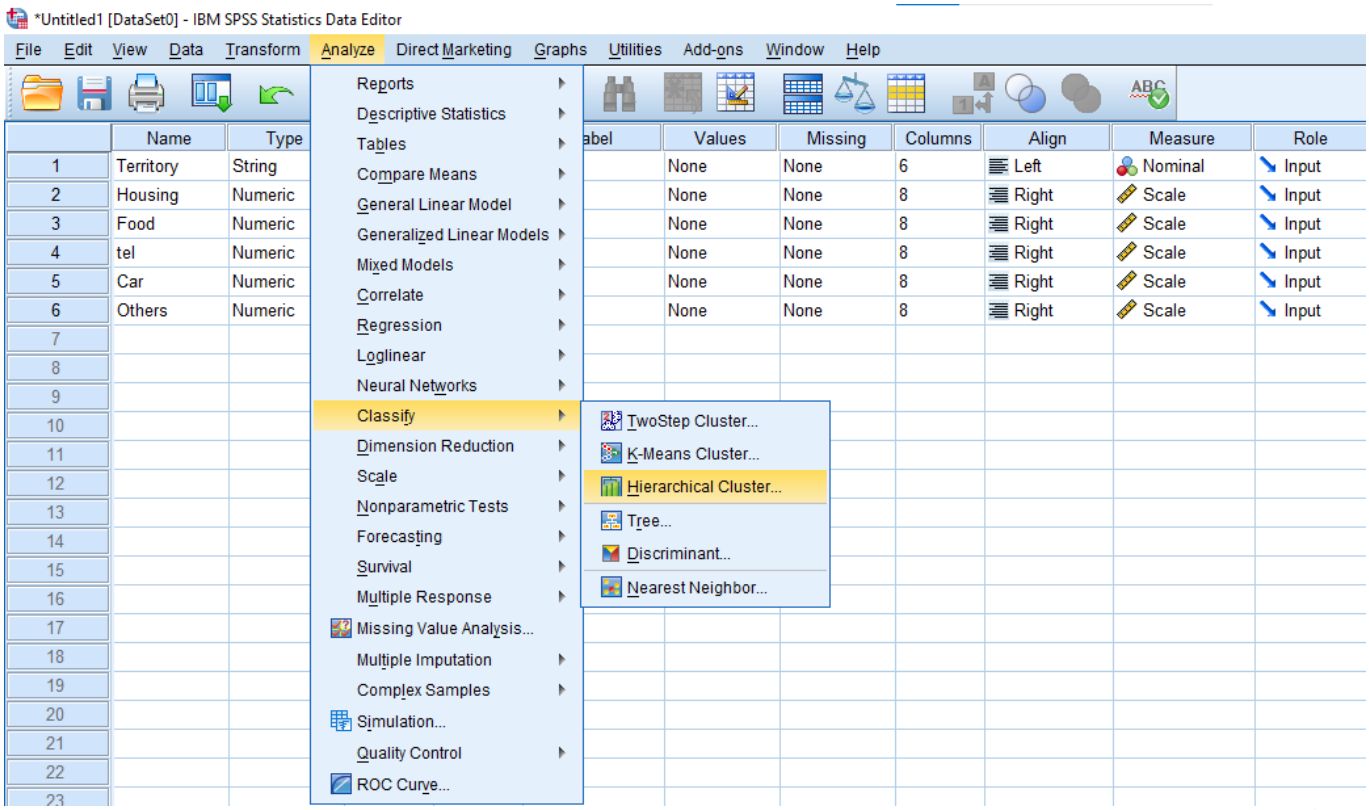
الخطوة (1): إدخال البيانات الواردة في المثال (انطلاقا من ملف Exel الأول المرفق) إلى نافذة البيانات SPSS (Data View)، ثم في نافذة المتغيرات Variable View، يتم تسمية 6 متغيرات: المنطقة Territory (نوعه: Chain، القياس Measure: اسمي Nominal)، السكن Housing، الغذاء Food، الاتصالات Tel، السيارات Car، أخرى Others، وهي كلها متغيرات عددية Numeric، والقياس Scale؛ وها تظهر صفحة البيانات كما يلي:

Territory	Housing	Food	Tel	Car	Others
A	200.00	100.00	60.00	110.00	80.00
B	150.00	90.00	40.00	10.00	60.00
C	400.00	250.00	30.00	200.00	30.00
D	420.00	210.00	40.00	180.00	80.00
E	80.00	40.00	10.00	20.00	10.00
F	390.00	220.00	30.00	190.00	170.00
G	140.00	170.00	80.00	60.00	20.00
H	250.00	200.00	80.00	120.00	110.00
I	270.00	180.00	50.00	60.00	90.00
J	350.00	230.00	30.00	180.00	50.00
K	110.00	70.00	40.00	90.00	70.00
L	140.00	90.00	70.00	20.00	30.00
M	240.00	200.00	60.00	120.00	40.00
N	240.00	120.00	30.00	120.00	200.00
O	320.00	210.00	20.00	210.00	70.00

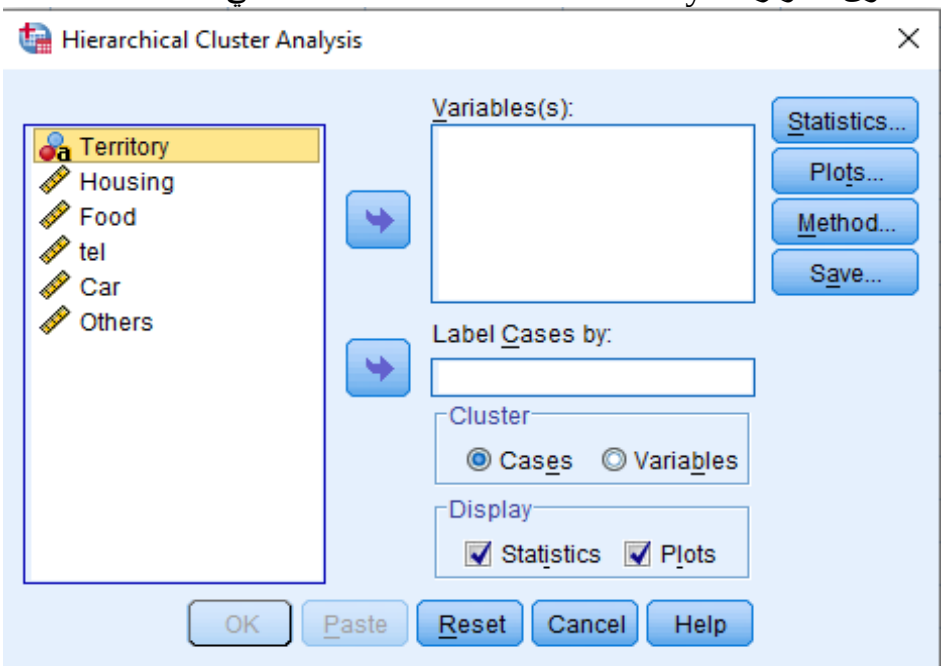
أما في نافذة المتغيرات Variable View فتظهر كما يلي:

Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
Territory	String	1	0		None	None	6	Left	Nominal	Input
Housing	Numeric	8	2		None	None	8	Right	Scale	Input
Food	Numeric	8	2		None	None	8	Right	Scale	Input
tel	Numeric	8	2		None	None	8	Right	Scale	Input
Car	Numeric	8	2		None	None	8	Right	Scale	Input
Others	Numeric	8	2		None	None	8	Right	Scale	Input

الخطوة (2): اختيار القائمة الرئيسة Analyze، ثم القائمة لفرعية Classify، ثم الضغط على Hierarchical Cluster؛



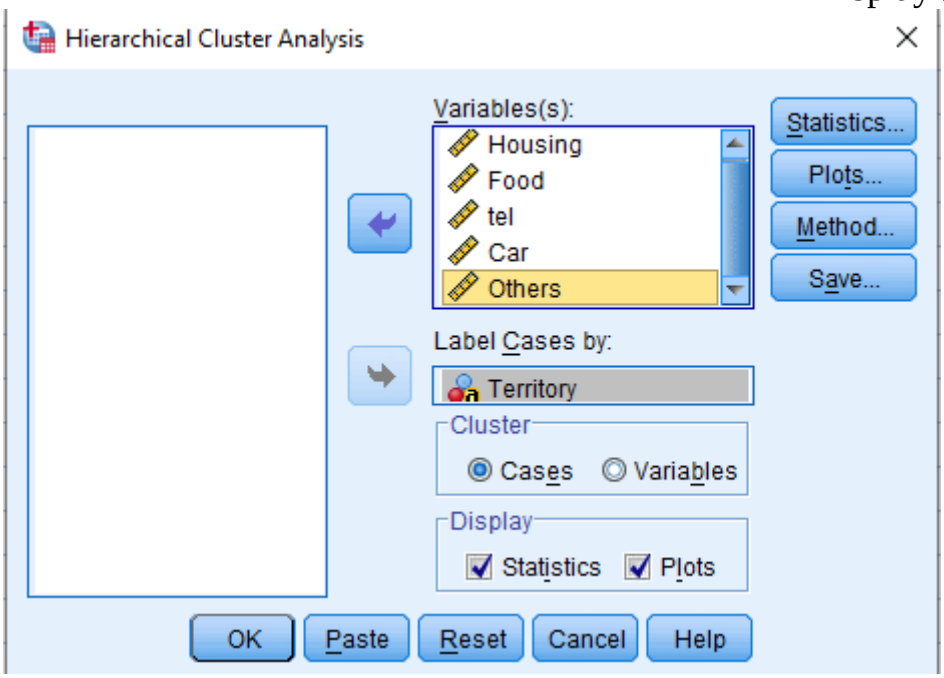
الخطوة (3): يظهر صندوق الحوار Hierarchical Cluster Analysis التالي:



يتم نقل المتغيرات الخمسة إلى داخل المستطيل المعنون Variable(s)، ونقل المتغير Territory إلى داخل المستطيل المعنون Label Cases by؛

الخطوة (4): اختيار عنقدة الحالات Cases تحت Cluster، حيث أن المطلوب تصنيف الحالات وليس المتغيرات Variables في هذا المثال؛

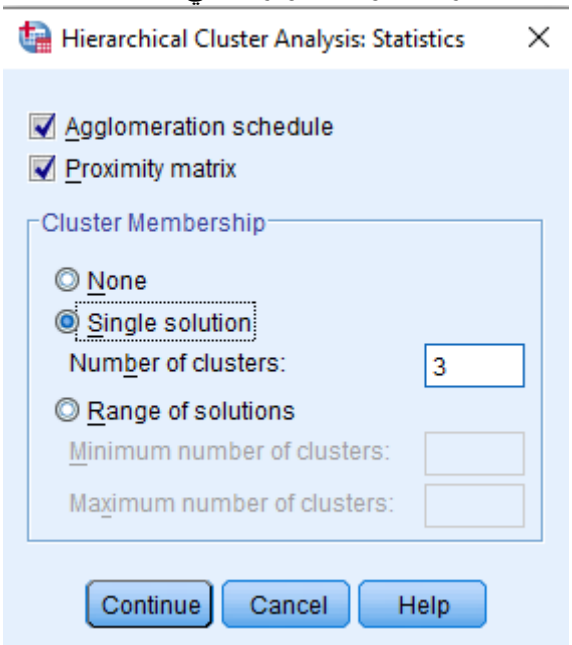
الخطوة (5): طلب عرض الاحصاءات Statistics والرسوم البيانية Plots في المخرجات من خلال التأشير على كل منهما تحت: عرض Display.



يوجد أربعة مفاتيح جانب صندوق الحوار:

المفتاح الأول Statistics:

الخطوة (6): بالضغط على Statistics يظهر صندوق الحوار التالي:

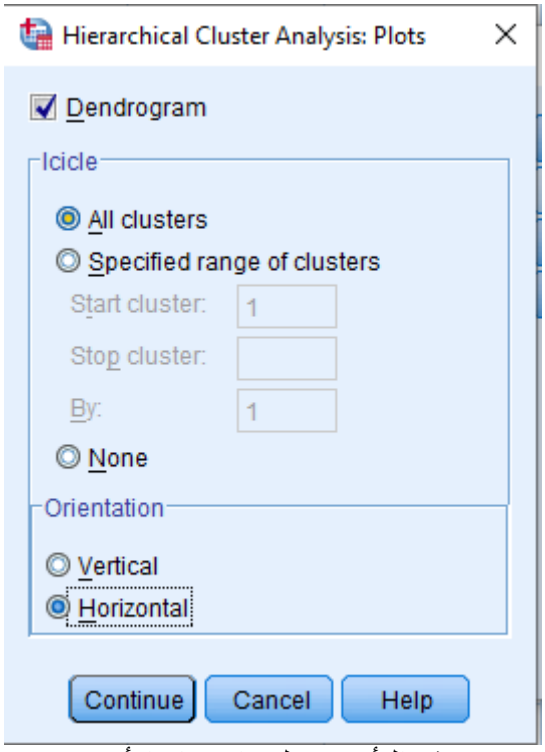


يتم التأشير على المربع الصغير أمام جدول التجميع أو التكتل Agglomeration schedule، والذي يبين المسافات بين الحالات غير المتشابهة، والتأشير على المربع الصغير أمام مصفوفة القرب Proximity Matrix، أما بالنسبة إلى عضوية العنقود Cluster Membership، فيمكن اختيار Single Solution وطباعة الرقم 3 أمام عدد العناقيد المطلوب.

اضغط Continue، لترجع إلى صندوق الحوار الرئيسي.

المفتاح الثاني Plots:

الخطوة (7): بالضغط على Plots يظهر صندوق الحوار التالي:

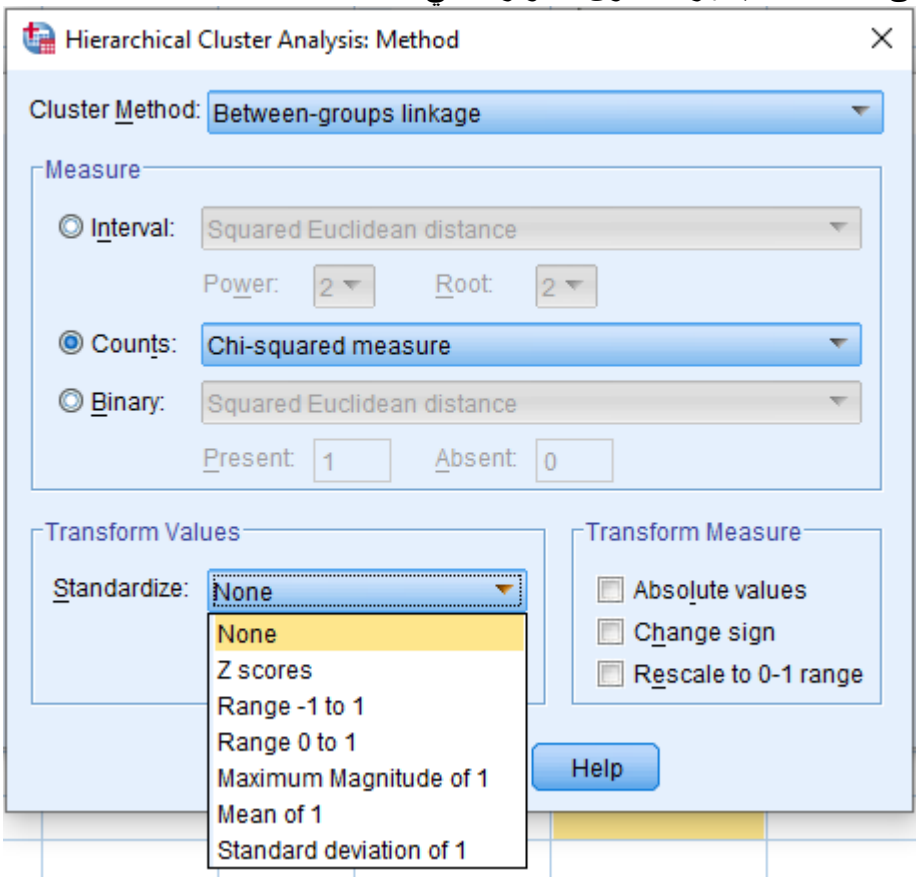


يتم اختيار المخطط الشجري Dendrogram، ثم التأشير على الاتجاه الأفقي Horizontal للمخطط الشجري، أما بالنسبة إلى Icicle (جليد متدلي)، فيتم اختيار كافة العناقيد All Clusters.

يتم الضغط على Continue للعودة إلى صندوق الحوار الرئيس.

المفتاح الثالث Method:

الخطوة (8): بالضغط على Method يظهر صندوق الحوار التالي:



يوفر صندوق حوار Method اختيار طريقة العنقدة، طريقة قياس المسافات، طريقة تحويل القيم وتحويل القياسات:

1) اختيار طريقة التحليل العنقودي Cluster Method:

هناك عدة طرق لدمج العناقيد من أهمها طريقة الربط المتوسط بين المجموعات Between-groups Linkage، حيث يقوم البرنامج بحساب المتوسط الأقل للمسافة بين كافة الأزواج Pairs، ويدمج المجموعتين الأكثر قرباً من

بعضهما، ثم يحسب البرنامج بعد عملية الدمج الأولى المتوسط الأقل للمسافة بين باقي الأزواج، ويقوم بدمج المجموعتين الأكثر قربا.. وهكذا، الخيار موضوع تلقائيا على Between-groups Linkage، يتم الإبقاء على الخيار كما هو؛

2) اختيار طريقة قياس المسافات بين الحالات Measure:

يوجد ثلاثة مستويات لقياس المسافات بين الحالات الممكن إجراء التحيل العنقودي عليها:

- **في حالة بيانات الفترات Interval:** من أمثلتها: درجات الحرارة، التوقيت بالساعة، التاريخ بالسنوات، متوسط مقياس ليكرت، الدخل، الوزن، الطول ...، وهناك عدة قياسات للمسافة لهذا النوع من البيانات أبرزها Euclidean Distance، الذي يساوي الجذر التربيعي لمجموع مربع الفروق بين قيم حالتين أو أكثر على كافة المتغيرات.
- **حالة البيانات التكرارية (العد) Counts:** يشير قياس التكرارات في برنامج SPSS إلى حساب عدد مرات ظهور قيم محددة عبر متغيرات متعددة لكل حالة (صنف)، أو حساب توزيعات التكرار للمتغيرات الفئوية. هناك قياسين للمسافة لهذا النوع من البيانات التي تكون في هيئة تكرارات هما: Chi-square measure, Phi-square measure، يتم اختيار Chi-square measure، والذي يعتمد على Chi-square لاختبار تساوي أو اختلاف المجموعتين.
- **حالة بينات ثنائية Binary:** وهي البيانات التي تأخذ قيمتين فقط، مثل ذكور/إناث، وهناك عدة قياسات على هذا النوع من البيانات، من أهمها اختبار Squared Euclidean Distance.

3) اختيار طريقة تحويل القيم Transform values:

إذا كانت الوحدات المستخدمة لقياس المتغيرات موحدة كما في مثالنا (الدينار)، فيتم الإبقاء على الخيار None، أما إذا كانت هنالك عدة وحدات لقياس المتغيرات، فيمكن استخدام الإحصاءات الموجودة مثل التوزيع المعياري Z، أو Range -1 to 1 أو غيرها، وهذه التحويلات لا تنطبق على المتغيرات/البيانات الاسمية الثنائية.

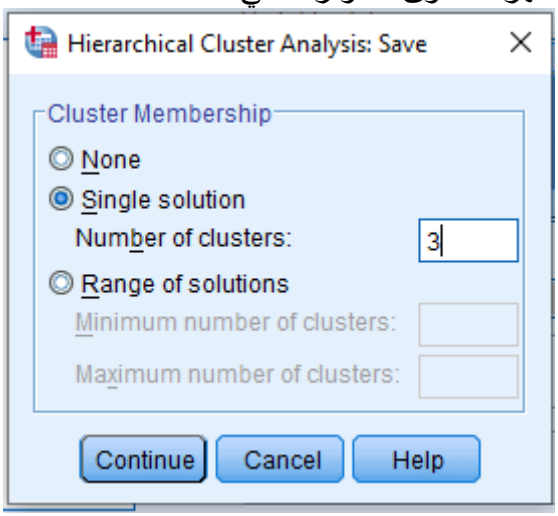
4) تحويل المقاييس Transform measures:

هنالك ثلاث خيارات (مربعات صغيرة) تتيح تحويل البيانات من خلال قياس المسافات إلى: قيم مطلقة Absolute values، تغيير إشارة البيانات Change sign، تحويلها إلى النطاق [0 , 1] Rescale to 0-1 range.

يتم الضغط على Continue للعودة إلى صندوق الحوار الرئيس.

المفتاح الرابع Save:

الخطوة (9): بالضغط على Save يظهر صندوق الحوار التالي:



هناك ثلاثة خيارات تحت عضوية العنقود Cluster Membership:

- عدم تحديد عدد العناقيد None، يصلح في حالة التحليل العنقوديالهرمي، وهو يؤدي إلى عدم حفظ أية متغيرات جديدة (متغير يحدد انتماء الحالات للعناقيد)، وعدم إظهار أي متغير جديد في شاشة تحرير البيانات؛
- الحل الواحد Single Solution، يتم تحديد عدد العناقيد مسبقا حسب المطلوب في السؤال، يلح في حالة التحليل العنقودي غير الهرمي (طريقة k-means مثلا)، وفي مثالنا يتم اختيار هذا الحل وطبع الرقم 3 أمام عدد العناقيد المطلوبة
- مدى من الحلول Range of Solutions، حيث تضع الحد الأدنى والحد الأقصى لعدد العناقيد المطلوبة. يمكن مثلا اختيار مدى من العناقيد بين 3-5.

يتم الضغط على Continue للعودة إلى صندوق الحوار الرئيس.

الخطوة (10): يتم الضغط على OK، فتظهر **المخرجات التالية:**

Cluster

1. ملخص معالجة الحالات Case processing summary

يعكس هذا الجدول عدد ونسب الحالات الموجودة والمفقودة، أو القيم السالبة، وكذلك عدد ونسب إجمالي هذه الحالات.

Case Processing Summary[a]

Cases							
Valid		Rejected				Total	
		Missing Value		Negative Value			
N	Percent	N	Percent	N	Percent	N	Percent
15	100.0	0	.0	0	.0	15	100.0

a. Average Linkage (Between Groups)

يتضح من الجدول السابق أن حجم العينة 15 حالة (منطقة بيعية)، مع غياب القيم المفقودة أو القيم السالبة أو المكررة، وبالتالي فالبيانات لا تحتاج لتنظيف.

2. مصفوفة القرب Proximity Matrix:

يطلق عليها كذلك مصفوفة التباعد/ عدم التشابه Dissimilarity matrix، مصفوفة المسافات distance matrix، وهي تبين أوجه الاختلاف أو أوجه التشابه Similarities بين أزواج من عناصر البيانات، مما يسهل عملية التجميع، كما هو الحال في التجميع الهرمي.

Proximity Matrix

Case	Chi-square between Sets of Frequencies														
	1:A	2:B	3:C	4:D	5:E	6:F	7:G	8:H	9:I	10:J	11:K	12:L	13:M	14:N	15:O
1:A	.000	7.50	10.4	6.58	4.69	6.55	8.75	3.69	6.44	8.27	2.69	7.430	6.409	6.912	8.03
2:B	7.50	.000	12.4	9.25	5.58	9.15	8.84	6.49	4.34	10.6	8.50	4.675	8.691	9.064	11.1
3:C	10.4	12.4	.000	5.38	3.70	9.81	10.6	10.8	10.3	2.73	11.4	12.0	6.176	14.49	5.20
4:D	6.58	9.25	5.382	.000	2.57	5.74	10.7	7.54	6.97	3.33	7.87	10.3	5.813	10.53	4.19
5:E	4.69	5.58	3.703	2.57	.000	4.73	5.40	4.60	3.20	3.20	6.08	4.796	3.322	6.534	4.60
6:F	6.55	9.15	9.812	5.74	4.73	.000	12.8	7.15	7.15	7.39	6.59	12.0	8.748	5.933	6.03
7:G	8.75	8.84	10.6	10.7	5.40	12.8	.000	7.08	8.11	9.89	9.15	5.511	5.205	13.55	11.5
8:H	3.69	6.49	10.8	7.54	4.60	7.15	7.08	.000	4.72	8.50	4.36	6.730	5.369	8.212	8.68
9:I	6.44	4.34	10.3	6.97	3.20	7.15	8.11	4.72	.000	8.35	7.66	6.162	6.549	8.831	9.20
10:J	8.27	10.6	2.726	3.33	3.20	7.39	9.89	8.50	8.35	.000	9.15	10.9	4.946	12.13	3.14
11:K	2.69	8.50	11.4	7.87	6.08	6.59	9.15	4.36	7.66	9.15	.000	8.586	7.485	5.732	8.34
12:L	7.43	4.68	12.0	10.3	4.80	12.0	5.51	6.73	6.16	10.9	8.59	.000	7.166	11.72	12.2
13:M	6.41	8.69	6.176	5.81	3.32	8.75	5.20	5.37	6.55	4.95	7.48	7.166	.000	11.62	6.75
14:N	6.91	9.06	14.5	10.5	6.53	5.93	13.6	8.21	8.83	12.1	5.73	11.7	11.62	.000	10.8
15:O	8.03	11.1	5.195	4.19	4.60	6.03	11.5	8.68	9.20	3.14	8.34	12.2	6.753	10.79	.000

This is a dissimilarity matrix

تقرر مصفوفة القرب التشابه أو عدم التشابه بين الحالات (المناطق البيعية)، ويعبر عنها بالمسافات المشتقة من خلال المتغيرات المعطاة، فالمسافة بين المنطقتين A وB مثلاً تساوي 7.5، والمسافة بين A وC تساوي 10.4 وهكذا...

3. جدول التجميع Agglomeration schedule

يوضح جدول التجميع في التجميع الهرمي عملية دمج الحالات خطوة بخطوة، مبينًا المجموعات التي يتم دمجها والمسافات بينها (أوجه التشابه). يُستخدم هذا الجدول لتحديد العدد الأمثل للمجموعات من خلال تحديد المرحلة التي يحدث فيها ارتفاع كبير في المعامل (المسافة)، مما يشير إلى دمج مجموعات غير متشابهة.

الربط بين المجموعات (أو الربط المتوسط) هو أسلوب تجميع هرمي يدمج المجموعات بناءً على متوسط المسافة بين جميع أزواج الملاحظات في مجموعتين مختلفتين. يُنشئ هذا الأسلوب بنية شجرية متوازنة (مخطط شجري dendrogram) من خلال مراعاة قرب جميع العناصر بدلاً من التركيز على أقربها (الربط الأحادي) أو أبعدها (الربط الكامل). يتميز هذا الأسلوب بحساسية أقل للقيم الشاذة مقارنةً بالربط الأحادي أو الكامل، وهو فعال في اكتشاف مجموعات متميزة وكثيفة.

الجدول التالي يوضح العناقيد المتكونة وفق طريقة الربط المتوسط (بين المجموعات): **Average Linkage (Between Groups):**

Average Linkage (Between Groups)

Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	4	5	2.566	0	0	4
2	1	11	2.694	0	0	5
3	3	10	2.726	0	0	4
4	3	4	3.903	3	1	6
5	1	8	4.023	2	0	11
6	3	15	4.279	4	0	13
7	2	9	4.336	0	0	9
8	7	13	5.205	0	0	12
9	2	12	5.419	7	0	12
10	6	14	5.933	0	0	11
11	1	6	6.858	5	10	13
12	2	7	7.477	9	8	14
13	1	3	8.361	11	6	14
14	1	2	8.503	13	12	0

وفقا لطريقة الربط بين المجموعات/ الربط المتوسط، يبين جدول التجميع/التكتل Agglomeration Schedule كيفية تكوين العناقيد عند كل مرحلة من مراحل التحليل، وبالنظر إلى الجدول نجد أن المرحلة الأولى (من قاعدة الشجرة) كانت بين الحالة رقم 4 ورقم 5 (أي بين المنطقة D,E) فتم اندماجهم تحت العنقود (D)، وإذا تابعنا العنقود (D) نجد أنه في المرحلة الرابعة تم اندماج العنقود (D)، والعنقود (C) المكون من الحالتين (C,J) لينتج عنهما العنقود (C)، وهكذا يمكن أن نحلل تلك المراحل أولاً بأول ومتابعة ذلك مع سير الشجرة الثنائية.

إن الهدف الأساسي من جدول التقارب هو المساعدة في تحديد عدد العناقيد. فإذا نظرنا إلى عمود المعاملات/المسافات Coefficients في الجدول نجد أن المعاملات/المسافات تتزايد بمعدلات قليلة ما بين المرحلتين الأولى والثانية وبين المرحلتين الثانية والثالثة، أما المعاملات ما بين المرحلتين الثالثة والرابعة فهناك قفزة كبيرة في معدلها، حيث ازدادت بصورة قفزة Jump كبيرة من 2.726 إلى 3.903، مما يشير إلى ضرورة وجود عنقودين في المثال، وإذا أمعنت النظر في قراءة المعاملات نجد هناك قفزة كبيرة ثانية ما بين المرحلتين العاشرة والحادية عشرة، مما يؤكد على ضرورة وجود عنقود ثالث في هذا المثال.

4. جدول انتماء الحالات إلى العناقيد Cluster Membership:

يُحدد العنقود/ المجموعة التي تنتمي إليها كل حالة/ مفردة، ويُنشئ هذا الجدول متغيرًا جديدًا، يُسمى غالبًا QCL_1 أو CLU_1، في نافذة عرض البيانات، يُظهر رقم العنقود/ المجموعة المُخصصة لكل حالة.

Cluster Membership

Case	3 Clusters
1:A	1
2:B	2
3:C	3
4:D	3
5:E	3
6:F	1
7:G	2
8:H	1
9:I	2
10:J	3
11:K	1
12:L	2
13:M	2
14:N	1
15:O	3

يوضح جدول انتماء الحالات إلى العناقيد، العنقود الذي تنتمي إليه كل حالة (منطقة) من الحالات (من A إلى O)، فمثلا الحالة 1 (A) تنتمي إلى عنقود 1، والحالة 2 (B) تنتمي إلى عنقود 2 وهكذا ...

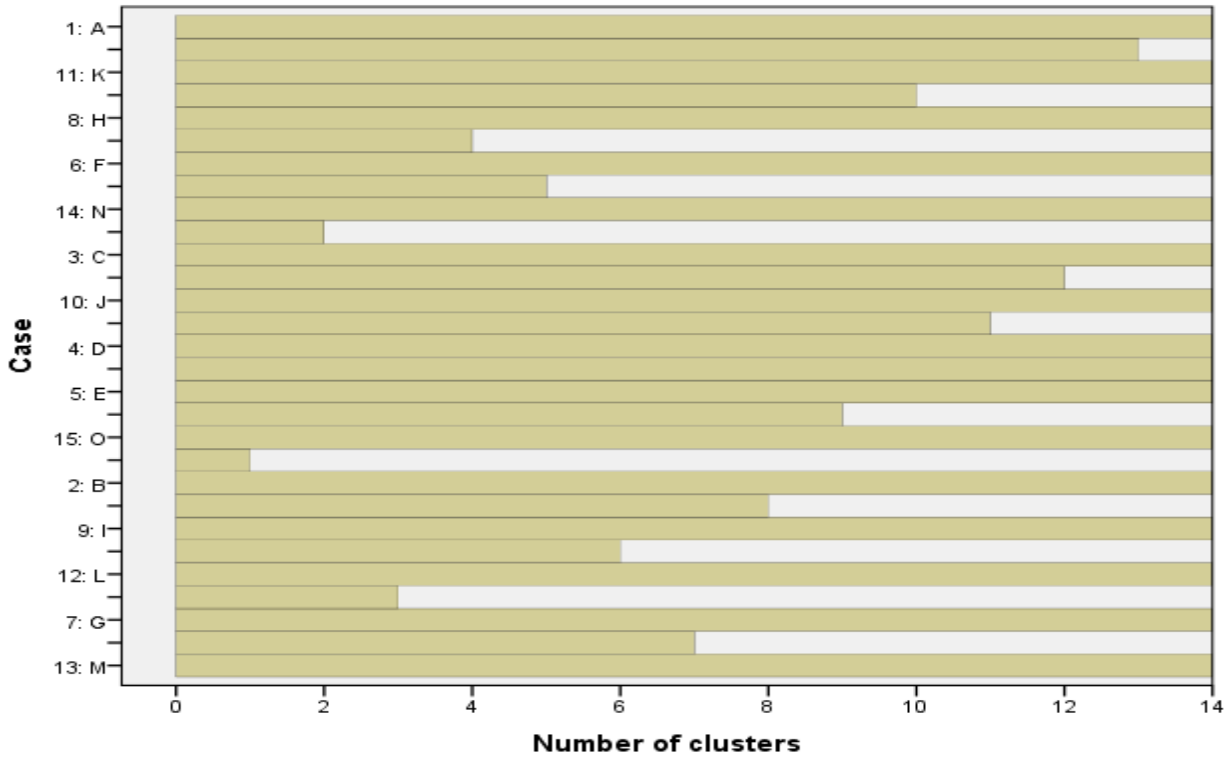
5. المخطط الجليدي Icicle Plots

تُعدّ كلٌّ من مخططات الجليد Icicle plots ومخططات التفرع الشجري dendrograms أدوات تصويرية للتجميع الهرمي، حيث تُظهر مخططات الجليد Icicle plots اندماج المجموعات من خلال أشرطة متراصة ومتدلية (تشبه الجليد)، بينما تستخدم مخططات التفرع الشجري dendrogram بنيةً شجرية متفرعة. في حين تُظهر مخططات التفرع المسافة/التشابه بوضوح من خلال أطوال الفروع، تتفوق مخططات الجليد في تمثيل انتماء المجموعة وحجمها في كل تكرار.

في بعض طبعات SPSS، يظهر المخطط الجليدي في شكل أعمدة أفقية أو عمودية يطلق عليها **الألواح الجليدية**، والذي يمثل نقاط التقاء الحالات Cases مع عدد العناقيد Number of clusters عند كل مرحلة من مراحل التحليل، وبالتالي فهو يوفر معلومات عن كيفية دمج الحالات في عناقيد في كل تكرار من التحليل.

وقد اطلق على هذا المخطط مسمى **الألواح الجليدية diagram Icicle** بسبب تشابهه مع الواح الجليد النازلة من اسطح المنازل عند ذوبانها. تمثل الأعمدة المفردات التي يراد تقسيمها، وتمثل الصفوف عدد القطاعات التي تم الحصول عليها. ويقرأ هذا الشكل من القاع إلى القمة.

المخطط الجليدي الأفقي Horizontal Icicle



يتضح من المخطط الجليدي Icicle Plots، أنه يتم تمثيل كل حالة (منطقة بيعية) بلوح أفقي بلون رمادي، يمتدان من 0 إلى نهاية المخطط (المرحلة 14 الأخيرة من العنقدة)، واللوح الذي يقع بينهما يمثل العنقود الذي يتشكل منهما، وهو ما يعني أن الحالات يتم عنقدتها/تجميعها معا، تظهر بجوار بعضها في المخطط الجليدي.

كما نلاحظ من المخطط السابق أنه يتم عنقدة المنطقتينD: 4 وE: 5 في المرحلة الأولى من العنقدة، لأن اللوح الذي يقع بينهما هو الأطول من بين كل الألواح الفاصلة بين مختلف الحالات/المناطق، ثم يظهر بعده تجميع/عنقدة المنطقتين A: 1 و K: 11، حيث طول اللوح بينهما أقصر من السابق. أما أقصر لوح أفقي فيقع بين الحالتين/المنطقتين B: 2 وO: 15، مما يدل على أنه تم تجميعهما/عنقدتهما في أخر مرحلة.

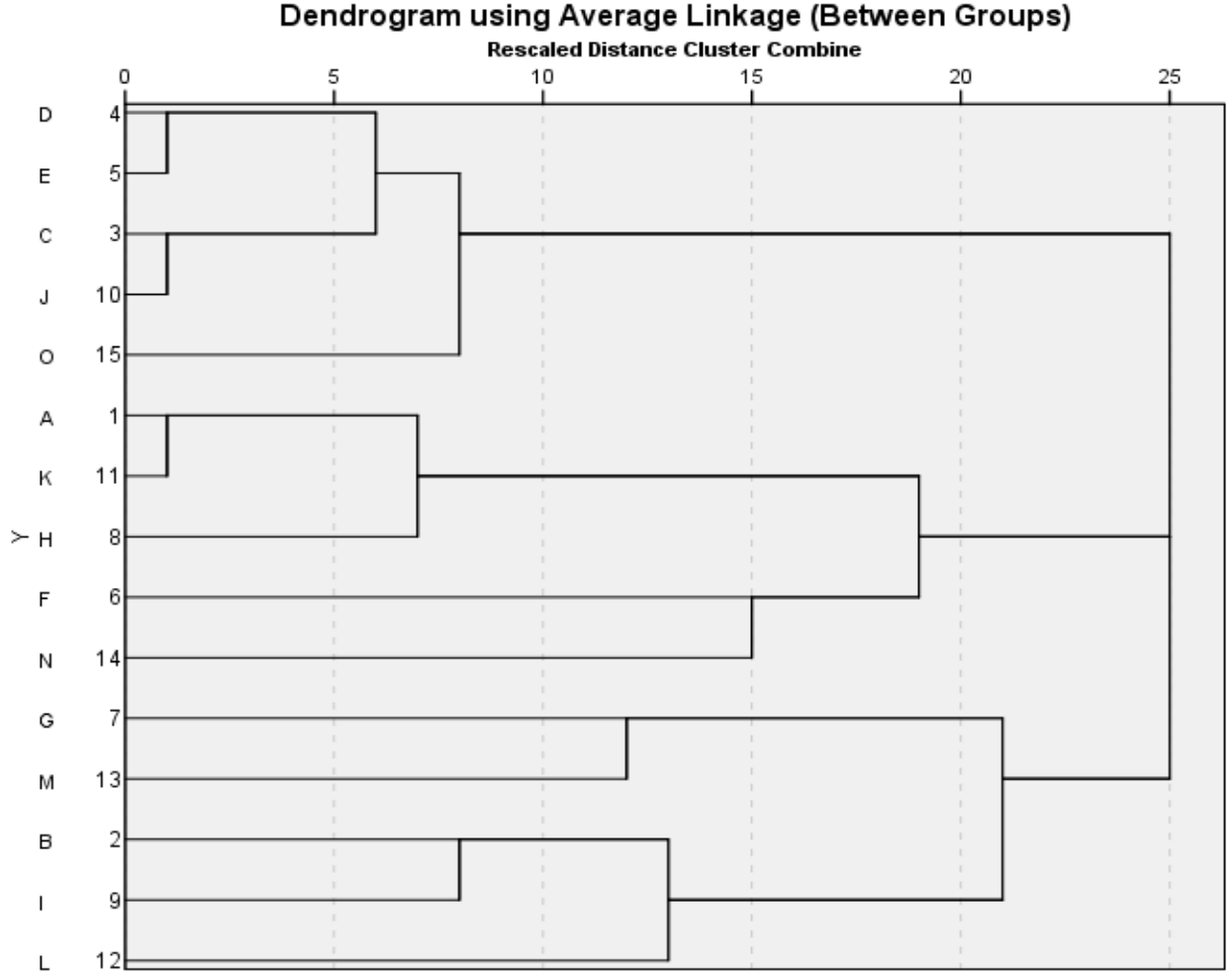
يقوم برنامج SPSS بتمثيل المخطط الجليدي وفق جدول، حيث الحالات على المحور العمودي، والعناقيد التي تنتمي إليها الحالات/المفردات على المحور الأفقي، وبدل رسم ألواح أفقية أو عمودية ملونة، يتم رسم صفوف أو أعمدة تحتوي x. ويبقى التفسير نفسه.

Horizontal Icicle

Case	Number of clusters													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
13:M	×	×	×	×	×	×	×	×	×	×	×	×	×	×
	×	×	×	×	×	×	×							
7:G	×	×	×	×	×	×	×	×	×	×	×	×	×	×
	×	×	×											
12:L	×	×	×	×	×	×	×	×	×	×	×	×	×	×
	×	×	×	×	×	×								
9:I	×	×	×	×	×	×	×	×	×	×	×	×	×	×
	×	×	×	×	×	×	×	×						
2:B	×	×	×	×	×	×	×	×	×	×	×	×	×	×
	×													
15:O	×	×	×	×	×	×	×	×	×	×	×	×	×	×
	×	×	×	×	×	×	×	×	×					
5:E	×	×	×	×	×	×	×	×	×	×	×	×	×	×
	×	×	×	×	×	×	×	×	×	×	×	×	×	×
4:D	×	×	×	×	×	×	×	×	×	×	×	×	×	×
	×	×	×	×	×	×	×	×	×	×	×			
10:J	×	×	×	×	×	×	×	×	×	×	×	×	×	×
	×	×	×	×	×	×	×	×	×	×	×	×		
3:C	×	×	×	×	×	×	×	×	×	×	×	×	×	×
	×	×												
14:N	×	×	×	×	×	×	×	×	×	×	×	×	×	×
	×	×	×	×	×									
6:F	×	×	×	×	×	×	×	×	×	×	×	×	×	×
	×	×	×	×										
8:H	×	×	×	×	×	×	×	×	×	×	×	×	×	×
	×	×	×	×	×	×	×	×	×	×				
11:K	×	×	×	×	×	×	×	×	×	×	×	×	×	×
	×	×	×	×	×	×	×	×	×	×	×	×	×	
1:A	×	×	×	×	×	×	×	×	×	×	×	×	×	×

6. المخطط الشجري Dendogram

مخطط التفرع الشجري هو رسم بياني يُستخدم لتصوير العلاقات الهرمية، ودمج، وتقسيم العناقيد/المجموعات في البيانات، قد يرسم بشكل عمودي أو أفقي، في حالة المخطط الأفقي تمثل الارتفاعات المسافة أو الاختلاف بين العناقيد. ويُستخدم بشكل أساسي في التجميع الهرمي (التجميعي)، حيث يعرض الحالات/ المفردات في الأسفل أو على اليمين ويُجمّعها، مما يسمح بتحديد العدد الأمثل للعناقيد عن طريق قطع الشجرة عند ارتفاعات محددة.



على المحور الأفقي من المخطط الشجري Dendogram، نجد الحالات (المناطق البيعية من A إلى O)، وعلى المحور الأفقي نجد قياسات المسافات التي تمتد إلى 25 وحدة قياس، حيث يشير الطول الكبير للخط إلى زيادة درجات عدم التشابه، وبالتالي فزيادة طول الخط بين متغيرين تشير إلى عدم التشابه بين بياناتهما، كما أن هناك عدة عقدات Nodes موجودة في الشجرة، حيث تمثل كل عقدة هدف معين يعكس اندماج حالتين أو أكثر.

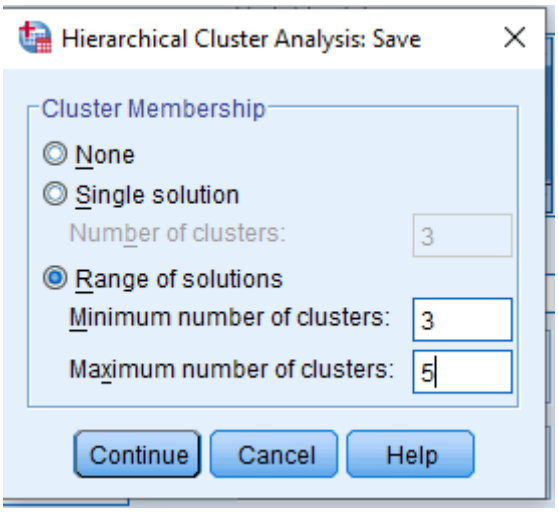
كما نلاحظ في أعلى الشجرة أن عملية العنقدة كانت بين الحالتين D وE في البداية، حيث المسافة بينهما هي أصغر مسافة، وفي أسفل الشجرة كانت بين B و I، ثم B وI مع L، وتستمر العملية هكذا حتى نصل إلى عنقودين فقط.

الخطوة (11): بعد إقفال شاشة المخرجات وبالرجوع إلى شاشة تحرير البيانات، تجد أنه قد تم إضافة متغيرا جديدا باسم CLU3_1:

	Territory	Housing	Food	Tel	Car	Others	CLU3_1
1	A	200.00	100.00	60.00	110.00	80.00	1
2	B	150.00	90.00	40.00	10.00	60.00	2
3	C	400.00	250.00	30.00	200.00	30.00	3
4	D	420.00	210.00	40.00	180.00	80.00	3
5	E	80.00	40.00	10.00	20.00	10.00	3
6	F	390.00	220.00	30.00	190.00	170.00	1
7	G	140.00	170.00	80.00	60.00	20.00	2
8	H	250.00	200.00	80.00	120.00	110.00	1
9	I	270.00	180.00	50.00	60.00	90.00	2
10	J	350.00	230.00	30.00	180.00	50.00	3
11	K	110.00	70.00	40.00	90.00	70.00	1
12	L	140.00	90.00	70.00	20.00	30.00	2
13	M	240.00	200.00	60.00	120.00	40.00	2
14	N	240.00	120.00	30.00	120.00	200.00	1
15	O	320.00	210.00	20.00	210.00	70.00	3

يوضح المتغير الجيد أعلاه وضع كل حالة أو منطقة بيعية فيما يتعلق بانتسابها إلى عضوية العنقود 1 أو 2 أو 3.

الخطوة (12): أما الآن فسوف نستعرض نتائج اختيارالحالة الثالثة من صندوق الخيار والمتعلقة بمدى من الحلول، لنفترض أنه بعد الضغط على Save، قمنا باختيار مدى الحلول Range of solutions، ووضعنا الرقم 3 كحد أدنى من العناقيد المطلوبة، والرقم 5 كحد أعلى لذلك، كما يلي:



بعد الضغط على Continue في صندوق الحوار الفرعي Save والضغط على OK في صندوق الحوار الرئيس Hierarchical Cluster Analysis، يظهر متغيرات ثلاثة جديدة في شاشة تحرير البيانات كالآتي:

Territory	Housing	Food	Tel	Car	Others	CLU5_1	CLU4_1	CLU3_1
A	200.00	100.00	60.00	110.00	80.00	1	1	1
B	150.00	90.00	40.00	10.00	60.00	2	2	2
C	400.00	250.00	30.00	200.00	30.00	3	3	3
D	420.00	210.00	40.00	180.00	80.00	3	3	3
E	80.00	40.00	10.00	20.00	10.00	3	3	3
F	390.00	220.00	30.00	190.00	170.00	4	1	1
G	140.00	170.00	80.00	60.00	20.00	5	4	2
H	250.00	200.00	80.00	120.00	110.00	1	1	1
I	270.00	180.00	50.00	60.00	90.00	2	2	2
J	350.00	230.00	30.00	180.00	50.00	3	3	3
K	110.00	70.00	40.00	90.00	70.00	1	1	1
L	140.00	90.00	70.00	20.00	30.00	2	2	2
M	240.00	200.00	60.00	120.00	40.00	5	4	2
N	240.00	120.00	30.00	120.00	200.00	4	1	1
O	320.00	210.00	20.00	210.00	70.00	3	3	3

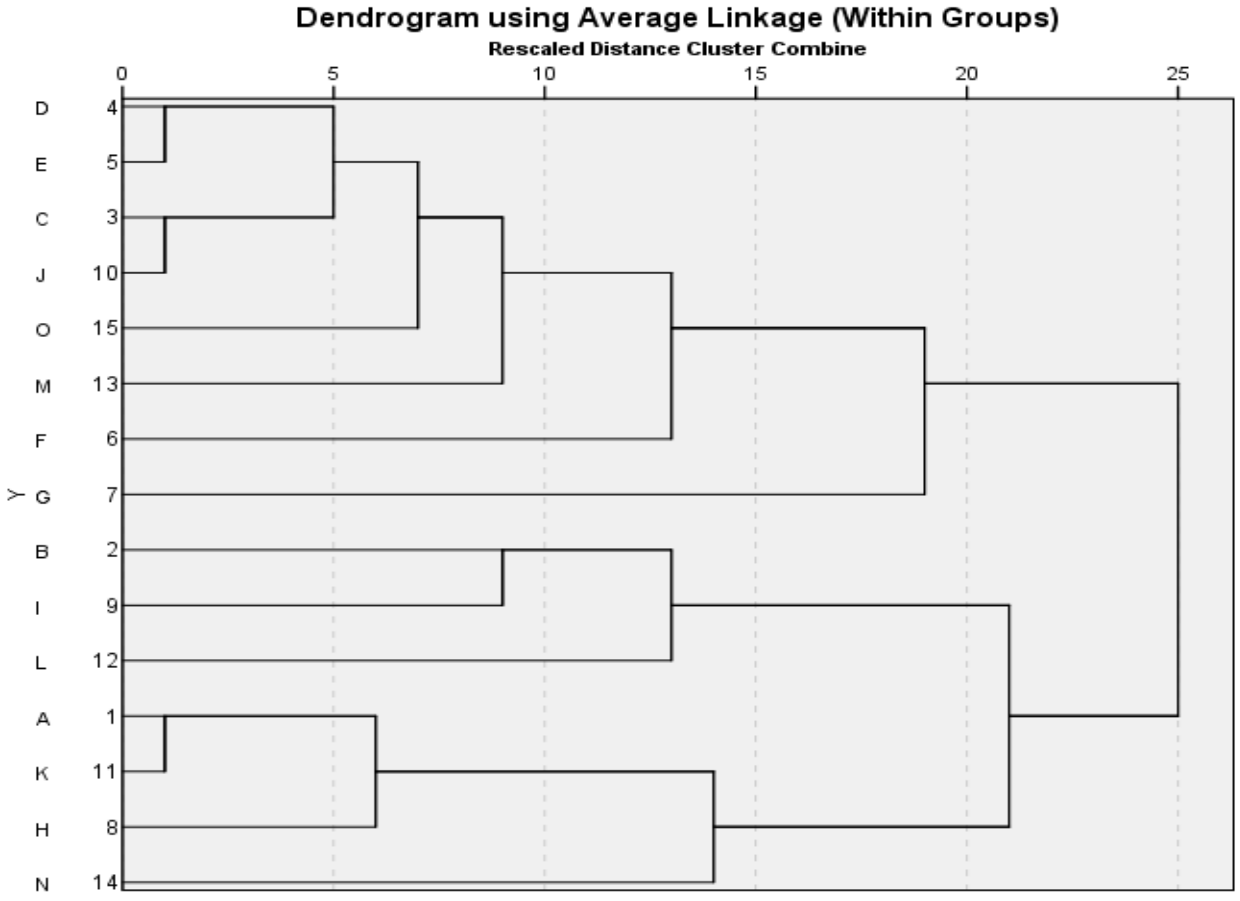
يبين المتغيرالجديد CLU3_1 تصنيف الحالات على أساس 3 عناقيد وعضوية كل حالة في هذه العناقيد، كما يبين المتغير CLU4_1 تصنيف الحالات على أساس 4 عناقيد، وعضوية كل حالة في هذه العناقيد، أما المتغير الجديد CLU5_1 فيبين تصنيف الحالات على أساس 5 عناقيد وكذلك عضوية كل حالة في هذه العناقيد.

الخطوة (13): بعد أن تم استعراض شكل المخطط الشجري وفقا لطريقة الربط المتوسط أو الربط بين المجموعات Between-groups Linkage، نعرض فيما يلي شكلين آخرين لنفس المعلومات الواردة في المثال، وذلك لأجل المقارنة بين أشكال المخطط الشجري تبعاً لطرق أخرى للربط المتبعة في التحليل العنقودي.

1. العنقدة وفق طريقة الربط داخل المجموعات Within-groups Linkage

الربط داخل المجموعة (أو الربط داخل العناقيد) هو تقنية تجميع هرمية تدمج عنقودين لتقليل متوسط المسافة بين كل زوج من العناصر (أو زيادة التشابه) داخل العنقود الجديد الناتج. على عكس طرق الربط بين المجموعات، يركز هذا الأسلوب فقط على تماسك المجموعة المدمجة بدلاً من المسافة بين العناقيد.

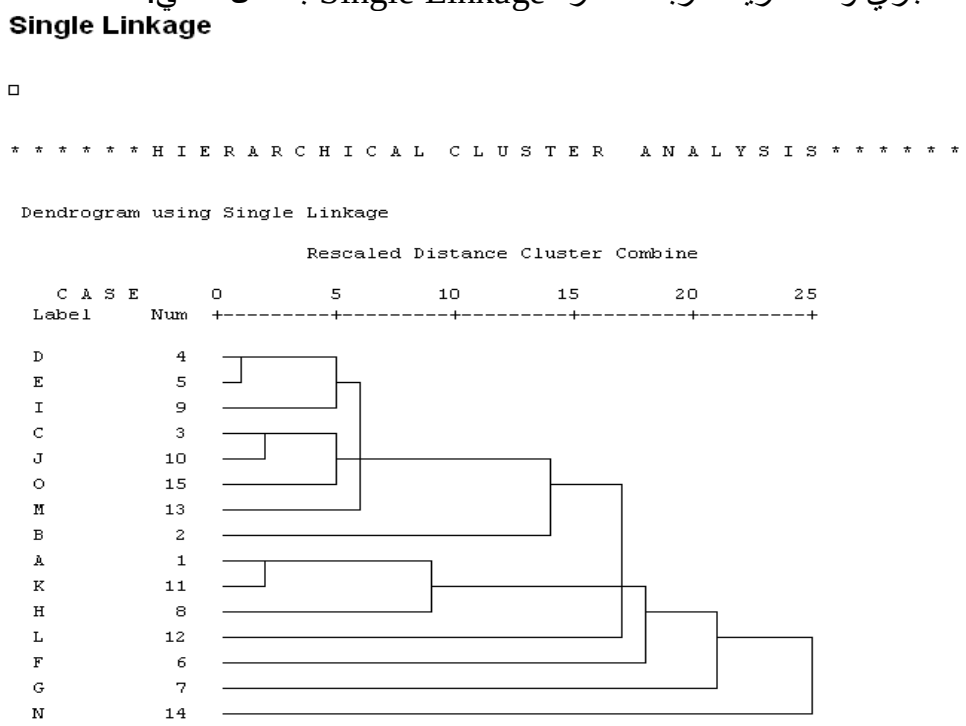
ويبدو المخطط الشجري Dendrogram وفقا لطريقة الربط داخل المجموعات Within-groups Linkage بالشكل التالي:



نلاحظ في أعلى الشجرة أن بداية عملية العنقدة كانت بين الحالتين D وE، وفي أسفل الشجرة كانت بين A وK، ثم AK مع H، ثم AKH مع N، وتستمر العملية هكذا حتى نصل إلى عنقودين فقط.

2. العنقدة وفق طريقة الربط المنفرد Single Linkage (أو الجوار الأقرب Nearest neighbor)

يبدو المخطط الشجري وفقا لطريقة الربط المنفرد Single Linkage بالشكل التالي:

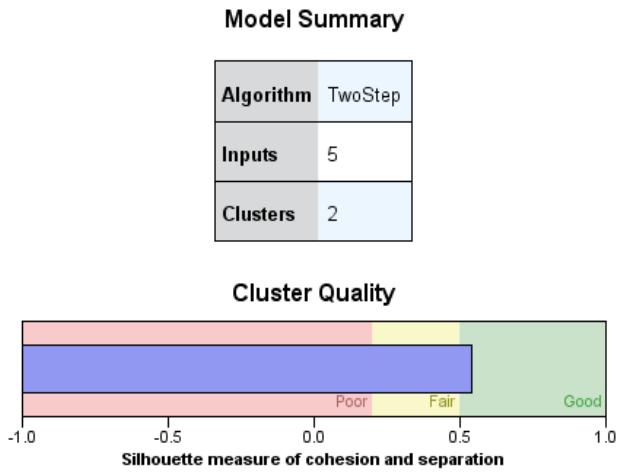


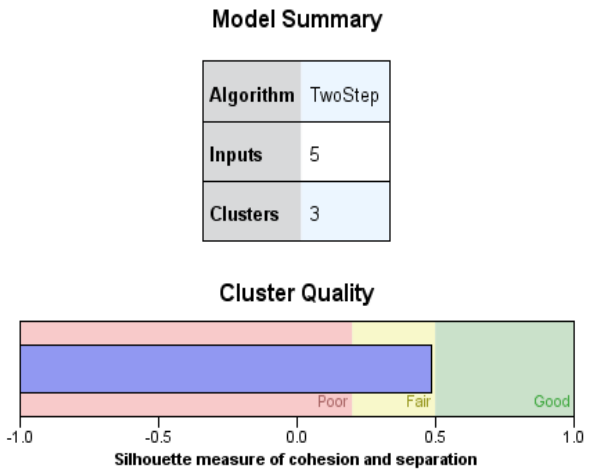
هنا تختلف الشجرة عن الحالة السابقة، فنجد في أعلى المخطط الشجري، أن عملية العنقدة كانت قد بدأت أيضاً بين الحالتين D وE، بينما في أسفل المخطط الشجري كانت بين مجموعة أخرى من الحالات كما يبين الشكل.

الخطوة (14): اختبار العدد الأمثل من العناقيد/ المجموعات

1. يتم الوصول إلى **العدد الأمثل من العناقيد** عن طريق عدة طرق، من أبرزها **مقياس ومخطط الظلية** في برنامج SPSS، وهذا بالاستعانة بطريقة التجميع على مرحلتين TwoStep Cluster، وذلك بتشغيل: Analyze> Classify> TwoStep Cluster.
2. يظهر صندوق حوار **TwoStep Cluster**، ننقل المتغيرات الخمسة إلى مربع Continuous Variables، لأنها كلها كمية مستمرة، وبما أنه لا توجد في المتغيرات فئوية، فإن المربع Categorical Variables يبقى فارغا.
3. في **قياس المسافة** نختار المسافة الإقليدية Ecledian، وفي عدد العناقيد Number ofClusters، نؤشر على Specify fixed، ونكتب رقم 2 (أي عنقودين) في المرحلة الأولى، ثم 3 في المرحلة الثانية، وهكذا 4، 5 ...، وفي كل مرة نقوم بتقييم جودة التجميع وفق كل عدد من العناقيد.
4. بالنسبة **لمفتاح Options**، يمكن أن نحول المتغيرات المستمرة إلى معيارية فنتركها في مربع To be standardized، أو نفترض أنها معيارية، هنا ننقلها إلى مربع Assumed Standardized.
5. بالنسبة **لمفتاح Output**، نؤشر فقط على رسوم وجداول في نموذج العارض Charts and Tables in model viewer.
6. بعد هذه العمليات نضغط على ok، فتظهر مخرجات التجميع على مرحلتين TwoStep Cluster من أجل k عنقود.
7. في المخرجات، ننقر نقرًا مزدوجًا على الرسم البياني "ملخص النموذج Model Summary" لفتح عارض التجميع Cluster Viewer.
8. تعرض لوحة "**جودة التجميع Cluster Quality**" مقياس الظل للتماسك والانفصال Silhouette measure of cohesion and separation. حيث تشير الدرجات الأكبر من 0.5 إلى فصل جيد بين المجموعات/العناقيد، بينما تشير الدرجات الأقل من 0.2 إلى بنية عنقودية ضعيفة.

من خلال اتباع الخطوات السابقة نحصل على مخطط لمقياس الظل للتماسك والانفصال Silhouette measure of cohesion and separation، وذلك تبعا لقيم مختلفة لعدد العناقيد k، يمثل الرسمان التاليان هذا المخطط في حالة (k=2) عنقودين و(k=3) ثلاث عناقيد:





نلاحظ من الرسم في الأعلى أنه من أجل عدد عناقيد k= مرتفعة (درجة الظلية أكبر من 0.5 في المخطط على اليسار: مستوى Good)، لكنها انخفضت قليلا إلى ما دون 0.5 بمجرد إضافة عنقود ثالث (k=3) في المخطط على اليمين: مستوى وسط Fair.

وبمقارنة جدولي أعضاء العناقيد في حالة عنقودين (k=2)، وثلاث عناقيد (k=3)، نجد أن عناصر العنقود الثاني بقيت بدون تغيير في في الحالتين k=1 وk=3، وهي المناطق: B؛ G؛ I؛ L وm، أما العنقود الأول في حالة k=2، فقد تم تقسيمه إلى عنقودين 1 و3، حيث صار يضم العنقود 1، فقط المناطق: A؛ F؛ H؛ K؛ وN، في حين شكلت المناطق: C؛ D؛ E؛ J وO العنقود 3.

في حالة ثلاث عناقيد (k=2)، وعند تفحص المخطط الشجري Dendogram، الذي أعددناه في الخطوة(10)، - 6، نجد أن هناك قفزة كبيرة في المسافة عند الانتقال من العنقود (1) إلى العنقود (2)، لكن في ثلاث عناقيد (k=3)، نجد أن هناك مسافة بسيطة عند الانتقال من العنقود (1) إلى العنقود (3)، وبالتالي فإضافة عنقود ثالث لا يضيف تباينا كبيرا بين العناصر. مما يؤكد أمثلية عنقودين (k=2).

Cluster Membership

Case	2 Clusters
1:A	1
2:B	2
3:C	1
4:D	1
5:E	1
6:F	1
7:G	2
8:H	1
9:I	2
10:J	1
11:K	1
12:L	2
13:M	2
14:N	1
15:O	1

Cluster Membership

Case	3 Clusters
1:A	1
2:B	2
3:C	3
4:D	3
5:E	3
6:F	1
7:G	2
8:H	1
9:I	2
10:J	3
11:K	1
12:L	2
13:M	2
14:N	1
15:O	3