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

السنة الجامعية: 2025/2024
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:

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.

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.

التحليل العنقودي الهرمي:

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 نقاط التقاء المتغيرات، لكن أكثرها شيوعا الطريقة الإقليدية Euclidean Distances، والتي تعتبر من أهم مقاييس التباعد في هذا المجال.

طريقة ك- متوسطات K-Means:

تقوم هذه الطريقة 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؛
4. تحديد متوسط قيم العناقيد (Centroids) بشكل مبدئي؛
5. حساب المسافات بين كل مفردة ومراكز المتوسطات؛
6. نقل المفردة إلى العنقود الذي يكون متوسطه أقرب إليها (استخدام مقياس Euclidean Distances)؛
7. إعادة حساب متوسطات قيم العناقيد الجديدة؛
8. تكرار نفس الخطوات (6-3) 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.

الحل:

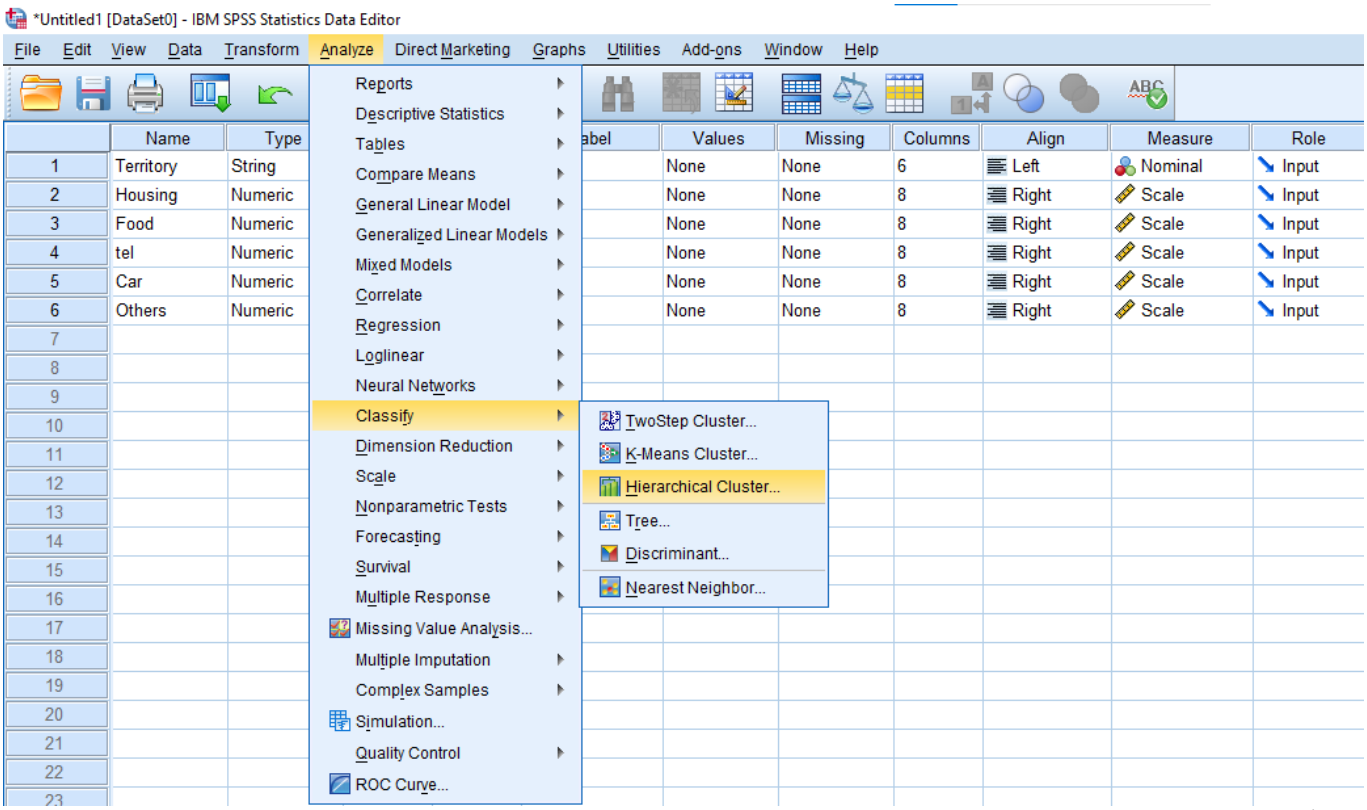
الخطوة (1): إدخال البيانات الواردة في المثال (انطلاقا من ملف Exel الأول المرفق) إلى نافذة البيانات SPSS (Data View)، ثم في نافذة المتغيرات Variable View، يتم تسمية 6 متغيرات: المنطقة Territory (وهو متغير نوعه: String القياس: اسمي Nominal)، السكن Housing، الغذاء Food، الاتصالات Tel، السيارات Car، أخرى 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

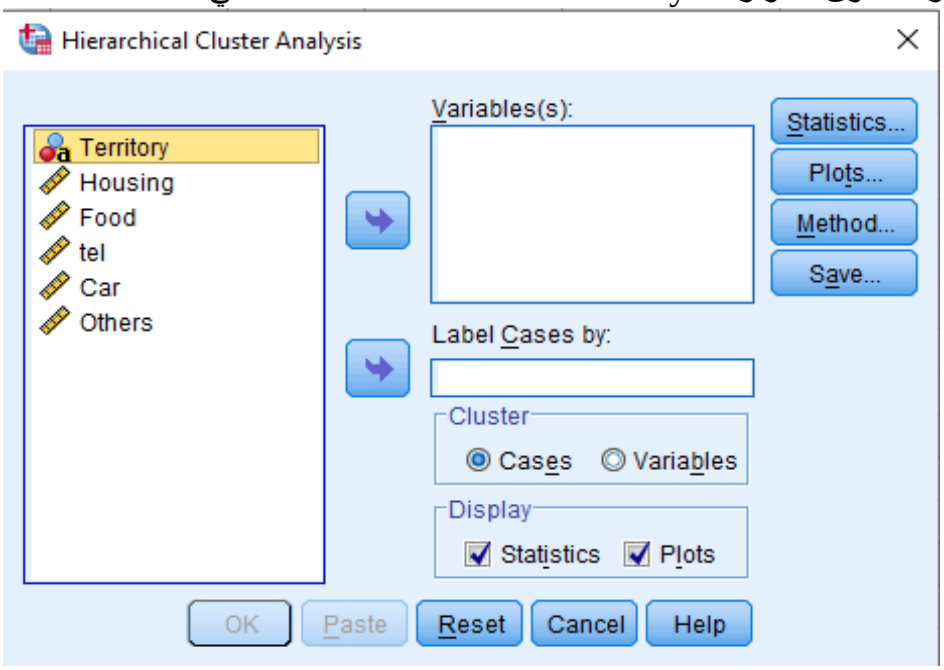
أما في نافذة المتغيرات Variable View فتظهر كما يلي:

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؛



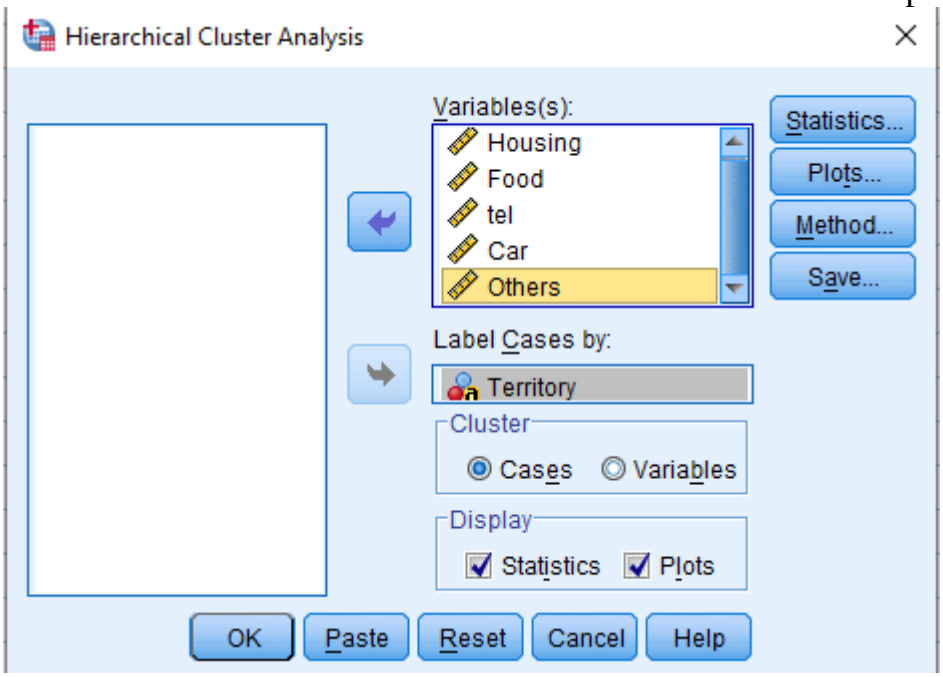
الخطوة (3): يظهر صندوق الحوار Hierarchical Cluster Analysis التالي:



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؛

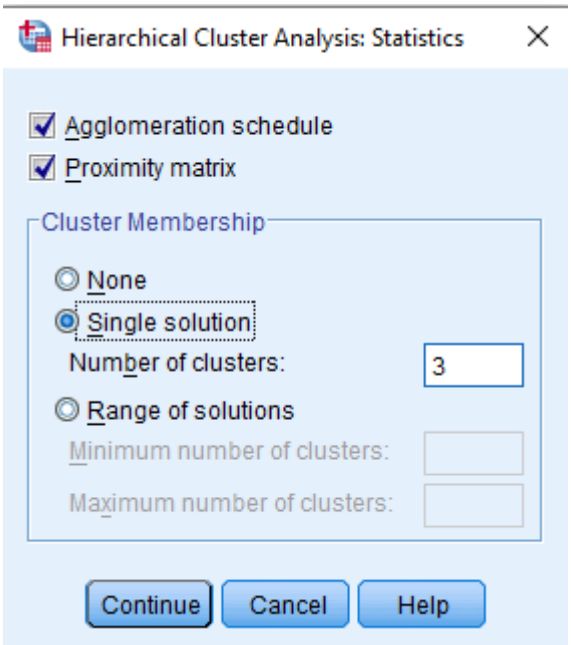
الخطوة (5): طلب عرض الاحصاءات Statistics والرسوم البيانية Plots في المخرجات من خلال التأشير على كل منهما تحت Display.



يوجد أربعة مفاتيح أسفل صندوق الحوار:

المفتاح الأول Statistics:

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

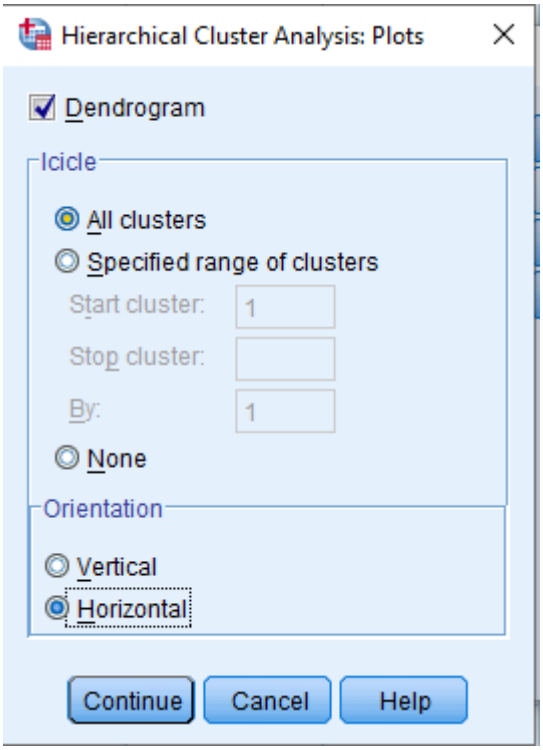


يتم التأشير على المربع الصغير أمام جدول Agglomeration، والذي يبين المسافات بين النقاط غير المتشابهة، والتأشير على المربع الصغير أمام مصفوفة القرب Proximity Matrix، أما بالنسبة إلى عضوية العنقود فيمكن اختيار Single Solution وطباعة الرقم 3 أمام عدد العناقيد المطلوب.

اضغط Continue، لترجع إلى صندوق الحوار الرئيسي.

المفتاح الثاني Plots:

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

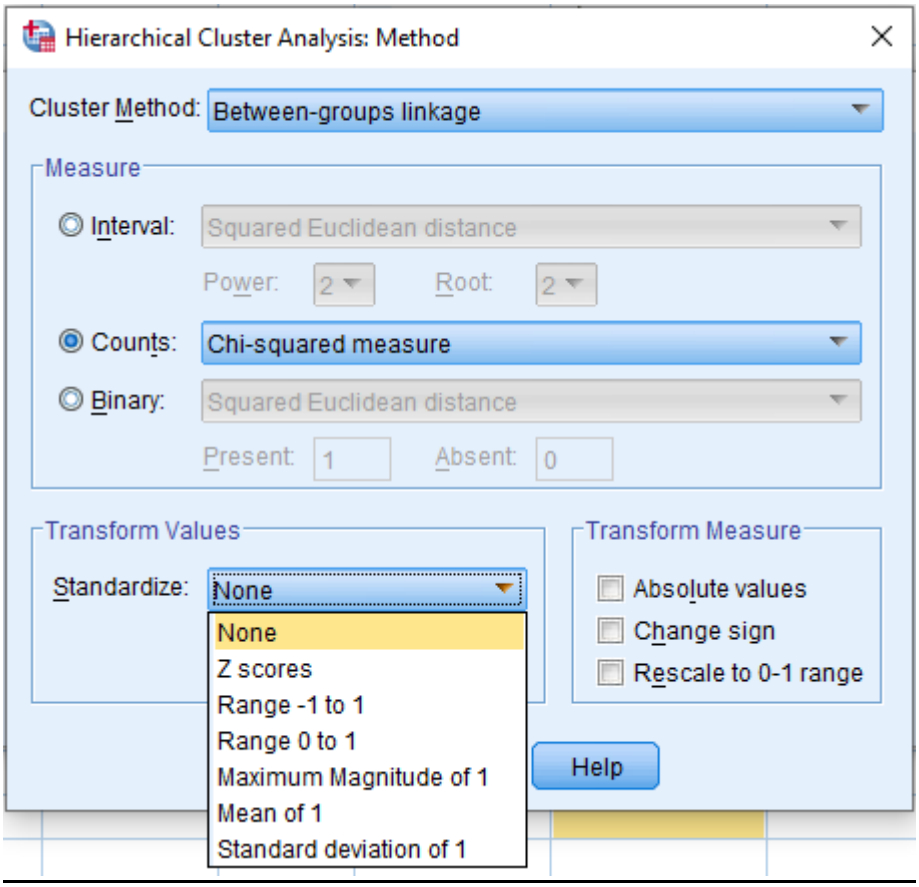


يتم اختيار الشجرة الثنائية Dendrogram، ثم التأشير على الاتجاه الأفقي Horizontal للشجرة، أما بالنسبة إلى Icicle فيتم اختيار كافة العناقيد All Clusters.

يتم الضغط على Continue للعودة إلى صندوق الحوار الرئيس.

المفتاح الثالث Method:

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



طريقة التحليل العنقودي Cluster Method: هناك عدة طرق لدمج العناقيد من أهمها 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؛

المقاييس Measure: يوجد ثلاثة مستويات للبيانات الممكن إجراء التحيل العنقودي الهرمي عليها:

- **المسافات المنتظمة أو الفترة Interval:** هناك عدة قياسات لهذا النوع من البيانات كاختبار Euclidean Distance، الذي يساوي الجذر التربيعي لمجموع مربع الفروق بين قيم حالتين أو أكثر على كافة المتغيرات، ومربع القياس السابق Squared Euclidean Distance.
- **الحالات التي تمثل بأعداد Counts:** هناك قياسين 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.
- **الثنائي Binary:** هناك عدة قياسات على هذا النوع من البيانات، من أهمها اختبار Squared Euclidean Distance

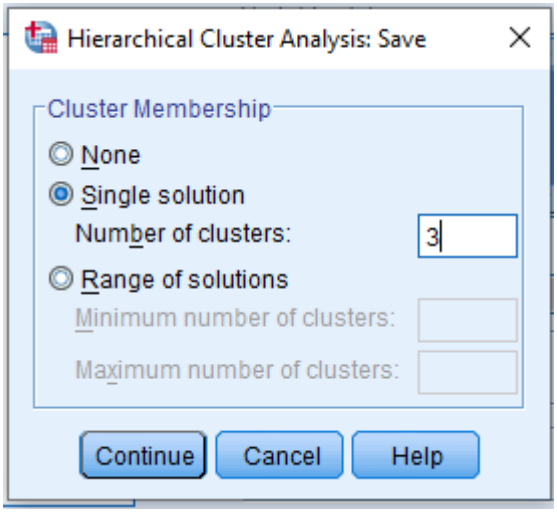
- **قيم التحويلات Transform values**، إذا كانت الوحدات المستخدمة لقياس المتغيرات موحدة كما في مثالنا (الدينار)، فيتم الإبقاء على الخيار None، أما إذا كانت هنالك عدة وحدات لقياس المتغيرات، فيمكن استخدام الإحصاءات الموجودة مثل اختبار Z، أو Range -1 to 1 أو غيرها، وهذه التحويلات لا تنطبق على البيانات الثنائية.

- **مقاييس التحويلات Transform measures**، هنالك ثلاث مربعات صغيرة: Absolute values, Change sign, Rescale to 0-1 range، حيث تتيح هذه الخيارات تحويل القيم المستخرجة من خلال قياس المسافات.

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 حفظ أية متغيرات جديدة None، وعدم إظهار أي متغير جديد في شاشة تحرير البيانات؛

الحل الواحد Single Solution، حسب المطلوب في السؤال، 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

Case Processing Summary[a]

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

a. Average Linkage (Between Groups)

- **الجدول الأول (السابق)** يعكس عدد ونسب الحالات الموجودة والمفقودة، وكذلك عدد ونسب إجمالي هذه الحالات.

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

- **مصفوفة القرب Proximity Matrix في الأعلى**، والتي تقرر التشابه أو عدم التشابه بين الحالات، ويعبر عنها بالمسافات المشتقة بين الأهداف المحددة، فالمسافة بين A,B مثلاً تساوي 7.5، والمسافة بين A,C تساوي 10.4 وهكذا...

Average Linkage (Between Groups)

Agglomeration Schedule

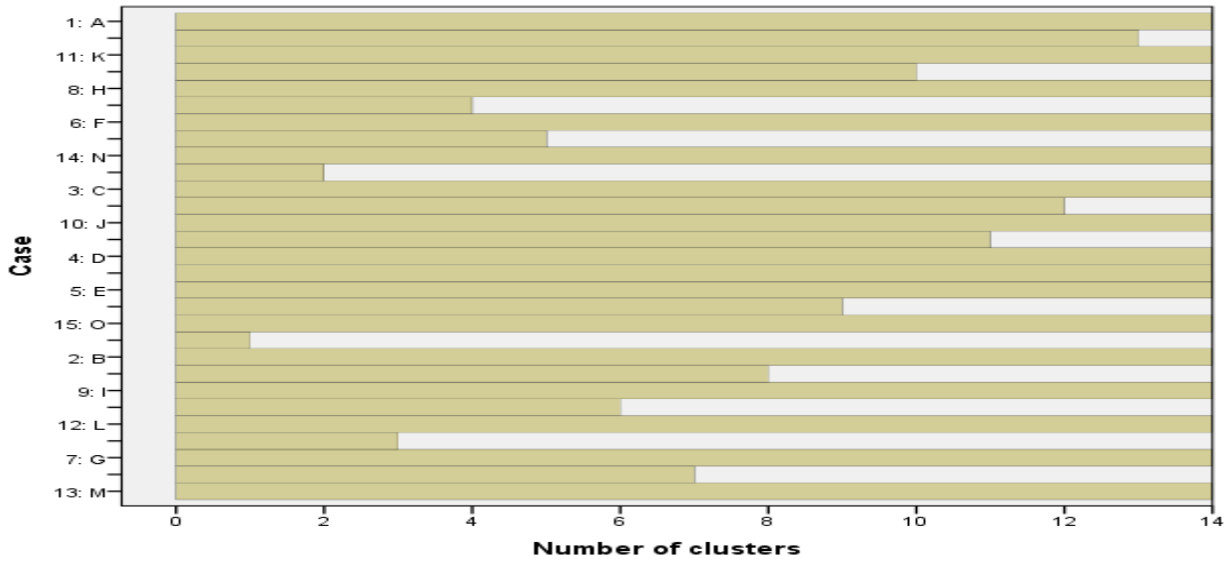
Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	4	5	2.566	0	0	4
2	1	11	2.694	0	0	5
3	3	10	2.726	0	0	4
4	3	4	3.903	3	1	6
5	1	8	4.023	2	0	11
6	3	15	4.279	4	0	13
7	2	9	4.336	0	0	9
8	7	13	5.205	0	0	12
9	2	12	5.419	7	0	12
10	6	14	5.933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

- **جدول التقارب Agglomeration Schedule في الأعلى** يبين 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 Coefficients في الجدول نجد أن المعاملات 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.

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

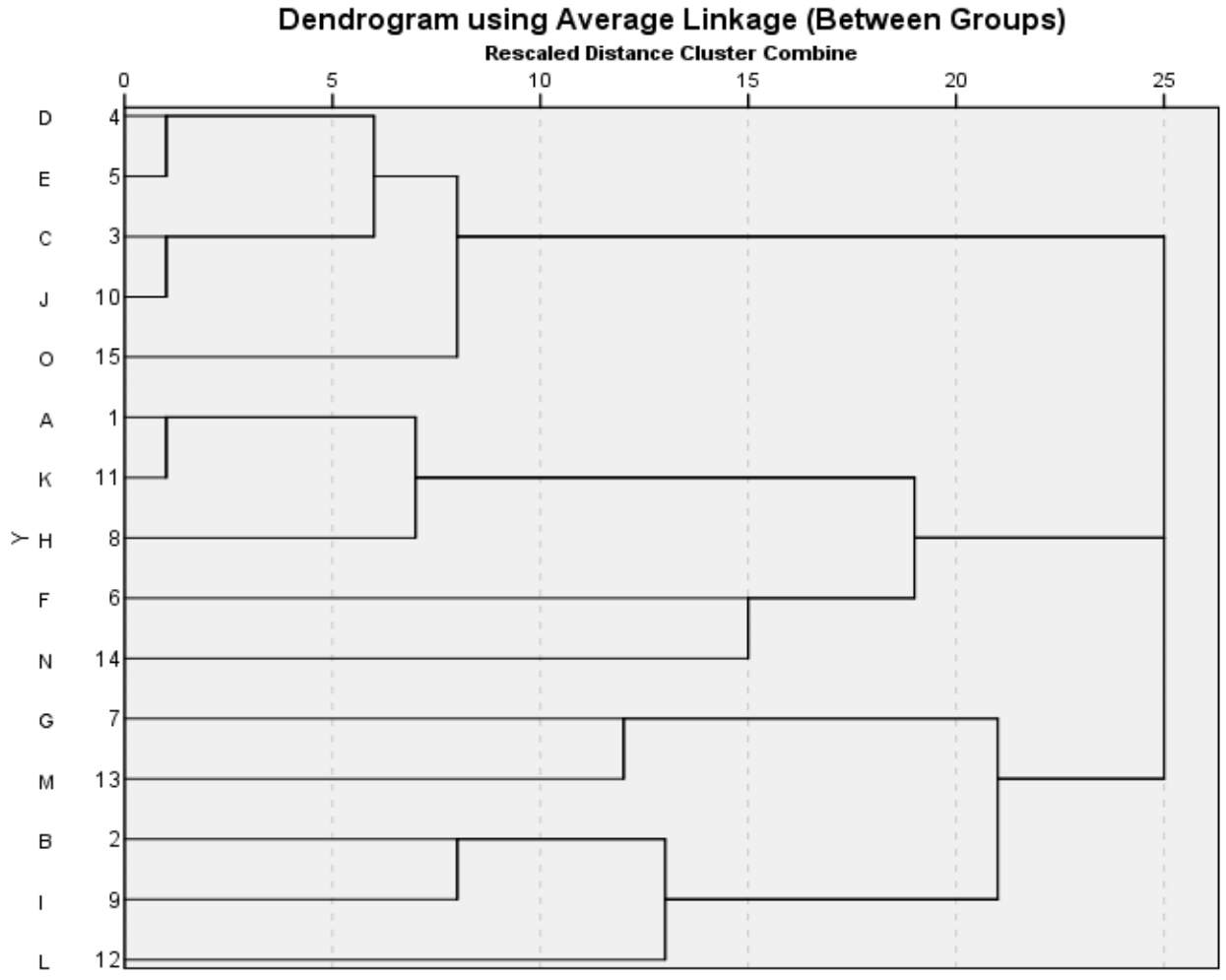
- **جدول انتماء الحالات إلى العناقيد Cluster Membership في الأعلى**، يوضح العنقود الذي تنتمي إليه كل حالة من الحالات (من A إلى O)، فمثلا الحالة 1 (A) تنتمي إلى عنقود 1، والحالة 2 (B) تنتمي إلى عنقود 2 وهكذا ...



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	×	×	×	×	×	×	×	×	×	×	×	×	×	×

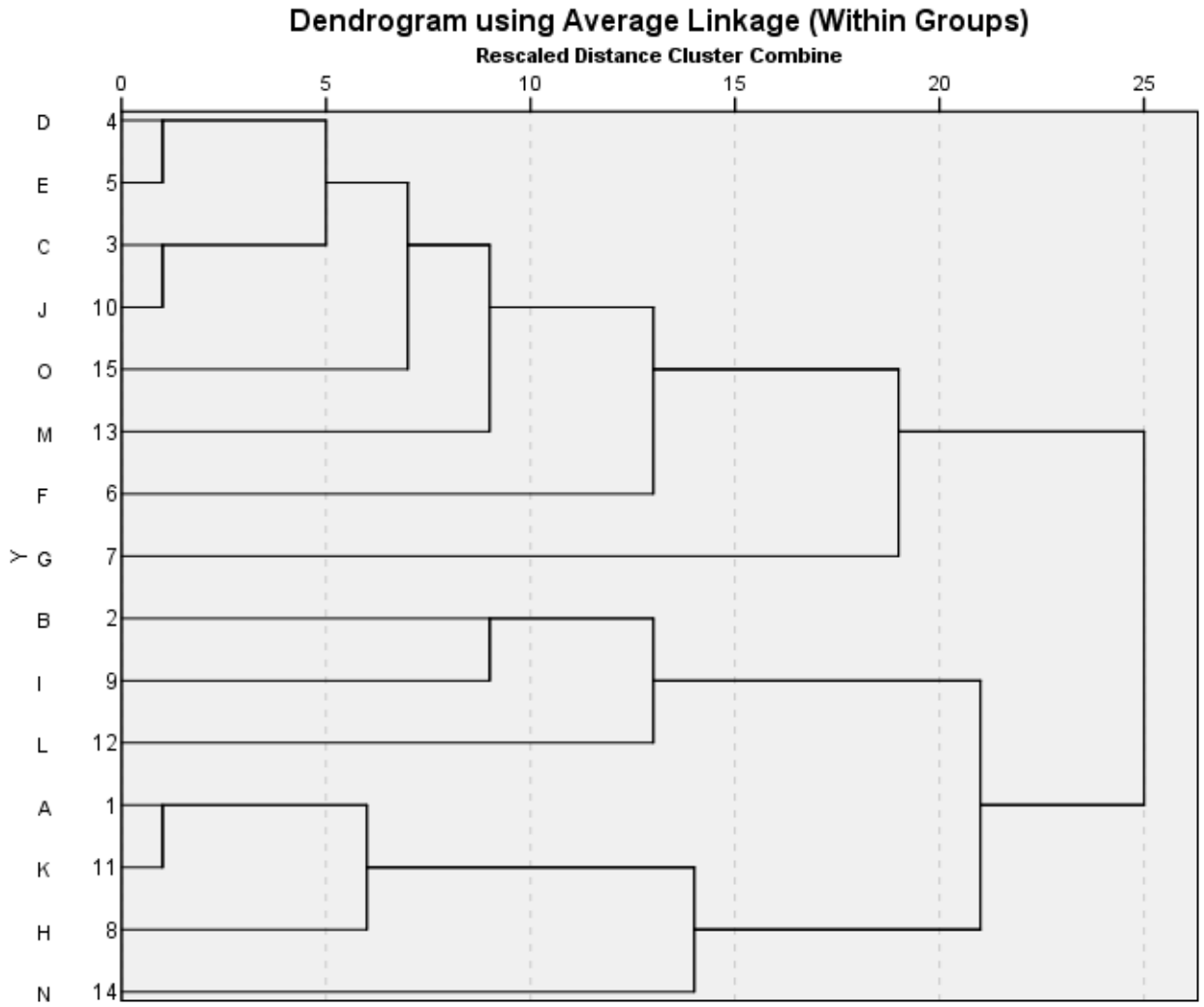
- **الجدول Icicle في الأعلى (Icicle تعني الجليد)**، في بعض طبعات SPSS يظهر في **شكل أعمدة يطلق عليها الألواح الجليدية**، والذي يمثل نقاط التقاء الحالات Cases مع عدد العناقيد Number of clusters عند كل مرحلة من مراحل التحليل، وبالتالي فهو يوفر معلومات عن كيفية دمج الحالات في عناقيد في كل تكرار من التحليل.



- **شكل الشجرة الثنائية Dendrogram**، وتضم الشجرة قياسات تمتد إلى 25 وحدة قياس حيث يشير طول الخط 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.
- نلاحظ في أعلى الشجرة أن عملية العنقدة كانت بين الحالتين D,E في البداية وفي أسفل الشجرة كانت بين B,I ثم B,I مع L، وتستمر العملية هكذا حتى نصل إلى عنقودين فقط.

بعد أن تم استعراض شكل الشجرة الثنائية وفقا لطريقة الربط بين المجموعات Between-groups Linkage، نعرض فيما يلي شكلين آخرين لنفس المعلومات الواردة في المثال، وذلك لأجل المقارنة بين أشكال الشجرة الثنائية تبعاً لطريقة التحليل العنقودي المتبعة.

الشجرة الثنائية وفقا لطريقة الربط داخل المجموعات Within-groups Linkage، وتبدو بالشكل التالي:



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

- الشجرة الثنائية وفقا لطريقة الربط المنفرد Single Linkage، وتبدو بالشكل التالي:

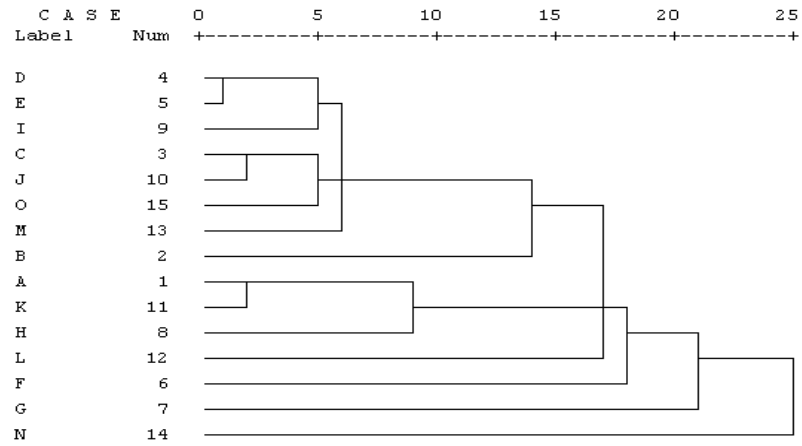
Single Linkage

□

* * * * * * H I E R A R C H I C A L C L U S T E R A N A L Y S I S * * * * * *

Dendrogram using Single Linkage

Rescaled Distance Cluster Combine



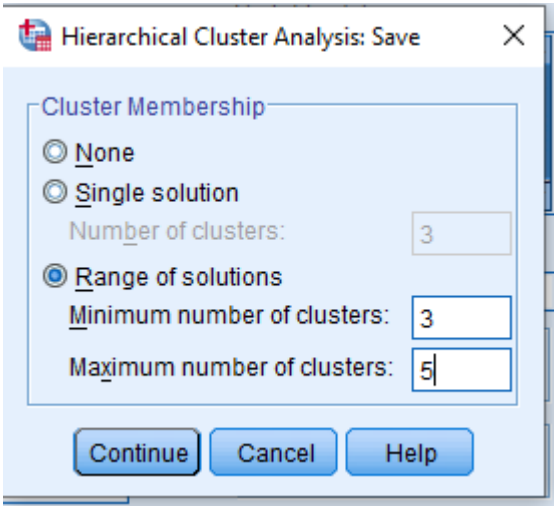
هنا تختلف الشجرة عن الحالة السابقة، فنجد في أعلى الشجرة أن عملية العنقدة كانت قد بدأت أيضاً بين الحالتين D,E، بينما في أسفل الشجرة كانت بين مجموعة أخرى من الحالات كما يبين الشكل.

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.

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 المتغيرالجديد CLU5_1 فيبين تصنيف الحالات على أساس 5 عناقيد وكذلك عضوية كل حالة في هذه العناقيد.

مثال تطبيقي (2):

أجرت إحدى الشركات الغذائية دراسة حول نسبة وجود البروتين والدهون والكالسيوم والبوتاسيوم في 100 جرام من كل نوع أنواع اللحوم التي تقدمها لعملائها. وفيما يلي نتائج الدراسة: **تعطى في جدول مرفق Exel رقم 02**

اللحوم	بروتينات	دهون	كالسيوم	بوتاسيوم
1	23	11	12	210
2	28	16	32	235
3	27	21	55	350
4	18	20	44	425
5	21	18	47	230
6	16	16	33	350
7	11	12	11	205

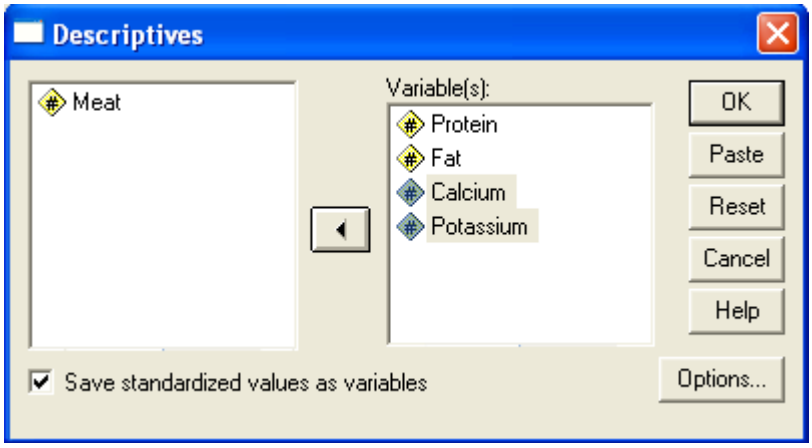
8	25	8	22	360
9	14	8	26	280
10	26	17	25	315
11	19	19	28	265
12	10	20	11	230
13	20	8	19	245
14	17	24	32	350
15	28	17	25	380

المطلوب: التحليل العنقودي على أساس تصنيف أنواع اللحوم الواردة تفاصيل مكوناتها أعلاه في عنقودين اثنين.

الحل:

الخطوة (1): إدخال البيانات في خمسة متغيرات كما يلي Meat, Protein, Fat, Calcium, Potassium، وهذا (انطلاقا من ملف Exel الأول المرفق إلى صفحة المعطيات SPSS)

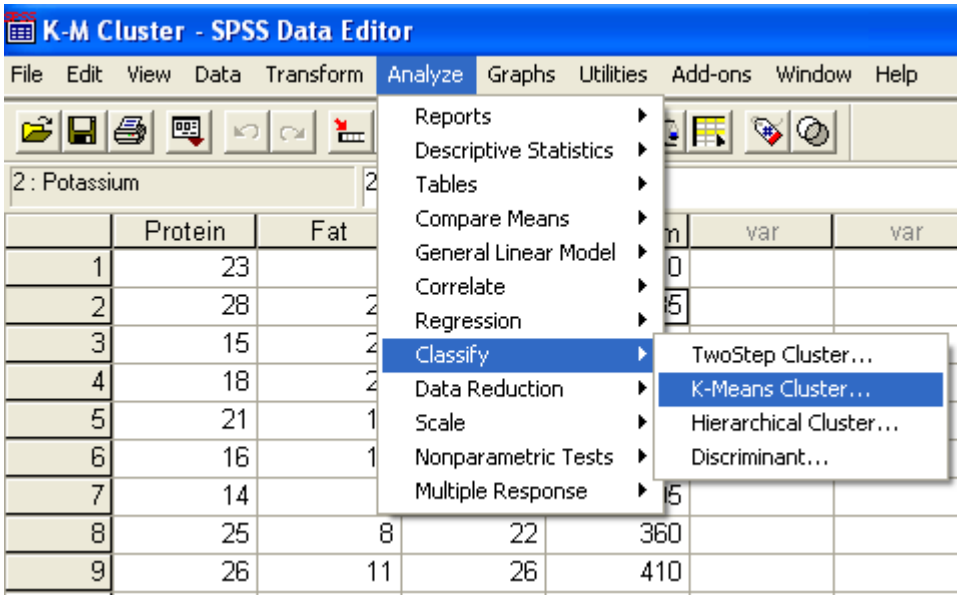
الخطوة (2): تحويل قيم المتغيرات إلى قيم معيارية، حيث أن وحدات قياس المتغيرات ليست موحدة، فالبروتين والدهون بـ: غ/100 غ، والكالسيوم والبوتاسيوم بـ: ملغ/ 100 غ، ويتم ذلك من خلال الأمر Descriptives، وإدخال المتغيرات الأربعة تحت Variable(s)، مع التأشير على المربع الصغير أمام Save standardized values as variables كما يلي:



الخطوة (3): الضغط على OK، واقفال صندوق الحوار بدون حفظ، فتظهر أربعة متغيرات جديدة في شاشة تحرير البيانات بالقيم المعيارية:

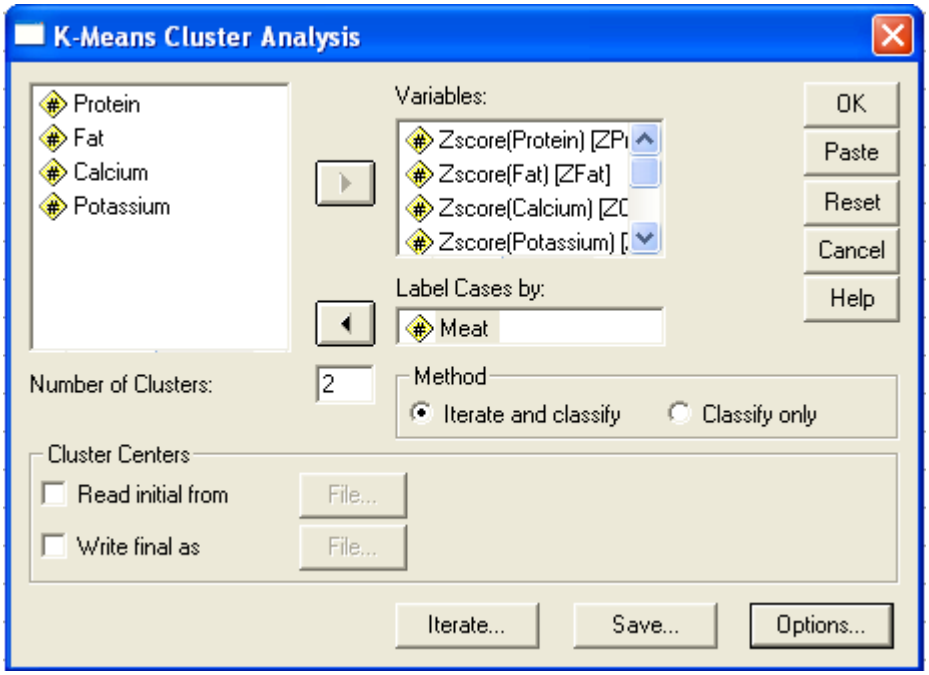
	Meat	Protein	Fat	Calcium	Potassium	ZProtein	ZFat	ZCalcium	ZPotassium
1	1	23	11	12	210	.47118	-.91105	-1.2385	-1.1353
2	2	28	16	32	235	1.31256	.06507	.29683	-.77448
3	3	27	21	55	350	1.14428	1.04120	2.06243	.88512
4	4	18	20	44	425	-.37021	.84597	1.21801	1.96746
5	5	21	18	47	230	.13462	.45552	1.44831	-.84663
6	6	16	16	33	350	-.70676	.06507	.37359	.88512
7	7	11	12	11	205	-1.5481	-.71582	-1.3152	-1.2074
8	8	25	8	22	360	.80773	-1.4967	-.47083	1.02943
9	9	14	8	26	280	-1.0433	-1.4967	-.16377	-.12507
10	10	26	17	25	315	.97601	.26030	-.24053	.38002
11	11	19	19	28	265	-.20193	.65075	-.01024	-.34154
12	12	10	20	11	230	-1.7164	.84597	-1.3152	-.84663
13	13	20	8	19	245	-.03366	-1.4967	-.70112	-.63016
14	14	17	24	32	250	-.53849	1.62687	.29683	-.55801
15	15	28	17	25	380	1.31256	.26030	-.24053	1.31805

الخطوة (4): بعد تحويل قيم المتغيرات إلى قيم معيارية (لاحظ إضافة الحرف Z بالإنجليزية قبل اسم كل متغير للدلالة على القيم المعيارية)، يتم اختيار القائمة الرئيسة Analyze، ثم القائمة الفرعية Classify، والضغط على K-Means Cluster كما يلي:



الخطوة (5): يظهر صندوق الحوار K-Means Cluster Analysis، انقل المتغيرات الأربعة التي تبدأ بالحرف Z إلى داخل المستطيل المعنون Variables وانقل المتغير Meat إلى داخل المستطيل المعنون Label Cases by

الخطوة (6): اختر الرقم 2 أمام المربع الصغير Number of Clusters كما يلي:



هناك طريقتان لترتيب الحالات تحت كل عنقود:

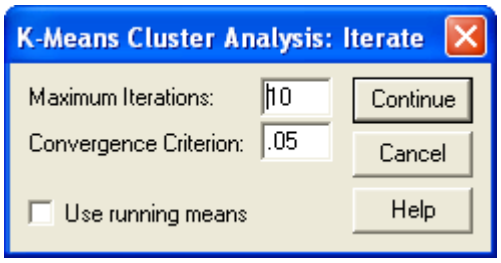
الطريقة الأولى Iterate and classify، فمن خلال هذا الاختيار يتم تفعيل المفتاح Iterate(إعادة)، وبالتالي يمكن فتح صندوق الحوار الخاص بذلك، وتحديد الخيارات الموجودة فيه.

الطريقة الثانية Classify only، حيث بناء على هذا الخيار يقفل الدخول إلى صندوق حوار Iterate and classify، اختر الطريقة الأولى Iterate and classify.

هناك ثلاثة مفاتيح تكون أسفل أو بجانب صندوق الحوار يتم الضغط عليها على التتالي

المفتاح الأول Iterate. (Iterate تعني إعادة)

الخطوة (7): اضغط على Iterate، ليظهر صندوق الحوار التالي:



- الحد الأعلى لتكرار العملية Maximum Iterations المحدد تلقائيا عشر مرات، يمكن تحديد رقم آخر، يتوقف تكرار العمليات بعد الوصول الى العدد المطلوب من المرات، يتم إبقاء الرقم 10 كما هو.

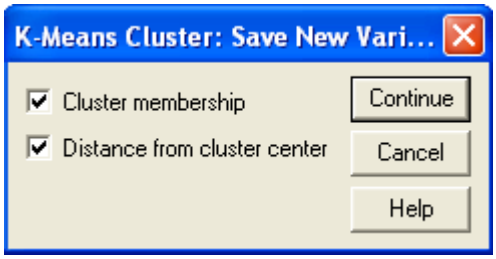
- أما فيما يتعلق بمعيار التقارب Convergence Criteria، فهو يمثل الحد الأدنى للمسافة بين مراكز العناقيد المبدئية، حيث تتوقف العمليات عندما يصل التغيير الأقصى المراد إجراؤه في أي مركز عنقود أقل من 2% من أقل مسافة بين العناقيد المبدئية، وبالتالي ينبغي أن يكون معيار التقارب أقل من واحد صحيح، اطبع الرقم 05 مثلاً.

- خيار استخدام الأوساط الجارية Use Running Means يقوم بتحديث مراكز العناقيد بعد تخصيص كل حالة إلى عنقود محدد، لا داعي لطلب هذا الخيار لأن البرنامج يقوم بحساب مراكز العناقيد بعد الانتهاء من تخصيص كافة الحالات.

يتم الضغط على Continue للعودة الى صندوق الحوار الرئيس.

المفتاح الثاني Save:

الخطوة (8): بالضغط على Save، يظهر صندوق الحوار التالي:



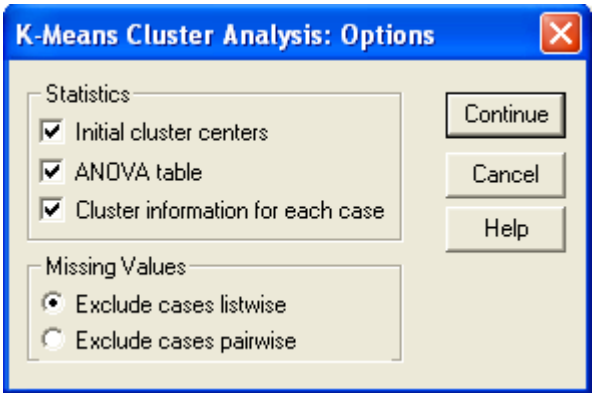
- اختيار الأمر المتعلق بعضوية العنقود Cluster Membership، وذلك لأجل اضافة متغير جديد في شاشة تحرير البيانات يشير الى رقم عضوية العنقود.

- اختيار الأمر المتعلق ببعد المسافة من مركز العنقود Distance from cluster center، لإضافة المتغير الجديد في شاشة تحرير البيانات لإظهار بعد المسافة، حسب اختيار Euclidean بين كل حالة وبين مركز العنقود.

يتم اضغط على Continue لتعود الى صندوق الحوار الرئيس.

المفتاح الثالث Options:

الخطوة (9): بالضغط على Options، يظهر صندوق الحوار التالي:



استبعاد القائمة: Exclude listwise؛ استبعاد الحالات الزوجية: Exclude cases pairwise

يمكن اختيار الإحصاءات المرغوب إظهارها في المخرجات، من خلال التأشير داخل المربع الصغير أمام كل من: مراكز العناقيد المبدئية Initial cluster centers، جدول تحليل التباين ANOVA table، بالإضافة إلى معلومات عن العنقود لكل حالة Cluster information for each case.

اضغط Continue، للعودة الى صندوق الحوار الرئيس.

الخطوة (10): اضغط OK في صندوق الحوار الرئيس ، فتظهر **المخرجات التالية**:

Quick Cluster

Initial Cluster Centers

	Cluster	
	1	2
Zscore(Protein)	-1.54815	-.37021
Zscore(Fat)	-.71582	.84597
Zscore(Calcium)	-1.31525	1.21801
Zscore(Potassium)	-1.20741	1.96746

- يشير جدول مراكز العناقيد المبدئية Initial Cluster Centers إلى متوسطات مراكز العناقيد التي تم تحديدها بشكل عشوائي لكل متغير من المتغيرات الأربعة، وذلك على أساس القيم المعيارية لهذه المتغيرات؛

Iteration History[a]

Iteration	Change in Cluster Centers	
	1	2
1	1.451	1.717
2	.589	.256
3	.000	.000

a. Convergence achieved due to no or small change in cluster centers. The maximum absolute coordinate change for any center is .000. The current iteration is 3. The minimum distance between initial centers is 4.508.

- يشير جدول Iteration History إلى أن هناك ثلاث مرات تمت فيهم عملية تخصيص الحالات للعناقيد، حيث بلغ الحد الأقصى للتغيير في مراكز العناقيد صفراً، وقد كان أقل حد من المسافة بين المراكز التي حددت مبدئياً 4.508؛

Cluster Membership

Case Number	Meat	Cluster	Distance
1	meat1	1	1.334
2	meat2	2	1.534
3	meat3	2	1.944
4	meat4	2	1.955
5	meat5	2	1.600
6	meat6	2	1.243
7	meat7	1	.955
8	meat8	2	2.233
9	meat9	1	1.295
10	meat10	2	.933
11	meat11	2	1.094
12	meat12	1	1.895
13	meat13	1	1.088
14	meat14	2	1.832
15	meat15	2	1.494

- يبين جدول عضوية العنقود Cluster Membership نوع اللحوم الخاضعة للدراسة ورقم العنقود الذي ينتمي إليه كل نوع من اللحوم ، بالإضافة إلى المسافة بين القيم المعيارية لكل نوع وبين مركز العنقود التابع له؛

Final Cluster Centers

	Cluster	
	1	2
Zscore(Protein)	-.77407	.38704
Zscore(Fat)	-.75487	.37743
Zscore(Calcium)	-.94677	.47339
Zscore(Potassium)	-.78891	.39445

- يوضح الجدول النهائي لمراكز العناقيد Final Cluster Centers متوسطات هذه المراكز لكل عنقود ولكل متغير بشكلها النهائي، كما يوضح جدول المسافة بين مراكز العناقيد النهائية Distances between Final Cluster Centers، والذي بلغ 2.459؛

Distances between Final Cluster Centers

Cluster	1	2
1		2.459
2	2.459	

ANOVA

	Cluster		Error			
	Mean Square	df	Mean Square	df	F	Sig.
Zscore(Protein)	4.494	1	.731	13	6.146	.028
Zscore(Fat)	4.274	1	.748	13	5.712	.033
Zscore(Calcium)	6.723	1	.560	13	12.010	.004
Zscore(Potassium)	4.668	1	.718	13	6.502	.024

The F tests should be used only for descriptive purposes because the cluste have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

- بناء على الطلب يمكن إضافة جدول التباين ANOVA، والذي يبين فيه قيمة F، ومستوى المعنوية .Sig لكل متغير، وقيمة اختبار F تستخدم فقط لأغراض الوصف، لأن اختيار العناقيد كان بهدف مضاعفة الفروق بين الحالات في العناقيد المختلفة؛

Number of Cases in each Cluster

Cluster	1	5.000
	2	10.000
Valid		15.000
Missing		.000

- الجدول الأخير Number of Cases in each Cluster يشير إلى إجمالي تصنيف الحالات أو أنواع اللحوم، حيث تم تصنيف ما مجموعه خمسة حالات في العنقود الأول، وعشرة حالات في العنقود الثاني.

الخطوة (11): بعد إقفال شاشة المخرجات وبالرجوع إلى شاشة تحرير البيانات، تجد أنه قد تم إضافة متغيرين اثنين جديدين:

	Meat	Protein	Fat	Calcium	Potassium	ZProtein	ZFat	ZCalcium	ZPotassium	QCL_1	QCL_2
1	1	23	11	12	210	.47118	-.9110	-1.238	-1.135	1	1.3342
2	2	28	16	32	235	1.3126	.06507	.29683	-.7745	2	1.5335
3	3	27	21	55	350	1.1443	1.0412	2.0624	.88512	2	1.9442
4	4	18	20	44	425	-.3702	.84597	1.2180	1.9675	2	1.9549
5	5	21	18	47	230	.13462	.45552	1.4483	-.8466	2	1.6002
6	6	16	16	33	350	-.7068	.06507	.37359	.88512	2	1.2429
7	7	11	12	11	205	-1.548	-.7158	-1.315	-1.207	1	.95480
8	8	25	8	22	360	.80773	-1.497	-.4708	1.0294	2	2.2325
9	9	14	8	26	280	-1.043	-1.497	-.1638	-.1251	1	1.2948
10	10	26	17	25	315	.97601	.26030	-.2405	.38002	2	.93300
11	11	19	19	28	265	-.2019	.65075	-.0102	-.3415	2	1.0941
12	12	10	20	11	230	-1.716	.84597	-1.315	-.8466	1	1.8947
13	13	20	8	19	245	-.0337	-1.497	-.7011	-.6302	1	1.0882
14	14	17	24	32	250	-.5385	1.6269	.29683	-.5580	2	1.8320
15	15	28	17	25	380	1.3126	.26030	-.2405	1.3181	2	1.4943

ملاحظة: يمكن استخدام المتغيرين الجديدين: رقم العنقود لكل حالة QCL_1، والمسافة بين الحالة وبين مركز العنقود QCL_2 في أية عمليات أخرى للتحليل الإحصائي.