مجلة بحوث التربية الخاصة والتعليم الشامل، المجلد الأول، العدد الثاني سبتمبر، 2023م، ص ص 34-87

التحليل التجميعي باستخدام برنامج SPSS وتطبيقاته النفسية والتربوية والاجتماعية

إعداد: أ.د. عبد الناصر أنيس عبد الوهاب

مستخلص

يمثل التحليل التجميعي أو العنقودي مهمة تجميع مجموعة من الكائنات بطريقة تجعل الكائنات الموجودة في نفس المجموعة (تسمى التجمع) أكثر تشابهًا (بمعنى ما) مع بعضها البعض عن تلك الموجودة في المجموعات (التجمعات) الأخرى. ويمثل التحليل التجميعي مهمة رئيسية لتحليل البيانات الاستكشافية، وهو فنية شائعة لتحليل البيانات الإحصائية، تُستخدم في العديد من المجالات، بما في ذلك التعرف على الأنماط، وتحليل الصور، واسترجاع المعلومات، والمعلوماتية الحيوية، وضغط البيانات، ورسومات الكمبيوتر، وتعلم الآلة، إلا أنه غير شائع الاستخدام في العلوم النفسية والتربوية والاجتماعية. يعتبر التحليل التجميعي في حد ذاته ليس خوارزمية واحدة محددة، رغم المهمة العامة التي يتعين حلها باستخدامه. ويمكن تحقيقه من خلال خوارزميات مختلفة تختلف اختلافًا كبيرًا في فهمها لما يشكل التجمع وكيفية العثور عليه بكفاءة. وتتضمن المفاهيم الشائعة للتجمعات أو العناقيد مجموعات ذات مسافات صغيرة بين أعضاء التجمع، أو مناطق كثيفة من مساحة البيانات، أو فترات أو توزيعات إحصائية معينة. لذلك يمكن صياغة التجميع على أنه مشكلة تحسين متعددة الأهداف. وتعتمد خوارزمية التجميع المناسبة وإعدادات المعلمات (بما في ذلك المعلمات مثل وظيفة المسافة المراد استخدامها أو عتبة الكثافة أو عدد المجموعات المتوقعة) على مجموعة البيانات الفردية والاستخدام المقصود للنتائج. وتتناول هذه الورقة العلمية أهداف التحليل التجميعي، والمفاهيم الأساسية له، وكيفية إجرائه باستخدام حزمة البرامج الإحصائية للعلوم الاجتماعية SPSS، وتقديم بعض التطبيقات مع العرض الجدولي والبياني لها في البحوث النفسية والتربوية والاجتماعية.

الكلمات المفتاحية:

التحليل التجميعي، التحليل التجميعي للحالات، التحليل التجميعي للمتغيرات، العلوم النفسية، التربية، العلوم الاجتماعية

Cluster Analysis Using SPSS and Its Psychological, Educational and Social applications

Prepared by: Prof. Dr. Abdel Nasser Anis Abdel Wahhab

Abstract

Clustering or cluster analysis is the task of grouping a cluster of objects in such a way that the objects in the same group (called a cluster) are more similar (in some sense) to each other than those in other groups (clusters). Cluster analysis is a major task of exploratory data analysis, and it is a popular technique for statistical data analysis, used in many fields, including pattern recognition, image analysis, information retrieval, bioinformatics, data compression, computer graphics, and machine learning, but it is not commonly used in psychological, educational and social sciences. Cluster analysis in itself is not one specific algorithm, despite the general task to be solved using it. It can be achieved by different algorithms that differ widely in their understanding of what constitutes a pool and how to find it efficiently. Common concepts of clusters include groups with small distances between members of the cluster, dense regions of data space, or particular statistical intervals or distributions. So clustering can be formulated as a multi-objective optimization problem. The appropriate clustering algorithm and parameter settings (including parameters such as the distance function to use, density threshold, or number of expected clusters) depend on the individual data set and the intended use of the results. This scientific paper deals with the objectives of cluster analysis, its basic concepts, and how to conduct it using the statistical software package for social sciences: SPSS, and presents some applications with tabular and graphic presentation in psychological, educational and social research.

Key words:

Meta-analysis, meta-analysis of cases, meta-analysis of variables, psychological sciences, education, social sciences

التحليل التجميعي باستخدام برنامج SPSS وتطبيقاته النفسية والتربوية والاجتماعية

إعداد: أ.د. عبد الناصر أنيس عبد الوهاب

مقدمة:

يمثل التحليل التجميعي أو العنقودي مهمة تجميع مجموعة من الكائنات بطريقة تجعل الكائنات الموجودة في نفس المجموعة (تسمى التجمع) أكثر تشابهًا (بمعنى ما) مع بعضها البعض عن تلك الموجودة في المجموعات (التجمعات) الأخرى. ويمثل التحليل التجميعي مهمة رئيسية لتحليل البيانات الاستكشافية، وهو فنية شائعة لتحليل البيانات الإحصائية، تُستخدم في العديد من المجالات، بما في ذلك التعرف على الأنماط، وتحليل الصور، واسترجاع المعلومات، والمعلوماتية الحيوية، وضغط البيانات، ورسومات الكمبيوتر، وتعلم الآلة، إلا أنه غير شائع الاستخدام في العلوم النفسية والتربوية والاجتماعية.

يعتبر التحليل التجميعي في حد ذاته ليس خوارزمية واحدة محددة، رغم المهمة العامة التي يتعين حلها باستخدامه. ويمكن تحقيقه من خلال خوارزميات مختلفة تختلف اختلافًا كبيرًا في فهمها لما يشكل التجمع وكيفية العثور عليه بكفاءة. وتتضمن المفاهيم الشائعة للتجمعات أو العناقيد مجموعات ذات مسافات صغيرة بين أعضاء التجمع، أو مناطق كثيفة من مساحة البيانات، أو فترات أو توزيعات إحصائية معينة. لذلك يمكن صياغة التجميع على أنه مشكلة تحسين متعددة الأهداف. وتعتمد خوارزمية التجميع المناسبة وإعدادات المعلمات (بما في ذلك المعلمات مثل وظيفة المسافة المراد استخدامها أو عتبة الكثافة أو عدد المجموعات المتوقعة) على مجموعة البيانات الفردية والاستخدام المقصود للنتائج.

وتتناول هذه الورقة العلمية أهداف التحليل التجميعي، والمفاهيم الأساسية له، وكيفية إجرائه باستخدام حزمة البرامج الإحصائية للعلوم الاجتماعية SPSS، وتقديم بعض التطبيقات مع العرض الجدولي والبياني لها في البحوث النفسية والتربوية والاجتماعية.

الأهداف والغايات:

تهدف هذه الورقة العلمية إلى ما يلي:

- المعرفة العملية بالطرق التي يمكن من خلالها قياس التشابه بين الحالات (على سبيل المثال، ارتباط فردي، وارتباط كامل وارتباط متوسط).
- المعرفة العملية بالطرق التي يمكن من خلالها قياس التشابه بين المتغيرات.
- القدرة على إنتاج وتفسير الجداول والمخططات الرسومية dendrograms التي

ينتجها برنامج SPSS عند استخدام التحليل التجميعي.

- العلم بأن طرق التجميع المختلفة ستنتج هياكل تجميعية (عنقودية) مختلفة.

ما هو التحليل العنقودي؟

يمكن بالفعل استخدام التحليل العاملي لتجميع المتغيرات وفقًا للتباين المشترك. في التحليل العاملي، نأخذ العديد من المتغيرات، ونفحص مقدار التباين الذي تشاركه هذه المتغيرات، ومقدار التباين الفريد، ثم متغيرات "التجمع أو المجموعة" معًا التي تشترك في نفس المتغيرات. باختصار، نقوم بتجميع المتغيرات معًا التي تبدو وكأنها تشرح نفس التباين. كان المثال في كتاب SPSS لفيلد (Field, 2013) عبارة عن استبيان لقياس القدرة على اختبار قلق استخدام SPSS، وكانت نتيجة التحليل العاملي عزل مجموعات من الأسئلة التي يبدو أنها تشترك في تباينها من أجل عزل الأبعاد المختلفة لقلق الإحصاء باستخدام SPSS.

إن الحديث عن التحليل العاملي عن تناول التحليلي التجميعي، أمر طبيعي؛ لأن، التحليل العنقودي في جوهره يعد أسلوبًا مشابهًا باستثناء أنه بدلاً من محاولة تجميع المتغيرات معًا، فإننا مهتمون بتجميع الحالات، ويمكننا أيضاً تجميع المتغيرات مثل التحليل العاملي.

عادة، في علم النفس على أي حال، هذا يعني أننا مهتمون بتجميع مجموعات من الناس في حالة تجميع الحالات. لذا، فهو بمعنى ما يعتبر عكس التحليل العاملي: بدلاً من تشكيل مجموعات من المتغيرات بناءً على استجابات العديد من الأشخاص لتلك المتغيرات، نقوم بدلاً من ذلك بتجميع الأشخاص بناءً على استجاباتهم للعديد من المتغيرات.

لذلك، كمثال، إذا قمنا بقياس الاحتفاظ، وعدد الأصدقاء والمهارات الاجتماعية، فقد نجد مجموعتين متميزتين من الأشخاص: محاضرين الإحصاء (الذين حصلوا على درجة عالية في الاحتفاظ ودرجة منخفضة في عدد الأصدقاء والمهارات الاجتماعية) والطلاب (الذين حصلوا على درجات منخفضة في احتباس الشرج وعالية في عدد الأصدقاء والمهارات الاجتماعية).

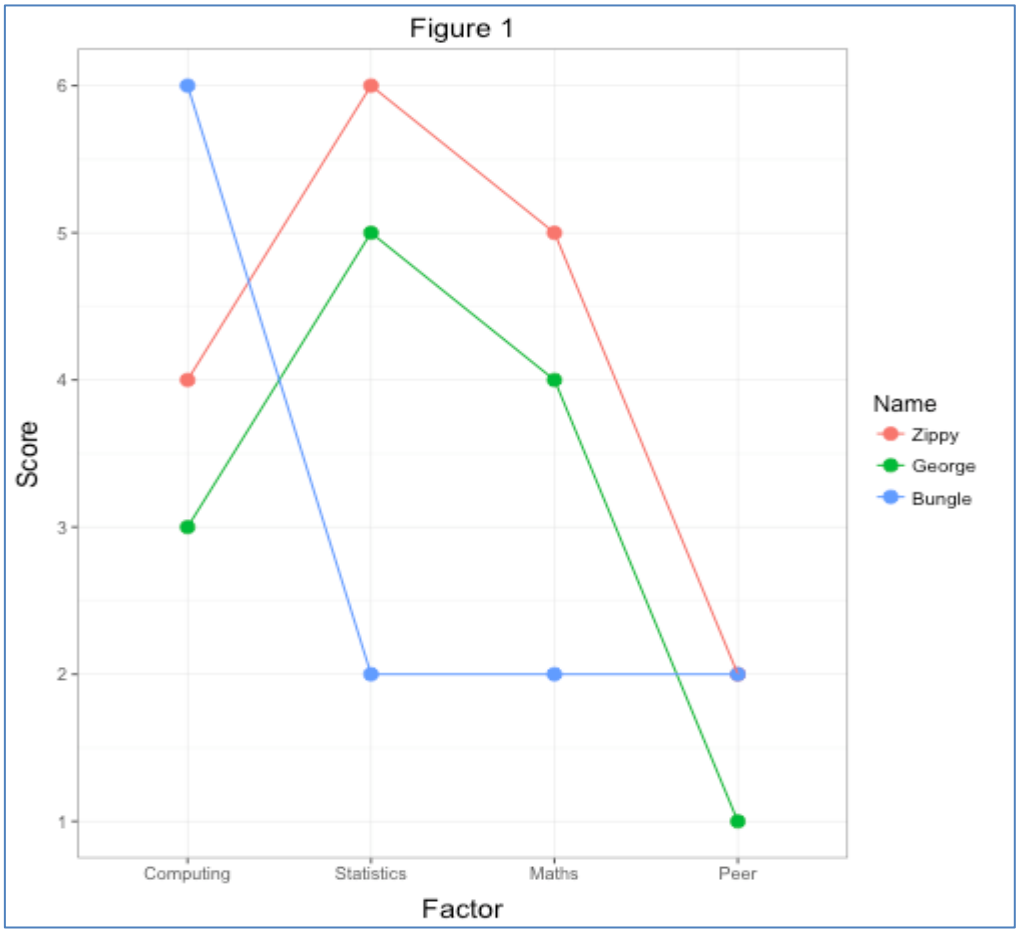
بإيجاز، يعتبر التحليل التجميعي طريقة لتجميع حالات البيانات بناءً على تشابه الاستجابات مع العديد من المتغيرات، وتجميع المتغيرات بناء على تشابه المتغيرات في العديد من الحالات.

كيف يعمل التحليل العنقودي؟

تخيل سيناريو بسيطًا قمنا فيه بقياس درجات ثلاثة أشخاص في استبيان قلق SPSS الخاص بفيلد (Field, 2013). نتج عن هذا الاستبيان أربعة عوامل: القلق الحسابي، القلق الإحصائي، القلق من الرياضيات والقلق المتعلق بالتقييم من

الأقران. يقوم الأشخاص الثلاثة بملء الاستبيان ومن التحليل العاملي نحصل على درجات العوامل لكل من هذه المكونات الأربعة. كمقياس بسيط للتشابه بين درجاتهم، يمكننا رسم بياني خطي بسيط يوضح العلاقة بين درجاتهم. يوضح الشكل 1 مثل هذا الرسم البياني.

بالنظر إلى الشكل 1، من الواضح جدًا أن لدى زيبي وجورج نمطًا متشابهًا جدًا من الاستجابات عبر العوامل الأربعة (في الواقع خطوطهم متوازية، مما يشير إلى أن الاختلاف النسبي في درجاتهم عبر العوامل هو نفسه). ومع ذلك، فإن لدى Bungle مجموعة مختلفة جدًا من الاستجابات. حصل على درجة مشابهة جدًا لزيبي وجورج لعامل "تقييم الأقران" ولكن بالنسبة للعوامل الثلاثة المتبقية، فإن درجاته مختلفة جدًا عن العاملين الآخرين. لذلك، يمكننا تجميع زيبي وجورج معًا بناءً على حقيقة أن بروفيل استجاباتهما متشابه جدًا.



شكل 1. رسم بياني خطي بسيط يوضح العلاقة بين درجات 4 أربعة أفراد على أربعة مقاييس

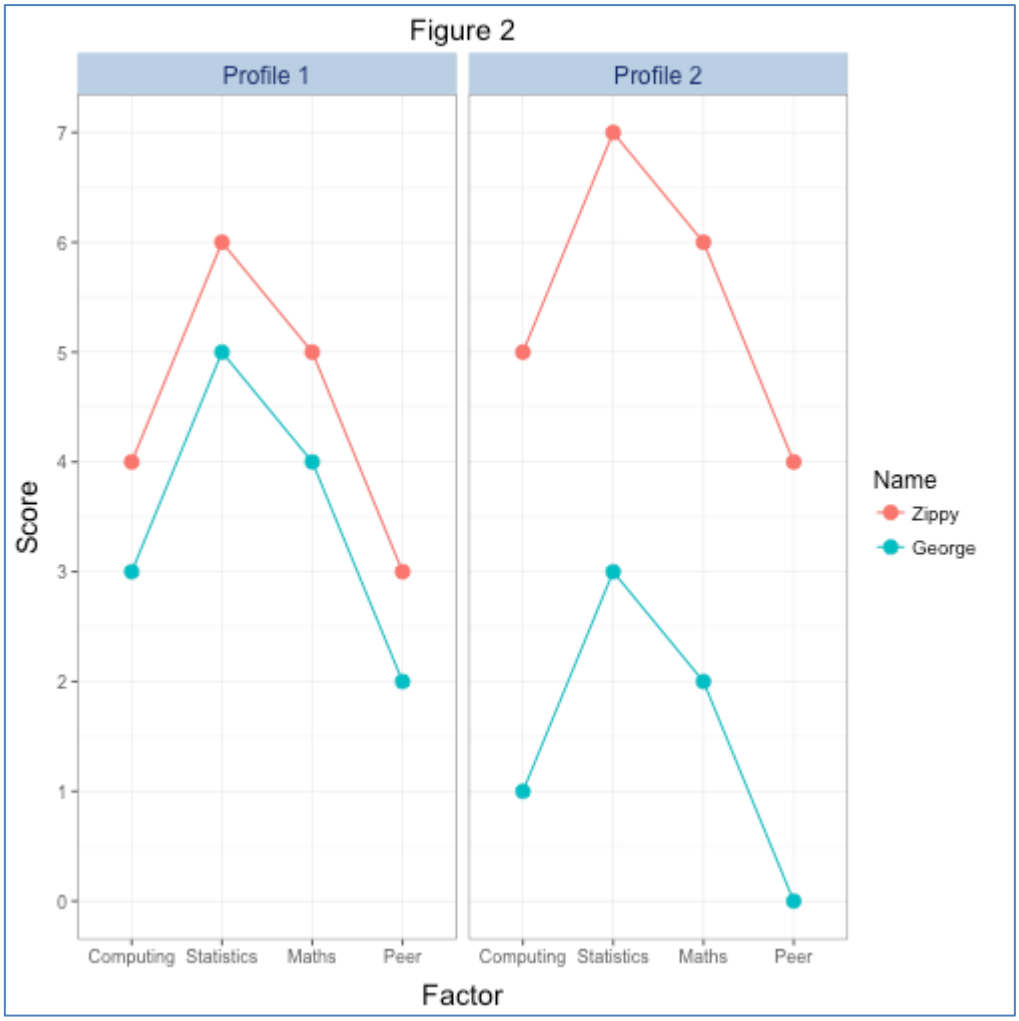
قياس التشابه أو التقارب بين الحالات:

من الواضح، النظر إلى الرسوم البيانية للاستجابات إذا كانت طريقة ذاتية للغاية لتحديد ما إذا كان لدى شخصين استجابات متشابهة عبر المتغيرات. بالإضافة إلى ذلك، في المواقف التي يكون لدينا فيها مئات الأشخاص والعديد من المتغيرات، فإن الرسوم البيانية للاستجابات التي نرسمها ستصبح مرهقة للغاية ويكاد يكون من المستحيل تفسيرها. لذلك، نحتاج إلى طريقة موضوعية لقياس درجة التشابه بين درجات الأشخاص عبر عدد من المتغيرات. هناك نوعان من القياس: معاملات التشابه ومعاملات الاختلاف. هل يمكنك التفكير في مقياس للتشابه بين متغيرين صادفتهما من قبل (مرات عديدة) يمكن تكييفهما لقياس تشابه الناس؟

معامل الارتباط، ر

معامل الارتباط Correlation Coefficient, r هو مقياس للتشابه للتقارب أو التشابه بين متغيرين (يخبرنا ما إذا كان متغير واحد يقترن بالتغيير بالتغييرات الأخرى بمقدار مماثل أم لا). من الناحية النظرية، يمكننا تطبيق معامل الارتباط على شخصين بدلاً من متغيرين لمعرفة ما إذا كان نمط الاستجابات لشخص واحد هو نفسه للآخر أم لا. يُعد معامل الارتباط مقياسًا موحدًا، ولذا فهو يتمتع بميزة أنه لا يتأثر باختلافات التشتت عبر المتغيرات (هذا يعني أنه إذا تم قياس المتغيرات التي نقارن الأشخاص عبرها بوحدات مختلفة، فلن يكون معامل الارتباط متأثراً كذلك). ومع ذلك، هناك مشكلة في استخدام معامل ارتباط بسيط لمقارنة الأشخاص عبر المتغيرات: فهو يتجاهل المعلومات حول ارتفاع الدرجات. وبالتالي، على الرغم من أن معامل الارتباط يخبرنا ما إذا كان نمط الاستجابات بين الأشخاص متشابهًا، إلا أنه لا يخبرنا بأي شيء عن المسافة بين البروفيلات الشخصية لشخصين.

يوضح الشكل 2 مثالين للاستجابات عبر عوامل استبيان قلق الإحصاء SAQ. في كلا المخططين، يمتلك الشخصان (زيبي وجورج) بروفيلات متشابهة (الخطوط متوازية). لذلك، سيكون معامل الارتباط الناتج للرسمين البيانيين متطابقًا (في الواقع، ستحصل على ارتباط مثالي 1). ومع ذلك، فإن المسافة بين البروفيلين أكبر بكثير في الرسم البياني الثاني (الارتفاع أعلى). لذلك، قد يكون من المعقول استنتاج أن الأشخاص الموجودين في الرسم البياني الأول أكثر تشابهًا من الاثنين في الرسم البياني الثاني، ومع ذلك فإن معامل الارتباط هو نفسه. على هذا النحو، فإن معامل الارتباط يفتقد إلى معلومات مهمة.



شكل 2. مثالان للاستجابات عبر عوامل استبيان قلق الإحصاء SAQ

المسافة الإقليدية، د

هناك مقياس بديل هو المسافة الإقليدية Euclidean Distance, d. والمسافة الإقليدية هي المسافة الهندسية بين شخصين (أو حالتين). لذلك، إذا كنا سنطلق على جورج الشخص الأول George subject i وزيي الشخص الثاني Zippy subject j، فيمكننا التعبير عن المسافة الإقليدية من حيث المعادلة التالية:

$$d_{ij} = \sqrt{\sum_{k=1}^{p} (x_{ik} - x_{jk})^2}$$

تعني هذه المعادلة ببساطة أنه يمكننا اكتشاف المسافة بين زيي Zippy وجورج George من خلال أخذ درجاتهم على متغير k، وحساب الفرق. الآن، بالنسبة لبعض المتغيرات، سيكون لدى زيي Zippy درجة أكبر من جورج George وبالنسبة

للمتغيرات الأخرى، سيكون لدى جورج George درجة أكبر من Zippy. لذلك، ستكون بعض الاختلافات إيجابية وبعضها الآخر سلبي. في النهاية، نريد جمع الفروق عبر عدد من المتغيرات، وبالتالي إذا كان لدينا فرق موجب وفرق سالب، فقد يتم حذفهما. لتجنب هذه المشكلة، نقوم ببساطة بتربيع كل فرق قبل جمعهم. لقد حصلنا حتى الآن على درجات زيبي Zippy وجورج George للمتغير k وقمنا بحساب الفرق وتربيعه. كل ما نفعله الآن هو الانتقال إلى المتغير التالي ونفعل الشيء نفسه. عندما نفعل الشيء نفسه لكل متغير نضيف جميع الاختلافات (الأمر يشبه حساب التباين تماماً). عندما أضفنا جميع الفروق التربيعية، نأخذ الجذر التربيعي (لأنه بتربيع الاختلافات قمنا بتغيير وحدات القياس إلى وحدات مربعة وهكذا بأخذ الجذر التربيعي نعود إلى وحدات القياس الأصلية). في الواقع، يتم استخدام متوسط المسافة الإقليدية (لذلك بعد جمع الفروق التربيعية نقسمها ببساطة على عدد المتغيرات) لأنها تسمح بالبيانات المفقودة.

مع المسافات الإقليدية، كلما كانت المسافة أصغر، كانت الحالات أكثر تشابهًا. ومع ذلك، يتأثر هذا المقياس بشدة بالمتغيرات ذات الحجم الكبير أو اختلافات التشتت. لذلك، إذا تمت مقارنة الحالات عبر متغيرات لها تباينات مختلفة جدًا (أي أن بعض المتغيرات أكثر انتشارًا من غيرها)، فإن المسافات الإقليدية ستكون غير دقيقة. على هذا النحو، من المهم توحيد الدرجات قبل متابعة التحليل. يعد توحيد الدرجات أمرًا مهمًا بشكل خاص إذا تم قياس المتغيرات على مستويات مختلفة.

تكوين التجمعات (العناقيد):

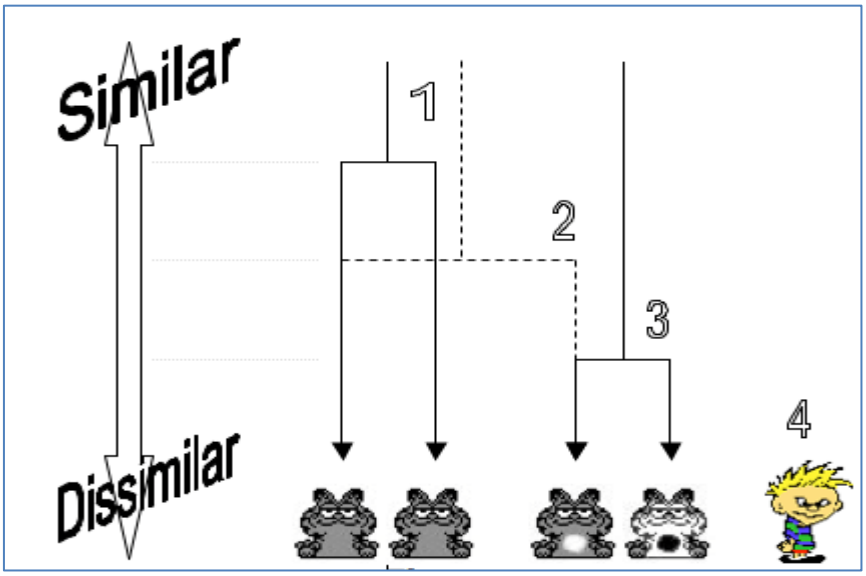
بمجرد أن يكون لدينا مقياس للتشابه بين الحالات، يمكننا التفكير في الطرق التي يمكننا من خلالها تجميع الحالات بناءً على تشابهها. وهناك عدة طرق لتجميع الحالات بناءً على معاملات التشابه. وتعمل معظم هذه الأساليب بطريقة هرمية. ويتشابه المبدأ الكامن وراء كل طريقة من حيث أنه يبدأ بمعالجة جميع الحالات كمجموعة في حد ذاتها، ثم يتم دمج المجموعات بناءً على معيار خاص بالطريقة المختارة. لذلك، في جميع الطرق، نبدأ بالتجمعات أو العناقيد بقدر عدد الحالات وننتهي بمجموعة واحدة فقط تحتوي على جميع الحالات. ومن خلال فحص تقدم الدمج التجميعي (العنقودي)، يمكن عزل مجموعات الحالات ذات التشابه الكبير.

الرابطة المنفردة Single Linkage **أو** SLINK **(الجار الأقرب** Nearest Neighbour**):**

تعتبر الرابطة المنفردة أبسط طريقة وبالتالي فهي نقطة انطلاق جيدة لفهم المبادئ الأساسية لكيفية تشكل التجمعات (والطبيعة الهرمية للعملية). وتتمثل الفكرة الأساسية في الآتي:

1. تبدأ كل حالة على شكل تجمع.

2. ابحث عن الحالتين / التجمعين الأكثر تشابهًا (مثل A و B) من خلال النظر إلى معاملات التشابه بين أزواج من الحالات (مثل الارتباطات أو المسافات الإقليدية). يتم دمج الحالات / المجموعات ذات التشابه الأعلى لتشكيل نواة تجمع أكبر.
3. الحالة / التجمع التالي الذي سيتم دمجه مع هذا التجمع الأكبر هو الحالة ذات أعلى معامل تشابه مع أي من A أو B.
4. الحالة التالية المدمجة هي الحالة ذات التشابه الأكبر مع A أو B أو C، وهكذا.



شكل 3. كيفية عمل طريقة الرابطة المنفردة

يوضح الشكل 3 كيفية عمل طريقة الرابطة المنفردة. إذا قمنا بتقييم 5 حيوانات بناءً على خصائصها الفيزيائية (اللون، وعدد الأرجل، والعينين وما إلى ذلك) وأردنا تجميع هذه الحيوانات بناءً على هذه الخصائص، فسنبدأ بالحيوانين الأكثر تشابهًا. أولاً، تخيل معامل التشابه كمقياس رأسي يتراوح من تشابه منخفض إلى مرتفع. في طريقة الرابطة المنفردة، نبدأ بالحالتين الأكثر تشابهًا.. لدينا حيوانان متشابهان جدًا بالفعل (في الواقع يبدوان متطابقين). وبالتالي فإن معامل التشابه بينهما مرتفع. الشوكة التي تنقسم عند نقطة على المقياس الرأسي تمثل معامل التشابه تمثل التشابه بين هذه الحيوانات. لذلك، نظرًا لأن التشابه مرتفع، فإن نقاط الشوكة طويلة جدًا. هذه الشوكة 1 في الرسم التخطيطي. بعد أن وجدنا أول حالتين لمجموعتنا، فإننا نبحث عن حالات أخرى. في هذه الحالة البسيطة، تبقى ثلاثة حيوانات. الحيوان الذي تم اختياره ليكون جزءًا من التجمع هو الأكثر تشابهًا مع أي من الحيوانات الموجودة بالفعل في التجمع. في هذه الحالة يوجد حيوان مشابه من جميع النواحي باستثناء أن بطنه أبيض. الحالتان الأخريان أقل تشابهًا (لأن إحداهما لون مختلف تمامًا والأخرى

صورة بشرية!). معامل التشابه للحيوان المختار أقل بقليل من معامل التشابه للحيوانين الأولين (لأن بطنه أبيض) وبالتالي تنقسم الشوكة (ممثلة بخط منقط) عند نقطة أقل على طول المقياس الرأسي. هذه المرحلة هي 2 في الرسم التخطيطي. بعد الإضافة إلى التجمع، ننظر مرة أخرى في الحالات المتبقية ونقيّم تشابهها مع أي من الحيوانات الثلاثة الموجودة بالفعل في التجمع. يوجد حيوان واحد مشابه إلى حد ما للحيوان الذي تمت إضافته للتو إلى التجمع أو العنقود. على الرغم من اختلاف لونه، إلا أنه يتميز بنفس النمط المميز على بطنه. لذلك، يضاف هذا الحيوان إلى العنقود على أساس تشابهه مع الحيوان الثالث في العنقود (على الرغم من أنه يختلف نسبيًا عن الحيوانين الآخرين). هذا يمثل المرحلة 3 في الرسم التخطيطي. أخيرًا، هناك حيوان واحد متبقي (صورة الإنسان) يختلف عن جميع الحيوانات الموجودة في العنقود، لذلك سيتم دمجه في النهاية في العنقود، لكن درجة التشابه الخاصة به ستكون منخفضة جدًا، في المرحلة 4.

هناك عدة نقاط مهمة هنا. الأول هو أن العملية هرمية. لذلك، ستعتمد النتائج التي نحصل عليها إلى حد كبير على الحالتين اللتين اخترناهما كنقطة انطلاق. ثانيًا، لا تحتاج الحالات في التجمع أو العنقود إلا إلى حالة واحدة أخرى في التجمع أو العنقود، وبالتالي، يمكن تقديم قدر كبير من الاختلاف بين الحالات عبر سلسلة من التحديدات. أخيرًا، يُعرف المخطط الذي رسمناه لربط الحالات باسم مخطط الشجرة dendrogram (أو tree diagram). يكون ناتج التحليل التجميعي (العنقودي) في شكل مثل هذا النوع من المخططات.

الرابطة الكاملة Complete Linkage أو CLINK (الجار الأبعد):

يُعرف الاختلاف في طريقة الرابطة المنفردة (الجار الأقرب) بالرابطة الكاملة (أو الجار الأبعد). هذه الطريقة هي النقيض المنطقي للرابطة المنفردة. للبدء بالإجراء هو نفس الارتباط البسيط حيث نبحث في البداية عن الحالتين الأكثر تشابهًا (من حيث ارتباطهما أو متوسط المسافة الإقليدية). هاتان الحالتان (A & B) تشكلان نواة التجمع أو العنقود. الخطوة الثانية هي حيث يكون الاختلاف في الطريقة واضحًا. بدلاً من البحث عن حالة جديدة تشبه أيًا من A أو B، فإننا نبحث عن حالة بها أعلى درجة تشابه لكل من A وB . تمت إضافة الحالة C ذات أعلى تشابه مع كل من A و B إلى التجمع. الحالة التالية التي سيتم إضافتها إلى التجمع هي الحالة ذات أعلى تشابه مع A و B و C. تقلل هذه الطريقة من الاختلاف داخل التجمع أو العنقود لأنها تستند إلى التشابه العام مع أعضاء التجمع (بدلاً من التشابه مع عضو واحد في التجمع). ومع ذلك، ستظل النتائج تعتمد إلى حد كبير على الحالتين اللتين تعتبرهما نقطة البداية.

رابطة المتوسط Average Linkage (بين المجموعات Between-Group):

هذه الطريقة هي شكل آخر من أشكال الرابطة البسيطة. مرة أخرى، نبدأ بإيجاد

الحالتين الأكثر تشابهًا (بناءً على ارتباطهما أو متوسط المسافة الإقليدية). هاتان الحالتان (A & B) تشكلان نواة التجمع. في هذه المرحلة يتم حساب متوسط التشابه داخل التجمع.. ولتحديد الحالة C التي تمت إضافتها إلى التجمع، نقارن تشابه كل حالة متبقية بمتوسط تشابه التجمع. الحالة التالية التي سيتم إضافتها إلى التجمع هي الحالة ذات أعلى تشابه مع متوسط قيمة التشابه للتجمع. وبمجرد إضافة هذه الحالة الثالثة، يتم إعادة حساب متوسط التشابه داخل التجمع. الحالة التالية (D) التي سيتم إضافتها إلى التجمع هي الحالة الأكثر تشابهًا مع هذه القيمة الجديدة لمتوسط التشابه.

طريقة وارد: Ward's Method

تعتمد جميع طرق الربط على مبدأ مشابه: هناك سلسلة من التشابه تؤدي إلى ما إذا كانت الحالة قد تمت إضافتها إلى تجمع أم لا. وتختلف القواعد التي تحكم هذه السلسلة من طريقة ربط إلى أخرى. هناك طريقة مختلفة وهي طريقة وارد Ward، وهي أكثر تعقيدًا بكثير من طريقة الربط البسيطة. الهدف من طريقة وارد هو ضم الحالات إلى تجمعات بحيث يتم تقليل التباين داخل التجمع. للقيام بذلك، تبدأ كل حالة كتجمع خاص بها. ثم يتم دمج التجمعات بطريقة تقلل التباين داخل التجمع. لنكون أكثر دقة، يتم دمج تجمعين إذا أدى هذا الاندماج إلى الحد الأدنى من الزيادة في مجموع الخطأ للمربعات. في الأساس، هذا يعني أنه في كل مرحلة يتم قياس متوسط تشابه التجمع. يتم حساب الفرق بين كل حالة داخل التجمع ومتوسط التشابه وتربيعه (تمامًا مثل حساب الانحراف المعياري). ويستخدم مجموع الانحرافات التربيعية كمقياس للخطأ داخل التجمع. ويتم تحديد الحالات للدخول إلى التجمع إذا كانت الحالة التي ينتج عنها إدراجها في التجمع أقل زيادة في الخطأ (كما تم قياسه بمجموع الانحرافات التربيعية).

حدود التحليل التجميعي:

هناك العديد من الأشياء التي يجب أن تكون على دراية بها عند إجراء التحليل التجميعي (العنقودي)، منها:

1. عادةً ما تعطي الطرق المختلفة للتجميع نتائج مختلفة جدًا. يحدث هذا بسبب المعيار المختلف لدمج المجموعات (بما في ذلك الحالات). من المهم أن تفكر مليًا في الطريقة الأفضل لما تهتم بالنظر إليه.
2. باستثناء الرابطة المنفردة، ستتأثر النتائج بطريقة ترتيب المتغيرات.
3. يكون التحليل غير مستقر عند إسقاط الحالات: يحدث هذا لأن اختيار حالة (أو دمج التجمعات) يعتمد على تشابه حالة واحدة مع التجمع. يمكن أن يؤثر إسقاط حالة واحدة بشكل كبير على المسار الذي يتقدم فيه التحليل.
4. الطبيعة الهرمية للتحليل تعني أنه لا يمكن تصحيح "الأحكام السيئة" المبكرة.

التحليل التجميعي باستخدام برنامج SPSS:

سوف نتمسك بمثال أساسي للغاية. تخيل أننا أردنا النظر في مجموعات من الحالات المحالة للعلاج النفسي. قمنا بقياس كل عميل على أربعة استبيانات: قائمة سبيلبرجر لقلق السمة Spielberger Trait Anxiety Inventory (STAI)، قائمة بيك للاكتئاب Beck Depression Inventory (BDI)، مقياس الأفكار المتطفلة والاجترار (Intrusive Thoughts and Rumination: IT) ومقياس الأفكار والتصرفات الاندفاعية (Impulsive Thoughts and Actions: Impulse). الأساس المنطقي وراء هذا التحليل هو أن الأشخاص الذين يعانون من نفس الاضطراب يجب أن يبلغوا عن نمط مماثل من الدرجات عبر المقاييس (لذلك يجب أن تكون ملامح استجاباتهم متشابهة). للتحقق من التحليل، طلبنا من اثنين من علماء النفس المدربين الموافقة على التشخيص بناءً على الدليل التشخيصي والاحصائي للاضطرابات النفسية – الإصدار الرابع DSM-IV. هذه البيانات موجودة في الجدول 1، وفي بروفيل التشخيص.

جدول 1. بيانات العملاء على أربعة استبيانات: قائمة سبيلبرجر لقلق السمة، قائمة بيك للاكتئاب، مقياس الأفكار المتطفلة والاجترار ومقياس الأفكار والتصرفات الاندفاعية

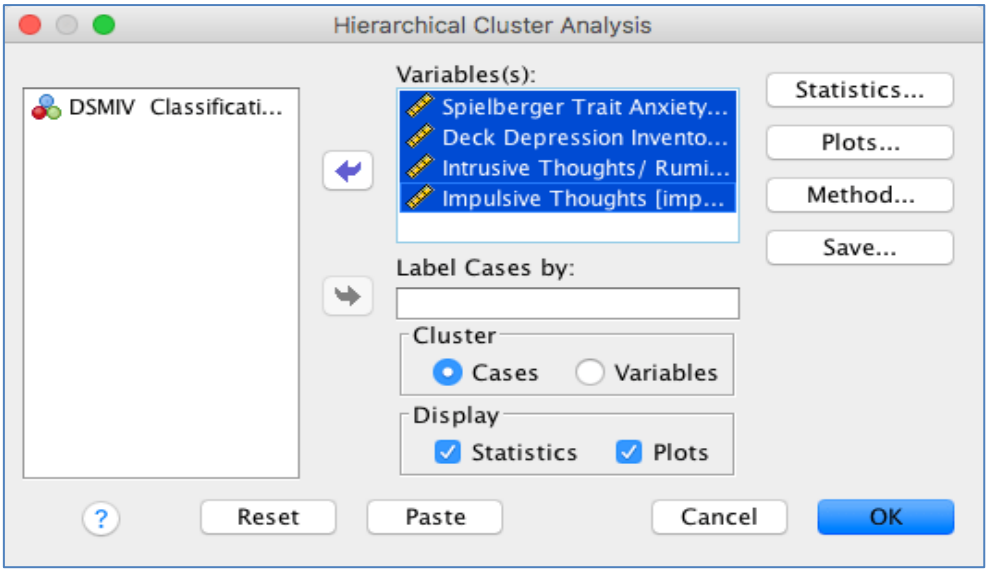
Table 1

dsm	stai	bdi	it	impulse
GAD	74	30	20	10
Depression	50	70	23	5
OCD	70	5	58	29
GAD	76	35	23	12
OCD	68	23	66	37
OCD	62	8	59	39
GAD	71	35	27	17
OCD	67	12	65	35
Depression	35	60	15	8
Depression	33	58	11	16
GAD	80	36	30	16
Depression	30	62	9	13
GAD	65	38	17	10
OCD	78	15	70	40
Depression	40	55	10	2

أول شيء يجب ملاحظته هو أنه مثل التحليل العاملي والانحدار، يتم وضع بيانات كل متغير في عمود منفصل. لذلك، يمثل كل صف في محرر البيانات بيانات شخص واحد.

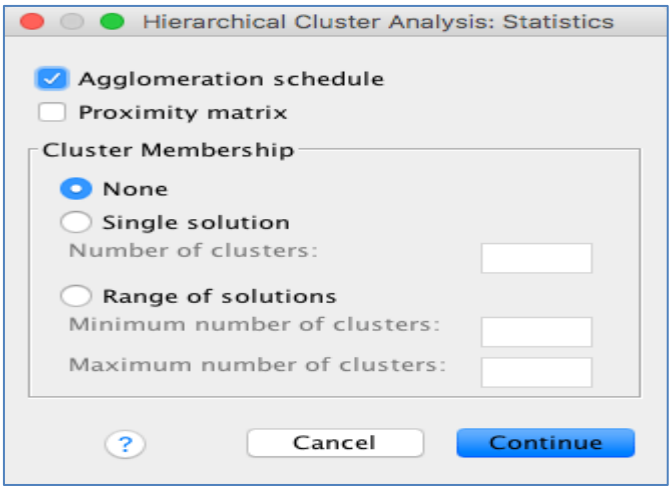
إجراء التحليل التجميعي:

يوضح الشكل 4 مربع الحوار الرئيسي لإجراء التحليل التجميعي. يتم الحصول على مربع الحوار هذا باستخدام مسار القائمة *Analyze-> Classify-> Hierarchical Cluster*. حدد الاستبيانات التشخيصية الأربعة من القائمة الموجودة على الجانب الأيسر واسحبها إلى المربع المسمى *المتغيرات*. يتم تضمين المتغير DSM في محرر البيانات فقط كطريقة للمساعدة في توضيح ما تعنيه مخرجات التحليل العنقودي، وبالتالي، لا نحتاج إلى تضمينها في التحليل.



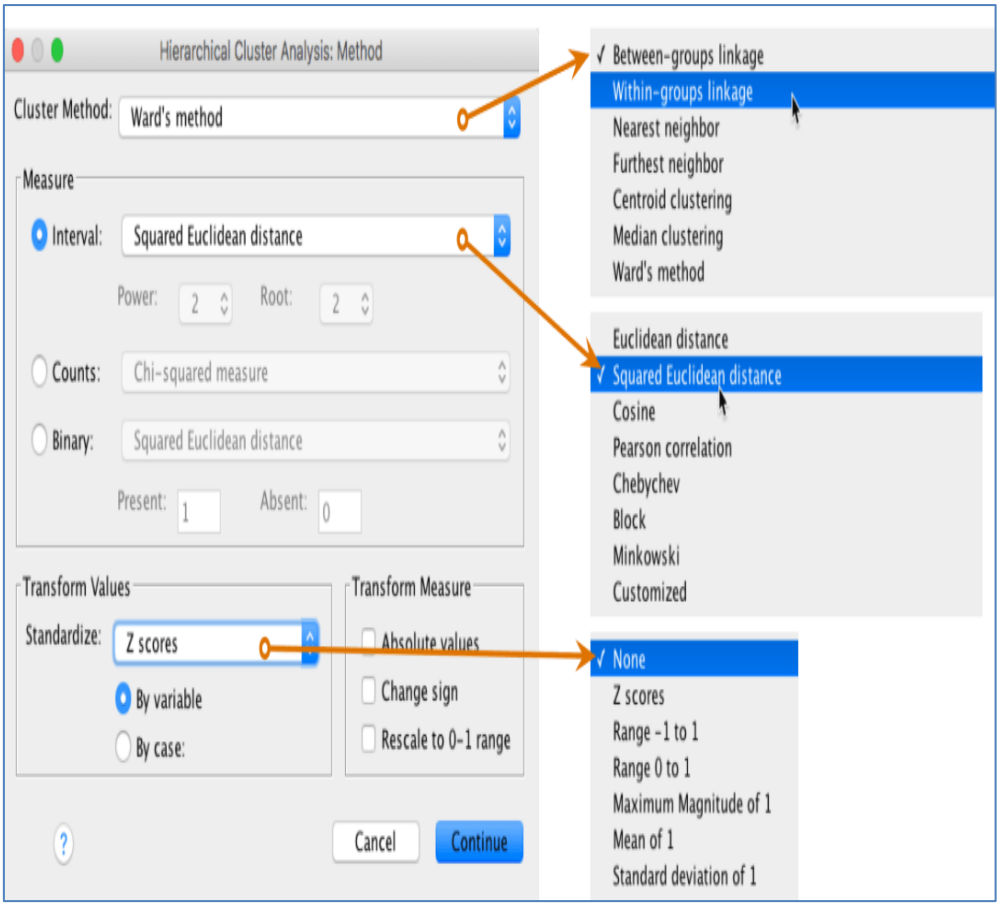
شكل 4. مربع الحوار الرئيسي لإجراء التحليل التجميعي

إذا قمت بالنقر فوق *الإحصائيات* Statistics في مربع الحوار الرئيسي، فسيظهر مربع حوار آخر (انظر الشكل 5). يتمثل الاستخدام الرئيسي لمربع الحوار هذا في تحديد عدد محدد من التجمعات. بشكل افتراضي، سيقوم برنامج SPSS ببساطة بدمج جميع الحالات في مجموعة واحدة ويتعين على الباحث فحص المخرجات لتحديد التجميعات الفرعية الموضوعية. ومع ذلك، إذا كانت لديك فرضية حول عدد التجمعات التي يجب أن تظهر، فيمكنك إخبار SPSS بإنشاء عدد محدد من التجمعات، أو إنشاء عدد من التجمعات داخل نطاق. في هذا المثال، اترك الخيارات الافتراضية كما هي وانتقل مرة أخرى إلى مربع الحوار الرئيسي بالنقر فوق *متابعة*.



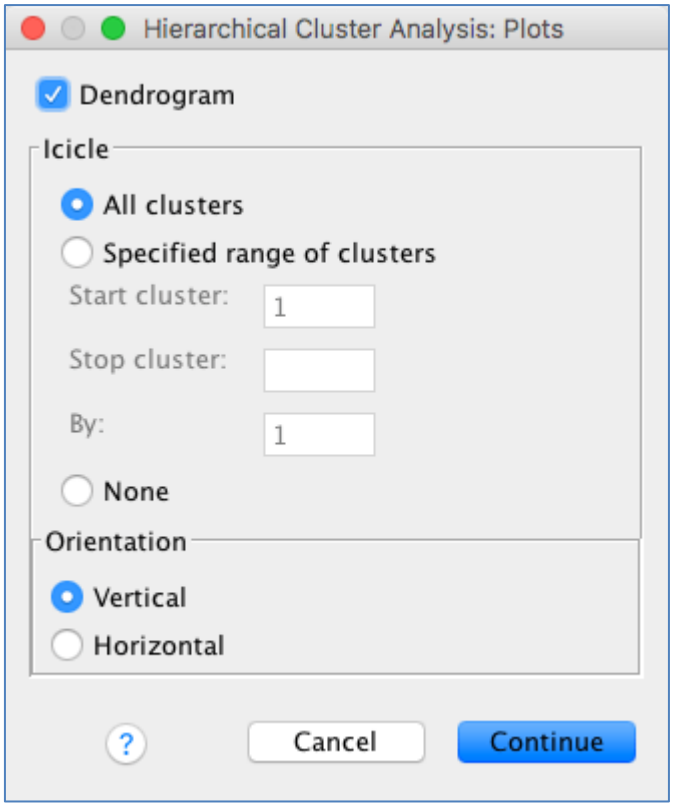
شكل 5. مربع حوار *الإحصائيات* Statistics المطلوبة للتحليل التجميعي

انقر على *الطريقة Method* ..... للوصول إلى مربع الحوار في الشكل 6. يمكنك استخدام مربع الحوار هذا لاختيار طريقة إنشاء المجموعات (تم وصف بعضها أعلاه). بشكل افتراضي، يستخدم SPSS طرق الربط بين المجموعات (أو الارتباط المتوسط). ومع ذلك، هناك العديد من الخيارات الأخرى المتاحة (على سبيل المثال، الجار الأقرب، الجار الأبعد وطريقة وارد). يمكن تحديد كل طريقة من خلال النقر على السهم لأسفل حيث يشير إلى طريقة التجميع العنقودي. بالنسبة لهذا التحليل، أقترح اختيار طريقة وارد، ولكن كممارسة أقترح العودة وتجربة بعض الطرق المختلفة: ستجد أنك تحصل على نتائج مختلفة تمامًا! أسفل اختيار الطريقة، هناك سلسلة من الخيارات اعتمادًا على ما إذا كنت تقوم بتحليل بيانات مقياس الفترة (كما نحن هنا)، أو بيانات التكرار (الأعداد) أو البيانات الثنائية (المتغيرات ثنائية التفرع مع استجابتين محتملتين فقط). يحتوي كل نوع من أنواع البيانات هذه على مجموعة مرتبطة من مقاييس التشابه. وصفت سابقًا المسافات الإقليدية ومعامل الارتباط. بشكل افتراضي، يستخدم برنامج SPSS المسافات الإقليدية (وهو خيار جيد للاستخدام). ومع ذلك، يمكنك اختيار مقياس مختلف للتشابه إذا لزم الأمر؛ حيث يقدم روميزبرج (Romesberg, 1984) وأيفرت (Everett, 1993) مزيدًا من التفاصيل عن الطرق الممكنة. أخيرًا، يوجد في أسفل مربع الحوار خيار توحيد بياناتنا. لقد ذكرت سابقًا أن توحيد البيانات فكرة جيدة (خاصة وأن بعض مقاييس التشابه حساسة للاختلافات في تباين المتغيرات) لذلك أوصي بهذا الخيار. وهناك عدد من الطرق التي يمكن من خلالها توحيد البيانات ولكن الأسهل فهمًا هو التحويل إلى الدرجة المعيارية Z؛ حيث يوصى بهذا الخيار. من الممكن التوحيد إما عن طريق المتغير، أو عبر حالة معينة. عند تجميع الحالات (كما نفعل هنا، المعروف باسم تحليل Q)، يجب علينا توحيد المتغيرات. إذا كنا نحاول تجميع المتغيرات (تحليل R)، فسنحتاج إلى توحيد المعايير عبر الحالات. لذلك، في هذا المثال، حدد الدرجات المعيارية Z-scores للمتغيرات وتابع بالنقر فوق متابعة Continue.



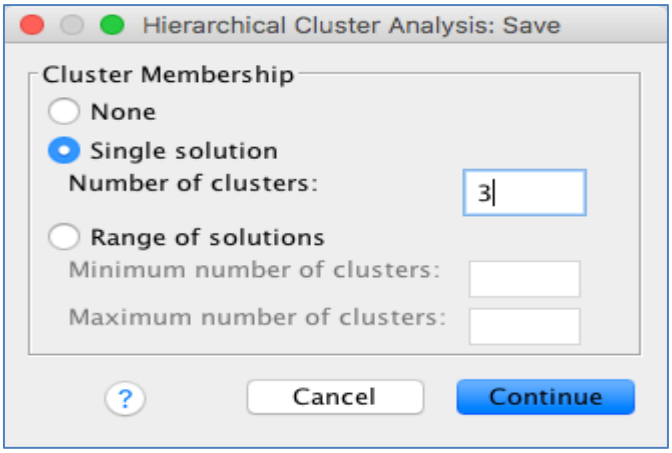
شكل 6. مربع حوار توصيف *الطريقة* المطلوبة للتحليل التجميعي

بمجرد العودة إلى مربع الحوار الرئيسي، يمكنك تحديد مربع حوار التمثيل البياني بالنقر فوق Plots. هناك نوعان من المخططات التي يمكنك طلبها من التحليل العنقودي. الخيار الافتراضي هو المخطط الجليدي icicle plot، ولكن الأكثر فائدة لأغراض التفسير هو المخطط الشجري. يوضح لنا المخطط الشجري dendrogram ويسمى أيضاً مخطط أسنان الشوكة (أو الروابط) بين الحالات وبنيتها التي تعطينا أدلة على الحالات التي تشكل تجمعات متماسكة. لذلك، من الضروري طلب هذا الخيار. بمجرد تحديد هذا الخيار، انقر فوق *متابعة*.



شكل 7. مربع حوار توصيف *المخططات الرسومية* المطلوبة للتحليل التجميعي

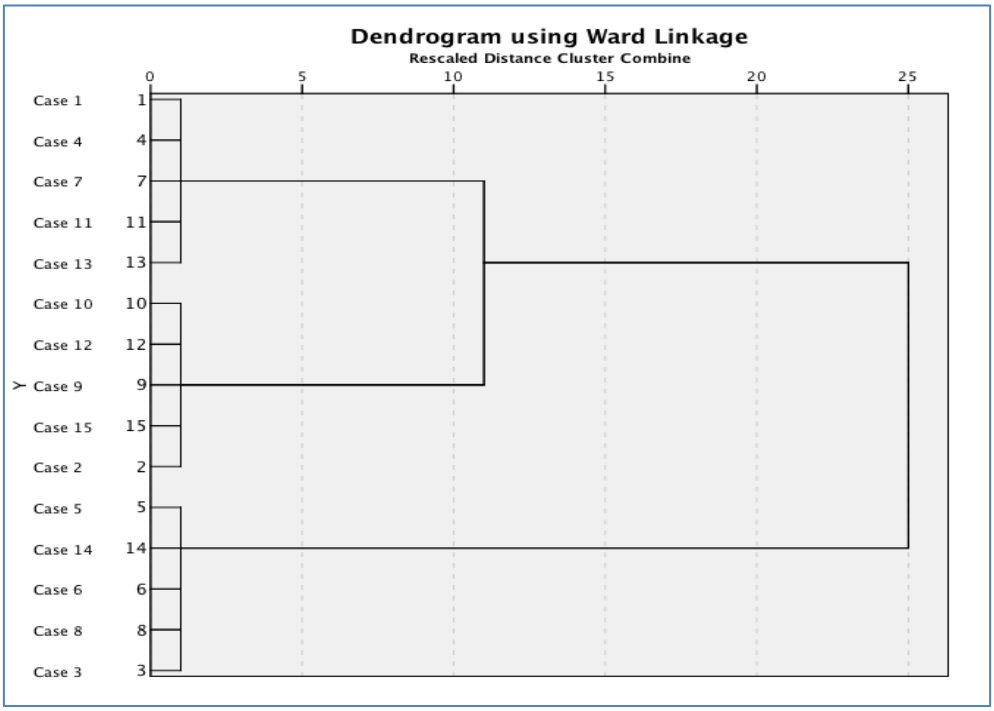
بمجرد العودة إلى مربع الحوار الرئيسي، يمكنك تحديد مربع حوار الحفظ بالنقر فوق *حفظ* Save. يتيح لنا مربع الحوار هذا حفظ متغير جديد في محرر البيانات يحتوي على قيمة ترميز تمثل العضوية في العنقود. على هذا النحو، يمكننا استخدام هذا المتغير لإخبارنا بالحالات التي تقع في نفس المجموعات. افتراضيًا، لا يُنشئ SPSS هذا المتغير. في هذا المثال، نتوقع ثلاث مجموعات من الأشخاص بناءً على تصنيفات الدليل التشخيصي والإحصائي للاضطرابات النفسية الإصدار الرابع DSM-IV (القلق العام GAD، والاكتئاب Depression، والوسواس القهري OCD) حتى نتمكن من تحديد حل واحد ثم كتابة 3 في المساحة الفارغة (انظر الشكل 8). في الواقع، ما سنفعله عادةً هو تشغيل التحليل العنقودي دون تحديد هذا الخيار ثم فحص مخطط الأسنان الشوكة الناتج لتحديد عدد المجموعات الموضوعية الموجودة داخل البيانات. بعد القيام بذلك، يمكننا إعادة إجراء التحليل، ونطلب من SPSS حفظ قيم الترميز لعدد المجموعات التي حددناها.



شكل 8. تحديد عدد الحلول في التحليل التجميعي

نواتج SPSS: مخطط أسنان الشوكة Dendrogram

الجزء الرئيسي من نواتج SPSS هو المخطط الشجري أو مخطط أسنان الشوكة (على الرغم من المفارقة أن هذا الرسم البياني يظهر فقط إذا تم تحديد خيار خاص). يتم تقديم مخطط أسنان الشوكة dendrogram لبيانات التشخيص في الناتج 1.



شكل 9. مخطط أسنان الشوكة dendrogram لبيانات التشخيص في الناتج 1

كما أوضحنا سابقًا، يعمل التحليل التجميعي صعودًا لوضع كل حالة في مجموعة واحدة. لذلك، ينتهي بنا الأمر بشوكة واحدة تنقسم عند مستويات أقل من التشابه. بالنسبة إلى هذه البيانات، يتم تقسيم الشوكة أولاً إلى حالات منفصلة 1 و 4 و 7 و 11 و 13 و 10 و 12 و 9 و 15 و 2 عن الحالات 5 و 14 و 6 و 8 و 3.

في الواقع، إذا نظرت إلى تصنيف الدليل التشخيصي والإحصائي للاضطرابات النفسية – الإصدار الرابع DSM-IV لهذه الحالات، فقد قسم هذا الفاصل الأول مرضى اضطراب القلق العام GAD والاكتئاب عن الوسواس القهري. من المحتمل أن يكون هذا قد حدث لأن كلا من مرضى اضطراب القلق العام والاكتئاب لديهم درجات منخفضة في الأفكار المتطفلة والأفكار والأفعال المندفعة، في حين أن أولئك الذين يعانون من الوسواس القهري يسجلون درجات عالية في كلا المقياسين. القسم الرئيسي الثاني هو تقسيم فرع واحد من هذه الشوكة الأولى إلى تجمعين أخرين. يفصل هذا التقسيم الحالات 1 و 4 و 7 و 11 و 13 عن 10 و 12 و 9 و 15 و 2. بالنظر إلى تصنيف الدليل التشخيصي والإحصائي للاضطرابات النفسية DSM، فقد فصل هذا التقسيم الثاني القلق العام GAD عن الاكتئاب. باختصار، كشف التحليل النهائي عن ثلاث مجموعات رئيسية، والتي يبدو أنها مرتبطة بالتصنيفات الناشئة عن الدليل التشخيصي والإحصائي للاضطرابات النفسية. على هذا النحو، يمكننا القول أن استخدام قائمة سبيلبرجر لقلق السمة STAI وقائمة بيك للاكتئاب BDI، ومقياس الأفكار المتطفلة والاجترار (Intrusive Thoughts and Rumination: IT) ومقياس الأفكار والتصرفات الاندفاعية (Impulsive Thoughts and Actions: Impulse) كمقاييس تشخيصية هي طريقة دقيقة لتصنيف هذه التجمعات الثلاث من المرضى (وربما تستغرق وقتًا أقل من التشخيص الكامل باستخدام الدليل التشخيصي والاحصائي للاضطرابات النفسية – الإصدار الرابع). من الواضح أن هذه البيانات مبسطة إلى حد ما وقد أدت إلى حل غير معقد للغاية. في الواقع، هناك الكثير من الذاتية التي ينطوي عليها تحديد التجمعات الموضوعية.

بعد فحص مخطط أسنان الشوكة وتحديد عدد التجمعات الموجودة، من الممكن إعادة إجراء التحليل بمطالبة SPSS بحفظ متغير جديد يتم فيه تعيين رموز التجمع للحالات (مع قيام الباحث بتحديد عدد التجمعات في البيانات). وبالنسبة لهذه البيانات، رأينا ثلاث تجمعات واضحة وبالتالي يمكننا إعادة إجراء التحليل وطلب ترميز مجموعة التجمعات لثلاث تجمعات (في الواقع، لقد أخبرتك أن تفعل ذلك كجزء من التحليل الأصلي).

يظهر الناتج 2 أدناه الرموز الناتجة لكل حالة في هذا التحليل. من الواضح تمامًا أن هذه الرموز تتطابق تمامًا مع تصنيفات الدليل التشخيصي والإحصائي للاضطرابات النفسية – الإصدار الرابع. على الرغم من أن هذا المثال بسيط للغاية، إلا أنه يوضح

لك كيف يمكن أن يكون التحليل العنقودي مفيدًا في تطوير أدوات التشخيص والتحقق من صدقها، أو في إنشاء مجموعات طبيعية من الأعراض لاضطرابات معينة. ويوضح شكل 10 مقارنة بين تصيف الدليل التشخيصي والإحصائي للاضطرابات النفسية – الإصدار الرابع والتصنيف الناتج عن التحليل التجميعي.

Case Summaries[a]

	DSMIV Classification	Ward Method
1	GAD	1
2	Depression	2
3	OCD	3
4	GAD	1
5	OCD	3
6	OCD	3
7	GAD	1
8	OCD	3
9	Depression	2
10	Depression	2
11	GAD	1
12	Depression	2
13	GAD	1
14	OCD	3
15	Depression	2
Total N	15	15

a. Limited to first 100 cases.

شكل 10. مقارنة بين تصيف الدليل التشخيصي والإحصائي للاضطرابات النفسية – الإصدار الرابع والتصنيف الناتج عن التحليل التجميعي.

تطبيقات التحليل التجميعي:

لم يكن ضمن قراءاتي بيانات بحثية كان التحليل التجميعي / العنقودي Cluster Analysis فيها أسلوبًا يُعتقد أنه مناسب لتحليل البيانات، إلى أن وجدت مقالاً لكارل وينيش (Wuensch, 2016) يتناول هذا النوع من التحليل لمجرد التسلية، حيث أجرى تحليلاً تجميعياً لملف بيانات لحالات أعضاء هيئة التدريس في قسم علم النفس بجامعة إيست كارولينا في شهر نوفمبر 2005، يشمل المتغيرات التالية:

- Name الاسم - على الرغم من أن رواتب أعضاء هيئة التدريس هي معلومات عامة بموجب قانون ولاية كارولينا الشمالية، إلا أنني على الرغم من ذلك فضلت تخصيص اسم وهمي لكل حالة.

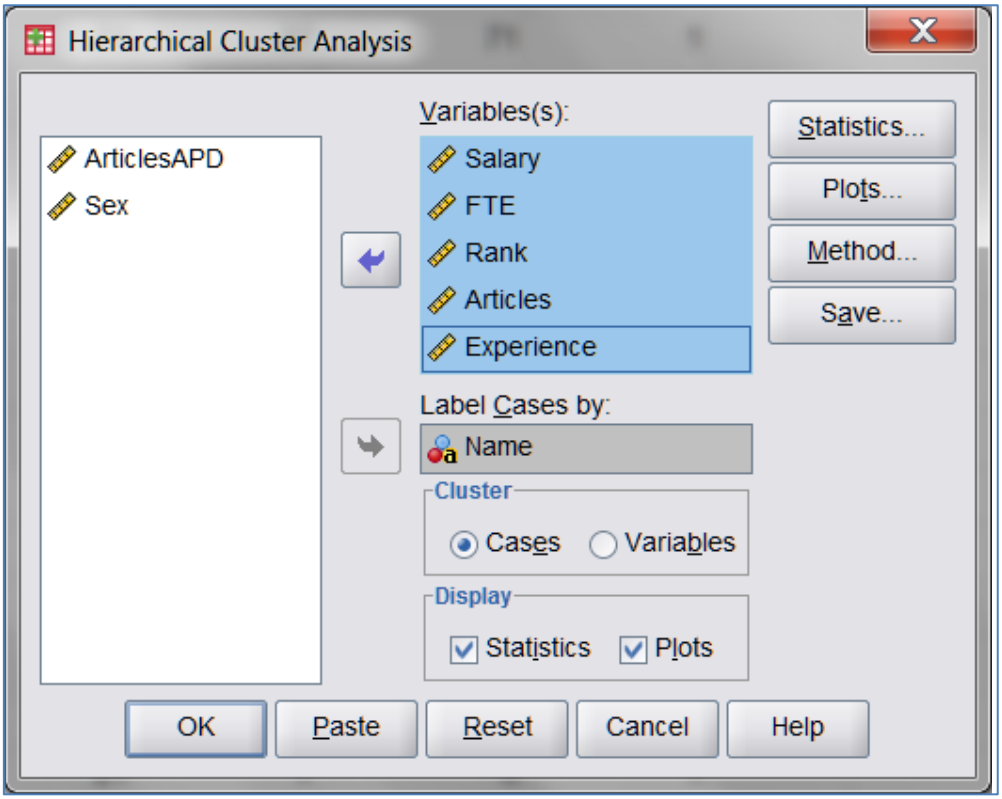
- Salary الراتب - الراتب السنوي بالدولار، من تقرير الجامعة.
- FTE - عبء العمل المكافئ لعضو هيئة التدريس بدوام كامل.
- Rank الرتبة - حيث 1 = هيئة معاونة (معيد ومدرس مساعد)، 2 = أستاذ زائر، 3 = أستاذ مساعد، 4 = أستاذ مشارك، 5 = أستاذ
- Articles المقالات - عدد المقالات العلمية المنشورة، باستثناء أشياء مثل التعليقات في النشرات الإخبارية، والملخصات في الاجتماعات، وما شابه ذلك. كان المصدر الأساسي لهذه البيانات هو حياة عضو هيئة التدريس عبر الإنترنت. وعندما لم يكن ذلك متاحًا، تم استخدام البيانات الموجودة في قاعدة بيانات المنشورات الأكاديمية بالجامعة، بعد إزالة المداخلات المكررة.
- Experience الخبرة - عدد سنوات العمل كعضو هيئة تدريس بدوام كامل في قسم علم النفس. إذا لم يكن لدى عضو هيئة التدريس معلومات التوظيف على صفحة الويب الخاصة به، فقد تم استخدام مصادر أخرى عبر الإنترنت - على سبيل المثال، من قاعدة بيانات المنشورات يمكنني تقدير عام أول وظيفة على أنها سنة النشر الأول.
- كان ضمن ملف البيانات ولكن لم يتم استخدامها في تحليل التجميعي أيضا:
- ArticlesAPD - عدد المقالات المنشورة كما هو مدرج في قاعدة بيانات المنشورات الأكاديمية بالجامعة. كان هناك الكثير من الأخطاء في قاعدة البيانات هذه، لكنني حاولت تصحيحها (على سبيل المثال، عن طريق ضبط المدخلات المكررة).
- Sex الجنس - لقد استنتجت الجنس البيولوجي من المظهر الجسدي.

وبناء على القراءة المتعمقة لهذا المقال، وما لدينا من خبرات في مجال التحليل الإحصائي فقد قمت بمحاكاة هذا النوع من التحليل وإثرائه بالمزيد من الأشكال والجداول والمعلومات والتعليقات عليها، وفيما يلي عرض لكيفية إجراء هذا التحليل سواء لتجميع الحالات أو تجميع المتغيرات، وعرضه نتائجه في البحوث النفسية والتربوية والاجتماعية:

أولاً: إجراء التحليل التجميعي للحالات:

ابدأ بجلب ملف البيانات ClusterAnonFaculty.sav في SPSS، وهو متوفر على الرابط: https://core.ecu.edu/wuenschk/SPSS/SPSS-Data.htm. الآن انقر فوق تحليل Analyze، تصنيف Classify، تجميع هرمي Hierarchical Cluster. حدد الاسم باعتباره المتغير الذي يمكن من خلاله تسمية الحالات label cases والراتب Salary، و عبء العمل المكافئ FTE، والرتبة الأكاديمية Rank، وعدد المقالات Articles، والخبرة Experience

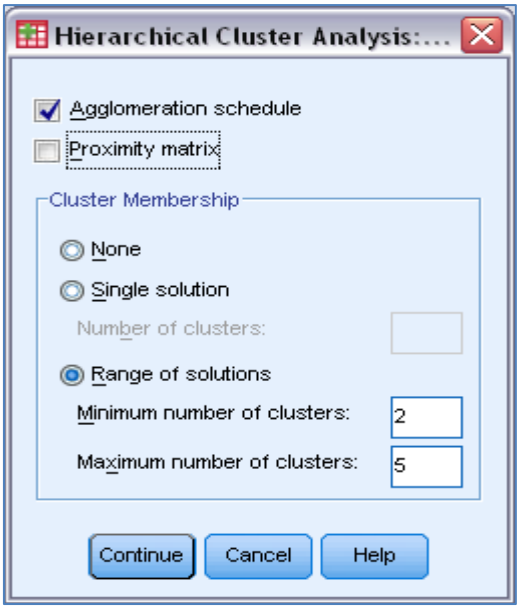
كمتغيرات. أشر إلى أنك تريد تجميع الحالات cluster cases بدلاً من المتغيرات cluster variables وتريد عرض كل من الإحصائيات statistics والمخططات الرسومية plots. كما في الشكل 11.



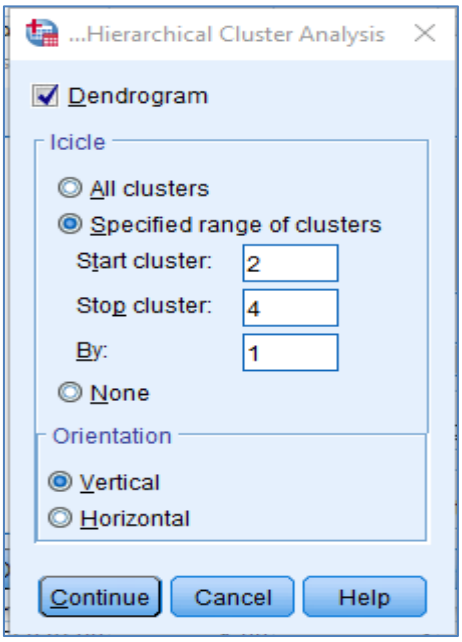
شكل 11. إدخال المتغيرات وتسمية الحالات ضمن التحليل التجميعي الهرمي

قد ترغب في فتح الإخراج في علامة تبويب جديدة أثناء قراءة هذا المستند. تم إدراج أجزاء من المخرجات في هذا المستند.

انقر فوق إحصائيات وقم بالإشارة إلى أنك تريد رؤية جدول تجميعي مع حلول التجميع 2 و 3 و 4 و 5 Agglomeration schedule with 2, 3, 4, and 5 cluster solutions. انقر فوق متابعة. انقر فوق Plots وقم بالإشارة إلى أنك تريد مخطط Dendogram ومخطط جليد عمودي مع حلول عنقودية (تجميعية) 2 و 3 و 4 vertical Icicle plot with 2, 3, and 4 cluster solutions. انقر فوق متابعة Continue. ويوضح الشكلان 12، و13 هذه المدخلات.



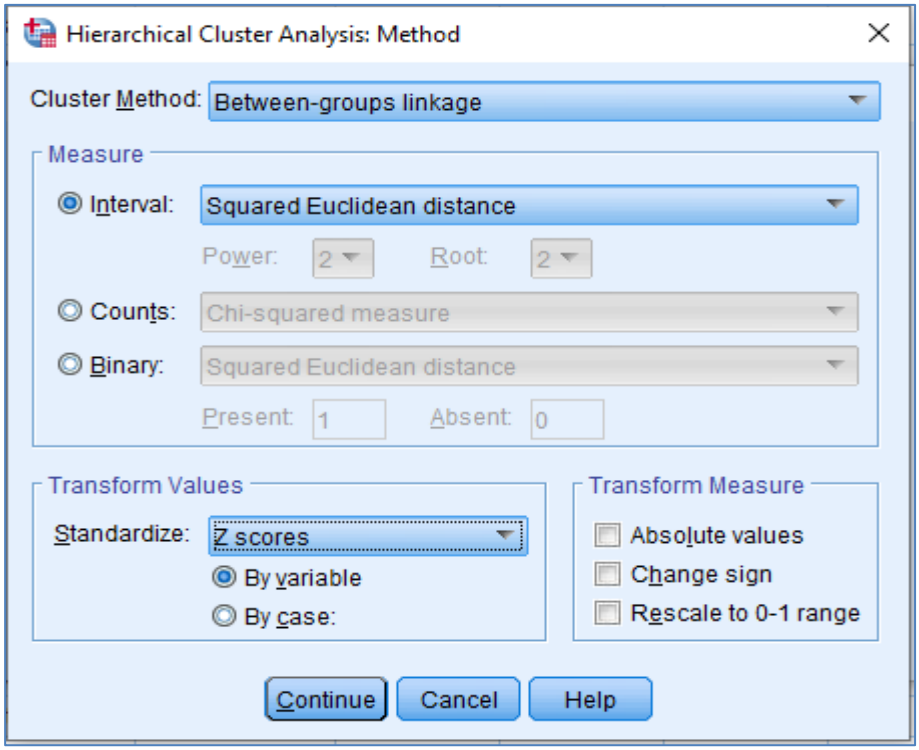
شكل 12. مدخلات جدولة التجميعات وعدد الحلول المطلوبة



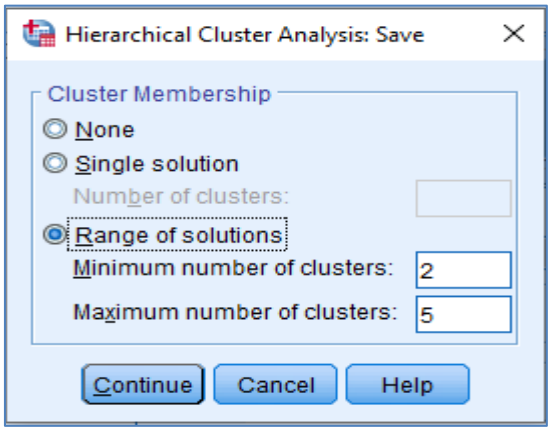
شكل 13. مدخلات المخططات الرسومية وعدد الحلول المطلوبة وتحديد اتجاه الرسم

انقر فوق الطريقة وقم بالإشارة إلى أنك تريد استخدام طريقة الربط بين المجموعات للتجميع Between-groups linkage method of clustering، والمسافات الإقليدية التربيعية quared Euclidian distances، والمتغيرات

المعيارية بدرجات variables standardized to z scores z (بحيث يساهم كل متغير بالتساوي). انقر فوق متابعة. ثم انقر فوق حفظ وقم بالإشارة إلى أنك تريد حفظ Save، لكل حالة، التجميع الذي تم تعيين الحالة له من أجل حلول التجمع 2 و 3 و 4. انقر فوق متابعة، ثم موافق. ويوضح الشكلين 14، و15.



شكل 14. توصيف طريقة التجميع الهرمي



شكل 15. توصيف عضوية التجميع ومدى عدد الحلول المطلوبة

يبدأ برنامج SPSS بمعايرة جميع المتغيرات لتكون بمتوسط قدره صفر،

وتباين قدره 1. ينتج عن هذا أن جميع المتغيرات تكون على نفس المقياس ويتم ترجيحها بالتساوي؛ أي يكون لها نفس الوزن ووحدة القياس.

يعرض برنامج SPSS الصيغة المستخدمة كالتالي:

```
DATASET DECLARE D0.2297918118361889.
PROXIMITIES   Salary FTE Rank Articles Experience
  /MATRIX OUT(D0.2297918118361889)
  /VIEW=CASE
  /MEASURE=SEUCLID
  /PRINT NONE
  /ID=Name
  /STANDARDIZE=VARIABLE Z.
```

ثم يعرض برنامج SPSS النواتج بداية بجدول ملخص معالجة الحالات كالتالي:

جدول 2. ملخص معالجة الحالات بناء على المتغيرات المقاسة (ن=44)

Case Processing Summary[a]					
Cases					
Valid		Missing		Total	
N	Percent	N	Percent	N	Percent
44	100.0%	0	0.0%	44	100.0%
a. Squared Euclidean Distance used					

ثم يعرض برنامج SPSS صيغة سيناريو التحليل الذي تم تحديده في الأشكال من 12 إلى 15 كالتالي:

```
CLUSTER
  /MATRIX IN(D0.2297918118361889)
  /METHOD BAVERAGE
  /ID=Name
  /PRINT SCHEDULE CLUSTER(2,5)
  /PLOT DENDROGRAM VICICLE(2,4,1)
  /SAVE CLUSTER(2,4).
```

في الخطوة الأولى، يحسب برنامج SPSS لكل زوج من الحالات المسافة الإقليدية المربعة بين الحالات. هذا بكل بساطة $\sum_{i=1}^{v}(X_i - Y_i)^2$، المجموع عبر المتغيرات (من I =1 إلى 7) للفرق التربيعي بين الدرجة على المتغير i للحالة الواحدة (Xi) والنتيجة على المتغير i للحالة الأخرى (Yi). يتم تحديد الحالتين اللتين

تفصلهما أصغر مسافة إقليدية ثم يتم تصنيفهما معًا في التجمع الأول. في هذه المرحلة، هناك مجموعة واحدة بها حالتان.

بعد ذلك، يقوم برنامج SPSS بإعادة حساب المسافات الإقليدية التربيعية بين كل كيان (حالة أو مجموعة) وكل كيان آخر. وعندما يكون أحد الكيانات المقارنة أو كليهما عبارة عن تجمع، يحسب برنامج SPSS متوسط المسافة الإقليدية التربيعية بين أعضاء الكيان الواحد وأعضاء الكيان الآخر. ويتم تصنيف الكيانين مع أصغر مسافة إقليدية مربعة معًا. ثم يعيد برنامج SPSS حساب المسافات الإقليدية التربيعية بين كل كيان وكل كيان آخر ويتم تصنيف الاثنين مع أصغر مسافة إقليدية مربعة معًا. ويستمر هذا حتى يتم تجميع جميع الحالات في مجموعة كبيرة واحدة.

انظر إلى جدول التجمعات. في الخطوة الأولى، الحالة المجمعة في SPSS 32 مع 33. المسافة الإقليدية المربعة بين هاتين الحالتين هي 0.000. في المراحل من 2 إلى 4، يُنشئ برنامج SPSS ثلاث مجموعات أخرى، تحتوي كل منها على حالتين. وفي المرحلة الخامسة، يضيف برنامج SPSS الحالة 39 إلى التجمع الذي يحتوي بالفعل على الحالتين 37 و 38. بحلول المرحلة الثالثة والأربعين، تم تجميع جميع الحالات في كيان واحد.

يعرض برنامج SPSS مراحل التحليل التجميعي وخصاص كل مرحلة كالتالي:

Cluster
Average Linkage (Between Groups)

جدول 3. مراحل التحليل التجميعي وخصاص كل مرحلة

Agglomeration Schedule						
	Cluster Combined			Stage Cluster First Appears		
Stage	Cluster 1	Cluster 2	Coefficients	Cluster 1	Cluster 2	Next Stage
1	32	33	.000	0	0	9
2	41	42	.000	0	0	6
3	43	44	.000	0	0	6
4	37	38	.000	0	0	5
5	37	39	.001	4	0	7
6	41	43	.002	2	3	27
7	36	37	.003	0	5	27
8	20	22	.007	0	0	11
9	30	32	.012	0	1	13

Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
10	21	26	.012	0	0	14
11	20	25	.031	8	0	12
12	16	20	.055	0	11	14
13	29	30	.065	0	9	26
14	16	21	.085	12	10	20
15	11	18	.093	0	0	22
16	8	9	.143	0	0	25
17	17	24	.144	0	0	20
18	13	23	.167	0	0	22
19	14	15	.232	0	0	32
20	16	17	.239	14	17	23
21	7	12	.279	0	0	28
22	11	13	.441	15	18	29
23	16	27	.451	20	0	26
24	3	10	.572	0	0	28
25	6	8	.702	0	16	36
26	16	29	.768	23	13	35
27	36	41	.858	7	6	33
28	3	7	.904	24	21	31
29	11	28	.993	22	0	30
30	5	11	1.414	0	29	34
31	3	4	1.725	28	0	36
32	14	31	1.928	19	0	34
33	36	40	2.168	27	0	40
34	5	14	2.621	30	32	35
35	5	16	2.886	34	26	37
36	3	6	3.089	31	25	38
37	5	19	4.350	35	0	39
38	1	3	4.763	0	36	41
39	5	34	5.593	37	0	42
40	35	36	8.389	0	33	43
41	1	2	8.961	38	0	42

Agglomeration Schedule						
	Cluster Combined			Stage Cluster First Appears		
Stage	Cluster 1	Cluster 2	Coefficients	Cluster 1	Cluster 2	Next Stage
42	1	5	11.055	41	39	43
43	1	35	17.237	42	40	0

ثم يعرض البرنامج جدولاً بعضوية كل تجمع أو عنقود كالتالي:

جدول 4. عضوية الحالات لكل تجمع أو عنقود ناتج عن التحليل التجميعي للحالات

Cluster Membership				
Case	5 Clusters	4 Clusters	3 Clusters	2 Clusters
1:Rosalyn	1	1	1	1
2:Lawrence	2	2	1	1
3:Sunila	1	1	1	1
4:Randolph	1	1	1	1
5:Mickey	3	3	2	1
6:Louis	1	1	1	1
7:Tony	1	1	1	1
8:Raul	1	1	1	1
9:Catalina	1	1	1	1
10:Johnson	1	1	1	1
11:Beulah	3	3	2	1
12:Martina	1	1	1	1
13:Marie	3	3	2	1
14:Ernest	3	3	2	1
15:Christopher	3	3	2	1
16:Ernie	3	3	2	1
17:Christa	3	3	2	1
18:Linette	3	3	2	1
19:Bo	3	3	2	1
20:Carla	3	3	2	1
21:Alberto	3	3	2	1
22:Christina	3	3	2	1
23:Jonah	3	3	2	1

Cluster Membership				
Case	5 Clusters	4 Clusters	3 Clusters	2 Clusters
24:Tucker	3	3	2	1
25:Shanta	3	3	2	1
26:Melissa	3	3	2	1
27:Jenna	3	3	2	1
28:Johnny	3	3	2	1
29:Cleatus	3	3	2	1
30:Jonas	3	3	2	1
31:Tad	3	3	2	1
32:Amaryllis	3	3	2	1
33:Nathan	3	3	2	1
34:Deanna	3	3	2	1
35:Willy	4	4	3	2
36:Deana	5	4	3	2
37:Dea	5	4	3	2
38:Claude	5	4	3	2
39:Amanda	5	4	3	2
40:Boris	5	4	3	2
41:Garrett	5	4	3	2
42:Stew	5	4	3	2
43:Bree	5	4	3	2
44:Karma	5	4	3	2

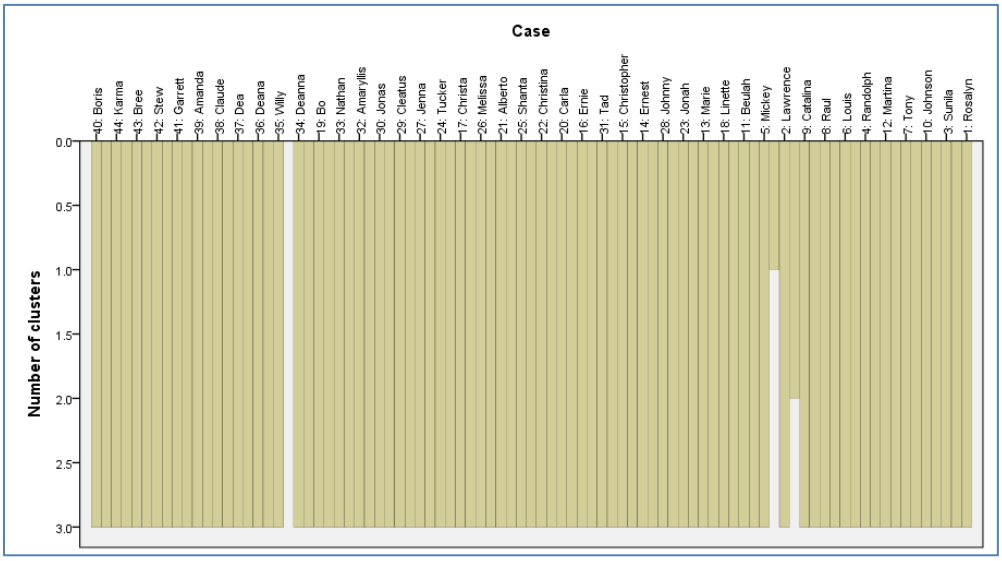
انظر إلى المخطط الجليدي العمودي Vertical Icicle. بالنسبة للحل التجميعي (العنقودي)، يمكنك أن ترى أن التجمع الواحد يتكون من عشر حالات (من Boris حتى Willy، متبوعًا بعمود أبيض). كان هؤلاء هم أعضاء هيئة التدريس الملحقة (بدوام جزئي part-time) (باستثناء حالة واحدة) ويتألف التجمع الثاني من أي شخص آخر.

بالنسبة إلى حل المجموعات الثلاثة، يمكنك أن ترى أن مجموعة أعضاء هيئة التدريس المساعدة تظل سليمة ولكن التجمع الأخر ينقسم إلى قسمين. كانت ديانا Deanna حتى ميكي Mickey هم أعضاء هيئة التدريس المبتدئين ولورانس Lawrence حتى روزالين Rosalyn أعضاء هيئة التدريس ذوي الرتبة الأعلى.

بالنسبة إلى حل المجموعات الأربع، يمكنك أن ترى أن حالة واحدة (لورانس

Lawrence) تشكل تجمع خاصة بها.

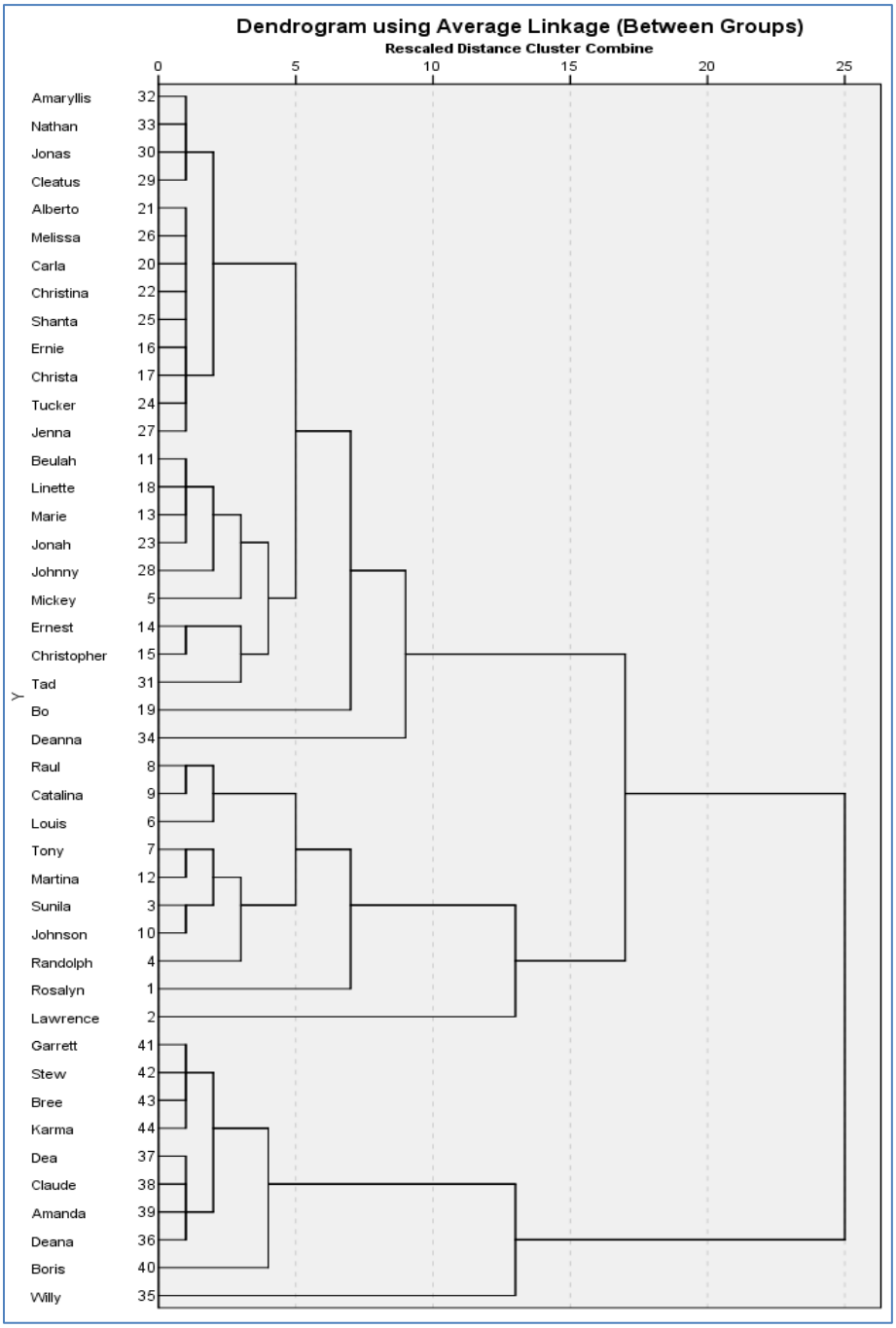
ثم يعرض للمخططات الرسومية للتحليل التجميعي بدء بالمخطط الجليدي العمودي كما في الشكل 16.



شكل 16. مخطط للحالات وعدد التجمعات الناتجة عن التحليل التجميعي الهرمي

ثم يعرض البرنامج للمخطط الشجري Dendrogram للحالة والدرجة المركبة لتجمع المسافات الموزونة في الشكل 17.

انظر إلى المخطط الشجري في شكل 17، حيث يعرض بشكل أساسي نفس المعلومات الموجودة في جدول التجمعات ولكن في شكل رسوم بيانية.



شكل 17. المخطط الشجري (مخطط أسنان الشوكة) Dendrogram للحالة والدرجة المركبة لتجمع المسافات الموزونة

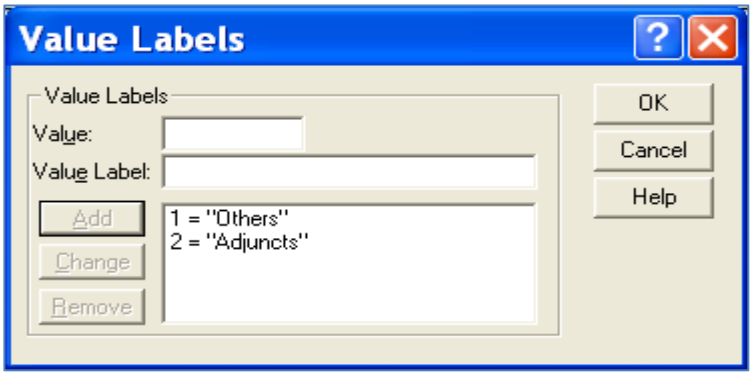
انظر للوراء في ورقة البيانات. ستجد ثلاثة متغيرات جديدة. CLU2_1 عبارة

عن عضوية مجموعة لحل المجموعات المكونة من مجموعتين، CLU3_1 لحل المجموعات الثلاث، و CLU4_1 لحل المجموعات الأربع، كما في الشكل 18.

CLU4_1	CLU3_1	CLU2_1
1	1	1
2	1	1
1	1	1
1	1	1
3	2	1
1	1	1
1	1	1
1	1	1
1	1	1
1	1	1
3	2	1
1	1	1
3	2	1
3	2	1
3	2	1
3	2	1
3	2	1
3	2	1
3	2	1

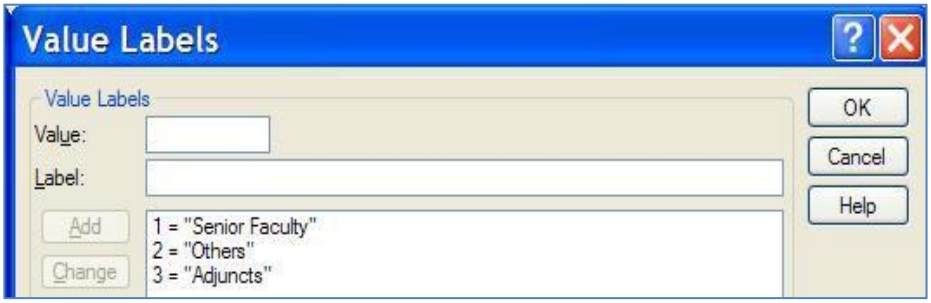
شكل 18. المتغيرات الجديدة لعضوية التجمع لحل المجموعات

قم بإزالة تسميات المتغيرات ثم قم بتسمية قيم CLU2_1، كما في الشكل 19.



شكل 19. تسمية قيم المتغير CLU2_1

قم بإزالة تسميات المتغيرات ثم قم بتسمية قيم CLU3_1، كما بالشكل 20.

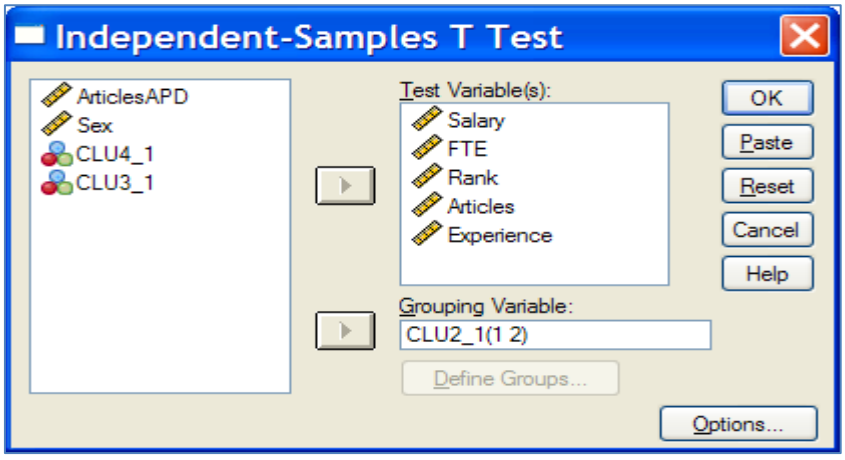


شكل 20. تسمية قيم المتغير CLU3_1

مقارنة التجمعات:

حل المجموعتين: الهيئة المعاونة مقابل الرتب الأكاديمية الأخرى Adjuncts vs others. لنرى كيف تختلف المجموعتان في حل المجموعتين عن بعضهما البعض في المتغيرات التي تم استخدامها لتجميعهما، كما في الشكل 21، باستخدام الصيغ التالية:

```
Dataset Close D0.32357303541392757.
T-TEST GROUPS=CLU2_1(1 2)
  /MISSING=ANALYSIS
  /VARIABLES=Salary FTE Rank Articles Experience
  /CRITERIA=CI(.95).
```



شكل 21. اختبار t للعينات المستقلة

تُظهر نواتج برنامج SPSS أن مجموعة "الهيئة المعاونة" لديها متوسطات رواتب أقل، ودوام كامل، ورتب أكاديمية أقل، ومقالات منشورة أقل، وسنوات خبرة أقل. ويوضح الجدول 4 الخصائص الإحصائية للمتغيرات المستهدفة من التحليل.

جدول 5. الخصائص الإحصائية للمتغيرات موزعة على فئات التجميع على المتغير الأول الجديد CLU2_1

Group Statistics					
	Average Linkage (Between Groups)	N	Mean	Std. Deviation	Std. Error Mean
Salary	1	34	60085.48	18665.114	3201.041
	2	10	5956.41	2101.013	664.399
FTE	1	34	1.00	0.000	0.000
	2	10	0.38	.132	.0417
Rank	1	34	3.53	1.134	0.195
	2	10	1.00	0.000	0.000
Articles	1	34	14.91	16.539	2.836
	2	10	1.90	4.771	1.509
Experience	1	34	12.79	11.335	1.944
	2	10	4.70	10.688	3.380

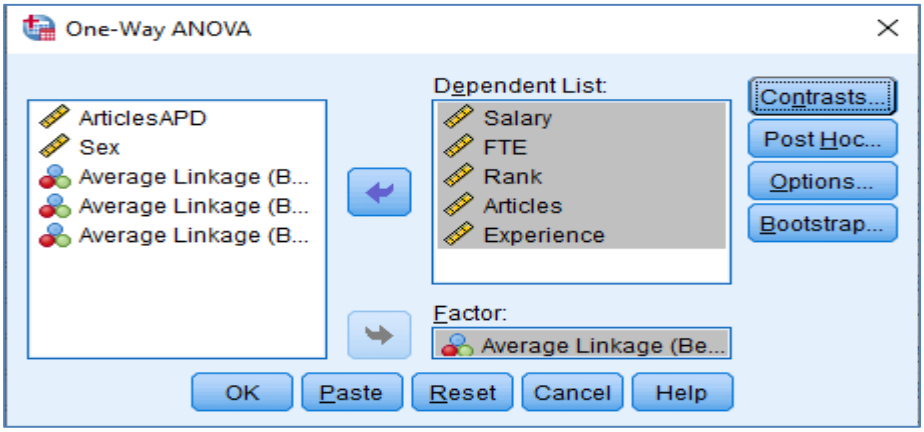
ثم يعرض برنامج SPSS لدلالة الفروق بين المجموعتين الناشئين عن المتغير الجديد CLU2_1 (الهيئة المعاونة، والرتب الأكاديمية الأخرى)، كما بالجدول 6.

جدول 6. دلالة الفروق بين المجموعتين الناشئين عن المتغير الجديد CLU2_1 (الهيئة المعاونة، والرتب الأكاديمية الأخرى)

Independent Samples Test						
		Levene's Test for Equality of Variances		t-test for Equality of Means		
		F	Sig.	t	df	Sig. (2-tailed)
Salary	Equal variances assumed	7.270	.010	9.079	42	.000
	Equal variances not assumed			16.557	35.662	.000
FTE	Equal variances assumed	.	.	28.484	42	.000

Independent Samples Test						
		Levene's Test for Equality of Variances		t-test for Equality of Means		
		F	Sig.	t	df	Sig. (2-tailed)
	Equal variances not assumed			15.000	9.000	.000
Rank	Equal variances assumed	29.711	.000	6.992	42	.000
	Equal variances not assumed			13.001	33.000	.000
Articles	Equal variances assumed	7.350	.010	2.440	42	.019
	Equal variances not assumed			4.050	41.990	.000
Experience	Equal variances assumed	1.544	.221	2.009	42	.051
	Equal variances not assumed			2.076	15.477	.055

بعد حل المجموعتين (التجمعان)، يعرض برنامج SPSS الحل الثلاث مجموعات (تجمعات): أعضاء هيئة التدريس، الهيئة المعاونة، الآخرون. قارن الآن المجموعات الثلاث من حل المجموعات الثلاث. سنستخدم تحليل التباين ANOVA أحادي الاتجاه، بناء على التوصيف الموضح بشكل 22.



شكل 22. توصيف تحليل التباين أحادي الاتجاه للمتغيرات المقاسة مع اعتبار أن المتغير الجديد لحل الثلاث مجموعات (أعضاء هيئة التدريس، الهيئة المعاونة، الرتب الأخرى) هو CLU3_1

يتم تنفيذ تحليل التباين الأحادي بناء على الصيغة التالية:

```
ONEWAY Salary FTE Rank Articles Experience BY CLU3_1
  /STATISTICS DESCRIPTIVES
  /MISSING ANALYSIS.
```

يعرض برنامج SPSS ناتج التحليل بإظهار الخصائص الإحصائية للمجموعات الثلاث في المتغيرات المقاسة، كما بالجدول 6.

جدول 7. الخصائص الإحصائية للمجموعات الثلاث في المتغيرات المقاسة وفق المتغير الجديد لحل الثلاث مجموعات

Descriptives

		N	Mean	Std. Deviation	Std. Error
Salary	1	10	80277.4080	18259.10829	5774.03702
	2	24	51672.1825	10875.28739	2219.90874
	3	10	5956.4080	2101.01288	664.39861
	Total	44	47783.4214	28192.46890	4250.17460
FTE	1	10	1.0000	.00000	.00000
	2	24	1.0000	.00000	.00000
	3	10	.3750	.13176	.04167
	Total	44	0.858	0.272	0.041
Rank	1	10	4.80	0.422	0.133
	2	24	3.00	0.885	0.181
	3	10	1.00	0.000	0.000
	Total	44	2.95	1.462	0.220
Articles	1	10	32.90	17.483	5.529
	2	24	7.42	8.577	1.751
	3	10	1.90	4.771	1.509
	Total	44	11.95	15.656	2.360
Experience	1	10	26.80	5.534	1.750
	2	24	6.96	7.178	1.465
	3	10	4.70	10.688	3.380
	Total	44	10.95	11.588	1.747

ثم عرض برنامج SPSS لدلالة الفروق بين المجموعات الثلاث في المتغيرات المقاسة كما بالجدول 8.

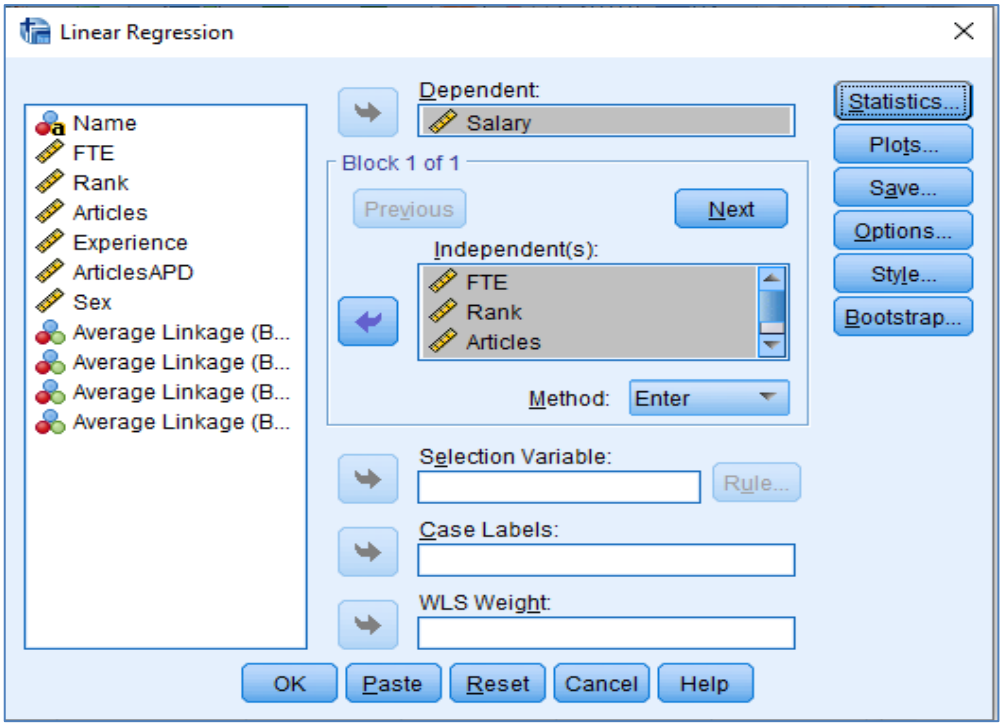
جدول 8. دلالة الفروق بين المجموعات الثلاث في المتغيرات المقاسة

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Salary	Between Groups	28416521260.677	2	14208260630.338	101.126	.000
	Within Groups	5760536757.805	41	140500896.532		
	Total	34177058018.481	43			
FTE	Between Groups	3.018	2	1.509	396.023	.000
	Within Groups	.156	41	.004		
	Total	3.175	43			
Rank	Between Groups	72.309	2	36.155	75.629	.000
	Within Groups	19.600	41	.478		
	Total	91.909	43			
Articles	Between Groups	5892.276	2	2946.138	25.990	.000
	Within Groups	4647.633	41	113.357		
	Total	10539.909	43			
Experience	Between Groups	3285.251	2	1642.625	27.062	.000
	Within Groups	2488.658	41	60.699		
	Total	5773.909	43			

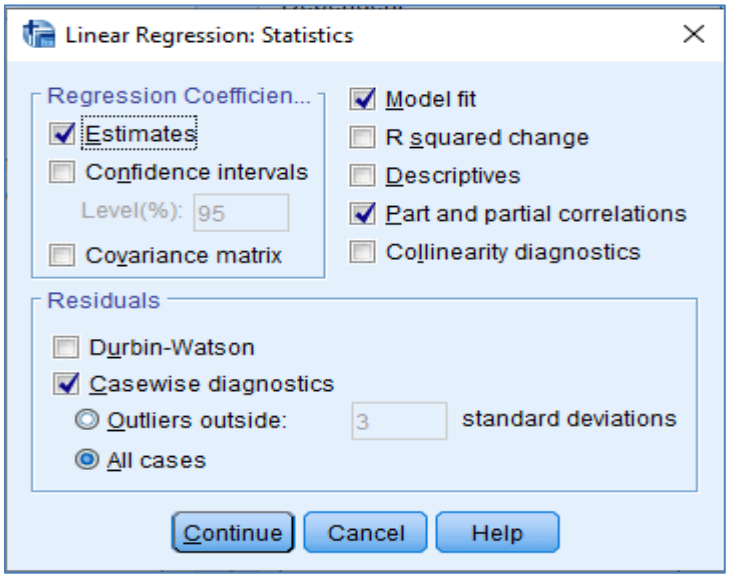
توقع الراتب من عب العمل المكافئ FTE، الرتبة الأكاديمية، عدد المنشورات، وعدد سنوات الخبرة:

كنوع من التحليل الإحصائي المكافئ، الذي يمكن من خلاله اختبار فرضية انحدار عدد من المتغيرات المقاسة للتنبؤ بمتغير ما، يمكن استخدام تحليل الانحدار المتعدد، لتفسير كيفية ارتباط رواتب أعضاء هيئة التدريس بدوام كامل، والرتبة الأكاديمية، وعدد المقالات المنشورة، وعدد سنوات الخبرة. وتوضح الأشكال 23، و24، و25 توصيف اختيارات تحليل الانحدار المتعدد في هذه الحالة.



شكل 23. توصيف متغيرات تحليل الانحدار المتعدد

اختر ارتباطات الجزء والارتباطات الجزئية وتشخيصات Casewise لجميع الحالات.

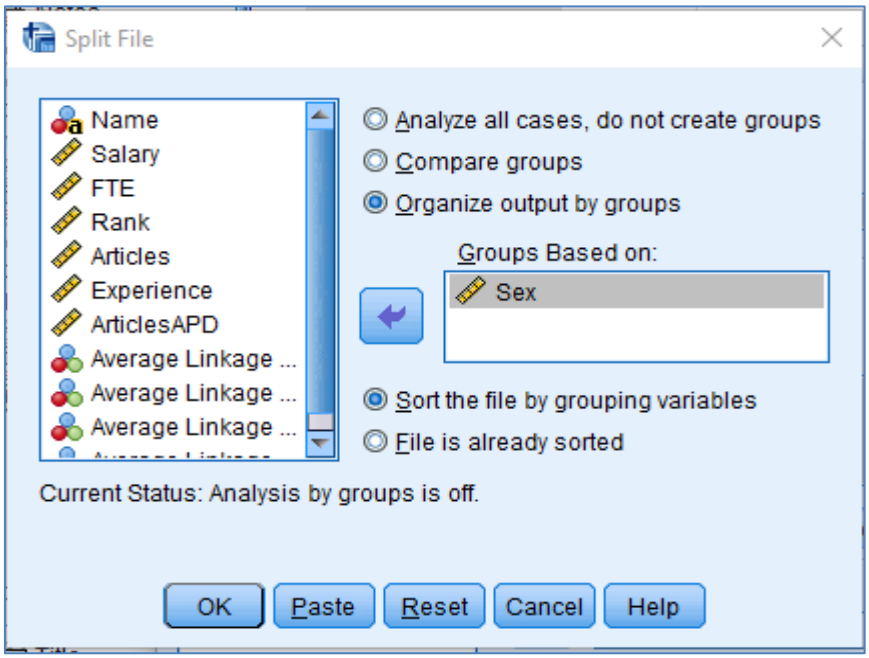


شكل 24. اختيار الإحصاءات المطلوبة لعرض تحليل الانحدار المتعدد

توضح نواتج برنامج SPSS لتحليل الانحدار المتعدد أن كل من المتغيرات المنبئة المحدد لها ارتباط رتبة-صفرية zero-order correlation إيجابي متوسط إلى كبير مع الراتب، ولكن متغير العبء المكافئ FTE والرتبة الأكاديمية فقط لهما تأثيرات جزئية كبيرة. في جدول Casewise التشخيصي، يتم إعطاؤك لكل حالة المتبقي القياسي (أعتقد أن أيًا منهما تتجاوز قيمته المطلقة 1 يستحق الفحص من قبل الأشخاص الذين حددوا رواتب أعضاء هيئة التدريس)، والراتب الفعلي، والراتب الذي تنبأ به النموذج، والفرق بالدولار بين الراتب الفعلي والراتب المتوقع.

إذا تم تقسيم الملف حسب الجنس وتكرار تحليل الانحدار، فسترى بعض الاختلافات المثيرة للاهتمام بين نموذج النساء ونموذج الرجال. التأثير الجزئي للرتبة أكبر بكثير بالنسبة للنساء منه بالنسبة للرجال. وبالنسبة للرجال، يكون التأثير الجزئي لعدد المقالات المنشورة إيجابيًا ودالاً إحصائياً، ولكن بالنسبة للنساء يكون تأثيرًا سلبيًا. أي، بين أعضاء هيئة التدريس، يتمثل الأثر الجزئي للنشر في خفض راتب النساء.

للعلم، يتم تقسيم الملف من خلال اختيار Data من القائمة الرئيسة لبرنامج SPSS ثم اختيار Split File من القائمة المنسدلة، ثم اختيار تنظيم النواتج بناء على المجموعات، ثم ادخال متغير النوع Sex في مربع تصنيف المجموعات بناء على: كما بالشكل 24.



شكل 24. توصيف عملية تقسيم الملف بناء على متغير النوع Sex لتنظيم نواتج التحليل المطلوب برنامج SPSS بناء على هذا المتغير

في هذه الحالة تكون صيغة تحليل الانحدار كالتالي:

```
SORT CASES  BY Sex.
SPLIT FILE SEPARATE BY Sex.
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA ZPP
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT Salary
  /METHOD=ENTER FTE Rank Articles Experience
  /CASEWISE PLOT(ZRESID) ALL.
```

وتظهر نواتج تحليل الانحدار المتعدد لكل نوع (النساء / الرجال) كما في الجداول (9، 10، 11، 12، 13، 14) بالنسبة للإناث، والجداول (15، 16، 17، 18، 19، 20) بالنسبة للذكور.

Regression
Sex = Female

جدول 9. المتغيرات المدخلة لتحليل الانحدار المتعدد

Variables Entered/Removed[a,b]			
Model	**Variables Entered**	**Variables Removed**	**Method**
1	**Experience, FTE, Articles, Rank[c]**	**.**	**Enter**
a. Sex = Female			
b. Dependent Variable: Salary			
c. All requested variables entered.			

يوضح جدول 9 أن المتغيرات المدخلة للنموذج كمتغيرات منبئة هي: الخبرة، والعمل المكافئ، وعدد المقالات المنشورة، والرتبة الأكاديمية.

جدول 10. ملخص نموذج الانحدار المتعدد

Model Summary[a,c]				
Model	**R**	**R Square**	**Adjusted R Square**	**Std. Error of the Estimate**
1	**0.954[b]**	**0.911**	**0.887**	**10420.466**
a. Sex = Female				
b. Predictors: (Constant), Experience, FTE, Articles, Rank				
c. Dependent Variable: Salary				

يوضح جدول 10 وجود نموذج 1 للانحدار المتعدد بمعامل ارتباط 0.954، ومعامل اسهام 0.911 للمتغيرات المنبئة عدد سنوات الخبرة Experience، وعبء العمل المكافئ FTE وعدد المقالات المنشورة Articles، والرتبة الأكاديمية Rank بالإضافة إلى الثابت (Constant) بالمتغير التابع وهو الراتب Salary.

جدول 11. دلالة معامل الارتباط المتعدد بمعلومية قيمة ف الناتجة عن تحليل التباين لمجموعة الإناث

ANOVA[a,b]						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	16660209183.111	4	4165052295.778	38.357	.000[c]
	Residual	1628791645.854	15	108586109.724		
	Total	18289000828.965	19			
a. Sex = Female						
b. Dependent Variable: Salary						
c. Predictors: (Constant), Experience, FTE, Articles, Rank						

يتضح من جدول 11 أن قيمة ف دالة إحصائياً عند مستوى دلالة 0.0001 مما يدل على أن قيمة معامل الانحدار المتعدد دالة عند نفس المستوى.

جدول 12. دلالة معامل الانحدار للمتغيرات المنبئة بمعلومية قيمة ت ومعاملات الارتباط الجزئية لكل متغير من المتغيرات المنبئة مع المتغير التابع باستبعاد تأثير بقية المتغيرات المنبئة

Coefficients[a,b]									
		Unstandardized Coefficients		Standardized Coefficients			Correlations		
Model		B	Std. Error	Beta	t	Sig.	Zero-order	Partial	Part
1	(Constant)	-18583.206	7999.142		-2.323	.035			
	FTE	15569.891	14658.068	.137	1.062	.305	.756	.264	.082
	Rank	18042.686	4817.433	.850	3.745	.002	.933	.695	.289
	Articles	-923.656	471.655	-.315	-1.958	.069	.676	-.451	-.151
	Experience	830.194	491.389	.286	1.689	.112	.796	.400	.130
a. Sex = Female									
b. Dependent Variable: Salary									

يتضح من جدول 12 معاملات الانحدار المعيارية لمتغيرات: عبء العمل

المكافئ، والرتبة الأكاديمية، وعدد المقالات المنشورة، وعدد سنوات الخبرة كانت 0.137، 0.850، -0.315، 0.286 وجميعها غير دالة باستثناء متغير الرتبة الأكاديمية، والثابت الذي بلغ مقداره -18583.206. كما يتضح أن معامل الانحدار لمتغير عدد المقالات المنشورة كان سالباً لدى الإناث بمعامل -0.315 عند مستوى دلالة قريب من 0.05 يبلغ 0.065. وكان أعلى معامل ارتباط جزئي لمتغير الرتبة الأكاديمية مع الراتب حيث بلغ 0.695 مع استبعاد بقية المتغيرات المنبئة، بعد أن كان 0.289 دون حدث عملية الاستبعاد.

ثم يعرض برنامج SPSS لبيانات الحالة التشخيصية للحالات المتشابهة كما بالجدول 13.

جدول 13. القيم المتوقعة للمتغير التابع (الراتب) بناء بيانات الحالة التشخيصية Casewise Diagnosticsa للحالات المتشابهة الناتجة عن معادلة الانحدار وبواقي هذه القيم للإناث

Casewise Diagnostics[a,b]				
Case Number	Std. Residual	Salary	Predicted Value	Residual
1	2.805	123600.00	94369.5538	29230.44615
2	-.597	83358.00	89575.3132	-6217.31317
3	-.037	68127.84	68518.1567	-390.31670
4	-.764	62476.80	70433.0092	-7956.20923
5	-1.289	62256.00	75683.5973	-13427.59731
6	-.332	61992.96	65451.8453	-3458.88526
7	.816	55550.88	47046.2737	8504.60634
8	-1.572	55072.80	71450.1265	-16377.32652
9	.484	55050.96	50004.1639	5046.79612
10	.252	53550.96	50927.8195	2623.14048
11	.426	52600.80	48156.8526	4443.94740
12	.010	52050.00	51944.9368	105.06319
13	-.678	49762.80	56832.6391	-7069.83914
14	.562	41424.00	35562.6390	5861.36103
15	-.068	14324.94	15029.3712	-704.43115
16	.085	8130.87	7244.4256	886.44444
17	-.027	6962.94	7244.4256	-281.48556
18	-.120	5999.14	7244.4256	-1245.28556
19	.024	3601.89	3351.9528	249.93723

20	.017	3528.90	3351.9528	176.94723
a. Sex = Female				
b. Dependent Variable: Salary				

كما يعرض جدول 13 الخصائص الإحصائية للبواقي الناتجة عن معادلة الانحدار.

جدول 14. الخصائص الإحصائية للبواقي الناتجة عن معادلة الانحدار

Residuals Statistics[a,b]					
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	3351.959	94369.555	45971.174	29611.706	20
Residual	-16377.326	29230.445	.000	9258.827	20
Std. Predicted Value	-1.439	1.634	.000	1.000	20
Std. Residual	-1.572	2.805	.000	.889	20
a. Sex = Female					
b. Dependent Variable: Salary					

يتضح من جدول 14 أن متوسط الراتب المتوقع يبلغ 45971.74 بانحراف معياري قدره 29611.706، بنما يبلغ متوسط البواقي صفراً.

وبالنسبة للذكور، توضح الجداول (15، 16، 17، 18، 19، 20) نتائج تحليل الانحدار لهذا التجمع.

Regression
Sex = Female

جدول 15. المتغيرات المدخلة لتحليل الانحدار المتعدد

Variables Entered/Removed[a,b]			
Model	Variables Entered	Variables Removed	Method
1	Experience, FTE, Articles, Rank[c]	.	Enter
a. Sex = Male			
b. Dependent Variable: Salary			
c. All requested variables entered.			

يوضح جدول 15 أن المتغيرات المدخلة للنموذج كمتغيرات منبئة هي: الخبرة، والعمل المكافئ، وعدد المقالات المنشورة، والرتبة الأكاديمية.

جدول 16. ملخص نموذج الانحدار المتعدد لمجموعة الذكور

Model Summary[a,c]				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.972[b]	.945	.933	6752.324
a. Sex = Male				
b. Predictors: (Constant), Experience, FTE, Articles, Rank				
c. Dependent Variable: Salary				

يوضح جدول 16 وجود نموذج 1 للانحدار المتعدد بمعامل ارتباط 0.972، ومعامل اسهام 0.945 للمتغيرات المنبئة عدد سنوات الخبرة Experience، وعبء العمل المكافئ FTE وعدد المقالات المنشورة Articles، والرتبة الأكاديمية Rank بالإضافة إلى الثابت (Constant) بالمتغير التابع وهو الراتب Salary.

جدول 17. دلالة معامل الارتباط المتعدد بمعلومية قيمة ف الناتجة عن تحليل التباين لمجموعة الإناث

ANOVA[a,b]						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	14901351390.482	4	3725337847.621	81.707	.000[c]
	Residual	866283647.110	19	45593876.164		
	Total	15767635037.592	23			
a. Sex = Male						
b. Dependent Variable: Salary						
c. Predictors: (Constant), Experience, FTE, Articles, Rank						

يتضح من جدول 17 أن قيمة ف دالة إحصائياً عند مستوى دلالة 0.0001 مما يدل على أن قيمة معامل الانحدار المتعدد دالة عند نفس المستوى.

جدول 18. دلالة معامل الانحدار للمتغيرات المنبئة بمعلومية قيمة ت ومعاملات الارتباط الجزئية لكل متغير من المتغيرات المنبئة مع المتغير التابع باستبعاد تأثير بقية المتغيرات المنبئة لمجموعة الذكور

Coefficients[a,b]					
Model	Unstandardized Coefficients	Standardized Coefficients	t	Sig.	Correlations

		B	Std. Error	Beta			Zero-order	Partial	Part
1	(Constant)	-17427.298	4757.743		-3.663	.002			
	FTE	42736.192	7498.594	.450	5.699	.000	.849	.794	.306
	Rank	8316.282	1768.487	.473	4.702	.000	.918	.733	.253
	Articles	270.221	107.346	.190	2.517	.021	.656	.500	.135
	Experience	23.032	142.830	.011	.161	.874	.431	.037	.009
a. Sex = Male									
b. Dependent Variable: Salary									

يتضح من جدول 18 معاملات الانحدار المعيارية لمتغيرات: عبء العمل المكافئ، والرتبة الأكاديمية، وعدد المقالات المنشورة، وعدد سنوات الخبرة كانت 0.450، 0.473، 0.190، 0.011 وجميعها دالة إحصائياً باستثناء متغير عدد سنوات الخبرة، وكان الثابت الذي بلغ مقداره -17427.298 دالاً أيضاً. كما يتضح أن معاملات الانحدار للمتغيرات المنبئة على الترتيب: 0.794، و0.733، 0.500، 0.037. لكل متغير والواضح أن أعلاه متغير عبء العمل المكافئ، يليه متغير الرتبة الأكاديمية، ثم عدد المقالات المنشورة، بنما لم يكن متغير عدد سنوات الخبرة ذي تأثيراً دالاً.

ثم يعرض برنامج SPSS لبيانات الحالة التشخيصية للحالات المتشابهة كما بالجدول 19.

جدول 19. القيم المتوقعة للمتغير التابع (الراتب) بناء بيانات الحالة التشخيصية Casewise Diagnosticsa للحالات المتشابهة الناتجة عن معادلة الانحدار وبواقي هذه القيم للذكور

Casewise Diagnostics[a,b]				
Case Number	Std. Residual	Salary	Predicted Value	Residual
21	1.496	96799.92	86697.8120	10102.10799
22	1.611	83235.84	72359.2528	10876.58719
23	.449	75040.80	72007.6186	3033.18137
24	-.895	74956.80	81000.1497	-6043.34970
25	.926	72225.84	65974.4188	6251.42121
26	-.937	72055.92	78383.9119	-6327.99189
27	-1.231	66157.92	74468.7942	-8310.87421
28	.132	58827.84	57939.0707	888.76929
29	.208	57499.92	56093.5905	1406.32954

Casewise Diagnostics[a,b]				
Case Number	Std. Residual	Salary	Predicted Value	Residual
30	.650	55999.92	51608.8408	4391.07924
31	.619	55060.80	50879.5929	4181.20709
32	.554	54000.00	50257.7382	3742.26178
33	-1.371	53262.00	62518.3284	-9256.32845
34	.006	52999.92	52959.9433	39.97670
35	-1.684	48159.84	59529.0306	-11369.19056
36	.806	47499.84	42056.6150	5443.22504
37	.369	44499.84	42010.5517	2489.28835
38	-1.013	41999.76	48841.3187	-6841.55866
39	-.087	41424.00	42010.5517	-586.55165
40	-1.020	10162.89	17047.3998	-6884.50980
41	-.806	6812.82	12257.0792	-5444.25923
42	.284	4919.85	2999.3881	1920.46188
43	.469	4738.86	1573.0313	3165.82874
44	.464	4705.92	1573.0313	3132.88874
a. Sex = Male				
b. Dependent Variable: Salary				

كما يعرض جدول 20 الخصائص الإحصائية للبواقي الناتجة عن معادلة الانحدار لمجموعة الذكور.

جدول 20. الخصائص الإحصائية للبواقي الناتجة عن معادلة الانحدار لمجموعة الذكور

Residuals Statistics[a,b]					
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	1573.0313	86697.813	49293.628	25453.582	24
Residual	-11369.190	10876.587	.000	6137.142	24
Std. Predicted Value	-1.875	1.470	.000	1.000	24
Std. Residual	-1.684	1.611	.000	.909	24
a. Sex = Male					
b. Dependent Variable: Salary					

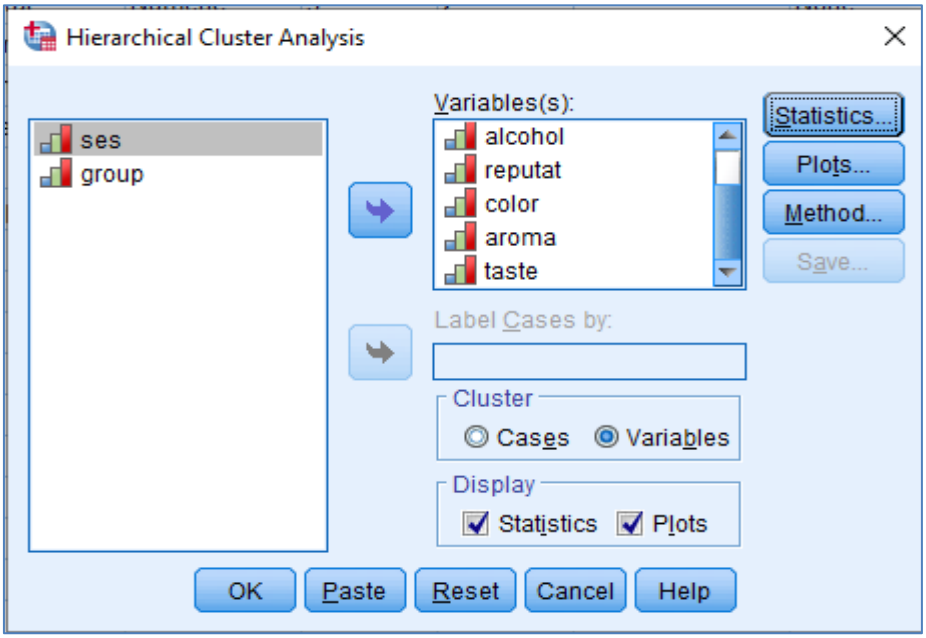
يتضح من جدول 20 أن متوسط الراتب المتوقع يبلغ 45971.74 بانحراف

معياري قدره 29611.706، بنما يبلغ متوسط البواقي صفراً.

ثانياً: إجراء التحليل التجميعي للمتغيرات:

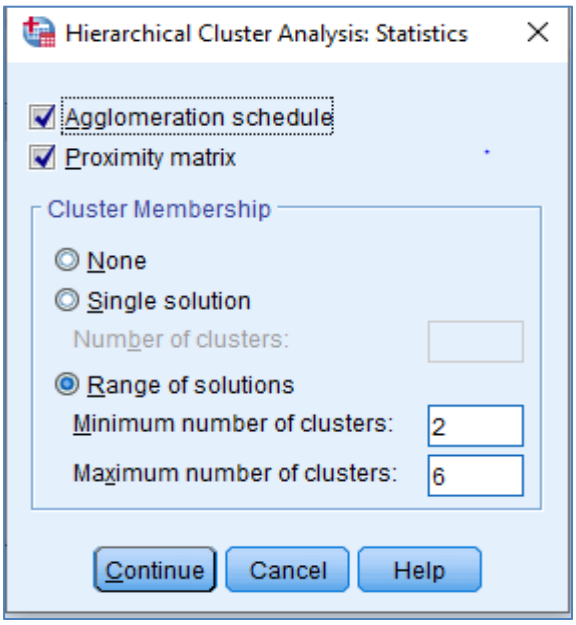
يمكن استخدام التحليل التجميعي (العنقودي) لتجميع المتغيرات بدلاً من الحالات. في هذه الحالة، يكون الهدف مشابهًا للهدف في التحليل العاملي Factor Analysis - للحصول على مجموعات من المتغيرات المتشابهة مع بعضها البعض. مرة أخرى، لم أطلع على أي من البحوث استخدم هذه الفنية، ولم أستخدمها أيضاً بعد في أي من بحوثي، لكنها تبدو. كما يقول وينيس Wuensch (2016) مثيرة للاهتمام.

سنستخدم نفس البيانات المستخدمة سابقًا للمكونات الرئيسية والتحليل العاملي، من خلال الملف المنشور على صفحة وينش (2021) باسم FactBeer.sav، وهذا الملف يمكن الحصول عليه من خلال الرابط: https://core.ecu.edu/wuenschk/SPSS/SPSS-Data.htm. ابدأ بالنقر فوق تحليل Analyze، تصنيف Classify، مجموعة هرمية Hierarchical Cluster. أدخل في مربع المتغيرات نفس المتغيرات السبعة التي استخدمناها في تحليل المكونات والتحليل العاملي. تحت اختيار "تجميع Cluster"، حدد "المتغيرات Variables"، كما بالشكل 25.

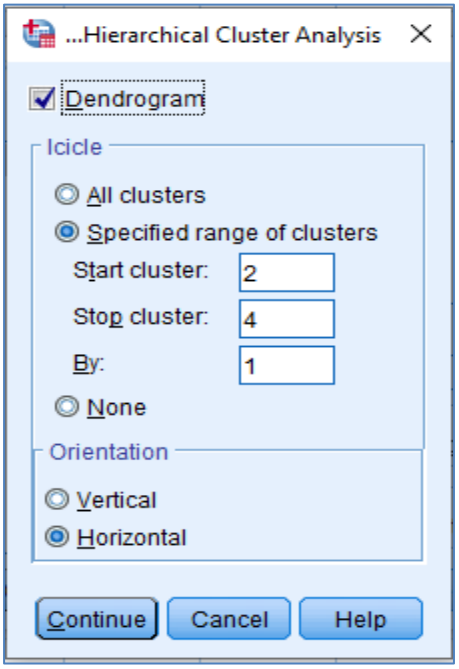


شكل 25. توصيف التحليل التجميعي الهرمي للمتغيرات

حدد الإحصائيات والمخططات الرسومية كما هو موضح بالشكلين 26، و27.

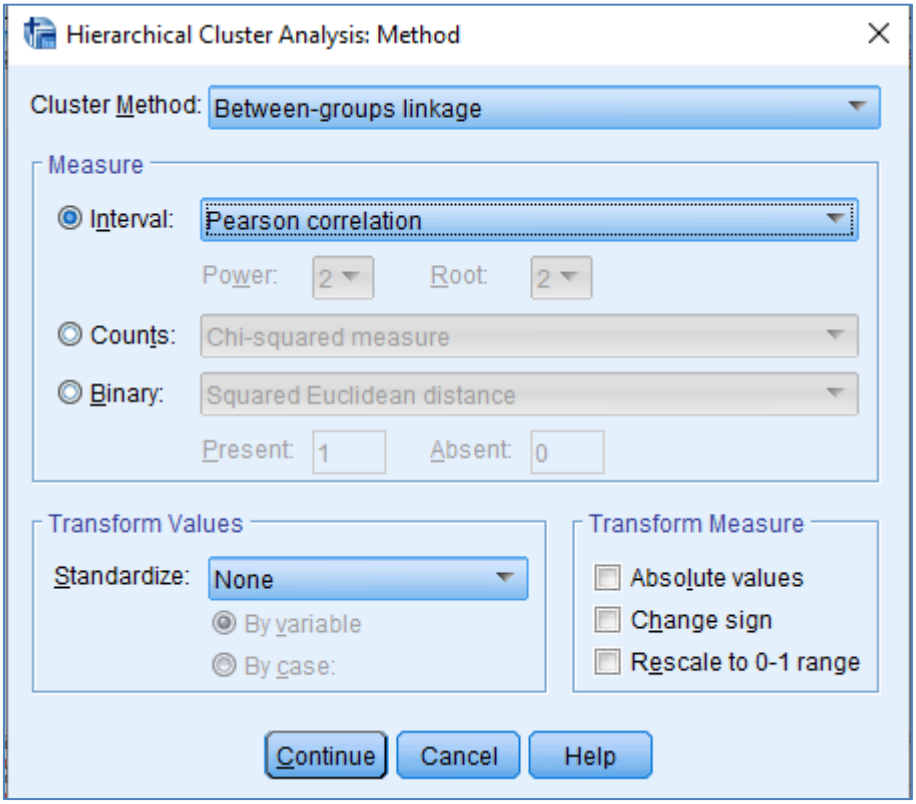


شكل 26. توصيف إحصاءات Statistics التحليل التجميعي الهرمي للمتغيرات



شكل 27. توصيف المخططات الرسومية Plots للتحليل التجميعي الهرمي للمتغيرات

ثم توصيف طريقة التحليل التجميعي الهرمي كما بالشكل 28.



شكل 28. توصيف طريقة التحليل التجميعي الهرمي للمتغيرات

تظهر صيغة التحليل وفق السيناريو التالي:

```
CLUSTER
  /MATRIX IN(D0.11839465930474014)
  /METHOD BAVERAGE
  /PRINT SCHEDULE CLUSTER(2,6)
  /PRINT DISTANCE
  /PLOT DENDROGRAM HICICLE(2,4,1).
```

يعرض برنامج SPSS لنواتج التحليل في شكل الجداول من 21 إلى 24 بداية من الجدول 21 لمصفوفة القرب Proximity Matrix.

Cluster

جدول 21. مصفوفة القرب أو التشابه

Proximity Matrix

Case	Matrix File Input						
	cost	size	alcohol	reputat	color	aroma	taste
cost	1.000	.832	.767	-.406	.018	-.046	-.064
size	.832	1.000	.904	-.392	.179	.098	.026
alcohol	.767	.904	1.000	-.463	.072	.044	.012
reputat	-.406	-.392	-.463	1.000	-.372	-.443	-.443
color	.018	.179	.072	-.372	1.000	.909	.903
aroma	-.046	.098	.044	-.443	.909	1.000	.870
taste	-.064	.026	.012	-.443	.903	.870	1.000

ثم يعرض لمراحل جدولة التجميع للمتغيرات كما بالجدول 22.

جدول 22. مراحل جدولة التجميع للمتغيرات

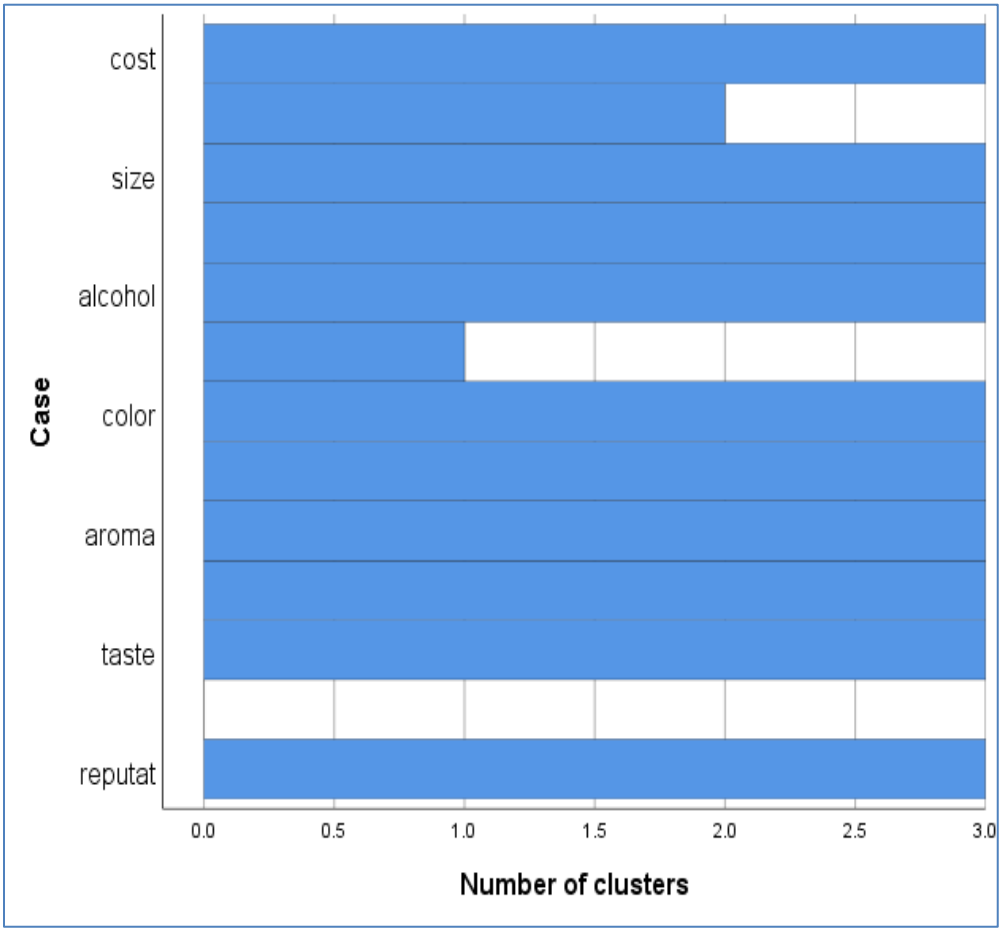
Agglomeration Schedule						
	Cluster Combined			Stage Cluster First Appears		
Stage	Cluster 1	Cluster 2	Coefficients	Cluster 1	Cluster 2	Next Stage
1	5	6	.909	0	0	3
2	2	3	.904	0	0	4
3	5	7	.887	1	0	5
4	1	2	.800	0	2	5
5	1	5	.038	4	3	6
6	1	4	-.420	5	0	0

يلي ذلك عرض لعضوية متغير لكل تجمع من هذه المتغيرات كما بالجدول 23.

جدول 23. عرض لعضوية كل متغير لكل تجمع من المتغيرات

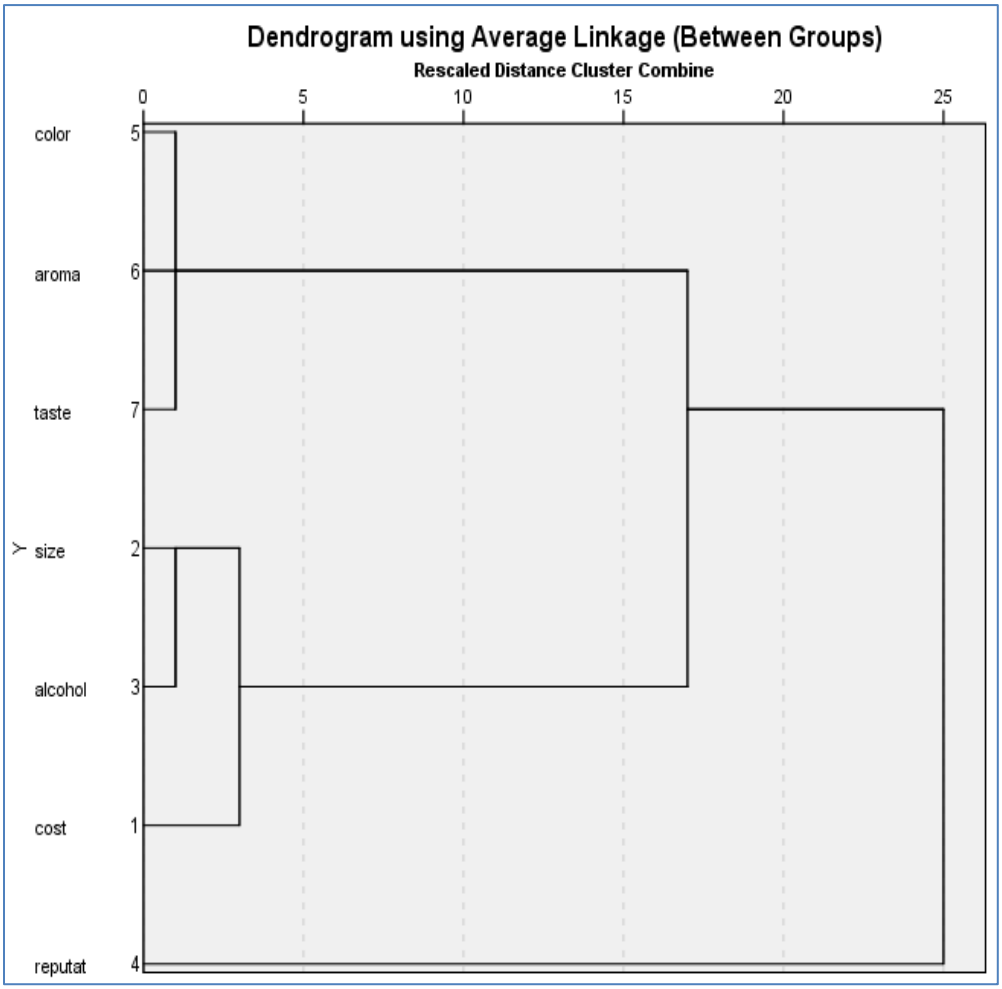
Cluster Membership					
Case	6 Clusters	5 Clusters	4 Clusters	3 Clusters	2 Clusters
cost	1	1	1	1	1
size	2	2	2	1	1
alcohol	3	2	2	1	1
reputat	4	3	3	2	2
color	5	4	4	3	1
aroma	5	4	4	3	1
taste	6	5	4	3	1

ثم يعرض للمخطط الجليدي كما بالشكل 29.



شكل 29. المخطط الجليدي لتجمعات المتغيرات المستهدفة من التحليل التجميعي

يلي ذلك عرض للمخطط الشجري (مخطط أسنان الشوكة) Dendrogram لتجمعات المتغيرات المستهدفة بالتحليل التجميعي، كما بالشكل 30.



شكل 30. المخطط الشجري Dendrogram لتجمعات المتغيرات المستهدفة بالتحليل التجميعي الهرمي

بالنظر إلى مصفوفة القرب، يتضح أنها تمثل ببساطة مصفوفة ارتباط. نبدأ مع كل متغير كعنصر خاص به يكون معامل الارتباط قدره 1.0. خطوتنا الأولى هي الجمع بين العنصرين الأقرب - أي المتغيرين الأكثر ارتباطًا. كما يمكنك من مصفوفة القرب، هذا هو اللون والرائحة (ر = 909). لدينا الآن ستة عناصر - مجموعة واحدة وخمسة متغيرات لم يتم تجميعها بعد.

في المرحلة 2، نقوم بتجميع أقرب عنصرين من العناصر الستة المتبقية. هذا هو الحجم size والكحول alcohol (ص = 0.903). انظر إلى جدول التجميع. كما ترى، تضمنت المرحلة الأولى تجميع المتغيرين 5 و 6 (اللون color والرائحة taste)، بينما تضمنت المرحلة الثانية تجميع المتغيرين 2 و 3 (الحجم size

والكحول alcohol).

في المرحلة 3، تمت إضافة المتغير 7 (التذوق taste) إلى التجمع الذي يحتوي بالفعل على المتغيرين 5 (اللون color) و 6 (الرائحة taste).

في المرحلة 4، تمت إضافة المتغير 1 (التكلفة cost) إلى التجمع الذي يحتوي بالفعل على المتغيرين 2 (الحجم size) و 3 (الكحول alcohol). أصبح لدينا الآن ثلاثة عناصر – وتجمعان (مجموعتان من المتغيرات)، كل واحدة بها ثلاثة متغيرات، ومتغير واحد لم يتم تجميعه بعد.

في المرحلة 5، تم دمج التجمعين (المجموعتين من المتغيرات)، لكن لاحظ أنهما ليسا متشابهين جدًا، معامل التشابه هو 0.38 فقط. في هذه المرحلة، لدينا عنصرين، متغير السمعة reputation وحده، والمتغيرات الستة المتبقية مجمعة في تجمع واحد.

يُظهر المخططان الرسوميان بشكلي 19، و20 إلى حد كبير نفس ما أوضحته مصفوفة القرب وجدولة التجميعات، ولكن في شكل يمكن أن يكون أكثر سهولة في الفهم.

أنا أفضل الحل العنقودي في مثل هذه الحالات. ولاحظ أن متغير السمعة reputation لم يتم تجميعه حتى الخطوة الأخيرة، حيث كان مرتبطاً سلبًا بالمتغيرات المتبقية كما هو موضح بجدول 21 الخاص بمصفوفة القرب أو التشابه.

المراجع:

Aldenderfer, M. S., & Blashfield, R. K. (1984). Cluster Analysis. Sage university paper series on quantitative applications in the social sciences, 07–044. Newbury Park, CA: Sage.

Everitt, B. (1993). Cluster analysis (3rd edition). London: Arnold.

Field, A. P. (2000). Postgraduate Statistics: Cluster Analysis. London: Sage. pp. 1-10. http://www.discoveringstatistics.com/docs/cluster.pdf

Field, A. P. (2013). Discovering statistics using IBM SPSS Statistics: And sex and drugs and rock 'n' roll (4th ed.). London: Sage.

Johnson, R. A., & Wichern, D. W. (1998). Applied multivariate statistical analysis (4th edition). New Jersey: Prentice Hall. Chapter 12: 12.1-12.4.

Romesburg, H. C. (1984). Cluster analysis for researchers. Belmont, CS: Lifetime Learning Publications.

Wuensch, Karl L. (2016, January 17). Cluster Analysis With SPSS. East Carolina University Department of Psychology Greenville, NC 27858-4353. https://core.ecu.edu/wuenschk/SPSS/ClusterAnalysis-SPSS.pdf

Wuensch, Karl L. (2021, March 16). Professor Karl L. Wuensch's SPSS-Data Page. East Carolina University Department of Psychology Greenville, NC 27858-4353. https://core.ecu.edu/wuenschk/SPSS/SPSS-Data.htm