الفصل الحادي والعشرون

تحليل التجمع

يتشابه تحليل التجمع من نواحي متعددة مع التحليل العاملي الذي ناقشناه في الفصل السابق. ولذلك ربما كان من المناسب أن نبدأ هذا الفصل بمقارنة بين تحليل التجمع والتحليل العاملي حتى يتضح الفرق بينهما وبالتالي يسهل فهم تحليل التجمع كأسلوب إحصائي يهدف إلى خفض ما لدينا من أفراد.

مقارنة بين تحليل التجمع والتحليل العاملي:

يتشابه تحليل التجمع مع التحليل العاملي في أن كلاهما يشتمل على إجراءات من شأنها خفض عدد كبير من الحالات أو المتغيرات إلى عدد أقل من العوامل (أو التجمعات) بناء على بعض التشابه في خصائص الأفراد في المجموعة الواحدة. إلا أن الإجراءات الإحصائية التي تشكل الأساس لكل نوع من التحليل وكذلك طريقة تفسير نتائج كل منهما مختلفة تماما. وفيما يلي أهم أوجه المقارنة بين كل منهما:

١- يستخدم التحليل العاملي لخفض عدد كبير من المتغيرات إلى عدد أقل من العوامل التي تصف هذه المتغيرات. أما تحليل التجمع فيستخدم لتجميع الأفراد في مجموعات. وبتحديد أكثر يمكن القول أن إجراءات SPSS قد صممت لاستخدام المتغيرات كمعايير لتجميع الأفراد أو الحالات (وليس المتغيرات) في مجموعات وذلك بناء على درجات الفرد (أو الحالة) في مجموعة معينة من المتغيرات. مثال ذلك إذا كان لدينا ٥٠٠ فرد تم اختبارهم بخمسة عشر اختبارا (متغيرا) مختلفا، فإن التحليل العاملي يمكن استخدامه لاستخلاص ثلاثة أو أربعة عوامل من هذه المتغيرات. أما تحليل التجمع فإنه يقوم بتجميع هؤلاء الأفراد في مجموعات بناء على خصائصهم في تلك المتغيرات. فبدلا من تجميع المتغيرات كما الحال في التحليل العاملي، يقوم تحليل التجمع بتجميع الأفراد في مجموعات حسب خصائصهم المشتركة. وقد ينتج عن ذلك ثلاث أو أربع مجموعات يتشابه أفرادها في بعض السمات والخصائص.

٢- يمكن القول أن ما ذكر في الفقرة السابقة ينطبق على تحليل التجمع في الإصدارات القديمة من SPSS، ذلك أن الإصدارات الأخيرة فقد أدت التحسينات التي شهدها نظام النوافذ إلى أن أصبح من السهل أيضا تجميع المتغيرات (وليس الأفراد أو الحالات). ولعل أهم تطور في النوافذ هو حجم الذاكرة المتاح للمستخدم والذي أصبح كبيرا جدا مقارنة بالإصدارات السابقة مما ساعد على استخدام تحليل التجمع في سهولة أكثر، فقد كان من الأمور التي تقف عقبة أمام المستخدم هو جنوح البرنامج إلى التوقف لعدم توفر ذاكرة عشوائية كافية. أما الآن مع التوسع في إمكانيات الذاكرة العشوائية أصبح أمام البرنامج إمكانيات لا حدود لها في تنفيذ تحليل التجمع. وأخير نجد أن تجمع المتغيرات كان عملية معقدة للغاية فيما سبق ولكنه الآن أصبح في سهولة ضغط زر من الأزرار. ونظرا لأن الاستخدام الأكثر شيوعا حتى الآن هو تجميع الأفراد وليس المتغيرات فإننا سوف نتناول في هذا الفصل الطرق المستخدمة في تجميع الأفراد، إلا سوف نستخدم أيضا عملية التحليل في تجميع المتغيرات.

٣- تختلف الإجراءات الإحصائية في كل من العمليتين اختلافا كبيرا. ففي التحليل العاملي يتم تحليل جميع المتغيرات في كل خطوة من خطوات استخلاص العوامل لحساب التباين الذي يسهم به كل متغير في العامل. أما تحليل التجمع فيقوم بحساب مقياس للتشابه أو المسافة بين كل فرد أو حالة وكل فرد أو حالة أخرى ثم يقوم بتجميع الحالتين أو الفردين اللذين يظهران أكبر تشابه أو أقل المسافات في مجموعة واحدة. ثم يقوم بحساب التشابه أو المسافة من جديد بين الفردين الأقرب لبعضهما البعض (إذا كانت المسافة قصيرة)، ومن ثم قد يضمهما لمجموعة أو واحدة أو قد يضم أحدهما أو كلاهما إلى المجموعة السابق تكوينها، مما يؤدي إلى تكوين تجمعين يتكون كل منهما من حالتين، أو تجمع واحد يتكون من ثلاث حالات. وتستمر هذه العملية إلى أن يتم تجميع جميع الحالات في تجمع واحد. ويحدد الباحث المرحلة التي تتوقف عندها عملية التجميع.

٤- من بين أوامر SPSS الخاصة بتحليل التجمع هو التجمع الهرمي **Hierarchical Cluster**. وهذا الأمر أكثر استخداما في مجال الأعمال والصناعة، أو علم الاجتماع، أو العلوم السياسية (مثال ذلك فئات أكثر أجهزة التليفزيون مبيعا، أو تصنيف ٤٠ مجتمعا محليا بناء على بعض الخصائص

السكانية، تجميع أكثر ٣٠ مدينة سكانا بناء على العمر والمستوى الاقتصادي والاجتماعي والانتماء الحزبي) منه استخدام في علم النفس. فالباحثون النفسانيون أكثر ميلا للبحث عن التشابه بين المتغيرات أو أسباب الظاهرات من العمل على تقسيم الأفراد في مجموعات بناء على تكوين معين. وعند رغبتهم في تقسيم الأفراد في مجموعات فإن التحليل التمييزي **Discriminant Analysis** ينفذ هذه العملية بفاعلية أكبر من تحليل التجمع.

٥- نظرا لسهولة استخدام تحليل التجمع الآن في تجميع المتغيرات في مجموعات يصبح من المفيد مقارنة نتائج تحليل التجمع بنتائج التحليل العاملي باستخدام نفس البيانات. وكما هو الحال في التحليل العاملي يوجد عدد من التنويعات للطريقة التي يمكن بها تنفيذ العناصر الأساسية لتحليل التجمع (وهذه تقوم أساسا على تحديد المسافات أو التشابه وعلى كيفية تجمع متغيرات الأفراد). وحيث أن الطريقة التي استخدامها في التحليل العاملي وتحليل التجمع يمكن أن تؤدي إلى تعديلات جوهرية في النتائج فقد يستخدم الباحثون جميع الطرق المتاحة ويختارون منها ما يناسبهم. وهذا استخدام غير سليم ولا يتسم بالأمانة العلمية. ذلك أن المفروض أن يستخدم الباحث الطريقة الأنسب لبياناته بعد التأكد من أي منهما (التحليل العاملي أو تحليل التجمع) أنسب لبياناته.

طرق إجراء تحليل التجمع:

يمر تحليل التجمع بعدة خطوات حتى نصل إلى النتيجة النهائية. وللمساعدة في فهم هذه العملية قبل البدء في إجراءات التحليل سنذكر مثالا أعد خصيصا لهذا التحليل. ويشتمل هذا المثال على خطوات تحليل التجمع. وفي معظم البيانات التي نحصل عليها يكون التركيز على المتغيرات وطبيعتها، ومن النادر أن نهتم بتجميع الحالات. ولكننا الآن نلتفت إلى بيانات وضعت خصيصا لتحليل التجمع. ويحتوي ملف البيانات على أفضل ٢١ نوعا من مسجلات الفيديو. وقد نظم هذا الملف بحيث للمساعدة في على بناء تجمع واضح وبذلك يخدم كمثال مقنع. ويلاحظ أن أسماء الأنواع في البيانات غير حقيقية.

خطوة ١: تخير المتغيرات المستخدمة كمعايير لتكوين التجمع. وسوف يتم تحليل التجمع في ملف VCR.SAV باستخدام المتغيرات التالية (بالترتيب الذي ذكرت به في شكل (٢١-١). وقد استخدمت متغيرات السعر، وجودة الصورة (خمسة مقاييس)، وجودة

الاستقبال (ثلاثة مقاييس)، وجودة الصوت (ثلاثة مقاييس)، سهولة البرمجة (مقياس واحد)، عدد الأحداث (مقياس واحد)، عدد أيام البرمجة (مقياس واحد)، والتحكم عن بعد (ثلاثة مقاييس)، والأشياء الإضافية (ثلاثة مقاييس).

خطوة ٢: تخير إجراء قياس المسافة أو التشابه بين كل حالة والتجمعات الأخرى (يلاحظ أن التحليل يبدأ باعتبار كل حالة تجمعا قائما بذاته، أي أن لدينا تجمعات من ٢١ نوعا من مسجلات الفيديو). ويطلق على الإجراء الافتراضي في SPSS لهذا المقياس مربع المسافة الإقليدية Squared Euclidean Distance، وهي مجموع مربع الفروق لكل متغير وكل حالة. مثال ذلك قد يكون للنوع أ الدرجات ٢، ٣، ٥، لثلاثة تقديرات سمعية. وقد يكون للنوع ب التقديرات ٤، ٣، ٢، لنفس المقاييس. ومربع المسافة الإقليدية لهذين النوعين هي $(٢ - ٤)^{٢} + (٣-٣)^{٢} + (٥-٢)^{٢} = ١٣$. وعند التحليل الفعلي تجمع مربعات هذه المسافات لكل المتغيرات الإحدى وعشرين حيث تعطي مقياسا رقميا بين كل زوج من الأجهزة. ويسمح برنامج SPSS باستخدام مقاييس أخرى غير مربع المسافة الإقليدية لتحديد المسافة أو التشابه بين التجمعات. ويشرح دليل SPSS كيفية إجراء هذه المقاييس.

وقد يتساءل البعض إذا ما كان إجراء التجمع صالحا إذا كانت موازين قياس المتغيرات مختلفة. معظم المقاييس في ملف VCR.SAV عبارة عن تقديرات خماسية المقياس من ١ (أضعف تقدير) إلى ٥ (أعلى تقدير). إلا أن الأسعار المذكورة تتراوح بين ١٠٠٠ جنيه إلى ٢٥٠٠ جنيه، كما أن الأحداث، والأيام والأشياء الإضافية هي القيم الفعلية المرتبطة بتلك المتغيرات. والحل الذي يقترحه SPSS هو تحويل كل قيم المتغيرات إلى درجات معيارية (درجات Z، متوسطها صفر وانحرافها المعياري ١). وهناك أنواع أخرى من التحويلات المعيارية إلا أن درجات 'ز' لها ميزة أنها مألوفة بين معظم الباحثين. وهذه العملية تجعل لكل متغير نفس ميزان المتغيرات الأخرى. وإذا كانت لجميع المتغيرات نفس الميزان فليس هناك داع للتحويل إلى درجات معيارية، ويمكن استخدام الدرجات الأصلية.

خطوة ٣: وهذه هي خطوة تكوين التجمعات. وهناك طريقتان أساسيتان لتكوين التجمعات. في الطريقة الأولى ويطلق عليها طريقة التجمع الهرمي التكتلي (Agglomerative hierarchical clustering) يضع SPSS الحالات في مجموعات يزيد حجمها بالتدريج إلى أن تصبح جميع الحالات في تجمع واحد كبير. أما الطريقة الثانية وهي عكس الطريقة الأولى فيطلق عليها طريقة التجمع الهرمي التقسيمي

(Divisive hierarchical clustering)، إذ توضع جميع الحالات في تجمع واحد كبير، ثم يقسم هذا التجمع إلى تجمعات أقل إلى نصل إلى العدد المرغوب من التجمعات. والطريقة الأولى هي الطريقة الافتراضية في SPSS.

ويوجد عدد من الاختيارات لتوحيد التجمعات في داخل الطريقة التكتلية. ويطلق على الإجراء الافتراضي الربط بين المجموعات أو متوسط الربط داخل المجموعات. ويقوم SPSS بحساب أصغر مسافة متوسطة بين كل أزواج مجموعات الحالات ويوحد بين أقرب مجموعتين لبعضهما البعض. ويلاحظ أنه في المرحلة المبدئية (عندما تكون كل التجمعات مجرد حالات فردية) يكون متوسط المسافة هو المسافة المحسوبة بين أزواج الحالات. ولا ينطبق مصطلح المسافة المتوسطة إلا عندما تتكون التجمعات الفعلية. ويبدأ هذا الإجراء بعدد من التجمعات يتساوى مع عدد الحالات (٢١ جهاز تسجيل فيديو في مثالنا الحالي). وفي الخطوة الأولى يتم تجميع الحالتين الأقرب لبعضهما البعض. ويحسب SPSS المسافة مرة أخرى ويجمّع الحالتين التاليتين من حيث القرب لبعضهما البعض. ويكون لدينا بعد الخطوة الثانية إما ١٨ حالة فردية وتجمعا من ٣ حالات، أو ١٧ حالة فردية وتجمعين في كل منهما حالتين. وتستمر هذه العملية حتى تتجمع كل الحالات في تجمع واحد كبير. وهناك طرق أخرى لتجميع الحالات جاء ذكرها في دليل SPSS.

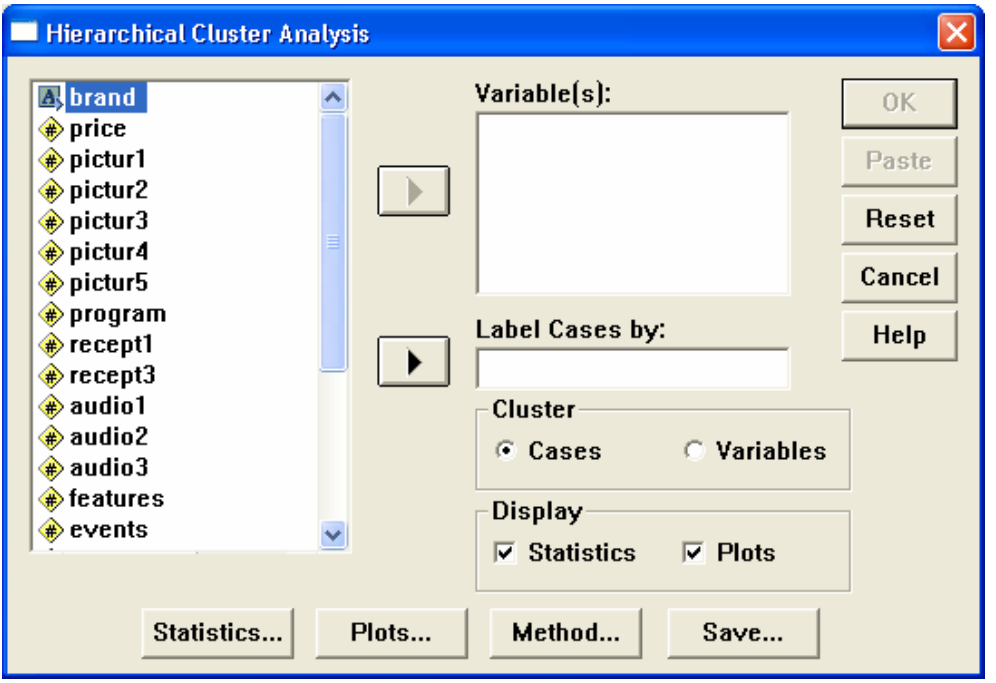
خطوة ٤: تفسير النتائج. كما هو الحال في التحليل العاملي فإن تفسير النتائج وعدد التجمعات التي يتم قبولها أمر يرجع في المرتبة الأولى للباحث. ويبدو أن أفضل عدد من التجمعات بالنسبة لملف VCR.SAV هو ثلاثة. ويبدو أن هناك ثلاث خصائص هي التي تميز بين المجموعات: فالمجموعة الأولى كانت أعلى في السعر (M = 2500)، وأعلى في جودة الصورة (M = 5)، وأكبر عدد من الصفات الإضافية (M = 10.8). وتتكون المجموعة الثانية من الأجهزة التي لها صفات متوسطة في السعر، وصورة متوسطة أو منخفضة، ولها عدد أقل من الصفات الإضافية (وكانت الدرجات ٢٠٠٠ و٣,٠ و٧,٨ على التوالي). أما المجموعة الثالثة فقد احتوت على الأجهزة رخيصة السعر (وكانت درجاتها ١٢٥٠ و ٢,٧٨ و٣,٠ على التوالي). ولا يبدو أن للخصائص (المتغيرات) الأخرى دور في التجمع بأي طريقة منتظمة.

ننتقل الآن إلى الخطوات التنفيذية لتحليل التجمع. وسوف نتبع في الخطوات الإجرائية للتحليل نفس الخطوات التي اتبعناها في وحدات SPSS الأخرى وسنبدأ أولا بطريقة التأشير والضغط.

طريقة التأشير والضغط:

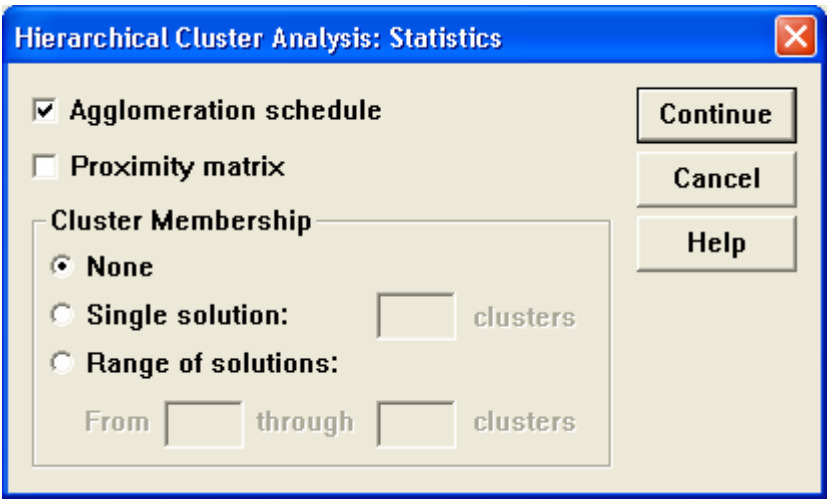
افتح ملف VCR.SAV في محرر البيانات إذا لم يكن مفتوحا، ثم نفذ خطوات التحليل التالية:

١- اضغط على **Statistics** (الإصدار الثامن) أو **Analyze** (الإصدار التاسع إلى الإصدار الحادي عشر).
٢- من القائمة المنسدلة اختبر **Classify** ثم Hierarchical Cluster.
٣- عندما يظهر مربع حوار تحليل التجمع الهرمي (شكل ٢١-١) تأكد من وجود علامة أمام Cases لأننا نريد أن يكون تحليل التجمع للحالات، وهذه سوف يتعرف عليها البرنامج من أسماء أنواع مسجلات الفيديو. ولذلك نقوم بنقل المتغير **brand** إلى مربع تسمية الحالات Label Cases by.



شكل ٢١-١ مربع حوار تحليل التجمع الهرمي

٤- في مربع الحوار الحالي اضغط على **Statistics** لإظهار مربع حوار جديد (شكل ٢١-٢)



شكل ٢١-٢ مربع حوار الإحصاء في تحليل التجمع الهرمي

ويلاحظ أن الجزء الأسفل من شكل ٢١-٢ يحتوي على ثلاثة اختيارات هي:

- لايوجد (None) ويعطي هذا الاختيار جميع التجمعات إلا أنه لا يحدد الأفراد الموجودين في كل تجمع.
- الحل المفرد، وهو يحدد عضوية التجمع بالنسبة لعدد معين من التجمعات. مثال ذلك إذا كتبت رقم ٣ في المربع ليحدد عدد التجمعات، فإن SPSS يطبع حلا لثلاثة تجمعات.
- مدى من الحلول: إذا كنت تريد أن تحصل على عدة حلول محتملة، أكتب قيمة أصغر عدد تريد أن يكون في المربع الأول، وقيمة أكبر عدد من التجمعات التي تريد أن تحصل عليها في المربع الثاني. مثال ذلك إذا كتبت ٣ و ٥ فإن SPSS سوف يبين العضوية لتجمع ثلاثي، ورباعي، وخماسي.

ويبين مربع الحوار الذي نحصل عليه بعد ذلك احتمالات إضافة أو تعديل الرسوم المصاحبة للبيانات. وإذا ضغطنا على زر الرسوم في شكل ٢١-١ يظهر مربع حوار آخر (شكل ٢١-٣). ويعطينا هذا الشكل معلومات مماثلة لتلك التي توجد في شكل ٢١-

٢ إلا أنها تضيف مقياسا نسبيا لحجم الفروق بين المتغيرات أو تجمعات كل خطوة في العملية. ويعطي SPSS رسما متدليا لعملية التجمع بأكملها وبشكل افتراضي. ونظرا لأن للخطوات الأولى في العملية تأثيرا محدودا على الحل النهائي فإن كثيرا من الباحثين يفضلون أن يظهروا مدى محدودا من التجمعات فقط، ويمكن تحقيق ذلك بالضغط على اختيار المدى الخاص للتجمعات Specific range of clusters متبوعا بكتابة أصغر عدد للتجمعات التي تهتم بها، وأكبر عدد من التجمعات التي تريدها كحل للمشكلة، والمسافات بين تلك القيم. مثال ذلك إذا كانت التي أدخلتها هي ٣، و٥، و١، سوف ترى حلولا ذات ثلاثة تجمعات، وأربعة تجمعات، وخمسة تجمعات. وإذا اخترت رقم ٢ و١٠، و٢ فإن SPSS يعطي حلولا لإثنين، وأربعة وستة وثمانية وعشرة تجمعات. وإذا ضعطنا على الاختيار **None** فإننا نلغي جميع الرسوم من النتائج. واختيار **Vertical** يؤدي عدد أكبر كثيرا من الحالات على صفحة واحدة (حوالي ٢٤) من اختيار **Horizontal** (حوالي ١٤). ومن الأفضل اختيار هذا الأخير عندما يكون لديك عدد كبير جدا من المتغيرات أو الحالات لوضعها في أعلى صفحة واحدة. وزر **Method في شكل ٢١-١** يفتح أكبر نافذة لتحليل التجمع اتساعا في هذا الشكل (شكل ٢١-٤). وهناك عدد مختلف من الاختيارات **لطريقة التجمع Cluster Method ومقياس المسافة Interval Method وتحويل القيم المقنن Transform Values: Standardize**.