استخدام أسلوب التحليل العنقودي Cluster Analysis في تصنيف محافظات جمهورية مصر العربية

د/ ماجي أحمد محمد خليل الحلواني
قسم الإحصاء والرياضة – كلية التجارة – جامعة عين شمس

الملخص:

تعتبر عملية تصنيف البيانات من أهم مراحل بناء النظريات العلمية، ويعتبر التحليل العنقودي (Cluster Analysis) أسلوبًا مناسبًا لتصنيف الظواهر، والذي يهتم بتقسيم وتصنيف عناصر بيانات المتغيرات إلى عدة مجموعات جزئية تكون متجانسة داخل المجموعة الواحدة (العنقود الواحد)، وتكون متباينة ومختلفة بالنسبة إلى المجاميع الأخرى (العناقيد الأخرى).

وقد كانت هذه الدراسة ذات فائدة كبيرة في تصنيف محافظات جمهورية مصر العربية (سبع وعشرين محافظة) إلى تجمعات متجانسة بالاستناد على مجموعة من الصفات والمتغيرات والخصائص التي تتميز بها هذه المحافظات (تم انتقاء ٣٥ متغيرًا يمثلون بعض خصائص محافظات مصر)، وتحديد مدى التقارب والتباعد بينها وكذلك دراسة العلاقة بين هذه التجمعات وأساليب التجميع وقد تم استخدام البرنامج الإحصائي (SPSS Version 20) في إجراء التحليل العنقودي على هذه البيانات.

وتم كذلك استخدام نفس الأسلوب، مرة أخرى، في تصنيف هذه الخصائص (الصفات) التي تميز محافظات جمهورية مصر العربية إلى مجموعات متجانسة وتحديد مدى التقارب والتباعد بين هذه الخصائص، وقد تم التوصل إلى النتائج التالية:

أولاً: من خلال استخدام أسلوب التحليل العنقودي في تصنيف محافظات جمهورية مصر العربية الـ ٢٧ إلى مجموعات متجانسة حسب خصائص (متغيرات) هذه المحافظات المتمثلة في ٣٥ متغيرات أمكن تصنيفها إلى أربعة عناقيد وهي كالآتي:

- العنقود الأول: شمل محافظة القاهرة فقط.
- العنقود الثاني: يشمل محافظات: الإسكندرية، الدقهلية، الشرقية، القليوبية، الغربية، البحيرة، والجيزة.
- العنقود الثالث: يشمل محافظات: بورسعيد والسويس، دمياط، الإسماعيلية، أسوان، الأقصر، البحر الأحمر، الوادي الجديد، مطروح، شمال سيناء، وجنوب سيناء.
- العنقود الرابع: يشمل محافظات: كفر الشيخ، المنوفية، بني سويف، الفيوم، المنيا، أسيوط، سوهاج، وقنا.

ثم في مرحلة تالية تم تجميعها في ثلاث عناقيد كالآتي:

- العنقود الأول: محافظة القاهرة فقط.
- العنقود الثاني: شمل نفس محافظات العنقود الأول في المرحلة السابقة وتم ضم محافظات العنقود الثاني إليه.
- العنقود الثالث: شمل نفس محافظات العنقود الثالث في المرحلة السابقة.

بعد ذلك تم تجميعها في عنقودين اثنين فقط كالآتي:

- العنقود الأول: شمل محافظة القاهرة، التي كانت تمثل العنقود الأول في المرحلة الأولى والثانية، وتم ضم محافظات العنقود الثاني في المرحلة الثانية إليه.
- العنقود الثاني: شمل محافظات العنقود الثالث في المرحلة الثانية ،التي كانت نفس محافظات العنقود الثاني في المرحلة الأولى.

ثانيًا: من خلال استخدام التحليل العنقودي في تصنيف خصائص (متغيرات) محافظات جمهورية مصر العربية المتمثلة في ٣٥ متغيرًا أمكن تصنيفها إلى أربعة عناقيد كالآتي:

- العنقود الأول: شمل ١٦ متغيرا من أصل ٣٥، أي ما نسبته ٤٥,٧% من المتغيرات.
- العنقود الثاني: شمل ١٣ متغيرا من أصل ٣٥ أي ما نسبته ٣٧,٢% من المتغيرات.
- العنقود الثالث: شمل متغيرين اثنين فقط أي ما نسبته ٥,٧% من المتغيرات.
- العنقود الرابع: شمل أربعة متغيرات أي ما نسبته ١١,٤%.

ثم في مرحلة تالية تم تجميعها في ثلاثة عناقيد كالآتي:

العنقود الأول: شمل نفس عناصر العنقود الأول في المرحلة الأولى وتم ضم عناصر العنقود الثاني إليه.

العنقود الثاني: شمل نفس عناصر العنقود الثالث في المرحلة الأولى.

العنقود الثالث: شمل نفس عناصر العنقود الرابع في المرحلة الأولى.

تم بعد ذلك تجميعها في عنقودين اثنين فقط كالآتي:

- العنقود الأول: شمل نفس عناصر العنقود الأول في المرحلة الثانية.

- العنقود الثاني: شمل نفس عناصر العنقود الثاني في المرحلة الثانية. وتم ضم عناصر العنقود الثالث في المرحلة الثانية.

الكلمات الأساسية: التحليل العنقودي، محافظات جمهورية مصر العربية، الألواح الجليدية Icicle Plot، مخطط الشجرة الثنائية Dendrogram

□قد□ـة:

في ظل التقدم التكنولوجي الهائل في كافة ميادين حياتنا الاجتماعية، لم تعد الدراسات والبحوث الإحصائية كافية لعرض المشاكل، ودراسة الظواهر، وتحديد الأسباب، واستخلاص النتائج، واتخاذ القرارات بطريقة سطحية مجردة، وبعيدة عن الموضوعية والقياس. وأصبح الاتجاه العام في مثل هذه البحوث هو استخدام طرق القياس الكمية، والمناهج الإحصائية، لتصنيف الظواهر العلمية وإبراز خصائصها، وتحليل العلاقات المتبادلة بين الظواهر على أساس موضوعي. وتعد عملية تصنيف البيانات من أهم مراحل بناء النظرية العلمية، بل وأصبح عدد من النظريات العلمية يعد في حقيقته تصنيفًا للملاحظات والمتغيرات المتعلقة بالظواهر موضوع الدراسة. ويؤدي اكتشاف أسس التصنيف وتحديدها إلى إقامة الفروض العلمية التي تختبر هذه الأسس والمتغيرات في الظاهرة ،ومنطق هذه المتغيرات، الذى يؤدى فى النهاية إلى صياغة القانون العلمي، وأسلوب التحليل الإحصائي (أسلوب تحليل متعدد المتغيرات)، وهو أحد الأساليب الإحصائية الهامة، ويستخدم في تنظيم ظاهرة جديدة نحتاج إلى التعرف على خصائصها ومتغيراتها، وهو ما يسعى إليه الباحث عندما يطرق مجالاً جديدًا لا يعرف كل متغيراته، أو مدى تعلق هذه المتغيرات المختلفة بالظاهرة الرئيسية. والنتيجة المباشرة لهذه الخطوات الاستكشافية هي إعادة دراسة المتغيرات الهامة وتناولها في الظاهرة وبناء الفروض التي تفسر العلاقات بين هذه المتغيرات.

ويعتمد أسلوب التحليل الإحصائي على نواة المشكلة محل القياس والتحليل، ونوع البيانات المتوافرة، علمًا بأن أسلوب التحليل، الذي يستخدم لدراسة مشكلة معينة، قد لا يكون مناسبًا لدراسة مشكلة أخرى. وذلك لاختلاف طبيعة المشكلات ونوع البيانات.

ونظرًا لتعدد أهداف البحوث والدراسات، وتعدد متغيراتها الاقتصادية والاجتماعية، فقد تم تحديد الأساليب الإحصائية التي تتفق مع طبيعة البيانات، وتحقق أهداف البحوث والدراسات التي من أهمها أسلوب التحليل الإحصائي متعدد المتغيرات "Multivariate Statistical Analysis" الذي يتضمن مجموعة من الأساليب التى أهمها:

١- أسلوب تحليل المكونات الرئيسية. Principal Components Analysis
٢- أسلوب التحليل العاملي. Factor Analysis
٣- أسلوب التحليل العنقودي (التجميعي). Cluster Analysis

وجدير بالذكر أن أسلوب التحليل الإحصائي متعدد المتغيرات بأساليبه المختلفة يعتمد على وصف وتحليل الظواهر ذات الأبعاد والمتغيرات المتعددة. فإذا كانت المشاهدات $[X_1, X_2, \ldots X_p]$ تشترك فيما بينها فى مجموعة من الخصائص والصفات بدرجات متفاوتة، فإن التحليل الإحصائي متعدد المتغيرات يتناول دراسة بيانات تلك المشاهدات والتعبير عنها من خلال أكثر المتغيرات تأثيرًا في الظاهرة محل الدراسة.

وسوف يتم من خلال هذا البحث التركيز على أسلوب التحليل العنقودي (التجميعي) (Cluster Analysis) باعتبار أن الكثير من الدراسات والبحوث تناول أسلوب المكونات الرئيسية، وأسلوب التحليل العاملي ، وأن النادر من الدراسات تناول أسلوب التحليل العنقودي (التجميعي).

هدف البحث:

يهدف هذا البحث إلى استخدام أحد أساليب التحليل الإحصائي متعدد المتغيرات Multivariate Analysis، وهو أسلوب التحليل العنقودي (التجميعي)، في تصنيف محافظات جمهورية مصر العربية إلى تجمعات متجانسة بالاستناد على مجموعة من الصفات والمتغيرات والخصائص التي تتميز بها هذه المحافظات، وتحديد مدى التقارب والتباعد بينها، وكذلك دراسة العلاقة بين هذه التجمعات وأساليب التجميع.

الجانب النظري للبحث:

١- التحليل العنقودى (التجميعى) Cluster Analysis:

ويعد أسلوب التحليل العنقودي الذي يعتبر أحد فروع التحليل الإحصائي متعدد المتغيرات، من الأساليب المهمة في تحليل البيانات ، إذ أنه يهتم بتجميع مفردات مجتمع البحث على شكل عنقود يبدأ متفرعًا وينتهي بغصن واحد. ويتم التجميع إما على أساس المشاهدات نفسها، في ضوء خصائص المتغيرات، وأما على أساس المتغيرات نفسها. وتبدأ فكرة التحليل العنقودي دون توفر معرفة مسبقة بعدد المجاميع أومعرفة أي من المفردات ينتمي لهذه المجموعة أو تلك. فالأسلوب يعد بالتالى استكشافياً (Exploration Approach).

ويعد التحليل العنقودى من أساليب التحليل الإحصائي التي تهتم بتقسيم وتصنيف عناصر بيانات المتغيرات إلى عدة مجموعات جزئية، تكون متجانسة داخل المجموعة الواحدة (العنقود الواحد)، وتكون متباينة ومختلفة بالنسبة إلى المجاميع الأخرى (العناقيد الأخرى). وبسبب هذه الخاصية التي يمتلكها التحليل العنقودي

(Cluster analysis)، شاع استخدام هذا النوع من التحليل الإحصائي في الكثير من المجالات التطبيقية. وعادة ما يكون الغرض من هذا التحليل هو اكتشاف نمط معين ينظم المشاهدات ، التي غالبًا ما تكون أفرادًا، ويقسمها إلى مجموعات تتمتع عناصرها بخواص مشتركة. فالتحليل العنقودي يهتم بجمع المعلومات عن المفردات، وربطها معًا بطريقة تتيح تصنيف هذه المفردات في مجموعات مفيدة. ويستخدم التحليل العنقودي، ليس فقط لتحديد التركيب الموجود في البيانات، ولكن يستخدم أيضًا لتحديد المجموعات التي ينبغي أن تقسم بطريقة عادلة. والتحليل العنقودي يختلف تمامًا عن التحليل التمييزي في أنه ينشيء المجموعات، بينما التحليل التمييزي يعزى المشاهدات إلى المجموعات التي حددت مسبقًا. وبينما يهتم التحليل العنقودي بتصنيف المفردات إلى فئات غير معروفة مسبقًا فإن التحليل العاملي يتجه نحو فحص العلاقات الارتباطية بين عدد من المتغيرات واستخلاص الأسس التصنيفية العامة بينها. و قد استخدم كثير من البحوث الأسلوبين معًا.

٢- المفاهيم الأساسية في التحليل العنقودي:

٢-١- العنصر Element:

العنصر (X_i) هو متجه لـ n من الأبعاد في فراغ المقياس

$$\underline{X_i} = (X_{i1}, X_{i2}, ..., X_{in})$$

وتعد العناصر بمثابة قيم عددية لكميات ممكنة القياس (الخواص).

٢-٢- المسافة (Distance (D:

وهي الحيز أو الفضاء الفاصل بين عنصرين، والعلاقة بين التشابه والمسافة هي علاقة عكسية، ويمكن إجراء التحليل العنقودي بالاستناد على أي منهما والصيغة الرياضية لها:

$$D_{ni}\left(\underline{X_i},\underline{X_j}\right)=\left[\sum_{w}\left(X_{iw}-X_{jw}\right)^{m}\right]^{\frac{1}{m}}$$

حيث أن:

X_i, X_j العنصران اللذان تم احتساب المسافة الأقليدية بينهما Euclidian distance (وهو مقياس المسافة الأكثر شيوعًا في الاستخدام، حيث أنه يستخدم نظرية فيثاغورث في قياس المسافة بين كل عنصرين).

X_{iw} هي المركبة w للعنصر I في فراغ القياس ذات (n) من الأبعاد.

٢-٣- العنقود Cluster:

هو عبارة عن مجموعة من العناصر elements المتجانسة، إلى حد ما، لوصف خاصية ما داخل العنقود الواحد ، والمختلفة عن العناصر داخل العناقيد الأخرى، ويعرف أيضًا بأنه مجموعة من الأشياء (Objects) المتجاورة لمجتمع إحصائي (مثل العائلة).

٢-٤- الشجرة البيانية Tree diagram OR dendogram:

هي الشكل الهرمي الناتج بعد إجراء عملية العنقدة (تكوين العناقيد) ويمكن الوصول إليها بطريقتين:

١- طريقة التجميع "Agglomerative": وتتكون هذه الطريقة من سلسلة من الخطوات التى يتم في كل خطوة منها ربط العناقيد والعناصر مع بعضها، بالاعتماد على معامل التشابه أو معامل المسافة. ويبدأ التحليل بعنقود واحد لكل حالة، ثم يتم تجميع العناقيد المتشابهة تدريجيًا حتى نصل إلى العدد المطلوب من العناقيد.

٢- طريقة التجزئة "divisive": تبدأ بتقسيم المجموعة الكبيرة التي تقع فيها العناصر إلى أجزاء صغيرة حتى تصل إلى آخر مجموعة، وهى التي تحتوي على عنصرين تم فصلهما إلى مكونات . وفي هذا النوع يؤخذ فى الأعتبار أن جميع الحالات تتجمع في عنقود واحد، ثم يتم بعد ذلك تصنيف الحالات فى عناقيد أصغر فأصغر.

وفي كلا الحالتين تكون النتائج، التي تظهرها الطريقتين، عبارة عن شجرة من التسلسل الهرمي، ويطلق علي بداية الشعبة الجذر، ويطلق علي نقاط التفرعات العقد. ولاتكون للعقدة النهائية أو الأخيرة تفرعات على الشجرة ، ويطلق عليها الأوراق، وهي تمثل العناصر التي اجتمعت مع بعضها. وكل واحدة من العقد في الشجرة ، ومن بينها الجذر، تمثل مجموعة نوعية لكل مايمكن الوصول إليه في تلك العقدة باتجاه المقدمة ومن خلال الشجرة.

٢-٥- طرق التعنقد:

إن عملية التعنقد تعني وضع العناصر في مجاميع، أي أنها تمثل الحصول على مجموعة واحدة أو أكثر من المجاميع التي تضم عددًا من العناصر التى على درجة كبيرة من التشابه فيما بينها، و تتم عملية الحصول على هذه المجاميع عن طريق تجزئة العدد الكلي للعناصر إلى مجاميع، و تضم كل مجموعة منها عددا من العناصر بالإعتماد على مصفوفة التشابه.

وهناك العديد من طرق التعنقد التي تم تطويرها لتصف شكل العلاقات بين العناصر. وتستند هذه الطرق بالأساس إلى مصفوفة المسافة أو مصفوفة الارتباط.

أولاً: طريقة الربط المنفرد Single Linkage Clustering:

تعتبر هذه الطريقة من أبسط طرق التعنقد، وتسمى أيضًا بطريقة الجوار الأقرب (The Nearest neighbor). وتعتمد هذه الطريقة بالأساس على اعتبار أن العنصرين الأكثر تشابهًا بين العناصر يشكلان نواة العنقود، ثم تضاف باقي الوحدات إلى هذه النواة بالتسلسل وحسب درجة التشابه مع عناصر نواة العنقود، حيث يضاف الأكثر تشابهًا ثم الأقل وبالتدريج. وفي حالة ربط مجموعة من العناقيد مع بعضها، يتم ذلك بالاستناد على أقرب المسافات بين عناصر العناقيد أو معاملات التماثل بين أزواج العناصر حسب الصيغة التالية:

$$D_{ij} = \text{Min}(d_{1j}, d_{2j})$$

حيث أن:

i, j : تمثل العناصر في العناقيد j , iعلى التوالي.

ثانيًا: طريقة الربط الشامل Complete Linkage Clustering:

وتسمى أيضًا بطريقة الجوار الأبعد (The farthest neighbor). وتعتمد هذه الطريقة أساسا على اعتبار أن العنصرين الأكثر تشابهًا بين العناصر، يشكل نواة العنقود، ثم يتشكل العنقود بعد ذلك بطريقة معاكسة للطريقة الأولى، حيث أنها تعتمد على الأقل تشابهًا بين العناصر، وبعبارة آخرى أن العنصر المرشح للدخول إلى العنقود، يدخل فقط إذا كانت المسافة بينه وبين أي من عناصر العنقود، تعد أكبر مسافة بينه وبين أي من عناصر العنقود. ويتم الربط بين عناصر العناقيد بناء على أبعد المسافات بينها، وهى التي تتحدد طبقًا للصيغة الآتية:

$$D_{ij} = Max(D_{ij}) \qquad i \in I\ ,\ j \in J$$

حيث أن:

i, j : تمثل العناصر في العناقيد $I\ ,\ J$ على التوالي.

ثالثًا: الطريقة الهرمية Ward Method:

هذه الطريقة تعتمد على أساس أقل فاقد من المعلومات Loss Information لعمل العنقدة، حيث إن فقدان المعلومات الناتج من تجميع المفردات إلى عناقيد، في أيه مرحلة، من مراحل التحليل، يمكن أن يقاس بواسطة المجموع الكلي لمربع انحرافات كل نقطة عن الوسط الحسابي للعنقود الذي تتبعه في كل خطوة من خطوات التحليل، ويتم الجمع بين كل عنقودين يؤدي دمجمها إلى زيادة في الحد الأدنى لمجموع مربعات الخطأ. فإذا كانت قيمة المتغير (j) للعنصر (i) في العنقود (k) يطلق عليها X_{ijk} فإن مركز العنقود (k) سيكون:

$$\left[\overline{X}_{.1k}, \overline{X}_{.2k}, ..., \overline{X}_{.nk}\right]$$

إذ إن:

$$\overline{X}_{.jk} = \frac{1}{M_k}\sum_{j=1}^{M_k} X_{ijk}$$

حيث : M_k تمثل عدد العناصر في العنقود (k)

وأهم ما يميز هذه الطريقة هو أنها لا تتطلب المعرفة المسبقة بعدد العناقيد التي سيتم تصنيف الحالات على أساسها. ويتم توزيع الحالات فى شكل عناقيد بالاعتماد على أقل مسافة بين الحالات، أو حسب معاملات الارتباط وفقًا لنوع التحليل، وهذا النوع من التحليل يناسب العينات الصغيرة نسبيًا.

وعليه فإن مجموع مربعات الخطأ الخاص بالعنقود (k) سيكون:

$$E_k = \sum_{i=1}\sum_{j=1}^{M_k}\left(X_{ijk} - \overline{X}_{.jk}\right)^2$$

ومجموع التشتت الكلي:

$$E_{total} = \sum_{k=1}^{k} E_k$$

حيث إن:

$E_{total})_t$ تمثل قياسا لمقدار المعلومات المفقودة في حالة الاستعاضة عن العناصر بمراكز العناقيد عند إجراء عملية الربط. :

و يساوي مجموع مربعات الخطأ صفرًا في بداية عملية التعنقد. ولكن نلاحظ أن مجموع مربعات الخطأ سيزداد عند ربط عنقودين.

رابعًا: طريقة الربط المتوسط

:Average Linkage method (Group Average Method)

إذا تم تعريف المسافة بين كل عنقودين بأنها متوسط المسافة بين كل زوج من المفردات التى يتم الحصول على كل واحد منها من كل عنقود، يصبح لدينا فى هذه الحالة طريقة الربط المتوسط للمجموعة. وهذه الطريقة مفضلة لدى كثير من الباحثين لأنها تميل إلى الجمع بين العناقيد ذات التباينات الصغيرة، وهى من الأساليب الأكثر استخدامًا في التحليل.

ويلاحظ أن هناك تعريفًا مختلفًا للمسافة في كل طريقة، ولذلك يتم الدمج في مستويات مختلفة. وتوجد طرق أخرى مثل الطريقة العنقودية المركزية Centroid Cluster Method والطريقة العنقودية الوسيطية Median Cluster Method، وطريقة لانس المرنة Lance and Williams Flexible Method

٢-٦- التصنيف Classification:

وهو ترتيب الأشياء استنادًا على ما بينها من تشابه أو اختلاف . و قد ترتب هذه الأشياء وفق أكثر من أسلوب،كما أنه يمكن وضع أكثر من ترتيب للعناصر أو الأشياء وذلك حسب صفة التشابه أو الاختلاف قيد الاهتمام.

خطوات التعنقد Clustering Steps وخرجات التحليل العنقودي:

١- حساب مصفوفة المسافة أو مصفوفة الإرتباط أو مصفوفة التشابه

مصفوفة التباينات الأولية The Data Matrix:

تستخدم في عملية التحليل العنقودي بيانات غالبًا ما تكون بيانات كمية (أعداد) أوبيانات نوعية أو بيانات مختلطة (Mixture data) من كلا النوعين. وتمثل البيانات بصورة أساسية مشاهدات لبعض العمليات الطبيعية في فضاء متعدد الأبعاد، إذ أن كل بعد من هذه الأبعاد يمثل خاصية معينة لوصف مشاهدات العناصر.

يمكن تمثيل مجموعة العناصر بمصفوفة ذات بعد (n×p)، حيث أن هناك n من الصفوف التى تمثل كل واحدة منها مشاهدات العناصر $Y = \{y_1, y_2, ..., y_n\}$ ، p من الأعمدة التى تمثل كل واحدة منها خاصية من خواص العناصر.

حيث أن R^p تمثل الأبعاد ، $Y_i \in R^p$ ، $y_i = (y_{i1}, y_{i2}, ... y_{ip})$ يمكن تنظيم مثل هذه البيانات في شكل المصفوفة n×p بحيث أن:

$$Y = \{y_{ij}\}, i = 1,...,n, j=1,...,p$$

ويسمى ذلك بمصفوفة التباينات الأولية، حيث الصفوف تناظر المفردات والأعمدة تناظر المتغيرات وذلك كما يلي:

$$Y = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1p} \\ y_{21} & y_{22} & \cdots & y_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{np} \end{bmatrix}$$

يتم في التحليل العنقودي تقسيم هذه المجموعة من المفردات إلى (m) من العناقيد.

و تتمثل الخطوة الأولى، عند تطبيق التحليل العنقودي، فى تحويل مصفوفة التباينات الأولية Y إلى مصفوفة (n×n) المشتقة منها، و تسمى مصفوفة القرابة (أو مصفوفة التشابه). وهى تعبر عن التشابهات وعدم التشابهات بين المفردات في مجموعة البيانات.

٢- **حساب مصفوفة القرابة Proximities Matrix:**

وهي عبارة عن مصفوفة متماثلة توضح مسافات التقارب أو التباعد بين كل زوج من المفردات ويمكن حسابها كما يلى:

أ- **للحالات (Cases):** تقدر مسافات التقارب أو التباعد بواسطة قانون يسمى مربع المسافة الإقليدية Euclidean Distance Squared ،الذي يحسب بالقانون الآتى:

$$D_{(i,j)} = \left(X_{1i} - X_{1j}\right)^2 + \left(X_{2i} - X_{2j}\right)^2$$

حيثأن X_{ij} تمثل قيم المتغيرات، ويتم تحويل البيانات الخاصة بالمتغيرات إلى قيم معيارية لتجريدها من وحدات القياس غير المتشابهة.

ب- **للمتغيرات (Variable)**: يتم هنا تقدير التباعد أو التقارب على أساس معامل الارتباط "بيرسون (Pearson Correlation Coefficient)" لجميع المتغيرات دون إجراء أي تحويل للبيانات.

فمثلاً إذا كان y_i , y_j يمثلان العنصرين ith , jth بالتعاقب فإن عناصر مصفوفة القرابة (p_{ij}) تمثل التشابه (s_{ij}) أو عدم التشابه (D_{ij}) بين y_i , y_j . ويشار إلى مصفوفة التشابه S (Similarity Matrix) ومصفوفة عدم التشابه D (Dissimilarity Matrix) بمصفوفة التقارب.

مصفوفة التشابه Similarity Matrix:

تعرف المصفوفة S ذات البعد (n×n) بمصفوفة التشابه إذا كانت متماثلة وتتصف عناصرها (S_{ij}) بالخواص الآتية:

$$S_{ii=1}$$

$$S_{ij} \le S_{ii} \qquad \text{لكل} \qquad i,j \in n \; ; \; i \neq j$$

مصفوفة المسافة Distance Matrix:

تعرف المصفوفة D ذات البعد (n×n) بمصفوفة المسافة إذا كانت المصفوفة متماثلة وتتصف عناصرها (d_{ij}) بالخواص التالية:

$$d_{ii} = 0$$

$$i,j \in n \; ; \; i \neq j \qquad \text{لكل} \qquad (d_{ij} \ge d_{ii})$$

٣- **جدول التجميع Agglomeration Schedule:**

يوضح المراحل السابقة واللاحقة لربط المفردات والمتغيرات المتقاربة حسب المسافات بينها والتي يتم إيجادها في مصفوفة القرابة.

٤- **جدول العناقيد Cluster Membership:**

يمثل توزيع المفردات والمتغيرات على عناقيد حسب الخصائص المشتركة بينها.

٥- **الألواح الجليدية Circle Plots:**

تمثيل كل مفردة بمستطيل معلق من الأعلى و تمثيل الروابط بأعمدة بين المتغيرات.

٦- **رسم الشجرة الثنائية Dendrogram:**

يصف (هذا الرسم) النتائج بشكل بياني، حيث يتم تمثيل كل مفردة بعقدة، وتمثل الفروع خطوة على سبيل توحيد مجموعتين جزئيتين تحتويان على هذه المفردة. كما يمثل هذا الرسم طول المسافة بين المجموعتين الجزئيتين عند تجمعهما.

٧- **جدول المتوسطات (المعدلات):**

ويوضح معدلات المفردات لكل مجموعة أوعنقود وفقًا للمتغيرات، وذلك لتوضيح خصائص كل عنقود وما يتميز به.

الجانب التطبيقي:

يتم في هذا الجزء من البحث، تطبيق أسلوب التحليل العنقودي على البيانات بعد أن تم التصنيف حسب محافظات جمهورية مصر العربية، بهدف إيجاد تجمعات

من المحافظات (وكذلك من الخصائص أو المتغيرات) تكون متجانسة فيما بينها من أجل التصنيف والمقارنة في آن واحد.

وعلى هذا الأساس تم تصنيف الـ (٣٥) متغير (خصائص المحافظات) حسب محافظات جمهورية مصر العربية والبالغ عددها (٢٧) محافظة.

وسوف تكون نتائج التحليل والمقارنة بين الحالات (المحافظات) قيد الدراسة بالاعتماد على المتغيرات (خصائص المحافظات) كما هي موضحة بجدول رقم (١)(*)، جدول رقم (٢)(*).

جدول رقم (١): المتغيرات المستخدمة ورموزها(خصائص المحافظات ورموزها)

	الرمز	المتغير	الرمز	المتغير	الرمز	المتغير	الرمز	المتغير	الرمز
قيمة ساعدات	X29	عدد التلاميذ		طالة		عدد وفيات		عدد السكان	

(*) استخدم الباحث نفس هذه البيانات في بحث سابق له بعنوان "استخدام التحليل العاملي في تحديد أهم خصائص محافظات جمهورية مصر العربية – المجلة العلمية للاقتصاد والتجارة – كلية التجارة – جامعة عين شمس ٢٠١٧.

وزارة التضا□ن		قبل الجا□عي	X22	□المنات إناث	X15	الرضع	X8	□الألف نسمة	X1
عدد المشروعات الصغيرة	X30	وحدات صحية حكو□ي	X23	□طالة □المنات ذكور	X16	المساحة الكلية □الكم٢	X9	عدد الإناث □الألف	X2
فرص عمل المشروعات الصغيرة	X31	وحدات صحية خاص	X24	عا□لون □القطاع الحكو□ي	X17	المساحة المأهولة □الكم٢	X10	عدد الذكور □الألف	X3
عدد المشروعات □تناهية الصغر	X32	قصور و□يوت الثقافة	X25	عدد عقود الزواج	X18	الكثافة المأهولة نسمة/ كم٢	X11	عدد المواليد	X4
فرص عمل □تناهية الصغر	X33	إجمالي□راكز الشباب والأندية	X26	عدد شهادات الطلاق	x19	قوة العمل □المنات	X12	عدد الوفيات	X5
المنصرف □المليون في □شروعات التنمية المجتمعية	X34	الجمعيات الأهلية العا□ة	X27	عدد المدارس	X20	قوة العمل □المنات إناث	X13	الزيادة الطبيعية	X6
فرص العمل في □شروعات التنمية المجتمعية	X35	عدد حالات □ساعدات وزارة التضا□ن	X28	عدد الفصول	X21	قوة العمل □المنات ذكور	X14	عدد وفيات الأطفال	X7

وسوف يتم استخدام البرنامج الإحصائي (SPSS version 20) في كل مراحل التحليل.

<u>أولاً: نتائج أسلوب التحليل العنقودي الهر□ى للمحافظات:</u>

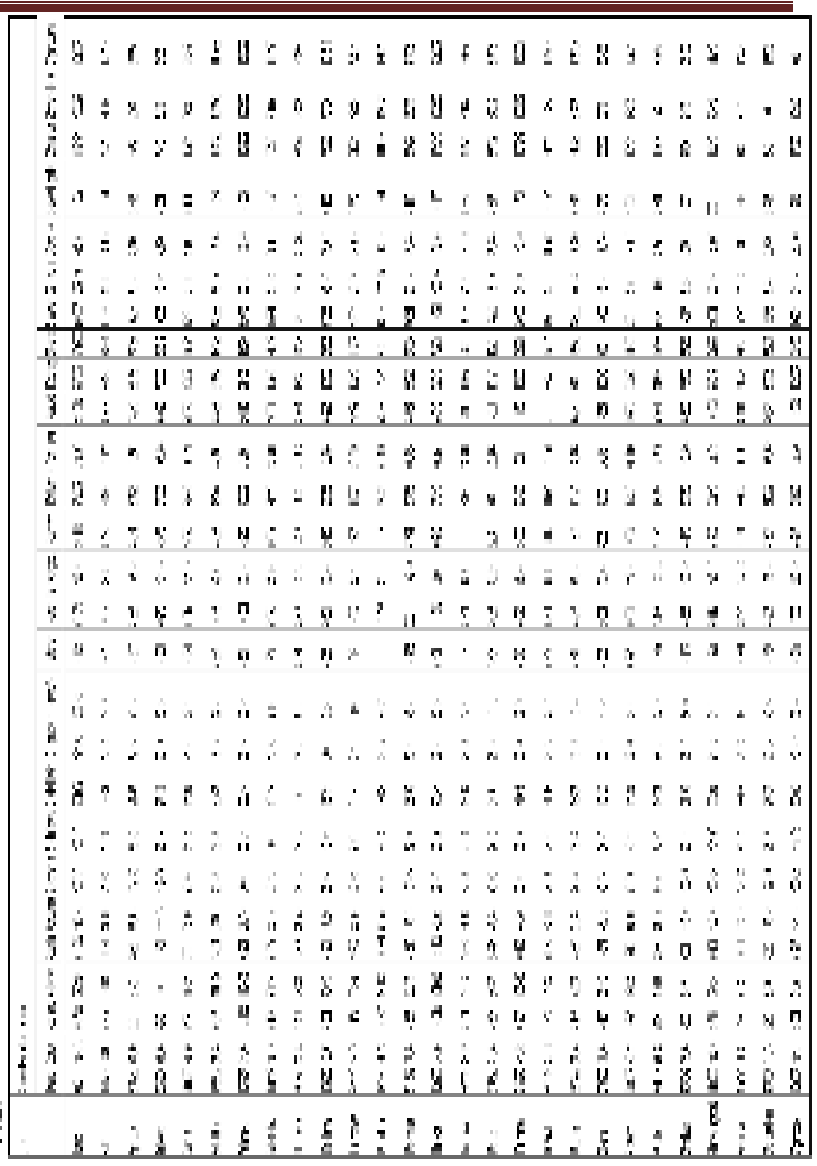
بعد تصنيف الـ (٣٥) متغير (خصائص المحافظات) حسب محافظات ج.م.ع البالغ عددها (٢٧) محافظة، تم تحويل هذه البيانات إلى شكلها القياسي (Standardization)، وذلك لجعل وحدات القياس واحدة لجميع المتغيرات ولكل المحافظات قيد الدراسة. وقد تم استخدام طريقة الربط الهرمية لإيجاد العلاقة بين المحافظات بالاعتماد على مقياس التشابه أو القرابة المستخدم Word's Method وهو مربع المسافة الإقليدية Squared Euclidian Distance

واستخدام Z scores لتحويل البيانات إلى معيارية.

والنتائج موضحة بالجدول رقم (٣-أ) وهى تمثل مصفوفة القرابة Proximity Matrix بين المحافظات.

جدول رقم (٢): بيانات المتغيرات

جدول رقم (٣-أ):□صفوفة القر□ة Proximity Matrix□ين□حافظات جمهورية□صر العر□ية



١- مصفوفة القرابة بين المحافظات:

توضح مصفوفة القرابة ذات البعد (n × n) مدى التشابه (التقارب) أو عدم التشابه (التباعد) بين كل محافظة وأخرى وفقًا لخصائص هذه المحافظات بناء على المسافات المشتقة بين كل محافظة والمحافظات الأخرى .وتم استخدام معادلة مربع المسافة الإقليدية بعد تحويل البيانات إلى القيم المعيارية لحساب هذه المسافات.

وهذه المصفوفة متماثلة نظرا لأن القيم أعلى القطر الرئيسي تتطابق مع القيم أسفل هذا القطر. وبما أن قيم المصفوفة عباره عن مسافات فإن المقياس يعبر عن عدم التشابه أو التقارب ين الوحدات. وكلما كانت المسافة كبيرة، فإنها تدل على وجود قليل من التشابه بين الوحدتين. فعلى سبيل المثال تتضح على رأس هذه المصفوفة المسافة بين محافظة القاهرة ومحافظة الإسكندرية، وهي تساوي 164.752 وتدل على تشابه إلى حد ما كبير بين المحافظتين. ويتضح ذلك عند مقارنة محافظة القاهرة مع سائر المحافظات (لاحظ قيم العمود الأول).

وبملاحظة عناصر العمود الأول أو الصف الأول (الذي يمثل المسافة بين محافظة القاهرة وسائر المحافظات) نجد ما يلي:

- أقل مسافة وهي 108.953 توجد بين محافظة القاهرة (1) ومحافظة الجيزة (14)، تليها محافظة الإسكندرية (2) المسافة 164.752، يليها محافظة الشرقية (7) المسافة 165.757، يليها محافظة الدقهلية (6) المسافة 182.592 يليها محافظة القليوبية (8) المسافة 183.117، يليها محافظة البحيرة (12) المسافة 187.432، ، ثم محافظة الغربية (10) المسافة 194.263، ثم محافظة المنيا (17) المسافة 216.973، ثم محافظة المنوفية (11) المسافة 222.443، ثم محافظة كفر الشيخ (9) المسافة 261.879، ثم محافظة قنا (20) المسافة

278.730، ثم محافظة الفيوم (16) المسافة 279.629 ثم محافظة بني سويف (15) المسافة 282.112، يليها محافظة أسوان (21) المسافة 330.067، ثم دمياط (5) المسافة 343.836، ثم الأقصر (22) المسافة 359.098، ثم الإسماعيلية (13) المسافة 371.872، ثم بورسعيد (3) المسافة 376.730، ثم السويس (4) المسافة 396.904، ثم البحر الأحمر (23) المسافة 405.342، ثم شمال سيناء (26) المسافة 408.397، ثم مطروح (25) المسافة 415.174، ثم الوادي الجديد (24) المسافة 440.566، وأخيرًا جنوب سيناء (27) المسافة 453.267.

وهكذا تتم المقارنة بين أي محافظتين في المصفوفة بناء على قدر المسافة بينهما.

ويلاحظ أن عناصر القطر الرئيسي لهذه المصفوفة = صفراً لأنه يمثل المسافة بين المحافظة مع نفسها.

وتمثل مصفوفة القرابة جدول التشابه النسبي Table of Relative Similarities بين الوحدات المستخدمة، ليتم استخدام المعلومات الناتجة من هذا الجدول لربط تلك المفردات في شكل مجموعات في الخطوات اللاحقة، ويطلق على طريقة ربط الوحدات في شكل مجموعات أسم طريقة التجميع (Clustering Algorithm)، لأنها تقوم بربط الوحدات المتشابهه مع بعضها في مجموعات منفصلة، أي أن القيم داخل هذه المصفوفة تمثل مقاييس القرابة (Measures of Similarity) أي الفروقبين أزواج الوحدات المختلفة، وهي حصيلة عدة مصفوفات متتالية.

٢- جدول التجميع Agglomeration Schedule:

جدول (٤-أ): خطوات التجميع لتحديد التقارب بهدف تكوين العناقيد بين المحافظات

Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	3	26	1.125	0	0	4
2	15	16	2.272	0	0	5
3	5	22	4.013	0	0	6
4	3	23	5.781	1	0	7
5	15	20	8.491	2	0	15
6	5	21	11.262	3	0	16
7	3	25	14.278	4	0	16
8	9	11	17.335	0	0	15
9	4	27	20.799	0	0	20
10	18	19	25.513	0	0	12
11	8	10	30.695	0	0	13
12	17	18	37.178	0	10	21
13	2	8	44.874	0	11	19
14	6	7	55.253	0	0	19
15	9	15	65.851	8	5	21
16	3	5	76.743	7	6	18
17	12	14	92.470	0	0	23
18	3	24	110.963	16	0	20
19	2	6	131.267	13	14	23
20	3	4	154.150	18	9	22
21	9	17	178.900	15	12	24
22	3	13	205.399	20	0	26
23	2	12	240.057	19	17	24
24	2	9	315.917	23	21	25
25	1	2	500.930	0	24	26
26	1	3	910.000	25	22	0

يتضمن جدول (٤-أ) خطوات التجميع حيث أنه يوضح مراحل تجميع المحافظات في عناقيد بالاستناد على المسافات بينها، التي تم إيجادها في مصفوفة القرابة. وهذا الجدول يتكون من خمسة أعمدة يوضح العمود الأول منها رقم الخطوة (Stage)، والثاني ربط المحافظات Cluster Combined، و يبين الثالث المسافة بين المحافظات المترابطة Coefficients.أما العمود الرابع فيوضح وجود أو عدم وجود المحافظة في عنقود سابق Stage Cluster First Appears.

ويمثل العمود الأخير الخطوة اللاحقة التي سوف تصادف وتظهر فيها محافظة جديدة لكل عنقود. وقد أظهرت نتائج التحليل العنقودي ، الذي يتحدد من خلاله التشابه أو عدم التشابه بين المفردات (المحافظات)، وهذا التشابه يعبر عنه بالمسافات المشتقة بين الأهداف المحددة . وقد قدر أقل فرق في المعاملات في الخطوة الأولى بين محافظة بورسعيد (المفردة 3) ومحافظة شمال سيناء (المفردة 26) حيث بلغ (1.125) وبالتالي ستقعان في العنقود الأول، ويوضح العمود الرابع عدم وجود أي من المحافظتين في أي خطوة سابقة. أما العمود الخامس فيبين أنه تم في الخطوة اللاحقة الرابعة ربط محافظة جديدة بإحدى هاتين المحافظتين وهو ما يتمثل فى ربط محافظة بورسعيد (المفردة 3) التي تم ربطها سالفًا بمحافظة شمال سيناء ومحافظة البحر الأحمر (مفردة 23).

وبالانتقال إلى الصف الثاني نجد التقارب والربط أصبح بين محافظة بني سويف (المفردة 15) ومحافظة الفيوم (مفردة 16) حيث بلغت المسافة (2.272) أي أن هاتين المحافظتين ستقعان بنفس العنقود. و يتضح من العمود الرابع عدم وجود أي من المحافظتين بأي خطوة سابقة. أما العمود الخامس فيوضح أنه سيتم فى

الخطوة اللاحقة الخامسة ربط وتقارب محافظة بني سويف (المفردة 15) بمحافظة قنا (المفردة 20).

وبالانتقال إلى الصف الثالث نجد التقارب والربط أصبح بين محافظة دمياط (المفردة 5) ومحافظة الأقصر (مفردة 22) وكان معامل الاقتراب (4.013) لتشكل عنقودًا جديدًا، ومن العمود الرابع يتضح عدم وجود أي من المحافظتين بأي خطوة سابقة، أما العمود الخامس فيظهر أنه سيتم الخطوة اللاحقة السادسة ربط وتقارب محافظة دمياط (التي ربطت سالفًا بمحافظة الأقصر) (المفردة 5) بمحافظة أسوان.

وبالانتقال إلى الصف الرابع نجد التقارب والربط بين محافظة بورسعيد (المفردة 3) ومحافظة البحر الأحمر (مفردة 23) بمعامل اقتراب (5.781) لتشكل عنقودًا جديدًا، ومن العمود الرابع يتضح وجود محافظة بورسعيد (المفردة ٣) بالخطوة رقم ١ السابق أما محافظة البحر الأحمر (مفردة 23) فلم تظهر في خطوة سابقة أما العمود الخامس فيظهر بأنه سيتم فى الخطوة اللاحقة السابعة ربط وتقارب محافظة بورسعيد (المفردة 3) (التي ربطت سالفًا بمحافظة شمال سيناء في الخطوة الأولى ومحافظة البحر الأحمر في الخطوة الرابعة) بمحافظة مطروح (مفردة 25).....

واستمرت عملية التعنقد من أبعد الأقل إلى أبعد الأكبر بشكل تصاعدي حيث كان أكبر معامل اقتراب هو (910.000) بين محافظتي القاهرة (مفردة 1)، بورسعيد (مفردة 3).

وهذا الجدول يوضح مقدار التجانس في المجموعات التي تم ضمها، فقيمة المعامل Coefficient الصغيرة تبين أن المجموعة متجانسة، بينما تبين القيمة

الكبيرة أن التجانس بين المجموعتين أقل، فعلى سبيل المثال نجد أن محافظتي بورسعيد (مفردة 3) وشمال سيناء (مفردة 26) أكثر تجانسًا من محافظة بورسعيد (مفردة 3) والبحر الأحمر (مفردة 23) لأن قيمة المعامل بين (3)، (26) هي (1.125) أقل من قيمة المعامل بين (3)، (23) وهي (5.781).

٣- جدول العناقيد Cluster Membership:

جدول (٥-أ): جدول العناقيد

Cluster Membership

Case	4 Clusters	3 Clusters	2 Clusters
1:cairo	1	1	1
2:alex	2	2	1
3:por said	3	3	2
4:elsewes	3	3	2
5:demiat	3	3	2
6:eldakhlia	2	2	1
7:elsharkia	2	2	1
8:elkaliobea	2	2	1
9:kafrelshekh	4	2	1
10:elgharbia	2	2	1
11:elmonofea	4	2	1
12:elbehera	2	2	1
13:elesmalia	3	3	2
14:elgeza	2	2	1
15:banyswif	4	2	1
16:elfaiome	4	2	1
17:elmenea	4	2	1
18:asiuwt	4	2	1
19:sohag	4	2	1
20:kena	4	2	1
21:aswan	3	3	2
22:elaksor	3	3	2
23:red sea	3	3	2
24:elwadielgaded	3	3	2
25:matrwh	3	3	2
26:shemalsena	3	3	2

Cluster Membership

Case	4 Clusters	3 Clusters	2 Clusters
1:cairo	1	1	1
2:alex	2	2	1
3:por said	3	3	2
4:elsewes	3	3	2
5:demiat	3	3	2
6:eldakhlia	2	2	1
7:elsharkia	2	2	1
8:elkaliobea	2	2	1
9:kafrelshekh	4	2	1
10:elgharbia	2	2	1
11:elmonofea	4	2	1
12:elbehera	2	2	1
13:elesmalia	3	3	2
14:elgeza	2	2	1
15:banyswif	4	2	1
16:elfaiome	4	2	1
17:elmenea	4	2	1
18:asiuwt	4	2	1
19:sohag	4	2	1
20:kena	4	2	1
21:aswan	3	3	2
22:elaksor	3	3	2
23:red sea	3	3	2
24:elwadielgaded	3	3	2
25:matrwh	3	3	2
26:shemalsena	3	3	2
27:ganobsena	3	3	2

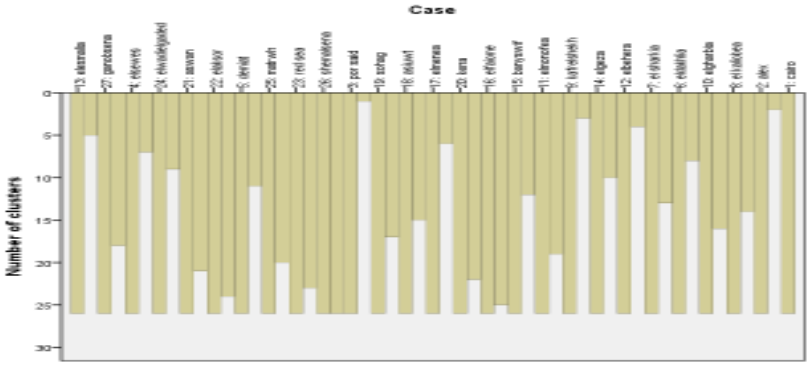
من جدول رقم (٥-أ) (جدول العناقيد) يتم توزيع المحافظات على العناقيد المتشكلة حيث يلاحظ الآتي:

- أن أهم ما يلفت النظر في عملية التعنقد هو خصوصية نمط محافظة القاهرة، حيث يلاحظ أنها لم تتفاعل مع أية محافظة أخرى خلال مراحل التعنقد. و يعكس هذا الاستنتاج طبيعة وأنماط المعيشة في هذه المحافظة التي تعد

العاصمة، وتتسم مظاهر الحياة الاقتصادية والاجتماعية والبيئية فيها بالتنوع بدرجة أكبر مما هو عليه في المحافظات الأخرى.

- وقد تجمعت سبع محافظات في عنقود واحد (العنقود الثاني) وهي محافظات: الإسكندرية، الدقهلية، الشرقية، القليوبية، الغربية، البحيرة، الجيزة، ويدل ذلك على تقارب خصائص هذه المحافظات.
- وقد تجمعت إحدى عشر محافظة في العنقود الثالث وهي محافظات: بورسعيد، السويس، دمياط، الإسماعيلية، أسوان، الأقصر، البحر الأحمر، الوادي الجديد، مطروح ، شمال سيناء، جنوب سيناء. مما يدل على تقارب خصائص هذه المحافظات.
- أما العنقود الرابع فقد ضم ثمانية محافظات وهي محافظات: كفر الشيخ، المنوفية، بني سويف، الفيوم، المنيا، أسيوط، سوهاج، قنا. مما يدل على تقارب خصائص هذه المحافظات.

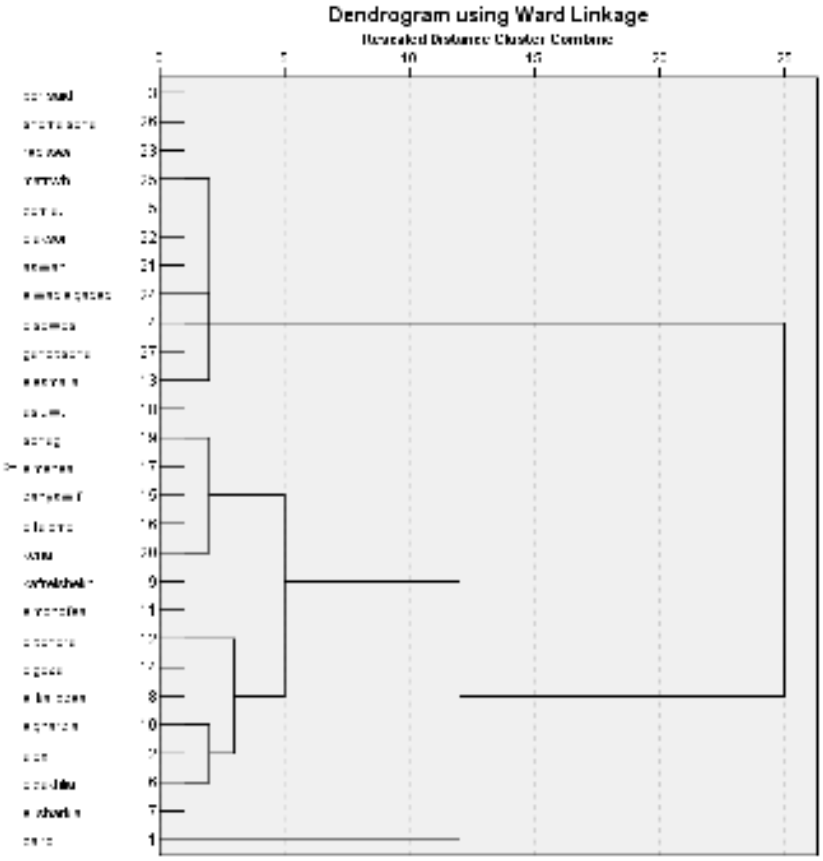
٤- <u>الألواح الجليدية Icicle Plot</u>:



شكل رقم (١-أ): الألواح الجليدية للمحافظات

الشكل رقم (١-أ) يمثل شكل الألواح الجليدية Icicle Plot .
و يتم في هذه الخطوة من التحليل تمثيل كل محافظة بمستطيل ملون (غامق) معلق من أعلى، وبأعمدة بين المتغيرات تمثل الروابط Linkage. وهو أحد مخرجات التحليل العنقودي، التي تظهر التقارب والتباعد بين مفردات التحليل، ويتضح من الشكل (١-أ) الذي يستعرض الترابط بين المحافظات أن تكوين المجموعات يتم من الأسفل. فنجد أنه قد تم في الخطوة الأولى (السطر ٢٦) ضم كل من محافظتي بورسعيد (3) وشمال سيناء (26) في مجموعة واحدة لأن بينهما أقصر مسافة، حيث أن المستطيل الملون (الغامق) بينهما يمتد إلى الأسفل إلى قاعدة الشكل، وفي الخطوة (٢٥) انضمت لهما مباشرة محافظة البحر الأحمر (٢٣) وفي الخطوة (٢٤) كل من محافظة بني سويف (١٥) والفيوم (١٦)، وفي الخطوة (٢٣) انضمت لهما محافظة قنا (٢٠) وفي الخطوة (٢٢) ... وهكذا.

٥- مخطط الشجرة الثنائية Dendrogram:



شكل رقم (٢-أ): مخطط الشجرة الثنائية للمحافظات

يوضح الشكل رقم (٢-أ) شكل الشجرة Dendrogram ،الذي يصور (يمثل) جدول خطوات التجميع Agglomeration Schedule الذي تم عرضه في

جدول رقم (٤-أ) والذي استخدمت فيه طريقة الربط الهرمية Word's Method حيث أن متابعة خطوات التجميع تكون

في أغلب الأحيان أكثر سهولة عن طريق الشجرة شكل رقم (٢-أ)، والشكل يظهر المحافظات التي تم ربطها معًا في كل خطوة من خطوات التحليل .ويتم تقسيم المسافات في أعلى هذا الشكل وقياسها بقيم تتراوح في الفترة (0 ، 25) وهي قياس لمدى قرب كل محافظة من الأخرى أو قرب المجموعات من بعضها البعض.

ويشير طول الخط إلى زيادة درجات عدم التشابه، وتوجد في الشجرة عدة عقد (Nodes)، تمثل كل عقدة منها حالة الاندماج بين المحافظات المتقاربة .و يتضح من شكل الشجرة الآتي:

- بُعد محافظة القاهرة (١) التي تمثل العنقود الأول عن باقي المحافظات.
- تليها محافظات العنقود الثاني التي تمثل، من حيث الأبعد فالأقرب، محافظة: الشرقية ثم الدقهلية، الإسكندرية، الغربية، القليوبية، الجيزة ثم البحيرة.
- تليها محافظات العنقود الرابع التي تمثل، من حيث الأبعد فالأقرب، محافظات: المنوفية، كفر الشيخ، قنا، الفيوم، بني سويف، المنيا، سوهاج ثم أسيوط.
- تليها محافظات العنقود الثالث التي تمثل، من حيث الأبعد فالأقرب، محافظات: الإسماعيلية، جنوب سيناء، السويس، الوادي الجديد، أسوان، الأقصر، دمياط، مطروح، البحر الأحمر، شمال سيناء، بورسعيد.

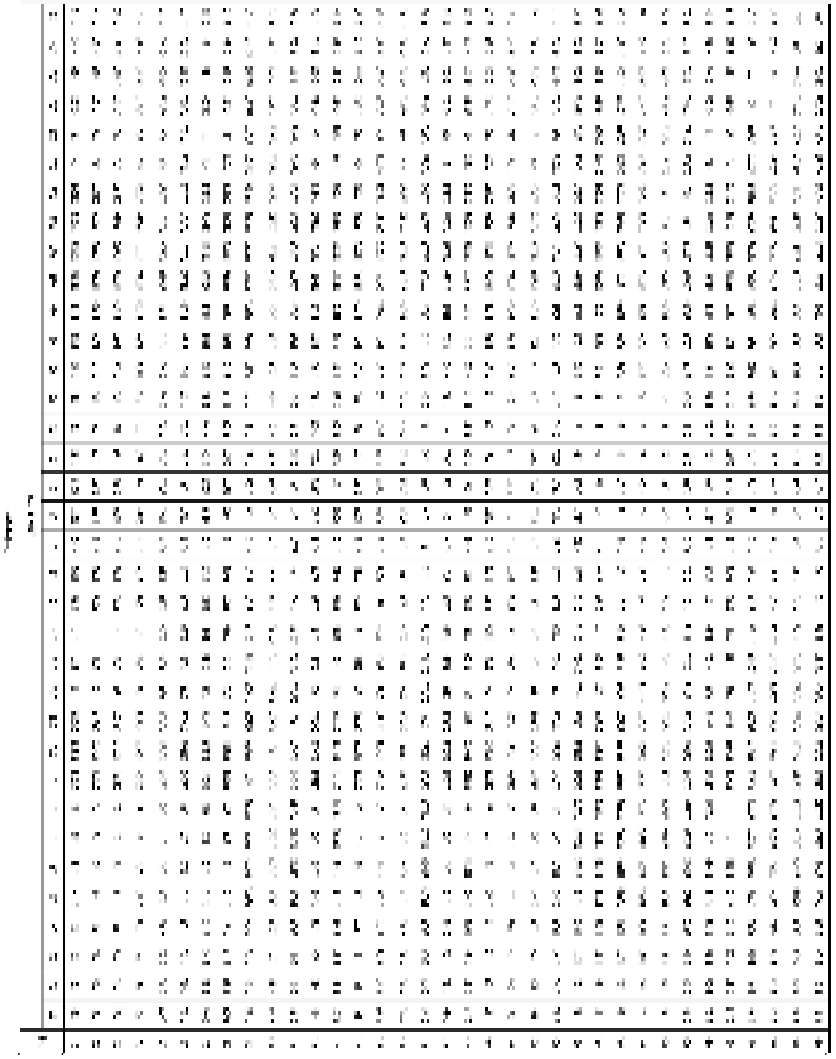
ثانيًا: نتائج أسلوب التحليل العنقودي الهرمي للمتغيرات (للخصائص):

تم إجراء التحليل بإستخدام نفس البيانات المدروسة ولكن على أساس المتغيرات (الخصائص) Variables وليس على أساس المحافظات (Cases).

١- <u>مصفوفة القرابة بين الخصائص (المتغيرات):</u>

٢- وهنا تبين المصفوفة مدى التقارب والتباعد بين الخصائص (المتغيرات) وفقًا لمربع المسافة الإقليدية، بعد تحويل البيانات إلى القيم المعيارية المقابلة (جدول ٣-ب).

جدول (٣-ب): مصفوفة القرابة بين الخصائص (المتغيرات) Proximity Matrix



١ – مصفوفة القرابة بين المتغيرات (الخصائص):

بملاحظة عناصر العمود الأول أو الصف الأول، الذي يمثل المسافة بين المتغير الأول X_1 (عدد السكان بالألف نسمة) وبين سائر المتغيرات نجد ما يلي:

- أقل مسافة وهي 0.004 بين X_1 (عدد السكان بالألف نسمة) وبين كل من X_2 (عدد الإناث بالألف)، X_3 (عدد الذكور بالألف).
- المسافة التالية وهي 0.13 بين X_1 (عدد السكان بالألف نسمة) وبين X_{14} (قوة العمل بالمئات ذكور)، يليها المسافات بين X_1 (عدد السكان بالألف نسمة) وكل من المتغيرات التالية على التوالي من الأقرب للأبعد: X_{22} (عدد التلاميذ قبل الجامعي) والمسافة 0.15، ثم X_{21} (عدد الفصول) والمسافة 0.21، ثم X_4 (عدد المواليد) والمسافة 0.55، ثم X_{12} (قوة العمل بالمئات) والمسافة 0.75، ثم X_{18} (عدد عقود الزواج) والمسافة 1.10، ثم X_{20} (عدد المدارس) والمسافة 1.19، ثم X_6 (الزيادة الطبيعية) والمسافة 1.71، ثم X_{16} (بطالة بالمئات ذكور) والمسافة 1.98 ثم X_5 (عدد الوفيات) والمسافة 2.41، ثم X_{15} (بطالة بالمئات إناث) والمسافة 5.18 ثم X_7 (عدد وفيات الأطفال) والمسافة 5.65، ثم X_{13} (قوة العمل بالمئات إناث) والمسافة 6.27، ثم X_8 (عدد وفيات الرضع) والمسافة 6.63، ثم X_{23} (وحدات صحية حكومية) والمسافة 7.19، ثم X_{19} (عدد شهادات الطلاق) والمسافة 7.22، ثم X_{31} (فرص عمل المشروعات الصغيرة) والمسافة 7.31، ثم X_{30} (عدد المشروعات الصغيرة) والمسافة 8.71، ثم X_{27} (الجمعيات الأهلية العامة) والمسافة 10.66، ثم X_{25} (قصور وبيوت الثقافة) والمسافة 11.49، ثم X_{24} (وحدات صحية قطاع خاص) والمسافة 12.90، ثم X_{26} (إجمالي مراكز الشباب والأندية) والمسافة 16.07، ثم X_{28} (عدد حالات مساعدة وزارة التضامن) والمسافة 17.35، ثم X_{17} (عاملون بالقطاع الحكومي) والمسافة 20.56 ثم X_{29} (قيمة مساعدات وزارة التضامن) والمسافة 21.08، ثم X_{11} (الكثافة

المأهولة نسمة/كم٢) والمسافة 26.39 ثم

X_{34} (المنصرف بالمليون في مشروعات التنمية المجتمعية) والمسافة 34.84، ثم X_{35} (فرص العمل في مشروعات التنمية المجتمعية) والمسافة 37.05، ثم X_{32} (عدد المشروعات متناهية الصغر) والمسافة 41.60، ثم X_{33} (فرص عمل متناهية الصغر) والمسافة 41.60، ثم X_{10} (المساحة المأهولة بالـ كم٢) والمسافة 62.56، ثم X_9 (المساحة الكلية بالـ كم٢) والمسافة 72.

وهكذا تكون المقارنة بين أي صفتين (عنصرين) في المصفوفة بناء على قيمة المسافة بينهما ويلاحظ أن عناصر القطر الرئيسي لهذه المصفوفة = صفر لأنه يمثل المسافة بين العنصر (الخاصية) وبين نفسه (نفسها).

٢- جدول التجميع Agglomeration Schedule:

جدول (٤- ب): خطوات التجميع لتحديد التقارب هدف تكوين العناقيدين المتغيرات (الخصائص)

طريقة الربط الهرمية Word's Method

Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	1	2	.002	0	0	2
2	1	3	.008	1	0	4
3	7	8	.056	0	0	22
4	1	14	.151	2	0	5
5	1	22	.286	4	0	6
6	1	21	.465	5	0	12
7	4	6	.665	0	0	18
8	5	18	1.202	0	0	14
9	12	20	1.815	0	0	12
10	28	29	2.470	0	0	31
11	30	31	3.336	0	0	26
12	1	12	4.314	6	9	18
13	32	33	5.300	0	0	29
14	5	16	6.341	8	0	20
15	19	24	7.557	0	0	24
16	11	17	8.925	0	0	28
17	34	35	10.379	0	0	29
18	1	4	12.165	12	7	21
19	15	23	14.610	0	0	20
20	5	15	17.825	14	19	22
21	1	13	22.877	18	0	27
22	5	7	28.145	20	3	24
23	25	26	33.555	0	0	25
24	5	19	41.400	22	15	26
25	25	27	50.086	23	0	27
26	5	30	61.113	24	11	28
27	1	25	78.781	21	25	31
28	5	11	96.999	26	16	32
29	32	34	121.286	13	17	33
30	9	10	149.728	0	0	33
31	1	28	180.117	27	10	32
32	1	5	224.529	31	28	34
33	9	32	277.382	30	29	34
34	1	9	409.510	32	33	0

وفقًا لهذا الجدول تم ربط (تجميع) كل من X_1 (عدد السكان بالألف نسمة)، X_2 (عدد الإناث بالألف) في العنقود الأول حيث سجلا أقصر مسافة بلغت نحو (0.002). والعمود الرابع يبين عدم ظهور هذين المتغيرين في أي خطوة سابقة. والعمود الخامس يظهر أنه في الخطوة الثانية سيتم ربط (تجميع) أحد هذين العنصرين بعنصر آخر، ومنه نجد ربط (تجميع) X_1 (عدد السكان بالألف نسمة)، X_3 (عدد الذكور بالألف) ليمثلا العنقود الثاني وبلغت المسافة بينهما (0.008)

وبالانتقال إلى الصف الثالث نجد التقارب والربط بين X_7 (عدد وفيات الأطفال)، X_8 (عدد وفيات الرضع) حيث بلغت المسافة (0.056) أي أن هذين المتغيرين يقعان في نفس العنقود، و يتضح من العمود الرابع عدم وجود أي من العنصرين بأي خطوة سابقة.أما العمود الخامس فيظهر بأنه يتم فى الخطوة اللاحقة (22) ربط (تقارب) X_7 (عدد وفيات الأطفال) بـ X_5 (عدد الوفيات) حيث أن المسافة بينهما (28.145) ومن العمود الرابع يتضح وجود المتغير X_5 في الخطوة رقم (20) والمتغير X_7 في الخطوة رقم (3). والعمود الخامس يظهر أنه يتم فى الخطوة اللاحقة (24) ربط (تجميع) المتغير X_5 (عدد الوفيات) بالمتغير X_{19} (عدد شهادات الطلاق) وهكذا واستمرت عملية التعنقد من أبعد الأقل إلى أبعد الأكبر بشكل تصاعدي حيث كان أكبر معامل اقتراب هو (409.510) بين X_1 (عدد السكان بالألف نسمة) وبين X_9 (المساحة الكلية بالـ كم^2).

٣- جدول العناقيد Cluster Membership:

جدول (٥-ب): توزيع المتغيرات (الخصائص) في عناقيد

Cluster Membership

Case	4 Clusters	3 Clusters	2 Clusters
X1	1	1	1
X2	1	1	1
X3	1	1	1
X4	1	1	1
X5	2	1	1
X6	1	1	1
X7	2	1	1
X8	2	1	1
X9	3	2	2
X10	3	2	2
X11	2	1	1
X12	1	1	1
X13	1	1	1
X14	1	1	1
X15	2	1	1
X16	2	1	1
X17	2	1	1
X18	2	1	1
X19	2	1	1
X20	1	1	1
X21	1	1	1
X22	1	1	1
X23	2	1	1
X24	2	1	1
X25	1	1	1
X26	1	1	1
X27	1	1	1
X28	1	1	1
X29	1	1	1
X30	2	1	1
X31	2	1	1
X32	4	3	2
X33	4	3	2

X34	4	3	2
X35	4	3	2

يتضح من جدول رقم (٥-ب) (جدول العناقيد) توزيع المتغيرات أو الخصائص على العناقيد المتشكلة حيث يلاحظ الآتي:

- أن العنقود الأول يشمل المتغيرات أو الخصائص الآتية:

X_1 , X_2 , X_3 , X_4 , X_6 , X_{12} , X_{13} , X_{14} , X_{20} , X_{21} , X_{22} , X_{25} , X_{26} , X_{27} , X_{28} , X_{29}

أى١٦ متغير من أصل ٣٥

- وأن العنقود الثاني يشمل المتغيرات أو الخصائص الآتية:

X_5 , X_7 , X_8 , X_{11} , X_{15} , X_{16} , X_{17} , X_{18} , X_{19} , X_{23} , X_{24} , X_{30} , X_{31}

أى ١٣ متغير من أصل ٣٥

- أن العنقود الثالث يشمل المتغيرات أو الخصائص الآتية: X_9 , X_{10}

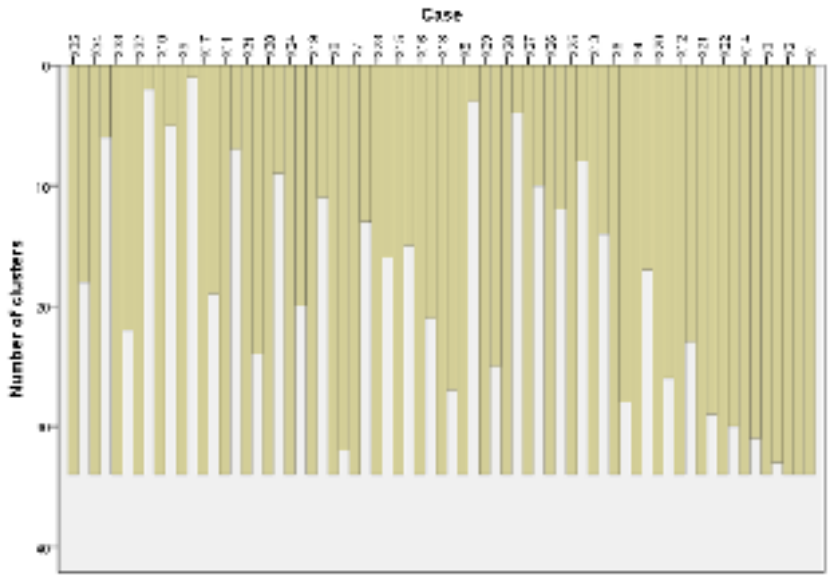
أى ٢ متغير من أصل ٣٥

- أن العنقود الرابع يشمل المتغيرات أو الخصائص الآتية:

X_{32} , X_{33} , X_{34} , X_{35}

أى ٤ متغير من أصل ٣٥

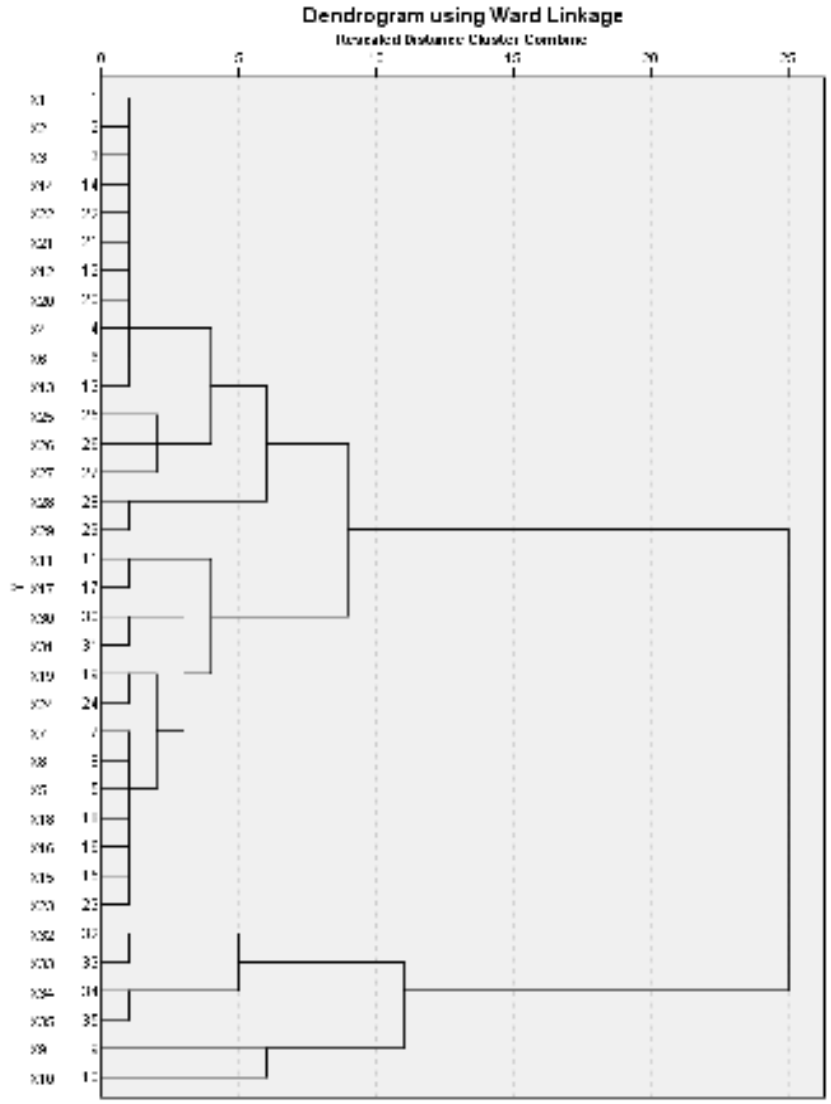
٤- الألواح الجليدية Icicle Plot:



شكل رقم (١-ب): الألواح الجليدية للمتغيرات

يتم تمثيل كل متغير بمستطيل ملون معلق من أعلى. وبالتالي يظهر هذا الشكل كما ورد سابقًا كيفية تكون العناقيد. ويتضح من الشكل (١-ب) ، الذي يستعرض الترابط بين المتغيرات (الخصائص) أنه تم تكوين المجموعات من الأسفل. فنجد أنه قد تم في الخطوة الأولى (السطر ٣٤) ضم كل من X_1 (عدد السكان بالألف نسمة)، X_2 (عدد الإناث بالألف) في مجموعة واحدة، أى عنقود واحد لأن بينهما أقصر مسافة حيث أن المستطيل الملون (الغامق) بينهما يمتد إلى الأسفل إلى قاعدة الشكل. وفي الخطوة الثانية انضم X_3 (عدد الذكور بالألف) (المستطيل الأبيض القصير) ثم X_7 ، X_8 ... وهكذا.

٥- مخطط الشجرة الثنائية Dendogram:


Dendrogram using Ward Linkage
Rescaled Distance Cluster Combine
0
5
10
15
20
25

شكل رقم (٢-ب): خطط الشجرة الثنائية للمتغيرات

في هذا المخطط أيضًا يظهر التقارب والتباعد بين الخصائص (المتغيرات) وفقًا للمسافات المحسوبة بينها ومنه نجد أن المتغيرات: X_{22} , X_{14} , X_3 , X_2 , X_1 , X_{13} , X_6 , X_4 , X_{20} , X_{12} , X_{21} أكثر قربًا من بعضها عن باقي المتغيرات، والمتغيرات X_{27} , X_{26} , X_{25} مع بعضها، و مدى بعد كل مجموعة أو قربها من المجموعات الأخرى.

وأن أبعد متغيرين (خاصيتين) هما: X_{10} , X_9 اللذين يمثلان العنقود الثالث.

الاستنتاجات والتوصيات:

من خلال الدراسة التطبيقية وهى تطبيق أسلوب التحليل العنقودي متعدد المتغيرات في تصنيف محافظات جمهورية مصر العربية إلى مجموعات متجانسة ،حسب خصائص (صفات) هذه المحافظات، وتحديد مدى التباعد والتقارب بينها.

ويتم كذلك استخدام نفس الأسلوب، مرة أخرى، في تصنيف هذه الخصائص (الصفات) التي تميز محافظات جمهورية مصر العربية، إلى مجموعات متجانسة، وتحديد مدى التباعد والتقارب بين هذه الخصائص . وقد توصل الباحث إلى جملة من الاستنتاجات والتوصيات على النحو الآتى:

أولاً :الاستنتاجات:

١- من خلال استخدام أسلوب التحليل العنقودي في تصنيف محافظات جمهورية مصر العربية الـ ٢٧ إلى مجموعات متجانسة حسب خصائص (متغيرات) هذه المحافظات المتمثلة في ٣٥ متغيرًا من المتغيرات الديموجرافية، والزراعية، والتعليمية، والصحية والخدمات الاجتماعية، والمرافق العامة، أمكن تصنيفها إلى أربعة عناقيد وهي كالآتي:

- العنقود الأول: يشمل محافظة القاهرة فقط.
- العنقود الثاني: يشمل محافظات: الإسكندرية، الدقهلية، الشرقية، القليوبية، الغربية، البحيرة، الجيزة.
- العنقود الثالث: يشمل محافظات: بورسعيد، السويس، دمياط، الإسماعيلية، أسوان، الأقصر، البحر الأحمر، الوادي الجديد، مطروح، شمال سيناء، جنوب سيناء.
- العنقود الرابع: يشمل محافظات: كفر الشيخ، المنوفية، بني سويف، الفيوم، المنيا، أسيوط، سوهاج، قنا.

٢- ثم في مرحلة تالية تم تجميعها في ثلاث عناقيد كالآتي:

- العنقود الأول: محافظة القاهرة فقط.
- العنقود الثاني: شمل نفس محافظات العنقود الأول في المرحلة السابقة وتم ضم محافظات العنقود الثاني إليه.
- العنقود الثالث: شمل نفس محافظات العنقود الثالث في المرحلة السابقة.

٣- بعد ذلك تم تجميعها في عنقودين اثنين فقط كالآتي:

- العنقود الأول: شمل محافظة القاهرة، التي كانت تمثل العنقود الأول في المرحلة الأولى والثانية، وتم ضم محافظات العنقود الثاني في المرحلة الثانية إليه.
- العنقود الثاني: شمل محافظات العنقود الثالث في المرحلة الثانية ،التي كانت نفس محافظات العنقود الثاني في المرحلة الأولى.

٤- تميزت بعض المحافظات عن غيرها باستقلاليتها نسبيًا عن غيرها من المحافظات، مثل محافظة القاهرة ، بحيث عكست خصوصية لها. ولكنها تمتاز بظروف اقتصادية واجتماعية تختلف عن غيرها من المحافظات.

٥- أما بالنسبة لاستخدام أسلوب التحليل العنقودي في تصنيف خصائص (متغيرات) محافظات جمهورية مصر العربية ،المتمثلة في ٣٥ متغيرًا من المتغيرات الديموجرافية، والزراعية، والتعليمية، والصحية، والخدمات الاجتماعية، والمرافق العامة أمكن تصنيفها إلى أربعة عناقيد كالآتي:

- العنقود الأول شمل المتغيرات (الخصائص) الآتية:

X_1 , X_2 , X_3 , X_4 , X_6 , X_{12} , X_{13} , X_{14} , X_{20} , X_{21} , X_{22} , X_{25} , X_{26} , X_{27} , X_{28} , X_{29}

أي شمل ١٦ متغيرا من أصل ٣٥، أي مانسبته ٤٥,٧% من المتغيرات.

- العنقود الثاني شمل المتغيرات (الخصائص) الآتية:

X_5 , X_7 , X_8 , X_{11} , X_{15} , X_{16} , X_{17} , X_{18} , X_{19} , X_{23} , X_{24} , X_{30} , X_{31}

أي شمل ١٣ متغيرا من أصل ٣٥ أي ما نسبته ٣٧,٢% من المتغيرات.

- العنقود الثالث شمل متغيرين اثنين فقط وهما: X_{10} , X_9 من المتغيرات أي مانسبته٥,٧% من المتغيرات.
- العنقود الرابع شمل أربعة متغيرات وهى: X_{35} , X_{34} , X_{33} , X_{32} أي شمل ٤ متغيرات من أصل ٣٥ أي ما نسبته ١١,٤%.

٦- ثم في مرحلة تالية تم تجميعها في ثلاثة عناقيد كالآتي:

- العنقود الأول: شمل نفس عناصر العنقود الأول في المرحلة الأولى وتم ضم عناصر العنقود الثاني إليه.
- العنقود الثاني: شمل نفس عناصر العنقود الثالث في المرحلة الأولى.
- العنقود الثالث: شمل نفس عناصر العنقود الرابع في المرحلة الأولى.

٧- تم بعد ذلك تجميعها في عنقودين اثنين فقط كالآتي:

- العنقود الأول: شمل نفس عناصر العنقود الأول في المرحلة الثانية.
- العنقود الثاني: شمل نفس عناصر العنقود الثاني في المرحلة الثانية. وتم ضم عناصر العنقود الثالث في المرحلة الثانية.

التوصيات:

١- يُعد التحليل العنقودي من الأساليب الإحصائية الجيدة ذات الفائدة العظيمة، ويمكن استخدامه على نطاق واسع في المجالات والأنشطة العلمية المختلفة فهو يساعد على تصنيف عدد المتغيرات الكبير في الدراسة من جهة، ويبين طبيعة العلاقة بين المتغيرات من جهة أخرى.

٢- لقد أُعدت هذه الدراسة لتشمل معظم المتغيرات (الخصائص) التي تميز محافظات جمهورية مصر العربية، بهدف تحليلها ومعرفة مدى أهمية وتأثير كل من هذه المتغيرات على كل تجمع من المحافظات.

٣- التأكيد على ضرورة استخدام الأساليب الإحصائية المتقدمة في مثل هذه الدراسات لما لهذه الدراسات من أهمية في الوصول إلى نتائج دقيقة تحقق أهدافا مرجوة لبناء مجتمع أفضل.

٤- ضرورة الاهتمام بالتدريب النظري وزيادته مع التطبيق للبرنامج الجاهز "SPSS" لما لهذا البرنامج من أهمية كبيرة في الدراسات الإحصائية من حيث اختصار الوقت، وتقليل الجهد وتحقيق نتائج أكثر دقة.

المراجـــع:

المراجع باللغة العربية:

١- أبو عساف وآخرون "دراسة تحليلية للأرقام القياسية لأسعار المستهلك في سورية وفقًا لمنهجية التحليل العنقودي". المجلة السورية للبحوث الزراعية ٤ (٢): ٣١-٥١/ يونيو ٢٠١٧.

٢- عبد الحميد محمد العباسي، "التنقيب في البيانات Data Mining تطبيقات باستخدام SPSS Modelr" معهد الدراسات والبحوث الإحصائية، جامعة القاهرة، ٢٠١٣.

٣- فريال محمود كاظم "تحليل إحصائي عن الأهداف الإنمائية للألفية باستخدام أسلوب التحليل الحاملي والتحليل العنقودي"، المعهد العربي للتدريب والبحوث الإحصائية، ٢٠٠٦.

٤- فريد الجاعوني – عدنان غانم "التحليل الإحصائي متعدد المتغيرات (التحليل التجميعي) في دراسة تحديد مستويات الهيكل الاقتصادي الاجتماعي لأسر المجتمع"، مجلة جامعة دمشق – المجلد السابع عشر – العدد الثاني – ٢٠٠١.

٥- فيصل ناجي نامق "أسلوب التحليل العنقودي لتصنيف الإنفاق على السلع والخدمات الأساسية وفقًا للمستوى البيئي (حضر وريف) للسنوات ١٩٧١-٢٠٠٧"، مجلة كلية بغداد للعلوم الاقتصادية، الجامعة العدد الخامس والعشرون، ٢٠١٠.

٦- كنان أحمد علي "فاعلية استخدام التحليل العنقودي والتحليل التمييزي في التحقق من الدلالة التمييزية لاختبارات الذكاء والشخصية" (دراسة ميدانية

مقارنة في محافظة دمشق) ماجستير، جامعة دمشق، كلية التربية، ٢٠١٤-٢٠١٥.

٧- كولالة دلشاد معروف "استخدام التحليل العنقودي في تحديد المتغيرات المؤثرة على تقييم نمو الأطفال عند مستويات العمرية المختلفة"، مجلة العلوم الاقتصادية والإدارية، العدد ٩٣ المجلد ٢٢ الصفحات (٤٦٨-٤٨٣)، ٢٠١٦.

٨- ماجي أحمد محمد خليل "استخدام التحليل العاملي في تحديد أهم خصائص محافظات جمهورية مصر العربية"، المجلة العلمية للاقتصاد والتجارة – كلية التجارة – جامعة عين شمس ٢٠١٧.

٩- مزمل الناير سومي "تحليل إمكانيات التنمية الإقليمية في السودان باستخدام التحليل العاملي والعنقودي" دكتوراه – جامعة السودان للعلوم والتكنولوجيا، كلية الدراسات العليا، ٢٠١٥.

١٠- نزار مصطفى "استخدام بعض طرق التحليل العنقودي في التصنيف مع تطبيق عملي" مجلة التقني، المجلد العشرون، العدد ٢-٢٠٠٧.

المراجع باللغة الإنجليزية:

1- Thanoon Y. Thanoon "Using Cluster Analysis and Discriminant Analysis Methods in classification with Application on factor scores Results" Tikrit journal of Administrative and Economics Sciences, Vol. (7), 2011.

2- G. David Garson Cluster Analysis "Statistical Associates Blue Book Series 24". www.amazon.con. North Carolins State University 2014.

3- NCSS Statistical Software. Chapter 445 "Hierarchial Clustering / Dendrograms.