تطبيق التحليل العاملي والتحليل العنقودي على الأرقام القياسية للأسعار الاستهلاكية في العراق: دراسة مقارنة

مروان عبد العزيز دبدوب نور نوزت سيالة

قسم الإحصاء / كلية علوم الحاسبات والرياضيات

جامعة الموصل

الاستلام القبول

٢٥ / ١٠ / ٢٠١٠ ٠٥ / ٠١ / ٢٠١١

Abstract

Factor analysis may be considered as a classifier analysis and mostly does the same function as cluster analysis. Cluster analysis and factor analysis are applied to the standard prices of certain nine consumer goods during twelve months in 2005. The results concluded from both analyses are the same.

المستخلص

يمكن أن يعد التحليل العاملي تحليلاً تصنيفياً يقترب في عمله من التحليل العنقودي، للتحقق من هذا الاعتقاد فقد تم إجراء كل من التحليلين على بيانات الأسعار القياسية للمستهلك العراقي لتسع سلع استهلاكية خلال اثني عشر شهراً لسنة ٢٠٠٥، وأعطى كل من التحليلي ن نتائج تصنيفية.

المقدمة

يعد التحليل العاملي أسلوباً إحصائياً يمكن استخدامه على نطاق واسع في العديد من المجالات والأنشطة العلمية لتحليل عدد كبير من المتغيرات بهدف الاختصار إلى اقل عدد من العوامل التي يُعزى إليها تباين هذه المتغيرات [1]. ويعد التحليل العنقودي أداة إحصائية تساعد الباحث في الكشف عن بنية البيانات متعددة المستويات والبحث عن الصفات المشتركة بين الوحدات المنفردة لإنشاء مجاميع بالاعتماد على التشابه بين أفراد كل مجمو عة ، كما يمكنه الكشف عن النقاط الشاذة في البيانات[2].

الهدف

أسلوبان قد يكون لهما نفس مفهوم التصنيف ، الأول أسلوب تصنيفي متعارف علية يدعى التحليل العنقودي ، والأسلوب الثاني هو التحليل العاملي الذي يقسم تباين الظاهر إلى مجاميع (عوامل) حسب ما يفسره كل عامل من ال تباين الكلي، في كل عامل متغيرات قد يكون لبعضها أهمية إحصائية . من هذا المفهوم وتصور العلاقة بين التحليلين وضعت فكرة احتمال إمكانية استخدام التحليل العاملي في التصنيف مقارنة بالتحليل العنقودي.

التحليل العاملي Factor analysis

إن غاية التحليل العاملي ، هي إنشا ء مجموعة جديدة من المتغيرات غير المرتبطة والمسماة بالعوامل (factors) المستخلصة من مجموعة مكونة من عدد من متغيرات مترابطة[3]. يمكن التعبير عن مجموعة المتغيرات كدالة في عدد من العوامل، إذ أن علاقة المتغيرات داخل العامل الواحد تكون أكثر ارتباطاً من علاقتها من المتغيرات في عوامل أخرى[4].

يمتاز التحليل العاملي بالمرونة، فهو يساعد على فهم العلاقة بين المتغيرات والظاهرة المدروسة من خلال عدد قليل من العوامل كما انه لا يتطلب و ضع أية فروض حول طبيعة المتغيرات قيد الدراسة ، ولهذا فقد مال الباحثون إلى استخدامه على نطاق واسع في الدراسا ت العملية لتحليل عدد كبير من المتغيرات وإرجاعها إلى عدد اقل من العوامل بحيث تفسر معظم الاختلافات الموجودة في البيانات لتعطي النموذج الملائم الذي يمثل تلك المشكلة[4].

نموذج التحليل العاملي The factor analysis model

لنكن n $(i=1,2,...,n)$ تمثل عدد المشاهدات ل. m $(j=1,2,...,m)$ من متجهات المتغيرات التوضيحية $\underline{x}_j$ ذوات متجه للأوساط الحسابية μ ومصفوفة تباين وتباين مشترك $\sum$. حتى يتم إجراء التحليل العاملي يجب أن يكون عدد أعمدة البيانات أقل من عدد الصفوف وأن تمركز مصفوفة البيانات طبقاً للأعمدة عند تحويلها إ لى مصفوفة قياسية بوسط حسابي يساوي واحد وتباين يساوي الصفر[5].

إن نموذج التحليل العاملي يحدد p من العوامل بحيث ان $(p<m)$ يرمز لها $f_1, f_2, ..., f_P$ وعامل وحيد u_j. إن كل استجابة للمتغيرات تتكون من قسم يأتي عن طريق العوامل العامة بشكل تركيبة خطية من هذه العوامل، وآخر عن طريق العامل الوحيد الذي يحتوي على التأثيرات الأخرى جميعاً الموجودة في العوامل العامة الأخرى التي عددها $(m-p)$ [6] .

$$x_j = a_{j1} f_1 + a_{j2} f_2 + ... + a_{jp} f_p + u \quad \text{..... (1)}$$

$f_1, f_2, ..., f_p$: العوامل العامة general factors التي تم اختيارها من m من المتغيرات.

a_{ip} : معامل العامل p للمتغير x_j ويسمى بتحميل (تشبع) العامل (factor loading).

u_j : العامل الوحيد.

كما يمكن صياغة النموذج باستخدام المصفوفات كالآتي:

$$\underline{x}_{m\times 1} = \underline{A}_{m\times p} \cdot \underline{f}_{P\times 1} + \underline{u}_{m\times 1} \quad \ldots\ldots\ldots(2)$$

حيث أن:

$\underline{x}$: متجه المتغيرات التوضيحية.

$\underline{A}$: مصفوفة تحميلات (تشبعات) العوامل.

$\underline{f}$: متجه العوامل العامة.

$\underline{u}$: متجه العوامل الخاصة.

تتراوح قيمة تحميل (تشبع) العامل بين الصفر والواحد، ويمثل أهمية وجود المتغير j في العامل p. وقد أشار Afifi and Clark[7] إلى كيفية اختبار الدلالة الإحصائية لتحميل العامل، وأشارا إلى اقتراب القيمة الحرجة للاختبار من ٠.٠٥.

من الأساليب الشائعة الإتباع في التحليل العاملي هو أسلوب المكونات الرئيسة بتطبيقه على مصفوفة الارتباط بين المتغيرات ، ويتم الحصول على m من المكونات الرئيسة. من ميزات المكونات الرئيسة أنها متعامدة وهذا يسهل تعين نسبة ما يفسر هـ كل مكون من التباين الكلي وبالتالي يساعد ذلك على اختيار عدد p من المكونات الرئيسة وهو بعينه اختيار عدد p من العوامل $(p<m)$. هنالك عدة طرائق لتعيين عدد العوامل المختارة ولكل منها ميزاتها وعيوبها [8]. وقد أشار Kline[9] ودبدوب والشكرجي[10] إلى أهمية رأي الباحث في اختبار عدد العوامل مع الأخذ بنظر الاعتبار ارتفاع نسبة التفسير من التباين الكلي.

يمكن الحصول على متجه $\underline{pc}_j$ يحتوي على n من قيم المكون الرئيس j:

$$\underline{pc}_j = \underline{Z}\underline{b}_j \quad \ldots\ldots(3)$$

حيث أن:

$\underline{Z}_{n\times m}$[11]: هي المصفوفة القياسية لمصفوفة البيانات $\underline{X}_{n\times m}$.

$\underline{b}_j$: هو المتجة المميز j (the j^{th} characterstic vector) لمصفوفة الارتباط بين المتغيرات.

كمّيات الشيوع Communalities

كميات الشيوع للمتغير عبارة عن مجموع مربعات تحميلات العوامل لذلك المتغير وتمثل نسبة التداخل بين المتغيرات والعوامل المشتركة[7]. ويرمز لها بـ h^2_j:

$$h_j^{\,2} = \sum_{p=1}^{m} a_{jp}^{\,2} \quad \ldots\ldots(٤)$$

فإذا كانت قيمة h_j^2 للمتغير j تقترب من الواحد ، فهذا يعني أن المتغير يتداخل بشكل كبير مع العوامل المختارة وبالتالي فله أهمية في تفسير الظاهرة المدروسة، والعكس صحيح[10].

التحليل العنقودي Cluster analysis

إن التحليل العنقودي يدقق البيانات الأولية وي ضمهم في مجاميع يطلق عليها العناقيد (Clusters)، والتي تتميّز بأن عناصرها متشابهة نسبياً، إلا أنها تختلف عن عناصر العناقيد الأخرى، وبتعبيرٍ أخر : يعمل على تصغير التباين داخل العنقود (minimization of within-group variation) وتعظيم تباين ما بين العناقيد (maximization of between-group variation)[12].

المسافة والتشابه Distance and similarity

يمكن تعريف مصطلح المسافة بأنه الفضاء الذي يفصل بين أي عنصرين، حيث أن زيادة المسافة بين أي عنصرين يعني زيادة الاختلاف بينهما وتضائل التشابه.

إن الخطوة الأولى من التحليل العنقودي هي حساب مصفوفة المسافة (distance matrix)، أعمدة هذه المصفوفة وصفوفها عبارة عن العناصر المراد إيجاد المسافة بينها، أما خلاياها $d_{jj'}$، فهي المسافة ألمقاسه بين أي زوجين من هذه العناصر، يمكن التعبير عن هذه المصفوفة بالشكل الآتي:

$$\underline{D} = [d_{jj'}] = \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1m} \\ d_{21} & d_{22} & \dots & d_{2m} \\ . & . & . & . \\ . & . & . & . \\ d_{m1} & d_{m2} & \dots & d_{mm} \end{bmatrix} \quad (j,j' = 1,2,\dots,m)$$

حيث إن $d_{jj'} = 0$ عندما $j=j'$ ، $d_{jj'} > 0$ عندما $j \neq j'$.

ويمكن حساب قيم عناصرها بالاعتماد على نظرية فيثاغورس : مربع الوتر (المسافة) يساوي مجموع مربعات الضلعين القائمين، ومن ثم اخذ الجذر ألتربيعي للصيغة الآتية[12]:

$$d_{jj'} = \sqrt{(x_{1j} - x_{1j'})^2 + (x_{2j} - x_{2j'})^2} = \sqrt{\sum_i^n (x_{ij} - x_{ij'})^2} \qquad \dots\dots (5)$$

حيث إن x_{ji} يرمز إلى المشاهدة i من المتغير j، و $x_{j'i}$ المشاهدة i من المتغير j' [6].

إن عكس مفهوم المسافة يطلق عليه معامل التشابه (similarity coefficient) وهو عبارة عن علاقة تقيس مدى التشابه بين مفردتين ، قيمته محصورة بين الصفر والواحد ، كلما

اقتربت قيمة معامل التشابه من الصفر تعني أن التشابه بين المفردتين تقترب من الانعدام وعند اقترابها من الواحد فان معامل التشابه يقترب من التشابه التام.

من طرائق الربط (linkage method) في مصفوفة المسافات هي طريقة التمركز الموزونة (wieghted piared group centroid method)، وهي تستخدم وزناً خاصاً بكل عنقود وذلك بأخذ حجم العنقود بنظر الاعتبار وتحسب وفق الصيغة الآتية[12]:

$$centriod\,(i,j) = \frac{n_i d(i,j) + n_j d(i,j)}{n_i + n_j} - \frac{n_i n_j d(i,j)}{(n_i + n_j)^2} \qquad \ldots\ldots(٦)$$

حيث إن:

ni: عدد العناصر في العنقود i.

n_j: عدد العناصر في العنقود j.

$d(i,j)$: قيمة العنصر في مصفوفة المسافات.

المخطط الشجري Dendogram

يطلق عليه أيضاً بالمخطط الهرمي (hierarchical tree diagram)، يتألف المخطط من سلسلة من الخطوات يتم في كل منها عنقدة مفردتين أو مفردة وعنقود أو عنقودين بالاعتماد على معامل التشابه أو معامل المسافة [1]، وتعرف طريقة العنقدة هذه بالتكتل (agglomerative Method). يمكن التوقف عن عملية التعنقد عندما تكون المسافات بين كل العناقيد تتخطى القيمة المخصصة d_o المسماة بمستوى عتبة اعتباطي (arbitrary threshold level)، أو عندما يحدث قفزة بين قي م المسافات أو قيم التشابه[1],[11].

جمع البيانات:

أخِذت البيانات من المجموعة الإحصائية السنوية لعامي ٢٠٠٥-٢٠٠٦ الصادرة عن الجهاز المركزي للإحصاء وتكنولوجي ا المعلومات[13]، وقد أُجري التحليلان العاملي والعنقودي على بيانات الأرقام القياسية للأسعار الاستهلاكية في العراق لتسعة أنواع من السلع خلال اثني عشر شهراً لسنة ٢٠٠٥، يبين الملحق (١) أنواع السلع (المتغيرات) وبياناتها.

استخدم لتحليل البيانات البرنامج الإحصائي Minitab v. 13.20[14]، ورتبت البيانات في البرنامج على شكل مصفوفة بنفس ترتيب الملحق (١). تمثل أعمدة مصفوفة البيانات بالتسع سلع (المتغيرات) وصفوفها باثني عشر شهرا (المشاهدات).

وجب الإشارة إلى أن عدد الأعمدة (المتغيرات أي السلع m=9) أقل من عدد الصفوف (المشاهدات أي الأشهر n=12)، ويجب أن تمركز البيانات حول المتغيرات عند تحويلها إلى وحدات قياسية عند تطبيق التحليل العاملي والتحليل العنقودي للمتغيرات.

التحليل العاملي

إن الغرض من التحليل اختيار عدد قليل من العوامل التي تفسر ما يكفي من التباين الكلي واعتبار هذه العوامل بمثابة مجاميع تصنيفية كل عامل يضم عدد من المتغيرات ذات الأهمية الإحصائية تحدد حسب قيمة تحميل ذلك العامل للمتغير المعني (factor loading) ويعتبر المتغير ذي دلاله إحصائية إذا ساوت أو زادت القيمة المطلقة للتحميل عن ٠.٥[7] وسيستخدم تدوير المحاور (varimax rotation).

نتيجة لتطبيق التحليل العاملي على مصفوفة البيانات فقد تم اختيار أربعة عوامل فسرت ٩٧.٤% من التباين الكلي (الجدول ١).

الجدول (١): نتائج التحليل العاملي وكميات الشيوع لأربعة عوامل مختارة.

المتغيرات	العامل الأول	العامل الثاني	العامل الثالث	العامل الرابع	الشيوع
x_1	0.440	-0.142	-0.132	-0.876*	0.998
x_2	0.155	-0.312	0.925*	0.120	0.992
x_3	0.916*	-0.147	0.128	-0.240	0.935
x_4	0.980*	0.053	0.044	-0.106	0.977
x_5	-0.031	-0.877*	0.430	-0.133	0.973
x_6	0.850*	-0.422	-0.006	-0.263	0.971
x_7	0.894*	0.310	0.182	-0.214	0.975
x_8	0.950*	0.017	0.002	-0.268	0.974
x_9	0.964*	-0.059	0.101	-0.191	0.980
var.%	**0.597**	**0.132**	**0.124**	**0.121**	**0.974**

*** تحميل العامل للمتغير ذو دلالة إحصائية.**

يلاحظ في الجدول (١) أن العامل الأول فسّر نسبة قدرها 59.7% من إجمالي التباين ، وقد ضمّ هذا العامل ستة سلع ذات دلالة إحصائية هي: الملابس والأحذية (X_3)، السلع المنزلية (X_4)، النقل والاتصالات (X_6)، الخدمات الطبية (X_7)، خدمات وسلع متنوعة (X_8)، والإيجار (X_9).

وفسّر العامل الثاني نسبة قدرها 13.2% من التباين الكلي ، وضم سلعة واحدة هي الوقود والإضاءة (X_5). أما العامل الثالث فقد بلغت نسبة تفسيره 12.4% من التباين الكلي

محتوياً على سلعة واحدة أيضا هي الدخان والكحوليات (X_2)، وتعدّ النسبة 12.1% هي النسبة التي فسرها العامل الرابع والذي بدوره احتوى على المواد الغذائية (X_1) فقط.

اجري التحليل العاملي مرة أخرى لاختيار ثلاثة عوامل ، نتائج التحليل تلاحظ في الجدول (٢) الذي يشير إلى أن العوامل الثلاث قد فسرت ٩٤.٤%من التباين الكلي ، و أن العامل الأول احتوى على ستة متغيرات ذات دلالة إحصائية هي : X_3 وX_4 وX_6 وX_7 وX_8 وX_9 وقد فسرت نسبة قدرها ٥٨.٩% من التباين الكلي، أما العامل الثاني فقد فسر ١٩.٤% من التباين الكلي واحتوى على متغيرين هما X2 وX5، وفسر العامل الثالث نسبة تباين قدرها ١٥.٨%، وإن X_1 هو المتغير الوحيد ذو الدلالة الإحصائية ضمن هذا العامل.

الجدول (2): تحميلات العوامل المدورة وكميات الشيوع لثلاثة عوامل.

المتغيرات	العامل الأول	العامل الثاني	العامل الثالث	كميات الشيوع
X_1	0.431	-0.026	-0.842*	0.895
X_2	0.232	0.883*	0.318	0.934
X_3	0.896*	0.201	-0.299	0.933
X_4	0.973*	0.007	-0.144	0.968
X_5	-0.112	0.916*	-0.331	0.961
X_6	0.763*	0.298	-0.496	0.918
X_7	0.964*	-0.083	-0.039	0.937
X_8	0.940*	-0.005	-0.300	0.973
X_9	0.950*	0.122	-0.236	0.976
Var.%	**0.589**	**0.197**	**0.158**	**0.944**

نتيجةً لإجراء التحليل العاملي للحصول على عاملين كانت النتائج في الجدول (٣).

الجدول (٣): تحميلات العوامل المدورة وكميات الشيوع لعاملين.

المتغيرات	العامل الأول	العامل الثاني	كميات الشيوع
X_1	٠.٦٨٥*	٠.٠٠٨	٠.٤٧٠
X_2	٠.١٠٩	٠.٨٦٥*	٠.٧٦٠
X_3	٠.٩٤٣*	٠.٢٠٥	٠.٩٣٢
X_4	٠.٩٦٦*	٠.٠٠٤	٠.٩٣٣
X_5	٠.٠٠٢	٠.٩٣١*	٠.٨٦٧
X_6	٠.٨٨٣*	٠.٣١٣	٠.٨٧٧
X_7	٠.٩٢٣*	-0.091	٠.٨٦٠
X_8	٠.٩٨٦*	-0.001	٠.٩٧٢
X_9	٠.٩٧٥*	٠.١٢٣	٠.٩٦٦

Var.%	٠.٦٥١	٠.١٩٨	٠.٨٤٩

وجِد أن العاملين قد فسرا نسبة ٨٤.٩% من التباين الكلي ، إذ فسر العامل الأول ٦٥.١% منه وكانت المتغيرات ذات الدلالة الإحصائية في العامل الأول هي : X_1 وX_3 وX_4 وX_6 وX_7 وX_8 وX_9، أما العامل الثاني فقد احتوى على المتغيرين X_2 و X_5 وبنسبة تباين قدرها ١٩.٨% من التباين الكلي.

يلخص الجدول (4) نتائج التحليل العاملي عند اختيار عدد من العوامل هي : أربعة ثم ثلاثة وأخيراً اثنان التي يشير إليها الجداول (1) و(2) و(3).

الجدول (4): ملخص نتائج ثلاثة تحاليل عاملية.

عدد العوامل	تسلسل العامل	المتغيرات ذات الأهمية
أربعة عوامل	العامل الأول	X_9،X_8،X_7،X_6،X_4،X_3
	العامل الثاني	X_5
	العامل الثالث	X_2
	العامل الرابع	X_1
ثلاثة عوامل	العامل الأول	X_9،X_8،X_7،X_6،X_4،X_3
	العامل الثاني	X_5،X_2
	العامل الثالث	X_1
عاملان	العامل الأول	X_9،X_8،X_7،X_6،X_4،X_3،X_1
	العامل الثاني	X_5،X_2

التحليل العنقودي

إن مصفوفة البيانات $\underline{X}_{12\times9}$ قد حُولت إلى مصفوفة قياسية $\underline{Z}_{12\times9}$ [11]على وفق الأعمدة التي تحتوي على السلع وسيطبق التحليل العنقودي على هذه المصفوفة بطريقتين هما:

أ- عَد السلع متغيرات وبذلك ستستخدم المصفوفة $\underline{Z}_{12\times9}$ في التحليل العنقودي للمتغيرات **(cluster variable).**

ب- عَد السلع مشاهدات وس يطبق التحليل العنقودي للمشاهدات (cluster observations) على المصفوفة $\underline{Z}_{9\times12}$.

وفي كلا التحليلين سيتم الاعتماد على أسلوب التكتل (agglomerative method) في العنقدة، وطريقة التمركز الموزونة على وفق المعادلة (6) في ربط العناقيد.

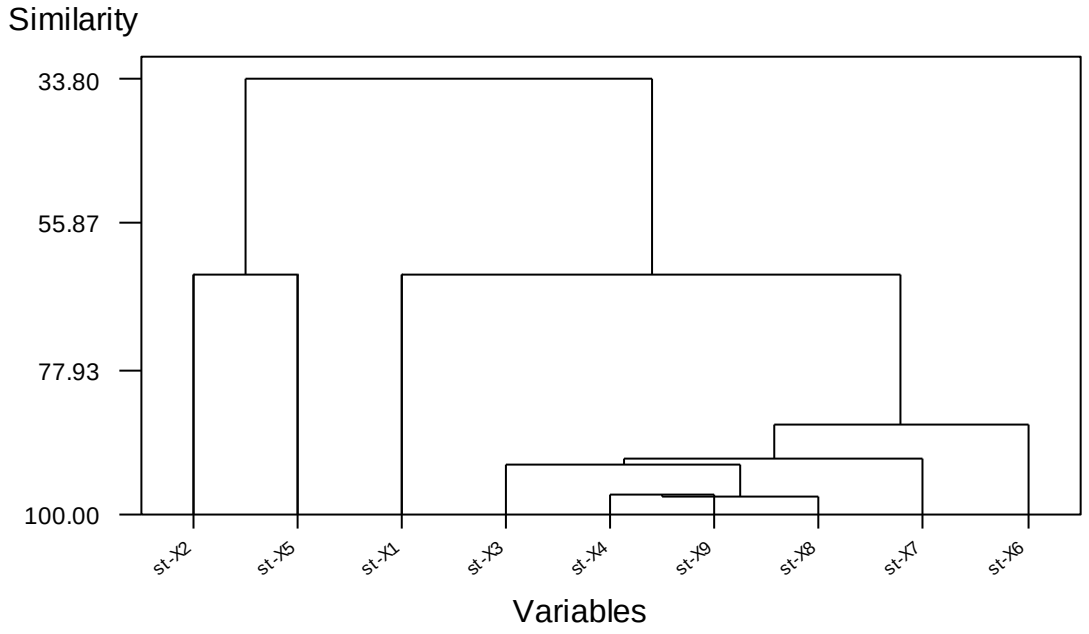
أ- التحليل العنقودي للسلع على اعتبارها متغيرات

من خلال التحليل العنقودي للمتغيرات وبا لاعتماد على Absolute correlation في حساب مصفوفة المسافة (الملحق ٢)، يوضح الجدول (5) عملية العنقدة للسلع من بداية العنقدة بوجود تسعة عناقيد (باعتبار كل سلعة منفردة هي عنقود) وحتى الحصول على عنقود واحد يحتوي على السلع جميعاً.

الجدول (5): نتائج التحليل العنقودي للسلع باستخدام (cluster variable).

رقم الخطوة	رقم العنقود	نسبة التشابه	نسبة المسافة	العنقودان المندمجين		العنقود الجديد	عدد المتغيرات في العنقود الجديد
				الأول	الثاني		
١	٨	٩٦.٩٦	٠.٠٣٠	٩	٤	٤	٢
٢	٧	٩٧.٤٦	٠.٠٢٥	٨	٤	٤	٣
٣	٦	٩٢.٦٠	٠.٠٧٤	٤	٣	٣	٤
٤	٥	٩١.٤٧	٠.٠٨٥	٧	٣	٣	٥
٥	٤	٨٦.٣٦	٠.١٣٦	٦	٣	٣	٦
٦	٣	٦٣.٦٨	٠.٣٦٣	٥	٢	٢	٢
٧	٢	٦٣.٥١	٠.٣٦٥	٣	١	١	٧
٨	١	٣٣.٨٠	٠.٦٦٢	٢	١	١	٩

يوضح الشكل الهرمي (١) سلسلة تكوين العناقيد مبتدأ بتسعة عناقيد حتى عنقود واحد.



الشكل الهرمي (١): التحليل العنقودي للسلع (متغيرات) وسلسلة تكوين العناقيد.

أنشأت أربعة عناقيد تقابل عدد العوامل المختارة في التحليل العاملي (الجدول ٥)، يلاحظ حدوث فجوة في نسبة التشابه عند الانتقال من الخطوة الخامسة التي كانت نسبة التشابه عندها تساوي ٨٦.٣٦% إلى الخطوة السادسة التي نسبة التشابه فيها يساوي ٦٣.٦٨%، عند هذه الفجوة يلاحظ الحصول على أربعة عناقيد الأول يضم المتغيرات : X_3 وX_4 وX_6 وX_7 وX_8 وX_9، والثلاث عناقيد التالية كل منها يضم متغير واحد وهي على التوالي : X_5 وX_2 وX_1.

عند تكوين ثلاثة عناقيد بقي العنقود ا لأول محافظاً على متغيراته: X_3 وX_4 وX_6 وX_7 وX_8 وX_9، أما العنقود الثاني فقد انضم المتغير X_2 إليه ليصبح حاوياً على متغيرين هما X_2 وX_5، وبقي العنقود الثالث يمثله المتغير X_1 منفرداً.

في حالة تكوين العنقودين احتوى العنقود الأول على المتغيرات :X_1 وX_3 وX_4 وX_6 وX_7 وX_8 وX_9، والعنقود الثاني ضم: X_2 وX_5.

يوضح الجدول (٦) ما تحتويه العناقيد من متغيرات عند تكوين أربعة عناقيد وثلاثة عناقيد وعنقودين.

الجدول (6): ملخص لثلاث نتائج للتحليل العنقودي بإتباع cluster variables.

عدد العناقيد	تسلسل العنقود	المتغيرات الظاهرة في العرقود
أربعة عناقيد	العنقود الأول	X_9،X_8،X_7،X_6،X_4،X_3
	العنقود الثاني	X_5
	العنقود الثالث	X_2
	العنقود الرابع	X_1
ثلاثة عناقيد	العنقود الأول	X_9،X_8،X_7،X_6،X_4،X_3
	العنقود الثاني	X_5،X_2
	العنقود الثالث	X_1
عنقودان	العنقود الأول	X_9،X_8،X_7،X_6،X_4،X_3،X_1
	العنقود الثني	X_5،X_2

بالمقارنة بين نتائج التحليل العاملي (الجدول ٤) ونتائج التحليل العنقودي (الجدول ٦) يلاحظ تطابق نتائجهما تماماً من حيث اختيار عدد العوامل المناظرة لعدد العناقيد ، فضلا عن أن كل عامل احتوى على عدد من المتغيرات ذ وات دلالة إحصائية هي ذات المتغيرات التي ضمتها العناقيد المناظرة، فكلا التحليلين أعطيا نفس النتائج عندما طبقا على المتغيرات.

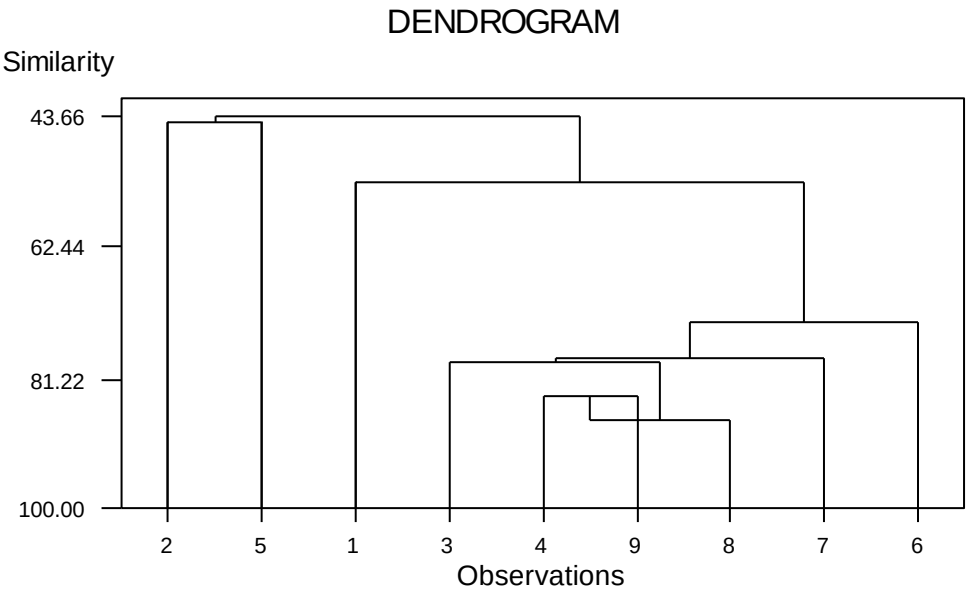
ب- التحليل العنقودي للسلع باعتبارها مشاهدات

من اجل تحليل السلع على أنها مشاهدات وللحفاظ على التمركز حول هذه السلع ، فلا بُد من استخدام مبدول مصفوفة البيانات القياسية ($\underline{Z}_{9\times12}$)[11] ومن ثم يطبق التحليل العنقودي باستخدام cluster observation عليها . كما في التحليلين السابقين ستكون أعداد العناقيد : أربعة ثم ثلاثة ثم اثنين، ستنشأ مصفوفة المسافة باستخدام المسافة الإقليدية (الملحق ٣)، وسيتم الربط بطريقة التمركز الموزونة (المعادلة ٦). يوضح الجدول (٧) نتائج التحليل العنقودي.

الجدول (7): يبين نتائج عمليات العنقدة للسلع على اعتبارها مشاهدات

رقم الخطوة	رقم العنقود	نسبة التشابه	نسبة المسافة	العنقودان المندمجين		العنقود الجديد	عدد المشاهدات في العنقود الجديد
				الأول	الثاني		
١	٨	٨٣.٩١	٠.٨١٨	٤	٩	٤	٢
٢	٧	٨٧.٢٧	٠.٦٤٧	٤	٨	٤	٣
٣	٦	٧٨.٩٥	١.٠٧١	٣	٤	٣	٤
٤	٥	٧٨.٢٨	١.١٠٤	٣	٧	٣	٥
٥	٤	٧٣.١٣	١.٣٦٦	٣	٦	٣	٦
٦	٣	٥٣.١٣	٢.٣٨٣	١	٣	١	٧
٧	٢	٤٤.٤١	٢.٨٢٧	٢	٥	٢	٢
٨	١	٤٣.٦٦	٢.٨٦٥	١	٢	١	٩

يوضح الشكل الهرمي (٢) سلسلة تكوين العناقيد مبتدأ بتسعة عناقيد حتى عنقود واحد.



الشكل الهرمي (٢): التحليل العنقودي للسلع (مشاهدات) وسلسلة تكوين العناقيد.

يلخص الجدول (٨) ما تحتويه العناقيد من متغيرات عند تكوين أربع عناقيد وثلاث عناقيد وعنقودين عند اعتبار السلع مشاهدات.

الجدول (8): يبين نتائج التحليل العنقودي عند اعتبار السلع مشاهدات

عدد العناقيد	تسلسل العنقود	المشاهدات الظاهرة في العنقود
أربعة عناقيد	العنقود الأول	X_9،X_8،X_7،X_6،X_4،X_3
	العنقود الثاني	X_5
	العنقود الثالث	X_2
	العنقود الرابع	X_1
ثلاثة عناقيد	العنقود الأول	X_9،X_8،X_7،X_6،X_4،X_3،X_1
	العنقود الثاني	X_5
	العنقود الثالث	X_2
عنقودان	العنقود الأول	X_9،X_8،X_7،X_6،X_4،X_3،X_1
	العنقود الثاني	X_5،X_2

إن نتائج التحليل العنقودي عند عد السلع مشاهدات (الجدول ٨) قد انحرفت في حالة تكوين ثلاثة عناقيد مقارنة بنتائج الجدولين (٤) و(٦)، قد يعزى السبب إلى ما يفرضه التحليل في طريقة حساب مصفوفة المسافة، فقد استُخدم الارتباط المطلَق في التحليل العنقودي للسلع كمتغيرات (ملحق ٢) بينما استخدمت المسافة الاقليدية عند عد السلع مشاهدات (الملحق ٣).

كما أن التحليلين العاملي والعنقودي أجريا على المتغيرات باتجاه الأعمدة، بينما التحليل العنقودي للسلع فقد حُللت باتجاه الصفوف.

يعد التقارب الكبير الحاصل في بعض نسب التشابه سبباً أخر من مسببات اختلاف نتائج التصنيف عند تغير اتجاهه . فلو لاحظت الخطوتين (٦) و(٧) في الجدول (٥) لتبين أن الفرق في نسبة التشابه بين عنقدة X_2 مع X_5 (الخطوة ٦) وعنقدة X_1 مع العنقود X_3 (الخطوة ٧) يساوي ٠.١٧%، إن هذا الفرق الضئيل مع المؤثرات على النتائج المشار إليها أعلاه سببت أن يحدث تبادل بين الخطوتين (٦) و(٧) عند الانتقال من الجدول (٥) إلى الجدول (٧)، ففي الجدول (٧) تصبح الخطوة (٦) هي العنقدة بين X_1 والعنقود X_3 والخطوة (٧) تعنقد فيها X_2 مع X_5. ومن مقارنة المخططين (١) و(٢) يمكن ملاحظة ما سببه التشابه من اختلاف عند تكوين ثلاثة عناقيد وذلك بملاحظة العلاقة بين X_1 وX_2 وX_5.

الاستنتاجات والتوصيات

اختصرت نتائج التحليل العاملي والتحليلين العنقوديين في الجدول (9). ونوصي بإمكانية استخدام التحليل العاملي في التصنيف حيث يعطي نفس نتائج التحليل العنقودي خاصة عند تطبيق التحليلين في اتجاه الأعمدة، إذ يجب أن يكون عدد الأعمدة أقل من عدد الصفوف. وعند تطبيق التحليل العنقودي بأسلوب cluster observation على الباحث ملاحظة قيم التشابه المتقاربة فقد تؤثر على النتائج عند تغيير اتجاه التحليل.

الجدول (9): ملخص نتائج التحاليل الثلاثة.

التحليل العاملي			التحليل العنقودي السلع متغيرات			التحليل العنقودي السلع مشاهدات		
عدد العوامل	رقم العامل	المتغيرات المعنوية	عدد العناقيد	رقم العنقود	المتغيرات المعنوية	عدد العناقيد	رقم العنقود	المتغيرات المعوية
٤	١	X_3، X_4، X_6، X_7، X_8، X_9	٤	١	X_3، X_4، X_6، X_7، X_8، X_9	٤	١	X_3، X_4، X_6، X_7، X_8، X_9
	٢	X_5		٢	X_5		٢	X_5
	٣	X_2		٣	X_2		٣	X_2
	٤	X_1		٤	X_1		٤	X_1
٣	١	X_3، X_4، X_6، X_7، X_8، X_9	٣	١	X_3، X_4، X_6، X_7، X_8، X_9	٣	١	X_1، X_3، X_4، X_6، X_7، X_8، X_9
	٢	X_2 ، X_5		٢	X_2، X_5		٢	X_5

X_2	٣		X_1	٣		X_1	٣	
X_1، X_3، X_4 ، X_6، X_7، X_8 ، X_9	١	٢	X_1، X_3، X_4 ، X_6، X_7، X_8 ، X_9	١	٢	X_1، X_3، X_4 ، X_6، X_7، X_8 ، X_9	١	٢
X_2، X_5	٢		X_2، X_5	٢		X_2 ،X_5	٢	

المصادر

١) يلدا، هدى قرداغ، (٢٠٠٠)، "استخدام التحليل العاملي لدراسة العوامل المؤثرة لبعض أمراض الجهاز العصبي"، رسالة ماجستير، قسم الإحصاء، كلية الإدارة والاقتصاد، جامعة صلاح الدين.

2) Wang, J. and Opsomer, J.C; (2001); "Cluster Analysis and its Application in the National Resources Inventory"; Department of Statistics, Iowa State University; U.S.A..

3) Johnson, D. E; (1998); "Applied Multivariate Methods for Data Analysis"; Kansas Stat University; USA.

٤) العلاف، مهدي محسن، (١٩٨٢)، "استخدام التحليل العاملي (طريقة الإمكان الأعظم) في تحليل وتفسير بعض نت ائج المسح الجيولوجي في العراق "، رسالة ماجستير، قسم الإحصاء، كلية الإدارة والاقتصاد، جامعة بغداد.

5) Agresti, A. and Franklin, C.; 2007; "Statistics The Art And Science Of Learning From Data"; Pearson Prentice-Hall; U.S.A..

٦) كريم، ژيان إحسان ، (١٩٩٢)، "استخدام التحليل العنقودي في تقييم الأراضي للأغراض الزراعية"، رسالة ماجستير، قسم الإحصاء، كلية الإدارة والاقتصاد، جامعة صلاح الدين.

7) Afifi, A.A and V. Clark; (1984); "Computer-Aided Multivariate Analysis"; life time learning publications; California; USA.

٨) الجبوري، ش‌لال حبيب وعبد ، صلاح حمزة، (٢٠٠٠)، "تحليل متعدد المتغيرات "، دار الكتب للطباعة والنشر، بغداد.

9) Kline, P.; 1994; "An Easy Guide To Factor Analysis"; TJ Press (Padstow) Ltd, Cornwall; G. B..

١٠) دبدوب، مروان عبد العزيز والشكرجي، ذنون يونس، (٢٠٠٧)، "الانتقال المختصر بين نتائج R-mode وQ-mode في التحليل العاملي"، المجلة العراقية للعلوم الإحصائية، كلية علوم الحاسبات والرياضيات، جامعة الموصل، العدد ١١، المجلد ٧، ص ١٥٣-١٧٤.

11) Berenson, M. L.; Levine, D. M. and Krehbiel, T. C.; 2006; "Basic Business Statistics Concepts and Applications"; 10th Edition; Pearson Prentice Hall; U.SA..

12) Rao, A. R. and Srinivas, V. V.; (2008); "Regionalization of Watersheds an approach Based on Cluster Analysis"; Springer Science Business Media B.V.; USA.

١٣) الجهاز المركزي للإحصاء وتكنولوجي ا المعلومات، (٢٠٠٧)، "المجموعة الإحصائية السنوية لعامي ٢٠٠٥-٢٠٠٦"، وزارة التخطيط والتعاون الإنمائي، جمهورية العراق.

14) Carver, R.; 1999; "Doing Data Analysis With Minitab"; Duxbury Press, U.S.A..

الملحق (1): الأرقام القياسية للأسعار الاستهلاكية في العراق حسب الأشهر لسنة 2005

الشهر / المادة	المواد الغذائية (x_1)	الدخان والكحوليات (x_2)	والملابس والأحذية (x_3)	السلع المنزلية (x_4)	الوقود والإضاءة (x_5)	النقل والاتصالات (x_6)	الخدمات الطبية (x_7)	خدمات و سلع متنوعة (x_8)	الإيجار (x_9)
كانون الثاني	5116.7	1281.3	3130.9	3398.5	115582.9	9068.6	33369.9	7409.1	43636.7
شباط	5321.4	1249.6	3286.9	3407.6	89103.4	9251.0	33571.3	7675.2	44490.5
آذار	5534.2	1184.5	3201.4	3426.5	47373.6	9071.1	33960.0	7675.2	44935.4
نيسان	5904.1	1121.1	3230.9	3564.5	40274.9	9326.4	34972.8	8417.7	46146.9
أيار	5449.2	1187.3	3396.3	3506.4	35744.2	9735.7	34380.3	8047.9	47208.3
حزيران	5436.6	1193.2	3370.8	3652.0	33627.4	9752.9	35765.3	8297.2	48430.6
تموز	4739.6	1232.4	3461.5	3671.8	36546.6	9798.9	35654.5	8467.6	49060.2
آب	3511.6	1249.7	3330.2	3687.2	64934.9	11226.0	34938.8	8472.4	50335.8
أيلول	5676.0	1208.9	3472.2	3667.4	60717.3	10074.5	35303.8	8459.0	50990.1
تشرين الأول	5864.9	1145.3	3637.8	3730.9	76562.6	14419.6	34950.8	8882.0	51958.9
تشرين الثاني	6259.0	1279.2	3789.5	3784.7	80571.9	13822.3	36550.5	9137.9	54089.2
كانون الأول	6084.2	1282.5	3856.0	3867.1	88897.4	15297.7	36851.7	9637.4	55657.0

الملحق(٢): مصفوفة مسافة الارتباط المطلق Matrix Of The Absolute Correlation Distance.

$$\underline{D}_{9\times9} = \begin{bmatrix} 0.00000000 & 0.88891850 & 0.38565487 & 0.48831069 & 0.83331491 & 0.33205204 & 0.49101213 & 0.35271436 & 0.40929413 \\ 0.88891850 & 0.00000000 & 0.71730463 & 0.84249508 & 0.36323644 & 0.76217087 & 0.82610189 & 0.89455721 & 0.75648398 \\ 0.38565487 & 0.71730463 & 0.00000000 & 0.11466866 & 0.83295399 & 0.08960348 & 0.16126177 & 0.08414395 & 0.05533375 \\ 0.48831069 & 0.84249508 & 0.11466866 & 0.00000000 & 0.97271927 & 0.17045745 & 0.06715201 & 0.03221422 & 0.03042388 \\ 0.83331491 & 0.36323644 & 0.83295399 & 0.97271927 & 0.00000000 & 0.64405090 & 0.82449565 & 0.99199393 & 0.91580751 \\ 0.33205204 & 0.76217087 & 0.08960348 & 0.17045745 & 0.64405090 & 0.0000000 & 0.31702173 & 0.14883789 & 0.11171014 \\ 0.49101213 & 0.82610189 & 0.16126177 & 0.06715201 & 0.82449565 & 0.31702173 & 0.00000000 & 0.08654909 & 0.11376737 \\ 0.35271436 & 0.89455721 & 0.08414395 & 0.03221422 & 0.99199393 & 0.14883789 & 0.08654909 & 0.0000000 & 0.03377032 \\ 0.40929413 & 0.75648398 & 0.05533375 & 0.03042388 & 0.91580751 & 0.11171014 & 0.11376737 & 0.03377032 & 0.00000000 \end{bmatrix}$$

الملحق (٣): مصفوفة المسافة الإقليدية Matrix Of Euclidean Distance.

$$\underline{D}_{9\times9} = \begin{bmatrix} 0.0000000 & 4.9440664 & 2.9128006 & 3.2776265 & 4.2816969 & 2.7028032 & 3.2866802 & 2.7856267 & 3.0007451 \\ 4.9440664 & 0.0000000 & 3.9724932 & 4.3052168 & 2.8268714 & 4.0948454 & 4.2631258 & 4.4362438 & 4.0795401 \\ 2.9128006 & 3.9724932 & 0.0000000 & 1.5883043 & 4.2807695 & 1.4040216 & 1.8835496 & 1.3605760 & 1.1033324 \\ 3.2776265 & 4.3052168 & 1.5883043 & 0.0000000 & 4.7539642 & 1.9365082 & 1.2154605 & 0.8418508 & 0.8181231 \\ 4.2816969 & 2.8268714 & 4.2807695 & 4.7539642 & 0.0000000 & 3.7641891 & 5.0853806 & 4.6716021 & 4.4886262 \\ 2.7028032 & 4.0948454 & 1.4040216 & 1.9365082 & 3.7641891 & 0.0000000 & 2.6409237 & 1.8095396 & 1.5676808 \\ 3.2866802 & 4.2631258 & 1.8835496 & 1.2154605 & 5.0853806 & 2.6409237 & 0.0000000 & 1.3798840 & 1.5820500 \\ 2.7856267 & 4.4362438 & 1.3605760 & 0.8417508 & 4.6716021 & 1.8095396 & 1.3798840 & 0.0000000 & 0.8619437 \\ 3.0007451 & 4.0796401 & 1.1033324 & 0.8181231 & 4.4886262 & 1.567808 & 1.5820500 & 0.8619437 & 0.0000000 \end{bmatrix}$$