المجلة العراقية للعلوم الاقتصادية / عدد خاص لوقائع المؤتمر العلمي الدولي/ السادس/ والسنوي/ السابع عشر/لسنة 2023
بعنوان / القيادة الرشيدة والتنمية المستدامة سبل الإصلاح الاقتصادي العراقي
Iraq Journal For Economic Sciences / ISSN:1812-8742 / ISSN ONLIN:2791-092X
https://doi.org/10.31272/IJES2024.80.S.S3

موائمة تحليل المكبات الرئيسية والتحليل العنقودي مع تطبيق عملي

Comfort Factor and Cluster Analysis with Practice Application

م. د. حيدر يحيى محمد
Hayder Yahya mohammed
haiderstatistic@yahoo.com
جامعة الكرخ للعلوم

أ. د. لقاء علي محمد
Lekaa Ali Muhamed
lekaa.ali.1968@gmail.com
كلية الادارة والاقتصاد / جامعة بغداد

الكلمات المفتاحية: تحليل المركبات الرئيسة ، التحليل العنقودي ، الموائمة.
Keywords: Major component analysis, cluster analysis, matching.

المستخلص

عند التعامل مع البيانات متعددة المتغيرات ذات الابعاد العالية غالبا ما نستخدم تحليل المركبات الرئيسية (principal component analysis(pca)) لتقليص الابعاد. يهدف البحث الى استخدام تحليل المركبات الرئيسة لمعرفة المتغيرات الاكثر تأثيرا على الظاهرة المدروسة ، ومن ثم يتم استعمال المركبات التي تم الحصول عليها على أنها متغيرات جديدة عددها بعدد المركبات وتمثل مشاهدات المركبات نسبة التأثير لكل متغير بالعامل ، وثم نطبق اسلوب عنقدة K من المتوسطات على المتغيرات الجديدة (المركبات) لمعرفة المشاهدات (نسبة التأثير لكل متغير بالعامل) الاقرب الى بعضها ذات التأثير المباشر على الظاهرة المدروسة ، وتم اللجوء الى اسلوب الموائمة للحصول على نتائج اكثر دقة والتخلص من اخطاء التصنيف في حالة البيانات الكبيرة بعدد المتغيرات والمشاهدات وتقليل زمن حساب النتائج ومن ثم سيتم استعمال تحليل المركبات الرئيسة لتحديد المركبات المؤثرة ومن ثم تصنيفها باستعمال اسلوب عنقدة Kmeans.

Abstract

When dealing with high-dimensional multivariate data, we often use principal component analysis (PCA) to reduce dimensionality.

The aim of this research is to Comfort (PCA) to identify the most effect variables on the studied ,The observations of the compounds represent the ratio of the effect of each variable in the factor, and then we apply the method K-means Clustering of averages to the new variables (factors) to find out what observation (the ratio of effect for each variable in the factor) closest to each The method of Comforting used to obtain more accurate results , Group similar variables, eliminate classification errors in the case of large data by the number of variables and observations and reduce the time of calculation of the results.

1- المقدمة Introduction

واحد من اكثر التقنيات شائعة الاستخدام في تحليل البيانات متعددة المتغيرات هو تحليل المركبات الرئيسية Principal Components Analysis (PCA) والذي يستخدم لتحليل عدد كبير من المتغيرات المرتبطة بهدف اختصارها الى عدد اقل من المركبات (العوامل) غير المرتبطة، يعزى لها تباين تلك المتغيرات[5]. بعد هذه المرحلة يتم تطبق التحليل العنقودي على المركبات الرئيسة باعتبارها متغيرات جديدة ، والمتغيرات الاصلية الظاهرة بالعوامل نعتبرها مشاهدات، حيث تبدأ الطريقة مع k من العناقيد العشوائية ومن ثم تحريك (نقل) المتغيرات بين هذه العناقيد مع العمل على تصغير الاختلافات

Iraq Journal For Economic Sciences / ISSN:1812-8742 / ISSN ONLIN:2791-092X
https://doi.org/10.31272/IJES2024.80.S.S3

داخل العناقيد وتعظيم الاختلافات بين العناقيد ، بتحديد عدد ثابت من العناقيد واختيار مراكز العناقيد الأولية عشوائياً، حيث يتم تجميع البيانات في k من العناقيد و تتم عملية العنقدة عن طريق التقليل من مجموع المسافات المربعة (المسافات الاقليدية) بين العناصر و مركز العنقود المقابل لها[8].

وتم اللجوء الى التوظيف للحصول على نتائج اكثر دقة ولافتقار المكتبات العربية والعراقية الى موضوعات التوظيف المركبات الرئيسة مع التحليل العنقودي.

جدول يمثل البيانات الحقيقية

	X1	X2	...	X14	X15
Observation 1					
Observation 2					
. . .					
Observation 22					

الجدول اعلا يمثل نتائج المركبات الرئيسة

كما في الجدولين السابقين نلاحظ ان الابعاد للمركبات الرئيسة تختلف عن ابعاد البيانات الحقيقية حيث كانت 15*22 واصبحت 15*15، يتم استعمال نتائج المركبات الرئيسة كبيانات جديدة بحجم 15*15 ونطبق التحليل العنقودي بالاعتماد عليها(15 متغير و 15 مشاهدة).

2- هدف البحث Aim of research: يهدف البحث الى دراسة وتحليل بيانات متعددة المتغيرات لمعرفة المتغيرات الاكثر تأثيرا على ظاهرة انتاج الخضروات من عام 1996 الى عام 2017 باستخدام التحليل العاملي، ومن ثم المواءمة مع التحليل العنقودي لمعرفة اهم المتغيرات التي تؤثر على الظاهرة المدروسة ثم التعرف على مدى التقارب بين تلك المتغيرات في المرحلة اللاحقة اي ايجاد اسلوب

	PC1	PC2	...	PC14	PC15
Loading 1					
Loading 2					
. . .					
Loading 15					

لغرض تجميع المتغيرات المتشابهة ، سيتم استعمال تحليل المركبات الرئيسة لتحديد العوامل المؤثرة ومن ثم تصنيفها باستعمال اسلوب عنقدة Kmeans.

3 - الجانب النظري Theoretical aspect

1-3 تحليل المكونات الرئيسة Principle Component Analysis: إن اول من اقترح فكرة المكونات الرئيسة هو (Karl person عام1901) ، وذلك حين استخدمها وسيله للوصول إلى ما سماه حينها بالمربعات الصغرى المتعامدة (orthogonal least squares). وبعد ذلك بسنوات وابتداءً من عام 1933 قام (Haroid Hoteling) بتطوير هذه الطريقة تطويراً لافتاً للنظر ليكون اساس عمل هذه الطريقة كما هو بين أيدينا. إن طريقة المركبات الرئيسة هي طريقة استكشافية يمكن الإفادة منها للتوصل الى تفسير او فهم العلاقات المتداخلة بين المتغيرات ، وهي تعالج مجموعة المتغيرات المرتبطة بتحويله الى متغيرات غير مترابطة فيما بينها (متعامدة orthogonal) والمتغيرات الاخيرة تدعى بالمركبات او المكونات الرئيسة، ويكون عددها بعدد المتغيرات المدروسة ، وإن كل مركبة رئيسة هي عبارة عن

Iraq Journal For Economic Sciences / ISSN:1812-8742 / ISSN ONLIN:2791-092X
https://doi.org/10.31272/IJES2024.80.S.S3

تداخل خطي للمتغيرات المدروسة يكون تباينها بمثابة مؤشر لتفسير جزء من التباين الكلي، لذلك فان الباحث حين يرغب في تقليل أبعاد المشكلة، كتقليل عدد المتغيرات المدروسة من دون فقدان كمية كبيرة من المعلومات يقوم باختيار المكونات الرئيسة الأولية ، يستطيع الباحث تحليل عدد قليل من المكونات الرئيسة المستقلة بدلاً من تحليل عدد كبير من المتغيرات الأصلية المرتبطة فيما بينها بعلاقات معقدة[6].

2-3 نموذج المكونات الرئيسة: Principal component model يهدف الى حساب العلاقة الرياضية الخطية بين جميع المتغيرات المدروسة $x_{j\ (j=1,2,\ldots,p)}$ تمثل مجموع المتغيرات المدروسة بعد ضربها بالمعاملات a_{ij} ويمكن تمثيلها كالاتي :

$$Y_i = a_{1i}x_1 + \cdots + a_{pi}x_p \qquad (1)$$

$$Y_i = \sum_{j=1}^{p} a_{ij}x_j \qquad (2)$$

إذ ان :

Y_i : المكون الرئيس i.

a_{ij} : معامل المتغير j في المكون i وتمثل قيمة المتجهات المميزة (a_i) (Characteristic Vectors) المرافق للجذور المميزة (Characteristic Roots) للمصفوفة المستخدمة. وحيث إن العلاقة السابقة يمكن ايجادها عن طريق مصفوفة التباين والتباين المشترك في حالة كون المتغيرات المدروسة لها وحدات القياس نفسها، اما اذا كانت وحدات القياس مختلفة فتحول المتغيرات الاصلية الى متغيرات جديدة معيارية (اي تتوزع توزيعاً طبيعياً قياسياً بوسط حسابي مقداره صفر وتباين مقداره واحد) عن طريق مصفوفة الارتباط[7]. المركبة الرئيسة الاولى First principal component هي تركيب خطي للمشاهدات x ومعادلتها هي :

$$Y_1 = a_{11}x_1 + \cdots + a_{p1}x_p \qquad (3)$$

$$Y_1 = a_1'x$$

Y_1 : تمثل المركبة الرئيسة الاولى.

a_1 : تمثل المتجه المميز الاول Eigen vector المرافقه للجذور المميز الاول Eigen value.

ويكون تباين المركبة الاولى هو :

$$S_{Y1}^2 = \sum_{i=1}^{p}\sum_{j=1}^{p} a_{i1}a_{j1}S_{ij} \qquad (4)$$

$$= a_1'Sa_1 = l_1$$

حيث l_1 هو الجذر المميز الاول First Eigen value.

S : مصفوفة التباين والتباين المشترك للمتغيرات x_j.

Iraq Journal For Economic Sciences / ISSN:1812-8742 / ISSN ONLIN:2791-092X
https://doi.org/10.31272/IJES2024.80.S.S3

ويمثل (S_{Y1}^2) التباين الاكبر مقارنة بالتباينات الاخرى ولتحقيق ذلك نستعين بمضروب لاكرانج بالقيد :

$$\mathrm{a}_1' a_1 = 1$$

ونشتق بالنسبة لـ a_1

$$\frac{\partial}{\partial a_1}\left(S_{Y1}^2 + l_1(1 - \mathrm{a}_1' a_1)\right) = 2(S - l_1 I)a_1 \qquad (5)$$

وبمساوات المعادلة (5) بالصفر :

$$(S - l_1 I)a_1 = 0 \qquad (6)$$

نحصل على المتجه المميز المرافق لأكبر جذر مميزة للمعادلة :

$$(S - l_1 I) = 0 \qquad (7)$$

وبضرب المعادلة (6) بـ a_1' نحصل على :

$$l_1 = a_1' S a_1 = S_{Y1}^2 \qquad (8)$$

إذ إن l_1 تمثل الجذر المميز الاكبر ويمتلك اكبر جزء من التباين المفسر الكلي.
المركبة الرئيسة الثانية معادلتها هي :

$$Y_2 = a_{12}x_1 + \cdots + a_{p2}x_p \qquad (9)$$

وباستعمال مضروب لاكرانج بالقيدين :

$$\mathrm{a}_2' a_2 = 1 \qquad \mathrm{a}_1' a_2 = 0$$

$$\frac{\partial}{\partial a_2}(\mathrm{a}_2' S a_2 + l_2(1 - \mathrm{a}_2' a_2) + M\mathrm{a}_1' a_2) = 2(S - l_2 I)a_2 + Ma_1 \qquad (10)$$

وبالأسلوب نفسه حل المعادلة (3) نحصل على :

$$(S - l_2 I)a_2 = 0 \qquad (11)$$

والذي يمثل المتجه المميز الثاني المرافق لثاني اكبر جذر مميز.

والان يمكن اعطاء الصيغة العامة لإيجاد باقي الجذور المميزة والمتجهات المرافقة لها.

$$\left(S - l_j I\right)a_j = 0 \qquad \text{ولكل } a_j \qquad (12)$$

وأن $0 < l_j < \cdots > l_2 > l_1$ وكذلك إن [6]:

$$\sum_{j=1}^{p} l_j = \sum_{j=1}^{p} S_{Yj}^2 = tr(s) \qquad (13)$$

Iraq Journal For Economic Sciences / ISSN:1812-8742 / ISSN ONLIN:2791-092X
https://doi.org/10.31272/IJES2024.80.S.S3

3-3 طريقة اختيار المكونات الرئيسة Method of selection principal component:
يتم اتخاذ قراراً بعدد المكونات الرئيسة المؤثرة لتلخيص البيانات عملياً وسوف نذكر بعض هذه الطرائق:
1. ان عدد المركبات الرئيسة المختارة يكون بعدد الجذور المميزة الاكبر من واحد $(l>1)$[2].
2. نحتفظ بالمكونات التي تفسر (80%)من التباين الكلي[15].
3. أشار Morrison (1976) إلى أن تفسير 75% من التباين الكلي يكون كافياً ، وعموماً كلما كانت نسبة التباين المفسر عالية وعدد المكونات المختارة قليلة كان ذلك أفضل من ناحية سهولة مناقشة النتائج وتفسيرها[4].

4-3 التحليل العنقودي Cluster Analysis:التحليل العنقودي هو اسلوب يهدف الى تقسيم بيانات متعددة المتغيرات الى مجاميع او عناقيد، المشاهدات في العنقود نفسه تكون متجانسة وتكون غير متجانسة مع العناقيد الاخرى[9]، وعرفRomesburg "H. Charles" التحليل العنقودي على انه اسم لمجموعة من الاساليب الرياضية للبيانات الضخمة التي يمكن استخدامها لتجميع المشاهدات في مجموعات متماثلة.[16]

5-3 الخطوات الاساسية للتحليل العنقودي:Basic step for cluster analysis
1. استكشاف مصفوفة البيانات Exploration the data matrix: أي معرفة البنية الاساسية للبيانات والهيكل الاجمالي لها وتفرعاتها المتجانسة فيبدأ التحليل العنقودي بمصفوفة البيانات، التي يتم الحصول عليها من البيانات المقسمة الى اعمدة تمثل الكائنات (المتغيرات) objects ممكن ان تكون اشخاص او قطع اراضي في الغابات، والصفوف تسمى السمات attributes قد تكون سمات الاشخاص او الاشجار التي تنمو على الغابات. وفي حالة مصفوفة البيانات الصغيرة فلا حاجة لاستعمال التحليل العنقودي حيث يمكننا اجراء الفحص البصري ببساطة ، عن طريق إلقاء نظرة على مصفوفة البيانات وإيجاد أشياء مماثلة ومختلفة لكائنين لهما القيم نفسها تقريبًا بالسمات مقارنة مع كائنين اخرين، لكن في حالة مصفوفة البيانات الكبيرة على سبيل المثال مائة كائن ومائة من السمات يفشل الفحص البصري ومن ثم نلجأ الى اسلوب التحليل العنقودي.
2. مصفوفة البيانات المعيارية Standardize the data matrix: مصفوفة البيانات المعيارية هي مصفوفة ناتجة من تحويل مصفوفة البيانات الى الصيغة المعيارية لغرض التخلص من اختلاف وحدات قياس المتغيرات ومن ثم ضمان صحة النتائج.
3. حساب المسافة: Compute the distance: حيث يتم قياس مدى التشابه بين ازواج المتغيرات ومن اهم هذه المقايس هي المسافة الاقليدية Euclidean distance[15].

$$d_i = \sqrt{\sum (x_i - y_i)^2}$$

4. تنفيذ اسلوب العنقدة Execute the clustering method: العنقدة في التحليل العنقودي تكون على نوعين هما:
1 - طريقة العنقدة الهرمية (Hierarchical clustering): في هذا الاسلوب لا يكون تقسيم المشاهدات الى عدد معين من العناقيد بل من سلسلة من التقسيمات والتي قد تعمل من عنقود واحدة يحتوي على جميع الأفراد، إلى عناقيد تحتوي كل منها على فرد واحد، تقنيات العنقدة الهرمية قد تنقسم إلى: الأساليب

Iraq Journal For Economic Sciences / ISSN:1812-8742 / ISSN ONLIN:2791-092X
https://doi.org/10.31272/IJES2024.80.S.S3

التجميعية والتي تتبعها سلسلة من الدمج للمشاهدات في العناقيد المتتالية حتى الوصول الى عنقود واحد يضم المشاهدات كافة. الأساليب التقسيمية والتي تفصل المشاهدات على التوالي إلى مجموعات أدق.
كلا النوعين من العنقدة الهرمية يمكن أن ينظر إليها على أنها محاولة للعثور على الخطوة المثلى[10].
2 - طرق العنقدة اللاهرمية (Non-Hierarchical clustering): تهدف العنقدة اللاهرمية الى تجزئه مجموعة البيانات الى (k) من المجاميع العنقودية المتجانسة وذلك بالاعتماد على فكرة النزعة المركزية لمجموعة العنقدة باستعمال (k) من المتوسطات. وهنالك العديد من طرائق العنقدة اللاهرمية بعضها يعتمد على الوسط الحسابي والبعض الاخر يعتمد على الوسيط او المنوال.
وسنشرح طريقة k-means بشكل مفصل بسبب استعمالها بالجانب التطبيقي[15].

6-3 طريقة عنقدة K من المتوسطات K-Means Clustering Method تهدف إلى تجميع المشاهدات المتماثلة مع بعضها البعض اعتمادا على خصائصها في K من العناقيد، وتتم عملية العنقدة من خلال تقليل المسافات بين المشاهدات ومركز العنقود. يبدأ أسلوب k-means بـ k من المجاميع تتكون كل منها من مشاهدة يتم اختيارها عشوائيا، ثم نضيف كل مشاهدة جديدة الى المجموعة التي تكون أقرب إليها من غيرها من ناحية المسافة، بعد اضافة المشاهدات الى المجاميع نقوم بإيجاد المتوسط المعدل لكل مجموعة وعده المركز الجديدة الذي تتوزع على اثره المشاهدات مرة اخرى ونستمر بهذا الاسلوب حتى الوصول الى عدم التغير بمراكز العناقيد، ولهذا السبب تم تسميتها بطريقة k-means[12].
نفترض ان البيانات تتكون من n من المشاهدات $X_1,X_2,\ldots,X_n$ ، خوارزميةk-means تقسم n من المشاهدات الى K من العناقيد $C_1,C_2,\ldots,C_K$ ،ويتم اختيار مراكز العناقيد $m_1,m_2,\ldots,m_K$ ، باستخدام المسافة الاقليدية علما أن خوارزمية k-means تعمل بالشكل الاتي:
1. اختيار k من مراكز العناقيد عشوائياً $(m_1, m_2,\ldots, m_k)$.
2. تعيين كل مشاهدة X_i $(1 \le i \le N)$ لأقرب مركز ، عن طريق حساب قيمة المؤشر $\delta(X_i, C_k)$ $(1\le k \le K)$ وكالاتي :

$$\delta(x_i, C_k) = \begin{cases} 1 & D(x_i, m_k) < D(x_i, m_j) for\ all\ j \neq k \\ 0 & otherwise \end{cases} \qquad (14)$$

اذ إن $D(x_i, m_k)$ هي المسافة الاقليدية وصيغتها :

$$D(x_i, m_k) = \|x_i - m_k\|^2$$

3.حساب مراكز العناقيد الجديدة ${m_k}^{new}$ ولكل عنقود من خلال حساب معدل المشاهدات في كل عنقود C_k كالاتي :

$${m_k}^{new} = \frac{1}{C_k}\sum_{i=1}^{N} \delta(x_i, C_k) x_i \qquad (15)$$

اذ إن $|C_k|$ تمثل عدد المشاهدات في العنقود C_k

$$C_k = \sum_{i=1}^{N} \delta(x_i, C_k) \qquad (16)$$

Iraq Journal For Economic Sciences / ISSN:1812-8742 / ISSN ONLIN:2791-092X
https://doi.org/10.31272/IJES2024.80.S.S3

4.اعادة الخطوة 2 و 3 حتى التقارب المتمثل بعدم تغير مراكز العناقيد.
5. اعادة حساب مراكز العناقيد m_k[17].

7-3 مواءمة التحليل العنقودي للتحليل العاملي comfort cluster analysis with factoranalysis

لدراسة وفهم افضل نتائج البيانات تم استخدام اسلوب المواءمة بين التحليل العاملي والتحليل العنقودي باستخدام اسلوب عنقدة K من المتوسطات، و التحليل العاملي باستخدام تحليل المركبات الرئيسة PCA. ففي البداية يستعمل PCA لتقليص عدد المتغيرات الكبيرة الى عدد أقل من العوامل غير المرتبطة تسمى المركبات اذ إن كل مركبة تحتوي على جميع المتغيرات لكن بنسب تأثير مختلفة ومن ثم تم تحديد اهم المتغيرات المؤثرة، ومن ثم يتم استعمال المركبات التي تم الحصول عليها على أنها متغيرات جديدة عددها بعدد المركبات، وتمثل مشاهدات المركبات نسبة التأثير لكل متغير بالعامل (وتسمى بالتحميلات للعوامل)، وسنطبق اسلوب عنقدة K من المتوسط على المتغيرات الجديدة لمعرفة ما المتغيرات الاقرب الى بعضها ذات التأثير المباشر على الظاهرة المدروسة.

الجانب التطبيقي The practical side

1-4 المقدمة Introduction

لغرض تحقيق الهدف من هذا البحث لمعرفة المتغيرات الاكثر تأثيرا من غيرها على الظاهرة المدروسة ومن ثم في مرحلة لاحقة تكمن المشكلة في كيفية معرفة ما المتغيرات الاقرب الى بعضها ذات التأثير المباشر على الظاهرة ؟ فقد آثرنا في هذا المبحث الى اللجوء الى دراسة للبيانات الحقيقية لعينة سحبت من بيانات خاصة بإنتاج الخضروات. اذ تم استعمال تحليل المركبات الرئيسة لمعرفة المتغيرات الاكثر تأثيرا، ومن ثم استعمال التحليل العنقودي على نتائج المركبات الرئيسة لمعرفة ما المتغيرات الاقرب الى بعضها ذات التأثير المباشر.

2-4 اطار المجتمع واسلوب جمع البيانات Community framework and data collection ethod

تم تحديد اطار المجتمع لبيانات إنتاج الخضروات ولـ 22 سنة (المشاهدات) من 1996 الى سنة 2017 ولـ 15 من المتغيرات، تم جمع البيانات من وزارة التخطيط - الجهاز المركزي للإحصاء. يتم التوصل إلى متوسط الانتاج بدلالة المساحة المزروعة والكمية المسوقة حيث يتم احتساب متوسط الإنتاجية على مستوى القضاء ويحتسب المستوى المرجح لإنتاجية الدونم الواحد على مستوى المحافظة باستخدام أوزان الإنتاج المتحقق في كل قضاء وكذلك الحال بالنسبة للمتوسط على مستوى العراق، بالتالي يتم احتساب تقديرات الإنتاج لكل محصول خلال الموسم وذلك من خلال المعادلة:

الإنتاج = المساحة المزروعة بالمحصول × متوسط الإنتاجية للدونم الواحد لكل قضاء

ثم على مستوى المحافظة والعراق ويتضمن التقرير بيانات تفصيلية عن المساحة المزروعة ومتوسط إنتاجية الدونم الواحد وكمية الإنتاج على مستوى العراق (الانتاج لكل 100 طن)، حيث نلاحظ هنالك تذبذب كبير في كمية الانتاج ولجميع البيانات وبنسب كبيرة من سنة الى اخرى.

تم تحديد اطار المجتمع لبيانات إنتاج الخضروات ولـ 22 سنة (المشاهدات) من 1996 الى سنة 2017 ولـ 15 من المتغيرات وهي: 1. البامية ، 2. الطماطم، 3. اللوبية الخضراء، 4. البطيخ، 5. الرقي، 6. لقرنبيط، 7. الهانة، 8. الخس، 9. البصل الأخضر، 10. البقلاء الخضراء، 11. الفاصوليا الخضراء 12. الباذنجان، 13. الفلفل الاخضر، 14. السبانغ، 15. الخيار

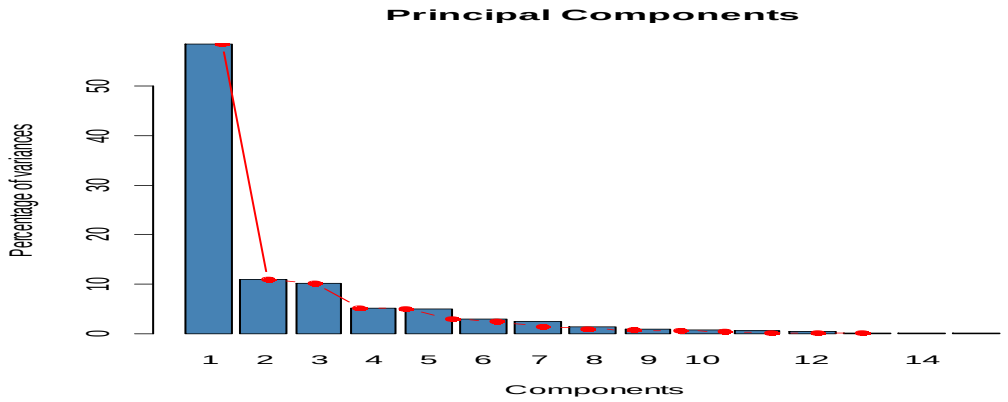
3-4 نتائج تحليل المركبات الرئيسة Result of principal component analysis: اظهرت النتائج ان طريقة المركبات الرئيسة فسرت (79.7) من التباين الكلي، اذ إن الجدول (4-1) يبين القيم المميزة و نسبة تفسير كل عامل من التباين المفسر فضلا عن التباين المفسر التجميعي للعوامل المؤثرة.

Iraq Journal For Economic Sciences / ISSN:1812-8742 / ISSN ONLIN:2791-092X
https://doi.org/10.31272/IJES2024.80.S.S3

الجدول (4-1) يمثل القيم المميزة ونسبة تفسير التباين لكل عامل من التباين الكلي لبيانات التطبيقية

PC	Eigen Value	Proportion Variance	Cumulative Variance
Pc1	8.77	58.51	58.5
Pc2	1.65	10.97	69.5
Pc3	1.53	10.24	79.7

الشكل (4-1)scree plot لطريقة المركبات الرئيسة للبيانات الحقيقية



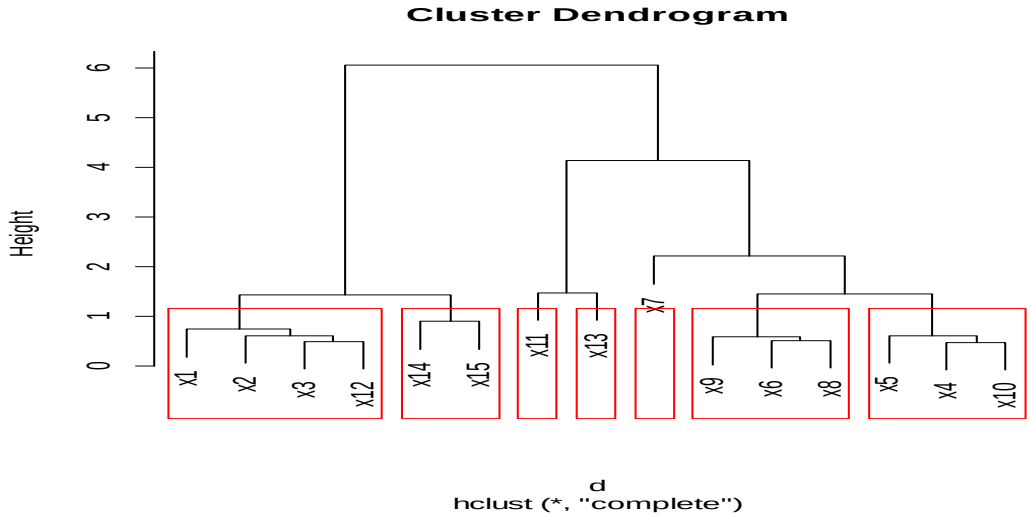
4-4نتائج الموائمة بين التحليل العاملي والعنقودي بأسلوبKmeans

Result of comfort cluster analysis with factor analysis

عدد العناقيد K=7

نلاحظ ظهور المتغيرات (X1,X2,X3,X12) في العنقود الاول، وظهور المتغيرين (X14,X15) في العنقود الثاني، وظهور المتغير(X11) في العنقود الثالث، و ظهور المتغير (X13) في العنقود الرابع، وظهور المتغير(X7) في العنقود الخامس، وظهور المتغيرات (X9,X6,X8) في العنقود السادس، و ظهور المتغيرات (X5,X4,X10) في العنقود السابع وفسرت 75.4% من التباين الكلي.

الشكل (4-2) توزيع المشاهدات للعناقيد لأسلوب عنقدة K means عندما k=7

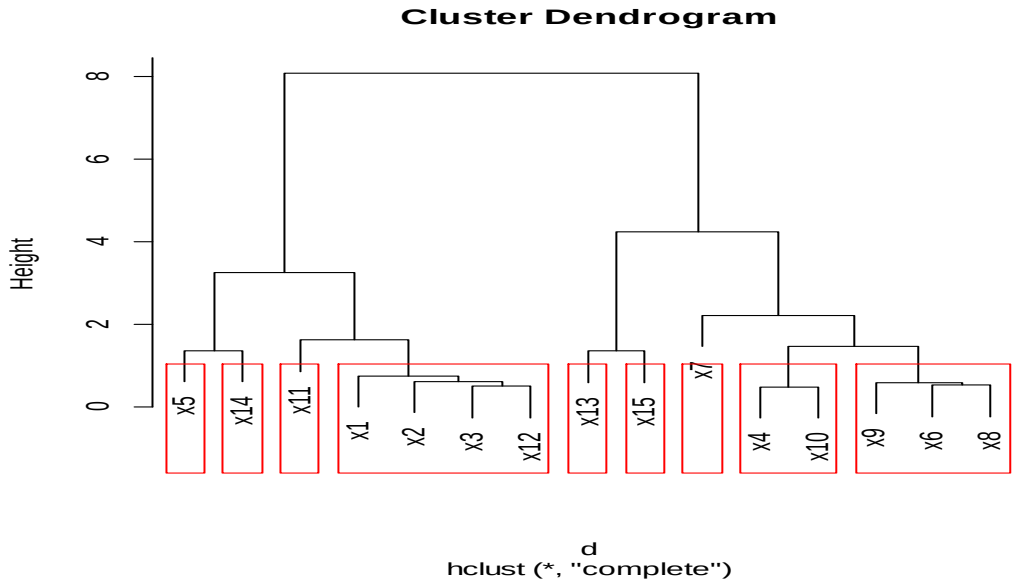


عدد العناقيد K=9

Iraq Journal For Economic Sciences / ISSN:1812-8742 / ISSN ONLIN:2791-092X
https://doi.org/10.31272/IJES2024.80.S.S3

نلاحظ ظهور المتغير (X5) في العنقود الاول، و ظهور المتغير(X4) في العنقود الثاني، و ظهور المتغير(X11) في العنقود الثالث، و ظهور المتغيرات (X1,X2,X3,X12) في العنقود الرابع، و ظهورالمتغير(X13) في العنقود الخامس، و ظهور المتغير(X15) في العنقود السادس، و ظهور المتغير (X7) في العنقود السابع، وظهور المتغيرين (X4,X10) في العنقود الثامن، و ظهور المتغيرات (X9,X6,X8) في العنقود التاسع وفسرت 84.5% من التباين الكلي.

الشكل (3-4) توزيع المشاهدات للعناقيد لأسلوب عنقدة K means عندما k=9



5- الاستنتاجات والتوصيات Conclusions and Recommendations

5-1 الاستنتاجات Conclusions

1. فسر تحليل المركبات الرئيسة (79.7) من التباين الكلي.
2. يعتمد أسلوب مواءمة التحليل العنقودي مع التحليل العاملي على عدد العناقيد حيث تؤدي زيادة تلك العناقيد الى زيادة نسبة التباين المفسر للتحليل العنقودي حيث ان النسبة كانت (75.4%) عند عدد عناقيد 7 ومن ثم عند زيادة عدد العناقيد الى 9 اصبحت (84.5%).
3. ظهور المتغيرات (البامية، الطماطم، اللوبية الخضراء، الباذنجان) و المتغيرات (البصل الاخضرالقرنبيط، الخس) لها نفس التأثير على الانتاج النباتي وبعدد العناقيد (9,7).

5-2 التوصيات Recommendations

1. استخدام المواءمة للتحليل العاملي مع التحليل العنقودي لمعرفة اهم المتغيرات التي تؤثر على الظاهرة المدروسة ثم التعرف على مدى التقارب بين تلك المتغيرات في المرحلة اللاحقة اي ايجاد اسلوب لغرض تجميع المتغيرات المتشابهة.
2. في حالة البيانات المستخدمة كبيرة فستقل دقة وأداء فضلا عن تعقد وقت التنفيذ للمركبات الرئيسة، ولحل هذه المشكلة سنلجأ الى تحليل المركبات الرئيسة ومن ثم استعمال عنقدة K من المتوسطات حيث اظهرت النتائج دقة أفضل وتعقيد الوقت تم خفضه أيضا.

Iraq Journal For Economic Sciences / ISSN:1812-8742 / ISSN ONLIN:2791-092X
https://doi.org/10.31272/IJES2024.80.S.S3

المصادر Reference

1) الجبوري، شلال حبيب وعبد، صلاح حمزة (2000)، " تحليل متعدد المتغيرات"، دار الكتب للطباعة والنشر، بغداد.
2) الراوي، خاشع محمود (1987)، "المدخل إلى تحليل الانحدار"، مديرية دار الكتب للطباعة والنشر جامعة الموصل ، الطبعة الأولى.
3) الشربجي، عبد الرزاق محمد صلاح (1981)، "الانحدار الخطي المتعدد"، وزارة التعليم العالي والبحث العلمي، جامعة الموصل، العراق.
4) الشكرجي، ذنون يونس ذنون (2005)، " استخدام مصفوفتي R-mode وQ-mode في التحليل العاملي"، رسالة ماجستير، كلية علوم الحاسبات والرياضيات، جامعة الموصل.
5) العلوي، لقاء علي محمد، (2003)، "مقارنة مقدرات التباين المشترك الحصينة في تحليل المركبات الرئسية "، أطروحة دكتوراه، كلية الإدارة والاقتصاد، جامعة بغداد.
6) محمد، حيدر يحيى، (2013)، دراسة العوامل المرتبطة بمرض السكري من خلال مواءمة اسلوبي التحليل العاملي وتحليل المسار، رسالة ماجستير، كلية الإدارة والاقتصاد، جامعة المستنصرية.
7) Anderson,T.W. (1984) " An introduction to multivariate statistical analysis " , second edition , john wile & sons , New York.
8) Combes, C., & Azema, J. (2013). Clustering using principal component analysis applied to autonomy–disability of elderly people. Decision Support Systems, 55(2), 578-586.
9) Ding, C., & He, X. (2004, July). K-means clustering via principal component analysis. In Proceedings of the twenty-first international conference on Machine learning (p. 29). ACM.
10) Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). Cluster Analysis. –John Wiley & Sons. Ltd., New York, 330.
11) Jiang, Y., Guo, H., Jia, Y., Cao, Y., & Hu, C. (2015). Principal component analysis and hierarchical cluster analyses of arsenic groundwater geochemistry in the Hetao basin, Inner Mongolia. Chemie der Erde-Geochemistry, 75(2), 197-205.
12) MacQueen, J. (1967, June). Some methods for classification and analysis of multivariate observations. In Proceedings of the fifth Berkeley symposium on mathematical statistics and probability(Vol. 1, No. 14, pp. 281-297).
13) Mohammed , L.A & Mohammed ,H.Y . (2019). On Robustness of Kernel Principal Component Analysis using FastHCS , INSTITUTE OF MECHANICS OF CONTINUA AND MATHEMATICAL SCIENCES, Vol-14, 292-303.
14) PITOY, C. (2017). EMPIRICAL STUDY ON BANDWIDTH OPTIMIZATION FOR KERNEL PCA IN THE K-MEANS CLUSTERING

Iraq Journal For Economic Sciences / ISSN:1812-8742 / ISSN ONLIN:2791-092X
https://doi.org/10.31272/IJES2024.80.S.S3

OF NON-LINEARLY SEPARABLE DATA. Journal of Theoretical & Applied Information Technology, 95(20).
15) Rencher , A.c. (2002) , " Method of multivariate analysis " , second edition , john wile & sons , N Y.
16) Romesburg, C. (2004). Cluster analysis for researchers. Lulu. com
17) Zhang, Z., Wang, G., Yeung, D. Y., & Kwok, J. T. (2004). Probabilistic kernel principal component analysis. Department of Compputer Science, The Hong Kong University of Science and Technology, Tech. Rep.