التحليل الى مركبات أساسية
L’analyse en composantes principales (ACP)

Principal compoment analysis (PCA)

تعريف التحليل الى مركبات أساسية :

استعملت هذه الطريقة لأول مرة من طرف karl pearson ، و أول من ضمها إلى الإحصاء سنة 1901 الرياضي Harold Hotelling سنة 1933، غير أنها لم تصبح واسعة الاستعمال إلا في التسعينات و ذلك لظهور الحاسوب و البرامج الإحصائية التي سهلت العمل بهذه التقنية .

كما يعد التحليل بالمركبات الأساسية أحد تقنيات تحليل البيانات و أحد طرق التحليل العاملي التي تختص في اختزال الجداول ذات الأبعاد الكبيرة أي اختزال عدد كبير من المتغيرات الخام إلى عدد أقل من المتغيرات الجديدة و المعروف بالمركبات، و الذي يكون عادة عددها أقل بكثير من المتغيرات الخام (الأصلية). على خلاف تحليل الانحدار أو تحليل التباين التي يعتبر متغير واحد تابع و البقية مستقلة فإن التحليل في مكونات أساسية كل متغير يقارن ببقية المتغيرات.

هذه المركبات الاساسية هي عبارة عن تراكيب خطية للمتغيرات الاصلية بشرط أن تفسر هذه المركبات الجزء الاكبر من التباين الكلي للمتغيرات الاصلية.

$$F_1=a_{11}X_1+a_{12}+X_2+....+a_{1n}X_n$$

$$F_2=a_{21}X_1+a_{22}+X_2+....+a_{2n}X_n$$

$$F_m=a_{m1}X_1+a_{m2}+X_2+....+a_{mn}X_n$$

حيث العامل الأول يكون أكثر ارتباطا بالمتغيرات أو أكثرها تفسيرا للتباين المشترك يليه العامل الثاني,

يعتمد التحليل بالمركبات الأساسية على إيجاد الجذور المميزة والمتجهات المميزة لمصفوفة التباين و التباين المشترك للمتغيرات التوضيحية، أو إيجاد الجذور المميزة لمصفوفة الارتباط، و هذا يعتمد على طبيعة البيانات فإن كانت وحدات القياس متشابهة نستخدم مصفوفة التباينات المشتركة أو مصفوفة الارتباط بين المتغيرات التوضيحية، أما إذا كانت وحدات القياس مختلفة فيمكننا استخدام مصفوفة الارتباطات فقط.

أهداف التحليل الى مركبات أساسية:

يكمن الهدف الرئيسي من استخدام تقنية التحليل الى مركبات أساسية في:

- اختزال العدد الكبير من المتغيرات في عدد أقل من المركبات (عوامل) وتكون مرتبطة بالمتغيرات الأصلية، بحيث تفسر هذه العوامل أكبر نسبة ممكنة من التباين في المتغيرات الاصلية.
- إظهار هيكل البيانات: نتيجة لتشعب العلاقات بين المتغيرات و تعددها فإنه يصعب في أحيان كثيرة قراءة الجداول واستنباط النتائج من خلالها خاصة إذا تحدثنا عن الدراسات الاجتماعية و العلوم الاقتصادية و علوم التسيير بوجه الخصوص التي يندر فيها تفسير الظواهر من خلال متغير أو اثنين، و من هنا يتم اللجوء إلى هذه التقنية لتحديد العلاقة بين المتغيرات و هيكلة البيانات.
- تسهيل قراءة الجداول ذات الأبعاد الكبيرة.

البيانات المستعملة :

– من خلال التحليل في مركبات أساسية يتم استخدام المتغيرات الكمية ذات قياس فئوي (echelle/scale) أو ذات قياس ترتيبي (Ordinale) و استخدام الجداول من نمط متغيرات مفردات إحصائية بالصيغة الموالية :

	التغيرة1 المتغيرة2 .....المتغيرةn
المفردة1 المفردة2 ..... المفردة m	

و يستحسن عموما أخذ عينة مكونة من 100 مشاهدة.

-**مراحل تطبيق التحليل الى مركبات أساسية**: يتم تطبيق التحليل الى مركبات أساسية من خلال المراحل التالية :

المرحلة الأولى: إلغاء البيانات المتطرفة والناقصة

قبل البدأ في تطبيق هذا التحليل لابد من القيام بفحص البيانات و تنقيحها بإلغاء القيم المتطرفة و الناقصة. فالقيم المتطرفة هي قيم تتجاوز بكثير القيم الموجودة أو تقل عنهم بكثير و يتم حذفها حتى لا تؤثر في تباين المركبات، لان المتغيرات التي تحتوي على قيم كبيرة مثلا سيكون تباينها أكبر و بالتالي يمكن أن تكون بمفردها مكون أو محور و هذا لا يعكس بالضرورة أهمية المتغير .

المرحلة الثانية: التحقق من توافر شروط تطبيق التحليل الى مركبات أساسية

ككل النماذج و التقنيات المستخدمة يجب التأكد من صلاحية البيانات لتطبيقها و من بين أهم المؤشرات المستعملة لذلك نجد :

أ- معاملات الارتباط: يعتمد التحليل في مركبات أساسية على مصفوفة الارتباط و التي تحتوي على معاملات الارتباط الثنائية بين المتغيرات، والذي يجب أن يكون قويا أي أكبر من 0.5 بالموجب أو السالب و بالتالي يجب أن تحتوي مصفوفة الارتباط على عدد مهم من معاملات الارتباط القوية.

ب- مؤشر KMO يشير هذا المؤشر الى كفاية عناصر العينة لاجراء الدراسة و تنص اغلب المراجع في هذا الشأن على أن تكون قيمته أكبر من 0.5.

ث- دلالة اختبار barttelet : يتم من خلاله اختبار الفرضية التالية :

- $H0$: كل معاملات الارتباط معدومة (معطيات لا تصلح للدراسة)
- $H1$:يوجد معاملات ارتباط غير معدومة.)معطيات تصلح للدراسة)

و بالتالي إذا كانت دلالة هذا الاختبار (الاحتمال) أقل من 0.05 (نسبة الخطأ المسموح به في الدراسات الاجتماعية) فإننا نرفض الفرضية الصفرية و نقبل الفرضية البديلة أي توجد معاملات ارتباط غير معدومة. و هذا ما يبرر استخدام تحليل بالمركبات الأساسية الذي يهدف إلى اختزال مجموعة من المتغيرات و تجميعها في محور واحد و لن يتم ذلك إلا إذا كان هنا ارتباط بينها .

و لتطبيق تقنية تحليل بالمركبات الأساسية يكفي توفر شرطين .

المرحلة الثالثة: تحديد نوعية تمثيل المتغيرات

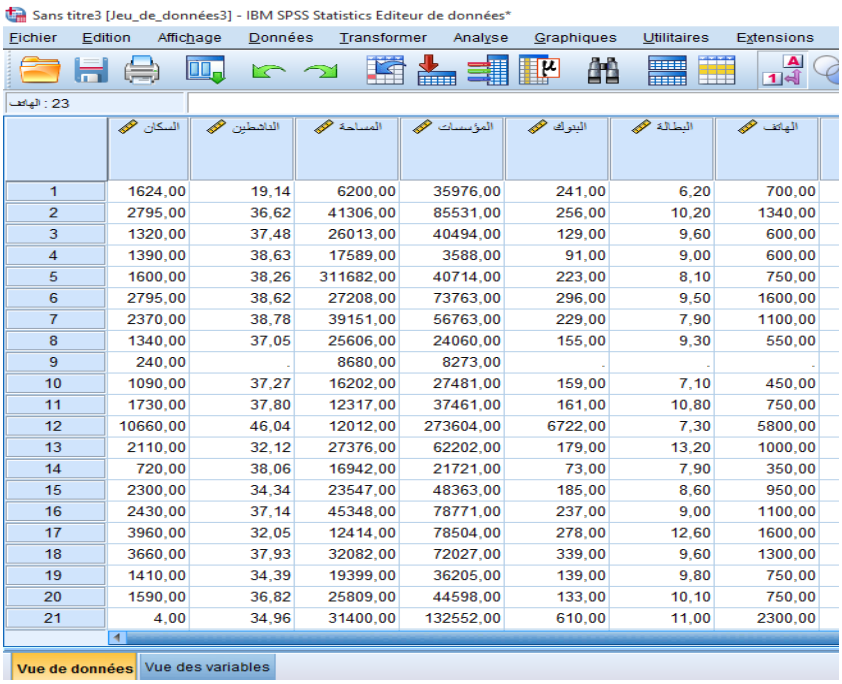
من خلال هذه المرحلة يتم اختبار نوعية أو جودة تمثيل المتغيرات المستعملة في الدراسة (درجة الشيوع)، فمن خلال هذه المرحلة يتم التأكد من أن المتغيرات المستعملة في الدراسة ذات جودة تمثيل مقبولة و التي ينص عادة على أن تكون أكبر من 0.4 . و يقصد بجودة التمثيل (درجة الشيوع) حجم المعلومات التي يساهم المتغير بها في الدراسة. حيث يتم إلغاء المتغيرات ذات جودة التمثيل أقل من 0.4 لأنها لا تساهم في الدراسة بشكل جيد. و في حال الإبقاء على هذه المتغيرات فإنها تتموقع قريبة من مبدأ

المنحنى أي لا يكون لها أي أثر في تسمية المحاور. و تجدر الإشارة أن المتغيرات ذات جودة تمثيل عالية تتموقع في محيط الدائرة التي نصف قطرها الواحد .

- التحليل الى مركبات أساسية باستخدام:SPSS

لتوضيح ما تم ذكره سابقا يمكن الاستعانة بالمثال التالي :

في المثال يوجد 22دولة يتم دراستها انطلاقا من سبعة متغيرات هي الكثافة السكانية، الكثافة السكانية النشطة، المساحة، عدد المؤسسات، عدد البنوك، معدل البطالة و عدد خطوط الهاتف. من المثال لا يمكن تحليل وضعية الدول وفقا للمتغيرات السبع ومقارنتها مع بعض فالأمر غير بسيط، و لذا نلجأ إلى تقنية التحليل الى مركبات أساسية.


Sans titre3 [Jeu_de_données3] - IBM SPSS Statistics Editeur de données*

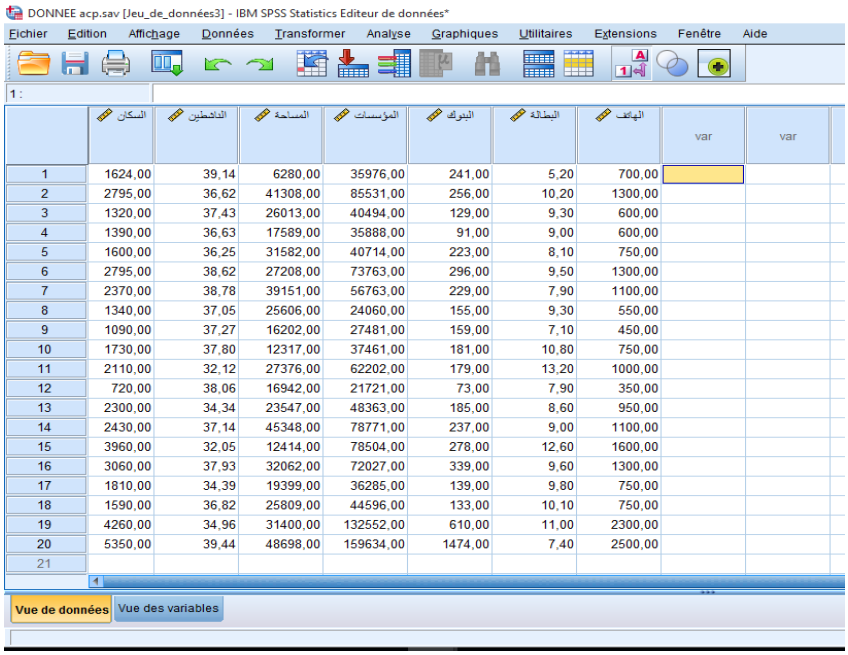
Fichier Edition Affichage Données Transformer Analyse Graphiques Utilitaires Extensions

الهاتف : 23

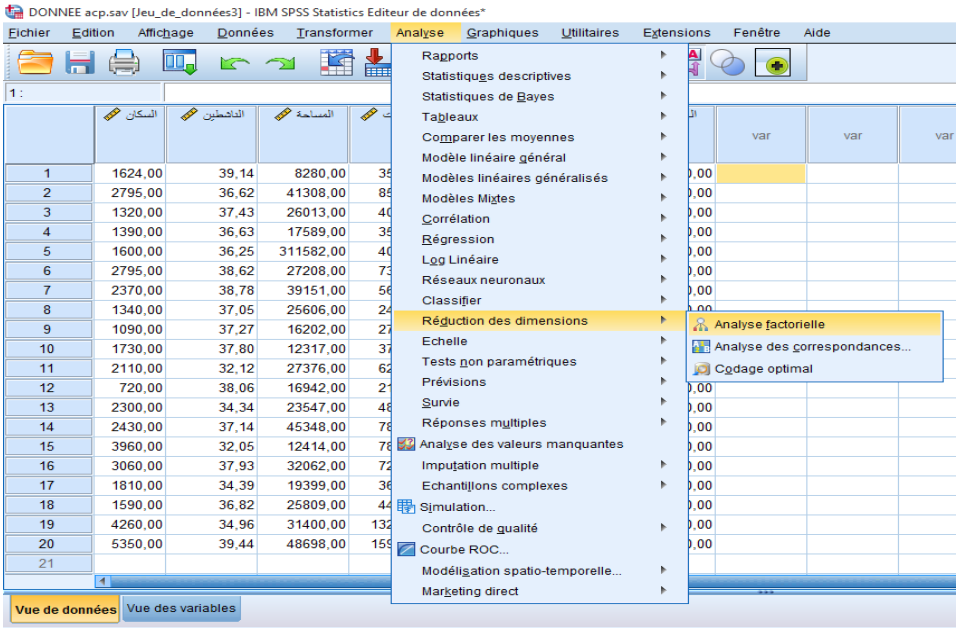
	السكان	الناشطين	المساحة	المؤسسات	البنوك	البطالة	الهاتف
1	1624,00	19,14	6200,00	35976,00	241,00	6,20	700,00
2	2795,00	36,62	41306,00	85531,00	256,00	10,20	1340,00
3	1320,00	37,48	26013,00	40494,00	129,00	9,60	600,00
4	1390,00	38,63	17589,00	3588,00	91,00	9,00	600,00
5	1600,00	38,26	311682,00	40714,00	223,00	8,10	750,00
6	2795,00	38,62	27208,00	73763,00	296,00	9,50	1600,00
7	2370,00	38,78	39151,00	56763,00	229,00	7,90	1100,00
8	1340,00	37,05	25606,00	24060,00	155,00	9,30	550,00
9	240,00	.	8680,00	8273,00	.	.	.
10	1090,00	37,27	16202,00	27481,00	159,00	7,10	450,00
11	1730,00	37,80	12317,00	37461,00	161,00	10,80	750,00
12	10660,00	46,04	12012,00	273604,00	6722,00	7,30	5800,00
13	2110,00	32,12	27376,00	62202,00	179,00	13,20	1000,00
14	720,00	38,06	16942,00	21721,00	73,00	7,90	350,00
15	2300,00	34,34	23547,00	48363,00	185,00	8,60	950,00
16	2430,00	37,14	45348,00	78771,00	237,00	9,00	1100,00
17	3960,00	32,05	12414,00	78504,00	278,00	12,60	1600,00
18	3660,00	37,93	32082,00	72027,00	339,00	9,60	1300,00
19	1410,00	34,39	19399,00	36205,00	139,00	9,80	750,00
20	1590,00	36,82	25809,00	44598,00	133,00	10,10	750,00
21	4,00	34,96	31400,00	132552,00	610,00	11,00	2300,00

Vue de données Vue des variables

و كمرحلة أولى و كما ذكرنا سابقا لابد من إلغاء البيانات المتطرفة و الناقصة إذ نلاحظ من الجدول أن للدولة التاسعة أربع قيم مفقودة من أصل سبعة و بالتالي يتم إلغاء هذه المفردة الاحصائية. كما نلاحظ أن الدولة رقم 12 بها قيم متطرفة أي تفوق بكثير بقية القيم في كل متغيرة (أكبر برقم) و أيضا يتم حذفها.

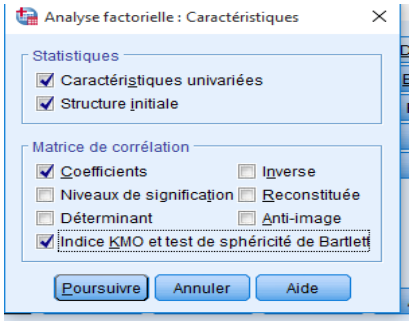


و بعد تنقيح البيانات نقوم بتطبيق التحليل الى مركبات أساسية وفقا للمراحل التالية :



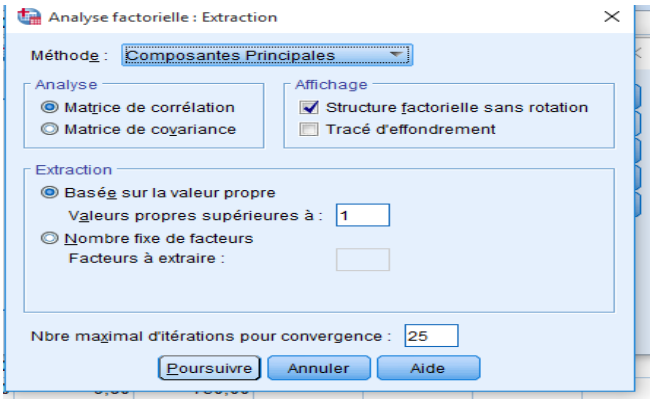
بعد ظهور صندوق الحوار التالي يتم إدخال جميع المتغيرات

بالتأشير على الامر Descriptivesيظهر صندوق الحوار التالي :

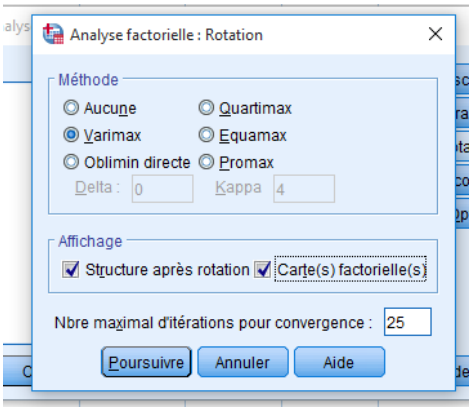


يخص صندوق الحوار هذا مؤشرات تحليل البيانات أحادية البعد المتمثلة أساسا في الانحراف المعياري و المتوسط الحسابي لوصف المتغيرات. كما يحتوي على شروط تطبيق التحليل في مكونات أساسية حيث نختار الأكثر استعمالا و التي سبق ذكرها.

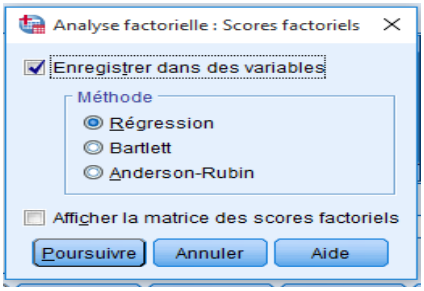
أما بالضغط على الامر Extraction (الاستخراج) فتظهر الاختيارات الثلاث لتحديد عدد المركبات المستخرجة من الدراسة كما ذكرنا سابقا .



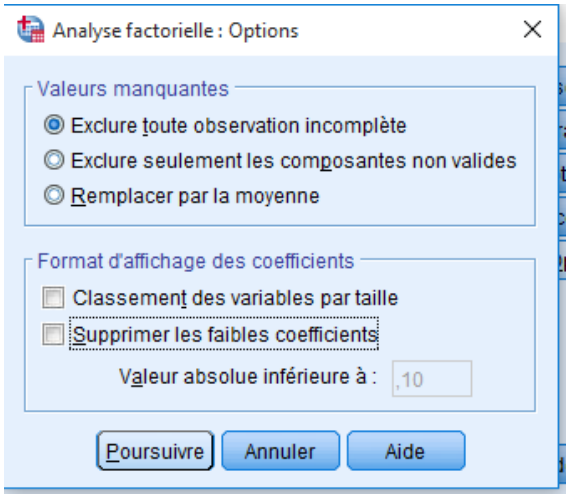
ثم انطلاقا من الامر Rotation (التدوير) يتم اختيار طريقة التدوير المستعملة و يجدر بالذكر أن التدوير المتعامد يسمح على خلاف بقية الطرق باستخراج محاور مستقلة. بالإضافة إلى البطاقة العاملية التي هي التمثيل البياني للمحاور المستخرجة .



أما في هذه المرحلة فإننا نقوم بحفظ إحداثيات المركبات المستخرجة في قاعدة البيانات الأصلية لاستعمالها في التمثيل البياني كما سنأتي على شرحه لاحقا .



أما المرحلة الموالية نقوم بالضغط على الأمر Option فيتم من خلالها معالجة المتغيرات المفقودة إما بحذف المتغير الذي يحتوي قيم مفقودة أو بحذف القيمة المفقودة فقط أو تعويضها بالمتوسط الحسابي للمتغيرة. كما يعرض خاصية اختيارية و هي حذف الإحداثيات الضعيفة للمتغيرات من كل محور و ذلك لتسهيل تحديد المتغيرات المنتمية للمحور و بالتالي تسهيل التسمية .



تحليل نتائج الدراسة:

بعد تطبيق هذه المراحل تظهر النتائج التالية ويعتبر تحليلها المرحلة الأهم على الإطلاق إذ أن التحليل يختلف من دراسة إلى أخرى ومن إشكالية إلى أخرى وهذا يتطلب الإلمام بالجانب النظري والتطبيقي في آن واحد حتى يتمكن الباحث من الوصول إلى النتائج الصحيحة والدقيقة. وهذا يتم كالاتي:

- الجدول الموالي يقدم وصف للمتغيرات المستعملة من خلال مؤشرات الإحصاء الوصفي والمتمثلة في المتوسط و الانحراف المعياري. و على سبيل المثال يمكن القول أن متوسط عدد السكان في عينة الدراسة يقدر ب2282.2000 نسمة بانحراف معياري قدره 1163.15 نسمة (أكثر من نصف المتوسط) أي بتشتت كبير نوعا ما، ما يعني أن هناك تفاوت في عدد السكان في مفردات العينة .

الجدول1: الإحصاء الوصفيStatistiques descriptives

	Moyenne	Ecart type	Analyse N
السكانية الكثافة	2282,2000	1163,15336	20
الناشطين السكانية الكثافة	36,6420	2,10155	20
المساحة	26312,5500	11366,04153	20
المؤسسات عدد	59639,3000	35579,88591	20
البنوك عدد	280,3500	303,50796	20
البطالة معدل	9,2800	1,84037	20
الهاتف خطوط عدد	1035,0000	569,18686	20

- و من خلال مصفوفة الارتباط التي هي مصفوفة قطرية متناظرة قطرها 1 التالية :

جدول 2 : مصفوفة الارتباط Matrice de corrélation

		الكثافة السكانية	السكانية الكثافة الناشطين	المساحة	عدد المؤسسات	البنوك عدد	معدل البطالة	خطوط عدد الهاتف
Corrélation	السكانية الكثافة	1,000	-,099	,510	,951	,827	,246	,984
	الناشطين السكانية الكثافة	-,099	1,000	,203	,001	,249	-,736	-,069
	المساحة	,510	,203	1,000	,645	,534	,018	,547
	المؤسسات عدد	,951	,001	,645	1,000	,863	,190	,978
	البنوك عدد	,827	,249	,534	,863	1,000	-,137	,836
	البطالة معدل	,246	-,736	,018	,190	-,137	1,000	,237
	الهاتف خطوط عدد	,984	-,069	,547	,978	,836	,237	1,000

من خلال الجدول نلاحظ أن عدد المعاملات المهمة (أكبر من 0.5 بالموجب و السالب) يقدر بـ11 من بين 21 أي يتجاوز النصف ومنه يتحقق الشرط الأول.

من خلال الجدول الموالي نلاحظ أن قيمة مؤشر KMO كانت 0.716 و قد تجاوزت 0.5 و بالتالي كفاية عناصر العينة لإجراء الدراسة و تحقق الشرط الثاني. و بالنسبة لدلالة اختبار Bartlett فيظهر الجدول أنها معدومة أي أقل0.05 و منه رفض الفرضية الصفرية و قبول الفرضية البديلة أي أنه يوجد معاملات ارتباط غير معدومة (البيانات تصلح للتحليل العاملي). و تحقق الشرط الثالث. و عليه و بما أن الشروط تحققت يمكن تطبيق التحليل الى مركبات أساسية .

Indice KMO et test de Bartlett

Indice de Kaiser-Meyer-Olkin pour la mesure de la qualité d'échantillonnage.		,716
Test de sphéricité de Bartlett	Khi-carré approx.	167,073
	Ddl	21
	Signification	,000

- أما من خلال جدول نوعية التمثيل فنلاحظ أن تمثيل كل المتغيرات أكبر من 0.4 و بالتالي كل المتغيرات تدخل في الدراسة و لا يتم استبعاد أي متغيرة. أما في حال وجود متغيرة بنوعية تمثيل أقل من 0.4 فمن المستحسن حذف المتغيرة و إعادة التحليل من جديد. كما نلاحظ أن كل المتغيرات ذات تمثيل عالي أي أن هذه المتغيرات ستتوضع في محيط الدائرة على عكس المتغيرة الممثلة للمساحة التي سيؤول توضعها إلى المبدأ دلالة على مساهمتها المتوسطة في الدراسة .

الجدول 3: Qualités de représentation
جودة لتمثيل/ درجة الشيوع

	Initiales	Extraction
السكانية الكثافة	1,000	,942
الناشطين السكانية الكثافة	1,000	,875
المساحة	1,000	,508
المؤسسات عدد	1,000	,979
البنوك عدد	1,000	,878
البطالة معدل	1,000	,863
الهاتف خطوط عدد	1,000	,964

Méthode d'extraction : Analyse en composantes principales.

- و يعد الجدول التالي الذي يمثل التباين المشروح جدول التحليل الى مركبات أساسية بامتياز. إذ من خلاله:

* يتم كخطوة أولى تحديد القيم الذاتية لكل المتغيرات المستعملة في الدراسة و في هذا المثال سبع متغيرات.

* ثم كخطوة ثانية و إن قمنا باختيار طريقة استخراج المحاور التي تعتمد على قيمة ذاتية أعلى من الواحد فإنه يتم إقصاء المركبات التي تحتوي على قيم ذاتية أقل من الواحد، وكل محور يمكن التعبير عنه بحجم البيانات المختزلة من خلاله ومن هذا الجدول نلاحظ أنه تم اختزال الجدول في محورين يضم المحور الأول أربع متغيرات بحجم بيانات قدره59.24 % و يضم الثاني متغيرين تقريبا بحجم معلومات قدره 26.59% .إجمالا المحورين اختزلا معا ما قيمته 85.84% من حجم المعلومات (6 متغيرات). تم التضحية بثلاث متغيرات بما يعادل 14.16 % .

Variance totale expliquée

	Valeurs propres initiales			Sommes extraites du carré des chargements			Sommes de rotation du carré des chargements		
Composante	Total	% de la variance	% cumulé	Total	% de la variance	% cumulé	Total	% de la variance	% cumulé
1	4,147	59,247	59,247	4,147	59,247	59,247	4,137	59,104	59,104
2	1,862	26,593	85,840	1,862	26,593	85,840	1,872	26,736	85,840
3	,604	8,634	94,474						
4	,235	3,350	97,824						
5	,113	1,620	99,444						
6	,032	,463	99,907						
7	,006	,093	100,000						

Méthode d'extraction : Analyse en composantes principales.

أما الجدولين المواليين فيمثلان احداثيات المتغيرات قبل و بعد التدوير

Matrice des composantes[a] مصفوفة
الاحداثيات قبل التدوير

	Composante	
	1	2
السكانية الكثافة	,961	-,132
الناشطين السكانية الكثافة	,024	,935
المساحة	,682	,207
المؤسسات عدد	,989	-,031
البنوك عدد	,894	,282
البطالة معدل	,167	-,914
الهاتف خطوط عدد	,976	-,107

Méthode d'extraction : Analyse en composantes principales.

a. 2 composantes extraites.

Rotation de la matrice des composantes[a]
مصفوفة الاحداثيات بعد التدوير

	Composante	
	1	2
السكانية الكثافة	,951	,196
الناشطين السكانية الكثافة	,086	-,931
المساحة	,694	-,162
المؤسسات عدد	,985	,096
البنوك عدد	,910	-,222
البطالة معدل	,106	,923
الهاتف خطوط عدد	,967	,172

Méthode d'extraction : Analyse en composantes principales.

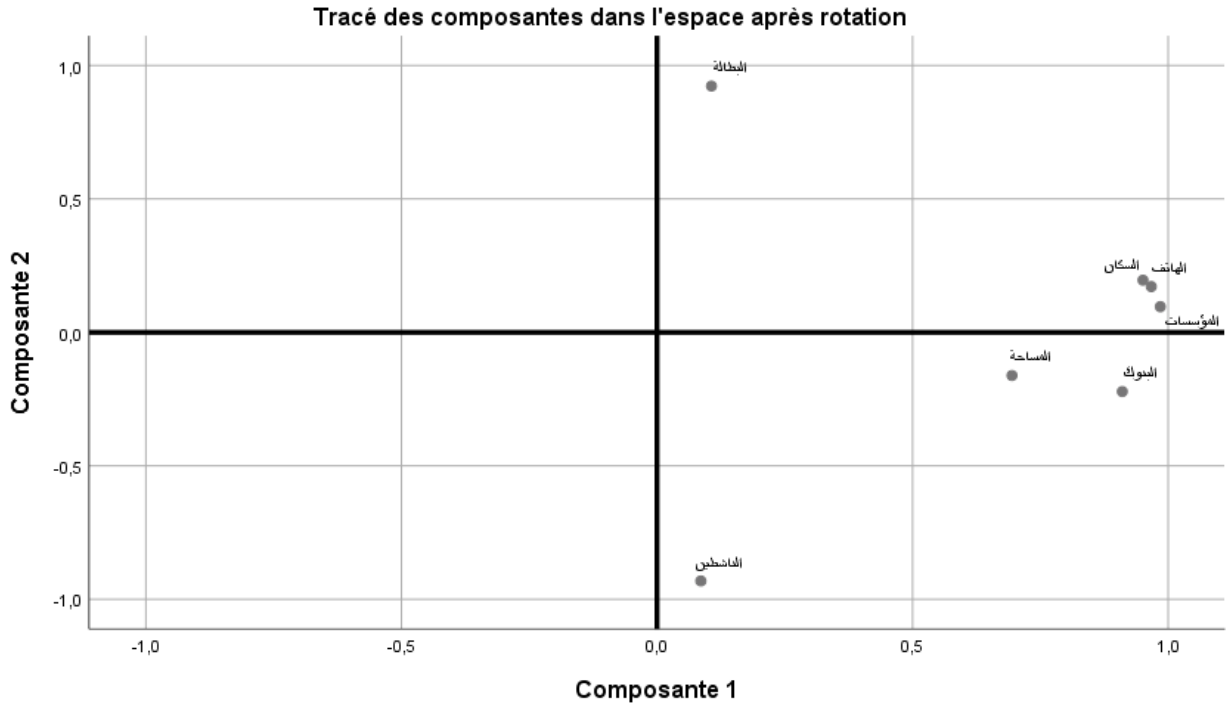
Méthode de rotation : Varimax avec normalisation Kaiser.

a. Convergence de la rotation dans 3 itérations.

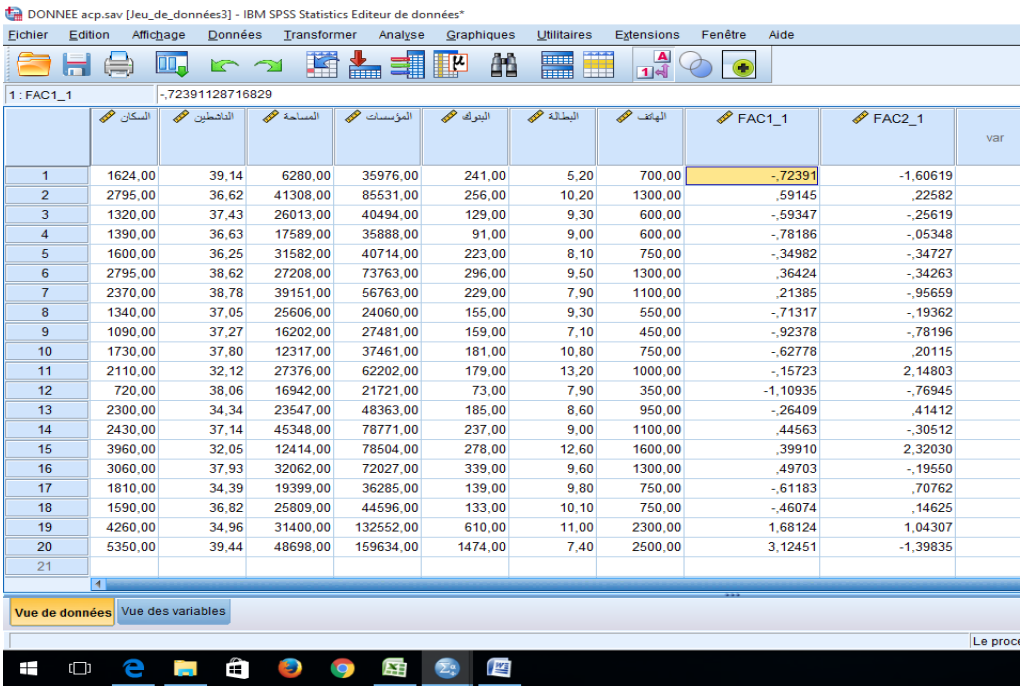
و يتم الاستعانة بالمنحنى الموالي لتسمية المحاور مع مراعاة أن المتغيرات الواقعة في محيط الدائرة ذات نصف قطر يساوي الواحد تعتبر ذات تمثيل جيد بعكس المتغيرات القريبة من مركز الدائرة:

-المحور الأول: يضم المتغيرات التالية الكثافة السكانية، عدد البنوك، عدد المؤسسات، عدد خطوط الهاتف والمساحة ، تمثل هذه المتغيرات في النظرية الاقتصادية عوامل الانتاج و من هنا يمكن تسمية هذا المحور "عوامل الإنتاج "

أما المحور الثاني يضم بالتعارض نسبة الكثافة السكانية النشطة (تعبر عن عدد السكان الإجمالي مطروحا منه عدد الأطفال و الشيوخ) و معدل البطالة و من هنا يمكن تسمية المحور الثاني " واقع العنصر البشري ."

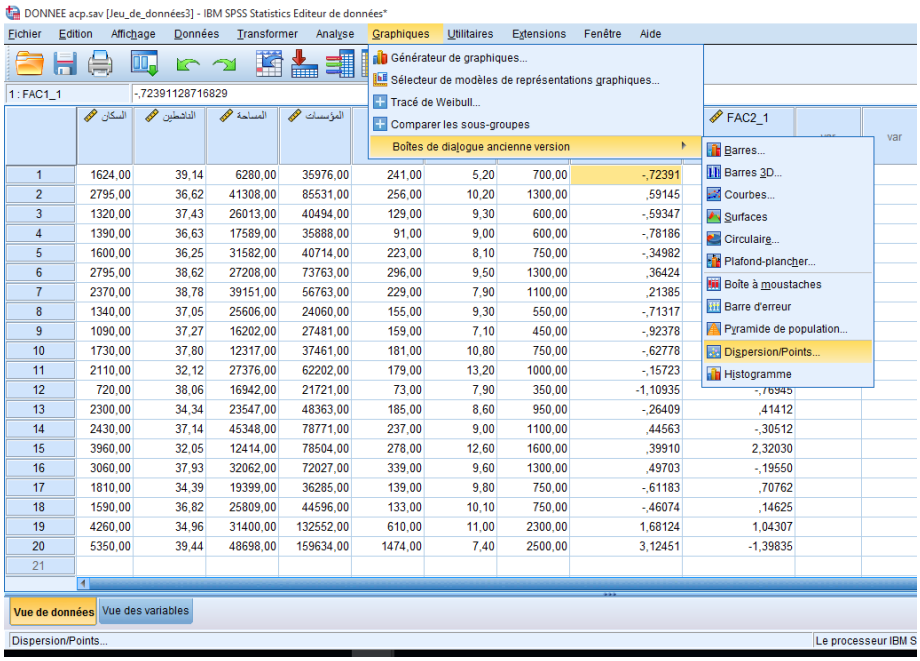


و خطوة أخيرة و بعد التعامل مع متغيرات الدراسة و تكوين المحاور (استخراج المركبات) يتم اسقاط المفردات الإحصائية (الدول في هذه الحالة) لدراستها و التي هي صلب البحث من خلال المراحل التالية: بعد حفظ احداثيات المركبات الأساسية كما سبق و ذكرنا و كما هو موضح في ما يلي :

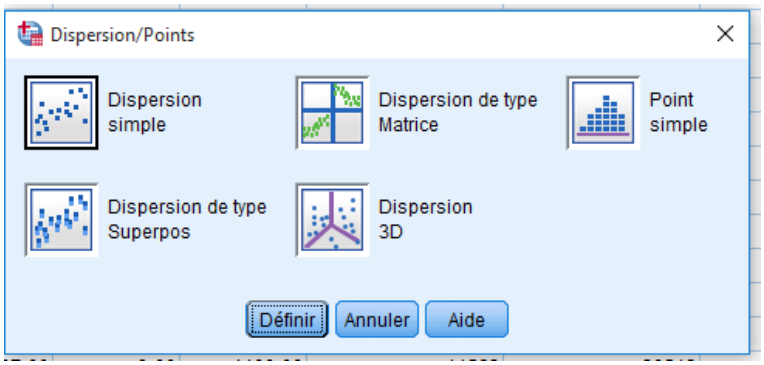


	السكان	الفاشطين	المساحة	المؤسسات	البنوك	البطالة	الهاتف	FAC1_1	FAC2_1	var
1	1624,00	39,14	6280,00	35976,00	241,00	5,20	700,00	-,72391	-1,60619	
2	2795,00	36,62	41308,00	85531,00	256,00	10,20	1300,00	,59145	,22582	
3	1320,00	37,43	26013,00	40494,00	129,00	9,30	600,00	-,59347	-,25619	
4	1390,00	36,63	17589,00	35888,00	91,00	9,00	600,00	-,78186	-,05348	
5	1600,00	36,25	31582,00	40714,00	223,00	8,10	750,00	-,34982	-,34727	
6	2795,00	38,62	27208,00	73763,00	296,00	9,50	1300,00	,36424	-,34263	
7	2370,00	38,78	39151,00	56763,00	229,00	7,90	1100,00	,21385	-,95659	
8	1340,00	37,05	25606,00	24060,00	155,00	9,30	550,00	-,71317	-,19362	
9	1090,00	37,27	16202,00	27481,00	159,00	7,10	450,00	-,92378	-,78196	
10	1730,00	37,80	12317,00	37461,00	181,00	10,80	750,00	-,62778	,20115	
11	2110,00	32,12	27376,00	62202,00	179,00	13,20	1000,00	-,15723	2,14803	
12	720,00	38,06	16942,00	21721,00	73,00	7,90	350,00	-1,10935	-,76945	
13	2300,00	34,34	23547,00	48363,00	185,00	8,60	950,00	-,26409	,41412	
14	2430,00	37,14	45348,00	78771,00	237,00	9,00	1100,00	,44563	-,30512	
15	3960,00	32,05	12414,00	78504,00	278,00	12,60	1600,00	,39910	2,32030	
16	3060,00	37,93	32062,00	72027,00	339,00	9,60	1300,00	,49703	-,19550	
17	1810,00	34,39	19399,00	36285,00	139,00	9,80	750,00	-,61183	,70762	
18	1590,00	36,82	25809,00	44596,00	133,00	10,10	750,00	-,46074	,14625	
19	4260,00	34,96	31400,00	132552,00	610,00	11,00	2300,00	1,68124	1,04307	
20	5350,00	39,44	48698,00	159634,00	1474,00	7,40	2500,00	3,12451	-1,39835	
21										

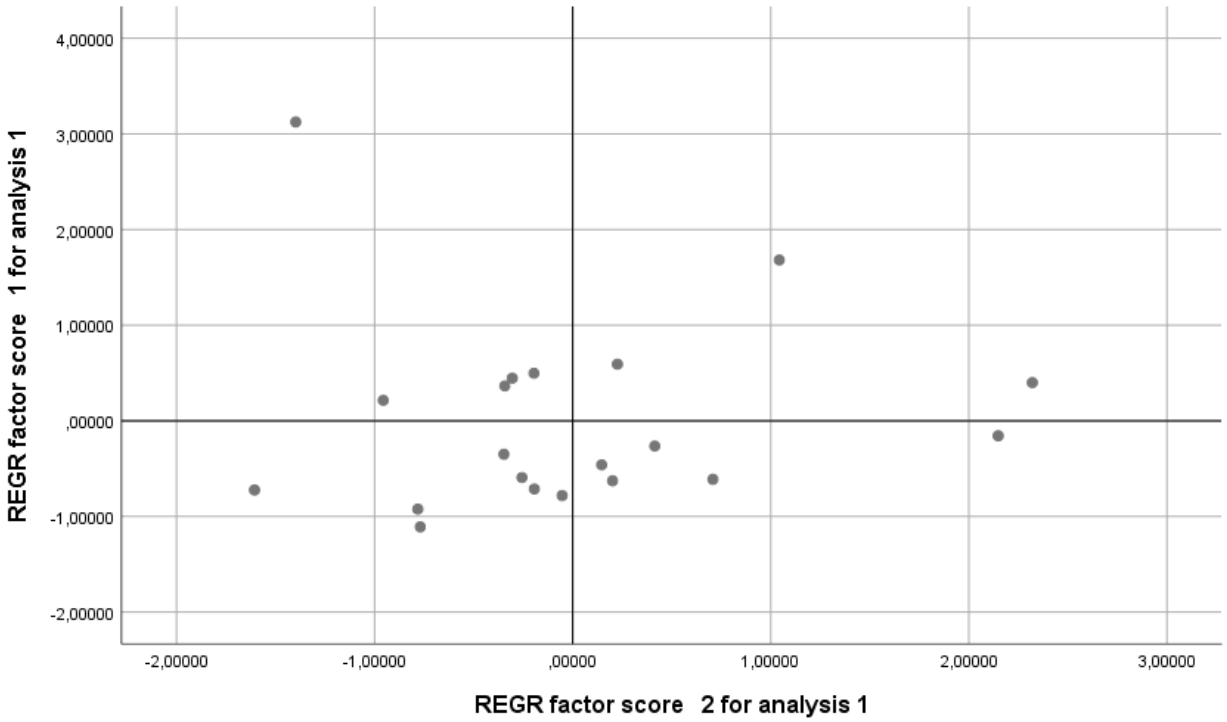
يتم الاستعانة بالتمثيل البياني بالخطوات التالية:



و من الصندوق الحوار التالي و بما أنه عندنا محورين نختار الانتشار البسيط:



حيث يتم ادخل المحاور في مكانهم المناسب:



من خلال هذا المنحنى يمكن دراسة الدول حسب مستوى عوامل الإنتاج المتوفرة (المحور 1) وواقع العنصر البشري (المحور2).