وزارة التعليم العالي والبحث العلمي
جامعة أكلي محند أولحاج – البويرة–
كلية العلوم الاقتصادية والتجارية وعلوم التسيير

فرقة البحث التكويني الجامعي "تأثير التسويق السياحي على اتجاهات السياح الجزائريين نحو المنتج السياحي المحلي – دراسة عينة من السياح الجزائريين–
بالتنسيق مع مخبر السياسات التنموية والدراسات الاستشرافية

تنظم يوم تكويني لطلبة السنة الأولى دكتوراه الطور الثالث L M D
موسوم بـ:

"طرق وأساليب جمع البيانات الاحصائية، وتحليلها باستخدام برامج التحليل الاحصائي"

يوم: الأربعاء 06 ديسمبر 2023

مداخلة بعنوان:

"طرق تحليل البيانات الإحصائية: التحليل العاملي والتصنيف الهرمي"

الأستاذ: حوشين يوسف
جامعة علي لونيسي البليدة2

1. تمهيد:

يعد تحليل البيانات الإحصائية من الأساليب الكمية الهامة في معالجة البيانات وتحليلها قصد استخراج معلومات مهمة لاتخاذ القرار في شتى المجالات الاقتصادية والاجتماعية. وتهتم أساليب تحليل المعطيات (Analyse des données) بالتحليل الإحصائي للجداول الإحصائية المتعددة المتغيرات، والتي لا تحتاج إلى فرضيات محددة للتحليل عكس النمذجة الإحصائية، فتحليل المعطيات في الحقيقة تعتمد في ذلك على مبادئ الجبر الخطي والهندسة في الفضاء.

وينقسم علم الإحصاء –بحسب الأبعاد التي يتناولها– إلى ثلاث (3) أقسام الإحصاء الأحادي البعد (Statistique Unidimensionnelle) والذي يدرس متغير واحد فقط، ويقوم التحليل الإحصائي على الإحصاء الوصفي، من مقاييس النزعة المركزية والتشتت وشكل التوزيع والتمثيلات البيانية الأحادية المتغير، ... إلخ. والإحصاء الثنائي البعد (Statistique Bidimensionnelle) والذي يدرس متغيرين اثنين، وبالإضافة إلى الإحصاء الأحادي البعد، فإن الإحصاء الثنائي البعد يسمح بتشكيل الجداول المزدوجة وحساب التغاير وحساب معامل الارتباط وإجراء اختبارات الارتباط والاستقلالية (الإحصاء الاستدلالي)، والتمثيل البياني بسحابة النقاط... إلخ. وأخيرا الإحصاء المتعدد الأبعاد (Statistique Multidimensionnelle) والذي يدرس عدة متغيرات (أكثر من متغيرين)، وبالإضافة إلى ما يتضمنه الإحصاء الأحادي البعد والإحصاء الثنائي البعد، فإن الإحصاء المتعدد الأبعاد يسمح بتبسيط وتلخيص العلاقات والارتباطات بين المتغيرات العديدة من خلال طرق تحليل المعطيات.

ونقصد بتحليل المعطيات مجموعة من الطرق والأساليب الجبرية والهندسية التي تسمح بالتحليل الإحصائي المتعدد الأبعاد (l'analyse statistique multidimensionnelles).

وتنقسم طرق تحليل المعطيات إلى قسمين رئيسيين، يحتوي كل قسم على مجموعة من الطرق، ويتم تحديد الطريقة المعتمدة في التحليل على أساس أهداف الدراسة ونوعية المعطيات. القسم الأول يعرف بالطرق العاملية (Méthodes Factorielles)، والتحليل العاملي هو أسلوب إحصائي للتحليل متعدد المتغيرات، يعمل على تجميع هذه المتغيرات في تركيبة متجانسة ومرتبطة فيما بينها لتكوين ما يسمى بالعامل (Facteur). وهي تسعى إلى الكشف على متغيرات غير مشاهدة (كامنة) تمثل تمثيلا كافيا للعلاقات البينية بين عدد من المتغيرات المشاهدة (المقاسة). فالعامل متغير كامن افتراضي مشتق من تحليل مجموعة من بيانات لمتغيرات مقاسة (مشاهدة). والهدف من التحليل العاملي هو تقليص وتلخيص العدد الكبير من المتغيرات المشاهدة إلى عدد أقل من المتغيرات الكامنة (العوامل) والتي تفسر نسبة كبيرة من تباين المتغيرات الأصلية دون فقدان الكثير من المعلومات. اما القسم الثاني فيعرف بتقنيات التصنيف (Techniques de Classification)، فبعد تطبيق إحدى الطرق العاملية، غالبا ما ندعم تحليلنا بتقنيات التصنيف، والتي تصنف الأفراد والمتغيرات في مجموعات، تم الحصول عليها من خلال التجميع المتتالي للعناصر المتشابهة.

وسنركز في هذه المداخلة على طريقة من طرق التحليل العاملي وهي طريقة المركبات الرئيسية، وأسلوب من أساليب التصنيف وهو التصنيف الهرمي التصاعدي.

I. تحليل المركبات الرئيسية (ACP)

سأتطرق أولا إلى بعض الجوانب النظرية لطريقة تحليل المركبات الرئيسية، ثم كيفية تطبيقها على برنامج SPSS.

1. مفهوم طريقة تحليل المركبات الرئيسية:

تعد طريقة تحليل المركبات الرئيسية من أهم وأشهر طرق تحليل المعطيات العاملية، بل تعد الطريقة "الأم" لمعظم الطرق الوصفية متعددة الأبعاد[1].

أ. تعريف طريقة تحليل المركبات الرئيسية:

تحليل المركبات الرئيسية هو أسلوب رياضي يقوم على أساس تحويل مجموعة من المتغيرات التوضيحية المترابطة فيما بينها إلى مجموعة جديدة من المتغيرات غير المترابطة (أو المتعامدة – Orthogonal) تدعى المركبات الرئيسية[2]. حيث كل مركبة رئيسية هي عبارة عن توليفة خطية (Combinaison linéaire) للمتغيرات الأصلية. يحتوي هذا الجدول على n فرد و p متغير، فكل فرد يمثل في فضاء ذو p بعد، وكل متغير يمثل في فضاء ذو n بعد.

تسمح طريقة تحليل المركبات الرئيسية من جمع المتغيرات في مجموعات محدودة، تسمى عوامل او مركبات (Facteurs ou composantes)[3]. كما تسمح بتعيين الأفراد المتشابهة والأفراد المختلفة، بالإضافة إلى تحديد الترابط بين المتغيرات، واكتشاف المتغيرات المسؤولة عن التشابه أو الاختلاف بين الأفراد.

ب. مبدأ طريقة تحليل المركبات الرئيسية:

يتمثل مبدأ ACP في إسقاط سحابة نقاط الأفراد والمتغيرات من فضاء ذو p (n) بُعد إلى محاور ومستويات من 2 إلى 3 أبعاد فقط، مع الحفاظ على المسافات بين هذه الأفراد (والمسافات بين المتغيرات) قدر الإمكان، أي العمل على عدم تشويه سحابة النقاط الأصلية لجميع المتغيرات عند اسقاطها (الحصول على تمثيل أقل تشوها لسحابة النقاط)، وبالتالي المحافظة بأفضل شكل ممكن على المسافات الأصلية، وتحقيق أفضل تمثيل للتنوع والتباين الأصلي. وذلك بالاعتماد على التباين الكلي (Inertie).

يعتمد الإسقاط على مبدأ المحافظة على المسافات بين الأفراد في المتوسط قدر المستطاع عند الإسقاط[4]:

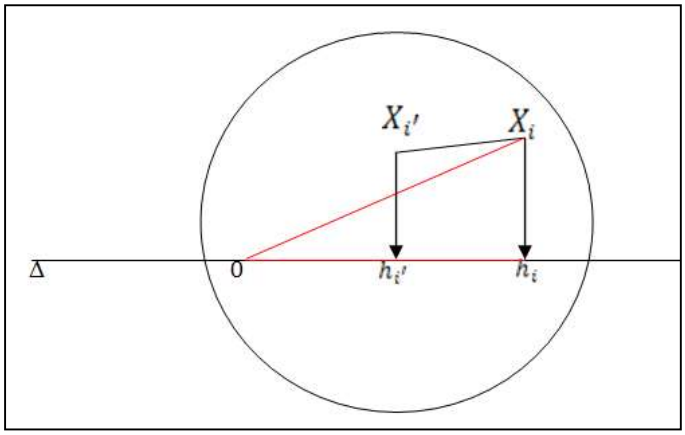
---

[1] Gilbert Saporta, « Probabilité, analyse des données, et statistique », Edition TECHNIP, 2ème Edition, Paris, 2006, p155.

[2] زياد رشاد الراوي، "طرق التحليل الإحصائي متعدد المتغيرات"، المعهد العربي للتدريب والبحوث الإحصائية، الطبعة الأولى، 2017، ص55.

[3] Jean Stafford, Paul Bodson, « L'analyse multivariée avec SPSS », Presse de l'Université du Québec, Canada, 2006, p58.

[4] Jean-Marie Bouruche, Gilbert Saporta, « L'analyse des données », Presses universitaires de France, 5ème Edition, 1992, p20.



أي: $d(X_i\ ,\ X_{i'}) \approx d(h_i\ ,\ h_{i'})$.

حيث: $X_i = (x_{i1} \quad x_{i2} \quad \dots \quad x_{ij} \quad \dots \quad x_{ip})$ X_i هي نقطة تمثيلية للفرد i في الفضاء ذو البعد p، أي لها p إحداثية) و h_i هي إسقاطها العمودي على المحور Δ.

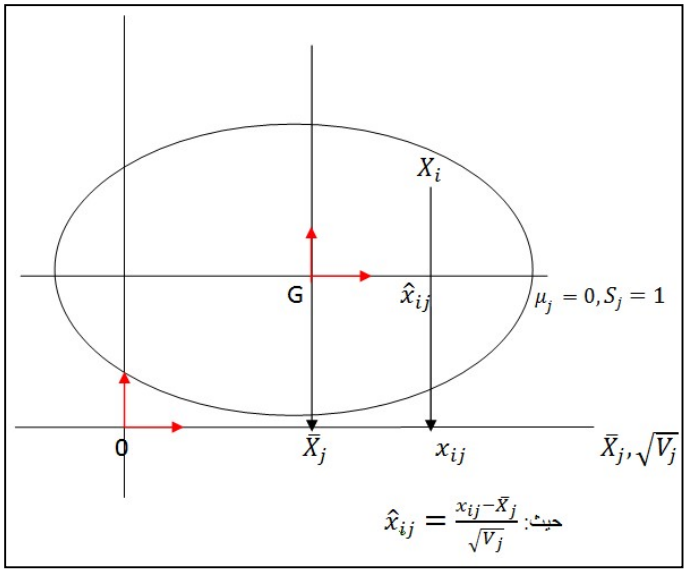
وحتى نتحصل على تمثيل جيد عند الإسقاط، يجب أن نجعل مسافة الإسقاط $0h_i$ أقرب ما يمكن من المسافة الحقيقة $0X_i$ (مع $0h_i \leq 0X_i$)، أي جعل $0h_i$ أكبر ما يمكن، وبالتالي فقدان أقل قدر ممكن من المعلومات.

عند الإسقاط على المستقيم Δ وللحصول على تمثيل أقل تشوه لسحابة النقاط، نأخذ مركز ثقل سحابة النقاط (G) كمركز للإسقاط، والذي يمثل متوسط المتغيرات (المبدأ 0 لا يعطي تمثيل جيد لتشتت سحابة النقاط).

عند جعل المتغيرات ممركزة (Centré) وذلك بطرح كل القيم من متوسطاتها، نكون قد حولنا المبدأ من 0 إلى مركز ثقل سحابة النقاط (G).

كما نقوم باختزال القيم (réduire) وذلك بالقسمة على الانحراف المعياري، وذلك لجعل القيم متجانسة (إهمال وحدات القياس)، حيث يصبح تباين القيم المختزلة هو 1 بالنسبة لجميع المتغيرات.

ويمكن توضيح ذلك بالشكل التقريبي التالي:



فيتضح من هذا التمثيل أن نقطة الإسقاط على المحور الجديد ما هي في الحقيقة إلا تمثيل للقيمة $\hat{x}_{ij}$.

2. خطوات إجراء طريقة تحليل المركبات الرئيسية:

لإجراء تحليل المركبات الرئيسية نتبع الخطوات التالية:

أ. تشكيل جدول البيانات الكمية X:

تدرس طريقة تحليل المركبات الرئيسية جدول البيانات الكمية (Tableau des données quantitative)، يضم أفراد ومتغيرات كمية مستمرة، قد تكون متجانسة أو غير متجانسة لكنها مترابطة فيما بينها[5].

تستخدم طريقة تحليل المركبات الرئيسية مع جدول البيانات الكمية، والذي يأخذ الشكل التالي:

المتغيرات / الأفراد	X_1	…	X_j	…	X_p
1	x_{11}				
2					
…					
i			x_{ij}		
…					
n					x_{np}

حيث: X_j يمثل المتغير j ($j = 1,2, \dots, p$).

و: الأعداد من 1 إلى n تمثل الأفراد.

و: x_{ij} تمثل قيمة المتغير X_j عند الفرد i.

نرمز لجدول البيانات الكمية بالرمز X، حيث: $X_{n\times p} = \begin{pmatrix} x_{11} & \cdots & x_{1j} & \cdots & x_{1p} \\ \vdots & \dots & \vdots & \dots & \vdots \\ x_{i1} & \dots & x_{ij} & \dots & x_{ip} \\ \vdots & \dots & \cdots & \cdots & \cdots \\ x_{n1} & \cdots & x_{nj} & \dots & x_{np} \end{pmatrix}$.

كل فرد i يُمثّل في فضاء ذو بعد p $\mathbb{R}^p$[6]: $X_i = (x_{i1} \quad x_{i2} \quad \cdots \quad x_{ip})$.

وكل متغير j يُمثّل في فضاء ذو بعد n $\mathbb{R}^n$: $X_j = \begin{pmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{nj} \end{pmatrix}$.

ب. حساب المتوسطات وتشكيل جدول البيانات الكمية الممركز $\tilde{X}$ (Tableau Centré):

جدول البيانات الكمية الممركز $\tilde{X}$ (X Tilde):

[5] Arnaud MARTIN, « L'analyse de données », Polycopié de cours ENSIETA, - Réf. : 1463, Septembre 2004, p23.

[6] Jean-Marie Bouruche, Gilbert Saporta, Op-cit, p26.

المتغيرات / الأفراد	$\tilde{X}_1$	…	$\tilde{X}_j$	…	$\tilde{X}_p$
1	$\tilde{x}_{11}$				
2					
…					
i			$\tilde{x}_{ij}$		
…					
n					$\tilde{x}_{np}$
G	$\bar{X}_1$		$\bar{X}_j$		$\bar{X}_p$

حيث: $\tilde{X}_j = X_j - \bar{X}_j$ $(j = 1,2, \dots, p)$. أي: $\tilde{x}_{ij} = x_{ij} - \bar{X}_j$.

و G هي نقطة مركز ثقل سحابة النقاط (إحداثياتها هي متوسطات المتغيرات)[7].

نرمز لجدول البيانات الكمية الممركز بالرمز $\tilde{X}$ ، حيث:

$$\tilde{X}_{n\times p} = \begin{pmatrix} \tilde{x}_{11} & \cdots & \tilde{x}_{1j} & \dots & \tilde{x}_{1p} \\ \vdots & \dots & \vdots & \dots & \vdots \\ \tilde{x}_{i1} & \cdots & \tilde{x}_{ij} & \dots & \tilde{x}_{ip} \\ \vdots & \dots & \cdots & \cdots & \cdots \\ \tilde{x}_{n1} & \cdots & \tilde{x}_{nj} & \dots & \tilde{x}_{np} \end{pmatrix}.$$

ج. حساب التباينات وتشكيل جدول البيانات الكمية المعياري $\hat{X}$ (Tableau normé):

جدول البيانات الكمية المعياري $\hat{X}$ (X Chapeau):

المتغيرات / الأفراد	$\hat{X}_1$	…	$\hat{X}_j$	…	$\hat{X}_p$
1	$\hat{x}_{11}$				
2					
…					
i			$\hat{x}_{ij}$		
…					
n					$\hat{x}_{np}$
G	$\bar{X}_1$		$\bar{X}_j$		$\bar{X}_p$
التباين	V_1		V_j		V_p

[7] Gilbert Saporta, Op-cit, p156.

حيث: $\hat{X}_j = \frac{X_j - \bar{X}_j}{\sqrt{V_j}}$ $(j = 1,2, \dots, p)$. أي: $\hat{x}_{ij} = \frac{x_{ij} - \bar{X}_j}{\sqrt{V_j}}$.

نرمز لجدول البيانات الكمية المعياري بالرمز $\hat{X}$، حيث:

$$\hat{X}_{n\times p} = \begin{pmatrix} \hat{x}_{11} & \cdots & \hat{x}_{1j} & \cdots & \hat{x}_{1p} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ \hat{x}_{i1} & \cdots & \hat{x}_{ij} & \cdots & \hat{x}_{ip} \\ \vdots & \cdots & \cdots & \cdots & \cdots \\ \hat{x}_{n1} & \cdots & \hat{x}_{nj} & \cdots & \hat{x}_{np} \end{pmatrix}.$$

د. حساب مصفوفة مصفوفة الارتباط R:

يمكن القيام بالتحليل بالمركبات الرئيسية إما بالاعتماد على مصفوفة التباين المشترك (Matrice de covariance)، أو مصفوفة الارتباط للمتغيرات التوضيحية، وإن نوع المصفوفة المفضل استخدامها يعتمد في الغالب على طبيعة المتغيرات قيد التحليل[8]. فإذا كانت للمتغيرات نفس وحدات القياس (تجانس الوحدات)، ونسمي في هذه الحالة التحليل بالمركبات الرئيسية غير معياري (ACP Non normé)، أو بالاعتماد على مصفوفة الارتباط R (Matrice de corrélation) وهذا إن لم تكن للمتغيرات نفس وحدات القياس (عدم تجانس الوحدات)، ونسمي في هذه الحالة التحليل بالمركبات الرئيسية معياري (ACP Normé).

مصفوفة الارتباط R تحسب بالاعتماد على الصيغة المصفوفية التالية:

$$R_{p\times p} = D_{p\times p} . \hat{X}^t_{p\times n} . \hat{X}_{n\times p}.$$

أي: $$R = \begin{pmatrix} \frac{1}{n} & \cdots & 0 & \cdots & 0 \\ \vdots & \cdots & \frac{1}{n} & \cdots & \vdots \\ 0 & \cdots & \vdots & \cdots & 0 \\ \vdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & \cdots & \frac{1}{n} \end{pmatrix} . \begin{pmatrix} \hat{x}_{11} & \cdots & \hat{x}_{i1} & \cdots & \hat{x}_{n1} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ \hat{x}_{1j} & \cdots & \hat{x}_{ij} & \cdots & \hat{x}_{nj} \\ \vdots & \cdots & \cdots & \cdots & \cdots \\ \hat{x}_{1p} & \cdots & \hat{x}_{ip} & \cdots & \hat{x}_{np} \end{pmatrix} . \begin{pmatrix} \hat{x}_{11} & \cdots & \hat{x}_{1j} & \cdots & \hat{x}_{1p} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ \hat{x}_{i1} & \cdots & \hat{x}_{ij} & \cdots & \hat{x}_{ip} \\ \vdots & \cdots & \cdots & \cdots & \cdots \\ \hat{x}_{n1} & \cdots & \hat{x}_{nj} & \cdots & \hat{x}_{np} \end{pmatrix}.$$

بعد الحساب، يمكن كتابة R على الشكل المصفوفي التالي: $$R = \begin{pmatrix} 1 & \cdots & r_{1j} & \cdots & r_{1p} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ r_{1j} & \cdots & 1 & \cdots & r_{jp} \\ \vdots & \cdots & \cdots & \cdots & \cdots \\ r_{1p} & \cdots & r_{jp} & \cdots & 1 \end{pmatrix}.$$

هـ. حساب القيم الذاتية والأشعة الذاتية لمصفوفة الارتباط R:

1. حساب القيم الذاتية:

لحساب القيم الذاتية لمصفوفة الارتباط R، نعتمد على العلاقة التالية:

$Ru = \lambda u$، حيث $\vec{u}$ هو الشعاع الذاتي للمصفوفة R.

النظام يمكن كتابته كما يلي:

$$Ru = \lambda u \Leftrightarrow (R - \lambda I)u = 0$$

ولكي يقبل النظام حل غير صفري لابد أن يكون: $|\mathrm{R} - \lambda \mathrm{I}| = 0$

[8] زياد رشاد الراوي، مرجع سابق، ص55.

وهي العلاقة التي تسمح بحساب القيم الذاتية للمصفوفة R.

2. إيجاد الأشعة الذاتية:

كل قيمة ذاتية يقابلها شعاع ذاتي، ولتحديد الأشعة الذاتية نعتمد على العلاقة التالية:

- ✓ إذا كانت λ_1 هي القيمة الذاتية الأولى، فإن الشعاع الذاتي الأول u_1 يحدد من العلاقة: $(R-\lambda_1 I)u_1=0$ ويجب أن يتحقق في هذا الشعاع: $u_1^t.u_1=1$.
- ✓ إذا كانت λ_2 هي القيمة الذاتية الثانية، فإن الشعاع الذاتي الثاني u_2 يحدد من العلاقة: $(R-\lambda_2 I)u_2=0$ ويجب أن يتحقق في هذا الشعاع[9]: $u_2^t.u_2=1$، و $u_1^t.u_2=0$.
- ✓ إذا كانت λ_3 هي القيمة الذاتية الثالثة، فإن الشعاع الذاتي الثالث u_3 يحدد من العلاقة: $(R-\lambda_3 I)u_3=0$ ويجب أن يتحقق في هذا الشعاع: $u_3^t.u_3=1$، و $u_1^t.u_3=0$، و $u_2^t.u_3=0$.
- ✓ ... وهكذا لبقية القيم الذاتية والأشعة الذاتية.

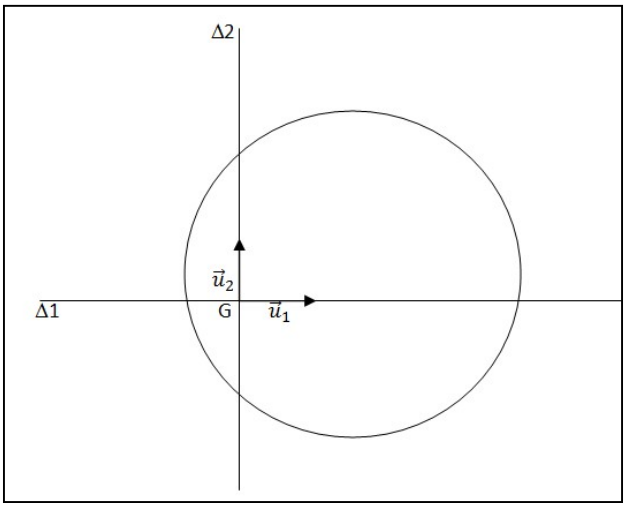
ملاحظة:

بعد ترتيب القيم الذاتية من الأكبر إلى الأصغر، القيمة الذاتية الأولى λ_1 تعطينا الشعاع الذاتي الأول u_1، والذي يعطينا بدوره المحور العاملي الأول (المستقيم: Δ_1)، وهذا الأخير يعطينا المركبة الرئيسية الأولى (C_1).

وبنفس الطريقة، القيمة الذاتية الثانية λ_2 تعطينا الشعاع الذاتي الثاني u_2، والذي يعطينا بدوره المحور العاملي الثاني (المستقيم: Δ_2)، وهذا الأخير يعطينا المركبة الرئيسية الثانية (C_2).

ثم المحورين الأوليين 1 و2 يشكلان لنا المستوي العاملي الأول (حيث يكون الشعاع الذاتي الأول متعامد مع الشعاع الذاتي الثاني، أي: $\vec{u}_1 \perp \vec{u}_2$، وهذا يعني: $\vec{u}_1.\vec{u}_2=0$)

ونشير إلى أن المركبات الرئيسية هي توليفة خطية للمتغيرات الأصلية[10] (مع مركبات الأشعة الذاتية)، وعدد المركبات الرئيسية هو نفسه عدد المتغيرات الأصلية، إلا أننا لا نستخدم إلا المركبة الأولى والثانية فقط، وفي بعض الأحيان نضيف الثالثة.



[9] زياد رشاد الراوي، مرجع سابق، ص58.

[10] Jean-Marie Bouruche, Gilbert Saporta, Op-cit, p21.

و. تحديد عدد المحاور التي تؤخذ في التحليل (Nombre d'axes à retenir):

هناك عدة معايير تسمح بتحديد عدد المحاور التي تؤخذ في التحليل، من بينها:

- ✓ **معيار كيزر Kaiser:** نأخذ كل القيم الذاتية الأكبر من 1 [11].
- ✓ **معيار نسبة المستوي العاملي الأول:** إذا كانت قيمة النسبة $\frac{\lambda_1+\lambda_2}{\sum_{i=1}^{p}\lambda_i} * 100$ أكبر من 80%، ففقدان المعلومات صغير، فلا داعي لإنشاء المستوي العاملي الثاني.
- ✓ **معيار نسبة المحور العاملي الثالث:** إذا كانت قيمة النسبة $\frac{\lambda_3}{\sum_{i=1}^{p}\lambda_i} * 100$ أكبر من 15%، فلا بد من إنشاء المستوي العاملي الثاني.

ز. الإسقاط على المحاور والمستويات:

1. إسقاط الأفراد:

إذا كان تحليل المركبات الرئيسية معياري (بالاعتماد على المصفوفة R)، فإن العلاقة التي تسمح بإسقاط الأفراد على المحاور العاملية الجديدة هي العلاقة التالية:

$$\boldsymbol{F_{n\times k} = \widehat{X}_{n\times p}.u_{p\times k}}.$$

حيث: n هو عدد الأفراد. p عدد المتغيرات. و k عدد المحاور التي تؤخذ في التحليل. و $u_{p\times k}$ هي الأشعة الذاتية للمصفوفة $R_{p\times p}$. إذا أخذنا محورين في التحليل: $u_{p\times 2} = (u_1 \;\; u_2)$.

2. إسقاط المتغيرات:

لاسقاط المتغيرات على المحاور العاملية، يمكن الاعتماد على العبارة التالية:

$$\boldsymbol{P_{p\times k} = (P_1 \quad P_2 \quad \cdots \quad P_k)}.$$

حيث $\boldsymbol{P_i = \sqrt{\lambda_i}.u_i}$ $(i = 1,2,\dots,k)$.

حيث $\boldsymbol{\lambda_i}$ هي القيم الذاتية للمصفوفة $\boldsymbol{R}$، و $\boldsymbol{u_i}$ هي الأشعة الذاتية الموافقة لها.

ح. التمثيل البياني للمتغيرات وللأفراد:

بعد تحديد المحاور والقيام بالإسقاط، نقوم بالتمثيل البياني للمتغيرات والأفراد.

1. التمثيل البياني للمتغيرات:

تمثل المتغيرات في دائرة مثلثية، ومن هذه الدائرة المثلثية يمكن استنتاج ما يلي:

- ➢ المسافات بين المتغيرات تدل على الارتباط (corrélation).
- ➢ الارتباط بين المتغيرات يكون أكثر دلالة إحصائية كلما كانت النقاط الممثلة لهذه المتغيرات بعيدة عن المبدأ (G).

[11] Alian Baccini, Philippe Besse, « Exploration Statistique », Institut National des Sciences Appliquées, Toulouse, Juin 2020, p41.

- إذا كان تمثيل المتغيرات في الدائرة المثلثية قريب من المبدأ (أي لم تكن قريبة من محيط الدائرة) [12]، **فلا يمكن تفسير علاقة الارتباط** بينهم، لأنهم يمثلون أفضل في محاور عاملية أخرى.
- إذا كان المتغير j قريب من المتغير 'j فهذا يدل على أن هذان المتغيران **مترابطان ايجابيا** (الزاوية المحصورة بينهما بالنسبة للمبدأ صغيرة وبالتالي تجب (cos) هذه الزاوية يكون قريب من 1).
- إذا كان المتغير j في الجهة المقابلة من المتغير 'j فهذا يدل على أن هذان المتغيران **مترابطان سلبيا** (الزاوية المحصورة بينهما بالنسبة للمبدأ كبيرة (قريبة من 180°) وبالتالي تجب (cos) هذه الزاوية يكون قريب من 1−).
- إذا كان المتغير j بعيد عن المتغير 'j فهذا يدل على أن هذان المتغيران **غير مترابطان** (الزاوية المحصورة بينهما بالنسبة للمبدأ قائمة (قريبة من 90°) وبالتالي تجب (cos) هذه الزاوية يكون قريب من 0).

2. التمثيل البياني للأفراد: من التمثيل البياني للأفراد يمكن استنتاج ما يلي:

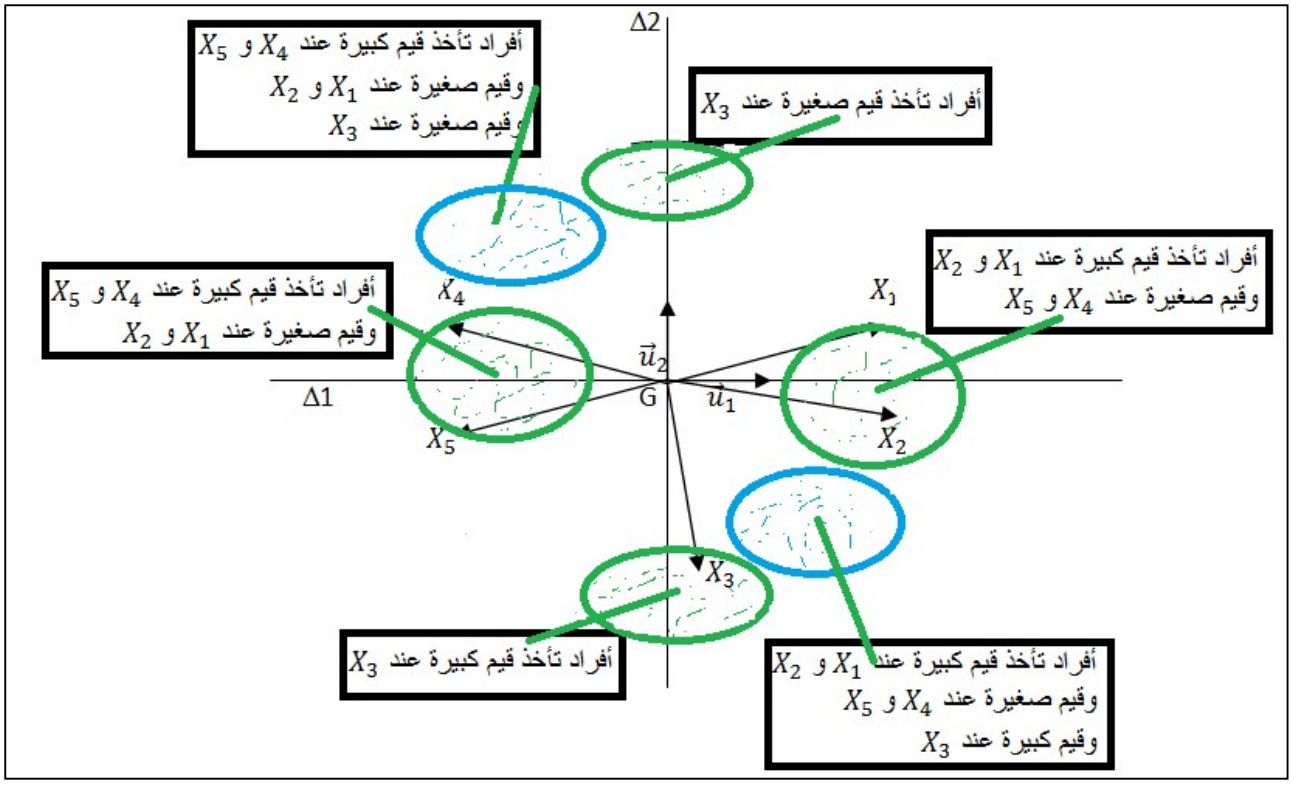
- المسافات بين الأفراد تدل على التشابه (Ressemblance).
- إذا كان الفرد i قريب من الفرد 'i فهذا يدل على أن هذان الفردان متشابهان، أي يأخذان قيما متقاربة على كل المتغيرات بشكل عام.
- إذا كان الفرد i بعيد عن الفرد 'i فهذا يدل على أن هذان الفردان مختلفان، أي يأخذان قيما متباعدة على كل المتغيرات بشكل عام.
- الأفراد القريبة من المبدأ (G) هي الأفراد التي قيمها عند جميع المتغيرات قريبة من المتوسط بشكل عام.

3. التمثيل البياني المشترك للأفراد والمتغيرات:

من التمثيل البياني المشترك للأفراد والمتغيرات يمكن استنتاج ما يلي:

- الفرد سيكون قريب (نفس الجانب) من المتغيرات التي له قيم كبيرة عندها، وبالعكس سيكون بعيد (من الجانب المقابل) من المتغيرات التي له قيم صغيرة عندها[13].
- تقارب نقطة فرد من نقطة متغير في نفس المحور العاملي، تدل على أن هذا المتغير له دور كبير في تفسير سلوك ذلك الفرد بشكل إيجابي، أما إذا كانا متقابلين على نفس المحور العاملي، فهذا يدل على أن هذا المتغير له دور كبير في تفسير سلوك ذلك الفرد كذلك لكن بشكل سلبي (عكسي).
- تفسير العلاقات بين الأفراد والمتغيرات تبعا لتمركزهم على المحاور العاملية وفق الشكل الموالي:

[12] Gilbert Saporta, Op-cit, p174.
[13] Arnaud MARTIN, Op-cit, p33.



مما يلاحظ كذلك أنه مثلا الأفراد الموجودة أسفل الشكل (عند X_3)، لا نفسر قيمها عند المتغيرات (X_1 و X_2 و X_4 و X_5) لأن هذه الأفراد تأخذ قيما متوسطة (valeurs moyennes) عند هذه المتغيرات (قيم قريبة من G).

3. تطبيق طريقة التحليل بالمركبات الرئيسية على برنامج SPSS

سنخصص هذا العنصر للدراسة التطبيقية، من خلال دراسة تحليلية إحصائية للعلاقة بين الفساد والجريمة والنمو الاقتصادي في عينة من الدول العربية، وقد تم اختيار 18 دولة عربية التي تتوفر بياناتها وهي: الإمارات، البحرين، الجزائر، مصر، العراق، الأردن، الكويت، لبنان، ليبيا، المغرب، عمان، قطر، السعودية، الصومال، السودان، سوريا، تونس، اليمن.

أ. التعريف بمتغيرات الدراسة:

- **مؤشر النمو الاقتصادي:** اعتمدنا لقياس النمو الاقتصادي على مؤشر نصيب الفرد من الناتج المحلي الخام، ونرمز له في هذه الدراسة بالرمز: GDP. والبيانات المتوفرة بالنسبة لهذا المؤشر للدول العربية الثمانية عشر تمتد من سنة 2000 إلى سنة 2020 (21 مشاهدة لكل دولة)، وتم أخذ متوسط هذه السنوات في التحليل. وقد تم الحصول على البيانات من قاعدة بيانات البنك الدولي[14].

[14] www.albankaldawli.org

❖ **مؤشر الفساد:** وهو المؤشر الصادر عن منظمة الشفافية الدولية، يقيس مستوياتها الفساد في القطاع العام. تم قياس مؤشر الفساد، وفق المعادلة التالية: CI=100−CPI. حيث CPI هو مؤشر مدركات الفساد –CPI) (Corruption Perceptions Index، والصادر عن منظمة الشفافية الدولية[15]. والبيانات المتوفرة بالنسبة لهذا المؤشر للدول العربية الثمانية عشر تمتد من 2010 إلى 2020 (11 مشاهدة لكل دولة)، وتم أخذ متوسط هذه السنوات في التحليل.

❖ **مؤشر الجريمة:** ونرمز له في هذه الدراسة بالرمز: CIR (Crime Index Rate). ويصدر مؤشر الجريمة (Crime Index Rate) عن قاعدة البيانات العالمية الصربية (نومبيو – Numbeo)[16]. والبيانات المتوفرة بالنسبة لهذا المؤشر للدول العربية الثمانية عشر تمتد من سنة 2012 إلى سنة 2022 (11 مشاهدة لكل دولة)، وتم أخذ متوسط هذه السنوات في التحليل.

ب. تطبيق طريقة تحليل المركبات الرئيسية (ACP) على البيانات:

تسمح طريقة تحليل المركبات الرئيسية باختزال الأبعاد وإعطاء تمثيل بياني للأفراد وللمتغيرات على مستوي عاملي، يمكن من خلاله تحليل مختلف العلاقات بين أفراد ومتغيرات الدراسة.

❖ **دراسة الارتباط بين الفساد والنمو الاقتصادي:**

Matrice de corrélation

		نصيب الفرد من الناتج المحلي الخام الحقيقي	مؤشر الفساد	مؤشر الجريمة
Corrélation	نصيب الفرد من الناتج المحلي الخام الحقيقي	1,000	-,800	-,805
	مؤشر الفساد	-,800	1,000	,897
	مؤشر الجريمة	-,805	,897	1,000

يتبين من مصفوفة الارتباط:

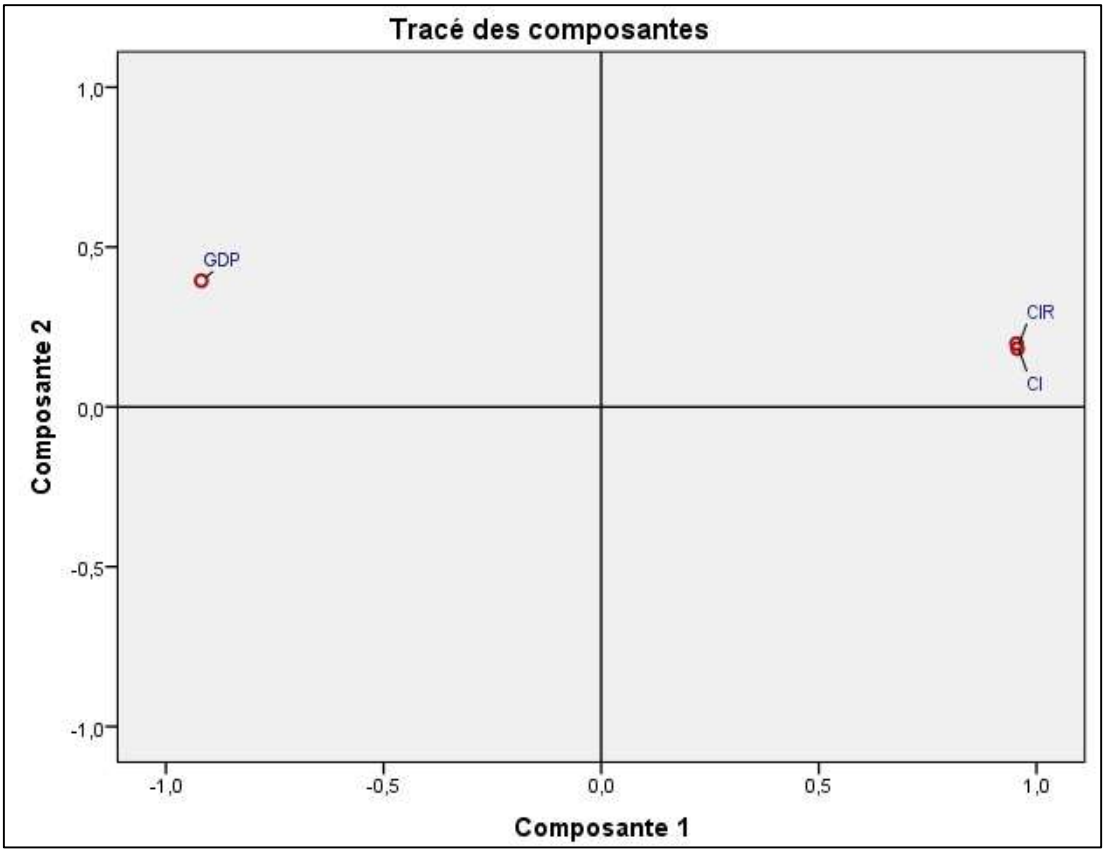
- قيمة معامل الارتباط بين نصيب الفرد من الناتج ومؤشر الفساد سالبة وقريبة من −1 (0,800−)، وهي تدل على وجود علاقة ارتباط عكسية قوية بين المؤشرين، فكلما انخفضت مستويات الفساد في البلدان العربية محل الدراسة كلما زادت نسب النمو الاقتصادي. والعكس، أي كلما ارتفعت مستويات الفساد انخفضت نسب النمو الاقتصادي.

[15] www.transparency.org/cpi
[16] www.numbeo.com

- قيمة معامل الارتباط بين نصيب الفرد من الناتج ومؤشر الجريمة سالبة وقريبة من −1 (−0,805)، وهي تدل على وجود علاقة ارتباط عكسية قوية كذلك بين المؤشرين، فكلما انخفضت مستويات الجريمة في البلدان العربية محل الدراسة كلما زادت نسب النمو الاقتصادي، والعكس.
- قيمة معامل الارتباط بين مؤشر الفساد ومؤشر الجريمة موجبة وقريبة من 1 (0,897)، وهي تدل على وجود علاقة ارتباط طردية قوية بين المؤشرين، فكلما ارتفعت مستويات الفساد ارتفعت مستويات الجريمة في البلدان العربية محل الدراسة، والعكس.

❖ **تحليل المتغيرات بعد الاسقاط على المحاور العاملية:**

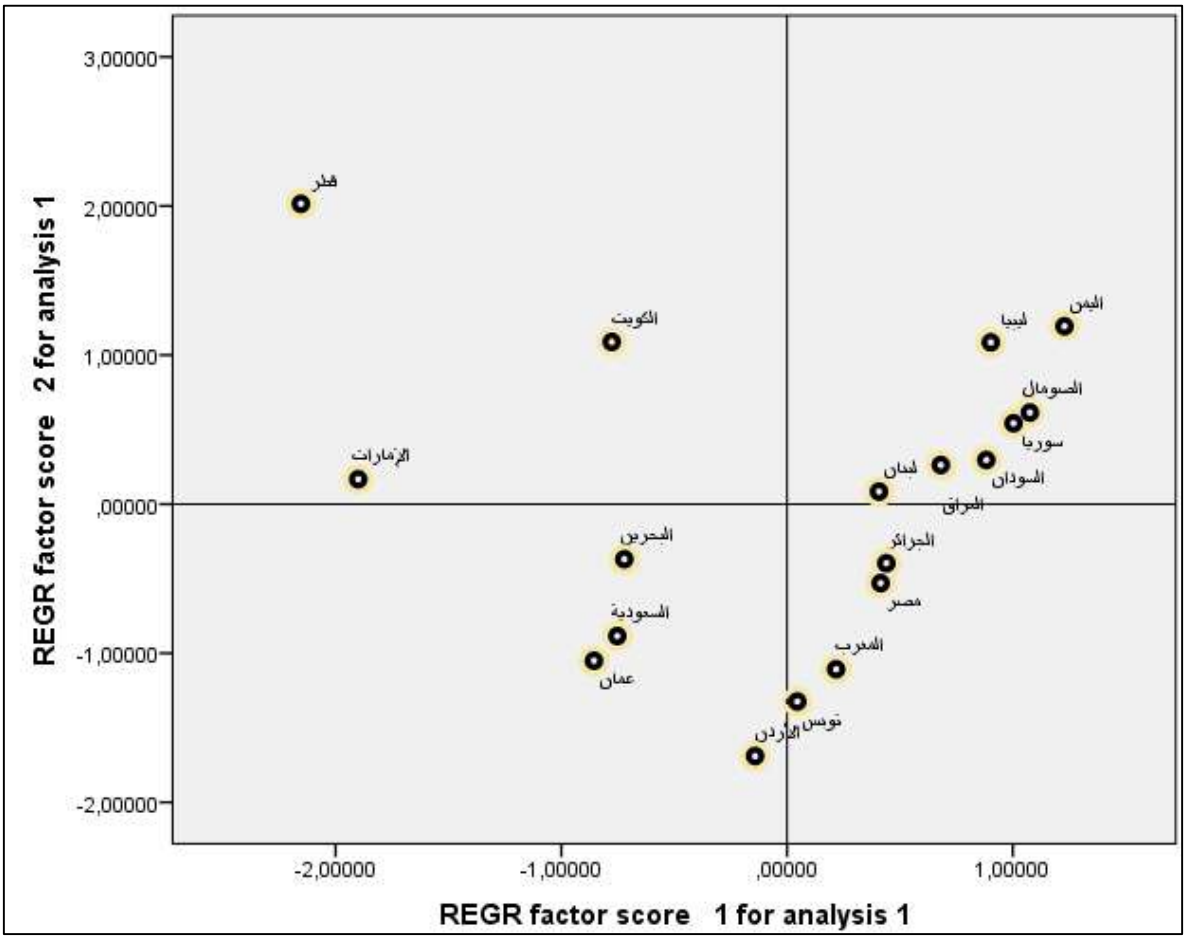
يمثل الشكل البياني أدناه اسقاط المتغيرات على المحورين العامليين الأول والثاني:



في الحقيقة هذا التمثيل البياني يؤكد نتائج تحليل الارتباط بين المتغيرات، فنلاحظ أن المتغيرات مرتبطة ارتباطا قويا بالمحور العاملي الأول، وهذا يؤكد الارتباط القوي لهذه المتغيرات وبما أن مؤشر الجريمة ومؤشر الفساد يقعان بالقرب من بعضها (المسافة بينهما قريبة جدا) وفي نفس جهة المحور الأول (يمين المحور) وبشكل مقابل لمؤشر نصيب الفرد من الناتج (يسار المحور)، فهذا كذلك يؤكد الارتباط الموجب بين مؤشري الجريمة والفساد في الدول العربية محل الدراسة، وارتباطهما السالب مع مؤشر نصيب الفرد من الناتج.

❖ **تحليل الأفراد (الدول) بعد الاسقاط على المحاور العاملية:**

يمثل الشكل البياني أدناه اسقاط الأفراد (الدول) على المحورين العامليين الأول والثاني:



يتبين من التمثيل البياني للدول (الأفراد) أن:

- ✓ بما أن المسافة بين الدول (قطر، الكويت، الإمارات) صغيرة، فهذا يدل على وجود تشابه كبير بين هذه الدول بالنسبة للثلاث مؤشرات المدروسة (الناتج، الفساد، الجريمة). كذلك فيما يخص الدول (البحرين، السعودية، عمان)، فقرب المسافة بينهما يدل على وجود تشابه كبير بينها من حيث المؤشرات المدروسة. وهذا ينطبق كذلك على الدول (اليمن، ليبيا، الصومال، سوريا، السودان، لبنان، العراق). نفس الملاحظة بالنسبة للدول (الجزائر، مصر، المغرب، تونس، الأردن) فهي متشابهة من حيث المؤشرات المدروسة.
- ✓ بما أن المسافة بين الدول (قطر، الكويت، الإمارات) والدول (اليمن، ليبيا، الصومال، سوريا، السودان، لبنان، العراق) كبيرة، فهذا يدل على وجود اختلاف كبير بين هاتين المجموعتين من الدول بالنسبة للثلاث مؤشرات المدروسة (الناتج، الفساد، الجريمة). كذلك فيما يخص الدول (البحرين، السعودية، عمان)، فبعد المسافة بينها وبين والدول (اليمن، ليبيا، الصومال، سوريا، السودان، لبنان، العراق) يدل على وجود اختلاف كبير بينها من حيث المؤشرات المدروسة.

❖ **تحليل الأفراد (الدول) والمتغيرات معا بعد الاسقاط على المحاور العاملية:**

يمثل الشكل البياني أدناه اسقاط الأفراد (الدول) والمتغيرات معا على المحورين العامليين الأول والثاني:

يتبين من التمثيل البياني المشترك أعلاه ما يلي:

- الدول (قطر، الكويت، الإمارات)، وبدرجة أقل (البحرين، السعودية، عمان)، لها مستويات مرتفعة في مؤشر نصيب الفرد من الناتج، وبالمقابل لها مستويات منخفضة في مؤشري الفساد والجريمة.
- الدول (الجزائر، مصر، المغرب، تونس، الأردن) لها مستويات متوسطة في جميع المؤشرات (نصيب الفرد من الناتج، الفساد، الجريمة).
- الدول (اليمن، ليبيا، الصومال، سوريا، السودان، لبنان، العراق) لها مستويات منخفضة في مؤشر نصيب الفرد من الناتج، وبالمقابل لها مستويات مرتفعة في مؤشري الفساد والجريمة.

يتبين من تطبيق طريقة تحليل المركبات الرئيسية وجود علاقة كبيرة بين مستويات الفساد والجريمة ومستويات النمو الاقتصادي، فالفساد والجريمة يؤثران سلبا على النمو فكلما كانت الدولة أقل فساد وأقل جريمة حققت مستويات نمو معتبرة، والعكس من ذلك فكلما ارتفعت مستويات الفساد والجريمة كانت مستويات النمو جد منخفضة. كما سمحت هذه الطريقة بتشكيل مجموعات للدول من حيث مؤشرات الدراسة، ولهذا فسنطبق في العنصر الموالي لهذه الورقة البحثية تقنية لتصنيف الدول، وهي طريقة التصنيف التسلسلي، وهذا بغية التحديد الدقيق للمجموعات المتجانسة.

II. التصنيف التسلسلي التصاعدي (CAH):

تركز الطرق العاملية على اختزال (تقليص) عدد المتغيرات، أما أساليب التصنيف فتركز في الغالب على تصنيف الأفراد في مجموعات (تقليص عدد الأفراد).

في الحقيقة يوجد تكامل بين الطرق العاملية وطرق التصنيف، فبعد تطبيق طريقة عاملية على البيانات الأصلية (كطريقة تحليل المركبات الرئيسية)، يتم اختزال المتغيرات في عدد محدود من المركبات (مركبتين أو ثلاث مركبات رئيسية)، وعلى أساس هذه المركبات نقوم بتصنيف الأفراد، والذي يسهل عملية التحليل والتفسير في الأخير.

1. مفهوم طريقة التصنيف التسلسلي التصاعدي:

سنخصص هذا العنصر لتعريف طريقة التصنيف التسلسلي التصاعدي ومبدؤها.

أ. تعريف طريقة التصنيف التسلسلي التصاعدي:

يقصد به تقسيم العناصر (الأفراد أو نادرا المتغيرات[17]) وترتيبها في مجموعات أو عناقيد (Classes, Clusters)، بحيث يكون المنتمين إلى نفس المجموعة متجانسين (متشابهين)، أما المنتمين إلى مجموعات مختلفة فهم غير متجانسين (مختلفين). والأسلوب العنقودي (التسلسلي) من شأنه تصنيف وحدات العينة إلى مجاميع (عناقيد أو مجموعات) غير معروفة مسبقا[18]، بالاعتماد على المسافات بينها. فكلما كانت المسافة بل فردين صغيرة فهذا يدل على أنهما متشابهان فيصنفان في نفس المجموعة. ويمكن أن تضم المجموعة فردا واحدا فقط أو عدة أفراد.

في البداية نكون أمام عدة عناصر أين يشكل كل عنصر مجموعة لوحده، فيتم ترتيبها بحيث نقوم بالجمع بين العناصر المتشابهة (القريبة) بشكل تسلسلي حتى تجتمع جميع العناصر في مجموعة واحدة فقط تمثل الجذر (la racine)[19] (الصعود من عدة مجموعات إلى مجموعة واحدة بشكل تسلسلي).

وطريقة التصنيف التسلسلي الصاعد تتناسب أفضل مع المتغيرات الكمية.

ب. مبدأ طريقة التصنيف التسلسلي التصاعدي:

يتمثل مبدأ التصنيف التسلسلي التصاعدي في التجميع التسلسلي للعناصر المتشابهة، بحيث نبدأ بتجميع الأفراد المتشابهة مثنى مثنى بالاعتماد على المسافات بينها (كل فردين قريبين هما متشابهان فيجمعان)، ثم نكرر العملية بشكل تسلسلي، وفي كل مرة نصعد إلى مستوى أعلى لضم أفراد أخرين للمجموعة، فيزيد عدد الأفراد في المجموعة الواحدة، ويقل عدد المجموعات بشكل تسلسلي تصاعدي.

وتقوم هذه الطريقة على مبدأ تعظيم التجانس (التشابه) بين الأفراد المنتمين لنفس المجموعة، وفي نفس الوقت تعظيم عدم التجانس (الاختلاف) بين المجموعات. ونشير إلى أن تجميع العناصر القريبة (المتشابهة) قد يكون جمع فرد مع فرد أو جمع مجموعة مع مجموعة أو جمع فرد مع مجموعة. والتصنيف التسلسلي يمثل بواسطة الشجرة التسلسلية أو شجرة التصنيف (Dendrogramme – Arbre de classification)[20].

---

[17] Arnaud MARTIN, Op-cit, p92.

[18] زياد رشاد الراوي، مرجع سابق، ص115.

[19] Alian Baccini, Philippe Besse, Op-cit, p84.

[20] Gilbert Saporta, Op-cit, p254.

لو نقوم بقطع (coupure) بخط أفقي على الشجرة التسلسلية، فإننا سنتحصل على تقسيم أو تجزئة (partition) للعناصر في مجموعات، فإننا كلما قمنا بالقطع أعلى الشجرة فسنتحصل على أقل عدد من المجموعات، وتكون هذه المجموعات أقل تجانسا[21]، وبالعكس كلما قمنا بالقطع أدنى الشجرة فسنتحصل على أكبر عدد من المجموعات، وتكون هذه المجموعات أكثر تجانسا.

2. خطوات إجراء طريقة التصنيف التسلسلي التصاعدي:

لإجراء تصنيف تسلسلي تصاعدي، يجب إتباع الخطوات التالية:

أ. جمع البيانات وتشكيل جدول البيانات:

بعد جمع البيانات حول الظاهرة المراد تصنيف عناصرها، نقوم بتشكيل جدول البيانات، والذي سنعتمد عليه في عملية التصنيف.

وجدول البيانات يضم n سطر تمثل الأفراد، و p عمود تمثل المتغيرات (غالبا ما تكون كمية).

المتغيرات / الأفراد	X_1	…	X_j	…	X_p
1	x_{11}				
2					
…					
i			x_{ij}		
…					
n					x_{np}

حيث: X_j يمثل المتغير j ($j = 1,2, \dots, p$).

و: الأعداد من 1 إلى n تمثل الأفراد (الحالات).

و: x_{ij} تمثل قيمة المتغير X_j عند الفرد i.

ب. حساب المسافات بين العناصر وتشكيل مصفوفة المسافات:

انطلاقا من جدول البيانات، نقوم بحساب المسافات بين العناصر ثم نشكل مصفوفة المسافات (Matrice des distances) (أو مصفوفة القرب Matrice de proximité)، ويستخدم في ذلك عدة مقاييس، ومن أشهر هذه المقاييس والأكثر استخداما المقياس الإقليدي (Euclidienne) والمقياس الاقليدي مربع.

الصيغ الرياضية لحساب المسافات وفقا لهاذين المقياسيين هي[22]:

✓ المسافة الاقليدية بين العنصرين i و i' هي: $d_e(i,i') = \sqrt{\sum_{j=1}^{p}\left(x_{ij} - x_{i'j}\right)^2}$.

[21] Ludovic Lebart, Alain Morineau, Marie Piron, op-cit, p155.

[22] صواليلي صدر الدين، مرجع سابق، ص113.

✓ المسافة الاقليدية مربع بين العنصرين i و i' هي: $d_e^2(i,i') = \sum_{j=1}^{p}\left(x_{ij} - x_{i'j}\right)^2$.

بعد حساب المسافات بين جميع العناصر، نشكل مصفوفة المسافات، والتي تضم المسافات بين العناصر (بين كل عنصرين).

ج. تحديد مؤشر التجميع:

يمكن حساب المسافة بين عنصرين بالصيغة السابقة بسهولة، لكن في حالة حساب المسافة بين عنصر ومجموعة او بين مجموعتين فنعتمد على بمؤشر التجميع (Indice d'agrégation) ويرمز له بالرمز δ. ومن بين أهم مؤشرات التجميع:

➢ **مؤشر التجميع للمسافة المتوسطة (Saut moyen):**

نقوم بتجميع العناصر وفقا لهذا المؤشر عنصر بعنصر، حيث نبدأ بجمع أقرب عنصرين، ثم أقرب العناصر بأخذ متوسط المؤشرات (Moy)، ... وهكذا. وذلك بالاعتماد على المسافات بين العناصر.

نحدد مؤشر التجميع كما يلي: $\delta(C_1, C_2) = \underset{\substack{x_i \in C_1 \\ x_j \in C_2}}{\text{Moy}}\{d^2(x_i, x_j)\}$.

حيث C_1 و C_2 هما مجموعتان من العناصر.

لضم أي عنصر من المجموعة C_2 (وليكن مثلا الفرد x_2) إلى مجموعة أخرى (ولتكن مثلا C_1)، نقوم بحساب المؤشر التالي: $\delta(\{C_1\}, x_2) = \underset{\substack{x_{i1} \in C_1 \\ x_2 \in C_2}}{\text{Moy}}\{d^2(x_{i1}, x_2)\}$.

➢ **مؤشر التجميع لـ "وارد" Ward:**

يعتبر مؤشر التجميع لوارد من أهم وأشهر مؤشرات التجميع، وهو الأكثر استخداما ويوصى به، ويقوم على مبدأ تصغير مجموع المربعات (التباين) لكل زوجين من العناصر الممكن تشكيلها في كل مرحلة[23]، من خلال تحليل التباين. ويحسب مؤشر التجميع من خلال مركز ثقل المجموعات.

تقوم هذه الطريقة على مبدأ تجميع الأفراد من خلال تصغير التباين داخل المجموعة وتعظيم التباين بين المجموعات[24]. وهذه الطريقة تندرج في إطار صيغة Lance et Williams [25].

مؤشر التجميع لعنصرين x_1 و x_2 هو: $\delta(x_1, x_2) = \frac{1}{2}d^2(x_1; x_2)$.

لضم العنصر x إلى مجموعة تضم مجموعتين جزئيتين (C_1, C_2)، نقوم بحساب:

$$\delta[(C_1, C_2), x] = \frac{(n_{C_1}+n_x)\delta(C_1,x)+(n_{C_2}+n_x)\delta(C_2,x)-n_x\delta(C_1,C_2)}{n_{C_1}+n_{C_2}+n_x}.$$

لضم المجموعة C_2 التي تضم عنصرين (x_1, x_2) إلى المجموعة C_1، نقوم بحساب:

$$\delta[C_2(x_1, x_2), C_1] = \frac{(1+n_{C_1})\delta(x_1,C_1)+(1+n_{C_1})\delta(x_2,C_1)-n_{C_1}\delta(x_1,x_2)}{n_{x_1}+n_{x_2}+n_{C_1}}.$$

[23] صواليلي صدر الدين، مرجع سابق، ص123.

[24] Arnaud MARTIN, Op-cit, p93.

[25] Gilbert Saporta, Op-cit, p259.

لضم العنصر x إلى مجموعة تضم عنصرين x_1 و x_2، نقوم بحساب:

$$\delta[(x_1, x_2), x] = \frac{2\delta(x_1,x)+2\delta(x_2,x)-\delta(x_1,x_2)}{3}.$$

د. تجميع العناصر بشكل تسلسلي تصاعدي وتكوين المجموعات:

تتم عملية التجميع بشكل تسلسلي تصاعدي كما يلي:

1) بالاعتماد على مصفوفة المسافات نقوم بجمع أقرب عنصرين (أصغر مسافة)، ليشكلا معا مجموعة واحدة (عنصر جديد)، فيكون هذا أول تجميع لـ $n-1$ مجموعة[26].
2) نقوم بتشكيل مصفوفة مسافات جديدة بعد تجميع العنصرين السابقين.
3) نعيد حساب المسافات بين المجموعة الجديدة وبقية العناصر بالاعتماد على مؤشر التجميع (المسافات الأولية بين العناصر خارج المجموعة المشكلة لا تتغير).
4) نقوم بجمع أقرب عنصرين (فرد أو مجموعة)، ليشكلا معا مجموعة جديدة.
5) .....
6) نكرر هذه العمليات للتجميع بشكل تسلسلي تصاعدي حتى تجمع جميع العناصر في مجموعة واحدة.
7) إذا كان هناك n عنصر فسنحتاج إلى $n-1$ عملية تجميع للعناصر.

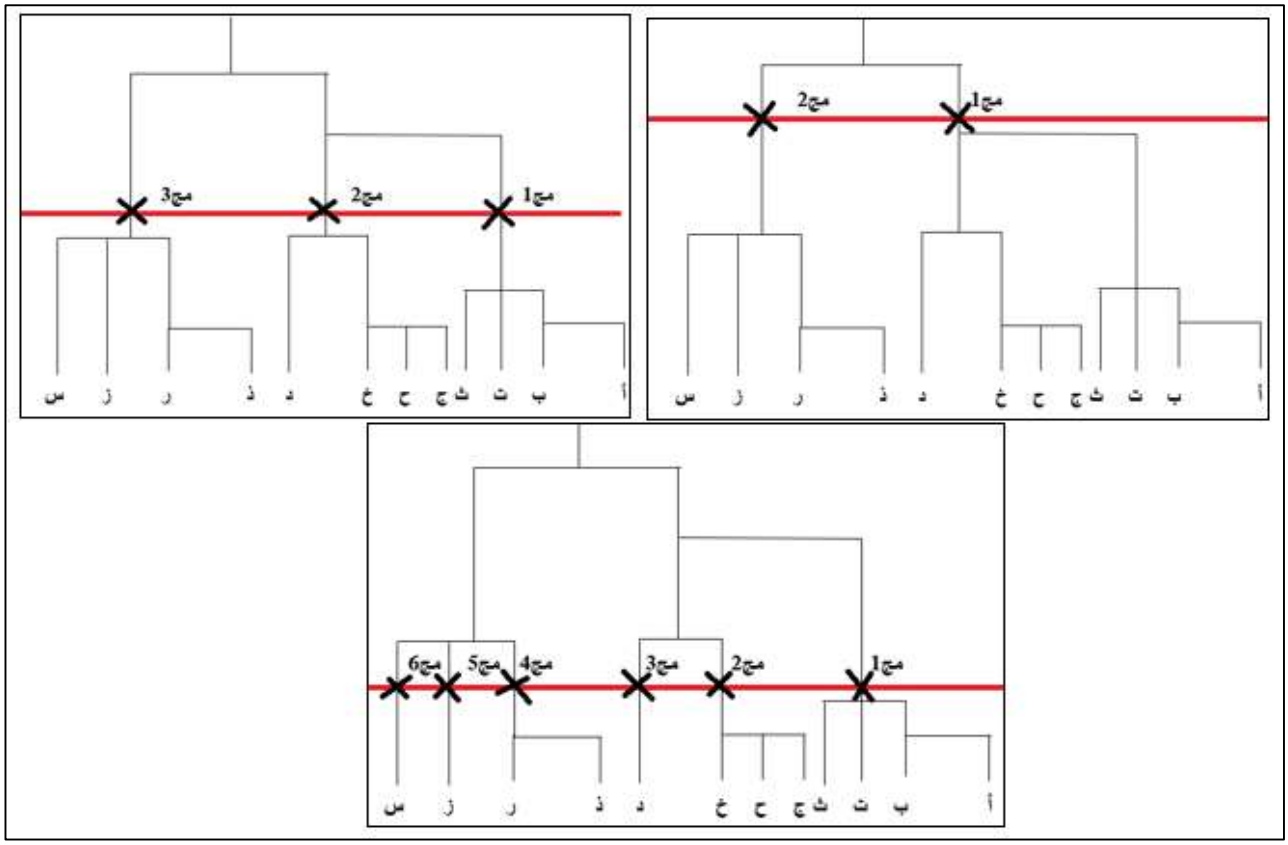
هـ. تحديد مستوى القطع (العدد الأنسب من المجموعات):

نقوم في هذه الخطوة بتحديد مستوى القطع في الشجرة، أي البحث عن التقسيم الأكثر أهمية في الشجرة، وذلك بالاعتماد على التباين، فالمستوى الذي يكون عنده التباين داخل المجموعات صغير والتباين بين المجموعات كبير هو أفضل مستوى للتقسيم.

ونستعين للقيام بتقسيم العناصر في مجموعات على بعض المحددات، منها: شجرة التصنيف (Dendrogramme)، ومنحنى المؤشرات (Courbe des indices)، وحجم العينة (عدد الأفراد)، وكيفية تفسير النتائج.

فلتحديد المجموعات على شجرة التصنيف نقوم برسم خط أفقي عند مستوى القطع المرغوب فيه، فيتحدد عدد المجموعات بحسب عدد مرات التقاطع بين المستقيم وفروع الشجرة، ونستعين في ذلك بعدة معايير، من أهمها ما يعرف بالقفزة (Le saut)، والتي تشير إلى تضاعف البعد بين المجموعات، فيتم القطع عند القفز الأعلى نسبيا (Le saut le plus élevé)، والذي يشير إلى ارتفاع كبير لمؤشر التجميع مقارنة بما قبله، والقطع يتم بعد التجميع الموافق لقيم صغيرة لمؤشر التجميع، وقبل التجميع الموافق لقيم كبيرة للمؤشر (وهو المستوى الذي يكون بعده خسارة كبيرة في المعلومات).

[26] Ludovic Lebart, Alain Morineau, Marie Piron, op-cit, p157.



نلاحظ من الشكل أعلاه أنه في نفس شجرة التصنيف، يختف عدد المجموعات بحسب موضع خط القطع، فقد تكون مجموعتين فقط إذا كان القطع أعلى الشجرة، وقد يرتفع إلى ستة مجموعات إذا نزلنا إلى مستويات دنيا من الشجرة.

و. تحليل وتفسير النتائج:

نعتمد لتحليل وتفسير النتائج بشكل أساسي على قراءة الشجرة التسلسلية وجدول التصنيف، وهذا بعد تصنيف جميع العناصر في مجموعات.

ويجب أن يتم التفسير من أعلى إلى أسفل، من أجل فحص وتحليل المجموعات التي تحتوي على عدد قليل من الفئات أولاً، ثم الخوض في التحليل الأكثر تفصيلاً (عدد كبير من المجموعات) [27].

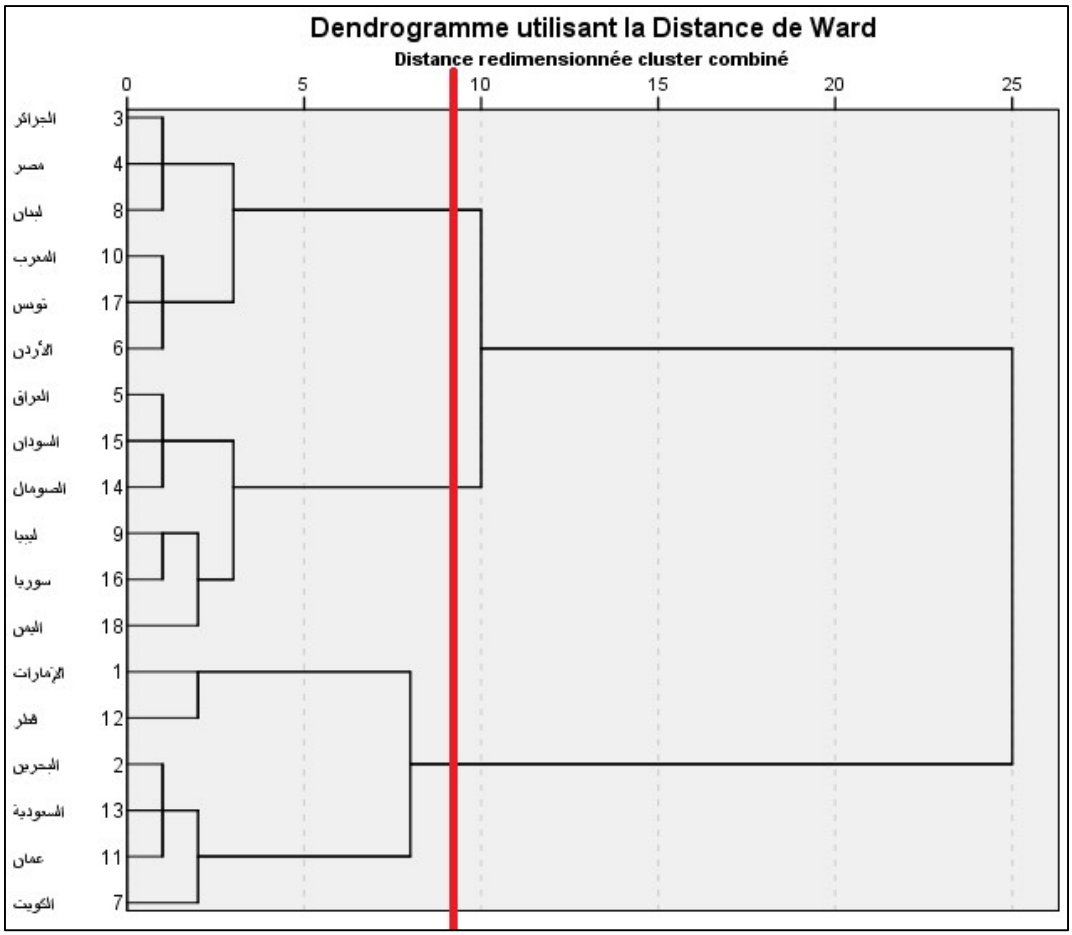
ولتفسير وتحليل النتائج نتبع الخطوات التالية:

- ✓ تحديد مستوى التقسيم (مستوى القطع في الشجرة).
- ✓ تحديد جميع المجموعات بناء على التقسيم السابق.
- ✓ تحديد جميع العناصر التي تنتمي لكل مجموعة.
- ✓ تشكيل جدول يلخص: المجموعات، الأفراد التي تنتمي لكل مجموعة، والمتغيرات الممثلة لكل مجموعة.

[27] Arnaud MARTIN, Op-cit, p98.

3. تطبيق طريقة التصنيف التسلسلي التصاعدي على برنامج SPSS:

تم تطبيق طريقة التصنيف التسلسلي التصاعدي على بيانات الدراسة قصد تصنيف الدول العربية محل الدراسة في مجموعات. يمثل الجدول وشجرة التصنيف أدناه تصنيف الدول في ثلاث مجموعات من حيث مؤشرات الدراسة (الناتج، الفساد، الجريمة):



Cluster(s) d'affectation

Observation	Clusters 3
1:الإمارات	1
2:البحرين	1
3:الجزائر	2
4:مصر	2
5:العراق	3
6:الأردن	2
7:الكويت	1
8:لبنان	2
9:ليبيا	3
10:المغرب	2
11:عمان	1
12:قطر	1
13:السعودية	1
14:الصومال	3
15:السودان	3
16:سوريا	3
17:تونس	2
18:اليمن	3

يتبين من الجدول وشجرة التصنيف وجود ثلاث مجموعات:

- ✓ المجموعة الأولى تضم الدول (6 دول): الإمارات، قطر، البحرين، السعودية، عمان، الكويت.

  وهي دول لها مستويات مرتفعة في مؤشر نصيب الفرد من الناتج، وبالمقابل لها مستويات منخفضة في مؤشري الفساد والجريمة. مع ملاحظة أن الدولتين قطر والإمارات لها مستويان نمو أفضل من الدول الأربعة المتبقية.
- ✓ المجموعة الثانية تضم الدول (6): الجزائر، مصر، لبنان، المغرب، تونس، الأردن.

  وهي دول لها مستويات متوسطة في جميع المؤشرات (نصيب الفرد من الناتج، الفساد، الجريمة).
- ✓ المجموعة الثالثة تضم الدول (6): العراق، السودان، الصومال، ليبيا، سوريا، اليمن.

  وهي دول لها مستويات منخفضة في مؤشر نصيب الفرد من الناتج، وبالمقابل لها مستويات مرتفعة في مؤشري الفساد والجريمة.

وهذا ما يؤكد مرة أخرى وجود علاقة عكسية قوية بين مستويات النمو المحققة وبين مستويات الفساد والجريمة، فالدول العربية الأقل فسادا وفيها مستويات جريمة ضعيفة هي الدول الأكثر نموا، أما الدول التي تفشى فيها الفساد والجريمة بمستويات مرتفعة فهي الدول الأقل نموا.

خاتمة:

كانت الغاية من هذه المداخلة ابراز أهمية تحليل البيانات الإحصائية بالاعتماد على الطرق العاملية وأساليب التصنيف كأسلوب كمي يسمح بالفهم الدقيق للظواهر الاقتصادية والاجتماعية. وقد تم التركيز على طريقة تحليل المركبات الرئيسية وأسلوب التصنيف التسلسلي التصاعدي.

وبعد التطرق إلى هاذين الأسلوبين من الناحية النظرية (التعريف، المبدأ، خطوات التطبيق)، تم تطبيق هاذين الأسلوبين على بيانات حقيقية خاصة بعينة من الدول العربية (18 دولة عربية)، وقد تم الاعتماد على ثلاث متغيرات هي: نصيب الفرد من الناتج المحلي الخام، مؤشر الفساد، مؤشر الجريمة.

ومن النتائج التي توصلنا إليه من خلال هذه الدراسة التطبيقية:

- ✓ وجود علاقة عكسية قوية بين مستويات النمو المحققة وبين مستويات الفساد والجريمة، فالدول العربية الأقل فسادا وفيها مستويات جريمة ضعيفة هي الدول الأكثر نموا، أما الدول التي تفشى فيها الفساد والجريمة بمستويات مرتفعة فهي الدول الأقل نموا.
- ✓ تصنف الدول العربية محل الدراسة إلى أربع مجموعات من حيث المؤشرات المدروسة (نصيب الفرد من الناتج، مؤشر الفساد، مؤشر الجريمة):
  - المجموعة الأولى تضم الدول (6 دول): الإمارات، قطر، البحرين، السعودية، عمان، الكويت.

وهي دول لها مستويات مرتفعة في مؤشر نصيب الفرد من الناتج، وبالمقابل لها مستويات منخفضة في مؤشري الفساد والجريمة. مع ملاحظة أن الدولتين قطر والإمارات لها مستويات نمو أفضل من الدول الأربعة المتبقية.

  - المجموعة الثانية تضم الدول (6): الجزائر، مصر، لبنان، المغرب، تونس، الأردن.

وهي دول لها مستويات متوسطة في جميع المؤشرات (نصيب الفرد من الناتج، الفساد، الجريمة).

  - المجموعة الثالثة تضم الدول (6): العراق، السودان، الصومال، ليبيا، سوريا، اليمن.

وهي دول لها مستويات منخفضة في مؤشر نصيب الفرد من الناتج، وبالمقابل لها مستويات مرتفعة في مؤشري الفساد والجريمة.

5. قائمة المراجع:

1) صواليلي صدر الدين، " تحليل المعطيات"، دار هومة للطباعة والنشر والتوزيع، 2011، الجزائر.

2) زياد رشاد الراوي، "طرق التحليل الإحصائي متعدد المتغيرات"، المعهد العربي للتدريب والبحوث الإحصائية، الطبعة الأولى، 2017.

3) المعجم الوسيط، مجمع اللغة العربية، مكتبة الشروق الدولية، 2004.

4) Alian Baccini, Philippe Besse, « Exploration Statistique », Institut National des Sciences Appliquées, Toulouse, Juin 2020.

5) Arnaud MARTIN, « L'analyse de données », Polycopié de cours ENSIETA, – Réf. : 1463, Septembre 2004.

6) Gilbert Saporta, « Probabilité, analyse des données, et statistique », Edition TECHNIP, 2ème Edition, Paris, 2006.

7) Jean Stafford, Paul Bodson, « L'analyse multivariée avec SPSS », Presse de l'Université du Québec, Canada, 2006.

8) Jean–Marie Bouruche, Gilbert Saporta, « L'analyse des données », Presses universitaires de France, 5ème Edition, 1992.

9) Ludovic Lebart, Alain Morineau, Marie Piron, « Statistique exploratoire multidimensionnelle », DUNOD, Paris, 1995.

10) الموقع الرسمي للبنك الدولي: https://data.albankaldawli.org/country