المحور الثالث: التحليل بالمركبات الأساسية **ACP:**

يهدف إلى تقليل الأبعاد مع الحفاظ على أكبر قدر ممكن من التباين، وذلك من خلال تحويل مجموعة من المتغيرات المترابطة فيها بينها إلى مجموعة من المتغيرات الجديدة غير المترابطة، تعرف بالمركبات الأساسية.

خطوات اجراء طريقة تحليل المركبات الأساسية:

1-تشكيل جدول البيانات الكمية:

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

كل فرد i يُمثّل في فضاء ذو بعد p $\mathbb{R}^p$ [8]: $X_i = (x_{i1} \quad x_{i2} \quad \cdots \quad x_{ip})$.

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

2-حساب المتوسطات وتشكيل جدول البيانات الممركزة:

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

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

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

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

3-توحيد البيانات: توجد عدة طرق أهمها التوحيد المعياري، وذلك من خلال العلاقة التالية:

$$X_{cr} = \frac{X - \mu}{\sigma}$$

ونحسب :

- **المصفوفة الممركزة X_c**

$$X_c = X - \mu$$

- **المصفوفة المعيارية X_{cr}**

$$X_{cr} = \frac{X - \mu}{\sigma}$$

4-إيجاد المصفوفة الأساسية للتحليل: وهناك نوعان أساسيان من المصفوفات:

مصفوفة التباين المشترك V:

مصفوفة التباين المشترك V تحسب بالاعتماد على الصيغة المصفوفية التالية:

$$.V = \begin{pmatrix} V_1 & \cdots & Cov(X_1, X_j) & \dots & Cov(X_1, X_p) \\ \vdots & \dots & \vdots & \dots & \vdots \\ Cov(X_1, X_j) & \dots & V_j & \dots & Cov(X_j, X_p) \\ \vdots & \dots & \dots & \dots & \dots \\ Cov(X_1, X_p) & \cdots & Cov(X_j, X_p) & \dots & V_p \end{pmatrix}$$

حيث يحسب تباين المتغير X بالعلاقة التالية:

$$V(x_i) = \frac{\sum_{j=1}^{n}(x_{ij} - \mu)^2}{n}$$

ويمكن حساب مصفوفة التباينات انطلاقا من مصفوفة البيانات الاصلية وفقا لعلاقة التالية:

$$\mathrm{V} = X_c^T . P . X_c = \frac{1}{n} X_c^T . X_c$$

مصفوفة الارتباط R:

مصفوفة الارتباط R تحسب بالاعتماد على الصيغة المصفوفية التالية:

$$.R = \begin{pmatrix} 1 & \cdots & r_{1j} & \cdots & r_{1p} \\ \vdots & \dots & \vdots & \dots & \vdots \\ r_{j1} & \cdots & 1 & \cdots & r_{jp} \\ \vdots & \dots & \cdots & \cdots & \cdots \\ r_{p1} & \cdots & r_{pj} & \dots & 1 \end{pmatrix}$$

ويمكن حساب مصفوفة الارتباطات انطلاقا من مصفوفة البيانات الاصلية وفقا لعلاقة التالية:

$$R = X_{cr}^{T}.P.X_{cr} = \frac{1}{n}X_{cr}^{T}.X_{cr}$$

5-حساب القيم الذاتية والمتجهات الذاتية :

كما تطرقنا اليها في الفصل السابق، حيث يكون حساب القيم الذاتية من خلال احدى العلاقتين:

$$P_V(\lambda) = det(V - \lambda I) = 0$$

$$P_R(\lambda) = det(R - \lambda I)$$

ويكون حساب المتجهات الذاتية (الاشعة الذاتية) بأحد العلاقتين:

$$(V - \lambda I)\vec{u} = 0$$

$$(R - \lambda I)\vec{u} = 0$$

6-اختيار عدد المركبات الرئيسية: وتوجد عدة طرق من أبرزها:

6-1- قاعدة نسبة التباين المفسر: وتحسب من خلال العلاقة التالية:

$$Expained\ Variance\ Ratio = \frac{\lambda_i}{\sum \lambda_i}$$

ونرتب القيم ترتيبا تنازليا في جدول القيم الذاتية كالتالي:

	F_1	F_2	F_3	...	F_k	Σ
القيم الذاتية	λ_1	λ_2	λ_3	...	λ_k	$\sum \lambda_i$
النسب	$\frac{\lambda_1}{\sum \lambda_i}$	$\frac{\lambda_2}{\sum \lambda_i}$	$\frac{\lambda_3}{\sum \lambda_i}$	...	$\frac{\lambda_k}{\sum \lambda_i}$	1
النسب التراكمية	$\frac{\lambda_1}{\sum \lambda_i}$	$\frac{\lambda_1 + \lambda_2}{\sum \lambda_i}$	$\frac{\lambda_1 + \lambda_2 + \lambda_3}{\sum \lambda_i}$	...	$\frac{\lambda_1 + \lambda_2 + \lambda_3 + \cdots + \lambda_k}{\sum \lambda_i}$	-

ونحدد عدد المركبات التي تفسر على الأقل نسبة تباين 75%، وذلك بالبحث في النسب التراكمية الأكبر من 75%.

6-2- قاعدة كايزر: وتنص على الاحتفاظ بالمركبات ذات القيم الذاتية الأكبر من متوسط القيم الذاتية، حيث:

$$\bar{\lambda} = \frac{\sum_{i=1}^{k} \lambda_i}{k}$$

7-اسقاط الافراد على المركبات الرئيسية:

ويتم ذلك من خلال ضرب مصفوفة البيانات الاصلية الممركزة أو المعيارية (حسب المصفوفة المعتمدة في التحليل) في مصفوفة الأشعة الذاتية الموحدة. من خلال أحد العلاقتين:

$$F = X_c . u$$

$$F = X_{cr} . u$$

8-تحميلات المركبات أو العوامل:

ويتم ذلك من خلال ضرب المتجهات الذاتية في جذر القيم الذاتية الموافقة لها، كما تبينه العلاقة التالية:

$$L = u . \sqrt{\lambda_i}$$

9- نسب تمثيل الافراد على المحاور:

ولحساب النسب نستخدم العلاقة التالية:

$$Cos_{ik}^2 = \frac{F_{ik}^2}{\sum_{j=1}^{p} F_{ij}^2}$$

10- نسب مساهمة الافراد في تشكيل المحاور:

ونميز بين نوعين من المساهمات:

المساهمة المطلقة: وتحسب من العلاقة التالية:

$$C_{ik} = \frac{F_{ik}^2}{\lambda_k}$$

المساهمة النسبية: وتحسب من العلاقة التالية:

$$C_{ik} = \frac{F_{ik}^2}{\sum_{i=1}^{n} F_{ik}^2}$$