المحور الرابع: التحليل العاملي للتوفيقات AFC

1. مقدمة

التحليل العاملي للتوفيقات $(Analyse\ factorial\ Correspondance)$ هو تقنية إحصائية تستخدم لتحليل البيانات النوعية الفئوية، تعتمد على تحليل الجداول المتقاطعة (المزدوجة) حيث تساعد في فهم العلاقات بين الفئات. يهدف التحليل العاملي للتوفيقات إلى:

- فهم العلاقات بين الفئات وتبسيط التباين عبر تحديد المحاور الرئيسية الأكثر تأثيرا.
- تمثيل الفئات في فضاء ثنائي أو ثلاثي الابعاد، مما يساعد في توضيح التشابهات والاختلافات بينها.

2. خطوات إجراء التحليل العاملي للتوفيقات

1.2. الجدول المتقاطع: هو عبارة عن جدول تكراري مزدوج يوضح العلاقة بين الفئات في الصفوف والأعمدة. يكون الجدول المتقاطع من الشكل التالي:

الجدول 01: الجدول المزدوج

	$Y_{.1}$	$Y_{.2}$	...	$Y_{.m}$	$\sum$
$X_{1.}$	n_{11}	n_{12}	...	n_{1m}	$n_{1.}$
$X_{2.}$	n_{21}	n_{22}	...	n_{2m}	$n_{2.}$
⋮	⋮	⋮	⋮	⋮	⋮
$X_{n.}$	n_{n1}	n_{n1}	...	n_{nm}	$n_{n.}$
$\sum$	$n_{.1}$	$n_{.12}$	...	$n_{.m}$	N

مثال: يمثل الجدول المتقاطع بين فئات المهن والجنس:

الجدول 02: الجدول المزدوج (المهنة والجنس)

	المهنة 01	المهنة 02	$\sum$
ذكر	20	15	35
أنثى	30	25	55
$\sum$	50	40	90

2.2. مصفوفة الاحتمالات: تمثل التكرارات النسبية وتحسب عن طريق قسمة التكرارات المطلقة على إجمالي التكرارات وفقا للعلاقة التالية:

$$f_{ij} = \frac{n_{ij}}{N}$$

الجدول 03: الجدول المزدوج النسبي

	Y_1	Y_2	...	Y_m	$\sum$
X_1	f_{11}	f_{12}	...	f_{1m}	R_1
X_2	f_{21}	ff_{22}	...	f_{2m}	R_2
⋮	⋮	⋮	⋮	⋮	⋮
X_n	f_{n1}	f_{n1}	...	f_{nm}	R_n
$\sum$	C_1	C_2	...	C_m	1

الجدول 04: الجدول المزدوج النسبي (المهنة والجنس)

	المهنة 01	المهنة 02	$\sum$
ذكر	$\frac{20}{90} = 0.22$	$\frac{15}{90} = 0.17$	0.39
أنثى	$\frac{30}{90} = 0.33$	$\frac{25}{90} = 0.28$	0.61
$\sum$	0.55	0.45	1

وعليه يمكن كتابة مصفوفة الاحتمالات أو ما تعرف بمصفوفة القيم الفعلية كما يلي:

$$F = \begin{pmatrix} 0.22 & 0.17 \\ 0.33 & 0.28 \end{pmatrix}$$

3.2. مصفوفة القيم المتوقعة: تحسب انطلاقا من التوزيع الهامشي للصفوف والأعمدة من خلال تطبيق العلاقة التالية:

$$E_{ij} = R_i \times C_j$$

حيث: $i = \overline{1,n}\,, \quad j = \overline{1,m}$

$$E = \begin{pmatrix} 0.39 \times 0.55 & 0.39 \times 0.45 \\ 0.61 \times 0.55 & 0.61 \times 0.45 \end{pmatrix} = \begin{pmatrix} 0.21 & 0.18 \\ 0.34 & 0.27 \end{pmatrix}$$

4.2. مصفوفة التباعد: تمثل الفرق بين القيم الفعلية والقيم المتوقعة:

$$D_{ij} = F_{ij} - E_{ij}$$

$$D = \begin{pmatrix} 0.22 & 0.17 \\ 0.33 & 0.28 \end{pmatrix} - \begin{pmatrix} 0.21 & 0.18 \\ 0.34 & 0.27 \end{pmatrix} = \begin{pmatrix} 0.01 & -0.01 \\ -0.01 & 0.01 \end{pmatrix}$$

5.2. مصفوفة التباينات والتباينات المشتركة:

$$\varepsilon = D^T . D$$

$$\boldsymbol{\varepsilon} = \begin{pmatrix} 0.01 & -0.01 \\ -0.01 & 0.01 \end{pmatrix} \times \begin{pmatrix} 0.01 & -0.01 \\ -0.01 & 0.01 \end{pmatrix} = \begin{pmatrix} 0.0002 & -0.0002 \\ -0.0002 & 0.0002 \end{pmatrix}$$

6.2. حساب القيم الذاتية $\boldsymbol{det(\varepsilon - \lambda I) = 0}$ (انظر المحور السابق التحليل بالمركبات الرئيسية).

7.2. حساب الأشعة الذاتية $\boldsymbol{(\varepsilon - \lambda I)\vec{u} = 0}$ (انظر المحور السابق التحليل بالمركبات الرئيسية).

من أجل تحليل العلاقة بين المتغيرين نلجأ إلى **اختبار كاي تربيع**، الذي يقوم على الفرضية الصفرية:

المتغيرين مستقلين : $\boldsymbol{H_0}$

حيث ترفض الفرضية الصفرية، إذا كانت إحصائية كاي تربيع المحسوبة أكبر من إحصائية كاي تربيع الحرجة:

$$\boldsymbol{\chi^2_{Calculated} > \chi^2_{Critical}}$$

تحسب إحصائية كاي تربيع بالعلاقة التالية:

$$\chi^2_{ij} = \sum_i \sum_j \frac{\left(F_{ij} - E_{ij}\right)^2}{E_{ij}}$$

$$\boldsymbol{\chi} = \begin{pmatrix} \dfrac{(0.22 - 0.21)^2}{0.21} & \dfrac{(0.17 - 0.18)^2}{0.18} \\ \dfrac{(0.33 - 0.34)^2}{0.34} & \dfrac{(0.28 - 0.27)^2}{0.27} \end{pmatrix} = \begin{pmatrix} 0.0005 & 0.0006 \\ 0.0003 & 0.0002 \end{pmatrix}$$

إحصائية كاي تربيع المحسوبة:

$$\boldsymbol{\chi^2 = 0.0005 + 0.0006 + 0.0003 + 0.0002 = 0.0016}$$

إحصائية كاي تربيع الحرجة:

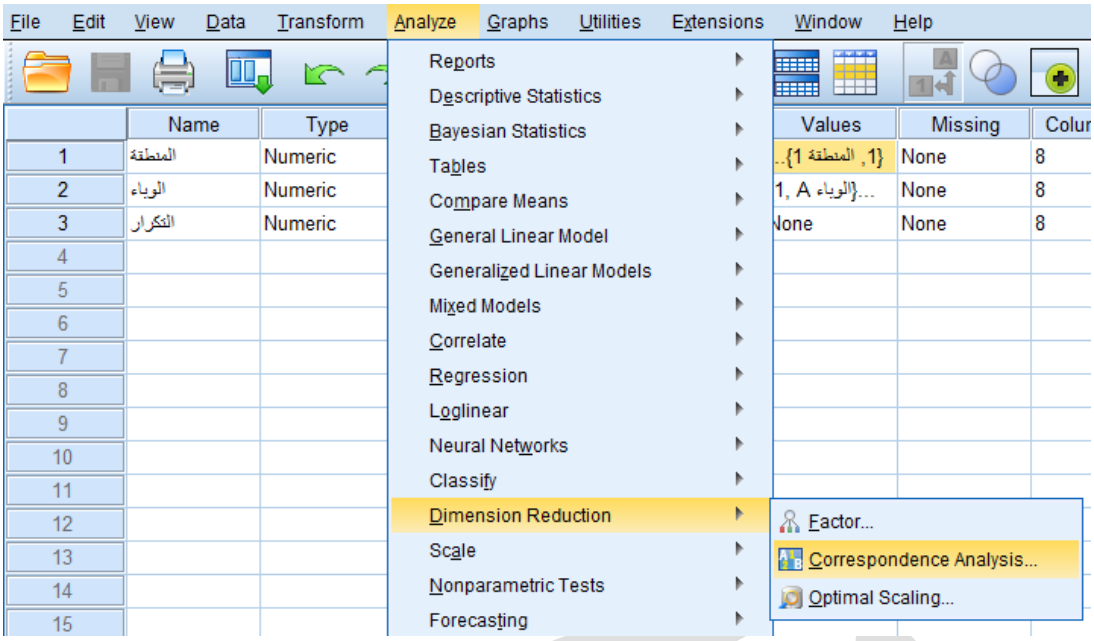
$$\boldsymbol{\chi^2_{((2-1)\times(2-1),\ 0.05)} = \chi^2_{(1,\ 0.05)} = 3.841}$$

بما أن إحصائية كاي تربيع المحسوبة (0.0016) أقل من إحصائية كاي تربيع الحرجة (3.841)، وعليه لا نرفض الفرضية الصفرية، أي أن المتغيرين مستقلين لا يوجد علاقة بين المهنة والجنس.

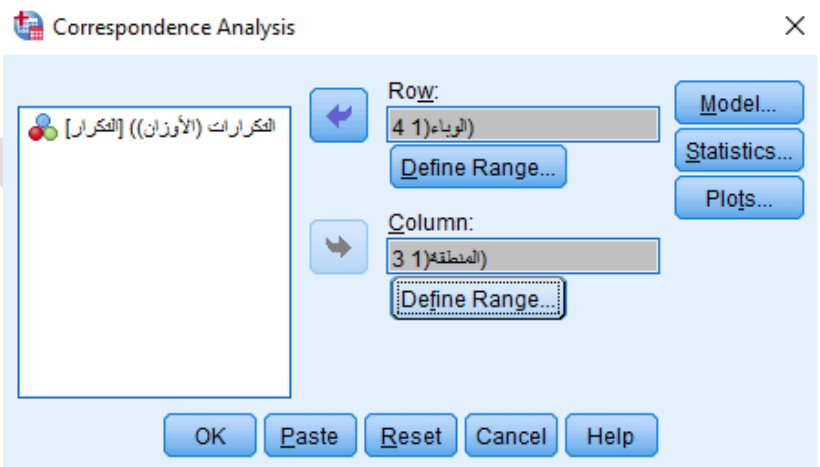
3 تطبيق عملي للتحليل العاملي للتوفيقات في *SPSS*

من أجل تطبيق التحليل العاملي للتوفيقات باستخدام البرنامج الاحصائي *SPSS*، يتم اتباع الخطوات التالية:

Analyze → Dimension Reduction → Correspondence Analysis



في النافذة التي تظهر يتم اختيار متغير في السطر *Row* ومتغير في العمود *Column* مع تحديد مدى أو مجال كل متغير *Define Range*.



بالضغط على *OK*، تظهر مخرجات التحليل أدناه انطلاقا من الجدول المتقاطع *Correspondence Table*:

الجدول 05: الجدول المزدوج (المناطق السكنية وأنواع الأوبئة)

	المناطق السكنية			
أنواع الأوبئة	المنطقة 1	المنطقة 2	المنطقة 3	Active Margin
الوباء A	5	5	30	40
الوباء B	5	25	5	35
الوباء C	15	5	5	25
الوباء D	15	5	0	20
Active Margin	40	40	40	120

يظهر من خلال الجدول 05 أن الوباء A منتشر بكثرة في المنطقة 3 والوباء B منتشر بكثرة في المنطقة 2 والوباء C والوباء D منتشرين بكثرة في المنطقة 1.

الجدول 06: نتائج تحليل AFC (القيم الذاتية، نسب التباين المفسر واختبار كاي تربيع)

				Proportion of Inertia		Confidence Singular Value	
						Standard	Correlation
Dimension	Inertia	Chi Square	Sig.	Accounted for	Cumulative	Deviation	2
1	.410			.618	.618	.069	.180
2	.253			.382	1.000	.086	
Total	.663	79.607	.000[a]	1.000	1.000		

قدرت القيمة الذاتية للمحور الأول بـ 0.41 والقيمة الذاتية للمحور الثاني بـ 0.253، أي أن المركبة الرئيسية الأولى تفسر 61.8% من التباين الإجمالي في البيانات أما المركبة الثانية فتفسر 38.2% من التباين الإجمالي في البيانات.

قدرت إحصائية كاي تربيع $(Chi\ Square)$ 79.607 باحتمالية $Sig = 0.000$، وهذه الأخيرة أصغر من مستوى المعنوية $\alpha = 0.05$، وعليه نرفض الفرضية الصفرية أي أن المناطق السكنية وأنواع الأوبئة مرتبطين (توجد علاقة بينهما)، بمعنى أن كل وباء ينتشر في منطقة معينة بخلاف المناطق الأخرى.

الجدول 07: مساهمة أنواع الأوبئة في تشكيل المحاور

		Score in Dimension			Contribution				
					Of Point to Inertia of Dimension		Of Dimension to Inertia of Point		
أنواع الأوبئة	Mass	1	2	Inertia	1	2	1	2	Total
الوباء A	.333	1.097	.148	.260	.626	.015	.986	.014	1.000
الوباء B	.292	-.397	-1.047	.190	.072	.636	.155	.845	1.000
الوباء C	.208	-.424	.638	.067	.058	.169	.359	.641	1.000
الوباء D	.167	-.968	.739	.146	.244	.181	.686	.314	1.000
Active Total	1.000			.663	1.000	1.000			

- تمثل $Mass$ نسب انتشار الأوبئة.

- تمثل $Score\ in\ Dimension$ إحداثيات الفئات (الوباء A، B، C وD) على المحورين 01 و02.
- تمثل $Inertia$ القيم الذاتية وهي تعبر عن مقدار التباين الذي يفسره كل وباء.
- تمثل $Contribution$ نسب مساهمة الأوبئة في تشكيل المحورين $Of\ Point\ to\ Inertia\ of\ Dimension$ ونسب تمثيل الأوبئة على المحورين $Of\ Dimension\ to\ Inertia\ of\ Point$.

يتضح من خلال الجدول 07 ما يلي:

- نسبة انتشار الأوبئة A، B، C وD هي 33.3%، 29.2%، 20.8% و16.7% على التوالي.
- الوباء A يفسر أكبر نسبة من التباين الإجمالي قدرت بـ $\frac{0.26}{0.663} \times 100 = 39.22\%$.
- الوباء C يفسر أصغر نسبة من التباين الإجمالي قدرت بـ $\frac{0.067}{0.663} \times 100 = 10.11\%$.
- الوباء A يساهم بشكل كبير في تشكيل المحور الأول، ويساهم بشكل ضعيف جدا في تشكيل المحور الثاني.
- الوباء A ممثل جيدا على المحور الأول على عكس تمثيله على المحور الثاني.
- الوباء B يساهم بشكل كبير في تشكيل المحور الثاني، ويساهم بشكل ضعيف جدا في تشكيل المحور الأول.
- الوباء B ممثل جيدا على المحور الثاني على عكس تمثيله على المحور الأول.
- الوبائين C وD يساهمان بشكل ضعيف في تشكيل كلا المحورين.

الجدول 08: مساهمة المناطق السكنية في تشكيل المحاور

		Score in Dimension			Contribution				
					Of Point to Inertia of Dimension		Of Dimension to Inertia of Point		
المناطق السكنية	Mass	1	2	Inertia	1	2	1	2	Total
المنطقة 1	.333	-.678	.803	.206	.240	.427	.476	.524	1.000
المنطقة 2	.333	-.445	-.922	.185	.103	.563	.229	.771	1.000
المنطقة 3	.333	1.124	.119	.272	.657	.009	.991	.009	1.000
Active Total	1.000			.663	1.000	1.000			

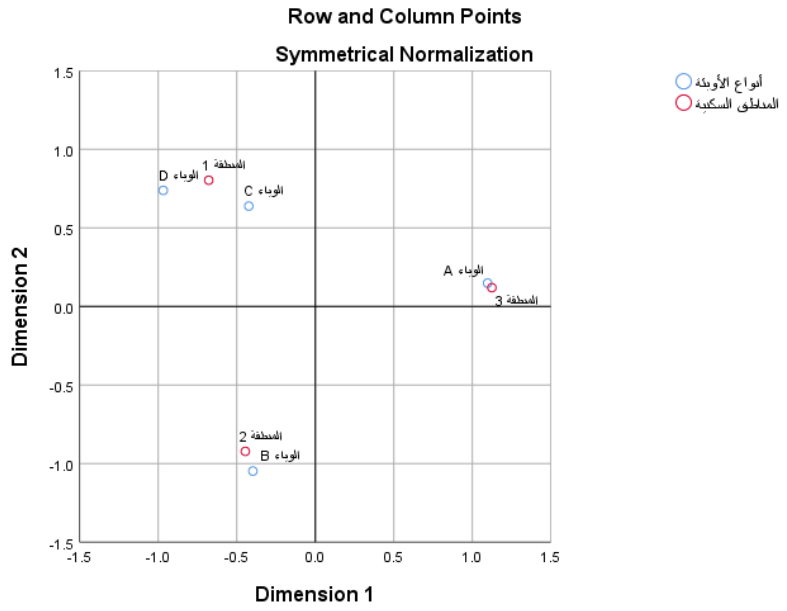
- تمثل $Mass$ التكرار النسبي للمناطق السكنية.
- تمثل $Score\ in\ Dimension$ إحداثيات الفئات (المناطق السكنية 1، 2 و3) على المحورين 01 و02.
- تمثل $Inertia$ القيم الذاتية وهي تعبر عن مقدار التباين الذي تفسره كل منطقة.

- تمثل $Contribution$ نسب مساهمة المناطق السكنية في تشكيل المحورين وكذلك نسب تمثيلها على المحورين $Of\ Point\ to\ Inertia\ of\ Dimension$ $Of\ Dimension\ to\ Inertia\ of\ Point$.

يتضح من خلال الجدول 07 ما يلي:

- تساوي نسب المناطق السكنية الثلاثة أي تساوي عدد أ أفراد العينة للمناطق الثلاثة.
- المنطقة 3 تفسر أكبر نسبة من التباين الإجمالي قدرت بـ $\frac{0.272}{0.663} \times 100 = 41.03\%$.
- المنطقة 2 تفسر أصغر نسبة من التباين الإجمالي قدرت بـ $\frac{0.185}{0.663} \times 100 = 27.9\%$.
- المنطقة 3 تساهم بشكل كبير في تشكيل المحور الأول، وتساهم بشكل ضعيف جدا في تشكيل المحور الثاني.
- المنطقة 3 ممثلة بشكل جيدا على المحور الأول على عكس تمثيلها على المحور الثاني.
- المنطقة 2 تساهم بشكل كبير في تشكيل المحور الثاني، وتساهم بشكل ضعيف في تشكيل المحور الأول.
- المنطقة 2 ممثلة جيدا على المحور الثاني على عكس تمثيلها على المحور الأول.

الشكل 01: تمثيل الأوبئة والمناطق السكنية على المحورين



يظهر من خلال الشكل أعلاه أن:

- الوباء A ينتشر بكثرة في المنطقة 3.
- الوباء B ينتشر بكثرة في المنطقة 2.
- المنطقة 1 ينتشر فيها الوباء C و D.