

ANÁLISIS FACTORIAL DE CORRESPONDENCIAS

MASTER EN ANÁLISIS AVANZADO DE DATOS MULTIVARIANTES Y BIG DATA

Partimos de los datos de una encuesta realizada a 1000 seguidores de fútbol del Barça, del Real Madrid, del Atlético de Madrid, del Athletic de Bilbao y del Valencia. Se les pidió que eligieran de entre esos cinco equipos dos preferentes. Uno, evidentemente será el propio equipo del que es seguidor pero el segundo debe ser el que más simpático les caiga o por el que menos antipatía sientan. Esto nos permitirá analizar la posición relativa de los seguidores de esos equipos respecto a esos equipos.

Los resultados obtenidos se presentan en la tabla siguiente:

SEGUIDORES	PREFERENCIAS				
	Barça	Real Madrid	Atlético de Madrid	Athletic de Bilbao	Valencia
Barça	100	0	30	60	10
Real Madrid	0	100	0	20	80
Atlético de Madrid	50	0	100	40	10
Athletic de Bilbao	70	5	5	100	20
Valencia	20	20	20	40	100

Se pide:

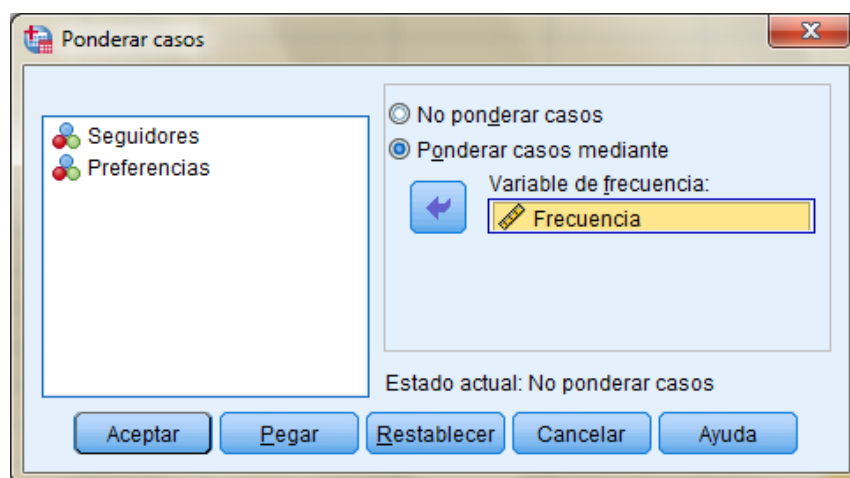
1- Estudiar las posibles relaciones entre las categorías fila y columna.

Para ello:

- *Comience explicando el modelo elegido y porqué*

Elegimos la normalización (modelo) SIMÉTRICO ya que lo que se nos pide es describir los patrones de asociación entre las dos variables fila y columna.

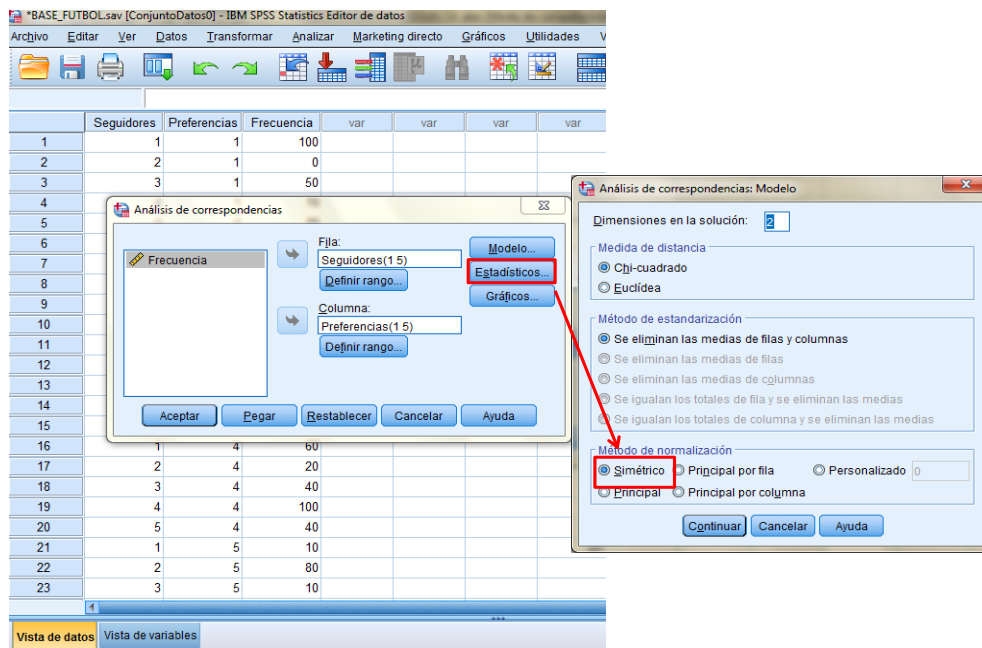
Datos → Ponderar casos → Ponderar casos mediante FRECUENCIA



ANÁLISIS FACTORIAL DE CORRESPONDENCIAS

MASTER EN ANÁLISIS AVANZADO DE DATOS MULTIVARIANTES Y BIG DATA

Analizar → Reducción de dimensiones → Análisis de correspondencias



- *Calcule la Inercia absorbida por cada eje factorial y decida si el primer plano factorial es suficiente para representar las nubes de puntos fila y columna. Razone su respuesta.*

Resumen

Dimensión	Valor singular	Inercia	Chi cuadrado	Sig.	Proporción de inercia		Valor singular de confianza	
					Contabilizado para	Acumulado	Desviación estándar	Correlación
								2
1	,760	,577			,663	,663	,017	,152
2	,437	,191			,220	,883	,033	
3	,287	,082			,095	,978		
4	,139	,019			,022	1,000		
Total		,870	870,289	,000 ^a	1,000	1,000		

a. 16 grados de libertad

Elegimos los dos primeros ejes factoriales ya que la absorción de inercia acumulada supera el 88%.

ANÁLISIS FACTORIAL DE CORRESPONDENCIAS

MASTER EN ANÁLISIS AVANZADO DE DATOS MULTIVARIANTES Y BIG DATA

- Estudie las contribuciones de las filas y de las columnas y díganos cuáles de las categorías columna (preferencias) reciben mayores contribuciones de eje 1 y cuáles de eje 2.

The screenshot shows the IBM SPSS Statistics interface. The main data table has columns: Seguidores, Preferencias, Frecuencia, and several empty columns labeled 'var'. The 'Análisis de correspondencias' dialog box is open, with 'Fila' set to 'Seguidores(1 5)' and 'Columna' set to 'Preferencias(1 5)'. The 'Estadísticos' sub-dialog box is also open, showing options for 'Tabla de correspondencias', 'Inspección de los puntos de fila', and 'Inspección de los puntos de columna', all of which are checked. A text box with an arrow points to the 'Inspección de los puntos de fila' checkbox, stating: 'Inspección de los puntos de fila y de columna nos permite obtener las contribuciones de fila y columna'.

Puntos de fila generales^a

		Puntuación en			Contribución				
		dimensión			Del punto en la inercia de		De la dimensión en la inercia del		
					dimensión		punto		
Seguidores	Masa	1	2	Inercia	1	2	1	2	Total
Barça	,200	-,725	,402	,109	,138	,074	,734	,130	,864
Real Madrid	,200	1,483	-,121	,353	,579	,007	,946	,004	,950
Atlético de Madrid	,200	-,810	-1,099	,208	,173	,553	,480	,509	,989
Athletic de Bilbao	,200	-,430	,892	,105	,049	,364	,267	,661	,928
Valencia	,200	,483	-,074	,095	,061	,002	,371	,005	,376
Total activo	1,000			,870	1,000	1,000			

a. Normalización simétrica

ANÁLISIS FACTORIAL DE CORRESPONDENCIAS

MASTER EN ANÁLISIS AVANZADO DE DATOS MULTIVARIANTES Y BIG DATA

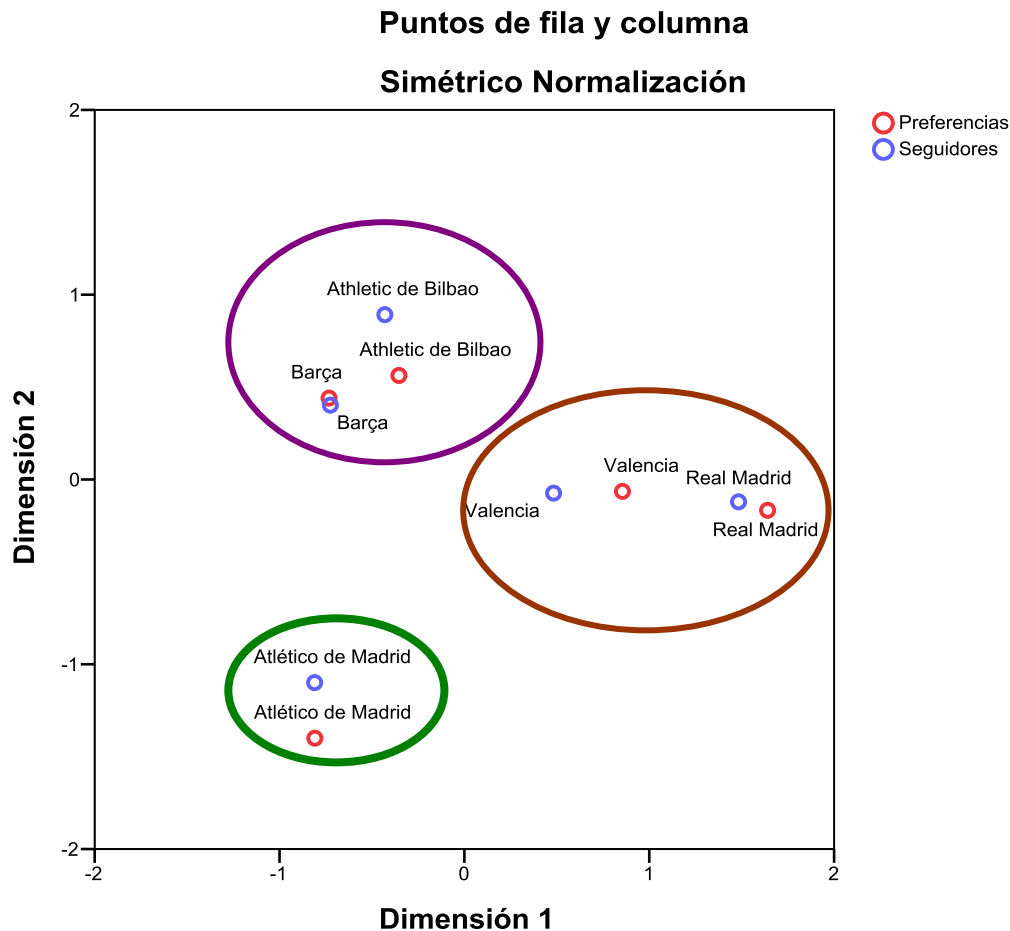
Puntos de columna generales^a

		Puntuación en dimensión			Contribución				
					Del punto en la inercia de dimensión		De la dimensión en la inercia del punto		
					1	2	1	2	Total
Preferencias	Masa	1	2	Inercia	1	2	1	2	Total
Barça	,240	-,732	,440	,131	,169	,106	,747	,156	,902
Real Madrid	,125	1,640	-,167	,292	,443	,008	,875	,005	,880
Atlético de Madrid	,155	-,809	-1,400	,210	,133	,695	,366	,632	,998
Athletic de Bilbao	,260	-,354	,563	,071	,043	,188	,350	,509	,859
Valencia	,220	,855	-,064	,166	,212	,002	,735	,002	,737
Total activo	1,000			,870	1,000	1,000			

a. Normalización simétrica

- *Analice el gráfico factorial y saque conclusiones*

The screenshot displays the IBM SPSS Statistics interface. In the background, a data table is visible with columns: Seguidores, Preferencias, Frecuencia, and several empty 'var' columns. The data rows show values for Seguidores (1, 2, 3) and Preferencias (1, 2, 3) with corresponding Frecuencia values (100, 0, 50). Overlaid on this is the 'Análisis de correspondencias: Gráficos' dialog box. In this dialog, under the 'Diagramas de dispersión' section, the 'Diagrama de dispersión biespacial' option is checked and highlighted with a red box. Below this, the 'Puntos de fila' and 'Puntos de columna' options are unchecked. The 'Ancho de la etiqueta de identificación para los diagramas de dispersión' is set to 20. In the 'Gráfico de líneas' section, both 'Categorías de fila transformadas' and 'Categorías de columna transformadas' are unchecked, and the 'Ancho de la etiqueta de identificación para los gráficos de líneas' is also set to 20. Under 'Dimensiones del gráfico', the 'Mostrar todas las dimensiones de la solución' option is selected. At the bottom of the dialog, the 'Modelo...' button is highlighted with a red box, and the 'Gráficos' button is also visible. The 'Aceptar', 'Pegar', 'Restablecer', 'Cancelar', and 'Ayuda' buttons are at the bottom of the dialog.



Como puede observarse en el gráfico el punto que hace referencia a los seguidores está muy cerca, en todos los casos, del punto de las preferencias. Además observamos:

*Los seguidores del Barça eligen con frecuencia al Atlético de Bilbao y viceversa. Son equipos próximos.

*Tanto el Barça como el Atlético de Madrid respecto al Real Madrid están muy alejados.

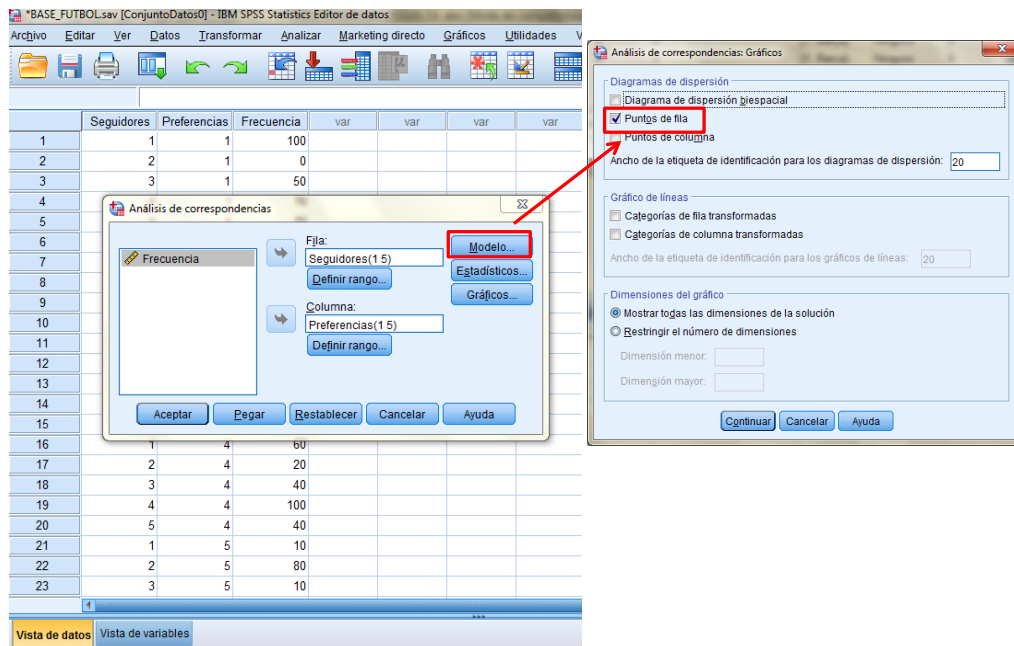
*El Valencia está más próximo al Real Madrid pero en realidad está bastante en el centro del gráfico, lo que significa que las distancias con todos los equipos está bastante repartida.

ANÁLISIS FACTORIAL DE CORRESPONDENCIAS

MASTER EN ANÁLISIS AVANZADO DE DATOS MULTIVARIANTES Y BIG DATA

- 2- Explorar la similitud/disimilitud entre los *seguidores* para averiguar si sería adecuado agrupar algunos de ellos y así simplificar el análisis.

Ahora se repetiría el estudio pero aplicando en vez del modelo simétrico el modelo Principal por filas (porque “seguidores” está en filas) ya que es con el que garantizamos que la distancia Chi cuadrado del hiperespacio esté bien proyectada en el plano principal que estamos estudiando.



Puntos de fila para Seguidores

