

PROXIMIDADES

Las medidas de proximidad sirven para medir el grado de semejanza o de diferencia existente entre dos variables o dos casos de un conjunto de datos. Así, las medidas de similitud toman valores tanto mayores cuanto más se parezcan los elementos comparados, y las medidas de disimilitud toman valores más grandes, cuanto más distintos sean los elementos comparados.

SPSS calcula las medidas de similitud y de disimilitud directamente, mediante el procedimiento **Distancias**. Para ejecutar dicho procedimiento, debemos seleccionarlo en el submenú de **Correlaciones** dentro del menú **Analizar**, como se indica en la Figura 1 (para que podamos ejecutar este procedimiento, es necesario tener abierto un fichero de datos en el editor de datos de SPSS, para lo cual abriremos previamente el fichero **Empresas_en_Espana.sav**).

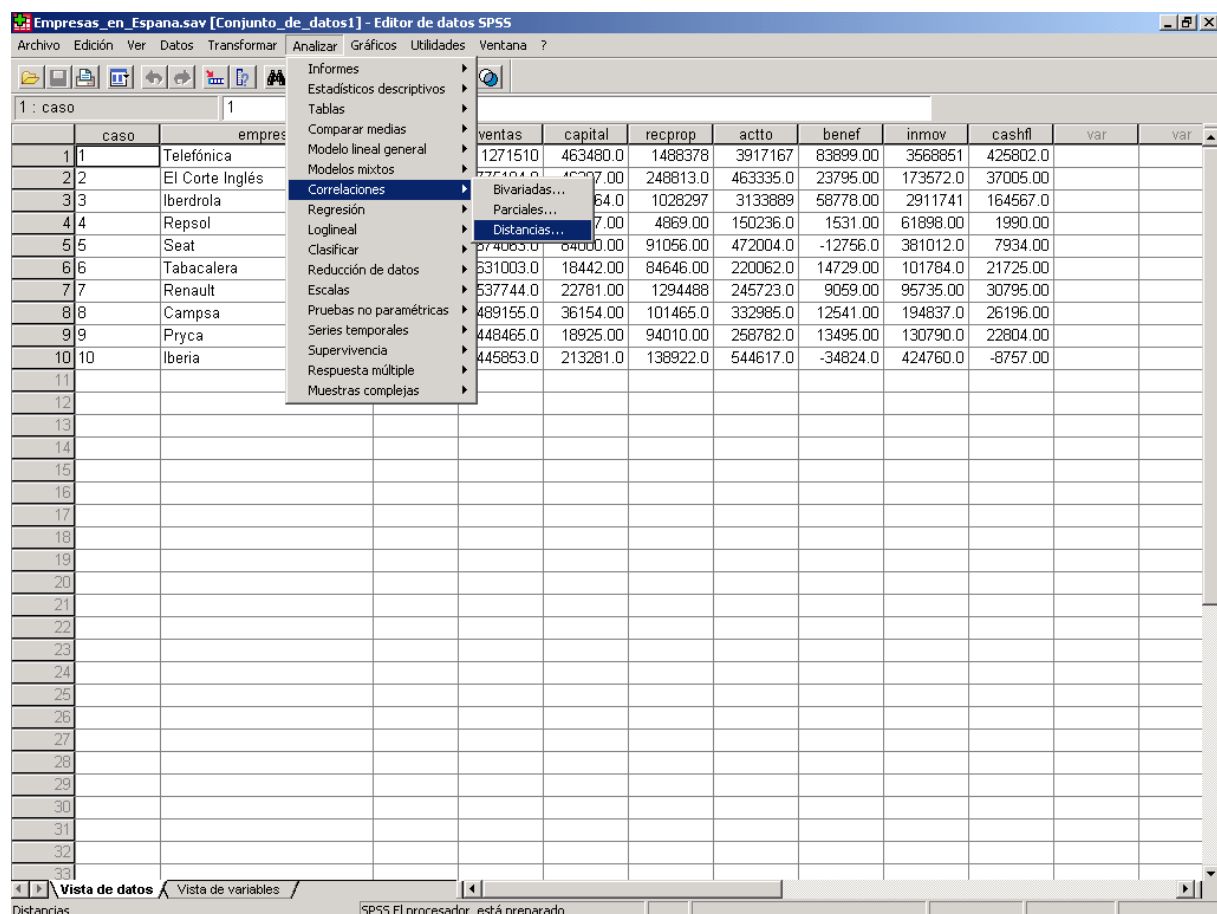


Figura 1: Selección del procedimiento **Distancias**.

Si se selecciona este procedimiento, aparece el siguiente cuadro de diálogo:

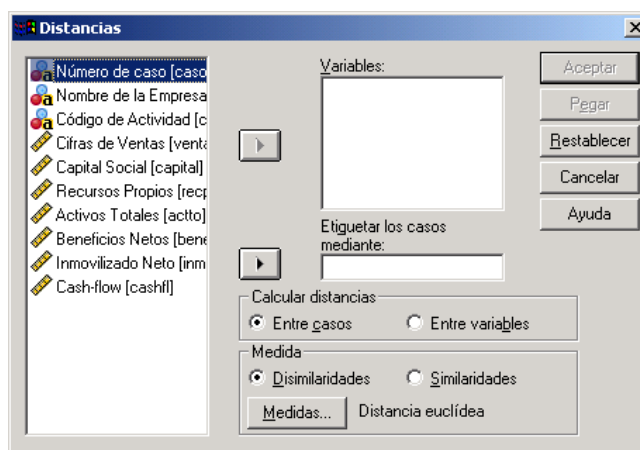


Figura 2: Cuadro de diálogo del procedimiento **Distancias**.

Dentro de este procedimiento se aprecia la posibilidad de calcular medidas de similitud o disimilitud para un determinado conjunto de variables o casos seleccionados. Por ejemplo, para las medidas de disimilitud en variables con escala de intervalo, las opciones son las siguientes:

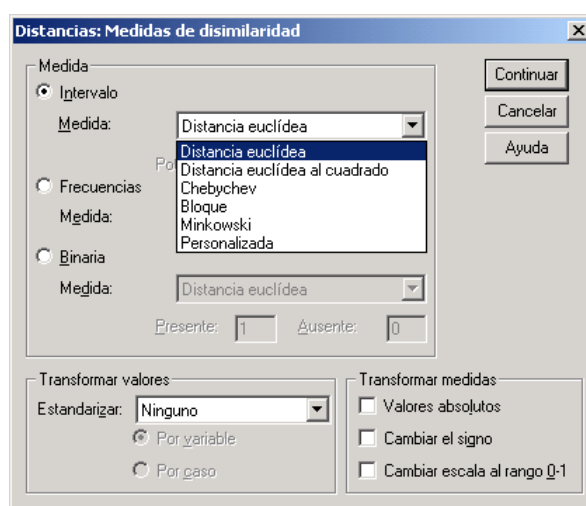


Figura 3: Opciones de las medidas de disimilaridad.

Las opciones para las medidas de similitud se pueden seleccionar dentro del siguiente cuadro de diálogo:

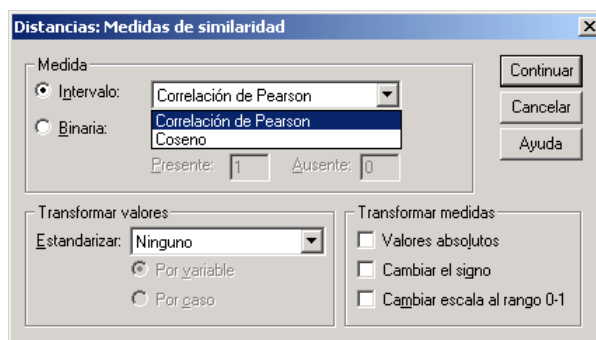


Figura 4: Opciones de las medidas de similitud.

Se aprecia que, en todos los casos, se puede realizar la transformación previa de los valores de cara a su homogeneización. Si, por ejemplo quisiéramos calcular las distancias euclídeas entre las variables del fichero **Empresas_en_Espana.sav**, sin realizar ninguna transformación de los valores de entrada, los resultados serían los siguientes:

Distancias

Resumen de procesamiento de los casos

Casos					
Valid		Perdidos		Total	
N	Porcentaje	N	Porcentaje	N	Porcentaje
10	100.0%	0	.0%	10	100.0%

Matriz de distancias

	Distancia euclídea						
	Cifras de Ventas	Capital Social	Recursos Propios	Activos Totales	Beneficios Netos	Inmovilizado Neto	Cash-flow
Cifras de Ventas	.000	1793716.149	1566468.3	3650812	2180781.624	3370765.495	1948231
Capital Social	1793716	.000	1754291.1	4459826	599810.607	3995928.478	362366.3
Recursos Propios	1566468	1754291.083	.000	3450854	2162346.744	3081849.141	1886876
Activos Totales	3650812	4459826.047	3450853.7	.000	5019693.766	597526.669	4686913
Beneficios Netos	2180782	599810.607	2162346.7	5019694	.000	4553978.216	360784.9
Inmovilizado Neto	3370765	3995928.478	3081849.1	597526.7	4553978.216	.000	4222063
Cash-flow	1948231	362366.349	1886876.1	4686913	360784.932	4222062.661	.000

Esta es una matriz de disimilaridades

Si ahora hiciéramos lo mismo, pero calculando las distancias entre los casos, los resultados serían los siguientes:

Distancias

Resumen de procesamiento de los casos

Casos					
Valid		Perdidos		Total	
N	Porcentaje	N	Porcentaje	N	Porcentaje
10	100.0%	0	.0%	10	100.0%

Matriz de distancias

	Distancia euclídea									
	1	2	3	4	5	6	7	8	9	10
1	.000	5056551	1254131	5423378	4966777	5332539	5145278	5207045	5309492	4901804
2	5056551	.000	3925262	423172.5	286028.8	336359.5	1097267	348232.9	418785.5	472893.7
3	1254131	3925262	.000	4277271	3812287	4182766	4073212	4043500	4144514	3725316
4	5423378	423172.5	4277271	.000	469340.7	136087.8	1304334	327190.2	298415.6	644831.8
5	4966777	286028.8	3812287	469340.7	.000	385463.1	1266556	302571.3	405120.8	281134.6
6	5332539	336359.5	4182766	136087.8	385463.1	.000	1213772	205311.2	189080.2	536863.8
7	5145278	1097267	4073212	1304334	1266556	1213772	.000	1201380	1204415	1257439
8	5207045	348232.9	4043500	327190.2	302571.3	205311.2	1201380	.000	107836.0	368469.7
9	5309492	418785.5	4144514	298415.6	405120.8	189080.2	1204415	107836.0	.000	459618.3
10	4901804	472893.7	3725316	644831.8	281134.6	536863.8	1257439	368469.7	459618.3	.000

Esta es una matriz de disimilaridades

SPSS tiene otras posibilidades para determinar el grado de asociación entre dos variables. Estas opciones se encuentran en el procedimiento **Bivariadas...** dentro del submenú **Correlaciones** del menú **Analizar**. Para ilustrarlo, vamos a calcular el coeficiente de correlación por rangos de Spearman (este ejemplo se podría haber desarrollado en el tema de Tablas de contingencia, pues sirve para medir el grado de asociación entre dos variables de tipo ordinal).

COEFICIENTE DE CORRELACIÓN POR RANGOS DE SPEARMAN

Para efectuar el cálculo del coeficiente de correlación por rangos de Spearman (ρ_s) debemos recuperar el fichero **Pruebas_habilidad.sav**, pues las variables que contiene representan la posición obtenida en distintas pruebas, siendo por tanto de tipo ordinal.

Ejecutamos el procedimiento **Analizar>Correlaciones>Bivariadas...**, según se muestra en la Figura 5:

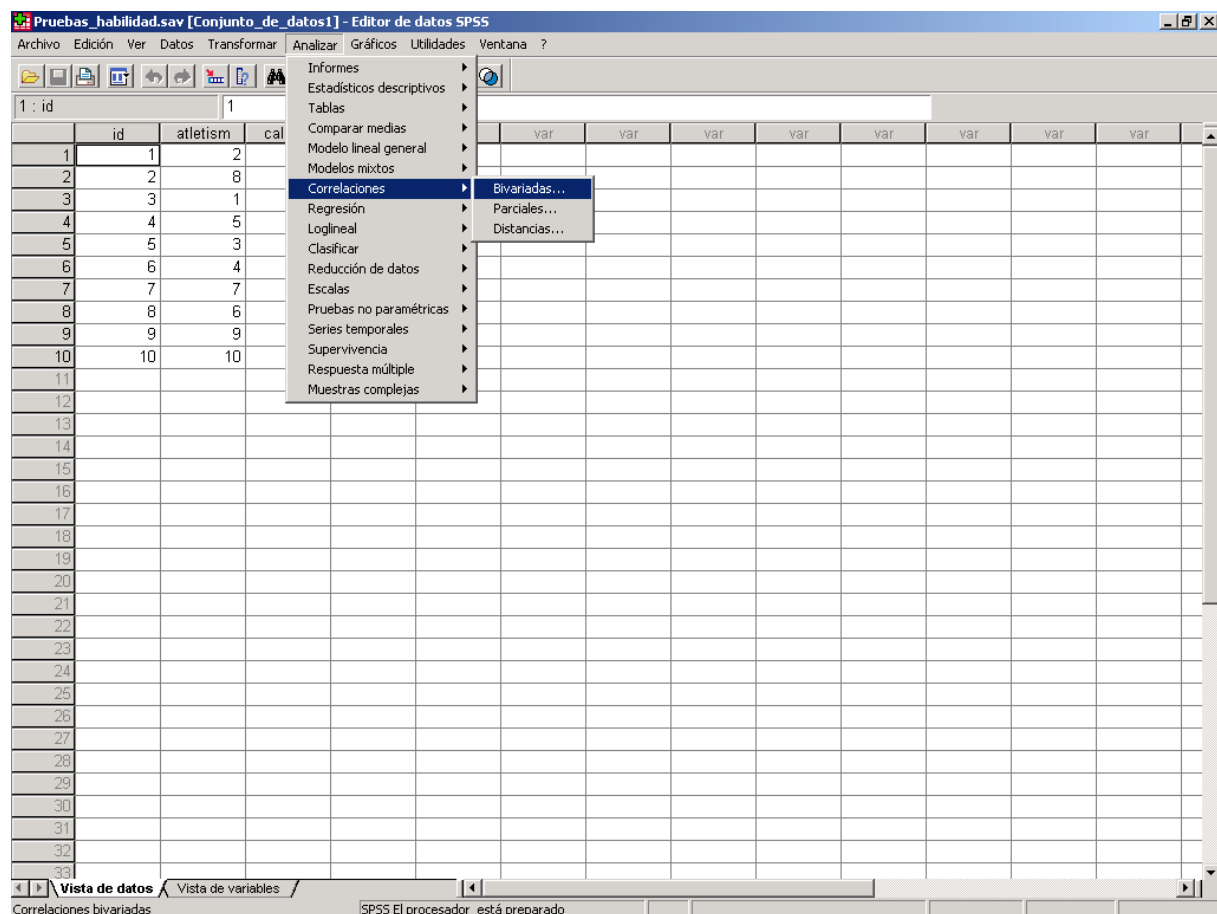


Figura 5: Selección del procedimiento **Correlaciones Bivariadas**.

El cuadro de diálogo de este procedimiento es el que se muestra en la Figura 6 (con las opciones correspondientes ya seleccionadas):

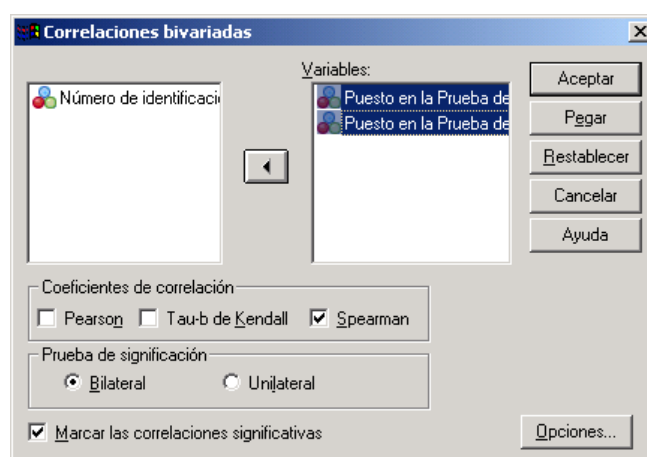


Figura 6: Cuadro de diálogo de **Correlaciones bivariadas**.

Al presionar el botón de Aceptar, el resultado es:

Correlaciones no paramétricas

Correlaciones

			Puesto en la Prueba de Atletismo	Puesto en la Prueba de Cálculo
Rho de Spearman	Puesto en la Prueba de Atletismo	Coefficiente de correlación	1.000	-.939**
		Sig. (bilateral)	.	.000
		N	10	10
	Puesto en la Prueba de Cálculo	Coefficiente de correlación	-.939**	1.000
		Sig. (bilateral)	.000	.
		N	10	10

** . La correlación es significativa al nivel 0,01 (bilateral).

Como se puede comprobar, el coeficiente de correlación por rangos de Spearman, calculado directamente a partir de las observaciones, es:

Atletismo	Cálculo	d_i	d_i^2
2	10	-8	64
8	3	5	25
1	8	-7	49
5	6	-1	1
3	9	-6	36
4	7	-3	9
7	5	2	4
6	4	2	4
9	1	8	64
10	2	8	64

SUMA: 320

La forma concreta de cálculo de este coeficiente es:

$$\rho_s = 1 - 6 \cdot \frac{\sum_{i=1}^{10} d_i^2}{N^3 - N} = 1 - 6 \cdot \frac{320}{1000 - 10} = -0.9394,$$

y, como se observa, es el mismo resultado al que SPSS llega en su procedimiento.