

Facultad de Ciencias Sociales de la Universidad de Chile

Correspondencias Múltiples en SPPS

Estadística IV

Ignacio Díaz e Isabel Garrido

15-5-2015

RESUMEN: CORRESPONDENCIAS MÚLTIPLES

El análisis de correspondencias es una técnica estadística para analizar la relación entre categorías de variables cualitativas. Es una técnica de "reducción de dimensiones en el contexto de tablas de contingencia. El propósito es representar gráficamente la estructura de relaciones de dos o más variables cualitativas mediante mapas de posicionamiento" (Vivanco, 1999:121)¹. El análisis de correspondencias múltiples, a diferencia del análisis de correspondencias simple, trabaja con más de 2 variables, lo que permite incluir variables objetos, que son aquellas donde cada categoría tiene frecuencia igual a 1 (como países o comunas).

Al igual que en el análisis de correspondencias simples, las relaciones entre las categorías de las variables, se analizan a través de los mapas perceptuales, donde cada categoría es ubicada en un punto. La distancia entre las distintas categorías nos permitirá analizar la relación entre ellas.

USOS:

1. Profundizar el análisis de la asociación entre variables cualitativas (nominales u ordinales)
2. Analizar la estructura de la relación entre categorías de variables
3. Identificar dimensiones que representen esquemas conceptuales de análisis
4. Corroborar dependencia entre dos variables cualitativas

CONDICIONES:

- A. **Cantidad de variables:** dos o más.
- B. **Variables admitidas:** nominales y ordinales. Si existen variables cuantitativas, éstas pueden recodificarse a ordinales (por ejemplo, una variable intervalar de ingreso, recodificada en tramos de ingreso).
- C. **Relación entre las variables:** de asociación según el estadígrafo chi². Si no hay relación entre las variables no tiene sentido realizar un análisis de correspondencias.

EJERCICIO DE APLICACIÓN

O. DISPONER LA BASE DE DATOS

La base de datos utilizada en este ejemplo es la Encuesta Nacional de Condiciones de Empleo, Trabajo, Salud y Calidad de Vida de los Trabajadores y Trabajadoras en Chile (ENETS) del año 2011. Se trabajará con 6 variables que pretenden abarcar distintos aspectos de la organización del trabajo en Chile.

La pregunta que guía al ejercicio es **¿Cómo se estructura la organización del trabajo en Chile?**

VARIABLES A UTILIZAR

Se eliminaron las categorías de respuesta NS/NR.

Ingreso: ordinal, cuyas categorías de respuesta son:

¹ Vivanco, M. (1999), Análisis Estadístico Multivariable, Editorial Universitaria, Santiago.

- 1,00 = "Menos de 136.000"
- 2,00 = "137.000 a 250.000"
- 3,00 = "251.000 a 450.000"
- 4,00 = "451.000 a 850.000"
- 5,00 = "851.000 a 1.250.000"
- 6,00 = "1.250.000 y más"

Nivel educacional: ordinal, cuyas categorías de respuesta son:

- 1,00 = "Básica"
- 2,00 = "Media"
- 3,00 = "Técnico-profesional"
- 4,00 = "Universitaria"
- 5,00 = "Postgrado"

Formalidad de contrato: nominal, cuyas categorías de respuesta son:

- 1,00 = "Contrato escrito"
- 2,00 = "Contrato verbal"
- 3,00 = "Sin contrato"

Escala de influencia en la toma de decisiones en el proceso de trabajo: ordinal, cuyas categorías de respuesta son:

- 1,00 = "Nunca decide"
- 2,00 = "Rara vez decide"
- 3,00 = "Algunas veces decide"
- 4,00 = "Casi siempre decide"
- 5,00 = "Siempre decide"

Duración del contrato: nominal, cuyas categorías de respuesta son:

- 1,00 = "Indefinido"
- 2,00 = "A plazo fijo"
- 3,00 = "Por obra"

Posición dentro de la empresa: nominal, cuyas categorías de respuesta son:

- 1,00 = "Directivo/Gerente"
- 2,00 = "Mandos medios"
- 3,00 = "Subordinado"

1. ESTADÍSTICOS DESCRIPTIVOS

Para llevar a cabo el Análisis de Correspondencias todas las categorías de respuesta deben tener al menos 20 casos. Analizamos las frecuencias en Análisis -> Estadísticos descriptivos -> Frecuencias.

Nombre	Tipo	Ar
657	rra58ba_1	N Numérico 8
658	rra58ca_1	N Numérico 8
659	re4	C Cadena 200
660	AñoEnt	N Numérico 11
661	FEXP_MAY15	N Numérico 11
662	Ingreso	N Numérico 8
663	Educ	N Numérico 8
664	Horas	N Numérico 8
665	Contrato	N Numérico 8
666	Duración	N Numérico 8
667	Autonomía	N Numérico 8
668	Posición	N Numérico 8
669	Formalidad	N Numérico 8
670	DecisiónCo...	N Numérico 8
671	EscalaDeci...	N Numérico 8
672	AmbSocial	N Numérico 8
673	AmbFísico	N Numérico 8
674	AmbienteCo...	N Numérico 8
675	EscalaAmbi...	N Numérico 8
676		
677		
678		
679		

EscalaDecisión

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
1,00	1673	17,6	17,6	17,6
2,00	2042	21,5	21,5	39,1
3,00	2038	21,4	21,4	60,5
4,00	1683	17,7	17,7	78,2
5,00	2067	21,8	21,8	100,0
Total	9503	100,0	100,0	

Ingreso

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Menos de 136.000	2047	21,5	22,1	22,1
137.000 a 250.000	4007	42,2	43,3	65,5
251.000 a 450.000	2088	22,0	22,6	88,0
Válidos 451.000 a 850.000	547	5,8	5,9	94,0
851.000 a 1.250.000	259	2,7	2,8	96,8
1.250.000 y más	300	3,2	3,2	100,0
Total	9248	97,3	100,0	
Perdidos Sistema	255	2,7		

Total	9503	100,0	
-------	------	-------	--

Nivel Educativo

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos				
Básica	3042	32,0	32,1	32,1
Media	3307	34,8	34,9	67,0
Técnico-profesional	2168	22,8	22,9	89,9
Universitaria	893	9,4	9,4	99,3
Postgrado	64	,7	,7	100,0
Total	9474	99,7	100,0	
Perdidos Sistema	29	,3		
Total	9503	100,0		

Duración del contrato

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos				
Indefinido	3739	39,3	76,6	76,6
A plazo fijo	639	6,7	13,1	89,7
Por obra	504	5,3	10,3	100,0
Total	4882	51,4	100,0	
Perdidos Sistema	4621	48,6		
Total	9503	100,0		

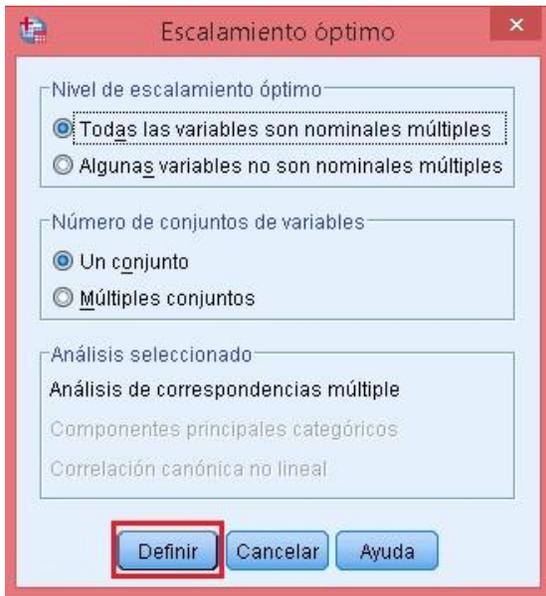
Los resultados muestran que se cumple el supuesto y por tanto se puede continuar con el análisis.

2. ANÁLISIS DE CORRESPONDENCIAS MÚLTIPLES

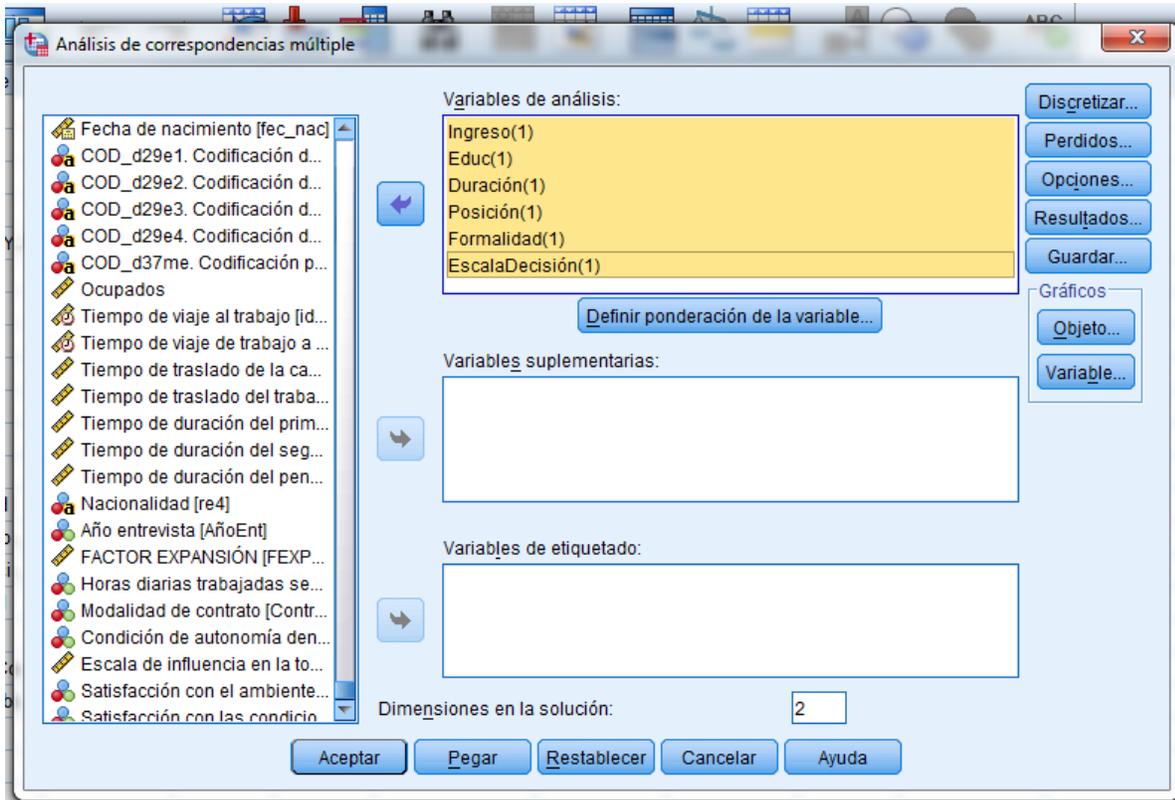
The screenshot shows the IBM SPSS Statistics interface. The 'Análisis' menu is open, and the 'Reducción de dimensiones' option is selected. A sub-menu is displayed, showing 'Análisis de correspondencias...' as the active option. The background shows a list of variables with their names, types, and scales.

Nombre	Tipo	Ar
rra58ba_1	Númérico	8
rra58ca_1	Númérico	8
re4	Cadena	200
AñoEnt	Númérico	11
FEXP_MAY15	Númérico	11
Ingreso	Númérico	8
Educ	Númérico	8
Horas	Númérico	8
Contrato	Númérico	8
Duración	Númérico	8
Autonomía	Númérico	8
Posición	Númérico	8
Formalidad	Númérico	8
DecisiónCo...	Númérico	8
EscalaDeci...	Númérico	8
AmbSocial	Númérico	8
AmbFísico	Númérico	8

Analizar -> Reducción de datos (o dimensiones, según la versión SPSS) -> Escalamiento óptimo...



- Nivel de escalamiento óptimo: Todas las variables son nominales múltiples
- Número de conjuntos de Variables: Un conjunto
- Análisis seleccionado: Análisis de correspondencias múltiple
- **DEFINIR**



En la nueva ventana, tenemos que tomar decisiones referentes al Análisis propiamente tal:

- **Variables de análisis:** Agregamos aquellas variables que nos interesa analizar sus relaciones.

Al agregar las variables de análisis aparece entre paréntesis un valor =1. Este indica el peso – proporcional- asignado a cada variable. Es decir, en nuestro modelo se le otorgará igual peso (importancia) a cada variable. Eventualmente, en “Definir ponderación de la variable” se le pueden asignar distintos pesos a las variables según la importancia para el modelo, pero en este caso, mantendremos todas las variables con peso =1.

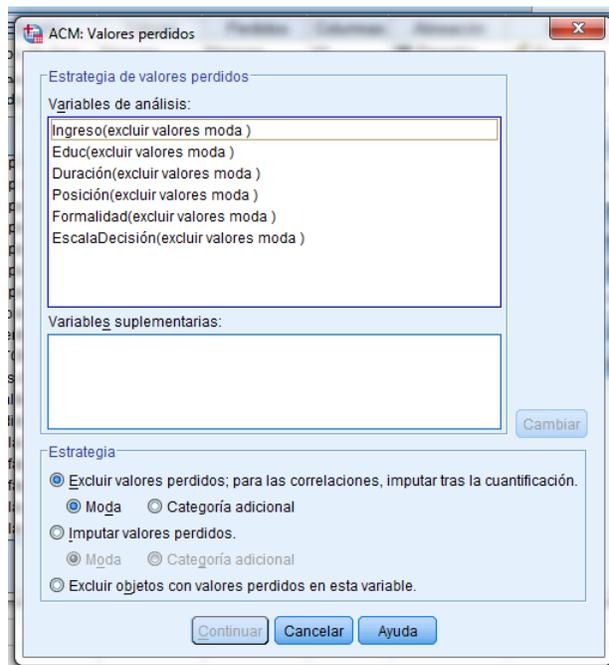
- **Variables suplementarias:** En esta sección se pueden agregar variables, que no serán parte del modelo, pero que nos interesa que sean igualmente ubicadas en el mapa generado por este. En este caso no ingresaremos variables suplementarias.
- **Variables de etiqueta:** Si quisiéramos etiquetar a todos los casos, es decir, si quisiéramos utilizar una variable objeto, la incluimos en este módulo. Sin embargo, esto solo es útil cuando se trabaja con pocos casos.
- **Dimensiones de la solución:** Aquí dejaremos el valor predeterminado 2, ya que la solución entregada en un plano (2 dimensiones) es la más sencilla para interpretar los resultados.

***Discretizar:** En este modulo podemos definir las categorías de cada variable. Esto se usa para transformar variables continuas (cadena), en discretas. En nuestro caso no es necesario.

Perdidos: Aquí se determina qué hacer con los valores perdidos. Hay tres opciones posibles:

1. Excluir casos perdidos: para las correlaciones imputar tras cuantificación: Se trata de excluir a los sujetos sólo en las variables en las que presentan valores perdidos, imputando la moda o un valor de categoría extra para el cálculo de correlaciones entre variables.
2. Imputar valores perdidos: Rellenar los casos perdidos con algún valor (moda u otra categoría).
3. Excluir a los objetos con valores perdidos en esta variable: Dejar fuera los sujetos con valores perdidos.

Dejaremos la opción predeterminada (1).



Opciones

Objetos suplementarios: Se puede asignar una condición de suplementariedad a un individuo u objeto. Con ello este es ignorado en la construcción del modelo. En este caso no se ignorará ningún individuo/objeto.

Método de Normalización:
Principal por variable.

Esta opción optimiza la asociación entre las variables, es útil cuando el interés principal está en la correlación entre las variables. Cuando se utilizan variables objeto, se recomienda utilizar el método Simétrico. Este distribuye la normalización entre variables y objetos.

Criterios: El criterio de convergencia e iteraciones máximas determinan hasta qué punto el programa debe intentar optimizar la solución. Dejamos las opciones predeterminadas.

- Convergencia: ,00001
- Iteraciones máximas: 100

Etiquetar gráficos con: Se puede optar por etiquetar según (1) etiquetas, (2) Nombres. Para el caso de la etiqueta hay que determinar la longitud máxima. Dejamos las opciones predeterminadas.

- Etiquetas de variable o de valor
- Límite de longitud de etiqueta: 20

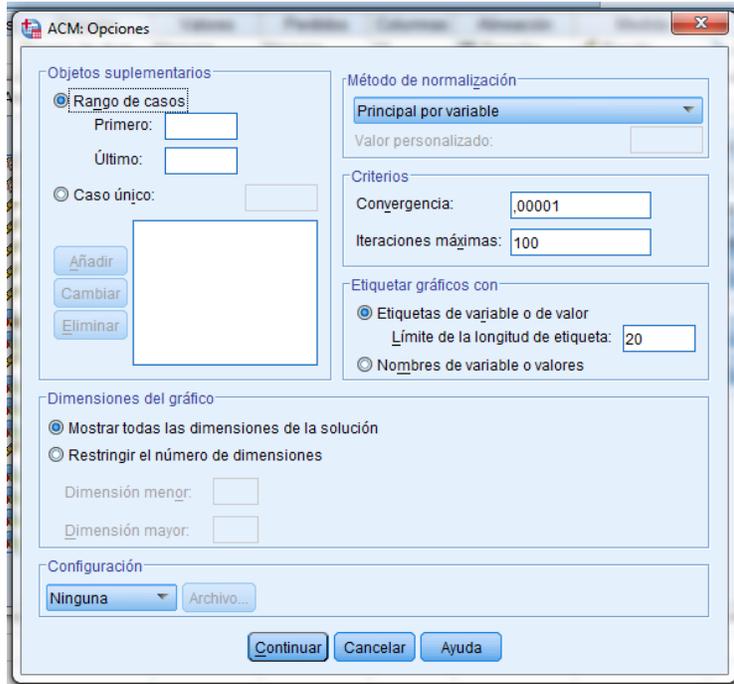
Dimensiones del gráfico: Cuando se trabaja con muchas dimensiones, se puede elegir graficar una menor cantidad de ellas. En este caso como trabajamos con 2, no es necesario. Por tanto, dejamos las opciones predeterminadas.

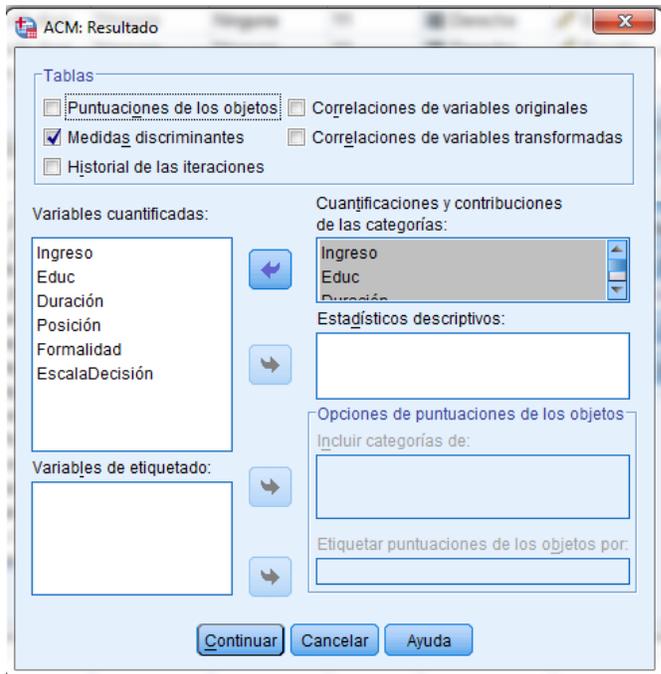
- Mostrar todas las dimensiones de la solución

Resultados

Tablas: La opción más importante son las Medidas Discriminantes, tabla que indica cuánto aporta cada variable a cada dimensión.

Cuantificaciones y contribuciones de las categorías: Indican las coordenadas de posicionamiento en el mapa perceptual de las categorías y su contribución a las dimensiones. Incluimos aquí todas las variables.



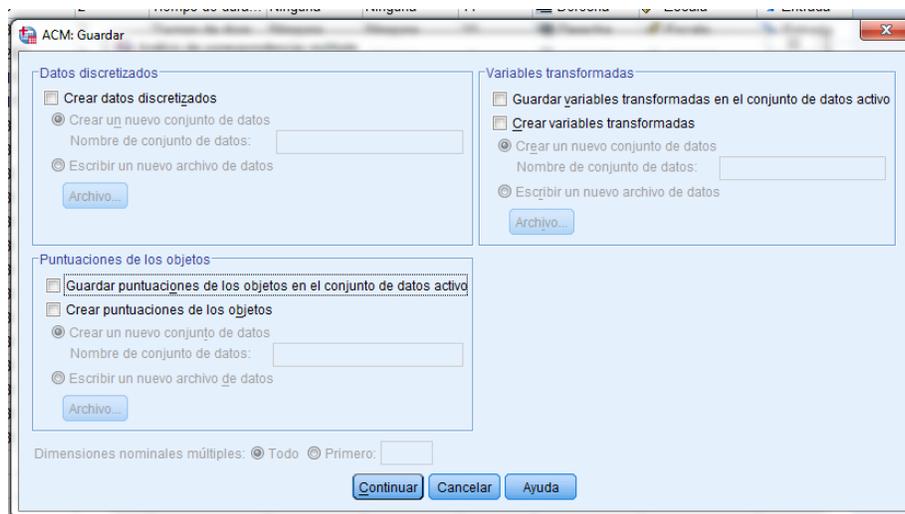


Estadísticos descriptivos: Ya obtuvimos las frecuencias antes.

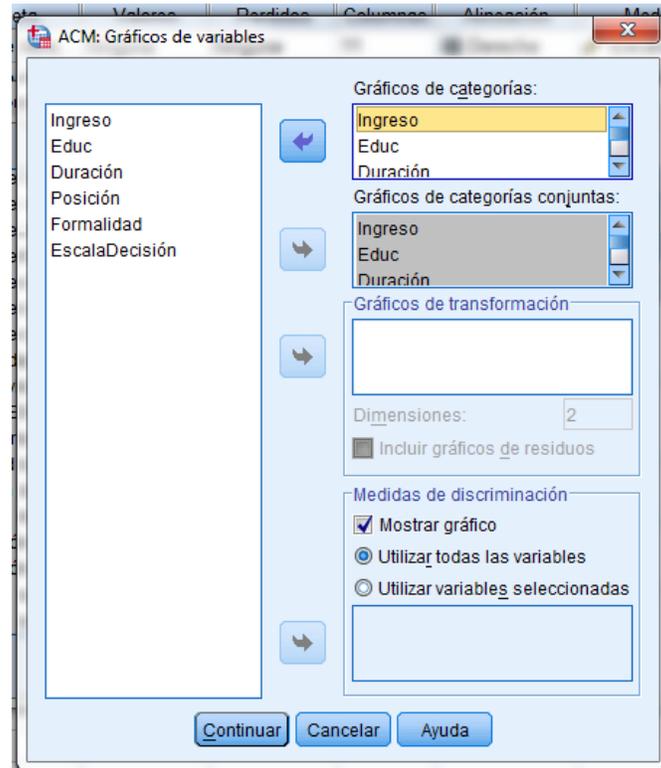
***Guardar:** Esta ventana puede ser útil cuando se quiere guardar información en la base de datos. En este caso no guardamos nada.

***Gráficos:**

Objetos: Se utiliza para graficar las categorías de las variables objetos. En este caso no es necesario, puesto que no usamos ese tipo de variables.



Variable: En Gráficos de categorías conjuntas se pide el mapa de posicionamiento conjunto de las categorías de las variables que participaron en la construcción del modelo. Este es la principal herramienta de interpretación del mismo. Incluir aquí todas las variables. Por su parte, los gráficos de categorías nos muestra la posición de las categorías de cada variable por sí sola en el mapa (facilita la lectura del gráfico de categorías conjuntas, por ejemplo cuando este tiene muchas variables).



PEGAR

Sintaxis:

```

MULTIPLE CORRES VARIABLES=Ingreso Educ Duración Posición Formalidad EscalaDecisión
/ANALYSIS=Ingreso(WEIGHT=1) Educ(WEIGHT=1) Duración(WEIGHT=1) Posición(WEIGHT=1)
Formalidad(WEIGHT=1) EscalaDecisión(WEIGHT=1)
/MISSING=Ingreso(PASSIVE,MODEIMPU) Educ(PASSIVE,MODEIMPU)
Duración(PASSIVE,MODEIMPU)
Posición(PASSIVE,MODEIMPU) Formalidad(PASSIVE,MODEIMPU)
EscalaDecisión(PASSIVE,MODEIMPU)
/DIMENSION=2
/NORMALIZATION=VPRINCIPAL
/MAXITER=100
/CRITITER=.00001
/PRINT=DISCRIM OBJECT
/PLOT=OBJECT(20) CATEGORY(Ingreso Educ Duración Posición Formalidad EscalaDecisión)
(20)

```

Ejecutar**3. RESULTADOS****Resumen del procesamiento de los casos**

Casos activos válidos	0
Casos activos con valores perdidos	9503
Casos suplementarios	0
Total	9503
Casos usados en el análisis	9503

La primera salida nos indica la cantidad de casos válidos (9503).

Historial de iteraciones

Número de iteraciones	Varianza explicada		Pérdida
	Total	Incremento	
26 ^a	2,933229	,000007	8,066771

a. Se ha detenido el proceso de iteración debido a que se ha alcanzado el valor de la prueba para la convergencia.

El historial de iteraciones muestra los pasos que fueron necesarios para llegar a la solución última. Observamos que el paso 26 se detuvo, dado que el incremento de varianza explicada dejó de ser significativo como para seguir iterando.

Resumen del modelo

Dimensión	Alfa de Cronbach	Varianza explicada	
		Total (Autovalores)	Inercia
1	,786	3,503	,318
2	,635	2,364	,215
Total		5,866	,533
Media	,725 ^a	2,933	,267

a. El Alfa de Cronbach Promedio está basado en los autovalores promedio.

La tabla de resumen del modelo, permite observar que se crearon 2 dimensiones (tal como nosotros señalamos). El autovalor da cuenta de la proporción de información del modelo que es explicada por cada dimensión; permite analizar de la importancia de cada una de ellas. Acá se puede observar que la primera dimensión es más importante para el modelo que la segunda. A su vez, la primera explica más inercia (0,318) que la segunda (0,215), lo cual es esperable puesto que las dimensiones se obtienen mediante un Análisis Factorial, en que, a mayor dependencia entre

variables, mayor inercia. Esto quiere decir que las categorías presentan mayor dispersión de varianza en la dimensión 1, sin embargo ambas dimensiones tienen un valor similar de inercia. El alfa de Cronbach indica también qué tan correlacionadas están las variables observables que componen las variables latentes, las dimensiones, por lo que ambos valores (alfa de Cronbach e inercia) tienen una relación directa.

Cuantificaciones

Las siguientes tablas muestran las coordenadas de cada una de las categorías en ambas dimensiones, es decir, la posición de las categorías en el eje X e Y. No solo sirve para conocer la posición, sino que a su vez entrega información respecto a cómo se relacionan con cada dimensión: puntajes altos indican que la categoría se asocia a dicha dimensión, puesto que mientras más lejos del punto de origen 0,0 más relación tiene con una dimensión.

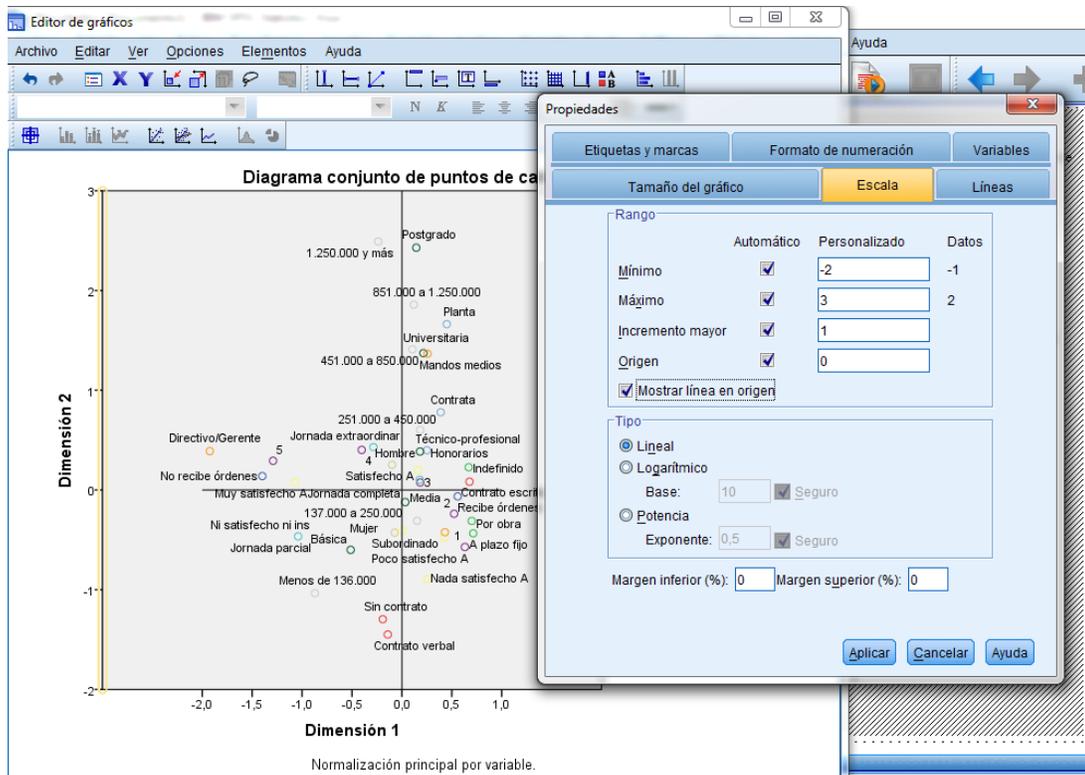
Posición dentro de la empresa

Puntos: Coordenadas

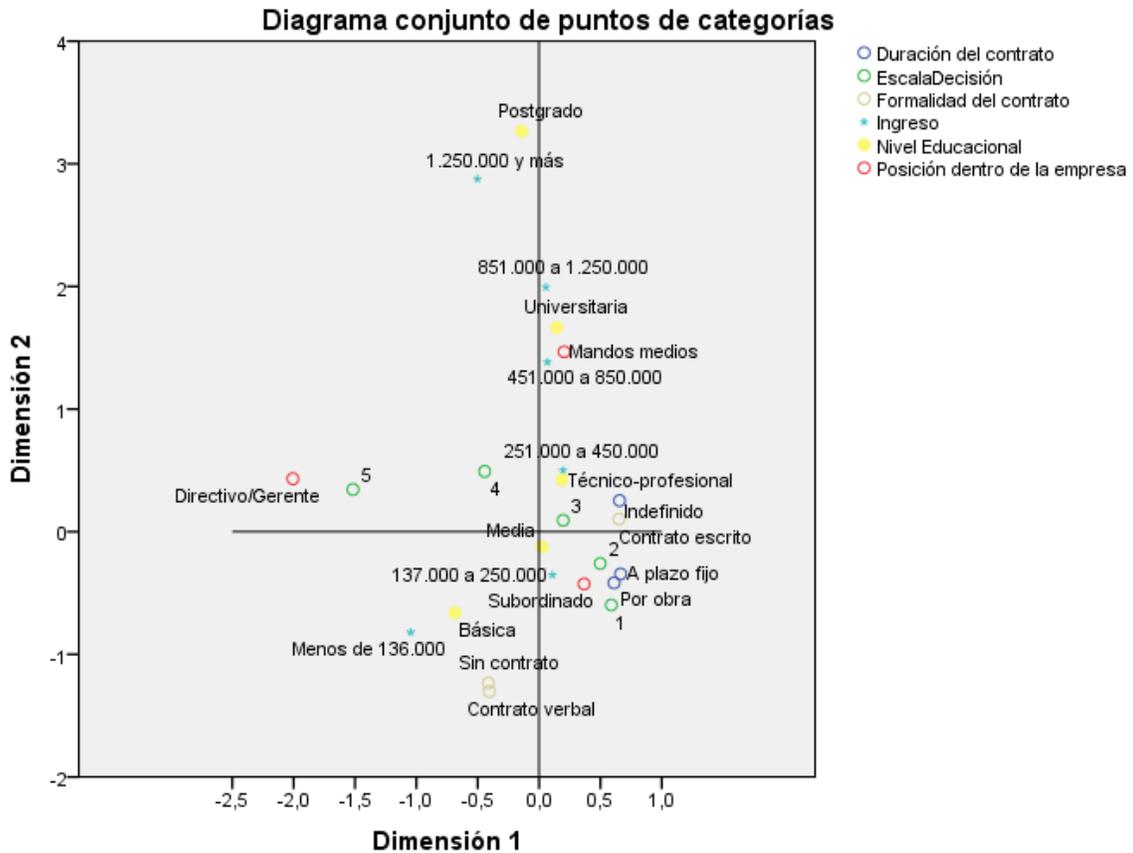
Categoría	Frecuencia	Coordenadas de centroide	
		Dimensión	
		1	2
Directivo/Gerente	1999	-2,006	,432
Mandos medios	1334	,205	1,467
Subordinado	6168	,366	-,425
Perdidos	2		

Normalización principal por variable.

Así por ejemplo, en la variable "Posición dentro de la empresa" se observa que la categoría "gerente" tiene fuerte relación con la dimensión 1, mientras que las otras con el 2.

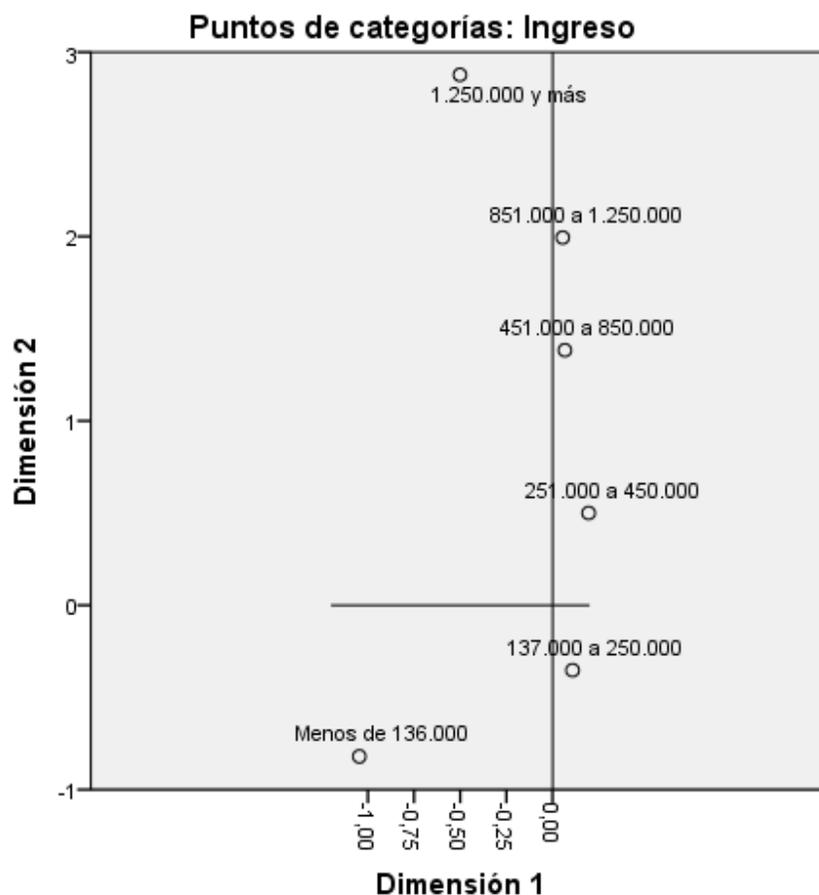


A continuación nos muestra el gráfico de categorías conjuntas. Para editarlo, **hacemos doble click en el gráfico** y se abrirá el editor de gráficos. Para que aparezcan las líneas de origen y sea más fácil leer el mapa, hacemos click en los signos "X" e "Y" y marcamos en "**mostrar línea de origen**". A su vez, podemos modificar la forma y color de la simbología de cada categoría **haciendo doble click en la figura de las variables** arriba a la derecha.



Normalización principal por variable.

El diagrama conjunto de puntos de categorías nos muestra el mapa de correspondencias propiamente tal, con todas las variables. Para interpretarlo podemos buscar patrones, grupos de categorías, analizar la cercanía o lejanía con el origen, ver la distribución a lo largo de cada eje (esto tiene sentido especialmente si se comprueba que los ejes son explicativos y con cuáles variables se relacionan más), analizar en qué cuadrante se encuentra cada punto. Es importante recalcar que la interpretación en torno a las dimensiones creadas depende de la capacidad explicativa (inercia) de cada una. En este caso ambas tienen una capacidad explicativa similar y suficiente para interpretar la cercanía de las categorías con cada eje que las representa, pero en caso de que no fuese así, hay casos en que alguna de las dimensiones no tiene sentido su interpretación debido a su escaso aporte interpretativo, o una de las dimensiones es mucho más importante que la otra a la hora de analizar la posición de las categorías. Aquí, observamos que las variables de caracterización socioeconómica (ingreso y educación) estructuran el eje Y, y las de nivel de autonomía en el trabajo con el X (posición dentro de la empresa, nivel de control).



Normalización principal por variable.

Los gráficos de puntos de cada categoría permiten leer mejor el comportamiento de cada variable por sí sola en el mapa. Esta información es útil cuando existen demasiados puntos en el gráfico. En el ejemplo, la variable "Ingreso" claramente se vincula con el eje Y.

Medidas de discriminación

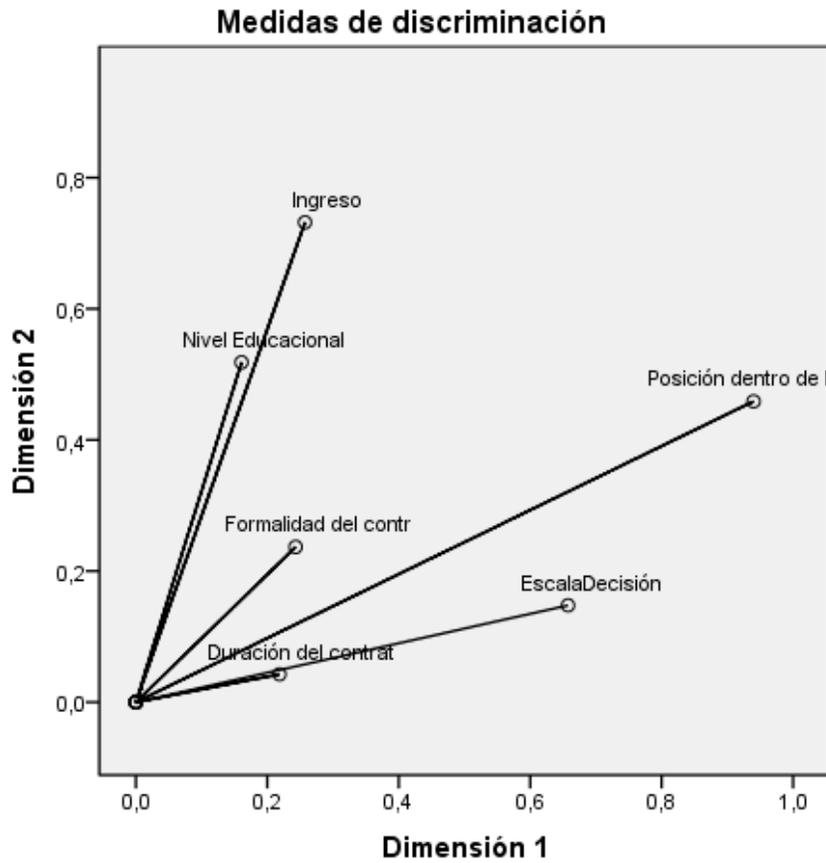
Esta tabla permite ver cuánto discrimina cada variable en cada dimensión; indican la importancia de cada variable para cada una de las dimensiones. Acá se ve que la dimensión 1 se encuentra explicada principalmente por la posición dentro de la empresa y el nivel de control sobre el proceso de trabajo, y la dimensión 2 por el ingreso y nivel educacional.

Medidas de discriminación

	Dimensión		Media
	1	2	
Ingreso	,257	,732	,495
Nivel Educacional	,161	,518	,340
Duración del contrato	,218	,043	,131
Posición dentro de la empresa	,940	,459	,699

Formalidad del contrato	,243	,237	,240
EscalaDecisión	,658	,148	,403
Total activo	2,477	2,136	2,307

El gráfico presenta la misma información que la tabla. Mientras más lejos del origen más explicativa es la variable, y la cercanía con una u otra dimensión dan cuenta de su relación con ésta. Se observa que el ingreso y el nivel educacional se relacionan con la dimensión 2, y así.



Normalización principal por variable.