

Fundamentos en Humanidades
Universidad Nacional de San Luis – Argentina
Año XI – Número II (22/2010) 75/87 pp.

Del análisis textual al análisis multidimensional

From textual to multidimensional analysis

Eduardo Escalante Gómez

Universidad del Aconcagua
Eduardoescalante2003@yahoo.com

(Recibido: 20/08/09 – Aceptado: 12/05/10)

Resumen

El propósito de este artículo es describir la aplicación del análisis de correspondencias a datos derivados de entrevistas en profundidad a docentes sobre sus conocimientos de los procesos de enseñanza y aprendizaje. Se aplicó una técnica de análisis descriptivo multidimensional para explorar los datos cualitativos. Este tipo de estudio es útil para mejorar la calidad del aprendizaje y la enseñanza. Una mayor comprensión de los procesos de enseñanza y aprendizaje permitirían mejorar la calidad de la enseñanza y el aprendizaje.

Abstract

The purpose of this article is to describe the correspondence analysis applied to the data gathered through in-depth interviews to teachers, inquiring about their knowledge of teaching and learning processes. The qualitative data were examined by a descriptive multidimensional analysis. This type of study would help to improve the quality of teaching and learning through a better understanding of these processes.

Palabras clave

análisis de correspondencias - entrevistas - aprendizaje - enseñanza
- perfiles

Key words

correspondence analysis - interviews - learning - teaching - profiles

La técnica del análisis de correspondencias

La finalidad del análisis de correspondencia es encontrar una representación dimensional de la dependencia entre categorías predeterminadas en una tabla de contingencia de doble entrada. Este tipo de análisis fue creado por Benzecri (1992).

Se puede conceptualizar de manera similar al análisis de componentes principales, con la calificación que el análisis de correspondencia es capaz de abordar datos de frecuencias. Algunos nombres de técnicas similares incluyen el escalamiento óptimo, escalamiento dual, análisis de correlación canónica y análisis de homogeneidad. Este tipo de análisis no está forzado a la consideración de que los datos tengan una distribución normal.

La fortaleza del análisis de correspondencia descansa en su interpretación de soluciones de dimensionalidad baja en representaciones gráficas que permiten al investigador hacer comparaciones entre los participantes, entre las variables, y entre los participantes y las variables en su ubicación relativa en un espacio de baja dimensión compartido. Una descripción más detallada puede hallarse en Greenacre (1992, 1994) y Lebart (1984).

Patrones de inter-dependencia

El análisis de correspondencia emplea las distancias ji cuadrado para calcular la disimilaridad (o similaridad) entre las frecuencias en cada celda de una tabla de contingencia. El concepto subyacente al cálculo de las distancias ji cuadrado es la interdependencia de las celdas. El análisis de correspondencia estandariza y transforma la frecuencia de los datos en la tabla de contingencia mediante el cálculo de las distancias de los perfiles de filas y columnas. El programa reduce la complejidad obtenida en los perfiles de columnas y filas mediante la creación de una representación de baja dimensionalidad de los perfiles de columnas y filas.

Normalización

El análisis de correspondencia requiere que el investigador elija entre diferentes métodos de normalización. El análisis de correspondencia que usa la normalización principal de columnas calcula las distancias euclídeas entre un punto de columna y el origen, el cual aproxima la distancia ji cuadrado entre la categoría columna y la categoría promedio de columna. En seguida, prepara la representación gráfica de las distancias de las coordenadas en el espacio de dimensionalidad baja. El análisis que usa la

normalización principal de filas emplea un procedimiento similar al referido anteriormente, pero para las filas de la tabla de contingencia. El análisis que usa la normalización simétrica distribuye la inercia (la correlación al cuadrado entre los puntajes de filas y columnas).

Las diferencias entre los métodos han sido bastante discutidas en la literatura respecto de la elección más apropiada de normalización y métodos de interpretación de las representaciones visuales (Gabriel, 2002). Se señala que los investigadores que tienen interés específico en las magnitudes actuales de diferencia entre filas (participantes) o columnas (variables) deberían elegir la normalización principal (filas o columnas). No obstante, los investigadores cuyos intereses consiste en comparar la orientación general de los puntos de fila y los puntos de columnas, en lugar de visualizar las magnitudes actuales, debería usar la normalización simétrica.

En este estudio la preocupación consiste en interpretar el significado de las dimensiones extraídas en la solución de baja dimensionalidad, e interpretar la ubicación de los participantes relativos a esas dimensiones. Por consiguiente, se eligió la normalización simétrica para la representación gráfica.

Método

Los participantes

Se trabajó con una muestra de ocho participantes a los cuales se aplicó un proceso de entrevistas en profundidad. Todos los participantes tienen actividades académicas de nivel universitario y sus edades fluctúan entre 29 y 40 años.

Las entrevistas

Se revisó la literatura relativa a la enseñanza y al aprendizaje en orden a configurar 18 preguntas que guiaron la dirección de cada entrevista. Uno de los estudios que sirvió para la elaboración del sistema de categorías y elaboración de las preguntas fue Trigwell, Prosser, y Waterhouse (1999). Las preguntas se categorizaron del siguiente modo:

- Categorías de enseñanza y aprendizaje
- La naturaleza de los entornos de aprendizaje
- La naturaleza de la enseñanza y el aprendizaje
- La naturaleza del aprendiz
- La naturaleza de la materia

Cada una de estas categorías se sustentaron teóricamente, por ejemplo, la teoría de la autoeficacia, la teoría de la auto-regulación, la teoría del contenido del currículo, la teoría del propósito del currículo, la teoría de la cognición y meta-cognición valores e intereses, entre otras.

A modos de ejemplo, se incluyó: la pregunta ¿Qué procesos de pensamiento usarán los alumnos en esta sesión/tópico/curso? (asociada a la cognición y a la meta-cognición); la pregunta ¿Cómo sabrá que han aprendido lo que usted piensa que deberían haber aprendido? (asociada a la auto-regulación); la pregunta ¿Por qué esta enseñando esto? (asociada al propósito del currículo).

La finalidad fue elicitare la comprensión de los participantes sobre la enseñanza y el aprendizaje en acción. Cada entrevista duró entre 20 y 90 minutos, con un promedio de 45 minutos. Las entrevistas fueron desgrabadas y digitas en Word como archivo de texto. Se incluyeron solamente los nombres de los docentes en orden a mantener la confidencialidad de la información.

Codificación de los textos de las entrevistas

Los textos de las entrevistas fueron codificados de modo de extraer las frecuencias que podían ser categorizadas en 38 variables. Estas variables se identificaron en una etapa inicial de la investigación usando temas aludidos por la literatura referida a la enseñanza y al aprendizaje. Para el procesamiento de los textos se usaron los programas *TextStat* y *Yoshikoder*. En este artículo no se ha incluido la información resultante a partir de la ejecución de estos programas, de modo de concentrarse en el análisis de correspondencia solamente.

Categorías de Enseñanza y Aprendizaje

Una muestra de las categorías empleadas para la codificación fue:

Tabla 1 - Categorías

Nivel 1	Nivel 2	Variables para el análisis de correspondencia
Naturaleza del aprendizaje	<p>Motivación Valor de la materia Útil y/o importante para el futuro</p> <p>Interesante Auto-eficacia Creencia en las capacidades</p>	<p>Útil interesante</p> <hr/> <p>Auto-eficacia fuerte Auto-eficacia débil</p>
Naturaleza de la enseñanza y aprendizaje	<p>Construcción de conocimiento Pertenencia a una comunidad de aprendizaje</p>	<p>Comunidad de aprendizaje</p>
Naturaleza del entorno de aprendizaje	<p>Aprendizaje auténtico Aprendizaje mediante práctica auténtica</p> <p>Aprendizaje a través de la interacción social Discusiones, representaciones</p>	<p>Práctica auténtica Aprendizaje social</p>
Naturaleza de la materia	<p>Propósito de la materia Propósito de las experiencias de aprendizaje</p>	<p>Propósito</p>

La tabla anterior permite apreciar que son cuatro las categorías que proveen la estructura sustantiva para organizar los datos en el nivel 1; en seguida en el nivel 2 se incluyeron las 38 variables derivadas del análisis de las teorías. Este procedimiento de codificación permitió capturar la riqueza de los datos. Además, permitió identificar finalmente 28 variables dado que algunas categorías no se pudieron establecer con claridad en las entrevistas. A partir de esta información se construyó la tabla de contingencia. En todo momento se tuvo presente que el modelo se tiene que adaptar a los datos y no vice-versa.

Tabla de contingencia

Para el procesamiento estadístico de los datos se usó el programa SPSS 15.0. La primera fase en un análisis de correspondencia es ingresar las respuestas de las frecuencias de los participantes en la forma de una tabla de contingencia. Se ha incluido los 8 participantes, pero solamente 6 de las variables del total de 28 variables, de modo de ilustrar el procedimiento de análisis.

En la tabla 2 se puede apreciar que Andrés, fila 1, hizo 24 ocurrencias referidas al propósito del aprendizaje, 119 referidos a la práctica auténtica y así sucesivamente, con un total de 653 ocurrencias. Si se leen las columnas sobre la cantidad de ocurrencias referidas por los 8 entrevistados respecto del propósito del aprendizaje, se tiene un total de 183 ocurrencias. Se obtuvo un total de 5474 ocurrencias al momento de incluir los actores (8 entrevistados) y 6 variables.

Tabla 2 - Tabla de correspondencias

Nombre	Categorías						Margen activo
	Propósito	Práctica auténtica	Aprendizaje social	Entorno apoyo	Meta cognición	Auto regulación	
Andrés	24	119	14	20	304	172	653
José	46	171	0	0	159	144	520
Juan	19	124	0	0	108	117	368
Raúl	38	249	21	0	167	294	769
Rosa	32	285	27	0	430	169	943
Sandra	15	206	31	3	513	493	1261
Jorge	9	297	15	5	175	227	728
Salvador	0	78	21	2	47	84	232
Margen activo	183	1529	129	30	1903	1700	5474

Perfiles de filas y columnas

Cada fila y columna de la tabla de contingencia se caracteriza por su perfil, el cual es un sistema de proporciones. El análisis de correspondencia empieza calculando los perfiles de fila, que son las proporciones relativas de cada variable dentro de todas las variables mencionadas por cada participante. Los perfiles de filas permiten una comparación de las variables en cada participante.

Tabla 3 - Perfiles de fila

Nombre	Categorías						
	Propósito	Práctica auténtica	Aprendizaje social	Entorno apoyo	Meta cognición	Auto regulación	Margen activo
Andrés	,037	,182	,021	,031	,466	,263	1,000
José	,088	,329	,000	,000	,306	,277	1,000
Juan	,052	,337	,000	,000	,293	,318	1,000
Raúl	,049	,324	,027	,000	,217	,382	1,000
Rosa	,034	,302	,029	,000	,456	,179	1,000
Sandra	,012	,163	,025	,002	,407	,391	1,000
Jorge	,012	,408	,021	,007	,240	,312	1,000
Salvador	,000	,336	,091	,009	,203	,362	1,000
Masa	,033	,279	,024	,005	,348	,311	

A modo de ejemplo se analizará el texto de Andrés, esto es, el perfil de fila con la proporción de cada variable que apareció en el texto de la entrevista. Un 0,037 está referido al propósito del aprendizaje; un 0,182 referido a la práctica auténtica y así sucesivamente. El total marginal de 1 es equivalente al 100 por ciento del total de las ocurrencias en el texto de Andrés.

El paso siguiente del análisis de correspondencias es el cálculo de los perfiles de columna. Estos son la proporción de cada variable mencionada por cada participante como total de las menciones de la variable por todos los participantes.

Tabla 4 - Perfiles de columna

Nombre	Categorías						
	Propósito	Práctica auténtica	Aprendizaje social	Entorno apoyo	Meta cognición	Auto regulación	Masa
Andrés	,131	,078	,109	,667	,160	,101	,119
José	,251	,112	,000	,000	,084	,085	,095
Juan	,104	,081	,000	,000	,057	,069	,067
Raúl	,208	,163	,163	,000	,088	,173	,140
Rosa	,175	,186	,209	,000	,226	,099	,172
Sandra	,082	,135	,240	,100	,270	,290	,230
Jorge	,049	,194	,116	,167	,092	,134	,133
Salvador	,000	,051	,163	,067	,025	,049	,042
Margen Activo	1,000	1,000	1,000	1,000	1,000	1,000	

Esto permite comparar a los participantes a partir de las variables de las columnas. La tabla anterior permite observar que Andrés contribuye con un 0,131, al examinar la variable propósito del aprendizaje, y Jorge con un 0,048, y en el caso de Salvador hay cero contribución para esta variable. El margen activo o margen total provee la masa, o peso, de la contribución de cada participante a la solución dimensional.

Las variables en la solución del análisis de correspondencia

La tabla 5, tomada del análisis de correspondencia (normalización simétrica), genera una salida que contiene los valores propios, la inercia, y la proporción de la variación explicada por 5 dimensiones. El valor propio indica la contribución relativa de cada dimensión a la explicación de la inercia, o varianza, en los perfiles de participantes y variables. Los valores singulares se pueden interpretar como la correlación entre las filas y las columnas de la tabla de contingencia, y son análogos a los coeficientes de correlación de *Pearson* en el análisis de correlación.

La primera dimensión explica el máximo de varianza posible, la segunda dimensión es ortogonal a la primera y muestra el máximo de varianza posible, y así sucesivamente. Se recomienda que los valores propios mayores que 0,2 indican que la dimensión debería ser incluida en el análisis, de todos modos hay que tener presente la varianza explicada por cada dimensión. El valor propio y la inercia están directamente relacionadas: la inercia es un indicador de cuánta de la variación en los datos originales se retiene en la solución dimensional.

Tabla 5

Resumen								
Dimensión	Valor propio	Inercia	Chi-cuadrado	Sig.	Proporción de Inercia		Confianza para el valor propio	
					Explicada	Acumulada	Desviación Típica	Correlación 2
1	,226	,051			,430	,430	,013	-,039
2	,174	,030			,257	,687	,013	
3	,135	,018			,155	,842		
4	,115	,013			,112	,954		
5	,074	,005			,046	1,000		
Total		,118	648,141	,000 ^a	1,000	1,000		

a. 35 grados de libertad

La representación gráfica

A continuación, se mostrará las representaciones visuales de las dimensiones identificadas en el análisis de correspondencia. Se representarán las dos dimensiones iniciales que explican el 68,7% de la proporción de inercia acumulada. Seguidamente, se incluyen los valores de las dos dimensiones referidas a los participantes (filas) y variables (columnas) que permitirán generar la representación gráfica. Esto es se incluyen las coordenadas de las dimensiones 1 y 2, con la dimensión 1 en el eje horizontal y la dimensión 2 en el eje vertical. Para interpretar cada dimensión es necesario considerar la contribución de las variables a esa dimensión.

Tabla 6

Examen de los puntos de fila ^a									
Nombre	Masa	Puntuación en la dimensión		Inercia	Contribución				
		1	2		De los puntos a la inercia de la dimensión		De la dimensión a la inercia del punto		Total
					1	2	1	2	
Andrés	,119	,761	,180	,023	,306	,022	,663	,029	,692
José	,095	-,368	,611	,013	,057	,203	,223	,476	,699
Juan	,067	-,367	,243	,004	,040	,023	,511	,173	,684
Raúl	,140	-,540	-,212	,012	,181	,036	,761	,091	,851
Rosa	,172	,220	,508	,017	,037	,255	,111	,461	,573
Sandra	,230	,416	-,467	,022	,177	,288	,412	,402	,814
Jorge	,133	-,536	-,082	,014	,169	,005	,609	,011	,620
Salvador	,042	-,418	-,830	,013	,033	,167	,129	,392	,521
Total Activo	1,000			,118	1,000	1,000			

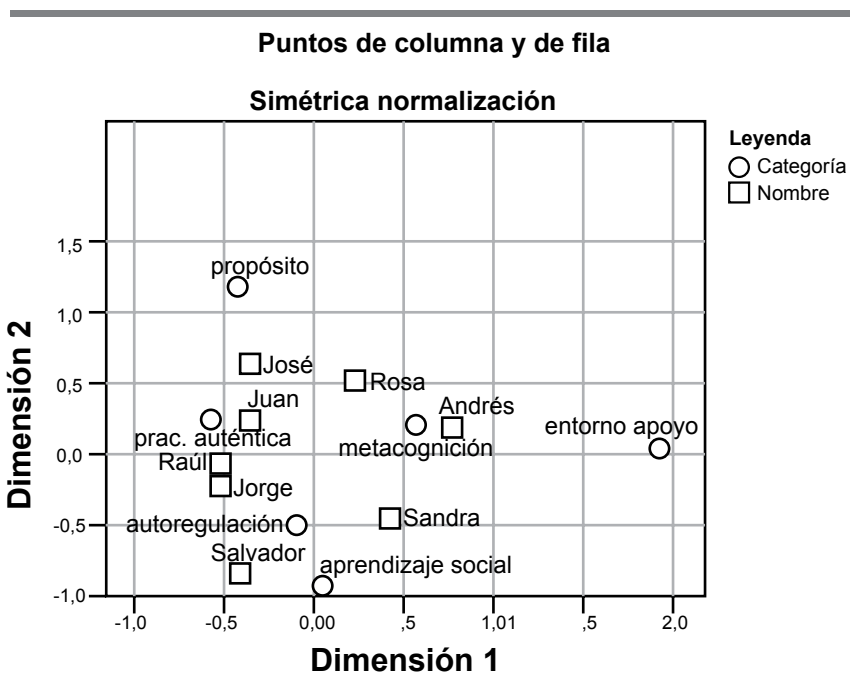
a. Normalización Simétrica

Tabla 7

Examen de los puntos columna ^a									
Categoría	Masa	Puntuación en la dimensión		Inercia	Contribución				
		1	2		De los puntos a la inercia de la dimensión		De la dimensión a la inercia del punto		Total
					1	2	1	2	
propósito	,033	-,428	1,175	,017	,027	,265	,083	,480	,563
prac. auténtica	,279	-,587	,236	,026	,399	,089	,767	,102	,870
aprendizaje social	,024	,046	-,948	,012	,000	,121	,001	,302	,303
entorno apoyo	,005	1,913	,024	,017	,089	,000	,268	,000	,268
metacognición	,348	,553	,206	,028	,471	,084	,863	,093	,955
autoregulación	,311	-,100	-,498	,018	,014	,441	,038	,734	,773
Total Activo	1,000			,118	1,000	1,000			

a. Normalización Simétrica

Gráfico 1



Raúl, Jorge, José y Salvador se agrupan a la izquierda de la dimensión 1, cercanos a la práctica auténtica y a la auto-regulación. Rosa, Andrés y Sandra se agrupan a la derecha de esta dimensión. Al considerar la dimensión 2, José y Rosa se asocian más a los propósitos del aprendizaje; en cambio Salvador y Sandra se asocian al aprendizaje social.

Dimensiones 1 y 2

El polo izquierdo de la dimensión 1 contiene variables como práctica auténtica (0,399). Esta variable dice relación con la naturaleza del entorno del aprendizaje. Otra variable que se ubica en este mismo polo es la meta-cognición (0,471). La variable entorno de apoyo (0,089) se ubica a la mano derecha de la dimensión 1.

En la dimensión 2 se ubican los propósitos del aprendizaje (0,265) y la auto-regulación (0,441).

Pero además el análisis de correspondencias permite ubicar los puntajes de los participantes en el mismo espacio multidimensional donde están ubicadas las variables.

El método de normalización simétrica seleccionado para el análisis de correspondencias, permite la ubicación de los puntajes de las variables y los participantes en la misma representación gráfica.

Un cuadro más completo de las variables

A continuación se incluye una representación gráfica de 15 de las 28 variables generadas por los procesos de análisis.

Gráfico 2 - Variables

Propósito	Interesante
Dificultad en estudios	Auto-realización
Auto-eficacia fuerte	Comunidad de Aprendizaje
Auto-eficacia débil	Diferencias individuales
Práctica auténtica	Entorno de apoyo
	Metacognición
Expectativas de éxito	Regulación externa
Auto-regulación	Aprendizaje social

Conclusiones

Este artículo describe la reducción sustancial de una muestra de 5.474 ocurrencias de un total de 18.000 codificadas a partir de las entrevistas a 8 docentes, identificándose dos dimensiones y representándose en un espacio dimensional. Estas dos dimensiones pueden resumirse como sigue:

Dimensión 1: Focalización del aprendizaje, diferenciando a los entrevistados en dos polos.

Dimensión 2: Motivación, diferencias individuales, separación de lo extrínseco e intrínseco.

Desde el punto de vista metodológico, el análisis de correspondencia permite la representación gráfica de los participantes y de las variables, siendo muy útil para cuando se dispone de frecuencias de datos que se pueden representar en tablas de contingencia. Además, los análisis desarrollados permiten elaboraciones estadísticas de bastante interesantes datos de origen cualitativo.

Mendoza (Argentina), 20 de marzo de 2009.

Referencias bibliográficas

Benzecri, J.-P. (1992). *Correspondence analysis handbook*. New York: Marcel Dekker.

Gabriel, K. R. (2002). Goodness of fit of biplots and correspondence analysis. *Biometrika*, 89(2), 423-436.

Greenacre, M. (1992). Correspondence analysis in medical research. *Statistical Methods in Medical Research*, 1, 97-117.

Greenacre, M. (1994). Correspondence analysis and its interpretation. En M. Greenacre y J. Blasius (Eds.), *Correspondence analysis in the social sciences* (pp. 3-22). London: Academic Press.

Lebart, L., Morineau, A. y Warwick, K. (1984) *Multivariate Descriptive Statistical Analysis. Correspondence analysis and related techniques for large matrices*. New York: J.Wiley & Sons.

Trigwell, K., Prosser, M. y Waterhouse, F. (1999). Relations between teachers' approaches to teaching and students' approaches to learning. *Higher Education*, 37, 57-70.