



# UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

---

INSTITUTO DE INVESTIGACIONES EN MATEMÁTICAS APLICADAS Y EN SISTEMAS

## APLICACIÓN DE UN ANÁLISIS DE CONGLOMERADOS EN LA SEGMENTACION DE EMPRESAS EDITORIALES EN MÉXICO

**T E S I S I N A**  
QUE PARA OBTENER EL TÍTULO DE:  
**E S P E C I A L I S T A**  
EN **ESTADÍSTICA APLICADA**  
P R E S E N T A:  
ACT. CLAUDIA ESTHER ALCALÁ ESCAMILLA

DIRECTOR DE TESINA: DR. IGNACIO MÉNDEZ RAMÍREZ

MÉXICO, D.F.

2012



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

*Agradezco sinceramente a:*

*Dr. Ignacio Méndez Ramírez,*

*por aceptar dirigir este trabajo, enriqueciéndolo con sus comentarios  
y sus conocimientos, y por su paciencia a lo largo de todo este tiempo,*

*Dra. Silvia Ruiz-Velasco Acosta,*

*M. en E. Adriana Margarita Ducoing Watty,*

*M. en C. Leticia Eugenia Gracia-Medrano Valdelamar*

*M. en C. Patricia Isabel Romero Mares*

*por su disposición a revisar este documento  
y por aportarme sus comentarios y su experiencia*

*Gracias a su invaluable apoyo y paciencia,  
este trabajo ha sido terminado*

# APLICACIÓN DE UN ANÁLISIS DE CONGLOMERADOS EN LA SEGMENTACIÓN DE EMPRESAS EDITORIALES EN MÉXICO

<b>INTRODUCCIÓN.....</b>	<b>3</b>
<b>1. SECTOR EDITORIAL EN MÉXICO.....</b>	<b>6</b>
1.1 <i>Composición del sector</i> .....	6
1.2 <i>Cifras principales</i> .....	7
1.2.1 Producción .....	7
1.2.2 Comercialización .....	9
1.2.3 Inversión .....	11
1.2.4 Empleo .....	11
1.3 <i>Definición de las variables de estudio</i> .....	13
<b>2. ANÁLISIS DE CONGLOMERADOS .....</b>	<b>15</b>
2.1 <i>MEDIDAS DE PROXIMIDAD</i> .....	16
2.1.1 Medidas de disimilitud.....	16
2.1.2 Medidas de similitud.....	18
2.2 <i>ALGORITMOS DE CONGLOMERACIÓN</i> .....	19
2.2.1 Métodos jerárquicos.....	19
2.2.2 Métodos de partición (no jerárquicos) .....	21
2.3 <i>Interpretación de los resultados</i> .....	23
<b>3. RESULTADOS Y DISCUSIÓN .....</b>	<b>24</b>
3.1 <i>Descripción de los conglomerados</i> .....	26
<b>4. CONCLUSIONES .....</b>	<b>34</b>
<b>ANEXO I .....</b>	<b>36</b>
<b>ANEXO II .....</b>	<b>38</b>
<b>ANEXO III .....</b>	<b>41</b>
<b>5. BIBLIOGRAFÍA .....</b>	<b>43</b>

## INTRODUCCIÓN

La Cámara Nacional de la Industria Editorial Mexicana (CANIEM) es un organismo gremial en el que se conjuntan empresas e instituciones cuyo giro principal es la edición de libros o de publicaciones periódicas. Su objetivo primordial es la defensa de los intereses de estos empresarios, así como la promoción del crecimiento profesional y económico del sector editorial en México.

Una de las tareas de este organismo es la elaboración anual del informe que da cuenta de la condición en la que se encuentra la producción y la comercialización del libro en México, llamado Actividad Editorial. En la realización de este estudio se utiliza la metodología elaborada por el Centro Regional para el Fomento del Libro en América Latina y el Caribe (CERLALC), en la que las empresas editoriales son clasificadas según dos características principales:

- Rango de tamaño. Categoría que se asigna a las editoriales en función del monto de facturación obtenido por la venta de libros en un año. Se contemplan 5 niveles: macro, grande, mediana, pequeña y micro.
- Subsector editorial. Categoría que se asigna a las editoriales en función de su producción temática predominante. Es excluyente y consta de 4 divisiones editoriales: Texto, Interés general, Científicos, Técnicos y Profesionales y Religiosos.

La combinación de estas características da lugar a 20 estratos distintos en los que se clasifica cada una de las editoriales en México.

Para el levantamiento de la información de las editoriales, que da forma a la Actividad editorial, se utiliza el cuestionario propuesto por el CERLALC, en el que se abordan temas de importancia central, como es la producción (por temática y por subsector editorial), las ventas por zonas geográficas, la venta por canales de comercialización, el comercio exterior (exportación e importación), los derechos de autor y la generación de empleo. En el caso de México se han incluido temas complementarios que son de interés en nuestro contexto particular, como las coediciones, el valor de producción y la inversión en publicidad.

En cuanto al número de empresas que deben ser entrevistadas, el CERLALC apunta que, dado que el número de editores en cada país es pequeño (en México en 2008 era de 227 empresas) y tomando en cuenta que se estudiarán en función de sus ventas y del subsector editorial, lo ideal sería que se pudiera aplicar la encuesta a todo el universo. Sin embargo la decisión de un esquema censal y uno muestral debe tomarse a la luz del costo-beneficio de la investigación. [Ardila 2003]

La experiencia del CERALC demuestra que el nivel de respuesta de las empresas micro y pequeñas es bajo, por lo que la recomendación es aplicar la encuesta a una muestra de las empresas que componen este rango de tamaño. Para la obtención de esta muestra se utiliza un esquema estratificado con asignación proporcional y la selección de la muestra en cada estrato es probabilística, utilizando el valor de facturación como variable guía. A partir de esta muestra, se estiman los totales de ejemplares producidos y vendidos, y los demás temas que son considerados centrales.

En el caso de que los estratos tengan pocos elementos, como ocurre con las empresas de tamaño Mediana, Grande y Macro, la consideración del CERALC es la de estudiar el estrato en su totalidad y no a través de una muestra. En caso de no poder obtener el total de empresas en el estrato, el CERALC indica que deberá recurrirse a ejercicios de imputación o estimaciones. [Ardila 2003]

En el caso particular de México, ocurre que no todas las empresas de rango Mediana, Grande o Macro responden el cuestionario, por lo que se debe recurrir a los ejercicios de estimación mencionados por el CERALC, los cuales se hacen en función del rango de tamaño y de la línea editorial de cada empresa.

Sin embargo, las editoriales de los tres rangos más grandes, que además ejercen una influencia significativa en la industria editorial de nuestro país, elaboran libros de una gran diversidad de materias que no se restringen a una sola línea temática, por lo que al clasificarlas en una sola división editorial (Texto, Interés General, Ciencia y Técnica, Religiosos) como indica la metodología del CERALC, es posible que no se manifieste la variedad de temas producidos por estas empresas, además de que las diferencias entre éstas no sólo provienen de las temáticas, sino también de la distribución en sus canales de comercialización, del número de empleados, de su inversión en publicidad, etc.

Por esto se piensa que cuando no se obtiene información de todas las empresas medianas, grandes o macro, las estimaciones de la Actividad Editorial pueden no reflejar los valores reales del sector, por lo que la hipótesis planteada en este trabajo es que la estratificación propuesta por el CERALC no refleja la realidad de las editoriales mexicanas, en particular de las más grandes. El objetivo es identificar la agrupación de las editoriales mexicanas tomando en cuenta otras variables distintas a las sugeridas por el CERALC (línea editorial y rango de facturación), a través de un análisis de conglomerados. Las variables que se utilizarán para este análisis serán: monto total de facturación, clasificación temática de la producción (16 temas diferentes), canales de comercialización (15 canales diferentes), nivel de empleo, e inversión en publicidad.

Con la elaboración de este trabajo, se podrá obtener información que permita averiguar si la estratificación propuesta por el CERALC es la adecuada en la conformación de estratos representativos de las características de las empresas de la industria editorial en México. Se podrá también hacer propuestas para investigar nuevos criterios de clasificación de las editoriales mexicanas.

Para poder alcanzar el objetivo planteado, la investigación se estructura en cuatro apartados, que se describen a grandes rasgos a continuación:

En primer lugar se presenta un resumen del sector editorial en México, según el informe de Actividad Editorial en 2008, publicada por CANIEM. En esta síntesis se introducen las cifras principales del sector: producción de ejemplares, venta de ejemplares y valor de facturación, inversión en publicidad y empleo. Esto desagregado por los estratos que sugiere el CERLALC: división editorial y rango de tamaño.

En segundo lugar se hace una breve introducción a los métodos de análisis de conglomerados que contempla los conceptos principales de esta técnica y los métodos que pueden ser utilizados para llevar a cabo este análisis.

Como tercer punto se exponen los resultados obtenidos de la aplicación de esta técnica, haciendo una descripción de cada uno de los conglomerados resultantes en función de las variables que se tomaron en cuenta para realizar el análisis y se comentan las diferencias observadas entre estos conjuntos.

Finalmente se encuentran las conclusiones a las que se llegaron, y se presenta una propuesta acerca de cuál podría ser el siguiente paso.

# 1. SECTOR EDITORIAL EN MÉXICO.

## 1.1 Composición del sector

En 2008 se registraron un total de 227 empresas editoriales con actividad industrial (facturación anual superior a los 250 mil pesos). Estas empresas son clasificadas en función de 2 características principales:

- Subsector editorial (según la línea temática predominante):
  - **Texto:** empresas que producen libros para la educación básica y media (preescolar, primaria, secundaria y bachillerato), materiales de apoyo didácticos y libros para educadores.
  - **Interés General:** empresas que producen libros para todo tipo de público de los más diversos temas, incluidas las empresas que editan enciclopedias, colecciones generales y publicaciones seriadas.
  - **Científicos, Técnicos y Profesionales (CTP ó Ciencia y Técnica):** empresas que editan libros para estudiantes de educación superior y profesionales de las distintas áreas del conocimiento.
  - **Religiosos:** empresas que producen todo tipo de libros sobre las distintas religiones y creencias.
  
- Rango de tamaño (según el monto de facturación en un año):
  - **Proyecto inicial /Autor-editor (A/E):** proyectos editoriales con una facturación anual neta de más de 250 mil pesos y hasta un millón.
  - **Micro:** empresas con una facturación anual neta de más de un millón de pesos hasta 6 millones.
  - **Pequeña:** empresas con una facturación anual neta de más de 6 hasta 21 millones de pesos.
  - **Mediana:** empresas con una facturación anual neta de más de 21 hasta 81 millones de pesos.
  - **Grande:** empresas con una facturación anual de más de 81 hasta 200 millones de pesos.
  - **Macro:** empresas con una facturación anual neta de más de 200 millones de pesos.

La distribución en 2008 de cada una de las editoriales en las categorías arriba mencionadas fue la siguiente:

Tabla 1. Distribución de las editoriales en los estratos definidos por CERALC  
Número de empresas

Subsector editorial	A/E	Micro	Pequeña	Mediana	Grande	Macro	Total general
Texto	13	12	8	6	5	6	50
Interés general	37	26	17	6	6	3	95
CTP	22	17	10	11	2	2	64
Religiosos	6	7	2	3	0	0	18
Total general	78	62	37	26	13	11	227

## 1.2 Cifras principales

### 1.2.1 Producción

La producción de ediciones propias de estas 227 empresas en 2008 fue de 19,493 títulos y 123,271,568 ejemplares, incluyendo las ediciones que el sector privado hace para los programas de gobierno de Libro de texto gratuito para secundaria y Bibliotecas de aula y escolares.

Tabla 2. Producción de ejemplares por subsector editorial y por rango de tamaño

Subsector editorial	Macro	Grande	Mediana	Pequeña	Micro	Total
Texto	49,731,506	11,161,477	5,281,974	597,483	1,009,226	67,781,666
Interés general	22,352,143	7,675,026	1,963,252	5,821,322	2,358,023	40,169,766
CTP	6,778,094	2,539,446	3,086,039	391,717	486,068	13,281,364
Religiosos	-	-	1,001,629	384,500	652,643	2,038,772
Total general	78,861,743	21,375,949	11,332,894	7,195,022	4,505,960	123,271,568

Más de la mitad de la producción total es elaborada por las editoriales de Texto (55%), la segunda aportación más grande proviene de las empresas de Interés general, con 32.6%. Las editoriales de Ciencia y técnica participan con 10.8%, y los religiosos producen 1.7% de los ejemplares.

En cuanto al rango de tamaño, la mayor producción también se agrupa en una categoría, que en este caso es la de macro, con 64% de los ejemplares.

La principal temática de producción es la de educación básica, que junto con las ediciones que el sector privado hace para el programa de libro de texto gratuito de secundaria, conjuntan 54.4% de la producción. La siguiente temática más producida es la de

Literatura, con 7.7%, le siguen los libros infantiles con 6.4% y las ediciones para el programa de Bibliotecas de aula y escolares, que representan 6.1% y de las que 78% son libros infantiles (para preescolar y primaria).

Tabla 3. Producción de títulos y ejemplares por clasificación temática

<b>Temática</b>	<b>Títulos</b>	<b>Ejemplares</b>	<b>%</b>
Libros infantiles	1,817	7,864,549	6.4
Libros juveniles	580	1,858,792	1.5
Bibliotecas escolares y de aula	156	7,572,857	6.1
Literatura	2,149	9,519,123	7.7
Lenguas	1,127	6,201,507	5.0
Religión	894	2,697,945	2.2
Filosofía y Psicología	1,820	5,386,335	4.4
Artes, cultura y deportes	323	757,258	0.6
Ciencias puras	450	1,020,596	0.8
Tecnología y ciencias aplicadas	2,399	4,610,670	3.7
Ciencias sociales	1,324	4,600,435	3.7
Educación básica	4,960	39,859,892	32.3
Texto gratuito para secundaria	234	27,243,670	22.1
Geografía e Historia	235	645,609	0.5
Generalidades	651	2,497,045	2.0
Otros	374	935,285	0.8
<b>Total</b>	<b>19,493</b>	<b>123,271,568</b>	<b>100</b>

El catálogo conjunto (número de títulos que una empresa produce y/o distribuye) en 2008 fue de 183,087, de los que 38.9% son ediciones propias de las editoriales, 5.7% son ediciones que se distribuyen de otros editores nacionales y 56.6% son títulos de otros editores extranjeros que las editoriales mexicanas distribuyen.

Tabla 4. Número de títulos disponibles en los catálogos de las editoriales

<b>Subsector editorial</b>	<b>Número de títulos</b>	<b>%</b>
Texto	27,592	15.1
Interés general	72,163	39.4
CTP	70,304	38.4
Religiosos	13,028	7.1
<b>Total general</b>	<b>183,087</b>	<b>100.0</b>

La mayor variedad de títulos en catálogo se encuentra con las editoriales de Interés general y de Ciencia y Técnica, con 39% y 38% respectivamente, y en ambos casos,

alrededor de 61% de los títulos son de otros editores extranjeros. Esta proporción es inversa en el caso de las editoriales de Texto, en las que 60% de los títulos son ediciones propias.

### 1.2.2 Comercialización

En 2008 se vendieron un total de 133,314 títulos y 138,689,028 ejemplares con un valor de 8,095,009,254 pesos.

Tabla 5. Venta de títulos, ejemplares y valor de facturación por subsector editorial

<b>Subsector editorial</b>	<b>Títulos vendidos</b>	<b>%</b>	<b>Ejemplares vendidos</b>	<b>%</b>	<b>Facturación</b>	<b>%</b>
Texto	19,427	14.6	67,422,517	48.6	3,450,645,621	42.6
Interés general	53,402	40.1	46,626,544	33.6	2,294,119,903	28.3
CTP	53,554	40.2	21,235,701	15.3	2,165,380,236	26.7
Religiosos	6,931	5.2	3,404,266	2.5	184,863,494	2.4
<b>Total</b>	<b>133,314</b>	<b>100.0</b>	<b>138,689,028</b>	<b>100.0</b>	<b>8,095,009,254</b>	<b>100.0</b>

La mayor participación se observa con las editoriales de Texto, con 48.6% de los ejemplares vendidos y 42.6% de la facturación. Sin embargo, la oferta de títulos más alta proviene de las empresas de Interés General y de Ciencia y Técnica, cuya aportación en el monto de facturación es de 28.3% y 26.7% respectivamente.

Tabla 6. Venta de ejemplares y valor de facturación por rango de tamaño

<b>Rango de tamaño</b>	<b>Ejemplares</b>	<b>%</b>	<b>Valor</b>	<b>%</b>
Macro	82,304,587	59.3	4,342,880,971	53.6
Grande	28,953,233	20.9	1,973,641,183	24.4
Mediana	14,536,673	10.5	1,138,500,056	14.1
Pequeña	8,657,203	6.2	439,394,719	5.4
Micro	4,237,332	3.1	200,592,325	2.5
<b>Total</b>	<b>138,689,028</b>	<b>100.0</b>	<b>8,095,009,254</b>	<b>100.0</b>

Por rangos de tamaño, las editoriales Macro participaron con 59% de los ejemplares y 53% del monto de venta.

La temática que más se vende es la de libros de educación básica, que aportan 45.1% de los ejemplares vendidos y 35.7% de la facturación, esto incluyendo las ediciones que se producen para el programa de gobierno de libro de texto gratuito de secundaria.

En segunda posición se encuentra la temática de Literatura, con una participación de 8% en los ejemplares. Sin embargo, a nivel de facturación, es la materia de Lenguas la segunda aportación más importante.

Tabla 7. Venta de ejemplares y valor de facturación por clasificación temática

<b>Temática</b>	<b>Ejemplares</b>	<b>%</b>	<b>Facturación</b>	<b>%</b>
Literatura infantil	9,367,691	6.8	284,714,235	3.5
Literatura juvenil	1,816,633	1.3	145,114,128	1.8
Bibliotecas de Aula y Escolares	7,331,217	5.3	142,670,222	1.8
Literatura	11,081,008	8.0	764,885,991	9.4
Lenguas	10,923,869	7.9	1,077,460,003	13.3
Religión	3,622,432	2.6	229,426,714	2.8
Filosofía y sicología	5,094,094	3.7	407,624,177	5.0
Artes, cultura y deportes	781,118	0.6	100,324,309	1.2
Ciencias puras	1,501,806	1.1	192,633,042	2.4
Tecnología, ciencias aplicadas	5,461,861	3.9	870,003,250	10.7
Ciencias sociales	4,659,117	3.4	426,002,110	5.3
Educación básica	34,768,403	25.1	1,999,346,129	24.7
Texto gratuito de secundaria	27,760,689	20.0	891,886,699	11.0
Geografía e historia	895,340	0.6	89,171,048	1.1
Generalidades	9,232,722	6.7	316,580,227	3.9
Otros	4,391,028	3.2	157,166,970	1.9
<b>Total</b>	<b>138,689,028</b>	<b>100.0</b>	<b>8,095,009,254</b>	<b>100.0</b>

Los programas de gobierno de bibliotecas de aula y escolares, así como el de libro de texto gratuito en secundaria, representan una cuarta parte de la venta total de ejemplares en el sector, aunque sólo 13% de la facturación total.

El principal canal de venta de ejemplares es el Gobierno, con 34.7% de la venta total, le sigue muy de cerca el canal de librerías con 30.1%. En cuanto al nivel de facturación, 43% del monto de venta proviene de librerías y en segunda posición se encuentra la venta a gobierno con 17.5%. Le sigue el canal de exportaciones en cuanto a ejemplares y respecto a facturación, la venta a escuelas es el siguiente con 10.2% del total.

Tabla 8. Venta de ejemplares y valor de facturación por canal de distribución

Canales de distribución	Ejemplares vendidos	%	Facturación	%
Ventas a gobierno	48,140,293	34.7	1,415,855,797	17.5
Librerías	41,704,640	30.1	3,488,322,665	43.1
Exportaciones	14,893,842	10.7	734,616,898	9.1
Tiendas de autoservicios y departamentales	11,012,886	7.9	572,831,199	7.1
Ventas a escuelas	9,821,607	7.1	822,842,523	10.2
Ventas a empresas privadas	2,972,251	2.1	227,124,141	2.7
Expendios propios	2,239,881	1.6	168,038,001	2.1
Otros canales	2,230,307	1.6	119,859,218	1.5
Puestos de periódicos	1,958,634	1.5	74,919,270	0.9
Ferias de libros	1,377,736	1.0	81,051,654	1.0
Ventas directas a crédito	1,094,396	0.8	145,994,252	1.8
Correo directo	577,121	0.4	146,186,052	1.8
Ventas a bibliotecas y centros de documentación	281,052	0.2	24,928,293	0.3
Otros eventos	186,331	0.1	18,256,333	0.2
Internet	128,033	0.1	44,943,744	0.6
Ventas a asociaciones de padres de familia	70,018	0.1	9,239,214	0.1
<b>Total</b>	<b>138,689,028</b>	<b>100.0</b>	<b>8,095,009,254</b>	<b>100.0</b>

### 1.2.3 Inversión

La inversión hecha en publicidad alcanzó el monto de 457,636,788 pesos, de la que 46% fue empleada por las editoriales de Texto y 31% por las de Interés general. Las empresas de Ciencia y Técnica utilizaron 23% de este total y la proporción correspondiente a editoriales religiosas ni siquiera alcanzó el 1% de la participación.

### 1.2.4 Empleo

El número de empleos generados por el sector fue de 8,222 puestos, de los que 92% fueron para empleados de base y 8% fueron contratos por tiempo determinado. En cuanto al empleo externo, se contrató a 3,012 personas para *free-lance* y 1,088 empresas.

Tabla 9. Número de plazas o contratos generados

<b>Subsector editorial</b>	<b>Base</b>	<b>Eventuales</b>	<b>Total</b>	<b>Personas</b>	<b>Empresas</b>	<b>Total</b>
Texto	3,127	227	3,354	1,279	169	1,449
Interés General	1,849	116	1,965	1,108	618	1,725
CTP	2,304	269	2,573	591	273	864
Religiosos	320	10	330	34	28	62
<b>Total</b>	<b>7,600</b>	<b>622</b>	<b>8,222</b>	<b>3,012</b>	<b>1,088</b>	<b>4,100</b>

Las empresas de Texto, junto con las de Ciencia y Técnica, fueron las que más empleados registraron en su nómina, con 3,127 y 2,304 personas respectivamente. Esto representa 71% de la nómina del sector. También estas empresas requirieron en mayor proporción de personal eventual, 80% fueron registrados por estas editoriales.

En servicios *free-lance*, nuevamente las editoriales de Texto contrataron la mayor cantidad de profesionales, y junto con la división de Interés general, aglomeran 79% de los contratos de *free-lance*.

En cuanto a servicios externos contratados, las editoriales de Interés general requirieron el apoyo de 57% de las empresas, mientras que las de Ciencia y técnica solicitaron 25%.

### 1.3 Definición de las variables de estudio.

Los datos utilizados en este ejercicio provienen de la encuesta que anualmente lleva a cabo la Cámara Nacional de la Industria Editorial Mexicana. En particular, se utilizaron las cifras del levantamiento de 2008.

El total de empresas en la muestra fue de 98. En el informe de la Actividad Editorial esta muestra se expande para representar el total del sector, pero en la elaboración de este análisis no se recurrió a este procedimiento.

Las variables utilizadas para llevar a cabo el análisis de conglomerados, fueron: número de ejemplares producidos por temática (16 temáticas distintas), número de ejemplares vendidos por canal de comercialización (15 canales de comercialización distintos), nivel de facturación, monto de inversión en publicidad, número de empleados de base, número de empleados eventuales, número de contratos *free-lance* con personas y número de contratos con empresas para servicios externos.

Tabla 10. Variables utilizadas en el análisis

Variable	clave	Descripción
<b>Producción por temática</b>		
Infantiles	p_infant	Número de ejemplares producidos
Juveniles	p_juv	Número de ejemplares producidos
Literatura	p_lit	Número de ejemplares producidos
Lenguas	p_leng	Número de ejemplares producidos
Religión	p_rel	Número de ejemplares producidos
Filosofía y psicología	p_fil	Número de ejemplares producidos
Arte cultura y deportes	p_arte	Número de ejemplares producidos
Ciencias puras	p_cpuras	Número de ejemplares producidos
Tecnología y ciencias aplicadas	p_tec	Número de ejemplares producidos
Ciencias sociales	p_csoc	Número de ejemplares producidos
Educación básica	p_ed_bas	Número de ejemplares producidos
Geografía e historia	p_geog	Número de ejemplares producidos
Generalidades	p_gen	Número de ejemplares producidos
Otros	p_otro	Número de ejemplares producidos
Libros de texto de secundaria (Programa de texto gratuito)	p_sec	Número de ejemplares producidos
Bibliotecas Infantiles y juveniles (Programa de Bibliotecas)	p_bib	Número de ejemplares producidos

Venta por canal		
Librerías	libreria	Número de ejemplares vendidos
Puestos de periódicos	puesto_per	Número de ejemplares vendidos
Tiendas de autoservicio y departamentales	autoserv	Número de ejemplares vendidos
Ferias de libros	ferias	Número de ejemplares vendidos
Ventas a gobierno	gobierno	Número de ejemplares vendidos
Ventas a escuelas, colegios, universidades	escuelas	Número de ejemplares vendidos
Ventas a bibliotecas y Centros de documentación	bibliotecas	Número de ejemplares vendidos
Ventas a empresas privadas	empresas	Número de ejemplares vendidos
Ventas a Asociaciones de Padres de familia o similares	asociac	Número de ejemplares vendidos
Ventas directas a crédito (puerta a puerta)	credito	Número de ejemplares vendidos
Correo directo	correo	Número de ejemplares vendidos
Expendios propios (puntos de venta de la editorial)	expendio	Número de ejemplares vendidos
Internet	internet	Número de ejemplares vendidos
Otros eventos, otros canales	otros	Número de ejemplares vendidos
Exportaciones	export	Número de ejemplares vendidos
Otras variables		
Facturación	fact	Monto de facturación por la venta de libros (pesos)
Publicidad	inversion	Monto de la inversión hecha en publicidad (pesos)
Empleo Base	emp_base	Número de empleados con contrato de nómina
Empleo eventual	emp_event	Número de empleados con contrato eventual
Personas <i>free lance</i>	pers_free	Número de personas con contrato de <i>free-lance</i>
Empresas	empre_free	Número de empresas contratadas para servicios externos.

## 2. ANÁLISIS DE CONGLOMERADOS

El análisis de conglomerados es una técnica exploratoria utilizada para generar clasificaciones de observaciones multivariadas mediante la agrupación de los elementos en función de sus similitudes o diferencias.

Con esta técnica se pueden analizar las siguientes situaciones:

**Partición de datos.** Se tiene un conjunto de observaciones multivariadas que se presumen son heterogéneas y se desea dividir las en conjuntos tales que: cada observación pertenezca a uno y sólo uno de los grupos, todas las observaciones queden clasificadas y cada grupo debe ser, internamente, homogéneo

**Construcción de jerarquías.** Se desea estructurar las observaciones de un conjunto de forma jerárquica por su similitud. Esta clasificación implica que las observaciones se ordenan por niveles, de manera que los niveles superiores contienen a los inferiores. Estos métodos definen la estructura de asociación en cadena que pueda existir entre los elementos y la jerarquía obtenida permite generar una partición de datos.

**Clasificación de variables.** El análisis de conglomerados también se puede realizar sobre las variables de un problema con el objetivo de dividir las en grupos cuando éstas son demasiadas. Esto sirve de orientación para plantear modelos que reduzcan la dimensión de las variables.

Al momento de llevar a cabo un análisis de conglomerados es necesario tomar varias decisiones que tienen que ver con el procedimiento del análisis:

**Selección de las variables.** Las variables deben seleccionarse adecuadamente para proporcionar tanta información como sea posible de los datos observados.

**Medida de proximidad.** Es la medida que cuantifica qué tan similares o diferentes son dos vectores de observaciones.

**Algoritmo de aglomeración.** Una vez adoptada la medida de proximidad, se selecciona el algoritmo específico que revele la estructura de conglomeración dentro de los datos. [Theodoridis; Koutroumbas 2003].

## 2.1 MEDIDAS DE PROXIMIDAD

Las medidas de proximidad se utilizan para representar la cercanía de dos observaciones. Dada la matriz de observaciones  $X(n,p)$  con mediciones de  $p$  atributos en  $n$  objetos:

$$X = \begin{pmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1p} \\ x_{21} & x_{22} & x_{23} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & x_{n3} & \dots & x_{np} \end{pmatrix} = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_n \end{pmatrix}$$

se puede cuantificar la proximidad o similitud que se presenta entre éstos. [Timm, 2002]

La matriz de proximidad  $D_{n \times n}$ , se encuentra definida por:

$$D = \begin{pmatrix} d_{11} & d_{12} & \dots & d_{1n} \\ d_{21} & d_{22} & \dots & d_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1} & d_{n2} & \dots & d_{nn} \end{pmatrix}$$

donde las  $d_{ij}$  representan la medida de proximidad entre los vectores de observaciones  $x_i$  y  $x_j$ . Esta matriz constituye la base para la formación de los conglomerados.

Estas medidas pueden ser de similitud (el valor de la medición se incrementa cuando las observaciones son más semejantes), o disimilitud (el valor de la medición decrece cuando dos observaciones son más similares). [Härdle, Simar 2007]

### 2.1.1 Medidas de disimilitud

Una matriz de disimilitudes es una matriz  $D_{n \times n}$  tal que para todo objeto  $i, j, k$ :

$$\begin{aligned} \rightarrow d_{i,j} &\geq 0 \\ \rightarrow d_{i,i} &= 0 \\ \rightarrow d_{i,j} &= d_{j,i} \end{aligned}$$

[Timm, 2002] La medida de disimilitud depende del tipo de escala en el que se miden los datos, las más comunes para variables continuas se presentan a continuación:

→ Distancia euclídea

$$d_{i,j} = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{ip} - x_{jp})^2}$$

→ Distancia de Manhattan

$$d_{i,j} = |x_{i1} - x_{j1}| + |x_{i2} - x_{j2}| + \dots + |x_{ip} - x_{jp}|$$

→ Distancia de Minkowski. Es la generalización de las dos distancias anteriores

$$d_{i,j} = (|x_{i1} - x_{j1}|^q + |x_{i2} - x_{j2}|^q + \dots + |x_{ip} - x_{jp}|^q)^{1/q}$$

para  $q=2$  se tiene la distancia euclidiana, con  $q=1$  se obtiene la distancia de Manhattan.

Cuando las variables que se usarán para el cálculo de la disimilitud son mediciones en diferentes unidades, lo más conveniente es estandarizarlas para evitar que las diferencias de las escalas no influyan en los cálculos de la distancia, con esto se persigue que todas las variables tengan un mismo peso. Sin embargo, es posible que algunas variables sean más importantes que otras en un contexto particular o que las unidades de medición se consideren significativas para el análisis, por lo que es probable que sea mejor no estandarizar los datos [Kaufman, Rousseeuw, 1990].

Otra forma de construir disimilitudes es designando calificaciones subjetivas de qué tanto difieren entre sí ciertos objetos, desde el punto de vista de uno o varios observadores. Este tipo de datos se encuentran típicamente en las ciencias sociales y en la mercadotecnia.

Una manera distinta de calcular las medidas de disimilitud entre variables es utilizando el coeficiente de correlación entre dos objetos, este coeficiente puede ser convertido en una medida de disimilitud con la siguiente fórmula:

$$d_{i,j} = \frac{1 - \text{corr}(i,j)}{2}$$

Esta magnitud tiene un rango de valores entre 0 y 1, las variables con una alta correlación positiva recibirán un coeficiente de disimilitud cercana a 0, mientras que un coeficiente con una fuerte correlación negativa estará cerca de 1 y las observaciones se consideraran muy diferentes entre sí. [Kaufman, Rousseeuw, 1990].

### 2.1.2 Medidas de similitud

También se pueden utilizar coeficientes de similitud en la creación de conglomerados, éstos indican qué tanto se parecen dos objetos entre sí. Esta medida toma valores entre 0 y 1, donde 0 significa que no hay semejanza entre los dos objetos y 1 refleja el máximo nivel de similitud. El coeficiente cumple con las siguientes condiciones:

$$\rightarrow 0 \leq s_{i,j} \leq 1$$

$$\rightarrow s_{i,i} = 1$$

$$\rightarrow s_{i,j} = s_{j,i}$$

Las medidas  $s_{i,j}$  se arreglan en una matriz de  $n \times n$  llamada matriz de similitud. Como en el caso de las disimilitudes, los coeficientes pueden ser resultados de juicios subjetivos.

Para calcular las similitudes entre dos objetos se puede recurrir, nuevamente, al coeficiente de correlación, aunque no puede ser utilizado directamente como medida de similitud ya que éste también toma valores negativos. Para que el coeficiente se mantenga entre 0 y 1 se puede utilizar la siguiente transformación:

$$s_{i,j} = \frac{1 + corr_{i,j}}{2}$$

[Kaufman, Rousseeuw, 1990]. Cuando se tienen vectores con variables medidas en diferentes escalas, lo mejor es adoptar medidas de proximidad adecuadas para variables continuas ya que los vectores con valores discretos pueden ser comparados de manera exacta con medidas para variables continuas, mientras que lo contrario no suele ser así. [Theodoridis, Koutroumbas, 2003]

## 2.2 ALGORITMOS DE CONGLOMERACIÓN

Esencialmente hay dos tipos de algoritmos de conglomeración que forman grupos excluyentes: jerárquicos y no jerárquicos (de partición).

Los algoritmos jerárquicos se pueden dividir en aglomerativos y divisivos. En los métodos aglomerativos se inicia con la partición más fina posible (cada observación es un conglomerado) y se van agrupando, mientras que los métodos divisivos inician con un solo conglomerado que contiene a todas las observaciones y el procedimiento va dividiendo este conjunto. [Härdle, Simar, 2007]

En los métodos no jerárquicos se dividen las observaciones en  $k$  conglomerados, lo cual puede hacerse con una partición inicial, y después se van reasignando las observaciones de acuerdo a un criterio de optimización. [Rencher, 2002]

La diferencia entre los algoritmos jerárquicos y los algoritmos de partición es que, en los primeros una vez que se ha asignado una observación a un conglomerado en particular ésta no puede ser cambiada; mientras que en los segundos, la asignación de las observaciones en los conglomerados puede ir cambiando durante el procedimiento. [Härdle, Simar, 2007]

### 2.2.1 Métodos jerárquicos

Los procedimientos jerárquicos son métodos muy utilizados para resumir la estructura de las observaciones. Las particiones se pueden visualizar a través del dendrograma o diagrama de árbol, que al seccionarlo en un nivel en particular produce una partición de  $k$  grupos.

#### Métodos aglomerativos

En los métodos aglomerativos, una observación ó un conjunto de observaciones se unen en un conglomerado en función de las disimilitudes entre éstos. Para medir estas disimilitudes se pueden utilizar diversos criterios y cada uno puede generar conglomerados diferentes. En cada paso del proceso se calculan las disimilitudes entre cada par de conglomerados y se unen los dos que sean más similares (la disimilitud más pequeña). [Webb, 2002]

Sean  $A$  y  $B$  los conglomerados que se van a combinar, los métodos más comunes para calcular las nuevas disimilitudes son:

→ Vecino más cercano. Este método es invariante ante transformaciones monótonas, es decir, se obtienen las mismas jerarquías aunque las distancias sean numéricamente distintas. Este procedimiento tiende a producir conjuntos alargados y puede incluir elementos muy diferentes en los extremos. Las nuevas disimilitudes se calculan con la siguiente fórmula:

$$d_{A,B} = \min_{\substack{i \in A \\ j \in B}} d_{i,j}$$

→ Vecino más lejano. Este criterio también es invariante ante transformaciones monótonas y tiende a producir conjuntos esféricos. La fórmula para calcular la nueva disimilitud es:

$$d_{A,B} = \max_{\substack{i \in A \\ j \in B}} d_{i,j}$$

→ Media de grupos. En este método se promedian las mediciones de disimilitud de cada conglomerado, por lo que este criterio no es invariante ante transformaciones monótonas. La fórmula para calcular las nuevas disimilitudes es:

$$d_{A,B} = \frac{1}{n_A n_B} \sum_{\substack{i \in A \\ j \in B}} d_{i,j}$$

Los siguientes criterios comienzan a partir de la matriz de observaciones original:

→ Método del centroide. Este método es adecuado para variables de intervalo (ó razón). La distancia entre los conglomerados se define como la distancia euclidiana entre los vectores de medias de éstos (llamados centroides):

$$d_{A,B} = d_{\bar{y}_A, \bar{y}_B}$$

Donde  $\bar{y}_A$  y  $\bar{y}_B$  son los vectores de medias de  $A$  y de  $B$ . Después de que se unen los conglomerados  $A$  y  $B$ , el centroide del nuevo conglomerado  $AB$  esta dado por el promedio:

$$\bar{y}_{AB} = \frac{n_A \bar{y}_A + n_B \bar{y}_B}{n_A + n_B}$$

→ Método de Ward. Este método utiliza las distancias (al cuadrado) dentro de los conglomerados y las distancias (al cuadrado) entre los conglomerados. El objetivo es minimizar la suma total de cuadrados dentro de los grupos. Si  $AB$  es el conglomerado obtenido al combinar los conjuntos  $A$  y  $B$ , entonces la suma de las distancias dentro de los conglomerados es:

$$SSE_A = \sum_{i=1}^{n_A} (y_i - \bar{y}_A)'(y_i - \bar{y}_A)$$

$$SSE_B = \sum_{i=1}^{n_B} (y_i - \bar{y}_B)'(y_i - \bar{y}_B)$$

$$SSE_{AB} = \sum_{i=1}^{n_{AB}} (y_i - \bar{y}_{AB})'(y_i - \bar{y}_{AB})$$

Donde  $y_i$  es el vector de observaciones  $i$ ,  $\bar{y}_A$ ,  $\bar{y}_B$  y  $\bar{y}_{AB}$  son los vectores de medias de  $A$ , de  $B$ , y de  $AB$ .

En este método se unen los dos conglomerados  $A$  y  $B$  que minimicen el incremento en la suma de los cuadrados definida como:

$$I_{AB} = SSE_{AB} - (SSE_A + SSE_B)$$

[Rencher 2002; Webb,2002]

## Métodos divisivos

Los métodos divisivos trabajan en sentido opuesto a los aglomerativos, es decir, inician con un único conjunto (el total de las observaciones) y lo dividen hasta tener un conglomerado por cada elemento de la muestra. [Kaufman, Rousseeuw, 1990]

### 2.2.2 Métodos de partición (no jerárquicos)

Los métodos de partición construyen  $k$  conglomerados, es decir, clasifican a los datos en  $k$  conjuntos distintos que satisfacen lo siguiente:

→ Cada conjunto contiene al menos un elemento. Esto significa que, a lo más, hay tantos conjuntos como elementos.

- Cada elemento debe pertenecer exactamente a un conjunto. Esto significa que dos conglomerados distintos no pueden tener ningún elemento en común. Además, la unión de los  $k$  conjuntos genera el conjunto total.

El número  $k$  de conglomerados es determinado por el investigador. Estos algoritmos intentan encontrar una “buena partición” en el sentido de que los elementos de un mismo conglomerado deben estar cerca o ser parecidos unos a otros, mientras que los objetos que se encuentran en conglomerados distintos tienen que estar alejados o ser muy diferentes. [Kaufman, Rousseeuw, 1990]

Una cualidad general de los algoritmos de partición es que permiten que los elementos se muevan de un conglomerado a otro, lo que no es posible con los métodos jerárquicos. [Rencher 2002]

### **Método de las k-medias**

Un algoritmo de partición muy utilizado es el de  $k$ -medias, en éste se seleccionan  $k$  individuos que sirven como semilla y son reemplazados después por los centroides de los conglomerados finales.

La semilla puede seleccionarse de manera aleatoria, o eligiendo los primeros  $k$  puntos en el conjunto de datos, o escogiendo los  $k$  puntos más alejados o encontrar los  $k$  puntos de mayor densidad. En estos casos, el número de  $k$  de conglomerados debe ser especificado.

Una vez que las semillas han sido determinadas, cada uno de los puntos restantes se asignan al conglomerado más cercano (que la distancia entre la observación y la semilla sea mínima) y en seguida se actualiza la semilla, que pasa a ser el centro del conglomerado (centroide).

Al momento en que todas las observaciones han sido asignadas a algún conjunto, cada observación es examinada para ver si es que se encuentra más cerca del centro de otro conglomerado que no es el propio, si esto ocurre, el elemento se reasigna y se vuelven a calcular los centroides. Este proceso continúa hasta que no es posible mejorar. [Rencher 2002]

### **Método de los k-medoides**

La técnica de  $k$ -medias tiene la desventaja de ser sensible a las observaciones extremas, por ejemplo cuando el conjunto de observaciones más cercanas a alguna semilla dada está vacío y por lo tanto los centroides no se van actualizando. En este caso es mejor alternativa la técnica de los  $k$ -medoides.

En el algoritmo de los  $k$ -medoides se obtiene una partición inicial seleccionando sucesivamente objetos representativos (elementos cuya suma de disimilitudes al resto de los elementos es la más pequeña) hasta que los  $k$  elementos son encontrados. Estas

observaciones son llamadas medoides. Una vez encontrado el conjunto de  $k$  medoides, los conglomerados se construyen asignando cada observación al objeto representativo más cercano. [Kaufman, Rousseeuw, 1990]

### **2.3 Interpretación de los resultados**

La selección de diferentes variables, medidas de proximidad, criterios de conglomeración y algoritmos pueden resultar en conglomerados totalmente diferentes, por lo que es aconsejable que, una vez terminado el proceso de conglomeración, los resultados sean revisados por un experto en el tema específico que se está tratando en el análisis, y permitir que éste determine cuál es el mejor conglomerado y que participe en la interpretación del resultado, pues la experiencia y conocimiento previo del experto enriquecerán el análisis final [Theodoridis, Koutroumbas, 2003].

### 3. RESULTADOS Y DISCUSIÓN

El análisis se aplicó a las 98 empresas editoriales que componen la muestra del Estudio de producción y comercialización del libro en México 2008. Se llevaron a cabo varios ejercicios, con distintas variables y distintas técnicas. El que se expone aquí es el que se consideró como el mejor resultado, se trata de un análisis con algoritmo jerárquico aglomerativo con liga de Ward.

Todas las variables utilizadas se estandarizaron para evitar la influencia de la escala de medición en la conformación de los conglomerados. El programa utilizado para realizar el análisis fue SPSS versión 17.0.

El número de estratos que se maneja en la metodología del CERALC es de 20 (5 rangos de tamaño por 4 divisiones editoriales), sin embargo para este ejercicio, 20 conglomerados se consideraron demasiados, por lo que se optó por tomar la mitad de conjuntos, lo cual además coincide con la fórmula utilizada por [Loureriro, Torgo 2004] para determinar un número de conglomerados que permita que las observaciones atípicas queden aisladas. La fórmula es la siguiente:

$$nc = \max\{2, n/10\}$$

donde  $n$  es el número de observaciones. En el caso de este ejercicio:

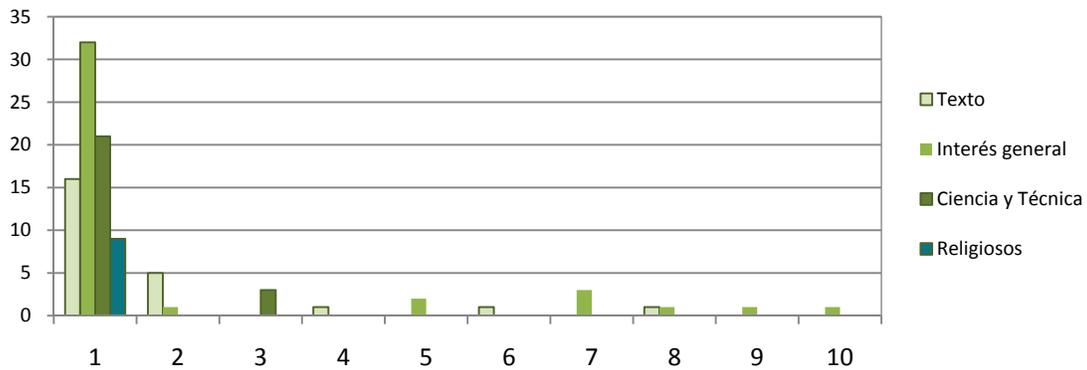
$$nc = \max\{2, 98/10\} = \max\{2, 9.8\} = 9.8$$

La distribución de las editoriales en los conglomerados se muestra en la Tabla 11:

Tabla 11. Distribución de las editoriales en 10 conglomerados

Conglomerado	1	2	3	4	5	6	7	8	9	10	Total
Número de elementos	78	6	3	1	2	1	3	2	1	1	98

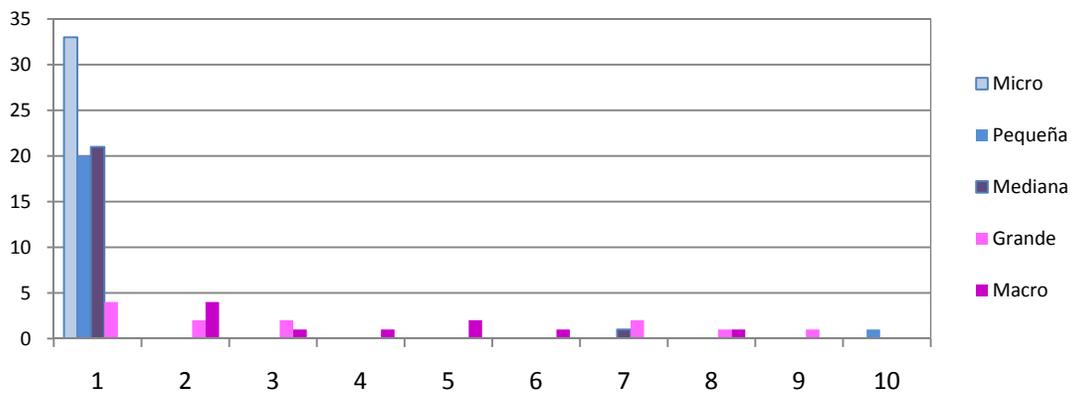
**Gráfico 1.** Distribución de las editoriales en los 10 conglomerados  
Por subsector editorial



En el gráfico 1 se encuentra la distribución de las editoriales dentro los 10 conglomerados desagregadas por división editorial, que es una de las variables que utiliza el CERLALC para estratificar el universo de editoriales.

En general parece no haber alguna segmentación representativa de la división editorial ya que en el conglomerado 1, donde se concentran la mayoría de las observaciones, se agruparon editoriales de las 4 categorías existentes. Aún así, se puede resaltar que todas las editoriales de Religiosos quedaron exclusivamente dentro del conglomerado 1.

**Gráfico 2.** Distribución de las editoriales en los 10 conglomerados  
Por rango de tamaño



En el gráfico 2 se presenta la distribución de las editoriales en los 10 conglomerados, desagregadas por el rango de tamaño. A diferencia del gráfico anterior, en éste se nota un perfil más definido en los conglomerados: todas las empresas de tamaño micro se agruparon en el conglomerado 1, lo mismo que todas las empresas pequeñas y medianas, con excepción de dos casos (una pequeña y una mediana), mientras que ninguna empresa de tamaño macro quedó dentro del conglomerado 1.

### **3.1 Descripción de los conglomerados.**

#### **Conglomerado 1**

El promedio de facturación de las editoriales en este conglomerado es de 22.8 millones de pesos, que después del registrado en el conglomerado 10 es el menor. Esto se debe a que en éste se agruparon todas las empresas de tamaño micro y pequeña. La producción promedio más alta corresponde a la clasificación de otras temáticas. El promedio de ventas en librerías es de 130 mil ejemplares, el más bajo de entre los 10 conglomerados, lo mismo que el promedio de venta a gobierno, el extranjero y los otros canales. La inversión promedio que se hace en este conglomerado apenas supera el millón de pesos y después del conglomerado 10, es el más bajo. La media de empleados contratados ya sea por nómina o eventuales es de 36, también el segundo más bajo.

#### **Conglomerado 2**

En este conglomerado se agruparon 6 empresas de tamaño grande o macro. La facturación promedio por editorial es de 277 millones de pesos al año, la producción promedio más alta es de libros educativos y de programas de gobierno, en ambos casos los tirajes promedios superan los 2 millones de ejemplares, mientras que para las ediciones de otras temáticas apenas alcanza los 182 mil ejemplares. La venta en librerías supera el millón de piezas, pero es menor que la venta promedio que estas editoriales hacen al gobierno. La inversión en publicidad es de 19.6 millones de pesos por editorial y es la tercera cantidad más alta.

#### **Conglomerado 3**

En este conglomerado se agruparon 3 empresas de Ciencia y Técnica con una facturación promedio de 286 millones de pesos, el promedio de producción por empresa más alto se observa en las ediciones de otras temáticas (ciencias puras, tecnología, ciencias sociales) con 630 mil ejemplares. El promedio de venta de ejemplares más alto se muestra en el canal de librerías con 1.3 millones de piezas, en seguida se presenta la venta en el extranjero con más de medio millón de ejemplares. Se advierte que el número promedio de ejemplares vendidos es superior al número de ejemplares producidos, esto puede deberse a la cantidad de importaciones que hacen estas empresas.

Tabla 12. Valores promedio por conglomerado con variables agrupadas

Conglomerado	Facturación	Producción de ejemplares			Venta de ejemplares por canal de comercialización				Inversión	Nómina-eventual	Freelance - empresas
		Educativos	Gobierno	Otra temática	Librería	Venta Gobierno	Extranjero	Otros			
1	22,847,676	76,215	77,896	114,700	130,745	77,896	19,467	88,700	1,057,362	36	15
2	277,563,177	2,251,461	2,366,565	182,564	1,157,140	2,366,565	415,382	632,989	19,645,249	204	81
3	286,437,423	418,009	397,340	630,727	1,327,784	397,340	674,717	443,002	16,085,248	272	105
4*	> 317,916,705	63.5%	25.9%	10.6%	42.2%	37.2%	0	20.6%	3,608,933	727	145
5	317,916,705	0	163,055	5,634,653	1,824,528	163,055	1,178,015	2,058,435	16,781,661	130	131
6*	> 317,916,705	67.0%	32.9%	0.1%	9.1%	49.7%	1.9%	39.4%	37,844,010	340	337
7	95,324,015	16,656	140,194	1,111,358	740,800	140,194	430,081	637,280	3,653,064	19	127
8	251,058,181	3,525,892	1,053,474	1,996,918	2,287,040	1,053,474	1,138,169	2,986,396	7,014,179	146	5
9	160,447,287	0	0	438,605	0	0	47,500	443,788	39,777,851	88	86
10	18,750,000	279,098	0	1,754,983	841,509	0	120,152	489,216	127,000	52	13

Libros educativos: Producción de ejemplares de educación básica y producción de ejemplares de lenguas

Gobierno: Producción de ejemplares para programas de Bibliotecas de Aula y Escolares y Programa de Texto gratuito en secundaria

Otra temática: Producción de libros Infantiles, Juveniles, Filosofía y religión, Artes cultura y deportes, Ciencias puras, Tecnología y ciencias aplicadas, Ciencias sociales, Geografía e historia, Generalidades y Otros.

Ventas a gobierno: En este particular caso la venta de ejemplares a gobierno iguala a la producción promedio de ejemplares producidos para los programas de gobierno.

Otros canales de comercialización: Puestos de periódicos, tiendas departamentales y de autoservicio, ferias de libros, escuelas y universidades, bibliotecas, empresas privadas, asociaciones de padres de familia, correo directo, expendios propios, internet.

\* Por razones de confidencialidad el valor de facturación y el número de ejemplares producidos y vendidos no se presentan en este cuadro. Los porcentajes se refieren a la distribución de los ejemplares producidos por temática y los ejemplares vendidos por canal.

## **Conglomerado 5**

En el conglomerado 5 se encuentran 2 editoriales que facturan en promedio 317 millones de pesos cada una y que además, no producen libros educativos, éste junto con el conglomerado 9, son los únicos que no registran ediciones de esta temática. La participación de estas editoriales en los programas de gobierno es muy pequeña (163 mil ejemplares). La mayor producción promedio de ejemplares en este conglomerado corresponde a otras temáticas, como literatura.

La venta promedio de ejemplares más numerosa se realiza a través de los otros canales, como tiendas departamentales, de autoservicio o incluso puestos de periódicos. Las librerías y las exportaciones también tienen una participación importante, de hecho, en este conglomerado se presenta el promedio de ejemplares exportados más alta de entre los 10 conglomerados. El número de empleados se encuentra nivelado entre los de contrato y los *freelance*: 130 y 131 respectivamente.

## **Conglomerado 7**

En este conglomerado la facturación media de las editoriales es de 95.3 millones de pesos. En promedio, la mayor producción es de ediciones de temáticas distintas a las de educación básica (filosofía y psicología, libros Infantiles), por lo que los canales a través de los cuales se venden más ejemplares son el de librerías y el de otros (tiendas de autoservicio o departamentales), aunque en ninguno de estos canales se supera el millón de ejemplares vendidos. El valor de inversión promedio es de 3.6 millones de pesos únicamente y en este conglomerado se registra, además, el promedio más bajo de empleados contratados: 19.

## **Conglomerado 8**

El promedio de facturación de las empresas en este conglomerado es de 251 millones de pesos, la producción promedio más alta corresponde a los libros educativos con 3.5 millones de ejemplares por editorial. Estas empresas también participan en los programas de gobierno con una producción de más de 1 millón de ejemplares como media. El canal de comercialización más importante para estas editoriales es el de tiendas departamentales y de autoservicio (otros canales), en el que se venden en promedio 2.9 millones de ejemplares, mientras que en librerías se comercializa 2.2 millones. La inversión que estas editoriales hacen en publicidad es de 7 millones de pesos. En este conglomerado se registran el menor número de empleados *free-lance* con 5.

## **Conglomerado 4, Conglomerado 6, Conglomerado 9, Conglomerado 10.**

En cada uno de estos conglomerados aparece segmentada una sola empresa de tamaño grande o macro, con excepción de la empresa del conglomerado 10.

En el **conglomerado 4** se encuentra una editorial de tamaño macro cuya producción principal son los libros educativos (63%), otro 26% son ediciones para los programas de gobierno y un 11% de la producción de esta empresa son ediciones de otras temáticas (libros juveniles, tecnología o ciencias sociales). En este conglomerado se presenta el segundo valor de facturación más alto, después del registrado en el conglomerado 6.

El canal más importante en este conglomerado es el de librerías, a través del cual se venden 42% de los ejemplares, además esta cifra de venta en librerías es la más alta de entre los 10 conglomerados. Este conjunto es el único en el que no se observa actividad de exportación, además tiene el promedio más alto de empleados en nómina y eventuales: 727.

En el **conglomerado 6** se presentan los valores más altos de facturación, producción de libros educativos y para gobierno, mientras que la producción de ejemplares de otra temática es la menor, de entre los 10 conglomerados. En éste se observa que casi la totalidad de ejemplares producidos son para la enseñanza básica. El principal canal de venta de este conglomerado es el gobierno a través del cual se venden aproximadamente 50% de los ejemplares. El siguiente canal en importancia, para este conjunto, es el de otros canales con 39% de los ejemplares vendidos (principalmente escuelas). La inversión que se hace en publicidad es de 37.8 millones, y es el segundo valor más alto, además de que en este conglomerado se encuentra el valor máximo de *freelance* y de empresas contratadas para servicios externos.

En el **conglomerado 9** se encuentra una empresa que también resulta muy particular. Por su nivel de facturación se encuentra clasificada en el rango de grande, la única temática de producción y de venta es la de literatura (otras temáticas), el principal canal de comercialización es otro distinto a la librería y al gobierno, de hecho, este conglomerado es el único en el que no hay venta a gobierno ni en librerías, además la inversión en publicidad es la más alta de todos los conglomerados: 39.7 millones de pesos.

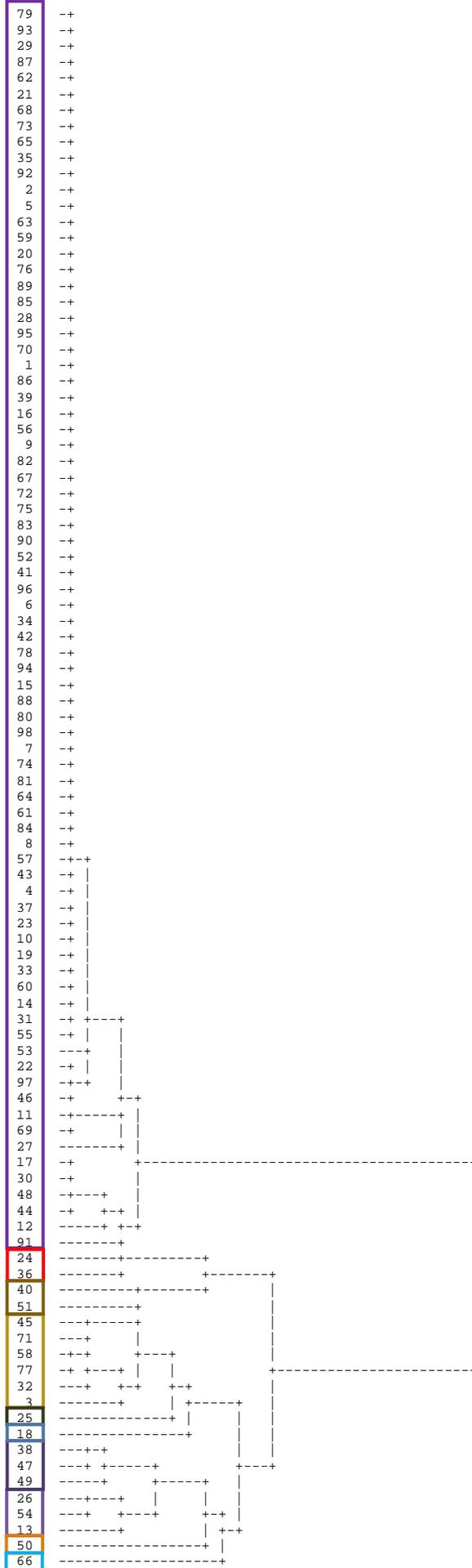
En el **conglomerado 10** se encuentra una empresa de interés general y de tamaño pequeña, y es la única editorial con este rango de tamaño que se encuentra fuera del conglomerado 1. La principal temática de producción en este conglomerado son los libros de psicología y filosofía. El canal de comercialización más importante de este conglomerado es el de librerías, además de otros canales (puestos de periódicos y tiendas de autoservicio y departamentales).

En el gráfico 3 se presenta el dendrograma resultante y los distintos conglomerados son señalados con cuadros de colores. En este gráfico se puede observar que la distancia a la que se unen los elementos del conglomerado 1 es la más pequeña, lo que evidencia una gran homogeneidad.

También se puede observar que la distancia a la que se integra la observación 66 a algún conglomerado es la más grande, lo que se puede interpretar como que es la más diferente. Esta observación corresponde a la segmentada en el conglomerado 10: empresa de interés general y de tamaño pequeña.

Rescaled Distance Cluster Combine

C A S E 0 5 10 15 20 25  
 Label Num +-----+-----+-----+-----+-----+



- Conglomerado 1
- Conglomerado 2
- Conglomerado 3
- Conglomerado 4
- Conglomerado 5
- Conglomerado 6
- Conglomerado 7
- Conglomerado 8
- Conglomerado 9
- Conglomerado 10

Con el resultado obtenido se puede inferir que la variable más importante en la aglomeración de las editoriales es la facturación anual, ya que en un solo conglomerado fueron agrupadas 90% de éstas, cada una de las cuales factura en promedio 22.8 millones de pesos al año, esta cifra apenas alcanza el valor de empresas de tamaño mediana. Aunado a esto, se encuentra que los promedios de empleo y publicidad son bajos, es decir, en este conglomerado permanecieron las empresas más pequeñas, no solamente en cuanto a facturación, sino también en cuanto a producción, venta y empleo.

Los restantes conglomerados son empresas macro o grande que se diferencian, además del nivel de facturación, por la producción temática predominante y por el canal de comercialización principal.

Tabla 13. Principal temática producida y principal canal de venta por conglomerado

Conglomerado	Facturación promedio (millones)	Temática predominante en la producción	Canal de venta principal
1	22.8	Religiosos, filosofía, técnicos	Librería
2	277.5	Gobierno	Gobierno
3	286.4	Ciencias puras, tecnología, ciencias sociales	Librería
4	>317.9	Educativos	Librería
5	317.9	Literatura, infantiles, filosofía	Librería, exportación
6	>317.9	Educativos, gobierno	Gobierno, escuelas
7	95.3	Literatura, filosofía	Librería, autoservicios
8	251.0	Educativos, infantiles	Librería, autoservicios
9	160.4	Literatura	Otro canal
10	18.7	Filosofía, entretenimiento	Librería

Cabe hacer especial mención de los conglomerados 4 y 6, ya que en ambos se producen y venden, en su mayoría, libros de texto, también en estos conglomerados la producción y venta de ediciones para gobierno participan de manera significativa. Sin embargo, mientras que en la empresa del conglomerado 4 el 10% de la producción es de otra temática y su venta principal se hace a librerías, en la empresa del conglomerado 6 la edición es casi exclusivamente de libros escolares para nivel básico (primaria y secundaria) y aproximadamente la mitad de su venta es para gobierno. Una diferencia importante es también la inversión en publicidad, ya que la editorial del conglomerado 4 invierte sólo 3.6 millones, mientras que la del conglomerado 6 invierte 37.8 millones. Estas diferencias son significativas si se considera que ambas empresas se encuentran en el mismo estrato (Texto-Macro) según la definición del CERLALC.

Debido al gran tamaño del conglomerado número 1, se procedió a hacer un segundo análisis con las 78 editoriales que quedaron dentro de éste, utilizando las mismas variables, el mismo algoritmo y la misma liga, con lo cual se obtiene la siguiente distribución de observaciones en 10 conglomerados:

Tabla 14. Análisis de conglomerados número 2  
Distribución de las 78 observaciones en el primer conglomerado

Subsector editorial	Conglomerados									
	1	2	3	4	5	6	7	8	9	10
<b>Micro</b>	32				1					
<b>Pequeña</b>	14	5		1						
<b>Mediana</b>	6	6	1		3	2	1		1	1
<b>Grande</b>		2	1					1		
<b>Total</b>	52	13	2	1	4	2	1	1	1	1

De nuevo el primer conglomerado es el más grande (66% de las editoriales) y contiene a las editoriales más pequeñas, mientras que todas las más grandes (rango de tamaño Grande) quedan fuera de éste. Tomando en cuenta esto, se hizo otro análisis de conglomerados utilizando las mismas variables, el mismo algoritmo, la misma liga, pero considerando 5 conglomerados y exclusivamente a las empresas pequeñas y micro.

Tabla 15. Análisis de conglomerados número 3  
Distribución de las 54 editoriales de tamaño pequeña o micro

Subsector editorial	Conglomerados				
	1	2	3	4	5
<b>Micro</b>	33				
<b>Pequeña</b>	15	3	1	1	1
<b>Total</b>	48	3	1	1	1

De nueva cuenta se presenta un primer conglomerado que contiene la mayor cantidad de elementos (88%) y el resto de los conglomerados con 3 o 1 observaciones en cada uno. Las editoriales más chicas (en este caso las de tamaño micro) son las que quedan en un solo conglomerado.

Estos resultados son un indicativo de que entre las empresas de rango de tamaño pequeño y micro hay un alto grado de homogeneidad, especialmente a nivel de facturación mientras que, en sentido contrario, las empresas que quedaron distribuidas en los restantes 9 conglomerados (del primer análisis) muestran un alto grado de heterogeneidad y se pueden señalar como observaciones atípicas dentro de la muestra de la Actividad Editorial.

Como se comentó al inicio del capítulo, se realizaron varios análisis con diversos acercamientos. Dos de estos ejercicios, presentados en los anexos, fueron realizados bajo la premisa de reducir el número de variables y de obtener conglomerados mejor distribuidos.

En el Anexo II se presenta un análisis de componentes principales del cual se tomaron las variables más importantes del primer componente y se utilizaron para generar los conglomerados. En el Anexo III se redujeron las variables desde un punto de vista conceptual al agrupar la producción por temática en una sola variable, mientras que las 15 variables de canales de comercialización se agruparon en cuatro categorías.

Aún cuando los conglomerados obtenidos en estos ejercicios no fueron mejores, cabe resaltar lo siguiente:

- En el Anexo II la variable con mayor importancia dentro del primer componente es el valor de facturación, con una correlación de 0.911, lo cual refuerza la suposición de que el nivel de facturación es la variable principal en la aglomeración de las editoriales.
- En el Anexo III se evidencia la importancia de considerar la clasificación temática dentro del análisis, al compararse el dendrograma del análisis principal (37 variables) con el de este anexo (9 variables).

## 4. CONCLUSIONES

De acuerdo al objetivo de este trabajo, se preparó un análisis de conglomerados para determinar las agrupaciones naturales de las editoriales mexicanas utilizando como variables el nivel de facturación, el número de ejemplares producidos por temática, número de ejemplares vendidos por canal de comercialización, inversión en publicidad, número de empleados de base, eventuales, *free-lance*, y empresas contratadas para servicios especiales. Esto en lugar de utilizar las variables sugeridas por el CERLALC: subsector editorial y rango de facturación.

Los resultados obtenidos permiten realizar las siguientes observaciones:

- La diferencia principal entre las editoriales proviene del nivel de facturación. Así parece indicar el método de conglomeración, ya que todas las empresas con rango de tamaño micro y pequeña (que facturan menos de 21 millones de pesos), además de la mayoría de las empresas de tamaño mediana, fueron agrupadas dentro del conglomerado 1.
- La distribución de los conglomerados sugiere un alto grado de similitud entre las empresas cuya facturación es inferior a los 21 millones de pesos. Por otro lado, y en sentido opuesto, las empresas de tamaño Macro presentan grandes diferencias entre sí.
- Las empresas que se encuentran en los conglomerados 2 al 10 son observaciones extremas y las diferencias existentes entre éstas se acentúan más debido al tamaño de las editoriales.
- El algoritmo nos da información acerca de qué otros criterios podrían considerarse en la segmentación de las empresas editoriales en México, especialmente en el caso de las empresas de tamaño Grande o Macro:
  - principal temática de producción, particularmente cuando las editoriales producen ediciones para los programas de gobierno,
  - canal de comercialización.

Este conocimiento de nuevos y diferentes criterios de estratificación puede ser de gran importancia en los casos en los que, en el estudio de la Actividad Editorial que prepara CANIEM, no se cuente con la participación de las editoriales de tamaño Grande o Macro. A través de estos criterios se pueden refinar los estratos a utilizar para estimar la información de las empresas faltantes.

En relación con la hipótesis planteada al inicio de este trabajo, en la que se considera la posibilidad de que la estratificación propuesta por el CERLALC para la elaboración de los estudios estadísticos no es adecuada para las editoriales mexicanas, se puede concluir que, en el caso de las empresas grandes y especialmente las macro, esta estratificación no es suficiente, ya que por ejemplo, dos empresas dentro de un mismo estrato (Texto-Macro) que se suponen homogéneas, resultan ser tan diferentes en su oferta editorial y en su organización (canales de comercialización, inversión de publicidad, empleo) que más bien, se pueden ver como opuestas (caso del conglomerado 4 y del conglomerado 6).

En cambio, para las empresas de tamaño pequeña y micro que están agrupadas dentro de un solo conglomerado, se puede pensar que la estratificación en función del subsector editorial (principal temática de producción) no es tan importante. Sin embargo, debe de considerarse que las empresas utilizadas en el presente trabajo forman parte de una muestra, por lo que el resultado obtenido puede variar al tomarse una muestra distinta.

También se percibe que este análisis muestra lo atípicos que pueden ser los 20 casos segmentados en los conglomerados distintos al 1, en particular las editoriales que se encuentran en los conglomerados de un solo elemento. Con esto se exhibe la importancia de la participación de estas empresas en las encuestas de actividad industrial que se hacen en este organismo.

Finalmente, a partir de los resultados obtenidos se puede proponer una nueva investigación, en la que se discuta la factibilidad de una nueva estratificación o un método de muestreo distinto, en el que se integre el nivel de facturación de la empresa (en lugar del rango de tamaño), la clasificación temática de la producción (en lugar del subsector editorial) y los canales de comercialización.

## ANEXO I

Promedios de las variables estandarizadas

	Fact	p_infant	p_juv	p_lit	p_leng	p_rel	p_fil	p_arte	p_cpuras	p_tec
Conglomerado 1	-.39034	-.22121	-.23711	-.16760	-.14862	.03299	-.21113	-.14529	-.18447	-.18233
Conglomerado 2	1.71309	.12956	.14125	-.18155	2.07663	-.21377	-.19460	-.11434	-.24617	-.34632
Conglomerado 3	1.78638	-.26679	-.25373	-.14872	.11808	-.30331	-.20036	.23478	5.30226	3.21157
Conglomerado 4	2.18373	-.05936	7.52437	-.19498	-.11705	-.30331	.36100	.21194	1.44768	2.01543
Conglomerado 5	2.04633	4.64637	2.76967	6.09476	-.15022	.12414	5.35635	-.21694	-.24617	.14750
Conglomerado 6	4.59751	-.26679	-.25373	-.19498	-.15022	-.30331	-.32726	-.21694	-.24617	-.38438
Conglomerado 7	.20817	.36560	.45607	.66437	-.13451	.20125	1.17644	.45255	.00177	-.29453
Conglomerado 8	1.49422	3.80696	.23473	.01267	.00229	-.30331	.64600	.85554	-.24617	1.62866
Conglomerado 9	.74596	-.26679	-.25373	.49645	-.15022	-.06885	-.26356	-.21694	-.01692	.52646
Conglomerado 10	-.42418	-.13293	4.01417	.29411	-.10495	.04918	2.93228	8.90166	-.24617	1.83852

Promedios de las variables estandarizadas (continuación)

	p_csoc	p_ed_bas	p_geog	p_gen	p_otro	p_bib	p_sec	libreria	puesto_per	autoserv
Conglomerado 1	-.06264	-.23288	-.18206	-.14093	-.13764	-.19231	-.26392	-.36321	-.17992	-.24518
Conglomerado 2	-.18637	.69155	-.15962	-.23486	-.16202	1.90375	2.09292	1.18697	-.18368	.41029
Conglomerado 3	1.53141	.19448	.39747	-.23265	.03564	-.21493	.14632	1.44470	-.18368	-.28193
Conglomerado 4	.62938	4.33899	1.21462	.73960	-.16202	.30296	2.94289	4.09050	-.18368	-.29213
Conglomerado 5	.51999	-.28561	1.01404	1.82096	1.05567	.49227	-.33267	2.19494	6.42408	2.67427
Conglomerado 6	-.26601	6.62194	-.29033	-.23486	.16563	.27926	5.68211	.78305	-.18368	-.32473
Conglomerado 7	.29725	-.27818	3.79277	.62425	.14018	.38005	-.33267	.55817	.88081	.78233
Conglomerado 8	-.26601	2.13283	-.29033	3.72972	-.16202	1.06594	.64664	2.89348	-.18368	5.23255
Conglomerado 9	-.23679	-.28561	.16603	-.14447	-.02993	-.30811	-.33267	-.56068	-.18368	-.32473
Conglomerado 10	-.11651	-.10302	.05016	-.23486	9.41942	-.30811	-.33267	.71027	1.11462	.28937

Promedios de las variables estandarizadas (continuación)

	ferias	gobierno	escuelas	bibliotecas	empresas	asociac	credito	correo	expendio	internet
Conglomerado 1	-.16227	-.26089	-.17468	-.13398	-.13166	.03044	-.04244	-.08305	-.07679	-.10485
Conglomerado 2	.01746	2.53536	.64814	-.26471	-.06357	-.11871	-.19425	-.13521	-.24527	-.13734
Conglomerado 3	.29691	.06820	.38457	3.39656	-.18886	-.11871	-.19425	-.13521	1.28181	.20370
Conglomerado 4	.32147	1.80535	.29894	.55280	.01064	-.11871	-.19425	-.13521	5.11446	-.13734
Conglomerado 5	.60240	-.20838	-.21389	-.26471	-.17276	-.11871	1.49923	-.13521	-.25988	-.13734
Conglomerado 6	-.44455	3.50847	8.89246	-.26471	-.21172	-.11871	-.19425	-.13521	-.23150	-.13734
Conglomerado 7	1.24760	-.21789	-.19332	1.04974	-.20149	-.11871	-.19425	-.12995	-.10628	-.10209
Conglomerado 8	.48328	.67672	.41389	-.26471	6.39586	-.11871	1.70996	-.13521	-.07460	-.13734
Conglomerado 9	-.34494	-.33218	-.21462	-.26471	-.21172	-.11871	-.19425	9.03087	-.01911	9.65886
Conglomerado 10	6.21554	-.33218	-.21462	-.26471	-.21172	-.11871	-.19425	-.13521	-.25988	-.13734

Promedios de las variables estandarizadas (final)

	otros	export	inversion	emp_base	emp_event	empre_free	pers_free
Conglomerado 1	-.16773	-.27518	-.35051	-.28479	-.11829	-.26041	-.29500
Conglomerado 2	-.18271	.75128	1.89072	1.15159	.93879	-.01229	.99520
Conglomerado 3	-.16425	1.42364	1.46147	1.93355	-.13262	1.46869	1.10219
Conglomerado 4	7.73819	-.32565	-.04286	5.96088	1.75400	-.61473	2.49646
Conglomerado 5	2.72183	2.72850	1.54544	.28658	2.20818	1.68205	1.56464
Conglomerado 6	-.03290	.15244	4.08503	2.60425	-.34224	.89136	6.05805
Conglomerado 7	.76626	.78938	-.03754	-.41389	-.34224	4.20474	.80539
Conglomerado 8	-.19388	2.62519	.36773	.73789	-.06274	-.23821	-.50605
Conglomerado 9	-.19388	-.20250	4.31820	.23487	-.34224	-.16290	1.15050
Conglomerado 10	-.19388	-.01414	-.46269	-.10362	-.34224	.36423	-.50605

## ANEXO II

El análisis original incluyó un total de 98 observaciones (empresas editoriales) medidas a través de 37 variables:

- 16 variables distintas de producción de ejemplares por clasificación temática
- 15 variables de venta de ejemplares a través de canales de comercialización
- Nivel de facturación anual
- Inversión en publicidad
- Número de empleados de base
- Número de empleados eventuales
- Número de contratos *freelance*
- Número de empresas contratadas para servicios externos

Con el propósito de reducir el número de atributos con los que se midieron a las editoriales, se agruparon las variables de producción de ejemplares y de venta de ejemplares. También se juntaron las variables de empleo de nómina y empleo de eventuales en una sola variable, así como los contratos *freelance* y los contratos con empresas se unieron en otra variable, quedando de la siguiente manera:

- Facturación. Monto de facturación por la venta de libros (pesos)
- Infantiles y juveniles (inf\_juv). Número de de ejemplares producidos
- Literatura y generalidades (lit\_gen). Número de ejemplares producidos
- Filosofía, psicología y religión (fil\_psi\_rel). Número de ejemplares producidos
- Ciencias puras, Tecnología y ciencias aplicadas y Ciencias sociales (tecnico). Número de ejemplares producidos
- Lenguas y educación básica (educacion). Número de ejemplares producidos
- Arte, Geografía y otros (otro). Número de ejemplares producidos
- Ejemplares de Bibliotecas de Aula, Bibliotecas escolares y programa de secundaria (gob). Número de ejemplares producidos
- Venta en librerías. Número de ejemplares vendidos
- Venta en autoservicios. Número de ejemplares vendidos
- Exportación. Número de ejemplares vendidos
- Escuelas. Número de ejemplares vendidos
- Puestos de periódicos, ferias, bibliotecas, empresas, asociaciones, crédito, correo expendios, internet, otros (otro\_acum). Número de ejemplares vendidos
- Publicidad (inversion). Monto de la inversión hecha en publicidad

- Personas *freelance* y empresas contratadas para trabajos determinados (*freelance*). Número de contratos
- Empleados en nómina y empleados eventuales (*empleo*). Número de empleados

Con estas nuevas variables se preparó un análisis de componentes principales con diferentes combinaciones de editoriales según el rango de tamaño (empresas de tamaño micro y pequeño, empresas de tamaño mediana y grande, empresas de tamaño mediana, grande y macro, etc.), y de éstos se eligió el análisis resultante con las empresas de rango de tamaño Macro, Grande y Mediana.

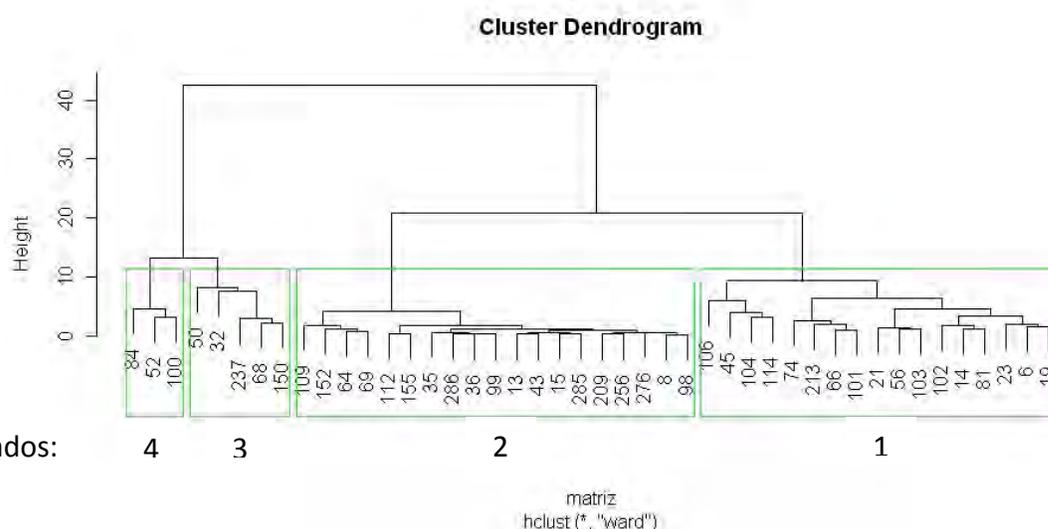
El propósito del análisis de componentes principales fue el de seleccionar las variables más significativas, se calcularon los 2 primeros componentes que acumulan 59.5% de la varianza. Como detalle particular se observó que las variables más relacionadas con el primer componente son las que tienen que ver con los libros de educación y con la producción que se hace para el gobierno, además del nivel de facturación, la inversión en publicidad y las variables de empleo. Por otro lado, las variables que tienen mayor relación con el segundo componente fueron las de producción de libros infantiles y juveniles, literatura y generalidades, filosofía y psicología, las otras temáticas y los otros canales, mientras que la facturación, la producción de libros de educación, la venta a gobierno presentan una correlación negativa con este componente.

Matriz de componentes

Variables	Componentes	
	1	2
Zscore(fact)	.911	-.242
Zscore(inf_juv)	.639	.620
Zscore(lit_gen)	.419	.711
Zscore(fil_psi_rel)	.197	.697
Zscore(tecnic)	.131	.031
Zscore(educacion)	.751	-.463
Zscore(otro)	.255	.519
Zscore(gob)	.688	-.586
Zscore(libreria)	.817	.182
Zscore(autoserv)	.412	.594
Zscore(exportacion)	.649	.329
Zscore(escuelas)	.539	-.590
Zscore(otro_acum)	.567	.605
Zscore(inversion)	.680	-.356
Zscore(freelance)	.633	-.261
Zscore(empleo)	.666	-.334
% Varianza acumulada	35.916	23.601

Para el análisis de conglomerados se utilizaron las variables que en el primer componente presentaran una correlación superior a 0.6 (excepto la de infantiles y juveniles que tiene el mismo grado de correlación en los dos componentes) y a las 44 empresas que facturan más de 21 millones de pesos (rango de tamaño Mediana o superior) y que participaron en el estudio de producción y comercialización del libro en México 2008. Estas empresas agrupan 91% de los ejemplares producidos y 92% del monto de facturación del sector durante 2008.

El algoritmo utilizado para generar los conglomerados fue jerárquico, con liga de Ward. Se utilizaron el programa SPSS versión 17.0 y el programa R. El número de conglomerados elegidos fue de 4 y se decidió a partir del análisis del dendrograma obtenido con R.



Conglomerados:

Valores promedio de las variables por conglomerado

Conglomerados				
Variables	1	2	3	4
Valor de facturación	143,940,806	50,461,347	366,588,208	432,037,542
Producción de educación	378,702	101,850	4,981,687	3,456,240
Producción para gobierno	470,906	64,703	3,374,675	1,629,122
Venta en librerías	792,633	282,917	1,371,744	2,685,219
Venta al extranjero	321,497	31,176	68,922	1,836,469
Inversión en publicidad	9,230,297	1,534,962	22,625,119	44,545,955
Empleo (nómina y eventuales)	119	41	385	283
Contratos externos (free-lance y empresas)	89	11	144	57
<b>Número de editoriales</b>	<b>17</b>	<b>19</b>	<b>5</b>	<b>3</b>

## ANEXO III

Otro ejercicio consistió en preparar un análisis sustituyendo las 18 variables de producción por temática por una sola que englobe la producción total, también se agruparon los 15 canales de comercialización en cuatro categorías: librerías, gobierno, exportación y otros. Así mismo se sumaron el empleo de nómina con el de eventual y el empleo *freelance* con las empresas, quedando 9 variables en lugar de las 38 originales:

- Producción de ejemplares: Número de ejemplares producidos por una editorial en un año
- Librerías: Número de ejemplares vendidos en librerías
- Gobierno: Número de ejemplares vendidos al gobierno
- Exportación: Número de ejemplares exportados
- Otros: Número de ejemplares vendidos en puestos de periódicos, tiendas de autoservicio y departamentales, ferias del libro, escuelas, bibliotecas, empresas privadas, asociaciones de padres de familia, directas a crédito, correo directo, expendios propios, internet, otros.
- Empleo: Número de empleados con contrato ya sea de nómina o eventual
- *Freelance*: Número de contratos con *freelance* ó con empresas para servicios externos
- Facturación: Monto de las ventas de libros en un año
- Inversión: Monto invertido en la publicidad y promoción

Utilizando las 98 editoriales que participaron en la Actividad Editorial de 2008 y con un análisis jerárquico con liga de Ward, se obtuvo la siguiente distribución de conglomerados:

Distribución de las 98 observaciones en 10 conglomerados

Rango de tamaño	1	2	3	4	5	6	7	8	9	10
Micro	32	1								
Pequeña	21									
Mediana	18	4								
Grande	1	8		1				1		1
Macro			1	1	1	2	1	1	1	2
Total	72	13	1	2	1	2	1	2	1	3

Se puede observar nuevamente como las editoriales chicas quedan agrupadas en el primer conglomerado, mientras que las empresas Macro y Grande se reparten en los restantes 9 conglomerados.

El dendrograma obtenido se muestra a continuación

C A S E	0	5	10	15	20	25
Label	Num					
	29	--+				
	62	--+				
	87	--+				
	21	--+				
	79	--+				
	93	--+				
	83	--+				
	75	--+				
	86	--+				
	39	--+				
	42	--+				
	81	--+				
	1	--+				
	78	--+				
	41	--+				
	63	--+				
	16	--+				
	56	--+				
	59	--+				
	90	--+				
	34	--+				
	76	--+				
	89	--+				
	2	--+				
	95	--+				
	72	--+				
	65	--+				
	73	--+				
	68	--+				
	70	--+				
	74	--+				
	85	--+				
	53	--+				
	61	--+				
	67	--+				
	82	--+				
	52	--+				
	88	--+				
	9	--+				
	98	--+				
	5	--+				
	64	--+				
	28	--+				
	37	--+				
	43	--+				
	91	--+				
	15	--+				
	57	--+				
	4	--+				
	20	--+				
	17	--+				
	35	--+				
	23	--+				
	92	--+				
	10	--+--+				
	19	--+				
	8	--+				
	55	--+				
	80	--+				
	7	--+				
	6	--+				
	22	--+				
	69	--+				
	84	--+				
	96	--+				
	44	--+				
	97	--+				
	33	--+				
	12	--+				
	30	--+				
	60	--+				
	66	--+				
	38	--+				
	47	--+				
	14	--+				
	54	--+				
	31	--+--+				
	46	--+				
	71	--+				
	3	--+				
	27	--+				
	94	--+				
	11	--+				
	48	--+				
	13	--+				
	18	----+--+				
	40	----+ +--+				
	26	--+--+				
	45	--+ +--+				
	24	--+--+				
	49	--+				
	36	--+--+				
	51	--+				
	58	--+				
	77	--+--+				
	50	--+ +--+--+				
	32	----+				
	25	-----+				

## 5. BIBLIOGRAFÍA

- Ardila, R.; Uribe, R. (2003). *Metodología para la realización de estudios estadísticos del libro*. Centro Regional para el Fomento del Libro en América Latina y el Caribe. Colombia.
- CANIEM (2009). *Actividad Editorial, 2008*. Cámara Nacional de la Industria Editorial Mexicana. México
- Härdle, W.; Simar, L. (2007). *Applied Multivariate Statistical Analysis*. Springer.
- Kaufman, L.; Rousseeuw, P. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons.
- Loureiro, A., Torgo, L., and Soares, C. (2004). "Outlier Detection Using Clustering Methods: a data cleaning application". En *Proceedings of KNet Symposium on Knowledgebased systems for the Public Sector*, M. May y D. Malerba, eds. (Citeseer), Alemania pp. 57-62. Disponible en:  
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.61.7266&rep=rep1&type=pdf>.
- Rencher, A. (2002). *Methods of Multivariate Analysis*. John Wiley & Sons.
- Theodoridis, S.; Koutroumbas K. (2003). *Pattern Recognition*. Elsevier Academic Press.
- Timm, Neil H. (2002) *Applied Multivariate Analysis*. Springer-Verlag New York, Inc.
- Webb, A. (2002). *Statistical Pattern Recognition*. John Wiley & Sons.