

Europa

Introducción

La estadística industrial es la parte de la estadística que busca implementar los procedimientos probabilísticos y estadísticos de análisis e interpretación de datos o características de un conjunto de elementos al entorno industrial. Para poder analizar cualquier realidad social o económica de interés es imprescindible el empleo de diferentes métodos estadísticos que permitan la observación del fenómeno y la recolección de datos. Con esto se pretende lograr una mayor comprensión de la realidad estudiada que facilite la toma de decisiones. De esta forma, en el presente trabajo se pretende emplear la estadística para profundizar en los conocimientos que actualmente se tienen sobre las industrias en Europa y determinar cuáles son los condicionantes de las mismas.

El objetivo principal de dicho estudio consiste en analizar los datos proporcionados por el *Euromonitor (1979), European Marketing Data and Statistics, London: Euromonitor Publications 76-79*, encontrando conjuntos de datos similares sobre empleo en diferentes industrias de países europeos, con el fin de volver a realizar un estudio pero con datos actuales, averiguando su evolución a lo largo del tiempo. Cabe mencionar que este fue el primer estudio relacionado con este tema.

Bajo este estudio podemos ver qué industrias en qué países están más desarrolladas.

En el presente trabajo se pretende emplear el Análisis Cluster para profundizar en los conocimientos que actualmente se tenían sobre los empleados en diferentes industrias en Europa.

Metodología

El estudio se basa en el Análisis Cluster, cuyo objetivo es buscar agrupar variables tratando de lograr la máxima homogeneidad en cada grupo y la mayor diferencia entre los grupos. Nos vamos a agrupar los países según la actividad industrial.

Al principio, establecemos un indicador que nos diga en qué distancia cada par de elementos se parecen entre sí. A continuación, creamos grupos, denominados clusters, de forma que cada uno de ellos, contenga las muestras más parecidas entre sí. Para esto, utilizamos el método jerárquico para definir este número de clusters y después, el método de partición para ver si es necesario que el punto cambiaría el cluster (con k-medias, k-modas y k-medoides).

Por último, vamos a hacer la validación de los clusters, es decir, la evaluación de la solución encontrada: describimos los grupos y los comparamos unos con otros.

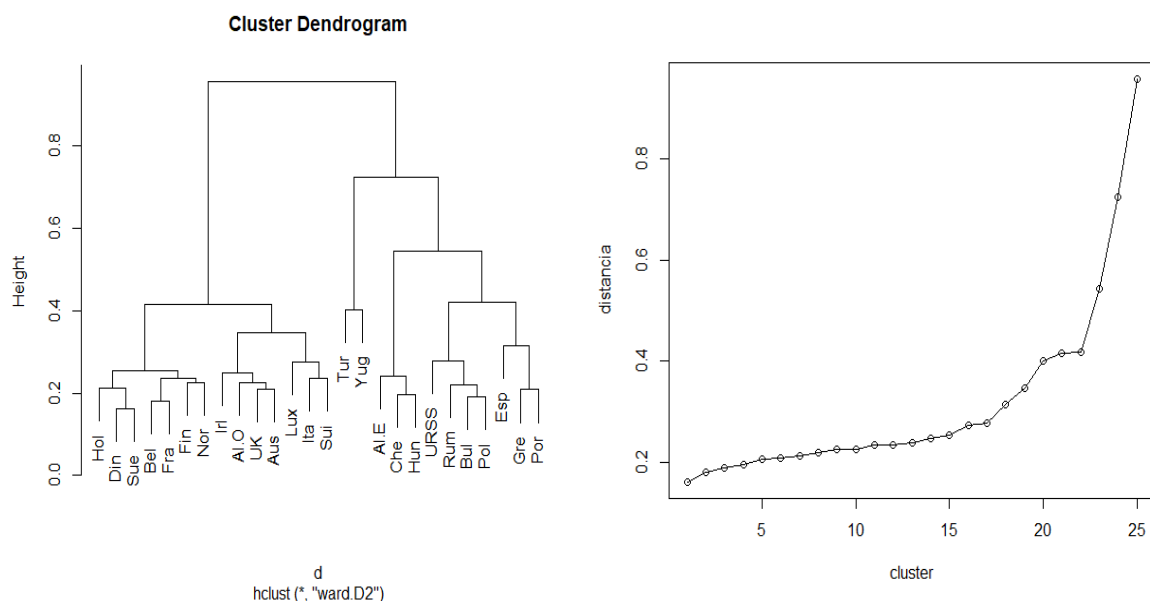
Para realizar este análisis se han utilizado el Software R, un lenguaje de código abierto con enfoque al análisis estadístico.

Los datos fueron sacados desde el Euromonitoring International que se incluye datos y análisis sobre industrias, países y consumidores, con los que generan soluciones a la medida de planificación estratégica y anual para desarrollar y expandir oportunidades comerciales e influir en la toma de decisiones.

Un estudio realizado sobre empleados recogió información de las siguientes variables: país (Country), porcentajes de empleados en agricultura (Agr), minería (Min), industria (Man), suministros industriales (PS), construcción (Con), servicios industriales (SI), finanzas (Fin), servicios sociales (SPS) y en comunicaciones y transporte (TC).

Análisis

El fichero *europa* contiene información sobre el porcentaje de empleados en 26 países europeos relativa a las siguiente variables: Agr, Min, Man, PS, Con, SI, Fin, Country, SPS y TC. Abajo aparece el dendrograma y el gráfico de distancias aplicando el método de Ward con la disimilitud de Gower, la que se utiliza para variables mixtas.



La selección del número de grupos es de $k = 3$ y se aplica sobre la matriz de disimilitudes d el algoritmo PAM con el número de clusters estimado por el criterio de optimalidad de anchura de

la silueta media global. Luego en las tablas aparecen los medoides de los 3 grupos y la información que nos proporcionan nos permite ya caracterizar a los grupos.

Grupo	Agr	Min	Man	PS	Con	SI	Fin	SPS	TC
Franci a	10.8	0.8	27.5	0.9	8.9	16.8	6.0	22.6	5.7
Bulga ria	23.6	1.9	32.3	0.6	7.9	8.0	0.7	18.2	6.7
Yugos lavia	48.7	1.5	16.8	1.1	4.9	6.4	11.3	5.3	4.0

Grupo	Tamaño
1	16
2	8
3	2

Agricultura

Grupo	Mín.	1 cuart.	Mediana	Media	3 cuatr.	Max
1	2.7000	6.6000	9.1000	11.5625	13.7250	27.8000
2	4.2000	20.4000	23.6500	24.6125	32.0000	41.4000
3	48.7000	53.2250	57.7500	57.7500	62.2750	66.8000

Minería

Grupo	Mín.	1 cuart.	Mediana	Media	3 cuatr.	Max
1	0.1000	0.3750	0.7000	0.8125	1.0250	3.1000
2	0.6000	1.7750	2.3000	2.1750	2.9000	3.1000
3	0.7000	0.9000	1.1000	1.1000	1.3000	1.5000

Industria

Grupo	Mín.	1 cuart.	Mediana	Media	3 cuatr.	Max
1	20.70000	24.00000	27.55000	27.48125	30.20000	37.80000
2	17.60000	25.77500	29.85000	29.72500	33.10000	41.20000
3	7.90000	10.12500	12.35000	12.35000	14.57500	16.80000

Suministros Industriales

Grupo	Mín.	1 cuart.	Mediana	Media	3 cuatr.	Max
1	0.50000	0.77500	0.85000	0.91875	1.07500	1.40000
2	0.60000	0.60000	0.75000	0.96250	1.22500	1.90000
3	0.10000	0.35000	0.60000	0.60000	0.85000	1.10000

Construcción

Grupo	Mín.	1 cuart.	Mediana	Media	3 cuatr.	Max
1	6.9000	7.4750	8.5000	8.6125	9.2750	11.5000
2	7.6000	8.0500	8.3000	8.3500	8.7000	9.2000
3	2.8000	3.3250	3.8500	3.8500	4.3750	4.9000

Servicios Industriales

Grupo	Mín.	1 cuart.	Mediana	Media	3 cuatr.	Max
1	9.70000	14.55000	16.80000	16.03125	17.62500	19.10000
2	5.90000	7.15000	8.60000	8.60000	9.85000	11.50000
3	5.20000	5.50000	5.80000	5.80000	6.10000	6.40000

Finanzas

Grupo	Mín.	1 cuart.	Mediana	Media	3 cuatr.	Max
1	1.600	4.675	5.400	5.175	6.050	8.500
2	0.500	0.850	0.900	1.100	1.225	2.400
3	1.100	3.650	6.200	6.200	8.750	11.300

Servicios Sociales

Grupo	Mín.	1 cuart.	Mediana	Media	3 cuatr.	Max
1	11.800	18.600	22.450	22.850	27.775	32.400
2	11.000	15.000	17.550	17.225	19.175	23.600
3	5.300	6.950	8.600	8.600	10.250	11.900

Comunicaciones y transporte

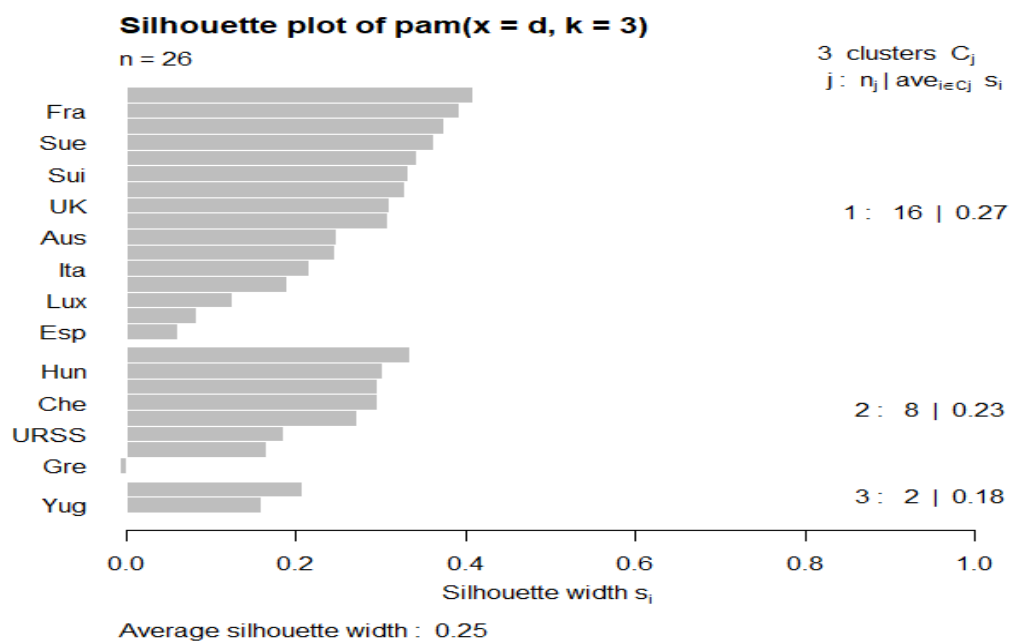
Grupo	Mín.	1 cuart.	Mediana	Media	3 cuatr.	Max
1	5.5000	5.7000	6.3000	6.5625	7.0250	9.4000
2	5.0000	6.7000	6.9500	7.2500	8.1000	9.3000
3	3.2000	3.4000	3.6000	3.6000	3.8000	4.0000

Al grupo 1 pertenecen países más desarrollados, con un nivel muy elevado en sectores sociales, industriales en comparación con los demás grupos, sin embargo, tiene menos empleados en sector agrario. Podremos identificarlos como países de Europa Central y Europa Occidental (Ej. Alemania, Bélgica, Francia)

Los países del grupo 2 son las que en su mayoría pertenecen a Europa Oriental, países en vía de desarrollo, en los que, por ejemplo el sector agrario sigue teniendo un nivel de empleados más elevado que en el grupo 1. Además, en lo que se refiere a financiación y construcción están un poco menos avanzados.

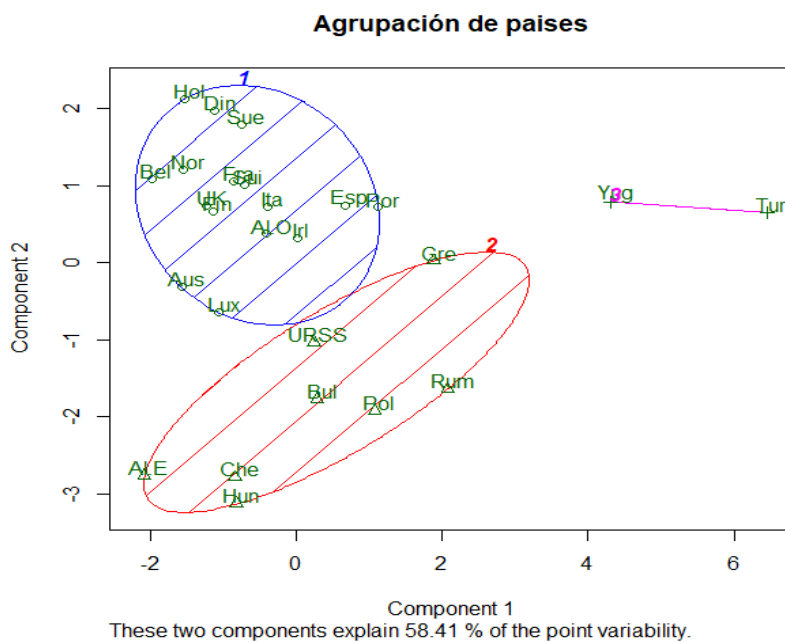
El grupo 3 es representado solamente por 2 países, pero tienden a mostrar un comportamiento parecido, son Yugoslavia y Turquía. Su sector agrario demuestra un mayor avance, en comparación con los demás grupos, mientras tanto el nivel de Transporte y Comunicación, Minería y Construcción queda por debajo de todos.

Vamos a aplicar en primer lugar el criterio de maximización de asw para seleccionar el número de *clusters* apropiado. Como salida nos da un número de 3 lo que nos indica que la proposición de que $k = 3$ se considera correcta. Abajo aparece el gráfico de siluetas con anchuras medias de siluetas para cada grupo: 0.27, 0.23, 0.18.



Aunque todos parecen ser bastante pequeños, se ha encontrado una estructura razonable en los datos. No se puede observar ningún valor negativo, con lo cual, no habría que reasignar ninguna observación al *cluster* vecino. Hacemos una validación con el coeficiente de Dunn y la anchura media global de la silueta que sale igual a $\text{Índ.Dunn} = 0.4822800$, $ASW = 0.2515949$, por lo tanto la asignación realizada por el método de las *k-medoides* se considera la más fiable, comprando con la de método Ward ($k = 4$, $\text{Índ.Dunn} = 0.5132209$ un poco más fiable, sin embargo $ASW = 0.1900979$ mucho menos apropiada).

Al fin y al cabo obtenemos una representación gráfica del AC sobre nuestro conjunto de datos.



Resultados

En definitiva, los datos sobre los países europeos se han dividido finalmente en 3 clusters: el primero se caracteriza por un nivel elevado en sectores sociales e industriales, el segundo por países en vía de desarrollo y el tercero por el nivel alto de sector agrario. Al final ha alcanzado un determinado número de iteraciones (marcadas por el investigador). Cabe destacar que el método de encadenamiento medio permitió seleccionar más países homogéneas que el método del centroide. También, el de Ward arrojó los peores resultados que los resultados del k-medoides.

Conclusiones y Recomendaciones

Como técnica de análisis multivariado el AC puede ser muy útil para reducir una compleja cantidad de información. Los métodos de aglomeración jerárquicos, no jerárquicos, las medidas de similitud, la forma como se escoge el número óptimo de grupos e incluso los elementos de procedimiento, hacen que se presenten múltiples soluciones y que la decisión final sea tomada más de manera subjetiva que de forma objetiva. En la literatura consultada en muchos casos se afirma que esta técnica es un estigma metodológico, ya que algunas veces los grupos que se identifican no reflejan condiciones reales sino que son “simples artificios estadísticos supeditados a variaciones numéricas aleatorias entre clústers”, cuestionando su valor como método aplicativo a la investigación debido a los resultados equívocos a los que lleva a veces su utilización.

Bibliografía

Applied multivariate statistics for the social sciences. Edición: 5th ed. Autor: Stevens, James Paul. Editorial: London [etc.] : Rouledge, 2009

Hair, J.F., Anderson, R.E., Tatham, R.L. y Black, W.C. (1999) Análisis Multivariante (5ª edición). Ed. Prentice Hall.

Methods of multivariate analysis. Edición: 3rd ed. Autor: Rencher, Alvin C. Editorial: Hoboken, NJ : John Wiley & Sons, 2012

Applied multivariate analysis [Recurso electrónico]. Edición: -. Autor: Timm, Neil H.. Editorial: New York : Springer, c2002.

Peña, D. (2002). Análisis de Datos Multivariantes. Mc Graw - Hill. España.

Páginas web:

<https://www.euromonitor.com/>

https://es.wikipedia.org/wiki/Estad%C3%ADstica_industrial