



TOCHE ENJALMADO (ROJO). (Ramphocelus flammigerus). 18cm. Generalmente está en grupos ruidosos. Se alimenta de frutas e insectos. Para cortejar a la hembra, el macho alza la cabeza hacia ella y muestra la parte inferior del pico ligeramente inflamada. Anida en arbustos bajos. La hembra pone dos huevos de color azul. La hembra es de alas negruzcas claras y de pecho y cola amarillo-naranja. También hay toches enjalmados de color amarillo y negro en el macho y café y amarillo en la hembra.

*Estructuración y aplicación del Análisis
de Componentes Principales
como criterio para la determinación
de los inputs y outputs
a considerar en un análisis financiero*

SÍNTESIS

Se describe la estructuración matemática del Análisis de Componentes Principales (ACP) y una aplicación en el sector metalmecánico en Colombia. Es detallado el procedimiento analítico y numérico empleado para encontrar los componentes principales. Se hace énfasis en el procedimiento "Algoritmo QR de matrices cuadradas" para el cálculo de valores y vectores propios que reduce las variables a analizar. El Análisis de Componentes Principales es un criterio estadístico que simplifica la subjetividad de los tomadores de decisiones empresariales, gracias a la aplicación del álgebra lineal y el análisis numérico en la reducción de variables a analizar en los estudios financieros. El comportamiento financiero de las empresas del sector metalmecánico 2009-2010 puede ser explicado por la Utilidad operacional, Pasivo no Corriente e Ingresos no Operacionales.

DESCRIPTORES: Ortogonalización, Análisis de Componentes Principales, inputs, outputs, Algoritmo QR, valor propio, vector propio, sector metalmecánico, análisis financiero.

ABSTRACT

This paper describes the mathematical structure of the Principal Component Analysis (PCA) and an application in the metal-mechanical sector in Colombia. The analytical and numerical procedure used, is detailed to find the principal components. It emphasizes in the "square matrix QR algorithm" procedure for calculating its own values and vectors which reduce the variables to be analyzed. The Principal Component Analysis is a statistical approach that simplifies the subjectivity of business decision makers, through the implementation of linear algebra and numerical analysis in reducing variables to analyze financial studies. The financial behavior of companies in the metal-mechanical sector 2009-2010 can be explained by the operating profit, non-current liabilities and non-operating income.

DESCRIPTORS: Orthogonalization, main component analysis, inputs, outputs, QR Algorithm, own value, own vector, metal-mechanical sector, financial sector.

Estructuración y aplicación del Análisis de Componentes Principales como criterio para la determinación de los inputs y outputs a considerar en un análisis financiero¹

Eduardo Arturo Cruz Trejos²
Sergio Aristizábal Hernández³

Structuring and implementation of the principal components analysis as criteria for determining the inputs and outputs to be considered in a financial analysis

Primera versión recibida: 13 de abril de 2011. Versión final aprobada el 19 de agosto de 2011

Para citar este artículo: Cruz Trejos, Eduardo Arturo, Aristizábal Hernández, Sergio (2011). "Estructuración y aplicación del Análisis de Componentes Principales como criterio para la determinación de los inputs y outputs a considerar en un análisis financiero". En: Gestión y Región., N.º. 12, (Julio-Diciembre, 2011); p. 67-88.

Los resultados mostrados en el presente documento tienen como objetivo evidenciar la utilización de la técnica de Análisis de Componentes Principales para la selección reducida de los indicadores financieros que afectan directamente la eficiencia de las empresas pertenecientes al sector metalmeccánico de Colombia en el año 2009-2010. Esto es logrado gracias a:

- La caracterización del sector metalmeccánico en Colombia referente a la cantidad de empresas (por tamaño) que hacen parte de la fuerza productiva de este sector en el país.
- El diseño de un modelo numérico para desarrollar la técnica de Análisis de Componentes Principales.
- La aplicación del modelo numérico en el sector metalmeccánico en Colombia para el año 2009-2010.

La información financiera: *inputs* y *outputs* de la empresa

Las empresas, como unidades de producción económica, generan constantemente un flujo de entradas y salidas de recursos, que son el resultado de las operaciones directas o movimientos ajenos a la actividad de la empresa. Estos movimientos, ya sean operativos o no-operativos, se denominan generalmente *inputs* (entradas) y *outputs* (salidas), y son los utilizados para determinar la eficiencia de una empresa, que se mide a través de la razón entre las salidas y las entradas de la empresa:

$$\text{Eficiencia} = \frac{\text{Outputs}}{\text{Inputs}}$$

¹ Producto derivado del proyecto de investigación "Modelo de análisis financiero, riesgo de crédito y eficiencia, construido y aplicado al sector metalmeccánico en Colombia", registrado ante el Centro de Investigaciones de la Universidad Libre Seccional Pereira.

² MSc en Administración Económica y Financiera, Ingeniero Industrial, Docente investigador Universidad Tecnológica de Pereira, Director Grupo de investigación Administración Económica y Financiera, Docente catedrático Universidad Libre Seccional Pereira.

³ Ingeniero Financiero, Investigador programa de Ingeniería Financiera Universidad Libre Seccional Pereira, Grupo de investigación Administración Económica y Financiera.

Igualmente se supone, por principio, que el movimiento financiero de la empresa se encuentra detallado en los estados financieros, como son: el balance general, el estado de resultados, estado de fuentes y aplicación de fondos y el flujo de efectivo. Determinar las entradas y salidas para determinar la eficiencia de una empresa, requiere experiencia por parte del investigador, así como también de fundamentos que solidifiquen su criterio para la selección de técnicas matemáticas pertinentes.

Weston y Brighman (1994), aseveran que cuando las empresas realizan sus operaciones productivas y financieras, deben registrar todos los movimientos en unos documentos que puedan ser vistos posteriormente y que en base a ellos, haya toma de decisiones de acuerdo a lo encontrado en tales informes, los cuales describen puntualmente las características fundamentales de una empresa. “En general puede decirse que un estado financiero es una relación de cifras monetarias vinculadas con uno o varios aspectos específicos del negocio y presentados con un ordenamiento determinado” (García, 1999, p.62). Apreciando esta afirmación: los estados financieros involucran todos los movimientos relacionados con el concepto de moneda, no haciendo referencia a lo que solo es representado físicamente; en el mercado hay títulos valores y marcas comerciales, por ejemplo.

Periódicamente, muchas personas internas o externas a la empresa, siempre están interesadas en ver la situación financiera de la empresa en un momento determinado. Esto implica que los estados financieros deben tener un orden, confiabilidad y claridad estricta, ya que de estos mismos dependen las tomas de decisiones de las directivas más importantes de la empresa. En este orden ideas se infiere que la labor de conseguir y tener una información financiera involucra un tedio y profundo proceso para poder generar estados financieros “estándar”. Ross (2009), hábilmente, afirma: “Una cosa muy obvia que se podría querer hacer con los estados financieros de una compañía es compararlos con los de otras empresas similares. Sin embargo, inmediatamente surgiría un problema. Es casi imposible comparar directamente los estados financieros de dos compañías debido a las diferencias en tamaño” (p.41). El anterior, es uno de los muchos ejemplos encontrados en la vida real en cuanto a estandarización de información financiera. Para el caso de la comparación entre empresas de diferente tamaño, la solución es trabajar con porcentajes de proporción.

Pero esto es un detalle menor si es comparado con la disparidad que existe en algunos momentos cuando las empresas tienen que presentar sus informes a un ente regulador, como por ejemplo, a la superintendencia de sociedades mercantiles. También es un problema cuando es requerido realizar investigaciones en el tema financiero. Por ejemplo, para construir una función matemática que describa el comportamiento de cierta característica de un sector de la economía, es indispensable que muchas empresas hagan parte del estudio;

uno de los problemas más intensos es el de la reorganización de la información, ya que cada una de las empresas presenta sus propios estados financieros.

Los estados financieros se componen de 5 informes principales (García, 1999):

- El estado de resultados.
- El balance general.
- Estado de utilidades retenidas.
- Estado de fuentes y aplicación de fondos.
- Estado de flujo de efectivo.

Pero también es considerado por otros autores como Gitman (2003):

- Estado de resultados.
- Balance general.
- Estado de utilidad retenidas.
- Estado de flujos de efectivo.

El segundo autor no considera el estado de fuente y aplicación de fondos. Inclusive, la teoría de información financiera de los Estados Unidos no considera el estado de fuentes y aplicación de fondos. Pero, García (1999) desarrolla sus apuntes en la economía colombiana. El estado de fuente y aplicación de fondos no deja de ser importante, ya que este, haciéndole un esbozo, dice de dónde se sacaron los recursos para la empresa y en dónde se usaron, ya sea en figura de gastos o inversiones (García, 1999).

A partir de la información que brindan los estados financieros, se inicia con el cálculo de una serie de indicadores, proporciones y comparaciones para vislumbrar el comportamiento de la empresa objeto de estudio. “Los signos vitales de una empresa son tres: la liquidez, la rentabilidad y el endeudamiento” (Cruz, Villareal & Rosillo, 2005). De esta manera se fundamenta la creación, utilización e interpretación de los indicadores financieros.

El análisis de componentes principales es una técnica estadística que, partiendo de un número dado de variables que explican un hecho, es capaz de disminuir este número de variables sin perder el poder de explicación que se tiene inicialmente con una cantidad mayor de variables (Guisande Gonzáles, Cástor *et. Al*, 2006). Esta técnica requiere la utilización de métodos numéricos para su solución si el número de variables a considerar es mayor de tres; se deja claro que en la práctica es innecesaria la disminución de variables si un hecho es explicado por solo tres variables. Sin embargo, podría presentarse redundancia o una alta correlación entre estas variables; de esta manera se deja entreabierto la dificultad que se daría si se pretendiera explicar analíticamente un hecho con más de tres incógnitas, a menos, eso sí, que fuera un problema netamente de programación lineal.

Análisis de Componentes Principales

El Análisis de Componentes Principales (ACP) es una técnica estadística que permite reducir la cantidad de variables que explican un hecho a costa de una disminución insignificante de información. El supuesto básico en el que se fundamenta la necesidad de aplicar esta técnica, radica en que es posible, al menos cuantitativamente hablando, que existan variables que se encuentren altamente correlacionadas y describan una misma situación (Smith, 2002). En la práctica, esto se traduce en disponer de información redundante que hace innecesaria la inclusión de una de estas variables, debido al tratamiento extra que se requeriría para realizar un análisis. Es decir, si un suceso puede ser explicado por un número de variables reducido, es preferible a que sea explicado por medio de un número mayor de variables.

Según Shlens (2009), el método más aproximado para encontrar los componentes principales de un conjunto de datos con sus respectivas mediciones en diferentes variables, requiere una serie de pasos ordenados que se enuncian a continuación:

- I. Determinar la relación existente entre cada una de las variables. Esto es, crear una matriz de correlaciones o una matriz de covarianzas.⁴
- II. Hallar los valores y vectores propios de la matriz de correlaciones o covarianzas creada.
- III. Obtener los componentes principales.
- IV. Seleccionar el número de componentes principales a analizar y las variables que explican en mayor medida cada uno de estos componentes.

En seguida, es detallado, puntualmente, cada uno de los pasos mencionados:

Matriz de correlaciones y matriz de covarianzas

Estas matrices se construyen con el fin de establecer cuál es la relación existente entre sí de cada una de las variables (Levin & Rubin, 2004).

$$Cov = \sigma_{xy} = \frac{\sum f_i(x_i - \bar{x})(y_i - \bar{y})}{N}$$

$$Corr = \rho_{xy} = \frac{\sigma_{xy}}{\sigma_x^2 \sigma_y^2}$$

Valores y vectores propios

Los valores propios de una matriz cuadrada $n \times n$ se pueden definir como las raíces complejas o reales de un polinomio característico (Villardón, 2004). Un valor propio, se representa con la letra griega λ . En otras palabras, λ es un valor propio de una matriz cuadrada A , si existe un vector, $v \neq 0$ en R^n , tal que:

$$Av = \lambda v$$

⁴ Para saber cuál de las dos matrices aplicar se establece si los datos son homogéneos o heterogéneos. Para los primeros, se utiliza la matriz de covarianzas; para los segundos, la matriz de correlaciones.

A: Es una matriz cuadrada (necesariamente).

λ : Es un valor escalar y es un valor propio de la matriz A.

V: es un vector propio de A correspondiente al valor propio λ de A.

Lo que significa que λ es un valor propio si al transformar el vector V mediante su multiplicación por la matriz cuadrada A, mantiene su misma dirección, tal vez su dirección, y/o cambie su sentido. Este vector V es el que denomina “vector propio”.

Procedimiento analítico para hallar los valores y vectores propios de una matriz 2 x 2

Así como el número uno siempre está multiplicando a los números reales, la matriz “identidad” siempre está multiplicando a las matrices $m \times n$. Se plantea esto debido a que se podría pensar inicialmente que la ecuación $Av = \lambda v$ se resolvería despejando así:

$$v = (A - \lambda) = 0$$

El despeje anterior es erróneo. El verdadero despeje quedaría de la forma:

$$v(A - I\lambda) = 0; v \neq 0 \text{ en } R^n$$

Esta igualdad generada es un sistema lineal homogéneo porque está igualado a cero; pero impide que V sea cero. Esto significa que hay variables libres. Si V no puede ser cero debe tener por lo menos una variable libre para que V no pueda ser cero; entonces la matriz de coeficientes del sistema lineal no tiene inversa (Grossman, 1996):

$$v(A - I\lambda)^{-1} = \text{no existe}$$

Esto implica que su determinante es igual a cero. Entonces:

λ es un valor propio de A si, y solo si,

$$\det(A - I\lambda) = 0$$

La expresión anterior es un polinomio de grado n y se denomina “polinomio característico de A (ecuación característica)”.

De esta manera, se cuenta un instrumento para el cálculo de valores y vectores propios de una matriz cuadrada.

De acuerdo con lo anterior, supóngase una matriz cuadrada 2 x 2:

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

Entonces,

$$A - \lambda I = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \\ = \begin{bmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{bmatrix}$$

Teniendo este resultado, se procede a hallar su determinante:

$$\det(A - \lambda I) = (a_{11} - \lambda)(a_{22} - \lambda) - a_{21}a_{12}$$

Se puede observar que, considerando esta expresión como un polinomio en λ , se encuentra con que es un polinomio de grado dos. Es decir, se podría pensar que es del tipo:

$$\lambda^2 + a_1\lambda + a_2$$

Cuando se calculan los valores de λ se está solucionando la ecuación característica; esta es el polinomio característico igualado a cero.

Para los vectores propios, se utiliza procedimiento de solución de un sistema de ecuaciones *compatible indeterminado*. Esto debido a que la solución del sistema tiene infinitas soluciones.³

Para encontrar una solución, se iguala una de las variables del sistema de ecuaciones a uno, con el fin de tener una solución única.

Valores y los vectores propios a través del algoritmo de descomposición QR de matrices

Este procedimiento es de carácter numérico y consiste en descomponer una matriz de dimensión cuadrada A, en una matriz de columnas ortonormales Q y otra matriz triangular superior R:

$$A = QR$$

Procedimiento:

Como lo enuncia Shlens (2009), el primer paso es hallar el conjunto de vectores ortonormales (matriz Q). Estos se hallan, entre otros, mediante el proceso de ortogonalización de *Gram – Schmidt*. Este, consiste en la transformación de un conjunto de vectores de un espacio R^n , en otro conjunto de vectores ortonormales que representen el mismo sub espacio vectorial. Para su solución se tiene que:

Sea A una matriz cuadrada compuesta de columnas linealmente independientes (a_1, a_2, \dots, a_n) . Este sería el conjunto de vectores a ortogonalizar:

³ El sistema de ecuaciones compatible indeterminado se debe solucionar tantas veces como valores propios hayan sido generados.

$$\begin{bmatrix} \mathbf{a1} & \mathbf{a2} & \mathbf{an} \\ a_{11} & a_{12} & a_{1n} \\ a_{21} & a_{22} & a_{2n} \\ a_{n1} & a_{n2} & a_{nn} \end{bmatrix}$$

Donde la solución se realiza de izquierda a derecha, es decir, primero se ortogonaliza el vector a_1 , después el vector a_2 , y así sucesivamente. Es importante recalcar que a medida que se trabaja con más variables, el grupo de ecuaciones a solucionar se va haciendo cada vez más pesado. Con esta matriz planteada se procede a hallar el conjunto de vectores ortogonalizados:

$$\begin{bmatrix} \mathbf{b1} & \mathbf{b2} & \mathbf{bn} \\ b_{11} & b_{12} & b_{1n} \\ b_{21} & b_{22} & b_{2n} \\ b_{n1} & b_{n2} & b_{nn} \end{bmatrix}$$

Esta matriz, compuesta por columnas linealmente independientes, se resuelve con:

$$b_1 = a_1$$

$$b_2 = a_2 - Proj_{b_1} a_2$$

$$b_k = a_k - \sum_{j=1}^{k-1} Proj_{b_j} a_k$$

Teniendo que:

$$Proj_{b_j} a_k = \frac{\langle a_k, b_j \rangle}{\|b_j\|^2} * b_j$$

Donde se realiza un producto punto entre el vector inicial k por el vector anteriormente ortogonalizado (b_j); este resultado se divide entre la norma al cuadrado de b_j y, finalmente, se multiplica el resultado por b_j (Larson & Edwards, 2002).

Finalmente, se ortonormalizan cada uno de los vectores (b_1, b_2, \dots, b_n). Esto se logra dividiendo cada uno de estos vectores por su respectiva norma.

$$Norma = \|b_j\| = \sqrt{b_j b_j}$$

Como resultado final, se tiene que:

$$Q = \begin{bmatrix} \mathbf{q1} & \mathbf{q2} & \mathbf{qn} \\ q_{11} & q_{12} & q_{1n} \\ q_{21} & q_{22} & q_{2n} \\ q_{n1} & q_{n2} & q_{nn} \end{bmatrix}$$

Luego de tener los resultados de la matriz Q , se procede a calcular el valor de la matriz R^3 , que se obtiene resolviendo la ecuación:

$$R = Q^T A$$

Donde Q^T es la traspuesta de la matriz Q .

Como se mencionó en el apartado anterior, el algoritmo QR es de carácter numérico, es decir, se hacen iteraciones numéricas hasta que se llegue a un resultado, de acuerdo con un criterio establecido. En este caso, el criterio para la solución final es la aproximación de R a una matriz triangular superior: que todos los elementos que se encuentren por debajo de la diagonal principal sean ceros.

Entonces, la matriz de valores iniciales se denomina A_0 . Para la matriz A_n sería:

$$A_n = R_{n-1} Q_{n-1}$$

Cuando se tiene finalmente una matriz R lo más próxima a triangular superior, se han obtenido los valores propios buscados (Nakos & Joyner, 1999). Estos se ubican en la diagonal principal de tal triangular superior.

En cuanto a los vectores propios, se requiere el mismo procedimiento de solución detallado en la solución analítica de valores y vectores propios ya descrita anteriormente.

Obtención de los componentes principales

En este punto sólo se ortonormalizan las columnas de la matriz solución A_n .

Selección de componentes principales y variables explicadoras

Como ya se ha mencionado, los valores propios se organizan de mayor a menor. La razón de esto es que se buscan los valores que expliquen en mayor medida el comportamiento de los datos (Rodríguez, 2009). Para esto se construye una tabla con tres columnas: una en la que se ingresen los valores propios de mayor a menor; la otra con la proporción de cada valor propio respecto a la suma total de los valores propios y, en la última columna, se calcula la proporción acumulada de dichos valores propios.

El número de componentes principales a elegir serán los que expliquen, preferiblemente, más del 90%, la dispersión de los datos⁷.

Las variables que explicarán el comportamiento de los datos serán las que representen un valor superior al 50% dentro del componente principal

6 Este orden es estricto: siempre se halla primero la matriz Q y enseguida la matriz R

7 La selección de los componentes principales se puede realizar en cualquier momento después de haber obtenido los valores propios.

considerado. No se considerarán los elementos (variables) que presenten iguales resultados.

Valores propios: solución analítica contra solución numérica

Sea la matriz cuadrada:

$$A_0 = \begin{bmatrix} 8 & 7 \\ 1 & 2 \end{bmatrix}$$

Analíticamente se resuelve:

$$A - \lambda I = \begin{bmatrix} 8 & 7 \\ 1 & 2 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} = \begin{bmatrix} 8 - \lambda & 7 \\ 1 & 2 - \lambda \end{bmatrix}$$

$$\det(A - \lambda I) = (8 - \lambda)(2 - \lambda) - 1 * 7$$

$$\det(A - \lambda I) = (8 - \lambda)(2 - \lambda) - 7$$

$$\det(A - \lambda I) = 16 - 8\lambda - 2\lambda + \lambda^2 - 7$$

$$\det(A - \lambda I) = \lambda^2 - 10\lambda + 9$$

Ahora, se hallan las raíces de este polinomio de segundo grado:

$$\lambda^2 - 10\lambda + 9$$

$$(\lambda - 9)(\lambda - 1)$$

$$\boxed{\lambda_1 = 9; \lambda_2 = 1}$$

Los valores propios hallados, son: 9 y 1.

Empleando el Algoritmo QR:

$$b_1 = a_1$$

$$b_1 = \begin{bmatrix} 8 \\ 1 \end{bmatrix}$$

$$b_2 = \begin{bmatrix} 7 \\ 2 \end{bmatrix} - \frac{\langle [7 \ 2], \begin{bmatrix} 8 \\ 1 \end{bmatrix} \rangle}{\| \begin{bmatrix} 8 \\ 1 \end{bmatrix} \|^2} \begin{bmatrix} 8 \\ 1 \end{bmatrix}$$

$$b_2 = \begin{bmatrix} 7 \\ 2 \end{bmatrix} - \frac{58}{65} \begin{bmatrix} 8 \\ 1 \end{bmatrix}$$

$$b_2 = \begin{bmatrix} 7 \\ 2 \end{bmatrix} - \begin{bmatrix} 464/65 \\ 58/65 \end{bmatrix}$$

$$b_2 = \begin{bmatrix} 7 \\ 2 \end{bmatrix} - \begin{bmatrix} 464/65 \\ 58/65 \end{bmatrix}$$

$$b_2 = \begin{bmatrix} -0,13 \\ 1,10 \end{bmatrix}$$

Ya se tienen los dos vectores ortogonalizados; ahora se procede a ortonormalizarlos:

$$q_1 = \frac{b_1}{\|b_1\|}$$

$$q_1 = \frac{\begin{bmatrix} 8 & 1 \end{bmatrix}}{\|\begin{bmatrix} 8 & 1 \end{bmatrix}\|}$$

$$q_1 = \begin{bmatrix} 0,99 \\ 0,12 \end{bmatrix}$$

Se procede de la misma manera con b_2 y queda:

$$q_2 = \begin{bmatrix} -0,12 \\ 0,99 \end{bmatrix}$$

$$Q_0 = \begin{bmatrix} 0,99 & -0,12 \\ 0,12 & 0,99 \end{bmatrix}$$

Ahora se resuelve la fórmula para R:

$$R_0 = \begin{bmatrix} 0,99 & 0,12 \\ -0,12 & 0,99 \end{bmatrix} * \begin{bmatrix} 8 & 7 \\ 1 & 2 \end{bmatrix}$$

$$R_0 = \begin{bmatrix} 8,06 & 7,19 \\ 0,00 & 1,11 \end{bmatrix}$$

La primera solución es A_1 y su resultado es:

$$A_1 = \begin{bmatrix} 8,06 & 7,19 \\ 0,00 & 1,11 \end{bmatrix} * \begin{bmatrix} 0,99 & -0,12 \\ 0,12 & 0,99 \end{bmatrix}$$

$$A_1 = \begin{bmatrix} 8,89 & 6,13 \\ 0,13 & 1,10 \end{bmatrix}$$

Este resultado sería la nueva referencia para iniciar nuevamente el “Algoritmo QR”. Como ya se mencionó, el procedimiento se detendrá cuando la matriz R_k se aproxime lo más posible a una triangular.

De esta manera, se encontró que: .

$$\lambda_1 \cong 9; \lambda_2 \cong 1$$

Descripción del problema

Cuando los empresarios se enfrentan a la toma de decisiones, deben realizar una ardua investigación acerca de las variables implicadas en los eventos que originan la toma de decisiones. Para esto acuden tanto a técnicas cualitativas como cuantitativas; para tener un espectro amplio en el cual poder fundamentar su

posición de respuesta. Se hace imperativo, entonces, que se desarrollen o mejoren modelos que permitan disminuir el alto número de variables a las que se enfrentan y se exponen actualmente las empresas. A través de los métodos cuantitativos se pueden determinar criterios para la selección reducida de variables que expliquen un hecho o evento sin producir, en ningún momento, pérdida de información en la dispersión de los datos.

El Análisis de Componentes Principales (ACP), técnica de la estadística multivariada, tiene como aplicación, la reducción de dimensiones a través de la manipulación de subespacios vectoriales; las dimensiones de las que se hablan en esta técnica representan a un grupo grande de variables que se quiere disminuir para evitar la redundancia e información innecesaria. Esto deja al descubierto que el ACP es una técnica de índole cuantitativa y que obliga a que los lectores o interesados en la optimización de procesos para la toma de decisiones empresariales, los cuales no son doctos en las demostraciones matemáticas, acudan a expertos o simplemente tomen decisiones sin tener fundamentos de peso. Lo anterior confluente a que técnicas como estas, que poseen un poder aplicativo amplio para la solución de problemas económicos o administrativos, no sean utilizadas ampliamente por los empresarios o tomadores de decisiones. A continuación se detallan los resultados en el diseño de un modelo de Análisis de Componentes Principales para la sección reducida de variables para realizar un análisis financiero.

Resultados y discusión

Selección de indicadores financieros utilizando el Análisis de Componentes Principales

Hasta el momento se han mostrado los mecanismos que se utilizan para desarrollar la técnica estadística ACP y así reducir el número de variables que explican un hecho. A continuación se verá la aplicación de ACP utilizando *Minitab16*®, el cual es un *software* estadístico que, dentro de sus aplicaciones, contiene un solucionador de componentes principales; su capacidad máxima de variables a reducir es de 20 (esto fue constatado por los autores en la versión de prueba).

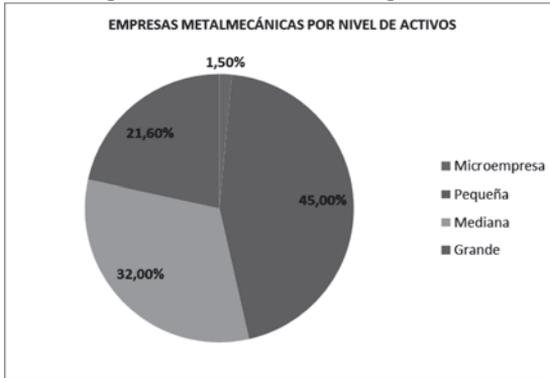
El sector metalmecánico: un caso de estudio

Sánchez, Hincapié & Quintero (2009) definen que el sector metalmecánico hace referencia a todas las industrias que producen equipos y piezas metálicas; también a las empresas que son proveedores de partes de las demás industrias, siendo el insumo básico los metales y aleaciones de hierro. El sector metalmecánico objeto de estudio se compuso de 338 empresas, entre las cuales se encontraron empresas de todos los tamaños (de acuerdo con el nivel de activos): micro, pequeñas, medianas y grandes (tabla y figura 1).

Tabla 1. Empresas metalmeccánicas 2009-2010 por nivel de activos

DISTRIBUCIÓN DEL TAMAÑO DE LAS EMPRESAS SEGÚN ACTIVOS TOTALES SECTOR METALMECÁNICO AÑOS 2009 -2010				
TAMAÑO	DESCRIPCIÓN	TOPE	No. EMPRESAS	FRECUENCIA RELATIVA
Microempresa	Hasta 500 SMMLV	267.800,00	5	1,5%
Pequeña	Superior a 500 y hasta 5.000	2.678.000,00	152	45,0%
Mediana	Superior a 5.000 y hasta 30.000	16.068.000,00	108	32,0%
Grande	Superior a 30.000	> 16.068.000,00	73	21,6%
TOTAL NÚMERO DE EMPRESAS SECTOR			338	100,0%
SMMLV 2011 (Cifra en miles de pesos)		535,60		

Figura 1. Empresas metalmeccánicas por nivel de activos



En la figura 1 se observa que, a nivel país y de acuerdo con las empresas de este sector que están registradas en la Superintendencia de Sociedades, la mayoría de empresas que se dedican a actividades de metalmeccánica se concentran en el tamaño “pequeñas”. El segundo grupo de empresas son de carácter medianas; tan sólo se evidenciaron cinco microempresas, dentro de este sector.

El ACP se fundamenta, principalmente, en lo que se refiere al tipo de datos: en los datos que son heterogéneos y los que son homogéneos. Los datos son heterogéneos cuando cada una de las observaciones⁸ difiere mucho de las demás en dimensión (Kolman y Hill, 2006), es decir, puede haber una empresa que tenga utilidades netas de 40 millones y otra que tenga utilidades netas de dos millones. Siguiendo este lineamiento se deduce que los datos homogéneos son los que tienen dimensiones concentradas en rangos reducidos y que el ACP me permite comparar datos heterogéneos sin que afecte negativamente los resultados generados por los componentes principales.

En el caso de la información recolectada acerca de las 348 empresas del sector metalmeccánico registradas en la Superintendencia de Sociedades, se encontró que los valores de las diferentes empresas estaban muy dispersos; entonces se pudo aseverar que se contaba con información heterogénea.

8 En este caso cada una de las cuentas de los estados financieros de las empresas consideradas.

Los estados financieros⁹ ya estandarizados poseían en total 68 cuentas que resumían la situación económica de la empresa. De todas estas cuentas se necesitaba escoger las que reflejaran la situación final de la empresa de manera simplificada. Las cuentas a considerar para la aplicación de la técnica del ACP se especifican en la tabla 2:

Tabla 2. Cuentas consideradas en el ACP

LISTA DE CUENTAS DE ESTADOS FINANCIEROS

1	Total Activo Corriente
2	Total Activo no Corriente
3	Total Activo
4	Total Pasivo Corriente
5	Total Pasivo no Corriente
6	Total Pasivo
7	Total Patrimonio
8	Total Pasivo y Patrimonio
9	Ingresos Operacionales
10	Costo de Ventas y Prestación de Servicios
11	Gastos Administración y Ventas
12	Utilidad Operacional
13	Ingresos no Operacionales
14	Gastos no Operacionales
15	Ganancias y Pérdidas

Gracias a la técnica estadística del análisis de componentes principales, este número de cuentas se puede disminuir sin perder poder de explicación significativo en el estudio del estado financiero de las empresas de un sector económico.

La aplicación de ACP contenida en *Minitab16*® es desarrollada a través del algoritmo QR de matrices. Aunque es importante resaltar que existen otros métodos numéricos diferentes al de *Gram-Schmidt* que se utilizan para tal fin.

Aplicación del ACP en las empresas metalmecánicas utilizando *Minitab16*®

Antes de analizar los resultados obtenidos para el sector metalmecánico, es necesario detallar un poco el significado de los informes estándares que genera un estudio de ACP.

Para el caso de *Minitab16*®, el informe generado después de utilizar la herramienta se compone de:

1. Los valores propios (eigenvalores) de cada componente principal.
2. La varianza explicada proporcional de cada uno de los componentes principales.
3. La varianza explicada acumulada consecutiva de los componentes principales.
4. La lista de los componentes principales con sus respectivos coeficientes factoriales.

⁹ Balance general, estado de resultados y flujo de efectivo.

Los primeros tres puntos son los que se analizan con el objetivo de determinar el número de componentes principales a elegir para el posterior análisis de las variables más significativas. En cuanto al cuarto punto, este permite ver los coeficientes factoriales de cada componente principal; lo que permite determinar cómo va a quedar reducido, finalmente, el conjunto de variables que representan en mayor medida el comportamiento del hecho objeto de estudio, lo cual se logra analizando la correlación entre cada componente principal y sus respectivos coeficientes factoriales.

Resultados y análisis

En el caso de estudio del sector metalmecánico (Anexo), después de realizar el ACP con los 15 indicadores financieros mostrados en la tabla 2, se encontró con que los 2 primeros componentes principales, los cuales presentaron los mayores eigenvalores, explicaban la varianza en 96,6%. Para determinar si estos componentes eran suficientes para estudiar sus coeficientes factoriales, se debió analizar la sensibilidad de la varianza explicada, acumulada al agregar o excluir un nuevo componente principal. Al caso, si se agregara un nuevo componente principal, la varianza explicada se aumentaría hasta el 98,6%, lo que es una subida de 2 puntos porcentuales; si se excluyera un componente principal de los que sumados explicaban una varianza de 96,6%, tal varianza se disminuiría en 4,6%, lo que dejaría una explicación de la varianza del 92%. Es importante recalcar que los componentes principales son ordenados de mayor a menor de acuerdo con el valor propio de cada uno de los componentes.

En este orden de ideas, finalmente se encontró que dos componentes principales eran los que podrían contener las variables explicadoras del comportamiento financiero de las empresas del sector metalmecánico, cada uno con un valor propio de 13,797 y 0,690 respectivamente.

Los indicadores financieros elegidos como explicadores del sector metalmecánico, se eligieron de acuerdo con el grado de relación con su respectivo componente. Para el primer componente se tuvieron correlaciones muy similares entre cada uno de los coeficientes factoriales con el componente: para las correlaciones positivas, estuvieron entre 0,269 y 0,257; para las negativas solo hubo dos valores -0,213 y -0,251. Para el segundo componente se encontraron 3 indicadores financieros con alta correlación con su respectivo componente:

- *Utilidad operacional* con una correlación de -0,623.
- *Total pasivo no corriente* con correlación de -0,487.
- *Ingresos no operacionales*, correlación de 0,374.

Metodología y selección de indicadores utilizada en la revista Dinero en el mes de Mayo de 2012 en su edición especial “5.000 empresas”

La revista Dinero en su edición especial “5.000 empresas”, la cual fue distribuida en los meses de mayo y junio del año 2012, muestra los resultados de un estudio (en el cual el Grupo Bancolombia participó) para determinar el *ranking* de las 5.000 empresas más grandes y representativas de los sectores económicos más consolidados del país, fundamentado en un diagnóstico financiero el cual a su vez se apoyó en indicadores de rentabilidad, liquidez, operación y endeudamiento de cada sector. En la definición de la metodología, la revista enuncia los siguientes indicadores como variables a considerar:

- Margen Bruto.
- Margen operacional.
- Margen neto.
- Utilidad sobre activos.
- Rotaciones (cuentas por pagar, cuentas por cobrar e inventarios).
- EBITDA.
- EBITDA sobre ventas.
- Obligaciones financieras a EBITDA.
- Ventas sobre activos.
- Pasivo frente a activo
- Utilidad sobre patrimonio.

Se puede apreciar en la lista anterior que para el estudio fueron elegidos 11 indicadores de desempeño y viabilidad. Pero, finalmente, en los resultados del *ranking*, el estudio de la revista sólo considera:

- Ventas.
- Activos.
- Utilidad neta.
- Utilidad/Patrimonio.
- Obligaciones financieras/EBITDA.

De las que definieron en la metodología sólo fueron utilizadas:

- Obligaciones financieras/EBITDA.

El estudio arroja unas tablas de *ranking* de acuerdo a los sectores, y queda determinada, como variable única a considerar en el *ranking*, el indicador Ventas. Finalmente, se puede apreciar que el estudio no deja evidencia alguna para saber los criterios empleados para seleccionar el grupo de variables que hicieron parte del estudio.

Conclusiones

Los resultados indican que el comportamiento financiero de las empresas del sector metalmeccánico puede ser explicado casi en la misma proporción con estas 3 variables que con el conjunto de las 15 variables de la tabla 2. Es decir, esto genera un criterio para los analistas o tomadores de decisiones empresariales, encaminado al diagnóstico de la empresa con unos indicadores clave que representan en sí todos los resultados de la compañía. Lo más importante es que detrás de este posible criterio existe todo un fundamento matemático y estadístico riguroso.

El ACP permite a los analistas financieros y a los tomadores de decisiones importantes en las compañías, determinar, con un criterio estadístico, simplificado y preciso, cuáles son los indicadores financieros que realmente evidencian el comportamiento de la empresa. Esto se logra a través de la exclusión de la información redundante y del conocimiento técnico que tenga el analista o tomador de decisiones en cuanto al significado de los indicadores financieros.

La subjetividad por parte de los que manipulan la herramienta de ACP es primordial, ya que los resultados estadísticos que se obtienen solo brindan unos valores numéricos que pueden ser interpretados de forma diferente, dependiendo del nivel de formación respecto a la disciplina financiera, en este caso.

Los fundamentos empíricos han sido los criterios más adoptados por parte de los investigadores para darle respuesta a los problemas de gestión empresarial, como se pudo evidenciar en la publicación de la revista Dinero. Pero es importante dilucidar la gran acogida que han venido teniendo las aplicaciones de las ciencias matemáticas. También se resalta que estas soluciones cuantitativas deben ser complementadas con los análisis fundamentales del entorno económico, político y social en general, con el propósito de una toma de decisiones correcta.

El Análisis de Componentes Principales es una herramienta estadística que se fundamenta en el álgebra lineal y en la solución numérica a través de las iteraciones y que, además, determina cuál es la cantidad mínima necesaria de variables para explicar acertadamente un hecho.

Los investigadores que aplican las teorías científicas en la solución de problemas triviales, deben ser conscientes de que es importante explicar los procedimientos tanto para la aplicación como para la demostración debido a que esta última queda referenciada para los expertos en una materia y la aplicación, para personas tomadoras de decisiones importantes en la economía y la sociedad.

Los algoritmos numéricos utilizados para encontrar las raíces de un sistema de ecuaciones son la mejor alternativa debido a la simplicidad que implica programarlos en un sistema informático automatizado.

Referencias

- Berenson, M. L., Levine, D. M. y Krehbiel, T. C. (2001). Regresión lineal simple y correlación. En: *Estadística para administración*. (2 ed.) (p. 504-505). México: Pearson Educación
- Cruz, J. S., Villareal, J. y Rosillo, J. (2005). Indicadores financieros como medidores de riesgo. En: *Finanzas Corporativas* (p. 599-626). Bogotá: Thomson Learning
- CSI /ITESM (2008). *Factorización QR*. Departamento de matemáticas. México: CSI
- García Serna, O. L. (1999). Revisión de los estados financieros. En: *Administración Financiera: Fundamentos y aplicaciones* (3 ed.) (p. 62-63). Bogotá: Textos y sistemas.
- Gitman, L.J. (2003). *Principios de administración financiera* (10 ed.) México: Pearson Educación
- Grossman, S. (1996). *Álgebra lineal* (5 ed.) Bogotá: Editorial Mc Graw Hill
- Guisande Gonzáles, C., Barreiro Felpeto, A., Maneiro Estraviz, I., Riveiro Alarcón, I. & Vergara Castaño, A. R. (2006). *Tratamiento de datos*. España: Ediciones Díaz de Santos
- Kolman, B. y Hill, D. R. (2006). Método de Gram-Schmidt; factorización QR de una matriz $m \times n$. En: *Álgebra lineal* (p. 356-379). México: Pearson Education
- Larson, R. y Edwards, B. H. (2002). *Introducción al álgebra lineal*. México: LIMUSA Noriega editores
- Levin, R. I. y Rubin, D. S. (2004). Regresión múltiple y modelado. En: *Estadística para administración y economía* (7 ed.) (p. 566-593). México: Pearson Educación
- Nakos, G. & Joyner, D. (1999). Productos punto e interno. En: *Álgebra lineal con aplicaciones*. (1 ed.) (p. 531-557). México: Thomson Editores
- Publicaciones Semana Dinero (2012, 05, 25). Cinco mil empresas. *Revista Dinero*. (398) (p. 170-256)

- Rodríguez Rojas, O. (2009). *Análisis en componentes principales*. San José: Universidad de Costa Rica
- Ross, S. A, Westerfield, R. W. y Jaffe, J. F. (2009). Análisis de estados financieros y planeación a largo plazo. En: *Finanzas Corporativas* (8 ed.) (p. 41-69). México: Mc Graw Hill
- Sánchez Castro, J. J., Hincapié R., Y. P. & Quintero Flórez, J. (2009). El capital intelectual en las empresas del sector metalmecánico Risaralda Colombia. *Scientia et Technica* Año XV. Universidad Tecnológica de Pereira (p.123-128)
- Shlens, J. (2009). *A tutorial on Principal Component Analysis*. New York: Center for neural Science, New York University
- Smith, L. (2002). *A tutorial on Principal Component Analysis*. USA: Cornell University
- Vicente Villardón, J. L. (2004). *Análisis de componentes principales*. Salamanca: Universidad de Salamanca: Departamento de Estadística
- Webster, A. L. (2000). Regresión múltiple y correlación. En: *Estadística aplicada a los negocios y la economía* (3 ed.) (p. 382-412). Bogotá: Irwin Mc Graw Hill
- Weston, J. F. y Brigham, E. F. (1994). Análisis de estados financieros. En: *Fundamentos de Administración Financiera* (10 ed.) (p. 53-104). México: Mc Graw Hill

ANEXO

Bienvenido a Minitab, presione F1 para obtener ayuda.

Análisis de componente principal: TOTAL ACTIVO. TOTAL ACTIVO. TOTAL ACTIVO. TOT

Análisis de los valores y vectores propios de la matriz de correlación

Valor propio	13,797	0,690	0,296	0,119	0,054	0,031	0,010	0,001	0,001
Proporción	0,920	0,046	0,020	0,008	0,004	0,002	0,001	0,000	0,000
Acumulada	0,920	0,966	0,986	0,993	0,997	0,999	1,000	1,000	1,000

Valor propio	0,000	0,000	0,000	-0,000	-0,000	-0,000
Proporción	0,000	0,000	0,000	-0,000	-0,000	-0,000
Acumulada	1,000	1,000	1,000	1,000	1,000	1,000

Variable	PC1	PC2	PC3	PC4	PC5	PC6
TOTAL ACTIVO CORRIENTE	0,267	-0,005	0,084	0,276	0,097	-0,118
TOTAL ACTIVO NO CORRIENTE	0,267	0,014	0,039	-0,156	-0,494	-0,164
TOTAL ACTIVO	0,269	0,004	0,063	0,068	-0,188	-0,141
TOTAL PASIVO CORRIENTE	0,264	-0,113	-0,091	0,490	0,064	-0,239
TOTAL PASIVO NO CORRIENTE	0,236	-0,487	-0,258	-0,619	0,136	-0,291
TOTAL PASIVO	0,264	-0,193	-0,128	0,270	0,081	-0,255
TOTAL PATRIMONIO	0,257	0,256	0,302	-0,194	-0,519	0,013
TOTAL PASIVO Y PATRIMONIO	0,269	0,004	0,063	0,068	-0,188	-0,141
41 INGRESOS OPERACIONALES	0,265	-0,085	0,290	0,037	0,260	0,053
61 MENOS COSTO DE VENTAS Y DE P	0,265	-0,077	0,256	0,030	0,276	-0,014
GASTOS ADMON Y VENTAS	0,264	0,159	0,178	0,047	0,124	0,448
UTILIDAD OPERACIONAL	-0,213	-0,623	0,600	0,060	-0,142	0,145
42 MAS INGRESOS NO OPERACIONALE	0,253	0,374	0,104	-0,280	0,349	0,004
53 MENOS GASTOS NO OPERACIONALE	0,265	-0,155	0,034	-0,217	0,138	0,382
59 GANANCIAS Y PERDIDAS	-0,251	0,236	0,498	-0,149	0,238	-0,585

Variable	PC7	PC8	PC9	PC10	PC11	PC12
TOTAL ACTIVO CORRIENTE	-0,112	-0,150	0,630	0,079	0,401	0,404
TOTAL ACTIVO NO CORRIENTE	0,074	0,039	-0,542	-0,039	0,370	0,374
TOTAL ACTIVO	-0,023	-0,059	0,067	0,022	0,103	-0,356
TOTAL PASIVO CORRIENTE	-0,041	-0,186	-0,234	0,030	-0,604	0,084
TOTAL PASIVO NO CORRIENTE	0,262	-0,074	0,209	-0,045	-0,155	0,022
TOTAL PASIVO	0,021	-0,167	-0,147	0,015	0,172	-0,147
TOTAL PATRIMONIO	-0,077	0,082	0,336	0,030	-0,446	-0,034
TOTAL PASIVO Y PATRIMONIO	-0,023	-0,059	0,067	0,022	0,116	-0,343
41 INGRESOS OPERACIONALES	0,070	0,460	-0,066	-0,085	-0,180	0,477
61 MENOS COSTO DE VENTAS Y DE P	-0,020	0,564	-0,065	-0,086	0,165	-0,436
GASTOS ADMON Y VENTAS	0,735	-0,315	-0,043	0,068	0,023	-0,061
UTILIDAD OPERACIONAL	-0,106	-0,283	-0,025	-0,262	0,010	-0,027
42 MAS INGRESOS NO OPERACIONALE	-0,361	-0,381	-0,135	-0,539	0,000	-0,000
53 MENOS GASTOS NO OPERACIONALE	-0,444	-0,141	-0,169	0,668	-0,000	0,000
59 GANANCIAS Y PERDIDAS	0,140	-0,144	-0,099	0,403	-0,000	0,000

Variable	PC13	PC14	PC15
TOTAL ACTIVO CORRIENTE	0,010	-0,164	-0,184
TOTAL ACTIVO NO CORRIENTE	0,009	-0,152	-0,170
TOTAL ACTIVO	0,701	0,461	-0,116
TOTAL PASIVO CORRIENTE	-0,001	-0,139	-0,353
TOTAL PASIVO NO CORRIENTE	-0,000	-0,036	-0,090
TOTAL PASIVO	-0,002	-0,173	0,776
TOTAL PATRIMONIO	-0,003	-0,269	0,267
TOTAL PASIVO Y PATRIMONIO	-0,713	0,447	-0,124
41 INGRESOS OPERACIONALES	-0,004	0,473	0,235
61 MENOS COSTO DE VENTAS Y DE P	0,003	-0,432	-0,215
GASTOS ADMON Y VENTAS	0,000	-0,060	-0,030
UTILIDAD OPERACIONAL	0,000	-0,027	-0,013
42 MAS INGRESOS NO OPERACIONALE	0,000	-0,000	-0,000
53 MENOS GASTOS NO OPERACIONALE	-0,000	0,000	0,000
59 GANANCIAS Y PERDIDAS	-0,000	0,000	0,000

