ANÁLISIS FACTORIAL

Dr. Cs. Ruiz Aranibar, Gustavo

gustavoruiz432@hotmail.com

RESUMEN

El Análisis Factorial (AF) es una técnica de reducción de datos, que sirve para encontrar grupos homogéneos de variables a partir de un conjunto numeroso de variables. Estos grupos homogéneos se forman con las variables que se correlacionan mucho entre si y procurando, inicialmente, que unos grupos sean independientes de otros.

El AF es una técnica de reducción de dimensionalidad de los datos; su propósito final, consiste en buscar un número mínimo de dimensiones capaces de explicar el máximo de información contenida en los datos. A diferencia de lo que ocurre en otras técnicas como el análisis de varianza o el de regresión, en el AF todas las variables del análisis cumplen el mismo papel, todas ellas son independientes en el sentido de que no existe a priori una dependencia conceptual de unas variables sobre otras.

PALABRAS CLAVE

Análisis factorial, factores, correlación, desviación estándar, valor propio, vector propio, variable aleatoria, varianza común, denormalizacion, iterar, rotación.

1. INTRODUCCIÓN

El Análisis Factorial (AF) es usada en ciencias sociales, mercadotecnia, gestión de producto, investigación operativa, medicina y otras ciencias aplicadas que tratan con grandes cantidades de datos. El objetivo, es explicar la mayor parte de la variabilidad entre varias variables aleatorias observables, en términos de un número menor de variables aleatorias no observables llamadas factores. Las variables aleatorias no observables se modelan como combinaciones lineales de los factores más términos de errores.

El AF es un método del análisis estadístico múltiple, utilizado para el estudio e interpretación de las correlaciones entre un grupo de variables, partiendo de la idea de que dichas correlaciones no son aleatorias, sino que se deben a la existencia de factores comunes entre ellas y el objetivo del AF es la identificación y cuantificación de dichos factores comunes.

De una manera general, el AF se puede aplicar cuando existen relaciones entre las variables, es decir para explicar las relaciones con la ayuda de un número limitado de factores independientes, estando basada en el análisis de correlaciones existentes entre las variables medidas sobre un grupo de “n” individuos u objetos.

Las variables observables deben ser homogéneas y correlacionadas, sin que alguna predomine sobre las demás. La información estadística en el AF es de carácter multidimensional, por lo tanto, la geometría, el cálculo matricial, el álgebra lineal y la estadística son fundamentales en el desarrollo del presente trabajo.

2. BREVE HISTORIA DE LOS PRECURSORES DEL AF.

Jacobi Carl Gustav, (Potsdam, actual Alemania, 1804-Berlín, 1851) Matemático alemán. Hijo de una familia de banqueros de origen judío, estudió en la Universidad de Berlín, donde se doctoró en 1825. Convertido al cristianismo, tuvo oportunidad de acceder a un puesto de profesor en la Universidad de Königsberg. Destacadísimo pedagogo, influyó en numerosas generaciones posteriores de matemáticos alemanes. Sus trabajos más relevantes se produjeron en el campo del álgebra matricial y lineal, en el que introdujo y desarrolló el concepto de determinante, aplicándolo asimismo al estudio de las funciones de variables múltiples. Entre 1826 y 1827 estableció, independientemente del noruego Niels Henrik Abel, los principios fundamentales de la teoría de las funciones elípticas. En el ámbito de la teoría de números, demostró el teorema de Bachet sobre el total de las descomposiciones posibles de un entero, y en el de la mecánica física, trató con profundidad y rigor el problema de los tres cuerpos. Su obra más notable es sobre la formación y propiedades de los determinantes (1841).

Charles Edward Spearman. (Londres, nació el día 10 de septiembre de 1863, falleció el 7 de septiembre de 1945). Psicólogo de profesión, estudió Estadística y logró desarrollar notables aplicaciones de la estadística en el campo de la Psicología, desarrollando el AF. Propuso la existencia de un factor general de inteligencia (Factor g) que subyace a las habilidades para la ejecución de las tareas intelectuales. Estudió en las universidades de Leipzig, donde alcanzó el título cuando tenía 40 años. Además, cursó estudios en Wurzburgo y Göttingen y enseñó e investigó en la Universidad de Londres (1907 – 1931). El AF tiene su origen en las investigaciones realizadas con los trabajos en Psicología, (1904).

Creó y desarrollo la metodología de los llamados experimentos factoriales para la estadística, que son aquellos experimentos en los que se estudia simultáneamente dos o más factores, y donde los tratamientos se forman por la combinación de los diferentes niveles de cada uno de los factores.

Por todo esto, es considerado uno de los grandes estadisticos de todos los tiempos. Su método, inscrito en las matemáticas experimentales, estudia las dimensiones del campo empírico. Sus aportes metodológicos, no sólo se han constituido en herramientas fundamentales para algunos ámbitos de la psicología, sino que son instrumentos para la ciencia estadística. El desarrollo de Spearman es útil en todas las ciencias sociales que requieran de técnicas de estadística correlacional para poder interpretar la información.

Thurstone Louis León. (Chicago, 1887 - Chapel Hill, 1955) Psicólogo estadounidense. Se doctoró en la Universidad de Chicago, donde dio clases la mayor parte de su vida. Especializado en Psicometría, desarrolló nuevas técnicas para medir las cualidades mentales. Realizó y publicó varias escalas de actitud que pretendían medir la influencia de la propaganda en los prejuicios del hombre; también se interesó por la medición del aprendizaje e intentó expresar a través de unidades absolutas el desarrollo del aprendizaje.

Fue uno de los primeros en utilizar el AF para medir la inteligencia; es por su trabajo sobre AF por lo que es más conocido Thurstone, ya que él y sus seguidores lo aplicaron a múltiples problemas, como el análisis de las capacidades perceptivas humanas o el desarrollo de nuevos tests de aptitudes. Se interesó también por las características de la personalidad y publicó un test de tendencias Psiconeuróticas.

En la evolución de la Psicometría, y en particular en el campo de la medición de la inteligencia, hay que destacar las aportaciones de Charles Spearman, que propuso en 1927 la distinción entre un "factor g" (factor general), común a todas las pruebas de medición y presente en cualquier tarea intelectual, y un "factor s" (específico), asociado a cada operación en particular. Thurstone dio un paso más en la distinción de Spearman al identificar en 1934, con la ayuda de las técnicas estadísticas de AF, siete aptitudes primarias incluidas en la inteligencia: comprensión verbal, fluidez verbal, aptitud numérica, visualización espacial, velocidad perceptiva, memoria y razonamiento. Ello llevó a Thurstone a concebir la inteligencia como una combinación de varias capacidades distintivas; de este modo, el factor general "g" debe entenderse como factor secundario, detectable únicamente gracias a las correlaciones entre las aptitudes primarias.

La técnica para la medición conocida como AF múltiple, que puede manejar varios factores de capacidad simultáneamente, se convirtió a partir del impulso que le dio Thurstone en un potente instrumento de análisis estadístico aplicado a la investigación psicológica, cuyas repercusiones se extienden hasta nuestros días. Sus trabajos sobre AF pudieron aplicarse a múltiples problemas, como el análisis de las capacidades perceptivas humanas o el desarrollo de nuevos tests de aptitudes. Se interesó por las características de la personalidad y elaboró un test de tendencias psiconeuróticas. Fundador y director de la revista Psicometrika, entre sus obras destacan The nature of intelligence (1924) y Vectors of the mind (1935).

Karl Pearson. (27 de marzo de 1857, 27 de abril de 1936). Es uno de los fundadores de la estadística, desarrollo en el ámbito de la investigación, las ideas de Galton F. sobre regresión y correlación. Hijo de una familia acomodada estudió en la University College School y posteriormente en Cambridge especializándose en matemáticas. Años más tarde viajó a Berlín y Heidelberg donde estudió Literatura alemana medieval, destacando en esta área y colaborando en Cambridge en los estudios germánicos. Estudió también abogacía en la asociación profesional de Inner Temple.

Karl Pearson fue uno de los miembros fundadores de la "Escuela Biométrica", que estudiaba la aplicación de la Estadística en materias como la biología, campo en el que Pearson poseía diversos conocimientos que sirvieron en futuras generaciones de investigadores y siempre será recordado por sus teorías positivistas radicales, su investigación estadística en materia de biología y por ser el fundador de la bioestadística.

En Julio de 1900, una de las más importantes contribuciones de Pearson a la Estadística fue presentada en la publicación de un artículo. Esta contribución era el Test de la $\chi2$. Pearson usó esta fórmula para obtener la distribución muestral de $\chi2$ en grandes muestras, las cuales estaba particularmente interesado en estudiar, como una función de k, la cual resultó ser una forma especial de la distribución de Pearson tipo 3, ahora conocida como "distribución $\chi2$ para K-1 grados de libertad". Además, daba una pequeña tabla de la integral de la distribución de $\chi2$ desde 1 a 70 y para k desde 3 a 20. Este test de la $\chi2$ de bondad del ajuste es una de las mayores y más útiles contribuciones de Pearson a los tests estadísticos.

Tras el reconocimiento por su trabajo sobre curvas de distribución, Pearson continuó recibiendo reconocimientos y honores. En 1893, comenzó su serie de 18 artículos titulados “Mathematical Contributions to the Theory of Evolution”, que contendrían parte de su trabajo más valioso. El mismo año que empezó estos artículos, Pearson acuñó el término “desviación estándar”. Entre 1906 y 1914 Pearson estuvo consagrado al desarrollo de un centro de postgrado para promover el desarrollo de la estadística como una rama de las matemáticas aplicadas. En el verano de 1933, tras una larga vida consagrada al avance estadístico, Pearson abandonó su trabajo en la Universidad.

El hecho de que tras la retirada de Pearson el departamento de estadística aplicada fuera dividido en dos unidades independientes muestra el importante trabajo soportado por Pearson. Incluso después de la muerte de Karl Pearson en 1936, su apellido continúa siendo uno de los más destacados en el campo de las matemáticas. No hay duda de que las contribuciones de Pearson a lo largo de su vida consolidaron la Estadística como una disciplina por derecho propio.

3. CONCEPTOS MATEMÁTICO-ESTADÍSTICOS NECESARIOS EN EL AF.

El análisis estadístico requiere del instrumento matemático, por esto los datos deben presentarse en una forma susceptible de tratamiento matemático de ahí que la aplicación del AF, requiere en principio que se disponga de una matriz de información de datos cuantitativos *X*, de orden *n*m*, (*n*=observaciones y *m*=variables), *n > m*:

Matriz de datos observados:

$$X = X_{ij} = \begin{bmatrix} x_{11} & \cdots & x_{1j} \\ \vdots & \ddots & \vdots \\ x_{i1} & \cdots & x_{ij} \end{bmatrix} \qquad \begin{matrix} i = 1,2,\dots,n \\ \mathrm{j} = 1,2,\dots,\mathrm{m} \end{matrix}$$

A partir de esta información, se tiene que aplicar las siguientes ecuaciones que son utilizadas para determinar los estadísticos de: los promedios, desviaciones estándar, suma de los productos cruzados de desviaciones de las medias y los coeficientes de correlación.

Promedios de cada columna:

$$\overline{X_j} = \frac{\sum_{i=1}^{m} X_{ij}}{n} \qquad j = 1,2,\dots,m$$

Media temporal para obtener un cálculo más exacto de :

$$T_j = \frac{\sum_{i=1}^{m} X_{ij}}{m}$$

Suma de los productos cruzados de las desviaciones:

$$S_{jk} = \sum_{i=1}^{n} (X_{ij} - T_j)(X_{ik} - T_k) - \frac{\sum_{i=1}^{n}(X_{ij} - T_j)\sum_{i=1}^{n}(X_{ik} - T_k)}{n}$$

$$i = 1,2,\dots,n \quad , \quad j = 1,2,\dots,m \quad , \quad k = 1,2,\dots,m$$

Coeficientes de correlación y su dominio:

$$r_{jk} = \frac{S_{jk}}{\sqrt{S_{jj}}\sqrt{S_{kk}}} \qquad -1 \le r_{jk} \le +1$$

$$j = 1,2,\dots,m \quad y \quad k = 1,2,\dots,m$$

Desviaciones estándar:

$$S_j = \frac{\sqrt{S_{jj}}}{\sqrt{n-1}} \qquad j = 1,2,\dots,m$$

Matriz identidad.

Una matriz de identidad es usada como una aproximación a R (matriz de correlación).

$$V_{ij} = I \qquad \begin{matrix} \forall\, i = j \;\; V_{ij} = 1 \\ \forall\, i \ne j \; V_{ij} = 0 \end{matrix} \quad y$$

Valores propios[1].

Se llama valor propio de una matriz simétrica A de orden n, las n soluciones de la ecuación:

$$|A - \lambda I| = 0 \;\; \lambda_1, \lambda_2, \dots, \lambda_n$$

El desarrollo de este determinante, da lugar en efecto al nacimiento a una ecuación de grado *n* en λ, llamada ecuación característica:

$$|A - \lambda I| = (-\lambda)^n + c_{n-1}(-\lambda)^{n-1} + \cdots + c_1(-\lambda) + c_0 = 0$$

Esta ecuación posee de una manera general *n* raíces. En particular, el coeficiente c_{n-1} es la suma de los elementos de la diagonal de la matriz inicial A, llamada traza de la matriz, mientras que el término independiente c_0 es el valor del determinante de A:

$$c_{n-1} = tr\ A = \sum_{i=1}^{n} a_{ii} \qquad c_0 = |A|$$

De una manera general, las raíces de la ecuación característica pueden ser reales o complejas, positivas, nulas o negativas, distintas o diferentes. En el caso de matrices simétricas, que interesan en forma particular, deben ser todas reales, teniéndose entonces:

$$\sum_{i=1}^{n} \lambda_i = tr\ A \qquad \prod_{i=1}^{n} \lambda_i = |A|$$

La determinación de los valores propios puede estar basado en el algoritmo de Rutishauer, que consiste en descomponer la matriz A dada, como producto de dos matrices B y H, tal que B es una matriz triangular inferior con todos los elementos de la diagonal principal igual a 1, y H es una matriz triangular superior. Por un proceso iterativo se consigue transformar la matriz A en una matriz triangular superior tal que, los valores propios buscados son los elementos de la diagonal principal, existiendo para esta determinación también otros métodos.

La determinación de los valores propios puede estar basado en el algoritmo de Rutishauer, que consiste en descomponer la matriz A dada, como producto de dos matrices B y H, tal que B es una matriz triangular inferior con todos los elementos de la diagonal principal igual a 1, y H es una matriz triangular superior. Por un proceso iterativo se consigue transformar la matriz A en una matriz triangular superior tal que, los valores propios buscados son los elementos de la diagonal principal, existiendo para esta determinación también otros métodos.

Vectores propios[2]

Sea A una matriz de orden n sobre un cuerpo K. Un escalar λ se denomina un valor propio de A si existe un vector (columna) no nulo v, es decir:

$$\lambda \in K \;\; \nu \in K^n \;\; A\nu = \lambda\nu$$

Todo vector que satisfaga esta relación se llama un vector propio de A perteneciente al valor propio λ_i.

Norma inicial:

$$\nu_I = 0$$

$$\nu_I = \left(\sum_{i \leq k} 2 * A_{ik}^2\right)^{1/2}$$

Norma final calculada en cada etapa de ese dominio:

$$\nu_F = \frac{\nu_I * 10^{-6}}{N}$$

1 Llamados, autovalores o valores característicos. En ingles: laten root, characteristic root o eigenvalue. En francés: valeurs propres o valeurs caractéristiques.

2 Llamados, autovectores o vectores característicos. En ingles: laten vector, characteristic vector o eigenvector. En francés: vecteurs propes o vecteurs caractéristiques.

Esta norma final es un conjunto suficientemente pequeña que requiere que cualquier elemento fuera de la diagonal debe ser más pequeño que V_F en magnitud absoluta define la convergencia del proceso.

Indicador inicializado. Este indicador posteriormente es utilizado para determinar cuándo cualquier elemento fuera de la diagonal ha sido encontrado, que son más grandes que en el presente inicio.

Cada elemento fuera de la diagonal es seleccionado en etapas y una transformación es ejecutada para eliminar el elemento fuera de la diagonal (pivote), como se muestra en las siguientes ecuaciones:

$$\lambda = -A_{1m}$$

$$\mu = \frac{A_{11} - A_{mm}}{2}$$

$$\omega = signo\ (\mu)\frac{\lambda}{\sqrt{\lambda^2 + \mu^2}}$$

$$sen\ \theta = \frac{\omega}{\sqrt{2*(1+\sqrt{1-\omega^2})}}$$

$$\cos\ \theta = \sqrt{1 - sen^2\theta}$$

$$B = A_{i1} * \cos\ \theta - A_{im} * sen\ \theta$$

$$C = A_{i1} * sen\ \theta - A_{im} * \cos\ \theta$$

$$B = R_{i1} * \cos\ \theta - R_{im} * sen\ \theta$$

$$R_{im} = R_{i1} * sen\ \theta + R_{im} * \cos\ \theta$$

$$R_{i1} = B$$

$$A_{11} = A_{i1} * \cos^2\theta + A_{mm} * sen^2\theta - 2 * A_{lm} * sen\theta * \cos\theta$$

$$A_{mm} = A_{11} * sen^2\theta + A_{mm} * \cos^2\theta + 2 * A_{lm} * sen\theta * \cos\theta$$

$$A_{lm} = (A_{11} - A_{mm}) * \sin\theta * \cos\theta + A_{lm} * (\cos^2\theta - sen^2\theta)$$

Los cálculos anteriores son repetidos hasta que todos los elementos del pivote sean menores que en el comienzo.

El método de diagonalización fue realizado por Jacobi y adaptado por von Neumann para las computadoras.

Se encuentra el valor de k, el número de valores propios que son más grandes o iguales que la constante especificada los valores propios determinados son arreglados en orden descendiente.

El porcentaje acumulado de esos k valores propios son:

$$d_j = \sum_{i=1}^{j} \frac{\lambda_i}{m} \qquad j = 1,\ 2, \ldots,\ k$$

m = número de valores propios (o variables) y $k \le m$

Se calcula los coeficientes de cada factor, multiplicando los elementos de los vectores propios normalizados por la raíz cuadrada de los correspondientes valores propios, es decir se determina una matriz factor de los valores propios asociada a los vectores propios:

$$a_{ij} = v_{ij} * \sqrt{\lambda_j}$$

$i = 1,\ 2, \ldots,\ m$ son las variables

$j = 1,\ 2, \ldots,\ k$ son los valores propios retenidos $k \le m$

Posteriormente se ejecuta la rotación ortogonal de la matriz factor de orden m por k, tal que:

$$\sum_j^k \left(m \sum_i^m \left(\frac{a_{ij}^2}{h_i^2} \right)^2 - \left(\sum_i^m \left(\frac{a_{ij}^2}{h_i^2} \right) \right)^2 \right)$$

$i = 1,\ 2, \ldots,\ m$ son las variables

$j = 1,\ 2, \ldots,\ k$ son los factores

a_{ij} es la matriz retenida o saturada, para la i-iem variable del j-ime factor

h_i^2 es la varianza común j-ime variable definida a continuación

Varianza común[3].

$$h_i^2 = \sum_{j=1}^{k} a_{ij}^2 \qquad i = 1,\ 2,\ldots\ldots,\ m$$

Matriz factor normalizada:

$$b_{ij} = \frac{a_{ij}}{\sqrt{h_i^2}} \qquad i = 1,\ 2,\ldots\ldots,\ m$$

$$j = 1,\ 2,\ldots\ldots,\ k$$

Varianza para la matriz factor:

$$V_c = \sum_{j=1}^{k} \frac{m\sum_{i=1}^{m}\left(b_{ij}^2\right)^2 - \sum_{i=1}^{m}\left(b_{ij}^2\right)^2}{m^2}$$

$$c = 1,\ 2,\ldots\ldots\ldots \qquad \text{(ciclo de iteración)}$$

La prueba de convergencia se ejecuta para cuatro operaciones sucesivas, se detiene la rotación y se ejecuta la ecuación de denormalización:

Si $V_c - V_{c-1} \le 10^{-7}$

De otra manera se repite la rotación de factores hasta que la prueba de convergencia sea satisfecha.

Rotación de dos factores: Se rota dos factores normalizados b_{ij} al mismo tiempo:

1 con 2, 1 con 3, …., 1 con k, 2 con 3,…..,2 con k,….,k-1 con k. Esto constituye un ciclo de iteración.

3 En inglés: communality. En Francés: varianza commune.

Asumiendo que x y y son factores a ser rotados, donde x es el número más bajo o el factor del lado izquierda, y siguiendo la notación para la rotación estos dos factores son usados:

$$\begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \ldots\ldots\ldots & \\ \ldots\ldots\ldots & \\ x_m & y_m \end{bmatrix} * \begin{bmatrix} \cos\ \phi & -sen\ \phi \\ sen\ \phi & \cos\ \phi \end{bmatrix} = \begin{bmatrix} X_1 & Y_1 \\ X_2 & Y_2 \\ \ldots\ldots\ldots & \\ \ldots\ldots\ldots & \\ X_m & Y_m \end{bmatrix}$$

Donde x_i y y_i son los valores almacenados que serán utilizables, y X_i y Y_i son los valores normalizados deseados, que son funciones de ϕ , el ángulo de rotación.

Las etapas de cálculo son las siguientes:

Cálculo de NUM y DEN

$$A = \sum_{i=1}^{m}(x_i + y_i)(x_i - y_i)$$

$$B = 2\sum_{i=1}^{m} x_i y_i$$

$$C = \sum_{i=1}^{m}\left((x_i + y_i)(x_i - y_i) + 2x_i y_i\right)\left((x_i + y_i)(x_i - y_i) - 2x_i y_i\right)$$

$$D = 4\sum_{i=1}^{m}(x_i + y_i)(x_i - y_i)x_i y_i$$

$$NUM = D - \frac{2AB}{m}$$

$$DEN = C - \frac{(A+B)(A-B)}{m}$$

Comparación de NUM y DEN:

Los cuatro casos siguientes se producen:

NUM < DEN determinar B1

NUM > DEN determinar B2

(NUM + DEN) $\geq \varepsilon^*$ determinar B3

(NUM + DEN) $= \varepsilon$ saltar a la próxima rotación

ε^* es un factor de tolerancia pequeño

B1: $tg\ 4\theta = |\ NUM\ | / |\ DEN\ |$

si $tg\ 4\theta$

i) Si tg $4\Theta < \varepsilon$ y DEN es positivo, salta a la próxima rotación

ii) DEN es negativo entonces:

$$\cos\ \theta = sen\ \theta = \frac{\sqrt{2}}{2}$$

Luego pasa a calcular E

si $tg\ 4\theta \geq \varepsilon$

Se calcula:

$$\cos\theta = \frac{1}{\sqrt{1 + tg^2 4\theta}}$$

$$\operatorname{sen}\theta = tg4\theta^* \cos 4\theta$$

y se calcula C

$B2: ctg\ 4\theta = |NUM|/|DEN|$

i) Si $ctg\ 4\theta < \varepsilon$
$\cos 4\theta = 0\ \ y\ \ sen\ 4\theta = 1$
se pasa a calcular

Si $ctg\ 4\theta \geq \varepsilon$ se calcula:

$$sen\ 4\theta = \frac{1}{\sqrt{1 + ctg^2 4\theta}}$$

$$\cos 4\theta = ctg4\theta * sen\ 4\theta$$

se calcula C

$$B3: cos\ 4\theta = sen\ 4\theta = \frac{\sqrt{2}}{2}$$

se calcula C

C: Determinación de $\cos\theta$ y $\operatorname{sen}\theta$

$$\cos 2\theta = \sqrt{\frac{1 + \cos 4\theta}{2}}$$

$$\operatorname{sen} 2\theta = \frac{\operatorname{sen} 4\theta}{2\ \cos 2\theta}$$

$$\cos\theta = \sqrt{\frac{1 + \cos 2\theta}{2}}$$

$$\operatorname{sen}\theta = \frac{\operatorname{sen} 2\theta}{2\ \cos\theta}$$

D: Determinación de $\cos\phi$ y $\operatorname{sen}\phi$

D1. Si DEN es positivo:

$\cos\phi = \cos\theta$

$\operatorname{sen}\phi = \operatorname{sen}\theta$

Ir a D2

Si DEN es negativo, calcular:

$$\cos\phi = \sqrt{\frac{2}{2}}\cos\theta + \sqrt{\frac{2}{2}}\operatorname{sen}\theta$$

$$\operatorname{sen}\phi = \left|\sqrt{\frac{2}{2}}\cos\theta - \sqrt{\frac{2}{2}}\operatorname{sen}\theta\right|$$

Ir a D2

D2.si NUM es positivo:

$\cos\phi = |\cos\phi|$

$\operatorname{sen}\phi = |\operatorname{sen}\phi|$

Ir a E

Si NUM es negativo:

$\cos\phi = |\cos\phi|$

$\text{sen}\,\phi = -|\text{sen}\,\phi|$

E: Rotación:

Posteriormente, un componente principal en la solución, es la rotación de la matriz factor denominada varimax, desarrollada por Kaiser J. B. Carrol (1958). Transforma la matriz factorial hasta conseguir la solución que verifique que la suma de las simplicidades de los factores sea máxima. (Simplicidad = varianza de los cuadrados de las saturaciones). Rotan los factores forzando a que unas saturaciones se aproximen más a uno y las otras a cero, para facilitar así su interpretación.

$$X_i = x_i \cos\phi + y_i\, sen\,\phi$$

$$Y_i = x_i\, sen\,\phi + y_i \cos\phi$$

$$i = 1, 2, \ldots.., m$$

Después de ser completado el ciclo *k(k-1)/2* las rotaciones son completadas, se vuelve a calcular la varianza para la matriz factor.

Denormalización:

$$a_{ij} = b_{ij} * h_i$$

$i = 1, 2, \ldots.., m$ $\quad j = 1, 2, \ldots.., k$

Prueba de la matriz común:

$$f_i^2 = \sum_{j=1}^{k} a_{ij}^2 \quad \text{Matriz común final}$$

$$d_i = h_i^2 - f_i^2 \quad \text{Diferencias } i = 1, 2, \ldots.., m$$

En este cálculo, se guarda en orden el número de dimensiones independientes tan pequeño como sea posible, solamente los valores propios (o coeficientes de correlación) más grandes o igual a 1 son retenidos en el análisis.

Es de hacer notar la diferencia entre: el análisis del componente principal que es usado para determinar el número mínimo de dimensiones independientes necesarias para contar al menos de la varianza en el conjunto original de variables; en cambio, la rotación varimax es usada para reducir las columnas (factores) y no las filas (variables) de la matriz factor.

4. APLICACIÓN

La técnica del AF es aplicada a cualquier rama del saber humano, ya sea a la Psicología, minería, geología, metalurgia, medicina, odontología, agronomía, etc., donde se tenga un determinado número de muestras con sus respectivas variables, todas las que deberán estar expresadas en forma cuantitativa. Por ejemplo, en educación para llegar a determinar el rendimiento académico de los estudiantes, se llegará a considerar las notas obtenidas en cada una de las asignaturas por el estudiante, las cuales constituyen las variables. En la Tabla 1., se tiene una determinada información base, y las tablas siguientes son las obtenidas de acuerdo al programa desarrollado, conforme a la teoría señalada anteriormente, en la cual se posee 25 muestras, cada una constituida por 7 variables:

Tabla 1
Valores de las 7 variables para las 25 muestras

X_{ij}	X_1	X_2	X_3	X_4	X_5	X_6	X_7
1	3,760	3,660	0,540	5,275	9,768	13,741	4,782
2	8,590	4,490	1,340	10,022	7,500	10,162	2,130
3	6,220	6,140	4,520	9,842	2,175	2,732	1,089
4	7,570	7,280	7,070	12,662	1,791	2,101	0,822
5	9,030	7,080	2,590	11,762	4,539	6,217	1,276
6	5,510	3,980	1,300	6,924	5,236	7,304	2,403
7	3,270	0,620	0,440	3,357	7,629	8,838	8,389
8	8,740	7,000	3,310	11,675	3,529	4,757	1,119
9	9,640	9,490	1,030	13,567	13,133	18,519	2,354
10	9,730	1,330	1,000	9,871	9,871	11,064	3,704
11	8,590	2,980	1,170	9,170	7,851	9,909	2,616
12	7,120	5,490	3,680	9,716	2,642	3,430	1,189
13	4,690	3,010	2,170	5,983	2,760	3,554	2,013
14	5,510	1,340	1,270	5,808	4,566	5,382	3,427
15	1,660	1,610	1,570	2,799	1,783	2,087	3,716
16	5,900	5,760	1,550	8,388	5,395	7,497	1,973
17	9,840	9,270	1,510	13,604	9,017	12,668	1,745
18	8,390	4,920	2,540	10,053	3,956	5,237	1,432
19	4,940	4,380	1,030	6,678	6,494	9,059	2,807
20	7,230	2,300	1,770	7,790	4,393	5,374	2,274
21	9,460	7,310	1,040	11,999	11,579	16,182	2,415
22	9,550	5,350	4,250	11,742	2,766	3,509	1,054
23	4,940	4,520	4,500	8,067	1,793	2,103	1,292
24	8,210	3,080	2,420	9,097	3,753	4,657	1,719
25	9,410	6,440	5,110	12,495	2,446	3,103	0,914

Fuente: Elaboración propia

En base a esta matriz de datos iníciales, aplicando el programa de AF desarrollado: se llega a obtener las siguientes tablas:

Tabla 2
Total de las columnas

$\sum X_j$	177,500	119,330	58,720	228,346	136,455	179,186	58,654

Fuente: Elaboración propia

Tabla 3
Promedios

$\overline{X}_j$	7,10000	4,77320	2,34880	9,13384	5,45820	7,16744	2,34616

Fuente: Elaboración propia

Tabla 4
Suma de cuadrados y productos cruzados

SCP_{ij}	750,097	492,280	231,787	959,858	604,324	784,220	235,410
	492,280	388,642	180,395	669,409	375,293	505,024	130,657
	231,787	180,395	117,809	324,953	128,429	165,650	58,114
	959,858	669,409	324,953	1256,672	749,150	979,738	286,960
	604,324	375,293	128,429	749,150	649,402	848,859	272,051
	784,220	505,024	165,650	979,738	848,859	1117,969	344,935
	235,410	130,657	58,114	286,960	272,051	344,935	165,419

Fuente: Elaboración propia

Tabla 5
Desviaciones estándar

S_j	2,3238	2,4178	1,6656	3,0178	3,2733	4,5581	1,6105

Fuente: Elaboración propia

Tabla 6
Coeficientes de correlación

R_{ij}	1,0000	0,5803	0,2011	0,9113	0,2833	0,2865	-0,5332
	0,5803	1,0000	0,3638	0,8337	0,1658	0,2611	-0,6087
	0,2011	0,3638	1,0000	0,4386	-0,7042	-0,6805	-0,6488
	0,9113	0,8337	0,4386	1,0000	0,1630	0,2023	-0,6755
	0,2833	0,1658	-0,7042	0,1630	1,0000	0,9902	0,4272
	0,2865	0,2611	-0,6805	0,2023	0,9902	1,0000	0,3571
	-0,5332	-0,6087	-0,6488	-0,6755	0,4272	0,3571	1,0000

Fuente: Elaboración propia

Tabla 7
Valores propios

$\lambda_{i=j}$	3,3946	2,8055	0,4373	0,2779	0,0810	0,0034	0,0003

Fuente: Elaboración propia

Tabla 8
Porcentajes acumulativos de valores propios

$\%\sum \lambda_{i=j}$	0,4849	0,8857	0,9482	0,9879	0,9995	1,0000	1,0000

Fuente: Elaboración propia

V_{ij}	0,4053	0,4316	0,3854	0,4939	-0,1277	-0,0968	-0,4809
	-0,2929	-0,2224	0,3559	-0,2323	-0,5751	-0,5800	-0,1303
	-0,6674	0,6980	0,1477	-0,1186	0,0294	0,1743	0,0176
	0,0888	-0,0338	0,6276	0,2103	0,1108	-0,0061	0,7353
	-0,2267	-0,4366	0,5121	-0,1054	0,3890	0,3549	-0,4553
	0,4098	0,1443	0,1875	-0,5878	-0,4232	0,5003	0,0332
	-0,2782	-0,2540	-0,1081	-0,5359	-0,5562	0,4975	0,0489

Fuente: Elaboración propia

Var. 1	0,7467	-0,4906	-0,4413	0,0468	-0,0645	0,0239	-0,0045
Var. 2	0,7952	.0,3726	0,4616	-0,0178	-0,1242	0,0084	-0,0041
Var. 3	0,7102	0,5961	0,0977	0,3309	0,1457	0,0109	-0,0018
Var. 4	0,9100	-0,3890	-0,0785	0,1109	-0,0300	-0,0343	0,0087
Var. 5	-0,2353	-0,9633	0,0195	0,0584	0,1107	-0,0247	-0,0090
Var. 6	-0,1784	-0,9715	0,1153	-0,0032	0,1010	0,0292	0,0081
Var. 7	-0,8861	-0,2182	-0,0116	0,3876	-0,1295	0,0019	0,0008

Fuente: Elaboración propia

Varianza d **ración**

Iteración Ciclo	Varianzas
0	0,2231
1	0,3592
2	0,3727
3	0,3730
4	0,3730
5	0,3730
6	0,3730
7	0,3730
8	0,3730
9	0,3730
10	0,3730

Fuente: Elaboración propia

Var. 1	0,9587	-0,1477	0,2018	-0,1333	0,0085	0,0208	0,0006
Var. 2	0,3763	-0,0976	0,9032	-0,1729	0,0655	0,0056	0,0007
Var. 3	0,2434	0,7707	0,3285	-0,0845	0,4813	-0,0011	-0,0002
Var. 4	0,8091	-0,0321	0,5415	-0,1579	0,1561	-0,0422	-0,0011
Var. 5	0,1555	-0,9760	0,0461	0,1419	0,0145	-0,0235	-0,0157
Var. 6	0,1271	-0,9762	0,1528	0,0812	-0,0049	0,0247	0,0157
Var. 7	-0,4439	-0,4298	-0,4002	0,6754	-0,0437	0,0006	-0,0002

Fuente: Elaboración propia

5. CONCLUSIONES

Los conceptos y definiciones matemático-estadísticos dados en el presente trabajo, fueron utilizados en el desarrollo del programa computacional, en la misma secuencia mencionada de operaciones, cuya aplicación y obtención de resultados se muestra mediante una aplicación numérica; observándose desde la matriz de datos que se ha utilizado, y los resultados parciales que se determinan, como: sumas, promedios, desviaciones estándar, suma de cuadrados y productos, coeficientes de la matriz de correlación, valores propios, porcentaje acumulativo de los valores propios, vectores propios, matriz factor, varianza de la matriz factor para cada ciclo de iteración, rotación de la matriz factor y verificación de las varianzas comunes.

La finalidad de esta investigación es llegar a que el interesado tenga un acceso inicial al conocimiento y comprensión del AF sin dificultad, pudiendo profundizar a futuro este tema con mayor facilidad, recomendándose los textos especializados 5, 7 y 8 de la bibliografía para su mejor comprensión.

6. COLABORACIÓN

Prof. Mary Nilda Avilés de Ruiz. Licenciada en Idiomas. Universidad Autónoma Gabriel René Moreno Santa Cruz – Bolivia (agosto, 2005)

BIBLIOGRAFÍA

Balestra Pietro, *Calcul Matriciel pour Économistes.* Éditions Castella, Lausanne, Suisse, 1972, pp. 242 – X.

Dagnelie Pierre, *Analyse Statistique á Plusieurs Variables. Les Presses Agronomiques de Gembloux*, Gembloux, Belgique, 1975 (2º éddition), pp. 362 – XIV.

Davis C. John, *Statistics and Data Analysis in Geology*. John Wiley & Sons, New York, USA, 1973, pp. 550 – VII.

International Business Machines Corporation, *System Reference Library, Scientific Subroutines*, New York, USA, 1967, pp. 191.

Joreskog j.k., klovan J. E. Reyment R. A., *Geological Factor Analysis. Elsevier Scientific Publishing Company*. Amsterdam, Holanda, 1976, pp 178 – VII.

Kendall S. Maurice, *Multivariate Analysis*. Charles Griffin & Co. Ltd., Londres, Inglaterra, 1975, pp. 210 – XI.

Harman Harry H., *Modern Factor Analysis*, The University of Chicago Press, Chicago, USA, 1967, pp. 474 – XVI.

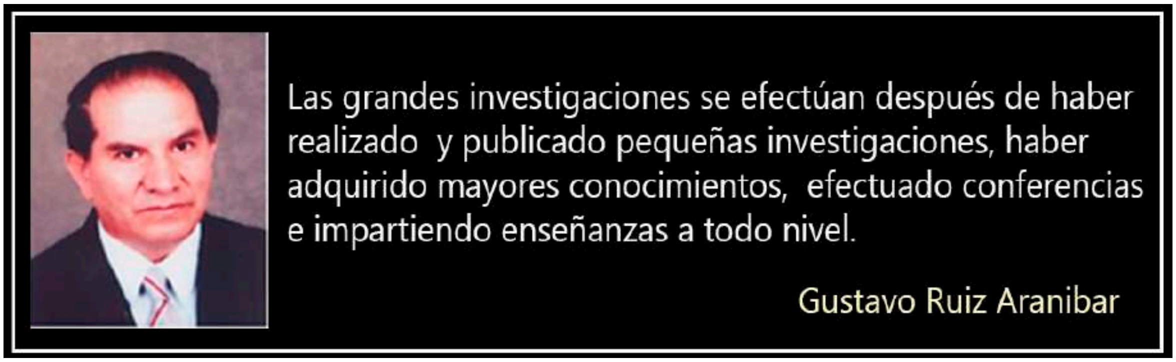
Horst Paul, *Factor Analysis of Data Matrices*. Copyright by Hold Rinehart and Wiston, Inc, Seattle, Washington, USA, 1965, pp. 730 – XXIV.

Lipschutz Seymour, *Algebra Lineal*. Editorial McGraw-Hill, Madrid, España, 1998, pp. 553 – XV.

Shafiro S. Miriam et al. *Courant Institute of Mathematical Sciences*, New York University, New York, USA, 1965, pp. 395.

Morice E., Bertrand M., *Penglaou C. Dictionnaire de Statistique*. Editorial Dunod, Paris, Francia, 1968, pp. 196.

Ruiz Aranibar Gustavo[4]. *Libreria Científica de Programas Informáticos*, La Paz –Bolivia.



Índice

4 Calle 20 y Av. Ballivian, Nº 8035, Calacoto, La Paz – Bolivia, Tel. 591-22772162 Cel. 67111778 gustavoruiz432@hotmail.com.bo ruizaranibargustavo@gmail.com.bo Blog: Gustavo Ruiz Aranibar