8
24

# UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

ESCUELA NACIONAL DE ESTUDIOS PROFESIONALES

"ACATLAN"

# EL ANALISIS FACTORIAL APLICANDO EL METODO DE COMPONENTES PRINCIPALES EN UN CASO PRACTICO " INDICES DE MARGINACION "



TESIS PROFESIONAL

QUE PARA OBTENER EL TITULO DE

LICENCIADO EN ACTUARIA

P R E S E N T A

AURELIA SALAZAR ESCOBAR

ASESOR: ACT. EFRAIN MEZA MORENO

UNAM ENEP ACATLAN

ACATLAN, EDO. DE MEXICO 1992

EL ANÁLISIS FACTORIAL APLICANDO EL METODO DE COMPONENTES PRINCIPALES EN UN CASO PRÁCTICO "ÍNDICES DE MARGINACIÓN"

# I N D I C E

# INTRODUCCIÓN

La inquietud de abordar el tema del análisis factorial en el presente trabajo, surge de la participación en la solución de algunos problemas en el área de trabajo, los cuales fueron: el análisis de información de una encuesta a nivel nacional sobre opinión pública; la obtención de indicadores altamente interrelacionados sobre el problema de la vivienda en el Estado de México, con la perspectiva de desarrollar conjuntos habitacionales; y la elaboración de la actualización y comparación de los 'indices de marginación' 1970-1980, para las entidades federativas de la República Mexicana.

En todos ellos se aplicó la técnica estadistica del análisis factorial, en forma un poco mécanica. A raiz de esto se crea el interés de investigar más ampliamente sobre esta técnica. Esto es, de dónde surge el método del análisis factorial, cuáles son sus bases matemáticas para el modelo, y entender su conceptualización y funcionamiento en la aplicación de investigaciones cientificas y sociales.

Inicialmente el análisis factorial se desarrolla por la necesidad de los psicólogos, para estudiar por ejemplo: los resultados de pruebas de inteligencia, el análisis factorial se utiliza para determinar las diferencias individuales, representadas por los puntajes de las pruebas. Asi también, la psicologia clinica tiene la necesidad de organizar y analizar un gran número de categorias diagnosticadas, para después clasificar el comportamiento de sus pacientes, aplicando el análisis factorial en la verificación del sistema de clasificación con datos empiricos de una población especifica.

Estos especialistas del comportamiento humano deseaban encontrar una solución a sus problemas planteados, la cual seria significativa, única y aplicada igualmente a la inteligencia, personalidad, medidas físicas o cualquier otra variable de interés. Derivándose de aquí las primeras aplicaciones del análisis factorial.

Los orígenes del método de análisis factorial se atribuye principalmente a Charles Spearman y Karl Pearson (1901), cuyos trabajos fueron orientados al desarrollo de la teoría psicológica.

Después de aproximadamente treinta años de investigaciones y trabajos realizados, por los estudiosos interesados en la técnica estadística de análisis factorial, hacia 1950 y con la accesibilidad de las computadoras, la aplicación de esta técnica se vuelve más popular, en otras áreas del conocimiento humano tales como: la Biología, Sociología, Economía, Medicina, Geología, etc. Utilizando el análisis factorial como instrumento analítico, para explorar y estructurar conceptos desconocidos o definir relaciones entre variables, también para verificar o modificar teorías e hipótesis acerca de datos empíricos.

El análisis factorial como un método del análisis multivariado es definido muy general e intuitivamente de la siguiente manera: es un instrumento o herramienta estadística, mediante la cual se puede descubrir la regularidad y el orden de los fenómenos.

Pueden manejarse simultáneamente más de un centenar de variables, y dar una resolución representada por la combinación lineal del conjunto de variables, en términos de un pequeño número de categorías o \`factores´. Esta solución es realizada por el análisis de la correlación entre las variables. Dichos factores relativamente independientes transportan toda la información esencial del conjunto de variables originales.

Así, el análisis factorial cumple con el objetivo de determinar factores comunes entre un conjunto de medidas o variables de una muestra o población determinada.

El objetivo general de la presente investigación es dar en forma sencilla y clara un marco conceptual del análisis factorial, los elementos matemáticos y estadísticos sobre los cuales se fundamenta el modelo matemático del método factorial, el funcionamiento de un programa computarizado para aplicar esta técnica estadística, y una aplicación práctica en la solución de un problema específico.

Además de describir paso a paso la metodología del análisis factorial que consiste en cuatro etapas fundamentales a saber: generación de la matriz de correlación, selección del método de extracción de los factores iniciales, rotación de los factores iniciales y determinación de puntajes factoriales.

Para que así, el investigador logre una buena utilización del método de análisis factorial, y en consecuencia obtenga excelentes y satisfactorios resultados en sus diferentes áreas de aplicación.

Como objetivo particular se persigue orientar y darle al investigador los elementos necesarios, para usar en los paquetes estadísticos computarizados la técnica estadística del análisis factorial, y poder interpretar adecuadamente los resultados o cuadros factoriales que resultan al ejecutar la subrutina FACTOR. Esto es muy importante para las áreas sociales, cuyos conocimientos matemáticos y estadísticos a veces son mínimos y se enfrentan a la problemática del análisis de resultados, para algunos métodos estadísticos.

Se desea que el trabajo desarrollado aquí, sea de gran utilidad para estos fines.

A continuación se resume el contenido de los capítulos tratados.

Capítulo I. Presenta una breve introducción acerca de la historia del análisis factorial, sobre la evolución de los métodos para determinar los factores iniciales. Se contempla la relación que tiene con otros métodos estadísticos multivariados y el manejo del concepto del análisis factorial.

Por otro lado, se da una amplia referencia sobre las aplicaciones del análisis factorial en las diferentes áreas sociales y científicas.

Finalmente se analiza la clasificación de las variables, para determinar el método de análisis de datos y tener criterios, para hacer una buena selección del método de análisis factorial.

Capítulo II. Se dan los elementos necesarios, para la interpretación algebraica y geométrica del modelo del análisis factorial, esto es, las definiciones y conceptos matemáticos y estadísticos como son: matrices, vectores, rotación de sistemas de coordenadas, variable, media, varianza, índice de correlación etc., representando las bases matemáticas del modelo.

Capítulo III. Se definen específicamente los términos especiales que más involucra el análisis factorial.

El modelo del análisis factorial es representado en forma geométrica y algebraica.

Básicamente se presenta el análisis de los cuatro pasos esenciales a seguir, en la metodología del análisis factorial.

Primer paso: Matriz de datos y generación de la matriz de correlación.

Segundo paso: Selección del método para determinar los factores iniciales, en este caso se describe el método de componentes principales.

Tercer paso: La rotación de los factores iniciales.

Cuarto paso: Determinación de los puntajes factoriales.

Capitulo IV. Se desarrolla una breve explicación de las subrutinas de métodos estadísticos que contiene el paquete SPSS/PC, así también se dan características generales sobre la programación de éste.

Describe ampliamente la subrutina FACTOR, señalando las diferentes formas de introducir los datos al paquete, en seguida se presenta la elaboración del programa para realizar un análisis factorial, y finalmente se hace la interpretación de los cuadros factoriales y gráficas más importantes, que resultan en la ejecución de la mencionada subrutina.

Capitulo V. Aquí se realiza una aplicación del método de análisis factorial, en un ejemplo práctico.

Se hace el planteamiento del problema, la definición de objetivos, selección de variables y recopilación de información. Después se utiliza el paquete estadístico SPSS/PC, para ejecutar el proceso del análisis factorial con los datos obtenidos, y por último se analizan los resultados de dicho análisis, resolviendo así el problema planteado.

# C A P I T U L O I

## MARCO CONCEPTUAL DEL ANÁLISIS FACTORIAL

### 1.1. Introducción al Análisis Factorial.

#### Breve historia del análisis factorial.

El análisis factorial es una rama de la ciencia estadística, pero por su desarrollo y uso extensivo en la psicología, esta técnica es frecuentemente interpretada y considerada como una teoría psicológica. El método viene a ser específicamente un proveedor de modelos matemáticos, para la explicación de las teorías psicológicas de habilidades humanas y de comportamiento. Entre las más frecuentes de tales teorías son las propuestas por Charles Spearman, Cyril Burt, Kelley, Thurstone, Karl Holzinger, Karl Pearson, J. C. Maxwell Garnett y Godfrey H. Thomson.

#### Desarrollo del análisis factorial en la psicología.

Los psicólogos y educadores probaron, para generalizar lo concerniente al comportamiento humano, características y conceptos bien conocidos como: memoria, imaginación, discernimiento, etc.

En el campo de pruebas mentales se tienen diversas pruebas de inteligencia general y de habilidades especializadas, tales como eclesiástico, mecánico, o aptitud médica. Con respecto a la inteligencia, el análisis factorial se aplicó en la determinación de diferencias individuales, representadas por puntajes de pruebas de inteligencia, que son atribuibles a un sólo origen de variación, o a la operación de una combinación de varias características mentales, tales como razonamiento, verbal y habilidades numéricas, las cuales pueden estar en varias combinaciones.

Similarmente en la psicología clínica gran número de categorías diagnosticadas son adoptadas por psiquiatras, psicólogos clínicos y otros, para clasificar neuróticos y psicópatas con desórdenes de comportamineto.

El rol del análisis factorial aquí, es para determinar si el sistema de clasificación puede verificarse con datos empíricos de poblaciones específicas, o si coinciden parcialmente en forma externa, para hacer un reordenamiento y redefinición de los conjuntos de categorías de diagnóstico.

Origenes del análisis factorial.

El origen del análisis factorial es generalmente atribuido a Charles Spearman. Por el monumental trabajo en desarrollo de la teoría psicológica, involucrando un singular factor general y un número de factores específicos.

La investigacón de 1904, fue el inicio del trabajo desarrollando la teoría de Dos-Factores, aplicados a la inteligencia el cual no está explícitamente en términos de `factores´.

Tal vez el artículo más crucial, naturalmente en la medida en cómo los aspectos estadísticos están contemplados, es el trabajo de Karl Pearson (1901), éste adelanta `El método de ejes principales´. Sin embargo, Spearman dedica otros cuarenta años de su vida al desarrollo del análisis factorial, y es considerado como el padre del análisis.

El primer periodo moderno incluyendo la activa y controvertida publicidad sobre el análisis factorial, viene después de 1925, con gran esfuerzo de investigación en 1930, y tuvo que volverse completamente aparente que la teoría de Dos-Factores de Spearman, conocida como dos factores para discriminar la inteligencia, no siempre es adecuada para describir una batería de pruebas de psicología.

Lo que actualmente acontece es que, la teoría de un factor general y específico en la forma original de Spearman, fue superada por la teoría de muchos grupos de factores, pero el primer método continúa siendo desarrollado para determinar este grupo de factores. Entonces siguiendo algunos trabajos y explorando la posibilidad de extracción de varios factores, directamente de la matriz de correlaciones entre pruebas, surge el concepto de análisis factorial múltiple en el trabajo de Garnett (1919).

El actual término del análisis factorial múltiple se atribuye a L. L. Thurstone, además del análisis del método Centroide, pero su remarcada contribución fue la generalización del criterio de cuatro diferencias de Spearman, para el rango de la matriz de correlación, como la base para determinar el número de factores comunes.

En particular los psicólogos tienen introducidas varias teorías, para encontrar una forma de solución, la cual será significativa, única y aplicada igualmente a la inteligencia, personalidad, medidas físicas o cualquier variable de interés. La teoría Bi-Factor de Holzinger y la teoría de estructura simple de Thurstone son de esta clase. Sobre todo la teoría muestral de Thomson es una teoría psicológica de la mente.

Los tipos de soluciones factoriales, están determinados sobre la base de dos principales generalidades: 1) La simplicidad estadística y 2) El significado científico.

Definición y objetivos del análisis factorial.

Para tener una idea más clara en cuanto al significado y objetivos del análisis factorial, como definiciones tentativas. Se mencionan varios enunciados sobre lo que representa este análisis, según estudiosos del tema:

B. Fruchter.

El análisis factorial inicia con un conjunto de observaciones obtenidas de una muestra dada, por medias de tales medidas a priori. Este método analiza el conjunto de observaciones desde sus intercorrelaciones, para determinar si la variación representada puede explicarse adecuadamente, por un número pequeño de categorias básicas, del conjunto que el investigador inició. Estos datos obtenidos con un gran número de medidas a priori, es explicado en términos de un pequeño número de variables de referencia.

En el análisis factorial las `diferencias individuales´ representadas por un gran número de medidas, que están dadas por una población singular, usualmente en un tiempo bajo un conjunto de condiciones estándar, son estudiadas para detectar posibles origenes comunes de variación.

B. J. Rummel.

El análisis factorial es un instrumento mediante el cual se puede descubrir la regularidad y el orden de los fenómenos. Este puede manejar simultáneamente más de un centenar de variables, compensar el error y la invalidez debidos al azar y desenmarañar las intercorrelaciones complejas de acuerdo a sus principales distintivas.

El análisis factorial hace lo siguiente : Toma miles y potencialmente millones de medidas y observaciones cuantitativas y las resuelve en distintas pautas de ocurrencia. Explicita y confiere mayor precisión a la construcción de conexiones entre los hechos que se realizan constantemente en la mente huamana.

Harry. H. Harman.

El principal interés del análisis factorial es la resolución de un conjunto de variables linealmente en términos de (usualmente) un pequeño número de categorías o `factores´. Esta resolución puede ser realizada por el análisis de la correlación entre las variables. Una satisfactoria solución producirá factores, los cuales transportan toda la información esencial del conjunto de variables.

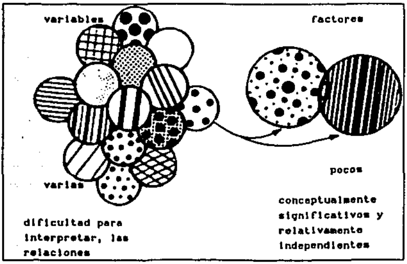
El objetivo de una completa descripción no puede ser teóricamente investigada, es aprovechada prácticamente en un limitado campo de investigación, donde un relativamente pequeño número de variables es considerado exhaustivo. En todos los casos, sin embargo el análisis factorial da una simple interpretación de un conjunto de datos, y así proporciona una descripción fundamental del particular conjunto de variables analizadas.

El análisis factorial tiene como meta la determinación de factores comunes entre un conjunto de medidas o variables, y si las variables no muestran correlaciones muy elevadas entre sí, no tiene sentido hacer el análisis.

Kleinbaum y Kupper.

El análisis factorial es un método multivariado que tiene como objetivo principal, la explicación de relaciones con dificultades para interpretar, variables correlacionadas en términos conceptualmente significativos de factores relativamente independientes.

Fig. 1.1. Representación gráfica del análisis factorial.




El dibujo de la Fig. 1.1. presenta un conjunto de circulos traslapados con distinto sombreado, que representan las variables a explicar y reconstituidos en dos circulos no traslapados con diferente patrón de sombreado, siendo estos los factores significativos obtenidos.

Métodos factoriales.

La acalorada e inesperada controversia acerca del 'mejor' método de análisis factorial están sobre Charles Spearman (1863-1945), L. L. Thurstone (1887-1955), y Karl J. Holzinger (1893-1954). Esta no es una controversia personal, pero ante la firme convicción de cada individuo, tienen dedicada la mayor parte de su vida al desarrollo del particular aprendizaje en el análisis factorial.

Todo los ensayos que aparecen durante los cuarentas y cincuentas abogan por estos 'métodos'. Sin embargo, una comprensión de las caracteristicas salientes de cada método y con un crecimiento eficiente de la computación, las diferencias entre los diversos métodos mucho tiempo no se perfilaron de manera inquietante, y los seguidores de un parecido particular son más tolerantes que los adherentes a un esquema alternativo.

Es evidente que los diferentes métodos factoriales, corresponden a las diferentes teorías matemáticas en la explicación de un problema científico particular.

El investigador antes de seleccionar un método factorial de acuerdo a sus características, debe sopesar sus ventajas y limitaciones de cualquier solución particular, para sus datos específicos.

## 1. La teoría de Dos-Factores de Spearman.

Spearman fue uno de los primeros en atacar el problema de los factores. Un primer trabajo con grupos de cuatro pruebas. Observando que pruebas de habilidades tienen intercorrelaciones positivas representadas en la tabla 1.1. No necesariamente es considerada para un singular origen de variación (o factor), si los coeficientes en cualquier combinación de dos columnas (evitando las celdas de la diagonal principal) son proporcionales.

Tabla 1.1 Intercorrelaciones de cuatro pruebas.

| test | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | | $r_{12}$ | $r_{13}$ | $r_{14}$ |
| 2 | $r_{21}$ | | $r_{23}$ | $r_{24}$ |
| 3 | $r_{31}$ | $r_{32}$ | | $r_{34}$ |
| 4 | $r_{41}$ | $r_{42}$ | $r_{43}$ | |

Así, en la columna 1 y 2, para satisfacer el `criterio de proporcionalidad´, se tienen las ecuaciones siguientes:

Columnas 1 y 2

$$\frac{r_{31}}{r_{32}} = \frac{r_{41}}{r_{42}} \qquad ó \qquad r_{31}\, r_{42} = r_{32}\, r_{41} \qquad (1.1)$$

Columna 1 y 3

$$\frac{r_{21}}{r_{23}} = \frac{r_{41}}{r_{43}} \qquad ó \qquad r_{21}\, r_{43} = r_{41}\, r_{23} \qquad (1.2)$$

Columna 1 y 4

$$\frac{r_{21}}{r_{24}} = \frac{r_{31}}{r_{34}} \qquad ó \qquad r_{21}\, r_{34} = r_{31}\, r_{24} \qquad (1.3)$$

Ecuaciones similares son escritas para las columnas 2y3, 2y4, 3y4. Cuando la igualdad de las Ecs. 1.1, 1.2, y 1.3 existe, el criterio de proporcionalidad es satisfecho, y las cuatro pruebas están asumiendo tener un factor en común.

Las ecuaciones 1, 2 y 3 pueden ser expresadas en la forma:

$$r_{31}\, r_{42} - r_{32}\, r_{41} = 0$$

$$r_{21}\, r_{43} - r_{23}\, r_{41} = 0$$

$$r_{21}\, r_{34} - r_{24}\, r_{31} = 0$$

Estas son las famosas cuatro diferencias de Spearman.

Spearman y otros fundan un considerable número de pruebas de habilidades y sus intercorrelaciones satisfacen el criterio de proporcionalidad, lo cual es suficiente para concluir que sus intercorrelaciones están considerando un factor singular. Este factor fue llamado g, el factor general-intelectual. La teoria Dos-Factores de Spearman postula que todos las pruebas de inteligencia, los cuales satisfacen el criterio de proporcionalidad contienen dos factores, g y s.

El g-factor es un factor general común para las pruebas intelectuales. El s-factor es especifico para cada prueba, y representa que proporción de la confianza de la varianza de una pruebas, que no está correlacionado con otras pruebas. Spearman desarrolla la hipótesis que g es una función de herencia, mientras s-factore representa la adquisición de la experiencia y aprendizaje especifico.

2. La teoria Bi-Factor de Holzinger.

Más recientemente Spearman y sus adherentes realizaron pruebas, que no cumplian el criterio de proporcionalidad, Spearman los ha denominado `disturbios´ son retenidos en la matriz de correlaciones, si son reconocidas que algunas de las pruebas tienen un factor en común, en adición al factor general que no es común a todas las pruebas. Estos factores comunes a grupos de pruebas, son llamados grupos-factores. El método Bi-Factor de Holzinger, el cual es una variación del método de Spearman Dos-Factores, obtiene un general y uno o más grupos de factores.

La varianza para la prueba j en valores estándar, si ésta asume ser perfectamente confiable, es dada por la ecuación:

$$g_j^2 + c_j^2 + s_j^2 = 1 \qquad (1.4)$$

donde c es cargado con un grupo de factores.

La correlación entre las pruebas es determinada por extensión, las cuales son cargados sobre los mismos factores y son calculadas desde el producto-cruzado de sus cargas de acuerdo a la ecuación:

$$r_{jk} = g_i g_k + c_{j1}c_{k1} + c_{j2}c_{k2} + \dots + c_{jr}c_{kr}. \qquad (1.5)$$

Una ventaja del método Bi-Factor sobre el de Dos-Factores, es que éste no insiste sobre un singular factor común, y permite que las pruebas con criterio de propocionalidad no conveniente, sean retenidas para formar parte de las varianzas comunes, bajo un grupo de factores.

Entre los grupos de factores que son identificados por investigadores usando el método Bi-Factor, están las habilidades verbales, habilidades numéricas, habilidades mecánicas, la atención, imaginación y factores de personalidad como perseverancia.

3. Método del Factor-Principal.

El método del Factor Principal es probablemente la teoria más usada en el análisis factorial.

El método requeria considerables cálculos, los cuales también consumian mucho tiempo antes de estar disponibles las computadoras electrónicas.

Los fundamentos para el `método de ejes pricipales´ fueron dados al inicio de siglo por Karl Pearson (1901). Sin embargo, hasta 1933 el método del Factor-Principal que ahora se conoce, fue desarrollado por Hotelling, siguiendo esencialmente el mismo proceso pero operando sobre la reducción de la matriz de correlaciones, empleando el modelo siguiente:

$$z_j = a_{j1}F_1 + a_{j2}F_2 + \dots + a_{jm}F_m + u_jY_j. \qquad (1.6)$$

$$( j = 1, 2, \dots, n ) \quad , \quad ( m < n ).$$

Donde cada una de las n variables observadas es descrita linealmente en términos de m factores comunes y un único factor. Los factores comunes cuentan para la correlación entre variables, mientras cada único factor cuenta para la varianza restante (incluyendo el error) de las variables.

Subsecuentemente, Kelley en 1935 desarrollo un proceso alternativo, el cual veinte años después, prueba ser uno de los más útiles para adaptarse a computadoras electrónicas de alta velocidad. La primera aplicación de computadoras a este problema en el análisis factorial fue hecha por Weigley y Nouhaus (1952-1955).

## 4. Método de Componentes Principales.

El método de Componentes Principales, o análisis de componentes, se apoya sobre el primer trabajo de Pearson (1901), con la adaptación especifica del análisis factorial sugerida por el trabajo de Hotelling (1933), quien desarrollo el procedimiento para el análisis de componentes de la matriz de correlaciones fundamentado en el modelo siguiente:

$$z_j = a_{j1}F_1 + a_{j2}F_2 + \cdots + a_{jn}F_n. \tag{1.7}$$

$( j = 1, 2, \ldots , n )$.

Donde cada una de las n variables observadas son descritas linealmente en términos de n nuevos componentes no correlacionados $F_1, F_2, \ldots , F_n$. Una importante propiedad en este modelo, en cuanto a como la sumarización de datos es afectada, es que cada componente en turno, hace una contribución máxima a la suma de las varianzas de las variables.

Una importante distinción entre componentes principales y factores principales, es que los componentes son inmediatamente expresados en términos de las variables observadas, en tanto que los factores significativos sólo acontecen indirectamente.

5. Método Centroide.

Este método de factorización a la matriz de correlación, provee un compromiso computacional para el método del Factor-Principal, antes que las computadoras generalmente fueran disponibles. Ahora el método Centroide es de interés histórico. La fórmula fundamental del método Centroide fue primero empleada por Burt (1917), pero aplicada al problema de determinación a un singular factor general del tipo Spearman. El Complemento de este método fue desarrollado por Thurstone (1933), en conjunción con el análisis de grandes baterías de pruebas psicométricas en varios factores comunes.

6. Método de Máxima-Verosimilitud.

Lawley (1940-1942), hace una fundamental contribución al análisis factorial, proporcionando bases estadísticas para juzgar adecuadamente el modelo (1.6), con un especifico número de factores para explicar una matriz de correlación empírica. Esta prueba estadística, para el número de factores comunes es dependiente sobre un particular tipo de factores solución, esto es la máxima verosimilitud estimada de los factores cargados. La cantidad de cálculos surgidos de este método, restringe el uso a pequeños problemas en los cuarentas y cincuentas. Actualmente las computadoras hacen posible el método de Máxima-Verosimilitud.

7. Método Minres.

Un método de análisis factorial el cual minimiza residuales (de aquí el nombre de `minres´), fue desarrollado por Harman y Jones en 1966. Específicamente, este método estima los factores cargados en tal forma, como hacer la suma de cuadrados de los residuales, fuera de la diagonal de la matriz de correlación minima. Reproduce las mejores correlaciones observadas. Este es un contraste para el método del Factor-Principal, el cual extrae la máxima varianza, sin embargo una solución Minres tiene la misma apariencia general, como la solución del Factor-Principal. El método Minres es dependiente sobre una estimación de los números de factores comunes; la comunalidad consistente con esta hipótesis, es obtenida por productos del método. Estas Propiedades son comunes para el método Minres y Máxima-Verosimilitud.

8. Método de Grupos Múltiples.

La distinguida caracteristica del método envuelve la factorización de una matriz de correlación, en varios factores múltiples simultáneamente.

Hortst (1937), anticipado al método de Grupos-Múltiples del análisis factorial, no lleva el trabajo teórico a la etapa de aplicación práctica. Similarmente Guttman (1944), presenta la teoria fuera de un proceso computacional, Holzinger (1944) y Thurstone (1945), presentan un simple proceso computado para el `Análisis factorial de grupos´ sin reconocimiento similar a la primera teoria desarrollada. Los diversos ‹métodos de grupos múltiples› independientes desarrollados para el análisis factorial, son comparados y sintetizados por Harman en 1954.

## 1.2. Aplicaciones del Análisis Factorial.

La aplicación de la técnica del análisis factorial ha sido principalmente en el campo de la psicologia, por el hecho de tener origen en ésta. Esta limitación no tiene fundamento para que la técnica de análisis, sea aplicada en otras diferentes áreas cientificas o sociales.

El análisis factorial se usa en la formulación de teorias en el comportamiento y las ciencias sociales, pero el `instrumento analitico´ (incluyendo el análisis factorial) no será confundido con la `ciencia´. Como un instrumento exploratorio (entre otros), el análisis factorial puede usarse para verificar o modificar teorias de nuevos experimentos, y nuevos datos sujetos a un análisis fresco, para proponer la clarificación o refinación de formulaciones previas. Por contraste `confirmatoriamente´ el análisis factorial, se usa para verificar o probar una concebida o dada hipótesis acerca de la estructura de datos empiricos.

El más reciente acercamiento es ilustrado por la teoria de Spearman `Todas las ramas de la actividad intelectual tienen en común una fundamental función (o grupos de funciones), puesto que los elementos especificos de la actividad aparecen en muchos casos, para ser totalmente diferentes de todos los demás´. Él muestra que si ciertas relaciones existen entre las correlaciones, todas las variables son resueltas en la expresión lineal involucrando sólo un factor general, y un adicional factor único para cada variable.

Estas relaciones suministran la verificación estadistica de la teoria Dos-Factores. Si un conjunto de variables psicológicas dan coeficientes de correlación, los cuales no satisfacen las relaciones predecidas, entonces una teoria más compleja es postulada. Esta puede requerir varios factores comunes en la descripción estadistica de las variables.

Durante la segunda guerra mundial con las pruebas de gran escala, clasificación y problemas de asignación, el análisis factorial fue aplicado extensamente en varias ramas de los servicios militares de los Estados Unidos.

En la actualidad los psicólogos continúan desarrollando y explorando esta técnica de análisis, para determinar relativamente pequeños números de pruebas y describir la mente humana.

Las aplicaciones del análisis factorial en otros campos que no sea la psicología, se volvieron muy populares desde 1950, con la accesibilidad de las computadoras. Estos campos incluyen tales variedades de disciplinas como la Meteorología y Medicina, la Ciencia Política y Taxonomía, Arqueología y Economía, Sociología y la Ciencia Regional. También hay muchos estudios individuales que es difícil asignarlos a una disciplina particular.

Se mencionan algunos estudios del gran número que existen, para dar una idea de la aplicación del análisis factorial:

Económicos: Evaluando la ejecución de sistemas, La inversión de decisiones bajo incertidumbre, La estructura de cambios de precios de seguridad, El Sistema de ecuaciones económicas empleando componentes principales, etc.

Medicina: Estudios en alergias, Estudios cardiovasculares, Análisis de electrocardiograma y electroencefalograma, Diagnóstico y Clasificación, etc.

Ciencias Físicas: El análisis factorial de datos sobre minerales pesados, Mapas geológicos, Predicciones en meteorología, etc.

Ciencias Políticas y Sociales: Relación entre variables politicas y sociales y el Gasto Público Nacional, Estudios de la Suprema Corte y la Asamblea General, Dimensiones de naciones, Análisis de datos sociométricos, etc.

Ciencias Regionales: Estructura y desarrollo económico de áreas urbanas, Dimensiones de gobernadores locales, etc.

Aplicaciones Taxonómicas: Clasificación de sistemas para reportes psicológicos, Clasificación biológica, Clasificación de levaduras, etc.

Aplicaciones Misceláneas: Técnicas analiticas de arqueologia, Valoración de edificios por arquitectos, Sistemas de hombre-máquina e ingenieria humana, etc.

El análisis factorial también se aplica generalmente para explorar un área de contenido, estructurar un campo, situar conceptos desconocidos, clasificar o reducir datos, iluminar nexos causales, proyectar o transformar los datos, definir relaciones, verificar hipótesis, formular teorias, controlar variables o hacer inferencias.

Cuando se aplica para discernir pautas[1] de perfiles similares entre individuos, grupos o naciones el análisis se llama análisis factorial Q. Aplicado a delinear pautas de variaciones en caracteristicas es llamado análisis factorial R.

[1] Significado de pautas: Los fenómenos co-ocurren en el espacio o en el tiempo, por lo que se pueden clasificar como pautados, y como estos fenómenos concurrentes son independientes entre si, se da una variedad de distintas pautas, constituyendo la esencia de conceptos comunes como, 'mesa', 'silla', y 'casa' y las pautas estructuran las teorias e hipótesis.

Usos del análisis factorial.[2]

Se subraya las aplicaciones del análisis que se relacionan con diversos intereses científicos y políticos.

1. Interdependencia y delineación de pautas.

Si un científico tiene un cuadro de datos -por ejemplo, votos en las Naciones Unidas, características personales o respuestas a un cuestionario- y si sospecha que esos datos están relacionados entre sí de un modo complejo, el análisis factorial puede ser usado para desentrañar las relaciones lineales antre las distintas pautas. Cada una de ellas aparecerá como un factor que delinea un grupo de datos intercorrelacionados.

2. Simplificación o reducción de los datos.

El análisis factorial puede ser útil para reducir una masa de información a una descripción más simple. Por ejemplo, los datos de cincuenta características para 300 naciones no pueden ser manipulados descriptiva o analíticamente. El manejo, análisis y comprensión de tales datos se hace más fácil si se los reduce a sus pautas factoriales cumunes. Estos factores concentran e indexan la información dispensada en los datos originales y pueden así reemplazar las cincuenta características sin que se pierda mucha información. Las naciones, por ejemplo, pueden ser comparadas y discutidas más fácilmente en lo que respecta a sus dimensiones de desarrollo, tamaño y totalitarismo que a los cientos de características que involucra cada dimensión.

[2] SCHWARTZMAN, Simón. et al. Técnicas avanzadas en cincias sociales. Buenos Aires, Ed. Nueva Visión, 1977. Págs. (44-54).

3. Estructura.

El análisis factorial puede ser empleado para descubrir la estructura de un campo. A propósito de un caso, un científico puede querer descubrir las líneas primarias independientes o dimensiones -tales como tamaño, liderazgo y edad- de variación en las características de grupo de conducta. Es posible que los datos reunidos en una amplia muestra de grupos ayuden, después de ser sometidos al análisis factorial, a descubrir esta estructura.

4. Clasificación o descripción.

El análisis factorial es un instrumento apto para desarrollar una tipología. Puede ser utilizado para agrupar variables interdependientes con el objeto de formar categorías descriptivas, tales como ideología, revolución, voto liberal y autoritarismo ; para clasificar los perfiles de naciones en tipos con características o conductas similares; o sobre matrices de datos de transacciones de elección social para mostrar, cómo los individuos o las naciones se agrupan mediante sus transacciones o sus elecciones recíprocas.

5. Formación de escalas.

El científico desea frecuentemente desarrollar una escala en la que puedan evaluarse y compararse individuos, grupos o naciones. La escala puede referirse a fenómenos tales como participación política, conducta electoral o conflicto. Un problema en el desarrollo de escalas es la ponderación de las características que se combinan.

El análisis factorial ofrece una solución dividiendo las características en fuentes independientes de variación (factores). Cada factor representa entonces una escala basada en la relación empírica entre las características. Como resultados adicionales, el análisis factorial dará el peso que deberá otorgar a cada caracteristica cuando se le combine en escalas. Los resultados del puntaje factorial son, de hecho tales escalas desarrolladas combinando las caracteristicas en función de sus pesos.

6. Comprobación de hipótesis.

Abundan las hipótesis relativas a las dimensiones de actitud, personalidad, grupo, conducta social, voto y conflicto. Puesto que el significado asociado corrientemente con ‹dimensión› es el de grupo o conjunto de caracteristicas o conductas altamente interrelacionadas, se puede usar el análisis factorial para comprobar su existencia empirica. También postula de antemano qué caracteristicas o conducta deberian estar relacionadas teóricamente, con qué dimensiones, y se pueden aplicar las pruebas estadisticas de significación a los resultados del análisis factorial.

Hay otro tipo de hipótesis que pueden ser comprobados, además de los que se relacionan con las dimensiones. Por ejemplo: si se trata de una relación entre desarrollo económico e inestabilidad, permaneciendo constantes las demás cosas, se puede hacer un análisis factorial de las variables de economia e inestabilidad justo con otras variables que pueden afectar (ocultar, interponerse, disminuir) su relación. Los factores resultantes pueden ser definidos (rotados) de tal manera que los primeros factores involucren las medidas intervenientes (hasta el máximo permitido por las relaciones empiricas).

Entonces puede calcularse un factor independiente restante, para definir mejor las relaciones postuladas entre las medidas de economia y de inestabilidad. El grado en que están involucradas ambas variables en la pauta, permite ver al cientifico si existe realmente una pauta desarrollo económico-inestabilidad cuando las demás permanecen constantes.

7. Tranformación de los datos.

El análisis factorial puede ser usado para transformar los datos a fin de adecuarlos a los supuestos de otras técnicas. por ejemplo, la aplicación de la técnica de regresión múltiple (si se tiene que aplicar las pruebas de significación a los coeficientes de regresión), supone que los predictores -las llamadas variables independientes- no están relacionados estadísticamente.

Si las variables predictoras están relacionadas entre sí, contrariamente a lo supuesto, el análisis factorial puede servir para reducirlas a un grupo menor de puntajes no relacionados. Éstos pueden utilizarse en el análisis de regresión en lugar de las variables originales, sabiendo que no se ha perdido la variación significativa en los datos originales. Igualmente, mediante el análisis factorial, se puede reducir un gran número de variables dependientes.

8. Exploración.

En un campo nuevo de interés científico, como la investigación sobre la paz, las complejas interrelaciones entre los fenómenos no han sido sometidas aún a una investigación sistemática. El campo desconocido puede ser explorado mediante el análisis factorial. Puede reducir relaciones complejas a una expresión lineal relativamente simple y puede descubrir insospechadas quizás, asombrosas relaciones.

Comúnmente, el científico social es incapaz de manipular variables en un laboratorio, y debe tratar con la múltiple complejidad de las conductas en su contexto social. El análisis factorial desempeña entonces algunas funciones propias del laboratorio, y permite al científico hallar relaciones, separar diferentes fuentes de variación, excluir y controlar parcialmente influencias indeseables en las variables de que se ocupa.

9. Delineación.

Además de facilitar la exploración, el análisis factorial permite también al científico delinear el terreno social. Entendiendo por delinear el intento sistemático de hacer un mapa de los principales conceptos empíricos y fuentes de variación. Estos conceptos pueden ser usados, para describir un campo o servir como insumo para ulteriores investigaciones. Algunos campos sociales, como las relaciones internacionales, la vida familiar y la administración pública carecen aún de tales mapas, aunque en algunas áreas diferentes, como la personalidad, las capacidades, las actitudes y el significado cognoscitivo, se ha realizado un considerable esfuerzo en este sentido.

10. Teoria.

Se puede construir la estructura analitica de teorias o modelos sociales a partir de la estructura geométrica o algebraica del análisis factorial. Los factores mismos pueden ser postulados. A partir de ellos, se puede derivar y comprobar deducciones operacionales con contenido empirico. El modelo factorial representa un formalismo matemático que proviene de las funciones de cálculo de la fisica clásica. La parte analitica del modelo factorial es semejante a la teoria cuántica. Los vectores y su posición, los operadores lineales y las dimensiones (factores) de un sistema constituyen el centro de interés.

Puesto que el análisis factorial incorpora las posibilidades analiticas como teoria, y las técnicas empiricas para conectar la teoria con los fenómenos sociales, su potencialidad promete un gran desarrollo teórico para las ciencias sociales.

Algunas aplicaciones concretas.

1. Se tiene información sobre catorce naciones y diez características. Las naciones fueron escogidas para representar los principales grupos regionales, políticos, económicos y culturales; las características reflejan diferentes facetas de cada nación incluyendo la inestabilidad interior y los conflictos exteriores. En total existen 140 piezas de información para 1955 y el análisis factorial debe responder a esta pregunta: ¿Cuáles son las pautas de relación entre esos datos?

Las pautas pueden ser consideradas desde dos perspectivas. Se puede ver la pauta de variabilidad de las naciones según sus características y agrupar después las naciones según la semejanza de su perfil. Se puede agrupar juntas las naciones que tienen un alto PBN per cápita, poco comercio, mucho poder, etcétera.

La otra perspectiva es enfocar las pautas de variaciones de las características. Por ejemplo, naciones de alto PBN per cápita muestran también poco comercio y poder. Hay una regularidad en los valores nacionales de esas tres características, y esta regularidad es descrita como una pauta de variación. Muchos de los conceptos sociales definen esas pautas. Por ejemplo, el concepto de ·desarrollo económico· involucra (entre otras cosas) PBN per cápita, alfabetización, urbanización, educación y comunicación; es una pauta porque estas características están muy rrelacionadas.[3]

Las variables contempladas en este ejemplo son: PBN per cápita, Comercio, Rango de poder, Estabilidad, Libertad de grupos de oposición, Conflictos exteriores, Acuerdos con E.E. U.U. en la ONU, Presupuesto de la Defensa, Porcentaje del PBN para Defensa y Aceptación de las leyes internacionales.

[3] Ibid., SCWARTZMAN, Simón. et al. Págs. 40 y 41.

2. Investigaciones anteriores [e. g., Stamler (1967) y Harburg et al. (1973)] han mostrado que los negros Americanos tienen más alto promedio del nivel de presión sanguinea que los blancos, y la tasa de muerte de hipertensión y desórdenes relacionados se considera más alta entre negros que entre blancos. Las razones de estas diferencias no son completamente entendibles, los investigadores atribuyen estas diferencias a combinaciones de factores genéticos y socio-ambientales, incluyendo el acceso a buenos cuidados médicos.

Un reciente estudio por James y Kleinbaum (1976), considerando el aspecto socio-ambiental, plantean la siguiente hipótesis; si los negros que viven en áreas de alto stress, tienen más alta hipertensión relacionada con la tasa de mortalidad, que otros negros en áreas de bajo-stress o blancos en áreas de alto o bajo stress.

Usando 86 condados del Norte de Carolina como unidades muestrales, James y Kleinbaum computaron en un periodo de tres años (1959-1961), las tasas de muerte relacionadas con la hipertensión para cada condado, relacionando estas tasas a un índice de stress-socioecológico para cada condado, los cuales fueron derivados del análisis factorial.

La cuestión de interés fue, si los condados con un alto-stress tienen significativamente más altas tasas de mortalidad que los condados de bajo-stress.

El índice de stress-socioecológico fue construido usando 15 variables (en 1960 los valores fueron obtenidos separadamente por raza para cada condado). Las que reflejan el bienestar económico y social de los condados durante el periodo (1959-1961).

También se identifican dos factores uno llamado status socio-económico (SSE) y otro inestabilidad social (IS), como resultado del análisis factorial, que en combinación definen el índice de stress-socioecológico.

Se supone que algunas variables están más relacionadas con un factor que con otro, y existen variables que están relacionadas con ambos factores.

En este estudio de dos factores de interés tuvieron que ser identificados conceptualmente a priori al del análisis factorial, tal como una identificación a priori no es prerequisito para la aplicación de este método. De hecho, algunas veces, el investigador no tiene con anterioridad factores resultantes y desea utilizar el análisis factorial, para ayudarse a caracterizar factores significativos describiendo los datos.[4]

Variables utilizadas en el estudio James-Kleinbaum.

(.) 1. Ingresos per cápita.
(.) 2. Años promedio de educación.
(x)(.) 3. Porcentaje de desempleo.
(.) 4. Porcentaje de familias que ganan arriba de 8,000 dólares al año.
(.) 5. Porcentaje de hombres obreros negros.
(.) 6. Porcentaje de hombres obreros blancos.
(x)(.) 7. Porcentaje de familias que ganan abajo de 3,000 dólares al año.
(x) 8. Porcentaje de mujeres separadas o divorciadas.
(x) 9. Índice de hombres delincuentes juveniles.
(x) 10. Índice de mujeres delincuentes juveniles.
(x) 11. Porcentaje de hombres en escuela correccional
(x) 12. Porcentaje de mujeres en escuela correccional.
(x) 13. Porcentaje de hombres en prisión.
(x) 14. Tasa de Homicidios.
(x) 15. Porcentaje de jóvenes menores de 18 años sin padres.

[4] KLEINBAUM G. David y Lawrence L. Kupper. Applied Regression Analysis and other Multivariable Methods. California, Ed. Duxburry Press, 1978. Págs. 377 y 378.

(.) relación con el factor status socioeconómico (SSE).
(x) relación con el factor inestabilidad social (IS).
(x)(.) relación con ambos factores.

3. Aplicación en el área científico-médica.[5]

Una investigación acerca de `Los cambios de comportamiento en pacientes de Parkinsonismo después de la cirugía´.

El objetivo de esta investigación es obtener la descripción adecuada, y la cuantificación de las capacidades psico-motoras del paciente, para poder emitir en base a datos precisos juicios sobre la terapéutica, ya sea quirúrgica o medicinal.

El experimento consiste en someter a pacientes con el mal del Parkinson antes y después de una cirugía, a una evaluación de la eficacia de la función motora, para después hacer un análisis del patrón de los síntomas motores asociados con el parkinsonismo.

El estudio se realiza con una muestra de 40 pacientes antes de la operación y 30 pacientes 3 meses después de la operación, considerando 28 variables.

Los datos están basados en una extensa investigación de los efectos de las lesiones terapéuticas del cerebro, sobre el comportamiento de los que padecen de la enfermedad de Parkinson.

Los resultados preoperatorios fueron sometidos a un análisis factorial, se identificaron los principales factores y a cada paciente se le asigno un tanto por cada factor. Este procedimiento es una buena descripción objetiva y empírica del patrón sintomático. Más tarde los resultados posoperatorios también

[5] CURTIS, HARDYCK y Lewis F. Petrinovich. Investigación en Ciencias Sociales. Tr. Dr. Pedro Rivero, México, Ed. Interamericana, 1977. Págs. (150-162).

fueron sometidos a un análisis factorial y se examinan las diferencias en la estructura de los factores en los dos análisis.

Variables calificadas para la evaluación del paciente.

1. el andar.
2. apertura del brazo derecho.
3. apertura del brazo izquierdo.
4. vuelta a la derecha.
5. vuelta a la izquierda.
6. postura de pie.
7. levantarse.
8. sentarse.
9. capacidad motora general.
10. capacidad para sujetar un vaso de agua y bebérselo con la mano derecha.
11. lo mismo con la mano izquierda.
12. capacidad para trazar un circulo, un triángulo y una espiral con la mano derecha.
13. lo mismo con la mano izquierda.
14. capacidad para escribir la frase ‹¿cómo está usted?› con la mano derecha.
15. capacidad para mover los dedos separados y ritmicamente de abajo para arriba `como si estuviera tocando el piano´, con la mano derecha.
16. lo mismo con la mano izquierda; capacidad para redondear los labios.
17. capacidad para alargar los labios.
18. mover los labios varias veces y con la mayor rapidez posible, como si estuviera diciendo `uuu...iii´.
19. capacidad para mover la lengua sacada, de un lado para otro y lo más rápidamente posible, varias veces (movimiento lateral de la lengua).
20. expresión facial derecha.
21. expresión facial izquierda.
22. inteligibilidad de la locución.

23. calificación de la eficiencia general.
24. temblor de la cara.
25. temblor del brazo derecho.
26. temblor del brazo izquierdo.
27. temblor de la pierna derecha.
28. temblor de la pierna izquierda.

Existen una infinidad de ejemplos sobre las aplicaciones del análisis factorial, los expuestos aquí tratan de dar una visión clara y sencilla de esta técnica estadística, en los diversos campos de la ciencia científica y social.

## 1.3. Clasificación de Variables y Selección del Método.

### Clasificación de variables.

Las variable se clasifican en diferentes formas. Tales clasififcaciones son usadas para determinar el método de análisis de datos. A continuación se describen tres formas de clasificar.

### Variables discretas y continuas.

La clasificación proyectará los llamados espacios, para determinar si hay o no vacios, entre valores observables sucesivamente de una variable. Si hay vacios entre observaciones la variable es declarada discreta; si no hay vacios entre observaciones la variable es continua.

Para ser más precisos una variable es discreta si, entre cualquier dos potencialmente valores observados, hay un valor que no es posible observar. Una variable es continua si, entre cualquier dos potencialmente valores observables, hay siempre otro potencialmente valor observable.

Ejemplos de variables continuas: edad, presión sanguínea, nivel de colesterol, estatura, peso, etc. Ejemplos de variables discretas: sexo, número de muertos, grupos identificados (1 grupo A y 2 grupo B), y estado de enfermedad (1 si el caso es coronario y 2 si el caso no es coronario).

En el comportamiento con datos, la distribución de frecuencias de una muestra para variables continuas son representadas diferentemente, que para variables discretas. Los datos de una variable continua son usualmente agrupados en clases de intervalos, y una distribución de frecuencias es determinado por observaciones continuas de cada intervalo. Tal distribución es generalmente representada por un histograma. Los datos sobre una variable discreta, no son agrupados pero están representados en su lugar por una linea trazada.

Es importante notar que las variables discretas algunas veces son tratadas como variables continuas, para un análisis. Esto es particularmente usual cuando los posibles valores de tales variables, no están muy alejadas y cubren un amplio rango de números. En tal caso los posibles valores técnicamente abiertos, muestran pequeños vacios entre los valores que en una representación visual se aproximan a un intervalo.

A menudo es usual tratar variables discretas como continuas, algunas variables que son fundamentalmente continuas, pueden ser agrupadas en categorias y tratadas como variables discretas en un análisis dado. Por ejemplo, la variable `edad´ se hace discreta agrupando los valores en dos categorias, `jóvenes´ y `viejos´.

## Orientación descriptiva.

Una segunda forma para clasificar las variables, está basada sobre, si una variable es para describir o para ser descrita por otras variables. Tal clasificación depende de los objetivos de estudio y orientación más bien, que sobre la estructura matemática inherente de la variable misma.

Si la variable bajo investigación es para ser descrita en términos de otras variables, ésta se llama variable dependiente. Si la variable es una que se usa en conjunción con otras variables, para describir una variable dependiente dada, entonces esa variable es llamada variables independiente.

Por ejemplo en 1972 Thompson estudia la relación de la percepción de embarazo del paciente y la satisfacción del paciente con los cuidados médicos, la variable percepción es independiente y la variable satisfacción es dependiente. Similarmente en el estudio de relación de la dureza del agua y la tasa de muerte repentina, la medida de la dureza del agua es una variable independiente, y la tasa de muerte repentina es una variable dependiente.

Usualmente, la distinción entre variables dependientes e independientes es clara como lo es en el ejemplo. Sin embargo, una variable considerada como dependiente para evaluar un estudio especifico, puede ser considerada como independinete, para evaluar un diferente objetivo. Por ejemplo, en el estudio de Thompson, en adición a determinar la relación de percepción como variable independiente para la satisfacción del paciente, fue también de interés a determinar la relación de clase social, edad y educación, para tratar la percepción como variable dependiente.

Los niveles de medición.

La tercera forma de clasificar es encontrar el nivel de precisión matemática de medición de las variables.

En la obtención de estadísticas se manipulan conjuntos con un determinado o indeterminado número de unidades, estas pueden ser (personas, objetos, animales, etc.), las cuales poseen determinadas caracteristicas. Un ejemplo de unidades en un conjunto son los estudiantes, entre sus múltiples caracteristicas se pueden señalar: sexo, edad, estatura, peso, lugar de nacimiento, estrato social, grado de escolaridad, promedios obtenidos, actitudes y opiniones entorno a muchas situaciones, coeficientes de inteligencia, problemas familiares etcétera.

Para elaborar estadisticas repecto a las unidades y sus caracteristicas es necesario contarlas, jerarquizarlas y medirlas.

Dentro de las caracteristicas señaladas, sólo es posible clasificar algunas; es decir, agruparlas de acuerdo a subclases o subconjuntos. Por ejemplo, en la caracteristica sexo se puede determinar cuantos son hombres y cuantos son mujeres, clasificando el conjunto de personas, en dos subconjuntos, hombres y mujeres.

En otro caso podemos jerarquizar las unidades; esto es, clasificarlas en un orden decreciente o creciente; por ejemplo en las caracteristicas coeficiente de inteligencia y estrato social al que pertenecen. Por último algunas caracteristicas pueden ser medidas en un sentido estricto, por ejemplo: ingreso, estatura, edad, peso, tiempo diario de trabajo, etc.

Se manifiesta lo anterior para que sea claro el hecho de que los datos numéricos, pueden diferir en cuanto al tipo de medición que es factible aplicarles, según sea una u otra medición la que admitan los datos; también será diferente el tipo de manipulación matemática a la que se sometan.

Por lo general se usan tres niveles de medición: nominal o clasificatorio, ordinal y de intervalo.

## Nivel nominal o clasificatorio.

En este nivel de medición las unidades (personas, objetos, animales, etc.) únicamente se clasifican de acuerdo a las caracteristicas que se pretenden analizar. Para distinguir los agrupamientos de unidades se emplean simbolos, letras o números. En el caso de que se empleen números para distinguir las clasificaciones o subclases, éstos tienen un valor simbólico y no numérico. Por ejemplo dos subclases, hombres y mujeres, se pueden distinguir con diferentes simbolos.

Hombres: H, A X $\overset{\uparrow}{O}$ 1

Mujeres: M, B Z O 2

Algunas clasificaciones son más fáciles de realizar que otras porque es fácil establecer los criterios de clasificación. En ocaciones resulta difícil establecer quién es un peón y quién un obrero; un obrero calificado de un obrero no calificado; una persona activa y una no activa económicamente.

La única relación entre las unidades que componen el conjunto es la equivalencia, simbolizada por el signo igual (=); es decir que los miembros del subconjunto deben ser equivalentes repecto a las propiedades o caracteristicas, que los constituyen como similares.

Nivel ordinal.

En este nivel las unidades de las subclases guardan una cierta relación entre sí, esto se pone de manifiesto cuando es posible establecer la relación mayor que ( > ) o menor que ( < ) respecto a las caracteristicas de las unidades escaladas. Por ejemplo cuando se clasifican a los miembros de una comunidad en estratos alto, medio y bajo, se puede establecer que: alto > medio > bajo; o a la inversa: bajo < medio < alto.

Muchas de las pruebas psicológicas de habilidades, o de aptitudes y las escalas de opinión, tienen la particularidad de construir escalas ordinales.

En algunos casos puede establecerse la relación > o < sólo para algunos pares de subclases. En este caso la escala se denomina parcialmente ordenada.

En el nivel ordinal la distancia entre dos unidades no es conocida. También los números que se asignan a las caracteristicas permiten determinar el orden o la posición jerárquica en una escala, pero no tiene significado en lo referente a ¿cuánto? o ¿cuántas veces?, porque son cualidades no aditivas.

Nivel de intervalo.

El nivel de intervalo tiene, además de las propiedades de la escala ordinal, la propiedad de que la distancia entre dos valores es una magnitud conocida, lo cual da a esta escala un mayor grado de perfección. En la escala de intervalo el punto cero y la unidad de medición son arbitrarios. La razón entre dos intervalos es siempre independiente del punto cero y de la unidad que se emplee en la medición.

Ejemplos de escalas de intervalos lo constituyen las escalas empleadas en la medición de la temperatura: grados centígrados, Fahrenheit y Kelvin; las escalas de actitudes, las puntuaciones (IQ) de inteligencia, altura, peso, presión sanguinea, etc.

De las escalas ya men c ionadas, nominal, ordinal y de intervalo, esta última es la primera realmente cuantitativa y admite todas las manipulaciones aritméticas.

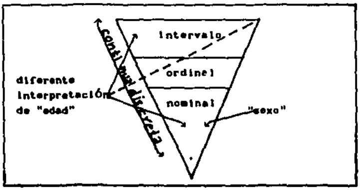
Otra forma de clasificar las variables es por el nivel de medición, la misma variable es considerada con un nivel de medición en un análisis y con diferente nivel en otro análisis. Así, la edad puede considerarse variable de intervalo en un análisis de regresión, y ser agrupada en categorias y considerada nominal en un análisis de varianza.

También se nota que varios niveles de precisión matemática son acumulativos. Una escala ordinal posee las propiedades de una escala nominal. Una escala de intervalo es también nominal y ordinal. Las acumulaciones de estos niveles permiten al investigador dejar apoyado en uno o más niveles de medición un análisis de datos. Así, una variable de intervalo es tratada como nominal u ordinal para un particular análisis, y una variable ordinal es analizada como nominal.

La superposición natural de las formas de clasificación.

Es importante señalar que las tres formas de clasificación descritas se traslapan en el sentido, que cualquier variable es etiquetada de acuerdo a cada proyecto. La clase social por ejemplo, es considerada como ordinal, discreta e independiente en un estudio dado; la presión sanguinea es considerada de intervalo, continua y dependiente en el mismo estudio u otros estudios.

Fig. 1.2. Diagrama de superposición en la clasificación de variables.




El diagrama no incluye la clasificación en variables dependientes e independientes, porque es enteramente en función del objetivo de estudio y no de la estructura misma de la variable. Leyendo el diagrama se considera cualquier variable como ser representada por algún punto dentro del triángulo. Si el punto cae debajo de la linea punteada dentro del triángulo, es clasificada como discreta, si caen arriba de esta linea, es continua. También un punto que caiga en el área marcada con `intervalo´ es clasificada como una variable de intervalo;

similarmente para los otros dos niveles de medición.

Se observa en la figura que cualquier variable nominal es discreta, una variable discreta puede ser nominal, ordinal o de intervalo. Una variable continua es ambas ordinal o de intervalo. Por ejemplo, \`sexo´ es nominal y discreta; \`edad´ puede ser de intervalo y continua, o si se agrupa en categorias, nominal y discreta: \`clase social´ depende sobre la forma que es medida y el punto de vista del investigador, puede ser ordinal y continua, ordinal y discreta, o nominal y discreta.

## Selección del método de análisis.

La estadistica generalmente usa el término de análisis multivariado, para describir un método, de quien su estructura teórica permite la consideración simultánea de diversas variables dependientes. Por otro lado, el investigador en la biomédica y la salud social quien no son estadisticos, miran este término como descubrir cualquier técnica estadistica incluyendo diversas variables, regularmente si sólo una variable dependiente es considerada en un tiempo.

Cualquier investigador orientado a la necesidad de analizar datos, requiere de un razonamiento para seleccionar entre alternativas de métodos de análisis. Diversas consideraciones se introducen en tal selección: 1) La propuesta del investigador, 2) La general caracteristica matemática de las variables involucradas, 3) Las suposiciones estadisticas hechas acerca de estas variables y 4) La manera con la cual los datos son recopilados (el procedimiento de la muestra). Los conocimientos de las primeras dos consideraciones son suficientes, para guiar al investigador en la elección del análisis apropiado.

La tabla 1.2 provee en lineas generales la guia para ayudar al investigador en la selección, del método de análisis donde diversas variables son involucradas.

Esta guia distingue varios Métodos multivariados entre los cuales esta el Análisis factorial que tiene relevancia en el desarrollo de este trabajo. Considera los tipos de conjuntos de variables usualmente asociados con cada método, y da una descripción general de la propuesta de cada método de análisis.

Tabla 1.2 Guia para métodos multivariados.[6]

| Nombre del Método | Clasificación de variables Dependientes | Clasifcación de varables Indepedientes | Propuesta General |
|---|---|---|---|
| Análisis de Regresión Múltiple | Continuas | Clásicamente todas continuas, pero, prácticamente cualquier tipo es usado. | Para describir la extensión, dirección y fuerza de las relaciones entre diversas variables independientes y dependientes continuas. |
| Análisis de Varianza | Continuas | Todas nominal | Para describir la relación entre variables continuas dependientes, y una o más variables independientes nominal. |
| Análisis de Covarianza | Continuas | Mixtas de variables nominal y variables continuas (la segunda será como variable de control)* | Para describir la relación entre una variable continua dependiente y una o más variables nominal independientes, controlando para el efecto de una o más variables continuas independientes. |
| Análisis Discriminante. | Nominal | Clásicamente todas continuas pero, práctica mente, pueden | Para determinar como una o más variables independientes pueden ser usadas,para discri- |

[6] Ibid., KLEINBAUM G. David y Lawrence L. Kupper. Págs. 11.

Continúa la tabla 1.2.

| Nombre del Método | Clasificación de variables Dependientes | Clasificación de variables Independientes | Propuesta general |
|---|---|---|---|
| | | tener una mixta de varios tipos, algunas son continuas. | minar entre diferentes categorías de una variable nominal dependiente. |
| ** Análisis Factorial. | (Las variables usadas en el análisis factorial son clásicamente continuas, perc, prácticamente, pueden ser de cualquier tipo. Estas variables no son claramente identificables como son ambas dependientes e independientes, los factores resultantes pueden ser usados como variables independientes o dependientes en un segundo análisis). | | Construcción de variables o reducción usando diversas variables, para definir una o más nuevas variables compuestas llamadas factores. |
| Análisis de datos por categorías usando modelos lineales. | Nominal | Principalmente nominal, pero algunas veces ordinal. | Para describir la relación entre una variable dependiente nominal y diversas variables nominal o indepndiente ordinal, aplicadas a situaciones enolviendo sólo variables dependientes. |

Generalmente hablando una variable ‹de control› es una variable, que debe considerarse que antes de cualquier relación de interés puede ser cuantificada; esto es porque una variable de control es una que posiblemente es relacionada a las variables de primer interés, es tomada en consideración en estudios de relación entre las variables primarias.

Por ejemplo, en la descripción de la relación entre presión sanguínea y actividad física, probablemente se consideren la edad y el sexo como variables de control, porque ellas están relacionadas a la presión sanguínea y actividad física, a no ser que se tome en cuenta, cualquier conclusión considerando la primera relación de interés.

** El proceso del análisis factorial como técnica estadística sirve de prueba independiente, para someter grupos de variables y ayudar a definir las relaciones funcionales de éstas, y así proporcionar información sobre lo correcto de una hipótesis de investigación.

Frecuentemente el análisis factorial es usado en conjunción con otros métodos multivariados, tales como el análisis de Regresión, análisis Discriminante, etc.

# C A P I T U L O  II

## ELEMENTOS MATEMÁTICOS Y ESTADÍSTICOS

### 2.1. Algebra de Matrices y vectores.

El análisis factorial es tratado generalmente combinando algebra lineal, geometría multidimensional y estadística. Muchas ideas involucradas no son familiares para algunas personas, sobre todo en las áreas sociales. El propósito de este capitulo es definir e ilustrar, las bases principales comprendidas en la estructura fundamental de esta técnica de análisis, facilitando así su entendimiento.

Los ejemplos presentados a manera de ilustración son retomados de la bibliografía señalada para el presente trabajo.

#### Matrices.

Una matriz se define como un sistema de mxn elementos $a_{ij}$ en un arreglo ordenado rectangular o cuadrado, con uno o más renglones o una o más columnas. Se representa de la siguiente forma:

$$A_{(mxn} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdot & \cdot & \cdot & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdot & \cdot & \cdot & a_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{m1} & a_{m2} & a_{m3} & \cdot & \cdot & \cdot & a_{mn} \end{bmatrix} \qquad (2.1)$$

Los elementos o componentes de una matriz se simbolizan por $(a_{ij})$, donde $i=1, \ldots, m$, y $j=1, \ldots, n$. El primer subíndice se refiere al i-ésimo renglón, y el segundo subíndice a la j-ésima columna. Así, los subíndices i, j señalan la posición de cualquier elemento relacionado con la matriz.

Una matriz es representada convencionalmente por letras mayúsculas A, B, C, etc.

El orden de la matriz es el tamaño de la matriz en términos de los números de renglones y columnas. El orden es denotado por medio de subíndices asociados a la letra que simboliza la matriz. Por ejemplo, la matriz $X_{(8x4)}$ es de orden 8 por 4 y su tamaño es de ocho renglones y cuatro columnas. También se puede saber el número total de elementos que tiene la matriz X, multiplicando los subíndices 8x4 = 32 elementos.

La dimensión de la matriz A será referida como de orden mxn, es decir m-renglones y n-columnas.

El renglón $i$ de la matriz A se denota por $A_i$ , y se define como:

$$A_i = \begin{bmatrix} a_{i1} , a_{i2} , \ldots , a_{in} \end{bmatrix} \tag{2.2}$$

y la columna $j$ se denota por $A^j$ y de define como:

$$A^j = \begin{bmatrix} a_{1j} \\ a_{2j} \\ . \\ a_{mj} \end{bmatrix} \tag{2.3}$$

Ejemplo de una matriz de datos:

Tabla 2.1. Matriz de datos de cinco variables socio-económicas.

| No. de Región | v a r i a b l e s | | | | |
|---|---|---|---|---|---|
| | Población Total | Promedio de Escolaridad | Población Empleada | Serv.Prof. Miscelaneos | Valor Prom de la casa |
| 1 | 5700 | 12.8 | 2500 | 270 | 25000 |
| 2 | 1000 | 10.9 | 600 | 10 | 10000 |
| 3 | 3400 | 8.8 | 1000 | 10 | 9000 |
| 4 | 3800 | 13.6 | 1700 | 140 | 25000 |
| 5 | 4000 | 12.8 | 1600 | 140 | 25000 |
| 6 | 8200 | 8.3 | 2600 | 60 | 12000 |
| 7 | 1200 | 11.4 | 400 | 10 | 16000 |
| 8 | 9100 | 11.5 | 3300 | 60 | 14000 |
| 9 | 9900 | 12.5 | 3400 | 180 | 18000 |
| 10 | 9600 | 13.7 | 3600 | 390 | 25000 |
| 11 | 9600 | 9.6 | 3300 | 80 | 12000 |
| 12 | 9400 | 11.4 | 4000 | 100 | 13000 |

La tabla 2.1 representa una matriz de datos de $M_{(12x5)}$, la cual contiene información sobre cinco variables socio-económicas para doce regiones.

Definición de vector.

Un vector se define como una matriz con un sólo renglón o columna. Esto es cualquier arreglo de 1xn elementos es llamado vector renglón y el arreglo de nx1 elementos se llama vector columna.

El vector columna de n-componentes se escribe de esta forma:

$$x = \begin{bmatrix} x_1 \\ . \\ . \\ x_n \end{bmatrix} \tag{2.4}$$

similarmente el vector renglón de n-componentes.

$$x' = \begin{bmatrix} x_1, & . & . & . & ,x_n \end{bmatrix} \tag{2.5}$$

el vector renglón x' representa la transpuesta del vector columna x.

Algunas veces se escribe $x = (x_i)$, para denotar el vector columna x con elementos $x_1$, $x_2$,..., $x_n$. Para denotar un vector se usan letras minúsculas.

Cualquier vector especifica las coordenadas de un punto en un espacio euclidiano de n-dimensiones.

Un ejemplo simple de un vector renglón son las notaciones (x,y) y (x,y,z), para los puntos en un plano y un espacio, los elementos de estos vectores son las coordenadas de los puntos.

Definición de escalares.

Un escalar puede distinguirse como una matriz de orden 1x1, es conocido como una constante o número que afecta a una matriz en forma de producto, los escalares se designan con letras minúsculas, ejemplo, el escalar k y la matriz A, se define (kA ó Ak), esto es a cada elemento de la matriz A le corresponde k veces dicho escalar.

Tipos comunes de matrices.

## Matriz transpuesta.

La transposición de una matriz significa intercambiar renglones por columnas y viceversa, en cualquier matriz dada.

En general, si $A_{(rxc)}$ es cualquier matriz su transpuesta es la matriz $A'_{(cxr)}$. La transpuesta de A se simboliza por A' `A prima´ o $A^t$.

Los elementos $A_{ij}$ de la matriz A vienen a ser elementos $A_{ji}$ en la transpuesta A'.

La transpuesta de un vector renglón es un vector columna o viceversa.

Si A es la matriz:

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1c} \\ \cdots & \cdots & \cdots \\ a_{r1} & \cdots & a_{rc} \end{bmatrix}, \text{ entonces } \underbrace{A' = \begin{bmatrix} a_{11} & \cdots & a_{r1} \\ \cdots & \cdots & \cdots \\ a_{1c} & \cdots & a_{rc} \end{bmatrix}}_{\text{transpuesta}} \quad (2.6)$$

El orden de la transpuesta es el reverso del orden de la matriz original, $A_{rxc}$ , $A'_{cxr}$ .

## Matriz cuadrada.

Una matriz con el mismo número de renglones y columnas es llamada una matriz cuadrada. Si los subíndices son iguales (i.e, i=j) estos forman la diagonal principal de la matriz. Son los elementos extendidos desde la esquina superior izquierda, hasta la esquina inferior derecha.

Ejemplo:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}_{3x3}$$

La diagonal principal esta formada por los elementos $(a_{11}, a_{22}, a_{33})$.

Los elementos de la diagonal son las varianzas de las variables, y los elementos fuera de la diagonal es la covarianza.

La traza de la matriz A se define como la suma de elementos de la diagonal, y es denotada por $tr = \sum_{i=1}^{n} a_{ij}$, con $i=j$.

Matriz simétrica.

Si una matriz cuadrada A es igual a su transpuesta A', esto es A = A', se dice que la matriz A es simétrica.

Ejemplo:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 3 & 5 & 6 \end{bmatrix} \qquad A' = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 3 & 5 & 6 \end{bmatrix}$$

A = A', entonces A es simétrica, esto es todos los elementos $a_{ij} = a_{ji}$ $(i, j = 1, 2, \ldots, n)$. También se dice que la matriz A es simétrica respecto a su diagonal principal. Puesto que abajo y arriba de la diagonal están los mismos elementos. Frecuentemente se omite la duplicidad de los elementos de abajo, y sólo se representan en una matriz simétrica los elementos de arriba de la diagonal.

Una tabla de intercorrelaciones es una matriz simétrica, con entradas no numéricas en la diagonal principal.

## Matriz diagonal y escalar.

Una matriz diagonal es una matriz simétrica cuadrada, la cual tiene todos sus elementos de la diagonal principal mayores que cero, mientras el resto de los elementos de la matriz son ceros, esto es $a_{ij} = 0$ si $i \neq j$.

El aspecto de una matriz diagonal es:

$$D = \begin{bmatrix} a_1 & 0 & . & . & . & 0 \\ 0 & a_2 & & . & . & 0 \\ \vdots & \vdots & & \vdots & . & . \\ 0 & 0 & & 0 & & a_n \end{bmatrix} \qquad (2.7)$$

La matriz diagonal será denotada por diag (A) = D $(a_{ii})$.

La matriz escalar es una especial instancia de la matriz diagonal. Los elementos de la diagonal principal en la matriz escalar son todos iguales, y tiene la forma siguiente:

$$K = \begin{bmatrix} k & 0 & . & . & . & 0 \\ 0 & k & . & . & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & . & . & . & k \end{bmatrix} \tag{2.8}$$

Si una matriz escalar es premultiplicada o posmultiplicada por cualquier matriz A del mismo orden K, las siguientes relaciones son evidentes. $KA = AK = kA$, lo cual no necesita demostración.

Matriz identidad.

La matriz identidad es también un caso especial de la matriz diagonal, está formada por unos en la diagonal principal, y el resto de los elementos son ceros. Es representada por la letra I. La matriz identidad tiene la similar función en algebra lineal como el número 1 en aritmética.

La matriz identidad tiene la siguiente forma:

$$I = \begin{bmatrix} 1 & 0 & . & . & . & 0 \\ 0 & 1 & . & . & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & . & . & . & 1 \end{bmatrix} \tag{2.9}$$

Con la propiedad, si A es cualquier matriz entonces.

$IA = AI = A$

Matriz Gramian.

Es una matriz de especial interés para el análisis factorial, frecuentemente es referida en la literatura psicológica como una matriz "Gramian" o una matriz con propiedades Gramian. Estas propiedades incluyen la simetria y la *semidefinición positiva*. Una matriz $A_{nxn}$ cuadrada simétrica, es declarada definida positiva si, todos los menores asociados con los elementos de la diagonal principal, son mayores que cero.

Todas las matrices de correlación con unidades en la diagonal principal son matrices gramian y la comunalidad estimada es reemplazada por los valores de la diagonal.

**Determinantes.**

El determinante es un escalar derivado de la operación sobre una matriz cuadrada. Un determinante de nxn es de orden n. El determinante de la matriz cuadrada A es denotado como $|A|$.

Para la matriz $A_{2x2}$, el determinante es de segundo orden y se obtiene en la siguiente forma:

$$|A| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21} \tag{2.10}$$

Para la matriz $A_{3x3}$, la expresión es:

$$|A| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} \\ a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} \tag{2.11}$$

La matriz con determinante igual a cero es llamada singular. Las matrices no singulares tienen determinante mayor que cero.

Menores de un determinante y cofactores.

Un menor es una clase especial de determinante. Dada una matriz A, el menor $M_{ij}$ es definido como el determinante de la matriz, formada al borrar el i-ésimo renglón y la j-ésima columna de A. Existe, por lo tanto un menor correspondiente por cada elemento de A. Entonces para un determinante nxn hay $n^2$ primeros menores.

Ejemplo:

$$A = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \quad \Rightarrow \quad a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}$$

El primer menor $a_{11}$ ( la intersección renglón-columna ) da como resultado el determinante de orden 2. La matriz A tiene $3^2=9$ primeros menores.

Los cofactores de $a_{ij}$ es el menor multiplicado por $(-1)^{i+j}$ y será escrito como $A_{ij} = (-1)^{i+j} M_{ij}$. (2.12)

Los signos algebraicos fijados a los menores para obtener los correspondientes cofactores, son alternadamente (+) y (-).

El determinante de una matriz cuadrada A puede ser expresado en términos de cofactores de los elementos de cualquier columna o renglón dado.

$$
\begin{aligned}
|A| &= a_{i1}\, A_{i1} + \,.\,.\,.\, + a_{in}\, A_{in}. \qquad i = 1, \,.\,.\,.\,, n \qquad (2.13)\\
&= a_{1j}\, A_{1j} + \,.\,.\,.\, + a_{nj}\, A_{nj}. \qquad j = 1, \,.\,.\,.\,, n
\end{aligned}
$$

Ejemplo:

El det A expresado en cofactores.

$$
\det A = \begin{vmatrix} \overset{+}{a_{11}} & \overset{-}{a_{12}} & \overset{+}{a_{13}} \\ \overset{-}{a_{21}} & \overset{+}{a_{22}} & \overset{-}{a_{23}} \\ \overset{+}{a_{31}} & \overset{-}{a_{32}} & \overset{+}{a_{33}} \end{vmatrix}
$$

$$
-a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} = -a_{12}\left( a_{21}a_{33} - a_{31}a_{23} \right)
$$

$$
+a_{22}\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} = +a_{22}\left( a_{11}a_{33} - a_{31}a_{13} \right)
$$

$$
-a_{32}\begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix} = -a_{32}\left( a_{11}a_{23} - a_{21}a_{13} \right)
$$

Por lo tanto $\det A = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{32}a_{21} - a_{13}a_{22}a_{31} - a_{23}a_{32}a_{11} - a_{33}a_{21}a_{12}$

Matriz inversa.

Si una matriz cuadrada A es no singular, esto es que el determinante de A sea diferente de cero $|A| \neq 0$, entonces existe otra matriz en la cual los $A_{kj}$ denotan los cofactores de los elementos de A, y la matriz de estos cofactores ( con $1/|A|$ factorizada ) es llamada la adjunta de la matriz A. La matriz simbolizada $A^{-1}$, con elementos denotados por $a^{jk}$, es llamada la inversa de A, y ella misma es una matriz no singular con la siguiente propiedad.

$$A\,A^{-1} = A^{-1}\,A = I \tag{2.14}$$

Sea

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \quad \text{con } |A| \neq 0,$$

entonces

$$A^{-1} = 1/|A| \begin{bmatrix} A_{11} & A_{21} & \cdots & A_{n1} \\ A_{12} & A_{22} & \cdots & A_{n2} \\ \cdot & \cdot & \cdots & \cdot \\ A_{1n} & A_{2n} & \cdots & A_{nn} \end{bmatrix} = \begin{bmatrix} a^{11} & a^{21} & \cdots & a^{n1} \\ a^{12} & a^{22} & \cdots & a^{n2} \\ \cdot & \cdot & \cdots & \cdot \\ a^{1n} & a^{2n} & \cdots & a^{nn} \end{bmatrix} \tag{2.15}$$

Rango de una matriz.

La forma de determinar el rango de una matriz es encontrar el orden del máximo menor, y debe ser diferente de cero. Por ejemplo, si el determinante de una matriz de 4x4 es evaluado y diferente de cero, entonces el rango de la matriz es 4. Si el determinante es igual a cero, la matriz es de rango 3 o menor.

Vector normal y ortonormal.

Vector normal es un vector con todos sus elementos iguales a 1.

Vector nulo es un vector con todos sus elementos iguales a 0.

Vectores normalizados.

Un vector es normalizado si tiene longitud de uno.

Cualquier vector x, puede ser normalizado dividiendo cada uno de sus elementos por la longitud del vector[1]:

$$x / ( x'x )^{1/2} \tag{2.16}$$

Vector ortogonal.

Dos vectores son ortogonales si $u'v = 0$, entonces $u$ y $v$ son ortogonales cada uno respecto al otro. Para que dos vectores, $u$ y $v$ tengan la propiedad de ortogonalidad, esto implica que los vectores esten separados por $90^0$.

[1] La longitud de un vector representada por $(x'x)^{1/2}$, es la raiz cuadrada de la suma de los elementos del vector al cuadrado, por ejemplo $V = [3 \;\; 4]$ la longitud del vector $V$ denotada por $|V| = [3^2 + 4^2]^{1/2} = 5$, esto es igual a $(V'V)^{1/2} = ([3 \;\; 4] \begin{bmatrix} 3 \\ 4 \end{bmatrix})^{1/2} = 9 + 16 = (25)^{1/2} = 5$.

Matriz ortogonal.

Una matriz cuadrada es ortogonal si y sólo si $A^{-1} = A'$. (2.17)

De esta definición se derivan dos casos:

Cuando el resultado del producto $A'A = I$ es la matriz identidad, entonces la matriz A es declarada ortonormal.

Si el producto $A'A = D$ resulta una matriz diagonal, entonces la matriz A es llamada ortogonal.

El determinante de una matriz ortogonal es mayor o menor que uno.

## 2.2. Conceptos Esenciales de Geometria.

### Dependencia lineal.

La natural relación entre vectores columna o renglón contenidos en una matriz, es descrita en términos de dependencia lineal e independencia lineal.

Un vector que es múltiplo escalar de otro vector, una poderación o una suma ponderada de un conjunto de vectores, se dice que son linealmente dependientes. Si sucede lo siguiente:

Para un vector, $y$ y un conjunto de vectores, $v_1$, $v_2$, . . . , $v_n$, la dependencia lineal se establese por la relación.

$$y = a_1v_1 + a_2v_2 + \ldots + a_mv_m \qquad (2.18)$$

con al menos un escalar $a_i \neq 0 \quad (i = 1, \ldots, m)$

Un vector es linealmente independiente de un conjunto de vectores, si no es un múltiplo escalar o una suma ponderada de cualquier combinación de los miembros del conjunto.

Para la relación

$$y = a_1v_1 + a_2v_2 + \ldots + a_mv_m \qquad (2.19)$$

Todos los $a_i = 0 \quad (i = 1, \ldots, m)$ en el orden que $y$ es linealmente independiente del conjunto de vectores $v_i$.

Transformaciones ortogonales.

El concepto de una transformación ortogonal se expresa más claramente en notación matricial.

Sea el conjunto de coordenadas iniciales y coordenadas finales, en el espacio N se representan por renglones en las siguientes matrices:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1N} \\ x_{21} & x_{22} & \cdots & x_{2N} \\ \vdots & \vdots & & \vdots \end{bmatrix} \quad y \quad Y = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1N} \\ y_{21} & y_{22} & \cdots & y_{2N} \\ \vdots & \vdots & & \vdots \end{bmatrix}$$

entonces, si existe T

$$T = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1N} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2N} \\ \vdots & \vdots & \cdots & \vdots \\ \alpha_{N1} & \alpha_{N2} & & \alpha_{NN} \end{bmatrix}$$

La transformación ortogonal se da con la relación:

$$Y = X\,T' \quad \text{o sea} \quad y_{ji} = \sum_{k=1}^{N} \alpha_{ij}\, x_{jk} \qquad (2.20)$$

siempre y cuando la matriz T sea ortogonal, cumpliendo la propiedad $T'T = I$ .

Interpretación geométrica de la correlación.

La interpretación geométrica de la matriz de correlación juega una parte importante en el análisis factorial.

Una medida de asociación entre dos variables es dada por el coeficiente de correlación, el cual es representado en términos de vectores que son expresados en forma de valores desviados, es decir la desviación del valor de la variable con respecto a la media de dicha variable. Entonces el coeficiente de correlación es representado por:

$$r_{ij} = y_i' \, y_j \,/\, \left[(y_i' \, y_i)^{1/2} \, (y_j' \, y_j)^{1/2}\right] \qquad (2.21)$$

$y_{nj} = x_{nj} - \bar{x}_j$ ; $y_{ni} = x_{ni} - \bar{x}_i$ desviación de $x_j$ , $x_i$.

Se puede ver que el coeficiente de correlación es el menor producto momento[2] de dos vectores, dividido por el producto de sus respectivas longitudes.

$y_i' \, y_j = s_{ij}$ es la covarianza de las variables $x_i$ y $x_j$ .[3]

[2] El producto menor es un escalar, el cual es la suma de productos de los correspondientes elementos de los vectores; esto es dado algebraicamente como: $x'y = \sum_{k=1}^{p} x_k \, y_k$ .

El menor producto es definido solamente por vectores de orden igual.

El menor producto momento es el menor producto de un vector columna pre-multiplicado por su transpuesta, y representa la suma de los elementos del vector al cuadrado.

[3] La matriz de covarianza es igual al producto de la matriz de datos por su transpuesta, expresada en la forma de desviaciones, y cada elemento didivido por N.

Esta ecuación es la misma que se usa para calcular el ángulo entre dos vectores.

$$\cos \theta = \frac{v'w}{(v'v)^{1/2}\ (w'w)^{1/2}}$$

$v'w$ es la longitud de los vectores.

Estas ecuaciones nos conducen a la interpretación de la correlación entre variables, como los ángulos entre vectores de variables en un espacio. La matriz de correlación R, contiene los cosenos de todos los posibles ángulos de separación entre todas las variables. El concepto se ilustra en la Fig. 2.1.

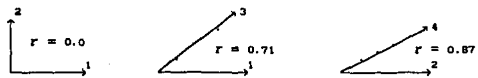



Fig. 2.1. Ejemplos de correlaciones interpretadas geométricamente.

Otra forma de interpretar el coeficiente de correlación es refiriéndose a un diagrama de dispersión, como un plano de puntos situados en el espacio de pares de variables estandarizadas. El coeficiente de correlación mide la tendencia de los puntos agrupados a lo largo de la línea.

## 2.3. Rotación de Sistemas de Coordenadas.

Una tabla de correlaciones tiene una sigular configuración de vectores, esta configuración puede ser descrita con respecto a un infinito número de localizaciones de la estructura de los ejes de coordenadas de referencia. La localización se obtiene por la rotación de los ejes de referencia alrededor del origen. Esta situación es considerada en el análisis factorial al rotar los factores iniciales.

La rotación de un conjunto de ejes de coordenadas, para una nueva localización es usulmente llamada rotación.

Dado un vector renglón $x' = [3 \quad 2]$ cuyos elementos son las coordenadas de un punto con respecto a los ejes coordenados, para dos variables. Suponemos que por algún razonamiento se necesita rotar rigidamente los ejes, a través de un ángulo con ciertos grados en una dirección contraria a las manecillas del reloj. Entonces el problema es encontrar las coordenadas de los puntos con referencia para los nuevos ejes, $y_1$ y $y_2$ .

Desde la trigonometria elemental, las coordenadas requeridas son dadas por las ecuaciones:

$$y_1 = \cos\phi \; x_1 + \sin\phi \; x_2 \qquad (2.22)$$

$$y_2 = - \cos\phi \; x_1 + \cos\phi \; x_2$$

Estas dos ecuaciones pueden ser escritas en forma matricial.

$$y' = x'T$$

$$\text{donde:} \quad T = \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} \qquad (2.23)$$

Para una colección de N vectores renglón en la matriz $X_{(Nxp)}$, en la ecuación $Y = X T$. (2.24)

Y representa las coordenadas de los N vectores renglón en términos de los p-ejes rotados rigidamente.

T es denominada matriz de transformación poniendo X en Y. T puede ser una matriz ortogonal .

Las distancias y ángulos entre vectores no cambiarán durante la rotación rigida de los ejes de referencia.

Interpretación de la rotación.

La multiplicación de una matriz de datos por una matriz ortogonal, se considera como un simple proceso mecánico de rotación rigida. Dos rasgos de este proceso son usuales, para hacerse una idea dentro de la manipulación de datos por la operación de matrices.

Primeramente, los nuevos ejes rotados pueden ser interpretados como variables. Los vectores columna en Y de la Ec. (2.24) son núevas variables las cuales son combinaciones lineales de los vectores columna (variables) de X. Los elementos en la matriz T vienen a ser los coeficientes en la combinación lineal.

El segundo rasgo, es que los vectores renglón de X dan la composición de las unidades de estudio en términos de las variables medidas originalmente. Los vectores renglón de Y dan la composición de las unidades de estudio u observaciones en términos de las nuevas variables derivadas.

## 2.4. Conceptos Estadisticos.

### Variables.

Un estudio estadistico tipicamente involucra un grupo de individuos con algunos atributos comunes. El término ‹individuos› se usa para estandarizar los objetos o entidades como personas, censos en regiones, fábricas etc. Las medidas hechas sobre tales individuos, o atributos de estas entidades son designadas como variables.

La letra N es usada para representar el número total de individuos, y la letra n para el número de variables. En particular una variable es denotada por $x_j$ , esta puede ser cualquiera de las n variables ($j = 1, \ldots, n$). El indice i se emplea para designar cualquiera de los N individuos o unidades de estudio, ($i = 1, \ldots, N$). Entonces, el valor de una variable $x_j$ para el individuo i es representado por $x_{ji}$ , el orden de los subindices son importantes.

Un particular $x_{ji}$ es llamado valor observado, el cual es medido desde un origen arbitrario y por una unidad arbitraria.

### La media.

Sea una muestra de N observacines y n variables $x_1, x_2, \ldots x_j$ , la media aritmética de cualquier variable $x_j$, para alguna caracteristica, es definida por:

$$\bar{X}_j = \sum_{i=1}^{N} x_{ji} \;/\; N \qquad (2.25)$$

Donde $x_{ji}$ es el valor observado en la unidad de estudio i de la variable $x_j$ .

N es el número total de observaciones en la muestra.

La media es una medida de tendencia central y representa el promedio de todos los valores observados $x_{ji}$ dentro de la muestra. Esto quiere decir que los valores $x_{ji}$ de la variable $x_j$ están por arriba o por abajo de este promedio $\overline{X}_j$ .

La desviación de un valor observado.

Los valores observados de las variables pueden ser transformados en una forma más conveniente, por fijación del origen y la unidad de medida. Cuando el origen es ocupado por la media muestral, un particular valor con la forma:

$$x_{ji} = x_{ji} - \overline{x}_j \qquad (2.26)$$

es llamada la desviación de $x_{ji}$ .

El promedio de las desviaciones de los valores observados es siempre cero, esto es:

$$\sum_{j=1}^{N} (x_{ji} - \overline{x}_j) \ / \ N = 0$$

Varianza.

La varianza de una variable $x_j$ es la medida de la dispersión de valores individuales acerca de la media. Es definida como el promedio de la suma del cuadrado de la desviación de. valores observados.

La varianza muestral de la variable $x_j$ es definida por:

$$S_j^2 = \sum (x_{ji} - \overline{x}_j)^2 \ / \ N \qquad (2.27)$$

$$S_j^2 = \sum x_{ji}^2 \ / \ N$$

El valor de la varianza indica que tan cerca o alejados están los valores observados respecto a la media.

Covarianza.

La covarianza expresa la relación entre dos variables, para cualquiera dos variables $x_k$ y $x_j$, la covarianza es definida como el promedio de la suma del producto de las desviaciones, para todas las observaciones.

$$S_{jk} = \sum x_{ji} x_{ki} / N \tag{2.28}$$

$$S_{jk} = \sum (x_{ji} - \bar{x}_j)(x_{ki} - \bar{x}_k) / N$$

Comparándola con la fórmula de la varianza, ésta es un caso especial de la fórmula para la covarianza.

Una matriz de covarianza es simétrica y cuadrada de orden p. Sus elementos en la diagonal principal son las varianzas de las variables.

Desviación estándar.

La desviación estándar es definida como la raiz cuadrada positiva de la varianza.

$$S_j = \sqrt{\sum (x_{ji} - \bar{x}_j)^2 / N} \tag{2.29}$$

La longitud de un vector se define como $|x| = (x'x)^{1/2}$, la desviación estándar de una variable es proporcional a la longitud del vector de observaciones sobre la variable. Así, la desviación estándar también se expresa como: $S_j = |x| N^{-1/2}$ .

Esta relación permite la interpretación de la desviación estándar, en términos de longitud de vectores de variables en forma de desviación.

La desviación estándar representa la varianza, pero en términos de unidades simples y no cuadráticos, para una mejor comparación con los valores de las variables.

Estandarización de variables o puntajes.

Frecuentemente es usual expresar las observaciones en términos de sus desviaciones desde la media, usando la desviación estándar como unidad de divergencia.

La estandarización de un puntaje, para un particular individuo o unidad de estudio, para la variable j es dada por:

$$z_j = (x_j - \bar{x}_j) / S_j \tag{2.30}$$

Las variables estandarizadas tienen una media de cero y una varianza, y desviación estándar, de uno.

La fórmula para estandarizar todas las variables en una matriz de datos es: $Z = Y D^{1/2}$

Donde D denota la matriz diagonal formada por los elementos de la diagonal de la matriz de covarianza S, y Y es la matriz de desviación de puntajes u observaciones.

Coeficiente de correlación.

La correlación entre todas las variables de un estudio son usualmente computadas como un paso inicial en un análisis factorial.

El coeficiente de correlación de Pearson (r de Pearson) es la principal estadística que se usa, para determinar si el cambio de un grupo de medidas está asociado con el cambio de otro grupo.

La medida de asociación entre dos variables es definida como el cociente de la covarianza de las variables entre el producto de sus desviaciones estándar:

$$r_{jk} = S_{jk} / S_j \, S_k \qquad (2.31)$$

Para variables en forma de desviación, la fórmula toma la forma:

$$r_{jk} = \frac{\sum_{i=1}^{N} x_{ji} \, x_{ki}}{\left[ \sum_{i=1}^{N} x_{ji}^2 \, \sum_{i=1}^{N} x_{ki}^2 \right]^{1/2}}$$

Para una matriz de correlación con datos estandarizados, la fórmula es convenientemente representada como: $R = Z'Z / N$ .

La matriz de correlaciones es una matriz simétrica cuadrada. El producto de la matriz Z por su transpuesta Z', dividido por el tamaño de la muestra N es igual a la matriz de correlación R.

Factores que intervienen en la interpretación del coeficiente.

La $r$ de Pearson puede variar en magnitud desde +1.00 hasta -1.00. Una $r$ de Pearson de +1.00 significa que los cambios en un grupo de medidas son exactamente proporcionales a los cambios en el otro grupo.

El inverso de este caso es cuando la $r$ de Pearson es -1.00, indicando que según aumenta en valor numérico una medida, la otra disminuye en una cantidad que ha de ser exactamente proporcional a lo largo de toda la extensión.

La circunstancia de ausencia total de relación, es decir el cambio de una variable carece totalmente de toda relación con el cambio en la otra variable. El patrón de los puntos forma casi un circulo perfecto. En esta condición, el valor de $r$ será entonces de 0.00.

El valor del coeficiente de correlación no es un valor que se pueda interpretar linealmente, esto es una $r$ de .60 no indica una relación doblemente estrecha respecto a una $r$ de .30.

Una manera muy provechosa para entender el significado de una $r$ obtenida es ver su cuadrado. Esto puede interpretarse como el porcentaje de asociación entre las dos variables.

Ahora ya se puede entender que una correlación de .30 es solamente una cuarta parte de una correlación de .60, ya que la primera sólo explica un 9 por ciento de la variación, mientras que la segunda explica el 36 por ciento de la misma.

Otro aspecto de la correlación que debe mencionarse. Es que, el error más común que se comete en la investigación de correlaciones, es precisamente el de establecer una causalidad a base de una relación de correlación.

Limitaciones de la correlación.

Si los datos no cumplen con ciertas condiciones bien determinadas, la mayoría de los coeficientes serán por lo menos inapropiados, y en algunos casos, darán una idea falsa del grado de asociación.

Las medidas que se correlacionan sin determinar sus trazos dispersos, son sospechosas, ya que invariablemente nos llevan a concluciones erróneas sobre el nivel de asociación.

Otra tendencia es la presentar todas las correlaciones, para hacer resaltar sólo las más importantes, sin tener en cuenta cuántas de ellas son importantes por mera casualidad.

La *r* de Pearson exige que el cambio sea lineal, es decir, que el grado de cambio sea constante a través de toda la gama de resultados.

## 2.5. Valores y Vectores Propios.[4]

El rango de la dimensionalidad de la matriz la cual determina el número de vectores linealmente independientes, necesarios para alcanzar el espacio que contienen los vectores de una matriz. Los valores y vectores propios de una matriz no sólo determinan el rango; también producen un conjunto de vectores básicos linealmente independientes.

Algunos sinónimos de lo que significan los valores y vectores propios: raiz latente y vectores, raiz caracteristica y vectores, valor caracteristico y vector caracteristico.

En el caso del análisis factorial las raices caracteristicas, las cuale son soluciones de la ecuación caracteristica, y esta a su vez dará los coeficientes de los factores comunes, para obtener los componentes principales. Estas raices caracteristicas son referidas como "valores propios" de la matriz, y los vectores caracteristicos asociados a las raices son llamados "vectores propios" de la matriz.

Se estudia sólo los valores propios de una matriz simétrica real R, puede ser por ejemplo: la matriz de datos por su transpuesta, la de covarianza o la matriz de correlación.

Un vector propio de la matriz R es un vector $u$ , con algunos elementos iguales a cero, tal que:

$$Ru = u\lambda \tag{2.32}$$

donde $\lambda$ es un vector no conocido.

[4] Son también conocidos como valores y vectores caracteristicos, algunos autores se refieren a ellos como eigenvalores y eigenvectores.

Se necesita encontrar un vector tal que el vector Ru sea proporcional a u .

La ecuación (2.32) es presentada en otra forma, representando así, un sistema de ecuaciones homogéneas.

$$Ru - u\lambda = 0 \qquad ó \qquad (R - \lambda I)u = 0 \qquad (2.33)$$

el 0 es un vector nulo.

Esto implica que el vector u desconocido es ortogonal a todos los vectores renglón de $(R - \lambda I)$ .

Por ejemplo, para una matriz R, de 2x2 el sistema sería el siguiente:

$$\begin{aligned} (r_{11} - \lambda)u_1 + r_{12}\, u_2 &= 0 \\ r_{21}\, u_1 + (r_{22} - \lambda)u_2 &= 0 \end{aligned} \qquad (2.34)$$

Se obtiene una solución no trivial del sistema si $(R - \lambda I)$ es singular, esto es el $|R - \lambda I| = 0$ .

La forma general de (2.34) se conoce como la ecuación característica y esta dada por:

$$\begin{vmatrix} r_{11} - \lambda & r_{12} \\ r_{21} & r_{22} - \lambda \end{vmatrix} = 0 \qquad (2.35)$$

Desarrollando el determinante tenemos:

$$(r_{11} - \lambda)\,(r_{22} - \lambda) - r_{12}\, r_{21} = 0$$

Multiplicando y reordenando los términos resulta:

$$\lambda^2 - \lambda( r_{11} + r_{22} ) + ( r_{11} r_{22} - r_{12} r_{21} ) = 0 \qquad (2.36)$$

La cual es una ecuación cuadrática, de la forma:

$$ax^2 + bx + c = 0 \ , \qquad (2.37)$$

Cuya solución general es:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Recordando que $r_{12} = r_{21}$ , la solución de (2.36) se obtiene por:

$$\lambda = 1/2\left[ -(r_{11} + r_{22}) \pm \{ (r_{11} + r_{22})^2 - 4(r_{11} r_{22} - r_{12}^2) \}^{1/2} \right] \qquad (2.38)$$

Dos valores de $\lambda$ , $\lambda_1$ y $\lambda_2$ , son obtenidos. Estas dos raices son determinadas los valores propios de la matriz R.

En el caso general de una matriz $R_{(pxp)}$ , la ecuación caracteristica será un polinomio de grado p y producirá p raices o valores propios, $\lambda_1,\lambda_2,\ldots,\lambda_p$ . Si R es una matriz simétrica cuadrada y real, estos valores propios son siempre reales. Sin embargo, los p valores propios pueden no ser todos diferentes y algunos ser iguales a cero. Si dos o más valores propios son iguales, se dice que son valores propios múltiplos.

Una vez obtenidos los valores propios, los vectores propios asociados se derivan de la relación (2.33). Pueden obtenerse varias soluciones para un vector propio. Por conveniencia, los vectores propios son siempre normalizados, éstos son contraídos para ser de longitud uno.

Con cada distinto valor propio hay asociado un vector propio único, excepto por el signo éste se multiplica por -1, y los vectores propios asociados con diferentes valores propios son ortogonales.

Si los valores propios $\lambda_i$ $(i = 1, \ldots, p)$ son colocados como los elementos de una matriz diagonal $\Lambda$, y los vectores propios coleccionados como columnas dentro de la matriz U, entonces la relación (2.32) se expresa en forma matricial: $RU = U\Lambda$.

La matriz U es cuadrada ortogonal, así que debe cumplir con: $U'U = UU' = I$.

Entonces cada vector propio es ortogonal a todos los otros en el conjunto.

Algunas propiedades de los valores propios.

1. La traza $\Lambda$ = traza R : la suma de los volores propios es igual a la suma de los elementos en la diagonal principal de la matriz.

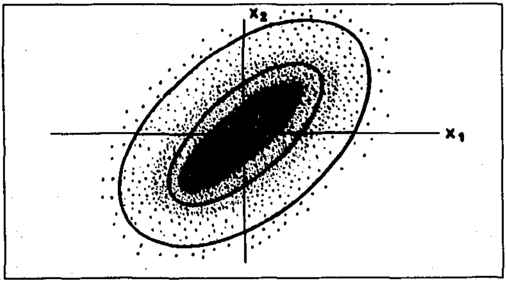
2. $\prod_{i=1}^{p} \lambda_i = |R|$ ; el producto de los valores propios es igual al determinante de la matriz. Si uno o más valores propios son cero, el determinante de R es cero, indicando que R es singular.

3. El número de los valores propios diferentes de cero es iguales al rango de R.

Interpretación geométrica de valores y vectores propios.

Para más de dos variables, los puntos de los datos forman una hiperelipsoide p-dimensional (p variables). Si las variables no están interrelacionadas, la elipse describe un circulo. Si están perfectamente relacionadas, éstas se degeneran en una linea. La Fig. 2.2 muestra un diagrama de dispersión de observaciones sobre dos variables.

Fig. 2.2. Diagrama de dispersión de dos variables, formando un elipsoide.




Pearson (1901) y después, Hotellin (1933), reconocen que el mayor y menor eje de la hiperelipsoide, son fundados desde los vectores propios de la matriz de correlación. La ecuación de la hiperelipsoide es:

$$w'R^{-1}\ w = c \tag{2.39}$$

Donde R es la matriz de correlación o covarianza de los datos, $w$ es un vector de coordenadas para los puntos sobre la elipsoide y c es una constante.

La interpretación de los valores y vectores propios de la matriz de correlación y covarianza, conduce a varias generalidades:

1. La localización de vectores propios a lo largo del eje principal de la hiperelipsoide, su posición coincide con la dirección de la máxima varianza. El vector propio asociado con el gran valor propio determina la dirección de máxima varianza de los datos; el vector propio asociado con el segundo gran valor propio sitúa la dirección de la máxima varianza ortogonal al primero, y así, sucesivamente.

2. La localización de vectores propios se puede ver como un problema de rotación o transformación. Los ejes originales, representan las medidas originales de las variables, y son rotados para una nueva posición.

Los renglones de la matriz de vectores propios son los coeficientes que rotan los ejes de las variables, a posiciones a lo largo de los ejes mayor y menor. Los elementos de los vectores se consideran como los cosenos de los ángulos de rotación del anterior sistema al nuevo sistema de ejes coordenados. Las columnas de la matriz de vectores propios da la dirección de los cosenos de los mismos vectores propios.

3. Los vectores propios son vectores linealmente independientes los cuales son combinaciones lineales de las variables originales. Ellos pueden ser vistos como \`nuevas´ variables con la propiedad que no están relacionados, y cuentan para la varianza de los datos en orden decreciente de importancia.

4. La suma de la proyección cuadrada de los datos, sobre los vectores propios es proporcional a la varianza a lo largo de los vectores propios. Esta varianza es igual a el valor propio asociado. La raiz cuadrada del valor propio es usada como la desviación estándar de la "nueva" variable, representada por el vector propio.

5. Un valor propio igual a cero indica que el correspondiente eje menor es de longitud cero. Esto sugiere que la dimensionalidad del espacio que contiene los datos o puntos, es menor que el espacio original.

# CAPITULO III

## EL MODELO DEL ANÁLISIS FACTORIAL

### 3.1. Terminología Básica del Análisis Factorial.

Existen varios términos analíticos factoriales, incluidos en la metodología general y específica en los propósitos que conciernen al análisis factorial, pero para entender más claramente este método, se tratan sólo cuatro términos básicos: 1) Cargas factoriales, 2) Cosenos de factores, 3) Pesos factoriales y 4) Puntajes factoriales.

Los términos mencionados anteriormente serán ejemplificados en el punto (4.4) del siguiente capítulo, donde se interpretan los cuadros o matrices factoriales.

#### 1) Cargas factoriales.

Las cargas factoriales describen la interrelación entre los factores que surgen del análisis factorial, y las variables originales usadas en la construcción de los factores. Estas cargas factoriales son independientes es decir ortogonales. -(ver la relación 2.17)- Es importante tener presente que la carga factorial es un coeficiente de correlación entre una variable y un factor.

#### 2) Cosenos de factores.

El coseno de factor representa también un coeficiente de correlación, pero en contraste a una carga factorial, relaciona un factor con otro factor. Este coeficiente es importante porque cuantifica el grado de los diferentes factores relacionados. Se espera idealmente que los factores sean relativamente no relacionados, esto es que los cosenos de factores estén cercanos a cero. Si esto sucede, cada factor puede ser una representación de un componente fundamental, de la información contenida en el conjunto original de variables.

Por otro lado, si dos factores están altamente relacionados, cada uno describiendo esencialmente el mismo componente de información, sólo uno de ellos debe ser considerado. Si dos factores son considerados sobre fundamentos empiricos o teóricos a ser necesariamente relacionados, se tolera la misma cantidad de asociación entre los dos factores.

3) Pesos factoriales.

Un peso factorial es un número asignado a una variable (usualmente en forma estandarizada), para usarlo en la determinación de puntajes para un factor. Comparando el peso factorial con la carga factorial en una matriz, se ve que las correspondientes entradas no son iguales. Sin embargo, en una columna especifica las altas cargas factoriales tienden a corresponder altos pesos factoriales, esta situación sucede generalmente en cualquier solución de análisis factorial.

Así, el peso factorial y la carga factorial dan una similar información, excepto que son medidos sobre diferentes escalas y son usados para diferente propósito, los pesos computan los puntajes factoriales y las cargas describen correlaciones. Es decir el grado de asociación entre el factor y la variable.

4) Puntajes factoriales.

El puntaje factorial es un valor especifico de un factor calculado para una particular unidad muestral, y formado como una suma de pesos de los valores de las variables originales (usualmente estandarizadas), para la unidad muestral.

Los puntajes de los factores se derivan de la siguiente manera: cada variable es ponderada proporcionalmente a su pertenencia a un concepto (o pauta)[1], cuanto más involucrada está una variable, mayor es el peso. Las variables que no están en manera alguna relacionadas con un concepto determinado, serán ponderadas con un valor cercano a cero.

Para determinar el puntaje de un caso en un concepto, los datos del caso (unidad muestral) para cada variable, serán multiplicados por el peso del concepto para esa variable.

La suma de esos productos de los datos para su ponderación para todas las variables, da el puntaje factorial. Los casos tendrán altos o bajos puntajes factoriales, según sean altos o o bajos sus valores en las variables que forman parte de un concepto o fenómeno.

Estos puntajes factoriales dan valores para cada caso, en las funciones definidas o estructuradas, de acuerdo al modelo factorial, y según el objetivo de la investigación.

1 El significado de pauta fue referido en el capitulo I, Pág. 22.
Los puntajes se derivan de la ponderación de variables, según la relación con la pauta. Por ejemplo, 'Desarrollo económico' implica ciertas pautas de características como el concepto de 'sistema político' el 'PBN per cápita' etc.

## 3.2. Definición Matemática del Modelo Factorial.

### El modelo factorial.

En la aplicación hay muchos modelos factoriales que difieren en aspectos importantes. El modelo aplicado más frecuentemente en la psicología es el denominado análisis del factor común. Los psicólogos reservan usualmente el término de ‹análisis factorial› sólo para este modelo. El análisis del factor común procura definir los conceptos o fenómenos de variación común entre un conjunto de variables. La variación que concierne a una sola variable es dejada de lado. En cambio, otro modelo factorial llamado análisis del factor componente trata de modelar todas las variaciones en un conjunto de variables, ya sean comunes o únicas.

Otros modelos factoriales son el análisis de imagen, el análisis canónico y el análisis alfa. El análisis de imagen tiene el mismo objetivo que el análisis del factor común, pero sus propiedades matemáticas son más elegantes. El análisis canónico define factores comunes a una muestra de casos que son las mejores estimaciones de la población; permite pruebas de significancia. El análisis alfa define factores comunes para una muestra de variables que son las mejores estimaciones de un universo.

Esto sugiere que de la variedad de métodos factoriales mencionados en el punto (1.1), para determinar los factores comunes, conlleva alguno de los muchos modelos factoriales existentes, fundamentados en las teorías matemáticas en la explicación de problemas científicos o sociales en particular.

### Modelo algebraico.

Un enfoque tradicional para expresar relaciones consiste en establecer la función matemática $f(x,y,w)$, que conecta una variable, $z$, con el conjunto de variables $x$, $y$, $w$. Esa función puede ser $z = 2x + 3y - 2w$ , o $z = 4xw/y$ . Las variables, tanto del primero como del segundo miembro de la ecuación, son conocidas, los datos son disponibles, y se trata sólo de determinar la mejor función para describir la relación.[2]

Digamos que se tiene cierto número de variables, $z_1$, $z_2$, $z_3$, pero que no se conocen las variables que corresponden al lado derecho de la ecuación, ni las funciones involucradas. Es decir, tenemos datos que queremos explicar matemáticamente, pero las variables que nos darian esta explicación son desconocidas o imposibles de medir. Para esto se recurre a un enfoque matemático no tradicional.

Suponemos que las variables z están relacionadas con cierto número de funciones que operan linealmente. Ver la relación (2.18). Esto es:

$$
\begin{aligned}
z_1 &= a_{11}F_1 + a_{12}F_2 + \dots + a_{1n}F_n \\
z_2 &= a_{21}F_1 + a_{22}F_2 + \dots + a_{2n}F_n \\
z_3 &= a_{31}F_1 + a_{32}F_2 + \dots + a_{3n}F_n \qquad (3.1) \\
&\vdots \\
z_n &= a_{n1}F_1 + a_{n2}F_2 + \dots + a_{nn}F_n
\end{aligned}
$$

Donde: z = una variable con datos conocidos.
a = una constante.

F = una función f( ), de algunas variables desconocidas.

2Aquí podrían utilizarse las técnicas, para ajustar curvas como el análisis de regresión múltiple lineal y curvilineal.

Es decisivo, para comprender el análisis factorial, recordar que F representa una función de variables y no una variable. Por ejemplo, las funciones pueden ser $F_1 = xw + 2y$ , y $F_2 = 3x^2y/\sqrt{w}$ . Las variables desconocidas que entran en cada función, F, de las ecuaciones (3.1) están relacionadas de maneras desconocidas, aunque las ecuaciones que relacionan las funciones son, en si mismas, lineales.[3]

Dentro de esta perspectiva algebraica, para z, ¿qué hace el análisis factorial?. Si se aplica a los datos conocidos de las variables z, el análisis factorial define las funciones F desconocidas. Las cargas que surgen de un análisis factorial son las constantes a . Los factores son las funciones F. El tamaño de cada carga para cada factor mide hasta qué punto esta función especifica está relacionada con z. Para cada una de las variables z de las ecuaciones (3.1). Las que se pueden generalizar de la siguiente forma:

$$Z_j = a_{j1}F_1 + a_{j2}F_2 + a_{j3}F_3 + \ . \ . \ . + a_{jn}F_n \qquad (3.2)$$

$$(j = 1,\ 2,\ .\ .\ .\ .\ ,\ n)$$

Este modelo se deriva del trabajo de Hotelling (1933), ver la relación (1.7). Donde las F representan los factores y las a representan las cargas. Se puede encontrar que algunas de las funciones de F son comunes a distintas variables. Son llamados factores de grupo, y el objeto del análisis factorial consiste frecuentemente en delinearlos.

Además de determinar las cargas, a, el análisis factorial generará también datos (puntajes) para cada caso (individuo, grupo o nación) de cada una de las funciones F no cubiertas. Se denominan puntajes factoriales a estos valores derivados para cada caso.

3La confunción con respecto a este punto ha dado lugar a muchas críticas infundadas al análisis factorial, como si describiera solamente relaciones lineales.

Junto con los datos sobre z y las ecuaciones (3.1), dan una relación matemática entre los datos tan útil e importante como las ecuaciones clásicas.

En el modelo de la ecuación (3.1) existe una parte de la variable z; que está influenciada por factores de determinantes compartidos llamados comunes o comunalidad, y por otra parte de determinantes únicos, no comunes a las otras variables. Así, la parte única de la variable no contribuye a las relaciones entre variables, y se llama factor único.

El modelo anterior, por tanto, puede ser expresado:

$$Z_j = a_{j1}F_1 + a_{j2}F_2 + \ldots + a_{jm}F_m + u_jY_j \quad [4] \qquad (3.3)$$

$(j = 1, 2, \ldots, n)$ , $(m < n)$.

Donde: $Z_j$ = variable expresada en forma estandarizada.
$F_i$ = factores.
$Y_j$ = factor "único" de la variable j
$a_{ji}$ = coeficiente estandarizado de regresión múltiple de la variable j sobre el factor i.

[4] La expresión (3.3) representa la regresión de $z_j$, sobre los factores $F_1$, $F_2$, . . . . , $F_m$ con el residual $u_jY_j$. La diferencia entre ésta y la regresión ordinaria es que los factores no son observados directamente.

Considerando p variables $y_1$, $y_2$, . . . . , $y_p$ sobre las cuales se tiene N observaciones. Un proceso común, para estudiar la interrelación lineal entre las variables, es considerando la regresión lineal de cada variable sobre todas las demás.

Suponemos p=4 y consideramos $y_1$. Entonces la regresión es dada por:

$$y_1 = b_{12}y_2 + b_{13}y_3 + b_{14}y_4 + u_1$$

donde $u_1$ es el residual y $b_{12}$, $b_{13}$ y $b_{14}$ son los coeficientes de regresión.

$u_j$ = coeficiente estandarizado de regresión de la variable j sobre su factor único j.

En este modelo cada una de las variables observadas es descrita linealmente en términos de m factores y un único factor. Los factores comunes cuentan para la correlación entre variables, mientras cada único factor cuenta para la varianza restante (incluyendo el error) de las variables. Los coeficientes de los cofactores son referidos a las cargas factoriales. El modelo fue empleado por Karl Pearson y Hotelling para desarrollar el Método del Fator-Principal, ver la relación (1.6). También es llamado el modelo clásico del análisis factorial.

Se hacen las siguientes suposiciones:

$$r(F_i, Y_j) = 0 \qquad i = 1, 2, \ldots, n \; ; \; j = 1, 2, \ldots, n$$

$$\text{con } i \neq j \quad y \quad j \neq k$$

$$r(Y_j, Y_k) = 0$$

Es decir, el factor único es ortogonal a todos los factores comunes y con los factores únicos asociados con otras variables. Se asume que las F's y las Y's tienen media cero y varianza uno.

**Definición matemática de un factor.**

Los factores pueden ser estimados como una combinación lineal de las variables originales. Por ejemplo, considerando una aplicación concreta del punto (1.2), sobre el factor status socio-económico (SSE) y el factor inestabilidad social (IS). El factor (SSE) es representado como:

$$SSE = W_1 IPC + W_2 APE + W_3 PD + \ldots + W_7 PFI3-$$

Estas son las posibles variables que contribuyen al factor status socio-económico. Ver ejemplo 2 del punto (1.2).

La expresión general para el i-ésimo factor $F_i$ es:

$$F_i = \sum_{j=1}^{p} W_{ij}X_j = W_{i1}X_1 + W_{i2}X_2 + \ldots + W_{ip}X_p \qquad (3.4)$$

con $i = 1, 2, \ldots, k$

Donde las W'son los pesos factoriales (son estimados desde los datos) y las X's son las variables originales (generalmente expresadas en forma estandarizada).

Se hace una distinción entre "factor" y "puntaje factorial". El término factor es referido a la expresión general para F dada anteriormente, los cuales toman un valor especifico para cada unidad muestral seleccionada. El término puntaje factorial se refiere a un valor especifico de los factores obtenidos, a partir de la expresión (3.4) para F con los valores de las variables originales, para una unidad muestral seleccionada.

Cuando las X's son estandarizadas (es decir se reduce su dimensión) de esto se deriva que el más alto de los pesos factoriales para una variable dada, contribuye más al total del puntaje factorial y a una correspondiente alta carga factorial.

El modelo factorial en notación matricial.

Modelo para variables observadas.

Se desarrolla el modelo factorial en relación a la matriz de datos, tratando el caso fijo; esto se refiere al uso del análisis factorial, solamente cuado el interés es dirigido hacia el análisis de una particular colección de datos (muestra). Cualquier factor obtenido en el análisis, sólo puede ser interpretado con repecto a los elementos o unidades en la muestra, y no hacer inferencia acerca de una población grande.

El modelo matemático es:

$$X_{(Nxp)} = F_{(Nxp)} A_{(kxp)} + E_{(Nxp)} \qquad (3.5)$$

Donde X es la matriz de datos, F la matriz de puntajes factoriales, A la matriz de cargas factoriales y E es la matriz de los residuales o término de error. Aquí k es un escalar denotando el número de factores a ser usados. (ver definición escalares punto (2.1)). Casi siempre es menor que p, el número de variables.

Para cualquier valor de la matriz de datos, declarando el n-ésimo renglón y la i-ésima columna, el modelo en notación escalar es dado por:

$$x_{ni} = \sum_{j1}^{k} f_{nj}\, a_{ij} + e_{ni} \qquad (3.6)$$

$$e_{ni} = u_j Y_j \quad \text{de (3.3)}$$

Para cualquier particular, $x'$de X se tiene:

$$x' = f'A' + e' \qquad (3.7)$$

Donde $f'$y $e'$son los vectores renglón correspondientes de F y E.

Finalmente tomando la transpuesta de (3.7) se obtiene:

$$x = Af + e \qquad (3.8)$$

[definición (2.6) del punto (2.1)].

$x$ es ahora un vector columna representando una unidad de estudio de la matriz de datos. La condición que cada variable observada (un elemento de $x$) es una suma de pesos de los factores, más un término de error o residual. El producto, $Af$, produce un vector de estimaciones de $x$, el vector $e$, representa las diferencias entre estas estimaciones y el vector observado. Estos residuales están asumiendo ser no relacionados con los factores.

El vector columna $f$ contiene los factores y los elementos de A, son los pesos para ser asignados a cada factor, para cada variable particular.

Las cargas factoriales de la matriz A, contiene los coeficientes que pueden ser usados, para combinar los factores dentro de la estimación de una variable particular. También, el vector columna de A puede ser considerado como el que contiene los coeficientes, que describe la composición del factor en términos de las variables originales.

## 3.3. Factor Patrón y Factor Estructura.

### Factor patrón.

Al ser descrita la composición de las variables en términos de factores, ahora es posible considerar en grandes lineas los objetivos del análisis factorial de datos. Para una conjunto de n variables, el modelo lineal (3.3) expresando cualquier variable $z_j$ en términos de m factores comunes y su único factor puede ser escrito en forma expandida com sigue:

$$
\begin{aligned}
z_1 &= a_{11}F_1 + a_{12}F_2 + \dots + a_{1m}F_m + u_1Y_1 \\
z_2 &= a_{21}F_1 + a_{22}F_2 + \dots + a_{2m}F_m \qquad + u_2Y_2 \qquad (3.9)\\
&\dots \\
z_n &= a_{n1}F_1 + a_{n2}F_2 + \dots + a_{nm}F_m \qquad\qquad + u_nY_m
\end{aligned}
$$

Tal conjunto de ecuaciones es llamado un *factor patrón*, o patrón. Para simplificar un patrón es presentado en forma tabular, en la cual solamente los coeficientes de los factores son listados, con la designación de factor como encabezado de las columnas. Algunas veces una tabla incluyendo sólo los coeficientes de los factores comunes puede ser referida como el factor patrón.

En el factor patrón (3.9), los factores comunes $F_p$ (p=1, 2, ..., m) pueden ser relacionados y no relacionados, pero los factores únicos $Y_j$(j=1, 2, ...,n) siempre asumen ser no relacionados entre ellos mismos y con los factores comunes. En la descripción lineal de una variable particular, el número de factores comunes involucrado es menor que m, algunos de los coeficientes son cero.

En el caso ortogonal, en el cual los factores son no relacionados, el factor patrón contiene elementos $a_{nj}$ , los cuales son la covarianza entre las variables y los factores;

$a_{nj}$ es la covarianza entre la variable $z_j$ y el factor $f_j$. Si las variables y los factores son estandarizados, entonces esta covarianza es una correlación.

Factor estructura.

En el caso oblicuo, los factores son ellos mismos interrelacionados y éstos pueden ser tomados en cuenta cuando se compute la covarianza entre factores y variables. La matriz que contiene las correlaciones entre variables y factores es llamada un *factor estructura*, o solamente estructura. En el caso oblicuo, ambos el factor patrón y el factor estructura son necesarios para una completa descripción de la solución factorial.

La funcional relación entre los elementos de una estructura y los coeficientes de un patrón son mostrados en la siguiente matriz.

Multiplicando cualquiera de las ecuaciones (3.9) por los respectivos factores, sumando sobre el número de observaciones N, y dividiendo por N, se produce:

$$
\begin{aligned}
r_{z_jF_1} &= a_{j1} + a_{j2}r_{F_1F_2} + \dots + a_{jp}r_{F_1F_p} + \dots + a_{jm}r_{F_1F_m} \\
&\dots \\
r_{z_jF_p} &= a_{j1}r_{F_pF_1} + a_{j2}r_{F_pF_2} + \dots + a_{jp} + \dots + a_{jm}r_{F_pF_m} \\
&\dots \\
r_{z_jF_m} &= a_{j1}r_{F_mF_1} + a_{j2}r_{F_mF_2} + \dots + a_{jp}r_{F_mF_p} + \dots + a_{jm}
\end{aligned}
\tag{3.10}
$$

$$r_{z_jY_j} = u_j \tag{3.11}$$

La ecuación (3.11) muestra que la correlación con los factores únicos (los elementos $r_{z_jY_j}$ de un factor estructura) son siempre idénticos con los coeficientes del único factor en el factor patrón. (3.9)

En la tabla de correlaciones de variables con factores comunes los $r_{z_j F_p}$, serán referidos como el factor estructura.

Formalmente, el sistema (3.10) es considerado como n conjuntos de m ecuaciones lineales desconociendo los coeficientes $a_{jp}$ (j=1, 2, ..., n; p=1, 2, ..., m), y los miembros del lado izquierdo son cantidades conocidas. Este es un sistema de ecuaciones posible de resolver.

Ahora definiendo la correlación con los factores comunes por:

$$s_{jp} = r_{z_j F_p} \quad , \quad (j=1, 2, \ldots, n \; ; \; p=1, 2, \ldots, m)$$

y el factor estructura es representado por:

$$S = \begin{bmatrix} s_{11} & s_{12} & \ldots & s_{1m} \\ s_{21} & s_{22} & \ldots & s_{2m} \\ \cdot & \cdot & \ldots & \cdot \\ s_{n1} & s_{n2} & \ldots & s_{nm} \end{bmatrix} \qquad (3.12)$$

Esta notación no debe confundirse con la definición (2.28) de la covarianza muestral.

La matriz de coeficientes de correlación entre los factores comunes es definida como:

$$\Phi = F\,F' \begin{bmatrix} 1 & r_{F_1F_2} & \dots & r_{F_1F_m} \\ r_{F_2F_1} & 1 & \dots & r_{F_2F_m} \\ \cdot & \cdot & \dots & \cdot \\ r_{F_mF_1} & r_{F_mF_2} & \dots & 1 \end{bmatrix} \qquad (3.13)$$

Entonces el factor estructura esta dado en forma matricial por:

$$S = A\,\Phi \qquad (3.14)$$

Partiendo del modelo (3.5)

(1) $X = A\,F$

(2) $X\,F' = A(F\,F')$

(3) $X\,F' = S$

Explicación de los pasos anteriores.

El paso 1) representa la parte común de las variables observadas $x$. ($x = z$)

La expresión 2) es la relación 1) posmultiplicada ambos lados por la transpuesta de la matriz de factores.

La relación 3) representa cómo cada elemento es un coeficiente de correlación. De acuerdo a la definición (2.31).

De aqui sustituyendo (3) y (3.13) en (2) resulta la relación (3.14), esto es: $S = A\Phi$

Esta es la fundamental relación entre el factor patrón A y un factor estructura S, la matriz estructura es igual a la matriz patrón posmultiplicada por la matriz de correlación entre factores.

Para el caso oblicuo, los elementos de A no son covarianzas; sólo puden ser interpretados como coeficientes de regresión. En el caso ortogonal, cuando $\Phi$ = 1, esto es los factores no son correlacionados. El factor patrón y el factor estructura, ambos son dados por la matriz A.

3.4. Generación de la Matriz de Correlación.

Pasos metodológicos en el análisis factorial.

El método generalmente consta de cuatro pasos.

El primer paso involucra la colocación de los datos para su entrada. En este punto se tienen descritos los datos como un conjunto de valores de las variables originales $x_1, x_2, \ldots, x_p$, para cada una de las unidades de estudio (en el estudio sobre "Indices de Marginación" a desarrollarse en el capitulo V, p=19 variables socio-económicas y las unidades de estudio son las 32 entidades federativas de la República Mexicana).

Actualmente, considerar los datos en esta forma es suficiente, pero el método del análisis factorial solamente requiere de la matriz de correlación entre las variables, por tanto el paso uno del método es la preparación de la matriz de correlación.

El segundo paso consiste en usar la matriz de correlación, para determinar un conjunto de factores iniciales. Esto es usualmente ejecutado por el método de componentes principales.

Un tercer paso, comprende la rotación de los factores iniciales que se requiere, para conseguir un significado conceptual.

Después de ejecutar la rotación, se obtendrá un conjunto de pesos factoriales para usarlos en la determinación de los puntajes factoriales, éstos son obtenidos sobre las unidades de estudio para cada factor derivado. Los cuales pueden ser usados como variables muevamente construidas en otro análisis. Todo esto se realiza en el cuarto paso de la metodologia.

Preparación de la matriz de correlación.

El análisis factorial trabaja sobre la matriz de correlación R de las variables originales, para obtener los factores deseados. Entonces la relación entre las variables, las cuales se busca brevemente describir, son representadas por la información en la matriz de correlación.

La matriz de correlación R tiene la forma:

$$R = \begin{array}{c c} & \begin{array}{cccc} x_1 & x_2 & & x_p \end{array} \\ \begin{array}{c} x_1 \\ x_2 \\ \\ x_p \end{array} & \begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{12} & 1 & \cdots & r_{2p} \\ \cdot & \cdot & \cdot & \cdot \\ r_{1p} & r_{2p} & \cdots & 1 \end{bmatrix} \end{array} \qquad (3.15)$$

Una matriz de correlación con datos estandarizados, es representada por la fórmula:

$$R = Z\,Z' / N$$

La ecuación matricial R W = L describe como la matriz R, está relacionada en el análisis factorial.

Donde W representa una matriz de pesos factoriales y L es una matriz de cargas factoriales. Esta ecuación deja ver claramente que el método analítico factorial, determina una matriz de pesos W que es aplicada a la matriz de correlación R, para obtener la matriz L de cargas factoriales deseable. (Esto significa que la matriz L refleje parsimonia, independencia aproximada, y significado conceptual). La parsimonia es la representación de la información contenida en las diversas variables originales en términos de un pequeño número de factores, y la independencia aproximada es realizada por la construcción de factores de tal forma que sean estadísticamente independientes.

La matriz L puede ser el resultado del proceso de dos etapas. La primera etapa (es equivalente, al segundo paso de la metodología del análisis factorial) involucra la determinación de una matriz $W_1$ de pesos iniciales y una matriz $L_1$ de cargas factoriales iniciales, tal que R $W_1 = L_1$. La segunda etapa (o tercer paso) comprende el uso de una rotación, a través de la matriz de transformación T (ver la relación 2.20), para generar la matriz de cargas factoriales deseada L ( = $L_1$ T) y la correspondiente matriz de pesos W (= $W_1$ T) .

De hecho solamente los elementos fuera de la diagonal, que son las relaciones entre las variables están reflejadas en la matriz R. Los elementos de la diagonal principal (los cuales son 1's en la matriz R) son superfluos, para los propósitos del análisis factorial. Esto sugiere una alternativa u opción concerniente a si remplazan los 1's o no, en la diagonal, por algún otro número (llamado comunalidad) y entonces del análisis factorial resulta una matriz de correlación "adjunta", el propósito de ésto es encontrar más parsimonia y más significado conceptual en la solución analítica factorial.

Características de la matriz de correlación.

1) La mayoría de los análisis factoriales empiezan con un cuadro de resultados. Con éste se computa y organiza un cuadro de correlaciones de manera que se conozca la *r* de Pearson entre todos los posibles pares de medidas.

Si al relacionar las variables o medidas resulta tener un alto nivel de interrelación, implica que tienen todas algo en común.

Los coeficientes de correlación, expresan el grado de relación lineal entre las variables de las filas y las columnas de la matriz. Cuanto más se acerca el coeficiente a cero, menor es la relación; cuanto más se acerca a uno, mayor es la relación. Un signo negativo indica que las variables están relacionadas inversamente. Ver relación (2.31).

Un análisis de factores puede llevarse a cabo con la misma facilidad, sobre correlaciones no válidas y sin contenido, que sobre datos perfectamente válidos, y no hay nada en el procedimiento matemático que indique esto, con la posible excepción de cierta dificultad para interpretar los resultados.

2) Para interpretar el coeficiente hay que elevarlo al cuadrado y multiplicarlo por 100. Esto dará el porcentaje de variación en común, para los datos en las variables.

3) El coeficiente de correlación entre dos variables es el coseno del ángulo entre las variables, como vectores establecidos en el sistema de coodenadas correspondiente a cada caso. Ver Fig. 2.1 (interpretación geométrica de la correlación).

4) La diagonal principal contiene habitualmente la correlación de una variable consigo misma y su valor es 1.0. Pero muchas veces, cuando la matriz de correlaciones es sometida al análisis factorial ( utilizando el modelo del análisis de factor común), la diagonal principal contendrá, por el contrario, estimaciones de comunalidad. Estas miden la variación de una variable en común con todas las demás juntas.

Muchas de la hipótesis sociales involucran relaciones entre dos variables, estas relaciones empiricas pueden ser descritas en una martiz de correlación.

### 3.5. Concepto de Comunalidad.

Discusión teórica de la comunalidad.

La comunalidad se incorpora a partir de la estructura de las variables originales $x_1, x_2, \ldots, x_p$. Considerando la definición de un factor en términos de estas variables, según la expresión (3.4).

$$F_i = \sum_{j=1}^{p} W_{ij} X_j = W_{i1}X_1 + W_{i2}X_2 + \ldots + W_{ip}X_p$$

Donde $F_i$ y $W_{ij}$ denota el i-ésimo factor y el peso factorial asociado con la j=ésima variable en el i-ésimo factor, repectivamente.

Ahora se desea considerar la estructura de tal forma que la variable sea descrita por los factores. En general el modelo del análisis factorial describe $X_j$ (en forma estandarizada) como:

$$X_j = (\lambda_{1j}F_1 + \lambda_{2j}F_2 + \ldots + \lambda_{kj}F_k) + U_j = C_j + U_j \qquad (3.16)$$

Donde $F_1, F_2, \ldots, F_k$ son los k factores (presentados en común en la expresión para todas las X's); $\lambda_{1j}$, $\lambda_{2j}$, . . . , $\lambda_{kj}$ son las cargas factoriales; y $U_j$ es un componente único aleatorio para $X_j$ y son (estadisticamente independiente) de las F's. En otras palabras cualquier variable es representada como la suma de una combinación lineal de las F's (factores) es decir $C_j$, y una cantidad aleatoria esto es $U_j$, única para cada variable.

Además la varianza de $X_j$ (var($x_j$)=1 , entonces $X_j$ esta en forma estandarizada) es escrita como la suma de dos varianzas.

$$\mathrm{Var}(X_j) = 1 = \sigma^2_{C_j} + \sigma^2_{U_j} \tag{3.17}$$

Donde $\sigma^2_{C_j}$ es la varianza de la combinación lineal $C_j$ de los factores comunes y $\sigma^2_{U_j}$ es la varianza del único componente $U_j$ .

La varianza $\sigma^2_{C_j}$ es llamada la comunalidad de las variables $X_j$, entonces es la cantidad medida de información (en términos de varianza) que la variable tiene en común (a través de los factores comunes) con todas las otra variables. Si los factores no son correlacionados, la comunalidad de la variable $X_j$ es la suma de los elementos cuadrados en el i-ésimo renglón de la matriz A de cargas factoriales. Ver relación (3.5).

Remplazando los 1's sobre la diagonal de la matriz de correlación R con algún otro número, puede ser mostrada una equivalencia teorica usando estimaciones de las $\sigma^2_{C_j}$. Asi, al usar comunalidad se pone atención a la parte común, esto es las C's de las variables y se ignoran los componentes únicos. Por otro lado al conservar los 1's en la diagonal principal se enfoca la atención sobre las variables originales, y se asume que no todas las variables tienen componente único.

La selección de si o no usar la comunalidad, depende sobre si el investigador desea considerar solamente la parte común de las variables, o si desea trabajar directamente con las variables originales.

Métodos para estimar la comunalidad.

Uno de los métodos más frecuentemente usado, para estimar la comunalidad es el coeficiente de correlación múltiple al cuadrado (CMC) una variable con todas las demás. (n-1 variables observadas).

El CMC es llamado la "comunalidad observada", mide la varianza común entre las correlaciones observadas. Estos valores son determinados únicamente, dentro de iteracciones. Para obtener el CMC en un conjunto de variables, es necesario calcular la inversa $R^{-1}$ de la matriz de correlación R (comunalidades en la diagonal), y entonces el CMC para la variable $X_j$ es dado por:

$$CMC_j = R^2_{(j=1,2,\ldots,n)} = 1 - \frac{1}{r^{jj}} \tag{3.18}$$

Donde $r^{jj}$ es el elemento diagonal en $R^{-1}$ correspondiente a la variable $X_j$. Estos valores son frecuentemente calculados como un paso inicial en el programa, para la solución del componente principal.

Una propiedad importante del CMC es:

$$R^2_{j(n-1)} \leq h^2_j \tag{3.19}$$

$h^2_j$ es una estimación de comunalidad.

Esto quiere decir que la comunalidad para una variable dada estará entre $R^2$ y 1 . Entonces la comunalidad es menor que uno.

La inserción de los CMC en la diagonal principal produce una matriz de correlación reducida.

Cualquier CMC mide la varianza común de la variable particular, y el resto (n-1) variables seleccionadas para el estudio.

El CMC multiplicado por 100 mide el porcentaje de variación que se puede producir (predecir, generar o explicar), para una variable teniendo en cuenta a todas las demás.

## 3.6. Método de Componentes Principales.

El método de componentes-principales, o análisis de componentes, se fundamenta sobre los primeros trabajos de Pearson (1901) con adaptaciones especificas del análisis factorial sugeridas por el trabajo de Hotelling (1933).

Mientras que el método de componentes-principales es proyectado desde el modelo (3.2), Thomson (1934) fue el primero en aplicar el modelo clásico del análisis factorial, en algún tiempo la aplicación sólo fue para la solución de Dos-factores de Spearman.

Generalmente el "método de factores principales" o ejes principales es indicado en la aplicación del método de componentes principales, para reducir la matriz de correlación R, esto es, con comunalidades en lugar de unos en la diagonal principal de R.

En el análisis de componentes toda la varianza de las variables es analizada en términos de los componentes principales, empleando el modelo (3.2). Mientras que en el modelo del análisis factorial (3.3) la comunalidad es analizada en términos de factores comunes.

Por otro lado el análisis de la matriz de correlación, con unos en la diagonal principal, lleva a obtener componentes principales. Mientras el análisis de la matriz de correlación con comunalidades en la diagonal resultan factores principales. En consecuencia podria decirse que los términos "componentes principales" y "factores principales" son similares, excepto por la distinción hecha anteriormente.

### Significado del método de componentes principales.

El conjunto de variables empleado en el análisis, son representadas por puntos en un lugar geométrico a través de la densidad de frecuencia uniforme, esencialmente concéntricos, similares y colocados en forma de elipsoides.

Los ejes de estos elipsoides corresponden a los componentes principales. El Método del análisis de componentes, entonces, involucra la rotación de ejes coordenados, para una nueva construcción de ejes de referencia en el espacio total de las variables, a través de una transformación ortogonal, donde cada una de las n variables originales es descrita en términos de los n nuevos componentes principales.

Una importante caracteristica de los nuevos componentes es que ellos cuentan, en turno, para una cantidad máxima de varianza de las variables. Más especificamente, el primer componente principal es esa combinación lineal de las variables originales las cuales contribuyen al máximo de su varianza total; el segundo componente principal, no está relacionado con el primero, y contribuye al máximo en la varianza residual; y así sucesivamente hasta que el total de la varianza es analizada. La suma de las varianza de todos los n componentes principales, es iguala la suma de las varianzas de las variables originales.

Entonces el método depende de la varianza total de las variables originales, es más conveniente cuando todas las variables son medidas en la misma unidad. De otra manera, por cambio de unidad o transformación lineal de las variables, las elipsoides serán comprimidas o extendidas, y sus ejes (los componentes principales) no tendrán significado especial. Por lo tanto, es usual expresar las variables en forma estándar, esto es seleccionar la unidad de medida para cada variable, de tal forma que la varianza muestral sea uno. Entonces el análisis es hecho sobre la matriz de correlación, con el total de la varianza igual a n.

Determinación de factores iniciales por el análisis de componentes principales.

Este es el paso número dos de la metodología a seguir en el análisis factorial (ver punto 3.4), el cual consiste en usar la matriz de correlación, para determinar un conjunto de factores iniciales.

Los factores iniciales son usualmente determinados por el método de componentes-principales. Este método es de suma importancia para los objetivos del presente trabajo.

El término de "componentes principales" es el nombre que se da al conjunto de factores, como resultado de la aplicación de este método.

La variación total en $x_1, x_2, \ldots, x_p$. $\quad n = p$

Para describir el método de componentes principales, primero es necesario introducir el concepto de la variación total en los datos con relación a las variables $x_1, x_2, \ldots, x_p$. La variación total es matemáticamente definida como la suma de la varianza muestral de las k variables. Como es usualmente, las variables están en forma estandarizada (así que $S_j^2=1$ para cada j), la variación total es simplemente igual a p, el número de variables.

$$\text{Variación total} = S_1^2 + S_2^2 + \ldots + S_p^2 \tag{3.20}$$

Donde $S_j^2$ es la varianza muestral de $x_j$, $j=1, 2, \ldots, p$.

Conceptualmente la variación total es una medida de la cantidad de "incertidumbre" asociada con las observaciones sobre todas las p variables. La incertidumbre se refiere a cómo muchas observaciones sobre la unidad de estudio difiere una de otra. Por ejemplo, si todas las observaciones sobre una variable dada son exactamente las mismas, no habría incertidumbre acerca del valor esperado para la variable, pero si las observaciones son completamente diferentes una de otra, habría incertidumbre acerca del valor esperado.

Definición de componentes-principales.

El propósito del análisis de componentes-principales es determinar factores (esto es, componentes-principales) en tal forma que expliquen mucha de la variación total en los datos, con pocos de estos factores o componentes como sea posible.

Partiendo del modelo clásico del análisis factorial (3.3), para determinar el coeficiente del factor común puede ser escrito en la forma:

$$z_j = a_{j1}F_1 + \ldots + a_{jp}\,F_p + \ldots + a_{jm}F_m \qquad (3.21)$$

$$(j = 1,\ 2,\ \ldots,\ n)$$

Donde el factor único es omitido, por simplicidad.

La suma del cuadrado de los coeficientes de los factores da la comunalidad de una variable en particular, mientras cualquier término $a_{jp}^2$, indica la contibución del factor $F_p$ a la comunalidad de $z_j$. La primera etapa del método factores-principales involucra la selección del coeficiente del primer factor, $a_{j1}$, de manera de hacer la suma de la contribución de ese factor a la máxima comunalidad total. Esta suma es dada por:

$$V_1 = a_{11}^2 + a_{21}^2 + \ldots + a_{n1}^2\ , \qquad (3.22)$$

y el coeficiente $a_{j1}$ puede ser seleccionado de tal manera que

$V_1$ sea la máxima bajo la condición:

$$r_{jk} = \sum_{p=1}^{m} a_{jp}\, a_{kp} \qquad (j, k = 1, 2, \ldots, n) \; , \qquad (3.23)$$

donde $r_{jk} = r_{kj}$ y $r_{jj}$ es la comunalidad $h_j^2$ de la variable $z_j$ .

El método de multiplicadores de Lagrange es empleado, para maximizar $V_1$, la cual es una función de n variables $a_{j1}$ bajo las 1/2n(n+1) condiciones (3.23) entre todos los coeficientes $a_{jp}$ .

Desarrollando la relación (3.22) y la condición (3.23) resulta la siguiente relación:

$$\sum_{=1}^{n} r_{jk}\, a_{k1} - \lambda_1\, a_{j1} = 0 \qquad (3.24)$$

La expresión (3.24) es un sistema de n ecuaciones, una para cada valor de j , y tiene la forma:

$$\begin{array}{lllllll}
(h_1^2 - \lambda)a_{11} + & r_{12}a_{21} + & r_{13}a_{31} + & \ldots + & r_{1n}a_{n1} & = 0 \\
r_{21}a_{11} + & (h_2^2 - \lambda)a_{21} + & r_{23}a_{31} + & \ldots + & r_{2n}a_{n1} & = 0 \\
r_{31}a_{11} + & r_{32}a_{21} + & (h_3^2 - \lambda)a_{31} + & \ldots + & r_{3n}a_{n1} & = 0 \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
r_{n1}a_{11} + & r_{n2}a_{21} + & r_{n3}a_{31} + & \ldots + & (h_n^2 - \lambda)a_{n1} & = 0 \, ,
\end{array} \qquad (3.25)$$

Donde el parámetro de (3.24) es designado por $\lambda$ .

Una condición suficiente y necesaria, para que el sistema (3.25) de n ecuaciones homogéneas tenga una solución no trivial, es que el determinante de los coeficientes de $a_{j1}$ sea igual a cero, esto es:

$$\begin{vmatrix} (h_1^2 - \lambda) & r_{12} & r_{13} & \ldots r_{1n} \\ r_{21} & (h_2^2 - \lambda) & r_{23} & \ldots r_{2n} \\ r_{31} & r_{32} & (h_3^2 - \lambda) & \ldots r_{3n} \\ \cdots & \cdots & \cdots & \cdots \\ r_{n1} & r_{n2} & r_{n3} & \ldots (h_n^2 - \lambda) \end{vmatrix} = 0 \qquad (3.26)$$

La ecuación (3.26) es conocida como la ecuación caracteristica, ver la expresión (2.35).

Para el análisis factorial, una de las propiedades importantes de las ecuaciones caracteristicas incluye el hecho de que todas las raices son reales, y las q-raices múltiples dobles sustituidas por λ en (3.26) reduce el rango del determinante a (n-σ).

Cuando las raices simples de la ecuación caracteristica es sustituida por λ en (3.25), un conjunto de ecuaciones lineales homogéneas de rango (n-1) es obtenido.

De todo este análisis resulta que el factor de proporcionalidad es $\lambda_1 = \sum_{j=1}^{n} a_{j1}^2$ , esta expresión es precisamente

$V_1$, la cantidad a ser maximizada. En otras palabras, $V_1$ es igual a una de las raices de la ecuación caracteristica (3.26), es decir, la raiz más grande $\lambda_1$. Asi se encuentra el coeficiente $a_{j1}$ del primer factor $F_1$, el cual contará para la comunalidad total lo mas posible. La más grande raiz $\lambda_1$ de (3.26) es sustituida en

(3.25), y cualquier solución $\alpha_{11}$, $\alpha_{21}$, . . . , $\alpha_{n1}$ es obtenida. Entonces para satisfacer la relación (3.22), estos valores son divididos por la raiz cuadrada de la suma de sus cuadrados y después multiplicado por $\sqrt{\lambda_1}$.

La cantidad resultante es:

$$a_{j1} = \alpha_{j1} \sqrt{\lambda_1} \Big/ \sqrt{(\alpha_{11}^2 + \alpha_{21}^2 + \ . \ . \ . + \alpha_{n1}^2)} \qquad (3.27)$$

(j=1, 2, . . . , n)

Los cuales son los coeficientes deseados de $F_1$ en el factor patrón (3.21), esta expresión escrita en forma expandida representa el factor patrón, ver la estructura (3.9).

Las raices (λ's) de la ecuación caracteristica (3.26) son llamadas raices caracteristicas (algunas veces referidas como valores propios), o simplemente raices de la matriz R. La solución para el conjunto de ecuaciones (3.25) para una raiz dada lleva a un vector (un conjunto de α's) la cual es definida como un vector caracteristico (algunas veces llamado vector propio) de la matriz R.

El problema matemático generalizado es expresado en la forma:

Encontrando un número λ y un n-dimensional vector $q \neq 0$ tal que:

$$Rq = \lambda q \qquad (3.28)$$

Cualquier número $\lambda_p$ que satisface esta ecuación es una raiz (o valor propio) de R y está asociado al vector $q_p = \{\alpha_{1p}, \alpha_{2p}, \ . \ . \ . \ , \alpha_{np}\}$ es un vector caracteristico (o vector propio) de R.

Para cada raiz $\lambda_p$ el sistema de ecuaciones (3.25) tiene una solución diferente de cero $q_p = \{\alpha_{1p}, \alpha_{2p}, \ldots, \alpha_{np}\}$. Además, las n raíces $\lambda_1, \lambda_2, \ldots, \lambda_n$ llevan a n vectores $q_1, q_2, \ldots, q_n$, así que (3.28) puede ser escrito:

$$R\ (q_1, q_2, \ldots, q_n) = (\lambda_1 q_1, \lambda_2 q_2, \ldots, \lambda_n q_n) \qquad (3.29)$$

o, sobre la constitución de una matriz Q de n vectores,

$$R\,Q = Q\,\Lambda \ , \text{ donde } \Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n). \qquad (3.30)$$

Cuando el análisis en términos de componentes principales (con unidades en lugar de comunalidades en la diagonal principal de R), entonces los vectores en Q son linealmente independientes, así que su determinante es diferente de cero y Q tiene una inversa. Entonces desde (3.30) resulta que:

$$Q^{-1} R\,Q = \Lambda \qquad (3.31)$$

Donde R lleva dentro de la diagonal principal los elementos de $\Lambda$ los cuales son las raíces y las columnas de Q son los vectores característicos.

El siguiente problema es encontrar un factor que contará, para el máximo de la comunalidad restante.

La conveniente notación para la correlación restante de $r_{jk}$ con s factores removidos es ${}_s r_{jk}$. Así, después de obtener el primer factor, el primer factor restante o residual viene dado por:

$${}_1 r_{jk} = r_{jk} - a_{j1}a_{k1} = a_{j2}a_{k2} + a_{j3}a_{k3} + \ldots + a_{jm}a_{km}. \qquad (3.32)$$

Más generalmente, la matriz del primer factor residual es expresado por:

$$R_1 = R - \hat{R}_1 \ ,$$

Donde $\hat{R}_1 = a_1 a'_1$ , representa la matriz simétrica nxn de productos de los coeficientes del primer factor, esto es, la correlación reproducida desde el primer factor.

En la determinación del coeficiente del segundo factor $F_2$, es necesario maximizar la cantidad:

$$V_2 = a_{12}^2 + a_{22}^2 + \ . \ . \ . + a_{n2}^2 \qquad (3.33)$$

La cual es la suma de la contribución de $F_2$ , para la comunalidad restante. Esta maximización es sujeta a la condición (3.32), que es análoga a la restricción (3.23) en el caso del primer factor. La teoría de la ecuación característica provee las bases para determinar el coeficiente del segundo factor y los factores subsecuentes. En otras palabras se requiere la máxima raíz de $R_1$, es de hecho, la segunda más grande raíz de la correlación original de la matriz R.

En la práctica todos los factores son computados simultáneamente.

Otra presentación conceptual en la definición del componente principal.

El componente principal, CP(1), es aquel peso de la combinación lineal de las variables, que cuentan para la más grande cantidad de la variación total en los datos. Esto es, CP(1) es la combinación lineal de las X's , definido como:

$$CP(1) = W_{(1)1}X_1 + W_{(1)2}X_2 + \ . \ . \ . + W_{(1)p}X_p$$

Donde los pesos $W_{(1)1}$, $W_{(1)2}$, ..., $W_{(1)p}$ tienen que ser seleccionados como la máxima cantidad de:

$$\frac{\text{Varianza de CP(1)}}{\text{Varianza total}}$$

Es decir, las otra combinaciones lineales de las X's no tienen la más grande varianza como CP(1).

Cuando las X's están en forma estandarizada (el análisis está basado sobre la matriz de correlación), la proporción de la variación total en los datos explicada por CP(1) es:

$$\frac{\text{Varianza de CP(1)}}{p}$$

También, los pesos son seleccionados sujetos a la restricción

$\sum_{j=1}^{p} W_{(1)j}^2 = 1$, la varianza de CP(1) no debe exceder la varianza rotal.

El segundo componente principal, CP(2), es aquel peso de la combinación lineal de las variables, el cual no está correlacionado con CP(1), y cuenta para la máxima cantidad de la variación total restante, no explicada anteriormente por el CP(1). Es decir:

$$CP(2) = W_{(2)1}X_1 + W_{(2)2}X_2 + . . . + W_{(2)p}X_p$$

Es la combinación lineal de las X's, que tiene la varianza más grande de todas las combinaciones lineales no correlacionadas con CP(1).

En general, el i-ésimo componente principal CP(i) es la combinación lineal de la forma:

$$CP(i) = W_{(i)1}X_1 + W_{(i)2}X_2 + \ldots + W_{(i)p}X_p \tag{3.34}$$

Que tiene la varianza más grande de todas las combinaciones lineales y además no están correlacionados con todos los previamente determinados i-1 componentes principales. Actualmente, es posible determinar muchos componentes principales como existen variables originales. Sin embargo, en la aplicación práctica, la variación total en los datos es usualmente explicada por los primeros componentes. Además, estos componentes son seleccionados para estar mutuamente no correlacionados.

Interpretación de los componentes principales.

A menudo es dificil interpretar los componentes principales directamente, y mayormente como resultado de la manipulación de éstos (via rotación).

En cualquier caso, es frecuente establecer que el primer componente principal CP(1) o $F_1$, representa una completa medida de la información contenida en todas las variables; el segundo componente principal CP(2) o $F_2$, frecuentemente puede ser interpretado como una diferencia entre un particular subconjunto de variables y el resto del subconjunto.

El indice general usualmente tiene la más grande carga factorial (en valor absoluto) sobre casi todas las variables. Consecuentemente, es dificil asegurar una etiqueta especifica sobre tal factor como la estimación más relacionada, para la interpretación de un subconjunto particular de variables que para otro.

Sin embargo, la trayectoria primaria del análisis factorial es la reducción de datos, por la construcción de indices, con no mayor énfasis sobre la interpretabilidad, la solución factorial final es verosimil o probable a la consistencia de los mismos componentes principales.

Por otro lado, si se desea encontrar el significado comprendido en los factores, que describen la variación en un conjunto de variables, entonces otro paso que involucra la técnica de la rotación, es generalmente requerido para lograr este objetivo.

### 3.7. Métodos de Rotación para Factores Iniciales.

La rotación de los factores iniciales representa el tercer paso de la metodologia en el análisis factorial.

#### Carácter de los factores rotados.

Un cientifico puede rotar factores para ver si existe un hipotético grupo de relaciones. Esto se puede hacer postulando las cargas de una hipotética matriz factorial y después rotando los factores, para ajustarlos mejor con esta matriz. La veracidad de la hipótesis se prueba mediante la diferencia entre las cargas factoriales y las hipotéticas.

Por otra parte, el cientifico puede rotar los factores para controlar ciertas influencias en los resultados. Puede rotar el primer factor para una variable o grupo de variables, y después rotar los factores siguientes hasta que queden en ángulos rectos (no relacionados con el primero). Esto quita los efectos de las variables altamente cargadas del primer factor y permite apreciar los conceptos o fenómenos independientes de ellos.

Lo más frecuente es que el cientifico rote sus factores hasta obtener una solución de *estructura simple*. Cuando una matriz factorial es llamada "de factores rotados", esto significa casi siempre una rotación de estructura simple.

#### Definición de estructura simple.

La estructura simple se define de la siguiente forma: si cada una de las variables originales están altamente relacionadas con sólo un factor, y además cada factor es identificado como la representación común de un pequeño grupo de variables. En otras palabras, la estructura simple se logra cuando las cargas factoriales para más variables son aproximadas a cero, para cada factor y el resto de las cargas factoriales son relativamente grandes. Así, un factor es concebido como el que describe la proporción de la variación común, para el subconjunto de variables con una alta relación a éste, y no describe la variación con otras variables.

Mediante la rotación la interpretación de los factores iniciales cambia de los factores no rotados, los cuales para definir un fenómeno involucran todas o casi todas las variables, y las variables pueden tener cargas moderadas o altas, para muchos conceptos. Resulta que el primer factor no rotado puede estar ubicado entre grupos independientes de variables interrelacionadas. Estos grupos no pueden ser distinguidos por sus cargas en el primer factor, aunque tengan cargas de signo diferente en los factores segundo y ulteriores. También determina la clasificación más comprensiva, la red de los lazos más amplia o el mayor orden en los datos.

Las dos mejores formas para describir como la rotación procura lograr la estructura simple son: 1) geométricamente, por rotación de los ejes coordenados; y 2) numéricamente, improvisando la estructura de las cargas factoriales.

Características de la estructura simple.

1) Cada variable está identificada con un factor o una pequeña parte de ellos. Si los factores son considerados como explicaciones, causas o influencias subyacentes, esto equivale a minimizar el número de agentes o condiciones que se necesitan para dar cuenta de la variación de los distintos grupos de variables.

2) El número de variables con altas cargas en el factor es minimizado. Esto hace que los conceptos de los factores no rotados dejen de ser generales al mayor número de variables, y se conviertan en conceptos y fenómenos que involucren grupos separados de variables. La rotación intenta definir un pequeño número de grupos distintos de fenómenos interrelacionados.

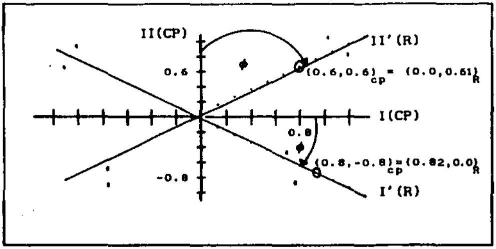
3) Una suposición ontológica importante que subyace al uso de la estructura simple es que, siempre que sea posible, el modelo de la realidad debe ser simplificado. Si los fenómenos pueden ser descritos de la misma manera utilizando factores más simples, entonces el principio de simplificación indica que se haga así.La estructura simple maximiza la sencillez, yendo de factores generales, que involucran todas las variables, a factores de grupo, que involucran diferentes conjuntos de variables.

4) Un objetivo de la investigación consiste en generalizar los resultados factoriales. La solución de factores no rotados, sin embargo, depende de todas las variables. Añadir o quitar una variable del estudio modifica los resultados. Por lo tanto, la solución no rotada debería ajustarse de manera que los factores sean invariables de las variables seleccionadas. Una solución factorial invariable delineará los mismos grupos de relaciones independientemente de las variables extrañas incluidas en el análisis.

Una de las principales justificaciones para la rotación de estructura simple, es que determina factores invariables. Esto permite una comparación de los resultados factoriales de diferentes estudios.

Ilustración geométrica de la rotación.

Fig. 3.1. Ilustración gráfica de la rotación.




El objetivo de rotar los ejes consiste primero; en hacer que cada punto se aproxime a sólo uno de los dos ejes, o sea lograr una (estructura simple) y segundo; estos puntos aproximados a los mismos ejes rotados definen un factor significativo.

La Fig. 3.1 ilustra geométricamente como la rotación puede ayudar a lograr una estructura simple. La figura describe esencialmente el resultado ideal de la rotación.

En la figura hay tantos puntos como hay variables, y cada punto corresponde a una particular variable. Las coordenadas asociadas con cada punto (o variable) son las dos cargas factoriales sobre esa variable, para los dos factores seleccionados a ser rotados. Los valores de estas coordenadas, están basados sobre los dos ejes usados para definir la escala de medidas, los cuales representan los dos factores considerados en la rotación. Si estos ejes son rotados, entonces se tendrán definidos dos nuevos factores rotados.

Entonces, los factores iniciales determinados son los componentes principales, la prerotación de ejes representa dos componentes principales. Estos son etiquetados en la Fig. 3.1 por I(CP) y II(CP), para el primer y segundo componente respectivamente.

El segundo conjunto de coordenadas dadas para cada uno de estos grupos de puntos son relativos a un nuevo par de ejes, etiquetados por I'(R) y II'(R). Estos dos ejes determinados por la rotación de los ejes originales según, las manecillas del reloj. (a través de un ángulo $\phi$)

La rotación tiene que realizar lo siguiente, cada uno de estos grupos de puntos señalados en la Fig. 3.1, son ahora aproximados a sólo uno de los dos nuevos ejes rotados. De hecho las coordenadas de los dos grupos de puntos son ahora cambiados como sigue:

$$(0.6,0.6)_{cp} \longrightarrow (0.0,0.61)_R$$

$$y\ (0.8,-0.8)_{cp} \longrightarrow (0.82,0.0)_R$$

Similarmente, las coordenadas de los otros puntos tienden a ser altos para una coordenada y cercanos a cero para la otra. Así, es lograda la estructura simple.

La importacia de estos resultados en término conceptual viene del hecho de que los puntos (o variables), ahora son vistos para ser agrupados dentro de dos subgrupos, un subgrupo situado cerca de un eje rotado y el otro subgrupo situado cerca del otro eje rotado. Entonces estos nuevos ejes representan nuevos factores rotados, ahora ya se tiene la capacidad para interpretar cada nuevo factor en términos de un particular subgrupo de variables situado cerca del factor.

Métodos de rotación.

En la ejecución de una rotación, hay dos formas en la cual los ejes pueden ser rotados. Primero, los ejes pueden mantenerse en la misma orientación, para uno y otro durante la rotación, así cuando los ejes quedan inmóviles en forma perpendicular después de la rotación. (Esto es, hay un ángulo de $90^{\circ}$ entre los dos nuevos ejes); a esta se le llama rotación ortogonal. Segundo, cada eje puede ser rotado independientemente, así que los ejes no son necesariamente perpendiculares, para uno y otro después de la rotación; esta forma es llamada rotación oblicua.

La fase de rotación del análisis factorial intenta transformar la matriz de factores iniciales, para facilitar la interpretación de los factores.

El objetivo de la rotación es transformar las matrices complicadas en una simple matriz.

Rotación de estructura simple ortogonal.

Un tipo importante de rotación de estructura simple es la estructura simple ortogonal. Los factores rotados en una estructura simple ortogonal son llamados, por lo general, simplemente ‹factores ortogonales›.

La ortogonalidad es una restricción colocada en la búsqueda de la estructura simple, para las agrupaciones de variables interdependientes. El conjunto total de factores rotado como un marco rígido, donde cada factor está fijado de manera inamovible al origen en ángulo recto (ortogonal) respecto a cualquier otro factor. Este sistema de factores es rotado alrededor del origen hasta que el sistema quede alineado en forma máxima con los grupos separados de variables.

Si todos los grupos resultan no correlacionados con los demás, cada factor ortogonal estará alineado con un grupo distinto. Sin embargo, cuanto más correlacionados están los grupos separados, menos claramente puede discriminarlos la rotación ortogonal. La estructura simple puede ser, entonces sólo aproximativa, no perfecta.

Existan o no agrupamientos correlacionados en los datos, la rotación ortogonal definirá los fenómenos no correlacionados de relaciones. Estos fenómenos pueden no superponerse completamente con los distintos grupos, pero la delineación de esos factores no correlacionados es útil. Los resultados que involucran fenómenos no correlacionados son más fáciles de comunicar, y las cargas pueden ser interpretadas como correlaciones. Además, los factores ortogonales son más manejables para el análisis y la manipulación matemática posterior.

De la rotación ortogonal resultan factores no correlacionados. Si bien esta es una atractiva propiedad, pero hay situaciones en las cuales se permite la correlación entre factores, resultando la matriz del factor patrón simple.

La matriz de correlación de factores en el caso de la rotación ortogonal es una matriz identidad.

Algoritmos computados para la rotación ortogonal.

Existe una variedad de algoritmos que se usan en la rotación ortogonal, para una estructura simple. Tres algoritmos de los más conocidos están disponibles como opciones en programas computados, para el análisis factorial, en el paquete SPSS/PC (Paquete Estadístico para las Ciencias Sociales, para computadora personal) y quizá otro paquete estadístico. Estos métodos son: varimax, quartimax y equamax. El varimax es el más frecuentemente usado de estos métodos.

El método varimax intenta lograr una estructura simple con respecto a las columnas de la matriz de cargas factoriales, esto es intenta minimizar el número de variables que tienen altas cargas sobre un factor. Esto mejorará la interpretabilidad de los factores.

El método quartimax procura lograr una estructura simple con respecto a los renglones de la matriz de cargas factoriales, quiere decir, enfatiza la interpretación simple de variables, entonces la solución minimiza el número de factores necesarios para explicar una variable. De una rotación quartimax frecuentemente resultan en general factores con altas y moderadas cargas sobre más variables.

El método equamax intenta conseguir la estructura simple con respecto a ambos, los renglones y las columnas de la matriz de cargas factoriales, es decir una combinación del método varimax, el cual simplifica los factores, y el método quartimax, el cual simplifica las variables.

Rotación de estructura simple oblicua.

Mientras que en la rotación de estructura simple ortogonal los factores finales están necesariamente no correlacionados, en la rotación oblicua se permite a los factores que se correlacionen. En la rotación ortogonal la estructura factorial global se mueve alrededor del origen como un marco rigido (como los radios de una rueda en torno del eje), para ajustarse a la configuración de los agrupamientos de variables interelacionadas.

En la rotación oblicua, para una estructura simple, los factores rotan individualmente para ajustarse a cada agrupamiento diferente. La relación entre los factores resultantes refleja, entonces la relación entre los agrupamientos.

La rotación ortogonal es un subconjunto de las rotaciones oblicuas. Si los grupos de relaciones son, de hecho, no correlacionados, entonces la rotación oblicua resultará en factores ortogonales. Por lo tanto, la diferencia entre la rotación ortogonal y la oblicua no está en discriminar factores correlacionados o no, sino en determinar si esta distinción es empírica o está impuesta a los datos por el modelo.

Se discute sobre si el mejor enfoque científico es la rotación ortogonal o la oblicua. Los propugnadores de la rotación oblicua generalmente argumentan, basándose en dos puntos: Primero, que genera información adicional; hay una definición más precisa de los límites de un agrupamiento, y las variables centrales en un agrupamiento pueden ser identificadas por sus altas cargas. Segundo, se obtienen las correlaciones entre los agrupamientos, y estas permiten al investigador computar el grado en que sus datos se aproximan a los factores ortogonales.

Además de dar más información, la rotación oblicua se justifica con fundamentos epistemológicos. Una justificación es que el mundo real no debe ser tratado como si los fenómenos se coagularan en agrupamientos no relacionados. De la misma manera que los fenómenos pueden interrelacionarse en agrupamientos, también los mismos agrupamientos pueden estar relacionados.

La rotación oblicua permite que esta realidad se refleje en las cargas de los factores y sus correlaciones. Una segunda justificación es que las correlaciones entre los factores permiten que la investigación científica, en busca de uniformidad sea realizada en el segundo orden.

Las mismas correlaciones de factores pueden ser sometidas al análisis factorial, para determinar las correlaciones más generales, más abstractas y más comprensivas, y las influencias más agudas que subyacen a los fenómenos.

La rotación oblicua preserva la comunalidad de las variables, como la rotación ortogonal.

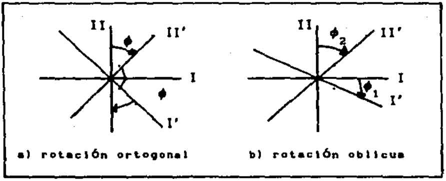
Algoritmos computados para la rotación oblicua.

Un gran número de algoritmos computados son desarrollados para ejecutar la rotación oblicua. Estos algoritmos más convenientemente disponibles son: el oblimin, quartimin, biquartimin y covarimin. Todos estos representan algoritmos designados, para satisfacer varios tipos de criterios de la estructura simple. Desafortunadamente, ningún algoritmo permite producir una solución superior, así varios algoritmos diferentes pueden necesitar ser probados sobre el mismo conjunto de datos.

El método más usado para la rotación oblicua es el oblimin, está disponible en el SPSS/PC. Un parámetro llamado $\delta$ (delta) controla la extensión de la oblicuidad. Cuando $\delta$ es cero, los factores son más oblicuos. Para valores negativos de $\delta$, los factores vienen menos oblicuos.

La diferencia entre estos dos tipos de rotación es descrita en términos de los ángulos de rotación. En general, hay dos de tales ángulos: en la Fig. 3.2 se muestra gráficamente la rotación ortogonal y oblicua, el ángulo entre el primer eje original (I) y el correspondiente eje rotado (I') y el ángulo entre el segundo eje original (II) y el eje rotado (II'). Bajo la rotación ortogonal (porque los ejes se mantienen perpendiculares a uno y otro durante la rotación), solamente un ángulo tiene que ser especificado, este es el ángulo $\phi$ mostrado en la Fig. 3.2. Bajo la rotación oblicua, dos ángulos necesitan ser especificados, estos son el ángulo $\phi_1$ (entre el eje I y I') y el $\phi_2$ (entre II y II') señalados en la Fig. 3.2.

Fig. 3.2. Ilustración gráfica de la rotación ortogonal y oblicua.




Una importante diferencia estadística entre la rotación oblicua y ortogonal es que los factores resultantes desde la rotación ortogonal de los componentes principales, permanecerán estadísticamente no correlacionados (esto es, los cosenos de factores serán todos cero), mientra que los factores resultantes desde una rotación oblicua serán usualmente correlacionados, para alguna extensión (esto es, algunos o todos los cosenos de factores serán diferentes de cero).

Para generalizar estadísticamente los factores no correlacionados es un objetivo deseable, primeramente porque de la ventaja asociada con la representación de un conjunto complejo de interrelaciones, entre diversas variables correlacionadas en términos de pocos índices no correlacionados. Otra propiedad deseada de la rotación ortogonal es que la cantidad de la variación total contada por los factores, bajo consideraciones no es afectada por la rotación.

Sin embargo, el uso de una justa rotación ortogonal, puede resultar en no encontrar el mejor conjunto de factores rotados. Frecuentemente, los investigadores pueden razonar sobre bases empíricas qué características de los factores serán medidas, y el uso de la rotación ortogonal muchas veces no es suficiente, para determinar factores con los atributos deseados. Entonces los primeros objetivos de la rotación son ‹estructura simple› y "factores significativos", estos objetivos son más probables de lograrse si se considera tanto la rotación oblicua como la ortogonal.

Esto significa lograr que cada factor tenga una carga diferente de cero para solamente algunas de las variables, y también se desea que cada variable tenga cargas diferentes de cero para solamente pocos factores, probablemente uno. Esto permite a los factores ser diferenciados cada uno de otro. Si varios factores tienen altas cargas sobre las mismas variables, es difícil acertar como los factores difieren. De hecho ejecutando los ejes factoriales para convertirse en oblicuos, es posible llegar a mucho más factores interpretables.

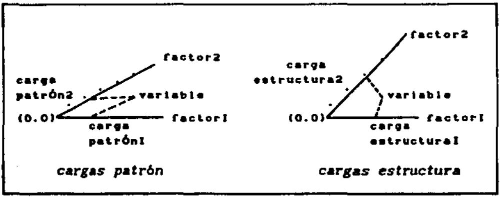
La rotación no afecta la bondad de lo apropiado de la solución factorial. Esto quiere decir que aunque la matriz de factores cambia, la comunalidad y el porcentaje de la varianza total explicada no cambia. El porcentaje de la varianza contada para cada uno de los factores está, sin embargo, cambiada. La rotación redistribuye la varianza explicada por los factores individuales.

Aunque tal vez, lo más simplificado para acercarse a la rotación será examinar los datos gráficamente, y entoces decidir sobre la apropiada rotación, visualmente esto involucra considerable subjetividad. Porque la rotación es generalmente ejecutada usando algoritmos computarizados, los cuales están fundamentados sobre criterios cuantitativos definidos. Sin embargo, se recomienda que el investigador considere mirar los datos gráficamente, y entonces un dibujo geométrico a menudo provee una visión adicional.

Con respecto al uso de algoritmos computados para el análisis factorial, también serán consistentes de alguna complejidad adicional de la rotación oblicua, no presentada para la rotación ortogonal. En particular, considerando los resultados de una rotación oblicua ejecutada a través de algún algoritmo computado, hay dos alternativas representadas para las cargas factoriales [esto es, las coordenadas de los puntos o (variables) con respecto a los ejes rotados] dependiendo de como cada punto es proyectado sobre el eje rotado.

Una alternativa, produce lo que se llama cargas patrón, que se funda sobre la proyección de cada punto sobre cada eje rotado por lineas paralelas a estos dos ejes, asi que las cargas factoriales son entonces, definidas como las dos coordenadas proyectadas. La otra alternativa, produce lo que se llama cargas estructura, que está hecha sobre la proyección de cada punto sobre cada eje rotado por lineas perpendiculares a estos dos ejes. La carga patrón y la carga estructura son representadas gráficamente en la Fig. 3.3 .

Fig. 3.3. Las cargas Patrón y Estructura.




Una importante diferencia entre cargas patrón y estructura es que las cargas patrón no son realmente coeficientes de correlación entre variables y factores, mientras que las cargas estructura representan tales correlaciones. Sin embargo, la matriz patrón es frecuentemente más usual que la matriz estructura con miras a la interpretación de los factores rotados.

## 3.8. Determinación de los Puntajes Factoriales.

La determinación de los puntajes factoriales, es el cuarto y último paso del proceso del análisis factorial. Los puntajes para cada uno de los factores pueden ser computados para cada caso o unidad de estudio. Estos puntajes pueden, entonces ser usados en una variedad de otros análisis.

Como previamente se describió, el puntaje factorial es un valor nunérico de un factor F obtenido, por la sustitución especifica de valores para las X's (estandarizadas) dentro de la expresión.

$$F = W_1X_1 + W_2X_2 + . . . + W_pX_p \qquad (3.35)$$

Existen algunos métodos propuestos para encontrar los puntajes factoriales. Dos de los métodos más usados son:

**Método de puntajes factoriales directo.**

La matriz de puntajes factoriales, es una matriz condensada de datos, las columnas son combinaciones lineales de las variables originales y los renglones son las unidades de estudio o casos en el análisis. Bajo ciertas condiciones la solución para la matriz de puntajes factoriales, puede ser relativamente simple y directa. Partiendo de la ecuación factorial básica, ver relación (3.5).

$$Z = F A' + E \quad , \text{ se deriva F desde: } F = Z(A')^{-1}$$

Esto asume que A es cuadrada, matriz no singular; contiene muchos factores como variables.

Asumiendo k factores y p variables, aproximadamente se supone E pequeña, entonces:

$$Z_{(Nxp)} = F_{(Nxk)} A'_{(kxp)}$$

Posmultiplicando por A se tiene:

ZA = FA'A finalmente, después posmultiplicando por $(A'A)^{-1}$, se obtiene:

$$F = ZA(A'A)^{-1} \qquad (3.36)$$

En el caso de haber pocos factores que variables, solamente una porción de la varianza total de las variables será tomada en cuenta. Por lo tanto, las observacines originales, elementos de Z, son considerados solamente como aproximaciones, para los valores necesarios, para computar los puntajes factoriales exactos en términos de esta varianza parcial. Los puntajes factoriales, en este caso, son aproximaciones a el factor medido en el espacio reducido delineado por los k-factores.

El otro método es el de regresión y es usado en SPSS/PC, dentro del análisis factorial, para estimar la matriz de coeficientes de puntajes factoriales.

# C A P I T U L O IV

## UTILIZACIÓN DEL PAQUETE ESTADÍSTICO SPSS/PC

### 4.1. Contenido y Funcionamiento del Paquete.

El paquete SPSS/PC (Paquete Estadistico para las Ciencias Sociales) descrito aqui, es para el uso de PC/XT, y se utiliza la versión 03.

El objetivo principal del sistema SPSS es el de proporcionar métodos estadisticos computarizados, para el análisis estadistico descriptivo e inferencial de datos muestrales o poblacionales.

#### Contenido.

Este paquete SPSS/PC consta de dos manuales los cuales contienen información sobre el Tutorial del SPSS, la Guía Estadistica, la Referencia de Comandos, Ejemplos y un Glosario entre los puntos más importantes.

#### Tutorial.

El sistema SPSS/PC incluye el tutorial para introducir al usuario al SPSS. El tutorial está en un diskette separado, es fácil instalarlo y operarlo. Es un conjunto de lecciones que contienen ejemplos acerca de los comandos del sistema SPSS/PC, permitiendo practicar el uso de éste.

Guía estadística.

Es una guía completa para el usuario, contiene teoría estadística sobre los procesos de análisis estadístico que comprende el SPSS, además el procedimiento de la utlización de los comandos para dichos procesos y ejemplos ilustrativos.

Procesos que describe la guía estadística:

1. Tabulación de datos: Proceso o subrutina FREQUENCIES.

2. Estadísticas descriptivas: Proceso FREQUENCIES y DESCRIPTIVES.

3. Tablas cruzadas y medidas de asociación:Proceso CROSSTABS.

4. Descripción de subpoblaciones: Proceso MEANS.

5. Pruebas diferentes entre dos medias: Proceso T-TEST.

6. Gráficando datos: Proceso PLOT.

7. Midiendo líneas de asociación: Proceso CORRELATION.

8. Una forma de análisis de varianza: Proceso ONEWAY.

9. Análisis de varianza: Proceso ANOVA.

10. Pruebas no-paramétricas: Proceso NPAR TESTS.

11. Regresión múltiple: Proceso REGRESSION.

12. Análisis de líneas logarítmicas: Proceso HILOGLINEAR.

13. Análisis factorial: Proceso FACTOR. Esta es la subrutina que se aplicará, en el caso práctico desarrollado en el

capítulo V.

14. Agrupando casos: Proceso CLUSTER.

15. Reporte de resultados: Proceso REPORT.

Estos son todos los procesos o subrutinas que conforman el paquete SPSS/PC.

Generalmente los temas comprendidos en la guía estadística son tratados de la siguiente manera:

1. Breve explicación de la teoria estadistica, sobre el tema que trata el proceso o la subrutina.

2. El procedimiento para poder ejecutar este proceso.

3. La especificación de las variables.

4. Las estadísticas opcionales para el proceso.

5. Ejemplos ilustrativos, con la sintaxis adecuada del comando a utilizar, según las necesidades del problema planteado.

6. Presentación de los resultados al ejecutar el punto anterior.

Referencia de comandos.

Es una detallada descripción de la sintaxis y operaciones de cada comando del SPSS/PC. Los comandos son presentados en orden alfabético.

Ejemplos.

Los ejemplos presentados en el manual, ilustran los usos típicos de los procesos analíticos del SPSS/PC.

La explicación de la entrada y salida son arregladas, para mostrar el conjunto de comandos, que lleva la terminación de una tarea completa de análisis de datos.

Los ejemplos tienen un contenido interpretativo, y conlleva la comprensión de la lógica de la estructura de comandos del SPSS/PC.

Glosario.

El glosario define los términos usados en el manual que no son familiares al usuario o lector.

Funcionamiento.

El SPSS/PC es un sistema basado en comandos, una forma de operar es iterativa y es como sigue:

. El SPSS/PC manda una señal a la pantalla en espera de un comando.

. Se introduce un comando.

. El SPSS/PC responde al comando.

. El SPSS/PC envía otra señal similar a la anterior en espera de otro comando.

Así, sucesivamente este proceso continúa hasta que finaliza la sesión del SPSS/PC.

El sistema ejecuta un comando de operación cada vez que lo encuentre.

Otra forma de operar del sistema SPSS/PC es mediante la creación de un archivo de datos y de lineas de comandos, en forma externa usando procesadores de texto como el WORD, WORD PERFECT, CHIWRITER, EDLIN, WORDSTAR, etc.

Para el funcionamiento del sistema SPSS/PC, existen tres tipos de comandos: Comandos de operación, Comandos de definición de datos y manipulación , y Comandos de procedimiento. Cada uno sirve para una diferente función, y el SPSS responde diferentemente para cada tipo.

A continuación se da el contenido y la definición de cada tipo de comando, en una forma muy general.

Comandos de operación.

Estos comandos proveen información acerca de la forma como el sistema SPSS/PC ejecuta una función.

Dichos comandos son los siguientes:

-HELP, SHOW y DISPLAY.
Cuyo objetivo es dar asistencia al usuario.

-SET.
Especifica opciones para operaciones y salidas.

-INCLUDE.
Incluye un conjunto de comandos que se encuentran en un archivo creado por fuera del paquete.

-FINISH.
Cuya finalidad es dar por terminada la sesión del paquete.

**Comandos de definición de datos y manipulación.**

Estos comandos nos dicen dónde y cómo se leen los datos, como se calculan nuevas variables, como se cambian valores de variables existentes, como se identifican valores perdidos, cuales casos se usan y como se etiquetan las salidas.

Los comandos de definición de datos y manipulación son los siguientes:

-DATA LIST, BEGIN DATA y END DATA, IMPORT y GET.
Nos permiten introducir datos (lectura de datos).

-RECODE, COMPUTE, IF y COUNT.
Realizan la transformación de datos.

-MISSING VALUE.
Define datos perdidos.

-SELECT IF, PROCESS IF, N ,SAMPLE y WEIGHT.
Cuyo objetivo es la selección de datos y ejecutar ponderaciones.

-TITLE, SUBTITLE, VARIABLE LABELS, VALUE LABELS y FORMAT.
Estos comandos dan lugar a titulos y especificar formatos de salida.

**Comandos de procedimiento:**

Con estos comandos se manda al SPSS/PC a hacer algunas cosas con los datos, tales como:

-Ejecutar un análisis estadistico.

-Producir un reporte, listar o gráficar.

-Salvar los datos en un archivo.

Cuando se introduce un comando de procedimiento, el SPSS/PC primero ejecuta los anteriores comandos de definición de datos y manipulación. Estos crean un archivo activo. El archivo activo contiene los datos que se mandaron al SPSS para leer, los resultados de cualquier transformación solicitada, y una serie de información que se tiene provista acerca de cada variable (nombres, etiquetas, valores perdidos, etc. ). El procesador entonces lee los datos desde el archivo activo, para producir el análisis requerido o reporte.

Se pueden especificar diferentes tipos de análisis, para el mismo archivo activo y puede modificarse el archivo activo en orden, para ejecutar análisis especificos sobre éste.

Los comandos disponibles en el SPSS/PC proveen un amplio rango de análisis estadisticos, reportes de datos desplegados y funciones de utileria. Estos son clasificados en siete categorias:

1. Datos desplegados. Los comandos que despliegan datos son: LIST, PLOT y REPORT.

2. Estadisticas descriptivas. Los comandos que proveen estadisticas descriptivas de variables simples son: DESCRIPTIVES y FREQUENCIES.

3. Estadisticas de categorias. Los comandos estadisticos para datos con categorias son: CROSSTABS y HILOGLINEAR.

4. Grupos comparativos. Los comandos estadisticos para grupos comparativos son: T-TEST, ONEWAY, MEANS y ANOVA.

5. Estadisticas multivariadas. Los comandos estadisticos para análisis multivariado son: CORRELATION, CLUSTER, REGRESSION y FACTOR.

6. Estadísticas no-paramétricas. El comando para estadísticas no-paramétricas es: NPAR TESTS.

7. Utilerías. Los comandos de utilería son: WRITE, SORT CASES, EXPORT y SAVE.

Ejemplos ilustrativos.

Estos ejemplos muestran el uso de algunos de los comandos descritos anteriormente.

Ejemplo 1:

```
DATA LIST FIXED /BI 3-7 PEASI 9-13 SUB 15-19 PR 21-25
/PA 27-31.
BEGIN DATA.
1 40.01 14.54 14.06 36.04 17.89
2 14.48 09.64 11.65 22.28 09.47
3 17.33 10.37 12.99 46.54 19.35
END DATA.
DESCRIPTIVES BI PEASI SUBE PR PA
/STATISTICS=1,2,5,6.
FINISH.
```

El comando DATA LIST define el nombre de las variables, y la localización por columna de las variables usadas en el análisis.

El BEGIN DATA indica el inicio de las lineas de datos y el END DATA señala el fin de las lineas de datos.

El comando DESCRIPTIVES requiere las estadisticas descriptivas, para las cinco variables.

La opción STATISTICS solicita que sólo se desplieguen las estadísticas señaladas con el número esto es, la media, el error estándar de la media, la desviación estándar y la varianza.

El FINISH significa fin de la sesión del SPSS/PC.

Ejemplo 2:

```
DATA LIST FILE='DATA80.DAT'1 FREE
/BI PEASI SUBE PR PA.
CORRELATION BI PEASI SUBE PR PA
/OPTION=6
/STATISTICS=1.
FINISH.
```

El DATA LIST FILE indica que datos son leidos desde un archivo externo, con un formato libre y también define el nombre de las variables.

El comando CORRELATION solicita la correlación entre las variables BI, PEASI, SUBE, PR y PA.

La opción 6 nos dice que la matriz debe ser desplegada en forma seriada.

La STASTISTICS 1 pide las estadisticas univariadas para el análisis.

El FINISH significa fin de la sesión.

[1]El archivo DATA80.DAT contiene datos de 19 indicadores socio-econÓmicos, para las 32 entidades federativas de la RepÚblica Mexicana, del censo de 1980.

4.2. Uso de la Subrutina FACTOR.

La subrutina FACTOR es un análisis estadístico, de los muchos disponibles en el SPSS/PC. Se clasifica dentro de los comandos de procedimiento, para análisis multivariados.

Esta subrutina FACTOR raliza los cuatro pasos analíticos fundamentales, para llevar a cabo un análisis factorial, los cuales son: generación de la matriz de correlación, extracción de factores iniciales, rotación de la matriz de factores iniciales, y la generación de la matriz de coeficientes de puntajes factoriales.

Funcionamiento del comando FACTOR.

Existe una variedad de técnicas de extracción y rotación disponibles en el comando FACTOR. Los métodos de extracción incluyen el Análisis de Componentes Principales, el método de factores de Máxima Probabilidad entre otros. Los métodos de rotación de factores son varimax, equamax, quartimax y oblimin.

También pueden gráficarse las cargas factoriales, para ayudar en la selección e interpretación de factores.

El comando FACTOR aceptará una matriz de correlación o una matriz de cargas factoriales como datos de entrada, que serán como los valores originales de las variables de cada unidad de estudio.

Subcomandos globales y bloques de análisis.

El comando FACTOR opera vía subcomandos. Hay dos tipos de subcomandos FACTOR. Los subcomandos globales son especificados una vez, y son para la entrada del procedimiento FACTOR. Los otros subcomandos son de análisis y están organizados en bloques.

Los subcomandos globales son: VARIABLES, MISSING y WIDTH.

VARIABLES: este subcomando lista un subconjunto de variables desde el archivo activo, que estarán disponibles para un análisis FACTOR. Si no se especifica enseguida el subcomando EXTRACTION o ROTATION, en forma implícita el análisis de componentes principales con la rotación varimax será producida.

VARIABLES puede ser situado antes de todos los otros subcomandos excepto MISSING y WIDTH.

Solamente las variables nombradas sobre el subcomando VARIABLES pueden ser referidas en los subsecuentes subcomandos. Se especifica sólo un subcomando VARIABLES sobre el comando FACTOR.

MISSING:provee varias alternativas, para el tratamiento de valores perdidos. Se usa este subcomando en el tratamiento de la matriz de correlación, en cuanto a sus valores perdidos. Si se omite el subcomando MISSING, o se incluye sin ninguna especificación, los valores perdidos en forma de lista son borrados.

Las siguientes especificaciones son, para el subcomando MISSING:

LISTWISE - Borra los valores perdidos que están en forma de lista. Sólo los casos con valores válidos, sobre todas las variables del subcomando VARIABLES son usados. Esto sucede también en forma implícita.

PAIRWISE - Borra los valores perdidos en forma de pares. Los casos con datos completos, sobre cada par de variables correlacionadas son usados.

MEANSUB - Reemplaza los valores perdidos con la media de las variables. Este incluye a ambos, uso de valores perdidos y sistema de valores perdidos.

INCLUDE - Incluye valores perdidos. Los casos con uso de valores perdidos, son tratados como observaciones válidas. El sistema de valores perdidos es excluido desde el análisis.

Sólo se especifica un subcomando MISSING, por un comando FACTOR. El MISSING puede ser colocado antes de todos los otros subcomandos excepto de VARIABLES y WIDTH.

WIDTH: Controla el ancho del desplegado, para la salida de los factores.

El valor sobre WIDTH es un entero. Este valor pasa por encima de la especificación hecha en el comando SET. Se especifica solamente un subcomando WIDTH, por un comando FACTOR. El WIDTH se puede poner en cualquier sitio.

Orden de los subcomandos.

El orden de los subcomandos de FACTOR es flexible. Los subcomandos globales VARIABLES y MISSING, pueden preceder cualquier subcomando de bloque de análisis.

Una vez iniciado un bloque de análisis, en forma implícita o explícita el orden de los subcomandos restantes del bloque de análisis es determinado, primeramente por el orden en el cual el análisis factorial es procesado, y por el análisis que se desea.

Los subcomandos de desplegado, PRINT, PLOT, DIAGONAL y FORMAT son declarados una vez, en cualquier sitio dentro de un bloque de análisis.

El PRINT, PLOT y DIAGONAL se aplican a un bloque entero de análisis, y sólo para ese bloque se usan. El FORMAT se aplica para una bloque de análisis, y también para todos los siguientes bloques.

Dentro de un bloque de análisis, la colocación del subcomando CRITERIA es importante, afecta toda la extracción y rotación que le sigue. Un CRITERIA es para el resto del procedimiento FACTOR, una vez especificado.

La forma en la cual los subcomandos del bloque de análisis, deben ser designados para que permanezca el efecto, será conveniente para hacer el análisis.

Bloques de análisis.

Inicia con el subcomando ANALYSIS que selecciona un subconjunto de variables, para analizarlas. Si se omite el subcomando ANALYSIS, el bloque de análisis es implícitamente inicializado con el subcomando VARIABLES, usando todas las variables nombradas. Si de VARIABLES sigue otro subcomando de bloque de análisis anterior a ANALYSIS, también implícitamente se inicia ese bloque.

El subcomando ANALYSIS permite ejecutar diferentes análisis, sobre el conjunto de variables nombradas en el subcomando VARIABLES, o sobre un subconjunto de estas variables.

La palabra TO en una lista de variables sobre el subcomando ANALYSIS, se refiere al orden de las variables en el subcomando VARIABLES, no al orden en el archivo. La palabra ALL se refiere a todas las variables listadas sobre el subcomando VARIABLES.

Métodos de extracción de factores.

La fase de extracción en el bloque de análisis, es inicializada explícitamente con el subcomando EXTRACTION. En forma implícita, FACTOR extraee factores por el método de análisis de componentes principales. Esto sucede si no está especificado el subcomando EXTRACTION. Para otener la extracción de factores por otro método, se usa el subcomando EXTRACTION con alguna de las siguientes especificaciones:

PC - Análisis de Componentes Principales.

PAF - Factores por el método Ejes Principales.

ML - Máxima Probabilidad.

ALPHA - Factores Alfa.

IMAGE - Factores Imagen.

VLS - Mínimo cuadrado no ponderado.

GLS - Mínimo cuadrado generalizado.

Se puede especificar más de un subcomando EXTRACTION. Esto es, subcomandos EXTRACTION múltiples en cada bloque de análisis, para producir salidas sobre diferentes métodos de extracción, para subconjuntos de variables nombradas en el subcomando VARIABLES.

Especificar los valores de la diagonal.

El subcomando DIAGONAL se usa, para especificar los valores iniciales de la diagonal en conjunción con la factorización ejes principales (EXTRACTION = PAF). DIAGONAL se especifica con una de las opciones siguientes:

Valuelist - Los valores de la diagonal son sustituidos solamente, para la factorización de ejes principales.

DEFAULT - Los 1's sobre la diagonal para componentes principales o comunalidad inicial, estimada sobre la diagonal para métodos de factores.

Se suple el mismo número de valores en la diagonal, como hay variables en el análisis.

Rotación de factores.

La fase de rotación es explicitamente inicializada con el subcomando ROTATION, el cual especifica el método de rotación a usarse.

Cuatro métodos son disponibles en FACTOR: varimax, equamax, quartimax y oblimin. Cuando los subcomandos EXTRACTION y ROTATION son omitidos los factores son rotados en forma implicita, usando el método varimax. Si se desea otro método de rotación o no rotar, debe usar el subcomando ROTATION.

Para especificar el método de rotación, usar una de las siguientes palabras reservadas sobre el subcomando ROTATION:

VARIMAX - Rotación varimax. Esta se da en forma implicita si EXTRACTION y ROTATION se omiten.

EQUAMAX - Rotación equamax.

QUARTIMAX - Rotación quartimax.

OBLIMIN - Rotación oblimin directa.

NO ROTATE - Este es implicito si EXTRACTION es especificado, pero no ROTATION.

Oblimin usa implicitamente el valor delta igual a cero.

La rotación no ocurre si EXTRACTION es especificada dentro de

ROTATION.

Se pueden especificar más de una rotación, para una extracción dada.

Criterios para la extracción y rotación.

El subcomando CRITERIA sirve para controlar el criterio de extracciones y rotaciones de factores.

Las siguientes especificaciones son disponibles, para el subcomando CRITERIA:

FACTORS (nf) - El número máximo de factores extraidos. En forma implicita es el número de eigenvalores[2] mayores que MINEIGEN.

MINEIGEN (eg) -Los eigenvalores minimos usados para controlar el número de factores. El valor implicito es 1.

ITERATE (ni) - Número de iteracciones para la solución de factores. El valor implicito es 25.

ECONVERGE (e1) - Criterio de convergencia por extracción. El valor implicito es .001.

RCONVERGE (e2) - Criterio de convergencia por rotación. El valor implicito es .0001.

KAISER - Normalización Kaiser en rotación. Esta es en forma implicita.

[2] El término eigenvalores lo usa el manual del SPSS/PC, para referirse a los valores propios o valores característicos, obtenidos de la ecuación característica.

NOKAISER - No a la normalización Kaiser.

DELTA (d) - Valor de delta para la rotación oblimin directa. El valor implicito es 0.

DEFAULT - Usa los valores implicitos para todos los criterios.

Una vez especificado el criterio permanece afectando el proceso, hasta que explicitamente se especifica de otra manera.

Estadisticas opcionales.

El subcomando PRINT requiere estadisticas que no son imprimidas implicitamente. Se usa un sólo PRINT para cada bloque de análisis. Las especificaciones siguientes están disponibles, para el PRINT:

UNIVARIATE - Número de observaciones válidas, medias y desviaciones estándar, para las variables nombradas sobre el subcomando ANALYSIS.

INITIAL - Comunalidades iniciales, eigenvalores y porcentajes de varianza explicada.

CORRELATION - Matriz de correlación para las variables nombradas en el subcomando ANALYSIS.

SIG - Nivel de significancia de la correlación. Estos son los dos extremos de probabilidad.

DET - El determinante de la matriz de correlación.

INV - La inversa de la matriz de correlación.

AIC - La anti-imagen de la matriz de covarianza y correlación.

KMO - La media de la muestra adecuada de Kaiser-Meyer-Olkin y la prueba de esfericidad de Bartlett.

EXTRACTION - Comunalidad, eigenvalores y cargas factoriales rotadas.

REPR - Correlaciones reproducidas y sus residuales.

ROTATION - Matrices de factores patrón y estructura rotadas, matriz de transformación de factores y la matriz de correlación de factores.

ESCORE - Matriz de coeficientes de puntajes factoriales. Están basados sobre una solución de regresión.

DEFAULT - Especifica INITIAL, EXTRACTION y ROTATION.

ALL - Especifica todas las estadisticas disponibles.

Al especificar el PRINT, solamente las opciones requeridas son producidas.

El subcomando FORMAT.

Se usa para reformatear el desplegado de las cargas factoriales y matriz de estructura, para ayudar a interpretar los factores.

El FORMAT tiene las siguientes opciones:

SORT - Ordena las cargas factoriales por magnitudes.

BLAN (n) - Suprime los coeficientes menores en valor absoluto que n.

DEFAULT - Pone blancos y sortea.

Puede usarse solamente un subcomando FORMAT en cada bloque de análisis.

Gráficas.

Para obtener gráficas de las cargas factoriales o eigenvalores, se usa el subcomando PLOT.

PLOT tiene las opciones siguientes:

EIGEN -Gráfica de eigenvalores en orden descendente.

ROTATION (n1 n2) - Gráfica de cargas factoriales.

La especificación n1 y n2 se refiere a los factores usados como ejes. Varios pares de factores en paréntisis pueden ser especificados, sobre un ROTATION.

Una gráfica es desplegada por cada par de número de factores encerrados en paréntesis.

Se usa sólo un subcomando PLOT en cada subcomando ANALYSIS.

Las gráficas son basadas sobre factores rotados. Para obtener una gráfica de factores no rotados, explicitamente debe especificarse NO ROTATE en el subcomando ROTATION.

Estructura general de la subrutina FACTOR.

```
FACTOR VARIABLES=
/MISSING=
/WIDTH=
/ANALYSIS=
```

```
/FORMAT=
/PRINT=
/PLOT=
/DIAGONAL=
/CRITERIA=
/EXTRACTION=
/ROTATION=
```

Después del signo igual se pone una o más de las opciones que se dieron anteriormente, para cada subcomando. Considerando también las restricciones y la flexibilidad del orden de los subcomandos.

Ejemplo:

El ejemplo sólo considera algunos de los subcomandos, presentados en la subrutina FACTOR.

```
DATA LIST FILE='DATA80.DAT'
/BI 4-8 PEASI 10-14 SUBE 16-20 PR 22-26 PA 28-32.
VARIABLE LABELS BI 'Bajos Ingresos' PEASI 'PEA Sin Ingresos'
/SUBE 'Subempleo' PR 'Población Rural' PA 'Población Agrico--
/la'.
LIST.
FACTOR VARIABLES=BI PEAS SUBE PR PA
/ROTATION=OBLIMIN
/PRINT=ALL
/PLOT=EIGEN ROTATION (1 2).
FINISH.
```

El comando DATA LIST define el nombre de las variables y la colocación de columnas, para las variables usadas en el análisis.

El comando VARIABLE LABELS asigna etiquetas descriptivas para las variables.

El comando LIST lista todas las variables con sus respectivos valores, para todos los casos.

FACTOR VARIABLES, este subcomando indica cuales variables son incluidas en el análisis factorial, y en los siguientes subcomandos.

El subcomando ROTATION solicita la rotación de los factores iniciales por el método oblimin.

El PRINT es el subcomando que especifica todos los posibles resultados factoriales.

El PLOT indica la graficación de los eigenvalores y de la rotación de los factores 1 y 2.

FINISH indica fin de la sesión del SPSS/PC.

4.3. Introducción de Datos y Elaboración de Programa.

La introducción de datos al paquete SPSS/PC.

El SPSS/PC permite introducir datos en varias formas: datos sin preparación que son definidos con los comandos de definición de datos; datos en forma de matriz; como un archivo de datos definido en el sistema SPSS; o como un archivo portable o externo.

Para definir los datos se usa el comando DATA LIST, sólo un DATA LIST se introduce por cada sesión del SPSS, y este opera sobre solamente un conjunto de datos en un mismo tiempo.

El archivo de datos puede tener valores para una o más variables, para uno o más casos. También, se introducen valores para estas variables directamente en el SPSS, o leerse desde un archivo de datos que se tiene creado en un programa editor.

Los datos son arreglados en formatos libre o fijo.

Definición de datos.

El primer comando de definición de datos es DATA LIST, el cual nos dice donde encontrar los datos y cómo leerlos. También, se pueden introducir y leer datos desde un archivo separado.

Si el archivo de datos no está definido dentro del sistema SPSS/PC, entonces con el subcomando FILE y el nombre del archivo, en el cual fueron almacenados los datos, se localiza el archivo de datos a utilizar, esto es:

DATA LIST FILE='DATA80.DAT'

El subcomando FILE es requerido cuando los datos están contenidos en un archivo externo.

Después de tener identificado el archivo de datos, se asignan nombres a cada una de las variables, y se da su localización sobre el archivo.

Ejemplo:

```
DATA LIST FILE='DATA80.DAT'
/BI PEASI SUBE PR PA.
```

Acompañado al nombre de las variables, se especifica la localización de las columnas sobre el archivo de datos. Todas las variables sobre la misma linea son identificadas al mismo tiempo.

Ejemplo:

```
DATA LIST FILE='DATA80.DAT'
/BI 4-8 PEASI 10-14
```

La variable Bajos Ingresos se localiza en las columnas de la 4 hasta la 8.

Tipos de variables.

Se definen dos tipos de variables, numéricas y alfanuméricas. Las numéricas sólo contienen números, y pueden ser enteras o con decimales. Las variables alfanuméricas contienen letras, números y caracteres especiales, y son identificadas con la letra (A) entre paréntesis, seguidas de la especificación de columnas sobre el comando DATA LIST.

Indicando el punto decimal en los datos.

En forma implicita el DATA LIST asume que el tipo de formato de los datos es numérico, y que los números son enteros, o que cualquier punto decimal es explicitamente codificado. Para indicar los valores no enteros cuando el punto decimal no es especificado en los datos, se especifica el número de lugares después del punto decimal entre paréntesis, en seguida de la especificación de columnas.

Ejemplo:

```
DATA LIST FILE='DATA80.DAT'
/BI 4-8(2) PEASI 10-14(2)
```

Indica que las variables BI y PEASI tienen dos decimales.

Entrada de datos por archivos libres.

Usando el formato libre, el SPSS/PC asigna valores a las variables secuencialmente. Asigna el primer valor a la primera variable para el primer caso, el segundo valor para la segunda variable, y así sucesivamente hasta que todas las variables son asignadas para el primer caso. Entonces empieza a asignar valores a las variables para el segundo caso y así, subsecuentemente.

Con el formato de archivo libre, sucesivamente los valores de los datos son simplemente separados por uno o más blancos o una coma. Las variables pueden estar en el mismo orden para cada uno de los casos, pero éstas no necesitan estar en la misma columna.

Al seleccionar de esta manera de introducir los datos, se especifica la palabra FREE después del comando DATA LIST. En este caso la localización de columnas no se especifica después del nombre de la variable. También, se incluye el punto decimal en los datos y la longitud de las variables alfanuméricas.

La mayor desventaja al introducir datos con archivos libres, es que inadvertidamente se omiten valores, y todos los valores para las subsecuentes variables y casos son incorrectos.

Ejemplo:

```
DATA LIST FREE
/BI PEASI SUBE PR PA.
```

Entrada de datos con formato fijo.

En el formato fijo, el SPSS/PC asigna valores a las variables de acuerdo a la posición fijada sobre la entrada del registro. Por ejemplo, se asigna el valor en la columna 1, para la primera variable y el valor en la columna 2 hasta la 4, para la segunda variable.

Ejemplo:

```
DATA LIST FILE='DATA80.DAT' FIXED
/CE 1-2 EF 4-13(A) SUBC 15-18 SUBH 20-24.
```

La palabra FIXED es opcional.

Introducción de datos en linea.

Si se prefiere introducir datos acompañando los comandos de SPSS/PC. En este caso, se separa las lineas de datos de otras lineas con los comandos BEGIN DATA y END DATA. El BEGIN DATA sigue de los comandos de definición de datos y preceden los datos, y el END DATA sigue después de la última linea de datos.

Ejemplo:

```
DATA LIST FREE
/MG MP VCLP.
BEGIN DATA.
1 6.54 2.9 21.12
. . . . . . . .
```

```
. . . . . . . . .
. . . . . . . . .
END DATA.
```

Entrada de datos en forma de matriz.

Muchos comandos de procedimiento analizan datos en forma de matriz. Por ejemplo, el procedimiento FACTOR puede leer una matriz de correlación y ejecutar un análisis sobre ésta.

Se usa la especificación MATRIX sobre el comando DATA LIST para leer datos en esta forma.

La estructura de definición de datos para leer una matriz de datos es:

```
DATA LIST [FILE ‹nombre del archivo›]MATRIX [FIXED]
                                            [FREE]
```

El SPSS/PC también, puede leer datos desde su mismo sistema de documentación, mediante un archivo creado por el comando SAVE. Este archivo incluye los datos como un diccionario de datos. El cual contiene nombre de variables y etiquetas, valores etiquetados y valores perdidos declarados para una variable.

Ejemplos:

Archivo de datos con formato libre.

```
DATA80.DAT

11 29.86 25.88 19.49 49.17 19.17
12 20.65 37.02 18.99 70.97 44.28
13 37.87 24.78 15.48 78.89 37.03
14 23.98 19.21 17.05 35.60 18.22
```

```
15 19.31 14.94 14.44 37.61 15.26
```

El archivo DATA80.DAT es externo creado en un programa editor.

```
DATA LIST FILE='DATA80.DAT' FREE
/BI PEASI SUBE PR PA.
```

Archivo de datos con formato fijo.

```
DATA80.DAT

16 MICHOACAN  6391 45.59 23.37 22.56 30.69
17 MORELOS    6271 32.18 23.14 14.53 25.38
18 NAYARIT    7122 48.05 24.64 13.84 38.23
19 NUEVO LEON 5761 16.35 12.97 06.31 23.77
20 OAXACA     6527 46.83 40.40 31.40 31.31
```

```
DATA LIST FILE='DATA80.DAT' FIXED
/CE 1-2 EF 4-13(A) SUBC 15-18(2) SUBH 20-24
/SUBL 26-30 ANALFA 32-36 PSP 38-42.
```

Datos en linea.

```
DATA LIST FREE /MG MP VCLP.
BEGIN DATA.
1 6.54 2.9 21.12
2 5.75 1.8 9.03
3 5.63 1.5 33.67
4 5.45 2.6 39.68
5 6.32 2.1 19.87
END DATA.
```

Elaboración de programa.

El Programa se elabora con las lineas de comandos y subcomandos de la subrutina FACTOR, comandos de definición de datos y otros comandos del sistema SPSS/PC.

Programa especifico.

Este programa se utiliza en el capitulo v, con el fin de obtener la ponderación de 19 variables o indicadores socio-económicos, a través del método de componentes principales, para las 32 entidades federativas y después calcular los "Indices de Marginación".

```
DATA LIST FILE='DATA80.DAT' FREE
/EDO I1 I2 I3 . . . I19.
VARIABLES LABELS I1 'Bajos Ingresos' I2 'PEA Sin Ingresos' I3
/'Subempleo ' . . . I19 'Población Indigena'.
LIST VARIABLES=EDO I1 I2 . . . I19.
FACTOR VARIABLES=I1 TO I19
/PRINT=ALL
/PLOT=EIGEN ROTATION(1 2).
FINISH.
```

El comando DATA LIST FILE lee los datos del archivo externo DATA80.DAT creado en el editor WORD, con un formato libre y define el nombre de las variables empleadas en el análisis.

El comando VARIABLE LABELS asigna etiquetas descriptivas, para cada una de las variables definidas.

Las etiquetas para las 19 variables involucradas en el análisis factorial son las siguientes:

| | |
|---|---|
| EDO | Clave del estado. |
| I1 | Bajos Ingresos. |
| I2 | PEA Sin Ingresos. |

I3 Subempleo.
I4 Población Rural.
I5 Población Agricola.
I6 Subconsumo de Carne.
I7 Subconsumo de Huevo.
I8 Subconsumo de Leche.
I9 Analfabetismo.
I10 Población sin Primaria.
I11 Mortalidad General.
I12 Mortalidad Preescolar.
I13 Vivienda Combustible Leña y Petróleo.
I14 Vivienda 1 Dormitorio con 5 y más ocupantes.
I15 Vivienda sin Agua.
I16 Vivienda con 1 y 2 cuartos.
I17 Vivienda sin Electricidad.
I18 Vivienda sin Drenaje.
I19 Población Indigena.

El comando LIST lista todas las variables con sus respectivos valores, para todos los casos.

El subcomando FACTOR VARIABLES nos indica cuales de las variables definidas son incluidas en el análisis factorial, y en los siguientes subcomandos. En el programa el procedimiento FACTOR aplica implicitamente el método de Componentes Principales, para la extracción de los factores iniciales y el método de rotación varimax, puesto que no hay otra especificación al respecto.

El subcomando PRINT especifica que en este análisis se impriman todas las estadisticas disponibles, para los resultados factoriales.

El subcomando PLOT indica gráficar los eigenvalores y la rotación de los factores 1 y 2.

El comando FINISH señala el fin de la sesión del sistema SPSS/CP.

## 4.4. Interpretación de los Cuadros Factoriales.

En este punto se describen, el formato y aspecto de los cuadros típicos que contienen los resultados factoriales.

Los principales cuadros factoriales a interpretar son:

1. Matriz de correlación
2. Matriz de factores no rotados.
3. Reproducción de la matriz de correlación.
4. Matriz de factores rotados.
5. Matriz de correlación de factores.
6. Matriz de coeficientes de los puntajes factoriales.

Para efectos de ejemplificar la descripción de estos seis cuadros factoriales, se analiza el siguiente cuadro de información socio-econónica, para diez entidades federativas, considerando cinco variables.

Cuadro 4.1. Cinco indicadores socio-económicos datos (1980).

| CLAVE EDO. | ENTIDAD FEDERATIVA | BAJOS INGRESOS % | PEA SIN INGRESOS % | SUB. EMPLEO % | POBL. RURAL % | POBL. AGRICOLA % |
|---|---|---|---|---|---|---|
| | | I1 | I2 | I3 | I4 | I5 |
| 1 | AGUSCALIENTES | 40.01 | 14.54 | 14.06 | 36.04 | 17.89 |
| 2 | BAJA CALIFORNIA | 14.48 | 9.64 | 11.65 | 22.28 | 9.47 |
| 3 | BAJA CALIFORNIA SUR | 17.33 | 10.37 | 12.99 | 46.54 | 19.35 |
| 4 | CAMPECHE | 31.11 | 15.47 | 16.56 | 47.13 | 31.87 |
| 5 | COAHUILA | 30.38 | 8.57 | 13.99 | 29.19 | 15.78 |
| 6 | COLIMA | 34.55 | 10.23 | 14.40 | 41.77 | 27.85 |
| 7 | CHIAPAS | 25.28 | 39.07 | 14.18 | 79.68 | 57.43 |
| 8 | CHIHUAHUA | 20.95 | 18.99 | 13.26 | 33.93 | 20.75 |
| 9 | DISTRITO FEDERAL | 15.14 | 7.96 | 12.63 | 00.00 | 6.13 |
| 10 | DURANGO | 28.74 | 25.74 | 16.42 | 61.66 | 30.88 |

NOTA: Cuando se aplica un análisis factorial[3] a una serie de variables y casos, resultan varios cuadros o matrices de resultados, pero en este caso sólo se evaluarán los resultados más relevantes para el objetivo del análisis.

1. Matriz de correlaciones.

Es la primera matriz de resultados factoriales que aparece. El cuadro 4.2 presenta la matriz de correlaciones de los datos del cuadro 4.1.

Cuadro 4.2. Matriz de correlación.

| | I1 | I2 | I3 | I4 | I5 |
|---|---|---|---|---|---|
| I1 | 1.00000 | | | | |
| I2 | .13038 | 1.00000 | | | |
| I3 | .64933 | .39175 | 1.00000 | | |
| I4 | .33505 | .83920 | .55987 | 1.00000 | |
| I5 | .32100 | .88368 | .54132 | .91808 | 1.00000 |

Habitualmente se presenta toda la matriz de correlaciones involucrada en el análisis factorial, si el número de variables analizadas no es excesivamente grande. En caso contrario sólo se presenta el triángulo inferior de la matriz de correlación, puesto que ésta es simétrica.

Frecuentemente la matriz se presenta sin comentarios. La matriz de correlaciones contiene muchos conocimientos útiles, por ejemplo, el de encontrar las relaciones entre pares de variables. Analizando la matriz del cuadro 4.2, se observa que existe una

3La aplicación del análisis factorial se refiere a procesar los datos a través del paquete SPSS/PC usando la subrutina FACTOR, y la interpretación de los cuadros factoriales, es de los que resultan de la ejecución de esta subrutina.

alta relación entre los indicadores población rural y PEA sin ingreso (.84); entre la población agrícola y PEA sin ingreso (.88); y entre población agrícola y población rural (.92).

La matriz de correlaciones también contiene los coeficientes de correlación de las variables analizadas. Este coeficiente de correlación expresa el grado de relación lineal, entre las variables de las filas y las columnas de la matriz, y puede tomar valores de 0, 1 y -1; si el coeficiente se acerca a cero indica que existe menor relación entre las variables, si se aproxima a uno hay mayor relación y si el signo es negativo indica, que las variables están relacionadas inversamente.

Para interpretar el coeficiente de correlación se eleva al cuadrado y multiplica por 100, esto dará el porcentaje de variación en común, para los datos en las dos variables.

El propósito del análisis factorial es obtener ‹factores› que ayuden a explicar estas correlaciones. (se consideran las variables altamente correlacionadas).

En el cuadro 4.2 la diagonal principal de la matriz de correlaciones, contiene habitualmente la correlación de una variable consigo misma, que siempre es 1.0. Esto sucede cuando se aplica el método de componentes principales, en el caso de emplear el método de ejes principales o factores principales, la diagonal principal contendrá, por el contrario, estimaciones de comunalidad.

2. Matriz de factores no rotados.

Antes de la matriz de factores se presenta una tabla de estadísticas iniciales.

Tabla 4.3. Estadisticas iniciales.

| Variable | Comunalidad | Factor | valor propio (Eigenvalor)[4] | Pct. de Var. | Pct.Acum |
|---|---|---|---|---|---|
| I1 | 1.00000 | 1 | 3.31484 | 66.3 | 66.3 |
| I2 | 1.00000 | 2 | 1.17451 | 23.5 | 89.8 |
| I3 | 1.00000 | 3 | .30502 | 6.1 | 95.9 |
| I4 | 1.00000 | 4 | .13378 | 2.7 | 98.6 |
| I5 | 1.00000 | 5 | .07184 | 1.4 | 100.0 |

La tabla 4.3 contiene el porcentaje de la varianza atribuible a cada uno de los factores. Para ayudar a decidir como muchos factores actualmente necesitan representar los datos, es útil examinar el porcentaje de la varianza total explicada por cada factor.

En la primera columna de la tabla se enumeran las variables utilizadas en el análisis. La segunda columna con el titulo de comunalidad da la proporción de la varianza contada por los factores comunes, o la comunalidad de una variable, es 1 para todas las variables, la varianza total es la suma de la varianza de cada uno de las variables. Por simplicidad, todas las variables y factores, son expresados en forma estandarizada con media 0 y desviación estándar 1. En el ejemplo la varianza total es igual a 5, porque se analizan cinco variables.

La varianza total explicada por cada factor es listada en la columna etiquetada valor propio. La siguiente columna contiene el porcentaje de la varianza total atribuible a cada factor. La última columna, el porcentaje acumulativo, indica el porcentaje de la varianza atribuible a ese factor y los que le preceden en la tabla.

En la tabla 4.3 se nota que los factores están arreglados en

[4] Este término es empleado por el paquete SPSS/PC, al presentar los cuadros de resultados.

orden descendente de la varianza explicada. También, se nota que el nombre de las variables y factores están desplegados sobre la misma línea, no hay correspondencia entre las dos partes de la tabla.

Las primeras dos columnas proveen información acerca de la variable individual, mientras las últimas cuatro columnas describen los factores. Por lo que se les llama estadísticas iniciales de los factores.

Diversos procedimientos son propuestos para determinar el número de factores, para usar en un modelo. Un criterio sugerido es observar el "valor propio mayor o igual a 1", y elegir solamente los factores que cuenten con varianza mayor que 1, para ser incluidos.

La tabla 4.4 muestra la matriz de factores, que contienen los coeficientes, que relacionan las variables con los factores.

Tabla 4.4. Matriz de factores iniciales.

| | FACTOR 1 | FACTOR 2 |
|---|---|---|
| I1 | .52163 | .78101 |
| I2 | .85417 | -.43948 |
| I3 | .73815 | .52184 |
| I4 | .93765 | -.20340 |
| I5 | .94290 | -.24019 |

Factores extraídos por el método de componentes principales.

Las columnas definen los factores, las filas se refieren a las variables. En la intersección de fila y columna se da la carga, para la variable de la fila de acuerdo con el factor de la columna. Esto es, cada renglón de la tabla 4.4 contiene los coeficientes usados, para expresar una variable en términos de los factores. Estos coeficientes son llamados cargas factoriales, entonces ellos indican como el peso es asignado para cada factor.

Por ejemplo, el indicador Bajos Ingresos es expresado en términos de los factores como:

Bajos Ingresos = .521F1 + .781F2

Los factores con grandes coeficientes (en valor absoluto), para una variable son absolutamente correlacionados a la variable. Por ejemplo, el factor 1 tiene grandes cargas para las variables I4 y I5. La matriz de las cargas factoriales es llamada la matriz de factor patrón.

Cuando los factores estimados no están correlacionados con cada uno de los otros (ortogonalmente), las cargas factoriales son también, la correlación entre los factores y las variables. Así, la correlación entre el indicador Subempleo y el factor 1 es (.738).

La matriz de correlaciones entre variables y factores es llamada la matriz factor estructura. Así, cuando los factores son ortogonales, la matriz factor estructura es equivalente a la matriz factor patrón.

Cuando los factores son ortogonales o no lo son, las cargas factoriales son los coeficientes de regresión estandarizados en la ecuación de regresión múltiple, con la variable original como la variable dependiente y los factores como las variables independientes. Si los factores son no correlacionados, los valores de los coeficientes no son dependientes sobre cada uno de los otros factores. Ellos representan la única contribución de cada factor, y son las correlaciones entre los factores y la variables.

Para juzgar como el modelo de factores describe las variables originales, se computa la proporción de la varianza de cada variables explicada por el modelo. Entonces los factores son no correlacionados, la total proporción de la varianza explicada es justo la suma de la varianza explicada por cada factor.

Considerando el ejemplo, el indicador PEA sin ingresos I2, en el factor 1 cuenta con el 72.9 % de la varianza para esta variable. Este es obtenido por el cuadrado del coeficiente de correlación, para el factor 1 y I2 $(.854)^2$ multiplicado por 100. El porcentaje total de la varianza en el indicador PEA sin ingreso cuenta para este modelo con dos factores, por lo tanto: (72.9 + 19.3) = 92.2 % . ( viene a ser la comunalidad de I2 en la tabla 4.5).

La proporción de la varianza explicada por los factores comunes es llamada la comunalidad de la variable.

Las comunalidades para las variables son mostradas en la tabla 4.5, junto con el porcentaje de la varianza, que cuenta, para cada uno de los factores retenidos.

Tabla 4.5. Estadisticas finales.

| Variable | Comunalidad | Factor | valor propio (Eigenvalor) | Pct. de Var. | Pct.Acum. |
|---|---|---|---|---|---|
| I1 | .88208 | 1 | 3.31484 | 66.3 | 66.3 |
| I2 | .92276 | 2 | 1.17451 | 23.5 | 89.8 |
| I3 | .81718 | | | | |
| I4 | .92057 | | | | |
| I5 | .94676 | | | | |

Esta tabla es llamada como "estadisticas finales" muestra las comunalidades y estadisticas de factores, después del número deseado de factores extraidos.

Cuando los factores son estimados usando el método de componentes principales, las estadisticas de factores son las mismas en la tabla etiquetada como inicial y final, sin embargo, las comunalidades son diferentes, y toda la varianza de las variables no es explicada.

La comunalidad está en el rango desde [0 a 1], con 0 indica que el factor común no explica ninguna de la varianza, y 1 indica que toda la varianza es explicada por los factores comunes. La varianza que no es explicada por los factores comunes es atribuida por el factor único.

Interpretación de los factores.

Un paso importante en la interpretación es la identificación de los factores, para identificar éstos, es necesario agrupar las variables que tengan cargas grandes, para los mismos factores. Esta es una forma de determinar los grupos de variables. Otra conveniente estrategia es sortear la matriz de factor patrón, así, que las variables con cargas altas sobre el mismo factor aparecen juntas. Las pequeñas cargas factoriales pueden ser suprimidas.

Caracteristicas de la matriz de factores no rotados. (Resumen)

1) El número de factores (las columnas) es el número de causas o influencias de relación, entre las variables independientes (no correlacionadas) que son sustantivamente significativas.

2) Las cargas a, miden qué variables están involucradas en qué causa factorial o factor, y hasta qué punto. [ver la ecuaciones (3.1) y (3.2), del punto 3.2]. Estas pueden ser

interpretadas como coeficientes de correlación (ver el punto 3.4). El cuadrado de la carga multiplicado por 100 equivale al porcentaje de la variación, que una variable tiene en común con un factor no rotado.

Se considera el porcentaje obtenido como el porcentaje de datos sobre una variable, que puede ser producido o predicho conociendo los valores de un caso. Otra perpectiva es la de que un determinado porcentaje es la confiabilidad de la predicción de una variable, a partir de un factor o de las variables en el mismo factor.

Si se comparan las cargas factoriales para todos los factores y variables, pueden definirse aquellas variables particulares involucradas en una influencia independiente, y pueden verse aquellas variables relacionadas más frecuentemente con la influencia o causa.

3) El primer factor no rotado describe la influencia de relaciones en los datos; el segundo factor decribe la que le sigue en magnitud, siendo independiente (no correlacionado) con el primero; el tercer factor describe la tercera influencia o causa en orden de magnitud, que es independiente del primero y segundo factor, y así sucesivamente. De este modo la variación en los datos descrita, para cada influencia o causa disminuye sucesivamente en cada factor; el primer factor define la mayor cantidad de variación; el último factor, la menor. Los factores no rotados no están relacionados unos con otros.

4) La comunalidad en algunos libros se presenta encabezada por la letra "$h^2$". Es la proporción de la varianza total de la variable, que está involucrada en los factores. El coeficiente (comunalidad) multiplicada por 100, da el porcentaje de variación de una variable en común con cada factor.

Esta comunalidad es considerada también como una medida de singularidad. Sustrayendo el porcentaje de variación en común con los factores de 100, se determina la singularidad de una variable. Esto indica hasta qué punto una variable no está relacionada con las demás, hasta qué punto los datos de una variable no pueden derivarse (predecirse) a partir de los datos de las demás variables.

El valor de $h^2$ para una variable se calcula sumando los cuadrados de las cargas de la variable, considerando los factores seleccionados para el modelo factorial.

La razón de la suma de los valores en la columna $h^2$ sobre el número de variables, multiplicada por 100, equivale al porcentaje de variación total en los datos, que están sometidos a los factores. Por ejemplo, en la tabla 4.5 la suma de las comunalidades es $(4.49 / 5)100 = 89.8$, esto equivale al porcentaje de variación total, que es igual a 89.8, ver tabla 4.5 columna Pct de Var. Así, se mide el orden, uniformidad o regularidad en los datos.

5) El porcentaje de variación total muestra el porcentaje de variación con un factor. Este número mide, así, la cantidad de datos de la matriz original que pueden ser reproducidos por un factor: mide la comprensividad y fuerza de una influencia o causa. (Tabla 4.5 columna Pct. de Var.).

El porcentaje de varianza total para un factor está determinado, por la suma de la columna de las cargas factoriales al cuadrado para un factor, dividido por el número de variables y multiplicado por 100. Por ejemplo, en la tabla 4.4 para el factor 1, la suma de las cargas al cuadrado de las cinco variables es $(3.315 / 5)100 = 66.3$, este valor es igual al porcentaje de varianza para el factor 1, mostrado en la tabla 4.5, en la columna encabezada por Pct. de Var.

6) El porcentaje de varianza común indica cómo cualquier regularidad que exista entre los datos, está dividida entre los factores, el porcentaje de la cantidad de varianza común mide cuánto de la variación perteneciente a todos los factores, está involucrada en cada factor. Este porcentaje se calcula sumando la columna de las cargas factoriales elevadas al cuadrado para un factor, y se divide por la suma de las comunalidades, multiplicado por 100.

Por ejemplo, de la tabla 4.4 la suma de las cargas elevadas al cuadrado para el factor 1, es 3.315, y la suma de las comunalidades de la tabla 4.5 es 4.49, entonces tenemos (3.315 / 4.49)100 = 73.83, esto es, el porcentaje de la varianza común para el factor 1.

7) Los valores propios equivalen a la suma de la columna de cargas factoriales al cuadrado, para cada factor. En el ejemplo desarrollado, en la tabla 4.4, para el factor 1 la suma de las cargas al cuadrado es 3.315, este valor es equivalente al primer valor 3.31484, de la columna valores propios de la tabla 4.5.

Los valores propios miden la cantidad de variación de la que puede descontar estadísticamente un factor. Dividiendo los valores propios, por el número de variables, o por la suma de las comunalidades, y multiplicando por 100 se determina el porcentaje, de la varianza total, o de la varianza común respectivamente. Por ejemplo, el valor propio(3.315/5)100 y (3.315/4.49), obtenemos 66,3 que es el porcentaje de la varianza total y 73.83 es el porcentaje de la varianza común.

3. La matriz de correlaciones reproducidas.

El modelo factorial y las correlaciones entre los factores y las variables pueden ser usados, para estimar la correlación entre las variables. Recordando una de las hipótesis básicas del análisis factorial, es que la correlación observada entre las variables está unida, para participar de los factores comunes. En general, si los factores son ortogonales, el coeficiente de correlación estimado para las variables i y j es:

$$r_{ij} = \sum_{f=1}^{k} r_{jf} r_{fj} = r_{1i} r_{1j} + r_{2i} r_{2j} + \ldots + r_{ki} r_{kj} \qquad (4.1)$$

Donde k es el número de factores comunes, y $r_{ij}$ es la correlación entre el f-ésimo factor y la i-ésima variable. Ver la Ec.(1.5) del punto 1.1. La tabla 4.6 representa la matriz de correlaciones reproducidas y residuales.

Tabla 4.6. Matriz de correlaciones reproducidas y residuales.

| | I1 | I2 | I3 | I4 | I5 |
|---|---|---|---|---|---|
| I1 | .88208* | .02806 | -.14328 | .00480 | .01674 |
| I2 | .10232 | .92276* | -.00942 | -.05111 | -.02729 |
| I3 | .79261 | .40117 | .81718* | -.02612 | -.02934 |
| I4 | .33025 | .89031 | .58599 | .92057* | -.01489 |
| I5 | .30426 | .91097 | .57066 | .93297 | .94676* |

El triángulo inferior izquierdo contiene la matriz de correlación reproducida; la diagonal, las comunalides; y el triángulo superior derecho, contiene los residuales entre la correlación observada y la correlación reproducida. Hay 2(20%) residuales (arriba de la diagonal) que son > 0.05.

Ejemplo: el coeficiente de correlación estimado para el indicador Subempleo I3 y Bajos Ingresos I1, basado sobre el modelo de dos-factores, es según la ecuación (4.1):

$r_{13} = (.52)(.74) + (.78)(.52) = 0.7904$

El cuadro 4.2 muestra que el coeficiente de correlación observado entre I3 y I1 es 0.64933, así la diferencia, entre el coeficiente de correlación observado y el estimado desde el modelo es de -0.1411. Esta diferencia es llamada residual.

La matriz de correlación reproducida es otro resultado factorial, y aparece después de la tabla de estadísticas finales.

Los residuales son listados arriba de la diagonal hasta el tope de la matriz, mientras que el coeficiente de correlación estimado está en la base de triángulo inferior izquierdo.

Los elementos con un asterisco sobre la diagonal son las comunalidades, de la tabla 4.5.

Abajo de la matriz está un mensaje indicando como muchos de los residuales son mayores que 0.05 en valor absoluto. Las magnitudes de los residuales indican como también, un apropiado modelo factorial reproduce las correlaciones observadas. Si los residuales son grandes, el modelo no es adecuado para los datos, y debe ser reconsiderado.

4. Matriz de factores rotados.

Los factores no rotados definen sucesivamente, las influencias o causas más generales de relación en los datos. No sucede lo mismo con los factores rotados. Estos delinean las distintas agrupaciones de relación, si existen. Significa transformar la matriz de factores iniciales, en otra que sea más fácil de interpretar.

La matriz de factores rotados se presenta después de la matriz de correlación reproducida, en los resultados factoriales. Esta matriz aparece en la siguiente tabla 4.7.

Tabla 4.7. Matriz de factores rotados.

| | FACTOR 1 | FACTOR 2 |
|---|---|---|
| I1 | .05606 | .93752 |
| I2 | .95921 | .05179 |
| I3 | .37379 | .82308 |
| I4 | .91210 | .29772 |
| I5 | .93521 | .26861 |

La matriz rotada tiene las siguientes caracteristicas:

1) Si una matriz rotada es ortogonal, esto se menciona en el titulo de la matriz (por ejemplo, "factores rotados ortogonalmente"), o si no por la palabra varimax o quartimax, que aparece en el titulo (éstas son técnicas para la rotación ortogonal).

2) Muchas caracteristicas de una matriz no rotada se canservan, en una matriz rotada ortogonalmente. Estas son sobre el número de factores que indican el número de influencias o causas; sobre la interpretación de las cargas factoriales; el porcentaje de varianza total y el porcentaje de varianza común.

3) Los valores de comunalidad ($h^2$) dados por los factores no rotados no cambian con la rotación ortogonal. De ahi que pueden ser dados ya sea con la matriz factorial rotada, o con la no rotada.

4) En la matriz no rotada los factores se ordenan, de acuerdo con la cantidad de variación de los datos de los que se da cuenta, definiendo el primero el mayor grado de relación en los datos. En

la matriz rotada ortogonalmente no se da importancia al orden de los factores.

5) Los factores no están relacionados.

6) Si la matriz rotada es oblicua, más que ortogonal, el título o descripción de la matriz lo indicará.

La matriz de factores rotados proporciona una estructura simple, obteniendo así cargas diferentes de cero, para algunas de las variables, y también cada variable puede tener cargas diferentes de cero para pocos factores. Esto permite diferenciar un factor de otro, y ayuda a una mejor interpretación de los factores.

Podemos ver que en la tabla 4.7, las cargas factoriales del factor 1 están mejor definidas, para las variables I1, I4 y I5, y en el factor 2 para las variables I1 y I3. Quiere decir que las cargas son altas para pocas variables. Esto no sucede en la matriz de factores de la tabla 4.4.

5. Matriz de correlación de factores.

La matriz de correlaciones de factores encontrados mediante la rotación ortogonal, es una matriz identidad. Esto es, hay 1's sobre la diagonal y 0's en otro lado.

En el caso de la matriz de correlaciones de factores oblicuos, obtenida por una rotación oblicua, es llamada a veces una matriz de cosenos de factores. Estos cosenos pueden ser interpretados como correlaciones entre factores, y viceversa.

Tabla 4.8. Matriz de correlación de factores
o transformación de factores.

| | FACTOR 1 | FACTOR 2 |
|---|---|---|
| FACTOR 1 | .86325 | .50478 |
| FACTOR 2 | -.50478 | .86325 |

Las caracteristicas de la matriz de correlaciones de las variables originales, descritas en el primer resultado factorial se aplican también aqui.

Las correlaciones distintas de cero entre dos factores, significa que las influencias o causas de los datos, tienen una relación hasta el punto medio por las correlaciones de los factores.

En ocasiones, la matriz de correlaciones de factores, puede ser a su vez sometida al análisis factorial, como lo fue la matriz de correlaciones de variables. Esto permitirá descubrir las influencias de relaciones entre los factores; la interpretación de esas influencias o causas no serán difernte de las halladas, para las correla c iones de variables. La reducción de las interrelaciones factoriales a sus factores, recibe el nombre de análisis factorial de orden superior.

6. Matriz de coeficientes de los puntajes factoriales.

La matriz factorial presenta las cargas a [coeficientes del modelo factorial (3.2)], mediante las que se puede conocer la existencia de una influencia o fenómeno para las variables. La matriz de puntajes factoriales da una puntuación para cada caso, a esas influencias representadas por los factores.

Los puntajes dan los valores para cada caso en las funciones F, de las ecuaciones (3.1) y (3.2) del punto 3.2. Al estar definida la constante a, por la matriz factorial y los puntajes factoriales que definen el valor de la función F, las ecuaciones factoriales quedan completamente especificadas.

En la tabla 4.9, se muestra la matriz de coeficientes de puntajes factoriales del ejemplo ilustrativo.

Tabla 4.9. Matriz de coeficientes de puntajes factoriales.

| | FACTOR 1 | FACTOR 2 |
|---|---|---|
| I1 | -.19982 | .65347 |
| I2 | .41132 | -.19294 |
| I3 | -.03205 | .49595 |
| I4 | .33160 | -.00671 |
| I5 | .34878 | -.03296 |

Esta matriz es el último de los resultados factoriales de un análisis factorial. El paquete SPSS/PC obtiene los puntajes, por medio del método de regresión.

Las variables representadas por los puntajes factoriales, pueden ser utilizadas con otros análisis o como un medio, para comparar casos sobre los factores. Pero los puntajes tienen ciertas características, que pueden no ser compartidas con otras variables. Encarnan fenómenos con una unidad funcional, fenómenos que están altamente interrelacionados en el tiempo o en el espacio.

Los puntajes están estandarizados, esto significa que han sido escalados de manera que tengan una media de cero, y que alrededor de dos tercios de los valores estén entre 1.00 y -1.00.

Los puntajes mayores que 1.00 o menores que 1.00 son, desusadamente altos o bajos.

Para definir los factores se sustituye el valor numérico del puntaje factorial de la tabla 4.9, en la expresión (3.4).

$$F_i = \sum_{j=1}^{p} W_{ij} X_j \qquad ; i = 1, \ldots, k$$

W's son los pesos factoriales.
X's son las variables originales.

Entonces según el ejemplo tenemos que:

$$F_1 = -.20(40.01) + .41(14.54) + (-.03)(14.06) + .33(36.04) + .35(17.89) = 15.69$$

$$F_2 = .65(40.01) + (-.19)(14.54) + .50(14.06) + (-.01)(36.04) + (-.03)(17.89) = 29.37$$

Ahora, ya calculados los factores para el ejemplo, podemos definir la relación de las funciones $F_1$ y $F_2$ mediante la ecuación del modelo factorial obtenido para este ejemplo:

$$Z_j = a_{j1}F_1 + a_{j2}F_2 \qquad \text{con } j = 1, 5. \qquad (4.2)$$

Considerando que se conocen las cargas factoriales (a) dadas en la matriz de la tabla 4.4.

Siguiendo el ejemplo, los indicadores o variables socio-económicas son expresadas en términos de los factores. esto es:

$$I1 = .52(15.69) + .78(29.37)$$

$$I2 = .85(15.69) + (-.44)(29.37)$$

$I3 = .74(15.69) + .52(29.37)$

$I4 = .94(15.69) + (-.20)(29.37)$

$I5 = .94(15.69) + (-.24)(29.37)$

Así, las cargas y los puntajes factoriales las cuales son encontradas a través del análisis factorial, describen la puntuación de los datos, haciendo posible expresar las relaciones de las variables en función de pocos factores, utilizando las ecuaciones del modelo factorial.

Después de encontrar los factores que representan, las influencias o causas y fenómenos de relación entre las variables, el investigador los estudiará y les pondrá una etiqueta adecuada. Dichas etiquetas facilitan la comunicación y discusión de los resultados.

# C A P I T U L O V

## UN CASO PRÁCTICO Y ANÁLISIS DE RESULTADOS

## "ÍNDICES DE MARGINACIÓN"

En este capítulo, finalmente se llega a poner en práctica el método o técnica estadística del análisis multivariado llamada análisis factorial. Que en los cuatro capítulos anteriores se desarrolló su teoría matemática en la cual se basa el modelo factorial que lo definen; y sus diversas aplicaciones en diferentes áreas. Así, también se describió el uso de un paquete computarizado, para utilizar dicha técnica, y cómo analizar los resultados.

Para efectos de utilizar el análisis factorial en un problema real, se plantea obtener el "índice de marginación", para las 32 entidades federativas de la República Mexicana. Con información socio-económica del Censo General de Población y Vivienda 1980.

En los puntos siguientes se desarrolla este caso práctico.

### 5.1. Planteamiento del Problema.

El problema consiste en el cálculo de los índices de marginación a nivel de entidades federativas. Son 32 entidades (unidades de estudio, casos o muestra).

Los indicadores de las necesidades esenciales, utilizadas en la definición del índice de marginación son 19 (variables).

Concepto de marginación.

La marginación caracteriza aquellos grupos que han quedado al margen de los beneficios del desarrollo nacional y de los beneficios de la riqueza generada, pero no necesariamente al margen de la generación de riqueza ni mucho menos de las condiciones que la hacen posible.

Asimismo, se aclara que si bien parte de la población marginada habita regiones de ecologia adversa, esto ha sido resultado de un largo proceso histórico en el cual la población ha sido despojada de las mejores tierras, y que la insastisfacción de las necesidades esenciales en México es debido a un resultado del bajo nivel de productividad y a la concentración de la riqueza social prevaleciente. El producto social global, adecuadamente distribuido, permitiria la satisfacción de las necesidades primarias de toda la población, y un ritmo adecuado de crecimiento del aparato productivo.

De esta manera, la satisfacción de dichas necesidades esenciales de los distintos grupos sociales, depende de su inserción en el sistema económico, de su organización y de su fuerza politica.[1]

Asi, la insastisfacción de las necesidades esenciales se localiza principalmente entre los grupos sociales que:

a) Están desposeidos de medios de producción, poseen únicamente su capacidad de trabajo no calificada y no han logrado encontrar trabajo, permanente en las ramas más productivas de la economia en condiciones de sindicalismo organizado; esto es, los trabajadores no organizados que trabajan fuera de las empresas de mayor productividad o que se encuentran desempleados o subempleados. Este gran grupo puede ser clasificado en las

[1] Coordinación General del Plan Nacional de Zonas Deprimidas y Grupos Marginados. Geografía de la Marginación. Volumen 6, México, Coplamar Siglo XXI, 1982, pp. 4-5.

siguientes subcategorías, todas ellas referidas a fuerzas de trabajo con bajo grado de calificación:

i) Jornaleros agropecuarios permanentes o eventuales.

ii) Asalariados eventuales urbanos, v. gr. en la industria de la construcción.

iii) Asalariados permanentes de actividades económicas urbanas de baja productividad: parte del comercio, industria tradicional y algunos servicios.

iv) Asalariados no organizados urbanos v.gr. del auto transporte urbano y suburbano.

v) Desempleados permanentes o temporales.

b) Poseen medios de producción (en forma individual o colectiva) de mala calidad o de carácter tradicional:

i) La inmensa mayoría de los campesinos de zonas de temporal que trabajan con técnicas tradicionales, y a los que parte de su excedente les es arrebatado por intermediarios agiotistas, caciques, etc.

ii) Población indígena que, despojada a lo largo de los siglos de las mejores tierras, vive ahora en las zonas de refugio.

iii) Artesanos debilitados por la competencia industrial y los oficiales y aprendices que les auxilian.

iv) Parte de los pequeños comerciantes.

v) Prestadores de servicios de baja productividad que requieren, empero, algunos medios de producción. Los vendedores ambulantes, lavadores y cuidadores de coches, etc. deben considerarse como parte de la población subocupada.

Como se puede apreciar, los marginados de los beneficios del desarrollo son principalmente los desempleados y subempleados, parte de la población asalariada ocupada, y por último, una parte de la población que labora por su cuenta en niveles familiar o comunal.

El problema de la marginación no es necesariamente un resultado del insuficiente desarrollo horizontal-geográfico y por ramas de actividad de la sociedad de mercado. Esta genera sus propias formas de marginación al tiempo que establece nuevas relaciones de desigualdad, con aquellos grupos marginados que no se ubican directamente en las relaciones de trabajo asalariado.

En México, el Estado ha reconocido estos hechos. Las garantías sociales contenidas en la Constitución son una manera de atemperar la vigilancia de las leyes económicas en una sociedad de mercado. La intervención del Estado en la economía y en la sociedad ha estado orientada a impulsar, el desarrollo tecnológico y a regular la distribución de los beneficios entre la población. Entre otras maneras de influir en este último aspecto, destacan la prestación gratuita de ciertos servicios (educación, asitencia social), la creación y apoyo a las instituciones de seguridad social, la reforma agraria, la política salarial y de control de precios, la creación de Infonavit, etc.

Sin embargo, buena parte de estas acciones están dirigidas principalmente al beneficio de la población asalariada urbana, y solamente la educación y la atención médica de la asistencia social se proporcionan a cualquier persona, independientemente de la clase social a la que pertenezca.[2]

Con la finalidad de econtrar un indicador, que englobe los diversos elementos de necesidades esenciales sociales consideradas. Se realiza este ejercicio estadístico, en la

[2] Op. cit. Geografía de la marginación p. 22, 23 y 24.

construcción del índice compuesto del nivel de vida de la población de los estados, con base en el método de componentes principales. La aplicación del análisis factorial dentro de este problema sería la formación de escalas (ver punto 1.2), esto es, la ponderación de las características que se combinan, ofreciendo una solución dividiendo las características en fuentes independientes de variación (factores). Así, el método proporciona una jerarquización para cada entidad, permitiendo ordenarlas de acuerdo a su nivel relativo de marginación.

La intención es vincular la posición relativa de marginación, entre las entidades federativas directamente con el grado de avance también relativo, que presentan en el desarrollo de sus actividades económicas. Permitiendo así, detectar entre otras cosas, el crecimiento o rezago de los estados en forma muy general, posibilitando la toma de decisiones orientadas hacia un desarrollo sano de las entidades.

Esto podría ser mediante el desarrollo de un sistema de indicadores macroeconómicos y microeconómicos de coyuntura; esto es, seleccionar y calcular continuamente los idicadores que están altamente relacionados con un alto índice de marginalidad, y así, poderlos evaluar y darles seguimiento dentro de su comportamiento económico en la entidad, para encontrar una solución y superar el problema.

## 5.2. Estructuración y Recolección de Datos.

La estructuración y selección de las variables a utilizar en el ejemplo para el análisis factorial, se hace de la siguiente manera.

En consideración con lo señalado al concepto de marginación, se considera que ésta se expresa en: 1) bajos niveles de ingreso de la población económicamente activa; 2) altos niveles de subempleo; 3) altos porcentajes de población rural agrícola; 4) alimentación adecuada; 5) bajos niveles de escolaridad; 6) bajos niveles de salud; y 7) viviendas inadecuadas y sin servicios.

Para medir estas características se construyeron 19 indicadores divididos con propósito de ordenamiento en: generales, de alimentación, de educación, de salud, de vivienda y sus servicios, y población indígena.

a) Generales.

1. Porcentaje de población económicamente activa (PEA) que recibe ingresos inferiores a 3611 pesos mensuales.[3]

2. Porcentaje de población económicamente activa (PEA) que no percibió ingrsos.

3. Porcentaje de población económicamente activa (PEA) subempleada.[4]

4. Porcentaje de población rural.[5]

[3] Ingresos inferiores al salario mínimo aproximado de esa fecha.

[4] Aquella que trabaja 32 horas o menos a la semana.

[5] La que habita localidades menores de 10,000 habitantes.

5. Porcentaje de población económicamente activa en el sector agropecuario.

b) Alimentación.

6. Porcentaje de población de cinco años de edad y menos, que consume carne tres o menos dias a la semana.

7. Porcentaje de población de cinco años de edad y menos, que consume huevo tres o menos dias a la semana.

8. Porcentaje de población de cinco años de edad y menos, que consume leche tres o menos dias a la semana.

c) Educación.

9. Porcentaje de población analfabeta de 10 años y más.[6]

10. Porcentaje de población de 15 años de edad y más sin primaria completa.[7]

d) Salud.

11. Tasa de mortalidad general.[8]

12. Tasa de motalidad preescolar.[9]

[6] Se consideró a la población que no sabe leer ni escribir de 10 y 14 años de edad, y a la población analfabeta de 15 años y más.

[7] Se tomó a la población de 15 años y más con enseñanza primaria, desde primero hasta quinto grado.

[8] Defunciones totales dividido entre población total, por 1000.

[9] Defunciones de niños de 1 a 4 años de edad entre población total respectiva por 1000.

e) Vivienda.

13. Porcentaje de viviendas que usa como combustible, para cocinar leña y petróleo.

14. Porcentaje de viviendas con un dormitorio con cinco y más ocupantes.

15. Porcentaje de viviendas sin agua entubada.

16. Porcentaje de viviendas de uno y dos cuartos.

17. Porcentaje de viviendas sin electricidad.

18. Porcentaje de viviendas sin drenaje.

f) Población indígena.

19. Porcentaje de población de cinco años de edad y más, que habla lengua indígena.

La relación detallada de los indicadores empleados, así como la forma de recolectarlos y construirlos, puede ser consultada en el anexo I página 211.

Los indicadores utilizados en la elaboración del índice de marginación tienen la siguiente fuente: X censo General de Población y Vivienda, 1980 y Anuario Estadístico de los Estados Unidos Mexicanos, 1983 y 1985.

Cabe mencionar que si bien los indicadores se refieren a 1980, el análisis que se realiza con ellos es de carácter estructural, ya que el ritmo de cambio de las variables consideradas es gradual y por consiguiente, el panorama del desequilibrio social interestatal que se obtiene, tiende a mantenerse por espacios prolongados de tiempo.

El cuadro 5.1 contiene los indicadores de la pobreza nacional a nivel estado (1980), clasificados anteriormente.

Cuadro 5.1. Indicadores de la Pobreza Nacional.
Nivel Estatal (Datos 1980).

| | | BAJOS INGRESOS | PEA SIN INGRESOS | SUB- EMPLEO | POBL. RURAL | POBL. AGRICOLA |
|---|---|---|---|---|---|---|
| CLAVE EDO. | ENTIDAD FEDERATIVA | % I1 | % I2 | % I3 | % I4 | % I5 |
| 1 | AGUASCALIENTES | 40.01 | 14.54 | 14.06 | 36.04 | 17.89 |
| 2 | BAJA CALIFORNIA | 14.48 | 9.64 | 11.65 | 22.28 | 9.47 |
| 3 | BAJA CALIFORNIA SUR | 17.33 | 10.37 | 12.99 | 46.54 | 19.35 |
| 4 | CAMPECHE | 31.11 | 15.47 | 16.56 | 47.13 | 31.87 |
| 5 | COAHUILA | 30.38 | 8.57 | 13.99 | 29.19 | 15.78 |
| 6 | COLIMA | 34.55 | 10.23 | 14.40 | 41.77 | 27.85 |
| 7 | CHIAPAS | 25.28 | 39.07 | 14.18 | 79.68 | 57.43 |
| 8 | CHIHUAHUA | 20.95 | 18.99 | 13.26 | 33.93 | 20.75 |
| 9 | DISTRITO FEDERAL | 15.14 | 7.96 | 12.63 | 00.00 | 6.13 |
| 10 | DURANGO | 28.74 | 25.74 | 16.42 | 61.66 | 30.88 |
| 11 | GUANAJUATO | 29.86 | 25.88 | 19.49 | 49.17 | 19.17 |
| 12 | GUERRERO | 20.65 | 37.02 | 18.99 | 70.97 | 44.28 |
| 13 | HIDALGO | 37.87 | 24.78 | 15.48 | 78.89 | 37.03 |
| 14 | JALISCO | 23.98 | 19.21 | 17.05 | 35.60 | 18.22 |
| 15 | MEXICO | 19.31 | 14.94 | 14.44 | 37.61 | 15.26 |
| 16 | MICHOACAN | 31.45 | 28.21 | 19.86 | 62.14 | 39.45 |
| 17 | MORELOS | 32.58 | 16.99 | 17.77 | 60.36 | 25.11 |
| 18 | NAYARIT | 42.97 | 15.86 | 18.25 | 60.94 | 40.35 |
| 19 | NUEVOLEON | 17.62 | 9.70 | 10.74 | 18.25 | 8.37 |
| 20 | OAXACA | 22.59 | 43.59 | 13.65 | 83.78 | 55.32 |
| 21 | PUEBLA | 34.08 | 27.22 | 15.67 | 63.59 | 41.37 |
| 22 | QUERETARO | 36.17 | 20.48 | 11.90 | 61.86 | 28.98 |
| 23 | QUINTANA ROO | 25.31 | 16.47 | 12.80 | 51.75 | 29.16 |
| 24 | SAN LUIS POTOSI | 35.49 | 24.29 | 15.23 | 62.44 | 34.08 |

Continúa el cuadro 5.1.

| | I1 | I2 | I3 | I4 | I5 |
|---|---|---|---|---|---|
| 25 SINALOA | 23.28 | 17.83 | 15.92 | 56.92 | 27.54 |
| 26 SONORA | 19.12 | 11.87 | 11.97 | 39.35 | 20.81 |
| 27 TABASCO | 33.37 | 17.17 | 13.91 | 72.10 | 38.92 |
| 28 TAMULIPAS | 27.16 | 13.23 | 14.44 | 30.94 | 17.99 |
| 29 TLAXCALA | 32.83 | 25.84 | 17.05 | 71.54 | 37.67 |
| 30 VERACRUZ | 32.80 | 20.83 | 13.53 | 62.95 | 37.75 |
| 31 YUCATAN | 40.41 | 14.57 | 20.35 | 46.13 | 31.36 |
| 32 ZACATECAS | 35.77 | 39.77 | 13.89 | 73.62 | 49.33 |
| MEDIA | 28.52 | 20.20 | 15.08 | 51.53 | 29.22 |
| DESV. ESTANDAR | 7.85 | 9.52 | 2.52 | 19.71 | 13.07 |
| PONDERACION | 0.47 | 0.83 | 0.35 | 0.93 | 0.95 |

| | SUBCONS. CARNE | SUBCON. HUEVO | SUBCONS. LECHE | ANALFA-BETISMO | POBL. S/PRIM | MORT. GRAL. P/MIL | MORT. PRESC. P/MIL |
|---|---|---|---|---|---|---|---|
| CLAVE EDO. | % I6 | % I7 | % I8 | % I9 | % I10 | % I11 | % I12 |
| 1 | 65.81 | 34.83 | 13.43 | 9.57 | 32.74 | 6.54 | 2.90 |
| 2 | 53.30 | 14.52 | 6.53 | 5.85 | 22.63 | 5.75 | 1.80 |
| 3 | 56.86 | 21.00 | 11.58 | 6.82 | 29.37 | 5.63 | 1.50 |
| 4 | 62.68 | 43.51 | 20.63 | 15.06 | 31.42 | 5.45 | 2.60 |
| 5 | 60.54 | 16.52 | 15.16 | 6.94 | 27.63 | 6.32 | 2.10 |
| 6 | 65.40 | 46.52 | 15.42 | 11.20 | 33.09 | 7.07 | 4.50 |
| 7 | 66.77 | 46.37 | 35.39 | 34.61 | 31.61 | 6.04 | 6.10 |
| 8 | 63.05 | 19.41 | 14.59 | 7.98 | 30.32 | 6.66 | 3.20 |
| 9 | 49.23 | 24.67 | 8.85 | 5.24 | 17.40 | 5.64 | 1.40 |
| 10 | 65.81 | 20.91 | 21.06 | 8.34 | 36.76 | 5.44 | 2.00 |
| 11 | 61.15 | 38.73 | 22.27 | 21.36 | 29.05 | 7.30 | 4.40 |

Continúa el cuadro 5.1.

| | I6 | I7 | I8 | I9 | I10 | I11 | I12 |
|---|---|---|---|---|---|---|---|
| 12 | 65.22 | 43.31 | 36.87 | 31.22 | 24.31 | 6.06 | 5.00 |
| 13 | 67.54 | 47.59 | 29.68 | 25.91 | 29.55 | 8.42 | 5.00 |
| 14 | 58.94 | 36.12 | 10.96 | 11.75 | 29.36 | 6.56 | 2.90 |
| 15 | 58.65 | 35.16 | 18.94 | 11.86 | 25.01 | 6.24 | 3.20 |
| 16 | 63.91 | 45.59 | 23.37 | 22.56 | 30.69 | 7.07 | 3.70 |
| 17 | 62.71 | 32.18 | 23.14 | 14.53 | 25.38 | 5.93 | 2.70 |
| 18 | 71.22 | 48.05 | 24.64 | 13.84 | 38.23 | 4.99 | 2.40 |
| 19 | 57.61 | 16.35 | 12.97 | 6.31 | 23.77 | 4.84 | 1.80 |
| 20 | 65.27 | 46.83 | 40.40 | 31.40 | 31.31 | 10.12 | 13.00 |
| 21 | 64.90 | 44.07 | 30.40 | 23.74 | 28.36 | 9.38 | 7.90 |
| 22 | 73.18 | 46.25 | 32.03 | 22.72 | 30.60 | 7.41 | 4.00 |
| 23 | 57.48 | 32.51 | 19.25 | 15.11 | 34.91 | 5.03 | 3.10 |
| 24 | 67.65 | 31.09 | 26.87 | 18.93 | 32.71 | 5.63 | 3.40 |
| 25 | 61.73 | 18.58 | 19.46 | 11.88 | 32.98 | 5.01 | 2.10 |
| 26 | 58.74 | 12.23 | 11.60 | 7.45 | 30.22 | 6.31 | 2.10 |
| 27 | 64.19 | 46.39 | 25.86 | 16.44 | 41.41 | 7.08 | 4.50 |
| 28 | 59.76 | 17.61 | 16.00 | 8.53 | 29.85 | 5.90 | 2.20 |
| 29 | 65.46 | 42.73 | 27.28 | 14.24 | 28.51 | 5.73 | 3.20 |
| 30 | 62.12 | 36.41 | 24.18 | 21.21 | 31.05 | 6.00 | 4.00 |
| 31 | 56.64 | 38.26 | 20.87 | 17.36 | 34.26 | 7.38 | 4.80 |
| 32 | 81.08 | 42.37 | 26.94 | 12.69 | 47.24 | 5.70 | 2.60 |
| MEDIA | 62.96 | 33.96 | 21.46 | 15.39 | 30.68 | 6.39 | 3.63 |
| DES.EST. | 5.94 | 11.92 | 8.38 | 8.00 | 5.54 | 1.21 | 2.23 |
| PONDERA. | 0.65 | 0.77 | 0.95 | 0.89 | 0.46 | 0.48 | 0.74 |

Continúa el cuadro 5.1.

| | VIVIEND. COMBUS. LEÑA Y P. | VIVIEND. 1 DORM. 5+ OCUP. | VIVIEND. S/AGUA | VIVIEND. 1 Y 2C. | VIVIEND. S/ELEC. | VIVIEND. S/DREN. | POBL. INDIG. |
|---|---|---|---|---|---|---|---|
| CLAVE EDO. | % I13 | % I14 | % I15 | % I16 | % I17 | % I18 | % I19 |
| 1 | 21.12 | 13.47 | 11.55 | 43.59 | 14.17 | 26.21 | 1.29 |
| 2 | 9.03 | 10.16 | 21.25 | 47.24 | 9.16 | 34.10 | 2.08 |
| 3 | 33.67 | 12.83 | 21.97 | 52.55 | 21.45 | 52.61 | 2.10 |
| 4 | 39.68 | 27.92 | 39.52 | 69.63 | 21.00 | 54.64 | 21.53 |
| 5 | 19.87 | 16.03 | 14.52 | 52.41 | 11.78 | 40.85 | 1.44 |
| 6 | 24.31 | 23.87 | 14.84 | 70.81 | 15.67 | 38.62 | 1.33 |
| 7 | 70.16 | 37.65 | 54.92 | 73.71 | 51.50 | 65.53 | 28.50 |
| 8 | 27.85 | 16.41 | 21.39 | 52.25 | 22.65 | 42.75 | 3.92 |
| 9 | 5.45 | 15.67 | 6.38 | 49.06 | 1.46 | 13.79 | 2.67 |
| 10 | 46.22 | 18.07 | 25.95 | 52.50 | 21.60 | 56.25 | 1.94 |
| 11 | 44.10 | 18.17 | 30.21 | 53.71 | 24.30 | 45.87 | 1.38 |
| 12 | 59.86 | 36.02 | 50.61 | 73.75 | 38.60 | 64.97 | 15.30 |
| 13 | 61.41 | 28.93 | 40.24 | 66.30 | 39.97 | 63.89 | 23.13 |
| 14 | 21.54 | 13.11 | 20.83 | 47.95 | 16.01 | 29.45 | 1.74 |
| 15 | 23.29 | 22.02 | 17.45 | 54.29 | 11.04 | 29.65 | 5.59 |
| 16 | 46.46 | 21.70 | 32.32 | 63.70 | 28.04 | 51.44 | 4.61 |
| 17 | 31.38 | 24.22 | 21.81 | 64.65 | 12.44 | 45.35 | 3.85 |
| 18 | 28.30 | 26.14 | 28.24 | 72.09 | 20.02 | 62.78 | 3.90 |
| 19 | 19.17 | 17.67 | 12.07 | 51.14 | 9.34 | 30.88 | 1.37 |
| 20 | 76.37 | 36.11 | 54.17 | 75.20 | 47.63 | 71.88 | 44.01 |
| 21 | 56.79 | 32.17 | 38.55 | 66.42 | 27.94 | 54.96 | 17.24 |
| 22 | 52.99 | 24.57 | 33.80 | 60.80 | 36.89 | 59.38 | 3.62 |
| 23 | 42.50 | 27.14 | 40.76 | 73.72 | 26.28 | 56.83 | 44.09 |
| 24 | 59.37 | 22.86 | 48.59 | 59.12 | 42.42 | 59.70 | 13.59 |
| 25 | 37.43 | 23.51 | 31.92 | 61.35 | 21.33 | 56.73 | 2.41 |
| 26 | 24.71 | 17.22 | 16.48 | 50.99 | 18.91 | 46.22 | 4.68 |
| 27 | 44.43 | 31.89 | 57.13 | 72.61 | 42.13 | 51.79 | 6.31 |

Continúa el cuadro 5.1.

| | I13 | I14 | I15 | I16 | I17 | I18 | I19 |
|---|---|---|---|---|---|---|---|
| 28 | 19.17 | 21.34 | 27.57 | 62.73 | 21.38 | 43.48 | 1.77 |
| 29 | 51.06 | 30.75 | 27.79 | 66.39 | 15.50 | 63.80 | 6.51 |
| 30 | 47.28 | 29.08 | 48.07 | 66.80 | 35.31 | 51.34 | 13.66 |
| 31 | 46.09 | 24.59 | 48.45 | 65.15 | 17.77 | 51.41 | 53.15 |
| 32 | 48.19 | 19.92 | 41.82 | 55.73 | 38.78 | 70.31 | 0.55 |
| MEDIA | 38.73 | 23.16 | 31.29 | 60.88 | 24.45 | 49.61 | 10.60 |
| DES.EST. | 17.55 | 7.26 | 14.25 | 9.34 | 12.45 | 13.87 | 14.00 |
| PONDERA. | 0.94 | 0.86 | 0.88 | 0.78 | 0.87 | 0.86 | 0.56 |

FUENTE: ELABORACION PROPIA CON DATOS DEL X CENSO GENERAL DE POBLACION Y VIVIENDA, 1980. RESUMEN GENERAL Y POR ENTIDAD FEDERATIVA. ANUARIO ESTADISTICO DE LOS ESTADOS UNIDOS MEXICANOS 1983 Y 1985.

5.3. Aplicación de Componentes Principales.

Dentro del análisis factorial existen varios métodos, para obtener los factores iniciales o matriz de factores. Estos métodos fueron mencionados en términos muy generales en el punto 1.1 del capitulo I. (Factor Pincipal, Componentes Principales, Máxima Verosimilitud, etc.). Especificamente el método de análisis de componentes principales tratado ampliamente en el capitulo III punto 3.6, es de gran interés, puesto que es parte del objetivo del desarrollo de este trabajo de tesis.

Asi, continuando con la aplicación práctica de esta técnica estadistica multivariada. Las 19 variables y 32 casos del cuadro 5.1, son sometidos a la subrutina FACTOR del paquete SPSS/PC, ésta procesa los datos de las variables utilizando el método de componentes principales y la rotación varimax.

Después de ejecutarse el procesamiento FACTOR, se obtienen los resultados factoriales analizados en el punto 4.4. Pero para el análisis en particular de este ejemplo, sólo se presentan dos cuadros de resultados factoriales: la matriz de factores iniciales y la tabla de estadisticas finales.[10]

Estos cuadros factoriales son elegidos, porque ellos contienen la estructura de los factores que están interrelacionados con los indicadores socio-económicos. Uno de estos factores será seleccionado de acuerdo a las cargas presentadas para cada variable. Por otro lado, el segundo resultado factorial es presentado para observar las estadisticas especificas sobre los factores extraidos, y asi poder evaluar de alguna forma el orden y la reguladidad en los datos utilizados.

[10] Para examinar todos los resultados factoriales del análisis factorial, del ejmplo práctico desarrollado en este capitulo, consultar el anexo II página 221, que contiene el listado de la corrida con el paquete SPSS/PC de las variables analizadas.

Tabla 5.2. Matriz de factores iniciales.

| | FACTOR 1 | FACTOR 2 | FACTOR 3 | FACTOR 4 |
|---|---|---|---|---|
| I1 | .47115 | .56190 | .53848 | .05688 |
| I2 | .83037 | -.02231 | -.20885 | .23521 |
| I3 | .35142 | .15407 | .67494 | -.21166 |
| I4 | .92957 | .19211 | -.08916 | .00980 |
| I5 | .95257 | .11482 | -.08682 | -.04661 |
| I6 | .64782 | .63259 | -.08418 | .31827 |
| I7 | .77269 | .07440 | .42716 | .09731 |
| I8 | .94973 | -.06079 | -.03049 | .06240 |
| I9 | .89310 | -.29292 | .06379 | .05311 |
| I10 | .46574 | .69690 | -.11453 | -.03653 |
| I11 | .48591 | -.43642 | .28563 | .59820 |
| I12 | .74474 | -.49874 | .07792 | .31229 |
| I13 | .94302 | -.08271 | -.12900 | .02998 |
| I14 | .86164 | -.22063 | .03549 | -.25910 |
| I15 | .87646 | -.06935 | -.15375 | -.24514 |
| I16 | .78338 | -.11754 | .12410 | -.43430 |
| I17 | .87134 | .01972 | -.35451 | .09736 |
| I18 | .85685 | .22065 | -.24519 | -.15591 |
| I19 | .56318 | -.46633 | .08642 | -.36920 |

Esta tabla muestra los cuatro factores iniciales, extraidos por el método de análisis de componentes principales, los cuales explican el 84.3% de la variación total de las variables analizadas. Estos cuatro factores estructuran el modelo factorial en la siguiente forma. Según Ec. (3.2).

$$Z_i = a_{i1}F1 + a_{i2}F2 + a_{i3}F3 + a_{i4}F4 \qquad (5.1)$$

$$(i = 1, 2, \ldots, 19)$$

Se observa que el factor 1, es el que especifica el patrón de cambio y de viraje en importancia de las variables, puesto que catorce de las 19 variables sometidas al análisis presentan una alta correlación con este factor. Esto es, el factor 1 agrupa la mayoría de las variables usadas en el análisis, y éstas tienen una cierta característica en común. El factor 1 explica el 59.8% de la variación total, lo que representa una alta proporción para una sola componente y un reflejo de la bondad del sistema de indicadores utilizados para medir la marginación a nivel entidad federativa.

Por lo tanto este primer componente se considera el vector de ponderación.

Tabla 5.3. Estadísticas finales.

| Variable | Comunalidad | Factor | valor propio (Eigenvalor) | Pct. de Var. | Pct.Acum. |
|---|---|---|---|---|---|
| I1 | .83091 | 1 | 11.35464 | 59.8 | 59.8 |
| I2 | .78896 | 2 | 2.15080 | 11.3 | 71.1 |
| I3 | .64758 | 3 | 1.35014 | 7.1 | 78.2 |
| I4 | .90905 | 4 | 1.16664 | 6.1 | 84.3 |
| I5 | .93028 | | | | |
| I6 | .92823 | | | | |
| I7 | .79452 | | | | |
| I8 | .91051 | | | | |
| I9 | .89031 | | | | |
| I10 | .71703 | | | | |
| I11 | .86600 | | | | |
| I12 | .90697 | | | | |
| I13 | .91366 | | | | |
| I14 | .85950 | | | | |
| I15 | .85672 | | | | |
| I16 | .83152 | | | | |
| I17 | .89477 | | | | |
| I18 | .86729 | | | | |
| I19 | .67841 | | | | |

En la tabla se muestra la comunalidad para cada variable y las estadísticas para cada factor, después de seleccionar los factores extraidos. La comunalidad nos da la medida del grado en que la variable es explicada por los factores comunes. Los valorespropios dados en la columna cuatro de la tabla 5.3 da la varianza total explicada por cada factor. Vemos que el valor del valor propio para el primer factor (11.35) es el más alto, que el de los otros tres factores, este es otro indicador para seleccionar el factor 1 y obtener el peso o la magnitud de cada indicador socio-económico, combinados independientemente en este factor. La siguiente columna contiene el porcentaje de la varianza total atribuible a cada factor.

El factor 1 cuenta con el 59.8% de la variación total. Esto significa que el primer factor explica la mayor parte de la varianza de las variables dento de la muestra.

A continuación se da una relación de las ponderaciones obtenidas mediante el factor 1, con los indicadores socio-económicos.

Cuadro 5.4. Ponderación de los indicadores de marginación a nivel estatal, 1980.

| Indicadores | Ponderaciones |
|---|---|
| 1. Porcentaje de PEA que percibe ingresos - inferiores a 3,611.00 pesos mensuales. | .47 |
| 2. Porcentaje de PEA sin ingreso. | .83 |
| 3. Porcentaje de PEA que labora desde menos de una hora hasta 32 horas a la semana. | .35 |
| 4. Porcentaje de población rural. | .93 |
| 5. Porcentaje de PEA en el sector agropecuario. | .95 |
| 6. Porcentaje de población de 5 años y menos que consume carne 3 ó menos días a la semana. | .65 |

Continúa el cuadro 5.4.

| Indicadores | Ponderadores |
|---|---|
| 7. Porcentaje de población de 5 años y menos que consume huevo 3 ó menos dias a la semana. | .77 |
| 8. Porcentaje de población de 5 años y menos que consume leche 3 ó menos dias a la semana. | .95 |
| 9. Porcentaje de analfabetas de 10 años y más de edad. | .89 |
| 10. Porcentaje de población de 15 años y más sin primaria completa. | .46 |
| 11. Tasa de mortalidad general. | .48 |
| 12. Tasa de motalidad prescolar. | .74 |
| 13. Porcentaje de viviendas que usan como combustible para cocinar leña y petróleo. | .94 |
| 14. Porcentaje de viviendas con un dormitorio con 5 y más ocupantes. | .86 |
| 15. Porcentaje de viviendas sin agua entubada. | .88 |
| 16. Porcentaje de viviendas de uno y dos cuartos. | .78 |
| 17. Porcentaje de viviendas sin eletricidad. | .87 |
| 18. Porcentaje de viviendas sin drenaje. | .86 |
| 19. Porcentaje de población de cinco años y más que habla lengua indigena. | .56 |

Fuente: Elaboración propia a partir de los 19 indicadores del cuadro 5.1, y el método estadistico análisis factorial (componentes principales).

## 5.4. Análisis de Resultados de los Índices de Marginación.

Después de obtener los valores de ponderación para cada variable (19 indicadores). Ahora podemos utilizarlos en la construcción de un índice que represente el comportamiento ordenado de las variables.

Aún y cuando las variables están expresadas en pocentajes, es necesario estandarizarlas con base en sus medias y sus desviaciones estándar, esto es:

$$Z_i = \frac{I_i - \bar{X_i}}{\sigma_i} \quad ; \; i = 1, 2, \ldots, 19 \qquad (5.2)$$

$I_i$ es el indicador i-ésimo.

Los valores estandarizados se ponderan con los del componente principal F1 seleccionado. Así, el índice de marginación queda expresado como:

$$IMJ = \sum_{i=1}^{n} F\,p_i \frac{X_{ij} - \bar{X}_i}{d_i} \quad ; \; j = 1, 2, \ldots, 32 \qquad (5.3)$$

Donde:

$IMJ$ = Indice de marginación del estado j.

$X_{ij}$ = Valor de la variable i, del estado j.

$\bar{X}_i$ = Valor de la media de la variable i.

$d_i$ = Desviación estándar de la variable i.

$Fp_i$ = Factor de ponderación de la variable i.

Al hacer todos los cáculos pertinentes, para elaborar el índice de marginación de acuerdo a la expresión (5.3). Tenemos como resultado el cuadro 5.5, donde se expresa el índice de marginación para cada entidad federativa, y agrupadas por estratos según su indice.

Cuadro 5.5. Indices de marginación de las entidades federativas agrupadas por estratos, 1980.

| Entidades Federativas | Indice de Marginación | Estratos | Grado de Marginación |
|---|---|---|---|
| Oaxaca | 25.12 | | |
| Chiapas | 19.94 | | |
| Guerrero | 14.76 | | |
| Hidalgo | 12.92 | 9.5 y más | Muy alto |
| Puebla | 11.52 | | |
| Zacatecas | 10.31 | | |
| Tabasco | 9.68 | | |
| | | | |
| Querétaro | 7.23 | | |
| San Luis Potosi | 6.77 | | |
| Veracruz | 5.96 | | |
| Michoacán | 5.62 | | |
| Tlaxcala | 5.26 | 0.0 a 9.49 | Alto |
| Yucatán | 5.02 | | |
| Nayarit | 4.62 | | |
| Campeche | 2.10 | | |
| Quintana Roo | 2.08 | | |
| Guanajuato | 0.12 | | |
| | | | |
| Durango | -1.87 | | |
| Morelos | -2.44 | | |
| Sinaloa | -2.95 | | |
| Colima | -3.00 | -0.01 a -9.49 | Medio |
| Tamaulipas | -8.39 | | |
| Chihuahua | -8.40 | | |
| México | -9.27 | | |
| | | | |
| Jalisco | -9.85 | -9.5 a -18.99 | Medio bajo |

Contiúa el cuadro 5.5.

| Entidades Federativa | Índice de Marginación | Estratos | Grado de Marginación |
|---|---|---|---|
| Aguascalientes | -10.00 | | |
| Baja California Sur | -10.62 | | |
| Sonora | -11.40 | -9.5 a -18.99 | Medio bajo |
| Coahuila | -12.33 | | |
| Nuevo León | -17.07 | | |
| Baja California | -18.94 | | |
| Distrito Federal | -22.49 | -19 y menos | Bajo |

Fuente: Elaboración Propia con base en los indicadores del X Censo de Población y Vivienda, 1980 y el método estadístico.

El índice de marginación de una unidad determinada no representa un valor absoluto de la marginalidad, sino la posición que ésta guarda en relación al resto de las unidades a partir de la situación, relativa que presentó el conjunto de los diecinueve indicadores.

El índice de marginación fue calculado de tal forma que, un alto índice indica una mayor posición de marginación para la entidad federativa, y un bajo índice significa menor posición de marginalidad entre los estados.

Interpretación de los resultados.

Los indicadores que determinaron en mayor medida el que una entidad presentara índices de marginación elevados, de acuerdo a los ponderadores de cada variable, fueron: población rural, población económicamente activa que trabaja en el sector agropecuario, subconsumo de leche, las condiciones y servicios de la vivienda y por último, las tasas de analfabetismo (Ver cuadro 5.4).

De acuerdo al ordenamiento de las entidades federativas (Ver cuadro 5.5, ‹Índices de marginación de las entidades›) en cinco estratos de marginación muy alta, alta, media, media baja y baja, destacan por observar situaciones de muy alta marginalidad los estados de Oaxaca, Chiapas, Guerrero, Hidalgo, Puebla, Zacatecas y Tabasco. Por el contrario, con excepción del Distrito federal, las entidades con niveles de marginación media baja eran Jalisco, Aguascalientes, Baja Clifornia Sur, Sonora, Coahuila, Nuevo León y Baja California.

Las siete entidades que comprenden el estrato de muy alta marginalidad, albergan a una población de 13,658,275 habitantes, el 20.43% del total del país. Las diez siguientes entidades, que corresponden al estrato de marginalidad alta, representaban el 24.98% de la población, con 16,669,100 habitantes. El nivel medio de marginación con siete estados, tenía una población de 15,819,877 habitantes, con el 23.66% de la población nacional. Los siete estados con niveles de marginación medio bajo, estaban poblados por 11,868,502 habitantes, el 17.75% del total. Por último, el estrato bajo, representado sólo por el Distrito Federal poseía una población de 8,831,079 habitantes y el 13.21% nacional, en 1980, según el X Censo de Población.

Por tanto, la suma de los estratos de muy alta y alta marginlidad, representaban una población total de 30,327,375 habitantes y el 45.37% a nivel nacional.

El estrato de muy alta marginalidad comprende, con excepción de Zacatecas, a entidades federativas situadas en la zona pacífico sur y centro este de la República Mexicana. Por su parte, en el estrato de marginalidad media baja, la mayoría de los estados están situados geográficamente al norte del país; salvo Jalisco y Aguascalientes que se encuentran más bien al centro y occidente de la República.

Con el propósito de establecer comparación de la situación de bienestar entre los estados, considerando exclusivamente los datos censales de 1980, se ha decidido presentar los datos de las tres entidades con muy alta marginación Oxaca, Chiapas y Guerrero; y los tres estados con menores niveles de marginación Baja California, Nuevo León y Coahuila. Considerar sólo seis entidades federativas permite exponer casos extremos y no abundar en demasiados datos.

En este sentido, algunos de los indicadores generales como población económicamente activa sin ingresos, población rural y población económicamente activa que trabaja en el sector agropecuario mostraron cambios, de acuerdo al estrato en que se ubica cada estado.(Ver el cuadro 5.1).

En primer término, el indicador de la población económicamente activa sin ingresos para los estados de Chiapas, Guerrero y Oaxaca, obtuvieron valores de 39.1% , 37.02% y 43.59% ; en cambio en Baja California, Coahuila y Nuevo León la PEA que no recibió percepciones sólo alcanzó el 9.64% , 8.57% y 9.40% respectivamente. Estos últimos valores quedan situados por debajo del promedio nacional de 20.19%.

Por su parte, el porcentaje de población rural en estas mismas entidades, también observa diferencias importantes, que dan cuenta de la presencia de situaciones de marginalidad en aquellos estados que cuentan con mayor población rural. Chiapas, Guerrero y Oaxaca albergan a un 79.68% , 70.97% y 83.78% de este tipo de población. Por el contrario, Baja California, Coahuila y Nuevo León tenian 22.28% , 29.19% y 18.25% respectivamente.

La población económicamente activa que trabaja en el sector agropecuario sólo representa el 9.47% , 15.78% y 8.37% para Baja California, Coahuila y Nuevo León; sin embargo en Chiapas, Guerrero y Oaxaca comprendian al 57.43% , 44.28% y 55.32% ; más de la mitad de PEA.

En el rubro de alimentación los indicadores de subconsumo de huevo y leche ofrecen la oportunidad de observar ciertos contrastes. Para los tres estados de muy alta marginación, ya menciomados con anterioridad, el subconsumo de huevo se expresó en 46.37% , 36.87% y 40.40% de población de cinco años y menos que consumió huevo tres o menos dias a la semana. En cambio, para los tres de la zona norte, estos indicadores tuvieron valores de 14.52% ,16.52% y 16.35% .

El subconsumo de leche fue de 6.53% , 15.16% y 12.97% para Baja California, Coahuila y Nuevo León; en tanto para Chiapas, Guerrero y Oaxaca representaba el 35.39% , 36.87% y 40.40%.

Estos porcentajes en materia de alimentación son de gran importancia, tomando en consideración que han afectado a la población infantil en aquellas entidades donde la situación es más critica.

En materia de educación, el indicador de analfabetismo, con un promedio de 15.39% de la población que no sabe leer ni escribir. Los niveles de las entidades de marginación media baja, Baja California, Coahuila y Nuevo León se encuentran muy por debajo de este promedio, 5.85% , 6.94% y 6.31% . En grave contraste, Chiapas, Guerrero y Oaxaca presentan niveles de analfabetismo de 34.61% , 31.22% y 31.40% respectivamente.

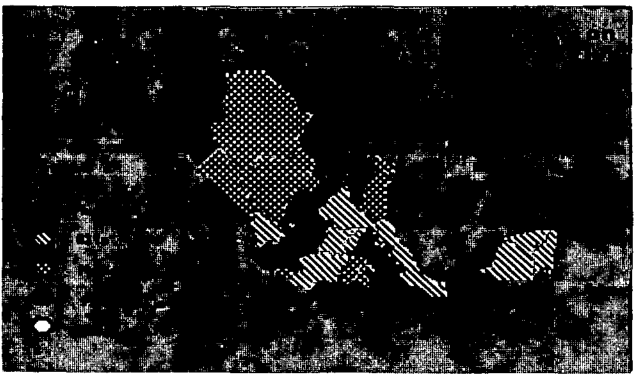
Entre los indicadores de salud, la mortalidad preescolar, para los estados con muy alta marginación alcanza los valores de 6.1 , 6.0 y 13.0 ; por el otro lado, los estados de la zona norte sólo llegaban a 1.8 , 2.1 y 1.8 . El promedio nacional se ubicó en 3.63.

En relación a las condiciones de vida de la población, por el tipo de servicios en la vivienda se tienen los indicadores de sin agua entubada y sin electricidad. En el caso de las viviendas sin agua entubada, Chiapas, Guerrero y Oaxaca tenian 54.92% , 50.61% y 54.17% ; más de la mitad de las viviendas en los estados. Baja California, Coahuila y Nuevo León, en cambio presentaron valores de 21.25% , 14.52% y 12.07% . Las viviendas sin electricidad también presentaron situación similar; con los datos siguientes: 51.50% , 38.60% y 47.63% por una parte; y 9.16% , 11.78% y 9.34% por otra.

En relación al porcentaje de población de cinco años y más, que habla lengua indígena, ésta se encuentra fundamentalmente en los estados de Oaxaca, Chiapas, Guerrero, Hidalgo, Puebla, San Luis Potosí, Veracruz, Yucatán, Campeche y Quintana Roo. Cabe destacar que estos diez estados se encuentran en los estratos de marginación muy alta y alta. En este sentido la población indígena, muy probablemente, es aquella que más difícil situación enfrenta en sus condiciones mínimas de bienestar.

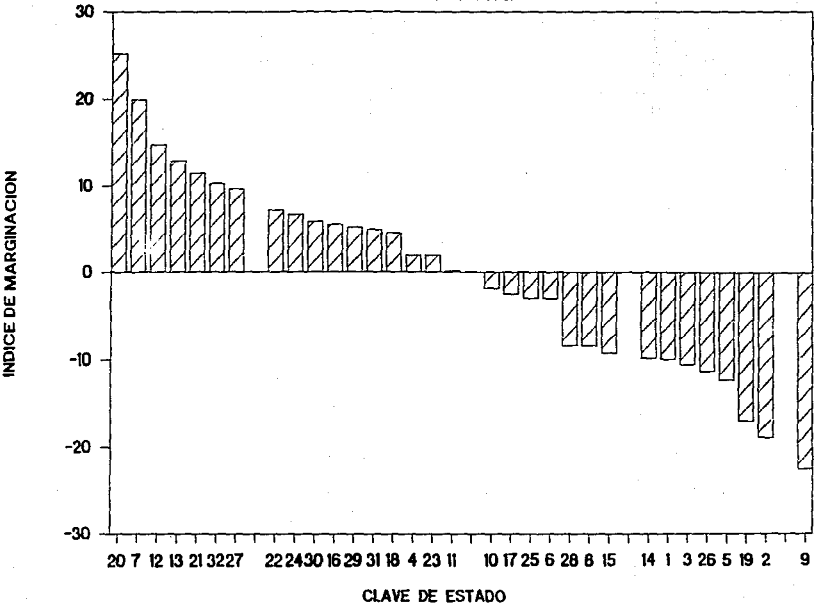
A continuación se presenta una gráfica y un mapa con los grados de marginación, para mostrar gráficamente los estratos de marginalidad, y apreciar el agrupamiento de las entidades federativas de acuerdo a su índice de marginación y estrato, por zona geográfica.

Mapa 5.6. Grados de Marginación por entidad federativa (1980)



# INDICE DE MARGINACION

NIVEL ESTADO (1980)




Gráfica 5.7.

# CONCLUSIONES GENERALES

Las conclusiones más importantes sobre el tema del análisis factorial, desarrollado en esta investigación son las siguientes:

Se logra obtener un concepto claro de lo que significa la técnica estadistica del análisis factorial. `Es un método que sirve para analizar un conjunto de variables, ver sus interrelaciones y definir sus conceptos o fenómenos de variación comunes o únicas, representadas por pocos factores expresados en una combinación lineal, por un subconjunto de dichas variables´.

El análisis factorial comprende varios métodos de extracción de factores iniciales. Entre ellos el análisis de Componentes Principales es el de mayor interés en el presente trabajo, y en la práctica es uno de los más aplicados.

Este método de Componentes Principales se fundamenta en el model matemático general del análisis factorial:

$$Z_j = a_{j1}F_1 + a_{j2}F_2 + \dots + a_{jm}F_m + u_jY_j$$

$$\text{y} \qquad F_i = \sum_{j=1}^{p} W_{ij}\, x_j$$

En estas dos relaciones se conjuga toda la teoria matemática, en la que se desarrolla el modelo factorial, del cual se deriva la estructura analitica de teorias o modelos sociales y cientificos, plasmados en aplicaciones teóricas y después prácticas.

Para entender en que consiste la técnica de componentes principales; esta puede verse como un elipsoide cuyos ejes corresponde a los componentes o factores, entonces se da una rotación de eje coordenados, para construir nuevos ejes de referencia en el espacio total de las variables, donde cada una de las n variables originales es definida en términos de los n nuevos componentes principales, y estos cuentan para una cantidad máxima de la varianza total de las variables.

Se tiene que el método de componentes principales, sólo forma parte de uno de los cuatro pasos fundamentales aplicados en el análisis factorial. Los otros tres pasos o fases de la metodología factorial son: la generación de la matriz de correlación, la rotación de los factores iniciales y la determinación de los puntajes factoriales. Estos fueron ampliamente descritos para comprender mejor la teoría del análisis. Los cuales deben ser realizados en forma ordenada.

Para ver con más precisión la gran utilidad en cuanto al uso de la técnica factorial. Se presentaron una serie de aplicaciones sobre el análisis factorial, comprendidas dentro de las áreas científicas y sociales, por ejemplo: en la simplificación o reducción de datos, la clasificación o descripción de conceptos, la formación de escalas ponderadas, la comprobación de hipótesis, etc.

Antes de iniciar cualquier análisis de variables es necesario considerar, su clasificación si es discreta o continua; determinar su orientación descriptiva, esto es cuál es la variable dependiente o independiente; así como la forma de recopilar los datos y definir el objetivo del estudio o la hipótesis. En su conjunto estas consideraciones guían al investigador en la selección del método de análisis, para las variables en cuestión.

De esto se deriva que en el método multivariado del análisis factorial se pueden usar todo tipo de variables, no están identificadas por dependientes e independientes y los factores resultantes pueden ser usados como variables en un segundo análisis. Esto representa una ventaja importante del método.

Así, esta técnica estadística sirve de prueba independiente, para el análisis de grupos de variables y ayuda a definir las relaciones funcionales de las mismas, proporcionando la información necesaria para probar la hipótesis propuesta.

Se alcanza otro objetivo propuesto en el trabajo, que es el de establecer los lineamientos para la utilización del paquete computarizado SPSS/PC, el cual proporciona varios métodos estadísticos para el análisis de datos, entre ellos la subrutina FACTOR que contiene el método de análisis factorial.

Se logra explicar el funcionamiento de la subrutina FACTOR a través de los subcomandos, regidos por una estructura general que está integrada por los cuatro pasos analíticos, para llevar a cabo un análisis factorial, como resultado se elabora el programa aplicado a la solución del caso práctico`índices de marginación´.

Otro aspecto importante es la interpretación de los resultados factoriales, estos se analizan por medio de un pequeño ejemplo considerando cinco variables socio-económicas y diez entidades federativas. Después de someter los datos a un análisis factorial utilizando la subrutina FACTOR con el paquete SPSS, resultan varios cuadros factoriales, que se analizan obteniendo ciertas conclusiones:

El indicador Población Rural y Población Agrícola están relacionados con el 68% al indicador PEA sin Ingresos y con el 30% al indicador Subempleo.

La varianza total es explicada en un 66.3% por el primer factor, y el 23.5% de la varianza restante es explicada por el segundo factor. En consecuencia ambos factores explican el 89.8% de la varianza total, esto indica que los factores encontrados son significativos.

Por tanto, el modelo factorial para el ejemplo particular es representado, por dos factores o componentes principales:

$$Z_j = a_{j1}F_1 + a_{j2}F_2$$

Y cada variable puede expresarse en términos de los factores extraídos.

Al relizarse una correcta interpretación de los cuadros factoriales, se logran obtener satisfactorios resultados del análisis realizado.

Con respecto al caso práctico de`indices de marginación´, se desprenden algunas observaciones:

a) En el estrato de marginalidad media baja, la mayoría de los estados se ubican geográficamente al norte del país, con excepción de Jalisco y Aguascalientes.

b) En aquellas entidades en donde los porcentajes de población rural y población económicamente activa en el sector agropecuario son muy elevados, superando en mucho los promedios nacionales, también se encuentra que el comportamiento de las variables de bienestar los ubica como de mayor marginación.

c) Entre los indicadores que mostraron mayores contrastes en las entidades de muy alta marginación, y los de media baja marginación se encuentran población económicamente activa sin ingreso, subconsumo de huevo y leche, analfabetismo, mortalidad preescolar, viviendas sin agua entubada y sin electricidad.

Para un número importante de los estados con mayor grado de marginación los porcentajes superaron en un 100 por ciento a la media nacional. Por ejemplo, en Oaxaca el 47.63% de las viviendas no contaban con eletricidad y el promedio nacional se situó en 21.78% . En cambio, en Nuevo León sólo el 9.34% de las viviendas carecía de este servicio.

d) En las entidades que observan los estratos de marginación muy alta y alta, Oaxaca, Chiapas, Guerrero, Hidalgo, Puebla, Zacatecas, Tabasco, Queretaro, San Luis Potosi, Veracruz, Michoacán, Tlaxcala, Nyarit y Yucatan vivia el 81.35% de la poblción de cinco años y más que habla lengua indigena.

e) Es posible concluir que el método estadistico, utilizado para calcular el indice de marginación, es un instrumento útil en la jerarquización y ordenación del comportamiento de un conjunto de indicadores socio-económicos, que miden el fenómeno estudiado y permite visualizar la prioridad de atención de las politicas de erradicación de la pobreza, y en la medida que la información con que se cuente tenga la actualidad y oportunidad requeridas para su aplicación, este instrumento puede ser utilizado como un valioso apoyo para este propósito.

Un papel muy importante juega el actuario en esta clase de investigaciones, porque actúa como analista de modelos matemáticos, estadisticos o de cualquier otra disciplina. Para ser aplicados en áreas interdiciplinarias como Economia, Sociologia, Psicologia, Biologia, Control de Calidad, etc.

Puede asesorar y ayudar, en cuanto al funcionamiento y comprensión, de la teoria estadistica y matemática de métodos estadisticos, en sus aplicaciones mediante sistemas computarizados y en la interpretación y análisis de resultados, a profesionales que tienen necesidades compatibles con estos conocimientos. Por tal motivo se desarrolló el presente trabajo esperando sea una ayuda útil y fructifera.

# P R O P U E S T A S

Se aportan algunas sugerencias o recomendaciones, para obtener el mayor provecho al usar el análisis factorial, logrando óptimos y útiles resultados en la solución de problemas reales.

1. Planteamiento y clasificación del problema a solucionar. Esto es, ubicar al problema dentro de las diversas aplicaciones que tiene el análisis factorial. Así, se podrá identificar que cuadros factoriales serán utilizados, para encontrar los resultados adecuados.

2. El investigador deberá tener en consideración: a) Su propuesta u objetivos, b) La característica matemática de las variables, c) La suposición estadística hecha acerca de estas variables y d) La manera de recolectar los datos. Para seleccionar en forma precisa y adecuada el método estadístico a emplear en el análisis de sus datos o variables. Este punto y el anterior se dan de manera simultánea.

3. En el análisis factorial como en otros métodos estadísticos, el número de sujetos es un elemento importante para el momento de evaluar los resultados. Un análisis factorial muy perfecto, llevado a cabo en diez sujetos, probablemente no vale mucho a causa de la muestra tan pequeña.

4. Otro elemento que hay que tener presente es que, para que el análisis sea significativo, el número de sujetos deberá ser bastante mayor que el de medidas; cuanto más se aproxime el número de medidas al de sujetos, menor será la confianza que tendremos en el valor de todo el análisis.

5. En la aplicación del análisis factorial a un conjunto de datos, siempre debe tomarse en cuenta la metodología de éste, que consiste en cuatro pasos fundamentales:

5.1) Generar la matriz de correlación con los datos recolectados, correspondientes a las variables definidas y efectuar las correlaciones entre estas variables.

5.2) Utilizar la matriz de correlación para determinar un conjunto de factores iniciales, por el método de componentes principales.

5.3) Realizar la rotación de los factores iniciales del paso anterior, para conseguir una estructura más simple de factores significativos, y hacer más fácil su interpretación.

5.4) Determinar los puntajes factoriales que se obtienen sobre las unidades de estudio para cada factor derivado.

6. El análisis factorial puede llevarse a cabo fácilmente sobre correlaciones no válidas y sin contenido, esto traería consecuencias en la interpretación de resultados, por lo que se recomienda hacer un análisis previo de las interrelaciones entre las variables de interés, para sólo seleccionar las variables que presentan una correlación significativa, éstas serán las que se sometan al análisis factorial.

7. El uso de un sistema computarizado para la aplicación de alguna subrutina FACTOR, en particular el paquete SPSS/PC requiere los cuatro pasos a seguir en el análisis factorial, para elaborar un programa específico, esto es:

```
FACTOR VARIABLES=
/ANALISIS=
/PRINT=
```

```
/PLOT=
/DIAGONAL=
/CRITERIA=
/EXTRACTION=
/ROTATION=
```

Son los subcomandos más importantes para ejecutar un análisis.

En la introducción de los valores de las variables (los datos) a un paquete estadístico SPSS, es conveniente definir los archivos con un formato fijo, para evitar errores de valores perdidos en el momento de la lectura.

Cuando el investigador tiene los mínimos conocimientos, sobre un paquete estadístico para utilizarlo. Puede emplear archivos de formato libre al introducir los datos de sus variables, cuidando de hacer una buena codificación de ellos.

8. Es recomendable crear archivos externos o separados, tanto de datos como para el programa de una proceso de análisis factorial. Para poder hacer las modificaciones necesarias a cualquiera de los archivos, de una manera fácil y rápida. Después se jalan estos archivos a través de los comandos de operación, de definición de datos y de manipulación, para ejecutarlos en el sistema SPSS/PC.

9.En la interpretación de los cuadros o resultados factoriales, deben analizarse detenidamente cada uno de ellos, de acuerdo a los criterios que se dieron en el desarrollo de este punto.

Los cuadros factoriales más importantes de analizar recomendados son: Matriz de correlación, Matriz de factores iniciales, La matriz de correlaciones reproducidas, La matriz de factores rotados, La matriz de coeficientes de puntajes factoriales y La tabla de estadísticas finales.

En base al análisis hecho de los resultados factoriales, el investigador deberá verificar sus hipótesis o resultados esperados de acuerdo a sus objetivos planteados en el problema de investigación.

10. Un criterio sugerido para determinar el número de factores significativos a usar en el modelo factorial, es observar el `valor propio mayor o igual a 1´ y elegir solamente los factores que cuenten con varianza mayor que 1, para ser incluidos.

11. Sin embargo, en la interpretación del significado positivo de los factores toca totalmente al investigador o analista. No hay nada intriseco en los métodos analiticos de factores, señalando que las medidas agrupadas juntas formen un patrón que tenga sentido para el investigador.

12. Sucede lo mismo para la interpretación de un conjunto de cargas factoriales la determina arbitrariamente el investigador. El punto en el cual se decide que la carga de factores deja de ser significativa, es también una decisión arbitraria que debe tomar el analista.

13.Una forma de verificar la regularidad y el orden de los datos, es comparando en la tabla de estadísticas finales, que la suma de las comunalidades entre el número de variables multiplicado por 100, sea igual o equivalente al porcentaje de variación total, o sea la suma de la columna etiquetada con el nombre de Pct. de Var.

14. Otra forma de evaluar si el modelo no es adecuado para los datos, y asi poderlo reconsiderar, es comprobando si la mayoria de los residuales que arroja la matriz de correlaciones reproducidas, son mayores que 0.05 en valor absoluto. Esto quiere decir que las magnitudes de los residuales indican que un apropiado modelo factorial reproduce las correlaciones esperadas.

## B I B L I O G R A F I A


1. ANDERSON, T. W. An Introduction to Multivariate Statistical Analysis. Canada, Ed. John Wiley & Sons, INC., 1958. 374p.

2. CATTEL, Raymond. Factor analysis: An introduction to Essentials.(1) The Purpose and Underlyng Models. (II) The Role of Factor Analysis in Research. Biometrics. 21pp. 190-215, 405-435. 1965.

3. Coordinación General del Plan Nacional de zonas Deprimidas y Grupos Marginados. Geografía de la marginación. México, COPLAMAR - Siglo XXI, 1983. 114p.

4. CURTIS, Hardyck y Lewis F. Petrinovich. Investigación en Ciencias Sociales. Tr. Dr. Pedro Rivera. México, Ed. Interamericana, 1977. 188p.

5. FRUCHTER, Benjamin. Introduction to Factor Analysis. 8o. ed., New Jersey, Ed. D. Van Nostrand Company, INC., 1968. 280p.

6. GUTTMAN, Louis. Some Necesary Conditions for Common - Factor Analysis. Psychometrika 19. pp. 149-161. 1964.

7. HARMAN, Harry H. Modern Factor analysis. 3a. ed., Chicago Press. Ed. The University of Chicago Press, 1976. 487p.

8. HOLGUIN QUIÑONES, Fernando. Estadística descriptiva (Aplicada a las ciencias sociales). 2a. ed., México, Ed. UNAM, Facultad de Ciencias Píiticas y Sociales, 1981. 474p. serie estudios 13.

9. Instituto Nacional de Estadística Geografía e Informática. X Censo General de Población y Vivienda 1980. Resumen General y por entidad federativa. México. INEGI, 1986.

10. JOHN C., Davis. Statistics and Data Analysis in Geology. Kansas, Ed. John Wiley & Sons, INC., 1973. 550p.

11. K. Hope. Manual práctico de estadística avanzada. Aplicaciones a las ciencias de la conducta y las ciencias médicas. 2a. ed., Tr. Nahum Martinez Reyes. México, Ed. Trillas, 1982. 111p.

12. K. G., Jöreskog, et al. Method in geomathematics 1, Geological Factor Analysis. Amsterdam, Ed. Elsevier Scientific Publishing Company, 1976. 179p.

13. KLEINBAUM G., David y Lawrence L. Kupper. Applied Regression Analysis and other Multivariate Methods. California, Ed. Duxbury Press, 1978. 556p.

14. LANG, Sarge. Algebra Lineal. 2a. ed., Tr. Miguel Lara A. México, Ed. Fonfo Educativo Interamericano, 1970. 350p.

15. MOOD, Graybill. Introducción a la Teoría de la Estadística. 4a. ed., Tr. Rafael Pro Bermejo. España, Ed. Aguilar,

1978. 536p.

16. MORRISON, Donal F. Multivariate Statistical Methods. 2a. ed., University of Pennsylvania, Ed. Mc Graw-Hill; series en probability and statistic, 1984. 415p.

17. NORUSIS J.,María. Statistical packge for social sciences, SPSS/PC para PC/XT. Chicago, Ed. Michigan Avanue, 1984. 307p.

18. ORDAZ ARANDA, F. J. Notas de Análisis Multivariado. 29.

19. ROY NATH, Samarendo, et al. Multivariate Analysis. New York, Ed. Paruchuri R. Krishnaiah, 1966. 592p.

20. RUMMEL R. J. Undertanding Factor Analysis. Conflict Resolution 11, pp. 444-480. 1967.

21. SCHWARTZMAN, Simón, et al. Técnicas avanzadas en ciencias sociales. Buenos Aires, Ed. Nueva Visión, 1977. 226p.

A N E X O S

# A N E X O I

Listado de variables utilizadas en el caso práctico:

Capitulo V "Indices de Marginación".

I1. Porcentaje de población económicamente activa que percibe ingresos inferiores a 3,611.00 pesos mensuales.

Fuente: X Censo General de Población y Vivienda, 1980 y Resumen General.
Volumen II, cuadro 33, pp. 1023-1040.

Datos: Población económicamente activa total por entidad federativa.
Población económicamente activa por entidad federativa en grupos de ingreso mensual de 1 a 3610 pesos.

Construcción: para cada entidad.

$$I1 = \frac{\text{PEA en grupos de ingreso mensual de 1 a 3610 pesos}}{\text{PEA Total}} \times 100$$

I2. Porcentaje de población económicamente activa sin ingreso.

Fuente: X Censo General de Población y Vivienda, 1980 y Resumen General.
Volumen II, cuadro 33, pp. 1023-1040.

Datos: Población económicamente activa que no recibe ingresos.
Población económicamente activa total por entidad federativa.

Construcción: para cada entidad.

$$I2 = \frac{\text{PEA que no recibe ingresos}}{\text{PEA total}} \times 100$$

I3. Porcentaje de población económicamente activa subempleada (PEA que trabaja 32 horas o menos a la semana).

Fuente: X censo General de Población y Vivienda, 1980 y Resumen General.
Volumen II, cuadro 35, pp. 1047-1092.

Datos: Población económicamente activa total por entidad federativa.
Población económicamente activa por horas trabajadas desde menos de una hora hasta 32 horas.

Construcción: para cada entidad.

$$I3 = \frac{\text{PEA por horas trabajadas desde menos de una hora hasta 32 hrs.}}{\text{PEA Total}} \times 100$$

I4. Porcentaje de población rural y (hasta 10,000 habitantes).

Fuente: X Censo General de Población y Vivienda, 1980 y Resumen General.
Volumen I, cuadro 3, pp. 33-111.

Datos: Población total por entidad federativa.
Población por entidad federativa por tamaño de localidades de 1 a 9,999 habitantes.

Construcción: para cada entidad.

$$I4 = \frac{\text{Población por tamaño de localidad de 1 a 9,999 habitantes.}}{\text{Población total}} \times 100$$

I5. Porcentaje de población económicamente activa en el sectore agropedcuario.

Fuente: X Censo General de Población y Vivienda, 1980
Volumen I, cuadro 20, pp. 435-455.

Datos: Población económicamente activa total por entidad federativa.
Población económicamente activa por rama de actividad económica agricultura; ganadería; caza, ect. por entidad federativa.

Construcción: Para cada entidad federativa.

$$I5 = \frac{\text{PEA por rama de actividad económica, agricultura; ganadería; caza, etc.}}{\text{PEA Total}} \times 100$$

I6. Porcentaje de población de cinco años y menos que consume carne tres o menos días a la semana.

Fuente: X Censo General de Población y Vivienda, 1980 y Resumen General.
Volumen II, cuadro 55, pp. 1469-1522.

Datos: Población total de cinco años y menos por entidad federativa.
Población de cinco años y menos por frecuencia de consumo

de carne de tres o menos días a la semana, por entidad federativa.

Construcción: Para cada entidad federativa.

$$I6 = \frac{\text{Población de cinco años y menos por frecuencia de consumo de carne de tres o menos días a la semana.}}{\text{Población total de cinco años y menos.}} \times 100$$

I7. Porcentaje de población de cinco años y menos que consume huevo tres o menos días a la semana.

Fuente: X Censo General de Población y Vivienda, 1980 y Resumen General.
Volumen II, cuadro 55, pp. 1469-1522.

Datos: Población total de cinco años y menos por entidad federativa.
Población de cinco años y menos por frecuencia de consumo de huevo, de tres o menos días a la semana, por entidad federativa.

Construcción: Para cada entidad federativa.

$$I7 = \frac{\text{Población de cinco años y menos por frecuencia de consumo de huevo de tres o menos días a la semana.}}{\text{Población total de cinco años y menos.}} \times 100$$

I8. Porcentaje de población de cinco años y menos que consume leche tres o menos días a la semana.

Fuente: X Censo General de Población y Vivienda, 1980 y resumen General.

Volumen II, cuadro 55, pp. 1469-1522.

Datos: Población total de cinco y menos por entidad federativa.
Población de cinco años y menos por frecuencia de consumo de leche de tres o menos días a la semana, ablactados y no ablactados, por entidad federativa.

Construcción: Para cada entidad federativa.

$$I8 = \frac{\text{Población de cinco años y menos por frecuencia de consumo de leche de tres o menos días a la semana.}}{\text{Población total de cinco años y menos.}} \times 100$$

I9. Porcentaje de población analfabeta de 10 años y más de edad.

Fuente: X Censo General de Población y Vivienda, 1980 y Resumen General.
Volumen I, cuadro 7, pp. 159-201.
Volumen I, cuadro 11, pp. 223-224.

Datos: Población total por entidad federativa de 15 años y más.
Población total por entidad federativa de 10 a 14 años.
Población que no sabe leer ni escribir de 10 a 14 años por entidad.
Población analfabeta de 15 años y más por entidad.

Construcción: Pra cada entidad federativa.

$$I9 = \frac{\text{Población que no sabe leer ni escribir de 10 a 14 años (más) Población analfabeta de 15 años y más.}}{\text{Población total de 15 años y más (más) Población de 10 a 14 años.}} \times 100$$

I10. Porcentaje de población de 15 años y más sin primaria completa.

Fuente: X Censo General de Poblción y Vivienda, 1980 y Resumen General.
Volumen I, cuadro 11, pp. 223-224.

Datos: Población total de 15 años y más por entidad federativa.
Población de 15 años y más con enseñanza primaria desde primer hasta quinto grado, por entidad federativa.

Construcción: Para cada entidad federativa.

$$I10 = \frac{\text{Población de 15 años y más con enseñanza primaria desde primero a quinto grado.}}{\text{Población de 15 años y más}} \times 100$$

I11. Tasa de mortalidad general

Fuente: X Censo General de Población y Vivienda, 1980.
Volumen I, cuadro 2, pp. 7-30.
Anuario Estadístico de los Estados Unidos Mexicanos, 1983.
Cuadro 11.2.5, pp. 164-169.

Datos: Población total por entidad federativa.
Defunciones totales por entidad federativa.

Construcción: Para cada estado.

$$I11 = \frac{\text{Defunciones}}{\text{Población}} \times 1000$$

I12. Tasa de motalidad preescolar.

Fuente: Anuario estadistico de los Estados Unidos Mexicanos, 1983.
Cuadro III.2.5, pp. 164-170.
X Censo General de Poblción y Vivienda, 1980.
Volumen I, cuadro 2, pp. 7.

Datos: Total de defunciones de población en el grupo de edad de 1 a 4 años, por entidad.
Total de población de 1 a 4 años por entidad.

Construcción: Para cada estado.

$$I12 = \frac{\text{Total de defunciones de población en el grupo de 1 a 4 años.}}{\text{Total de población de 1 a 4 años.}} \times 1000$$

I13.Porcentaje de viviendas que usa como combustible, para cocinar leña o petróleo.

Fuente: X Censo General de Población y Vivienda, 1980 y Resumen General.
Volumen II, cuadro 73, pp. 1811-1829.

Datos: Total de viviendas particulares por entidad federativa.
Total de viviendas que usan como combustible, para cocinar leña o petróleo.

Construcción: Por cada estado.

$$I13 = \frac{\text{Total de viviendas que usan como combustible para cocinar leña o petróleo.}}{\text{Total de viviendas}} \times 100$$

I14. Porcentaje de viviendas con un dormitorio y más de cinco ocupantes. (Tasa de hacinamiento)

Fuente: X Censo General de Población y Vivienda, 1980 y Resumen General.
Volumen II, cuadro 66, pp. 1705-1717.

Datos: Total de viviendas por entidad federativa.
Total de viviendas con un dormitorio y más de cinco ocupantes.

Construcción: Para cada entidad federativa.

$$I14 = \frac{\text{Total de viviendas con un dormitorio y más de cinco ocupantes.}}{\text{Total de viviendas}} \times 100$$

I15. Porcentaje de viviendas sin agua entubada.

Fuente: X Censo General de Población y Vivienda, 1980 y Resumen General.
Volumen II, cuadro 70, pp. 1769-1780.

Datos: Total de viviendas particulares por entidad federativa.
Total de viviendas particulares por entidad federativa, que no disponen de agua.

Construcción: Por cada entidad federativa.

$$I15 = \frac{\text{Total de viviendas que no disponen de agua entubada}}{\text{Total de viviendas particulares}} \times 100$$

I16. Porcentaje de viviendas de uno y dos cuartos.

Fuente: X Censo General de Población y Vivienda, 1980 y Resumen General.
Volumen II, cuadro 64, pp. 1681-1693.

Datos: Total de viviendas por entidad federativa.
Total de viviendas de uno y dos cuartos por entidad.

Construcción: Por cada entidad federativa.

$$I16 = \frac{\text{Total de viviendas de uno y dos cuartos.}}{\text{Total de viviendas}} \times 100$$

I17. Porcentaje de viviendas sin electricidad.

Fuente: X Censo General de Población y Vivienda, 1980 y Resumen General.
Volumen II, cuadro 70, pp. 1769-1780.

Datos: Total de viviendas particulares por entidad federativa.
Total de viviendas particulaes por entidad, que no disponen de enrgia eléctrica.

Construcción: Para cada entidad federativa.

$$I17 = \frac{\text{Total de viviendas sin energia eléctrica.}}{\text{Total de viviendas particulares.}} \times 100$$

I18. Porcentaje de viviendas sin drenaje.

Fuente: X Censo General de Población y Vivienda, 1980 y Resumen General.

Volumen II, cuadro 70, pp. 1769-1780.

Datos: Total de viviendas particulares por entidad.
Total de viviendas particulares que no tienen drenaje, por entidad.

Construcción: Para cada entidad federativa.

$$I18 = \frac{\text{Total de viviendas particulares sin drenaje.}}{\text{Total de viviendas particulares.}} \times 100$$

I19. Porcentaje de población de 5 años y más que habla lengua indígena.

Fuente: X Censo General de Población y Vivienda, 1980 y Resumen General.
Volumen II, cuadro 52, pp. 1419-1424.

Datos: Población de cinco años y más por entidad federativa que habla lengua indígena.
Población total de cinco años y más, por entidad federativa.

Construcción: Para cada entidad federativa.

$$I19 = \frac{\text{Población de cinco años y más que habla lengua indígena.}}{\text{Población total de cinco años y más.}} \times 100$$

## A N E X O II

**Listado de resultados al ejecutar el paquete SPSS/PC**

**. con los datos. (Indicadores socio-económicos 1980)**

```
The raw data or transformation pass is proceeding
SPSS/PC has written       32 cases to the active file

The VARIABLES are listed in the following order:

Line   1: EDO I1 I2 I3 I4 I5 I6

Line   2: I7 I8 I9 I10 I11 I12 I13

Line   3: I14 I15 I16 I17 I18 I19
```

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| EDO: | 1.00 | 40.01 | 14.54 | 14.06 | 36.04 | 17.89 | 65.81 |
| I7: | 34.83 | 13.43 | 9.57 | 32.74 | 6.54 | 2.90 | 21.12 |
| I14: | 13.47 | 11.55 | 43.59 | 14.17 | 26.21 | 1.29 | |
| EDO: | 2.00 | 14.48 | 9.64 | 11.65 | 22.28 | 9.47 | 53.30 |
| I7: | 14.52 | 6.53 | 5.85 | 22.63 | 5.75 | 1.80 | 9.03 |
| I14: | 10.16 | 21.25 | 47.24 | 9.16 | 34.10 | 2.08 | |
| EDO: | 3.00 | 17.33 | 10.37 | 12.99 | 46.54 | 19.35 | 56.86 |
| I7: | 21.00 | 11.58 | 6.82 | 29.37 | 5.63 | 1.50 | 33.67 |
| I14: | 12.83 | 21.97 | 52.55 | 21.45 | 52.61 | 2.10 | |
| EDO: | 4.00 | 31.11 | 15.47 | 16.56 | 47.13 | 31.87 | 62.68 |
| I7: | 43.51 | 20.63 | 15.06 | 31.42 | 5.45 | 2.60 | 39.68 |
| I14: | 27.92 | 39.52 | 69.63 | 21.00 | 54.64 | 21.53 | |
| EDO: | 5.00 | 30.38 | 8.57 | 13.99 | 29.19 | 15.78 | 60.54 |
| I7: | 16.52 | 15.16 | 6.94 | 27.63 | 6.32 | 2.10 | 19.87 |
| I14: | 16.03 | 14.52 | 52.41 | 11.78 | 40.85 | 1.44 | |
| EDO: | 6.00 | 34.55 | 10.23 | 14.40 | 41.77 | 27.85 | 65.40 |
| I7: | 46.52 | 15.42 | 11.20 | 33.09 | 7.07 | 4.50 | 24.31 |
| I14: | 23.87 | 14.84 | 70.81 | 15.67 | 38.62 | 1.33 | |
| EDO: | 7.00 | 25.28 | 39.07 | 14.18 | 79.68 | 57.43 | 66.77 |
| I7: | 46.37 | 35.39 | 34.61 | 31.61 | 6.04 | 6.10 | 70.16 |
| I14: | 37.65 | 54.92 | 73.71 | 51.50 | 65.53 | 28.50 | |
| EDO: | 8.00 | 20.95 | 18.99 | 13.26 | 33.93 | 20.75 | 63.05 |
| I7: | 19.41 | 14.58 | 7.98 | 30.32 | 6.66 | 3.20 | 27.85 |
| I14: | 16.41 | 21.39 | 52.25 | 22.65 | 42.75 | 3.92 | |
| EDO: | 9.00 | 15.14 | 7.96 | 12.63 | 0.0 | 6.13 | 49.23 |
| I7: | 24.67 | 8.85 | 5.24 | 17.40 | 5.64 | 1.40 | 5.45 |
| I14: | 15.67 | 6.38 | 49.06 | 1.46 | 13.79 | 2.67 | |
| EDO: | 10.00 | 28.74 | 25.74 | 16.42 | 61.66 | 30.88 | 65.81 |
| I7: | 20.91 | 21.06 | 8.34 | 36.76 | 5.44 | 2.00 | 46.22 |
| I14: | 18.07 | 25.95 | 52.50 | 21.60 | 56.25 | 1.94 | |
| EDO: | 11.00 | 29.86 | 25.88 | 19.49 | 49.17 | 19.17 | 61.15 |
| I7: | 38.73 | 22.27 | 21.36 | 29.05 | 7.30 | 4.40 | 44.10 |
| I14: | 18.17 | 30.21 | 53.71 | 24.30 | 45.87 | 1.38 | |
| EDO: | 12.00 | 20.65 | 37.02 | 18.99 | 70.97 | 44.28 | 65.22 |
| I7: | 43.31 | 36.87 | 31.22 | 24.31 | 6.06 | 5.00 | 59.86 |
| I14: | 36.02 | 50.61 | 73.75 | 38.60 | 64.97 | 15.30 | |
| EDO: | 13.00 | 37.87 | 24.78 | 15.48 | 78.89 | 37.03 | 67.54 |
| I7: | 47.59 | 29.68 | 25.91 | 29.55 | 8.42 | 5.00 | 61.41 |
| I14: | 28.93 | 40.24 | 66.30 | 39.97 | 63.89 | 23.13 | |
| EDO: | 14.00 | 23.98 | 19.21 | 17.05 | 35.60 | 18.22 | 58.94 |
| I7: | 36.12 | 10.96 | 11.75 | 29.36 | 6.56 | 2.90 | 21.54 |
| I14: | 13.11 | 20.83 | 47.95 | 16.01 | 29.45 | 1.74 | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| EDO: | 15.00 | 19.31 | 14.94 | 14.44 | 37.61 | 15.26 | 58.65 |
| I7: | 35.16 | 18.94 | 11.86 | 25.01 | 6.24 | 3.20 | 23.29 |
| I14: | 22.02 | 17.45 | 54.29 | 11.04 | 29.65 | 5.59 | |
| EDO: | 16.00 | 31.45 | 28.21 | 19.86 | 62.14 | 39.45 | 63.91 |
| I7: | 45.59 | 23.37 | 22.56 | 30.69 | 7.07 | 3.70 | 46.46 |
| I14: | 21.70 | 32.32 | 63.70 | 28.04 | 51.44 | 4.61 | |
| EDO: | 17.00 | 32.58 | 16.99 | 17.77 | 60.36 | 25.11 | 62.71 |
| I7: | 32.18 | 23.14 | 14.53 | 25.38 | 5.93 | 2.70 | 31.38 |
| I14: | 24.22 | 21.81 | 64.65 | 12.44 | 45.35 | 3.85 | |
| EDO: | 18.00 | 42.97 | 15.86 | 18.25 | 60.94 | 40.35 | 71.22 |
| I7: | 48.05 | 24.64 | 13.84 | 38.23 | 4.99 | 2.40 | 28.30 |
| I14: | 26.14 | 28.24 | 72.09 | 20.02 | 62.78 | 3.90 | |
| EDO: | 19.00 | 17.62 | 9.70 | 10.74 | 18.25 | 8.37 | 57.61 |
| I7: | 16.35 | 12.97 | 6.31 | 23.77 | 4.84 | 1.80 | 19.17 |
| I14: | 17.67 | 12.07 | 51.14 | 9.34 | 30.88 | 1.37 | |
| EDO: | 20.00 | 22.59 | 43.59 | 13.65 | 83.78 | 55.32 | 65.27 |
| I7: | 46.83 | 40.40 | 31.40 | 31.31 | 10.12 | 13.00 | 76.37 |
| I14: | 36.11 | 54.17 | 75.20 | 47.63 | 71.88 | 44.01 | |
| EDO: | 21.00 | 34.08 | 27.22 | 15.67 | 63.59 | 41.37 | 64.90 |
| I7: | 44.07 | 30.40 | 23.74 | 28.36 | 9.38 | 7.90 | 56.79 |
| I14: | 32.17 | 38.55 | 66.42 | 27.94 | 54.96 | 17.24 | |
| EDO: | 22.00 | 36.17 | 20.48 | 11.90 | 61.86 | 28.98 | 73.18 |
| I7: | 46.25 | 32.03 | 22.72 | 30.60 | 7.41 | 4.00 | 52.99 |
| I14: | 24.57 | 33.80 | 60.80 | 36.89 | 59.38 | 3.62 | |
| EDO: | 23.00 | 25.31 | 16.47 | 12.80 | 51.75 | 29.16 | 57.48 |
| I7: | 32.51 | 19.25 | 15.11 | 34.91 | 5.03 | 3.10 | 42.50 |
| I14: | 27.14 | 40.76 | 73.72 | 26.28 | 56.83 | 44.09 | |
| EDO: | 24.00 | 35.49 | 24.29 | 15.23 | 62.44 | 34.08 | 67.65 |
| I7: | 31.09 | 26.87 | 18.93 | 32.71 | 5.63 | 3.40 | 59.37 |
| I14: | 22.86 | 48.59 | 59.12 | 42.42 | 59.70 | 13.59 | |
| EDO: | 25.00 | 23.28 | 17.83 | 15.92 | 56.92 | 27.54 | 61.73 |
| I7: | 18.58 | 19.46 | 11.88 | 32.98 | 5.01 | 2.10 | 37.43 |
| I14: | 23.51 | 31.92 | 61.35 | 21.33 | 56.73 | 2.41 | |
| EDO: | 26.00 | 19.12 | 11.87 | 11.97 | 39.35 | 20.81 | 58.74 |
| I7: | 12.23 | 11.60 | 7.45 | 30.22 | 6.31 | 2.10 | 24.71 |
| I14: | 17.22 | 16.48 | 50.99 | 18.91 | 46.22 | 4.68 | |
| EDO: | 27.00 | 33.37 | 17.17 | 13.91 | 72.10 | 38.92 | 64.19 |
| I7: | 46.39 | 25.86 | 16.44 | 41.41 | 7.08 | 4.50 | 44.43 |
| I14: | 31.89 | 57.13 | 72.61 | 42.13 | 51.79 | 6.31 | |
| EDO: | 28.00 | 27.16 | 13.23 | 14.44 | 30.94 | 17.99 | 59.76 |
| I7: | 17.61 | 16.00 | 8.53 | 29.85 | 5.90 | 2.20 | 19.17 |
| I14: | 21.34 | 27.57 | 62.73 | 21.38 | 43.48 | 1.77 | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| EDO: | 29.00 | 32.83 | 25.84 | 17.05 | 71.54 | 37.67 | 65.46 |
| I7: | 42.73 | 27.28 | 14.24 | 28.51 | 5.73 | 3.20 | 51.06 |
| I14: | 30.75 | 27.79 | 66.39 | 15.50 | 63.80 | 6.51 | |
| EDO: | 30.00 | 32.80 | 20.83 | 13.53 | 62.95 | 37.75 | 62.12 |
| I7: | 36.41 | 24.18 | 21.21 | 31.05 | 6.00 | 4.00 | 47.28 |
| I14: | 29.08 | 48.07 | 66.80 | 35.31 | 51.34 | 13.66 | |
| EDO: | 31.00 | 40.41 | 14.57 | 20.35 | 46.13 | 31.36 | 56.64 |
| I7: | 38.26 | 20.87 | 17.36 | 34.26 | 7.38 | 4.80 | 46.09 |
| I14: | 24.59 | 48.45 | 65.15 | 17.77 | 51.41 | 53.15 | |
| EDO: | 32.00 | 35.77 | 39.77 | 13.89 | 73.62 | 49.33 | 81.08 |
| I7: | 42.37 | 26.94 | 12.69 | 47.24 | 5.70 | 2.60 | 48.19 |
| I14: | 19.92 | 41.82 | 55.73 | 38.78 | 70.31 | .55 | |

Number of cases read = 32 Number of cases listed = 32

--- F A C T O R   A N A L Y S I S ---

Analysis Number 1 Listwise deletion of cases with missing values

| | Mean | Std Dev | Label |
|---|---|---|---|
| I1 | ?8.52000 | 7.85460 | |
| I2 | 20.19781 | 9.52138 | |
| I3 | 15.07875 | 2.51733 | |
| I4 | 51.53500 | 19.71096 | |
| I5 | 29.21625 | 13.06692 | |
| I6 | 62.95625 | 5.94461 | |
| I7 | 33.95844 | 11.91788 | |
| I8 | 21.45656 | 8.38573 | |
| I9 | 15.39531 | 8.00467 | |
| I10 | 30.67906 | 5.54454 | |
| I11 | 6.39469 | 1.20694 | |
| I12 | 3.62813 | 2.22902 | |
| I13 | 38.72656 | 17.55139 | |
| I14 | 23.16281 | 7.26506 | |
| I15 | 31.28656 | 14.25500 | |
| I16 | 60.88563 | 9.33771 | |
| I17 | 24.45219 | 12.45352 | |
| I18 | 49.60813 | 13.87338 | |
| I19 | 10.60188 | 14.00300 | |

Number of Cases = 32

Correlation Matrix:

| | I1 | I2 | I3 | I4 | I5 | I6 |
|---|---|---|---|---|---|---|
| I1 | 1.00000 | | | | | |
| I2 | .16210 | 1.00000 | | | | |
| I3 | .44519 | .29767 | 1.00000 | | | |
| I4 | .46913 | .78968 | .34335 | 1.00000 | | |
| I5 | .43168 | .83406 | .31347 | .91264 | 1.00000 | |
| I6 | .63754 | .61623 | .15262 | .70691 | .67982 | 1.00000 |
| I7 | .59804 | .55962 | .43500 | .67119 | .71418 | .56994 |
| I8 | .38988 | .83514 | .31194 | .88125 | .87674 | .64245 |
| I9 | .27628 | .78392 | .35352 | .76790 | .79209 | .41621 |
| I10 | .57998 | .32345 | .12485 | .53452 | .54601 | .67608 |
| I11 | .22284 | .39400 | .11277 | .35182 | .35066 | .16669 |
| I12 | .14810 | .65904 | .12266 | .57423 | .65047 | .25518 |
| I13 | .34266 | .83776 | .29100 | .90179 | .87137 | .55739 |
| I14 | .29287 | .61205 | .25031 | .75341 | .81639 | .38076 |
| I15 | .31474 | .67770 | .26107 | .78412 | .82138 | .40644 |
| I16 | .34438 | .43770 | .30226 | .68072 | .75569 | .32448 |
| I17 | .27130 | .77357 | .04278 | .82244 | .82736 | .60738 |
| I18 | .37461 | .71619 | .26422 | .89504 | .86147 | .66859 |
| I19 | .14370 | .33549 | .16597 | .36947 | .48203 | -.06538 |

--- FACTOR ANALYSIS ---

| | I7 | I8 | I9 | I10 | I11 | I12 |
|---|---|---|---|---|---|---|
| I7 | 1.00000 | | | | | |
| I8 | .71737 | 1.00000 | | | | |
| I9 | .73001 | .89609 | 1.00000 | | | |
| I10 | .34252 | .28788 | .10003 | 1.00000 | | |
| I11 | .46174 | .45669 | .53453 | -.00470 | 1.00000 | |
| I12 | .57540 | .72306 | .76431 | .08744 | .82886 | 1.00000 |
| I13 | .62528 | .91703 | .87403 | .35437 | .46521 | .70779 |
| I14 | .68360 | .86765 | .81519 | .19470 | .34496 | .67521 |
| I15 | .56707 | .78687 | .78329 | .45936 | .30914 | .59434 |
| I16 | .66224 | .72441 | .68498 | .29253 | .23796 | .55695 |
| I17 | .54416 | .81407 | .80200 | .47799 | .38761 | .61307 |
| I18 | .48172 | .80658 | .64659 | .54967 | .19942 | .46056 |
| I19 | .35709 | .45772 | .55251 | .12502 | .36529 | .60364 |

| | I13 | I14 | I15 | I16 | I17 | I18 |
|---|---|---|---|---|---|---|
| I13 | 1.00000 | | | | | |
| I14 | .77728 | 1.00000 | | | | |
| I15 | .84476 | .77275 | 1.00000 | | | |
| I16 | .64410 | .91136 | .72383 | 1.00000 | | |
| I17 | .86902 | .67571 | .86838 | .58155 | 1.00000 | |
| I18 | .85309 | .68127 | .76914 | .67432 | .77436 | 1.00000 |
| I19 | .56834 | .57700 | .64101 | .58096 | .41148 | .43028 |

| | I19 |
|---|---|
| I19 | 1.00000 |

WARNING 11302
THE CORRELATION MATRIX IS ILL-CONDITIONED FOR FACTOR PROCESSING.

Determinant of Correlation Matrix = .0000000

--- F A C T O R   A N A L Y S I S ---

Inverse of Correlation Matrix:

| | I1 | I2 | I3 | I4 | I5 |
|---|---|---|---|---|---|
| I1 | 9.59433 | | | | |
| I2 | 9.59350 | 21.30277 | | | |
| I3 | -2.88314 | -4.98559 | 5.99206 | | |
| I4 | -1.12775 | .02125 | .51683 | 17.15529 | |
| I5 | -4.16483 | -12.08687 | .14750 | -3.65923 | 24.86514 |
| I6 | -10.28622 | -9.66754 | 4.61170 | 4.70079 | .70570 |
| I7 | 1.22979 | .91621 | .90143 | -1.39717 | -2.92913 |
| I8 | -1.68569 | -5.77610 | -2.19322 | -2.56660 | 12.10947 |
| I9 | -1.89902 | -1.95014 | -6.50450 | -2.80329 | .02139 |
| I10 | -1.08166 | .02544 | -2.83008 | -4.37553 | -1.20758 |
| I11 | -2.08249 | 2.24941 | -1.74509 | -1.01395 | 2.44970 |
| I12 | 1.85368 | -4.68397 | 3.75407 | 1.37977 | -4.67322 |
| I13 | -4.78312 | -4.79455 | -1.77654 | -7.01255 | 5.25398 |
| I14 | 2.42474 | 3.22207 | 5.05611 | -1.89243 | -12.04496 |
| I15 | -2.99323 | -1.58202 | -2.67411 | 1.76940 | 1.27359 |
| I16 | -.50637 | 6.05652 | -4.41501 | 1.00933 | 1.07385 |
| I17 | 4.40197 | 1.05042 | 9.27664 | 1.30022 | -4.49321 |
| I18 | 6.58459 | 5.35181 | -.15355 | -6.32569 | -8.90207 |
| I19 | -2.32058 | -1.48863 | 2.75316 | 4.65279 | -.18727 |

| | I6 | I7 | I8 | I9 | I10 |
|---|---|---|---|---|---|
| I6 | 26.76833 | | | | |
| I7 | -5.36712 | 7.88787 | | | |
| I8 | -14.01326 | -.29363 | 41.17713 | | |
| I9 | .79709 | -6.56433 | -2.98153 | 31.01003 | |
| I10 | -6.49167 | -1.44355 | 6.98714 | 11.83965 | 11.19416 |
| I11 | 1.16132 | -2.57589 | 2.75464 | 2.46517 | 1.86650 |
| I12 | 1.93795 | 3.06302 | -8.86450 | -4.02612 | -3.63868 |
| I13 | 8.94481 | -4.27381 | -4.22268 | 3.16912 | 2.14751 |
| I14 | -.34054 | 4.61783 | -19.54871 | -3.60770 | -.40143 |
| I15 | 10.46748 | -3.09840 | -5.13069 | 4.49923 | -2.65781 |
| I16 | 4.13398 | -6.34642 | 4.25746 | 2.32263 | .03110 |
| I17 | -7.10347 | 7.73101 | -.49405 | -23.17378 | -8.22670 |
| I18 | -12.18335 | 7.72030 | -6.81010 | 1.04746 | 1.17312 |
| I19 | 4.90354 | .09516 | .74330 | -5.15894 | -3.14117 |

| | I11 | I12 | I13 | I14 | I15 |
|---|---|---|---|---|---|
| I11 | 8.60324 | | | | |
| I12 | -11.05702 | 18.91763 | | | |
| I13 | 1.91866 | -3.24766 | 30.62519 | | |
| I14 | -2.17478 | 5.72584 | -10.01008 | 32.75147 | |
| I15 | 2.05310 | -.44473 | 5.88299 | -6.71552 | 17.22332 |
| I16 | 4.86898 | -8.04886 | 12.51917 | -19.76456 | 5.20636 |
| I17 | -5.59374 | 7.23567 | -15.24677 | 13.89654 | -16.57954 |
| I18 | -3.48254 | 6.61854 | -13.68287 | 15.04662 | -6.04002 |

--- F A C T O R  A N A L Y S I S ---

| | I11 | I12 | I13 | I14 | I15 |
|---|---|---|---|---|---|
| I19 | .51909 | -1.50776 | -4.57990 | 1.28106 | -2.47152 |

| | I16 | I17 | I18 | I19 |
|---|---|---|---|---|
| I16 | 22.89477 | | | |
| I17 | -11.41294 | 37.45408 | | |
| I18 | -13.95834 | 8.80984 | 24.24429 | |
| I19 | -1.38339 | 6.37275 | -2.95399 | 6.55766 |

Kaiser-Meyer-Olkin Measure of Sampling Adequacy = .79576

Bartlett Test of Sphericity = 798.91904, Significance = .00000

There are 0 ( 0.0%) off-diagonal elements of AIC Matrix > 0.09

Anti-Image Covariance Matrix:

| | I1 | I2 | I3 | I4 | I5 |
|---|---|---|---|---|---|
| I1 | .10423 | | | | |
| I2 | .04694 | .04694 | | | |
| I3 | -.05015 | -.03906 | .16689 | | |
| I4 | -.00685 | .00006 | .00503 | .05829 | |
| I5 | -.01746 | -.02282 | .00099 | -.00858 | .04022 |
| I6 | -.04005 | -.01695 | .02875 | .01024 | .00106 |
| I7 | .01625 | .00545 | .01907 | -.01033 | -.01493 |
| I8 | -.00427 | -.00658 | -.00889 | -.00363 | .01183 |
| I9 | -.00638 | -.00295 | -.03501 | -.00527 | .00003 |
| I10 | -.01007 | .00011 | -.04219 | -.02278 | -.00434 |
| I11 | -.02523 | .01227 | -.03385 | -.00687 | .01145 |
| I12 | .01021 | -.01162 | .03312 | .00425 | -.00993 |
| I13 | -.01628 | -.00735 | -.00968 | -.01335 | .00690 |
| I14 | .00772 | .00462 | .02576 | -.00337 | -.01479 |
| I15 | -.01811 | -.00431 | -.02591 | .00599 | .00297 |
| I16 | -.00231 | .01242 | -.03218 | .00257 | .00189 |
| I17 | .01225 | .00132 | .04133 | .00202 | -.00482 |
| I18 | .02831 | .01036 | -.00106 | -.01521 | -.01477 |
| I19 | -.03688 | -.01066 | .07007 | .04136 | -.00115 |

| | I6 | I7 | I8 | I9 | I10 |
|---|---|---|---|---|---|
| I6 | .03736 | | | | |
| I7 | -.02542 | .12678 | | | |
| I8 | -.01271 | -.00090 | .02429 | | |
| I9 | .00096 | -.02684 | -.00233 | .03225 | |
| I10 | -.02166 | -.01635 | .01516 | .03411 | .08933 |
| I11 | .00504 | -.03796 | .00778 | .00924 | .01938 |
| I12 | .00383 | .02053 | -.01138 | -.00686 | -.01718 |

--- F A C T O R  A N A L Y S I S ---

| | I6 | I7 | I8 | I9 | I10 |
|---|---|---|---|---|---|
| I13 | .01091 | -.01769 | -.00335 | .00334 | .00626 |
| I14 | -.00039 | .01788 | -.01450 | -.00355 | -.00109 |
| I15 | .02270 | -.02281 | -.00723 | .00842 | -.01379 |
| I16 | .00675 | -.03514 | .00452 | .00327 | .00012 |
| I17 | -.00709 | .02617 | -.00032 | -.01995 | -.01962 |
| I18 | -.01877 | .04037 | -.00682 | .00139 | .00432 |
| I19 | .02793 | .00184 | .00275 | -.02537 | -.04279 |

| | I11 | I12 | I13 | I14 | I15 |
|---|---|---|---|---|---|
| I11 | .11624 | | | | |
| I12 | -.06794 | .05286 | | | |
| I13 | .00728 | -.00561 | .03265 | | |
| I14 | -.00772 | .00924 | -.00998 | .03053 | |
| I15 | .01386 | -.00136 | .01115 | -.01191 | .05806 |
| I16 | .02472 | -.01858 | .01786 | -.02636 | .01320 |
| I17 | -.01736 | .01021 | -.01329 | .01133 | -.02570 |
| I18 | -.01670 | .01443 | -.01843 | .01895 | -.01446 |
| I19 | .00920 | -.01215 | -.02280 | .00596 | -.02188 |

| | I16 | I17 | I18 | I19 |
|---|---|---|---|---|
| I16 | .04368 | | | |
| I17 | -.01331 | .02670 | | |
| I18 | -.02515 | .00970 | .04125 | |
| I19 | -.00921 | .02595 | -.01858 | .15249 |

--- FACTOR ANALYSIS ---

Anti-Image Correlation Matrix:

| | I1 | I2 | I3 | I4 | I5 | I6 |
|---|---|---|---|---|---|---|
| I1 | .61431 | | | | | |
| I2 | .67104 | .83299 | | | | |
| I3 | -.38025 | -.44128 | .41231 | | | |
| I4 | -.08790 | .00111 | .05098 | .93256 | | |
| I5 | -.26965 | -.52517 | .01208 | -.17717 | .90161 | |
| I6 | -.64186 | -.40484 | .36413 | .21936 | .02735 | .70434 |
| I7 | .14137 | .07068 | .13112 | -.12011 | -.20915 | -.36936 |
| I8 | -.08481 | -.19502 | -.13963 | -.09657 | .37844 | -.42209 |
| I9 | -.11010 | -.07587 | -.47717 | -.12154 | .00077 | .02767 |
| I10 | -.10437 | .00165 | -.34555 | -.31574 | -.07238 | -.37502 |
| I11 | -.22922 | .16616 | -.24305 | -.08346 | .16749 | .07653 |
| I12 | .13759 | -.23333 | .35260 | .07659 | -.21547 | .08612 |
| I13 | -.27904 | -.18771 | -.13114 | -.30594 | .19039 | .31241 |
| I14 | .13679 | .12198 | .36092 | -.07984 | -.42208 | -.01150 |
| I15 | -.23285 | -.08259 | -.26323 | .10294 | .06154 | .48750 |
| I16 | -.03417 | .27424 | -.37694 | .05093 | .04501 | .16699 |
| I17 | .23222 | .03719 | .61923 | .05129 | -.14724 | -.22434 |
| I18 | .43173 | .23549 | -.01274 | -.31017 | -.36257 | -.47825 |
| I19 | -.29256 | -.12595 | .43921 | .43867 | -.01467 | .37010 |

| | I7 | I8 | I9 | I10 | I11 | I12 |
|---|---|---|---|---|---|---|
| I7 | .79675 | | | | | |
| I8 | -.01629 | .89997 | | | | |
| I9 | -.41972 | -.08344 | .84160 | | | |
| I10 | -.15362 | .32544 | .63547 | .66691 | | |
| I11 | -.31269 | .14635 | .15093 | .19020 | .65056 | |
| I12 | .25075 | -.31761 | -.16623 | -.25004 | -.86671 | .78196 |
| I13 | -.27498 | -.11891 | .10284 | .11598 | .11820 | -.13493 |
| I14 | .28730 | -.53232 | -.11320 | -.02097 | -.12956 | .23003 |
| I15 | -.26583 | -.19266 | .19468 | -.19141 | .16866 | -.02464 |
| I16 | -.47226 | .13866 | .08717 | .00194 | .34693 | -.38675 |
| I17 | .44979 | -.01258 | -.67998 | -.40177 | -.31162 | .27183 |
| I18 | .55828 | -.21554 | .03820 | .07121 | -.24114 | .30905 |
| I19 | .01323 | .04523 | -.36177 | -.36662 | .06911 | -.13537 |

| | I13 | I14 | I15 | I16 | I17 | I18 |
|---|---|---|---|---|---|---|
| I13 | .86323 | | | | | |
| I14 | -.31607 | .79775 | | | | |
| I15 | .25615 | -.28275 | .85148 | | | |
| I16 | .47279 | -.72178 | .26218 | .75080 | | |
| I17 | -.45018 | .39677 | -.65278 | -.38974 | .74370 | |
| I18 | -.50215 | .53397 | -.29558 | -.59246 | .29236 | .76493 |
| I19 | -.32318 | .08741 | -.23256 | -.11290 | .40663 | -.23428 |

--- F A C T O R   A N A L Y S I S ---

| | I19 |
|---|---|
| I19 | .72413 |

Measures of sampling adequacy (MSA) are printed on the diagonal.

Correlation Significance Matrix:
' . ' is printed for diagonal elements.

| | I1 | I2 | I3 | I4 | I5 |
|---|---|---|---|---|---|
| I1 | . | | | | |
| I2 | .18771 | . | | | |
| I3 | .00534 | .04900 | . | | |
| I4 | .00338 | .00000 | .02718 | . | |
| I5 | .00681 | .00000 | .04032 | .00000 | . |
| I6 | .00004 | .00009 | .20217 | .00000 | .00001 |
| I7 | .00015 | .00043 | .00642 | .00001 | .00000 |
| I8 | .01370 | .00000 | .04110 | .00000 | .00000 |
| I9 | .06292 | .00000 | .02358 | .00000 | .00000 |
| I10 | .00025 | .03547 | .24798 | .00081 | .00061 |
| I11 | .11011 | .01283 | .26944 | .02416 | .02455 |
| I12 | .20927 | .00002 | .25181 | .00029 | .00003 |
| I13 | .02744 | .00000 | .05307 | .00000 | .00000 |
| I14 | .05190 | .00010 | .08352 | .00000 | .00000 |
| I15 | .03967 | .00001 | .07447 | .00000 | .00000 |
| I16 | .02680 | .00612 | .04634 | .00001 | .00000 |
| I17 | .06655 | .00000 | .40809 | .00000 | .00000 |
| I18 | .01732 | .00000 | .07197 | .00000 | .00000 |
| I19 | .21634 | .03025 | .18199 | .01871 | .00261 |

| | I6 | I7 | I8 | I9 | I10 |
|---|---|---|---|---|---|
| I6 | . | | | | |
| I7 | .00033 | . | | | |
| I8 | .00004 | .00000 | . | | |
| I9 | .00891 | .00000 | .00000 | . | |
| I10 | .00001 | .02749 | .05505 | .29297 | . |
| I11 | .18092 | .00390 | .00430 | .00081 | .48983 |
| I12 | .07934 | .00029 | .00000 | .00000 | .31708 |
| I13 | .00046 | .00007 | .00000 | .00000 | .02330 |
| I14 | .01578 | .00001 | .00000 | .00000 | .14280 |
| I15 | .01049 | .00036 | .00000 | .00000 | .00409 |
| I16 | .03500 | .00002 | .00000 | .00001 | .05211 |
| I17 | .00011 | .00064 | .00000 | .00000 | .00283 |
| I18 | .00001 | .00262 | .00000 | .00003 | .00056 |
| I19 | .36110 | .02241 | .00422 | .00052 | .24769 |

--- FACTOR ANALYSIS ---

| | I11 | I12 | I13 | I14 | I15 |
|---|---|---|---|---|---|
| I11 | . | | | | |
| I12 | .00000 | . | | | |
| I13 | .00365 | .00000 | . | | |
| I14 | .02658 | .00001 | .00000 | . | |
| I15 | .04257 | .00017 | .00000 | .00000 | . |
| I16 | .09485 | .00046 | .00003 | .00000 | .00000 |
| I17 | .01419 | .00010 | .00000 | .00001 | .00000 |
| I18 | .13692 | .00399 | .00000 | .00001 | .00000 |
| I19 | .01990 | .00013 | .00035 | .00027 | .00004 |

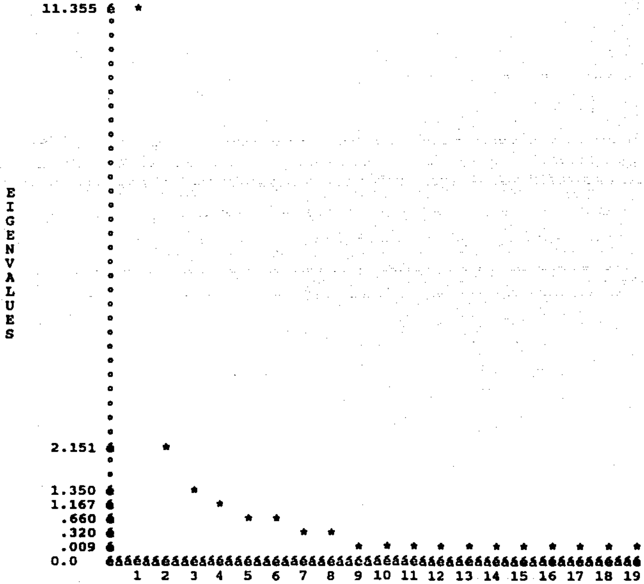
| | I16 | I17 | I18 | I19 |
|---|---|---|---|---|
| I16 | . | | | |
| I17 | .00024 | . | | |
| I18 | .00001 | .00000 | . | |
| I19 | .00024 | .00965 | .00698 | . |

Extraction 1 for Analysis 1, Principal-Components Analysis (PC)

Initial Statistics:

| Variable | Communality | Factor | Eigenvalue | Pct of Var | Cum Pct |
|---|---|---|---|---|---|
| I1 | 1.00000 | 1 | 11.35464 | 59.8 | 59.8 |
| I2 | 1.00000 | 2 | 2.15080 | 11.3 | 71.1 |
| I3 | 1.00000 | 3 | 1.35014 | 7.1 | 78.2 |
| I4 | 1.00000 | 4 | 1.16664 | 6.1 | 84.3 |
| I5 | 1.00000 | 5 | .88837 | 4.7 | 89.0 |
| I6 | 1.00000 | 6 | .65956 | 3.5 | 92.5 |
| I7 | 1.00000 | 7 | .33913 | 1.8 | 94.3 |
| I8 | 1.00000 | 8 | .32015 | 1.7 | 95.9 |
| I9 | 1.00000 | 9 | .26345 | 1.4 | 97.3 |
| I10 | 1.00000 | 10 | .14153 | .7 | 98.1 |
| I11 | 1.00000 | 11 | .09648 | .5 | 98.6 |
| I12 | 1.00000 | 12 | .07238 | .4 | 99.0 |
| I13 | 1.00000 | 13 | .05675 | .3 | 99.3 |
| I14 | 1.00000 | 14 | .04072 | .2 | 99.5 |
| I15 | 1.00000 | 15 | .03401 | .2 | 99.7 |
| I16 | 1.00000 | 16 | .02442 | .1 | 99.8 |
| I17 | 1.00000 | 17 | .01719 | .1 | 99.9 |
| I18 | 1.00000 | 18 | .01467 | .1 | 100.0 |
| I19 | 1.00000 | 19 | .00898 | .0 | 100.0 |

--- F A C T O R   A N A L Y S I S ---




PC Extracted 4 factors.

--- FACTOR ANALYSIS ---

Factor Matrix:

| | FACTOR 1 | FACTOR 2 | FACTOR 3 | FACTOR 4 |
|---|---|---|---|---|
| I1 | .47115 | .56190 | .53848 | .05688 |
| I2 | .83037 | -.02231 | -.20885 | .23521 |
| I3 | .35142 | .15407 | .67494 | -.21166 |
| I4 | .92957 | .19211 | -.08916 | .00980 |
| I5 | .95257 | .11482 | -.08682 | -.04661 |
| I6 | .64782 | .63259 | -.08418 | .31827 |
| I7 | .77269 | .07440 | .42716 | .09731 |
| I8 | .94973 | -.06079 | -.03049 | .06240 |
| I9 | .89310 | -.29292 | .06379 | .05311 |
| I10 | .46574 | .69690 | -.11453 | -.03653 |
| I11 | .48591 | -.43642 | .28563 | .59820 |
| I12 | .74474 | -.49874 | .07792 | .31229 |
| I13 | .94302 | -.08271 | -.12900 | .02998 |
| I14 | .86164 | -.22063 | .03549 | -.25910 |
| I15 | .87646 | -.06935 | -.15375 | -.24514 |
| I16 | .78338 | -.11754 | .12410 | -.43430 |
| I17 | .87134 | .01972 | -.35451 | .09736 |
| I18 | .85685 | .22065 | -.24519 | -.15591 |
| I19 | .56318 | -.46633 | .08642 | -.36920 |

Final Statistics:

| Variable | Communality | Factor | Eigenvalue | Pct of Var | Cum Pct |
|---|---|---|---|---|---|
| I1 | .83091 | 1 | 11.35464 | 59.8 | 59.8 |
| I2 | .78896 | 2 | 2.15080 | 11.3 | 71.1 |
| I3 | .64758 | 3 | 1.35014 | 7.1 | 78.2 |
| I4 | .90905 | 4 | 1.16664 | 6.1 | 84.3 |
| I5 | .93028 | | | | |
| I6 | .92823 | | | | |
| I7 | .79452 | | | | |
| I8 | .91051 | | | | |
| I9 | .89031 | | | | |
| I10 | .71703 | | | | |
| I11 | .86600 | | | | |
| I12 | .90697 | | | | |
| I13 | .91366 | | | | |
| I14 | .85950 | | | | |
| I15 | .85672 | | | | |
| I16 | .83152 | | | | |
| I17 | .89477 | | | | |
| I18 | .86729 | | | | |
| I19 | .67841 | | | | |

--- F A C T O R  A N A L Y S I S ---

Reproduced Correlation Matrix:

| | I1 | I2 | I3 | I4 | I5 |
|---|---|---|---|---|---|
| I1 | .83091* | -.11751 | -.15835 | -.02933 | -.03225 |
| I2 | .27961 | .78896* | .20004 | .00115 | .03846 |
| I3 | .60354 | .09763 | .64758* | .04934 | .00976 |
| I4 | .49846 | .78853 | .29401 | .90905* | -.00218 |
| I5 | .46392 | .79560 | .30371 | .91482 | .93028* |
| I6 | .63345 | .61626 | .20094 | .73435 | .68221 |
| I7 | .64141 | .57364 | .55071 | .69543 | .70296 |
| I8 | .40044 | .81103 | .29060 | .87449 | .89744 |
| I9 | .29357 | .74731 | .30054 | .76876 | .80909 |
| I10 | .54726 | .38651 | .20146 | .57667 | .53531 |
| I11 | .17155 | .49428 | .16969 | .34824 | .36008 |
| I12 | .13037 | .68672 | .17137 | .59258 | .63083 |
| I13 | .33007 | .81890 | .22524 | .87251 | .89859 |
| I14 | .28637 | .65205 | .34760 | .75287 | .80443 |
| I15 | .27724 | .70379 | .24543 | .81271 | .85170 |
| I16 | .34517 | .52506 | .43287 | .69031 | .74220 |
| I17 | .23625 | .82003 | .04936 | .84632 | .85851 |
| I18 | .38679 | .72111 | .20262 | .85922 | .87009 |
| I19 | .02885 | .37317 | .26254 | .42260 | .49262 |

| | I6 | I7 | I8 | I9 | I10 |
|---|---|---|---|---|---|
| I1 | .00409 | -.04337 | -.01056 | -.01728 | .03272 |
| I2 | -.00003 | -.01402 | .02411 | .03661 | -.06306 |
| I3 | -.04832 | -.11571 | .02134 | .05298 | -.07661 |
| I4 | -.02743 | -.02423 | .00676 | -.00086 | -.04215 |
| I5 | -.00239 | .01121 | -.02071 | -.01700 | .01070 |
| I6 | .92823* | .02729 | .04322 | .01140 | -.06450 |
| I7 | .54265 | .79452* | -.00501 | .02930 | -.01672 |
| I8 | .59923 | .72238 | .91051* | .02871 | -.11329 |
| I9 | .40481 | .70071 | .86738 | .89031* | -.10254 |
| I10 | .74058 | .35924 | .40117 | .20257 | .71703* |
| I11 | .20506 | .52322 | .51664 | .61180 | -.13240 |
| I12 | .25979 | .60202 | .75473 | .83277 | -.02105 |
| I13 | .57899 | .67032 | .90645 | .85980 | .39524 |
| I14 | .33317 | .63931 | .81449 | .82266 | .25294 |
| I15 | .45884 | .58254 | .82601 | .78025 | .38643 |
| I16 | .28447 | .60731 | .72027 | .71892 | .28459 |
| I17 | .63778 | .53278 | .84322 | .75497 | .45660 |
| I18 | .66568 | .55859 | .79811 | .67669 | .58661 |
| I19 | -.05493 | .40145 | .53754 | .62547 | -.05910 |

| | I11 | I12 | I13 | I14 | I15 |
|---|---|---|---|---|---|
| I1 | .05130 | .01774 | .01259 | .00650 | .03749 |
| I2 | -.10028 | -.02768 | .01886 | -.04001 | -.02609 |
| I3 | -.05692 | -.04871 | .06576 | -.09729 | .01564 |

--- FACTOR ANALYSIS ---

| | I11 | I12 | I13 | I14 | I15 |
|---|---|---|---|---|---|
| I4 | .00357 | -.01836 | .02928 | .00054 | -.02859 |
| I5 | -.00942 | .01964 | -.02722 | .01195 | -.03031 |
| I6 | -.03836 | -.00462 | -.02160 | .04759 | -.05240 |
| I7 | -.06148 | -.02662 | -.04504 | .04428 | -.01547 |
| I8 | -.05994 | -.03167 | .01058 | .05316 | -.03913 |
| I9 | -.07727 | -.06846 | .01423 | -.00747 | .00304 |
| I10 | .12771 | .10850 | -.04087 | -.05824 | .07293 |
| I11 | .86600* | .04025 | -.01020 | -.02515 | .04355 |
| I12 | .78861 | .90697* | -.03507 | .00163 | -.00444 |
| I13 | .47541 | .74286 | .91366* | -.04116 | .00002 |
| I14 | .37011 | .67359 | .81845 | .85950* | -.05581 |
| I15 | .26559 | .59879 | .84473 | .82855 | .85672* |
| I16 | .20760 | .51608 | .71944 | .81786 | .78214 |
| I17 | .37177 | .64186 | .86870 | .70862 | .79296 |
| I18 | .15676 | .46029 | .81673 | .72131 | .81160 |
| I19 | .28100 | .54344 | .54744 | .68687 | .60316 |

| | I16 | I17 | I18 | I19 |
|---|---|---|---|---|
| I1 | -.00079 | .03505 | -.01218 | .11484 |
| I2 | -.08735 | -.04646 | -.00492 | -.03768 |
| I3 | -.13061 | -.00659 | .06160 | -.09657 |
| I4 | -.00958 | -.02388 | .03582 | -.05314 |
| I5 | .01349 | -.03116 | -.00862 | -.01059 |
| I6 | .04001 | -.03039 | .00291 | -.01045 |
| I7 | .05493 | .01137 | -.07687 | -.04436 |
| I8 | .00414 | -.02915 | .00847 | -.07982 |
| I9 | -.03394 | .04703 | -.03010 | -.07296 |
| I10 | .00794 | .02138 | -.03694 | .18412 |
| I11 | .03036 | .01584 | .04266 | .08429 |
| I12 | .04087 | -.02879 | .00028 | .06020 |
| I13 | -.07534 | .00032 | .03636 | .02090 |
| I14 | .09350 | -.03291 | -.04003 | -.10986 |
| I15 | -.05830 | .07542 | -.04246 | .03785 |
| I16 | .83152* | -.01244 | -.00826 | -.08610 |
| I17 | .59399 | .89477* | -.04833 | -.00346 |
| I18 | .68259 | .82269 | .86729* | .01424 |
| I19 | .66706 | .41494 | .41603 | .67841* |

The lower left triangle contains the reproduced correlation matrix; The diagonal, communalities; and the upper right triangle, residuals between the observed correlations and the reproduced correlations.

There are 45 (26.0%) residuals (above diagonal) that are > 0.05

--- F A C T O R   A N A L Y S I S ---

Varimax Rotation 1, Extraction 1, Analysis 1 - Kaiser Normalization.

Varimax converged in 10 iterations.

Rotated Factor Matrix:

| | FACTOR 1 | FACTOR 2 | FACTOR 3 | FACTOR 4 |
|---|---|---|---|---|
| I1 | .00938 | .47757 | .07068 | .77314 |
| I2 | .48021 | .56983 | .48170 | -.04019 |
| I3 | .22202 | .00283 | .04244 | .77232 |
| I4 | .58333 | .67535 | .27285 | .19553 |
| I5 | .66197 | .61716 | .27712 | .18546 |
| I6 | .03478 | .91540 | .18156 | .23685 |
| I7 | .39957 | .34319 | .43437 | .57307 |
| I8 | .65642 | .49812 | .45346 | .16083 |
| I9 | .69925 | .26532 | .55354 | .15673 |
| I10 | .09419 | .78182 | -.19620 | .24170 |
| I11 | .12079 | .00349 | .91281 | .13481 |
| I12 | .52097 | .09991 | .79043 | .02832 |
| I13 | .69212 | .50588 | .41727 | .06789 |
| I14 | .83593 | .22931 | .26527 | .19436 |
| I15 | .80472 | .41708 | .17173 | .07549 |
| I16 | .83230 | .18335 | .06982 | .31672 |
| I17 | .59261 | .63878 | .34529 | -.12776 |
| I18 | .63896 | .66959 | .07397 | .07216 |
| I19 | .77927 | -.15467 | .18965 | .10607 |

Factor Transformation Matrix:

| | FACTOR 1 | FACTOR 2 | FACTOR 3 | FACTOR 4 |
|---|---|---|---|---|
| FACTOR 1 | .70075 | .54166 | .40365 | .22941 |
| FACTOR 2 | -.38548 | .71431 | -.47916 | .33402 |
| FACTOR 3 | -.13160 | -.35974 | .19847 | .90215 |
| FACTOR 4 | -.58569 | .25875 | .75371 | -.14807 |

Horizontal Factor 1 Vertical Factor 2

```
                     6
                     °
                     °10
                     °         418
                     °        2
                     1           81315
                     °        7
                     °             914
                     °              16
                     °         12
áááááááááááááááááááááéááááááááááááááááááááá
                     °11 3        19
                     °
                     °
                     °
                     °
                     °
                     °
                     °
                     °
                     °
```

| Symbol | Variable | Coordinates | |
|---|---|---|---|
| 1 | I1 | .009 | .478 |
| 2 | I2 | .480 | .570 |
| 3 | I3 | .222 | .003 |
| 4 | I4 | .583 | .675 |
| 5 | I5 | .662 | .617 |
| 6 | I6 | .035 | .915 |
| 7 | I7 | .400 | .343 |
| 8 | I8 | .656 | .498 |
| 9 | I9 | .699 | .265 |
| 10 | I10 | .094 | .782 |
| 11 | I11 | .121 | .003 |
| 12 | I12 | .521 | .100 |
| 13 | I13 | .692 | .506 |
| 14 | I14 | .836 | .229 |
| 15 | I15 | .805 | .417 |
| 16 | I16 | .832 | .183 |
| 17 | I17 | .593 | .639 |
| 18 | I18 | .639 | .670 |
| 19 | I19 | .779 | -.155 |

Factor Score Coefficient Matrix:

| | FACTOR 1 | FACTOR 2 | FACTOR 3 | FACTOR 4 |
|---|---|---|---|---|
| I1 | -.15267 | .07823 | .00747 | .44937 |
| I2 | -.04248 | .14002 | .15575 | -.15609 |
| I3 | .03455 | -.15885 | -.05936 | .50888 |
| I4 | .02671 | .13408 | -.01653 | -.01221 |
| I5 | .07007 | .09637 | -.03459 | -.01502 |
| I6 | -.22498 | .33401 | .07535 | .01469 |
| I7 | -.05614 | -.03066 | .13655 | .30024 |
| I8 | .04115 | .04708 | .08314 | -.01854 |
| I9 | .07473 | -.05989 | .14070 | .00844 |
| I10 | -.06666 | .27608 | -.17914 | .04574 |
| I11 | -.21995 | -.06519 | .54296 | .05698 |
| I12 | -.02902 | -.08161 | .35079 | -.04997 |
| I13 | .07055 | .05854 | .05235 | -.08379 |
| I14 | .21934 | -.09909 | -.08239 | .03974 |
| I15 | .20457 | .00537 | -.13437 | -.06468 |
| I16 | .27535 | -.13106 | -.20831 | .13562 |
| I17 | .03591 | .16417 | .03737 | -.22857 |
| I18 | .11550 | .14490 | -.15546 | -.09247 |
| I19 | .29526 | -.23292 | -.10191 | .04356 |