

ANÁLISIS FACTORIAL MÚLTIPLE Y METODOLOGÍA STATIS EN EL ANÁLISIS DE LA TRANSICIÓN DEMOGRÁFICA EN EL PERÚ. 1995-2010

Emma N. Cambillo M.¹, Ysela D. Aguero P.², Dora F. Rafael H.³.

Resumen: Se comparan el análisis factorial múltiple y la metodología STATIS, utilizados para encontrar una estructura o espacio común subyacente a un conjunto de datos, expresado como una tabla múltiple ("Multiway"). Se aplican ambos métodos en el estudio de la transición demográfica en las 25 regiones del Perú en el periodo 1995 a 2010, considerando indicadores de fecundidad (Tasa bruta de natalidad y tasa global de fecundidad) y tres indicadores de mortalidad (Tasa bruta de mortalidad, tasa de mortalidad infantil y esperanza de vida al nacer). Se utiliza la tipología elaborada por el CELADE (1992), respecto a las etapas de la transición demográfica. Se identifican cuatro grupos de regiones de acuerdo al comportamiento de sus indicadores de mortalidad y fecundidad, obteniéndose una visión sintética de los cambios en el tiempo y las perspectivas futuras de las regiones del Perú. Los datos se procesaron utilizando los paquetes estadísticos SPAD, STATISTICA y SPSS.

Palabras Claves: Tablas múltiples, Análisis factorial múltiple, Metodología STATIS, Transición demográfica, Fecundidad, Mortalidad, Esperanza de vida al nacer.

MULTIPLE FACTOR ANÁLISYS AND STATIS METODOLOGY IN THE DEMOGRAPHIC TRANSICION ANALISYS IN PERU. 1995-2010

Abstract: The multiple factor analysis and STATIS methodology used to find a common space or underlying structure expressed like multiway data, are compared. Both methods are applied in the study of the demographic transition in the 25 regions of Peru during the period 1995-2010, considering fertility indicators (Crude birth rate and total fertility rate) and three indicators of mortality (Crude death rate, rate infant mortality and life expectancy at birth). CELADE typology (1992) about to the stages of the demographic transition is used. Four groups of regions according to the behavior of their mortality and fertility indicators are identified for to give a synthetic view of the changes over time and future prospects of the peruvian regions of Peru. The data were processed using the SPAD statistical packages, SPSS and STATISTICA.

Key Words: Multiple Tables, Multiple Factor Analysis, Methodology STATIS, Demographic Transition, fertility, mortality, life expectancy at birth.

1. Introducción

El análisis estadístico de tablas múltiples (Multiway), en particular, de arreglos tridimensionales, donde las entradas pueden ser individuos (i), variables (j) y ocasiones (k) tiene por objeto encontrar una estructura o espacio común subyacente al conjunto de datos (Figura 1).

Una tabla múltiple tridimensional, se puede representar como una matriz X de orden $(I \times J) \times K$.

$$X = \begin{bmatrix} X_{111} & X_{1j1} & X_{1J1} \\ X_{i11} & X_{ij1} & X_{iJ1} \\ X_{I11} & X_{Ij1} & X_{IJ1} \end{bmatrix} \cdots \begin{bmatrix} X_{11K} & X_{1jK} & X_{1JK} \\ X_{i1K} & X_{ijK} & X_{iJK} \\ X_{I1K} & X_{IjK} & X_{IJK} \end{bmatrix}$$

¹UNMSM, Facultad de Ciencias Matemáticas, e-mail: ecambillom@unmsm.edu.pe

²UNMSM, Facultad de Ciencias Matemáticas, e-mail: yaguerop@unmsm.edu.pe

³UNMSM, Facultad de Ciencias Matemáticas, e-mail: litaperu895p@hotmail.com

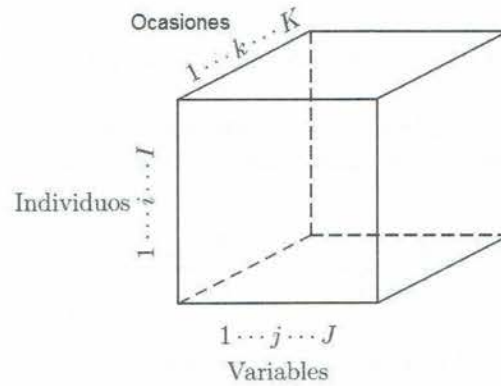


Figura 1: Arreglo de tres dimensiones

De manera compacta, se representa como: $X = [X_1 | \dots | X_k | \dots | X_K]$.

El interés cuando se analizan tablas múltiples es que, además de proporcionar un panorama global de los cambios, sea posible conocer la evolución de los individuos a través del tiempo. La elección del método de análisis de los datos depende de los objetivos de la investigación, de si las J variables son continuas o discretas, y si la matriz de datos es un arreglo de tablas múltiples o un conjunto de tablas múltiples.

Existen varios métodos estadísticos para el análisis de tablas múltiples, tales como; el Análisis de Componentes Principales de grupos comunes, el Análisis de Componentes Principales Triple, el Análisis Procrustes Generalizado, la Metodología STATIS y el Análisis Factorial Múltiple. Estos métodos usualmente, son utilizados con fines de exploración ya que permiten identificar patrones de comportamiento a partir del análisis de la estructura interna del arreglo, generando representaciones gráficas que permiten una mayor comprensión de los resultados. En particular, la Metodología STATIS y el Análisis Factorial Múltiple, han sido desarrolladas específicamente para el tratamiento de tablas múltiples ("multiway"), como se puede observar en las investigaciones realizadas utilizando estas técnicas.

En un estudio realizado por Di Cesare (2007), se aplicó la metodología STATIS con el objetivo de estudiar los perfiles de la transición demográfica y epidemiológica de Nicaragua y la interacción entre ambos procesos. Álvarez R. (1996), realizó una investigación para caracterizar la Comunidad Autónoma de Castilla y León según la producción agrícola (cereales, legumbres, remolacha, patatas y hortalizas) y la producción cárnica (ovino, bovino, porcino y caprino) durante el periodo 1985 -1993; utilizando la metodología STATIS. Cabezas (2008), utilizó el Análisis Factorial Múltiple para caracterizar los departamentos del Perú, proporcionando una visión global de los indicadores demográficos y su evolución en el periodo 1996 - 2000.

En otro estudio realizado por Arredondo, J. (2006) se aplicó el Análisis Factorial Múltiple para estudiar la situación de América Latina y El Caribe Hispano entre los años 1990 y 2002, mediante el Índice de Desarrollo Humano (IDH) del Programa de las Naciones Unidas para el Desarrollo.

En los estudios mencionados se observa que tanto la metodología STATIS como el Análisis Factorial Múltiple, buscan una estructura común entre diferentes tablas múltiples, aunque por caminos diferentes y el interés es conocer si se pueden considerar como métodos alternativos o complementarios.

En primer lugar se describe brevemente en que consiste cada uno de los dos métodos:

- a) La metodología STATIS, permite el tratamiento simultáneo de tablas múltiples. El análisis se realiza en tres fases (o etapas). En la fase 1 (Inter estructura): se analizan las ocasiones globalmente. En la fase 2 (compromiso), se obtiene la estructura media de las variables que determina el plano de compromiso, que será utilizado en la fase 3 (intra estructura), donde se proyectan las unidades y las variables.
- b) El Análisis Factorial Múltiple, permite la evaluación simultánea de varias matrices de datos correspondientes a un mismo conjunto de individuos.

Es importante resaltar que las variables incluidas en la tabla múltiple deben ser todas cuantitativas o todas categóricas.

El análisis se realiza en dos etapas:

Primera etapa.- Se aplica el método de componentes principales en cada sub tabla, X_k , de $X = [X_1 | \dots | X_k | \dots | X_K]$ y se obtiene una nueva tabla múltiple de variables ortogonales $Z = [Z_1 | \dots | Z_k | \dots | Z_K]$.

Segunda etapa.- Cada sub tabla Z_k es ponderada por el inverso de la raíz cuadrada de su mayor autovalor, (λ_1^k) , obteniéndose una tabla múltiple de la forma:

$$Z = \left[\frac{1}{\sqrt{\lambda_1^1}} Z_1, \dots, \frac{1}{\sqrt{\lambda_1^k}} Z_k, \dots, \frac{1}{\sqrt{\lambda_1^K}} Z_K \right]$$

El siguiente paso es, aplicar el análisis de componentes principales a la tabla múltiple Z y reducir su dimensión seleccionando m sub tablas ($m < K$).

El objetivo del presente trabajo es explorar la transición demográfica en las 25 regiones del Perú en el periodo 1995 a 2010, utilizando indicadores de Fecundidad y Mortalidad, además se comparan los resultados obtenidos a partir de los dos métodos de análisis antes mencionados.

2. Metodología

La transición demográfica es una teoría demográfica que permite explicar los cambios en el tamaño de la población, desde niveles elevados de mortalidad y natalidad, a un descenso de la mortalidad y posterior descenso de la natalidad lo cual significa un incremento en el tamaño de la población con las respectivas implicancias de este incremento.

En el análisis de la transición demográfica de las regiones, los indicadores generalmente utilizados en el proceso de operacionalización de la fecundidad y mortalidad son las tasas brutas de natalidad (TBN) y mortalidad (TBM).

En este estudio se utilizan la Tasa bruta de natalidad (TBN), la Tasa Global de Fecundidad (TGF), la Tasa bruta de mortalidad (TBM), la Tasa de Mortalidad Infantil (TMI) y la Esperanza de Vida al Nacer (EVN).

El CELADE (1992), ha propuesto una clasificación de las regiones de acuerdo a la etapa de la transición demográfica en que se encuentran. Se consideran cuatro grupos:

Grupo I: Transición incipiente.- Regiones con altas tasas de fecundidad y mortalidad, tienen una estructura por edades muy joven y una elevada relación de dependencia.

Grupo II: Transición moderada.- Representado por regiones con altas tasas de fecundidad y mortalidad moderada. El descenso de la mortalidad, sobre todo durante el primer año de vida, se ha traducido en un rejuvenecimiento de la estructura por edades, lo que también lleva a una elevada relación de dependencia.

Grupo III: En plena transición.- Formado por regiones con fecundidad moderada y mortalidad moderada o baja. Al descender la fecundidad, disminuye la relación de dependencia.

Grupo I: Transición avanzada.- Transición avanzada.- Considera regiones con fecundidad y mortalidad baja. Con una baja relación de dependencia y una elevada esperanza de vida al nacer.

La estructura de los datos para el estudio corresponde a un arreglo de tablas múltiples triple donde los individuos corresponden a las 25 regiones ($I = 25$), las condiciones son los periodos de tiempo 1995, 2000, 2005, 2010 ($K = 4$) y las variables corresponden a los cinco indicadores demográficos obtenidos para cada periodo de tiempo 1 ($J = 5$).

Los datos fueron obtenidos del boletín: "Perú: Estimaciones y Proyecciones de Población Departamental por Años Calendario y Edades Simples 1995 – 2025", Instituto Nacional de Estadística e Informática INEI. Dirección Técnica de Demografía e Indicadores Sociales, en Noviembre 2010.

El procesamiento de los datos se realizó utilizando los paquetes SPSS versión 20, STATISTICA versión 7.0 y el módulo para análisis simultáneo de varias tablas rectangulares de datos del paquete estadístico SPAD4.5 (Systeme Pour l Analyse des Données). En este módulo, se incluye el Análisis Factorial Múltiple y la metodología STATIS (Structuration des Tableaux a Tríos Indices de la Statistique).

3. Resultados y Discusión

El conjunto de datos corresponden a un arreglo de dimensión $25 \times 5 \times 4$. Las variables son los $j = 5$ indicadores demográficos: dos de fecundidad (TBN, TGF) y tres de mortalidad (TBM, TMI, EVN)

observados en las $I = 25$ regiones políticas del Perú, en $K = 4$ periodos de tiempo (1995, 2000, 2005 y 2010).

La matriz de dispersión relacionando los indicadores demográficos de las 25 regiones, para el periodo 2005 se muestra en la Figura 2. Las tendencias en los periodos siguientes son similares.

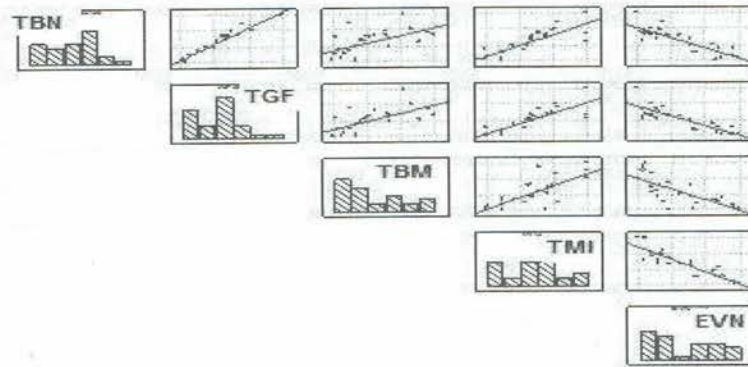


Figura 2: Asociaciones bivariadas entre los indicadores demográficos. Periodo 2005.

En los cuatro periodos de tiempo, se observan coeficientes de correlación de Pearson elevadas y positivas entre indicadores de la misma componente de la dinámica demográfica (Ej. TBN.TGF, TBM-TMI) o negativas como en el caso de la Esperanza de vida al nacer (EVN) que presenta coeficientes de correlación elevados y negativos con todos los demás indicadores (Ver Cuadro 1).

Cuadro 1: Coeficientes de correlación de Pearson entre los indicadores demográficos (variables) para las 25 regiones. Periodos 1995, 2000, 2005 y 2010.

VARIABLES	Años			
	1995	2000	2005	2010
TBN - TGF	0,97	0,98	0,98	0,98
TBN - TBM	0,74	0,66	0,57	0,50
TBN - TMI	0,80	0,83	0,83	0,77
TBN - EVN	-0,76	-0,81	-0,84	-0,83
TGF - TBM	0,84	0,76	0,64	0,52
TGF - TMI	0,87	0,85	0,82	0,78
TGF - EVN	-0,82	-0,83	-0,84	-0,85
TBM - TMI	0,88	0,86	0,76	0,66
TBM - EVN	-0,83	-0,79	-0,74	-0,66
TMI - EVN	-0,88	-0,88	-0,88	-0,88

3.1. Análisis factorial de los indicadores demográficos

Este método proyecta los indicadores demográficos y las regiones, pasando de un espacio de dimensión mayor a otro bidimensional, constituido por los factores 1 y 2.

En los cuatro periodos de tiempo, se observa que el análisis factorial de los indicadores demográficos identifica dos dimensiones o factores que explican entre el 92,6% y el 94,4% de la variabilidad total de los indicadores demográficos (Ver Cuadro 2).

Cuadro 2: Porcentaje de variación explicada por los dos primeros factores. Análisis factorial de las variables demográficas. Periodos 1995,2000, 2005, 2010.

FACTOR	Porcentaje de la variabilidad explicada			
	1995	2000	2005	2010
1	87,3	86,1	83,6	80,0
2	94,4	94,3	93,7	92,6

Las relaciones entre los indicadores de natalidad, mortalidad y esperanza de vida al nacer se mantienen en todo el periodo de observación, sin embargo las posiciones de las regiones en las nuevas dimensiones (planos de proyección) varían dependiendo de los cambios y tendencias de los cinco indicadores demográficos a través del tiempo.

En el año 1995, Huancavelica, Ayacucho, Pasco, Loreto, Junín, Ucayali, Cajamarca y Madre de Dios, presentan tasas de fecundidad elevadas; Cusco, Apurímac, Puno, Huánuco, Amazonas, San Martín y Ancash, muestran tasas más altas de mortalidad y La Libertad, Lambayeque, Ica, Lima, Arequipa, Tumbes, Tacna y Moquegua, tienen esperanzas de vida más elevadas.

En el año 2000, Huancavelica, Pasco, Loreto, Ucayali, Cajamarca, San Martín y Madre de Dios, muestran tasas fecundidad elevadas; Ayacucho, Junín, Cusco, Apurímac, Puno, Huánuco, Amazonas y Ancash, presentan tasas de mortalidad más elevadas y La Libertad, Piura, Lambayeque, Ica, el Callao, Lima, Arequipa, Tumbes, Tacna y Moquegua, tienen esperanzas de vida más elevadas.

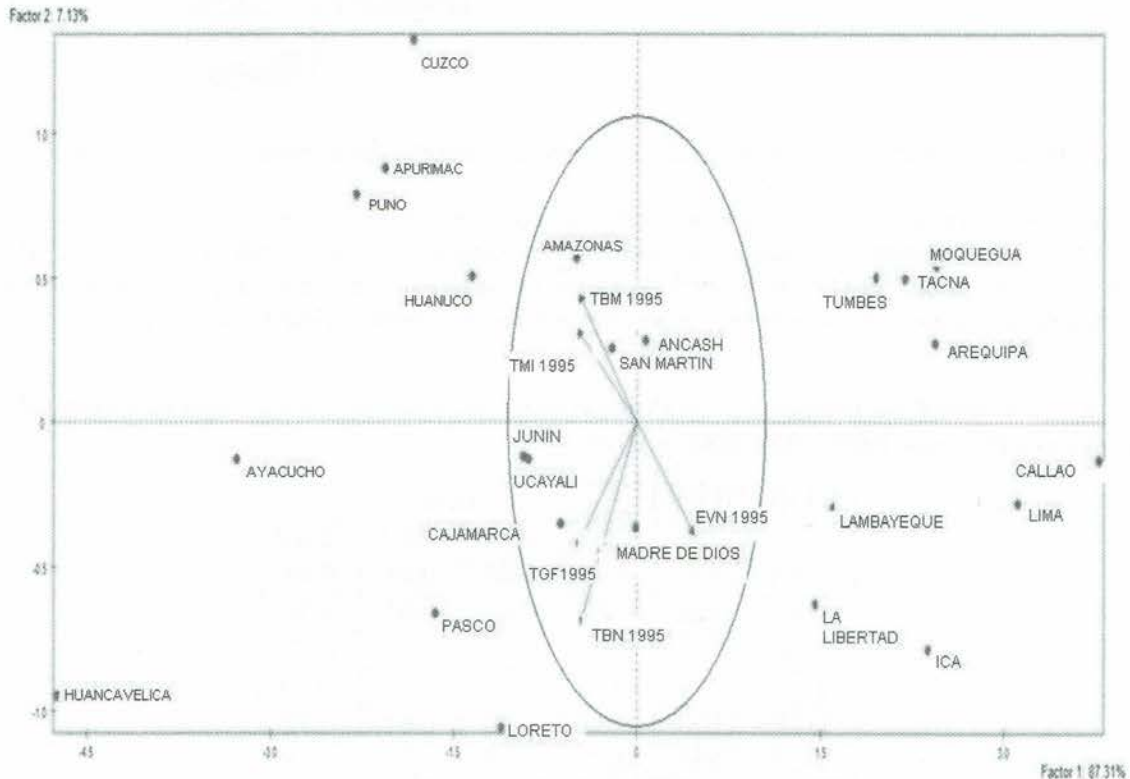


Figura 3: Plano factorial de las regiones del Perú. Año 1995.

En el año 2005, Huancavelica, Huánuco, Pasco, Loreto, Ucayali, Cajamarca, San Martín y Madre de Dios, presentan tasas más elevadas de fecundidad; en Ayacucho, Junín, Cusco, Apurímac, Puno, Huánuco, Amazonas y Ancash, se observa elevadas tasas de mortalidad y en La Libertad, Piura, Lambayeque, Ica, el Callao, Lima, Arequipa, Tumbes, Tacna y Moquegua, se observan esperanzas de vida al nacer más elevadas.

En el año 2010, Huancavelica, Huánuco, Pasco, Loreto, Ucayali, Cajamarca, San Martín y Madre de Dios, presentan tasas de fecundidad elevadas; Ayacucho, Junín, Cusco, Apurímac, Puno, Huánuco, Amazonas y Ancash, muestran tasas de mortalidad más elevadas, y La Libertad, Piura, Lambayeque, Ica, el Callao, Lima, Arequipa, Tumbes, Tacna y Moquegua, presentan esperanzas de vida al nacer más elevadas.

El siguiente paso es, identificar las trayectorias de las unidades y las variables en las diferentes ocasiones. El estudio de las trayectorias de las unidades (regiones) y las variables (indicadores) en las diferentes ocasiones (periodos) (Figura 4) muestra la evolución y el comportamiento de las 25 regiones a partir del comportamiento de los cinco indicadores demográficos, en los cuatro periodos.

La transición demográfica a partir del Análisis factorial múltiple

En la figura 4 se observa que entre las regiones en transición incipiente se encuentran Puno, Cusco y dos de las regiones que conforman el trapecio andino (Apurímac, Huancavelica), que presentan

indicadores de fecundidad y mortalidad elevados con una esperanza de vida al nacer relativamente baja).

Las regiones, Ayacucho, Loreto, Pasco, Madre de Dios, Pasco, Ucayali, Huánuco, Amazonas, Junín, Cajamarca, Piura, San Martín y Ancash presentan trayectorias relativamente más cortas, lo que indica que se encuentran en un proceso de transición moderada.

Las regiones que se encuentran en plena transición con tasas de fecundidad y mortalidad moderadas son, La Libertad, Lambayeque, Tumbes, Tacna, Moquegua y Arequipa.

FACTOR 2: 9.15%

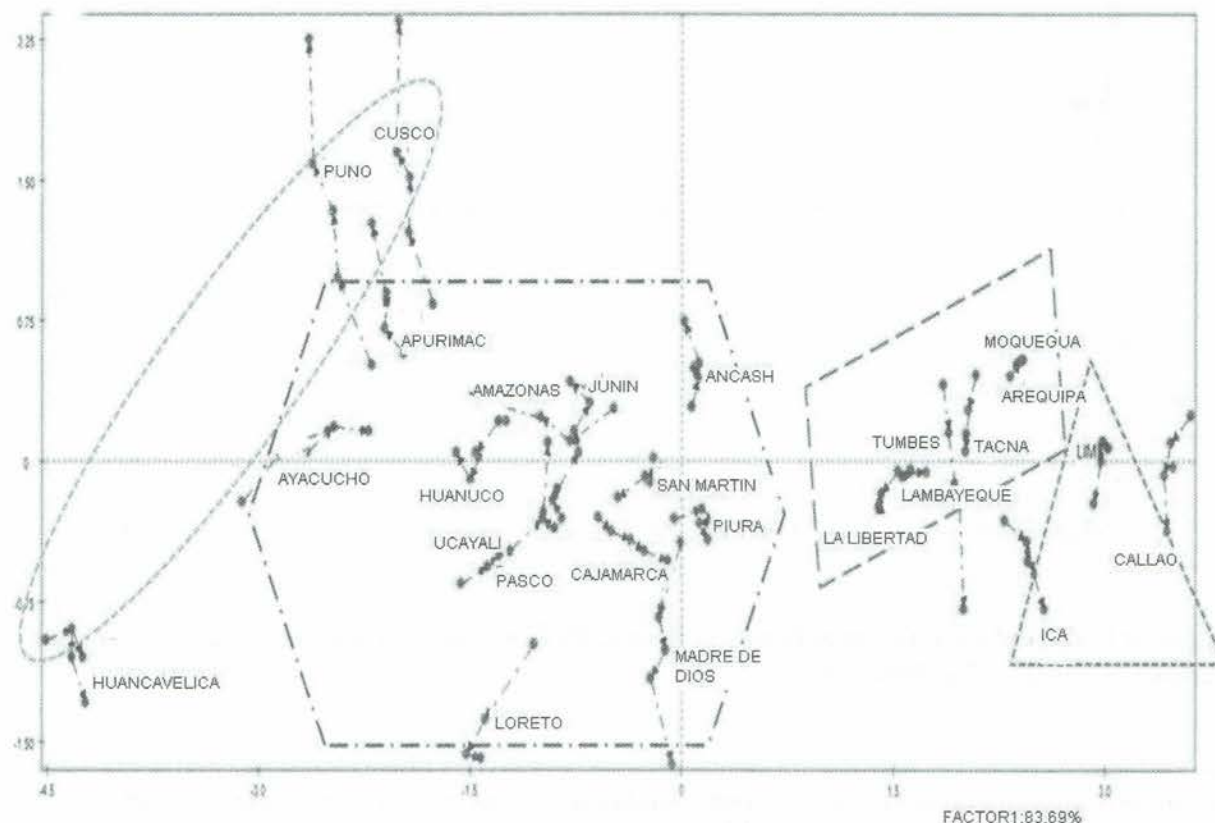


Figura 4: Plano factorial de las trayectorias de las regiones del Perú. Periodo 1995-2010.

Finalmente, las regiones de Ica, el Callao y Lima, presentan bajas tasas de mortalidad y fecundidad y esperanza de vida al nacer elevada, lo cual indica que se encuentran en una etapa de transición avanzada.

3.2. Metodología STATIS aplicado a de los indicadores demográficos

Este método, analiza los datos en tres etapas (ICI: Inter estructura-Compromiso-Intra estructura).

Etapa 1. Análisis de la Inter estructura.

Para la comparación global de las 25 regiones se construye una matriz superpuesta de los indicadores demográficos en cada uno de los años, obteniéndose una matriz de orden 125x4, (25 regiones por 5 indicadores). Se realiza el análisis de componentes principales y se obtiene un nuevo espacio de proyección de los años.

Etapa 2. Compromiso.

Para la etapa de compromiso se construye una matriz de correlación global entre los indicadores demográficos ponderada, con el autovalor correspondiente para cada año y se aplica análisis de componentes principales, obteniéndose dos nuevas dimensiones, en las cuales se observan correlaciones directas entre los indicadores de natalidad y mortalidad e inversas entre la esperanza de vida al nacer y todos los demás indicadores. El cambio significativo se puede observar en la tasa bruta de mortalidad y la tasa de mortalidad infantil.

Etapa 3. Análisis de la Intra estructura.

En esta etapa se representan las regiones en la nube compromiso. Las correlaciones de las regiones con

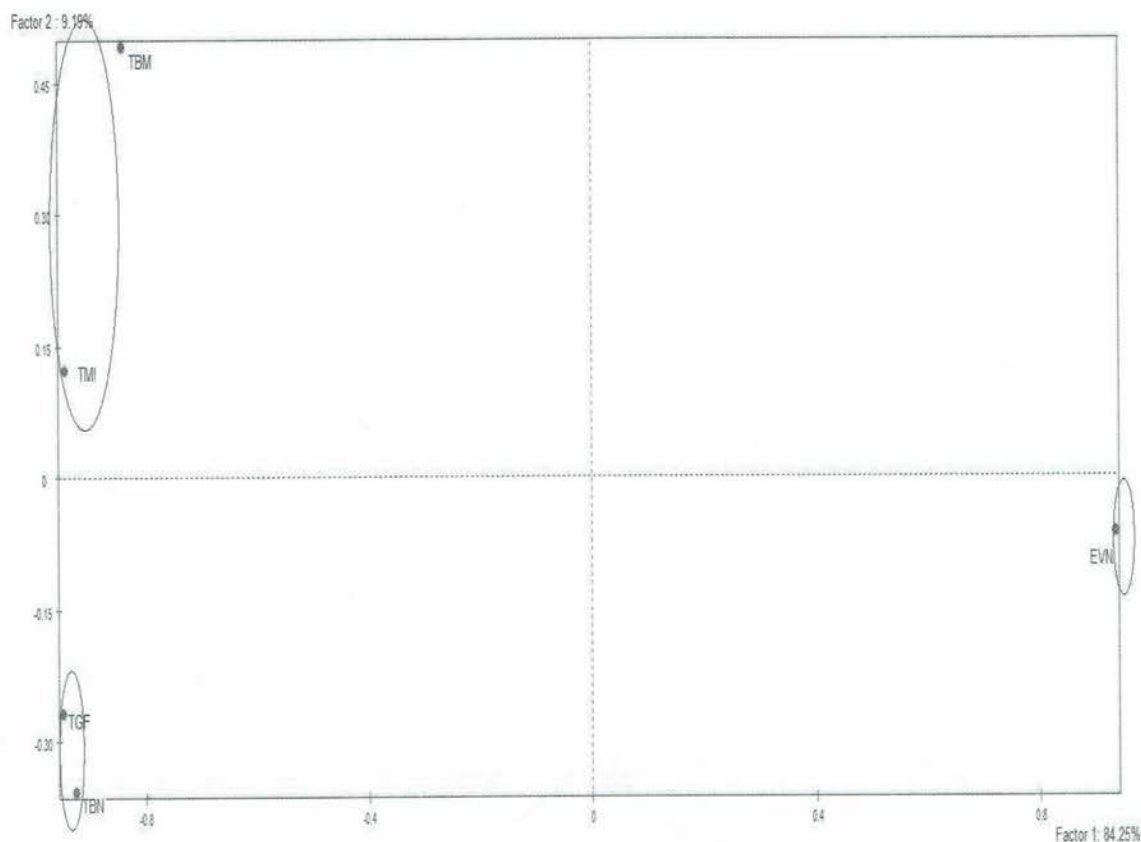


Figura 5: Representación de los indicadores demográficos en las dimensiones obtenidas de la matriz de correlación global. Años 1995-2010.

los dos factores permiten dotarlos de significado, definiendo dos dimensiones relevantes que configuran su estructura. Casi todos los puntos correspondientes a una misma región para los distintos años se agrupan según los factores, lo cual facilita la interpretación (Figura 4).

La transición Demográfica a partir de la metodología STATIS

En general, los resultados del análisis de datos mediante esta metodología permiten concluir que en el periodo 1995 a 2010, se produjo una reducción de las tasas de mortalidad y fecundidad y se incrementó la esperanza de vida al nacer (Figura 5).

En el grupo de transición incipiente (Grupo I) se encuentran las regiones de Huancavelica, Cusco y Puno, que por su elevada fecundidad, tienen una estructura por edades joven y una elevada relación de dependencia.

Entre las regiones con elevadas tasas de fecundidad, mortalidad moderada (Grupo II) se encuentran Ayacucho, Loreto, Pasco, Huánuco, Apurímac, Cajamarca, Junín, San Martín, Piura, Madre de Dios, Amazonas, Ucayali, Ancash y Lambayeque. El descenso de la mortalidad, sobre todo durante el primer año de vida, se traduce en un rejuvenecimiento de la estructura por edades, lo que también lleva a una elevada relación de dependencia.

Las regiones con fecundidad moderada y mortalidad moderada o baja (Grupo III) son, La Libertad, Tumbes, Tacna, Arequipa y Moquegua. Con el descenso de la fecundidad, se produjo un descenso de la relación de dependencia.

Las regiones que se encuentran en la etapa de transición avanzada, son Ica, Lima y El Callao, que presentan tasas de fecundidad y mortalidad bajas y en consecuencia una baja relación de dependencia y una elevada esperanza de vida al nacer.

En resumen, las regiones que se ubican en los grupos I y II presentan una estructura por edades jóvenes, con más del 50% de la población por debajo de los 15 años; las regiones del grupo III, tienen una estructura que podría denominarse "plena transición", con aproximadamente un tercio de la

FACTOR2:3.74%

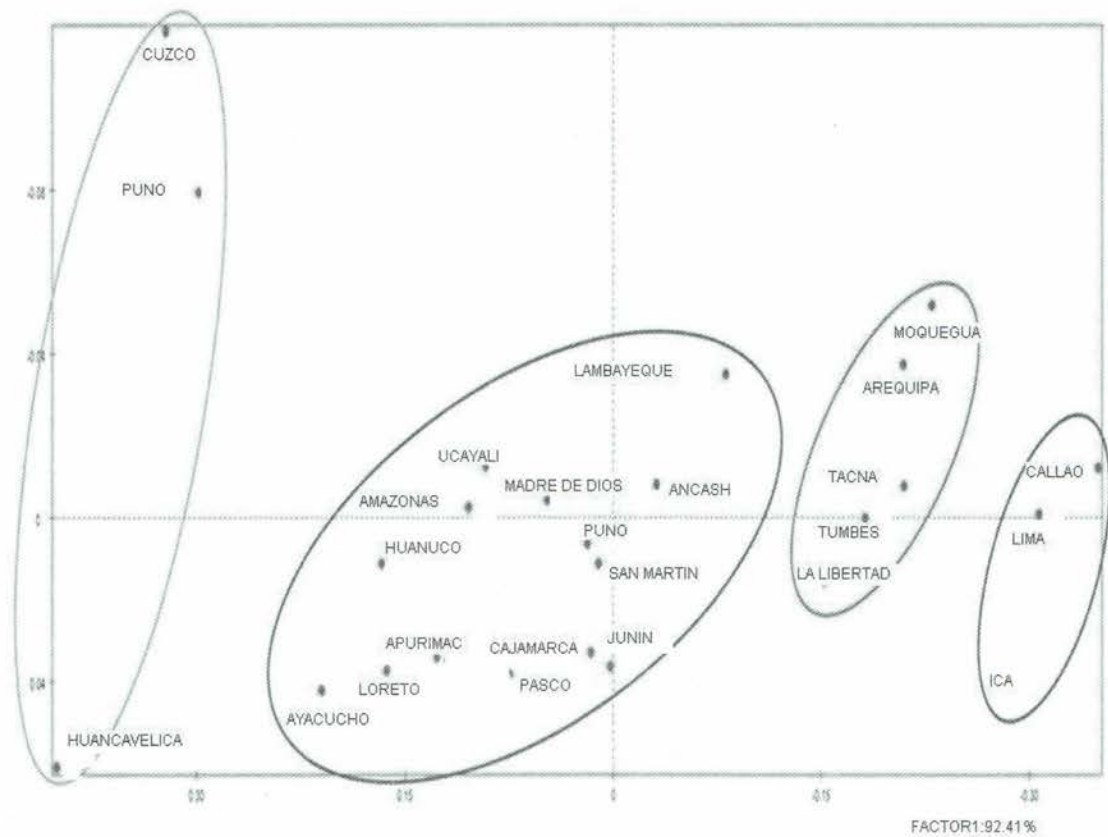


Figura 6: Regiones del Perú proyectada en el plano de compromiso determinado por los indicadores a través de los años.

población por debajo de los 15 años (32 a 36%). Finalmente, las regiones del grupo IV, se caracterizan por presentar una tendencia al envejecimiento, con menos del 30% de su población joven (Figura 6).

4. Conclusiones

La Metodología STATIS y el Análisis Factorial Múltiple, permiten realizar comparaciones a nivel de ocasiones y evalúan la estructura interna de las tablas múltiples.

El análisis factorial evalúa el comportamiento de cada región con respecto al promedio en cada una de las ocasiones, mientras que la metodología STATIS se centra en el estudio de la estructura común y de las ocasiones.

Al ser usados para realizar el análisis de la transición demográfica de las 25 regiones del Perú, se observó que ambos métodos estadísticos clasifican las regiones de manera muy similar, aunque los procedimientos difieren, y este hecho enriquece el análisis exploratorio. Por lo tanto, podrían ser considerados como complementarios y se recomienda su uso en el análisis de datos.

La simplificación de la estructura subyacente en los indicadores en las regiones a través del tiempo (1995–2010), utilizando los dos métodos, permite identificar grupos de regiones según la clasificación de CELADE (1992) en Transición incipiente, transición moderada, plena transición y transición avanzada.

REFERENCIAS BIBLIOGRÁFICAS

- [1] Álvarez, R. (1996). *Rasgos Básicos del Sector Agrario Castellano - Leones. Un Nuevo Planteamiento Espacial y Temporal: Metodología STATIS*. Artículo presentado en el 5º Congreso de Economía Regional. 23-37. Universidad de León- España. Recuperado de: [http : //www.jcyl.es/jcyl/cee/dgeae/congresos_coreg/CERCL/CER2544.HTM](http://www.jcyl.es/jcyl/cee/dgeae/congresos_coreg/CERCL/CER2544.HTM).
- [2] Arredondo, J. Carrasco, S. (2006). *Aportación del Análisis Factorial Múltiple (AFM) al Análisis de las Políticas Públicas. Una Visión de América Latina y el Caribe, 1990-2002*. Estudios Económicos de Desarrollo Internacional, 6(2), 91-118. Recuperado de: [http : //www.usc.es/economet/reviews/eedi625.pdf](http://www.usc.es/economet/reviews/eedi625.pdf).
- [3] Cabezas, G. (2008). *Caracterización Demográfica de los Departamentos del Perú: Usando Análisis Factorial Múltiple*. Monografía para Licenciado en Estadística. UNMSM, Lima.
- [4] CELADE (1992). *La Transición Demográfica en América Latina*. Centro Latinoamericano y Caribeño de Demografía. (CELADE / CEPAL). Recuperado de: [http : //www.eclac.cl/Celade/SitDem/DEsitDemTransDemDoc00e.html](http://www.eclac.cl/Celade/SitDem/DEsitDemTransDemDoc00e.html).
- [5] Di Cesare, M. (2007). *Las Interacciones entre transición demográfica y epidemiológica en Nicaragua: Implicancias para las Políticas Públicas en Salud*. Centro Latinoamericano y Caribeño de Demografía (CELADE)-División de Población de la CEPAL Recuperado de: <http://www.eclac.org/publicaciones/xml/7/32067/lcl2822-P.pdf>.