

UNIVERSIDAD NACIONAL DE INGENIERIA

**FACULTAD DE INGENIERÍA ECONÓMICA Y
CIENCIAS SOCIALES**



**DISEÑO DE MAPAS DE SALUD Y DE TENDENCIA
DE INVERSIÓN EN EL SECTOR SALUD DEL
PERÚ, MEDIANTE EL ANÁLISIS FACTORIAL**

TESIS

PARA OPTAR EL TÍTULO PROFESIONAL DE :

LICENCIADO EN ESTADÍSTICA

RICARDO JAVIER RODRÍGUEZ CASAS

LIMA - PERU

2000

SUMARIO

En la presente década se ha difundido y desarrollado la información georeferencial, por lo que se debe considerar, paralelamente el desenvolvimiento de la **estadística georeferenciada**, en la actualidad no se conoce trabajos que hayan empleado técnicas científicas estadísticas como puede ser el análisis multivariante para específicamente determinar componentes o factores que indican algún tipo de variable no observable fácilmente, en el desarrollo de este trabajo lo que se pretende es elaborar mapas de análisis factorial con referencia a la situación de salud y a la distribución de recursos e infraestructura sanitaria, en el Perú . En términos médicos esto sería como tomar una radiografía o una ecografía de dos variables: "nivel de salud" y "tendencia de inversión en salud".

El procedimiento adoptado es la técnica estadística del Análisis Factorial aplicado para la elaboración de mapas temáticos referentes a las dos variables anteriormente mencionadas.

Los resultados son importantes, se ha detectado que el mapa de la "tendencia de inversión en salud" contando con todas las instituciones , no corresponde a lo que se necesitaría, según el mapa de "nivel de salud". Similar resultado se observa tomando solo inversiones del Ministerio de Salud, además cabe señalar que el mapa de "tendencia de inversión en salud" involucrando todas las instituciones no contradice al mapa de "tendencia de inversión en salud" del Ministerio de Salud.

Al estudiar estos resultados se puede planificar, modificar y corregir de algún modo la distribución o dirección que deberá tomar nuevas inversiones en salud.

ESQUEMA DE TESIS

TITULO: "DISEÑO DE MAPAS DE SALUD Y DE TENDENCIA DE LA INVERSIÓN EN EL SECTOR SALUD DEL PERÚ, MEDIANTE EL ANÁLISIS FACTORIAL"

I.- INTRODUCCIÓN

1.-Tematización o Marco Situacional

El Sector Salud en la última década se ha desarrollado, mejorando el nivel de salud en muchos sectores de la población, esto se ha hecho a través de las inversiones realizadas casi en su totalidad por el Banco Interamericano de Desarrollo a través del Ministerio de Salud, también existe contribución por parte de la inversión privada, esto se puede verificar observando la variación de algunos indicadores de salud.

Por otro lado, actualmente se está viendo mucho el desarrollo de tecnología para cuantificar la información, una de ellas es la representación georeferenciada que es una de las nuevas formas de representar los datos o bases de datos, en esto se ve de algún modo comprometido lo que corresponde el tratamiento estadístico. La estadística aplicada puede en gran medida reforzar la interpretación de los datos, en este caso lo que se trata es de elaborar mapas de análisis factorial con referencia a la situación de salud y a la distribución de recursos e infraestructura en el Sector Salud del Perú .

2.- Problematización

Uno de los principales problemas que surge al momento de hacer nuevas inversiones es ubicar los lugares de mayor prioridad para realizarlas, por lo que se requiere urgentemente de un método científico que nos ayude a identificar las unidades con mayor prioridad para las nuevas inversiones, esta sería la principal razón de ser del presente trabajo.

En el presente no se cuenta con un diseño que utilice métodos analíticos estadísticos para elaborar un mapas de salud y de recursos e infraestructura sanitaria en el Perú. Hasta el momento no se conoce trabajos que hayan empleado técnicas científicas estadísticas como por ejemplo el análisis multivariante para específicamente determinar componentes o factores que indican algún tipo de variable no observable fácil o usualmente.

Este estudio es de gran importancia, y radica en la optimización de la distribución de recursos, con el previo análisis de cómo esta la situación de salud y como se ha hecho hasta la fecha la distribución de los recursos de salud, sobre todo considerando que conociendo la dirección que ha tomado la inversión de los recursos humanos y de la infraestructura en salud, y contrastando con los mapas de salud, veremos si se está en lo correcto o hay que modificar y corregir de algún modo el rumbo que ha tomado la inversión en salud.

II.- OBJETIVOS

General :

Proporcionar un *modelo* de decisión para la distribución de recursos e infraestructura en el Sector Salud utilizando el Análisis Factorial.

Específicos

1. Obtener mapas mediante el análisis factorial, que nos ayude a determinar el **nivel de salud** y la **tendencia de inversión en salud** utilizando el análisis factorial.
2. Comparar los mapas anteriores y verificar si tienen correspondencia es decir si ha

entidades geográficas donde el nivel de salud es baja se esta dando mayor prioridad para nuevas inversiones.

3. Determinar una metodología para obtener un ranking de prioridad para nuevas inversiones en el Sector Salud a nivel departamental y/o provincial del Perú.

III.- SUPUESTO A VERIFICAR

La distribución de los recursos del Sector Salud no se esta realizando equitativamente, es decir zonas donde la necesidad es alta, se realiza poca inversión y también, por el contrario donde se ve que el "nivel de salud" ha mejorado se sigue invirtiendo en salud, existiendo lugares de mayor prioridad.

IV.- ASPECTOS METODOLÓGICOS

Se ha considerado realizar un análisis factorial para el Mapa de Salud Departamental 1991 y el Mapa de Salud Provincial 1993, y principalmente se aplicará el análisis factorial a la información obtenida en el 2do. Censo de Infraestructura Sanitaria y Recursos del Sector Salud de 1996, el análisis de los datos se realizara a nivel provincial y departamental. Como los datos están a nivel de establecimiento de salud, lo primero que se hará es la construcción de los datos totalizados a nivel de provincia y departamento, luego a este resultado se le aplicará el análisis factorial con el fin de identificar y cuantificar la *tendencia de inversión en salud*, para finalmente construir los mapas correspondientes con los resultados obtenidos.

Aspectos que comprende la presente tesis:

1. Desarrollo de la teoría estadística
2. Levantamiento de información
3. Construcción de variables
4. Desarrollo de Metodología y Análisis
5. Presentación de resultados y elaboración de mapas temáticos
6. Crear una metodología para obtener un ranking de prioridad para nuevas inversiones

V.- MARCO TEÓRICO

1.- Antecedentes

Como antecedentes se puede indicar que se han elaborado algunos trabajos para medir la situación de pobreza del Perú, como el realizado por el Banco Central de Reserva con una primera aproximación de la medición del estado de salud de la población peruana, para 1981 y con datos del Censo Nacional de Población y Vivienda de ese año. Luego el Ministerio de Salud con financiamiento de UNICEF elaboró un Mapa de Salud en 1991 a nivel departamental con una metodología de promedios, por último se tiene información de un mapa de situación de pobreza elaborado por el INEI con los datos del Censo Nacional de Población y Vivienda del año 1993.

2.- Soporte teórico de la investigación

El desarrollo del presente trabajo está basado en el análisis estadístico multivariante, específicamente el análisis factorial.

Las variables que se emplearan son seleccionadas básicamente de los datos que han sido registrados en el 2° Censo de Infraestructura Sanitaria y Recursos del Sector Salud de 1996, además se emplearan datos del Ministerio de Salud publicado en el Mapa de Salud de 1991.

3.- Óptica de la investigación

La preparación de los Mapas de Salud y de Tendencia de Inversión en salud será un importante resultado en el esfuerzo orientado a la modernización tecnológica en el Sector Salud. El trabajo de investigación se realizará con información numérica y cartográfica a nivel departamental y provincial, bajo la filosofía de los Sistemas de Información Geográfica.

Los Sistemas de Información Geográfica (SIG) permiten, bajo un criterio espacial, el manejo integrado de análisis estadístico, bases de datos, procesamiento de imágenes y mapas. El análisis estadístico complementado con el sistema de información geográfico, constituyen una herramienta valiosa y el caso particular de este estudio servirá para mejorar la acción del Estado en cuanto a la focalización de sus acciones, a la asignación y distribución de recursos.

VI.- SELECCIÓN DE VARIABLES

VARIABLES CONSIDERADAS EN EL MAPA DE SALUD DEPARTAMENTAL 1991 (MAPA 01 Y 02):

- 1.- LUZAGUA = Porcentaje de viviendas con luz, agua y desagüe
- 2.- POBURBAN = Porcentaje de población urbana
- 3.- TBM = Tasa bruta de mortalidad (por mil habitantes)
- 4.- TMI = Tasa de Mortalidad Infantil (por mil nacidos vivos)
- 5.- TGF = Tasa global de fecundidad
- 6.- ESPERANZA = Esperanza de vida al nacer (años)
- 7.- ANALFAB = Tasa de analfabetismo
- 8.- MÉDICOS = Médicos por habitantes (10 mil hab.)
- 9.- CAMAS = Camas por habitantes (10 mil hab.)
- 10.- GASTO = Gasto per cápita en salud

VARIABLES CONSIDERADAS EN EL MAPA DE SALUD PROVINCIAL 1993 (MAPA 03 Y 04):

- 1.- POBURBAN = Porcentaje de población urbana
- 2.- TMI = Tasa de mortalidad infantil
- 3.- TANALFFE = Tasa de analfabetismo
- 4.- TMEDICOS = Médicos por 10 mil habitantes
- 5.- TCAMAS = Camas por 10 mil habitantes
- 6.- DESNINO = Porcentaje de desnutrición en niños de primaria.
- 7.- SAGUA = Porcentaje de viviendas sin agua
- 8.- SDESAGUE = Porcentaje de viviendas sin desagüe

VARIABLES CONSIDERADAS EN EL MAPA DE TENDENCIA DE INVERSIÓN EN SALUD 1996, ÍNDICES POR 10 MIL HABITANTES Y/O PORCENTAJES DE ESTABLECIMIENTOS, COMO SE INDICA (MAPA 05 A 20):

- 1.- NRO_ESTA = Número de establecimientos
- 2.- NRO_MEDI = Número de médicos
- 3.- OTR_PROF = Número de otros profesionales de la salud
- 4.- TECN_ASI = Número de técnicos asistenciales
- 5.- AREA_CON = Área construida
- 6.- NUM_CAMA = Número de camas
- 7.- NUM_LABO = Número de laboratorios
- 8.- N_ESTAGU = Núm. (ó %) de establec. con servicio de agua de red pública
- 9.- N_ESTDES = Núm. (ó %) de establec. con desagüe en red pública
- 10.- N_ESTLUZ = Núm. (ó %) de establec. según tipo de energía eléctrica
- 11.- N_TELECO = Núm. (ó %) de establec. con telecomunicaciones
- 12.- N_CERCOP = Núm. (ó %) de establec. con cerco perimétrico
- 13.- N_EQSERF = Número de equipos en servicios finales
- 14.- N_EQSERI = Número de equipos en servicios intermedios
- 15.- N_EQAPOY = Número de equipos en servicios de apoyo

ÍNDICE GENERAL

PORTADA	i
DEDICATORIA	ii
AGRADECIMIENTOS	iii
SUMARIO	iv
ESQUEMA DE TESIS	v
ÍNDICE	viii
INTRODUCCIÓN	1
CAPÍTULO 1. INFORMACION ESTADÍSTICA GEOREFERENCIADA	2
1.1 Concepto de la Información estadística	3
1.2 Sistema de Información Geográfica (SIG)	4
1.3 Integración de la información	5
CAPÍTULO 2. COMPONENTES PRINCIPALES	7
2.1 Introducción	8
2.2 Componentes Principales Poblacionales	9
CAPÍTULO 3. ANÁLISIS FACTORIAL	16
3.1 Introducción	17
3.2 El Modelo de Factor Ortogonal	18
3.3 Métodos de Estimación	23
3.4 Rotación Factorial	32
3.5 Scores Factoriales	33
3.6 Perspectivas y Estrategia para Análisis Factorial	39
CAPÍTULO 4. ANÁLISIS DE RESULTADOS	41
4.1 Introducción	42
4.2 Metodología	43

4.3 Análisis de Resultados.....	44
---------------------------------	----

A. MAPAS DE SALUD

Mapa Departamental de Salud del Perú 1991.

Mapa 01: Análisis factorial con el método de Componentes Principales	44
---	----

Mapa 02: Análisis factorial con el método de Máxima Verosimilitud	47
--	----

Mapas Provincial de Salud 1993

Mapa 03: Análisis Factorial con el método de Componentes Principales	50
---	----

Mapa 04: Análisis Factorial con el método de Máxima Verosimilitud	53
--	----

B. MAPAS DE TENDENCIA DE INVERSIÓN EN SALUD

Mapas de Tendencia de la Inversión en Salud 1996

Mapa Departamental: **Todas las Instituciones**

Análisis Factorial con el método de Componentes Principales

Mapa 05: Uso de solo índices o tasas por habitantes	56
--	----

Mapa 06: Uso de índices por habitantes y porcentajes	59
---	----

Análisis Factorial con el método de Máxima Verosimilitud

Mapa 07: Uso de solo índices o tasas por habitantes	62
--	----

Mapa 08: Uso de índices por habitantes y porcentajes	65
---	----

Mapa Departamental: **Ministerio de Salud**

Análisis Factorial con el método de Componentes Principales

Mapa 09: Uso de solo índices o tasas por habitantes	68
--	----

Mapa 10: Uso de índices por habitantes y porcentajes	71
---	----

Análisis Factorial con el método de Máxima Verosimilitud

Mapa 11: Uso de solo índices o tasas por habitantes	74
--	----

Mapa 12: Uso de índices por habitantes y porcentajes	77
---	----

Mapa Provincial: **Todas las Instituciones**

Análisis Factorial con el método de Componentes Principales		
Mapa 13:	Uso de solo índices o tasas por habitantes	80
Mapa 14:	Uso de índices por habitantes y porcentajes	83
Análisis Factorial con el método de Máxima Verosimilitud		
Mapa 15:	Uso de solo índices o tasas por habitantes	86
Mapa 16:	Uso de índices por habitantes y porcentajes	89
Mapa Provincial: Ministerio de Salud		
Análisis Factorial con el método de Componentes Principales		
Mapa 17:	Uso de solo índices o tasas por habitantes	92
Mapa 18:	Uso de índices por habitantes y porcentajes	95
Análisis Factorial con el método de Máxima Verosimilitud		
Mapa 19:	Uso de solo índices o tasas por habitantes	98
Mapa 20:	Uso de índices por habitantes y porcentajes	101
4.4	Balance de Mapas y su Aplicación	106
CONCLUSIONES Y RECOMENDACIONES		111
ANEXOS		
Anexo 1	Datos para la construcción de Mapas	117
Anexo 2	Matriz de Correlaciones	142
Anexo 3	Ficha de Recojo de Información, Hospitales-Clínicas. 2° CISRESSA	146
Anexo 4	Diagrama de Flujo del Modelamiento	159
Anexo 5	Rutinas SQL en Visual FoxPro	161
Anexo 6	Rutinas SPSS para Análisis Factorial	164
Anexo 7	Distribución de la Población. Perú : 1996	166
Anexo 8	Gráficos de variables ordenados por el primer factor	168
BIBLIOGRAFIA		173

INTRODUCCIÓN

El presente trabajo titulado "Diseño de Mapas de Salud y de Tendencia de Inversión en el Sector Salud del Perú, mediante el Análisis Factorial", tiene como propósito determinar factores que intervienen en el Sector Salud a nivel nacional y diseñar mapas temáticos con dichos factores.

El análisis factorial es una poderosa herramienta que nos permite identificar variables no observables fácilmente, trabajando con variables observables, en la actualidad afortunadamente se ha desarrollado software que nos permite realizar dicho análisis como en esta tesis.

Los objetivos están orientados específicamente a proporcionar un modelo de decisión para la distribución de recursos e infraestructura en el sector salud utilizando el análisis factorial y obtener mapas temáticos a nivel departamental y provincial que nos ayude a determinar una situación de la oferta de salud en el país, esto es, identificar el factor de *tendencia de inversión en salud*, con cuyos valores se puede etiquetar a cada provincia o departamento, para luego elaborar un ranking. Esto se puede comparar con un *mapa de salud* elaborado por el mismo método y verificar si tienen el mismo sentido, es decir a una mala situación de salud le corresponde una menor tendencia en la inversión, y si a una buena situación de salud le corresponde una mayor tendencia en la inversión.

Se ha considerado realizar un análisis factorial de la información obtenida en el 2do. Censo de Infraestructura Sanitaria y Recursos del Sector Salud de 1996, el análisis de los datos se realizara a nivel provincial y departamental. Como los datos están a nivel de establecimiento de salud, lo primero que se hará es la construcción de los datos totalizados a nivel de provincia y departamento, luego a este resultado se le aplicara el análisis factorial con el fin de identificar y cuantificar el *nivel de inversión en salud*, para finalmente construir los mapas correspondientes con los resultados obtenidos.

CAPÍTULO 1

LA INFORMACIÓN

ESTADÍSTICA

GEOREFERENCIADA

1.1 CONCEPTO DE INFORMACIÓN ESTADÍSTICA

La información estadística es un medio, una herramienta que permite cuantificar aspectos de la realidad, en un momento dado o a través de un período, de un fenómeno o problema determinado, a fin de que con dicho conocimiento se pueda inferir conclusiones que permitan adoptar un plan de acción específico.

Es así como, si se desea por parte de las autoridades educativas que la instrucción básica cubra las necesidades del máximo de la población del país, y para ello se cuenta con recursos económicos limitados, se deberá disponer de información acerca del número de personas que requieren dicha instrucción, de su distribución por edad, sexo, región y otras características que permitan crear establecimientos donde sus beneficios alcancen al máximo de personas, logrando con ello el aprovechamiento óptimo de los escasos recursos existentes.

Si se desea distribuir en forma óptima los recursos de salud en el Perú, lo primero que se debe conocer es cuál es el nivel de salud de los peruanos, esto es; con los datos observables que se tiene en salud como son tasa de mortalidad infantil, número de médicos, esperanza de vida, tasa global de fecundidad, etc. se puede construir un indicador del "nivel de salud" , luego con los recursos ya distribuidos se podrían elaborar indicadores de "tendencia de inversión", luego comparando los dos resultados se puede hacer una mejor distribución de recursos a nivel nacional. Justamente lo que se pretende en el presente trabajo es determinar un "mapa de salud del Perú " y un "mapa de la tendencia de la inversión" y compararlas para obtener conclusiones y poder tomar decisiones.

1.2 SISTEMA DE INFORMACIÓN GEOGRÁFICA (SIG)

Definición

El Sistema de información geográfica es un conjunto de reglas y principios enlazados entre si que nos enteran de alguna cosa, los cuales son asistidos por computadora para la entrada, manipulación y despliegue de datos espaciales; lógicamente es un conjunto de operadores que manipulan una base de datos espaciales y constituye un puente de soporte entre el "mundo real" y el usuario.

Se define al sistema de información geográfico como un conjunto de bases informatizadas de datos con algún tipo de componente espacial, la información entre ellas almacenada está referenciada geográficamente tales como mapas, estadísticas o datos sobre un territorio completo. Este sistema las combina en forma diversa, según las necesidades, ampliando las posibilidades de análisis que brindan los mapas convencionales, además de facilitar el almacenamiento organizado.

Objetivos del Sistema

La aplicación de los sistemas de información geográfica persiguen los siguientes objetivos:

- Tener la ubicación espacial del problema en estudio.
- Proporcionar un almacenamiento coherente de la información espacial, pudiendo ser actualizada o manipulada con el mínimo esfuerzo
- Permitir la obtención de modelos cartográficos a partir de la transformación o combinación de diversas variables.
- Facilitar la presentación gráfica de los resultados mediante diversos periféricos de salida (impresora, etc.).

Funciones Del Sistema

Los objetivos descritos determinan la funcionalidad del sistema, debiendo contar con procedimientos de orden modular que a continuación se describen:

- a) Entrada de datos, cartográficas y estadísticas.
- b) Organización, almacenamiento y procesamiento de la base de datos, que permita una estructura relacional.
- c) Análisis, que permita efectuar cualquier operación que transforme las variables originales para un objetivo determinado.
- d) Presentación gráfica, debe permitir la obtención de resultados del análisis en formatos de alta calidad cartográfica.

1.3 INTEGRACIÓN DE LA INFORMACIÓN

Con las definiciones anteriores, dados en el punto 1.1 y 1.2, ya se puede pensar en un sistema de información geo-estadístico, que sería simplemente la fusión de ambos conceptos. El sistema de información geo-estadístico, estaría integrada por el sistema de información geográfico y técnicas estadísticas aplicadas a dicha información lo cual nos permitiría presentar los resultados por medio de mapas temáticos.

En el siguiente gráfico se presenta como se integra el análisis estadístico aplicado, las bases de datos relacionales, y el sistema de información geográfica:

Sistema de Información Geográfico

Bases de Datos Relacionales

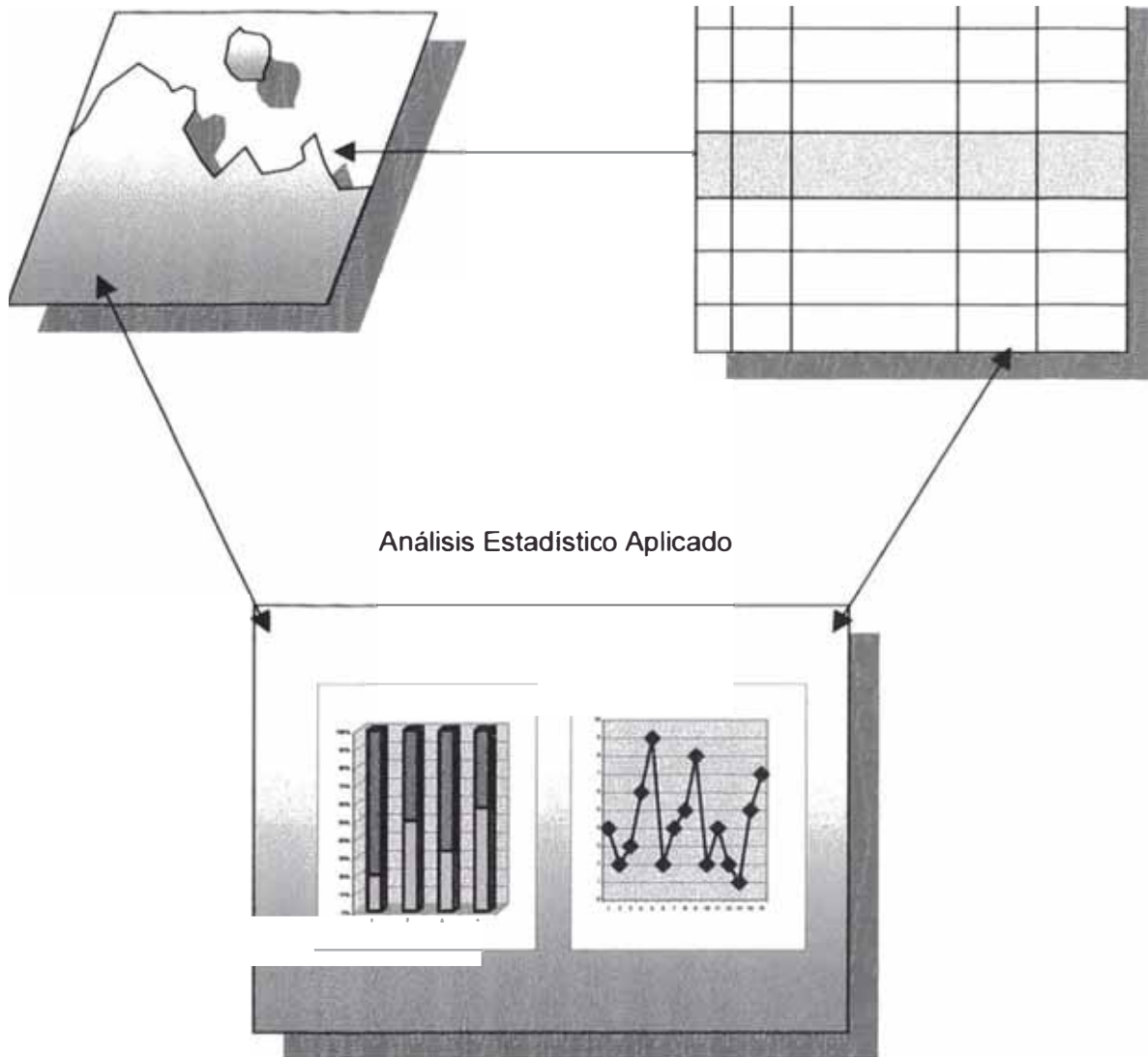


Figura 1.1 Esquema Básico de un Sistema de Información Geo-Estadístico

CAPITULO 2

COMPONENTES PRINCIPALES

2.1 INTRODUCCION

El análisis de Componentes Principales estudia la estructura de la matriz de varianzas y covarianzas a través de unas cuantas combinaciones lineales de las variables originales. Sus objetivos principales son (1) *reducción de datos* (2) *Interpretación*.

Aunque p componentes son necesarios para reproducir el total de variabilidad del sistema, frecuentemente mucho de la variabilidad puede ser considerado por un pequeño número, k , de los componentes principales. Si hay tanta mas información en las k componentes como hay en las originales p variables. Las k componentes principales pueden entonces reemplazar las iniciales p variables, y el conjunto original de datos consistiendo de n medidas en p variables, es reducido a uno el cual consiste de n medidas en k componentes principales ($k < p$).

Un análisis de componentes principales frecuentemente revela relaciones funcionales que no son previamente sospechadas y con eso permite interpretaciones que no pueden resultar ordinariamente.

El análisis de componentes principales son mas bien un medio que un fin en si mismos porque ellos sirven en pasos intermedios en muchas investigaciones mayores. Por ejemplo componentes principales pueden ser ingresado a un análisis de múltiple regresión o análisis de cluster, también pueden ser utilizados en análisis factorial.

2.2 COMPONENTES PRINCIPALES POBLACIONALES

Algebraicamente, componentes principales son combinaciones particulares de las p variables aleatorias X_1, X_2, \dots, X_p . Geométricamente estas combinaciones lineales representan la selección de un nuevo sistema de coordenadas obtenida al rotar el sistema original con X_1, X_2, \dots, X_p como los ejes de coordenadas. Los nuevos ejes representan las direcciones con máxima variabilidad y proporciona un simple y moderada descripción de la estructura de la covarianza.

Tal como veremos, los componentes principales depende solamente de la matriz de covarianzas Σ (o la matriz de correlación ρ) de X_1, X_2, \dots, X_p . Su desarrollo no requiere asumir una distribución normal multivariante. Por otro lado los componentes principales derivados para poblaciones normal multivariante tienen sus interpretaciones en términos de elipsoides constantes de densidad. Adicionalmente, se pueden hacer inferencias de las componentes muestrales donde la población en normal multivariante.

Sea el vector aleatorio $\mathbf{X}' = [X_1, X_2, \dots, X_p]$ tiene la matriz de covarianza Σ con valores propios $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$.

Consideremos las combinaciones lineales

$$\begin{aligned} Y_1 &= \ell_1' \mathbf{X} = \ell_{11}X_1 + \ell_{21}X_2 + \dots + \ell_{p1}X_p \\ Y_2 &= \ell_2' \mathbf{X} = \ell_{12}X_1 + \ell_{22}X_2 + \dots + \ell_{p2}X_p \\ &\vdots \\ Y_p &= \ell_p' \mathbf{X} = \ell_{1p}X_1 + \ell_{2p}X_2 + \dots + \ell_{pp}X_p \end{aligned} \tag{2-1}$$

Entonces usando la propiedad (2-45) de [1]

$$\text{Var}(Y_i) = \ell_i' \Sigma \ell_i \quad i = 1, 2, \dots, p \quad (2-2)$$

$$\text{Cov}(Y_i, Y_k) = \ell_i' \Sigma \ell_k \quad i, k = 1, 2, \dots, p ; i \neq k \quad (2-3)$$

Las componentes principales son esas combinaciones lineales incorrelacionadas Y_1, Y_2, \dots, Y_p cuyas varianzas en (2-2) son tan grandes como sea posible.

La primera componente principal es la combinación lineal con máxima varianza. Esto es maximizar $\text{Var}(Y_1) = \ell_1' \Sigma \ell_1$. Es claro que $\text{Var}(Y_1) = \ell_1' \Sigma \ell_1$ puede ser incrementado multiplicando cualquier ℓ_1 por alguna constante. Para eliminar esta indeterminación, es conveniente restringir la atención a coeficientes de vectores con longitud uno. Nosotros por tanto definimos:

Primera componente principal = combinación lineal $\ell_1' \mathbf{X}$ que maximiza

$$\text{Var}(\ell_1' \mathbf{X}) \text{ sujeto a } \ell_1' \ell_1 = 1$$

Segunda componente principal = combinación lineal $\ell_2' \mathbf{X}$ que maximiza

$$\text{Var}(\ell_2' \mathbf{X}) \text{ sujeto a } \ell_2' \ell_2 = 1 \text{ y}$$

$$\text{Cov}(\ell_1' \mathbf{X}, \ell_2' \mathbf{X}) = 0$$

En el i-ésimo paso

i-ésima componente principal = combinación lineal $\ell_i' \mathbf{X}$ que maximiza

$$\text{Var}(\ell_i' \mathbf{X}) \text{ sujeto a } \ell_i' \ell_i = 1 \text{ y}$$

$$\text{Cov}(\ell_i' \mathbf{X}, \ell_k' \mathbf{X}) = 0 \quad \text{para } k < i$$

Teorema 2.1

Sea Σ la matriz de covarianzas asociada con el vector aleatorio $\mathbf{X}' = [X_1, X_2, \dots, X_p]$. Sea Σ que tiene los pares de valores propios y los vectores propios $(\lambda_1, \mathbf{e}_1), (\lambda_2, \mathbf{e}_2), \dots, (\lambda_p, \mathbf{e}_p)$ donde $\lambda_1 > \lambda_2 > \dots > \lambda_p \geq 0$. Entonces la i-ésima componente principal está dada por :

$$Y_i = \mathbf{e}_i' \mathbf{X} = e_{i1}X_1 + e_{i2}X_2 + \dots + e_{ip}X_p \quad i = 1, 2, \dots, p \quad (2-4)$$

Con estas elecciones

$$\text{Var}(Y_i) = \mathbf{e}_i' \Sigma \mathbf{e}_i = \lambda_i \quad i=1, 2, \dots, p \quad (2-5)$$

$$\text{Cov}(Y_i, Y_k) = \mathbf{e}_i' \Sigma \mathbf{e}_k = 0 \quad i \neq k$$

Si algunos λ_i son iguales, la elección de los correspondientes coeficientes de los vectores \mathbf{e}_i , y por tanto Y_i no son únicos.

Prueba

Nosotros conocemos por (2-51) de [1] con $\mathbf{B} = \Sigma$,

$$\max_{\ell \neq 0} \frac{\ell' \Sigma \ell}{\ell' \ell} = \lambda_1 \quad \text{consiguiendo cuando } \ell = \mathbf{e}_1$$

Pero $\mathbf{e}_1' \mathbf{e}_1 = 1$ ya que los vectores propios están normalizados. Entonces

$$\max_{\ell \neq 0} \frac{\ell' \Sigma \ell}{\ell' \ell} = \lambda_1 = \frac{\mathbf{e}_1' \Sigma \mathbf{e}_1}{\mathbf{e}_1' \mathbf{e}_1} = \mathbf{e}_1' \Sigma \mathbf{e}_1 = \text{Var}(Y_1)$$

Similarmenete usando (2-52) de [1]

$$\max_{\ell \perp \mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_k} \frac{\ell' \Sigma \ell}{\ell' \ell} = \lambda_{k+1} \quad k = 1, 2, \dots, p-1$$

Para la elección $\ell = \mathbf{e}_{k+1}$, donde $\mathbf{e}_{k+1}' \mathbf{e}_k = 0$, $k = 1, 2, \dots, p-1$,

$$\mathbf{e}_{k+1}' \Sigma \mathbf{e}_{k+1} / \mathbf{e}_{k+1}' \mathbf{e}_{k+1} = \mathbf{e}_{k+1}' \Sigma \mathbf{e}_{k+1} = \text{Var}(Y_{k+1})$$

pero $\mathbf{e}_{k+1}' (\Sigma \mathbf{e}_{k+1}) = \lambda_{k+1} \mathbf{e}_{k+1}' \mathbf{e}_{k+1} = \lambda_{k+1}$ así como $\text{Var}(Y_{k+1}) = \lambda_{k+1}$.

Falta mostrar que \mathbf{e}_i es perpendicular a \mathbf{e}_k , (esto es, $\mathbf{e}_i' \mathbf{e}_k = 0$, $i \neq k$) dado $\text{Cov}(Y_i, Y_k) = 0$.

Ahora los vectores propios de Σ son ortogonales si todos los valores propios $\lambda_1, \lambda_2, \dots, \lambda_p$ son distintos. Si los valores propios no son todos distintos, los vectores propios que

corresponden a valores propios comunes pueden ser elegidos para ser ortogonales. Por tanto para cualquiera dos vectores \mathbf{e}_i y \mathbf{e}_k , $\mathbf{e}_i'\mathbf{e}_k = 0$, $i \neq k$. Ya que $\Sigma\mathbf{e}_k = \lambda_k\mathbf{e}_k$, multiplicando por \mathbf{e}_i' nos da

$$\text{Cov}(Y_i, Y_k) = \mathbf{e}_i'\Sigma\mathbf{e}_k = \mathbf{e}_i'\lambda_k\mathbf{e}_k = \lambda_k \mathbf{e}_i'\mathbf{e}_k = 0$$

Para cualquier $i \neq k$ y la prueba esta completa.

Lqgd.

Del teorema anterior, los componentes principales son incorrelacionados y tienen varianza igual a los valores propios de Σ .

Teorema 2.2

Sea $\mathbf{X}' = [X_1, X_2, \dots, X_p]$ con matriz de covarianzas Σ , con los pares de valores propios y vectores propios $(\lambda_1, \mathbf{e}_1), (\lambda_2, \mathbf{e}_2), \dots, (\lambda_p, \mathbf{e}_p)$ donde $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$. Sea $Y_1 = \mathbf{e}_1'\mathbf{X}$, $Y_2 = \mathbf{e}_2'\mathbf{X}$, ..., $Y_p = \mathbf{e}_p'\mathbf{X}$ las componentes principales. Entonces

$$\sigma_{11} + \sigma_{22} + \dots + \sigma_{pp} = \sum_{i=1}^p \text{Var}(X_i) = \lambda_1 + \lambda_2 + \dots + \lambda_p = \sum_{i=1}^p \text{Var}(Y_i)$$

Prueba

Por definición 2A.28 de [1] $\sigma_{11} + \sigma_{22} + \dots + \sigma_{pp} = \text{tr}(\Sigma)$. De (2-20) de [1] con $\mathbf{A} = \Sigma$, nosotros podemos escribir $\Sigma = \mathbf{P}\Lambda\mathbf{P}'$ donde Λ es la matriz diagonal de valores propios y $\mathbf{P} = [\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_p]$ también $\mathbf{P}\mathbf{P}' = \mathbf{P}'\mathbf{P} = \mathbf{I}$. Usando 2A.12(c) de [1], nosotros tenemos

$$\text{tr}(\Sigma) = \text{tr}(\mathbf{P}\Lambda\mathbf{P}') = \text{tr}(\Lambda\mathbf{P}'\mathbf{P}) = \text{tr}(\Lambda) = \lambda_1 + \lambda_2 + \dots + \lambda_p$$

Entonces

$$\sum_{i=1}^p \text{Var}(X_i) = \text{tr}(\Sigma) = \text{tr}(\Lambda) = \sum_{i=1}^p \text{Var}(Y_i)$$

Lqgd.

En otras palabras

$$\text{Varianza total de la población} = \sigma_{11} + \sigma_{22} + \dots + \sigma_{pp} = \lambda_1 + \lambda_2 + \dots + \lambda_p \quad (2-6)$$

Y consecuentemente la proporción del total de varianza debido a (explicada por) la k-ésima

componente principal es

$$\left(\begin{array}{l} \text{Proporción del total} \\ \text{de varianza poblacional} \\ \text{debido a la } k\text{-ésima} \\ \text{componente principal} \end{array} \right) = \frac{\lambda_k}{\lambda_1 + \lambda_2 + \dots + \lambda_p} \quad ; \quad k = 1, 2, \dots, p \quad (2-7)$$

Si bastante (por ejemplo 80 a 90%) de la varianza total de la población, para grandes valores de p , puede ser atribuido a las primeras una, dos o tres componentes, entonces estas componentes pueden *reemplazar* las originales p variables sin mucha pérdida de información.

Cada componente de los coeficientes del vector $\mathbf{e}'_i = [e_{1i}, \dots, e_{ki}, \dots, e_{pi}]$ también merece una revisión. La magnitud de e_{ki} es proporcional al coeficiente de correlación entre Y_i y X_k .

Teorema 2.3

Si $Y_1 = \mathbf{e}'_1 \mathbf{X}$, $Y_2 = \mathbf{e}'_2 \mathbf{X}$, ..., $Y_p = \mathbf{e}'_p \mathbf{X}$ son las componentes principales obtenidas desde la matriz de covarianzas Σ , entonces

$$\rho_{Y_i, X_k} = \frac{e_{ki} \sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}} \quad i, k = 1, 2, \dots, p \quad (2-8)$$

Son los coeficientes de correlación entre Y_i y las variables X_k . Aquí $(\lambda_1, \mathbf{e}_1), (\lambda_2, \mathbf{e}_2), \dots, (\lambda_p, \mathbf{e}_p)$ son los pares de valores propios y vectores propios de Σ .

Prueba

Sea $\mathbf{e}'_k = [0, \dots, 0, 1, 0, \dots, 0]$ también que $X_k = \mathbf{e}'_k \mathbf{X}$ y $\text{Cov}(Y_k, Y_i) = \text{Cov}(\mathbf{e}'_k \mathbf{X}, \mathbf{e}'_i \mathbf{X}) = \mathbf{e}'_k \Sigma \mathbf{e}_i$ de acuerdo a (2-45) de [1]. Ya que $\Sigma \mathbf{e}_i = \lambda_i \mathbf{e}_i$, $\text{Cov}(X_k, Y_i) = \mathbf{e}'_k \lambda_i \mathbf{e}_i = \lambda_i e_{ki}$. Entonces $\text{Var}(Y_i) = \lambda_i$ [ver (2-5)] y $\text{Var}(X_k) = \sigma_{kk}$ produciendo

$$\rho_{Y_i, X_k} = \frac{\text{Cov}(Y_i, X_k)}{\sqrt{\text{Var}(Y_i)}\sqrt{\text{Var}(X_k)}} = \frac{\lambda_i e_{ki}}{\sqrt{\lambda_i} \sqrt{\sigma_{kk}}} = \frac{e_{ki} \sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}} \quad i, k = 1, 2, \dots, p$$

Lqgd.

Componentes Principales obtenidos de Variables Estandarizadas

Las componentes principales pueden también ser obtenidas de variables estandarizadas

$$\begin{aligned} Z_1 &= \frac{(X_1 - \mu_1)}{\sqrt{\sigma_{11}}} \\ Z_2 &= \frac{(X_2 - \mu_2)}{\sqrt{\sigma_{22}}} \\ &\vdots \\ Z_p &= \frac{(X_p - \mu_p)}{\sqrt{\sigma_{pp}}} \end{aligned} \quad (2-9)$$

En notación matricial,

$$\mathbf{Z} = (\mathbf{V}^{1/2})^{-1}(\mathbf{X} - \boldsymbol{\mu}) \quad (2-10)$$

Donde la matriz de desviación estándar diagonal $\mathbf{V}^{1/2}$ es definido en (2-35) de [1].

Claramente

$$E(\mathbf{Z}) = \mathbf{0} \text{ y } \text{Cov}(\mathbf{Z}) = (\mathbf{V}^{1/2})^{-1} \boldsymbol{\Sigma} (\mathbf{V}^{1/2})^{-1} = \boldsymbol{\rho} \text{ , por (2-37) de [1]}$$

La componente principal de \mathbf{Z} puede ser obtenido desde los vectores propios de la matriz de correlación $\boldsymbol{\rho}$ de \mathbf{X} . Todo de nuestro previo resultado son aplicables, con algunas simplificaciones ya que la varianza de cada Z_i es la unidad. Continuando con el uso de la notación Y_i para referirnos a la i -ésima componente principal y $(\lambda_i, \mathbf{e}_i)$ para el par de valores propios y vectores propios. Sin embargo, estas cantidades derivadas desde $\boldsymbol{\Sigma}$ no son, en general, los mismo que los derivados de $\boldsymbol{\rho}$.

Teorema 2.4

La i -ésima componentes principal de las variables estandarizadas $\mathbf{Z}' = [Z_1, Z_2, \dots, Z_p]$, con

$\text{Cov}(\mathbf{Z}) = \rho$, es dado por

$$Y_i = \mathbf{e}_i' \mathbf{Z} = \mathbf{e}_i' (\mathbf{V}^{1/2})^{-1} (\mathbf{X} - \boldsymbol{\mu}), \quad i = 1, 2, \dots, p$$

Además

$$\sum_{i=1}^p \text{Var}(Y_i) = \sum_{i=1}^p \text{Var}(Z_i) = p \quad (2-11)$$

y

$$\rho_{Y_i, Z_k} = \mathbf{e}_{ki} \sqrt{\lambda_i}, \quad i, k = 1, 2, \dots, p$$

En este caso $(\lambda_1, \mathbf{e}_1), (\lambda_2, \mathbf{e}_2), \dots, (\lambda_p, \mathbf{e}_p)$ son los pares de valores propios y vectores propios para ρ con $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$.

Prueba

El teorema 2.4 se sigue del teorema 2.1, teorema 2.2, y teorema 2.3, con Z_1, Z_2, \dots, Z_p en lugar de X_1, X_2, \dots, X_p y ρ en lugar de Σ . Lqgd.

Nosotros vemos de (2-11) que el total (de variables estandarizadas) de la varianza poblacional es simplemente p , la suma de los elementos diagonales de la matriz ρ . Usando \mathbf{Z} en lugar de \mathbf{X} en (2-7), la proporción de varianza total explicada por la k -ésima componente principal de \mathbf{Z} es

$$\left(\begin{array}{l} \text{Proporción de (estandarizadas)} \\ \text{la varianza poblacional debido a} \\ \text{la } k \text{ - ésima componente principal} \end{array} \right) = \frac{\lambda_k}{p} \quad k = 1, 2, \dots, p \quad (2-12)$$

Donde los λ_k 's son los valores propios de ρ .

CAPITULO 3

ANALISIS FACTORIAL

3.1 INTRODUCCION

El análisis factorial ha provocado más bien controversia turbulenta en toda su historia. Fue desarrollado por Karl Pearson, Charles Spearman y otros para definir y medir *inteligencia*. Debido a esta primera asociación construyendo medidas de inteligencia, el análisis factorial fue iniciado y desarrollado primeramente por científicos interesados en medidas psicometricas. Argumentos de las interpretaciones psicológicas de varios primeros estudios y la falta de potentes computadoras y sus facilidades impidieron su inicial desarrollo como un método estadístico. La llegada de computadoras altamente veloces fue generando un renovado interés en la teoría y aspectos computacionales del análisis factorial. Mucho de las técnicas originales han sido abandonadas y primeras controversias resueltas en el camino del desarrollo reciente. Sin embargo es verdad que cada aplicación de las técnicas debe ser examinado sobre sus méritos para determinar su éxito.

El esencial propósito de análisis factorial es para describir, si es posible, la relación de las covarianzas en medio de muchas variables, en términos de unas cuantas bases, pero no observables cantidades aleatorias llamadas *factores*. Básicamente el modelo factorial es motivado por el siguiente argumento. Suponga que las variables pueden ser agrupadas por sus correlaciones. Esto es todas las variables dentro de un grupo particular son altamente correlacionadas entre si mismás pero tienen relativa pequeñas correlaciones con variables en un diferente grupo. Es concebible que cada grupo de variables representa una única base de construcción, o factor, que es responsable de las correlaciones observadas. Por

ejemplo, correlaciones de un grupo de pruebas francés, inglés, matemáticas y música coleccionada por Spearman sugiere una base de factor *inteligencia*. Un segundo grupo de variables, representando notas de ajuste físico, si esta disponible, significa que corresponde a otro factor. Este tipo de estructuras es lo que el análisis factorial busca confirmar.

El análisis factorial puede ser considerado como una extensión de análisis de componentes principales. Ambos pueden ser vistos como intentos para aproximar la matriz de covarianzas Σ . Sin embargo la aproximación basada en el modelo del análisis factorial es más elaborada. La primera cuestión en análisis factorial es si los datos son consistentes con una estructura pre-escrita.

3.2 EL MODELO DE FACTOR ORTOGONAL

El observable vector aleatorio \mathbf{X} con p componentes, que tiene media μ y matriz de varianzas y covarianzas Σ . El modelo factorial postula que \mathbf{X} es linealmente dependiente de unas cuantas variables aleatorias no observables F_1, F_2, \dots, F_m , llamados *factor común*, y p adicionales fuentes de variación $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$, llamados errores o algunas veces *factores específicos*. En particular el modelo de análisis factorial es

$$\begin{aligned} X_1 - \mu_1 &= l_{11}F_1 + l_{12}F_2 + \dots + l_{1m}F_m + \varepsilon_1 \\ X_2 - \mu_2 &= l_{21}F_1 + l_{22}F_2 + \dots + l_{2m}F_m + \varepsilon_2 \\ &\vdots \\ X_p - \mu_p &= l_{p1}F_1 + l_{p2}F_2 + \dots + l_{pm}F_m + \varepsilon_p \end{aligned} \tag{3-1}$$

O en notación matricial,

$$\begin{matrix} \mathbf{X} - \boldsymbol{\mu} &= & \mathbf{L} & \mathbf{F} &+ & \boldsymbol{\varepsilon} \\ (p \times 1) & & (p \times m) & (m \times 1) & & (p \times 1) \end{matrix} \tag{3-2}$$

El coeficiente l_{ij} es llamada la ponderación de la i -ésima variable en el j -ésimo factor,

también la matriz L es la matriz de *ponderación factorial*. Notar que el i -ésimo factor específico ε_i es asociada solamente con la i -ésima respuesta X_i . Las p desviaciones $X_1 - \mu_1, X_2 - \mu_2, \dots, X_p - \mu_p$ son expresadas en términos de $p + m$ variables aleatorias $F_1, F_2, \dots, F_m, \varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$ los cuales son no observables.

Con muchas cantidades no observables, una directa verificación de el modelo factorial de observaciones sobre X_1, X_2, \dots, X_p es preocupante. Sin embargo con alguna adicional asumción acerca de los vectores aleatorios F y ε , el modelo en (3-2) implica cierta relación de covarianzas, los cuales pueden ser chequeados.

Nosotros asumimos que

$$\begin{aligned}
 E(\mathbf{F}) &= \mathbf{0} & \text{Cov}(\mathbf{F}) &= E(\mathbf{F}\mathbf{F}') = \mathbf{I} \\
 & \text{(m \times 1)} & & \text{(m \times m)} \\
 \\
 E(\boldsymbol{\varepsilon}) &= \mathbf{0}_{p \times 1}, & \text{Cov}(\boldsymbol{\varepsilon}) &= E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \boldsymbol{\Psi} = \begin{pmatrix} \psi_1 & 0 & \dots & 0 \\ 0 & \psi_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \psi_p \end{pmatrix} \\
 & & & \text{(p \times p)} & & \text{(3 - 3)}
 \end{aligned}$$

y que F y ε son independientes por tanto

$$\text{Cov}(\boldsymbol{\varepsilon}, \mathbf{F}) = E(\boldsymbol{\varepsilon}\mathbf{F}') = \mathbf{0} \quad \text{(p \times m)}$$

Estas asumciones y la relación (3-2) constituye el **modelo de factor ortogonal**

Definición (3 - 4)

El modelo de factor ortogonal con m Factores Comunes

$$\mathbf{X} = \boldsymbol{\mu} + \mathbf{L} \mathbf{F} + \boldsymbol{\varepsilon}$$

$(p \times 1) \quad (p \times 1) \quad (p \times m) \quad (m \times 1) \quad (p \times 1)$

μ_i = media de la variables i

ε_i = i -ésima factor específico

F_j = j -ésimo factor común

ℓ_{ij} = ponderacion de la i -ésima variables en el j -ésimo factor

Los vectores aleatorios no observables \mathbf{F} y ε satisfacen:

\mathbf{F} y ε son independientes

$$E(\mathbf{F}) = \mathbf{0}, \text{Cov}(\mathbf{F}) = \mathbf{I}$$

$E(\varepsilon) = \mathbf{0}$, $\text{Cov}(\varepsilon) = \Psi$, donde Ψ es una matriz diagonal

El modelo de factor ortogonal implica una estructura en la matriz de covarianzas para \mathbf{X} de el modelo en (3-4)

$$\begin{aligned} (\mathbf{X} - \mu)(\mathbf{X} - \mu)' &= (\mathbf{LF} + \varepsilon)(\mathbf{LF} + \varepsilon)' \\ &= (\mathbf{LF} + \varepsilon)((\mathbf{LF})' + \varepsilon') \\ &= \mathbf{LF}(\mathbf{LF})' + \varepsilon(\mathbf{LF})' + \mathbf{LF}\varepsilon' + \varepsilon\varepsilon' \end{aligned}$$

por tanto

$$\begin{aligned} \Sigma &= \text{Cov}(\mathbf{X}) = E(\mathbf{X} - \mu)(\mathbf{X} - \mu)' \\ &= \mathbf{LE}(\mathbf{FF}')\mathbf{L}' + E(\varepsilon\mathbf{F}')\mathbf{L}' + \mathbf{LE}(\mathbf{F}\varepsilon') + E(\varepsilon\varepsilon') \\ &= \mathbf{LL}' + \Psi \end{aligned}$$

de acuerdo a (3-3).

También por el modelo en (3-4) $(\mathbf{X} - \mu)\mathbf{F}' = (\mathbf{LF} + \varepsilon)\mathbf{F}' = \mathbf{LFF}' + \varepsilon\mathbf{F}'$, así como

$$\text{Cov}(\mathbf{X}, \mathbf{F}) = E(\mathbf{X} - \mu)\mathbf{F}' = \mathbf{LE}(\mathbf{FF}') + E(\varepsilon\mathbf{F}') = \mathbf{L}$$

Estructuras de las covarianzas del Modelo de Factor Ortogonal:

1. $\text{Cov}(\mathbf{X}) = \mathbf{LL}' + \Psi$

$$\text{ó } \text{Var}(X_i) = \ell_{i1}^2 + \dots + \ell_{im}^2 + \psi_i$$

$$\text{Cov}(X_i, X_k) = \ell_{i1}\ell_{k1} + \dots + \ell_{im}\ell_{km} \tag{3-5}$$

2. $\text{Cov}(\mathbf{X}, \mathbf{F}) = \mathbf{L}$

$$\text{ó } \text{Cov}(X_i, F_j) = \ell_{ij}$$

El modelo $\mathbf{X} - \mu = \mathbf{LF} + \varepsilon$ es *lineal* en los factores comunes. Si las p respuestas \mathbf{X} son de hecho, relacionado a factores pero la relación es no lineal tal como en $X_1 - \mu_1 = \ell_{11}F_1F_3 +$

$\varepsilon_1, X_2 - \mu_2 = \ell_{21}F_2F_3 + \varepsilon_2$ y así sucesivamente entonces la estructura de la covarianza $LL' + \Psi$ dado por (3-5) puede no ser adecuado. La asunción muy importante de linealidad es inherente a la formulación del modelo tradicional de factores.

Esta porción de la varianza de la i -ésima variable contribuida por los m factores comunes es llamada la i -ésima *comunalidad*. Esta porción de la varianza $\text{Var}(X_i) = \sigma_{ii}$ debido a los factores específicos es a veces llamado la *unicidad* o varianza específica. Denotando la i -ésima comunalidad por h_i^2 , nosotros vemos de (3-5) que

$$\underbrace{\sigma_{ii}}_{\text{Var}(X_i)} = \underbrace{\ell_{i1}^2 + \ell_{i2}^2 + \dots + \ell_{im}^2}_{\text{comunalidad}} + \underbrace{\psi_i}_{\text{varianza específica}}$$

o

$$h_i^2 = \ell_{i1}^2 + \ell_{i2}^2 + \dots + \ell_{im}^2 \tag{3-6}$$

y

$$\sigma_{ii} = h_i^2 + \psi_i \quad i = 1, 2, \dots, p$$

La i -ésima comunalidad es la suma de cuadrados de las ponderaciones de la i -ésima variable en los m factores comunes.

El modelo factorial asume que las $p+p(p-1)/2-p(p+1)/2$ varianzas y covarianzas de \mathbf{X} pueden ser reproducidas desde las pm ponderaciones de factor ℓ_{ij} y las p varianzas específicas ψ_i . Cuando $m = p$, cualquier matriz de covarianzas puede ser reproducido exactamente como LL' [ver (3-11)], por tanto Ψ puede ser la matriz cero. Sin embargo cuando m es relativamente menor que p el análisis factorial es de lo más útil. En este caso el modelo factorial provee una "simple" explicación de la covariación de \mathbf{X} con pocos parámetros que los $p(p+1)/2$ parámetros en Σ .

Desafortunadamente para el análisis factorial, muchas matrices de covarianzas no pueden ser factorizadas como $LL'+\Psi$, en que el número de factores m es mucho menor que p .

Cuando $m > 1$, hay siempre alguna ambigüedad inherente asociada con el modelo factorial. Para ver esto sea T cualquier $m \times m$ matriz ortogonal, tal que $TT' = T'T = I$. La expresión en (3-2) puede ser escrita:

$$X - \mu = LF + \varepsilon = LTT'F + \varepsilon = L^*F^* + \varepsilon \quad (3-7)$$

Donde

$$L^* = LT \quad \text{y} \quad F^* = T'F$$

Ya que

$$E(F^*) = T'E(F) = 0$$

y

$$\text{Cov}(F^*) = T'\text{Cov}(F)T = TT' = I$$

(mxm)

es imposible, sobre la base de observación de X , distinguir las ponderaciones L de las ponderaciones L^* . Esto es, los factores F y $F^* = T'F$ tienen las mismas propiedades estadísticas, y siempre aunque las ponderaciones L^* son, en general, diferente de las ponderaciones L , ambos generan la misma matriz de covarianzas Σ . Esto es,

$$\Sigma = LL' + \Psi = LTT'L' + \Psi = (L^*)(L^*)' + \Psi \quad (3-8)$$

Esta ambigüedad provee el razonamiento de "rotación factorial", ya que matrices ortogonales corresponden a rotaciones (y reflexión) de el sistema de coordenadas para X .

<< Las ponderaciones factoriales L son determinadas solamente para una matriz ortogonal T . De este modo, las ponderaciones

$$L^* = LT \quad \text{y} \quad L$$

dan la misma representación. Las comunales, dadas por los elementos diagonales de $LL' = (L^*)(L^*)'$, son también inafectos a la elección de T . >>

(3-9)

El análisis de el modelo factorial avanza por impresionantes condiciones que permiten uno y solo un estimador de L y Ψ . La matriz de ponderaciones es entonces rotada (multiplicada por una matriz ortogonal), donde la rotación es determinada por algún criterio

de "comodidad de interpretación". Ya que las ponderaciones y las varianzas específicas son obtenidas, factores son identificados y estimado valores para los factores mismos (llamados marcas de factor) son frecuentemente construidos.

3.3 MÉTODOS DE ESTIMACIÓN

Dadas las observaciones $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ en p generalmente correlacionadas variables, el análisis factorial busca respuesta a la siguiente pregunta, ¿Dado el modelo factorial en (3-4), con un pequeño número de factores, se puede representar adecuadamente los datos?. En esencia, nosotros forcejamos el problema del modelo estadístico por difícil para verificar la relación de covarianza en (3-5).

La matriz de covarianza muestral \mathbf{S} es un estimador de la matriz de covarianza poblacional Σ .

Si los elementos fuera de la diagonal de \mathbf{S} son pequeños o aquellos de la matriz de correlación muestral \mathbf{R} esencialmente cero, las variables no están relacionadas y un análisis factorial no puede ser útil. En estas circunstancias, los factores específicos juegan un rol dominante, por cuanto el análisis factorial apunta a determinar unos cuantos importantes factores comunes.

Si Σ aparece desviarse significativamente de una matriz diagonal, entonces un modelo factorial puede ser ingresado y el inicial problema es uno de estimación de las ponderaciones factoriales ℓ_{ij} y varianzas específicas ψ_i . Nosotros consideraremos dos de las más populares métodos de estimación de parámetros el método de *componentes principales* y el método de *máxima verosimilitud*. La solución de estos métodos pueden ser rotados para simplificar la interpretación de factores como describimos en la sección (3-4). Es prudente probar más de un método de solución. Si el modelo factorial es apropiado para el problema por un lado, la solución puede ser consistente con algún otro.

Los concurrentes métodos de estimación y rotación requieren cálculos iterativos que deben ser hechos en una computadora. Varios programas de computadora están ahora disponibles para este propósito.

El método de Componentes Principales y/o Factor Principal (CP)

La descomposición espectral de (2-20) de [1] nos proporciona una factorización de la matriz de covarianzas Σ . Sea Σ que tiene los valores y vectores propios $(\lambda_i, \mathbf{e}_i)$ con $\lambda_1 > \lambda_2 > \dots > \lambda_p > 0$. Entonces

$$\Sigma = \lambda_1 \mathbf{e}_1 \mathbf{e}_1' + \lambda_2 \mathbf{e}_2 \mathbf{e}_2' + \dots + \lambda_p \mathbf{e}_p \mathbf{e}_p' \quad (3-10)$$

$$= \begin{bmatrix} \sqrt{\lambda_1} \mathbf{e}_1 & \sqrt{\lambda_2} \mathbf{e}_2 & \dots & \sqrt{\lambda_p} \mathbf{e}_p \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} \mathbf{e}_1' \\ \sqrt{\lambda_2} \mathbf{e}_2' \\ \vdots \\ \sqrt{\lambda_p} \mathbf{e}_p' \end{bmatrix}$$

Este ajuste prescribe la estructura de covarianzas para el modelo del análisis factorial, teniendo tantos factores como variables ($m = p$) y varianzas específica $\psi_i = 0$ para todo i . Las matriz de ponderaciones tiene la j -ésima matriz de ponderaciones dado por $\sqrt{\lambda_j} \mathbf{e}_j$. Esto es nosotros podemos escribir:

$$\Sigma = \underset{(p \times p)}{\mathbf{L}} \underset{(p \times p)}{\mathbf{L}'} + \underset{(p \times p)}{\mathbf{0}} = \underset{(p \times p)}{\mathbf{L}} \underset{(p \times p)}{\mathbf{L}'} \quad (3-11)$$

Aparte de el factor escalar $\sqrt{\lambda_j}$, las ponderaciones factoriales en la j -ésima factor son los coeficientes de la j -ésima componente principal poblacional.

Aunque la representación del análisis factorial de Σ en (3-11) es exacta, esto no es particularmente útil. Esto da empleo tanto a muchos factores comunes como variables hay y no permite alguna variación de los factores específicos ϵ en (3-4). Nosotros preferimos modelos que explican la estructura de la covarianza en términos de justamente unos

cuantos factores comunes. Una aproximación cuando las últimas $p - m$ valores propios son pequeños, es para despreocuparse de la contribución de $\lambda_{m+1} \mathbf{e}_{m+1} \mathbf{e}_{m+1}' + \dots + \lambda_p \mathbf{e}_p \mathbf{e}_p'$ para Σ en (3-10). Despreocupandose de esta contribución nosotros obtenemos la aproximación

$$\Sigma \cong \begin{bmatrix} \sqrt{\lambda_1} \mathbf{e}_1 & \sqrt{\lambda_2} \mathbf{e}_2 & \dots & \sqrt{\lambda_m} \mathbf{e}_m \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} \mathbf{e}_1' \\ \sqrt{\lambda_2} \mathbf{e}_2' \\ \vdots \\ \sqrt{\lambda_m} \mathbf{e}_m' \end{bmatrix} = \underset{(p \times m)}{\mathbf{L}} \underset{(m \times p)}{\mathbf{L}'} \quad (3-12)$$

La representación aproximada en (3-12) asume que los factores específicos ε en (3-4) son de menor importancia y pueden también ser ignorados en la factorización de Σ . Si factores específicos son incluidos en el modelo, sus varianzas pueden ser tomadas de los elementos diagonales de $\Sigma - \mathbf{L}\mathbf{L}'$, donde $\mathbf{L}\mathbf{L}'$ es definido en (3-12).

Siguiendo para factores específicos, la aproximación resulta

$$\Sigma \cong \mathbf{L}\mathbf{L}' + \Psi$$

$$\Sigma \cong \begin{bmatrix} \sqrt{\lambda_1} \mathbf{e}_1 & \sqrt{\lambda_2} \mathbf{e}_2 & \dots & \sqrt{\lambda_m} \mathbf{e}_m \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} \mathbf{e}_1' \\ \sqrt{\lambda_2} \mathbf{e}_2' \\ \vdots \\ \sqrt{\lambda_m} \mathbf{e}_m' \end{bmatrix} + \begin{bmatrix} \psi_1 & 0 & \dots & 0 \\ 0 & \psi_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \psi_p \end{bmatrix} \quad (3-13)$$

donde $\psi_i = \sigma_{ii} - \sum_{j=1}^m \ell_{ij}^2$ para $i = 1, 2, \dots, p$.

Para aplicar esta aproximación a un conjunto de datos $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$, se acostumbra primero centrar las observaciones, sustrayendo la media muestral $\bar{\mathbf{x}}$. Las observaciones centradas tienen la misma matriz de covarianzas, \mathbf{S} , que las observaciones originales.

$$\mathbf{x}_j - \bar{\mathbf{x}} = \begin{bmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{pj} \end{bmatrix} - \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix} = \begin{bmatrix} x_{1j} - \bar{x}_1 \\ x_{2j} - \bar{x}_2 \\ \vdots \\ x_{pj} - \bar{x}_p \end{bmatrix}, \quad j = 1, 2, \dots, n \quad (3-14)$$

En casos donde las unidades de las variables no son conmensurables, es usualmente deseable trabajar con variables estandarizadas

$$z_j = \begin{bmatrix} \frac{(x_{1j} - \bar{x}_1)}{\sqrt{s_{11}}} \\ \frac{(x_{2j} - \bar{x}_2)}{\sqrt{s_{22}}} \\ \vdots \\ \frac{(x_{pj} - \bar{x}_p)}{\sqrt{s_{pp}}} \end{bmatrix}, \quad j=1,2,\dots,n$$

Cuya matriz de covarianza muestral es la correlación muestral, \mathbf{R} , de las observaciones $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$. La estandarización elude el problema de tener una variable con varianza grande influenciando indevidamente la determinación de las ponderaciones factoriales.

La representación en (3-13), cuando es aplicado a la matriz de covarianzas muestrales \mathbf{S} o a la matriz de correlaciones \mathbf{R} , es conocida como la *solución de componentes principales*. El nombre se sigue del hecho que las ponderaciones factoriales son los coeficientes escalares de las primeras componentes principales.

Solución por Componentes Principales del modelo Factorial

El análisis factorial de componentes principales de la matriz de covarianzas muestrales \mathbf{S} , es especificada en términos de sus pares de valores propios y vectores propios $(\hat{\lambda}_1, \hat{\mathbf{e}}_1), (\hat{\lambda}_2, \hat{\mathbf{e}}_2), \dots, (\hat{\lambda}_p, \hat{\mathbf{e}}_p)$ donde $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_p$. Sea $m < p$ el número de factores comunes la matriz ponderaciones factoriales estimadas $\{\tilde{\ell}_{ij}\}$ esta dada por

$$\tilde{\mathbf{L}} = \left[\sqrt{\hat{\lambda}_1} \hat{\mathbf{e}}_1 \mid \sqrt{\hat{\lambda}_2} \hat{\mathbf{e}}_2 \mid \dots \mid \sqrt{\hat{\lambda}_m} \hat{\mathbf{e}}_m \right] \quad (3-15)$$

las varianzas específicas estimadas, son proporcionadas por los elementos de la diagonal

de la matriz $\mathbf{S} - \tilde{\mathbf{L}}\tilde{\mathbf{L}}'$, esto es

$$\tilde{\Psi} = \begin{bmatrix} \tilde{\Psi}_1 & 0 & \dots & 0 \\ 0 & \tilde{\Psi}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \tilde{\Psi}_p \end{bmatrix} \quad \text{con } \tilde{\Psi}_i = s_{ii} - \sum_{j=1}^m \tilde{\ell}_{ij}^2 \quad (3-16)$$

las comunalidades son estimadas como

$$\tilde{h}_i^2 = \tilde{\ell}_{i1}^2 + \tilde{\ell}_{i2}^2 + \dots + \tilde{\ell}_{im}^2 \quad (3-17)$$

la componente principal del análisis factorial de la matriz de correlación muestral es obtenida comenzando con \mathbf{R} en vez de \mathbf{S} .

Para la solución de componentes principales, la estimación de las ponderaciones factoriales para un factor dado, no cambia cuando el número de factores es incrementado. Por ejemplo, si $m=1$, $\tilde{\mathbf{L}} = [\sqrt{\hat{\lambda}_1} \hat{\mathbf{e}}_1]$ y si $m=2$, $\tilde{\mathbf{L}} = [\sqrt{\hat{\lambda}_1} \hat{\mathbf{e}}_1 \mid \sqrt{\hat{\lambda}_2} \hat{\mathbf{e}}_2]$, donde $(\hat{\lambda}_1, \hat{\mathbf{e}}_1)$ y $(\hat{\lambda}_2, \hat{\mathbf{e}}_2)$ son los dos primeros valores propios y vectores propios de \mathbf{S} ó \mathbf{R} .

Por la definición de $\tilde{\Psi}_i$, los elementos diagonales de \mathbf{S} , son iguales a los elementos diagonales de $\tilde{\mathbf{L}}\tilde{\mathbf{L}}' + \tilde{\Psi}$. Sin embargo, los elementos fuera de la diagonal de \mathbf{S} no son usualmente reproducidos por $\tilde{\mathbf{L}}\tilde{\mathbf{L}}' + \tilde{\Psi}$. ¿Cómo entonces, nosotros seleccionamos el número de factores m ?

Si el número de factores comunes, no es determinado por consideraciones a priori, tal como en teoría o como por otros investigadores, la elección de m puede ser basado en los valores propios estimados de la misma manera como en componentes principales.

Consideremos la *matriz residual*

$$\mathbf{S} - (\tilde{\mathbf{L}}\tilde{\mathbf{L}}' + \tilde{\Psi}) \quad (3-18)$$

Resultando de la aproximación de \mathbf{S} por la solución de componentes principales. Los elementos diagonales son cero y si los otros elementos son también pequeños, nosotros podemos tomar subjetivamente los m factores del modelo apropiadamente. Analíticamente, nosotros tenemos (ver Ejercicio 9.5 de [1])

$$\text{Suma de cuadrados entrados de } (\mathbf{S} - (\tilde{\mathbf{L}}\tilde{\mathbf{L}}' + \tilde{\Psi})) < \hat{\lambda}_{m+1}^2 + \dots + \hat{\lambda}_p^2 \quad (3-19)$$

Consecuentemente, un pequeño valor para la suma de cuadrados de la valores propios abandonados implica un pequeño valor para la suma de cuadrados de errores de aproximación.

Idealmente, la contribución de los primeros cuantos factores a las varianzas muestrales de las variables pueden ser mayores. La contribución de la varianza muestral s_{ii} desde los primeros factores comunes es ℓ_{i1}^2 . La contribución para el *total* de vananza muestral $s_{11} + s_{22} + \dots + s_{pp} = \text{Tr}(\mathbf{S})$, de el primer factor común es entonces

$$\tilde{\ell}_{11}^2 + \tilde{\ell}_{21}^2 + \dots + \tilde{\ell}_{p1}^2 = (\sqrt{\hat{\lambda}_1} \hat{\mathbf{e}}_1)' (\sqrt{\hat{\lambda}_1} \hat{\mathbf{e}}_1) = \hat{\lambda}_1$$

Ya que el vector propio $\hat{\mathbf{e}}_1$ tiene longitud uno. En general

$$\left(\begin{array}{l} \text{Proporción del total} \\ \text{de varianza muestral} \\ \text{debido al } j\text{-ésimo factor} \end{array} \right) = \begin{cases} \frac{\hat{\lambda}_j}{s_{11} + s_{22} + \dots + s_{pp}} & \text{para el análisis factorial de } \mathbf{S} \\ \frac{\hat{\lambda}_j}{p} & \text{para el análisis factorial de } \mathbf{R} \end{cases} \quad (3-20)$$

El criterio (3-20) es frecuentemente usado como un dispositivo heurístico para determinar el número apropiado de factores comunes. El número de factores comunes retenidos en el modelo es incrementado hasta que una "adecuada proporción" de la varianza total muestral haber sido explicada.

Otra convención, frecuentemente encontrada en paquetes de programa de computadoras, es hacer m igual al número de valores propios de \mathbf{R} mayores que uno si la matriz de correlación muestral es factorizada, o igual al número de valores propios positivos de si la matriz de covarianza muestral es factorizada. Esta reglas de truncamiento no puede ser utilizada indiscriminadamente.

El Método de Máxima Verosimilitud (MV)

Si los factores comunes \mathbf{F} y los factores específicos ε pueden ser asumidos a ser distribuidos normalmente, entonces la estimación de máxima verosimilitud de las ponderaciones factoriales y varianzas específicas pueden ser obtenidas. Cuando \mathbf{F}_j y ε_j son conjuntamente normal, las observaciones $\mathbf{X}_j - \mu = \mathbf{L}\mathbf{F}_j + \varepsilon$ son entonces normales, y por (4-16) de [1], la verosimilitud es

$$\begin{aligned}
 L(\mu, \Sigma) &= (2\pi)^{-np/2} |\Sigma|^{-n/2} \exp \left[-\frac{1}{2} \text{tr}[\Sigma^{-1}(\sum_{j=1}^n (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})' + n(\bar{\mathbf{x}} - \mu)(\bar{\mathbf{x}} - \mu)')] \right] \\
 &= (2\pi)^{-(n-1)p/2} |\Sigma|^{-(n-1)/2} \exp \left[-\frac{1}{2} \text{tr}[\Sigma^{-1}(\sum_{j=1}^n (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})')] \right] \\
 &\quad \times (2\pi)^{-p/2} |\Sigma|^{-1/2} \exp \left[-\frac{n}{2} (\bar{\mathbf{x}} - \mu)' \Sigma^{-1} (\bar{\mathbf{x}} - \mu) \right]
 \end{aligned}
 \tag{3-21}$$

El cual depende de \mathbf{L} y Ψ a través de $\Sigma = \mathbf{L}\mathbf{L}' + \Psi$. Este modelo aún no es bien definido debido a la multiplicidad de la elección para \mathbf{L} hecho por transformaciones ortogonales. Es deseable hacer a \mathbf{L} bien definido imponiendo la computacionalmente conveniente *condición de unicidad*

$$\mathbf{L}' \Psi^{-1} \mathbf{L} = \Delta \quad \text{una matriz diagonal}
 \tag{3-22}$$

Los estimadores máximo verosímiles de \mathbf{L} y Ψ deben ser obtenidas por aproximación numérica de (3-21). Afortunadamente en la actualidad existen eficientes programas de computadora que permiten tener la estimación máximo verosímil fácilmente.

Teorema 3.1. Sea $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ sea una muestra aleatoria de $N_p(\mu, \Sigma)$, donde $\Sigma = \mathbf{L}\mathbf{L}' + \Psi$ es la matriz de covarianzas para los m factores comunes del modelo (3-4). Los estimadores máximo verosímiles $\hat{\mathbf{L}}, \hat{\Psi}$, y $\hat{\mu} = \bar{\mathbf{x}}$ maximiza (3-21) sujeto a $\hat{\mathbf{L}}'\hat{\Psi}^{-1}\hat{\mathbf{L}}$ siendo una diagonal.

Los estimadores máximo verosímiles de las comunalidades son

$$\hat{h}_i^2 = \hat{\ell}_{i1}^2 + \hat{\ell}_{i2}^2 + \dots + \hat{\ell}_{im}^2 \quad \text{para } i = 1, 2, \dots, p \quad (3-23)$$

también

$$\left(\begin{array}{l} \text{Proporción de la varianza muestral} \\ \text{total debido al } j\text{-ésimo factor} \end{array} \right) = \frac{\hat{\ell}_{1j}^2 + \hat{\ell}_{2j}^2 + \dots + \hat{\ell}_{pj}^2}{s_{11} + s_{22} + \dots + s_{pp}} \quad (3-24)$$

Prueba

Por la propiedad de invarianza de estimadores máxima verosimilitud (ver sección 4.3 de [1]), funciones de \mathbf{L} y Ψ son estimados por las mismas funciones de $\hat{\mathbf{L}}$ y $\hat{\Psi}$. En particular, las comunalidades $h_i^2 = \ell_{i1}^2 + \ell_{i2}^2 + \dots + \ell_{im}^2$ tiene estimadores máxima verosimilitud $\hat{h}_i^2 = \hat{\ell}_{i1}^2 + \hat{\ell}_{i2}^2 + \dots + \hat{\ell}_{im}^2$. Lqqd.

Si, como en (2-10), las variables son estandarizadas tal que $\mathbf{Z} = \mathbf{V}^{-1/2}(\mathbf{X}-\mu)$, la matriz de covarianza ρ , de \mathbf{Z} tiene la representación

$$\rho = \mathbf{V}^{-1/2}\Sigma\mathbf{V}^{-1/2} = (\mathbf{V}^{-1/2}\mathbf{L})(\mathbf{V}^{-1/2}\mathbf{L})' + \mathbf{V}^{-1/2}\Psi\mathbf{V}^{-1/2} \quad (3-25)$$

De este modo ρ tiene una factorización análoga a (3-5) con matriz de ponderaciones

$\mathbf{L}_z = \mathbf{V}^{-1/2}\mathbf{L}$ y matriz de varianza específica $\Psi_z = \mathbf{V}^{-1/2}\Psi\mathbf{V}^{-1/2}$. Por la propiedad de invarianza de los estimadores máximo verosímiles, el estimador máximo verosímil de ρ es

$$\begin{aligned} \hat{\rho} &= (\hat{\mathbf{V}}^{-1/2}\hat{\mathbf{L}})(\hat{\mathbf{V}}^{-1/2}\hat{\mathbf{L}})' + \hat{\mathbf{V}}^{-1/2}\hat{\Psi}\hat{\mathbf{V}}^{-1/2} \\ &= \hat{\mathbf{L}}_z\hat{\mathbf{L}}_z'\hat{\Psi}_z \end{aligned} \quad (3-26)$$

donde $\hat{\mathbf{V}}^{-1/2}$ y $\hat{\mathbf{L}}$ son los estimadores máximo verosímiles de $\mathbf{V}^{-1/2}$ y \mathbf{L} , respectivamente (vea suplemento 9-A de [1]).

Como una consecuencia de la factorización de (3-26), siempre que el análisis de máxima verosimilitud pertenezca a la matriz de correlaciones, nosotros llamamos

$$\hat{h}_i^2 = \hat{\ell}_{i1}^2 + \hat{\ell}_{i2}^2 + \dots + \hat{\ell}_{im}^2 \quad i = 1, 2, \dots, p \quad (3-27)$$

los estimadores máximo verosímiles de las comunalidades y evaluamos la importancia de factores en base a

$$\left(\begin{array}{l} \text{Proporción de la varianza muestral} \\ \text{total (estandarizada) debido al } j\text{-ésimo factor} \end{array} \right) = \frac{\hat{\ell}_{1j}^2 + \hat{\ell}_{2j}^2 + \dots + \hat{\ell}_{pj}^2}{p} \quad (3-28)$$

para evitar más tediosa notación, los $\hat{\ell}_{ij}$'s denota los elementos de $\hat{\mathbf{L}}_z$.

Comentario

Ordinariamente las observaciones son estandarizadas y la matriz de correlación muestral es analizada factorialmente. La matriz de correlación muestral, \mathbf{R} , es insertada por $[(n-1)/n]\mathbf{S}$ en la función de verosimilitud de (3-21) y los estimadores máximo verosímiles $\hat{\mathbf{L}}_z$ y $\hat{\Psi}_z$ son obtenidos por computadoras usando los procedimientos esbozados en el suplemento 9A de [1]. Aunque la verosimilitud en (3-21) es apropiado para \mathbf{S} , no \mathbf{R} , sorprendentemente, esta práctica es equivalente a obtener los estimadores máximo verosímiles $\hat{\mathbf{L}}$ y $\hat{\Psi}$ basado en la matriz de covarianza muestral \mathbf{S} , haciendo $\hat{\mathbf{L}}_z = \hat{\mathbf{V}}^{-1/2}\hat{\mathbf{L}}$ y $\hat{\Psi}_z = \hat{\mathbf{V}}^{-1/2}\hat{\Psi}\hat{\mathbf{V}}^{-1/2}$. Aquí $\hat{\mathbf{V}}^{-1/2}$ es la matriz diagonal con el recíproco de las desviaciones estándar muestral (computado con el divisor \sqrt{n}) en la diagonal principal. Esta equivalencia debe haber sido aparentemente confuso en varias discusiones publicadas de análisis factorial. (Los detalles matemáticos son contenidos en Suplemento 9A de [1].)

3.4 ROTACIÓN FACTORIAL

Como indicamos en la sección 3.2, todas las ponderaciones obtenidas desde la ponderación inicial, por una transformación ortogonal tiene la misma habilitación para reproducir la matriz de covarianza (o correlación) [ver ecuación (3-8)]. Por el álgebra matricial conocemos que una transformación ortogonal corresponde a una rígida rotación (o reflexión) del eje de coordenadas. Por esta razón una transformación ortogonal de las ponderaciones factoriales y la implicada transformación ortogonal de los factores es llamada *rotación factorial*.

Si L es la $p \times m$ matriz estimador de las ponderaciones factoriales obtenida por cualquier método (componentes principales, máxima verosimilitud, etc) entonces

$$\hat{L}^* = \hat{L}T, \text{ donde } TT' = T'T = I \quad (3 - 29)$$

Es una matriz $p \times m$ de ponderaciones "rotadas". Además, la matriz de covarianza estimada (o correlación) permanece invariable, ya que

$$\hat{L}\hat{L}' + \hat{\Psi} = \hat{L}TT'\hat{L}' + \hat{\Psi} = \hat{L}^*\hat{L}^{*'} + \hat{\Psi} \quad (3 - 30)$$

La ecuación (3-30) indica que la matriz residual $S_n - \hat{L}\hat{L}' - \hat{\Psi} = S_n - \hat{L}^*\hat{L}^{*'} - \hat{\Psi}$

tiene restos inalterables. Además las varianzas específicas $\hat{\psi}_i$, y por tanto las comunalidades \hat{h}_i^2 , son inalterables. De este modo desde el punto de vista matemático, es indiferente si \hat{L} o L^* son obtenidas.

Kaiser [2] ha sugerido una medida analítica de estructura simple conocida como el *criterio varimax*. Define $\tilde{\ell}_{ij}^* = \hat{\ell}_{ij}^*/\hat{h}_i$ como los coeficientes escalares finales rotados por la raíz cuadrada de las comunalidades. El (normal) varimax procedimiento selecciona la transformación ortogonal T que hace

$$V = \frac{1}{p} \sum_{j=1}^m \left[\sum_{i=1}^p \tilde{\ell}_{ij}^{*4} - \left(\sum_{i=1}^p \tilde{\ell}_{ij}^{*2} \right)^2 / p \right] \quad (3 - 31)$$

tan grande como sea posible.

Los coeficientes rotados $\hat{\ell}_{ij}^*$ tienen el efecto de dar variables con pequeñas communalidades, relativamente más ponderadas en la determinación de una estructura simple. Después la transformación T es determinada, las ponderaciones $\tilde{\ell}_{ij}^*$ son multiplicadas por h_i tal que la comunalidad original es preservada.

Aunque mirando (3-31) se ve muy complicado y confuso, esto tiene una simple interpretación. En palabras

$$V \propto \sum_{j=1}^m \left(\begin{array}{l} \text{Varianza de cuadrados de (escalares)} \\ \text{ponderaciones para el } j\text{-ésimo factor} \end{array} \right) \quad (3-32)$$

Efectivamente, maximizando V corresponde a "esparcir" los cuadrados de las ponderaciones en cada factor tanto como sea posible. Por tanto esperamos encontrar grupos de mayor e insignificante coeficientes en alguna columna de la matriz de ponderaciones rotadas L^* .

Algoritmos computacionales existen para maximizar V y programas de computadoras de análisis factorial proporcionan rotación varimax.

3.5 MARCAS, TANTEADOR, SEÑAL O RAZÓN FACTORIAL (SCORES)

En análisis factorial, el interés está usualmente centrado en los parámetros del modelo factorial, sin embargo los valores estimados de los factores comunes, llamados *scores factoriales* pueden también ser requeridos. Estas cantidades son a veces usadas para propósitos de diagnóstico tanto como para la entrada a un subsecuente análisis.

Los scores factoriales no son estimaciones de parámetros desconocidos en el sentido usual. Más bien ellos son estimaciones de los valores de los vectores aleatorios factoriales no observados F_j , $j=1,2,\dots,n$. Esto es

$$\hat{f}_j = \text{estimado del valor } f_j \text{ obtenido por } F_j \text{ (j-ésimo caso)}$$

La situación de la estimación es complicada por el hecho que las cantidades no observables f_j y ε_j excede en número los observados x_j . Para superar esta dificultad haber sido avanzadas, algunas mas bien heurísticos, pero razonables, aproximaciones de el problema de estimación de valores factoriales. Se describirá dos de estas aproximaciones.

Ambas aproximaciones del score factorial tiene dos elementos en común.

1. Ellos tratan a los estimadores de las ponderaciones factoriales $\hat{\ell}_{ij}$, y varianzas específicas $\hat{\psi}_i$, como si ellos fueran los valores verdaderos.
2. Ellos involucran transformaciones lineales de los datos originales, quizás centrados o estandarizados. Típicamente. Los estimados *rotados* de las ponderaciones mas bien que los originales ponderaciones estimadas son usadas para calcular los scores factoriales. Las formulas computacionales, tanto dados en esta sección , no cambian cuando ponderaciones rotadas son sustituidas por ponderaciones no rotadas y no se diferenciará entre ellos.

El Método de Ponderaciones Mínimas Cuadradas

Suponga primero que el vector media μ , las ponderaciones factoriales, L , y las varianzas específicas Ψ , son conocidas para el modelo factorial

$$\begin{matrix} \mathbf{X} & - & \boldsymbol{\mu} & = & \mathbf{L} & \mathbf{F} & + & \boldsymbol{\varepsilon} \\ (p \times 1) & & (p \times 1) & & (p \times m) & (m \times 1) & & (p \times 1) \end{matrix}$$

adicional, atención a los factores específicos $\boldsymbol{\varepsilon}' = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p]$ como errores. Ya que $\text{Var}(\varepsilon_i) = \psi_i$, $i = 1, 2, \dots, p$, necesita no ser igual, Bartlett [3] ha sugerido que las ponderaciones mínimos cuadráticos pueden ser usadas para estimar los valores de factores comunes.

La suma de cuadrados de errores, ponderadas por el reciproco de sus varianzas, es

$$\sum_{i=1}^p \frac{\varepsilon_i^2}{\psi_i} = \boldsymbol{\varepsilon}' \boldsymbol{\Psi}^{-1} \boldsymbol{\varepsilon} = (\mathbf{x} - \boldsymbol{\mu} - \mathbf{L}\mathbf{f})' \boldsymbol{\Psi}^{-1} (\mathbf{x} - \boldsymbol{\mu} - \mathbf{L}\mathbf{f}) \quad (3-33)$$

Bartlett propone elegir los estimadores \mathbf{f} de \mathbf{f} para minimizar (3-33). La solución es (ver

ejercicio 7.3 de [1])

$$\hat{\mathbf{f}} = (\mathbf{L}'\hat{\Psi}^{-1}\mathbf{L})^{-1}\mathbf{L}'\hat{\Psi}^{-1}(\mathbf{x} - \hat{\boldsymbol{\mu}}) \quad (3-34)$$

Motivados por (3-34), nosotros tomamos los estimadores $\hat{\mathbf{L}}$, $\hat{\Psi}$, y $\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}}$ como los valores verdaderos y obtenemos los scores factoriales para el j-ésimo caso como

$$\hat{\mathbf{f}}_j = (\hat{\mathbf{L}}'\hat{\Psi}^{-1}\hat{\mathbf{L}})^{-1}\hat{\mathbf{L}}'\hat{\Psi}^{-1}(\mathbf{x}_j - \bar{\mathbf{x}}) \quad (3-35)$$

Cuando $\hat{\mathbf{L}}$ y $\hat{\Psi}$ son determinados por el método de máxima verosimilitud, estos estimados deben satisfacer la condición de unicidad $\hat{\mathbf{L}}'\hat{\Psi}^{-1}\hat{\mathbf{L}} = \hat{\Delta}$, una matriz diagonal. Nosotros tenemos lo siguiente

**Scores Factoriales Obtenidos por las Ponderaciones Mínimas Cuadradas
de los estimadores Máximo Verosímiles**

$$\hat{\mathbf{f}}_j = (\hat{\mathbf{L}}'\hat{\Psi}^{-1}\hat{\mathbf{L}})^{-1}\hat{\mathbf{L}}'\hat{\Psi}^{-1}(\mathbf{x}_j - \hat{\boldsymbol{\mu}}) = \hat{\Delta}^{-1}\hat{\mathbf{L}}\hat{\Psi}^{-1}(\mathbf{x}_j - \mathbf{x}) \quad j = 1, 2, \dots, n$$

o si la matriz de correlación es factorizada, (3-36)

$$\hat{\mathbf{f}}_j = (\hat{\mathbf{L}}_z'\hat{\Psi}_z^{-1}\hat{\mathbf{L}}_z)^{-1}\hat{\mathbf{L}}_z'\hat{\Psi}_z^{-1}\mathbf{z}_j = \hat{\Delta}_z^{-1}\hat{\mathbf{L}}_z\hat{\Psi}_z^{-1}\mathbf{z}_j \quad j = 1, 2, \dots, n$$

donde $\mathbf{z}_j = \mathbf{D}^{-1/2}(\mathbf{x}_j - \mathbf{x})$ y $\hat{\rho} = \hat{\mathbf{L}}_z\hat{\mathbf{L}}_z' + \hat{\Psi}_z$

Los scores factoriales generados por (3-36) tiene vector media muestral $\mathbf{0}$ y matriz de covarianzas muestrales cero (ver ejercicio 9.15 de [1]).

Si ponderaciones rotadas $\hat{\mathbf{L}}^* = \hat{\mathbf{L}}\mathbf{T}$ son usadas en lugar de las ponderaciones originales en (3-36), las subsecuentes scores factoriales $\hat{\mathbf{f}}_j^*$, son relacionadas a $\hat{\mathbf{f}}_j$ por

$$\hat{\mathbf{f}}_j^* = \mathbf{T}'\hat{\mathbf{f}}_j \quad j = 1, 2, \dots, n$$

Comentario. Si las ponderaciones factoriales son estimadas por el método de componentes principales, es costumbre generar scores usando un no ponderado (ordinario) procedimiento

de mínimos cuadrados. Implícitamente esto asciende a asumir los ψ_i son iguales o casi iguales. Los scores factoriales son entonces

$$\hat{\mathbf{f}}_j = (\hat{\mathbf{L}}' \hat{\mathbf{L}})^{-1} \hat{\mathbf{L}}' (\mathbf{x}_j - \bar{\mathbf{x}})$$

$$\text{o } \hat{\mathbf{f}}_j = (\hat{\mathbf{L}}_z' \hat{\mathbf{L}}_z)^{-1} \hat{\mathbf{L}}_z' \mathbf{z}_j$$

Para datos estandarizados. Ya que

$$\mathbf{L} = \left[\sqrt{\hat{\lambda}_1} \hat{\mathbf{e}}_1 \mid \sqrt{\hat{\lambda}_2} \hat{\mathbf{e}}_2 \mid \dots \mid \sqrt{\hat{\lambda}_m} \hat{\mathbf{e}}_m \right]$$

[ver (3-15)], tenemos

$$\hat{\mathbf{f}}_j = \begin{bmatrix} \frac{1}{\sqrt{\hat{\lambda}_1}} \hat{\mathbf{e}}_1' (\mathbf{x}_j - \bar{\mathbf{x}}) \\ \frac{1}{\sqrt{\hat{\lambda}_2}} \hat{\mathbf{e}}_2' (\mathbf{x}_j - \bar{\mathbf{x}}) \\ \vdots \\ \frac{1}{\sqrt{\hat{\lambda}_m}} \hat{\mathbf{e}}_m' (\mathbf{x}_j - \bar{\mathbf{x}}) \end{bmatrix} \quad (3-37)$$

Para estos scores factoriales

$$\frac{1}{n} \sum_{j=1}^n \hat{\mathbf{f}}_j = \mathbf{0} \quad (\text{media muestral})$$

Y

$$\frac{1}{n-1} \sum_{j=1}^n \hat{\mathbf{f}}_j \hat{\mathbf{f}}_j' = \mathbf{I} \quad (\text{covarianza muestral})$$

Comparando (3-51) con (8-21) de [1], nosotros vemos que las $\hat{\mathbf{f}}_j$ es nada mas que las primeras m (escalares) componentes principales evaluadas en la \mathbf{x}_j .

El Método de Regresión

Comenzando otra vez con el modelo factorial original $\mathbf{X} - \mu = \mathbf{L} \mathbf{F} + \varepsilon$, inicialmente

tratamos la matriz de ponderaciones \mathbf{L} y la matriz de varianza específica Ψ como conocida. Cuando los factores comunes \mathbf{F} y los factores específicos (o errores) ε son conjuntamente normalmente distribuidos con medias y covarianzas dadas por (3-3) la combinación lineal $\mathbf{X} - \mu = \mathbf{L}\mathbf{F} + \varepsilon$ tiene una distribución $N_p(\mathbf{0}, \mathbf{L}\mathbf{L}' + \Psi)$ (ver Resultado 4.3 de [1]). Además, la distribución conjunta de $(\mathbf{X} - \mu)$ y \mathbf{F} es $N_{(m+p) \times p}(\mathbf{0}, \Sigma^*)$, donde

$$\Sigma^*_{(m+p) \times (m+p)} = \begin{bmatrix} \Sigma = \mathbf{L}\mathbf{L}' + \Psi & \mathbf{L} \\ \text{-----} & \text{-----} \\ \mathbf{L}' & \mathbf{I} \end{bmatrix} \quad (3-38)$$

$(p \times p)$ $(p \times m)$
 $(m \times p)$ $(m \times m)$

y $\mathbf{0}$ es una $(m+p) \times p$ matriz de ceros. Usando el Resultado 4.6 de [1] la distribución condicional de \mathbf{F}/\mathbf{x} es normal multivariante con

$$\text{media} = E(\mathbf{F} / \mathbf{x}) = \mathbf{L}' \Sigma^{-1} (\mathbf{x} - \mu) = \mathbf{L}' (\mathbf{L}\mathbf{L}' + \Psi)^{-1} (\mathbf{x} - \mu) \quad (3-39)$$

y

$$\text{covarianza} = \text{Cov}(\mathbf{F} / \mathbf{x}) = \mathbf{I} - \mathbf{L}' \Sigma^{-1} \mathbf{L} = \mathbf{I} - \mathbf{L}' (\mathbf{L}\mathbf{L}' + \Psi)^{-1} \mathbf{L} \quad (3-40)$$

Las cantidades $\mathbf{L}' (\mathbf{L}\mathbf{L}' + \Psi)^{-1}$ en (3-39) son los coeficientes en una (multivariante) regresión de los factores sobre las variables. Estimadores de estos coeficientes produce scores factoriales que son análogos a los estimadores de la condicional valor medio en análisis de regresión multivariante. Consecuentemente, dado cualquier vector de observaciones \mathbf{x}_j y tomando los estimadores máximo verosímiles de $\hat{\mathbf{L}}$ y $\hat{\Psi}$ como los valores verdaderos, el j -ésimo score vectorial factorial esta dado por

$$\hat{\mathbf{f}}_j = \hat{\mathbf{L}}' \hat{\Sigma}^{-1} (\mathbf{x}_j - \bar{\mathbf{x}}) = \hat{\mathbf{L}}' (\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\Psi})^{-1} (\mathbf{x}_j - \bar{\mathbf{x}}), \quad j = 1, 2, \dots, n \quad (3-41)$$

El calculo de $\hat{\mathbf{f}}_j$ en (3-41) puede ser simplificado usando la matriz identidad (ver ejercicio 9.6 de [1])

$$\hat{\mathbf{L}}' (\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\Psi})^{-1} = (\mathbf{I} + \hat{\mathbf{L}}' \hat{\Psi}^{-1} \hat{\mathbf{L}})^{-1} \hat{\mathbf{L}}' \hat{\Psi}^{-1} \quad (3-42)$$

$(m \times p)$ $(p \times p)$ $(m \times m)$ $(m \times p)$ $(p \times p)$

Esta identidad nos permite comparar los scores factoriales en (3-41), generado por el argumento de regresión, con estos generados por los procedimientos de mínimos cuadrados ponderados [ver (3-36)]. Temporalmente, nosotros denotaremos el anterior por $\hat{\mathbf{f}}_j^R$ y la posterior por $\hat{\mathbf{f}}_j^{LS}$. Entonces, usando (3-42),

$$\hat{\mathbf{f}}_j^{LS} = (\hat{\mathbf{L}}' \hat{\Psi}^{-1} \hat{\mathbf{L}})^{-1} (\mathbf{I} + \hat{\mathbf{L}}' \hat{\Psi}^{-1} \hat{\mathbf{L}}) \hat{\mathbf{f}}_j^R = (\mathbf{I} + (\hat{\mathbf{L}}' \hat{\Psi}^{-1} \hat{\mathbf{L}})^{-1}) \hat{\mathbf{f}}_j^R \quad (3-43)$$

Para estimadores máximo verosímiles $(\hat{\mathbf{L}}' \hat{\Psi}^{-1} \hat{\mathbf{L}})^{-1} = \hat{\Delta}^{-1}$ y si los elementos de esta matriz diagonal son ceros, la regresión y el método generalizado de mínimos cuadrados dan casi los mismos scores factoriales.

En un intento para reducir los efectos de una (posible) incorrecta determinación de el número de factores, los que ejercen la profesión tienden a calcular los scores factoriales en (3-41) usando \mathbf{S} (la original matriz de covarianza muestral) en vez de $\hat{\Sigma} = \hat{\mathbf{L}} \hat{\mathbf{L}}' + \hat{\Psi}$. Y tenemos lo siguiente.

Scores factoriales obtenidos por Regresión

$$\hat{\mathbf{f}}_j = \hat{\mathbf{L}}' \mathbf{S}^{-1} (\mathbf{x}_j - \mathbf{x}) \quad j=1,2, \dots, n$$

O si la matriz de correlación es factorizada (3-44)

$$\hat{\mathbf{f}}_j = \hat{\mathbf{L}}_z' \mathbf{R}^{-1} \mathbf{z}_j \quad j=1,2, \dots, n$$

Donde

$$\mathbf{z}_j = \mathbf{D}^{-1/2} (\mathbf{x}_j - \mathbf{x}) \quad y \quad \hat{\rho} = \hat{\mathbf{L}}_z \hat{\mathbf{L}}_z' + \hat{\Psi}_z$$

Otra vez, si ponderaciones rotadas $\hat{\mathbf{L}}^* = \hat{\mathbf{L}} \mathbf{T}$ son usadas en lugar de las ponderaciones en (3-44), el subsecuente score factorial $\hat{\mathbf{f}}_j^*$ son relacionados a $\hat{\mathbf{f}}_j$ por

$$\hat{\mathbf{f}}_j^* = \mathbf{T}' \hat{\mathbf{f}}_j, \quad j=1,2, \dots, n$$

3.6 PERSPECTIVAS Y ESTRATEGIA PARA ANÁLISIS FACTORIAL

Hay muchas decisiones que deben hacerse en un estudio factorial analítico. Probablemente la más importante decisión es la elección de m , el número de factores comunes. Aunque un test muestral de mayor adecuabilidad esta disponible para un m dado, esto es adecuado solamente para datos que son aproximadamente normalmente distribuidos. Además el test permitirá seguramente rechazar lo adecuado del modelo para pequeños m , si el número de variables y observaciones es grande. Sin embargo esta es la situación cuando el análisis factorial proporciona una aproximación útil. Frecuentemente, la elección final de m esta basado en alguna combinación de: (1) La proporción de la varianza explicada, (2) sujeto al conocimiento de la materia, y (3) lo "razonable" de los resultados.

La elección del método de solución y tipo de rotación son decisiones menos cruciales. En efecto los más satisfactorios análisis factorial son aquellos donde rotaciones son probados con más de un método y todos los resultados substancialmente confirman la misma estructura factorial.

En el presente, el análisis factorial todavía mantiene el sabor de un arte y no de una estrategia singular que puede sin embargo ser cincelado en piedra. Se sugiere e ilustra una opción razonable.

1. Realizar un análisis factorial por el método de componentes principales. Este método es particularmente apropiado para un primer paso a través de los datos (esto no requiere que R o S sean no singulares).
 - a) Verificar observaciones sospechosas en la representación gráfica de los scores factoriales. También se puede calcular scores estandarizados para cada observación y distancias cuadradas como los descritos en la Sección 4.6 de [1].
 - b) Probar la rotación varimax.
2. Realizar un análisis factorial por el método de máxima verosimilitud incluyendo una

rotación varimax.

3. Comparar las soluciones del análisis factorial
 - Observar si las ponderaciones están agrupadas de la misma manera.
 - Representar gráficamente los scores factoriales obtenidos por componentes principales contra scores del análisis de máxima verosimilitud.
4. Repita los primeros tres pasos para otro número de factores comunes m . Verificar si los factores extras necesariamente contribuyen en el entendimiento e interpretación de los datos.
5. Para conjuntos grandes de datos partílos en la mitad y realizar un análisis factorial en cada parte. Compare las dos soluciones con el otro y que obtiene desde el conjunto completo de datos para chequear la estabilidad de la solución.(Los datos pueden ser divididos en forma aleatoria o eligiendo primero los casos enumerados con números impares en el primer grupo y los casos pares en el otro grupo.)

CAPÍTULO 4

ANÁLISIS DE RESULTADOS

4.1 Introducción

Los datos a los que se tuvieron acceso para la elaboración de la presente tesis fueron de diferentes años, para ser mas precisos de los años 1991, 1993 y 1996.

Se construyeron dos tipos de mapas referentes a salud por medio de la técnica del análisis factorial :

- a) Mapas de Salud.

- b) Mapas de Tendencia de Inversión en Salud.

Para el primer caso, los Mapas de Salud, se trabajó a nivel departamental para el año 1991 con los datos de la publicación de 1993 (ver bibliografía [4]). El Mapa de Salud a nivel provincial para el año 1993 se hizo con datos publicados por el INEI elaborados en base al Censo Nacional de Población y Vivienda de 1993.

Para el segundo caso, en los Mapas de Tendencia de Inversión en Salud se empleó datos del 2° Censo de Infraestructura Sanitaria y Recursos del Sector Salud realizada en el año de 1996 por el Ministerio de Salud, estudio que registro datos como Recursos Humanos, Infraestructura física, Equipamiento, etc. de todas las instituciones públicas y privadas que prestan servicios de salud , en ese sentido se tuvo la ventaja de contar con la base de datos creada para registrar dicha información, conjuntamente con sus libros de códigos, lo cual nos permitió la construcción de datos consolidados que podían ser a nivel departamental, a nivel provincial, consolidados de todas las instituciones de salud, consolidados referentes sólo al Ministerio de Salud, para finalmente elaborar índices, como tasas por 10 mil habitantes y/o porcentajes de establecimientos con ciertos servicios básicos y de seguridad.

4.2 Metodología

Para el caso de los Mapas de Salud no se tuvo que trabajar para obtener datos consolidados pues los datos ya estaban listos, en lo correspondiente a los Mapas de Tendencia de Inversión en el Sector Salud para obtención de los datos consolidados se trabajó con un software para bases de datos relacionales y los comandos utilizados se muestran en el Anexo 5. Luego en una hoja electrónica de cálculo se elaboraron los índices por 10 mil habitantes y/o porcentajes de establecimientos con servicios.

Una vez que ya se tenía los datos consolidados (índices y/o porcentajes) se procedió a trabajar con un software estadístico usando comandos para el análisis factorial (ver el Anexo 6)

Se obtuvieron los diferentes tipos de resultados de acuerdo a lo establecido en lo siguiente:

- **Mapa de Salud :**

Nivel Departamental del Perú : 1991 (Mapa 01 y Mapa 02).

Nivel Provincial del Perú : 1993 (Mapa 03 y Mapa 04).

- **Mapa de Tendencia de Inversión en Salud :**

Nivel Departamental del Perú : 1996 (de Mapa 05 a Mapa 12).

Nivel Provincial del Perú : 1996 (de Mapa 13 a Mapa 20).

Para el análisis factorial se trabajó con el método de componentes principales (abreviado como **CP**) y el método de máxima verosimilitud (abreviado como **MV**).

Los scores factoriales se obtuvieron con el método de regresión (ver Anexo 1).

Para la interpretación de los resultados se tiene adjunto el cuadro de ponderaciones factoriales, comunalidad y porcentaje de varianza explicada con el respectivo gráfico de las variables en los dos primeros factores. Los datos correspondientes y la matriz de correlaciones se encuentran listados en los Anexos 1 y 2 al final de la tesis.

4.3 Análisis de Resultados

Resultados de Mapa 01

En el primer mapa a nivel departamental, desarrollada con el método de análisis factorial en la modalidad de componentes principales, vemos que, el **Cuadro A1** (Anexo 1) tiene casi el mismo orden de el cuadro "Indicadores de Salud Standarizados Ordenados en base al Promedio" de la publicación de Mapa de Salud del Perú, indicando la validez de ambos resultados. Al verificar la matriz la correlaciones, **Cuadro B1** (Anexo 2) para las diez variables se puede ver que tienen altos valores, oscilando alrededor de 0.8, la mas alta correlación se obtiene entre Esperanza de Vida y la Tasa de Mortalidad infantil con $-.995$.

En el **Cuadro 4.1.1** el primer factor tiene un alto porcentaje de varianza de explicación con un 77.24 %, también en el listado de las ponderaciones del primer factor en las diez variables se ve que es de alto valor, la mayoría alrededor de 0.9, teniendo en cuenta que se ha factorizado la matriz de correlaciones, trabajadose con datos estandarizados.

En la **Figura 4.1.1** es de gran importancia pues allí se ve como son las ponderaciones en el Primer factor que denominaremos el "Nivel de Salud", vemos como influye negativamente la Tasa de Mortalidad Infantil, Tasa Bruta de Mortalidad, la Tasa Global de Fecundidad y tasa de analfabetismo y como influye positivamente el gasto percapita, la tasa de camas por habitante y médicos por habitante, la tasa de población urbana luz y agua y esperanza de vida al nacer, es decir estos últimos influyen positivamente en el factor de "Nivel de Salud".

En la **Figura 4.1.2**, el mapa No. 1, muestra como esta estructurado en nivel de salud del Perú de manera muy similar al de la publicación del mapa de salud del Perú de 1991, en ellos se ve que los que se encuentran en peor situación de Salud son los departamentos de Huancavelica, Apurimac, Ayacucho, Cuzco y Cajamarca y los que se encuentran en Mejor situación de salud son los departamento de Lima y Callao, Moquegua Ica y Tacna.

Cuadro 4.1.1 Mapa de Salud Nacional por Departamentos 1991
Método de Análisis Factorial - CP

Variables	Ponderaciones factoriales estimadas		Comunalidad
	F1	F2	
LUZAGUA	0.830	-0.174	0.720
POBURBAN	0.918	-0.057	0.845
TBM	-0.925	0.165	0.883
TMI	-0.925	0.326	0.962
TGF	-0.954	0.100	0.921
ESPERANZ	0.923	-0.311	0.948
ANALFAB	-0.894	-0.029	0.801
MEDICOS	0.846	0.281	0.794
CAMAS	0.813	0.442	0.856
GASTO	0.737	0.557	0.852
Porcentaje acumulado del total de varianza muestral (estandarizada) explicada	77.241	85.820	

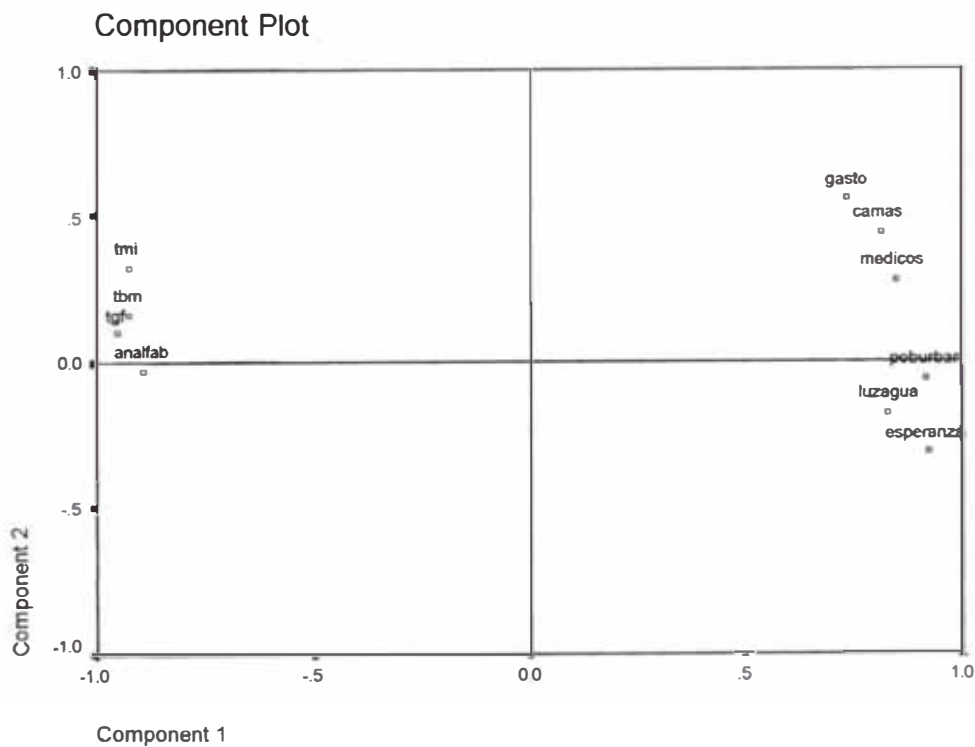


Figura 4.1.1 Ponderaciones de variables en los dos primeros factores (Mapa 01)

MAPA No. 1
 MAPA DE SALUD
 PERU : 1991



Elaboración Propia

Mapa_01.wor

Figura 4.1.2 MAPA No. 1 : Nivel Departamental. Método de Análisis Factorial - CP

Resultados de Mapa 02

En este mapa a nivel departamental, desarrollada con el análisis factorial y con el método de máxima verosimilitud, vemos que, en el **Cuadro A1** (Anexo 1) se tiene valores del score correspondiente al primer factor que producen un ordenamiento un poco diferente del cuadro "Indicadores de Salud Standarizados Ordenados en base al Promedio" de la publicación de Mapa de Salud del Perú, por lo que estos resultados pueden ser interpretados con precaución. La matriz de correlaciones es la misma que para el mapa anterior **Cuadro B1** (Anexo 2).

En el **Cuadro 4.2.1** el primer factor tiene un alto porcentaje de varianza de explicación con un 67.66 %, como también en el listado de las ponderaciones del primer factor en las diferentes diez variables se ve que es de alto valor, la mayoría alrededor de 0.9, teniendo en cuenta que se ha factorizado la matriz de correlaciones, es decir se ha trabajado con datos estandarizados.

La **Figura 4.2.1** es de gran importancia pues allí se ve como influyen sus ponderaciones en el Primer factor que denominaremos el "Nivel de Salud", vemos como influye negativamente la Tasa de Mortalidad Infantil, Tasa bruta de Mortalidad, la tasa global de fecundidad y tasa de analfabetismo y como influye positivamente el gasto percapita, la tasa de camas por habitante y médicos por habitante, la tasa de población urbana luz y agua y esperanza de vida al nacer, es decir estos últimos influyen positivamente en el factor de "Nivel de Salud". Además se observa una ligera rotación con respecto a la **Figura 4.1.1**.

En la **Figura 4.2.2**, el mapa No. 2, muestra como esta estructurado en nivel de salud del Perú de manera no muy similar al de la publicación del mapa de salud del Perú de 1991, por lo que hay que tomarlo con cuidado, en ellos se ve que los que se encuentran en peor situación de Salud son los departamentos de Huancavelica, Apurímac, Ayacucho, Cuzco y Puno y los que se encuentran en Mejor situación de salud son los departamento de

Cuadro 4.2.1 Mapa de Salud nacional por departamentos 1991
Método de Análisis Factorial - MV

Variables	Ponderaciones factoriales estimadas		Comunalidad
	F1	F2	
LUZAGUA	0.780	0.212	0.654
POBURBAN	0.823	0.377	0.820
TBM	-0.943	-0.137	0.907
TMI	-0.999	0.011	0.999
TGF	-0.882	-0.325	0.884
ESPERANZ	0.995	-0.006	0.990
ANALFAB	-0.805	-0.387	0.798
MEDICOS	0.717	0.436	0.705
CAMAS	0.626	0.607	0.760
GASTO	0.517	0.652	0.694
Porcentaje acumulado del total de varianza muestral (estandarizada) explicada	67.660	82.119	

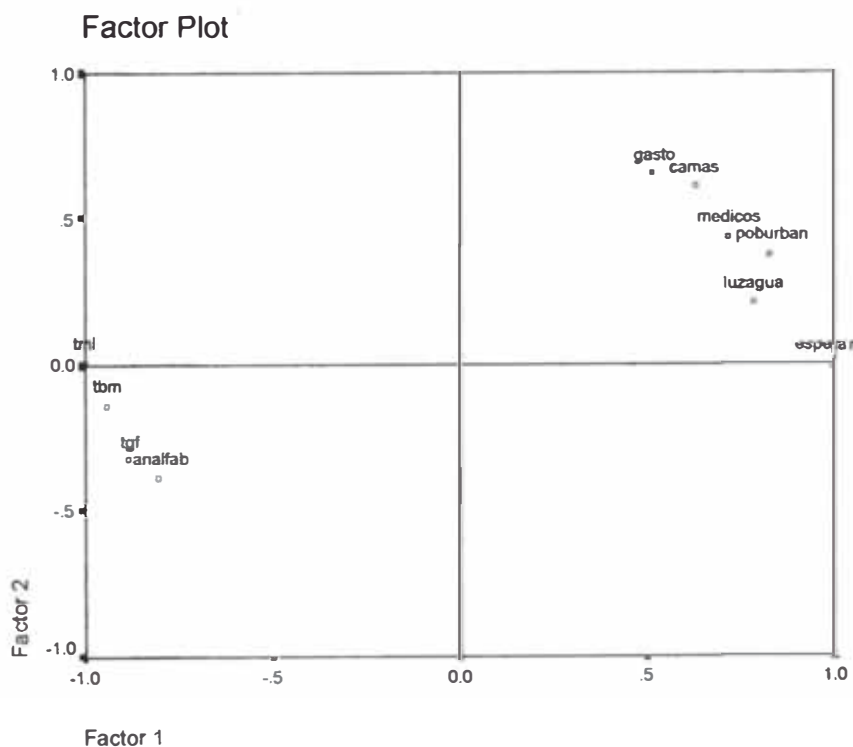


Figura 4.2.1 Ponderaciones de las variables en los dos primeros factores (Mapa 02)

MAPA No. 2
MAPA DE SALUD
PERU : 1991



Elaboración Propia

Mapa_02.wor

Figura 4.2.2 MAPA No. 2 : Nivel Departamental. Método de Análisis Factorial - MV

Lima y Callao, Ica, Arequipa, y Lambayeque

Resultados de Mapa 03

En este mapa de Salud a nivel Provincial, desarrollada con el análisis factorial y con el método de Componentes Principales, vemos que, el **Cuadro A2** (Anexo 1) el orden que tienen las diferentes 188 provincias que en 1993 tenía el Perú, estos resultados confirman o están de acuerdo con el mapa No.1 que está diseñado a nivel departamental, pues la zona más oscura que indica la peor situación de salud, se encuentra justamente en el trapecio andino, es decir en los departamentos de Apurímac, Huancavelica, Ayacucho y Cuzco.

Al verificar la matriz de correlaciones, **Cuadro B2** (Anexo 2) para las diez variables se puede ver que tienen altos valores, oscilando alrededor de 0.7, la más alta correlación se obtiene entre las variables de Desnutrición de niños y porcentaje de viviendas sin desagüe con 0.801

En el **Cuadro 4.3.1** el primer factor tiene un alto porcentaje de varianzas de explicación con un 71.24 %, como también en el listado de las ponderaciones del primer factor en las diferentes ocho variables se ve que es de alto valor, la mayoría alrededor de 0.8, teniendo en cuenta que se ha factorizado la matriz de correlaciones, es decir con datos estandarizados.

En la **Figura 4.3.1** es de gran importancia pues allí se ve como influyen sus ponderaciones en el Primer factor que denominaremos el negativo del "Nivel de Salud", vemos como influye en él, la Tasa de Mortalidad Infantil, Tasa de analfabetismo, Porcentaje de desnutrición, Porcentaje de viviendas sin agua y sin desagüe, y en el otro sentido están el porcentaje de población urbana, la tasa de médicos y tasa de camas, es decir influyen negativamente en el negativo del "Nivel de Salud".

En la **Figura 4.3.2** el mapa No. 3, muestra como está estructurado en nivel de salud

Cuadro 4.3.1 Mapa de salud nacional por provincia 1993
Método de Análisis Factorial - CP

Variables	Ponderaciones factoriales estimadas		Comunalidad
	F1	F2	
POBURBAN	-0.887	-0.052	0.789
TMI	0.783	0.368	0.748
TANALFFE	0.861	0.294	0.827
TMEDICOS	-0.850	0.368	0.857
TCAMAS	-0.758	0.543	0.87
DESNINO	0.907	0.117	0.836
SAGUA	0.821	0.040	0.676
SDESAGUE	0.876	-0.001	0.767
Porcentaje acumulado del total de varianza muestral (estandarizada) explicada	71.237	79.615	

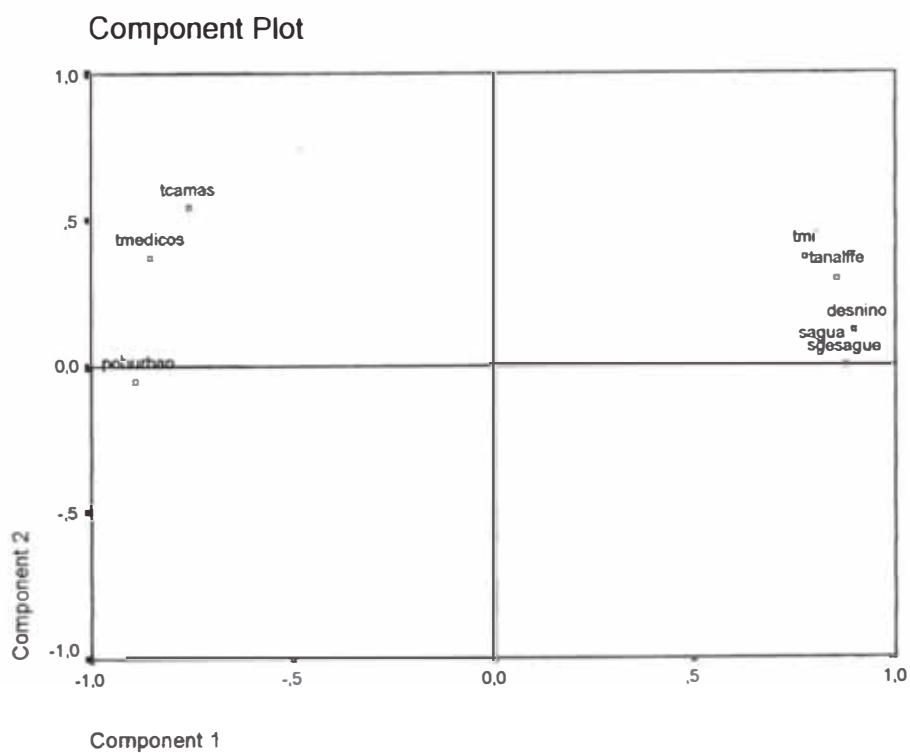
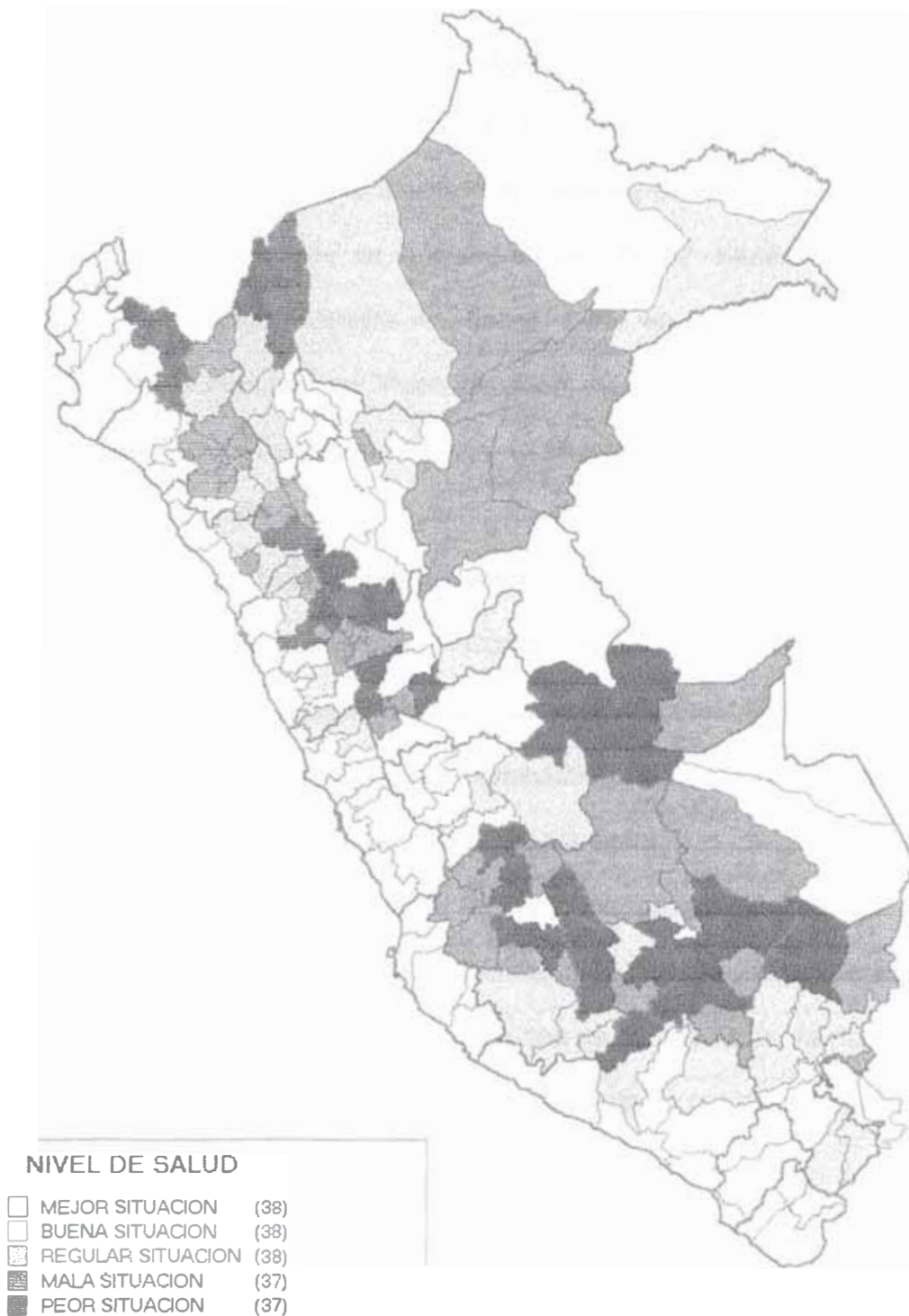


Figura 4.3.1 Ponderaciones de las variables en los dos primeros factores (Mapa 03)

MAPA No. 3
MAPA DE SALUD PROVINCIAL
PERU : 1993



Elaboración Propia.

Mapa_03.wor

Figura 4.3.2 MAPA No. 3 : Nivel Provincial. Método de Análisis Factorial - CP

del Perú de manera muy similar al mapa 1, pues la zona del trapecio andino (departamentos de Huancavelica, Ayacucho, Apurímac, Cuzco) y la zona norte de Puno se encuentran en Peor situación en el Nivel de Salud, se observa también que en la Provincia del Cuzco, capital de departamento esta en "Mejor situación" del nivel de Salud, contrastando con lo anteriormente dicho. Luego en la Sierra Norte también se ve una "Mala" y "Peor Situación" de Salud. Las provincias que se encuentran en "Mejor" situación de salud son Ilo en Moquegua, Callao, Jorge Basadre en Tacna, Lima en Lima y así sucesivamente. Las provincias que se encuentran en "Peor Situación" de Salud son Cotabambas en Apurímac, Chumbivilcas en Cuzco, Canas en Cuzco, y así sucesivamente.

Resultados de Mapa 04

En este mapa a nivel Provincial, desarrollada con el análisis factorial y con el método de máxima verosimilitud, vemos que, el **Cuadro A2** (Anexo 1) en la columna de Score 1° Factor MV no cordina con el ordenamiento dado por el Score 1° Factor CP, por lo que estos resultados pueden ser interpretados con precaución. La matriz la correlaciones **Cuadro B2** (Anexo 2) para las diez variables es igual al caso anterior.

En el **Cuadro 4.4.1**, el primer factor tiene un alto porcentaje de varianza de explicación con un 67.518 %, como también en el listado de las ponderaciones del primer factor en las diferentes ocho variables se ve que es de alto valor, la mayoría alrededor de 0.8, teniendo en cuenta que se ha factorizado la matriz de correlaciones, es decir con datos estandarizados.

En la **Figura 4.4.1** es de gran importancia pues allí se ve como influyen sus ponderaciones en el Primer factor que denominaremos el negativo del "Nivel de Salud", vemos como influye positivamente la Tasa de analfabetismo, la TMI, Desnutrición de niños, porcentaje de viviendas sin agua y porcentaje de viviendas sin desagüe, por el lado negativo tenemos las siguientes variables: la Tasa de médicos , tasa de camas por habitante,

Cuadro 4.4.1 Mapa de salud nacional por provincia 1993
Método de Análisis Factorial - MV

Variables	Ponderaciones factoriales estimadas		Comunalidad
	F1	F2	
POBURBAN	-0.857	-0.169	0.763
TMI	0.727	0.245	0.588
TANALFFE	0.820	0.323	0.776
TMEDICOS	-0.889	0.340	0.907
TCAMAS	-0.753	0.271	0.641
DESNINO	0.888	0.203	0.829
SAGUA	0.772	0.143	0.616
SDESAGUE	0.852	0.076	0.732
Porcentaje acumulado del total de varianza muestral (estandarizada) explicada	67.518	73.141	

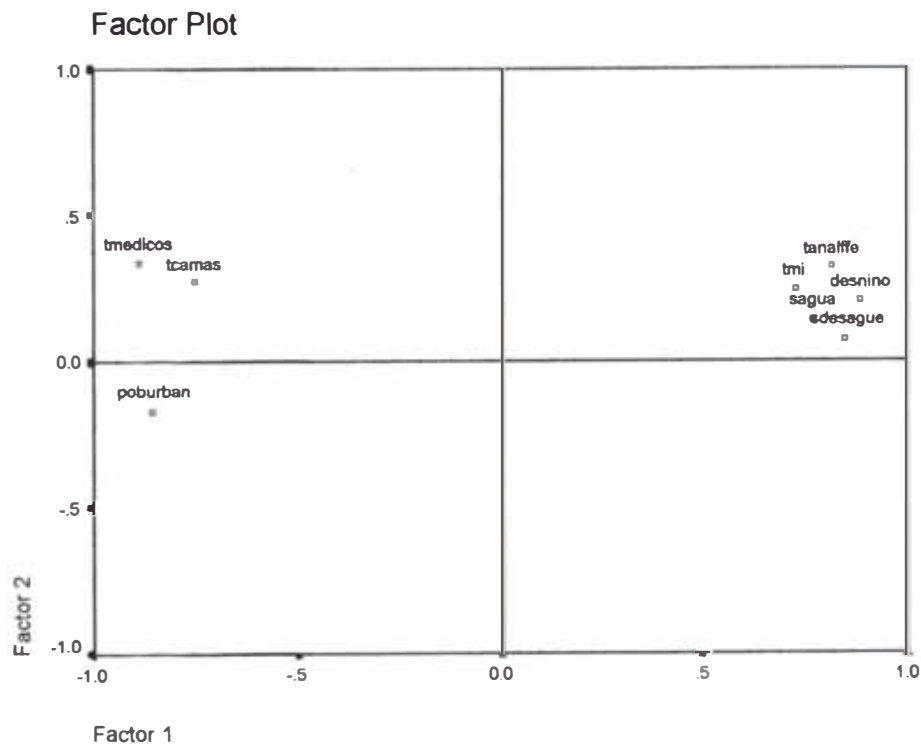
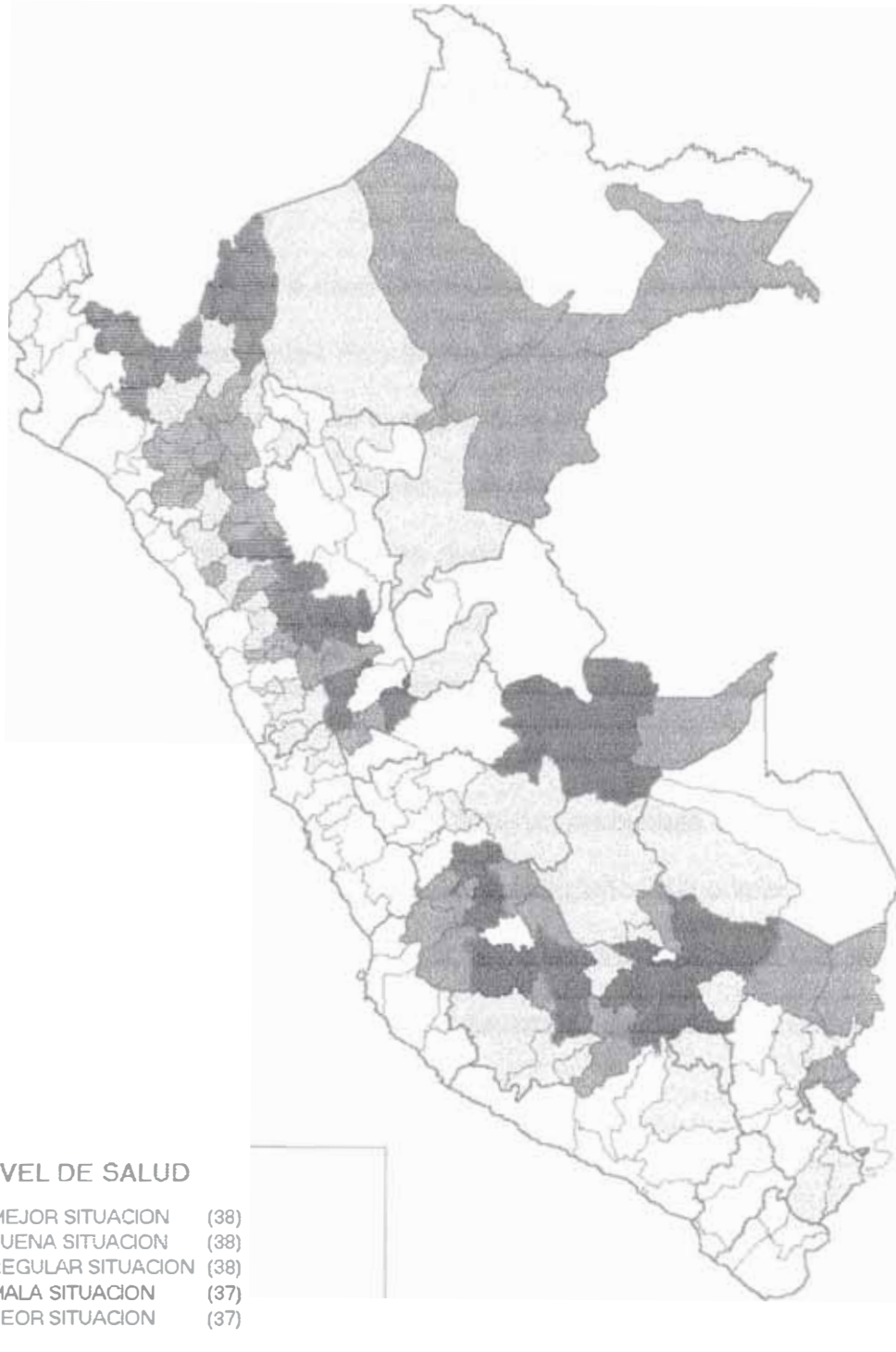


Figura 4.4.1 Ponderaciones de las variables en los dos primeros factores (Mapa 04)

MAPA No. 4
MAPA DE SALUD PROVINCIAL
PERU : 1993



Elaboración Propia.

Mapa_04.wor

Figura 4.4.2 MAPA No. 4 : Nivel Provincial. Método de Análisis Factorial (MV)

porcentaje de población urbana.

El mapa No. 4 (**Figura 4.4.2**), muestra como esta estructurado en nivel de salud del Perú de manera muy similar que en el caso anterior, indicando una "Peor Situación" de salud en la zona del trapecio andino y en la zona de la sierra norte. Por otro lado se sigue observando una "Mejor Situación" de salud en la zona de la costa.

Resultados de Mapa 05

En este mapa a nivel Departamental, desarrollada con el análisis factorial y con el método de Componentes Principales, en el diseño de este mapa se han empleado datos expresados todos como una tasa por diez mil habitantes, y donde se emplea datos pertenecientes a todas las instituciones que prestan servicios en salud, así tenemos el Ministerio de salud, el IPSS que hoy se denomina EsSalud, Fuerzas armadas y policiales, Clínicas Privadas lucrativas y otros. Vemos que, el **Cuadro A3** (Anexo 1). La matriz la correlaciones (**Cuadro B3**, Anexo 2) para las quince variables tienen gran amplitud variando desde 0.04 hasta 0.895.

En el **Cuadro 4.5.1** el primer factor tiene un porcentaje de varianza de explicación de 58.5 %, como también en el listado de las ponderaciones del primer factor en las diferentes quince variables se ve que es de alto valor, la mayoría alrededor de 0.8, teniendo en cuenta que se ha factorizado la matriz de correlaciones, es decir que se trabaja con datos estandarizados.

La Figura 4.5.1 es de gran importancia pues allí se ve como influyen sus ponderaciones en el Primer factor que denominaremos "Nivel de Inversión", vemos como influye positivamente las quince variables.

El mapa No. 5 (**Figura 4.5.2**), muestra como esta estructurado el "Nivel de Inversión en Salud" del Perú que es como llamamos a nuestro primer factor, todo esto a nivel departamental, así pues los departamentos con mejor nivel de inversión en salud, en

Cuadro 4.5.1 Mapa de Tendencia de Inversión en Salud
 Nivel Nacional por Departamento - Todas las Instituciones
 Método de Análisis Factorial - CP

Variables	Ponderaciones factoriales estimadas		Comunalidad
	F1	F2	
NRO_ESTA	0.587	0.689	0.820
NRO_MEDI	0.467	-0.809	0.872
OTR_PROF	0.747	-0.517	0.826
TECN_ASI	0.827	-0.163	0.710
AREA_CON	0.838	-0.200	0.742
NUM_CAMA	0.884	-0.270	0.854
NUM_LABO	0.857	-0.096	0.744
N_ESTAGU	0.565	0.673	0.772
N_ESTDES	0.809	0.398	0.812
N_ESTLUZ	0.831	0.337	0.805
N_TELECO	0.838	0.048	0.704
N_CERCOP	0.571	0.530	0.607
N_EQSERF	0.865	-0.214	0.795
N_EQSERI	0.820	0.284	0.752
N_EQAPOY	0.804	-0.486	0.883
Porcentaje acumulado del total de varianza muestral (estandarizada) explicada	58.545	77.99	

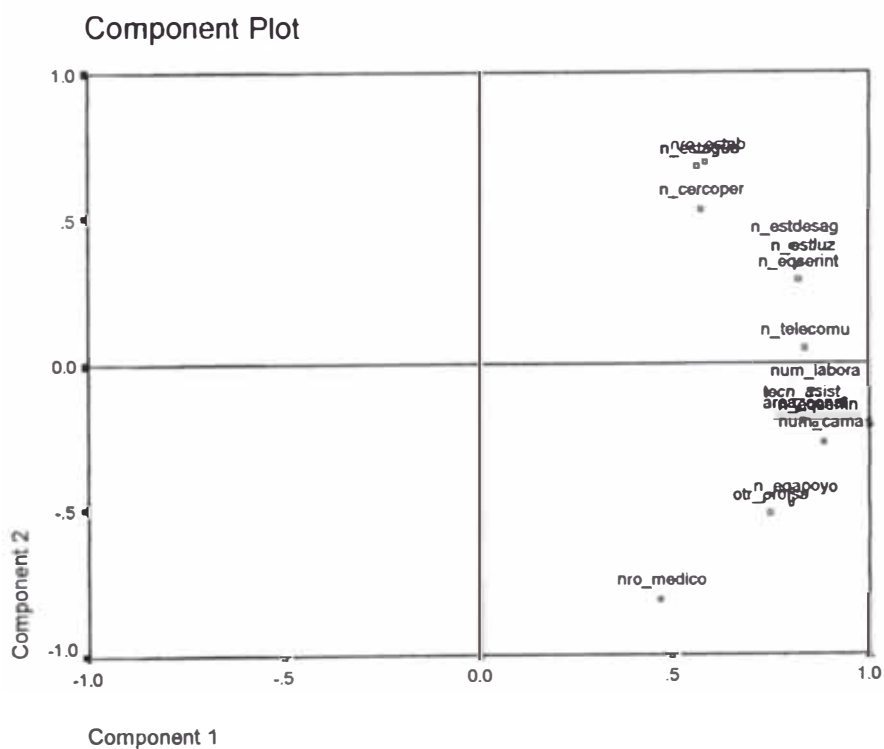
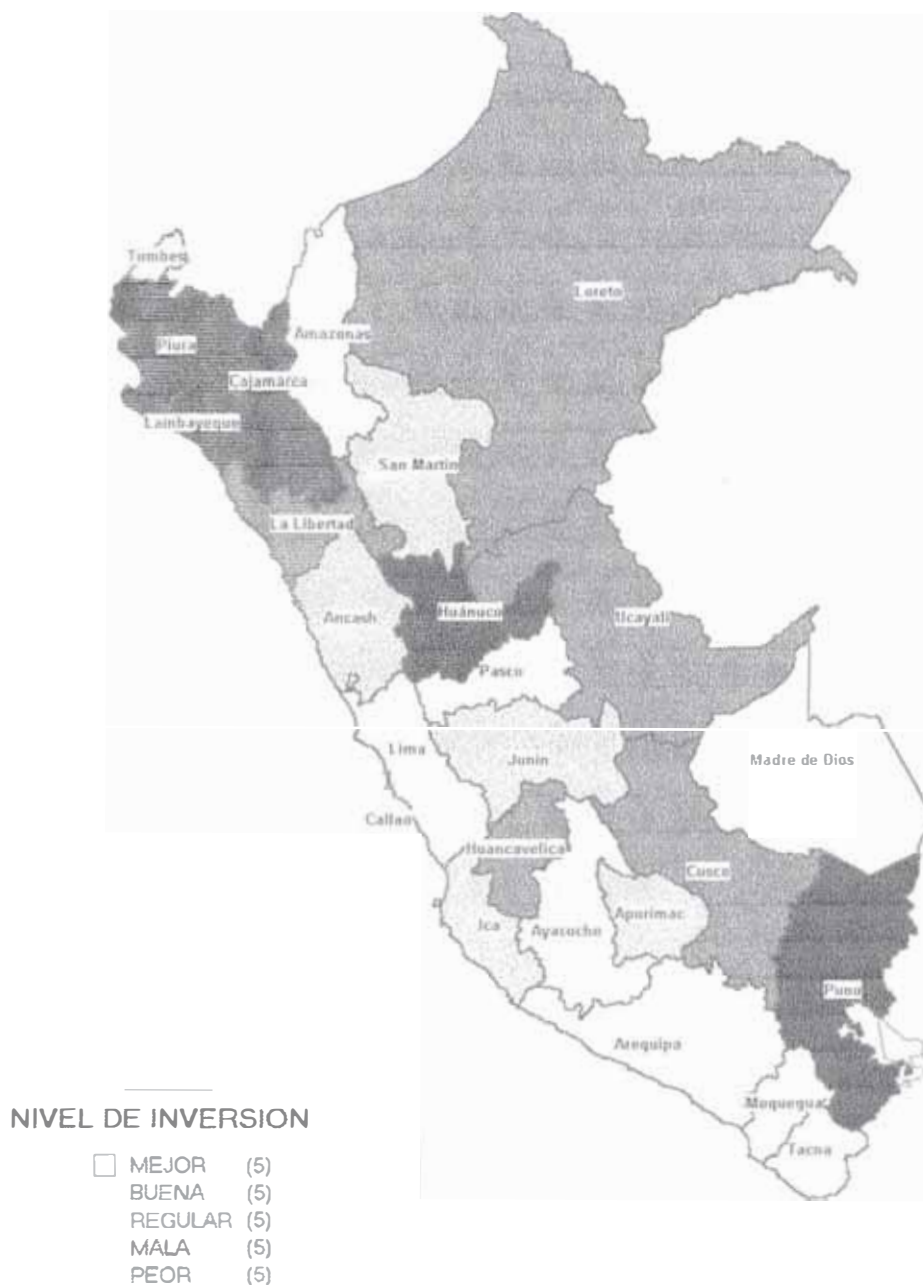


Figura 4.5.1 Ponderaciones de las variables en los dos primeros factores (Mapa 05)

MAPA No. 5
 MAPA DE TENDENCIA DE INVERSION EN SALUD
 PERU : 1996



Elaboración Propia

Mapa_05.wor

Figura 4.5.2 MAPA No. 5 : Nivel Departamental - Todas las Instituciones. Método de Análisis Factorial - CP (Con índices por 10 mil habitantes)

concordancia con el **Cuadro A3** son Madre de Dios, Moquegua, Pasco, Callao y Tacna, el departamento de Lima se encuentra en el segundo estrato, y los departamentos que tiene Peor "Nivel de inversión en Salud son, de acuerdo al índice Cajamarca, Huánuco, Lambayeque, Piura y Puno.

Resultados de Mapa 06

En este mapa a nivel Departamental, desarrollada con el análisis factorial y con el método de Componentes Principales y donde se emplea datos pertenecientes a todas las instituciones que prestan servicios en salud, así tenemos el Ministerio de salud, el IPSS que hoy se denomina EsSalud, Fuerzas armadas y policiales, Clínicas Privadas lucrativas y otros, tiene la variante que a diferencia del mapa anterior se utilizan datos expresados en porcentajes como se indica en el **Cuadro A4**, tiene ordenamiento determinado por el score obtenido con método de regresión. La matriz de correlaciones (**Cuadro B4**) para las quince variables tienen gran amplitud variando desde 0.001 hasta 0.895.

En el **Cuadro 4.6.1** el primer factor tiene un porcentaje de varianza de explicación de 50.9 %, como también en el listado de las ponderaciones del primer factor en las diferentes quince variables se ve que es de alto valor, la mayoría alrededor de 0.8, teniendo en cuenta que se ha factorizado la matriz de correlaciones, es decir que se trabaja con datos estandarizados.

En la **Figura 4.6.1** es de gran importancia pues allí se ve como influyen sus ponderaciones en el Primer factor que denominaremos "Nivel de Inversión", vemos como influye positivamente las quince variables.

El mapa No. 6, (**Figura 4.6.2**) muestra como está estructurado el "Nivel de Inversión en Salud" del Perú que es como llamamos a nuestro primer factor, todo esto a nivel departamental, así pues los departamentos con mejor nivel de inversión en salud, en concordancia con el **Cuadro A4** son Madre de Dios, Moquegua, Lima, Callao y Tacna, y los

Cuadro 4.6.1 Mapa de Tendencia de Inversión en Salud
 Nivel Nacional por Departamento - Todas las Instituciones
 Método de Análisis Factorial - CP

Variables	Ponderaciones factoriales estimadas		Comunalidad
	F1	F2	
NRO_ESTA	0.051	0.930	0.868
NRO_MEDI	0.847	-0.290	0.801
OTR_PROF	0.913	-0.015	0.833
TECN_ASI	0.736	0.527	0.819
AREA_CON	0.787	0.389	0.770
NUM_CAMA	0.853	0.343	0.846
NUM_LABO	0.719	0.571	0.843
N_ESTAGU	0.422	-0.730	0.711
N_ESTDES	0.651	-0.679	0.885
N_ESTLUZ	0.694	-0.582	0.820
N_TELECO	0.785	-0.457	0.824
N_CERCOP	0.455	-0.638	0.613
N_EQSERF	0.843	0.193	0.747
N_EQSERI	0.470	0.781	0.831
N_EQAPOY	0.919	0.096	0.854
Porcentaje acumulado del total de varianza muestral (estandarizada) explicada	50.926	80.441	

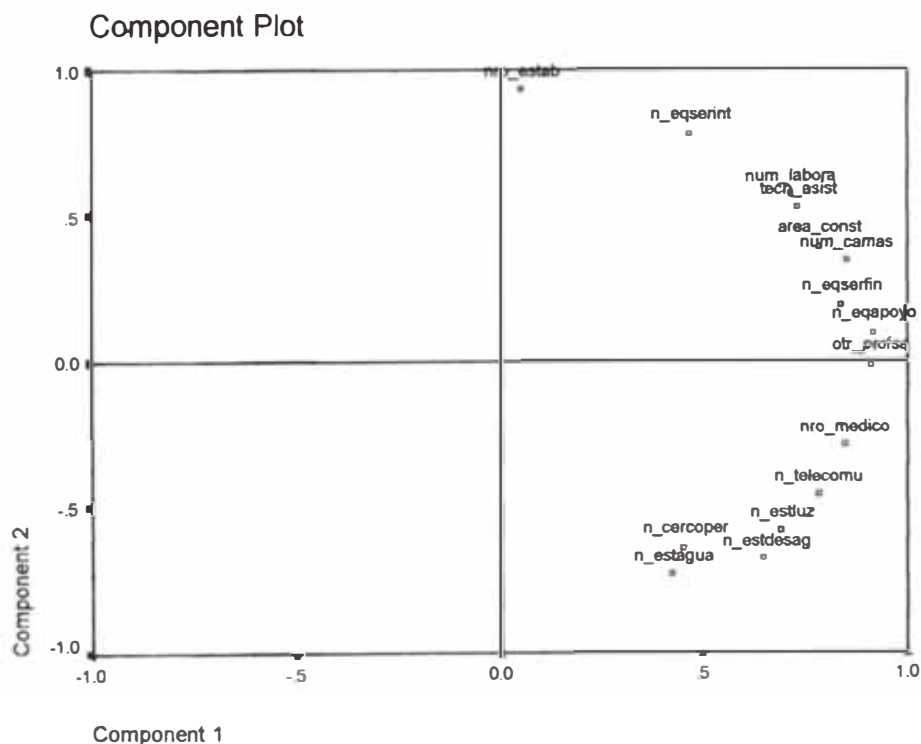
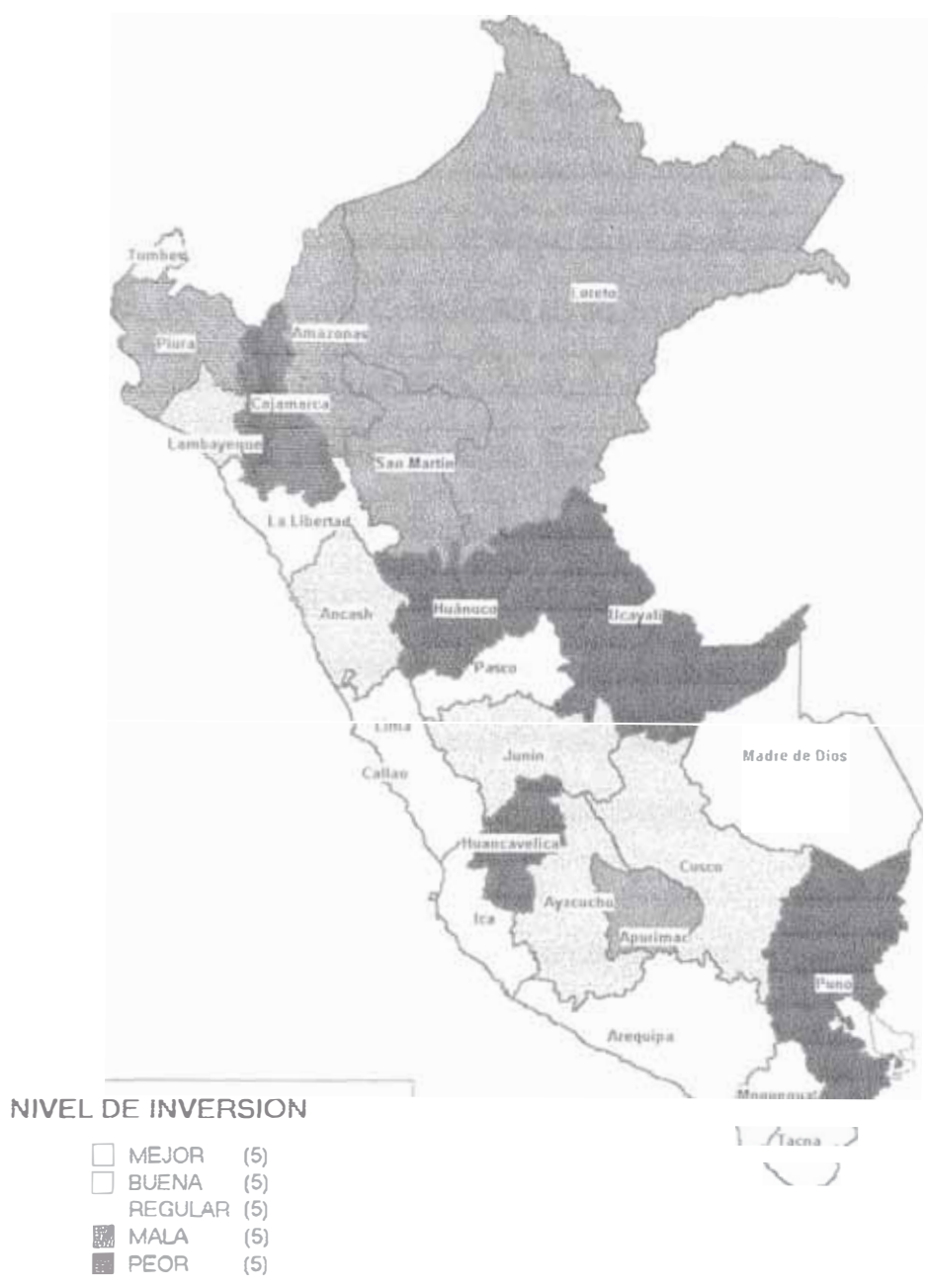


Figura 4.6.1 Ponderaciones de las variables en los dos primeros factores (Mapa 06)

MAPA No. 6
MAPA DE TENDENCIA DE INVERSION EN SALUD
PERU : 1996



Elaboración Propia

Mapa_06 wor

Figura 4.6.2 MAPA No. 6 : Nivel Departamental - Todas las Instituciones. Método de Análisis Factorial - CP (Con índices por 10 mil habitantes y Porcentajes de Establecimientos)

departamentos que tiene Peor "Nivel de inversión en Salud" son, de acuerdo al índice Cajamarca, Huanuco, Huancavelica, Puno, Ucayali.

Resultados de Mapa 07

En este mapa a nivel Departamental, desarrollada con el análisis factorial y con el método de Máxima Verosimilitud, y donde se emplea datos pertenecientes a todas las instituciones que prestan servicios en salud, así tenemos el Ministerio de salud, el IPSS que hoy se denomina EsSalud, Fuerzas armadas y policiales, Clínicas Privadas lucrativas y otros, para los datos de las mismas variables que son expresados en índices por diez mil habitantes como se indica en el **Cuadro A3** (Anexo 1) se ve que los scores dados por este método son similares a los del mapa 05.

La matriz la correlaciones (**Cuadro B3**, Anexo 2) para las quince variables es lo mismo que en el caso correspondiente al Mapa 05.

En el **Cuadro 4.7.1** el primer factor tiene un porcentaje de varianza de explicación de 56.767 %, así también en el listado de las ponderaciones del primer factor en las diferentes quince variables se puede ver que es de alto valor, la mayoría alrededor de 0.8, teniendo en cuenta que se ha factorizado la matriz de correlaciones, es decir se trabajó con datos estandarizados.

En la **Figura 4.7.1** es de gran importancia pues allí se puede ver como influyen las variables en el Primer factor que denominaremos "Nivel de Inversión", se verifica una influencia positiva para las quince variables.

El mapa No. 7 (**Figura 4.7.2**), muestra como está estructurado el "Nivel de Inversión en Salud" del Perú que es como llamamos a nuestro primer factor, todo esto a nivel departamental, así pues los departamentos con mejor nivel de inversión en salud son Madre de Dios, Moquegua, Pasco, Callao y Tacna, y los departamentos que tiene Peor "Nivel de inversión en Salud" son, de acuerdo al índice Cajamarca, Huánuco, Lambayeque, Piura y

Cuadro 4.7.1 Mapa de Tendencia de Inversión en Salud
 Nivel Nacional por Departamento - Todas las Instituciones
 Método de Análisis Factorial - MV

Ponderaciones factoriales
estimadas

Variables	F1	F2	Comunalidad
NRO_ESTA	0.637	-0.740	0.952
NRO_MEDI	0.409	0.717	0.681
OTR_PROF	0.704	0.549	0.796
TECN_ASI	0.822	0.013	0.675
AREA_CON	0.817	0.162	0.694
NUM_CAMA	0.857	0.279	0.813
NUM_LABO	0.857	0.021	0.734
N_ESTAGU	0.562	-0.493	0.559
N_ESTDES	0.796	-0.263	0.702
N_ESTLUZ	0.835	-0.313	0.795
N_TELECO	0.822	0.172	0.705
N_CERCOP	0.549	-0.285	0.383
N_EQSERF	0.832	0.389	0.844
N_EQSERI	0.845	-0.286	0.795
N_EQAPOY	0.779	0.572	0.934

Porcentaje acumulado del total de varianza muestral (estandarizada) explicada

	56.767	73.759
--	--------	--------

Factor Plot

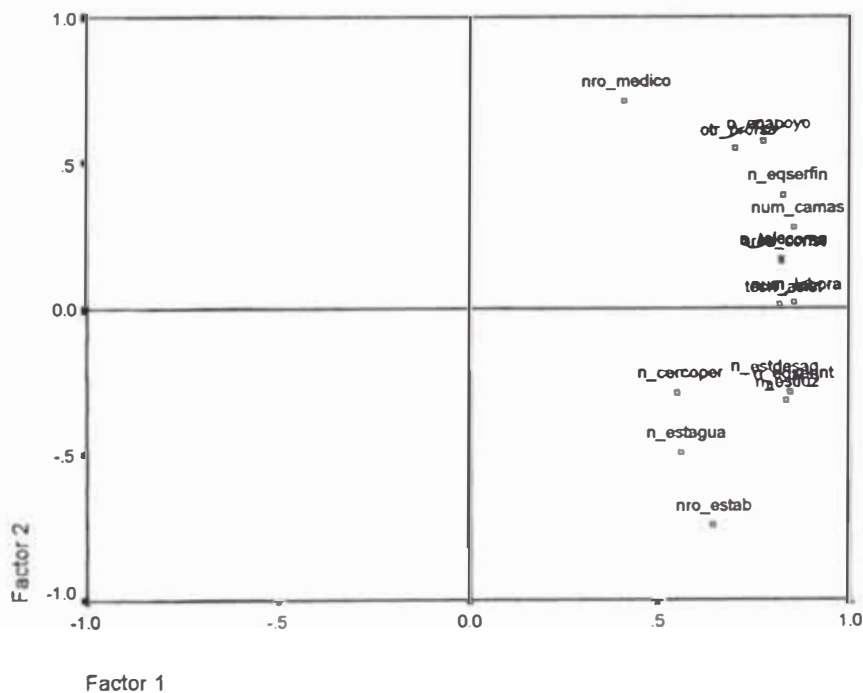
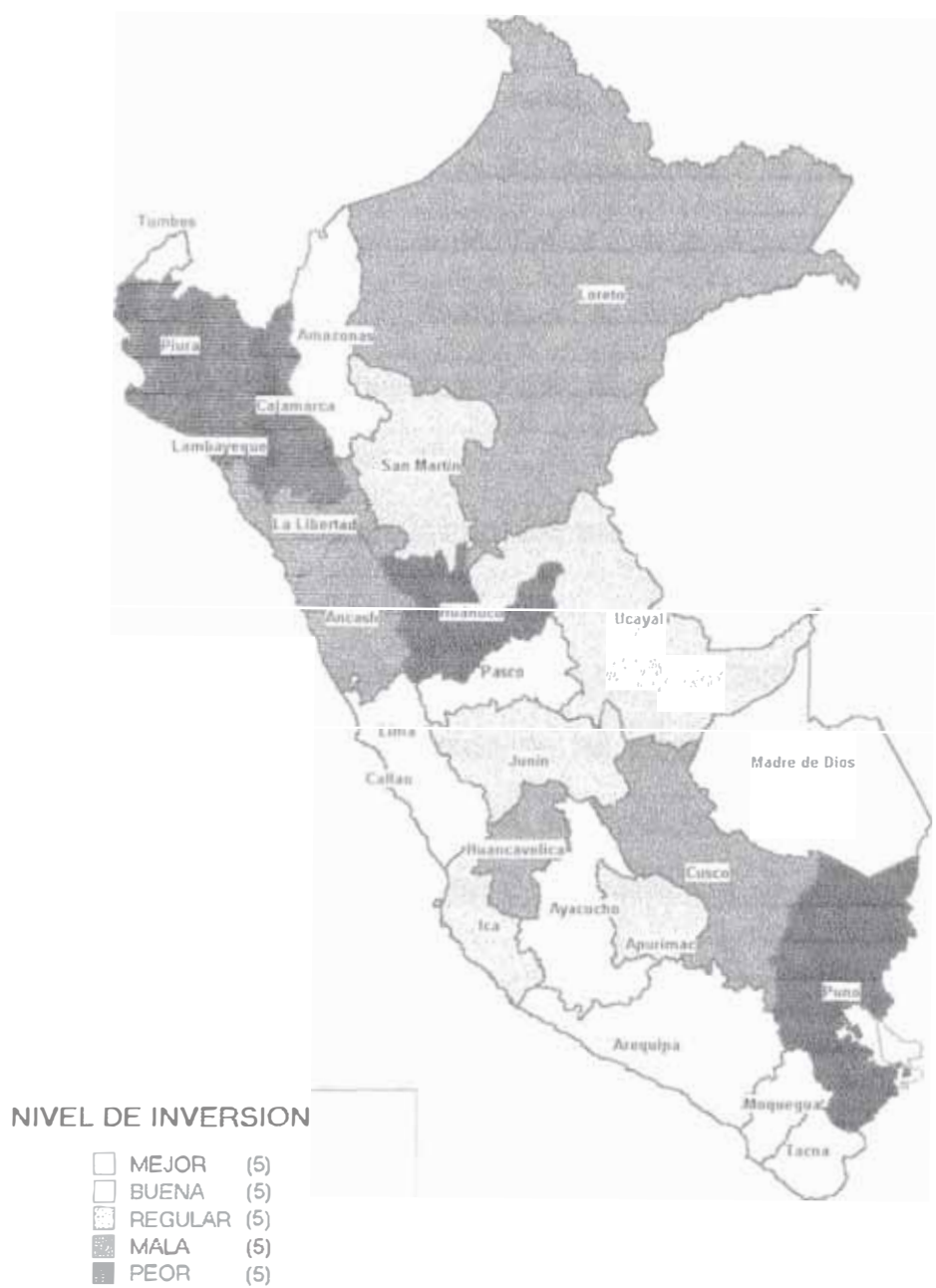


Figura 4.7.1 Ponderaciones de las variables en los dos primeros factores(Mapa 07)

MAPA No. 7
 MAPA DE TENDENCIA DE INVERSION EN SALUD
 PERU : 1996



Elaboración Propia

Mapa_07 wor

Figura 4.7.2 MAPA No. 7 : Nivel Departamental - Todas las Instituciones. Método de Análisis Factorial - MV (Con índices por 10 mil habitantes)

Puno.

Resultados de Mapa 08

En este mapa a nivel Departamental, desarrollada con el análisis factorial con el método de Máxima Verosimilitud, y donde se emplea datos pertenecientes a todas las instituciones que prestan servicios en salud, así tenemos el Ministerio de salud, el IPSS que hoy se denomina EsSalud, Fuerzas armadas y policiales, Clínicas Privadas Lucrativas y otros, tiene la variante que a diferencia del mapa anterior se utilizan datos expresados en porcentajes como se indica en el **Cuadro A4**, el cual tiene ordenamiento determinado por el score titulado "Score 1 Factor CP" del mapa 06, obtenido con método de regresión, estos varían desde 2.179, que es el mayor valor del índice que es denominado "Nivel de Inversión de Salud", hasta -1.265, el menor valor.

La matriz de correlaciones (**Cuadro B4**) para las quince variables es lo mismo que en el caso correspondiente al Mapa 06.

En el **Cuadro 4.8.1** el primer factor tiene un porcentaje de varianza de explicación de 49.481 %, como también en el listado de las ponderaciones del primer factor en las diferentes quince variables se ve que es de alto valor, la mayoría alrededor de 0.9, teniendo en cuenta que se ha factorizado la matriz de correlaciones, es decir que se ha trabajado con datos estandarizados.

En la **Figura 4.8.1** es de gran importancia pues allí se ve como influyen sus ponderaciones en el Primer factor que denominamos "Nivel de Inversión".

El mapa No. 8 (**Figura 4.8.2**), muestra como está estructurado el "Nivel de Inversión en Salud" del Perú que es como llamamos a nuestro primer factor, todo esto a nivel departamental, así pues los departamentos con mejor nivel de inversión en salud son Callao, Moquegua, Madre de Dios, Lima y Tacna, y los departamentos que tienen peor "Nivel de inversión en Salud" son, de acuerdo al índice Cajamarca, Huánuco, Huancavelica, Puno y

Cuadro 4.8.1 Mapa de Tendencia de Inversión en Salud Nivel Nacional por Departamento - Todas las Instituciones
Método de Análisis Factorial - MV

Variables	Ponderaciones factoriales estimadas		Comunalidad
	F1	F2	
NRO_ESTA	0.073	0.924	0.859
NRO_MEDI	0.836	-0.329	0.807
OTR_PROF	0.897	-0.028	0.806
TECN_ASI	0.749	0.480	0.791
AREA_CON	0.780	0.341	0.724
NUM_CAMA	0.848	0.300	0.808
NUM_LABO	0.735	0.539	0.831
N_ESTAGU	0.388	-0.687	0.623
N_ESTDES	0.631	-0.685	0.867
N_ESTLUZ	0.674	-0.599	0.813
N_TELECO	0.761	-0.478	0.807
N_CERCOP	0.411	-0.586	0.512
N_EQSERF	0.823	0.186	0.711
N_EQSERI	0.488	0.776	0.840
N_EQAPOY	0.907	0.063	0.826
Porcentaje acumulado del total de varianza muestral (estandarizada) explicada	49.481	77.499	

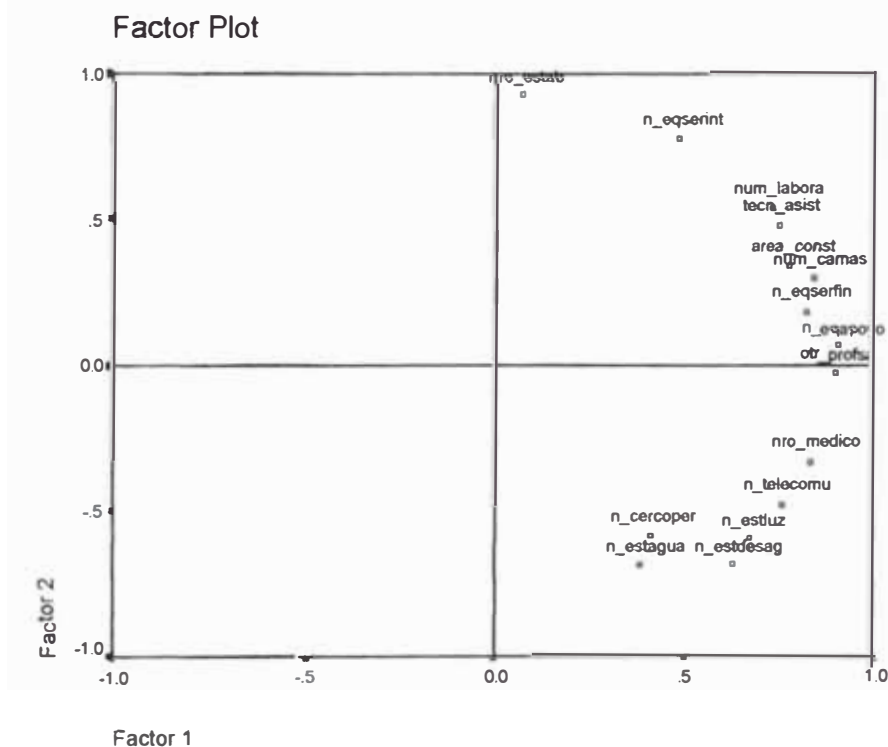
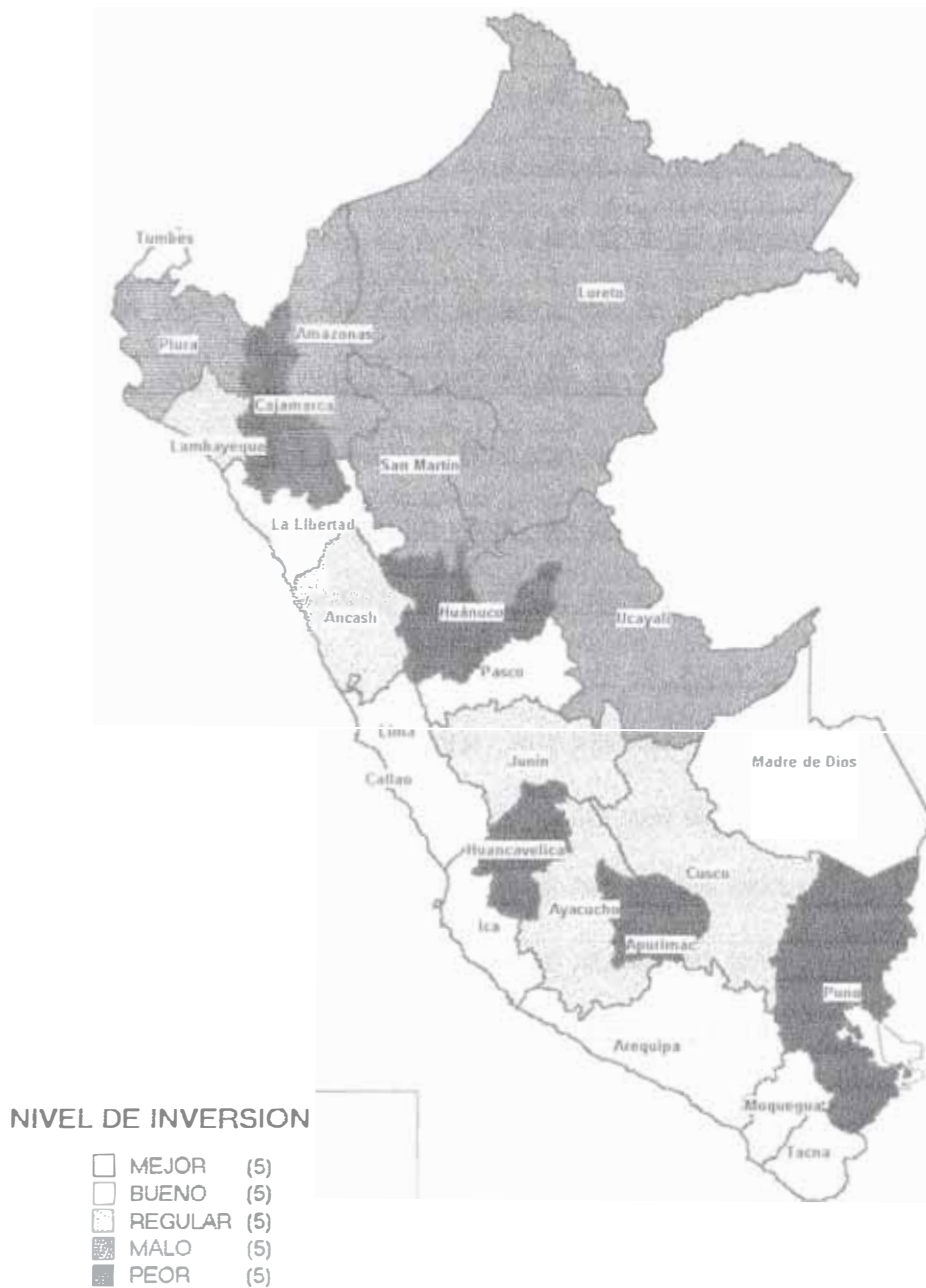


Figura 4.8.1 Ponderaciones de las variables en los dos primeros factores (Mapa 08)

MAPA No. 8
 MAPA DE TENDENCIA DE INVERSION EN SALUD
 PERU : 1996



Elaboración Propia

Mapa_08 wor

Figura 4.8.2 MAPA No. 8 : Nivel Departamental - Todas las Instituciones. Método de Análisis Factorial - MV (Con índices por 10 mil habitantes y Porcentajes de Establecimientos)

Apurímac.

Resultados de Mapa 09

En este mapa a nivel Departamental, desarrollada con el análisis factorial y con el método de Componentes Principales, y donde se emplea datos pertenecientes solo al Ministerio de salud, para el diseño se empleo datos expresados todos como una tasa por diez mil habitantes, vemos que, el **Cuadro A5** tiene ordenamiento determinado por el score obtenido con método de regresión, que es el método que se utiliza para calcular los scores, y su variación esta entre 3.489 y -1.435. La matriz la correlaciones (**Cuadro B5**) para las quince variables tienen gran amplitud variando desde 0.039 negativo hasta 0.912.

En el **Cuadro 4.9.1** el primer factor tiene un porcentaje de varianza de explicación de 57.9 %, como también en el listado de las ponderaciones del primer factor en las diferentes quince variables se ve que es de alto valor, la mayoría alrededor de 0.8, teniendo en cuenta que se ha factorizado la matriz de correlaciones, es decir se ha trabajado con datos estandarizados.

En la **Figura 4.9.1** es de gran importancia pues allí se ve como influyen las ponderaciones de las variables en el Primer factor que denominaremos "Nivel de Inversión".

El mapa No. 9 (**Figura 4.9.2**), muestra como esta estructurado el "Nivel de Inversión en Salud" del Perú que es como llamamos a nuestro primer factor, todo esto a nivel departamental, así pues los departamentos con mejor nivel de inversión en salud, son Madre de Dios, Moquegua, Pasco, Ayacucho y Tacna, y los departamentos que tiene Peor "Nivel de inversión en Salud son, de acuerdo al índice Lambayeque, Piura, Cajamarca, La Libertad y Huanuco.

Cuadro 4.9.1 Mapa de Tendencia de Inversión en Salud Nivel Nacional por Departamento - Ministerio de Salud
Método de Análisis Factorial - CP

Variables	Ponderaciones factoriales estimadas		Comunalidad
	F1	F2	
NRO_ESTA	0.826	-0.390	0.835
NRO_MEDI	0.352	0.811	0.781
OTR_PROF	0.670	0.562	0.765
TECN_ASI	0.903	0.076	0.822
AREA_CON	0.806	-0.185	0.684
NUM_CAMA	0.879	-0.018	0.774
NUM_LABO	0.788	0.251	0.683
N_ESTAGU	0.659	-0.611	0.808
N_ESTDES	0.776	-0.346	0.722
N_ESTLUZ	0.914	0.000	0.836
N_TELECO	0.555	0.078	0.314
N_CERCOP	0.518	-0.521	0.540
N_EQSERF	0.877	0.105	0.780
N_EQSERI	0.887	0.019	0.788
N_EQAPOY	0.753	0.462	0.780
Porcentaje acumulado del total de varianza muestral (estandarizada) explicada	57.906	72.736	

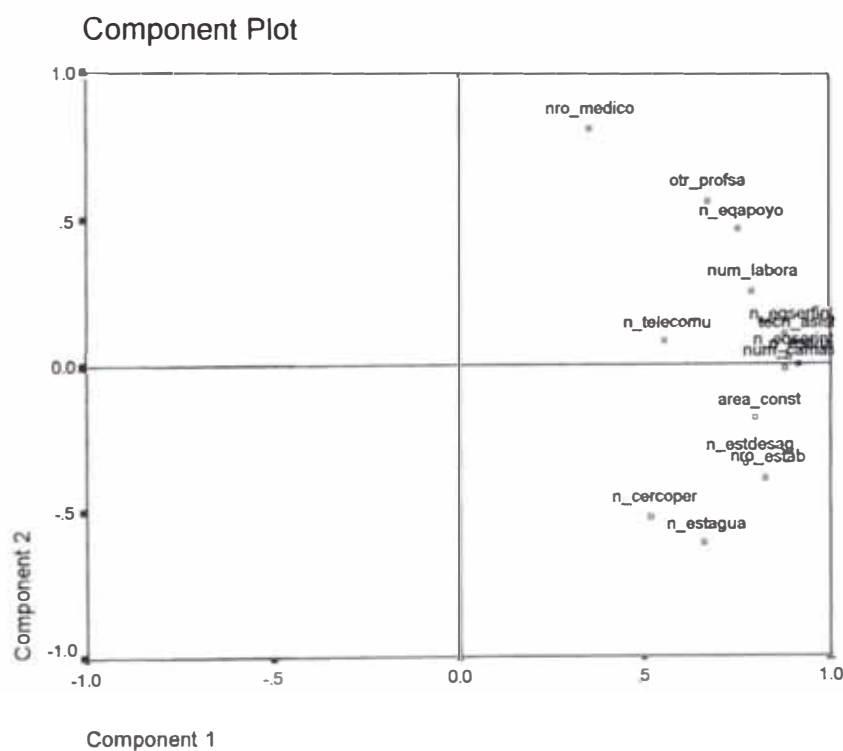


Figura 4.9.1 Ponderaciones de las variables en los dos primeros factores (Mapa 09)

MAPA No.9
MAPA DE TENDENCIA DE INVERSION EN SALUD
PERU : 1996



Elaboración Propia

Mapa_09.wor

Figura 4.9.2 MAPA No.9 : Nivel Departamental - Ministerio de Salud. Método de Análisis Factorial - CP (Con índices por 10 mil habitantes)

Resultados de Mapa 10

En este mapa a nivel Departamental, desarrollada con el análisis factorial y con el método de Componentes Principales y donde se emplea datos pertenecientes solo al Ministerio de salud, tiene la variante que a diferencia del mapa anterior se utilizan datos expresados en porcentajes como se indica en el **Cuadro A6**, tiene ordenamiento determinado por el score obtenido con método de regresión. La matriz la correlaciones (**Cuadro B6**) para las quince variables tienen gran amplitud variando desde 0.003 negativo hasta 0.912.

En el **Cuadro 4.10.1** el primer factor tiene un porcentaje de varianza de explicación de 45.7 %, como también en el listado de las ponderaciones del primer factor en las diferentes quince variables se ve que es de diversos valores, sin embargo, la mayoría alrededor de 0.8, teniendo en cuenta que se ha factorizado la matriz de correlaciones, es decir se trabajo con datos estandarizados.

En la **Figura 4.10.1** es de gran importancia pues allí se ve que el "Nivel de Inversión" esta obviamente repartido entre las dos primeros factores.

El mapa No. 10 (**Figura 4.10.2**), muestra como esta estructurado el "Primer factor del modelo", todo esto a nivel departamental. En este caso, por lo mencionado en el parrafo anterior no se puede afirmar nada del primer factor el cual solo tendria poco porcentaje del "Nivel de inversión" en salud.

Cuadro 4.10.1 Mapa de Tendencia de Inversión en Salud Nivel Nacional por Departamento - Ministerio de Salud
Método de Análisis Factorial - CP

Variables	Ponderaciones factoriales estimadas		Comunalidad
	F1	F2	
NRO_ESTA	0.896	-0.293	0.888
NRO_MEDI	0.248	0.840	0.767
OTR_PROF	0.535	0.716	0.799
TECN_ASI	0.958	0.084	0.926
AREA_CON	0.823	0.011	0.677
NUM_CAMA	0.891	0.093	0.802
NUM_LABO	0.868	0.113	0.767
N_ESTAGU	-0.483	0.654	0.661
N_ESTDES	-0.446	0.787	0.818
N_ESTLUZ	-0.287	0.824	0.761
N_TELECO	-0.171	0.809	0.684
N_CERCOP	-0.429	0.673	0.637
N_EQSERF	0.765	0.382	0.732
N_EQSERI	0.921	0.043	0.849
N_EQAPOY	0.635	0.626	0.795
Porcentaje acumulado del total de varianza muestral (estandarizada) explicada	45.715	77.084	

Component Plot

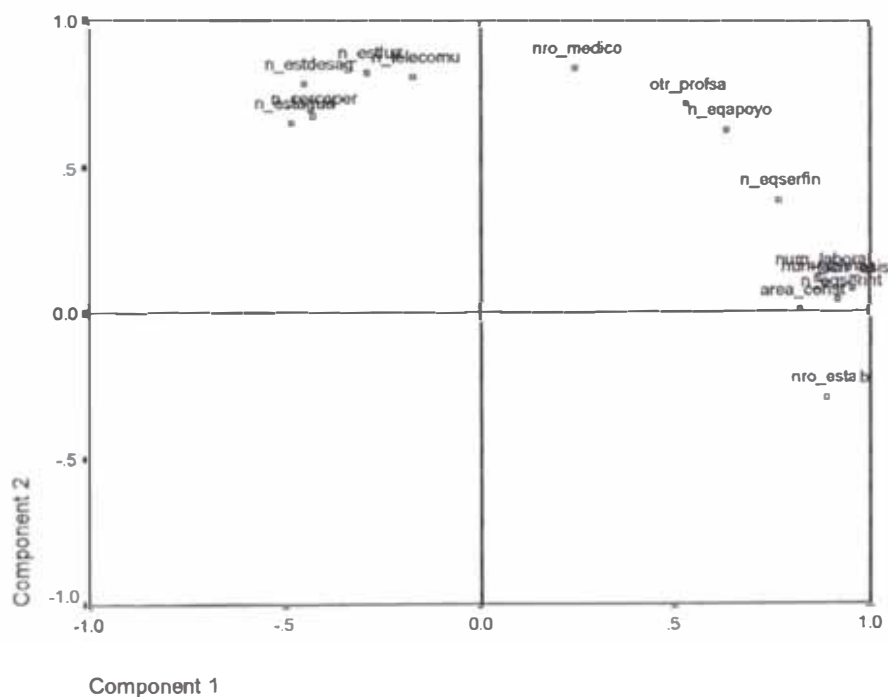
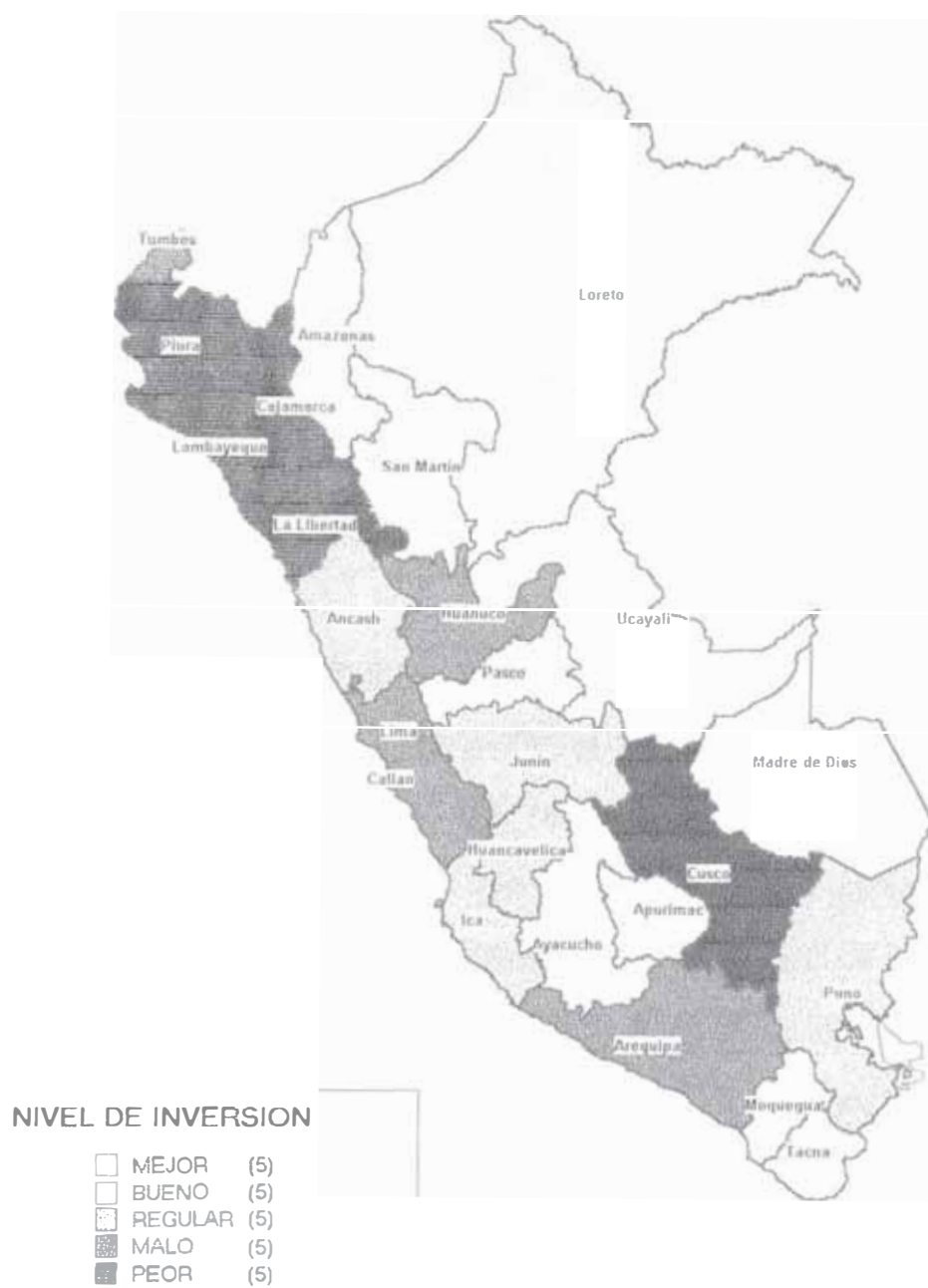


Figura 4.10.1 Ponderaciones de las variables en los dos primeros factores (Mapa 10)

MAPA No.10
 MAPA DE TENDENCIA DE INVERSION EN SALUD
 PERU : 1996



Elaboración Propia

Mapa_10.wor

Figura 4.10.2 MAPA No.10 : Nivel Departamental - Ministerio de Salud. Método de Análisis Factorial - CP (Con índices por 10 mil habitantes y Porcentajes de Establecimientos)

Resultados de Mapa 11

En este mapa a nivel Departamental, desarrollada con el análisis factorial y con el método de Máxima Verosimilitud para los datos referentes solo al Ministerio de Salud de las mismas variables que son expresados en índices por diez mil habitantes como se indica en el **Cuadro A5**, los score respectivos varían desde 3.396, que es el mayor valor del índice que es denominado "Nivel de Inversión de Salud", hasta -1.041, el menor valor.

La matriz de correlaciones (**Cuadro B5**) para las quince variables es lo mismo que en el caso correspondiente al Mapa 09.

En el **Cuadro 4.11.1** primer factor tiene un porcentaje de varianza de explicación de 43.727 %, así también en el listado de las ponderaciones del primer factor en las diferentes quince variables se puede ver que es de alto valor, la mayoría alrededor de 0.7, teniendo en cuenta que se ha factorizado la matriz de correlaciones, es decir se trabajó con datos estandarizados.

La **Figura 4.11.1** es de gran importancia pues allí se puede ver como influyen las variables en el Primer factor que denominaremos "Nivel de Inversión".

El mapa No. 11 (**Figura 4.11.2**), muestra como está estructurado el "Nivel de Inversión en Salud" del Perú que es como llamamos a nuestro primer factor, todo esto a nivel departamental, así pues los departamentos con mejor nivel de inversión en salud son Madre de Dios, Pasco, Amazonas, Huancavelica y Ayacucho, y los departamentos que tienen peor "Nivel de inversión en Salud" son, de acuerdo al índice Callao, Lima, Lambayeque, La Libertad, Ica.

Cuadro 4.11.1 Mapa de Tendencia de Inversión en Salud
 Nivel Nacional por Departamento - Ministerio de Salud
 Método de Análisis Factorial - MV

Ponderaciones factoriales
 estimadas

Variables	F1	F2	Comunalidad
NRO_ESTA	0.999	-0.014	0.999
NRO_MEDI	-0.028	0.760	0.579
OTR_PROF	0.246	0.875	0.826
TECN_ASI	0.864	0.380	0.891
AREA_CON	0.784	0.238	0.671
NUM_CAMA	0.767	0.401	0.749
NUM_LABO	0.709	0.414	0.674
N_ESTAGU	0.696	-0.011	0.484
N_ESTDES	0.722	0.203	0.563
N_ESTLUZ	0.803	0.430	0.830
N_TELECO	0.219	0.552	0.353
N_CERCOP	0.454	0.092	0.214
N_EQSERF	0.588	0.633	0.745
N_EQSERI	0.819	0.372	0.809
N_EQAPOY	0.354	0.846	0.841
Porcentaje acumulado del total de varianza muestral (estandarizada) explicada	43.727	68.199	

Factor Plot

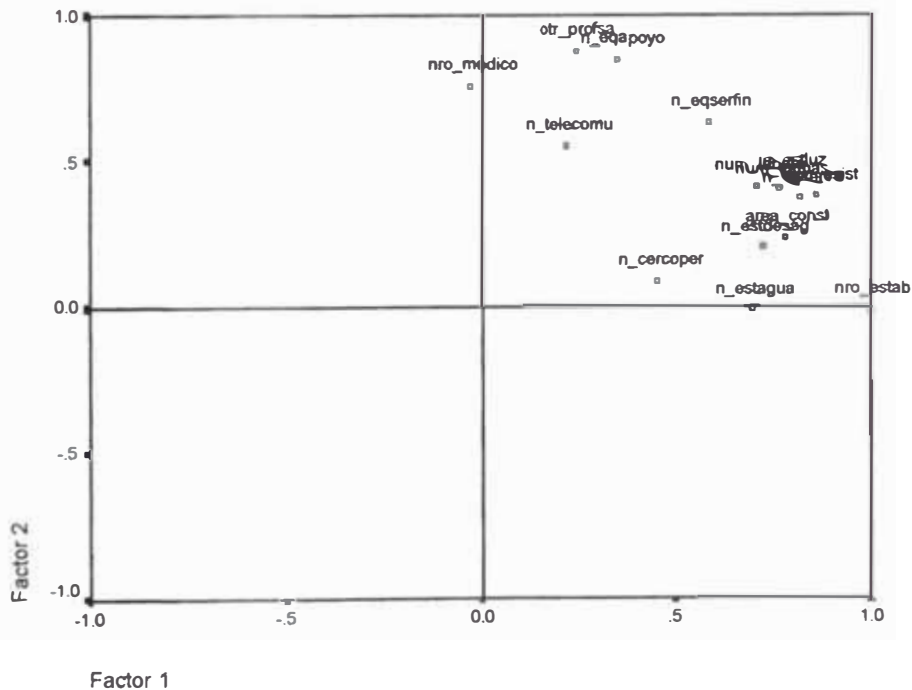
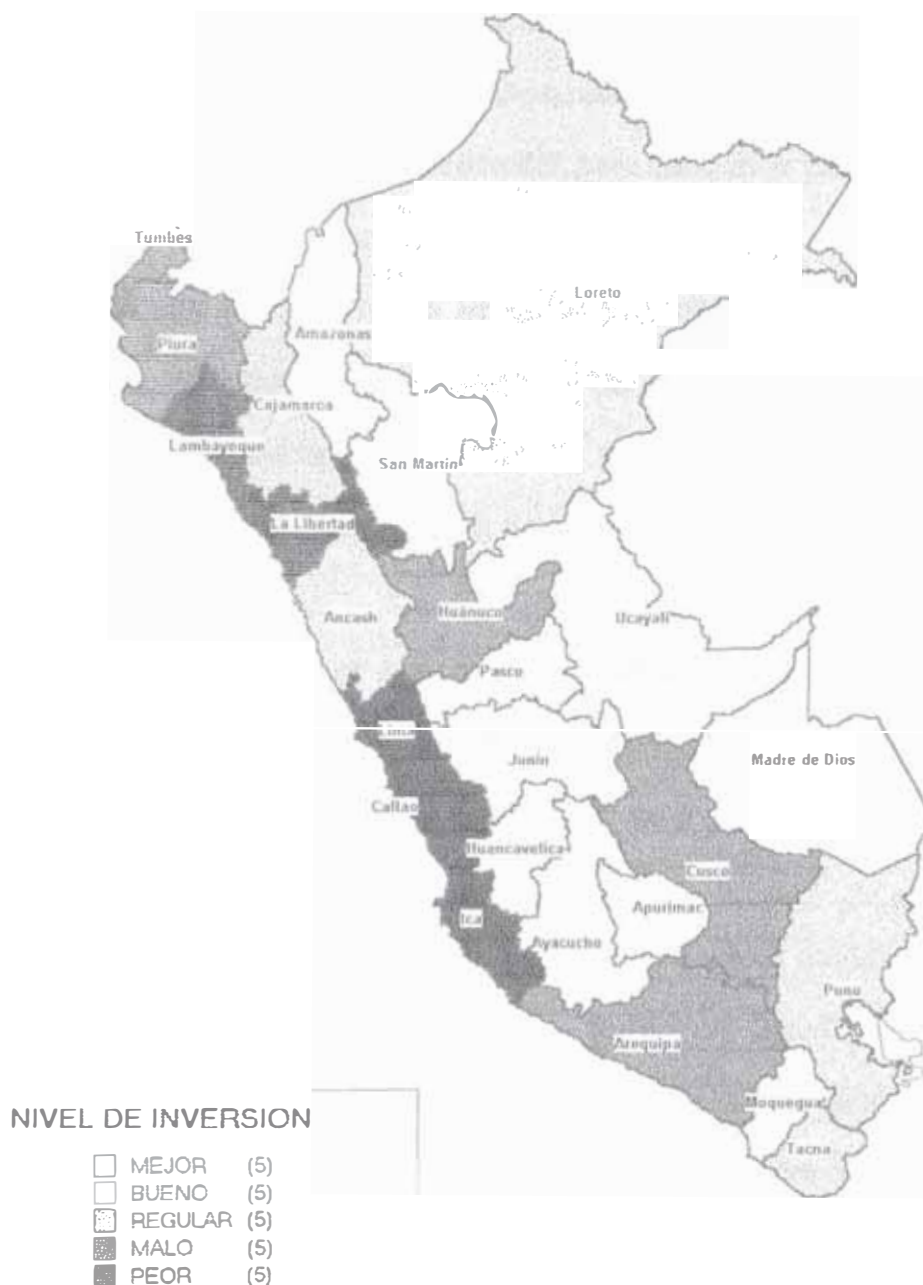


Figura 4.11.1 Ponderaciones de las variables en los dos primeros factores (Mapa 11)

MAPA No. 11
 MAPA DE TENDENCIA DE INVERSION EN SALUD
 PERU : 1996



Elaboración Propia

Mapa_11.wor

Figura 4.11.2 MAPA No. 11 : Nivel Departamental - Ministerio de Salud. Método de Análisis Factorial - MV (Con índices por 10 mil habitante)

Resultados de Mapa 12

En este mapa a nivel Departamental, desarrollada con el análisis factorial y con el método de Máxima Verosimilitud , solo con datos del Ministerio de Salud, tiene la variante que a diferencia del mapa anterior se utilizan datos expresados en porcentajes como se indica en el **Cuadro A6**, los scores correspondientes al método de MV varían desde 4.021, que es el mayor valor del índice, hasta -1.190, el menor valor.

La matriz la correlaciones (**Cuadro B6**) para las quince variables es lo mismo que en el caso correspondiente al mapa 10.

En el **Cuadro 4.12.1** el primer factor tiene un porcentaje de varianza de explicación de 44.057 %, como también en el listado de las ponderaciones del primer factor en las diferentes quince variables se ve que es de alto valor, la mayoría alrededor de 0.8, teniendo en cuenta que se ha factorizado la matriz de correlaciones, es decir que se ha trabajado con datos estandarizados.

En la **Figura 4.12.1** es de gran importancia pues allí se ve como influyen sus ponderaciones en el Primer factor que denominamos con precaución "Nivel de Inversión", vemos que las diez variables influye positivamente y cinco influyen negativamente algo contradictorio.

El mapa No. 12 (**Figura 4.12.2**), muestra como esta estructurado el primer factor denominado con precaución el "Nivel de Inversión en Salud" del Perú que es como llamamos a nuestro primer factor, todo esto a nivel departamental, así pues los departamentos con mejor nivel de inversión en salud son Madre de Dios, Pasco, Ayacucho, Moquegua, Tacna, y los departamentos que tiene Peor "Nivel de inversión en Salud" son, de acuerdo al índice Lambayeque, La Libertad, Cajamarca, Cusco, Piura.

Cuadro 4.12.1 Mapa de Tendencia de Inversión en Salud
Nivel Nacional por Departamento - Ministerio de Salud
Método de Análisis Factorial - MV

Variables	Ponderaciones factoriales estimadas		Comunalidad
	F1	F2	
NRO_ESTA	0.900	-0.316	0.910
NRO_MEDI	0.278	0.855	0.808
OTR_PROF	0.511	0.678	0.721
TECN_ASI	0.975	0.084	0.957
AREA_CON	0.811	-0.013	0.657
NUM_CAMA	0.855	0.074	0.737
NUM_LABO	0.875	0.133	0.783
N_ESTAGU	-0.469	0.606	0.587
N_ESTDES	-0.397	0.778	0.764
N_ESTLUZ	-0.272	0.834	0.770
N_TELECO	-0.155	0.762	0.605
N_CERCOP	-0.411	0.608	0.539
N_EQSERF	0.715	0.310	0.608
N_EQSERI	0.915	0.020	0.837
N_EQAPOY	0.603	0.577	0.696
Porcentaje acumulado del total de varianza muestral (estandarizada) explicada	44.057	73.195	

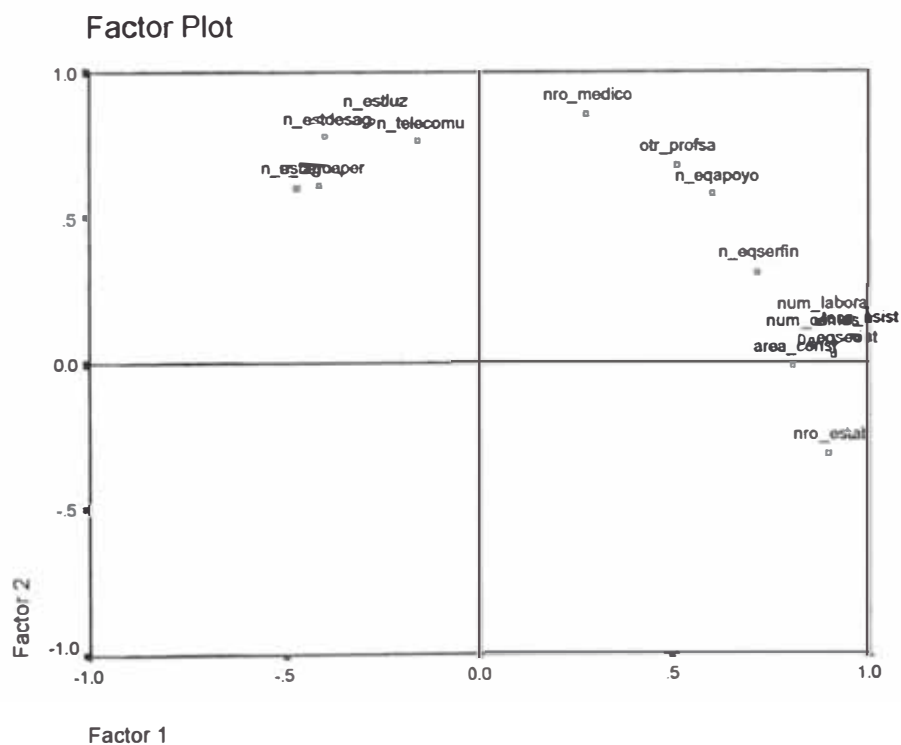


Figura 4.12.1 Ponderaciones de las variables en los dos primeros factores (Mapa 12)

MAPA No.12
 MAPA DE TENDENCIA DE INVERSION EN SALUD
 PERU : 1996



Elaboración Propia

Mapa_12.wor

Figura 4.12.2 MAPA No.12 : Nivel Departamental - Ministerio de Salud. Método de Análisis Factorial - MV (Con índices por 10 mil habitantes y Porcentajes de Establecimientos)

Resultados de Mapa 13

En este mapa a nivel Provincial, desarrollada con el análisis factorial y con el método de Componentes Principales, el diseño ha empleado datos expresados todos como una tasa por diez mil habitantes, y donde se emplea datos pertenecientes a todas las instituciones que prestan servicios en salud, así tenemos el Ministerio de salud, el IPSS que hoy se denomina EsSalud, Fuerzas armadas y policiales, Clínicas Privadas lucrativas y otros, vemos que, el **Cuadro A7** tiene ordenamiento determinado por el score obtenido con método de regresión. La matriz la correlaciones (**Cuadro B7**) para las quince variables tienen gran amplitud variando desde 0.04 negativo hasta 0.784.

En el **Cuadro 4.13.1** el primer factor tiene un porcentaje de varianza de explicación de 50.59 %, como también en el listado de las ponderaciones del primer factor en las diferentes quince variables se ve que es de alto valor, la mayoría alrededor de 0.7, teniendo en cuenta que se ha factorizado la matriz de correlaciones, es decir se ha trabajado con datos estandarizados.

En la **Figura 4.13.1** es de gran importancia pues allí se ve como influyen sus ponderaciones en el Primer factor que denominaremos "Nivel de Inversión", se observas como influye positivamente las quince variables.

El mapa No. 13 (**Figura 4.13.2**), muestra como esta estructurado el "Nivel de Inversión en Salud" del Perú que es como llamamos a nuestro primer factor, todo esto a nivel Provincial, así pues las provincias con mejor nivel de inversión en salud, en concordancia con el **Cuadro A7** son , Tahuamanu, Jorge Basadre, Gral Sanchez Cerro, Paucar del Sara Sara, Tambopata entre otras y las Provincias que tiene Peor "Nivel de inversión en Salud son, de acuerdo al índice Julcan, Pachitea, Chumbivilcas, Lauricocha, Yarowilca entre otras.

Cuadro 4.13.1 Mapa de Tendencia de Inversión en Salud
 Nivel Nacional por Provincia - Todas las Instituciones
 Método de Análisis Factorial - CP

Variables	Ponderaciones factoriales estimadas		Comunalidad
	F1	F2	
NRO_ESTA	0.638	0.568	0.730
NRO_MEDI	0.581	-0.642	0.749
OTR_PROF	0.759	-0.415	0.748
TECN_ASI	0.816	-0.024	0.666
AREA_CON	0.774	-0.119	0.613
NUM_CAMA	0.743	-0.391	0.705
NUM_LABO	0.667	-0.335	0.557
N_ESTAGU	0.679	0.620	0.845
N_ESTDES	0.632	0.448	0.601
N_ESTLUZ	0.766	0.332	0.697
N_TELECO	0.628	0.001	0.394
N_CERCOP	0.479	0.568	0.552
N_EQSERF	0.886	-0.190	0.821
N_EQSERI	0.820	0.228	0.724
N_EQAPOY	0.692	-0.433	0.666
Porcentaje acumulado del total de varianza muestral (estandarizada) explicada	50.599	67.132	

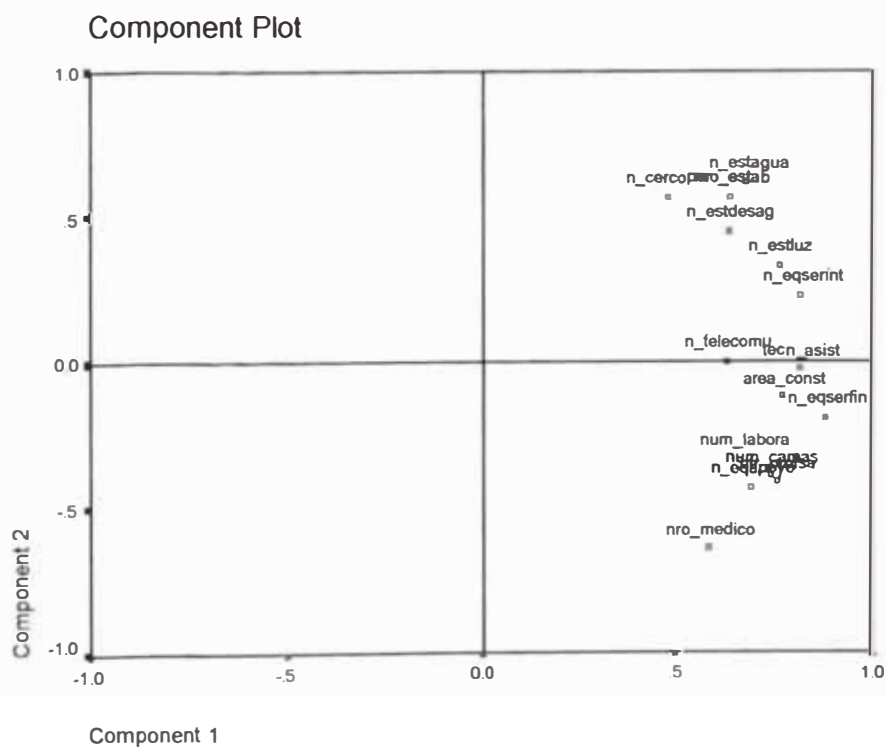
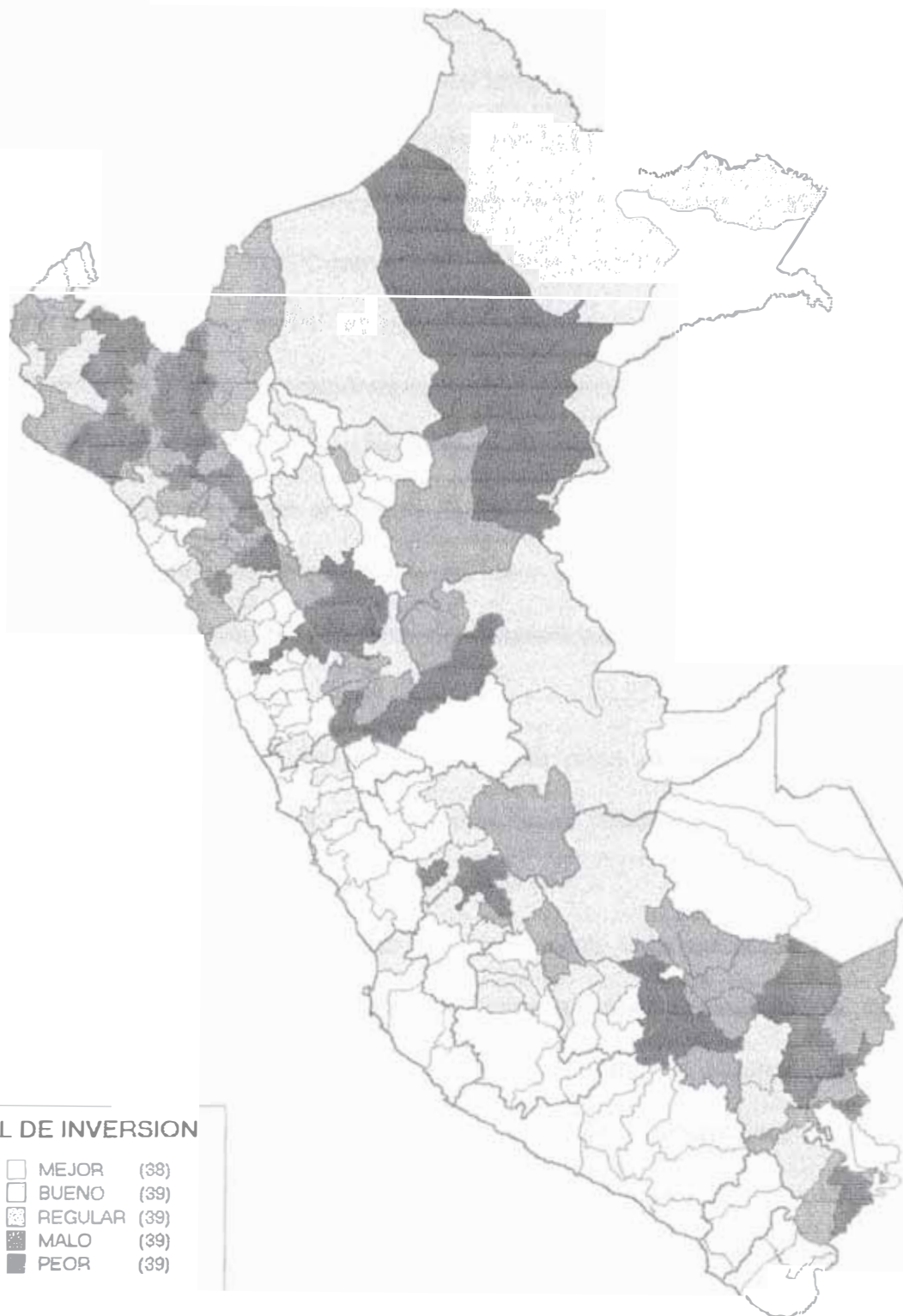


Figura 4.13.1 Ponderaciones de las variables en los dos primeros factores(Mapa 13)

MAPA No. 13
 MAPA DE TENDENCIA DE INVERSION EN SALUD
 PERU : 1996



Elaboración Propia

Mapa_13.wor

Figura 4.13.2 MAPA No. 13 : Nivel Provincial - Todas las Instituciones. Método de Análisis Factorial - CP (Con índices por 10 mil habitantes)

Resultados de Mapa 14

En este mapa a nivel Provincial, desarrollada con el análisis factorial y con el método de Componentes Principales y donde se emplea datos pertenecientes a todas las instituciones que prestan servicios en salud, así tenemos el Ministerio de salud, el IPSS que hoy se denomina EsSalud, Fuerzas armadas y policiales, Clínicas Privadas lucrativas y otros, tiene la variante que a diferencia del mapa anterior se utilizan datos expresados en porcentajes como se indica en el **Cuadro A8**, tiene ordenamiento determinado por el score obtenido con método de regresión. La matriz la correlaciones (**Cuadro B8**) para las quince variables tienen gran amplitud variando desde 0.008 negativo hasta 0.736.

En el **Cuadro 4.14.1** el primer factor tiene un porcentaje de varianza de explicación de 42.533 %, como también en el listado de las ponderaciones del primer factor en las diferentes quince variables se ve que es de alto valor, la mayoría alrededor de 0.7, teniendo en cuenta que se ha factorizado la matriz de correlaciones, es decir se trabaja con datos estandarizados.

En la **Figura 4.14.1** es de gran importancia pues allí se ve como influyen sus ponderaciones en el Primer factor que denominaremos "Nivel de Inversión", vemos como influye positivamente las quince variables.

El mapa No. 14 (**Figura 4.14.2**), muestra como está estructurado el "Nivel de Inversión en Salud" del Perú que es como llamamos a nuestro primer factor, todo esto a nivel Provincial, así pues las provincias con mejor nivel de inversión en salud, en concordancia con el **Cuadro A8** son Jorge Basadre, Tahuamanu, Mariscal Nieto, Tambopata, Gral. Sanchez Cerro, entre otras y las provincias que tiene Peor "Nivel de inversión en Salud" son, de acuerdo al índice Julcan Lauricocha, Yarowilca, Pachitea, Cutervo entre otras.

Cuadro 4.14.1 Mapa de Tendencia de Inversión en Salud
 Nivel Nacional por Provincia - Todas las Instituciones
 Método de Análisis Factorial - CP

Variables	Ponderaciones factoriales estimadas		Comunalidad
	F1	F2	
NRO_ESTA	0.336	-0.786	0.730
NRO_MEDI	0.796	0.318	0.735
OTR_PROF	0.859	0.101	0.748
TECN_ASI	0.743	-0.506	0.807
AREA_CON	0.757	-0.262	0.642
NUM_CAMA	0.818	-0.102	0.680
NUM_LABO	0.720	-0.210	0.562
N_ESTAGU	0.241	0.668	0.505
N_ESTDES	0.380	0.743	0.696
N_ESTLUZ	0.503	0.672	0.705
N_TELECO	0.508	0.579	0.594
N_CERCOP	0.097	0.465	0.225
N_EQSERF	0.892	-0.138	0.816
N_EQSERI	0.666	-0.498	0.692
N_EQAPOY	0.787	0.085	0.626
Porcentaje acumulado del total de varianza muestral (estandarizada) explicada	42.533	65.091	

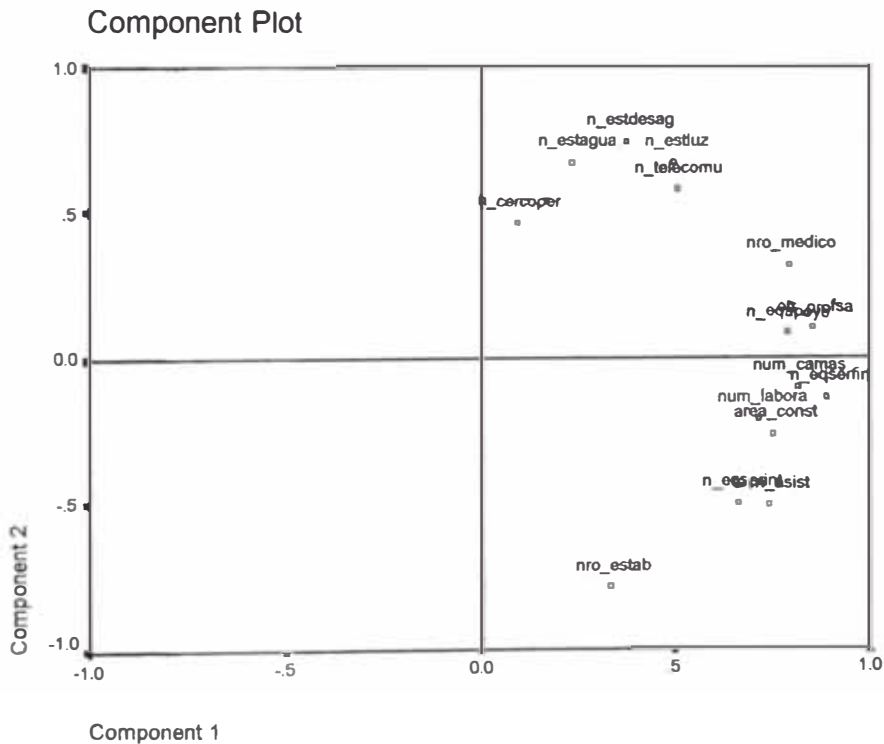
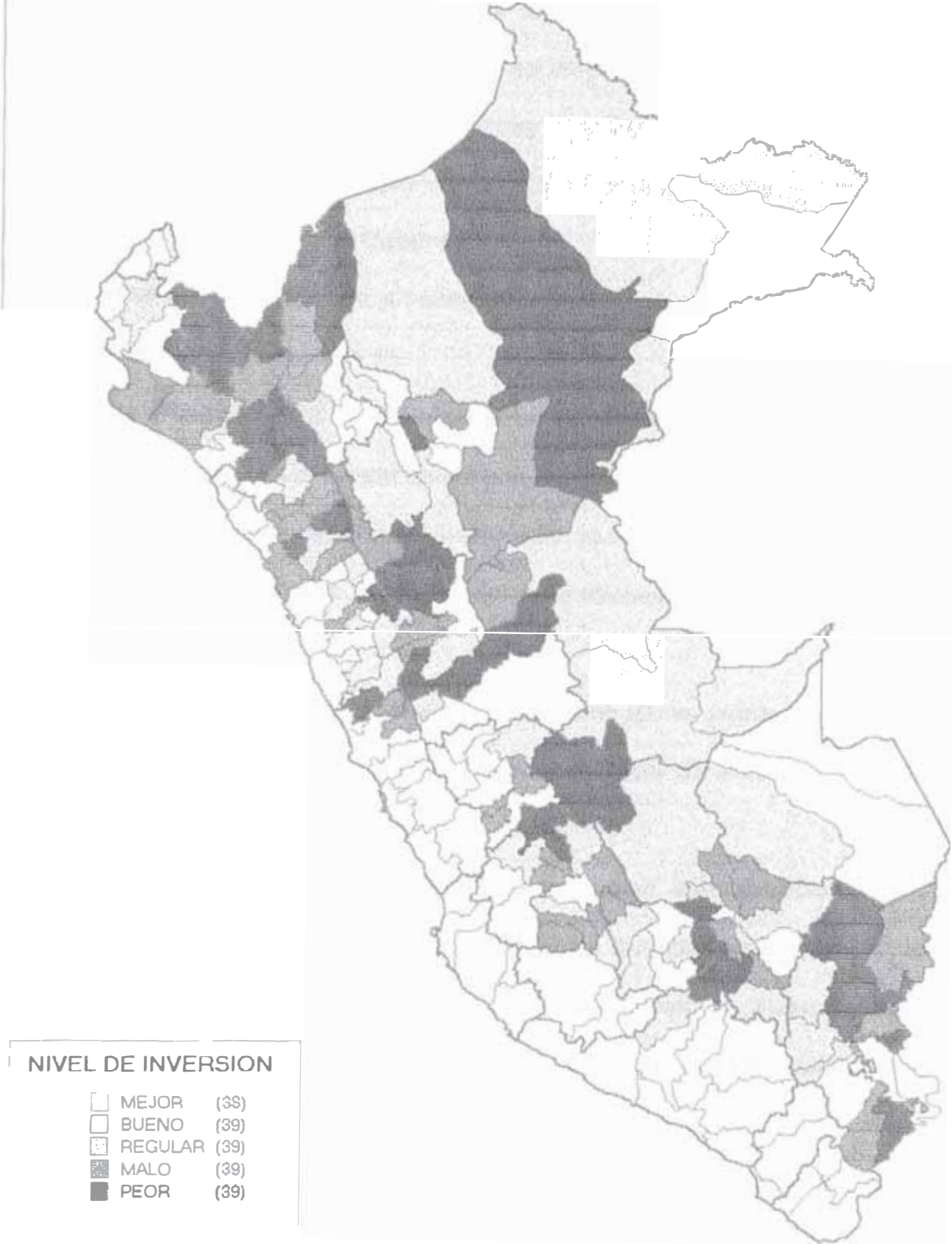


Figura 4.14.1 Ponderaciones de las variables en los dos primeros factores(Mapa 14)

MAPA No. 14
MAPA DE TENDENCIA DE INVERSION EN SALUD
PERU : 1996



Elaboración Propia

Mapa_14.wor |

Figura 4.14.2 MAPA No. 14 : Nivel Provincial - Todas las Instituciones. Método de Análisis Factorial - CP (Con índices por 10 mil habitantes y Porcentajes de Establecimientos)

Resultados de Mapa 15

En este mapa a nivel Provincial, desarrollada con el análisis factorial y con el método de Máxima Verosimilitud, y donde se emplea datos pertenecientes a todas las instituciones que prestan servicios en salud, así tenemos el Ministerio de salud, el IPSS que hoy se denomina EsSalud, Fuerzas armadas y policiales, Clínicas Privadas lucrativas y otros, para los datos de las mismas variables que son expresados en índices por diez mil habitantes como se indica en el **Cuadro A7**.

La matriz de correlaciones (**Cuadro B7**) para las quince variables es lo mismo que en el caso correspondiente al Mapa 13.

En la **Cuadro 4.15.1** el factor tiene un porcentaje de varianza de explicación de 48.242 %, así también en el listado de las ponderaciones del primer factor en las diferentes quince variables se puede ver que es de alto valor, la mayoría alrededor de 0.7, teniendo en cuenta que se ha factorizado la matriz de correlaciones, es decir se trabajó con datos estandarizados.

La **Figura 4.15.1** es de gran importancia pues allí se puede ver como influyen las variables en el Primer factor que denominaremos "Nivel de Inversión", se observa una influencia positiva en las quince variables.

El mapa No. 15 (**Figura 4.15.2**), muestra como está estructurado el "Nivel de Inversión en Salud" del Perú que es como llamamos a nuestro primer factor, todo esto a nivel Provincial, así pues las provincias con mejor nivel de inversión en salud, en concordancia con el **Cuadro A7** son Tahuamanu, Jorge Basadre, Gral. Sánchez Cerro, Paucar de Sara Sara, entre otras y las provincias que tienen peor "Nivel de inversión en Salud" son, de acuerdo al índice Julcan Pachitea, Chumbivilcas, Lauricocha, Yarrowilca entre otras.

Cuadro 4.15.1 Mapa de Tendencia de Inversión en Salud
 Nivel Nacional por Provincia - Todas las Instituciones
 Método de Análisis Factorial - MV

Variables	Ponderaciones factoriales estimadas		Comunalidad
	F1	F2	
NRO_ESTA	0.632	-0.538	0.688
NRO_MEDI	0.554	0.618	0.690
OTR_PROF	0.742	0.415	0.723
TECN_ASI	0.788	0.039	0.623
AREA_CON	0.749	0.128	0.577
NUM_CAMA	0.712	0.392	0.661
NUM_LABO	0.621	0.308	0.480
N_ESTAGU	0.695	-0.582	0.822
N_ESTDES	0.636	-0.398	0.563
N_ESTLUZ	0.760	-0.291	0.663
N_TELECO	0.590	0.017	0.349
N_CERCOP	0.469	-0.459	0.430
N_EQSERF	0.879	0.215	0.818
N_EQSERI	0.810	-0.188	0.691
N_EQAPOY	0.665	0.407	0.609
Porcentaje acumulado del total de varianzas muestrales (estandarizada) explicada	48.242	62.58	

Factor Plot

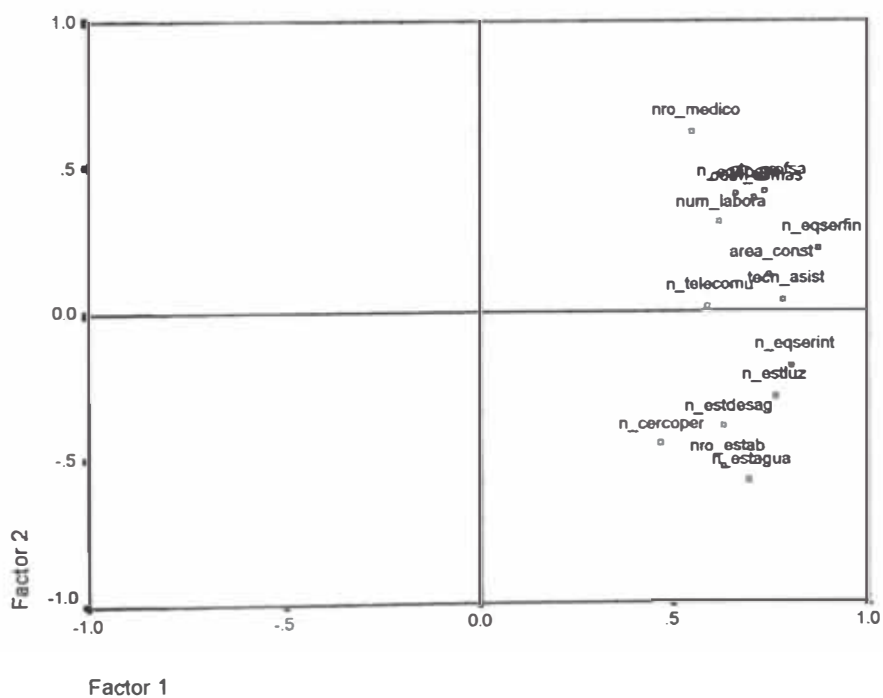
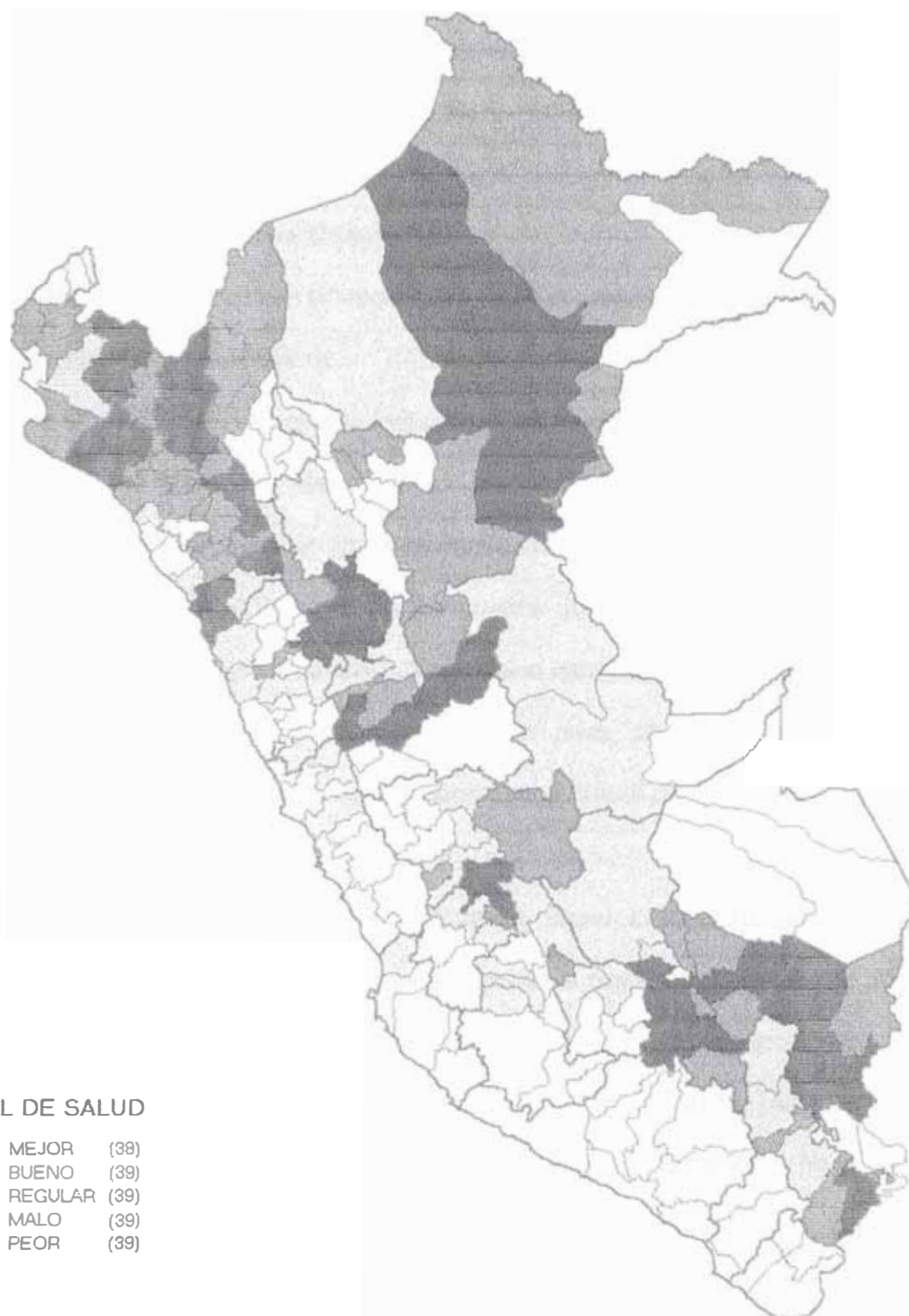


Figura 4.15.1 Ponderaciones de las variables en los dos primeros factores (Mapa 15)

MAPA No. 15
MAPA DE TENDENCIA DE INVERSION EN SALUD
PERU : 1996



Elaboración Propia

Mapa_15.wor

Figura 4.15.2 MAPA No. 15 : Nivel Provincial - Todas las Instituciones. Método de Análisis Factorial - MV (Con índices por 10 mil habitantes)

Resultados de Mapa 16

En este mapa a nivel Provincial, desarrollada con el análisis factorial con el método de Máxima Verosimilitud, y donde se emplea datos pertenecientes a todas las instituciones que prestan servicios en salud, así tenemos el Ministerio de salud, el IPSS que hoy se denomina EsSalud, Fuerzas armadas y policiales, Clínicas Privadas lucrativas y otros, tiene la variante que a diferencia del mapa anterior se utilizan datos expresados en porcentajes como se indica en el **Cuadro A8**.

La matriz de correlaciones (**Cuadro B8**) para las quince variables es lo mismo que en el caso correspondiente al Mapa 14.

En el **Cuadro 4.16.1** el primer factor tiene un porcentaje de varianza de explicación de 39.99 %, que dicho sea de paso es bajo, como también en el listado de las ponderaciones del primer factor en las diferentes quince variables se ve que es de alto valor, la mayoría alrededor de 0.7, teniendo en cuenta que se ha factorizado la matriz de correlaciones, es decir que se ha trabajado con datos estandarizados.

La **Figura 4.16.1** es de gran importancia pues allí se ve como influyen sus ponderaciones en el Primer factor que denominamos "Nivel de Inversión", vemos que las quince variables influyen positivamente.

El mapa No. 16 (**Figura 4.16.2**), muestra como está estructurado el "Nivel de Inversión en Salud" del Perú que es como llamamos a nuestro primer factor, todo esto a nivel Provincial, así pues las provincias con mejor nivel de inversión en salud, en concordancia con el **Cuadro A8** son Tahuamanu, Jorge Basadre, Tambopata, Mariscal Nieto y Gral. Sánchez Carrion, entre otras y las provincias que tienen peor "Nivel de inversión en Salud" son, de acuerdo al índice Julcan Pachitea, Chumbivilcas, Cutervo, Yarrowilca entre otras.

Cuadro 4.16.1 Mapa de Tendencia de Inversión en Salud
 Nivel Nacional por Provincia - Todas las Instituciones
 Método de Análisis Factorial - MV

Variables	Ponderaciones factoriales estimadas		Comunalidad
	F1	F2	
NRO_ESTA	0.441	-0.746	0.751
NRO_MEDI	0.730	0.445	0.731
OTR_PROF	0.827	0.227	0.736
TECN_ASI	0.799	-0.391	0.791
AREA_CON	0.756	-0.138	0.590
NUM_CAMA	0.795	0.046	0.634
NUM_LABO	0.703	-0.067	0.499
N_ESTAGU	0.152	0.569	0.347
N_ESTDES	0.266	0.697	0.557
N_ESTLUZ	0.392	0.695	0.637
N_TELECO	0.402	0.610	0.534
N_CERCOP	0.044	0.352	0.125
N_EQSERF	0.890	-0.017	0.792
N_EQSERI	0.719	-0.423	0.696
N_EQAPOY	0.730	0.217	0.579
Porcentaje acumulado del total de varianza muestral (estandarizada) explicada	39.998	59.997	

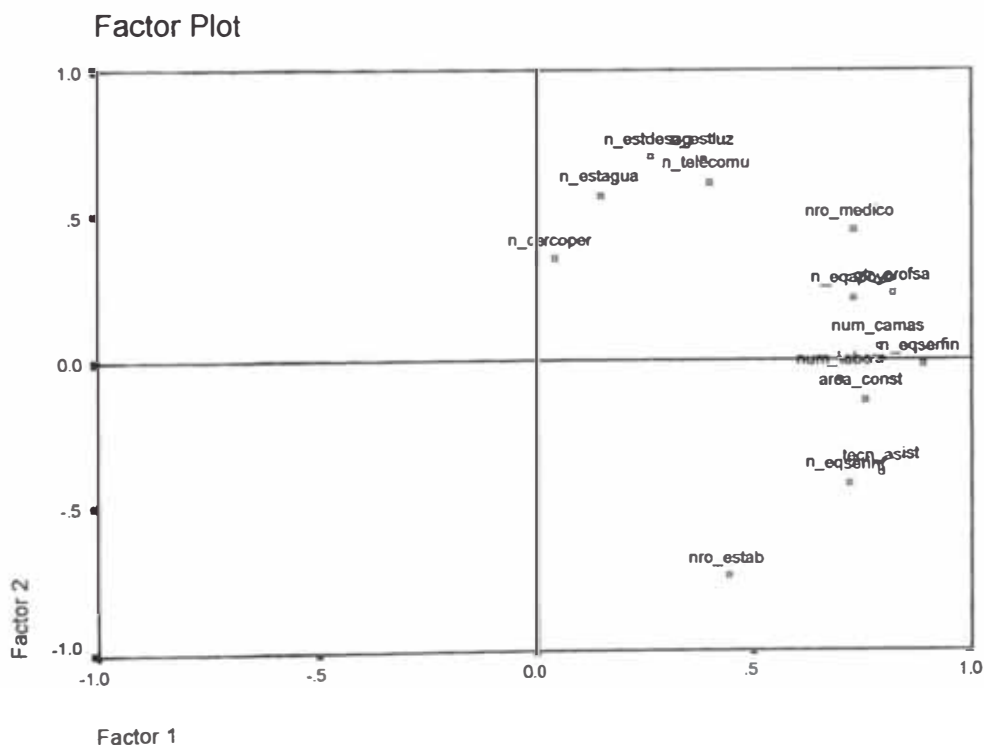
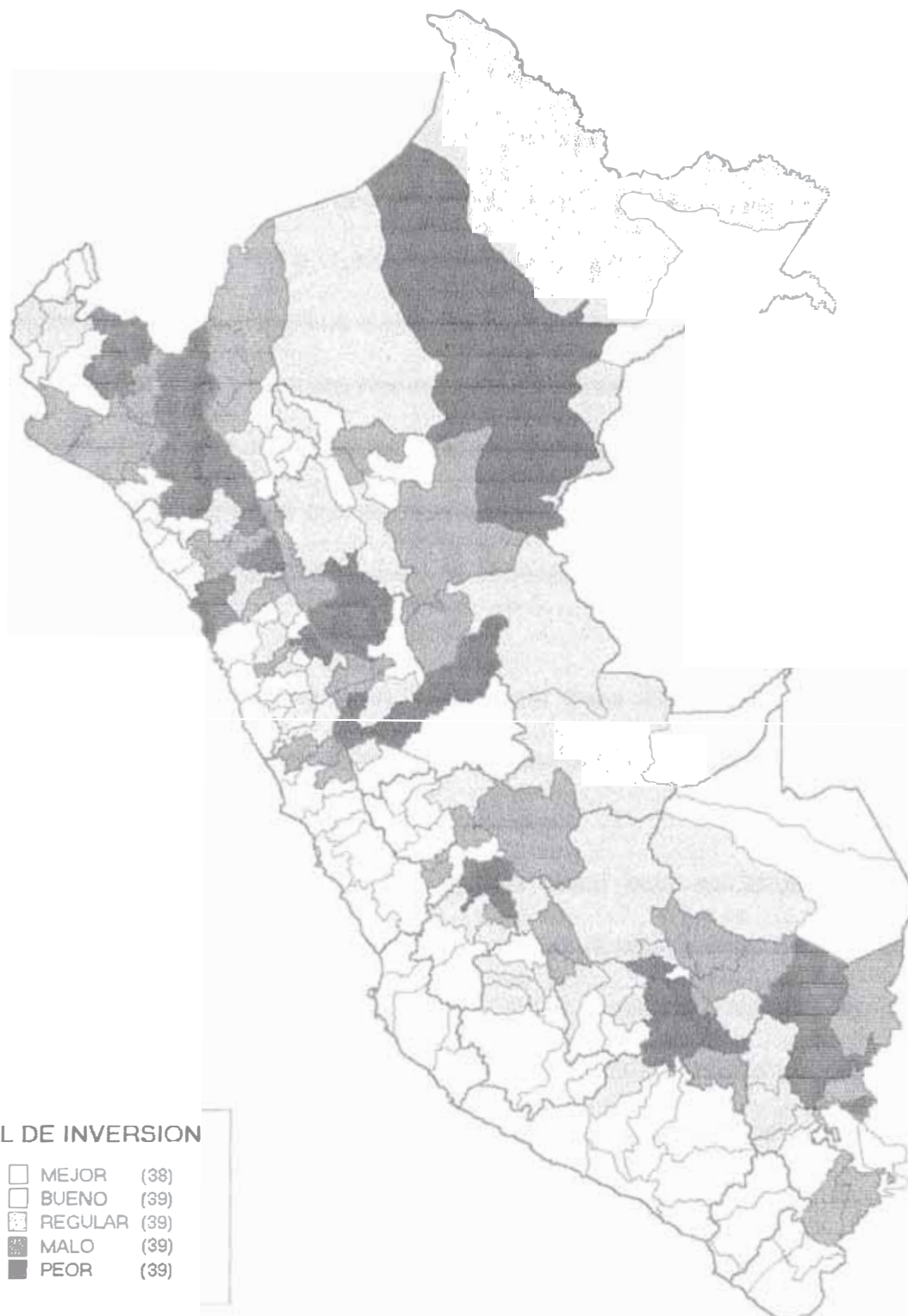


Figura 4.16.1 Ponderaciones de las variables en los dos primeros factores (Mapa 16)

MAPA No. 16
MAPA DE TENDENCIA DE INVERSION EN SALUD
PERU : 1996



Elaboración Propia

Mapa_16 wor

Figura 4.16.2 MAPA No. 16 : Nivel Provincial - Todas las Instituciones. Método de Análisis Factorial - MV (Con índices por 10 mil habitantes y Porcentajes de establecimientos)

Resultados de Mapa 17

En este mapa a nivel Provincial, desarrollada con el análisis factorial y con el método de Componentes Principales, y donde se emplea datos pertenecientes solo al Ministerio de salud, para el diseño se empleo datos expresados todos como una tasa por diez mil habitantes, vemos que, el **Cuadro A9** tiene ordenamiento determinado por el score obtenido con método de regresión, que es el método que se utiliza para calcular los scores, y su variación esta entre 6.753 y -1.324. La matriz la correlaciones (**Cuadro B9**) para las quince variables tienen gran amplitud variando desde 0.100 negativo hasta 0.912.

En el **Cuadro 4.17.1** el primer factor tiene un porcentaje de varianza de explicación de 54.5 %, como también en el listado de las ponderaciones del primer factor en las diferentes quince variables se ve que es de alto valor, la mayoría alrededor de 0.7, teniendo en cuenta que se ha factorizado la matriz de correlaciones, es decir se ha trabajado con datos estandarizados.

La **Figura 4.17.1** es de gran importancia pues allí se ve como influyen las ponderaciones de las variables en el Primer factor que denominaremos "Nivel de Inversión", vemos como influye positivamente las quince variables.

El mapa No. 17 (**Figura 4.17.2**), muestra como esta estructurado el "Nivel de Inversión en Salud" del Perú que es como llamamos a nuestro primer factor, todo esto a nivel Provincial, así pues las provincias con mejor nivel de inversión en salud, en concordancia con el **Cuadro A9** son Tahuamanu, Gral.Sanchez Cerro, Paucar del Sara Sara, Canta, Tarata entre otras, y las provincias que tiene Peor "Nivel de inversión en Salud son, de acuerdo al índice Talara Julcan Pachitea, Chiclayo, Chumbivilcas entre otros.

Cuadro 4.17.1 Mapa de Tendencia de Inversión en Salud
 Nivel Nacional por Provincia - Ministerio de Salud
 Método de Análisis Factorial - CP

Variables	Ponderaciones factoriales estimadas		Comunalidad
	F1	F2	
NRO_ESTA	0.800	-0.200	0.681
NRO_MEDI	0.630	0.485	0.632
OTR_PROF	0.732	0.370	0.672
TECN_ASI	0.846	0.179	0.747
AREA_CON	0.761	0.033	0.580
NUM_CAMA	0.667	0.312	0.543
NUM_LABO	0.577	0.399	0.492
N_ESTAGU	0.800	-0.483	0.873
N_ESTDES	0.650	-0.524	0.697
N_ESTLUZ	0.776	-0.237	0.659
N_TELECO	0.560	0.002	0.313
N_CERCOP	0.611	-0.452	0.578
N_EQSERF	0.867	0.114	0.765
N_EQSERI	0.865	-0.047	0.751
N_EQAPOY	0.827	0.119	0.698
Porcentaje acumulado del total de varianza muestral (estandarizada) explicada	54.541	64.534	

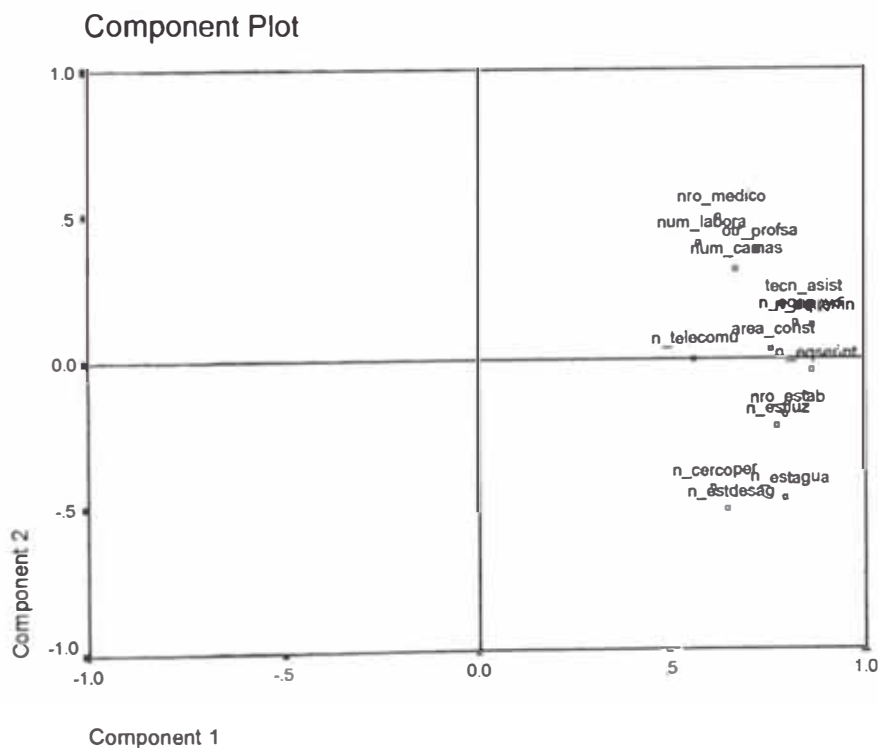
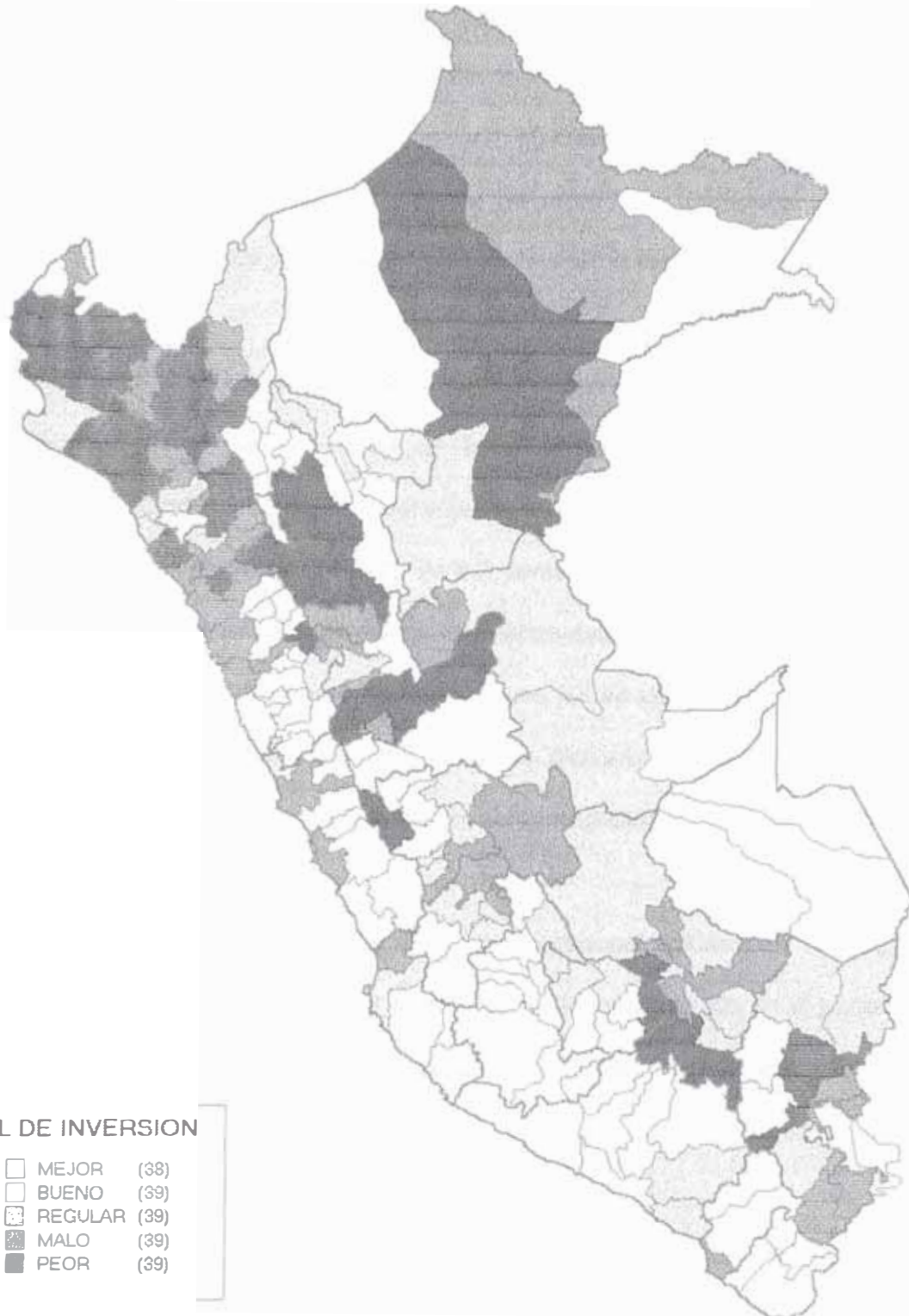


Figura 4.17.1 Ponderaciones de las variables en los dos primeros factores (Mapa 17)

MAPA No. 17
MAPA DE TENDENCIA DE INVERSION EN SALUD
PERU : 1996



NIVEL DE INVERSION

- MEJOR (38)
- BUENO (39)
- REGULAR (39)
- MALO (39)
- PEOR (39)

Elaboración Propia

Mapa_17.wor

Figura 4.17.2 MAPA No. 17 : Nivel Provincial - Ministerio de Salud. Método de Análisis Factorial - CP (Con índices por 10 mil habitantes)

Resultados de Mapa 18

En este mapa a nivel Provincial, desarrollada con el análisis factorial y con el método de Componentes Principales y donde se emplea datos pertenecientes solo al Ministerio de salud, tiene la variante que a diferencia del mapa anterior se utilizan datos expresados en porcentajes como se indica en el **Cuadro A10**, tiene ordenamiento determinado por el score obtenido con método de regresión. La matriz la correlaciones (**Cuadro B10**) para las quince variables tienen gran amplitud variando desde 0.003 negativo hasta 0.817 .

En el **Cuadro 4.18.1** el primer factor tiene un porcentaje de varianza de explicación de 41.18 %, que dicho sea de paso es baja, como también en el listado de las ponderaciones del primer factor en las diferentes quince variables se ve que es de diversos valores, sin embargo, la mayoría alrededor de 0.7, teniendo en cuenta que se ha factorizado la matriz de correlaciones, es decir se trabajo con datos estandarizados.

La **Figura 4.18.1** es de gran importancia pues allí se ve que el "Nivel de Inversión" esta obviamente repartido entre las dos primeros factores, en el caso de la primera componente las variables influyen positivamente como negativamente, lo mismo ocurre en el segundo factor.

El mapa No. 18 (**Figura 4.18.1**), muestra como esta estructurado el "Primer factor del modelo", todo esto a nivel Provincial. En este caso, por lo mencionado en el parrafo anterior no se puede afirmar nada del primer factor el cual solo tendria poco porcentaje del "Nivel de inversión" en salud.

Cuadro 4.18.1 Mapa de Tendencia de Inversión en Salud
 Nivel Nacional por Provincia - Ministerio de Salud
 Método de Análisis Factorial - CP

Variables	Ponderaciones factoriales estimadas		Comunalidad
	F1	F2	
NRO_ESTA	0.778	-0.386	0.755
NRO_MEDI	0.670	0.435	0.639
OTR_PROF	0.752	0.348	0.686
TECN_ASI	0.903	-0.208	0.858
AREA_CON	0.782	-0.073	0.616
NUM_CAMA	0.728	0.006	0.530
NUM_LABO	0.654	-0.102	0.438
N_ESTAGU	-0.107	0.738	0.555
N_ESTDES	-0.149	0.791	0.648
N_ESTLUZ	-0.046	0.793	0.631
N_TELECO	0.016	0.648	0.420
N_CERCOP	-0.051	0.509	0.262
N_EQSERF	0.870	0.144	0.778
N_EQSERI	0.858	-0.074	0.741
N_EQAPOY	0.802	0.301	0.734
Porcentaje acumulado del total de varianza muestral (estandarizada) explicada	41.184	61.944	

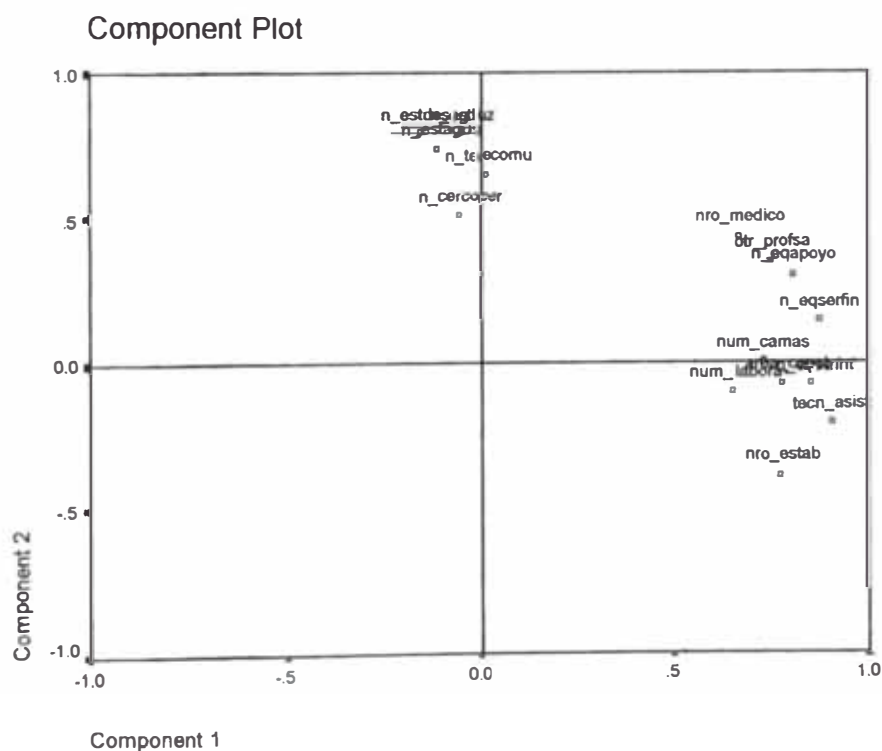
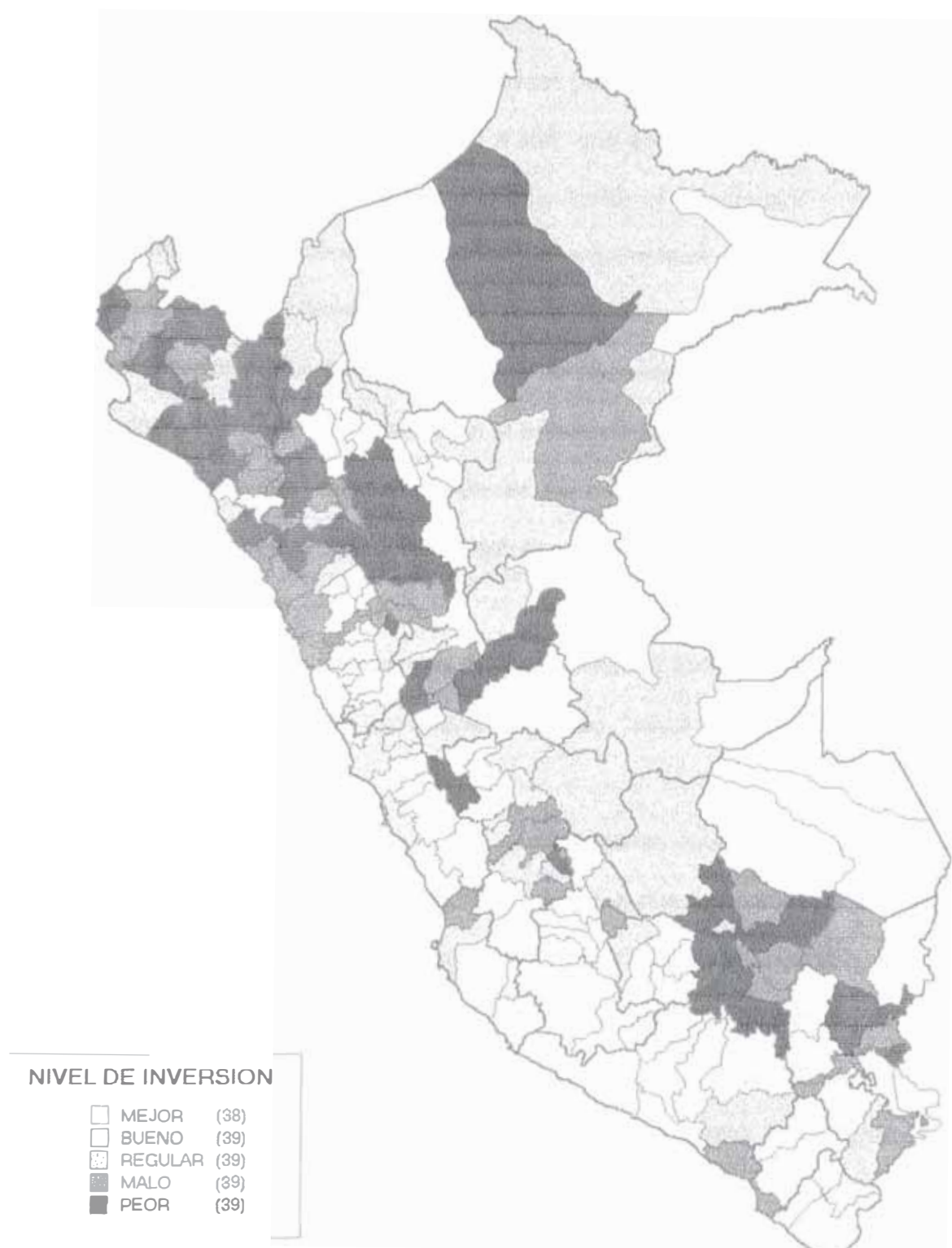


Figura 4.18.1 Ponderaciones de las variables en los dos primeros factores (Mapa 18)

MAPA No. 18
MAPA DE TENDENCIA DE INVERSION EN SALUD
PERU : 1996



Elaboración Propia

Mapa_18.wor

Figura 4.18.2 MAPA No. 18 : Nivel Provincial - Ministerio de Salud. Método de Análisis Factorial - CP (Con índices por 10 mil habitantes y Porcentajes de establecimientos)

Resultados de Mapa 19

En este mapa a nivel Provincial, desarrollada con el análisis factorial y con el método de Máxima Verosimilitud para los datos referentes solo al Ministerio de Salud de las mismas variables que son expresados en índices por diez mil habitantes como se indica en el **Cuadro A9**, estos scores varían desde 4.365, que es el mayor valor del índice que es denominado "Nivel de Inversión de Salud", hasta -1.192, el menor valor.

La matriz de correlaciones (**Cuadro B9**) para las quince variables es lo mismo que en el caso correspondiente al Mapa 17.

En el **Cuadro 4.19.1** primer factor tiene un porcentaje de varianza de explicación de 37.87 %, valor que es bajo así también en el listado de las ponderaciones del primer factor en las diferentes quince variables se puede ver que es de alto valor, la mayoría alrededor de 0.6, teniendo en cuenta que se ha factorizado la matriz de correlaciones, es decir se trabajó con datos estandarizados.

La **Figura 4.19.1** es de gran importancia pues allí se puede ver como influyen las variables en el Primer factor que denominaremos "Nivel de Inversión", como influye positivamente las quince variables.

El mapa No. 19 (**Figura 4.19.2**), muestra como está estructurado el "Nivel de Inversión en Salud" del Perú que es como llamamos a nuestro primer factor, todo esto a nivel Provincial, así pues las provincias con mejor nivel de inversión en salud, en concordancia con el **Cuadro A9** son Canta, Candarave, Paucar del Sara Sara, Tahuamanu, Gral. Sánchez Cerro entre otros, y las provincias que tienen peor "Nivel de inversión en Salud" son, de acuerdo al índice son Maynas, Puerto Inca, Loreto, Cnel Portillo, El Dorado entre otros.

Cuadro 4.19.1 Mapa de Tendencia de Inversión en Salud
Nivel Nacional por Provincia - Ministerio de Salud
Método de Análisis Factorial - MV

Variables	Ponderaciones factoriales estimadas		Comunalidad
	F1	F2	
NRO_ESTA	0.744	0.358	0.682
NRO_MEDI	0.351	0.520	0.394
OTR_PROF	0.471	0.529	0.502
TECN_ASI	0.560	0.686	0.784
AREA_CON	0.584	0.472	0.564
NUM_CAMA	0.378	0.582	0.482
NUM_LABO	0.239	0.638	0.463
N_ESTAGU	0.999	-0.009	0.999
N_ESTDES	0.782	-0.013	0.612
N_ESTLUZ	0.719	0.274	0.593
N_TELECO	0.441	0.268	0.266
N_CERCOP	0.641	0.082	0.418
N_EQSERF	0.614	0.612	0.751
N_EQSERI	0.674	0.550	0.757
N_EQAPOY	0.601	0.508	0.620
Porcentaje acumulado del total de varianza muestral (estandarizada) explicada	37.87	59.242	

Factor Plot

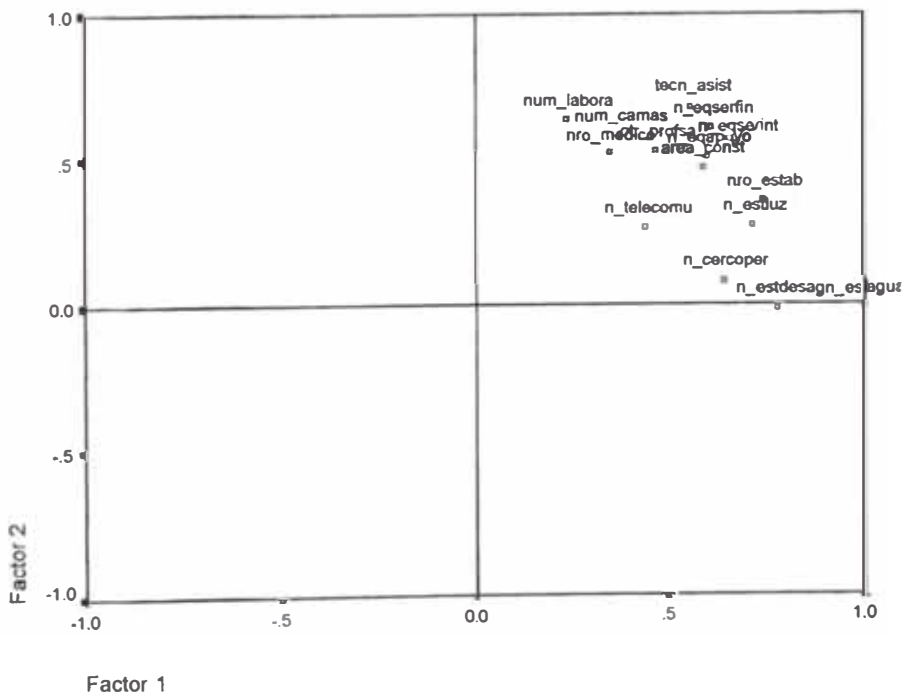


Figura 4.19.1 Ponderaciones de las variables en los dos primeros factores (Mapa 19)

MAPA No. 19
MAPA DE TENDENCIA DE INVERSION EN SALUD
PERU : 1996



Elaboración Propia

Mapa_19.wor

Figura 4.19.2 MAPA No. 19 : Nivel Provincial - Ministerio de Salud. Método de Análisis Factorial - MV
(Con índices por 10 mil habitantes)

Resultados de Mapa 20

En este mapa a nivel Provincial, desarrollada con el análisis factorial y con el método de Máxima Verosimilitud , solo con datos del Ministerio de Salud, tiene la variante que a diferencia del mapa anterior se utilizan datos expresados en porcentajes como se indica en el **Cuadro A10**, estos scores varían desde 7.875, que es el mayor valor del índice, hasta -1.358, el menor valor.

La matriz la correlaciones (**Cuadro B10**) para las quince variables es lo mismo que en el caso correspondiente al mapa 18.

En el **Cuadro 4.20.1** el primer factor tiene un porcentaje de varianza de explicación de 38.928 %, como también en el listado de las ponderaciones del primer factor en las diferentes quince variables se ve que es de alto valor, la mayoría alrededor de 0.6 incluyendo valores negativos, teniendo en cuenta que se ha factorizado la matriz de correlaciones, es decir que se ha trabajado con datos estandarizados.

La **Figura 4.20.1** es de gran importancia pues allí se ve como influyen sus ponderaciones en el Primer factor que denominamos con precaución "Nivel de Inversión", vemos que las diez variables influye positivamente y cinco influyen negativamente algo contradictorio.

El mapa No. 20 (**Figura 4.20.2**), muestra como esta estructurado el primer factor denominado con precaución el "Nivel de Inversión en Salud" del Perú que es como llamamos a nuestro primer factor, todo esto a nivel Provincial, así pues las provincias con mejor nivel de inversión en salud, en concordancia con el **Cuadro A10** son Tahuamanu, Gral. Sanchez Cerro, Paucar del Sara Sara, Tambopata, Tarata entre otras, y las provincias que tiene Peor "Nivel de inversión en Salud" son, de acuerdo al índice Talara, Julcan, Pachitea, Chiclayo, Chumbivilcas, entre otras.

Cuadro 4.20.1 Mapa de Tendencia de Inversión en Salud
Nivel Nacional por Provincia - Ministerio de Salud
Método de Análisis Factorial - MV

Variables	Ponderaciones factoriales estimadas		Comunalidad
	F1	F2	
NRO_ESTA	0.806	-0.352	0.773
NRO_MEDI	0.599	0.469	0.580
OTR_PROF	0.692	0.404	0.642
TECN_ASI	0.915	-0.170	0.866
AREA_CON	0.752	-0.022	0.566
NUM_CAMA	0.681	0.052	0.467
NUM_LABO	0.621	-0.053	0.389
N_ESTAGU	-0.131	0.648	0.437
N_ESTDES	-0.170	0.681	0.493
N_ESTLUZ	-0.094	0.730	0.541
N_TELECO	-0.021	0.561	0.315
N_CERCOP	-0.066	0.399	0.164
N_EQSERF	0.848	0.207	0.762
N_EQSERI	0.859	-0.030	0.739
N_EQAPOY	0.764	0.354	0.708
Porcentaje acumulado del total de varianza muestral (estandarizada) explicada	38.928	56.275	

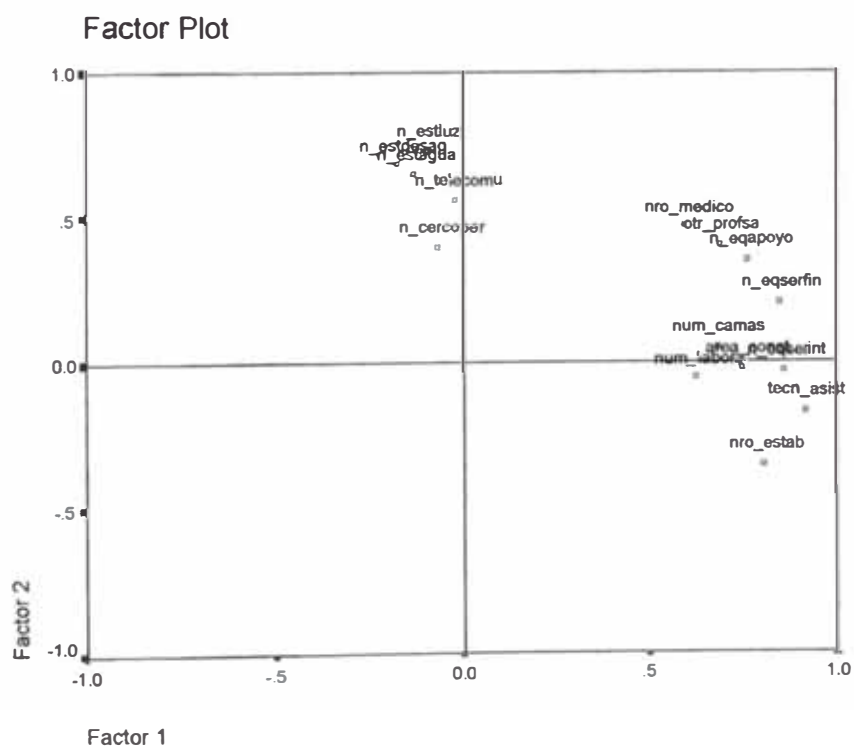
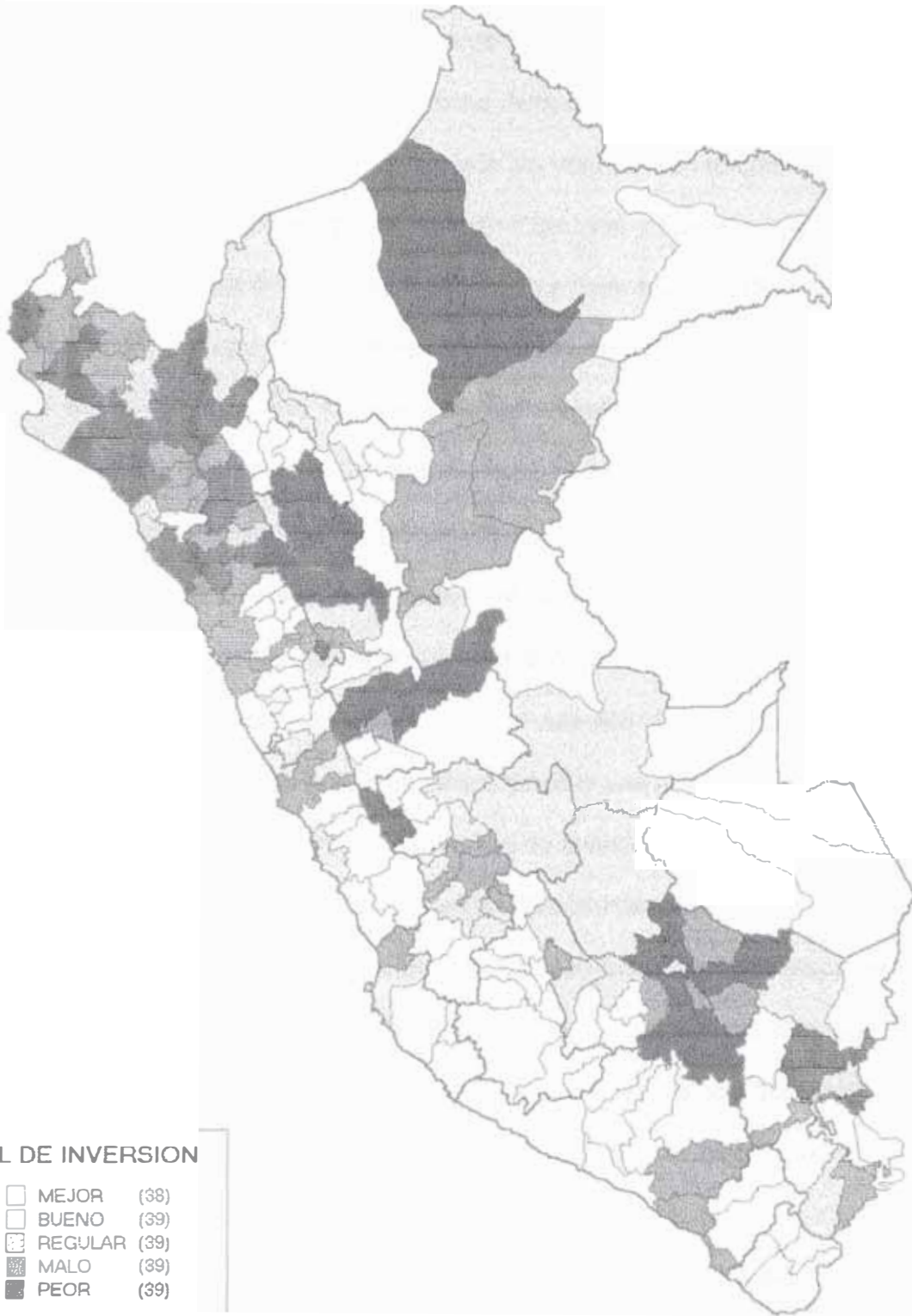


Figura 4.20.1 Ponderaciones de las variables en los dos primeros factores (Mapa 20)

MAPA No. 20
MAPA DE TENDENCIA DE INVERSION EN SALUD
PERU : 1996



Elaboración Propia

Mapa_20.wor

Figura 4.20.2 MAPA No. 20 : Nivel Provincial - Ministerio de Salud. Método de Análisis Factorial - MV (Con índices por 10 mil habitantes y Porcentajes de establecimientos)

OBSERVACIONES GENERALES DE LOS RESULTADOS ANTERIORES :

- El factor **Nivel de Salud** se identifica fácilmente debido a que todas las variables que se emplean son referentes al estado de salud de cada entidad (departamento o provincia) y es lógico que en todas intervengan dicho factor, esto también es comprobado al observar los gráficos de las ponderaciones de las variables en los dos primeros factores verificándose, los que influyen positivamente y los que influyen negativamente. Como ejemplo de influencia positiva podemos tomar entre otras a la variable de Médicos por 10 mil habitantes, Esperanza de vida al Nacer ubicadas en un extremo de la representación gráfica y en el otro extremo los que influyen negativamente como son la Tasa de Mortalidad Infantil, Tasa Bruta de Mortalidad, etc.
- El factor **Nivel de Tendencia de la Inversión en Salud** también se identifica fácilmente porque todas las variables que se emplean son de inversiones en el área de salud. Así se verifica que las variables que se emplean están en un extremo en la representación gráfica de las ponderaciones de las variables en los dos primeros factores , es decir el factor tiene influencia positiva en las variables, pues si una variable aumenta entonces el primer factor conocido como nivel de Tendencia de Inversión en Salud también aumenta.
- El mapa No. 1 y mapa No.2 sirve para verificar que el modelo que se expone aquí tiene un resultado muy similar a los resultados obtenidos en la publicación "Mapa de Salud del Perú" de 1991, donde se aplica un método más simple o sencillo que es el de promedios. Con el método del análisis factorial obtenemos un ordenamiento óptimo según las escalas de los scores del primer factor el cual lo hemos identificado como "nivel de salud".
- En el mapa Provincial de Salud del Perú - 1993 es decir mapa No.3 y mapa No.4 se obtiene un espectro de cómo están los niveles de salud a nivel provincial, también se observa que en algunos departamentos vistos como un ente global tienen mala situación

de salud, pero al hacer el estudio en su interior es decir a nivel provincial, la provincia capital de departamento tiene alto nivel de salud como es el caso del Cuzco evidenciando un caso de centralismo. El resultado nos permitirá contraponer al mapa de inversiones y verificar que la tendencia de la inversión esta en correcta dirección o se tiene que modificar la política de distribución de recursos en el sector salud, es decir si una provincia tiene un bajo nivel de salud, lo mejor que podría pasar es que a esta provincia se le este dando mayor apoyo en las inversiones en salud, lo que en el futuro llevaría a equilibrar su nivel de salud con respecto a las demás provincias.

- Con la base de datos del 2do. CISRESSA se elaboró dos tipos de resultados o indicadores, en uno se trabajó con indicadores por diez mil habitantes y en el otro con algunos indicadores y también con porcentajes de establecimientos con ciertas características como se indica en los respectivos cuadros, esto nos permitió verificar físicamente que los resultados obtenidos con solo indicadores por diez mil habitantes, según los cuadros respectivos, tienen mayor porcentaje de explicación de las varianzas de los datos estandarizados (matriz de correlaciones) .
- El mapa No. 5 nos muestra que existe buena inversión en el sur pero, hay otra observación, hay casos como el departamento de Madre de Dios que con poca población (74,067 hab.) tiene el mas alto valor inversión.
- El mapa No. 9 tiene mucha similitud al mapa No. 5, lo que se observa principalmente es que el Ministerio de Salud si ha invertido lijeramente más sus recursos, donde aparece la mayor necesidad indicada por un bajo *nivel de salud*, según el mapa No.1, como es el caso del departamento de Ayacucho.
- En el mapa No. 13 se observa los niveles de salud en el interior del departamento, pues esta elaborado a nivel de provincias.
- El mapa No. 17 muestra detalladamente los resultados del Mapa No. 9.

4.4 Balance de Mapas y su Aplicación

Lo que se tiene hasta este punto son los mapas de salud y los mapas de tendencia de inversión en salud con sus respectivos parámetros , lo siguiente que corresponde hacer es elaborar una forma de evaluar o medir la relación que existe entre ambos tipos de mapas.

Una dificultad que se ve para la realización del balance es que se tendría que hallar una relación entre una magnitud y otra de su misma especie, lo que no se da en este caso, ya que aquí con lo que se cuenta son los tanteadores (llamado también scores factoriales) dados por el método de regresión del análisis factorial, uno midiendo el *Nivel de Salud* y el otro midiendo el *Nivel de tendencia de inversión* en el Sector Salud.

Una forma de salvar esta dificultad es empleando los rangos, igualando de este modo la unidad de medición de dos magnitudes diferentes, con el objetivo de verificar si existe correspondencia o no, es mas, esta solución nos permitiría, además, obtener un ordenamiento de las entidades geográficas para las nuevas inversiones en Salud priorizando unas sobre otras, como veremos a continuación. Con todo esto lo que hacemos es solucionar dos problemas con un solo procedimiento.

Con los mapas construidos anteriormente, entonces, ya se puede realizar una aplicación muy importante, el cual sería elaborar un ranking de prioridades de entidades geográficas para nuevas inversiones en el Sector Salud, el cual , de alguna manera, nos dirá si existe o no alguna especie de engranaje o reciprocidad entre el Mapa de Salud y el Mapa de Tendencia de Inversión en Salud.

Para esto, se empleó los scores generados por el análisis factorial sobre el cual se realiza un proceso de etiquetar a cada entidad geográfica con un determinado rango o número de orden que le correspondería.

Esta aplicación se puede hacer a nivel de cualquier entidad geográfica que cuente con sus respectivos datos, así en nuestro caso se puede hacer por departamentos, o por

provincias.

Como caso ilustrativo se aplicó el algoritmo para el nivel departamental tal como lo describimos en las líneas siguientes.

Procedimiento a nivel departamental

El procedimiento es el siguiente:

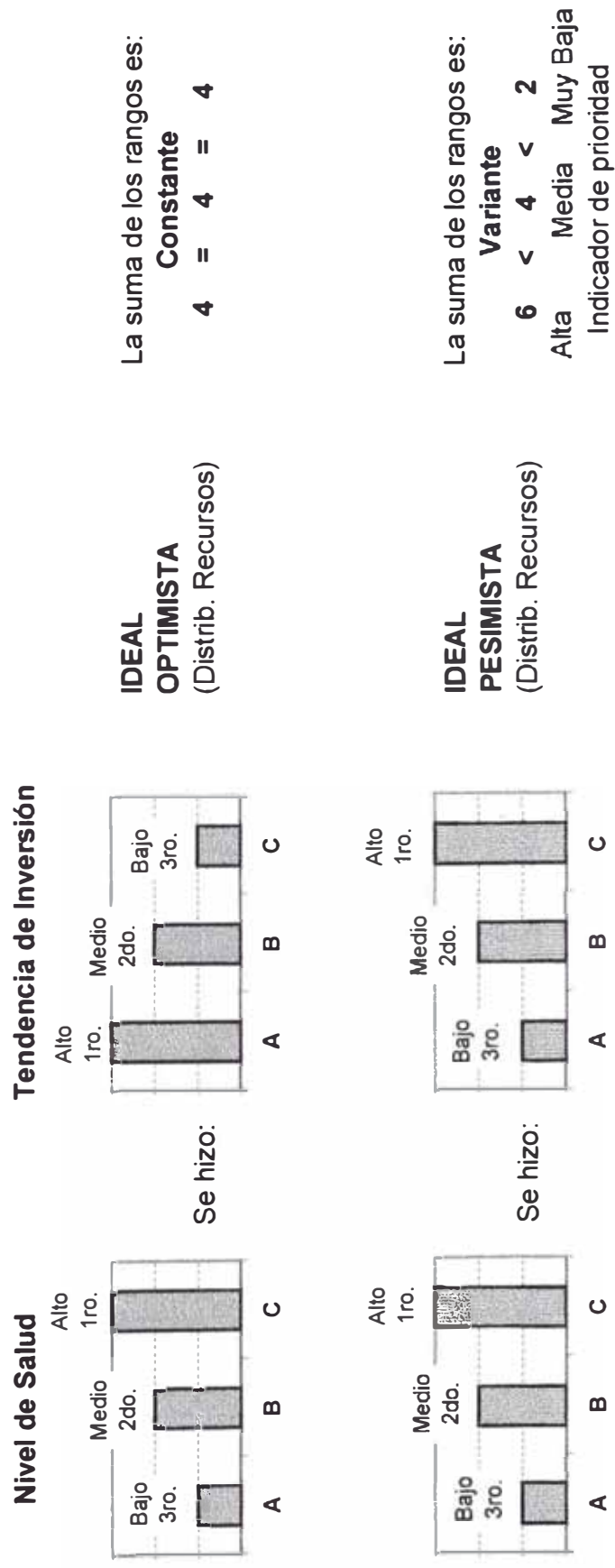
1. Asignamos un número de orden de menor a mayor según que el Nivel de Salud vaya disminuyendo, es decir para altos Niveles de Salud le corresponda números de rango bajos (1°,2°,3°,...) y para bajos Niveles de Salud le corresponde números de rango altos (...23°,24°,25°).
2. Asignamos un número de orden de menor a mayor según que el Nivel de Tendencia de Inversión en Salud vaya disminuyendo, es decir para altos Niveles de Tendencia de Inversión en Salud le corresponda números de rango bajos (1°,2°,3°,...) y para bajos Niveles de Tendencia de Inversión en Salud le corresponde números de rango altos (...23°,24°,25°).
3. Para cada departamento sumar los rangos dados anteriormente, así se obtiene un valor que denominaremos **Indicador de Prioridad**, el cual se interpreta de la siguiente manera, para valores altos del Indicador de Prioridad se esta haciendo referencia a mayor necesidad de inversión.

Como generalización del método, se puede decir que, el procedimiento de los tres pasos descritos anteriormente puede ser aplicado para cualquier entidad geográfica en estudio, es decir, que puede ser a nivel departamental, provincial y hasta distrital.

Primero se ilustra la lógica de procedimiento mediante el gráfico de barras, **Figura 4.21.1** y luego se aplica a nivel departamental, visualizandose, el ranking departamental para las nuevas inversiones en salud en el **Cuadro 4.21.1**, y finalmente se ilustra el mapa de priorización para nuevas inversiones, **Figura 4.21.2**.

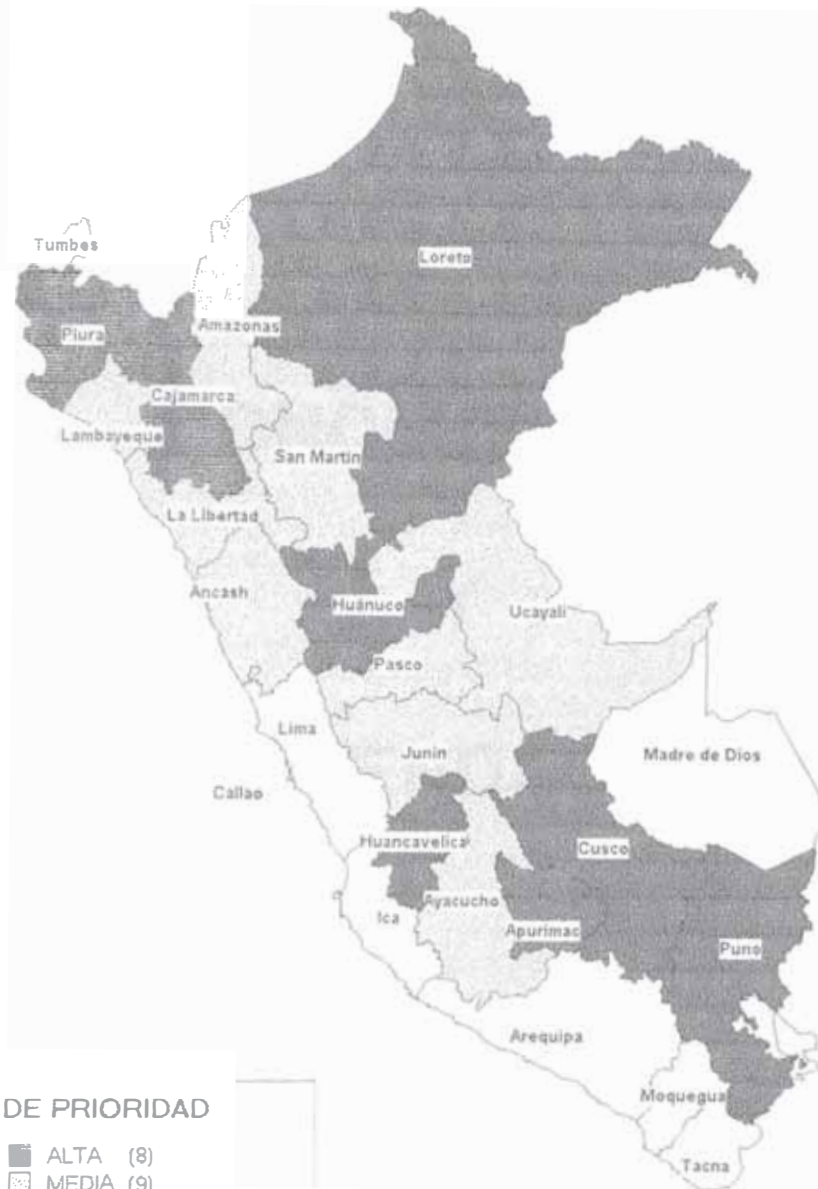
Figura 4.21.1

REPRESENTACION GRÁFICA PARA LA OPTIMIZACIÓN DE LA DISTRIBUCIÓN DE LOS RECURSOS DEL SECTOR SALUD



Elaboración Propia

MAPA No. 21
 MAPA DE PRIORIZACION DE INVERSION EN SALUD
 PERU : 1996



Elaboración Propia - RRC

Mapa_Resultado wor

Figura 4.21.2 MAPA No. 21 : Prioridad a Nivel Departamental. Tres estratos.

CONCLUSIONES

Y

RECOMENDACIONES

Conclusiones

Conclusión General

- El modelo de decisión presentado aquí es, entonces, consistente en la recopilación de información para la tabla de datos, que nos permita la construcción de índices por habitante, proceso con el cual comienza el modelamiento (ver Anexo 4), luego se aplica la técnica del Análisis Factorial a los índices elaborados, que nos proporciona entre otros parámetros, un tanteador (llamado también score), que nos indica el Nivel de Salud y otro tanteador que nos indica el Nivel de Tendencia de Inversión en el Sector Salud. Y finalmente con el procedimiento descrito aquí (ver 4.3), obtener, una evaluación de la reciprocidad de ambos mapas y al mismo tiempo un ordenamiento en la priorización de entidades geográficas para futuros proyectos de inversiones. Este modelo permite llegar a los resultados presentados aquí.
- Los resultados fueron sorprendentes, se ha detectado que el mapa de la "**Tendencia de Inversión en Salud**" involucrando todas las instituciones que prestan servicios de salud, no corresponde a la necesidad presentada por el mapa de "**Nivel de Salud**".

Conclusiones Específicas

- En el mapa No. 1 y mapa No.2 se verifica que el modelo que se expone aquí tiene un resultado muy similar que el que se aplica en la publicación "Mapa de Salud del Perú" de 1991, lo cual nos permite concluir que mediante un método más científico podemos

elaborar un mapa de salud y obtener óptimos resultados.

- En el mapa Provincial de Salud del Perú - 1993 es decir mapa No.3 y mapa No.4 se obtiene un espectro de cómo están los niveles de salud a nivel provincial.
- Se verificó físicamente que los resultados obtenidos con solo indicadores por diez mil habitantes, tienen mayor porcentaje de explicación de las varianzas de los datos estandarizados (matriz de correlaciones). Con este antecedente se puede asegurar que un resultado óptimo se obtiene con los indicadores por diez mil habitantes unicamente excluyendo los datos de indicadores porcentuales.
- En nuestro caso, es decir para este trabajo prácticamente los mejores resultados se obtiene aplicando en el método factorial la técnica de los componentes principales.
- El mapa No. 5 nos muestra una primera aproximación de lo que es la distribución de la inversión en el Sector Salud a nivel departamental.
- El mapa No. 9 es muy similar al mapa No. 5, pero hay una observación, el Ministerio de Salud si ha invertido mas, según el mapa donde hacia falta, como es el departamento de Ayacucho, pues en el Mapa No 1 se verifica un bajo nivel de salud.
- El mapa No. 13 nos muestra mas detalladamente los resultados del Mapa No. 5, observandose ademas baja tendencia de inversión en salud en el norte del pais, específicamente en los departamentos de Cajamarca, Piura, Lambayeque y parte de Amazonas y la Libertad.
- El mapa No. 17 nos muestra mas detalladamente los resultados del Mapa No. 9 y se observa además un espectro similar al Mapa No. 13.

Por los puntos anteriores, sus altos porcentajes de explicación de varianza y sus gráficos de ubicación de variables en el primer factor (Figuras 4.1.1, 4.3.1, 4.5.1, 4.9.1, 4.13.1, 4.17.1), se puede concluir que los mapas mas óptimos son el mapa No. 1 , mapa No. 3, mapa No. 5, el mapa No. 9 , el mapa No.13 y el Mapa No. 17.

- Como aplicativo del estudio desarrollado aquí, se ha elaborado una metodología que consiste en la utilización de los tanteadores factoriales (scores factoriales) generados por el análisis factorial, sobre el cual se realiza el proceso de etiquetar a cada entidad geográfica con un determinado rango o número de orden correspondiente, como se explica en el punto 4.3 del Capítulo 4, es así como se obtuvo el siguiente ranking por departamentos para nuevas inversiones en Sector Salud del Perú:

Orden de Prioridad para Nuevas Inversiones en el Sector Salud
Nivel Nacional por Departamentos
Perú : 1996

Orden para Nuevas Inversiones	Código	Departamento
1°	06	Cajamarca
2°	09	Huancavelica
3°	10	Huánuco
4°	21	Puno
5°	03	Apurimac
6°	08	Cusco
7°	20	Piura
8°	16	Loreto
9°	05	Ayacucho
10°	25	Ucayali
11°	14	Lambayeque
12°	01	Amazonas
13°	02	Ancash
14°	22	San Martin
15°	12	Junin
16°	13	La Libertad
17°	19	Pasco
18°	24	Tumbes
19°	11	Ica
20°	04	Arequipa
21°	17	Madre de Dios
22°	23	Tacna
23°	15	Lima
24°	07	Callao
25°	18	Moquegua

Recomendaciones

- Las entidades prestadoras de servicios de salud en el Perú, ya sea del Estado o del sector privado tienen el compromiso de participar en la modernización tecnológica y emplear técnicas científicas estadísticas, y lo que pretende esta tesis, en ese sentido, es brindar luces del moderno tratamiento estadístico, sugiriendo un nuevo enfoque del análisis de información.
- Según los resultados obtenidos en esta tesis, hay que hacer un cambio en la distribución de los recursos de salud, pues se tiene que primero identificar las necesidades de las diferentes entidades geográficas como son provincias o departamentos (Demanda en Salud), para luego satisfacerlas al momento de invertir (Oferta de Salud).
- El modelo de decisión presentado aquí se puede aplicar a los datos que se obtendrían con el censo de población que ha programado el INEI y si de ser posible al Tercer Censo Sanitario que también se haría muy posiblemente en el corto plazo.
- Se podría hacer otros trabajos de investigación para elaborar un listado de las provincias y/o departamentos ordenados por el orden de priorización para nuevas inversiones específicas como puede ser en equipamiento, recursos humanos, capacitación, etc. utilizando nuevos instrumentos de recojo de información.
- Explorar algunos otros factores que intervienen en el área de salud.
- Como sugerencia técnica, en la aplicación de este modelo, para la elaboración de la tabla de datos, sería mejor el empleo de índices por 10 mil habitantes en caso de ser posible, con el fin de homogenizar la información.

BIBLIOGRAFÍA

- [1] Johnson y Wichern. *Applied Multivariate Statistical Analysis*, New Jersey, Prentice-Hall. 1982.
- [2] Kaiser, H.F., "The Varimax Criterion for Analytic Rotation in Factor Analysis," *Psychometrika*, 23 (1958), 187-200
- [3] Bartlett, M. S., "The Statistical Conception of Mental Factors," *British Journal of Psychology*, 28 (1937), 97-104
- [4] Mapa de Salud del Perú : 1991. Ministerio de Salud-UNICEF. Oficina de Estadística e Informática. 1993.
- [5] Manual de Llenado de Hospitales-Clínicas, 2° Censo de Infraestructura Sanitaria y Recursos del Sector Salud. 2° CISRESSA - 1996. Ministerio de Salud