

CARACTERIZACIÓN DE LAS PROVINCIAS DEL PERÚ

Emma Cambillo Moyano

RESUMEN.- *Se presenta una aplicación del Análisis Factorial en la caracterización de las provincias del Perú; obteniéndose dos indicadores, los cuales fueron denominados: Indicador de desarrollo e Indicador de las condiciones de vida y necesidades básicas insatisfechas. Se obtiene un indicador de desarrollo global, el cual permite jerarquizar las provincias.*

1. INTRODUCCIÓN

Habitualmente para estratificar el país, se recurre al conocimiento subjetivo de las condiciones socio económicas de la población considerando la demarcación política, y se perfila la agrupación de las provincias en estratos teniendo en cuenta el conocimiento subjetivo de las provincias y algunas veces asociados a una o dos variables como puede ser el ingreso familiar. Al igual que la mayor parte de los países subdesarrollados, el Perú concentra sus capas poblacionales pobres en el campo, en las zonas andinas y en los pueblos alejados de las grandes ciudades. Para una adecuada caracterización de las provincias es necesario el conocimiento preciso del territorio, a través del análisis de un conjunto de variables socio económicas observadas espacialmente.

El análisis multivariante se define como un conjunto de métodos estadísticos que analizan de forma simultánea dos a más variables observadas, permitiendo una visión en conjunto del problema, desentrañando la compleja interacción de factores que afectan los fenómenos regionales, y en ese sentido, constituyen un avance frente a los estudios clásicos univariantes. El aporte de los métodos multivariantes es mayor, dado que existen muchos fenómenos, sobre todo aquellos que tienen una dimensión espacial, que no pueden medirse mediante una variable, sino como resultado de un conjunto de características.

2. MÉTODO

Los métodos de análisis multivariante son relativamente sencillos, puesto que utilizan esencialmente el álgebra lineal. En particular, el objetivo del análisis factorial [1], es estudiar la estructura de una nube de puntos resultante de un conjunto de mediciones realizadas en los individuos. Para conseguir dicho objetivo, se busca una representación simplificada, es decir una síntesis de la nube, en primer lugar, las tendencias centrales y luego las de dispersión. Las grandes líneas de dispersión de la nube serán los ejes factoriales .

Con el análisis factorial se trata de resumir y sintetizar la información contenida en una matriz de datos de manera que, con una pérdida de información mínima, se produzca una ganancia en interpretación. Pero ¿cómo medir esa información?. Existen diversas formas de cuantificar información de una matriz de datos, por ejemplo la variación total, que es uno de los criterios claves utilizado en el análisis factorial para medir la cantidad de información sintetizada.

En resumen, el objetivo del análisis factorial, será buscar un subespacio de variables de menor dimensión que contenga la mayor cantidad posible de información existente en la nube original, es decir, el que mejor se ajuste a la nube de puntos y la deforme lo menos posible.

3. APLICACIÓN DEL ANÁLISIS FACTORIAL A LA CARACTERIZACIÓN DE LAS PROVINCIAS DEL PERÚ

Los métodos multivariantes tiene por objetivo resumir la información contenida en un gran número de variables. Por lo tanto, para definir la matriz de datos inicial de la investigación, es conveniente justificar la elección de los elementos objetos de estudio y las variables que van a caracterizarlos. La elección de las provincias como unidad de análisis para este estudio se debió a que es el ámbito espacial más reducido establecido por el INEI. Las fuentes estadísticas utilizadas fueron: el Censo de Población y Vivienda 1993; Censo de Municipalidades 1995 y el Censo Agropecuario 1994.

La selección de variables, se realizó teniendo en cuenta que el propósito de la investigación es estudiar el desarrollo del territorio, por lo que estas deben caracterizar significativamente la situación de cada unidad de análisis, así como responder a indicadores de centralidad en un sentido amplio: demográfico, económico, de carácter urbano, etc.

El análisis exploratorio permitió:

- Seleccionar variables expresadas en porcentaje, porque ponen de manifiesto la estructura provincial respecto al departamento.
- Excluir variables con información repetida.
- Crear variables de participación, como el cociente entre el valor provincial y el departamental, debido a que expresan el grado de dominación de cada provincia sobre su entorno departamental.

La aplicación del análisis factorial implica el cálculo de la matriz de correlación, su diagonalización y obtención de los vectores y valores propios [5].

El proceso de análisis en conjunto, se realizó en varias etapas y sobre diferentes series de indicadores, hasta que se llegó a un resultado final considerado como adecuado para los objetivos de síntesis e interpretación. A modo de resumen señalaremos los criterios principales, que se utilizaron en las etapas sucesivas.

1. Se eliminaron variables:

- Con baja comunalidad.
- Con poca carga factorial en las primeras dimensiones.
- Con información repetida, observando los coeficientes de correlación simple, el indicador K.M.O. y el test de esfericidad de Bartlett.

- Se excluyeron del análisis aquellas variables que provocaban la aparición de un sólo factor principal, con valor propio pequeño, y apenas contribuían al objetivo de síntesis.
2. Se procuró utilizar el mayor número de variables expresadas en porcentaje o en algún tipo de participación, porque ponen en evidencia la estructura provincial respecto de la departamental.

Luego de realizados los ensayos factoriales, las variables seleccionadas de acuerdo a los objetivos de la investigación resultaron siendo treinta y dos, observándose correlaciones significativas entre variables relacionadas a educación, empleo, vivienda y fecundidad.

El cuadro A1, presenta medidas que ayudan a interpretar la adecuación del modelo de análisis factorial, el test de esfericidad de Barlett, indica que el nivel de relación existente entre las variables amerita la utilización de un método multivariado. El indicador de adecuación simple K.M.O. [2], con el valor 0.899; indica que es meritoria la aplicación del análisis factorial. Evaluando, el mismo indicador K.M.O. para cada variable (cuadro A2) se observan valores que superan 0.687, indicando de que el análisis factorial es adecuado.

La extracción de los factores, se realizó mediante el método de Componentes Principales [3] y se muestran en el cuadro A3, conjuntamente con los valores propios, porcentajes de variación explicada por cada factor y porcentajes acumulados. Considerando aquellos factores con valores propios mayores e iguales a uno, concluimos que la matriz de correlación puede resumirse en cinco factores.

Por otro lado, el análisis de las comunalidades indica que excepto el porcentaje de la población rural con respecto al total provincial (41.6%) las variables restantes tienen comunalidades superiores al 50%, por lo que el análisis factorial puede considerarse adecuado.

El criterio de Catell, [3], sugiere seleccionar tres factores, obteniéndose una variación del 76.2%.

Una vez evaluada la calidad del análisis factorial, mediante los criterios anteriormente descritos, y extraído el número de factores que van a sintetizar la cantidad de observación contenida en los datos, los pesos o cargas de los factores se obtuvieron mediante rotación Varimax, (estos miden la relación entre las variables originales y los factores (Cuadro A4)).

La rotación Varimax permitió caracterizar dos factores, incorporando aquellas variables cuya contribución supera el 50%. De este modo identificamos los siguientes factores

- ◆ El factor 1 podría caracterizar un indicador de desarrollo, y resumir el 45.5% de la variación total; permite diferenciar a las provincias cuya población es mayormente rural, con bajos niveles educativos, tasa global de fecundidad elevada, escasas fuentes de trabajo, en sectores no agrícolas.
- ◆ El factor 2 está relacionado con las condiciones de vida y necesidades básicas insatisfechas agrupa a las provincias con viviendas inadecuadas, con hacinamiento, sin servicios de agua y desagüe, y ubicados en zonas rurales.

Las provincias que alcanzaron puntajes más altos en el primer factor fueron: Cotabambas, Carlos Fitzcarrald, Chumbivilcas, Pachitea, Angaraes, Paruro, Cangallo, Acobamba, Vilcas Huaman y La Mar, las que en su mayoría están ubicadas en los departamentos de la sierra del Perú, son las mas alejadas y tal vez las menos atendidas dentro de cada departamento. En la tabla N° 01 se presenta las provincias menos desarrolladas del Perú según el análisis factorial.

Tabla N° 01. Jeraquización de las provincias del Perú de acuerdo al indicador de desarrollo

Provincias	Departamento
Cotabambas	Apurímac
Carlos F. Fitzcarrald	Ancash
Chumbivilcas	Cusco
Pachitea	Huánuco
Angaraes	Huancavelica
Paruro	Cusco
Cangallo	Ayacucho
Acobamba	Huancavelica
Vilcas Huamán	Ayacucho
La Mar	Ayacucho

Construcción del Indicador Global

Dado que el objetivo es de caracterizar a las provincias, es conveniente obtener un indicador global, que mida el desarrollo de las provincias, tomando en cuenta el nivel de educación, empleo, condiciones de salud, vivienda, tasa de fecundidad, etc. El indicador así obtenido permitirá una clasificación de las provincias según la combinación de las características mencionadas.

El indicador global es dado por [4], como la media ponderada de los factores principales,

$$\text{Indice} = \text{puntaje factor 1} * \sqrt{\text{autovalor 1}} + \text{puntaje factor 2} * \sqrt{\text{autovalor 2}}$$

$$\text{Indice} = \text{puntaje factor 1} * \sqrt{17.1} + \text{puntaje factor 2} * \sqrt{5.009}$$

Utilizando el indicador global, las provincias que alcanzaron puntajes más altos son las que tienen bajo nivel de desarrollo. Estas son Carlos Fitzcarrald, Paruro, Vilcas Huaman, Huacaybamba, Asunción, Acomaya, Víctor Fajardo, Marañón, Mariscal Luzuriaga y Cañas.

En la tabla N° 02, se presentan las 10 provincias con menor nivel de desarrollo.

Tabla N° 02. Provincias con menor nivel de desarrollo global

Orden	Provincias	Departamento
1	Carlos F. Fitzcarrald	Ancash
2	Paruro	Cusco
3	Vilcas Huáman	Ayacucho
4	Huacaybamba	Huanuco
5	Asunción	Ancash
6	Acomaya	Cusco
7	Víctor Fajardo	Ayacucho
8	Marañón	Huanuco
9	Mariscal Luzurriaga	Ancash
10	Cañas	Cusco

Las provincias que alcanzaron puntajes más bajos tienen alto nivel de desarrollo. Estas provincias fueron: Callao, Tacna, Tumbes, Lima, Arequipa, Ica, Chiclayo, Lambayeque, Ilo y Trujillo, las cuales están ubicadas en la costa y en su mayoría son capitales de departamento. En la tabla N° 3 se presentan las 10 provincias con mayor nivel de desarrollo.

Tabla N° 03. Provincias con mayor nivel de desarrollo global

Orden	Provincias	Departamento
189	Trujillo	Trujillo
190	Ilo	Ilo
191	Lambayeque	Lambayeque
192	Chiclayo	Lambayeque
193	Ica	Ica
194	Arequipa	Arequipa
195	Lima	Lima
196	Tumbes	Tumbes
197	Tacna	Tacna
198	Callao	Prov. Constituc. Callao

4. CONCLUSIONES

Es posible caracterizar las provincias del Perú, utilizando el análisis factorial, con el objetivo de resumir y sintetizar información contenida en una matriz de datos; de manera que, con una pérdida mínima información, se produzca una ganancia en interpretación. Fue posible deducir la matriz de datos de las provincias del Perú, de un espacio de 32 dimensiones (número de variables) a uno de dos dimensiones o factores, los cuales fueron interpretados como indicadores de desarrollo (factor 1) y de condiciones de vida y necesidades básicas insatisfechas (factor 2). Finalmente, fue posible la construcción de un indicador global de desarrollo a partir de los puntajes calculados para las provincias en cada uno de los factores.

Mediante el indicador global de desarrollo, la jerarquización de las provincias, considerando aquellas con bajo nivel de desarrollo fueron: Carlos Fitzcarrald, Paruro, VilcasHuaman Huacaybamba, Asunción, Acomaya, Victor Fajardo, Marañon, Mariscal Luzuriaga y Cañas. Las provincias con niveles aceptables de desarrollo fueron: Callao, Tacna, Tumbes, Lima, Arequipa, Ica, Chiclayo, Lambayeque, Ilo, y Trujillo.

BIBLIOGRAFÍA

- [1] Abascal y Grande. *Métodos multivariantes aplicados a la investigación comercial*. Editorial ESIC, (1985).
- [2] Comrey, A. *Manual de Análisis Factorial*. Academic Press, INC., (1985).
- [3] Jhonson, R; Wichern, D. *Applied Multivariate Statistical Analysis*. Prentice Hall INC, (1992).
- [4] Herrera Prieto. *Desarrollo Económico Municipal*, Universidad de Valladolid. (1995).
- [5] Mardia, K.; Kent, J.; Bibby, J. *Multivariate Analysis*. Academic Press, INC. (1979).

APENDICE

CUADRO A1. Medida de Adecuación KMO y Prueba de Esfericidad de Bartlett

Medida de adecuación Muestral (KMO) Kaiser-Meyer-Olkin	0.899
Prueba de esfericidad	
Aproximación chi-cuadrado	117.21.297
Grados de libertad	496
Nivel de Significación	.000

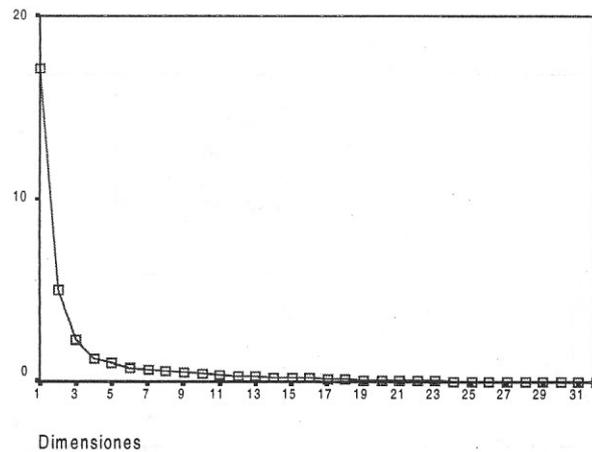
CUADRO A2.
Anti Imagen de la Matriz de Correlaciones – Elemento Diagonal

Variables	KMO
% de viviendas inadecuadas en zona rural / total de departamento	.911
% niños en zona rural que no asisten a la escuela / total de departamento	.808
Tasa de dependencia zona rural / total de departamento	.787
% de hogares con NBI en zona rural / total de departamento	.878
% de la población de la provincia / población del departamento	.881
% población con NBI% población con una NBI en el departamento	.848
% población urbana / total del departamento	.923
% de población con NBI en la zona urbana	.962
% población rural / población total del departamento	.930
% población rural / población urbana en el departamento	.935
Tasa de analfabetismo en población de 15 y más años	.836
Tasa de analfabetismo en mujeres de 15 y más años	.895
% población de 15 y más años con primaria incompleta	.917
% población femenina de 15 y más años con secundaria completa	.874
promedio de años de estudios aprobados población de 15 años	.896
% de niños de 6 a 12 años que no asisten a la escuela	.919
% de niños de 13 a 17 años que no asisten a la escuela	.687
% de niños de 9 a 15 años con atraso escolar	.920
Tasa de actividad de la población económicamente activa	.873
% de la población de 15 y más años ocupada con primaria incompleta	.931
% de la población de 15 y más años ocupada en la agricultura	.920
% de la población de 15 y más años ocupada en servicios	.928
% de la población de 15 y más años ocupada en establec. con 5 trabajadores	.922
% de la población de 15 y más años asalariada	.944
Tasa de crecimiento poblacional de 1981 a 1993	.717
% de la población de 15 y más años	.931
% de mujeres de 30 a 49 con más de cuatro hijos	.930
Tasa de mortalidad infantil	.928
% de niños en primero de primaria con desnutrición crónica	.981
% de hogares sin artefactos eléctricos	.943
% de jefe de hogar analfabetos	.858
% de jefes de hogar con primaria incompleta	.881

CUADRO A3.
Porcentajes de Variación Explicada por los Factores

Factor	Autovalor			Variación Explicada			Rotación		
	Total	%	%	Total	%	% acum.	Total	%	% acum.
1	17.1	53.488	53.488	17.1	53.488	53.488	14.6	45.502	45.502
2	5.009	15.654	69.142	5.009	15.654	69.142	7.062	22.069	67.571
3	2.256	7.050	76.192	2.256	7.050	76.192	2.759	8.621	76.192
4	1.257	3.928	80.121						
5	1.039	3.248	83.369						
6	.741	2.317	85.686						

GRAFICO A1. Autovalores Versus Número de Dimensiones



CUADRO A4. Peso de los Factores – Rotación Varimax

VARIABLES	KMO	Factor 2
% de viviendas inadecuadas en zona rural / total de departamento	-	.780
% niños en zona rural que no asisten a la escuela / total de departamento	-	.897
Tasa de dependencia zona rural / total de departamento	-	.903
% de hogares con NBI en zona rural / total de departamento	-	.939
% de la población de la provincia / población del departamento	-	.877
% población con NBI% población con una NBI en el departamento	-	.924
% población urbana / total del departamento	-	.761
% de población con NBI en la zona urbana	-	.682
% población rural / población total del departamento	-	.511
% población rural / población urbana en el departamento	.665	-
Tasa de analfabetismo en población de 15 y más años	.933	-
Tasa de analfabetismo en mujeres de 15 y más años	.937	-
% población de 15 y más años con primaria incompleta	.955	-
% población femenina de 15 y más años con secundaria completa	-.858	-
promedio de años de estudios aprobados población de 15 años	-.936	-
% de niños de 6 a 12 años que no asisten a la escuela	.630	-
% de niños de 13 a 17 años que no asisten a la escuela	-	-
% de niños de 9 a 15 años con atraso escolar	.864	-
Tasa de actividad de la población económicamente activa	-	-
% de la población de 15 y más años ocupada con primaria incompleta	.943	-
% de la población de 15 y más años ocupada en la agricultura	.812	-
% de la población de 15 y más años ocupada en servicios	-.784	-
% de la población de 15 y más años ocupada en establec. con 5 trabajadores	.785	-
% de la población de 15 y más años asalariada	-.756	-
Tasa de crecimiento poblacional de 1981 a 1993	-	-
% de la población de 15 y más años	.737	-
% de mujeres de 30 a 49 con más de cuatro hijos	.838	-
Tasa de mortalidad infantil	.783	-
% de niños en primero de primaria con desnutrición crónica	.833	-
% de hogares sin artefactos eléctricos	.720	-
% de jefe de hogar analfabetos	.905	-
% de jefes de hogar con primaria incompleta	.915	-

GRAFICO A2. Factor 1 Versus Factor 2

