

UNIVERSIDAD NACIONAL AGRARIA

LA MOLINA

FACULTAD DE ECONOMÍA Y PLANIFICACIÓN



**“MODELO DE FACTORES ASOCIADO AL DESARROLLO INFANTIL
TEMPRANO DE NIÑOS PERUANOS CON INDICADORES DEL
INSTITUTO NACIONAL DE ESTADISTICA E INFORMATICA”**

**TRABAJO DE SUFICIENCIA PROFESIONAL
PARA OPTAR EL TÍTULO DE
INGENIERO ESTADÍSTICO E INFORMÁTICO**

MIGUEL ALONSO GUZMÁN ALANYA

LIMA – PERÚ

2021

**La UNALM es titular de los derechos patrimoniales de la presente investigación
(Art. 24 - Reglamento de Propiedad Intelectual)**

**UNIVERSIDAD NACIONAL AGRARIA
LA MOLINA
FACULTAD DE ECONOMÍA Y PLANIFICACIÓN**

**“MODELO DE FACTORES ASOCIADO AL DESARROLLO INFANTIL
TEMPRANO DE NIÑOS PERUANOS CON INDICADORES DEL INSTITUTO
NACIONAL DE ESTADISTICA E INFORMATICA”**

**Presentado por:
MIGUEL ALONSO GUZMÁN ALANYA**

**TRABAJO DE SUFICIENCIA PROFESIONAL PARA OPTAR EL TÍTULO DE
INGENIERO ESTADÍSTICO E INFORMÁTICO**

SUSTENTADO Y APROBADO ANTE EL SIGUIENTE JURADO:

**Mg. Iván Dennys Soto Rodríguez
PRESIDENTE**

**M.A. Fernando René Rosas Villena
ASESOR**

**Dr. César Higinio Menacho Chiok
MIEMBRO**

**MS. Grimaldo José Febres Huamán
MIEMBRO**

LIMA – PERÚ

2021

DEDICATORIA

Dedicado a mi familia y a todas aquellas personas que siempre están apoyándome.

AGRADECIMIENTO

Agradezco a las personas que hicieron posible el desarrollo de este trabajo, personas de mi centro laboral, profesores y compañeros de mi carrera y en especial a mi asesor por ser la guía en esta parte de mi formación profesional.

ÍNDICE GENERAL

I.	INTRODUCCIÓN	8
1.1.	Problemática	8
1.2.	Objetivos	9
1.2.1.	Objetivo general	9
1.2.2.	Objetivos específicos	9
II.	MARCO TEÓRICO	10
2.1.	Antecedentes	10
2.2.	Metodología del análisis factorial	11
2.2.1.	Paso 1: Adecuación de los datos	12
a)	Objetivo	12
b)	Modelo	12
c)	Supuestos	13
d)	Justificación	14
2.2.2.	Paso 2: Extracción de factores	15
a)	Métodos de extracción de factores	15
b)	Criterios de extracción de factores	17
2.2.3.	Paso 3: Rotación de factores	18
a)	Matriz de factores no rotados	18
b)	Matriz de factores rotados	19
2.2.4.	Paso 4: Interpretación de factores	21
a)	Evaluación de factores	21
b)	Puntuaciones de los factores	22
III.	MARCO METODOLÓGICO	24
3.1.	Tipo de monografía	24
3.2.	Delimitación temporal y geográfica	24
3.3.	Hipótesis	24
3.4.	Variables	25
3.5.	Población y muestra	25
3.6.	Técnicas e instrumentos de recolección de datos	26
3.7.	Técnicas de procesamiento y análisis de datos	26
IV.	RESULTADOS Y DISCUSIÓN	27

V.	CONCLUSIONES Y RECOMENDACIONES	38
5.1.	Conclusiones	38
5.2.	Recomendaciones	38
VI.	REFERENCIAS BIBLIOGRÁFICAS	40
VII.	ANEXOS	42

ÍNDICE DE TABLAS

Tabla 1: Variables usadas en la metodología.....	25
Tabla 2: Matriz de correlaciones	28
Tabla 3: Prueba de KMO y Bartlett	29
Tabla 4: Comunalidades y extracción.....	29
Tabla 5: Matriz de componentes no rotados.....	30
Tabla 6: Autovalores y varianza explicada de factores no rotados.....	30
Tabla 7: Matriz de correlaciones según rotación oblicua	32
Tabla 8: Matriz de componentes según métodos ortogonales	33
Tabla 9: Autovalores y varianza explicada de factores rotados.....	34
Tabla 10: Puntuaciones de los Factores.....	36

ÍNDICE DE FIGURAS

Figura 1: Gráfico de sedimentación de los factores	31
Figura 2: Gráfico de componentes no rotados	32
Figura 3: Gráfico de componentes rotados.....	34

ÍNDICE DE ANEXOS

Anexo 1: Tabla de datos utilizados para el análisis factorial.....	42
Anexo 2: Sintaxis en SPSS del análisis factorial.....	43

RESUMEN

En el puesto de analista de procesamiento de información en el Ministerio de Desarrollo e Inclusión Social se realizaron las siguientes funciones: apoyar en el cálculo de indicadores, generar reportes con las bases de datos de los sectores relacionados con salud, educación y agua y asistir técnicamente al equipo del Fondo de Estímulo al Desempeño y Logro de Resultados Sociales. En los últimos cinco años, se desarrollaron diversas tareas que pudieron ser realizadas adecuadamente con los contenidos de los siguientes cursos: Estadística Aplicada, Técnicas de Muestreo, Bases de datos y Técnicas Multivariadas. En el Ministerio de Desarrollo e Inclusión Social, en el área del Fondo de Estímulo al Desempeño y Logro de Resultados Sociales, hasta el año 2019, no se contaba con una metodología capaz de reducir el gran número de indicadores relacionados con el Desarrollo Infantil Temprano, con la finalidad de ser usados posteriormente para el cálculo de los incentivos monetarios a los gobiernos regionales. En esta área se contaba con el apoyo de especialistas temáticos en salud, agua y educación, que explicaban la finalidad y utilidad de cada indicador, pero solo contaban con su criterio y experiencia para seleccionar los indicadores. Por lo tanto, evaluada la necesidad del área, se decidió a realizar un análisis factorial exploratorio, utilizando indicadores de agua, educación y salud publicados por el Instituto Nacional de Estadística e Informática relacionado con el Desarrollo Infantil Temprano, el cual permitió obtener un número reducido de indicadores con sustento técnico. El trabajo monográfico se encuentra dividido en seis partes, una vez presentado la introducción se pasará al marco teórico, seguido del marco metodológico, posteriormente, los resultados y discusión, luego las conclusiones y recomendaciones y finalmente las referencias bibliográficas.

Palabras clave: análisis factorial, componentes, factores, rotación, varimax.

ABSTRACT

In the position of information processing analyst in the Ministry of Development and Social Inclusion, the following functions were performed: support in the calculation of indicators, generate reports with the databases of health-related sectors, education and water, and provide technical assistance to the Performance Stimulation and Achievement of Social Results Fund. In the last five years, many tasks were developed that could be properly performed with the contents of the following courses: Applied Statistics, Sampling Techniques, Databases and Multivariate Techniques. In the Ministry of Development and Social Inclusion, in the area Performance Stimulation and Achievement of Social Results Fund, until 2019, there was no methodology capable of reducing the large number of indicators related to Early Childhood Development, in order to be used subsequently for the calculation of monetary incentives to regional governments. In this area, thematic specialists in health, water and education were available to explain the purpose and usefulness of each indicator, but they only had their own criteria and experience in selecting indicators. Therefore, assessing the need for the area, it was decided to carry out an exploratory factor analysis, using water, education and health indicators published by the National Institute of Statistics and Information related to Early Childhood Development, which produced a small number of indicators with a technical basis. The monographic work is divided into six parts, once presented the introduction will be passed to the theoretical framework, followed by the methodological framework, subsequently, the results and discussion, then the conclusions and recommendations and finally the bibliographic references.

Keywords: factor analysis, components, factors, rotation, varimax.

I. INTRODUCCIÓN

1.1. Problemática

El Fondo de Estímulo al Desempeño y Logro de Resultados Sociales es un mecanismo de incentivos para impulsar el logro de resultados en materia de Desarrollo Infantil Temprano, por medio de Convenios de Asignación por Desempeño entre el Ministerio de Desarrollo e Inclusión Social, Ministerio de Economía y Finanzas y cada Gobierno Regional, donde se establecen las metas a alcanzarse y se indica el monto máximo a ser transferido por cumplirse con las metas.

Inicialmente, los desembolsos de incentivos monetarios a los gobiernos regionales se llevaban a cabo considerándose únicamente el cumplimiento de un solo indicador: porcentaje de niños menores de 5 años que cuentan con anemia. Este procedimiento fue observado por los directores del área de Fondo de Estímulo al Desempeño y Logro de Resultados Sociales, porque no contemplaba los requerimientos reales de cada región. Para realizar una mejor distribución de los incentivos monetarios recomendaron utilizar todos los indicadores que pudieran estar relacionados con el Desarrollo Infantil Temprano. En opinión de los especialistas de salud, agua y educación, el número de indicadores asociados con el Desarrollo Infantil Temprano eran tantos que su incorporación podría perjudicar más la forma de distribuir los montos a cada región. Por esta razón, surgió la necesidad de encontrar una metodología estadística que pudiera reducir el número de indicadores a uno más pequeño, con la finalidad de optimizar la asignación de los incentivos monetarios.

Se propuso la técnica estadística multivariada del Análisis Factorial Exploratorio, el cual permitiría obtener un número reducido de dimensiones asociados con el Desarrollo Infantil Temprano. Estas dimensiones o factores serían capaces de explicar al máximo la información subyacente contenida en los datos.

1.2. Objetivos

1.2.1. Objetivo general

Identificar el número de dimensiones que forman parte de un modelo de factores asociado al desarrollo infantil temprano de niños peruanos considerando como variables los indicadores del Instituto Nacional de Estadística e Informática.

1.2.2. Objetivos específicos

- Identificar los indicadores asociados a las dimensiones del modelo factorial del desarrollo infantil temprano de niños peruanos.
- Calcular el porcentaje de variancia acumulada que optimiza la solución factorial del modelo.
- Identificar los gobiernos regionales que tienen valores atípicos para cada una de las dimensiones del modelo factorial.

II. MARCO TEÓRICO

2.1. Antecedentes

El Análisis Factorial Exploratorio es una técnica estadística multivariada que permite obtener de una gran cantidad de variables un número más pequeño de variables denominado factores.

Avalo (2012) indica que el Análisis Factorial Exploratorio tiene sus orígenes con los estudios de Galton en el año 1889, quien expuso el concepto de rasgo latente para explicar cómo un conjunto de variables se encuentra relacionadas. La posibilidad de relación de dos variables, basada en que ambas tienen algo en común y algo que las diferencia.

Pearson (1901) desarrolla los principios básicos del Análisis Factorial Exploratorio, aportando con sus resultados de investigación el coeficiente de correlación y aproximaciones de principios sobre lo que se apoya el análisis factorial de ejes principales.

López y Fachelli (2015) señalan que el análisis factorial tiene sus orígenes en el siglo XIX, a partir de los trabajos de Spearman en 1904 donde adopta su primera formulación, vinculándolo principalmente con los trabajos realizados en el campo de la psicología ante la problemática de medir los factores de la inteligencia humana.

Hotelling (1933) planteó los principios metodológicos y matemáticos para el Análisis Factorial de Componentes Principales, donde no se realizaba la distinción entre factores comunes y específicos, cada factor finalmente explicará una parte del total de la varianza inicial aportada por las variables introducidas.

Burt (1940) y Thurstone (1947) perfeccionan el método del Análisis Factorial de Factores Principales por medio de un modelo lineal y a partir de un grupo extenso de variables iniciales permitía la obtención de un grupo de factores comunes y otros factores específicos que incluían las características propias de cada variable y un error aleatorio.

Kaiser (1958) introdujo el procedimiento de la rotación varimax que implicaba la rotación ortogonal, con el objetivo de ganar interpretabilidad en los factores obtenidos.

Benzecri (1973) desarrolló la extensión de estos procedimientos que dio lugar al Análisis Factorial de Correspondencias.

Pérez (2004) indica que la técnica estadística del Análisis Factorial permite analizar la dimensionalidad latente en un conjunto de n variables observables, expresada a través de unos factores comunes, siguiendo unos criterios de estructura simple, tomando como información principal la matriz de correlaciones. Esta forma de análisis ha predominado hasta los años sesenta, bajo la influencia de Thurstone, es conocida con el nombre de Análisis Factorial Exploratorio.

López y Fachelli (2015) mencionan que el Análisis Factorial Exploratorio no busca identificar los factores a partir de un modelo prefijado o cerrado que se impone en los datos, sino a partir de la tarea interpretativa y atribuyendo un significado a posteriori a los factores.

López y Gutiérrez (2019) formulan que el Análisis Factorial Exploratorio tiene como finalidad hallar la estructura subyacente de un grupo de datos cuantitativos definiendo un pequeño número de dimensiones latentes comunes que expliquen la mayor parte de la varianza observada en un grupo más amplio de variables.

2.2. Metodología del análisis factorial

López y Fachelli (2015) proponen una metodología del Análisis Factorial que contiene cuatro pasos: adecuación de los datos, la extracción de factores o componentes, la rotación de los factores y por último la interpretación de los factores.

2.2.1. Paso 1: Adecuación de los datos

En el primer paso de la adecuación de los datos se toman en cuenta cuatro aspectos importantes: objetivo, modelo, supuestos y justificación.

a) Objetivo

Pérez (2004) indica que el análisis factorial tiene como objeto simplificar las múltiples y complejas relaciones que puedan existir entre un conjunto de variables observadas $X_1, X_2 \dots X_p$, mediante factores que vinculen a las aparentemente no relacionadas variables. En otras palabras, trata de detectar un conjunto de $k < p$ factores no directamente observables $F_1, F_2 \dots F_k$ que expliquen lo suficiente las variables observadas perdiendo el mínimo de información, de manera que sean fácilmente interpretables (Principio de interpretabilidad) y que sean los menos posibles, en otras palabras, un k pequeño (Principio de parsimonia). Además, los factores extraídos deben resultar independientes entre sí, es decir, ortogonales.

El aspecto más característico del análisis factorial lo constituye su capacidad de reducción de datos. Las relaciones entre las variables observadas $X_1, X_2 \dots X_p$ vienen dadas por su matriz de correlaciones, de modo que se puede partir de una serie de coeficientes de correlación para el conjunto de variables observadas y estudiar si subyace algún patrón de relaciones tal que los datos puedan ser reordenados a un conjunto menor de factores que podemos considerar como variables que recogen y resumen las interrelaciones observadas en los datos.

Por lo tanto, el Análisis Factorial se caracteriza por encontrar grupos homogéneos de variables a partir de un conjunto numeroso de variables, con el propósito de encontrar la manera de resumir la información contenida en el conjunto de variables originales en una pequeña serie de dimensiones, capaz de explicar el máximo de información.

b) Modelo

Pérez (2004) menciona que las variables observables $X_1, X_2 \dots X_p$, se define en el modelo factorial de la siguiente forma:

$$\begin{aligned}
X_1 &= a_{11}F_1 + a_{12}F_2 + \dots + a_{1m}F_m + e_1 \\
X_2 &= a_{21}F_1 + a_{22}F_2 + \dots + a_{2m}F_m + e_2 \\
&\vdots \\
X_p &= a_{p1}F_1 + a_{p2}F_2 + \dots + a_{pm}F_m + e_p
\end{aligned}$$

Donde $F_1, F_2 \dots F_m$ son los factores comunes; $e_1, e_2 \dots e_p$ son los factores únicos o factores específicos y a_{pm} es el peso del factor m en la variable p, denominado carga factorial o saturación de la variable p en el factor m. Según la expresión del modelo, cada una de las p variables observables es una combinación lineal de m factores comunes a todas las variables ($m < p$) y de un factor único para cada variable. En consecuencia, todas las variables originales están influenciadas por todos los factores comunes, entre tanto que para cada variable existe un factor único que es específico para esa variable. Tanto los factores comunes como los factores específicos son variables no observables. El modelo factorial en forma matricial se expresaría:

$$\begin{array}{ccccccc}
X_1 & & a_{11} & a_{12} & & F_1 & e \\
\begin{bmatrix} X_2 \\ \vdots \\ X_p \end{bmatrix} & = & \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pm} \end{bmatrix} & \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_m \end{bmatrix} & + & \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_p \end{bmatrix}
\end{array}$$

o que es lo mismo:

$$X = aF + e$$

c) Supuestos

Chávez (2017) señala que se pueden obviar los supuestos de normalidad y homocedasticidad, siendo conscientes de que su incumplimiento produce una reducción en las correlaciones observadas. De hecho, sólo es necesaria la normalidad cuando se aplica una prueba estadística a la significancia de los factores; pero raramente se utilizan estas pruebas. En realidad, es deseable que haya cierto grado de multicolinealidad.

Si al observar la matriz de correlaciones no notamos un número sustancial de valores mayores que 0.30 entonces el Análisis Factorial es probablemente inapropiado.

d) Justificación

Existen varios indicadores que permiten evaluar si se debe proseguir con el Análisis Factorial:

- Test de Bartlett

Es una prueba estadística para homogeneidad de varianzas, es decir, proporciona la probabilidad estadística de que la matriz de correlación de las variables que estamos ingresando al modelo sea una matriz identidad.

$$H_0: R = I$$

$$H_1: R \neq I$$

La expresión del estadístico de la prueba de Bartlett está basada en una transformación del determinante de la matriz de correlaciones R y sigue una distribución chi-cuadrado, con la siguiente fórmula, a partir de n casos y p variables:

$$X^2 = - \left(n - 1 - \frac{2p + 5}{6} \right) \cdot \ln|R|$$

Con $v = \frac{p^2 - p}{2}$ grados de libertad.

Si se rechaza la hipótesis nula de independencia o de esfericidad, el Análisis Factorial es aplicable.

- Medida de suficiencia de la muestra Kaiser-Meyer-Oikin (KMO)

Índice que se encuentra entre de 0 y 1, llegando a 1 cuando cada variable es perfectamente predicha sin error por las otras variables.

$$KMO = \frac{\sum_{i \neq j}^n r_{ij}^2}{\sum_{i \neq j}^n r_{ij}^2 + \sum_{i \neq j}^n d_{tag}^2}$$

Donde:

r_{ij} son dos variables diferentes, r_{ij} es el coeficiente de correlación simple y $diag$ es la correlación parcial.

Según los sugerentes calificativos de Kaiser (1974) podemos establecer una escala con la siguiente valoración:

- 0.9: Maravilloso
- 0.8: Meritorio
- 0.7: Intermedio
- 0.6: Mediocre
- 0.5: Miserable
- Menos 0.5: Inaceptable

2.2.2. Paso 2: Extracción de factores

En el segundo paso se considera dos aspectos importantes: métodos de extracción de factores y criterios de extracción de factores.

a) Métodos de extracción de factores

Existen varios métodos de extracción de factores como lo son: el método de Thurstone, método del factor principal, método Alpha, método del centroide, método de los componentes principales, método de componentes principales iteradas o ejes principales, método de máxima verosimilitud y el método de MINRES, ULS y GLS.

En este trabajo se utilizó el método de los componentes principales, Pérez (2004) indica que al tener una muestra de tamaño n acerca de p variables $X_1, X_2 \dots X_p$ inicialmente correlacionadas, para posteriormente adquirir a partir de ellas un número $k \leq p$ de variables incorrelacionadas $Z_1, Z_2 \dots Z_p$ que sean combinación lineal de las variables iniciales y que expliquen la mayor parte de su variabilidad, tendremos:

$$Z_1 = u_{11}X_1 + u_{12}X_2 + \dots + u_{1p}X_p$$

$$Z_2 = u_{21}X_1 + u_{22}X_2 + \dots + u_{2p}X_p$$

⋮

$$Z_p = u_{p1}X_1 + u_{p2}X_2 + \dots + u_{pp}X_p$$

Pero este sistema de ecuaciones es reversible, siendo posible expresar las variables X_j en función de los componentes principales Z_j de la siguiente forma:

$$X_1 = u_{11}Z_1 + u_{12}Z_2 + \dots + u_{1p}Z_p$$

$$X_2 = u_{21}Z_1 + u_{22}Z_2 + \dots + u_{2p}Z_p$$

⋮

$$X_p = u_{p1}Z_1 + u_{p2}Z_2 + \dots + u_{pp}Z_p$$

La matriz de coeficientes de este segundo sistema es la matriz transpuesta de la matriz de coeficientes del sistema anterior, pudiendo utilizarse este segundo sistema para la estimación de los factores. El único problema que podría presentarse es que las componentes Z_j no estén tipificadas, condición que sí se ha exigido a los factores. Este problema se salva utilizando componentes principales tipificadas, definidas por:

$$Y_j = \frac{Z_j}{\sqrt{\lambda_j}} \quad j=1, 2, \dots, p$$

Entonces, en el segundo sistema sustituimos los Z_j por $Y_j\sqrt{\lambda_j}$, resultando la ecuación j -ésima del sistema de la siguiente forma:

$$X_j = u_{1j}Y_1\sqrt{\lambda_1} + u_{2j}Y_2\sqrt{\lambda_2} + \dots + u_{pj}Y_p\sqrt{\lambda_p}$$

Pero de la teoría de componentes principales sabemos que $u_{hj}\sqrt{\lambda_h}$ es el coeficiente de correlación entre la variable j -ésima y la componente h -ésima, lo que permite escribir la ecuación como:

$$X_j = r_{1j}Y_1 + r_{2j}Y_2 + \dots + r_{pj}Y_p$$

pudiéndose separar en esta última ecuación sus últimos p-k términos, lo que permite escribirla como:

$$X_j = r_{1j}Y_1 + r_{2j}Y_2 + \dots + r_{kj}Y_k + (r_{k+1,j}Y_{k+1} + \dots + r_{pj}Y_p)$$

Comparando esta ecuación con la ecuación del modelo factorial:

$$X_j = l_{j1}F_1 + l_{j2}F_2 + \dots + l_{jk}F_k + e_j$$

Se observa que los k factores F_h se estiman mediante los k primeras componentes principales tipificadas Y_h y la estimación de los coeficientes l_{jh} viene dada por:

$$\hat{l}_{j1} = r_{1j}, \hat{l}_{j2} = r_{2j}, \dots, \hat{l}_{jk} = r_{kj}$$

Pudiéndose estimar la comunalidad de la variable X_j como:

$$\hat{h}_j^2 = \hat{l}_{j1}^2 + \hat{l}_{j2}^2 + \dots + \hat{l}_{jk}^2$$

y el factor único e_j se estimará como:

$$e_j = r_{k+1,j}Y_{k+1} + r_{m+2,j}Y_{k+1} + \dots + r_{pj}Y_p$$

y la especificidad o parte de la varianza debida al factor único se estima como:

$$\hat{\omega}_j^2 = 1 - \hat{h}_j^2$$

b) Criterios de extracción de factores

Según López y Fachelli (2015), existen cuatro criterios en un análisis factorial:

- a. Considerar todos aquellos factores que tienen un valor propio superior a 1, pues supone considerar un factor que mejora la varianza proporcionada en un inicio para cada variable sola.
- b. Considerar el número de ejes que acumulan en torno al 70% de la varianza total, cantidad que se considera equilibrada entre la pérdida de información (del 30%) y la ganancia en significación (el 70% retiene los principales factores de variabilidad).
- c. Representar gráficamente los distintos factores y los valores propios asociados y observar el comportamiento de la curva resultante (gráfico de sedimentación). El número de ejes a retener viene determinado por el cambio de pendiente de la curva, donde está presente el cambio de continuidad de la curva, llamado, scree test, en términos más coloquiales “test del codo”, donde se sitúa el codo del brazo imaginario que dibuja la forma de la curva es el punto que determina el número de componentes.
- d. El cuarto criterio es el razonamiento, la interpretabilidad y la pertinencia substantiva de los ejes obtenidos que en un momento pueden llevarnos a considerar más o menos en función del propio contenido de los factores y sus implicaciones en el análisis del fenómeno.

2.2.3. Paso 3: Rotación de factores

En el tercer paso se considera dos aspectos importantes: matriz de factores no rotados y matriz de factores no rotados.

a) Matriz de factores no rotados

Pérez (2004) indica que al realizar la representación geométrica del modelo factorial, puede pasar que las proyecciones de la mayoría de las variables sobre los factores no sean lo suficientemente grandes como para que la interpretación del modelo resulte adecuada.

Si la representación geométrica resulta difusa, se puede realizar una rotación de los factores que clarifique las proyecciones de las variables sobre ellos.

b) Matriz de factores rotados

Pérez (2004) menciona que los factores comunes deben tener una interpretación clara, de esa forma, analizar mejor las interrelaciones existentes entre las variables originales. Sin embargo, en muy pocas ocasiones resulta fácil encontrar una interpretación adecuada de los factores iniciales, independientemente del método que se haya utilizado para su extracción. Por ello, los procedimientos de rotación han sido ideados para obtener factores que sean fácilmente interpretables, cada una de las variables originales tendrá una correlación lo más próxima a 1 con uno de los factores y correlaciones próximas a 0 con el resto de los factores. De esta forma, cada factor tendrá una correlación alta con un grupo de variables y baja con el resto de variables.

Existen dos formas básicas de realizar la rotación de factores: rotación ortogonal y rotación oblicua. En la rotación ortogonal, los ejes se rotan de forma que no exista correlación entre los factores, los nuevos ejes son perpendiculares de igual forma que lo son los factores sin rotar, para ello se tienen los métodos Varimax, Equamax y Quartimax.

Pérez (2004) indica que en la rotación oblicua los ejes no son ortogonales y los factores tendrán cierto grado de correlación, perdiendo una propiedad que en un principio es deseable que cumplan los factores. Sin embargo, en ocasiones puede compensarse esta pérdida, si se consigue una mejor asociación de cada una de las variables con el factor correspondiente. Los métodos de rotación oblicua más conocidos son: Oblimin, Oblimax, Promax, Quartimin, Biquartimin y Covarimin.

López y Fachelli (2015) indican que el procedimiento de rotación oblicua y sus procedimientos particulares son más útiles con modelos previos, es decir, en una lógica de análisis factorial confirmatorio.

Para esta investigación se utilizó el método de Varimax, según Pérez (2004), este método obtiene los ejes de los factores maximizando la suma de varianzas de las cargas factoriales al cuadrado dentro de cada factor. Se define la simplicidad de un factor por la varianza de los cuadrados de sus cargas factoriales en las variables observables. La simplicidad S_i^2 del factor F_i será por lo tanto:

$$S_i = \frac{1}{p} \sum_{j=1}^p (l_{ji})^2 - \left(\frac{1}{p} \sum_{j=1}^p l_{ji}^2 \right)^2$$

La metodología de rotación Varimax procura hallar $B=LT$ de modo que la suma de las simplicidades de todos los factores sea máxima, lo que trae consigo la maximización de:

$$S = \sum_{i=1}^k S_i = \sum_{i=1}^k \left[\frac{1}{p} \sum_{j=1}^p (l_{ji})^2 - \left(\frac{1}{p} \sum_{j=1}^p l_{ji}^2 \right)^2 \right]$$

El problema que propone la expresión anterior es que las variables con mayores comunalidades poseen una mayor influencia en la solución final. Para resolver este problema se realiza la normalización de Kaiser, en la que cada carga factorial al cuadrado se divide por la comunalidad de la variable correspondiente. La función a maximizar será ahora:

$$SN^2 = \sum_{i=1}^k \left[\frac{1}{p} \sum_{j=1}^p \left(\frac{l_{ji}^2}{h_j^2} \right)^2 - \left(\frac{1}{p} \sum_{j=1}^p \frac{l_{ji}^2}{h_j^2} \right)^2 \right]$$

En su forma definitiva, la metodología Varimax halla la matriz B maximizando:

$$W = p^2 SN^2 = p \sum_{i=1}^k \sum_{j=1}^p \left(\frac{l_{ji}^2}{h_j^2} \right)^2 - \sum_{i=1}^k \left(\sum_{j=1}^p \frac{l_{ji}^2}{h_j^2} \right)^2$$

Para realizar la maximización se calcula la matriz $T = \begin{pmatrix} \cos(\varphi) & \text{sen}(\varphi) \\ -\text{sen}(\varphi) & \cos(\varphi) \end{pmatrix}$ que

realiza la rotación de factores de forma que su suma de simplicidades sea máxima.

Repetiendo esto para los $p(p-1)/2$ pares posibles de factores, se tiene:

$$B = LT_{11}T_{12}T_{13} \dots T_{m-1,m}$$

Cuando la rotación es de más de dos factores se efectúa un procedimiento iterativo. El primer y segundo factor se giran según el ángulo φ definido por el procedimiento anterior. El nuevo primer factor se gira con el tercer factor, y se sigue así hasta que todos los $k(k-1)/2$ pares de factores hayan sido girados. Esta sucesión de rotaciones se denomina ciclo, se itera los ciclos hasta completar uno en que todos los ángulos de giro sean menores a un cierto valor prefijado.

Una propiedad importante del método Varimax es que una vez aplicado ésta queda inalterada, tanto la varianza total explicada por los factores, como la comunalidad de las variables. La nueva matriz corresponde también a factores ortogonales y tiende a simplificar la matriz factorial por columnas, siendo muy apropiado cuando el número de factores es pequeño.

2.2.4. Paso 4: Interpretación de factores

En este paso se considera dos aspectos importantes: la evaluación de factores y la puntuación de los factores.

a) Evaluación de factores

En la evaluación de los factores se determinará la matriz de componentes con la cual trabajaremos: la matriz de componentes no rotados o la matriz de componentes rotados, aquella donde se vea una estructura simple con variables que estén relacionados con un único factor y a su vez factores que tengan un número reducido de variables.

López y Fachelli (2015) indican que la interpretación puede hacerse a partir de la matriz de componentes originales o también por medio de la matriz de componentes rotados.

De la Fuente (2011) menciona que la matriz de componentes factoriales debe reunir

tres características:

- En cada factor deben existir puntuaciones altas y bajas.
- Cada variable no debe estar relacionada con más de un factor.
- No deben existir factores con la misma distribución, es decir, dos factores distintos deben presentar diferentes puntuaciones.

Lloret, Ferreres, Hernández y Tomás (2014) indican que si se va a utilizar la matriz rotada es aconsejable utilizar rotación oblicua debido a que casi todos los fenómenos en las ciencias sociales y de la salud están más o menos interrelacionados entre sí, pero si las correlaciones entre factores fueran consistentemente bajas se propone repetir el análisis utilizando una solución ortogonal.

b) Puntuaciones de los factores

Pérez (2004) afirma que el análisis factorial es en muchas situaciones un paso previo a otros análisis, seleccionando los factores obtenidos en vez de las variables originales. Por ello, es necesario calcular estas puntuaciones.

Los procedimientos más conocidos según Pérez (2004) son:

- **Regresión:** con este método las puntuaciones tienen media cero y una varianza igual al cuadrado de la correlación múltiple entre las puntuaciones factoriales estimadas y los valores factoriales verdaderos. Estas puntuaciones pueden correlacionarse incluso si los factores son ortogonales.
- **Barlett:** en este método las puntuaciones tienen media cero y se minimiza la suma de cuadrados de los factores únicos sobre el rango de las variables.
- **Anderson-Rubin:** este método asegura la ortogonalidad de los factores estimados, teniendo las puntuaciones resultantes una media cero, una desviación típica de uno y no correlacionan entre sí.

En esta investigación se obtuvo las puntuaciones mediante la estimación de regresión.

Pérez (2004) indica que considerándose la regresión múltiple del factor F_i sobre las

variables X_1, \dots, X_p :

$$\hat{F}_i = \hat{\beta}_1 X_1 + \dots + \hat{\beta}_p X_p = \hat{\beta}_i' X$$

F_i verifica que $E[(F_i - \hat{F}_i)^2]$ es mínimo, y los coeficientes $\hat{\beta}$ se obtienen de la relación $\hat{\beta}_i = R^{-1} \delta_i$ siendo δ el vector columna con las correlaciones entre el factor F_i y las variables X . Estimando F_i mediante \hat{F}_i tendremos:

$$\hat{F}_i = \delta_i' R^{-1} X$$

Y considerando los m factores comunes tendremos

$$f = SR^{-1}x$$

Siendo $S=LT$, las columnas de T contienen las cargas factoriales de los factores oblicuos respecto a los ortogonales, la matriz de la estructura factorial. En el caso de los factores ortogonales $S=L$ y tenemos

$$f = L'R^{-1}x$$

III. MARCO METODOLÓGICO

3.1. Tipo de monografía

El tipo de monografía utilizado es de experiencias, porque a través de ella se trata de analizar una experiencia en la solución de una problemática laboral mediante la aplicación de una técnica estadística. El tipo de investigación utilizado es aplicado porque se sirve de la investigación básica, pura o fundamental a través de las ciencias fácticas o formales para generar problemas o hipótesis.

3.2. Delimitación temporal y geográfica

La aplicación de la técnica estadística del Análisis Factorial Exploratorio se realizó en el año 2018.

La información utilizada en el modelo de factores asociado al desarrollo infantil temprano de niños peruanos se recolectó de los 25 departamentos del Perú. Las variables asociadas a los factores considerados en el modelo factorial propuesto se describen en la Tabla 1.

3.3. Hipótesis

El modelo de factores asociado al desarrollo infantil temprano de niños peruanos está conformado por los factores: características y servicios básicos del hogar e indicadores de salud infantil y acceso a la educación

3.4. Variables

Tabla 1: Variables usadas en la metodología

Variables	Sub variables	Escala de medición
Porcentaje de niñas y niños de 6 a 35 meses de edad con prevalencia de anemia.	Proporción de niños y niñas menores de 60 meses de edad que acceden a agua clorada para consumo humano (cloro residual en muestra de agua de consumo ≥ 0.5 mg/l).	Porcentaje
	Dispersión: % de población que vive en centros poblados con menos de 400 viviendas.	Porcentaje
	Porcentaje de menores de 5 años con desnutrición crónica.	Porcentaje
	Tasa de Pobreza Total.	Porcentaje
	Porcentaje de nacidos en los últimos 5 años anteriores a la encuesta con bajo peso al nacer (< 2.5 kg).	Porcentaje
	Porcentaje de viviendas particulares que consume agua potable proveniente de red pública.	Porcentaje
	Porcentaje de viviendas particulares con ocupantes presentes que tienen tipo de servicio higiénico adecuado (sistema de alcantarillado por red pública o letrina/tanque séptico).	Porcentaje
	Tasa de cobertura del ciclo II de la educación básica regular (EBR), para niños y niñas de 3 años de edad en distritos de quintiles de pobreza 1 y 2 del departamento.	Porcentaje
	Prevalencia de diarrea en niñas y niños menores de cinco años de edad durante las dos semanas que precedieron la encuesta.	Porcentaje
	Porcentaje de viviendas particulares que no cuentan como tipo de piso predominante la tierra.	Porcentaje

FUENTE: Elaboración propia

3.5. Población y muestra

En el caso de la Encuesta Demográfica y de Salud Familiar (ENDES) el cuestionario se aplica de forma individual a las mujeres de 15 a 49 años de edad para y en el caso de la Encuesta Nacional de Hogares y el Censo Nacional de Población y vivienda al jefe del hogar.

La población de la investigación fueron todos los departamentos del Perú.

La muestra no aplica en esta investigación.

3.6. Técnicas e instrumentos de recolección de datos

La técnica de recolección de información fue la entrevista y el instrumento de recolección de información fue la encuesta, elaborados por el Instituto Nacional de Estadística e Informática relacionados con los indicadores socio demográficos, salud y hogares del año 2018 y el censo nacional de población y vivienda del año 2017.

3.7. Técnicas de procesamiento y análisis de datos

La técnica de procesamiento de datos fue realizada con el programa IBM SPSS versión 21 y el análisis de datos se realizó con la técnica estadística multivariada Análisis Factorial Exploratorio.

IV. RESULTADOS Y DISCUSIÓN

Paso 1: Adecuación de los datos

En este paso se consideran cuatro aspectos importantes: objetivo, modelo, supuestos y justificación.

Objetivos:

El objetivo principal de la aplicación del Análisis Factorial fue la reducción de los indicadores asociados al Desarrollo Infantil Temprano que se tenía en el año 2019.

Modelo:

El modelo de Análisis Factorial contó con 11 indicadores y 25 observaciones (número de gobiernos regionales) con escalas de medición en porcentaje:

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_{11} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{1m} \\ a_{21} & a_{22} & a_{2m} \\ \vdots & \vdots & \vdots \\ a_{11,1} & a_{11,2} & F_m \\ & a_{11,m} & \end{bmatrix} \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_m \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_{11} \end{bmatrix}$$

Donde $X_1, X_2 \dots X_{11}$ son los indicadores relacionados con el Desarrollo Infantil Temprano, $F_1, F_2 \dots F_m$ los factores obtenidos; $e_1, e_2 \dots e_p$ los factores específicos y $a_{11,m}$ el peso del factor.

Supuestos:

No se realizaron pruebas de Normalidad y Homocedasticidad en los datos, solo se observó la matriz de correlaciones, la cual está en la Tabla 2.

En la Tabla 2 se observa que la mayoría de las correlaciones son mayores a 0.30.

Tabla 2: Matriz de correlaciones

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11
X1	1	-0.308	-0.623	-0.720	-0.647	-0.581	-0.680	-0.647	0.061	-0.247	-0.627
X2	-0.308	1	0.331	0.400	0.326	0.265	0.512	0.590	0.163	0.494	0.283
X3	-0.623	0.331	1	0.866	0.810	0.266	0.816	0.706	-0.302	0.238	0.899
X4	-0.720	0.400	0.866	1	0.824	0.524	0.848	0.721	-0.119	0.403	0.753
X5	-0.647	0.326	0.810	0.824	1	0.478	0.710	0.697	-0.170	0.347	0.793
X6	-0.581	0.265	0.266	0.524	0.478	1	0.524	0.453	0.113	0.309	0.075
X7	-0.680	0.512	0.816	0.848	0.710	0.524	1	0.836	-0.096	0.335	0.685
X8	-0.647	0.590	0.706	0.721	0.697	0.453	0.836	1	0.020	0.428	0.685
X9	0.061	0.163	-0.302	-0.119	-0.170	0.113	-0.096	0.020	1	0.219	-0.309
X10	-0.247	0.494	0.238	0.403	0.347	0.309	0.335	0.428	0.219	1	0.198
X11	-0.627	0.283	0.899	0.753	0.793	0.075	0.685	0.685	-0.309	0.198	1

El indicador agua clorada tiene una relación significativa con los indicadores dispersión, desnutrición crónica, pobreza, bajo peso, agua potable, servicio higiénico y piso tierra; el indicador anemia tiene una relación significativa con los indicadores agua potable y servicio higiénico; el indicador dispersión tiene una relación significativa con los indicadores agua clorada, desnutrición crónica, pobreza, agua potable, servicio higiénico y piso tierra; el indicador desnutrición crónica tiene una relación significativa con los indicadores agua clorada, dispersión, pobreza, bajo peso, agua potable, servicio higiénico y piso tierra; el indicador pobreza tiene una relación significativa con los indicadores agua clorada, dispersión, desnutrición crónica, agua potable, servicio higiénico y piso tierra; el indicador bajo peso tiene una relación significativa con los indicadores agua clorada, desnutrición crónica y agua potable; el indicador agua potable tiene una relación significativa con los indicadores agua clorada, anemia, dispersión, desnutrición crónica, pobreza, bajo peso, servicio higiénico y piso tierra, el indicador servicio higiénico tiene una relación significativa con los indicadores agua clorada, anemia, dispersión, desnutrición crónica, pobreza, agua potable y piso tierra; el indicador piso tierra tiene una relación significativa con los indicadores agua clorada, dispersión, desnutrición crónica, pobreza, agua potable y servicio higiénico; y finalmente los indicadores matrícula y diarrea no están relacionadas con ninguna variable.

Justificación:

En la Tabla 3 se presentan las pruebas Barlett y Kaiser-Meyer-Olkin. El valor KMO = 0.809, permite afirmar que el Análisis Factorial es válida. El Test de Barlett, con $\alpha = 0.0$, se rechaza la hipótesis de que la matriz de correlación de las variables es de identidad, por lo tanto, existe correlación entre los indicadores asociados al Desarrollo Infantil Temprano.

Tabla 3: Prueba de KMO y Bartlett

Medida Kaiser-Meyer-Olkin de adecuación de muestreo		.809
	Aprox. Chi-cuadrado	215.581
Prueba de esfericidad de Bartlett	Gl	55
	Sig.	.000

Paso 2: Extracción de Factores

Se consideran los métodos de extracción de factores y criterios de extracción de factores.

Métodos de extracción de factores:

De la Tabla 4 se observa que el indicador que explica mejor el modelo factorial es la dispersión (0.905), seguido de la desnutrición crónica (0.868) y los hogares que no cuentan como tipo de piso predominante la tierra (0.853).

Tabla 4: Comunalidades y extracción

	Inicial	Extracción
agua_clorada	1	0.633
no_anemia	1	0.532
no_dispersion	1	0.905
no_dci	1	0.868
no_pobreza	1	0.792
no_bajo_peso	1	0.461
agua_potable	1	0.834
servicio_higienico	1	0.791
Matricula	1	0.609
no_diarrea	1	0.537
no_piso_tierra	1	0.853

Método de extracción: análisis de componentes principales.

En la Tabla 5 se pueden observar los indicadores que guardan mayor asociación con las componentes 1 y 2.

Tabla 5: Matriz de componentes no rotados

	Componente	
	1	2
agua_clorada	-0.796	-0.004
no_anemia	0.536	0.494
no_dispersion	0.890	-0.336
no_dci	0.931	-0.050
no_pobreza	0.879	-0.139
no_bajo_peso	0.546	0.404
agua_potable	0.913	0.024
servicio_higienico	0.876	0.156
Matricula	-0.127	0.770
no_diarrea	0.454	0.575
no_piso_tierra	0.824	-0.417

Método de extracción: análisis de componentes principales.
a. 2 componentes extraídos.

Criterios de extracción de factores:

En la Tabla 6 se determina el número de factores que forman parte del modelo factorial.

Tabla 6: Autovalores y varianza explicada de factores no rotados

Componente	Autovalores iniciales			Sumas de cargas al cuadrado de la extracción		
	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado
1	6.151	55.923	55.923	6.151	55.923	55.923
2	1.665	15.137	71.059	1.665	15.137	71.059
3	0.956	8.691	79.750			
4	0.632	5.744	85.494			
5	0.572	5.199	90.693			
6	0.341	3.100	93.793			
7	0.262	2.382	96.175			
8	0.234	2.127	98.301			
9	0.077	0.704	99.006			
10	0.070	0.636	99.641			
11	0.039	0.359	100.000			

Método de extracción: análisis de componentes principales.

De acuerdo al primer criterio del autovalor superior a 1, se aprecia que a partir del tercer componente los autovalores son inferiores al valor mínimo permitido, por lo que sería recomendable quedarnos con dos factores. De acuerdo al segundo, se consideran los

componentes que acumulen el 70% de la varianza total, esto se puede observar en el segundo componente, lo cual confirma el hecho de tomar dos factores.

En la Figura 1 se observa que el codo del brazo imaginario se sitúa en el segundo factor y con ello la confirmación de considerar dos factores en el modelo factorial.

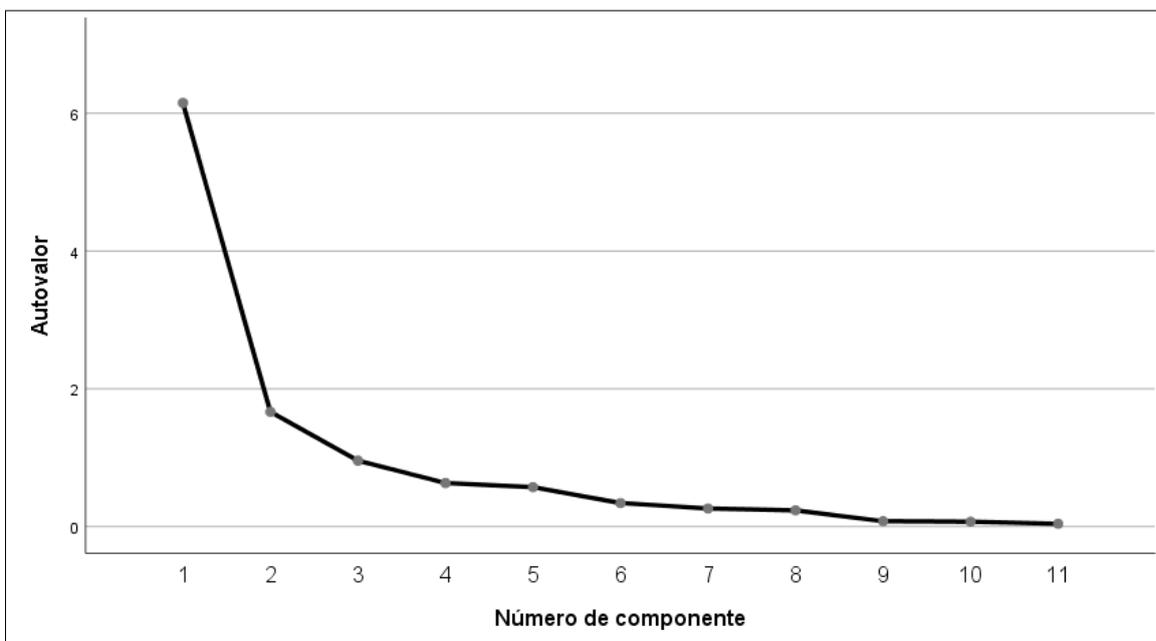


Figura 1: Gráfico de sedimentación de los factores

Paso 3: Rotación de Factores

En este paso se considera dos aspectos importantes: matriz de factores no rotados y matriz de factores no rotados.

Matriz de Factores No Rotados:

En el Figura 2 se pueden apreciar dos grupos diferenciados de variables, el primer grupo se encuentran cercanos de forma positiva al factor uno las variables servicio higiénico, agua potable, desnutrición crónica, pobreza, dispersión, piso tierra y bajo peso, mientras de forma negativa está solo la variable agua clorada. Por otro lado, para el factor dos, la variable matrícula se encuentra cercana. Sin embargo, queda la variable diarrea, la cual no se puede visualizar si pertenece al factor uno o al factor dos a simple vista.

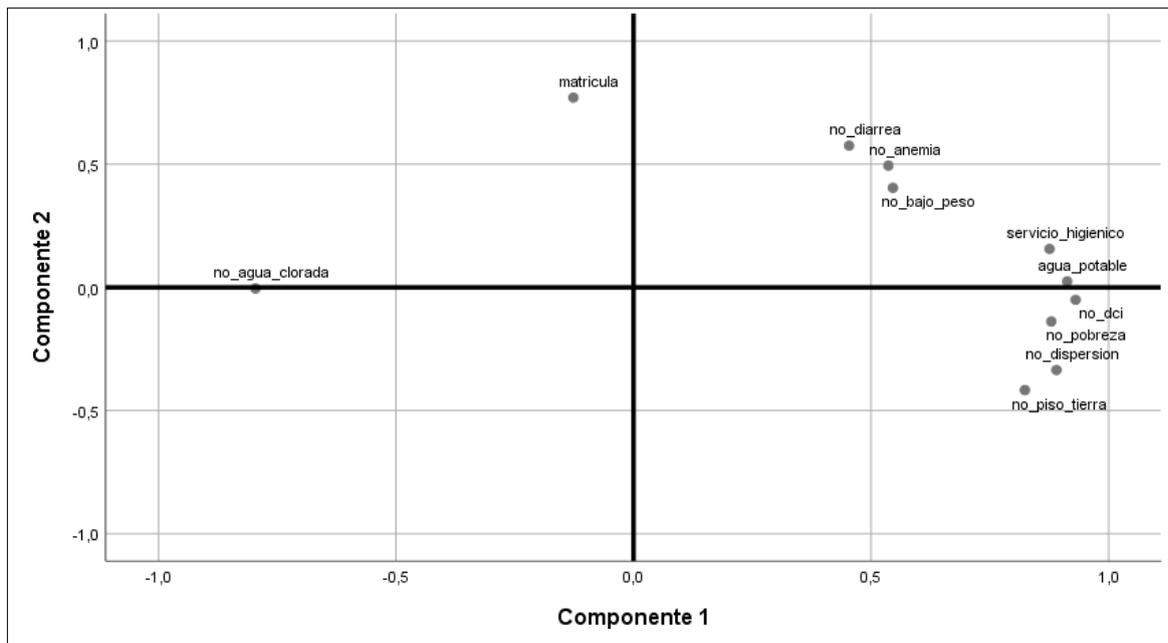


Figura 2: Gráfico de componentes no rotados

Matriz de Factores Rotados:

La recomendación de muchos autores es trabajar con el método de rotaciones oblicuas, siempre y cuando las correlaciones entre los factores sean significativas. En la Tabla 7 podemos observar el resultado de las correlaciones de los factores con los métodos Oblimin y Promax.

Tabla 7: Matriz de correlaciones según rotación oblicua

	Oblimin		Promax	
	Factor 1	Factor 2	Factor 1	Factor 2
Factor 1	1.000	0.211	1.000	0.347
Factor 2	0.211	1.000	0.347	1.000

Los dos métodos arrojaron correlaciones bajas por lo que no sería recomendable trabajar con estos métodos de rotación en esta investigación, por ello sería mejor seguir el análisis por medio de las rotaciones ortogonales.

Al calcular la matriz de la estructura factorial rotada por medio de los métodos ortogonales, la cual podemos apreciar en la Tabla 8, el mejor método que se ajusta a los datos utilizados fue Varimax.

Tabla 8: Matriz de componentes según métodos ortogonales

	Varimax		Quartimax		Equamax	
	Factor	Factor	Factor	Factor	Factor	Factor
	1	2	1	2	1	2
no_agua_clorada	-0.730	-0.316	-0.791	-0.088	-0.730	-0.316
no_anemia	0.299	0.665	0.482	0.548	0.299	0.665
no_dispersión	0.951	0.041	0.920	-0.241	0.951	0.041
no_dci	0.876	0.319	0.931	0.047	0.876	0.319
no_pobreza	0.863	0.218	0.889	-0.046	0.863	0.218
no_bajo_peso	0.344	0.586	0.501	0.459	0.344	0.586
agua_potable	0.830	0.380	0.905	0.119	0.830	0.380
servicio_higienico	0.744	0.487	0.854	0.247	0.744	0.487
Matricula	-0.419	0.659	-0.207	0.753	-0.419	0.659
no_diarrea	0.192	0.707	0.391	0.619	0.192	0.707
no_piso_tierra	0.921	-0.06	0.863	-0.329	0.921	-0.06

Un criterio importante para seleccionar la matriz de componentes es que cada variable debe estar correlacionada sólo con un factor, en el caso de la matriz de correlaciones sin rotar, podemos observar que cada variable pertenece a solo un factor, solo la variable no_anemia por poco está relacionado con ambos factores.

El resultado obtenido con el método Varimax muestra una mejor distribución de las variables en cada factor, además que se puede diferenciar fácilmente qué variable pertenece a cada factor sin ningún problema.

En el caso de la metodología por Quartimax, podemos ver que la variable no_bajo_peso tiene correlaciones bajas con respecto a otros métodos de rotación y el método Equamax arroja los mismos resultados que Varimax.

Por lo tanto, el método de rotación recomendable para los indicadores de Desarrollo Infantil Temprano con los indicadores del Instituto Nacional de Estadística e Informática del 2018 es Varimax.

Además, como podemos ver en la Tabla 9, la información referente al porcentaje acumulado tras la rotación de los factores no cambia.

Tabla 9: Autovalores y varianza explicada de factores rotados

Componente	Autovalores iniciales			Sumas de cargas al cuadrado de la rotación		
	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado
1	6.151	55.923	55.923	5.460	49.640	49.640
2	1.665	15.137	71.059	2.356	21.419	71.059
3	0.956	8.691	79.750			
4	0.632	5.744	85.494			
5	0.572	5.199	90.693			
6	0.341	3.100	93.793			
7	0.262	2.382	96.175			
8	0.234	2.127	98.301			
9	0.077	0.704	99.006			
10	0.070	0.636	99.641			
11	0.039	0.359	100.000			

Manteniendo el criterio del número de factores que se recomienda elegir, sin afectar lo mencionado en los pasos anteriores.

En la Figura 3 se muestra la composición de los dos factores, confirmando lo que se tiene en la Tabla 6 de componentes factoriales rotados.

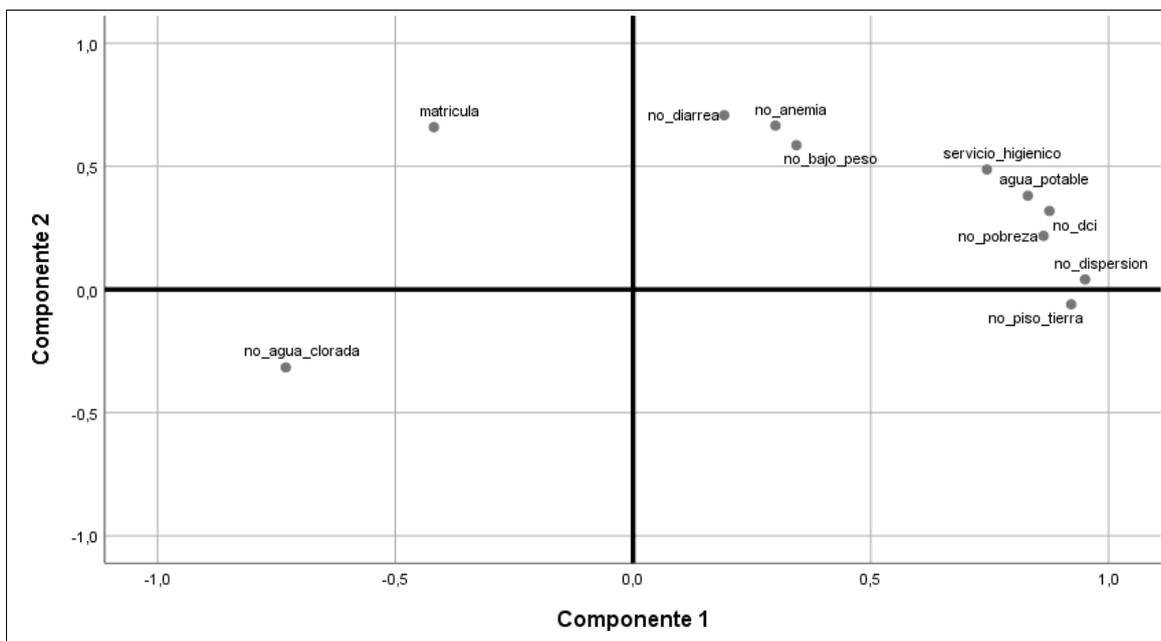


Figura 3: Gráfico de componentes rotados

Las variables agua clorada, dispersión, desnutrición crónica, pobreza, agua potable, servicio higiénico y piso tierra están más cercanos al eje del primer componente y las variables anemia, bajo peso, matrícula y diarrea al segundo componente.

Paso 4: Interpretación de Factores

En este paso se considera dos aspectos importantes: la evaluación de factores y la puntuación de los factores.

Evaluación de Factores:

Con el sustento en los pasos previos la composición de los factores sería:

- El primer factor tiene un poder explicativo del 49.64% de la varianza, está compuesto por agua clorada (-0.73), dispersión (0.951), pobreza (0.863), agua potable (0.83), servicio higiénico (0.744) y piso tierra (0.921), estas variables representan al factor de características y servicios básicos del hogar y se recomienda agregar un indicador de salud infantil el cual es desnutrición crónica (0.876) debido a que tiene fuerte relación con este factor.
- El segundo factor tiene un poder explicativo del 21.419% de la varianza, está compuesto por anemia (0.665), bajo peso al nacer (0.586) y diarrea (0.707), estas variables representan al factor de indicadores de salud infantil y también se le debería agregar el indicador de acceso a la educación denominado matrícula, ya que tiene una relación considerable con un valor de 0.659 con el factor.

Puntuaciones de los Factores:

En la Tabla 10 podemos apreciar las puntuaciones de los 25 gobiernos regionales.

Tabla 10: Puntuaciones de los Factores

Departamento	Factor 1	Factor 2
Amazonas	-1.108	0.870
Ancash	-0.124	0.622
Apurímac	-1.036	0.997
Arequipa	1.345	-0.767
Ayacucho	-0.528	0.847
Cajamarca	-1.335	1.370
Callao	1.714	0.293
Cusco	-0.204	0.010
Huancavelica	-1.719	1.082
Huánuco	-0.928	0.568
Ica	1.045	0.254
Junín	-0.488	-1.101
La Libertad	0.223	-0.440
Lambayeque	0.623	0.001
Lima Región	0.718	0.059
Loreto	-0.803	-2.786
Madre de Dios	0.747	-0.470
Moquegua	1.435	1.518
Pasco	-0.950	-0.480
Piura	-0.074	-0.207
Puno	-1.006	-0.391
San Martín	-0.221	-1.172
Tacna	1.691	1.210
Tumbes	0.771	-0.557
Ucayali	0.214	-1.328

Las puntuaciones factoriales para cada gobierno regional con los indicadores del Instituto Nacional de Estadística e Informática asociados al desarrollo infantil temprano de niños peruanos fueron:

Factor 1: Amazonas (-1.108), Ancash (-0.124), Apurímac (-1.036), Arequipa (1.345), Ayacucho (-0.528), Cajamarca (-1.335), Callao (1.714), Cusco (-0.204), Huancavelica (-1.719), Huánuco (-0.928), Ica (1.045), Junín (-0.488), La Libertad (0.223), Lambayeque (0.623), Lima Región (0.718), Loreto (-0.803), Madre de Dios (0.747), Moquegua (1.435), Pasco (-0.950), Piura (-0.074), Puno (-1.006), San Martín (-0.221), Tacna (1.691), Tumbes (0.771) y Ucayali (0.214).

Factor 2: Amazonas (0.870), Ancash (0.622), Apurímac (0.997), Arequipa (-0.767), Ayacucho (0.847), Cajamarca (1.370), Callao (0.293), Cusco (0.010), Huancavelica (1.082), Huánuco (0.568), Ica (0.254), Junín (-1.101), La Libertad (-0.440), Lambayeque (0.001), Lima Región (0.059), Loreto (-2.786), Madre de Dios (-0.470), Moquegua (1.518), Pasco (-0.480), Piura (-0.207), Puno (-0.391), San Martín (-1.172), Tacna (1.210), Tumbes (-0.557) y Ucayali (-1.328).

Las regiones que tienen puntuaciones altas para el primer factor son Callao y Tacna con valores de 1.714 y 1.691 respectivamente y puntuación baja en Huancavelica con -1.719.

En el caso del segundo factor, la puntuación más alta es para la región Moquegua con 1.518 y la puntuación más baja en Loreto con -2.786.

De esta forma ya tenemos identificados las regiones que tienen valores atípicos y tenerlo presente para análisis posteriores.

V. CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusiones

- Son dos las dimensiones que forman parte del modelo de factores asociado al desarrollo infantil temprano de niños peruanos: (1) características y servicios básicos del hogar e (2) indicadores de salud infantil y acceso a la educación.
- La dimensión características y servicios del hogar está compuesto por los indicadores: agua clorada, dispersión, pobreza, agua potable, servicio higiénico, piso tierra y desnutrición crónica.
- La dimensión indicadores de salud infantil y acceso a la educación está compuesto por los indicadores: anemia, bajo peso al nacer, diarrea y matrícula.
- El porcentaje de variancia acumulada que optimiza la solución factorial es 71.059%.
- Callao, Tacna y Huancavelica tienen valores atípicos para la dimensión características y servicios básicos del hogar.
- Moquegua y Loreto tienen valores atípicos para la dimensión indicadores de salud infantil y acceso a la educación.

5.2. Recomendaciones

- Para contar con un escenario alternativo de análisis de los indicadores de Desarrollo Infantil Temprano se recomienda usar la matriz de componentes no rotados, el uso de esta matriz también es válido de acuerdo a los estadísticos obtenidos.

- Para análisis posteriores con modelos de aprendizaje supervisados se recomienda retirar los valores atípicos encontrados en las puntuaciones factoriales.
- Para comprobar y validar los resultados obtenidos en este análisis se recomienda realizar posteriormente un Análisis Factorial Confirmatorio.

VI. REFERENCIAS BIBLIOGRÁFICAS

- Avalo, J. (2014). Aplicación de análisis factorial exploratorio para la validación de un modelo de innovación empresarial en Perú. (Tesis de Grado). Piura, Perú: Universidad de Piura.
- Benzécri, J. (1973). *L'Analyse des données. II. L'analyse des correspondances*. Paris: Dunod.
- Burt, C. (1940). *The factors of the mind: An introduction to factor analysis in psychology*. Londres: University of London.
- Chávez, M. (2017). Introducción a los métodos multivariantes, análisis factorial y de componentes principales. Recuperado de <https://rpubs.com/>
- De la Fuente, S. (2011). *Análisis factorial*. Universidad Autónoma de Madrid.
- Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24, 417-441 y 498-520.
- Kaiser, H. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23, 187-200.
- Lloret, S.; Ferreres, A.; Hernández, A. y Tomás, I. (2014). El análisis factorial exploratorio de los ítems: una guía práctica, revisada y actualizada. *Anales de Psicología*, 30(3), 1151-1169.
- López, P. y Fachelli, S. (2015). *Metodología de la investigación social cuantitativa*. Universidad Autónoma de Barcelona. Recuperado de <https://ddd.uab.cat/record/129382>

López, M. y Gutiérrez, L. (2019). Cómo realizar e interpretar un análisis factorial exploratorio utilizando SPSS. *Reire revista de innovación y revista en educación*, 12(2), 1-14.

Pérez, C. (2004). *Técnicas de análisis multivariante de datos. Aplicaciones con SPSS*. Madrid, España: Universidad Complutense de Madrid. pp. 121-154.

Pearson, K. (1901). On lines and planes of closet fit to systems of point in space. *Philosophical Magazine*, 6, 559-572.

Thurstone, L. (1947). *Multiple-factor analysis*. Chicago: University of Chicago Press.

VII. ANEXOS

Anexo 1: Tabla de datos utilizados para el análisis factorial

Departamento	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11
Amazonas	92.4	55.5	49.8	79.6	66.5	91.7	27.6	54.6	97.1	89.8	37.9
Ancash	73.3	54.3	63.1	83.6	79.7	94.5	60.5	70.4	99.4	85.4	50.4
Apurímac	85.4	46.8	37.8	79.9	68.2	93.5	39.8	60.9	97.9	88.4	30.3
Arequipa	49.4	61.6	91.0	94.7	91.4	93.5	88.2	80.6	75.7	91.1	83.7
Ayacucho	70.5	50.7	52.5	79.8	62.5	93.2	74.3	70.2	98.8	86.7	37.6
Cajamarca	92.5	68.1	37.0	72.6	58.1	90.2	23.9	58.4	94.1	91.6	35.0
Callao	30.9	63.0	100.0	94.0	84.1	93.7	96.3	93.5	91.9	91.7	94.3
Cusco	72.1	45.8	52.9	86.0	77.1	93.7	54.6	76.0	86.9	88.0	44.8
Huancavelica	86.4	44.2	20.6	68.0	61.3	93.2	20.9	58.5	96.3	88.3	23.3
Huánuco	77.8	56.2	40.2	77.6	70.1	92.0	29.0	64.3	94.8	87.2	45.1
Ica	81.3	56.9	93.5	95.0	96.9	93.6	77.5	80.7	95.4	91.1	79.5
Junín	77.9	43.0	62.1	80.8	78.5	90.4	42.5	69.6	82.8	88.4	61.0
La Libertad	81.4	62.1	79.5	85.2	79.1	91.8	67.2	74.1	80.6	89.5	62.1
Lambayeque	84.9	59.0	83.0	90.7	88.0	92.3	73.1	74.8	84.5	93.4	69.9
Lima Región	69.3	59.5	80.9	93.3	85.8	93.5	76.2	74.4	90.8	89.2	73.7
Loreto	73	42.6	69.2	80.0	67.3	89.0	45.7	60.4	77.0	80.8	75.8
Madre de Dios	31.5	45.2	81.3	93.0	96.8	92.8	61.1	63.3	94.6	90.2	83.6
Moquegua	35	66.1	80.7	97.7	91.3	94.6	75.0	88.4	99.7	93.2	71.5
Pasco	96.3	41.3	66.7	80.7	65.0	88.7	29.6	67.0	98.7	88.7	67.6
Piura	71.4	55.8	78.6	86.9	72.5	93.0	62.6	65.8	87.8	86.9	49.9
Puno	80.5	32.3	57.7	85.0	63.0	93.3	39.8	49.8	84.8	90.4	35.4
San Martín	64.6	49.9	67.4	89.5	75.1	93.6	46.3	61.8	80.4	83.8	57.3
Tacna	21.8	65.1	87.8	98.7	86.4	96.0	86.9	87.1	91.9	93.6	72.6
Tumbes	50.3	52.8	96.4	91.9	89.1	92.8	74.6	77.1	93.5	86.0	70.0
Ucayali	58.4	43.6	81.0	82.2	89.4	94.3	42.7	70.5	77.3	87.7	71.4

Anexo 2: Sintaxis en SPSS del análisis factorial

**Obtención de Correlaciones, Prueba KMO y Barlett e información de la matriz no rotada

DATASET ACTIVATE ConjuntoDatos1.

FACTOR

/VARIABLES agua_clorada no_anemia no_dispersion no_dci no_pobreza no_bajo_peso
agua_potable servicio_higienico matricula no_diarrea no_piso_tierra

/MISSING LISTWISE

/ANALYSIS agua_clorada no_anemia no_dispersion no_dci no_pobreza no_bajo_peso
agua_potable servicio_higienico matricula no_diarrea no_piso_tierra

/PRINT INITIAL CORRELATION KMO EXTRACTION

/PLOT EIGEN ROTATION

/CRITERIA MINEIGEN(1) ITERATE(25)

/EXTRACTION PC

/ROTATION NOROTATE

/SAVE REG(ALL)

/METHOD=CORRELATION.

**Obtención de matriz rotada: Oblimin

FACTOR

/VARIABLES agua_clorada no_anemia no_dispersion no_dci no_pobreza no_bajo_peso
agua_potable servicio_higienico matricula no_diarrea no_piso_tierra

/MISSING LISTWISE

/ANALYSIS agua_clorada no_anemia no_dispersion no_dci no_pobreza no_bajo_peso
agua_potable servicio_higienico matricula no_diarrea no_piso_tierra

/PRINT INITIAL ROTATION

/CRITERIA MINEIGEN(1) ITERATE(25)

/EXTRACTION PC

/CRITERIA ITERATE(25) DELTA(0)

/ROTATION OBLIMIN

/METHOD=CORRELATION.

**Obtención de matriz rotada: Promax

FACTOR

/VARIABLES agua_clorada no_anemia no_dispersion no_dci no_pobreza no_bajo_peso
agua_potable servicio_higienico matricula no_diarrea no_piso_tierra

/MISSING LISTWISE

```
/ANALYSIS agua_clorada no_anemia no_dispersion no_dci no_pobreza no_bajo_peso  
agua_potable servicio_higienico matricula no_diarrea no_piso_tierra
```

```
/PRINT INITIAL ROTATION
```

```
/CRITERIA MINEIGEN(1) ITERATE(25)
```

```
/EXTRACTION PC
```

```
/CRITERIA ITERATE(25)
```

```
/ROTATION PROMAX(4)
```

```
/METHOD=CORRELATION.
```

**Obtención de matriz rotada: Varimax

FACTOR

```
/VARIABLES agua_clorada no_anemia no_dispersion no_dci no_pobreza no_bajo_peso  
agua_potable servicio_higienico matricula no_diarrea no_piso_tierra
```

```
/MISSING LISTWISE
```

```
/ANALYSIS agua_clorada no_anemia no_dispersion no_dci no_pobreza no_bajo_peso  
agua_potable servicio_higienico matricula no_diarrea no_piso_tierra
```

```
/PRINT INITIAL ROTATION
```

```
/PLOT ROTATION
```

```
/CRITERIA MINEIGEN(1) ITERATE(25)
```

```
/EXTRACTION PC
```

```
/CRITERIA ITERATE(25)
```

```
/ROTATION VARIMAX
```

```
/METHOD=CORRELATION.
```

**Obtención de matriz rotada: Quartimax

FACTOR

```
/VARIABLES agua_clorada no_anemia no_dispersion no_dci no_pobreza no_bajo_peso  
agua_potable servicio_higienico matricula no_diarrea no_piso_tierra
```

```
/MISSING LISTWISE
```

```
/ANALYSIS agua_clorada no_anemia no_dispersion no_dci no_pobreza no_bajo_peso  
agua_potable servicio_higienico matricula no_diarrea no_piso_tierra
```

```
/PRINT INITIAL ROTATION
```

```
/PLOT ROTATION
```

```
/CRITERIA MINEIGEN(1) ITERATE(25)
```

```
/EXTRACTION PC
```

```
/CRITERIA ITERATE(25)
```

/ROTATION QUARTIMAX

/METHOD=CORRELATION.

**Obtención de matriz rotada: Equamax

FACTOR

/VARIABLES agua_clorada no_anemia no_dispersion no_dci no_pobreza no_bajo_peso
agua_potable servicio_higienico matricula no_diarrea no_piso_tierra

/MISSING LISTWISE

/ANALYSIS agua_clorada no_anemia no_dispersion no_dci no_pobreza no_bajo_peso
agua_potable servicio_higienico matricula no_diarrea no_piso_tierra

/PRINT INITIAL ROTATION

/PLOT ROTATION

/CRITERIA MINEIGEN(1) ITERATE(25)

/EXTRACTION PC

/CRITERIA ITERATE(25)

/ROTATION EQUAMAX

/METHOD=CORRELATION.