



TERCER INFORME DE LA INVESTIGACIÓN

Título: “Análisis comparativo de los indicadores sintéticos de la pobreza, desarrollo humano, competitividad y niveles de vida entre los departamentos en el año 2017”

EL INDICADOR SINTÉTICO DE NIVEL DE VIDA



**Mg. Renán Quispe
Llanos**

2018
INSTITUTO DE INVESTIGACIONES ECONÓMICAS Y
SOCIALES
UNI FIEECS

ÍNDICE

1. RESUMEN DEL AVANCE DEL PROYECTO	3
2. OBJETIVOS E HIPÓTESIS	4
3. INDICADORES DE NIVELES DE VIDA	4
3.1. Demográficos	4
3.2. Educativos	5
3.3. Vivienda	6
3.4. Salud	7
3.5. Empleo e Ingreso	7
4. LOS MÉTODOS MULTIVARIANTES LOS INDICADORES SINTÉTICOS Y EL ÍNDICE DE NIVEL DE VIDA	8
4.1. Aplicación del análisis factorial para obtener el indicador sintético	8
4.2. Elaboración de la Matriz de Correlaciones	10
b) El Test de Esfericidad de Bartlett	11
c) El Índice Kaiser-Meyer-Olkin	12
4.3. Extracción de los Factores Iniciales	13
a) Obtención de las componentes principales	15
b) Obtención de la primera componente	15
4.4. El Gráfico de Sedimentación	19
4.5. Comunalidades	19
4.6. Rotación de los Factores Iniciales	20
5. RESULTADOS DE LOS INDICADORES SINTÉTICOS DEL NIVEL DE VIDA - 2017	23
5.1. INDICADORES DE EDUCACIÓN	23
5.2. INDICADORES DEMOGRÁFICOS	29
5.3. INDICADORES DE VIVIENDA	30
5.4. INDICADORES DE SALUD	31
5.5. INDICADORES DE EMPLEO E INGRESO	32
5.6. Indicadores sintéticos de Niveles de vida	33
BIBLIOGRAFÍA	34
ANEXOS	36

1. RESUMEN DEL AVANCE DEL PROYECTO

En los dos primeros informes se han realizado una presentación conceptual de las metodologías utilizadas, así como los resultados para el año 2017, a nivel departamental en lo referente a la pobreza, el indicador de competitividad regional, y el índice de desarrollo humano de las Naciones Unidas. Asimismo, se efectuó un análisis comparativo de los indicadores, a través de los rankings generados. En la última parte, se identificaron, conceptualizaron y elaboraron y recopilaron un conjunto de indicadores relacionados con las condiciones de vida y nivel de vida de la población; agrupados en grandes áreas de análisis tales como: aspectos demográficos, aspectos educativos, vivienda salud, empleo e ingreso.

En esta oportunidad, el informe contiene el detalle de la metodología referida a la aplicación del análisis factorial y el de componentes principales como instrumento para el análisis multivariado como paso previo para obtener indicadores sintéticos. A continuación se aplica la metodología para cada área de análisis con información para el año 2017. En particular se presenta el proceso detallado para obtener el indicador sintético de nivel de vida de los aspectos relacionados con la educación.

Finalmente se presenta una comparación de los diferentes índices por tema a nivel departamental, que facilite el análisis de los niveles de vida entre departamentos según tema.

Palabras claves: Nivel de vida, Análisis Factorial, componentes principales.

2. OBJETIVOS E HIPÓTESIS

2.1. Objetivo General

A partir de un conjunto de indicadores demográficos, sociales y económicos, por departamento elaborar un indicador que sintetice las condiciones y calidad de vida de la población peruana en un índice que permita realizar un ranking del nivel de vida a nivel departamental, a fin de efectuar un análisis relacional complementario, respecto a otras metodologías que igualmente miden en forma sintética la pobreza, el índice de desarrollo humano y competitividad en lo referente al posicionamiento interdepartamental, identificando las principales variables que inciden en su evolución.

2.2. Hipótesis General

El Nivel de vida de la población tiene muchos componentes que se pueden agrupar en base a las principales características de la población relacionadas con su situación demográfica, así como aspectos relacionados con las condiciones de vida relacionados con la educación, vivienda, salud, empleo e ingreso. Las técnicas de análisis factorial y de componentes principales permite la elaboración de un Indicador sintético combinando los diferentes indicadores de la calidad y condiciones de vida de la población de los ítems mencionados para obtener un solo valor, con el cual se establecerá obtener un ranking de nivel de vida a nivel departamental.

3. INDICADORES DE NIVELES DE VIDA

En base a diferentes fuentes bibliográficas sobre los indicadores que más se utilizan para evaluar, la calidad o condiciones de vida se realiza una propuesta inicial de un conjunto de indicadores los cuales se utilizarán para medir en términos comparativos el nivel de vida departamental. En efecto a continuación se presentan los indicadores estructurales claves sobre los principales aspectos que están relacionados en el nivel demográfico social y económico regional.

3.1. Demográficos

- **Esperanza de vida al nacer:** Es una estimación del número promedio de años que le restaría vivir a un recién nacido si las condiciones de mortalidad actuales permanecen invariables.

- **Tasa de mortalidad infantil:** Es la probabilidad que tiene un recién nacido de morir antes de cumplir un año de vida. En la práctica, se define como el cociente entre las defunciones de los niños menores de un año ocurridas en un período dado y los nacimientos ocurridos en el mismo lapso, multiplicado por mil.

$$TMI = \frac{d_z}{B_z} \times 1000$$

Dónde:

d_z = Número de defunciones de menores de 1 año en el año z.

B_z = Número de nacidos vivos en el año z.

- **Tasa global de fecundidad:** Es el promedio de hijos nacidos vivos que tendrían las mujeres durante toda su vida reproductiva, si las tasas de fecundidad por edad se mantuviesen invariables en el tiempo. Están expresadas en nacimientos por mujer.
- **% de la población menor de 15 años de edad:** Indica que porcentaje representa la población de 15 años respecto a la población total.
- **Tasa de dependencia demográfica:** Es la relación entre las personas en edades "dependientes" (menores de 15 años y mayores de 64 años de edad) con las personas en edades "económicamente productivos" de 15 a 64 años, expresado en porcentaje.

3.2. Educativos

- **Tasa de Analfabetismo:** Según las NN.UU.: “Analfabeto es toda persona de 15 y más años de edad, que no sabe leer ni escribir o sabiéndolo no pueden interpretar un texto básico”.

$$TA = \frac{\text{Población analfabeta de 15 y más años}}{\text{Población de 15 y más años}} \times 100$$

- **Años promedios de estudios de la población de 15 años y más:** Corresponde al número de años de estudio promedio de la población mayor a 15 años.
- **Tasa de asistencia a secundaria de la población de 12 a 16 años de edad:**
 - **Tasa bruta de asistencia escolar de la población de 12 a 16 años de edad:** Se refiere al porcentaje de asistencia de la población de este grupo de edad a algún grado o año del Sistema de Enseñanza Regular, respecto de la población del mismo grupo de edad, se expresa en porcentaje.
 - **Tasa neta de asistencia escolar de la población de 12 a 16 años de edad:** Se refiere al porcentaje de asistencia de la población de este grupo de edad a algún año o grado de educación secundaria, respecto de la población del mismo grupo de edad, se expresa en porcentaje.

- **% de la Población de 6 a 11 años que asiste a primaria con atraso escolar:** Representa al porcentaje de la población de este grupo de edad que asiste a primaria con atraso escolar, respecto de la población del mismo grupo de edad.
- **% de la Población de 12 a 16 años de edad que asiste a secundaria con atraso escolar:** Es el porcentaje de la población de este grupo de edad que asiste a secundaria con atraso escolar, respecto de la población del mismo grupo de edad.
- **% de niños evaluados en el segundo grado de primaria con nivel satisfactorio de comprensión lectora:** Es el porcentaje de la población de segundo de primaria con nivel satisfactorio de comprensión lectora, nivel satisfactorio se refiere al estudiante que logró los aprendizajes esperados para el grado y está listo para seguir aprendiendo.
- **% de niños evaluados en el segundo grado de primaria con nivel satisfactorio de razonamiento matemático:** Es el porcentaje de la población de segundo de primaria con nivel satisfactorio de razonamiento matemático, nivel satisfactorio se refiere al estudiante que logró los aprendizajes esperados para el grado y está listo para seguir aprendiendo.
- **% de la población de 15 años y más con formación superior:** Es el porcentaje de la población de este grupo de edad con formación superior, respecto de la población del mismo grupo de edad.

3.3. Vivienda

- **% de viviendas con acceso a servicios básicos agua, servicios higiénicos y energía eléctrica:**
 - **% de viviendas que cuentan con abastecimiento de agua potable por red pública:** Es el porcentaje de viviendas que cuentan con agua potable en red pública (comprende viviendas con acceso a red pública dentro y fuera de la vivienda y pilón de uso público), respecto al total de viviendas particulares.
 - **% de viviendas con acceso de servicio higiénico:** Es el porcentaje de viviendas que poseen acceso al servicio higiénico conectado a red pública de desagüe, dentro o fuera de la vivienda, respecto al total de viviendas particulares.
 - **% de viviendas con acceso a red de alumbrado eléctrico por red pública:** Es el porcentaje de viviendas que cuentan con servicio de alumbrado eléctrico conectado a red pública, respecto al total de viviendas particulares.

- **% de viviendas inadecuadas:** Es el porcentaje de viviendas cuyas paredes han sido hechas de quincha, estera, piedra con barro y otro tipo de material de mala calidad o inadecuado, respecto al total de viviendas particulares.
- **% de viviendas con piso de cemento:** Es el porcentaje de viviendas cuyos pisos han sido hechos con cemento, respecto al total de viviendas particulares.
- **% de viviendas con techo de concreto:** Es el porcentaje de viviendas cuyos techos han sido construidos a base de concreto, respecto al total de viviendas particulares.

3.4. Salud

- **Número de habitantes por cada médico:** Se refiere a la cantidad de población por cada médico en el Perú.

$$\text{Numero de habitantes por médico} = \frac{\text{Población Total}}{\text{Cantidad Total de médicos}}$$

- **% de la población que reportó padecer algún problema de salud crónica:** Se refiere al porcentaje de la población que reporto algún problema de salud crónico, respecto al total de la población.
- **Tasa de desnutrición crónica de niños menores de cinco años de edad, según ámbito geográfico, 2017 (Patrón OMS):** Se refiere al porcentaje de la población de este grupo de edad que presento desnutrición crónica, respecto a la población del mismo grupo de edad.

3.5. Empleo e Ingreso

- **% de la PEA que tiene formación secundaria:** Se refiere al porcentaje de la población de este grupo de edad que pertenece a la PEA y tiene formación secundaria, respecto de la población de la PEA.
- **% de la PEA que tiene formación superior:** Se refiere al porcentaje de la población de este grupo de edad que pertenece a la PEA y tiene formación superior, respecto de la población de la PEA.
- **% de la PEA ocupada con seguro de Salud:** Se refiere al porcentaje de la población que pertenece a la PEA ocupada y posee seguro de salud, respecto de la población de la PEA ocupada.
- **% de la PEA ocupada que es independiente:** Se refiere al porcentaje de la población que pertenece a la PEA ocupada y es independiente, respecto de la población de la PEA ocupada.

- **% de la PEA ocupada que es agropecuaria:** Se refiere al porcentaje de la población que pertenece a la PEA ocupada y trabaja en el sector agropecuario, respecto de la población de la PEA ocupada.
- **Ingreso promedio de los ocupados (soles):** Corresponde al promedio de todos los ingresos de la PEA ocupada.

4. LOS MÉTODOS MULTIVARIANTES LOS INDICADORES SINTÉTICOS Y EL ÍNDICE DE NIVEL DE VIDA

La reconocida naturaleza multidimensional de los fenómenos socio-económicos requiere un tratamiento estadístico que facilite la combinación de variables. El análisis multivariado o multivariante es un conjunto de técnicas estadísticas cuyos objetivos se pueden resumir en dos: Reducción y Clasificación. Una de las alternativas de uso de estas técnicas sugiere identificar y construir indicadores sintéticos que capten el carácter multidimensional del índice del nivel de vida, que resuman cuantitativamente la realidad, desde diferentes ángulos y que tengan la capacidad de mostrar las desigualdades sociales.

Un indicador sintético es la función matemática de indicadores parciales los cuales pueden ser simples o componentes. El indicador resume y mide las manifestaciones, causas características o efectos de un determinado fenómeno económico social.

Se entiende por componente de un indicador sintético o global aquel indicador que aporta alguna información con independencia de la apreciación subjetiva acerca del objetivo buscado con el indicador.

A diferencia de otros métodos de construcción de indicadores globales, que asignan la importancia de las variables o indicadores según el criterio del investigador, el análisis factorial por medio del método de componentes principales, permite construir el índice global en la forma de una suma ponderada de las proyecciones sobre cada una de las primeras componentes principales.

4.1. Aplicación del análisis factorial para obtener el indicador sintético

Es una modalidad del Análisis Multivariado que nos permite reducir una serie de variables que están correlacionadas entre sí a un conjunto menor de factores con los

cuales comparten esa relación. Estos factores contienen la mayor parte de la información y son suficientes para explicar el modelo.

El objetivo, consiste en identificar un número reducido de factores, inferior al número de variables (indicadores de nivel de vida), mediante los cuales se pueda describir el fenómeno observado de forma simplificada, con una pérdida mínima de información. En síntesis se trata de reducir la información recogida.

El modelo matemático del análisis factorial es parecido al de la regresión múltiple. Sin embargo en este caso cada variable (cada indicador) se expresa como una variable dependiente a través de una combinación lineal de los factores calculados que no son directamente observables.

$$X_{ij} = F_{1i}a_{i1} + F_{2i}a_{i2} + \dots + F_{ki}a_{ik} + \mu_i$$

Donde: X_{ij} es el valor que tiene el indicador “j” en la unidad de análisis “i”; los F_{ij} constituyen los coeficientes factoriales; los “a” son puntuaciones factoriales que hacen el papel de ponderaciones de los coeficientes F_{ij} y “ μ_i ” es el factor único.

El investigador mide estas variables sobre n individuos, obteniéndose la siguiente matriz de datos:

Sujetos	Variables			
	X ₁	X ₂	...	X _p
1	X ₁₁	X ₁₂	...	X _{1p}
2	X ₂₁	X ₂₂	...	X _{2p}
...
n	X _{n1}	X _{n2}	...	X _{np}

El modelo del Análisis Factorial viene dado habitualmente por las ecuaciones:

$$X_1 = a_{11}F_1 + a_{12}F_2 + \dots + a_{1k}F_k + u_1$$

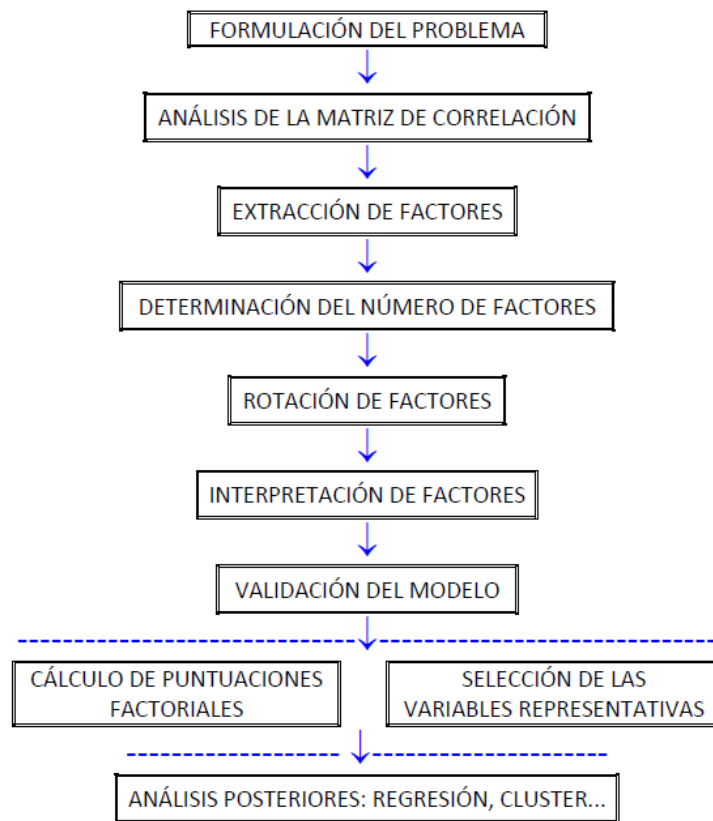
$$X_2 = a_{21}F_1 + a_{22}F_2 + \dots + a_{2k}F_k + u_2$$

.....

$$X_p = a_{p1}F_1 + a_{p2}F_2 + \dots + a_{pk}F_k + u_p$$

Sin embargo los factores identificados que explican estas variables observables, son construcciones que no son directamente observables. Ello constituye por lo tanto la principal limitación de esta técnica reduccionista para el análisis de los niveles de vida. Para solucionar ese problema, se acude al análisis de componentes principales que en combinación con el análisis factorial permite la estimación del modelo, obtener sus resultados y facilitar la interpretación como se verá más adelante. A continuación se

presenta se presenta la secuencia para estimar el modelo para estimar los resultados que permita obtener el indicador sintético.



4.2. Elaboración de la Matriz de Correlaciones

El primer paso en el análisis factorial, consiste en calcular la matriz de correlaciones entre todas las variables que entran en el análisis a partir de datos originales. A continuación se examina para comprobar si sus características son adecuadas para realizar un análisis factorial. Uno de los requisitos que deben cumplirse para que el análisis factorial tenga sentido es que las variables estén altamente intercorrelacionadas.

A continuación se realizará unos tests del grado de asociación entre las variables que nos indicarán si es pertinente, desde el punto de vista estadístico, llevar a cabo el Análisis Factorial con la data y muestras disponibles. Todos ellos cumplen la misma función y suelen llegar a la misma conclusión, Entre los principales se tiene:

a) **El valor del determinante**

El determinante de la matriz de correlaciones es un indicador del grado de las intercorrelaciones. Un determinante muy bajo significa que hay variables con intercorrelaciones muy altas. Pero no se tiene el grado de confianza.

b) **El Test de Esfericidad de Bartlett**

Se utiliza para probar la Hipótesis Nula que afirma que las variables no están correlacionadas en la población. Es decir, comprueba si la matriz de correlaciones es una matriz de identidad. Una matriz identidad es aquella que tiene unos en la diagonal principal y los valores restantes son ceros.

$$X_j = \begin{matrix} & 1 & 0 & 0 & \dots & 0 \\ & 0 & 1 & 0 & \dots & 0 \\ & 0 & 0 & 1 & \dots & 0 \\ & \dots & \cdot & \dots & \dots & \dots \\ & 0 & 1 & 0 & \dots & 1 \end{matrix} = I$$

Se puede dar como válidos aquellos resultados que nos presenten un valor elevado del test y cuya fiabilidad sea menor a 0.05. En este caso se rechaza la Hipótesis Nula y se continúa con el Análisis.

El test de esfericidad de Bartlett consiste en la estimación de un estadístico “B” que tiene una distribución ji-cuadrado a partir de una transformación del determinante de la matriz de correlaciones de la muestra observada como estimador del determinante de la matriz de correlación de la población a la que pertenece la muestra analizada. Un determinante próximo a cero indica que una o más variables podrían ser expresadas como una combinación lineal de otras variables.

$$H_a: R=I \text{ ó } |R| = 1$$

Si se confirma la H_a significa que las variables no están intercorrelacionadas. Por lo tanto, la nube de puntos en el espacio formaría una esfera (esfericidad). Se ha

demostrado que estos determinantes siguen una distribución de ji cuadrado. La expresión matemática de la prueba de Bartlett es:

$$B = X^2 = - [n - 1 - (1 / 6 * (2P + 5))] \ln |R|$$

Donde n es el número de individuos de la muestra y P el número de variables incluidas en la matriz de correlaciones. La expresión $\ln |R|$ se refiere al logaritmo neperiano del determinante de la matriz de correlaciones.

Los grados de libertad vienen dados por $P = \frac{1}{2} (P^2 - P)$ cuyo valor teórico en la tabla X^2 sirve para rechazar o no la Hipótesis nula con un nivel de confianza del 95%.

Si con el test de Bartlett se obtienen valores altos del B con distribución X^2 superior a su correspondiente valor teórico, definido por los grados de libertad se rechaza la hipótesis nula con un cierto grado de significación normalmente menor a 0.05, en concordancia con el nivel de confianza del 95%. En caso de no rechazarse la hipótesis nula significaría que las variables no están intercorrelacionadas. En este supuesto debería reconsiderarse la aplicación de un análisis factorial.

Para aplicar esta prueba se requiere que los datos procedan de una población que sigue la distribución normal de multivariable.

c) **El Índice Kaiser-Meyer-Olkin**

Indica qué tan apropiado es aplicar el Análisis Factorial. Los valores entre 0.5 y 1 indican que es apropiado aplicarlo.

La medida de adecuación de la muestra KMO, de Kaiser - Meyer- Olkim, es un índice para comparar las magnitudes de los coeficientes de correlación observados con las magnitudes de correlación parcial.

Valores bajos en el índice KMO desaconseja la aplicación del análisis factorial, puesto que las correlaciones entre pares de variables no pueden ser explicadas por las otras variables.

Kaiser (1974) ofrece un rango de valores del índice para aceptar que la matriz de datos es adecuada para proceder al análisis factorial.

Una manera de cuantificar este hecho es con la **Media de Adecuación de la Muestra KMO** propuesta por Kaiser-Meyer-Olkin:

$$KMO = \frac{\sum_{i \neq j} \sum_{i \neq j} r_{ij}^2}{\sum_{i \neq j} \sum_{i \neq j} r_{ij}^2 + \sum_{i \neq j} \sum_{i \neq j} r_{ij(p)}^2} \quad 0 \leq KMO \leq 1$$

Kaiser-Meyer-Olkin para realizar un Análisis Factorial, proponen:

1	≥	KMO	>	0.90	muy bueno
0.90	≥	KMO	>	0.80	meritorio
0.80	≥	KMO	>	0.70	mediano
0.70	≥	KMO	>	0.60	mediocre
0.60	≥	KMO	>	0.50	bajo
KMO	≤	0.50			inaceptable

4.3. Extracción de los Factores Iniciales

Se dispone de varios métodos para extraer los Factores Iniciales. Los más habituales son los siguientes:

Componentes principales, Ejes principales o factor principal, Máxima Verosimilitud, Mínimos Cuadrados no ponderados o generalizados, factorización alfa, Factorización imagen.

El más utilizado y el que se emplea en este estudio es el de “Componentes Principales”.

La utilidad principal del análisis de componentes principales reside, en que permite estudiar un fenómeno multidimensional, para examinar la interdependencia de algunas o muchas variables comprendidas en el estudio.

El Análisis de Componentes Principales tiene como objetivo es explicar la mayor parte

de la variabilidad total de un conjunto de variables con el menor número de componentes posible. Para ello, se obtienen combinaciones lineales de variables representativas de cierto fenómeno multidimensional, con la propiedad de que exhiban varianza máxima y que a la vez estén incorrelacionadas entre sí.

Se dispone de muchos métodos para extraer los Factores Iniciales. El más utilizado y el que empleamos en este estudio es el de “Componentes Principales”.

La utilidad principal del análisis de componentes principales reside, en que permite estudiar un fenómeno multidimensional, cuando algunas o muchas de las variables comprendidas en el estudio están correlacionadas entre sí, en mayor o menor grado.

El Análisis de Componentes Principales tiene como objetivo el hallar combinaciones lineales de variables representativas de cierto fenómeno multidimensional, con la propiedad de que exhiban varianza máxima y que a la vez estén incorrelacionadas entre sí.

La varianza de la componente es una expresión de la cantidad de información que lleva incorporada. Es decir cuanto mayor sea su varianza, mayor será la cantidad de información incorporada en dicha componente. Por ésta razón las sucesivas combinaciones o variantes o componentes se ordenan en forma descendente de acuerdo a la proporción de la varianza total presente en el problema, que cada una de ellas explica.

La primera componente es por lo tanto, la combinación de máxima varianza; la segunda es otra combinación de variables originarias que obedece a la restricción de ser ortogonal a la primera y de máxima varianza, la tercer componente es aún otra combinación de máxima varianza, con la propiedad de ser ortogonal a las dos primeras; y así sucesivamente.

Por sus propiedades de ortogonalidad, las sucesivas componentes después de la primera se pueden interpretar como las combinaciones lineales de las variables originarias que mayor varianza residual explican, después que el efecto de las precedentes ha sido ya

removido y así sucesivamente hasta que el total de varianza ha sido explicado.

Es posible que unas pocas primeras componentes logren explicar una alta proporción de la varianza total; en este caso que ocurre cuando las variables están correlacionadas en mayor grado, las componentes pueden sintéticamente sustituir a las múltiples variables originarias. Ello permitiría resumir en unas pocas variantes o componentes no correlacionadas gran parte de la información originaria.

Desde este punto de vista, el método de componentes principales es considerado como un método de reducción, ya que puede reducir la dimensión del número de variables que inicialmente se han considerado en el análisis.

En síntesis, este procedimiento busca el factor que explique la mayor cantidad de la varianza en la matriz de correlación. Este recibe el nombre de “factor principal”. Esta varianza explicada se resta de la matriz original produciéndose una matriz residual. Luego se extrae un segundo factor de esta matriz residual y así sucesivamente hasta que quede muy poca varianza que pueda explicarse. Los factores así extraídos no se correlacionan entre ellos, por esta razón se dice que estos factores son ortogonales.

a) **Obtención de las componentes principales**

En primer lugar, es necesario indicar que la obtención de componentes principales es un caso típico de cálculo de raíces y vectores característicos de una matriz simétrica. Esta matriz simétrica es la conformada por la matriz de correlaciones de las variables consideradas en el estudio del fenómeno multidimensional.

b) **Obtención de la primera componente**

Consideremos que disponemos de una muestra de n observaciones sobre el comportamiento de p variables, que explican un fenómeno multidimensional. Estas variables deben de estar expresadas en desviaciones respecto a la media o en variables tipificadas.

La primera componente al igual que las siguientes, se expresa como combinación lineal de las variables originales, es decir:

$$Z_{1i} = \mu_{11}X_{1i} + \mu_{12}X_{2i} + \dots + \mu_{1p} X_{pi}$$

Dónde: Z_{1i} es el valor de la componente Z_1 para la observación i . Asimismo, $X_{1i}, X_{2i}, X_{3i}, \dots, X_{pi}$ son los valores de las variables $X_1, X_2, X_3, \dots, X_p$, en la observación i . Finalmente $\mu_{11}, \mu_{12}, \mu_{13}, \dots, \mu_{1p}$ es el vector de pesos o coeficientes cuyo valor queremos determinar.

Para el conjunto de las n observaciones muestrales esta ecuación se puede matricialmente de la siguiente forma:

$$\begin{matrix} Z_{11} & X_{11} & X_{21} & X_{31} & \dots & X_{p1} & \mu_{11} \\ Z_{12} & X_{12} & X_{22} & X_{32} & \dots & X_{p2} & \mu_{12} \\ Z_{13} = & X_{13} & X_{23} & X_{33} & \dots & X_{p3} & \mu_{13} \\ \dots & & \dots & & & & \\ Z_{1n}, & X_{1n} & X_{2n} & X_{3n} & \dots & X_{pn} & \mu_{1p} \end{matrix} \quad \mathbf{X}$$

Si la expresamos en forma vectorial tenemos

$$\mathbf{Z}_1 = \mathbf{X} \boldsymbol{\mu}_1$$

La primera componente es obtenida de forma que su varianza sea máxima, sujeta a la restricción de que la suma de los pesos al cuadrado, sea igual a la unidad.

La varianza de la primera componente, dada por:

$$Var(Z_1) = (\sum Z_{1i}^2)/n = (\mathbf{Z}'_1 \mathbf{Z}_1)/n = (\boldsymbol{\mu}'_1 \mathbf{X}' \mathbf{X} \boldsymbol{\mu}_1)/n = \boldsymbol{\mu}'_1 (\mathbf{X}' \mathbf{X}) / n \boldsymbol{\mu}_1$$

Ya que la media de las Z_j es igual a 0 debido a que las X_j variables están expresadas en desviaciones con respecto a la media o están tipificadas.

Si las variables están expresadas en desviaciones respecto a la media $(1/n) \mathbf{X}' \mathbf{X}$ es la matriz de covarianza de las X_j variables. Si las variables están tipificadas $(1/n) \mathbf{X}' \mathbf{X}$ es

igual a la matriz de correlaciones.

Si denominamos a la matriz $X'X$ como V , la varianza de la primera componente sería:

$$\text{Var}(Z_1) = \mu_1' V \mu_1$$

Si queremos que Z_1 tenga la máxima varianza tenemos que poner una restricción porque si no la varianza podría ser infinita. La restricción es que la suma de cuadrados de los coeficientes sea 1.

$$\sum \mu_{1i}^2 = \mu_1' \mu_1 = 1$$

Si incorporamos la restricción podemos formar el siguiente lagrangiano

$$L = \mu_1' V \mu_1 - \lambda (\mu_1' \mu_1 - 1)$$

Para maximizar el valor del lagrangiano derivamos respecto a μ_j e igualamos a 0:

$$\partial L / \partial \mu_1 = 2V\mu_1 - 2\lambda \mu_1 = 0$$

Es decir:

$$(V - \lambda I) \mu_1 = 0$$

Al resolver la ecuación $(V - \lambda I) \mu_1 = 0$, se obtienen p raíces características. Si toma la raíz característica mayor (λ_1), se puede hallar el vector característico asociado aplicando:

$\mu_1' \mu_1 = 1$. Luego el vector de ponderaciones o pesos que se aplica a las variables originales para obtener la primera componente principal es el vector característico asociado a la raíz característica mayor de la matriz V . Obtención de las restantes componentes

La segunda y las demás componentes Z , expresada en forma genérica y matricial será:

$$Z_h = X\mu_h$$

Para su obtención, además de la restricción de que $\sum_{i=1}^h \mu_{ih}^2 = 1$, se imponen

restricciones adicionales para asegurar que la componente h-esima sea ortogonal a todos los vectores característicos calculados previamente.

$$\mu_h \mu_1 = \mu_h \mu_2 = \mu_h \mu_3 = \dots = \mu_h \mu_{h-1} = 0$$

En resumen las p componentes principales que se pueden calcular son una combinación lineal de las variables originales, en las que los coeficientes de ponderación son los correspondientes vectores característicos asociados a la matriz V. Las varianzas de los componentes vienen dadas por las raíces características es decir:

$$\text{Var } Z_h = \mu'_h V \mu_h = \lambda_h$$

Adoptando como medida de variabilidad de las variables originales, la suma de sus varianzas, la prop. de la componente h-esima en la variabilidad total será: $\lambda / \sum \lambda_h$.

En el caso de variables tipificadas, como la matriz de covarianzas es la matriz de correlaciones, la prop. de la componente h-esima en la variabilidad total será: λ_h / p .

En síntesis:

la técnica de componentes principales transforma las variables en otras incorrelacionadas, de media cero, que pueden escribirse como combinaciones lineales de las primeras y que se llaman factores o componentes principales, las cuales pueden ordenarse por la magnitud de su varianza la cual está dada por un valor propio de la matriz.

1. Las primeras r componentes principales bastan para describir en alto porcentaje la variabilidad total de las variables originales. Con frecuencia r vale 2 o 3, siendo el primero de ellos el caso más deseable.
2. Cuando el porcentaje de variabilidad explicado por dos componentes principales es alto (70%) se puede realizar una representación gráfica de las variables originales y de los individuos de la muestra (mapas perceptuales) que muestran algunas relaciones de correlación o semejanza entre ellos.

3. Aunque todas las variables originales entran en la composición de cada componente principal, algunas son más importantes que otras. Estas, las más importantes, determinan la naturaleza de cada componente

4.4. El Gráfico de Sedimentación

Representa en el eje de las “x” el número de orden de los factores y en el eje de las “y”, los valores propios (eigenvalues). Muestra la forma en que van disminuyendo los valores propios, graficando el hecho de que el primer factor es el que más varianza explica.

4.5. Comunalidades

La “comunalidad”, es la cantidad de varianza que una variable comparte con las demás variables consideradas. Los “eigenvalue” (valores propios) pueden interpretarse como la cantidad de varianza explicada por cada factor. La “carga factorial” es la correlación entre las variables y los factores.

Se denomina "comunalidad" a la proporción de varianza explicada por los factores comunes. Las comunalidades iniciales en el análisis de componentes principales son siempre iguales a uno. Por tanto este dato no aporta ninguna información relevante. En los demás métodos de extracción de factores (que no sean componentes principales), la comunalidad inicial (initial statistics) coincide con el coeficiente de correlación múltiple entre cada variable y todas las demás. En los otros métodos, por tanto, la comunalidad inicial es un índice de la adecuación de la matriz. La comunalidad (h^2) de cada variable se calcula a partir de la matriz factorial, y es igual a la suma de los cuadrados de las ponderaciones factoriales de cada variable. Matemáticamente:

$$H^2_i = F^2_{1j} + F^2_{2j} + \dots + F^2_{kj}$$

La comunalidad puede oscilar entre cero y uno. El cero indica que los factores comunes no explican nada en absoluto de la variabilidad de una variable. Mientras que el uno indica que -la variable queda totalmente explicada por los factores comunes. La varianza que no queda explicada por los factores comunes se atribuye al factor único.

Se da la siguiente igualdad. $1 = h^2 + U^2$

Siendo h^2 la comunalidad y U el factor único. En la matriz factorial aparecen solamente los factores comunes; el factor único no se incluye

4.6. Rotación de los Factores Iniciales

Con frecuencia es difícil interpretar los factores iniciales, por lo tanto, la extracción inicial se rota con la finalidad de lograr una solución que facilite la interpretación. Hay dos sistemas básicos de rotación de factores: los métodos de rotación ortogonales (mantienen la independencia entre los factores rotados: varimax, quartimax y equamax) y los métodos de rotación no ortogonales (proporcionan nuevos factores rotados que guardan relación entre sí). En el presente estudio se aplicará el Método de Rotación Varimax. Éste es, actualmente, uno de los métodos más utilizados.

a) El método de rotación Varimax

El método de rotación varimax, propuesto por Káiser (1958: 187-200), consiste en maximizar la varianza de los factores: Cada columna de la matriz factorial rotada produce algunos pesos muy altos y los otros se aproximan a cero. Este método tiende a minimizar el número de variables que tienen saturaciones altas en un factor.

De ello resulta una mayor facilidad de interpretación de los resultados.

Como conclusión teórica merece citar lo manifestado por Namakforoosh (1995) respecto del Análisis Factorial: Reduce la multiplicidad de pruebas y medidas hasta lograr una sencillez notable. Indica qué pruebas y medidas pertenecen al mismo grupo y cuáles miden prácticamente lo mismo. Por lo tanto, reduce el número de variables y ayuda a localizar o identificar unidades o propiedades fundamentales en que se deben basar las pruebas.

b) Representación matemática

- **Método Varimax.**- Es un método de rotación que minimiza el número de variables con cargas altas en un factor, mejorando así la interpretación de factores.

En consecuencia, el *método Varimax* determina la Matriz B de forma que maximice la suma de las varianzas:

$$V = p \sum_{i=1}^k \sum_{j=1}^p \left(\frac{b_{ij}}{h_j} \right)^2 - \sum_{i=1}^k \left(\sum_{j=1}^p \frac{b_{ij}^2}{h_j^2} \right)^2$$

De ello resulta una mayor facilidad de interpretación de los resultados.

Una vez determinado el número de factores a conservar y realizadas las rotaciones factoriales, se efectúan las puntuaciones factoriales que permiten determinar en qué medida los factores se dan en los departamentos. El cálculo de las puntuaciones factoriales se realiza a partir de la matriz factorial rotada y se basa en el modelo de la regresión múltiple, de acuerdo con la fórmula:

$$F_{ij} = F_{i1}Z_1 + F_{i2}Z_2 + \dots + F_{ip}Z_p = \sum_{i=1}^p F_{i1}Z_1$$

Siendo F_{ij} la puntuación factorial del departamento j en el factor i . Cada F_{ij} es la ponderación factorial de la variable I en el factor i . Las puntuaciones individuales en cada variable, en puntuaciones estandarizadas, vienen expresadas por Z_1 . La puntuación factorial individual F_{ij} es, por tanto, una nota Z que se refiere a la puntuación que el departamento j hubiera obtenido en el factor i .

El indicador propuesto es una combinación lineal de indicadores parciales en las áreas de Educación, Demografía, Vivienda Salud y empleo e ingreso. El indicador sintético sería el resultado de la suma ponderada de estos indicadores parciales.

$$I_{NV} = \beta_1 I_{edu} + \beta_2 I_{demo} + \beta_3 I_{viv} + \beta_4 I_{salud} + \beta_5 I_{ei}$$

Los indicadores parciales I_{edu} , I_{demo} , I_{viv} , I_{salud} , I_{ei} son las primeras componentes principales en las áreas Educación, Demográfico, Vivienda y Salud y empleo e ingreso respectivamente. Los coeficientes β_1 , β_2 , β_3 , β_4 y β_5 constituyen los autovalores correspondientes a las respectivas componentes principales.

Finalmente se realizará el análisis comparativo de los resultados con los otros indicadores a fin de establecer las similitudes y diferencias entre los departamentos según las cifras registradas por los 4 métodos.

5. RESULTADOS DE LOS INDICADORES SINTÉTICOS DEL NIVEL DE VIDA - 2017

Un indicador sintético es la función matemática de indicadores parciales los cuales pueden ser simples o componentes. El indicador resume y mide las manifestaciones, causas, características o efectos de un determinado fenómeno económico social.

Se entiende por componente de un indicador sintético o global aquel indicador que igualmente sintetiza el comportamiento de un conjunto de variables y que igualmente tiene un significado en forma independiente, al indicador global.

A diferencia de otros métodos de construcción de indicadores globales, que asignan la importancia de las variables o indicadores según el criterio del investigador, el análisis factorial por medio del método de componentes principales, permite construir el índice global en la forma de una suma ponderada de las primeras componentes principales.

Para el procesamiento de los indicadores se realizará la verificación de la correlación de las variables respecto de los niveles de vida. Por ello se modificará los datos estableciendo el siguiente criterio:

Si en caso que la variable se relacione directamente con los niveles de vida, los datos de la variable no serán modificadas, si la variable se relaciona inversamente con los niveles de vida, entonces se aplicará la inversa a los datos de dicha variable.

En los siguientes resultados se muestran las variables de acuerdo a los criterios anteriormente mencionados. Siguiendo dicha lógica además, los nombres de las variables serán modificadas.

5.1. INDICADORES DE EDUCACIÓN

Para desarrollar el siguiente índice se recoge la información de los siguientes indicadores:

- Inversa de Tasa de Analfabetismo
- Años promedios de estudios de la población de 15 años y más
- Tasa bruta de asistencia escolar de la población de 12 a 16 años de edad
- Tasa neta de asistencia escolar de la población de 12 a 16 años de edad
- Inversa de Población de 6 a 11 años que asiste a primaria con atraso escolar
- Inversa de Población de 12 a 16 años de edad que asiste a secundaria con atraso escolar.
- Niños evaluados en el segundo grado de primaria con nivel satisfactorio de comprensión lectora.
- Niños evaluados en el segundo grado de primaria con nivel satisfactorio de

razonamiento matemático.

- Población de 15 años y más con formación superior

Antes del realizar el análisis factorial, uno debe plantearse los siguiente: ¿Están correlacionadas entre si las variables originales? Si no lo estuvieran, no existirían factores comunes y, por lo tanto, no tendría sentido aplicar el análisis factorial.

Para dar respuesta a esta cuestión se examina el contraste de esfericidad de Bartlett y la medida de adecuación muestral de Kaiser, Meyer y Olkin (KMO).

El índice de KMO nos indica si la muestra se adecua para realizar un análisis factorial. Se espera obtener valores de KMO cercanos a 1 para considerar que la muestra se adecua excelentemente al modelo de análisis factorial, en cambio, para valores menores a 0.5, se considera a los datos inaceptables o muy malos para realizar el análisis factorial.

La prueba de Bartlett realiza una prueba estadística, donde la hipótesis nula a contrastar es que la matriz de correlaciones es igual a la matriz identidad, es otras palabras, que todos los coeficientes de correlación teóricos entre cada par de variables son nulos.

Prueba de KMO y Bartlett

Medida Kaiser-Meyer-Olkin de adecuación de muestreo		,733
Prueba de esfericidad de Bartlett	Aprox. Chi-cuadrado	198,192
	gl	36
	Sig.	,000

En este caso observamos un valor relativamente alto (0.733) para el índice KMO, el cual nos indica que los datos de la muestra se adecuan aceptablemente al modelo de un Análisis Factorial.

La prueba de Bartlett rechaza la hipótesis nula de que todos los coeficientes de correlación teóricos entre cada variable son nulos, lo que nos indica que existe un número significativo de correlaciones que son significativas.

Por lo tanto el supuesto de que las variables están correlacionadas entre si se cumple.

Comunalidades

	Inicial	Extracción
Inversa de Tasa de Analfabetismo	1,000	,734
Años promedios de estudios de la población de 15 años y más (%)	1,000	,906
Tasa bruta de asistencia escolar de la población de 12 a 16 años de edad	1,000	,805
Tasa neta de asistencia escolar de la población de 12 a 16 años de edad	1,000	,865
Inversa de Población de 6 a 11 años que asiste a primaria con atraso escolar (%)	1,000	,587
Inversa de Población de 12 a 16 años de edad que asiste a secundaria con atraso escolar (%)	1,000	,720
Niños evaluados en el segundo grado de primaria con nivel satisfactorio de comprensión lectora (2016) (%)	1,000	,820
Niños evaluados en el segundo grado de primaria con nivel satisfactorio de razonamiento matemático (2016) (%)	1,000	,618
Población de 15 años y más con formación superior (%)	1,000	,311

Método de extracción: análisis de componentes principales.

La comunalidad se define como la parte de la varianza que es debida a los factores comunes. Para las variables analizadas será:

- El 73.4% de la varianza de la variable *Inversa de Tasa de Analfabetismo* es explicada por los factores comunes retenidos.
- El 90.6% de la varianza de la variable *Años promedio de estudios de la población de 15 años y más* es explicada por los factores comunes retenidos.
- El 80.5% de la varianza de la variable *Tasa bruta de asistencia escolar de la población de 12 a 16 años de edad* es explicada por los factores comunes retenidos.
- El 86.5% de la varianza de la variable *Tasa neta de asistencia escolar de la población de 12 a 16 años de edad* es explicada por los factores comunes retenidos.
- El 58.7% de la varianza de la variable *Inversa de Población de 6 a 11 años que asiste a primaria con atraso escolar* es explicada por los factores comunes retenidos.
- El 72.0% de la varianza de la variable *Inversa de Población de 12 a 16 años de edad que asiste a secundaria con atraso escolar* de la población es explicada por los factores comunes retenidos.
- El 82.0% de la varianza de la variable *Niños evaluados en el segundo grado de primaria con nivel satisfactorio de comprensión lectora* es explicada por los factores comunes retenidos.

- El 61.8% de la varianza de la variable *Niños evaluados en el segundo grado de primaria con nivel satisfactorio de razonamiento matemático* es explicada por los factores comunes retenidos.
- El 31.1% de la varianza de la variable *Población de 15 años y más con formación superior* es explicada por los factores comunes retenidos.

Varianza total explicada

Componente	Total	Autovalores iniciales		Sumas de cargas al cuadrado de la extracción		
		% de varianza	% acumulado	Total	% de varianza	% acumulado
1	4,988	55,422	55,422	4,988	55,422	55,422
2	1,378	15,312	70,733	1,378	15,312	70,733
3	,935	10,394	81,127			
4	,714	7,928	89,055			
5	,501	5,562	94,617			
6	,316	3,509	98,126			
7	,094	1,041	99,167			
8	,062	,691	99,858			
9	,013	,142	100,000			

Método de extracción: análisis de componentes principales.

Uno de los métodos para la extracción de factores es el Método de Componentes Principales¹. Cuando se trabaja con datos tipificados o con la matriz de correlación, uno de los criterios que usualmente se utiliza, es retener los factores cuyo autovalor² (varianza del factor) sea mayor a 1.

Para el caso de los Indicadores de Educación se puede observar que se retienen los 2 primeros componentes con autovalores iguales a 4.988 y 1.378. Estos dos factores explican el 70.733% de la variabilidad total.

¹ El Método de Componentes Principales supone que el número de factores comunes es igual al número de variables originales.

² SPSS lo llama Eigenvalue

Matriz de componente^a

	Componente	
	1	2
Inversa de Tasa de Analfabetismo	,604	,608
Años promedios de estudios de la población de 15 años y más (%)	,848	,432
Tasa bruta de asistencia escolar de la población de 12 a 16 años de edad	,622	-,647
Tasa neta de asistencia escolar de la población de 12 a 16 años de edad	,864	-,345
Inversa de Población de 6 a 11 años que asiste a primaria con atraso escolar (%)	,713	-,281
Inversa de Población de 12 a 16 años de edad que asiste a secundaria con atraso escolar (%)	,776	,343
Niños evaluados en el segundo grado de primaria con nivel satisfactorio de comprensión lectora (2016) (%)	,906	,007
Niños evaluados en el segundo grado de primaria con nivel satisfactorio de razonamiento matemático (2016) (%)	,744	-,254
Población de 15 años y más con formación superior (%)	,536	,155

Método de extracción: análisis de componentes principales.

a. 2 componentes extraídos.

La Matriz de Componentes nos muestra las cargas factoriales, es decir, la correlación de cada variable con los factores retenidos.

Se espera que cada variable original este correlacionado fuertemente con un solo factor, para que se pueda dar una clara interpretación de los factores. Caso contrario, se procede a realizar una rotación de factores, que obtiene a partir de la solución inicial, unos factores que sean fácilmente interpretables.

Para el caso de los Indicadores de Educación se observa que los factores comunes tienen una interpretación clara ya que se puede ver con cuales variables está correlacionados cada factor. Podemos observar que todas las variables están relacionadas con el primer factor, en mayor medida la variable *Niños evaluados en el segundo grado de primaria con el nivel satisfactorio de comprensión lectora*, mientras que dicha variable es la única que no se relaciona fuertemente con el segundo factor.

Con esta matriz podemos hallar las puntuaciones de los factores retenidos de la siguiente manera:

$$\text{Factor 1} = 0.121 X_1 + 0.170 X_2 + 0.125 X_3 + \dots + 0.107 X_9$$

$$\text{Factor 2} = 0.441 X_1 + 0.314 X_2 - 0.469 X_3 + \dots + 0.113 X_9$$

Donde:

X_1 : Inversa de Tasa de Analfabetismo

X_2 : Años promedios de estudios de la población de 15 años y más

X_3 : Tasa bruta de asistencia escolar de la población de 12 a 16 años de edad

X_4 : Tasa neta de asistencia escolar de la población de 12 a 16 años de edad

X_5 : Inversa de Población de 6 a 11 años que asiste a primaria con atraso escolar

X_6 : Inversa de Población de 12 a 16 años de edad que asiste a secundaria con atraso escolar

X_7 : Niños evaluados en el segundo grado de primaria con nivel satisfactorio de comprensión lectora

X_8 : Niños evaluados en el segundo grado de primaria con nivel satisfactorio de razonamiento matemático

X_9 : Población de 15 años y más con formación superior

Donde las variables X_i son las variables originales tipificadas.

Luego se calcula el Índice de Educación como una combinación lineal de los factores retenidos ponderados por su respectiva varianza, de la siguiente manera:

$$\text{Índice de Educación} = 4.988 (\text{Factor1}) + 1.378 (\text{Factor2})$$

Tabla de Puntuaciones de los Factores y Valor del Índice de Educación

Departamento	Factor 1	Factor 2	IE=4.988*F1+1.378*F2	IE*=(IE+20)	ÍNDICE FINAL
Moquegua	1.86	0.08	9.4	29.4	100.00
Tacna	1.89	-0.50	8.8	28.8	97.82
Ica	1.32	1.01	8.0	28.0	95.08
Callao	1.07	1.57	7.5	27.5	93.55
Arequipa	1.42	-0.23	6.8	26.8	91.01
Lima	0.86	0.87	5.5	25.5	86.62
Pasco	0.44	-0.57	1.4	21.4	72.75
Tumbes	0.09	0.40	1.0	21.0	71.41
Puno	0.30	-0.42	0.9	20.9	71.16
Lambayeque	-0.24	1.55	0.9	20.9	71.14
Madre de Dios	0.02	0.56	0.9	20.9	71.06
Junín	0.10	-0.17	0.3	20.3	68.90
Cusco	0.22	-0.89	-0.1	19.9	67.59
Ayacucho	0.09	-1.48	-1.6	18.4	62.66
La Libertad	-0.52	0.62	-1.8	18.2	62.05
Piura	-0.69	0.95	-2.1	17.9	60.82
Apurímac	0.00	-1.71	-2.4	17.6	59.97
Áncash	-0.48	-0.62	-3.2	16.8	56.99
Amazonas	-0.64	-0.45	-3.8	16.2	55.04
San Martín	-1.17	1.22	-4.2	15.8	53.84
Huancavelica	-0.44	-1.65	-4.5	15.5	52.84
Ucayali	-1.23	1.19	-4.5	15.5	52.67
Huánuco	-1.09	-0.91	-6.7	13.3	45.25
Cajamarca	-1.24	-1.12	-7.7	12.3	41.72
Loreto	-1.92	0.72	-8.6	11.4	38.79

Elaboración: propia

Dónde: IE = Índice de Educación

Los procedimientos anteriores se realizan para los indicadores restantes, en los siguientes cuadros se muestran los resultados de los demás indicadores.

5.2. INDICADORES DEMOGRÁFICOS

Para desarrollar el siguiente índice se recoge la información de los siguientes indicadores:

- Esperanza de vida al nacer (años de vida)
- Inversa de Tasa de mortalidad infantil
- Inversa de Tasa global de Fecundidad
- Inversa de Población menor de 15 años de edad
- Inversa de Tasa de dependencia demográfica

Tabla de Puntuaciones de los Factores y Valor del Índice Demográfico

Departamento	Factor 1	ID=4.560*F1	ID*=(ID+20)	ÍNDICE FINAL
Callao	2.02	9.21	29.21	100.0
Lima	1.61	7.35	27.35	93.6
Moquegua	1.56	7.10	27.10	92.8
Arequipa	1.29	5.88	25.88	88.6
Ica	1.12	5.11	25.11	86.0
Tacna	0.95	4.33	24.33	83.3
Tumbes	0.85	3.89	23.89	81.8
Lambayeque	0.61	2.78	22.78	78.0
La Libertad	0.49	2.24	22.24	76.1
Áncash	-0.05	-0.24	19.76	67.6
Madre de Dios	-0.09	-0.42	19.58	67.0
Piura	-0.22	-1.00	19.00	65.0
Cajamarca	-0.27	-1.23	18.77	64.3
San Martín	-0.37	-1.67	18.33	62.8
Cusco	-0.50	-2.26	17.74	60.7
Ucayali	-0.55	-2.51	17.49	59.9
Junín	-0.60	-2.73	17.27	59.1
Pasco	-0.66	-2.99	17.01	58.2
Amazonas	-0.80	-3.66	16.34	55.9
Huánuco	-0.84	-3.84	16.16	55.3
Puno	-0.85	-3.89	16.11	55.1
Loreto	-0.89	-4.05	15.95	54.6
Ayacucho	-1.00	-4.58	15.42	52.8
Apurímac	-1.02	-4.64	15.36	52.6
Huancavelica	-1.79	-8.18	11.82	40.4

Elaboración: propia

Dónde: ID = Índice Demográfico

5.3. INDICADORES DE VIVIENDA

Para desarrollar este índice se recoge, en todos los departamentos del Perú, la información de los siguientes indicadores:

Tabla de Puntuaciones de los Factores y Valor del Índice de Vivienda

Departamento	Factor 1	Factor 2	$IV=3.751 \cdot F1+1.009 \cdot F2$	$IV^*=(IV+20)$	ÍNDICE FINAL
La Libertad	0.95	4.17	7.78	27.78	100.00
Lima	1.48	-0.19	5.36	25.36	91.27
Callao	1.41	-0.19	5.08	25.08	90.27
Arequipa	1.51	-0.83	4.81	24.81	89.31
Tacna	1.28	-0.46	4.34	24.34	87.61
Ica	1.32	-0.73	4.20	24.20	87.10
Moquegua	0.99	-0.55	3.17	23.17	83.41
Lambayeque	0.76	-0.02	2.84	22.84	82.22
Áncash	0.50	-0.13	1.75	21.75	78.28
Tumbes	0.22	-0.80	0.03	20.03	72.11
Junín	0.07	-0.26	-0.02	19.98	71.93
Piura	0.06	-0.44	-0.23	19.77	71.16
Cusco	-0.23	0.36	-0.49	19.51	70.21
Apurímac	-0.48	1.11	-0.68	19.32	69.54
Ayacucho	-0.34	0.21	-1.06	18.94	68.16
San Martín	-0.30	-0.52	-1.63	18.37	66.13
Madre de Dios	-0.26	-0.67	-1.67	18.33	65.98
Huancavelica	-1.05	0.93	-3.00	17.00	61.19
Cajamarca	-0.85	0.15	-3.03	16.97	61.10
Pasco	-0.93	0.42	-3.06	16.94	60.97
Puno	-0.89	-0.21	-3.55	16.45	59.22
Amazonas	-1.03	-0.19	-4.06	15.94	57.39
Huánuco	-1.11	-0.06	-4.22	15.78	56.82
Ucayali	-1.13	-0.58	-4.81	15.19	54.67
Loreto	-1.96	-0.49	-7.84	12.16	43.77

Elaboración: propia

Dónde: IV = Índice de Vivienda

- Viviendas que cuentan con abastecimiento de agua potable por red pública (%)
- Viviendas con acceso de servicio higiénico (%)
- Viviendas con acceso a red de alumbrado eléctrico por red pública (%)
- Viviendas con hacinamiento (%)

- Viviendas inadecuadas (%)
- Viviendas con piso de cemento (%)
- Viviendas con techo de concreto (%)

5.4. INDICADORES DE SALUD

Para desarrollar este índice se recoge, en todos los departamentos del Perú, la información de los siguientes indicadores:

- Inversa de Número de Habitantes por cada médico (Cant. de personas)
- Inversa de Población que reportó padecer algún problema de salud crónico (%)
- Inversa de Tasa de desnutrición crónica de niños menores de cinco años de edad

Tabla de Puntuaciones de los Factores y Valor del Índice de Salud

Departamento	Factor 1	IS=1.777*F1	IS*=(IS+20)	ÍNDICE FINAL
Tacna	2.63	4.67	24.67	100.0
Lima	1.77	3.15	23.15	93.9
Arequipa	1.61	2.86	22.86	92.7
Moquegua	1.04	1.85	21.85	88.6
Callao	0.90	1.60	21.60	87.6
Ica	0.88	1.57	21.57	87.4
La Libertad	0.65	1.16	21.16	85.8
Lambayeque	0.30	0.53	20.53	83.2
Tumbes	0.05	0.09	20.09	81.4
Cusco	-0.11	-0.20	19.80	80.3
Áncash	-0.17	-0.31	19.69	79.8
Junín	-0.27	-0.48	19.52	79.1
Puno	-0.31	-0.56	19.44	78.8
Apurímac	-0.37	-0.66	19.34	78.4
San Martín	-0.45	-0.79	19.21	77.8
Pasco	-0.46	-0.81	19.19	77.8
Piura	-0.50	-0.89	19.11	77.5
Amazonas	-0.52	-0.93	19.07	77.3
Huánuco	-0.64	-1.14	18.86	76.4
Loreto	-0.70	-1.24	18.76	76.0
Ucayali	-0.72	-1.28	18.72	75.9
Madre de Dios	-0.92	-1.63	18.37	74.5
Ayacucho	-1.11	-1.97	18.03	73.1
Cajamarca	-1.14	-2.02	17.98	72.9
Huancavelica	-1.45	-2.57	17.43	70.7

Elaboración: propia

Dónde: IS = Índice de Salud

5.5. INDICADORES DE EMPLEO E INGRESO

Para desarrollar este índice se recoge, en todos los departamentos del Perú, la información de los siguientes indicadores:

- PEA que tiene formación secundaria (%)
- PEA que tiene formación superior (%)
- PEA ocupada con seguro de salud (%)
- Inversa de PEA ocupada que es independiente (%)
- Inversa de PEA ocupada que es agropecuaria (%)
- Ingreso promedio de los ocupados (soles)

Tabla de Puntuaciones de los Factores y Valor del Índice de Empleo e Ingreso

Departamento	Factor 1	Factor 2	$IEI=3.751 \cdot F1+1.018 \cdot F2$	$IEI^*=(IEI+20)$	ÍNDICE FINAL
Callao	2.53	3.78	13.32	33.32	100.00
Lima	1.25	-0.20	4.48	24.48	73.45
Ica	1.37	-0.65	4.46	24.46	73.42
Arequipa	1.31	-0.99	3.92	23.92	71.78
Tacna	1.12	-1.50	2.68	22.68	68.05
Moquegua	0.85	-0.88	2.29	22.29	66.89
Madre de Dios	0.91	-1.17	2.24	22.24	66.74
Junín	0.25	-0.29	0.64	20.64	61.95
Lambayeque	0.14	0.04	0.57	20.57	61.73
La Libertad	0.19	-0.43	0.29	20.29	60.88
Ucayali	0.12	-0.40	0.06	20.06	60.20
Tumbes	-0.14	0.13	-0.39	19.61	58.85
Áncash	-0.19	0.12	-0.58	19.42	58.27
Pasco	-0.32	0.30	-0.89	19.11	57.35
Piura	-0.30	-0.29	-1.42	18.58	55.75
Loreto	-0.56	0.00	-2.09	17.91	53.75
Puno	-0.41	-0.67	-2.22	17.78	53.36
Apurímac	-0.93	0.81	-2.67	17.33	52.01
Cusco	-0.61	-0.39	-2.70	17.30	51.92
San Martín	-0.81	0.19	-2.84	17.16	51.49
Ayacucho	-0.89	0.46	-2.87	17.13	51.41
Huancavelica	-1.27	1.25	-3.48	16.52	49.59
Huánuco	-1.02	0.13	-3.68	16.32	48.96
Amazonas	-1.12	0.29	-3.90	16.10	48.32
Cajamarca	-1.49	0.36	-5.23	14.77	44.33

Elaboración: propia

Dónde: IS = Índice de Empleo e Ingreso

5.6. Indicadores sintéticos de Niveles de vida

- Demográfico
- Educación
- Vivienda
- Salud
- Empleo e Ingreso

Departamento	Índice de Educación	Índice Demográfico	Índice de Vivienda	Índice de Salud	Índice de Empleo Ingreso
Moquegua	100.00	92.77	83.41	88.57	66.89
Tacna	97.82	83.27	87.61	100.00	68.05
Ica	95.08	85.96	87.10	87.42	73.42
Callao	93.55	100.00	90.27	87.55	100.00
Arequipa	91.01	88.60	89.31	92.68	71.78
Lima	86.62	93.63	91.27	93.85	73.45
Pasco	72.75	58.23	60.97	77.77	57.35
Tumbes	71.41	81.76	72.11	81.41	58.85
Puno	71.16	55.14	59.22	78.81	53.36
Lambayeque	71.14	77.97	82.22	83.21	61.73
Madre de Dios	71.06	67.02	65.98	74.46	66.74
Junín	68.90	59.12	71.93	79.14	61.95
Cusco	67.59	60.74	70.21	80.25	51.92
Ayacucho	62.66	52.78	68.16	73.08	51.41
La Libertad	62.05	76.12	100.00	85.78	60.88
Piura	60.82	65.03	71.16	77.46	55.75
Apurímac	59.97	52.60	69.54	78.38	52.01
Áncash	56.99	67.64	78.28	79.82	58.27
Amazonas	55.04	55.93	57.39	77.31	48.32
San Martín	53.84	62.75	66.13	77.85	51.49
Huancavelica	52.84	40.45	61.19	70.66	49.59
Ucayali	52.67	59.89	54.67	75.90	60.20
Huánuco	45.25	55.32	56.82	76.45	48.96
Cajamarca	41.72	64.26	61.10	72.88	44.33
Loreto	38.79	54.61	43.77	76.05	53.75

BIBLIOGRAFÍA

1. F. Tusell (2016). Análisis Multivariante.
2. Joseph F. Hair, Rolph E. Anderson, Ronald L. Tatham, William C. Black. (1999). ANÁLISIS MULTIVARIANTE (5° Edición). Madrid: Prentice Hall International.
3. R. Ramirez Anormaliza, E. Carrasquero Rodríguez (2017). Análisis Multivariante: Teoría y Práctica de las principales Técnicas.
4. E. Uriel Jimenez y J. Aldás Manzano. (2005). Componentes Principales. En Análisis Multivariante Aplicado(365-394). España: Thomson.
5. E. Uriel Jimenez y J. Aldás Manzano. (2005). Análisis Factorial. En Análisis Multivariante Aplicado(407-440). España: Thomson.
6. J. Montenegro Fernandez (2018). Manual abreviado de Análisis Estadístico Multivariante.
7. C. M. Cuadras. (2018). Nuevos Métodos de Análisis Multivariante. Barcelona: CMC Editions.
8. Martin D. Farrell (1991). El Nivel de Vida
9. Selim Jahan, Eva Jespersen (2015). Programa de las Naciones Unidas para el Desarrollo (PNUD). Informe sobre Desarrollo Humano 2015.
10. M. Joseph Sirgy (2001). Medición y método de evaluación de la calidad de vida de la comunidad.
11. Instituto Nacional de Estadística e Informática (INEI) 2018. Evolución de la Pobreza Monetaria 2007 – 2017.
12. Grande Esteban, I. & Abascal, E. (1996) Fundamentos y Técnicas de Investigación Comercial. Madrid: ESIC.
13. Lerner, M. (1994) Métodos y procedimientos de Investigación de Mercados. Lima: CIUP.
14. Malhotra, N. (1997) Investigación de Mercados: Un enfoque práctico. México: Prentice Hall.
15. McDaniel, C. & Gates, R. (1999) Investigación de Mercados Contemporánea. México: Thomson.

16. Miquel, S. & Bigné, E. & Cuenca, A. & Miquel, J. & Lévy, J. (1999) Investigación de Mercados. Madrid: McGraw Hill.
17. Namakforoosh, M. (1995) Metodología de la Investigación. México: Limusa.
18. Instituto Nacional de Estadística e Informática (INEI). Encuesta Nacional a Hogares 2017.
19. Banco Mundial, “Una mirada a la evolución reciente de la pobreza en el Perú: avances y desafíos”, Lima, 2009.
20. Quispe Llanos Renán Jesús “Análisis de la Pobreza y distribución del Ingreso 2009 -2016.
21. Quispe Llanos Renán Jesús “Análisis de la Pobreza y distribución del Ingreso 2009 -2016.

ANEXOS

		INDICADORES DEMOGRÁFICOS				
Indicadores		Esperanza de vida al nacer (años de vida)	Inversa de Tasa de mortalidad infantil (%)	Inversa de Tasa global de Fecundidad (%)	Inversa de Población menor de 15 años de edad (%)	Inversa de Tasa de dependencia demográfica (%)
Departamento						
Amazonas		71.46	0.05	0.37	0.03	0.02
Áncash		74.67	0.06	0.42	0.04	0.02
Apurímac		71.22	0.05	0.34	0.03	0.02
Arequipa		76.98	0.08	0.53	0.04	0.03
Ayacucho		71.80	0.05	0.34	0.03	0.02
Cajamarca		73.84	0.06	0.42	0.03	0.02
Callao		78.31	0.13	0.56	0.04	0.03
Cusco		71.28	0.04	0.42	0.04	0.02
Huancavelica		70.82	0.04	0.28	0.03	0.02
Huánuco		72.53	0.05	0.36	0.03	0.02
Ica		77.58	0.11	0.48	0.04	0.03
Junín		72.69	0.06	0.37	0.03	0.02
La Libertad		76.37	0.08	0.45	0.04	0.02
Lambayeque		76.78	0.06	0.48	0.04	0.03
Lima		77.80	0.11	0.53	0.04	0.03
Loreto		72.65	0.05	0.37	0.03	0.02
Madre de Dios		73.18	0.05	0.42	0.04	0.03
Moquegua		76.51	0.08	0.53	0.04	0.03
Pasco		72.41	0.05	0.37	0.03	0.02
Piura		74.79	0.06	0.40	0.03	0.02
Puno		71.33	0.04	0.38	0.03	0.02
San Martín		72.34	0.06	0.40	0.03	0.02
Tacna		75.02	0.08	0.50	0.04	0.03
Tumbes		74.87	0.09	0.48	0.04	0.03
Ucayali		71.88	0.05	0.38	0.03	0.02

Elaboración: propia

INDICADORES DE EDUCACIÓN (%)									
Indicadores Departamento	Inversa de Tasa de Analfabetismo	Años promedios de estudios de la población de 15 años y más	Tasa de asistencia a secundaria de la población de 12 a 16 años de edad		Inversa de Población de 6 a 11 años que asiste a primaria con atraso escolar	Inversa de Población de 12 a 16 años de edad que asiste a secundaria con atraso escolar	Niños evaluados en el segundo grado de primaria con nivel satisfactorio de comprensión lectora (2016)	Niños evaluados en el segundo grado de primaria con nivel satisfactorio de razonamiento matemático (2016)	Población de 15 años y más con formación superior
			Tasa bruta de asistencia escolar de la población de 12 a 16 años de edad	Tasa neta de asistencia escolar de la población de 12 a 16 años de edad					
Amazonas	0.11	8.75	92.48	79.61	0.03	0.04	40.40	38.70	20.24
Áncash	0.10	9.51	95.91	85.67	0.03	0.04	37.30	26.20	24.95
Apurímac	0.06	9.18	96.59	87.21	0.04	0.04	38.00	35.10	22.25
Arequipa	0.22	10.93	97.66	90.45	0.04	0.05	59.00	38.00	38.74
Ayacucho	0.08	9.21	97.13	86.37	0.03	0.04	52.10	48.60	22.91
Cajamarca	0.08	8.21	93.72	81.00	0.03	0.03	34.10	31.90	15.53
Callao	0.42	11.06	92.68	84.78	0.03	0.04	60.70	44.60	35.12
Cusco	0.08	9.59	96.65	84.36	0.03	0.05	46.50	37.00	27.73
Huancavelica	0.07	8.73	97.08	87.36	0.03	0.04	41.70	40.50	16.57
Huánuco	0.07	8.59	96.11	76.89	0.03	0.03	31.90	28.30	20.32
Ica	0.36	10.99	97.21	90.35	0.03	0.06	52.10	39.70	35.65
Junín	0.16	9.83	95.74	84.23	0.03	0.04	47.80	40.30	28.62
La Libertad	0.16	9.72	91.40	79.38	0.03	0.04	39.80	30.50	28.74
Lambayeque	0.15	10.02	87.22	79.22	0.03	0.05	48.30	35.80	28.05
Lima	0.40	11.24	96.04	88.26	0.03	0.05	53.30	35.80	0.00
Loreto	0.11	9.08	91.18	73.81	0.02	0.03	17.70	12.40	20.88
Madre de Dios	0.19	9.91	93.91	85.13	0.03	0.05	41.30	26.60	27.72
Moquegua	0.20	10.80	98.10	90.95	0.03	0.06	69.20	53.70	39.18
Pasco	0.12	9.87	95.82	84.78	0.04	0.05	45.00	35.50	26.41
Piura	0.12	9.40	87.56	77.86	0.03	0.04	45.80	37.80	24.53
Puno	0.10	9.82	95.72	85.97	0.03	0.05	47.20	38.80	26.25
San Martín	0.11	8.65	86.38	71.87	0.03	0.05	38.50	30.90	20.40
Tacna	0.26	10.67	98.71	90.04	0.03	0.05	76.80	64.30	33.79
Tumbes	0.26	9.99	95.51	85.27	0.03	0.04	33.60	21.40	29.75
Ucayali	0.19	9.38	89.18	76.28	0.03	0.04	25.60	15.90	22.03

Elaboración: propia

Indicadores	INDICADORES DE VIVIENDA (%)					
	viviendas con acceso a servicios básicos agua, servicios higiénicos y energía eléctrica (2016)			Inversa de Viviendas inadecuadas	Viviendas con piso de cemento (2016)	Viviendas con techo de concreto (2016)
	Departamento	Viviendas que cuentan con abastecimiento de agua potable por red pública	Viviendas con acceso de servicio higiénico			
Amazonas	85.60	54.00	79.00	0.05	34.60	7.70
Áncash	95.60	72.90	95.70	0.22	46.30	25.50
Apurímac	89.80	52.30	93.10	0.68	20.40	10.20
Arequipa	94.30	84.00	97.20	0.25	71.70	68.90
Ayacucho	94.00	57.30	89.60	0.17	28.60	17.10
Cajamarca	83.00	46.60	86.60	0.24	30.50	13.70
Callao	95.80	87.20	99.70	0.41	55.70	60.30
Cusco	91.40	72.20	90.70	0.24	25.00	10.80
Huancavelica	87.00	42.00	87.70	0.47	14.00	7.10
Huánuco	74.60	43.50	84.30	0.21	32.80	15.50
Ica	91.50	82.20	97.80	0.28	69.10	59.60
Junín	92.10	64.30	92.10	0.09	42.20	23.60
La Libertad	91.00	75.50	95.50	3.33	48.60	38.80
Lambayeque	90.10	79.80	96.50	0.44	50.30	40.00
Lima	96.10	93.50	99.30	0.33	49.70	67.00
Loreto	55.30	41.60	77.90	0.05	32.60	0.80
Madre de Dios	84.90	49.60	91.10	0.13	62.00	7.10
Moquegua	95.10	83.70	92.90	0.21	59.40	51.80
Pasco	78.20	57.50	86.80	0.36	23.80	5.50
Piura	84.80	65.20	95.50	0.08	46.70	16.60
Puno	65.80	54.10	88.40	0.20	32.70	21.40
San Martín	88.10	49.60	91.50	0.04	50.70	11.00
Tacna	94.30	88.40	95.90	0.34	62.20	57.40
Tumbes	81.70	67.60	97.30	0.05	60.00	12.40
Ucayali	67.50	44.10	87.10	0.05	43.50	4.40

Elaboración: propia

INDICADORES DE SALUD (%)			
Indicadores	Inversa de Número de Habitantes por cada médico (2016)	Inversa de población que reportó padecer algún problema de salud crónico (2016)	Inversa de Tasa de desnutrición crónica de niños menores de cinco años de edad (2016)
Departamento			
Amazonas	0.001	0.025	0.052
Áncash	0.001	0.024	0.058
Apurímac	0.001	0.026	0.050
Arequipa	0.004	0.030	0.159
Ayacucho	0.001	0.043	0.053
Cajamarca	0.001	0.041	0.038
Callao	0.002	0.023	0.169
Cusco	0.002	0.032	0.068
Huancavelica	0.001	0.050	0.030
Huánuco	0.001	0.032	0.052
Ica	0.002	0.025	0.142
Junín	0.001	0.028	0.049
La Libertad	0.003	0.023	0.082
Lambayeque	0.002	0.025	0.085
Lima	0.004	0.026	0.185
Loreto	0.001	0.034	0.042
Madre de Dios	0.001	0.060	0.120
Moquegua	0.002	0.020	0.222
Pasco	0.001	0.023	0.040
Piura	0.001	0.033	0.065
Puno	0.001	0.025	0.061
San Martín	0.001	0.027	0.083
Tacna	0.003	0.029	0.435
Tumbes	0.001	0.025	0.136
Ucayali	0.001	0.033	0.040

Elaboración: propia

INDICADORES DE EMPLEO E INGRESO						
Indicadores Departamento	PEA que tiene formación secundaria (%)	PEA que tiene formación superior (%)	PEA ocupada con seguro de salud (%)	Inversa de PEA ocupada que es independiente (%)	Inversa de PEA ocupada que es agropecuaria (%)	Ingreso promedio de los ocupados (soles)
Amazonas	34.82	19.35	81.97	0.02	0.02	967.11
Áncash	43.42	25.29	76.83	0.03	0.03	1090.04
Apurímac	40.31	19.91	87.39	0.02	0.02	900.81
Arequipa	42.26	41.82	60.62	0.03	0.07	1545.02
Ayacucho	40.39	22.84	84.48	0.02	0.02	902.23
Cajamarca	33.07	14.39	81.23	0.02	0.02	843.35
Callao	51.35	38.99	73.28	0.04	0.85	1598.49
Cusco	37.81	25.71	73.95	0.02	0.02	1064.71
Huancavelica	41.70	15.16	90.96	0.02	0.02	709.05
Huánuco	35.40	20.41	78.90	0.02	0.02	933.14
Ica	50.26	38.34	62.02	0.03	0.06	1363.68
Junín	42.26	29.25	68.97	0.03	0.03	1135.65
La Libertad	39.89	29.70	68.95	0.03	0.04	1256.48
Lambayeque	44.65	29.62	74.53	0.03	0.05	1113.18
Lima	49.03	35.21	71.19	0.03	0.03	1548.27
Loreto	42.95	22.77	79.67	0.02	0.03	1167.29
Madre de Dios	48.10	28.61	60.12	0.02	0.05	1669.27
Moquegua	39.83	41.41	71.50	0.03	0.04	1689.68
Pasco	44.14	25.94	81.17	0.02	0.02	1051.09
Piura	40.26	27.24	72.25	0.02	0.03	1047.49
Puno	43.64	24.94	64.63	0.02	0.02	818.76
San Martín	35.24	21.53	81.31	0.02	0.02	1128.75
Tacna	45.87	33.86	49.32	0.03	0.06	1354.23
Tumbes	44.30	32.24	83.88	0.02	0.06	1260.84
Ucayali	48.90	24.88	68.69	0.02	0.05	1166.85

Elaboración: propia

