

## **Regional Comparative Evaluation: Synthetic Regional Development Index (RDI) for Peru**

**Margaret Coaquira**

Faculty of Economic Engineering, National University of Altiplano, Puno, Peru

E-mail: [mcoaquirav@est.unap.edu.pe](mailto:mcoaquirav@est.unap.edu.pe)

**Juan Tudela**

Faculty of Economic Engineering, National University of Altiplano, Puno, Peru

E-mail: [jtudela@unap.edu.pe](mailto:jtudela@unap.edu.pe)

**Miguel Jiménez**

Faculty of Industrial Engineering, Academic Department of Operations Research, National University of Piura, Piura, Peru

E-mail: [mjimenezc@unp.edu.pe](mailto:mjimenezc@unp.edu.pe)

### **Abstract**

*The research's aim was to comparatively evaluate regional development in Peru in the period 2015-2019, through the multidimensional elaboration of a Regional Development Index (RDI) that involves the four interdependent components of sustainable development (economic, social, environmental and institutional). For this, it has been rigorously followed the Principal Component Analysis (PCA), the Jenks Natural Breaks method, the sigma convergence and the Cronbach's Alpha coefficient. The results show that certain regions (Callao, Ica, Moquegua, Lima) triple the development of others (Cajamarca, Huancavelica, Puno, Loreto), despite the fact that the latter have large sources of development; one reason is due to the low progress in the institutional dimension. In hierarchical order, the dimensions with the greatest contribution to the RDI (degree of association of the dimensions with respect to the RDI) are social (95.20%), environmental (95.98%), economic (84.78%) and institutional (49.42%). In addition, the results show the existence of a regional sigma convergence, which indicates that regional disparities decreased in the 2015-2019 period. Likewise, Cronbach's Alpha Coefficient (0.90) reports the existence of internal reliability in the methodology used. Finally, the design of a web platform prototype that iteratively shows Peru regional development is included, using intelligent algorithms, big data, web scraping and geospatial information.*

**Keywords:** Dimensions, Economic Development, Gaps, Regions, Sustainable Development

**JEL Classification:** C38, O13, O18, O32, O47

Received: 13 July 2022; Received in revised form: 31 August 2022; Accepted: 01 September January 2022

## 1. Introduction

The political relevance in the development metrics of the economies has been going through a process of change, since the fact that development is multidimensional and cannot be measured through a single indicator is accepted and debated with greater preponderance by the international community. Under this premise, it is commendable to go back to the mid-1930s, when a “new national accounting” appeared to measure the development of economies, since after the Great Depression, more than 80 years ago, the Gross Domestic Product (GDP) was the indicator with the greatest presence when assessing the income -or wealth creation- of an economy. Formulated by the American economist Simon Kuznets, the GDP -an indicator that measures the production generated by a country taking into account a certain period of time and geographical borders- at that time became a widely accepted measure of the prosperity of a nation. In the same vein, the GDP per capita (GDP pc) arises, connoting itself as a more pressing indicator for the quantification of development, since it measures the production among the total number of inhabitants of a nation.

Nevertheless; despite the fact that both indicators enjoyed -and still enjoy- great acceptance in academia, in recent decades several analysts have criticized GDP. For example, as a consequence of the financial crisis of 2008, the economist Stiglitz et al. (2010, p.15) presented their report entitled “Mis-measuring our lives”, in which the limitations of GDP as an indicator of well-being are revealed and they explicitly state “(...) if we use the wrong metrics, we will also make wrong decisions”. As well as this work, there are two other initiatives that share the same conception, one carried out by the OECD (Global project for measuring the progress of societies) and another carried out by the European Commission through a communication from the Commission to Council and the European Parliament (Beyond the GDP).

The fact that development is multidimensional and cannot - neither should - be estimated using a single indicator, such as GDP, is accepted by the international community, especially by the United Nations with the formulation of the 2030 Agenda for Sustainable Development. The 2030 Agenda, awarded by the United Nations Development Programme-UNDP (2015), represents an important achievement, which -so far- has managed to forge a global consensus on development priorities. However, the scope and complexity of the 17 Sustainable Development Goals (SDG) and the 169 targets together with their indicators set out in the Agenda can be overwhelming for policymakers to get their bearings. In the same vein, Stiglitz et al. (2018, p.20) in their section included in “For Good Measure” (the continuation of the report “Mis-measuring our lives” by Stiglitz et al. (2010)) indicates that in order to achieve the SDG it is necessary to select a delimited panel of multidimensional indicators in order to guide the planning and elaboration of policies in the countries.

For Peru, multidimensionality to measure development is also important, given that it is still a developing country, it needs to have an accurate diagnosis of its level of development, so that, based on it, the proposed public policies can be controlled from time to time based on reliable and aggregate indicators according to the 2030 Agenda. Besides, since the fact that this country does not have a composite index that includes the four dimensions of sustainable development and that there is also no proposal on the web to be able to estimate it iteratively, makes the estimate even more interesting of a composite index.

There are initiatives from various organizations that attempt to measure development through a “summary” of relevant indicators. The Human Development Index (HDI), for example, is a synthetic indicator that since 1990 measures human development through three variables (income, health and education). Also, from 2010 to 2014, four international composite indicators came to light to measure poverty (Multidimensional Poverty Index), inequality (Human Development Adjusted Inequality Index), gender empowerment (Gender Inequality Index) and gender development (Gender Development Index). Another clear example is the Regional Competitiveness Index of the European Union prepared by the European Commission (2019, p.3-15), which evaluates the conditions levels of 268 European Union, through 74 indicators reflected in 11 dimensions of competitiveness, which -in turn- are organized into three large categories: basic factors (institutions, macroeconomic stability, infrastructure, health and basic education), efficiency (advanced education, learning by outcomes, workforce efficiency and market size), and innovation (technological readiness, sophisticated business and innovation). These indices are of a national nature, it means, their official analysis is based on the comparison of countries, leaving aside the internal comparison (interdepartmental, interregional, intermunicipal, as it may apply). Therefore, the new challenges of evaluating development internally in a country have been constituting a new field of research.

Thus, studies such as Zaman and Goschin (2014, p.217-225) emerge: “A new classification of Romanian counties based on a composite index of economic development”, which covered the performance of regional development in Romania in the period 2001-2012, where under a methodology of normalization and weighting of variables, a regional development index was built considering two influential factors of development, the accession of Romania to the European Union and the financial economic crisis of 2008. Continuing with the analysis in Romania, Goschin (2015, p.103-110) presented the study entitled “Regional divergence in Romania based on a new index of economic and social development”, in which, having recognized the problem of spatial disparities, the research objective was to build a multidimensional index that captures the economic and social aspect of the 42 cities in Romania, to later evaluate the regional convergence in the long term, this through a normalization of data and arithmetic calculations concerning the proposed variables, as well as the sigma ( $\sigma$ ) convergence, whose method includes Augmented Dickey Fuller and Phillips Perron tests.

With regard to Latin America, Vial (2019) developed in its third edition the Regional Development Index (known in Spanish as IDERE) for Chile; built as a tool that measures development at the territorial level from a multidimensional perspective through a geometric measure of normalized indices between 0 and 1 (where 0 means the minimum development and 1 the maximum), this index considers 32 variables and 7 key dimensions that have been necessary to verify the territorial inequalities and existing gaps in Chile, these 7 dimensions are education, health, socioeconomic well-being, economic activity, connectivity, security, and sustainability and environment. In the same vein, Aboal et. al (2018, p.9-30) prepared for Uruguay the study entitled “Analysis of territorial inequities based on synthetic indicators”, which -also- had the objective of building a Departmental Development Indicator (known in Spanish as IDD) that evaluates the territorial disparities of Uruguay. The dimensionality of the index considered 4 dimensions: citizen security and reliable legal system; prepared and healthy influential society; efficient and dynamic factor markets; and physical and technological infrastructure; which in turn were made up of 18 simple indicators. This synthetic

index was built by adopting the methodology of the Mexican Institute for Competitiveness (known in Spanish as IMCO) in the elaboration of the State Competitiveness Index (ICE) -another multidimensional index- that includes a weighting of the simple indicators by expert opinion and by a multidimensional analysis (ACP).

In reference to Peru, the Peruvian Institute of Economy (2019) (known in Spanish as IPE) presents since 2012 the Regional Competitiveness Index (known in Spanish as INCORE), the only regional measuring instrument currently available in Peru. However, the main weakness of this proposal is not to collect information on the environmental dimension and, as previous lines pointed out, one of the fundamental pillars for sustainable development is environmental conservation. It should be noted that this environmental conception is considered as a priority area by the SDG of the UNDP and by the Well-Being Framework proposed by the OECD, so it is worth estimating a synthetic index considering this environmental dimension.

Then, taking into account that in Peru there is still no a recent composite indicator that considers the four dimensions of sustainable development, and that it can also have an automatic execution scope on the web through an interactive platform, the purpose of this research is to contribute with the creation of a multidimensional index with these characteristics. Therefore, the objective of this article is to comparatively evaluate regional development in Peru, through the multivariate elaboration of a synthetic indicator, called the Regional Development Index (RDI or also known in Spanish as IDR), around 4 dimensions; economic, social, environmental and institutional. The reference period for the estimation of the RDI captures the data from 2015-2019. In addition, with the design and estimation of the RDI, a comparative analysis is considered at the regional level with emphasis on dimensional contrasts, levels of regional convergence, and methodological validation. Also, taking into account the above, in the research it is agreed to involve the design of a prototype of an iterative web platform, which serves as a didactic tool that generates valuable evaluation – reflected through the RDI and its dimensions-, mainly, for the public politics makers.

The present paper has the following organization: in section (1) the introduction beside the literature review is presented, through the background at global, regional and national (for Peru) levels. In section (2), corresponding to the materials and methods, the data and the methodology used in each specific objective is described. In section (3) are the results and, finally, in section (4) are the conclusions.

## **2. Materials and methods**

The research had a quantitative approach with a deductive method, basic, longitudinal and retrospective. In terms of scope, 25 units of analysis are included. The information sources are detailed below.

Table 1 - Variable Information Sources (INEI)

Source	Information obtained
<i>National Population and Housing Census 2017</i>	Census population
National Household Survey (known in Spanish as ENAHO)	
• Module I: Characteristics of the home and household	Number of household members
• Module II: Characteristics of household members	Age, gender, marital status, geographic domain, natural region, residence area, department.
• Module III: Education	Education level
• Module V: Employment and income	Informality levels, underemployment levels, company size, household income earners, branch of activity, occupational category, average monthly income from work
• Module XXXIV: <i>Sumarias</i>	Minimum referential income

Source: Author's elaboration.

## 2.1. Variables selection (simple indicators)

Taking into account the "Manual for the construction of composite indicators" prepared by Nardo et al. (2008, p.23-68) and published by the OECD, and the "Methodological Guide: Design of composite indicators of sustainable development" prepared by Schuschny and Soto (2009, p.27-33) and published by ECLAC, the selection of variables (21 simple indicators) had three main criteria: (i) concomitance with the dimensions (each dimension must have relevant variables), (ii) relationship with the objectives and, consequently, goals and indicators of the 2030 Sustainable Development Goals, since within the framework of research, the construction of the RDI is concomitant with the 2030 Agenda and (iii) availability of information (period 2015-2019 and 25 units of analysis).

Table 2 - Selected variables

Dimension	Variable (simple indicator)	Connotation	Relation with SDG	Availability
Economic	Real Gross Domestic Product per capita	PIBpc	SDG 8	Information Available
	Average monthly income from work	In_pm	SDG 8, 10	Information Available
	Occupancy rate	T_ocu	SDG 8	Information Available
	% PEA employed properly employed	PEA_ae	SDG 8	Information Available
	% Employed PEA affiliated with retirement pension systems	Pens	SDG 1, 2, 8	Information Available
	% Pop over 18 years of age who registers having at least one financial product	Prd_fnc	SDG 1,9,10	Information Available
			SDG 4	Information Available
Social	Illiteracy rate of the pop. from 15 years to more	T_anf	SDG 4	Information Available
	Average years of study achieved by the pop. 15 years of age and older	Añ_esc	SDG 4	Information Available

	Chronic malnutrition rate of children under 5 years of age (NCHS reference pattern)	T_dsnt	SDG 3, 2	Information Available
	% Children 6-59 months of age with total anemia	Anm	SDG 3, 2	Information Available
	Births attended by specialized health personnel	Prt_espc	SDG 3	Information Available
	% Dwellings with overcrowding (unsatisfied basic need 2)	Hcina	SDG 1, 11	Information Available
	% Pop in dwellings without drainage of any kind (unsatisfied basic need 3)	Sin_dsg	SDG 6	Information Available
	% Households with internet access	Acs_e	SDG 9, 17	Information Available
	% Urban households that properly dispose of their inorganic solid household waste	Ad_rs	SDG 11, 13	Information Available
	% Pop. with sustainable access to improved sources of water supply	Acs_stagu	SDG 6, 11, 13	Information Available
	% Municipality that have environmental management instruments	Mun_instam	SDG 13, 15, 17	Information Available
	% Municipality that carried out actions to encourage environmental conservation	Mun_incam	SDG 13, 15, 17	Information Available
Institutional	% Municipality that have a transparency portal	Mun_portr	SDG 16	Information Available
	% Municipality that have computerized systems implemented to support the management	Muni_sistinf	SDG 9, 16	Information Available
	% Municipality who have reported having implemented the Municipal Office for Attention to Persons with Disabilities	Muni_ofid	SDG 10, 11, 16	Information Available

Note: The selection of 21 variables (simple indicators) had the three main criteria indicated.

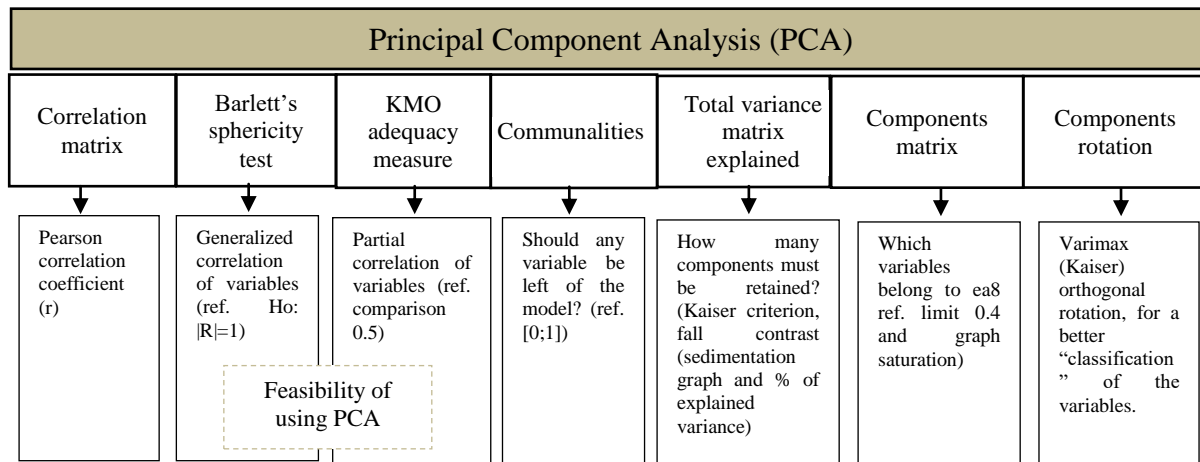
## 2.2. Methodology for RDI elaboration

The method chosen for the RDI elaboration corresponds to Principal Component Analysis (PCA), which aims to capture the highest possible variance in the variables (simple indicators) with the fewest possible components. The process presented below (Figure 1) requires the transformation of the set of selected variables into a reduced set of "new" synthetic variables (called main factors or components). As well as this method, there are other techniques that have the purpose of summarizing several variables in just one; nevertheless, the PCA method was chosen because in addition to representing one of the most standard techniques in different research fields (biology, medicine, economics, etc.), it also represents one of the techniques with the most detailed phases in its estimation, which makes the results more exact. Likewise, this method is frequently used in the construction of composite indicators in large well-known organizations (Business Climate Indicator,



General Indicator of Science and Technology and Relative Intensity of Regional Problems in the Community).

Figure 1 - Principal Component Analysis Phases



Source: Nardo et al. (2008, p.23-68) y Greco et al. (2018, p.61-94).

## 2.3. Methodology for comparative analysis: Natural Breaks and sigma convergence

### 2.3.1. Jenks Natural Breaks optimization method

In order to facilitate the comparison of the analysis units around their RDI's, three levels of relative development are established: high, medium and low. To delimit the extension of the ranges (intervals) the Jenks natural breaks method (cuts or natural thresholds) is used, which enables a better arrangement of data in different classes, thus ensuring that the classification of the data depends on their nature and distribution. Thus, it pursues the dual purpose of extracting classes with great internal homogeneity and with maximum differences between classes for the number of intervals previously specified (Jenks, 1967).

### 2.3.2. Sigma convergence

To capture the trend of regional disparities of the units of analysis and, in this way, determine the fluctuations in the behavior of the regional gaps, the sigma ( $\sigma$ ) convergence method is used. Formulated by Barro and Sala-i-Martin (1995, p.1-3), it aims to capture the trend of regional disparities based on the territorial dispersion of significant development indicators. According to León (2013, p.68), sigma convergence indicates that the dispersion of income distribution tends to decrease over time and, therefore, what is expected is that the differences or disparities between different economies will also decrease.

## 2.4. Cronbach's Alpha Coefficient Methodology

Cronbach's alpha coefficient (1951, p.297-334) reveals a value that measures internal consistency, that is, it indicates how well the information of various variables is represented in a single composite indicator. The coefficient takes values included in the interval [0,1]. Cronbach's alpha coefficient is calculated through:

$$\alpha = \frac{p}{p-1} \left( \frac{\sigma_I^2 - \sum_{i=1}^p \sigma_{x_i}^2}{\sigma_I^2} \right) \quad (1)$$

where:

$\alpha$ : Cronbach's Alpha coefficient.

I: Composite indicator.

p: Variables.

$\sigma_I$ : Variance of the indicator (I).

$\sigma_{x_i}$ : Variance of each of the p variables.

In this way, the estimator measures the fraction of total variability of the sample of variables as a result of their correlation. In equation (1), if there is no correlation and the variables are independent of each other, the value of  $\alpha$  will be null (0). Therefore, the closer the estimator  $\alpha$  is to 1, the reliability of the selection of proposed variables is affirmed to be better, while if this estimator is closer to 0, the opposite is affirmed. It should be noted that an acceptable reliability is considered from 0.70.

## 3. Results

### 3.1. Regional Development Index (RDI)

#### 3.1.1. Multivariate analysis (PCA)

Regarding to the first step in PCA, that is, with respect to the correlation matrix estimated with the Pearson coefficient for the period 2015-2019, on average, the variables show strong correlations with each other with significance levels between 0.01 and 0.05. Among the strongest and most significant linear correlations, the following variables stand out: PBIpc, In\_pm, T\_ocu, PEA\_ae, Pens and Prd\_fnc. The correlations between these variables agree with the dimensional classification they share with each other (economic). On the other hand, variables related to the social dimension such as: T\_anf and Añ\_esc –as expected– present a strong negative linear correlation significant at 0.05; while the variables related to health (T\_dsnt and Anm) present a positive but not significant correlation. Regarding the variables concomitant to housing (Hcina, Sin\_dsg and Acs\_e), all –without exception– present strong and significant correlations at 0.05. Regarding the environmental dimension, all the variables present linear correlations, however, the correlation that stands out due to its significance at 0.05 occurs between the variables: Mun\_instam and Mun\_incam Regarding the variables belonging to the institutional dimension, Mun\_portr and Mun\_ofid were found to have a strong and positive correlation with a significance level of 0.05.



According to the second step in PCA, with the Bartlett's sphericity test, through Table 3, it is observed that the value of Chi square is high and the significance (P-value) is less than 0.05 in all years, which is why the use is feasible. from PCA. Regarding the KMO sample adequacy measure, the analyzed coefficient takes values greater than 0.60 in all years, exceeding the statistical fence of 0.50 and indicating that the partial correlations between the variables are sufficiently small, which is why the PCA application is appropriated.

Table 3 - Barlett's sphericity test and KMO sample adequacy measure

Bartlett and KMO test		Years				
		2015	2016	2017	2018	2019
Barlett's sphericity	Aprox. Chi-squared	507.076	591.999	584.206	552.484	539.758
	Gl	210	210	210	210	210
	Sig.	0.000	0.000	0.000	0.000	0.000
Kaiser-Meyer-Olkin Measure		0.634	0.673	0.622	0.613	0.601

Source: Author's calculations.

Regarding to the third step, in order to verify the relevance of the selected variables, the communalities associated with each variable were estimated for all years. The extracted communalities, presented in Table 4, mostly have values greater than 0.75, so the existence of common factors that explain the variabilities of the variables is inferred, that is, the model reproduces on average more than 75% of the original variability of all variables.

Table 4 - Extracted communalities

Variables	Initial	Years				
		2015	2016	2017	2018	2019
PIBpc	1.000	0.617	0.674	0.599	0.570	0.799
In_pm	1.000	0.872	0.892	0.884	0.891	0.880
T_ocu	1.000	0.617	0.670	0.639	0.617	0.709
PEA_ae	1.000	0.858	0.851	0.898	0.895	0.897
Pens	1.000	0.890	0.878	0.889	0.900	0.912
Prd_fnc	1.000	0.498	0.619	0.762	0.637	0.696
T_anf	1.000	0.824	0.827	0.868	0.899	0.875
Añ_esc	1.000	0.838	0.885	0.899	0.899	0.920
T_dsnt	1.000	0.799	0.762	0.817	0.870	0.884
Anm	1.000	0.722	0.759	0.735	0.662	0.786
Prt_esp	1.000	0.892	0.924	0.891	0.810	0.907
Hcina	1.000	0.753	0.876	0.709	0.714	0.827
Sin_dsg	1.000	0.804	0.846	0.787	0.772	0.868
Acs_e	1.000	0.920	0.937	0.903	0.899	0.947
Ad_rs	1.000	0.671	0.753	0.737	0.791	0.828
Acs_stagu	1.000	0.829	0.842	0.835	0.795	0.810
Mun_instam	1.000	0.767	0.854	0.922	0.790	0.746
Mun_incam	1.000	0.841	0.871	0.916	0.793	0.858
Mun_portr	1.000	0.783	0.789	0.816	0.807	0.737
Mun_sistinf	1.000	0.719	0.500	0.853	0.602	0.628
Mun_ofid	1.000	0.850	0.875	0.848	0.832	0.812

Note: Extraction method: PCA

Source: Author's calculations.

The fourth phase of PCA is based on determining the number of factors that should be retained in the study. For this, the variance matrix explained by the components was constructed and, since the PCA was applied for each year in the analysis period (2015-2019), five matrices were estimated, from which the averages presented in Table 5 were obtained. To determine the number of retained components, the Kaiser criterion (latent root), the fall contrast (Castell's elbow) and the percentage of explained variance were jointly considered. According to the first criterion, four components with eigenvalues that exceed unity are retained, making sense to retain components that explain more variance than a single variable can contain.

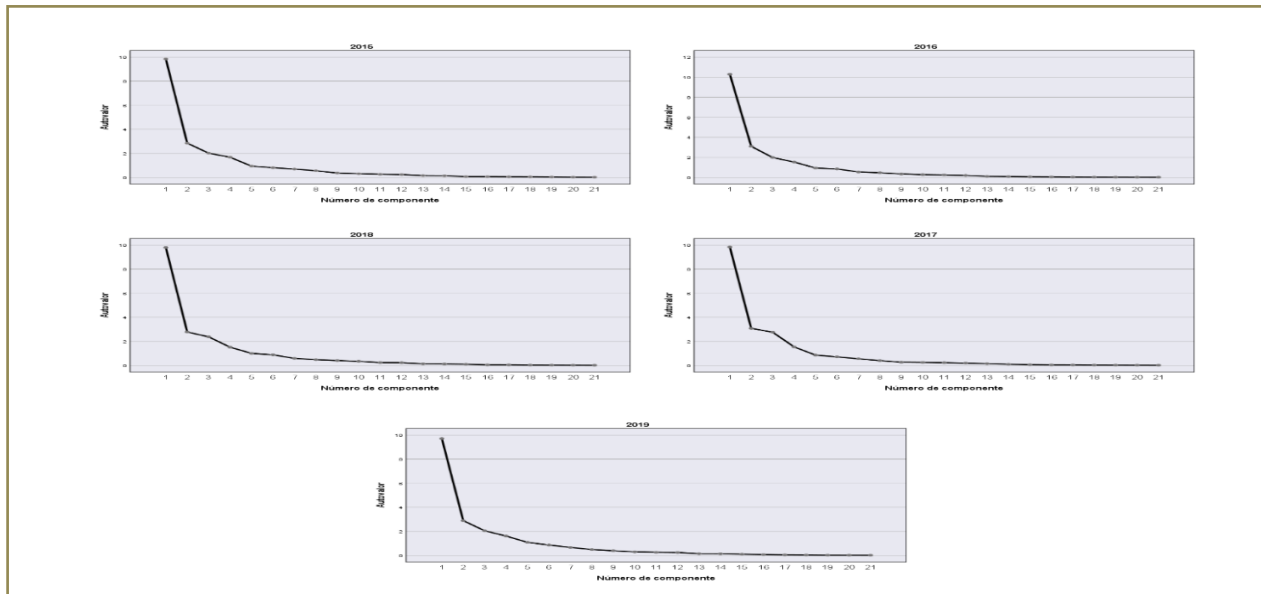
Table 5 - Average total explained variance matrix, 2015-2019

Component	Initial eigenvalues			Sums of squared extraction loads		
	Total	% Variance	% Accumulated	Total	% Variance	% Accumulated
1	9.896	47.126	47.126	9.896	47.126	47.126
2	2.934	13.971	61.097	2.934	13.971	61.097
3	2.229	10.616	71.713	2.229	10.616	71.713
4	1.569	7.470	79.183	1.569	7.470	79.183
5	0.958	4.563	83.746			
6	0.815	3.883	87.628			
7	0.601	2.862	90.490			
⋮	⋮	⋮	⋮			
19	0.014	0.068	99.947			
20	0.007	0.034	99.981			
21	0.004	0.019	100.000			

Source: Author's calculations.

Regarding the second criterion, the sedimentation graphs show –for all the years– inflection points after consigning the fourth component (Figure 2), ratifying the selection of the first criterion.

Figure 2 - Sedimentation graph, 2015-2019



Source: Author's calculations.

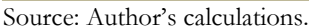
According to the third criterion, through Table 5, it is observed that the average accumulated percentage of the variance when considering four components is 79.18%, that is, the selected components explain –together– 79.18% of the total variance observed in twenty-one initial indicators throughout the five years and since it exceeds 60% (reference percentage in research of a social nature), so, once again, the selection of four components is ratified.

The fifth phase comprises a matrix that shows the correlation between the selected components and the original variables in order to assign a component to each of them. The classification of the components can be done from the component matrix taking into account the highest correlative values (minimum threshold of 0.40); however, through the estimations, it was observed that most of the factor loads of the original variables are correlated with the first component, leaving aside the other three components and not allowing a clear interpretation of these, which is why estimated the component matrix with rotation for each year.

So, as a result of the transformation, the global accumulated variance for the selected components remains invariable with the value of 79.18%; however, what does change is the distribution of the variance between the components, since it changes from 9.89 (47.13%) to 6.73 (32.02%), from 2.93 (13.97%) to 4.06 (19.34%), from 2.23 (10.62 %) to 3.16 (15.04%) and from 1.57 (7.47%) to 2.68 (12.77%) for the first, second, third and fourth components, respectively.

Also, the saturation graphs (Figure 3) consider the components as the axes from which the load values for each variable are projected in relation to one of the components. Regarding component 1-2, it is observed that at one end of the (horizontal) axis of component one the variables appear: PIB\_pc, In\_pm, PEA\_ae, Pens, Añ\_Esc, and Acs\_e, while at the other end are: T\_anf, T\_dsnt and T\_ocu; showing the contrast between not developing human capital and its "effects" on the

Figure 3 - Component Two-dimensional Saturation plot (2015-2019)



Based on the rotated component matrices, the variables that make up each principal component are:

- First component (32.02%): PIB\_pc, In\_pm, T\_ocu, PEA\_ae, Pens, T\_anf, Añ\_esc, T\_dsnt, Acs\_e.
- Second componente (19.34%): Prt\_esp, Hcina, Sin\_dsg, Acs\_stagu
- Third componente (15.14%): Mun\_instam, Mun\_incam, Mun\_portr, Mun\_sistinf, Mun\_ofid
- Fourth componente (12.77%): Prd\_fnc, Anm, Ad\_rs

After that, a data normalization is performed, this one was carried out through the re-scaling method. Given the nature of the data, the two normalization functions Min- Max and Max-Min, which took into account whether the target value of the variable was to reach a maximum (the higher, the better) or whether the target value for the variable was to reach a minimum (the lower, the better). Then, regarding the information weighting, the extract of the total explained variance matrix with the retained components, as well as their variance percentages, is presented in Table 6 below.

Table 6 -. Average total variance (2015-2019)

Component	Sums of loads squared of rotation		
	Total	% Variance	% Accumulated
1	6.725	32.023	32.023
2	4.061	19.340	51.363
3	3.159	15.044	66.407
4	2.683	12.775	79.183

Source: Author's calculations.

Based on the accumulated percentages, the distribution of percentages by components and variables is as follows in Table 7.

Table 7 - Weighting of variables according to retained components

Component	% Variance	% Variance accumulated	Variables	% RDI	% Accumulated RDI	% Per variable in RDI
1	32.023%	32.023%	PIB_pc, In_pm, T_ocu, PEA_ae, Pens, T_anf, Añ_esc, T_dsnt, Acs_e	40.442%	40.442%	4.494%
2	19.340%	51.363%	Prt_esp, Hcina, Sin_dsg, Acs_stagu	24.425%	64.867%	6.106%
3	15.044%	66.407%	Mun_instam, Mun_incam, Mun_portr, Mun_sistinf, Mun_ofid	18.999%	83.866%	3.800%
4	12.775%	79.182%	Prd_fnc, Anm, Ad_rs	16.134%	100.00%	5.378%

Source: Author's calculations.

The following procedure was an ordering according to the conceptual dimensions of the RDI, that is, a classification of the variables was carried out (with their weights assigned as a result of the entire process) according to the categories (dimensions) that they originally represent.

Table 8 - Weighting of variables according to dimensions of the RDI

Dimension	Variable	Individual % in RDI	Total % in RDI
Economic	PIBpc	4.494%	27.846%
	In_pm	4.494%	
	T_ocu	4.494%	
	PEA_ae	4.494%	
	Pens	4.494%	
	Prd_fnc	5.378%	
Social	T_anf	4.494%	41.671%
	Añ_esc	4.494%	
	T_dsnt	4.494%	
	Anm	5.378%	
	Prt_espc	6.106%	
	Hcina	6.106%	
	Sin_dsg	6.106%	
	Acs_e	4.494%	
Environmental	Ad_rs	5.378%	19.084%
	Acs_stagu	6.106%	
	Mun_instam	3.800%	
	Mun_incam	3.800%	
Institutional	Mun_portr	3.800%	11.400%
	Mun_sistinf	3.800%	
	Mun_ofid	3.800%	
TOTAL		100.00%	100.00%

Source: Author's calculations.

The last stage in the construction of the RDI was the aggregation of information of the weighted variables for each unit of analysis (regions of Peru) and for each year (2015-2019). For this, the technique used was linear, specifically, the weighted arithmetic mean.

## 3.2. Regional comparative analysis

### 3.2.1. RDI Global Comparison

Through Table 9, according to the estimated RDI, the ranking of the Peru's regions is presented.



Table 9 - Global RDI ranking, 2015-2019

Department	2015		2016		2017		2018		2019		Average	
	Rank	IDR	Rank	IDR	Rank	IDR	Rank	IDR	Rank	IDR	Rank	IDR
Callao	1	0.873	1	0.874	1	0.886	1	0.880	1	0.855	1	0.873
Ica	3	0.763	2	0.775	2	0.777	2	0.787	2	0.780	2	0.776
Moquegua	2	0.771	4	0.757	4	0.757	3	0.773	4	0.729	3	0.757
Lima	4	0.763	3	0.757	3	0.758	4	0.745	3	0.756	4	0.756
Arequipa	6	0.675	7	0.677	5	0.725	5	0.707	5	0.701	5	0.697
Tacna	5	0.675	8	0.674	7	0.676	6	0.695	6	0.698	6	0.684
Lambayeque	8	0.603	5	0.687	6	0.682	7	0.690	7	0.671	7	0.667
Tumbes	7	0.651	6	0.684	8	0.659	8	0.655	8	0.609	8	0.652
Piura	9	0.562	9	0.580	9	0.609	9	0.617	9	0.598	9	0.593
Madre de Dios	11	0.555	10	0.564	10	0.593	10	0.593	11	0.552	10	0.572
La Libertad	12	0.531	11	0.555	11	0.585	11	0.584	10	0.555	11	0.562
Cusco	10	0.558	12	0.542	14	0.528	12	0.527	14	0.497	12	0.530
Áncash	14	0.493	14	0.502	15	0.508	15	0.501	12	0.529	13	0.506
San Martín	16	0.455	13	0.529	13	0.533	14	0.504	13	0.510	14	0.506
Junín	13	0.507	15	0.496	12	0.536	16	0.499	15	0.488	15	0.505
Pasco	18	0.428	16	0.469	16	0.480	13	0.505	17	0.456	16	0.468
Ayacucho	15	0.466	18	0.462	17	0.466	18	0.441	19	0.443	17	0.456
Ucayali	17	0.440	17	0.463	19	0.453	17	0.459	20	0.397	18	0.443
Apurímac	19	0.412	19	0.415	18	0.459	19	0.430	16	0.464	19	0.436
Amazonas	22	0.330	21	0.368	20	0.394	20	0.404	18	0.452	20	0.389
Huánuco	20	0.375	22	0.363	21	0.392	21	0.392	22	0.374	21	0.379
Cajamarca	21	0.339	24	0.353	22	0.389	22	0.376	21	0.389	22	0.369
Huancavelica	23	0.325	20	0.382	23	0.380	24	0.327	23	0.339	23	0.351
Puno	24	0.320	23	0.357	24	0.359	23	0.355	24	0.328	24	0.344
Loreto	25	0.288	25	0.251	25	0.262	25	0.273	25	0.326	25	0.280
Perú		0.526		0.541		0.554		0.549		0.540		0.542

Source: Author's calculations.

Regarding the classification of the level of regional development: low, medium, high; Table 10 shows these three levels, along with their RDI intervals, which were estimated using the Jenks Natural Breaks method.

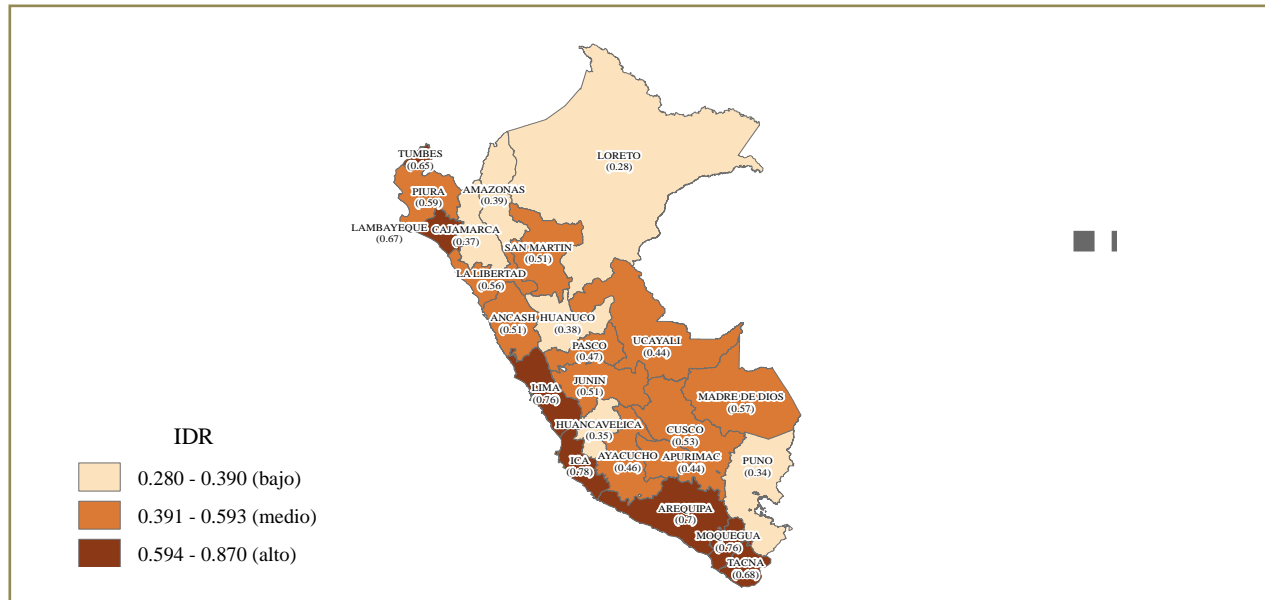
Table 10 - Average RDI levels

Level	RDI (Jenks Natural Breaks)
Alto	0.594 - 0.870
Medio	0.391 - 0.593
Bajo	0.280 - 0.390

Source: Author's calculations.

From the RDI spatial distribution, Through Figure it can be seen that the greatest concentration of development is located on the southern and central coast of the country, while the jungle and some regions of the sierra are the areas that display the worst positions of regional development.

Figure 4 - Spatial distribution of RDI average (2015-2019)

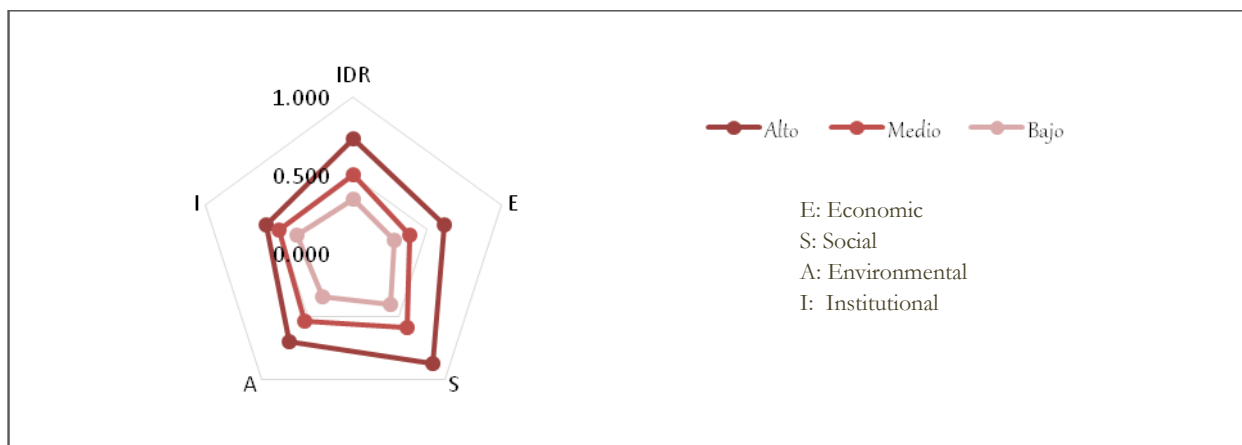


Source: Author's elaboration in ArcGIS 10.8.

### 3.2.2. Comparison by dimensions of the IDR (economic, social, environmental and institutional)

Figure 5 shows a radial graph that considers the average RDIs (2015-2019) of the levels of development (high, medium and low) according to the dimensions that make up the RDI. It is observed that the indices of the institutional dimension are closer to each other, while those of the social dimension are farther from each other. These two premises indicate that there is a smaller gap in the institutional dimension compared to the other dimensions. especially the social dimension.

Figure 5 - RDI and average dimensions by development levels (high, medium, low)



Source: Author's calculations.

### 3.2.3. Correlational analysis

Regarding the correlational analysis, Table 11 shows that Lambayeque was the region that obtained the best values (high statistically significant correlations), and also stands out with the highest correlational value in the environmental dimension (0.988) compared to the other regions, so part of its development is largely due to waste disposal (Ad\_rs), human habitat (Acs\_stagu) and environmental management (Mun\_instam, Mun\_incam). Piura, on the other hand, has the highest significant correlational value in the economic field (0.999), which shows that its development was increased by the economic structure (GDPpc), employment (In\_pm, T\_ocu, PEA\_ae), pension system (Pens) and financial system (Prd\_fnc). In the social aspect, Arequipa was the region with the highest correlational prevalence (0.961), which indicates progress in education (T\_anf, Añ\_esc), health (T\_dsnt, Anm, Prt\_Espc) and housing (Hcina, Sin\_dsg, Acs\_e). For its part, Ucayali has the highest significant correlation in the institutional dimension (0.983), which shows that its development was enhanced by management capacity (Mun\_portr), ICT infrastructure (Mun\_sistinf) and social inclusion (Mun\_ofid). Regions with correlations with negative signs are also observed, for example, for Áncash a limitation in its development is the institutional dimension (-0.815), for which policy makers must improve their actions in management capacity, municipal ICT infrastructure and social inclusion in municipal services.

Table 11 - Correlations between RDI and its dimensions (average for period 2015-2019)

Department	E	S	A	I
Amazonas	0.4339	0.8118*	0.9395**	0.635
Áncash	0.9388**	0.1466	0.8365*	*-0.8145*
Apurímac	0.9409**	0.7991	0.8582*	-0.0805
Arequipa	0.8098*	0.9606**	0.1528	0.5376
Ayacucho	-0.5191	0.2183	0.501	0.796
Cajamarca	0.9644**	0.9138**	-0.7758	0.5747
Callao	-0.4892	-0.4557	0.9617**	-0.0268
Cusco	0.7313	0.7473	0.8301*	0.3899
Huancavelica	0.2623	0.9462**	-0.0921	0.7763
Huánuco	0.5838	-0.267	-0.0706	0.9133**
Ica	0.7288	0.9096**	-0.2186	-0.4264
Junín	0.435	0.9358**	-0.3759	0.5336
La Libertad	0.9474**	-0.4697	0.0317	0.767
Lambayeque	0.9625**	0.8463*	0.9883**	0.9588**
Lima	-0.9211**	0.5923	0.862*	0.2215
Loreto	0.4207	0.2572	0.9176**	0.7079
Madre de Dios	0.7416	-0.0151	0.8940**	0.8970**
Moquegua	0.771	-0.5657	0.4663	0.9649**
Pasco	0.8529*	0.498	0.5909	-0.0267
Piura	0.9998**	0.5222	0.4761	0.5198
Puno	0.6744	0.341	0.6166	0.8873**
San Martín	0.6041	0.7876	0.6604	0.9263**

Tacna	0.7407	0.6228	0.6503	-0.4211
Tumbes	0.9538**	-0.6869	0.4578	0.6515
Ucayali	0.8219**	-0.0356	0.4896	0.9828**
Peru	0.8478**	0.9520**	0.8598**	0.4942**

\*\*The correlation is significant at 95% confidence.

\*The correlation is significant at 90% confidence.

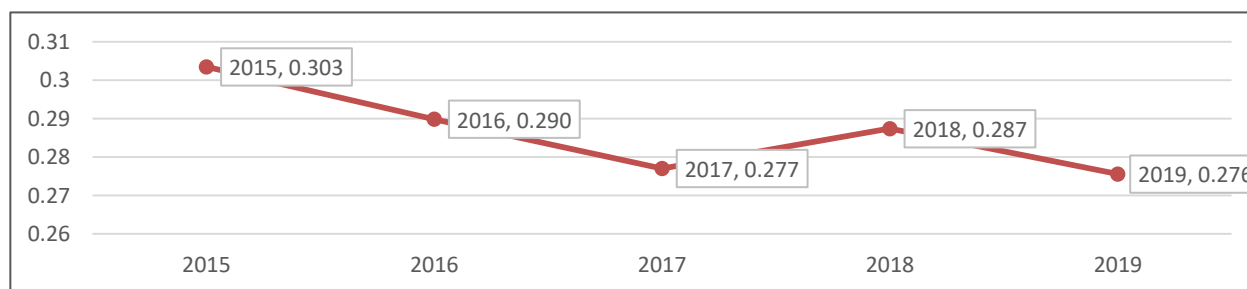
Source: Author's calculations.

It is highlighted that the social dimension has allowed to improve the levels of development in almost all the regions, because in addition to demonstrating the highest and statistically significant regional correlations, from the national level it is the highest (0.952). The next dimension that drives regional development is environmental (0.859), economic (0.848) and institutional (0.494), in hierarchical order. Likewise, the four dimensions are statistically significant at 5%. Regarding the institutional dimension, this ratifies what was found in the previously estimated institutional ranking, where the regions that occupied the last positions in the RDI came to light with medium developments in the institutional aspect, in the same way, the correlation found (0.494) for the institutional dimension is moderate and the lowest with respect to the other three dimensions.

### 3.2.4. Sigma convergence

In order to establish the level of regional gaps, the sigma ( $\sigma$ ) convergence methodology was incorporated, which captured the trend of regional disparities, indicating the fluctuations in the behavior of regional gaps in Peru. Figure 6 shows that the sigma value ( $\sigma$ ) computed for the 25 analysis units indicates a general trend of decay, thus demonstrating a regional sigma convergence, this means that regional disparities decreased in the period 2015-2019.

Figure 6 - Regional sigma convergence in Peru, 2015-2019



Source: Author's elaboration.

In order to support these results, we evaluated whether the calculated sigma is consistent with stationarity (Table 12 and Table 13). For this, the Dickery-Fuller Augmented (DFA) and Phillips Perron (PP) tests were used. Testing the null hypothesis ( $H_0$ ) in both tests, referring to the existence of a unit root and, consequently, non-stationarity of the time series, no unit root was found due to the t-statistic, so the preliminary results were validated. statistically, that is, the time series is stationary and, therefore, there is no sigma divergence, but regional sigma convergence.

Table 12 - Dickey Fuller Augmented test

Dickey-Fuller test for unit root				
	Test Statistic	Interpolated Dickey Fuller		
		1% Critical Value	5% Critical Value	10% Critical Value
Z(t)	-1.856	-3.75	-3	-2.63
MacKinnon approximate p-value for			Z(t) = 0.3528	
Source: Author's calculations in Stata 16.0.				

Table 13 - Phillips-Perron test

Phillips-Perron test for unit root				
	Test Statistic	Interpolated Dickey Fuller		
		1% Critical Value	5% Critical Value	10% Critical Value
Z(rho)	-3.384	-17.2	-12.5	-10.2
Z(t)	-1.856	-3.75	-3	-2.63
MacKinnon approximate p-value for			Z(t) = 0.3528	
Source: Author's calculations in Stata 16.0.				

### 3.3. Internal methodological reliability and iterative web platform prototype

#### 3.3.1. Internal reliability analysis

Through the estimation of Cronbach's Alpha coefficient, it was reported that the computed values for each year showed to be close to 1.0 and greater than 0.7 (minimum acceptable reliability). Table 14 shows these coefficients, indicating that the RDI collects objective information and, in addition to being reliable, makes stable and consistent measurements.

Table 14 - Cronbach's Alpha coefficient, 2015-2019

Year	Cronbach's Alpha coefficient
2015	0.910
2016	0.902
2017	0.904
2018	0.910
2019	0.882
Average	0.902

Source: Author's calculations.

#### 3.3.2. Iterative web platform prototype

The platform will have two main elements, the first corresponds to the selection of characteristics: (a) selection of geographic scope (regions, macro regions and/or Peru-general), (b) year of selection (2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022, etc.), (c) situational status (RDI, dimensions: economic, social, environmental and/or institutional) and (d) presentation of results (table, choropleth map, linear graph of evolution and /or radial graph); while the second element is based

on the illustration and presentation of the results according to the selection made within the first element. Regarding the transformation of variables, once the selection of characteristics has been made in the first element, the design of the platform will be subject to the present investigation, specifically to the estimated weightings of the variables, since, through the PCA, it was possible to calculate the weights of the variables that make up the RDI, these estimated weights will be used to estimate the following RDI's and dimensions on the platform. Likewise, regarding the illustration and presentation of results, through analytical and artificial intelligence, automatic update algorithms will be used through a web design scraping with Python language, in such a way that the extraction of information is obtained directly from the databases of INEI (microdata-modules), recoded and transformed. In this way, the automatic collection and updating of public data will make it possible to build reports, maps and graphs that show trends and comparisons of the regions of Peru around the RDI and its variables, according to year and region. The platform's design is presented in the Appendix.

According to the results found, as well as this research, there are other works that lead to results in the same line or in a different line. For example, considering the scale of [0 and 1]; where, the closer the coefficient is to 0, it represents a low level of development and, similarly, the closer it is to 1, it represents a higher level of development; it was found that the regions that occupied the first four positions are: Constitutional Province of Callao (0.86), Ica (0.78), Lima (0.76), Moquegua (0.73) and Arequipa (0.70); this result is similar to that presented by the Peruvian Institute of Economy (2020) in INCORE, since its report that, on scale of 0 to 10, the same regions are those with the highest levels of development for 2019: Lima (7.7, including the Constitutional Province of Callao), Moquegua (6.8), Tacna (6.7) and Arequipa (6.6). The same happens with respect to the regions with low levels of development in both indices (IDR, INCORE).

On the other hand, regarding the association analysis of the RDI and its dimensions, to verify which is the dimension that gives the most contribution (greatest correlation), as in the study by Correa and Morocho (2012), carried out for the period 2004-2010 in which only economic activity, physical capital, human capital and financial resource management are considered as dimensions, it is found that the social dimension (95.20%) and physical capital (97.6%) are the ones that have the greatest association with the RDI. Regarding the institutional dimension (49.42%) and management of financial resources (-0.1%), they are characterized by having the lowest, and even an inverse association with the RDI. In this context, taking into account the different period of analysis in both studies (2004-2010 for the research of Correa and Morocho (2012) and 2015-2019 for the present research), it is inferred that over the years, in Peru still has the institutional framework as a critical aspect, since the fact that this dimension has a moderate and even negative degree of association with development (reflected in the RDI) leads policymakers to place greater emphasis and attention in that dimension.

Likewise, in the research some limitations are recognized in the information provided by statistics and informatics institutions, which is why it is limited to 21 variables, since one of the criteria for the development of the study was that the variables have availability in the regions and in the proposed period of time.



## 4. Conclusions

The RDI elaboration, built through the Principal Components Analysis (PCA), allowed estimating with a multidimensional approach (economic, social, environmental and institutional) the level of development reached by the 24 regions of the country and the Constitutional Province of Callao, in the period 2015-2019. Regarding the general comparative analysis, the analysis units that ranked in the first four positions in the global RDI ranking (2015-2019) were the constitutional province of Callao (0.873), Ica (0.776), Moquegua (0.757) and Lima (0.756), while the regions that occupied the last four places were Cajamarca (0.369), Huancavelica (0.351), Puno (0.344) and Loreto (0.280). Regarding the analysis by dimensions, it was found that the regional indices of the institutional dimension are closer to each other, in contrast to those of the social dimension, thus indicating that in the institutional dimension there was a smaller gap between the development of regions with respect to the other dimensions, especially to the social dimension. In addition, by hierarchical order, the dimensions that contribute the most to sustainable development were: social (95.20%), environmental (85.98%), economic (84.78%) and institutional (49.42%). On the other hand, the sigma convergence analysis reported a general trend of decline, which evidenced a regional sigma convergence, that is, the regional disparities of Peru decreased in the 2015-2019 period. Finally, through Cronbach's Alpha coefficient, it was reported that the values computed for each year under analysis showed to be close to unity and on average revealed a coefficient of 0.902, higher than 0.700, so it is concluded the existence of internal methodological reliability in the RDI. Regarding the iterative web platform, it is expected that with the proposed design it can be implemented through genetic algorithms and artificial intelligence on the web, since, in this way, Peru would have a pioneering tool known within the indices of new generation.

In line with the proposed multidimensional development; there are main policy implications; for example, it is recommended to encourage access to financial products, since this variable represented a weakness for regions with medium and low levels in the economy development. Likewise, in the social aspect, given that anemia rate represents the main weakness for regions with medium social development, health policy initiatives are recommended around this variable, with great emphasis on this group of regions. Regarding the estimated correlational analysis, given that negative sign correlations of regions were observed, it is recommended to improve these values; for example, it can be seen that for the region Ancash, a limitation in its development is the institutional dimension (-0.815), so that the policymakers of this region must improve their actions in management capacity, municipal ICT infrastructure and social inclusion in municipal services. Similarly, the other regions with negative and significant correlations should improve their actions according to the dimension of the corresponding value.

Finally, within the investigative field of Principal Component Analysis (PCA), it is encouraged to develop various studies taking into account the multidimensionality of development in Peru; thus, the institutions responsible for reporting and disseminating statistics (INEI) are encouraged to continuously expand their results.

## Acknowledgment

This article was financed by CONCYTEC-FONDECYT in within the framework of the contest “Maria Reiche Mentoring Scholarships 2021-01” [contract number 015-2021].

## References

- Aboal D, Lanzilotta B, Pereyra M, Queraltó M (2018). Análisis de las inequidades territoriales a partir de indicadores sintéticos. Oficina de Planeamiento y Presupuesto. Programa Uruguay Integra, 9-30.
- Ballester S (2019). Convergencia en renta per cápita y productividad aparente del trabajo entre las Comunidades Autónomas españolas y el caso de las Islas Canarias en el periodo 2000-2017 (Memoria de trabajo de fin de grado). Universidad de La Laguna.
- Barro R, Sala-i-Martin X (1995). Technological diffusion, convergence and growth. National Bureau of Economic Research, 1-3.
- Camacho M, Horta R (2020). Metodologías para la construcción de índices compuestos. Departamento de Administración y Finanzas. Universidad Católica Del Uruguay, 1.
- European Comission (2019). The EU Regional Competitiveness Index. Piublicaciones Oficiales de La Unión Europea, 3-15.
- Cronbach J (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*. 16: 297–334.
- Goschin Z (2015). Regional divergence in Romania based on a new index of economic and social development. *Procedia: Economics and Finance*. 32(1): 103-110.
- Greco S, Ishizaka A, Tasiour M, Torrisi G (2018). On the methodological framework of composite indices: A review of the issues of weighting, aggregation and robustness. *Social Indicators Research*. 141(1): 61–94.
- National Institute of Statistics and Informatique. (2020). Sistema de Información Regional para la Toma de Decisiones (SIRTOD).
- Economic Peruvian Institute. (2019). Índice de Competitividad Regional (INCORE) de Perú. Instituto Peruano de Economía.
- Jenks G (1967). *The data model concept in statistical mapping*. International Yearbook of Cartography, 186–190.
- León G (2013). Crecimiento y convergencia económica: Una revisión para Colombia. *Revista Dimensión Empresarial*. 11(1): 61–76.

Nardo M., Saisana M., Saltelli A., Tarantola S, Anders H., Giovannini E (2008). Handbook on Contrusting Composite Indicators: Methodology and user guide. OECD Statistics Working Paper, 2005(3): 23-68.

United Nations Development Program. (2015). Agenda 2030 para el desarrollo sostenible. Naciones Unidas.

United Nations Development Program. (2019). El Reto de la Igualdad: Una lectura de las dinámicas territoriales en el Perú. Depósito Legal En La Biblioteca Nacional Del Perú.

United Nations Development Program. (2020). Plataforma de Análisis para el Desarrollo (PAD) para México.

Riesta L (2018). Las dimensiones del desarrollo sostenible como paradigma para la construcción de las políticas públicas en Venezuela. *Tekhné: Revista de La Facultad de Ingeniería*. 21(1):24-33.

Schuschny A, Soto H (2009). Guía metodológica: Diseño de indicadores compuestos de desarrollo sostenible. CEPAL-Colección de Documentos de Proyectos, 27-33.

Stiglitz J, Fitoussi J-P, Durand M (2018). For Good Measure: Advancing Research on Well-being Metrics Beyond GDP, 20.

Stiglitz J, Sen A, Fitoussi J-P (2010). *Mis-measuring our lives*. The New Press, 15.

Tello M (2021). Convergencia en crecimiento y brechas de productividad regional en el Perú: 2000-2020. Instituto Nacional de Estadística e Informática.

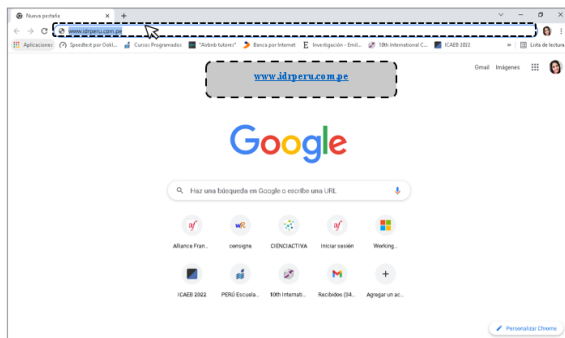
Vial C (2019). Índice de desarrollo regional para Chile (IDERE). Universidad Autónoma de Chile.

Zaman G, Goschin Z (2014). A new classification of Romain counties based on a composite index of economic development. *Annals of Faculty of Economics*. 1(1): 217–225.

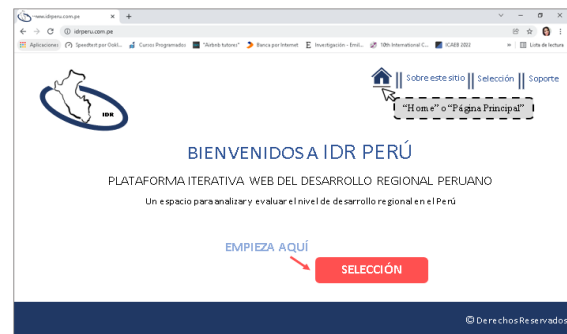
## Appendix

### Iterative web platform prototype

**1. Enter the platform:** To access the RDI platform, enter the following link in the browser of your choice



**2. Main page (Home):** As the first presentation within the website is the "home" or "main page" of the platform, where the user is welcomed and invited to make the selection of geographical area, year of analysis, IDR/dimension, etc.



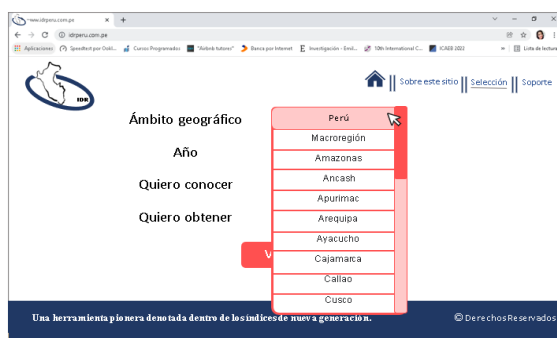
**3. About the site:** As a second window within the website is the "About the site" tab, where a brief description of the origin and development of the web platform is given.



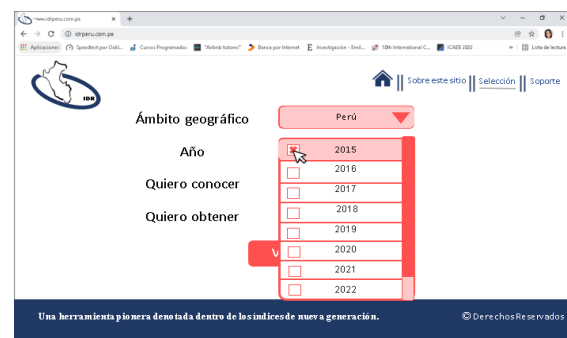
**4. Selection:** As a third window within the website is the "Selection" tab, where you can make the selection of (i) geographical area (regions, macro-regions and/or Peru-general), (ii) year of analysis (2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022, etc.), (iii) situational status/I want to know (IDR, dimensions: economic, social, environmental and/or institutional) and (iv) presentation of results (table, choropleth map, linear graph of evolution and/or radial graph).



**4.1 Selection-Geographic Scope:** In this section you can make the selection of (i) geographical scope (regions, macro-regions and/or Peru-general)



**4.2. Selection-Year of analysis:** In this section you can select the (ii) year of analysis (2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022, etc.).



**4.3. Selection-I want to know:** In this section it will be possible to make the selection of (iii) situational status/I want to know (IDR, dimensions: economic, social, environmental and/or institutional, simple indicators).

**4.4. Selection-I want to get:** In this section it will be possible to make the selection of (iv) presentation of results (table, choropleth map, linear graph of evolution and/or radial graph).

**4.5. Selection-View Results:** In this section you can see the results according to what was previously selected (according to geographical area, year of analysis, situational status/I want to know, presentation of results/I want to obtain).

Additionally, there is an option to export the results obtained, depending on the form of presentation (Excel and/or jpg).

**5. Support:** As a fourth window within the website is the "Support" tab, a contact space for the public interested in the platform.