CAPÍTULO 1

APPLICATION OF FACTOR ANALYSIS AND PRINCIPAL COMPONENT ANALYSIS IN THE CONSTRUCTION OF INDICATORS

Data de aceite: 01/09/2023

Carlos Alberto Paraguassú-Chaves
PhD in Health Sciences - University of Brasília - UnB, Brazil; PhD in Science - University of Havana (Cuba); Post-Doctor in Health Sciences - UnB and Degli Studi D'Aquila University - IT. Professor at the Federal University of Rondônia, Brazil

Josefa Lourdes Ramos
Mestre em Desenvolvimento Regional e Meio Ambiente pela Universidade Federal de Rondônia – UNIR; Researcher of the Institute of Health Sciences and the Amazon environment - AICSA.

Fabrício Moraes de Almeida
PhD in Physics (UFC), with post-doctorate in Scientific Regional Development (DCR/CNPq). Researcher of the Doctoral and Master Program in Regional Development and Environment (PGDRA/UNIR). Leader GEITEC — Federal University of Rondonia, Brazil

Fábio Robson Casara Cavalcante
4PhD and researcher linked to the graduate program in administration and the academic Department of Social and Environmental Sciences (DACSA) - Federal University of Rondônia, Brazil.

ABSTRACT - The objective of the research: to identify the significance of epidemiological, socioeconomic and coverage indicators of health services in Rondônia, applying the multivariate statistical technique. The method: multivariate statistical technique, using Factor Analysis and Principal Component Analysis method. A sample with 121 variables was used, divided into two scenarios: Scenario 1, with 121 epidemiological, socioeconomic and health service coverage variables in the State of Rondônia; and Scenario 2, with 42 epidemiological, socioeconomic and health service coverage variables with an explanation greater than 95%. Results: The statistical inferences show the following findings: factor (1) called epidemiological indicators, presented 69 variables with explanation above 70%, factor (2) called socioeconomic indicators highlighted 4 variables with explanation above 70% and factor (3) indicators of coverage of health services highlighted only 2 (two) variables with explanation above 70%. Scenario 2, composed of 42 variables with explanation above 95% in the application of Factor Analysis and Principal Component Analysis obtained 34 variables with explanation above 70%. Of these, 25 are of the factor

(1) epidemiological indicators and presented explanation above 95%. The variable with the greatest explanatory power is the first with 99.66% (referring to the % of children aged 10 to 14 years, with more than 1 year of school delay). The correlation between factors (1), (2) and (3) is explained by 66.53% for epidemiological indicators, 18.72% for socioeconomic indicators and 7.53% for service coverage indicators. From this study it is possible to infer the almost absolute predominance of epidemiological variables and, with less emphasis, of socioeconomic variables, which measure access to elementary education in the State of Rondônia. The variables of coverage of health services did not present significance. Conclusions: the study showed that of the 39 indicators agreed between the State Department of Health of Rondônia and the municipalities, 33 (thirty-three) are included among the indicators with statistical significance, which validates the choice of method.
KEYWORDS: Statistical and Health; Indicators; Multivariate Analysis; Rondônia; Brazil

1 I INTRODUCTION

In the process of building the Brazilian Unified Health System (SUS), develop research projects that have in their core the interest of researching political alternatives to foster the allocative policy of financial resources of the SUS, with a view to mitigating regional inequalities of access to health services, configures as essential, mainly to raise to the national manager the need to adopt financing policies that focus on the health needs and living conditions of the population, making prevail the maximum constitutional postulate of the Constitution of Brazil [1] “health a right of all and duty of the State”.

Given the proportions of the SUS system, public health policies have demonstrated the capacity for innovation with the implementation of the Basic Care Floor, despite the fact that the historical and cultural behavior of resource allocation based on production prevails, maintaining a process of distributive inequities. We seek to correct this situation through instruments such as the Integrated Agreed Programming (PPI), implemented since 2000, which uses a historical series of production to allocate resources in medium and high complexity actions, which comprise the SUS.

This new model of health care is governed by principles that seek to guarantee universal, integral, egalitarian, equitable and free access to health, consisting of a network organized in a regionalized and hierarchical way, which advocates a single command at each level of government, based on a strategy of administrative and operational decentralization of health actions and services and their legal provisions constitute the main instruments of regulation of the system for the transfer of resources to the health area between the three levels of government [2], [3], [4], [5], [6], [7], [8].

Considering the magnitude of the Brazilian State, the allocation of financial resources in health without observing regional inequalities, especially in the North and Northeast Regions, which are doubly peripheral, significantly accentuates the regional inequalities that reside in different levels of income concentration, low demographic density and low living conditions, pointed out through health indicators [9]. Thus, health inequalities are expressed in the different possibilities of life expectancy at birth, how to live, the ways of getting sick and dying for the same class in contrast to the rates of other regions of the country [9]. It is necessary to consider the regional differences regarding the characteristics of the installed service network, access to information and strategic planning instruments for the better development of health actions and services, and the investment capacity of the different States and regions in technological and managerial innovations, among other decisive factors in the performance of the critical areas evaluated in the process of reviewing financial ceilings [10].

It is evident the need to develop studies that point to public financing strategies in health that share with the particularities of the peripheral states of the country, and can from then on serve as a basis for (re)discussing the allocation of resources in the SUS, respecting the geographical, populational, socioeconomic and cultural aspects of each region of the country. The peculiar characteristics of the geographical space and the demographic distribution, as well as the inevitable problems of access and locomotion, to meet differentiated demands it is necessary to foresee modes of remuneration in line with the degrees of difficulties and time required in the execution of the work. A typical example to demonstrate the importance of this argument is in the vaccination action, essential for all Brazilians: for the sake of justice, it cannot be remunerated in the same amount and per capita criteria in the Northern region as in the other regions of the country, especially if compared with the South, Southeast and Midwest regions of Brazil. That is, we must consider the existence of a permanent challenge for managers at the three levels of government, in seeking alternatives to conduct the SUS guidelines, in the perspective of reducing regional inequalities, which imply in the distribution/allocation of federal financial resources that reach the differences and state and regional specificities in the management of the system. These challenges must be overcome, concomitantly with the exercise of the capacity to identify and solve the main health problems that affect the local population. And that it can modify the perspective of results, observing equal treatment for equal needs, equal access and equal health. In this sense, Escoda [11] considers the difficulties of the SUS as inherent to a social process and points out its cultural, political and technological dimensions. Highlighting that due to its social nature, it is a process still under construction. In this process of construction, there are so many laws, norms and normative instructions, where controls are still being implemented, and where there is a natural lack of service providers in the health area, there is also a lack of more consistent information and reliable indicators that adequately reflect reality. The correct situational diagnosis can be extremely decisive for the population of the State of Rondônia, and may mean the difference between access to treatment (life) and complete abandonment, resulting in deaths.

What can be deduced is that there is a multiplicity of factors involved: its human significance, the volume of resources, its social impact and its uniqueness. The manager of national public health policies has focused on strategic lines of discussion for the construction of a management pact, pact for life and pact in defense of the SUS, in this context the strongest point discussed within the collegiate instances is the reduction of regional inequities. Given the breadth and complexity of the problem, the present research was limited in an exploratory study of epidemiological, socioeconomic and coverage indicators of health services in the State of Rondônia.

To this end, we sought to apply the multivariate statistical technique, using Factor Analysis and the Principal Component Analysis method, with emphasis on identifying the significance of the indicators. According to Mingoti [12] Factor Analysis has as its main objective to describe the original variability of the random vector X, in terms of a smaller number m of random variables, called common factors "and that are related to the original vector X through a linear model". In this model, part of the variability of Xi attributed to common factors, with the remainder of the variability of, Y attributed to variables that were not included in the model, that is, to random error. In general, what is expected is that the original variables Xj, i=l,2,...,p are grouped into subsets of new variables that are mutually uncorrelated, and the factor analysis would aim to find these clustering factors. In cases where there is a large number of variables measured and correlated with each other, it would be possible, from the factor analysis, to identify a smaller number of new alternative variables, uncorrelated and that somehow summarize the main information of the original variables. These new alternative variables are called factors or latent variables. On the other hand, Principal Component Analysis, from the moment the factors are identified, their numerical values, called scores, can be obtained for each sample element. Consequently, these scores can be used in other analyses involving other statistical techniques, such as regression analysis or analysis of variance [12]. Although Factor Analysis can be applied to the original variables contained in the X vector, to facilitate understanding we prefer to introduce the main concepts of this technique using the original variables Xi, standardized by the respective mean and standard deviation.

Given this reasonable argumentation, the objective of the research is to identify the significance of the epidemiological, socioeconomic and coverage indicators of health services of the 52 municipalities of the State of Rondônia, applying Factor Analysis and Principal Component Analysis. What is expected with the use of these two statistical tools can guide future situational diagnoses in health, in order to effectively guide the allocation and use of resources.

2 | MATERIALS AND METHOD

2.1 Materials

The object of study in question is the epidemiological, socioeconomic and coverage indicators of health services, with emphasis on a contribution to the planning of actions based on situational diagnosis in health. The choice of indicators was based on the notes of Mingoti [12] to ensure the quality of the sample data “most multivariate statistical techniques use only complete observations, that is, if for a sample element, if the value of any variable has been lost, it is eliminated from the analysis process”. The 52 (fifty-two) municipalities of the State of Rondônia were considered, described in alphabetical order and acronym: Alta Floresta D’Oeste (ALFL); Alto Alegre dos Parecis (ALALG); Alto Paraíso (ALPA); Alvorada D’Oeste (ALV); Ariquemes (ARQUEMES); Buritis (BUR); Cabixi (CAB); Cacaulândia (CAUC); Cacoal (CACOAL); Campo Novo de Rondônia (CNRO); Candeias do Jamari (CJA); Castanheiras (CAST); Cerejeiras (CERJ); Chupinguaia (CHUP); Colorado D’Oeste (COLOR); Corumbiara (COR); Costa Marques (CMARQ); Cujubim (CUJB); Espigão D’Oeste (ESPIG); Gov. Jorge Teixeira (GOVJTEIX); Guajará-Mirim (GMRIM); Itapuã D’Oeste (ITAPUÃ); Jaru (JARÚ); Ji-Paraná (JI-PR); Machadinho D’Oeste (MACH); Ministro Andreazza (MANDREAZ); Mirante da Serra (MSERRA); Monte Negro (MNEGRO); Nova Brasilândia (NBRA); Nova Mamoré (NMA); Nova União (NU); Novo Horizonte (NHO); Ouro Preto D’Oeste (OPRETO); Parecis (PARECIS); Pimenta Bueno (PBUENO); Pimenteiras D’Oeste (PIMEN); Porto Velho (PVH); Presidente Médici (PMÉD); Primavera de Rondônia (PRIMARO); Rio Crespo (RCRESPO); Rolim de Moura (RLM); Santa Luzia D’Oeste (SLUZIA); São Felipe D’Oeste (SFELIPE); São Francisco do Guaporé (SFCO); São Miguel do Guaporé (SMIGUEL); Seringueiras (SERING); Teixeirópólis (TEIX); Theobroma (THOB); Urupá (URUPÁ); Vale do Anarí (VANARÍ); Vale do Paraíso (VPARAÍSO); Vilhena (VILHENA).

In order to establish fidelity, the data collection was concentrated in the main databases considered as official in the operationalization of the Health System at the national level, being: National Registry of Health Establishments (CNES), Department of Informatics of SUS (Datasus), Integrated Agreed Programming (PPI), Integrated Health Information Network (RIPSA), Public Health Budget Information System (SIOPS), Brazilian Institute of Geography and Statistics (IBGE) and United Nations Development Program (UNDP), considering the information present between the period of 6 (six) years. The construction of the database took place through isolated capture in each information system through online access via ADSL with connectivity and home accessibility. Each variable presented was extracted from the isolated and main system, and grouped in a table of the Microsoft Excel program.

The data capture period was approximately one (1) year due to the large number of variables. Once the process of capturing and exploring the information systems was exhausted, only the variable that presented continuous information in the 52 (fifty-two) municipalities of Rondônia was attributed as useful. A total of 57 socioeconomic variables, 54 epidemiological variables and 10 variables of health service coverage were collected from the databases of the 52 municipalities of the state of Rondônia. The computer instrument of choice for the treatment of data statistically was the software STATISTICA version (10) for presenting analytical capacity of considerable expression.

2.2 Method

2.2.1 Description of the search procedures

The first procedure to arrive at the statistical method was to carefully observe the data collected in order to identify distortions in the data collected, thus ensuring continuity of the information contained in the variables in the 52 municipalities. The research in question is classified into quantitative and qualitative in view of the presence of information in the form of data and the representative components of each variable, but it is an exploratory research in the sense that the intention is only to explore the statistical amplitude of each variable in the scenario State of Rondônia.

Data were statistically treated using the Factor Analysis technique and the Principal Component Analysis method. The procedure at first was to apply the technique and the method in the 121 grouped variables and after applying only in the group of variables with explanation greater than 95%, divided into two scenarios: Scenario 1: 121 variables; epidemiological, socioeconomic and coverage of health services in the State of Rondônia; and Scenario 2: 42 variables; epidemiological, socioeconomic and coverage of health services with an explanation greater than 95% in the State of Rondônia.

2.2.2 Factor Analysis

Mingoti [12] proposes:

2.2.2.1 Factor Analysis Model via Correlation Matrix

Be X_{pxi} a random vector with vector of means μ, where $\mu_i = (\mu_1, \mu_2 \ldots \mu p)`$, covariance matrix Σ_{pxp} and correlation matrix P_{pxp}. Are $Z_i = [(X_i - \mu_i) / \sigma_i]$ the original standardized variables, where μ_i and σ_i , represent respectively the mean and standard deviation of the variable X_i, $i = 1,2 \ldots p$. In this case, the matrix P_{pxp} is the matrix of covariances of the random vector $Z = (Z_1 Z_2 \ldots Z_p)'$.

The factor analysis model constructed from the theoretical correlation matrix P_{pxp} it is a model that linearly relates the standardized variables and the common factors that, at first, are unknown. The equations of the model are given by:

$$
\begin{aligned}
Z_1 &= l_{11} F_1 + l_{12} F_2 + \ldots + l_{1m} F_m + \varepsilon_1 \\
Z_2 &= l_{21} F_1 + l_{22} F_2 + \ldots + l_{2m} F_m + \varepsilon_2 \\
&\vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots \\
Z_p &= l_{p1} F_1 + l_{p2} F_2 + \ldots + l_{pm} F_m + \varepsilon_p
\end{aligned}
\tag{4.1}
$$

$Z_{1} = l_{ii} F_{i} + l_{12} F_{2} + \ldots + l_{im} F_{m} + \delta_{i}$
$Z_{2} = l_{2i} F_{1} + l_{22} F_{2} + \ldots + l_{2m} F_{m} + \delta_{2}$

In matrix notation, the model (4.1) can be expressed by:

$$D(X - \mu) = LF + \varepsilon \quad (4.2)$$

where:

$(X - \mu)_{px1} = [X_1 - \mu_1][X_2 - \mu_2][\vdots]$ $\varepsilon_{pxl} = [\varepsilon_1][\varepsilon_2][\vdots]$ $F_{mx1} = [F_1][F_2][\vdots]$ $L_{pxm} =$

$$\begin{bmatrix} l_{11} & l_{12} & \cdots & l_{1m} \\ l_{21} & l_{22} & \cdots & l_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ l_{p1} & l_{p2} & \cdots & l_{pm} \end{bmatrix}$$

$$D_{pxp} = \begin{bmatrix} 1/\sigma_1 & 0 & 0 & \cdots & 0 \\ 0 & 1/\sigma_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1/\sigma_p \end{bmatrix}$$

In this model, based on Mingoti [12], F_{mx1} is a random vector containing ***m*** factors, also called latent variables, that describe the elements of the population under study and are not observable, 1≤m≤p, that is, they cannot be measured a priori. Therefore, the factor analysis model assumes that the Z_i variables are linearly related to new random variables F_j, $j = 1, 2, \ldots, m$, that will need to be identified in some way. The vector ε_{px1} is a vector of random errors and corresponds to measurement errors and Z_i, variation, which is not explained by ordinary factors F_j, $j = 1,2,\ldots m$, included in the template. The coefficient of l_{ij}, commonly called loading, is the coefficient of the i-th standardized variable Z_i in the j-th factor Fj and represents the degree of linear relationship between Z_i and F_j, $j = 1,2,\ldots m$.

2.2.2.2 Orthogonal factor model

Some assumptions are necessary to operationalize the estimation of the model in (4.1). Let's assume that:

(i) $E[F_{mx1}] = 0$, which implies that $E[F_j] = 0$, $j = 1,2,\ldots, m$, that is, all factors have an average equal to zero:

(ii) $Var[F_{mx1}] = I_{mxm} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \vdots & 0 \\ \vdots & \cdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$

that is, all F_i factors are uncorrelated factors and have variances equal to 1;

(iii) $E[\varepsilon_{px1}] = 0$, which implies that $E[\varepsilon_j] = 0,\ j = 1,2,\ldots,p$, that is, all errors have averages equal to zero;

(iv) $Var[\varepsilon_{pxp}] = \psi_{pxp} = \begin{bmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & \vdots \\ \vdots & \cdots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_p \end{bmatrix}$

that is, $Var[\varepsilon_j] = \psi_j$ **and** $Cov(\varepsilon_i, \varepsilon_j) = 0,$ *"i ≠ j* which is to say that the errors are uncorrelated with each other and do not necessarily have the same variance;

(v) The vectors ε_{px1} and F_{mi1} are independent. So, $Cov(\varepsilon_{px1}, F_{mx1}) = E(\varepsilon F`) = 0$

Assumption (v) implies that the vectors ε and F represent two distinct sources of variation, related to the standardized variables Z_i, there is no relationship between these sources of information. A factorial model with assumptions **(i)-(v)** It is called orthogonal, where orthogonality refers to the fact that the m factors are orthogonal to each other.

An immediate consequence of assumptions **(i)-(v)** is related to the structure of the theoretical correlation matrix P_{pxp}. When the orthogonal model is assumed, the matrix P_{pxp} can be reparameterized in the form:

$$P_{pxp} = LL' + \psi \qquad (4.3)$$

This comes from the fact that:

$$P_{pxp} = Var(Z) = Var(LF + \varepsilon)$$
$$P_{pxp} = Var(LF) + Var(\varepsilon) = LIL` + \psi = LL' + \psi$$

where I is the dimension identity matrix pxp.

The objective of factor analysis according to Wíchern; Johnson [13] is to find the matrices L_{pxm} and Ψ_{pxp}, that can represent the matrix P_{pxp} for a given value of m, less than the number of original variables p. Unfortunately, there are many correlation matrices P_{pxp} which cannot be broken down into form $LL' + \Psi$ for a value of m much less than p.

In (4.4) it is possible to better visualize the format of the matrices involved in the decomposition given in (4.3) of the correlation matrix.

$$P_{pxp} = \begin{bmatrix} \sum_{j=1}^{m} l_{1j}^2 & \sum_{j=1}^{m} l_{1j} l_{j2} & \cdots & \sum_{j=1}^{m} l_{1j} l_{jp} \\ \sum_{j=1}^{m} l_{2j} l_{j1} & \sum_{j=1}^{m} l_{2j}^2 & \cdots & \sum_{j=1}^{m} l_{2j} l_{jp} \\ \vdots & \vdots & \cdots & \vdots \\ \sum_{j=1}^{m} l_{pj} l_{j1} & \cdot & \cdots & \sum_{j=1}^{m} l_{pj}^2 \end{bmatrix} + \begin{bmatrix} \psi_1 & 0 & \cdot & \cdot & 0 \\ 0 & \psi_2 & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & 0 & \psi_p \end{bmatrix} \quad (4.4)$$

The implications of decomposition (4.3) are presented below:

(i_1) $\mathrm{Var}(Z_i) = l_{i1}^2 + l_{i2}^2 + \ldots + l_{im}^2 + \psi_i = h_i^2 + \psi_i$, where $h_i^2 = l_{i1}^2 + l_{i2}^2 + \ldots + l_{im}^2$, *i = 1,2..., p.*

which means that the variance of $\boldsymbol{Z_i}$, it is decomposed into two parts. The first, denoted by h_i^2, is the variability of Z_i explicada pelos *m* fatores incluídos no modelo fatorial. This part of the variability is called “commonality”, a name given in function of the fact that the factors F_j, *j = 1, 2, ...,* ***m*** appear in all model equations (4.1) and variables Z_i have a common source of variation. The second, denoted by ψ_i, is the variability part of Z_i associated only with random error ε_i, which is specific to each variable Z_i. This part of the variability is called “uniqueness” or “specific variance.” As the variables Z_i have variances equal to 1, it follows that $h_i^2 + \psi_i = 1$.

(i_2) $Cov(Z_i Z_k) = l_{i1} l_{k1} + l_{i2} l_{k2} + \ldots + l_{im} l_{km}$, *i, k = 1, 2, ..., p, i ≠ k.*

(i_3) *Cov* (Z, F) = L_{pxm} and therefore, *Cov* $(Z_i F_j)$ = Corr $(Z_i F_j) = l_{ij}$, *i = 1, 2, ..., p*;
j = 1,2,...,m.

This comes from the fact that

Cov (Z, F) = Cov (LF+ε, F) = Cov (LF, F)+ *Cov* (ε,F) *= Cov* (LF, F) *= L*

Thus, one can use the L_{pxm} matrix in the search for understanding and interpretation of the factors F_j, *j = 1, 2, ..., m*

To operationalize factor analysis in practice, we must first use mechanisms to estimate the value of ***m***. From the estimated value of ***m*** we can then estimate the matrices L_{pxm} and ψ_{pxp}.

(i_4) Regarding the total variance, the proportion explained by the factor F_j is given by:

$$PVTE_{F_j} = \frac{\sum_{i=1}^{p} l_{ij}^2}{p} \quad (4.5)$$

and the most representative factors in the model are those with higher values of (4.5). It is common to express the values in (4.5) in percentage.

4.2.2.3 Estimation of the Number of Factors

The first step in conducting factor analysis is to estimate the theoretical correlation matrix $\boldsymbol{P}_{pxp}$ through the sample correlation matrix $\boldsymbol{R}_{pxp}$, as was done in principal component analysis. For the estimation of ***m***, it is enough to extract the eigenvalues from the $\boldsymbol{R}_{pxp}$ matrix and sort them in descending order.It is observed, then, which eigenvalues are the most important in terms of numerical quantity, using the following criteria:

Criterion 1: the analysis of the proportion of total variance related to each eigenvalue λ_i, given by λ_i**/p, i = 1, 2, ..., p**. Those eigenvalues that represent higher proportions of the total variance remain, and therefore the value of ***m*** will be equal to the number of eigenvalues retained; **Criterion 2:** the comparison of the numerical value of λ_i with the value **1, i = 1, 2, ..., p**. The value of ***m*** will be equal to the number of eigenvalues λ_i greater than or equal to **1**. The basic idea of this criterion is to maintain in the system new dimensions that represent at least the variance information of an original variable. This criterion was proposed by Kaiser (1958); **Criterion 3:** observation of Cattell's scree-plot [14], which shows the values of $\hat{l}_i$ ordered in descending order. By this criterion, the graph looks for a "jump point", which will be representing a decrease in importance in relation to the total variance. The value of ***m*** would then be equal to the number of eigenvalues prior to the "jump point". This criterion is equivalent to Criterion 1.

Suppose, for example, that we had p=6 and the eigenvalues λ_1 ***=2,24,*** λ_2***=1,38,*** λ_3***=*** *1,21,* λ_4***=0,63,*** λ_5***=0,41,*** λ_6***=0,13.***

In this case, by Criterion 2, ***m*** would be estimated to be equal to 3. The same suggestion would be indicated by the scree-plot shown in Figure 1.

The criteria described take into account only the numerical magnitude of the eigenvalues. An adequate choice of the value of ***m*** must, however, take into account the interpretability of the factors and the principle of parsimony, i.e. the description of the variability structure of the random vector ***Z*** with a small number of factors.

It is important to emphasize that the orthogonal factorial model should only be applied in situations in which the original variables are correlated with each other, because, otherwise, each factor will be related to only one original variable, making the value of ***m*** equal to ***p***.

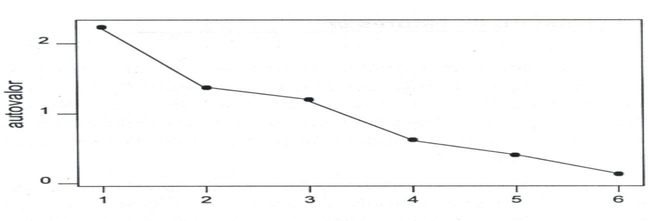

Figure 1: Scree-plot.

Source: Mingoti [12].

2.2.2.4 Methods of Estimation of Matrices L_{pxm} and ψ_{pxp}.

Once the value of ***m*** is chosen, it is possible to estimate the matrices $\boldsymbol{L_{pxm}}$ **e** $\boldsymbol{\psi_{pxp}}$. Initially, the principal components method will be introduced, commonly used as an exploratory analysis of the data, in terms of the underlying factors, because it does not require information or assumptions about the probability distribution of the random vector ***Z***. Subsequently, the principal factors method will be presented, which is a refinement of the principal components method, and the maximum likelihood method, which is indicated only when the random vector ***Z*** has multivariate normal distribution.

2.2.2.4.1 Principal component method for estimating matrices L_{pxm} and ψ_{pxp}

The principal components method works as follows: for each eigenvalue λ_1, **i =** ***1,2..., m*** retained in the estimation of the value of ***m***, as discussed in section 3, is the corresponding normalized eigenvector e where e= The matrices $\boldsymbol{L_{pxm}}$ and $\boldsymbol{\psi_{pxp}}$

shall be estimated respectively by:

$$\hat{L}_{pxm} = [illegible] \qquad (4.6)$$

$$\hat{\psi}_{pxp} = diag\, [illegible] \qquad (4.7)$$

where diag (.) denotes the diagonal matrix. Thus, the $\boldsymbol{\psi_{pxp}}$ matrix has the main diagonal equal to the elements of the main diagonal of the matrix [illegible]

The basic idea of this procedure lies in the application of the spectral decomposition theorem to the R_{pxp} matrix. By this theorem, the sample correlation matrix can be decomposed as a sum of ***p*** matrices, each related to an eigenvalue of the matrix R_{pxp}. For a fixed ***m*** value one has that:

$$R_{pxp} = \sum_{i=1}^{p} \hat{\lambda}_i \hat{e}_i \hat{e}_i' = \sum_{i=1}^{m} \hat{\lambda}_i \hat{e}_i \hat{e}_i' + \sum_{i=m+1}^{p} \hat{\lambda}_i \hat{e}_i \hat{e}_i' \qquad (4.8)$$

Thus, an approximation to the matrix *LL'* will be given by:

$$LL' = \sum_{i=1}^{m} \hat{\lambda}_i \hat{e}_i \hat{e}_i' = [illegible]$$

To build the matrix $\hat{\psi}_{pxp}$, one can consider using the following matrix:

$$\sum_{i=m+1}^{p} \hat{\lambda}_i \hat{e}_i \hat{e}_i' = \boldsymbol{R}_{pxp} - LL' \qquad (4.9)$$

Since the matrix in (4.9) is not diagonal, it cannot be used completely for the estimation of ψ_{pxp}. However, one can consider its diagonal. Thus, the matrix of specific variances is estimated as given in (4.7).

Considering this form of estimation, the original sample correlation matrix R_{pxp} is being approximated by:

$$\boldsymbol{R} \approx LL' + \psi \qquad (4.10)$$

and the residual matrix from the adjustment of the factorial model will be given by:

MRES = R - (4.11)

The residual matrix can serve as a criterion for evaluating the quality of fit of the factorial model. Ideally, your values should be close to zero. However, this matrix is only null when the value of ***m*** is equal to ***p***, which in practice is not the desired solution.

2.2.3 Principal factors method for estimating matrices L_{pxm}

From Mingoti [12] it is perceived that another method can be used for the estimation of the matrices of loadings and specific variances introduced by Thompson (1934) and is called the method of principal factors or iterative principal components. For it to be used, it is necessary that the value of ***m*** has already been estimated by some criterion. The basic idea is to proceed to a refinement of the estimates of L_{pxm} *and* ψ_{pxp} generated by the method of principal components. Considering the model $P = LL' + \psi$, where P is the theoretical correlation matrix of the random vector of interest X. Then we have:

$$LL' = P - \psi = \begin{bmatrix} h_1^2 & \rho_{12} & \rho_{13} & \cdots & \rho_{1p} \\ \rho_{21} & h_2^2 & \rho_{23} & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \rho_{p3} & \cdots & h_p^2 \end{bmatrix} \qquad (4.12)$$

onde $\boldsymbol{h_i^2 = 1 - \psi_i}$***, i = 1, 2. ..., p*** são as comunalidades. Suponha que se estime a matriz LL' por R^* dada por:

$$R^*_{pxp} = \begin{bmatrix} h_1^{*2} & r_{12} & r_{13} & \cdots & r_{1p} \\ r_{21} & h_2^{*2} & r_{23} & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & r_{p3} & \cdots & h_p^{*2} \end{bmatrix} \qquad (4.13)$$

where $(h_1^{*2}, h_2^{*2} \cdots h_p^{*2})$ are initial estimates of commonalities $(h_1^2, h_2^2 \cdots h_p^2)$. Using the principal components method, one has that:

L^* = (4.14)*

where are eigenvalues of ***R**** e are the respective normalized eigenvectors. From the matrix ***L"***, we have, therefore, new estimates of commonalities $(h_1^2, h_2^2 \cdots h_p^2)$, which are then placed on the main diagonal of the matrix in (4.13), and the procedure of estimating the matrix ***L****, by the principal components method, is repeated again. The algorithm is proceeded until such time as the differences between the commonalities of two successive interactions are negligible. Problems may occur during the execution of the algorithm, making it difficult for the iterative process to converge.

For example, eventually at some stage of the procedure, some eigenvalues of the ***R**** matrix may be negative, as these depend on the initial estimates of the commonalities. This creates an inconsistency with the fact that one is estimating a defined positive matrix, as is the case with the correlation matrix.

There may, also according to Rencher [15] be the Heywood problem, occasioned by the fact that at some stage of the iterative process some estimate h_i^2 can be greater than 1, which generates a negative estimate of ψ_1, and thus inconsistent with the definition of variance.

2.2.3.1 Principal Component Analysis

The use of Principal Component Analysis can under the developed conditions be understood as a process of analysis of statistical data, since it allows the methodological "walk" from beginning to end, that is, it allows the researcher to generate the expectation of results. Starting from the need to evidence the technique and method applied to the process of analysis of statistical data, and in attention to the objectives of the research, in order to leave it as a proposal for a real contribution to the Health System in Rondônia - more precisely to the planning area. From the task of explaining the method applied follows the demonstration of the statistical mode of how to obtain statistically results in a given sample.

Mingoti [12] proposes:

Be $\boldsymbol{X=(X_1, X_2 \dots X_p)'}$ a random vector with measure vector $\boldsymbol{\mu = (\mu_1\ \mu_2 \dots \mu_p)'}$ and matrix of covariances $\boldsymbol{\Sigma_{pxp}}$.

Are $\boldsymbol{\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p}$ the eigenvalues of the array $\boldsymbol{\Sigma_{pxp}}$, with their standard eigenvectors $\boldsymbol{e_1, e_2, \dots, e_p}$, that is, the eigenvectors $\boldsymbol{e_i}$ satisfy the following conditions:

(i) $\boldsymbol{e_i' e_j = 0}$ for all $\boldsymbol{i \neq j}$;

(ii) $\boldsymbol{e_i' e_i = 1}$ for all $\boldsymbol{i = 1,2, \dots, p}$;

(iii) $\boldsymbol{\Sigma_{pxp}\, e_i = \lambda_i\, e_i}$, for all $\boldsymbol{i = 1,2, \dots, p}$

where the eigenvector e_i is denoted by $\boldsymbol{e_i = (e_{i1} e_{i2} \dots, e_{ip})'}$. Consider the random vector $Y = O'X$, where $\boldsymbol{O_{pxp}}$ is the orthogonal matrix of dimension ***pxp***, consisting of the normalized eigenvectors of the matrix $\boldsymbol{\Sigma_{pxp}}$, that is,

$$\boldsymbol{O_{pxp}} = \begin{bmatrix} e_{11} & e_{21} & \cdot & \cdot & \cdot & e_{p1} \\ e_{12} & e_{22} & \cdot & \cdot & \cdot & e_{p2} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ e_{1p} & e_{2p} & \cdot & \cdot & \cdot & e_{pp} \end{bmatrix} = \lfloor e_1 e_2 \cdots e_p \rfloor \qquad (3.1)$$

The vector ***Y*** is composed of ***p*** linear combinations of the random variables of the vector ***X***, has vector means equal to and $\boldsymbol{O'\mu}$ and matrix of covariances $\boldsymbol{\Lambda_{pxp}}$, which is a diagonal matrix, whose elements are equal to $a_{ii} = \lambda_i, \; i = 1, 2, ..., p$ that is.

$$\Lambda_{pxp} = \begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \lambda_p \end{bmatrix}$$

Therefore, the random variables that constitute the Y vector are uncorrelated with each other. Thus, the idea of using linear combinations in ***Y*** arises as an alternative way of representing the covariance structure of the vector ***X***, trying to obtain a reduction of the space of variables, going from the dimension ***p***, to a dimension ***K*** less than ***p***. Therefore, instead of using the original random vector in the data analysis, using the K main linear combinations. The random vectors, ***X*** and ***Y***, have the same total variance and the same generalized variance, and the ***Y*** sector has the advantage of being composed of uncorrelated random variables, thus facilitating the joint interpretation of these. The following are some important definitions.

Definition 1: The j-th principal component of the matrix $\Sigma_{pxp}, j = 1,2,...,p$ is defined as:

$Y_j = e_j` X = e_{j1} X_1 + e_{j2} X_2 + ... + e_{jp} X_p$ (3.2)

The hope and variance of the component Y_j are, respectively, equal to:

$E[Y_j] = e_j ‘\mu = e_{j1} \mu_1 + e_{j2} \mu_2 + ... + e_{jp} \mu_p$

$Var [Y_j] = e_j ‘ \Sigma pxp \; e_j = \lambda_j$

being $Cov [Y_j, Y_k] = 0, j \neq k$. Each eigenvalue λj represents the variance of a principal component Y_j. Since the eigenvalues are ordered in descending order, the first component is the one with the greatest variability and the p-th is the one with the lowest.

Definition 2: The proportion of the total variance of ***X*** that is explained by the j-th principal component is defined as:

$$\frac{Var[Y_j]}{Variância\ Total\ de\ X} = \frac{\lambda_j}{Traço(\Sigma_{pxp})} = \frac{\lambda_j}{\sum_{i=1}^{p} \lambda i} \quad (3.3)$$

By the spectral composition theorem, the total and generalized variances of the random vector ***X*** can be described through the total variance and generalized variance of the random vector ***Y*** since,

$$traço\ (\Sigma_{pxp}) = \sum_{i=1}^{p} \sigma_{ii} = \sum_{i=1}^{p} \lambda_i \text{, where } \sigma_{ii} = Var[X_i], i = 1,2,..., p$$

Thus, in terms of these two global measures of variation, the vectors ***X*** and ***Y*** are equivalent. In general, the ratio in (3.3) is multiplied by 100, indicating the result as a percentage. It is evident that the first principal component has the highest proportion of explanation of the total variance of ***X***.

Definition 3: The proportion of the total variance that is explained by the first ***k*** compon

$$\frac{\sum_{j=1}^{k} Var[Y_J]}{Variância\ Total\ de\ X} = \frac{\sum_{j=1}^{k} \lambda_j}{Traço\left(\sum_{pxp}\right)} = \frac{\sum_{j=1}^{k} \lambda_j}{\sum_{i=1}^{k} \lambda_i} \quad (3.4)$$

If the first ***k*** principal components explain a large part of the total variance of the vector X, it may be restricted the focus of attention only to the random vector $(Y_1\ Y_2 \ldots Y_k)'$. In this way, a k-dimensional set of random variables can be examined, rather than a ***p***-dimensional set, without losing much information about the original structure of variances and covariances of the ***X*** vector.

By the spectral decomposition theorem, by restricting the focus of attention only to the first ***k*** principal components, the matrix of covariances Σ_{pxp} is being approximated by the formula in (3.5):

$$\sum_{pxp} \approx \sum_{j=1}^{k} \lambda_j e_j e_j` \quad (3.5)$$

Each portion of the sum in (3.5) involves a matrix of dimension ***pxp*** corresponding only to the information of the j-th principal component, $j = 1,2,\ldots,k$. Thus, the original variability system of the vector X is being approximated by the sum of ***k*** matrices, each representing the variability system related to a component. When $k=p$, one has that the covariance matrix Σ_{pxp} is accurately reproduced by the sum of matrices rationed to the principal components, that is,

$$\sum_{pxp} = \sum_{j=1}^{p} \lambda_j e_j e_j`$$

Definition 4: Another way to define the principal components is presented below. Consider the following system of linear combinations of ***X*** consisting of p equations of the type:

$Y_i = a'_i X = a_{i'} X_1 + a_{i2} X_2 + \ldots + a_{ip} X_{p'}\ i = 1,2,\ldots,p$

Thus, one has to:

$Var(Y_i) = a'_i \Sigma_{pxp} a_i$

$Cov(Y_{i'}, Y_j) = a'_i \Sigma_{pxp} a_j,\ i \neq j,\ i, j = 1,2,\ldots,p$

Suppose one wants to find the values of the coefficients aij such that $a'_i a_j = 1$, so that the linear combinations $Y_1, Y_2, \ldots, Y_p$ were uncorrelated with each other and had maximum variance.

So it can be demonstrated by Wichern; Johnson [13] that the maximum variance of $Y_1 = a'_1 X$, under the restriction $a'_1 a_1 = 1$, is equal to λ_1 and is obtained when $a_1 = e_1$, that is, the normalized eigenvector corresponding to λ_1, being Y_1 called the first principal component.

The maximum variance of $Y_2 = a'_2 X$, under the restrictions $a'_2 a_2 = 1$ and $cov(Y_1, Y_2)=0$, is equal to λ_2 and is obtained when $a_2 = e_2$, the normalized eigenvector corresponding to λ_2 being Y_2 called the second main component. The maximum variance of $Y_3 = a'_3 X$, under the restrictions $a'_3 a_3 = 1$ and $cov(Y_1, Y_3) = cov(Y_2, Y_3) = 0$, is equal to λ_3, being Y_3 called the third main component. Following the procedure, the maximum variance of $Y_i = a'_i X$, under the restrictions $a'_i a_i = 1$ e $cov(Y_i, Y_j) = 0 \quad j < i$, is obtained when $a_i = e_i$, that is, the normalized eigenvector corresponding to the eigenvalue λ_i, being Y_i called the i-th principal component. Thus, the principal components p are constructed, being unique, except for the signal exchange of all their coefficients. So, for example, if Yi is a principal component, $--Y_i$ will also be a principal component of order i.

Due to the very form of construction, the first principal component is always the most representative in terms of total variance and the p-th is always the least representative. Figure 2 shows a graphical illustration of the principal components in the case of two variables. As can be seen in the said figure, each point in the coordinate system Y_1 and Y_2.

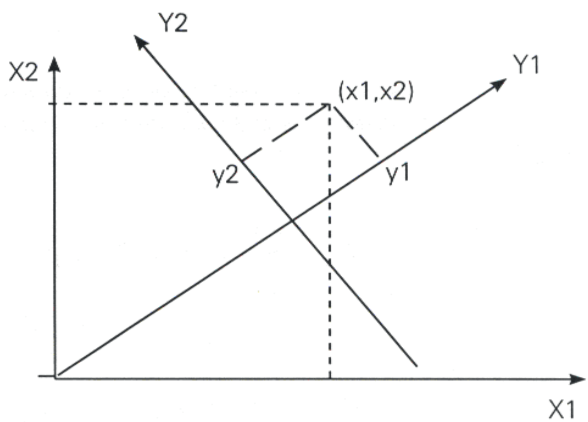

Figure 2: Geometric interpretation of principal component analysis – *p = 2*.

Source: Mingoti [12].

2.2.3.2 Principal Component Estimation: Covariance Matrix

In practice, the matrix Σ_{pxp} is unknown and needs to be estimated through the sample data collected. In general, the matrix Σ_{pxp} is estimated by the matrix of sample covariances S_{pxp}.

Let $\lambda_1, \lambda_2, ..., \lambda_p$ be the eigenvalues of the Spxp matrix, and let $\hat{e}_1, \hat{e}_2, ..., \hat{e}_p$, be the corresponding normalized eigenvectors. Then, the estimated j-th principal component is defined by:

$$Y_j = \hat{e}_j' X = \hat{e}_{j1} X_1 + \hat{e}_{j2} X_2 + ... + \hat{e}_{jp} X_p, \ j = 1,2,...,p \qquad (3.6)$$

Some properties of the sample principal components are presented below..

Property 1: The estimated variance of Y_j is equal to λ_j , $j = 1,2,...,p$

Property 2: The covariance between the components Y_j and Y_k is equal to zero, for all $k \neq j$, which means that these components are uncorrelated.

Property 3: The total variance explained by the j-th sample component is given by:

$$\frac{Var(Y_j)}{\text{variância total estimada de } X} = \frac{\hat{\lambda}_j}{traço(S_{pxp})} = \frac{\hat{\lambda}_j}{\sum_{i=1}^{p} \hat{\lambda}_i}$$

Property 4: The estimated correlation between the j-th main sample component and the random variable X_i, $i = 1, 2,...,p$ is given by:

$$r_{\hat{Y}_j, X_i} = \frac{\hat{e}_{ji}\sqrt{\hat{\lambda}_j}}{\sqrt{S_{ii}}}$$

where S_{II} is the sample variance of the random variable X_I.

Property 5: By the spectral decomposition theorem, the matrix of covariances S_{pxp} can be expressed as:

$$S_{pxp} = \sum_{j=1}^{p} \hat{\lambda}_j \hat{e}_j \hat{e}_j'$$

or approximated by (3.7) if only the first k sample principal components are used.

$$S_{pxp} \approx \sum_{j=1}^{k} \hat{\lambda}_j \hat{e}_j \hat{e}_j' \qquad (3.7)$$

In practical terms, to make use of the k main sample components considered most relevant in data analysis, it is necessary to calculate their numerical values for each sample element, values called component scores.

3 | RESULTS AND DISCUSSION

3.1 SCENARIO 1 = 121 EPIDEMIOLOGICAL, SOCIOECONOMIC AND HEALTH SERVICE COVERAGE VARIABLES IN THE STATE OF RONDÔNIA

In the application of the Factor Analysis technique and the Principal Component Analysis method, the findings combine with the theoretical foundations and demonstrate the importance of developing scientific studies in contribution to the management of the health system and in contribution to the decision-making process at the time of distribution of technical resources, financial and human.

In the application of the Factor Analysis technique and the Principal Component

Analysis method in the variables, with 121 variables grouped together. The statistical inferences show the following findings: factor (1), with an explanation greater than 70%, epidemiological indicators presented 46 variables, socioeconomic indicators 17 variables and indicators of health service coverage 6 variables. Therefore, with the application of the Factor Analysis technique and the Principal Component Analysis method in the variables, with 121 grouped variables, 69 variables with an explanation greater than 70% were found.

Factor (2) socioeconomic indicators highlighted 4 variables extracted from the application of Factor Analysis and Principal Component Analysis: Percentage of children aged 10 to 14 years with more than 1st year of school delay; Percentage of people in households with piped water; Percentage of children aged 10 to 14 years with less than 4 years of schooling; Percentage of children aged 7 to 14 out of school. The factor (3) indicators of coverage of health services highlighted only 2 (two) variables extracted from the application of Factor Analysis and Principal Component Analysis, higher than 70%: Percentage of government income; Percentage of people with more than 50% of government income.

Statistically in the context of factorial exploration, the factor (1) epidemiological indicators presented a variance explanation of 69.90%, the factor (2) socioeconomic indicators of 16.47% and factor (3) indicators of coverage of health services 6.46%. The set of factors accumulated eigenvalues of 92.8% of explanation extracted from the Principal Component Analysis. Considering what Mingoti [16] points out "the eigenvalues are ordered in descending order, the first component is the one with the greatest variability and the very bad one is the one with the lowest", the first main component is revealed in the study as the epidemiological indicators, the second the socioeconomic indicators and the third the indicators of coverage of health services.

Considering also the notes of Mingoti [12] the first main component is placed in the multivariate analysis in this study as of greater representativeness, and by inserted context of greater importance in view of the object of study. Table 1 shows the statistical findings.

Values	Eigenvalue	% Total Variance	Cumulative Eigenvalue	Cumulative %
F (1) Epidemiological Indicators	69,90479	57,77255	69,90479	57,7725
F (2) Socioeconomic Indicators	16,47150	13,61281	86,37628	71,3853
F (3) Health Services Coverage Indicators	6,46530	5,34322	92,84158	76,72858

Table 1: Demonstration of the eigenvalues in the application of PCA in the 121 variables in the year of study

Source: Prepared from secondary data.

The eigenvalues of correlation can also be explained through the total variance that has the ability to synthesize the global variance of the multivariate distribution, since this is the sum of the variances of all the variables involved in the X vector.

Another finding that draws attention is the coordination of factors with cases, which for this study were analyzed 121 variables and 52 cases, demonstrates that, considering the 52 cases, there are in factor (1) 13 cases with negative inferences, in factor (2) 27 cases and in factor (3) 24 cases. It is worth mentioning the health regions hosted by the following municipalities: Porto Velho, Ariquemes, Ji-Paraná, Cacoal, Rolim de Moura and Vilhena, emphasizing that only Porto Velho (capital) presents itself negatively in a decreasing situation in the three factors.

Cases	F (1) Epidemiological Indicators	F (2) Socioeconomic Indicators	F (3) Health Services Coverage Indicators
5 ARQUEMES	-7,8181	4,2101	2,92478
9 CACOAL	-8,8542	5,4915	2,88012
24 JI-PR	-11,4610	4,6260	2,01097
37 PVH	-47,8678	-11,4336	-2,23120
41 RLM	-4,9203	6,8001	0,65169
52 VILHENA	-9,0195	8,5311	1,33950

Table 2: Demonstration of negative and positive cases (city of health region) based on correlations, extracted from Principal Component Analysis, with 121 variables.

Source: Prepared from secondary data.

3.2 SCENARIO 2 = 42 EPIDEMIOLOGICAL, SOCIOECONOMIC AND HEALTH SERVICE COVERAGE VARIABLES WITH AN EXPLANATION ABOVE 95% IN THE STATE OF RONDÔNIA

From the exploration of the 121 variables, it was necessary to consider the notes of Hair et al [16], who propose, in the use of factor analysis, to adopt criteria of percentage of variance in order to obtain practical significance for the determined factors, considering it desirable that the level of explanation of variance be 95%.

According to Hair et al [16], if in the interpretation of factor analysis it is inferred that the:

1) factorial matrix: works as an aid in the process of choosing the number of factors, this non-rotated matrix, demonstrates the particular combination of original variables explaining more the variation in the data as a whole than any other linear combination of variables. Therefore, the first factor can be seen as the best summary of linear relationships displayed in the data;

1) the second factor is orthogonal to the first: (...) thus, the second factor can be defined as the linear combination of variables that explains most of the residual variance after the effect of the first factor has been removed from the data;

2) Factor loadings: is the correlation of each variable with each factor. The loads indicate the degree of correspondence between the variable and the factor, and larger

loads make the variable representative of the factor.

Scenario 2 represents the findings of the application of Factor Analysis and Principal Component Analysis in variables with explanation above 95% in the State of Rondônia. Scenario 2 is extracted from Scenario 1, in order to obtain greater clarity of data exploration and to envision a greater possibility of contributing to the situational diagnosis in health in the planning area.

The variables with an explanation greater than 95 % are shown in Table 3, as follows: They were found 25 Socioeconomic Variables, 8 Epidemiological Variables and 3 Health Services Coverage Variables.

25 Socioeconomic Variables	
% children aged 10 to 14 years older than 1 year behind in school (CUMATES)	% children aged 10 to 14 years with less than 4 years of schooling (CQUATES)
% of children aged 7-14 years with 1 year of delay in school (CÇ7-14C1AATRASADO)	% children aged 7 to 14 years in primary school (CENFUN)
% children in households with per capita income less than R$ 75.50 (CDVIQMS)	% child aged 7-14 years out of school (CÇ7-14AFORAESC)
% of people with more than 50% of government income (PRMGOV)	% of government income (RENDAGOV)
% of people in households with electricity and TV (PDENERG)	% of people in households with piped water (PDAGENC)
Women 25 years of age and older (MULAC25ANOS)	Women 15 years of age and older (NMAC15ANOS)
Population aged 10 to 14 years (POP10-14ANOS)	Population up to 1 year of age (POP1ANO)
Population 25 years of age or older (POPAC25ANOS)	Population 15 years of age and older (POPAC15ANOS)
Population 65 years of age or older (POPAC65ANOS)	5-year-old population (POP<5ANOS)
Urban population (POPTOTAL)	Total population (POPRURAL)
Probability of survival up to 60 years (PROBSOBAC65ANOS)	Probability of survival up to 40 years (PROBSOBAC40ANOS)
Municipal Human Development Index Longevity (IDHM-LONG)	Municipal Human Development Index (IDHM) Life expectancy at birth (ESPVIDAONASC)
8 Epidemiological Variables	
Hospitalization infectious-parasitic diseases DOENINFEC-PARASIT	Hospitalization for digestive diseases D.DIGESTIVA
Hospitalizations for pregnancy complications COMPGRAVIDEZ	Injuries from external causes LESCAUEXTER
General death ÓBITOGERAL	Infant death ÓB-INF
Death from External Causes OB-CAUS-EXTER	Infant death by residence OB-INF-RES
3 Health Services Coverage Variables	
Examination of the cervix EXPREVENTIVO	Prenatal consultation of 7 and/+ NVC/7CONSPN
Medical visit number MÉDIAVISDOM	

Table 3: Demonstration of socioeconomic variables with explanation greater than 95% in the application of Factor Analysis and Principal Component Analysis for the State of Rondônia.

Source: Prepared from secondary data.

Factors	Explanation	Variables
F (1) epidemiological	95% - 99%	25
F (2) socioeconomic	70% - 82%	08
F (3) health service coverage	70% - 80%	03

Table 4: Demonstration of the concentration of variables by factor and percentage of explanation.

Source: Prepared from secondary data.

As for the relationship between the first and the last variable, according to Mingoti [12] “variance serves to measure the degree of linear relationship between two variables”. Thus, in the application of PCA in the 42 variables above 95%, the variable with the greatest explanatory power is the first with 99.66% (% children aged 10 to 14 years with more than 1 year of school delay; CUMATES), from the second and successively until the 8th, 99.99% of explanation is obtained, as shown in Table 5.

he variability associated with the random error $\varepsilon_{i,}$ which is specific to each variable, can also be highlighted.

Variables	Eigenvalue	% TOTAL Variance	Cumulative Eigenvalue	Cumulative %
1 CUMATES	1,661123E+11	99,66815	1,661123E+11	99,6681
2 CQUATES	5,354738E+08	0,32129	1,666478E+11	99,9894
3 CENFUN	1,149565E+07	0,00690	1,666593E+11	99,9963
4 CDVIQMS	3,275427E+06	0,00197	1,666626E+11	99,9983
5 RENDAGOV	1,776124E+06	0,00107	1,666643E+11	99,9994
6 PRMGOV	4,844556E+05	0,00029	1,666648E+11	99,9997
7 PDAGENC	3,033109E+05	0,00018	1,666651E+11	99,9998
8 PDENERG	1,187135E+05	0,00007	1,666653E+11	99,9999

Table 5: Demonstration of covariance eigenvalues of matrices (factors)

Source: Prepared from secondary data.

There is a correlation between factors (1), (2) and (3), which explains 66.53% for epidemiological indicators, 18.72% for socioeconomic indicators and 7.53% for service coverage indicators. The finding corroborates what Hair et al [16] say: “the first factor can be seen as the best summary of linear relationships displayed in the data”.

According to Reis [17] and Mingoti [12], commonality is “the total amount of variance that an original variable shares with all the other variables included in the analysis” or “the variables have a source of variation in common”. In Scenario (1) the set of 121 variables (epidemiological, socioeconomic and coverage of health services) presented 69 variables with commonalities above 70%. In Scenario (2), the set of 42 variables (epidemiological, socioeconomic and coverage of health services with explanation above 95%) presented 16 variables with commonalities above 95%. It is noteworthy that the variables of Scenario (2) are included in scenario (1).

Variables	F (1) Epidemiological Indicators	F (2) Indicators Socioeconomic	F (3) Health Services Coverage Indicators	Multiple R-Square
NMAC15ANOS	0,983867	0,996511	0,996612	1,000000
MULAC25ANOS	0,985501	0,996481	0,996542	1,000000
POP1ANO	0,965979	0,994573	0,995047	0,999977
POP10-14ANOS	0,978427	0,996382	0,996495	0,999994
POPAC15ANOS	0,982552	0,995959	0,996183	1,000000
POPAC25ANOS	0,984372	0,995766	0,995977	1,000000
POP<5ANOS	0,974340	0,995863	0,996087	0,999988
POPAC65ANOS	0,979914	0,980371	0,980497	0,999872

Table 6: Demonstration of commonalities in variables above 95% explanation

Source: Prepared from secondary data.

4 I CONCLUSIONS

There is an almost absolute predominance of epidemiological variables in factor (1), which represents an explanation above 70%. This fact is external, thus the affinity of the object of the work in the face of being represented by the health problems, especially the causes of hospitalizations and deaths.

Regarding factor (2) of socioeconomic variables, there was a predominance of 04 variables that tend to express the quality of the education system of Rondônia in view of the presence of indicators that measure access to elementary education. On the other hand, factor (3) - of the variables of coverage of health services - did not present significance for the study.

For the 52 municipalities in question, the situation of the Municipality of Porto Velho (Capital of the State of Rondônia) stands out, which presents a negative correlation of (-47.86%), which can be justified by the presence of diseases common to the other municipalities, for example, dengue, tuberculosis, leprosy, injuries due to external causes, among others, here observing only numerical values. Only the municipality of Porto Velho (Capital) contributes with 72.84% in Factor (1) and 17.64% in Factor (2). This fact can be explained by the characteristic of the city in concentrating the services of medium and high complexity in health and being references for all the municipalities of the State in the following services: urgency and emergency in adults (Hospital Estadual Pronto Socorro João Paulo – II), major surgical treatment (Hospital de Base Dr. Ari Pinheiro), pediatric urgency and emergency (Cosme e Damião Children's Hospital) and in infectious-contagious diseases (Hospital Centro de Medicina Tropical de Rondônia). In this case, some variables such as hospitalization and death can directly influence the contribution of the municipality.

Scenario (2) aimed to stratify the result of Scenario (1) in order to ensure better clarity to the study. In the application of factor analysis and principal component analysis in 42 variables with explanation above 95%, confirmed the ability to represent the Factor (1) epidemiological indicators, pointing out, that 08 (eight) variables influenced the set of variables with explanation between 99.66% - 99.99%. These variables meet the group of access to elementary education, income, non-health public goods and services.

It can be said that the statistical attribute commonality directly influences the result of the research, since it was highly present in the first Scenario (1) with 69 variables and in the second Scenario (2) with 16 variables. Among the variables that presented commonality for this study, the following can be considered as the most important: number of cancer preventive exams, pregnancy complication and number of deaths in the period corresponding to 4 years of studies.

As for the agreed indicators, 07 (seven) are part of the indicators of service coverage and 26 of the epidemiological indicators. However, one can attribute a criticism to the non-agreement of socioeconomic indicators, which demonstrates low power of articulation in the field of intersectoral policy between the segments of education, public safety, environment and others that infers scope in the area of public health.

REFERENCES

1. [1]. BRASIL. CONSTITUIÇÃO DA REPÚBLICA FEDERATIVA DO BRASIL DE 1988. Promulgada em 05 de outubro de 1988. Versão eletrônica, revista e atualizada. Disponível em <http://www.planalto.gov.br/ccivil_03/Constituicao/Constitui%C3%A7ao.htm>, acesso em 17/10/2017.

[2]. BRASIL. Lei nº 8.080, de 19 de setembro de 1990. Lei Orgânica de Saúde. Dispõe sobre as condições para a promoção, proteção e recuperação da saúde, a organização e o funcionamento dos serviços correspondentes e dá outras providências. Disponível em <http://www.planalto.gov.br/ccivil/leis/L8080.htm>, acesso em 17/10/2017.

[3]. BRASIL. Lei nº 8.142, de 28 de dezembro de 1990. Dispõe sobre a participação da comunidade na gestão do Sistema Único de Saúde (SUS} e sobre as transferências intergovernamentais de recursos financeiros na área da saúde e dá outras providências. Disponível em <http://www.planalto.gov.br/ccivil_03/LEIS/L8142.htm>, acesso em 17/10/2017.

[4]. BRASIL. MINISTÉRIO DA SAÚDE. Norma Operacional da Assistência à Saúde - NOAS-SUS 01/2001. Disponível em <http://www.sespa.pa.gov.br/Sus/Legisla% C3%A7%C3%A3o/NOAS01_PT95.htm>, acesso em 17/10/2017.

[5]. BRASIL. CONASS - Conselho Nacional de Secretários de Saúde. SUS: avanços e desafios. (Org. SCOTTI, Ricardo F; autores: MENDES, Eugênio Vilaça; MÜLLER, Júlio; SANTOS, René). Brasília: CONASS, 2006.164 p.

[6]. BRASIL. MINISTÉRIO DA SAÚDE. Secretaria de Gestão Estratégica e Participativa. A construção do SUS: histórias da Reforma Sanitária e do processo participativo. Brasília: Ministério da Saúde, 2006.

[7]. BRASIL. MS. SUS. OPAS. Painel de Indicadores do SUS. Ano 1 – nº 1 – out/2006. Brasília: Ministério da Saúde, 2006.

[8]. BRASIL. MS/SCTIE/DES/SIOPS/FNS. (Ministério da Saúde. Secretaria de Ciência, Tecnologia e Insumos Estratégicos. Sistema de Informações sobre Orçamentos Públicos em Saúde. Fundo Nacional de Saúde), 2005. Relatório de Gestão 2005. Departamento de Economia da Saúde. Disponível em: <http://dtr2001.saude.gov.br/editora/produtos/livros/ pdf/06_0283_M.pdf>, acesso em 17/10/2017.

[9]. ALBERTO PARAGUASSÚ-CHAVES, Carlos. RAMOS, Josefa Lourdes. TRINDADE, Carla Dolezel. AZMAR FILHO, Simão. ALMEIDA, Fabrício Moraes de. DANTAS, Lenita Rodrigues Moreira. LUZ NETO, Leonardo Severo da. MACHADO NETO, Edmundo. SALTON, Ronaldo André Bezerra. . ANJOS, Osvaldo dos and SALTON, Gisely Beck Gonçalves. Multivariate Analysis of Health Indicators in the State of Rondônia, Western Amazon, Brazil. Acta Scientifci Nutritional Health, v. 5, p. 65-75, 2021.

[10]. LUCCHESE, Patrícia Tavares Ribeiro. Processo Alocativo e Redução da Desigualdade Regionais. Equidade na Gestão do Sistema Único de Saúde. 2003. Relatório Final, ENSP/FIOCRUZ, Rio de Janeiro.

[11]. ESCODA, Maria do Socorro Quirino. Iniquidade em Saúde. Revista: Espaço para Saúde, ISSN 1517-30, versão online, vol. V, nº II, UEL/NESCO-PR, 2004. Disponível em <http://www.ufrnet.br/~scorpius/20-Iniquidade%20em%20saude.htm>, acesso em 17/10/2017.

[12]. MINGOTI, Sueli Aparecida. Análise de Dados através de Métodos de Estatística Multivariada: Uma abordagem aplicada. Belo Horizonte: Editora UFMG, 2005.

[13]. WICHERN, D. W; JOHNSON, R, A, Applied Multivariate Statistical Analysis. Londres: 3=" Ed, Englewood (N.J,): Prentice-Hall, Inc, 1998.

[14]. CATTELL, R. B. (1966). The Scree Plot Test for the Number of Factors. Multivariate Behavioral Research, 1, 140-161. http://dx.doi.org/10.1207/s15327906mbr010210.

[15]. RENCHER, ALVIN C. Methods of Multivariate Analysis. 2nd ed. p. cm. — (Wiley series in probability and mathematical statistics) "A Wiley-Interscience publication." Printed in the United States of America.

[16]. HAIR Jr., Joseph F.; ANDERSON, Rolph E.; TATHAM, Ronald L.; BLACK, William C. Análise Multivariada de Dados. 5ª ed., Trad. Adonai Schlup Sant´Anna e Anselmo Chaves Neto. Porto Alegre: Bookman, 2005.

[17]. REIS, Elizabeth. Estatística Multivariada Aplicada. 2ª ed. Revista e corrigida. Lisboa: Edições Silabo, 2001.