Acta Scientific NUTRITIONAL HEALTH (ISSN:2582-1423)

Volume 5 Issue 5 May 2021

Research Article

Health Economics Study: Epidemiological, Socioeconomic and Health Service Coverage Indicators in a State in the Western Amazon (Brazil)

Carlos Alberto Paraguassú-Chaves[1], Josefa Lourdes Ramos[2], Carla Dolezel Trindade[3], Simão Aznar Filho[4], Fabrício Moraes de Almeida[5]*, Edmundo Machado Neto[6], Leonardo Severo da Luz Neto[7], Ronaldo André Bezerra Salton[8], Fábio Robson Casara Cavalcante[9], Lenita Rodrigues Moreira Dantas[10] and Osvaldo dos Anjos[11]

[1]PhD in Health Sciences, University of Brasília - UnB, Brazil and PhD in Science - University of Havana, Cuba and Post-Doctor in Health Sciences, UnB and Degli Studi D'Aquila University, IT, Professor at the University Institute of Rio de Janeiro, IURJ, Brazil

[2]Master in Regional Development and Environment, Federal University of Rondônia, UNIR, Brazil, Voluntary researcher at the Higher Institute of Health and Environmental Sciences of the Amazon, AICSA, "In Memoriam", Brazil

[3]PhD in Law - Universidad Nacional de Lomas de Zamora, Argentina and Post-doctorate, Universita deli Studi di Messina,Italy and Full Professor at the University Institute of Rio de Janeiro, IURJ, Brazil

[4]PhD in Law, Universidad Nacional de Lomas de Zamora, Argentina and Post-doctorate, Universita deli Studi di Messina, Italy and Full Professor at the University Institute of Rio de Janeiro, IURJ, Brazil

[5]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), Brazil

[6]Master in Regional Development and Environment from the Federal University of Rondônia, UNIR, Volunteer researcher at the Institute of Health Sciences and the Environment of the Amazon, AICSA, Brazil

[7]Master in School Psychology and Human Development, University of São Paulo, USP, Brazil and Professor at the Federal University of Rondônia, UNIR, Brazil

[8]Master in Public Administration, Federal University of Rondônia, UNIR, Brazil and Volunteer researcher at the Institute of Health Sciences and the Environment of the Amazon, AICSA, Brazil

[9]PhD in Sciences: Socioenvironmental Development, Federal University of Pará, UFPA, Brazil and Associate Professor at the Federal University of Rondonoa, UNIR, Brazil

[10]Geographer Expert and Lawyer, Researcher of the Institute of Health Sciences and the Amazon Environment, AICSA, Brazil

[11]Master in Environmental Sciences, San Carlos University, Paraguay and Volunteer researcher at the Institute of Health Sciences and the Environment of the Amazon, AICSA, Brazil

***Corresponding Author:** 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), Brazil.

Received: March 29, 2021

Published: April 27, 2021

Citation: Fabrício Moraes de Almeida., *et al.* "Health Economics Study: Epidemiological, Socioeconomic and Health Service Coverage Indicators in a State in the Western Amazon (Brazil)". *Acta Scientific Nutritional Health* 5.5 (2021): 43- 64.

Abstract

Objective: The objective of the research was to identify the significance of the epidemiological, socioeconomic and health service coverage indicators in Rondônia, Western Amazon, Brazil.

Method: Multivariate statistical technique, using Factor Analysis (PA) and the Principal Component Analysis method (PCA), with emphasis on identifying significance. A sample with 121 variables was used, divided into two scenarios: Scenario 1 with 121 variables; and Scenario 2 with 42 variables with an explanation above 95%.

Results: Factor 1 epidemiological indicators showed 69 variables with explanation above 70%, factor 2 socioeconomic indicators highlighted 4 variables with explanation above 70% and factor 3 indicators of health service coverage highlighted only 2 (two) variables with explanation above 70%. Statistically, in the context of factorial exploration, it was found that factor (1) presented a variance of 69.90%, factor (2) 14.47% and factor (3) 6.46%. Scenario 2 composed of 42 variables with explanation above 95% in the application of PA and ACP obtained 34 variables with explanation above 70%. Of these, 24 are of factor 1 and had an 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, socioeconomic variables, which measure access to basic education in the State of Rondônia. The health service coverage variables, however, were not significant. Porto Velho (Capital) presented a negative correlation of (-47.86%) which can be justified by the presence of diseases common to other municipalities, for example, dengue, tuberculosis, leprosy, injuries due to external causes, etc. And, it contributed with 72.84% in Factor 1 and 17.64% in Factor 2, which can be explained by the condition of being the state reference in health.

Conclusions: The study showed that of the 39 indicators agreed between the State Department of Health of Rondônia and the 52 municipalities, 33 (thirty-three indicators) are included among the indicators with statistical significance, which validates the choice of the method.

Keywords: Economics and Health; Indicators; Multivariate Analysis; Rondônia; Brazil

Introduction

The Unified Health System (SUS) of Brazil, as a new model of health care, is governed by principles that seek to guarantee universal, integral, equal, equitable and free access to health, constituted by 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 [1]. The model is characterized by a design consisting of three arrangements that make up the SUS in Brazil (the municipal, state and national systems) that during the construction process were composed by implementation governed by ministerial standards: the Basic Operating Standards (NOBs) and Health Assistance Operational Standard (NOAS). These legal provisions are the main regulatory instruments of the system for the transfer of resources to the health area between the three levels of government [2-8]. The SUS financing logic, implemented through NOBs 91, 92 and 93, constitutes a dynamic process that regulates the production payment system, redefines resource allocation criteria and establishes a financing mechanism for participation and decentralization of health actions and services. NOB 01/96 becomes one of the most important, as it changes the perception of States and Municipalities in the decentralization process, which started to qualify according to the level of financial autonomy and management model [2], [3-8]. In this respect, the most significant point was the change in the system's remuneration mode, which ceases to be per production and becomes per capita for primary care actions, through the implementation of the Basic Care Floor (PAB) and health care programs. strengthening primary care, such as the Family Health Program (PSF), the Community Agents Program (PAC) and the health surveillance program [9,10]. According to Piola and Biasoto Júnior [11], NOB 01/96 expanded the mechanism for transferring funds directly from the national federal health fund to state and municipal health funds.

For Lucchese [12], it is necessary to consider 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 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 revising financial ceilings. In recent years, the process of financing in a decentralized and regionalized way of health actions and services has taken place in an emergent way, motivated by the search for a health network that would make it possible to guarantee the principle of equity and universality, as it has basically passed to observe management capacity, installed capacity and loco-regional aspects, as population aggregations, by certain municipalities, given their micro and macro regional positions. According to Escoda [13] the principle of equity, consists, more than what is contained in the legal or linguistic disposition described by the researcher Ferreira (1995), in equally recognizing the right of each one, in a feeling of justice, therefore, inside out a strict treatment criterion in strictly legal form. Equity in health comprises a framework of revolutionary principles of equality, rectitude and equanimity of the citizen in the face of their living conditions, regardless of social security contribution or income level for access to comprehensive health actions, according to their medical care needs.

Given 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 regional inequalities that reside in different levels of income concentration, low demographic density and low living conditions, indicated by health indicators. Thus, health inequalities are expressed in the different possibilities of life expectancy at birth, in how to live, in the ways of falling ill and dying for the same class in contrast to rates in other regions of the country. According to Ugá and Santos [14] until the mid-nineties, public spending on health kept the mark of the period before the new Constitution, characterized by a strong centralization of the health system and its financing at the federal level. According to Escoda [13], SUS regulation was conceived, more as a social advance and less as a result of political decision. The Law that regulates it, in addition to the doctrinal principles of the health system, establishes the organizational principles, from resolvability to regionalization, hierarchization, decentralization, complementarity and social control. Escoda [13] considers the difficulties of SUS as inherent to a social process and, in this sense, punctuates its cultural, political and technological dimensions. It highlights that due to its social nature, it is a process under construction.

In view of so many laws, norms and normative instructions, it should be a matter of concern to know what actually reaches the other end, in the beneficiary of this whole apparatus: the citizen. This issue is even more serious and relevant in a federation unit where, due to its own condition of being a very new state, where controls are still being implemented, and where there is a natural shortage of health service providers, it also lacks 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, which can mean the difference between access to treatment (life) and complete abandonment, resulting in deaths. Given the breadth and complexity of the problem, it is necessary to delimit this research in order to make it feasible. Thus, the present research is delimited in the exploratory study of epidemiological, socioeconomic and health service coverage indicators in the State of Rondônia. The objective of the research is to identify the significance of the epidemiological, socioeconomic and health service coverage indicators in Rondônia, applying the multivariate statistical technique. It is hoped, therefore, that the criteria for choosing priority health items agreed between the State of Rondônia and the Federal Government can be validated, as well as guiding future situational health diagnoses in the State, in order to effectively guide the allocation and the use of resources. This research is justified taking into account the multiplicity of factors involved: its human significance, the volume of resources, its social impact and its originality. 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 SUS, in this context the strongest point discussed within the collegiate bodies is the reduction of regional inequities. Porto [15] stresses that for the effective reach of more equitable societies, public policies are needed that go beyond the sectorial scope and are able to jointly reduce the currently existing inequalities, without losing sight of efficiency and effectiveness".

Materials and Methods

Materials

The object of study in question are the epidemiological, socioeconomic and health service coverage indicators, with an emphasis on a contribution to the planning of actions based on situational health diagnosis. The choice of indicators was based on the notes of Mingoti [16] to ensure the quality of the sample data "most of the multivariate statistical techniques use only the 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". Thus, we considered the indicators that showed the time of data collection continuous information in the online information systems.

For the study in question, the 52 (fifty-two) Municipalities of the State of Rondônia were then considered, represented in table 1.

County	Acronym	County	Acronym
Alta Floresta D'Oeste	ALFL	Mirante da Serra	MSERRA
Alto Alegre dos Parecis	ALALG	Monte Negro	MNEGRO
Alto Paraíso	ALPA	Nova Brasilândia	NBRA
Alvorada D'Oeste	ALV	Nova Mamoré	NMA
Ariquemes	ARQUEMES	Nova União	NU
Buritis	BUR	Novo Horizonte	NHO
Cabixi	CAB	Ouro Preto D'Oeste	OPRETO
Cacaulândia	CAUC	Parecis	PARECIS
Cacoal	CACOAL	Pimenta Bueno	PBUENO
Campo Novo de Rondônia	CNRO	Pimenteiras D'Oeste	PIMEN
Candeias do Jamari	CJA	Porto Velho	PVH
Castanheiras	CAST	Presidente Médici	PMÉD
Cerejeiras	CERJ	Primavera de Rondônia	PRIMARO
Chupinguaia	CHUP	Rio Crespo	RCRESPO
Colorado D'Oeste	COLOR	Rolim de Moura	RLM
Corumbiara	COR	Santa Luzia D'Oeste	SLUZIA
Costa Marques	CMARQ	São Felipe D"oeste".	SFELIPE
Cujubim	CUJB	São Francisco do Guaporé	SFCO
Espigão D'Oeste	ESPIG	São Miguel do Guaporé	SMIGUEL
Gov. Jorge Teixeira	GOVJTEIX	Seringueiras	SERING
Guajará-Mirim	GMRIM	Teixeirópólis	TEIX
Itapuã D'Oeste	ITAPUÃ	Theobroma	THOB
Jaru	JARÚ	Urupá	URUPÁ
Ji-Paraná	JI-PR	Vale do Anarí	VANARÍ
Machadinho D'Oeste	MACH	Vale do Paraíso	VPARAÍSO
Ministro Andreazza	MANDREAZ	Vilhena	VILHENA

Table 1: Statement of municipalities surveyed by acronyms.

Source: Research adaptation.

In order to establish loyalty, data collection was concentrated in the main databases considered to be official in the operationalization of the Health System at the national level, being: National Registry of Health Establishment (CNES), SUS Informatics Department (Datasus), Integrated Pacted 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 six (6) year period.

The construction of the database took place through capture in an isolated way in each information system through online access via ADSL with connectivity and domestic accessibility. Each variable presented was extracted from the main and isolated system, and grouped in a table in the Microsoft Excel program.

Table 2-4 show the variables considered for the study in question.

The computer instrument of choice for the treatment of the statistical data was the STATISTICA software version (7) for presenting analytical capacity of considerable expression.

Socioeconomic indicators	Acronym
% children 10 to 14 years old with more than 1 year of school delay	CUMATES
% children aged 10 to 14 years with less than 4 years of schooling	CQUATES
% children aged 5 to 6 years in literacy	CNAL
% children aged 7 to 14 years in elementary school	CENFUN
Intensity of indigence	INTIND
Intensity of poverty	INTENPOB
% children in households with per capita income less than R $ 75.50	CDVIQMS
% children in households with per capita income less than R $ 37.75	CDVIMSM
% people with per capita income below R $ 37.75	PRPCUQSM
% people with per capita income below R $ 75.50	PRPCMDSM
% of income from work	RENDATRAB
% of government income	RENDAGOV
% of people with more than 50% of government income	PRMGOV
Per capita income	RENPC
% of people in households with running water	PDAGENC
% of people in households with a car	PDCAR
% of people in households with a computer	PDCOMP
% of people in households with electricity and TV	PDENERG
% of people in households with a telephone	PDTELF
% of people in urban households with garbage collection service	PDSERCOLIX
% of people in households with more than 2 people per bedroom	PDDUASPD
Women aged 15 and over	NMAC15ANOS
Women 25 years of age and over	MULAC25ANOS
Population up to 1 year of age	POP1ANO
Population 10 to 14 years of age	POP10-14ANOS
Population aged 15 years and over	POPAC15ANOS
Population 25 years of age and over	POPAC25ANOS
55-year-old population	POP<55ANOS
Population 65 years of age and over	POPAC65ANOS
Total population	POPRURAL
Urban population	POPTOTAL

Annual average study of people aged 25 or over	POPURBANA
% illiterate adolescents aged 15 to 17 years	MEDESTAC25ANOS
% adolescents aged 15 to 17 years out of school	ADOL15-17AANALF
% adolescents aged 15 to 17 years at school	ADOL15-17AFORAESC
% female adolescents aged 15 and 17 with children	ADOL15-17ANAESC
% children aged 10 to 14 years illiterate	ADOLFEM15-17ACFIL
% children aged 10 to 14 who work	CÇ10-14AANALF
% children aged 7 to 14 years with more than 1 year of school delay	ADOL10-14ATRAB
% children aged 7 to 14 years out of school	CÇ7-14C1AATRASADO
% female children. 10 and 14 years with children	CÇ7-14AFORAESC
% people aged 65 and + + living alone	ADOLFEM10-14CFILH
% of persons aged 25 and/+ illiterate	PESAC65ARESSOZ
% of persons aged 25 and/+ with less than 8 years of study	PESAC25AANALF
% of persons aged 25 and/+ with less than 8 years of study	PESAC25AS/ENFUND
% of people in families with 75% dependency	PESC/75%DEPEN
% of elementary school teachers with higher education	PROFC/ENSUPERIOR
% women heads of families without spouse and with children <15 years old	MULCHFAMIL+FILH
Probability of survival up to 40 years	PROBSOBAC40ANOS
Probability of survival up to 60 years	PROBSOBAC65ANOS
IDHM	IDHM
IDHM-E	IDHM-E
IDHM-LONG	IDHM-LONG
IDHM-R	IDHM-R
Gross school attendance rate	TAXFREQESC
Literacy rate	TAXANALF
Life expectancy at birth	ESPVIDAONASC

Table 2: Statement of the 57 socioeconomic variables, captured from the databases, of the 52 municipalities in the State of Rondônia.

Source: Adapted from the Human Development Atlas in Brazil (UNDP).

Epidemiological indicators	Acronym
Hospitalization of infectious and parasitic diseases	DOENINFEC-PARASIT
Hospitalization for Hansen's disease	AGRAVOHANS
Tuberculosis Hospitalization	AGRAVOTUBERC
Hospitalization for bacterial diseases	DOENÇABACTERIANA
Hospitalization for dengue	AGRAVODENGUE
Hospitalization in 5-year-old children	HOSPCÇS5ANOS
Hospitalization for dehydration	HOSPCÇ/DESIDRAT
Hospitalization for Hepatitis	HEPATITES
HIV hospitalization	AGRAVOHIV
Hospitalization for Mycoses	MICOSES

Hospitalization for Malaria	AGRAVOMALÁRIA
Hospitalization for Leishmaniasis	LESH
Cancer Hospitalization	AGRAVOCA
Hospitalization Diabetes Mellitus	DIABMELITUS
Hospitalization for anemias	ANEMIAS
Malnutrition hospitalization	DESNUTRIÇÃO
Hospitalization for mental illness	D.MENTAIS
Hospitalization for neurological diseases	D.NEUROLOGICA
Hospitalization for ophthalmic diseases	D.OFTALM
Hospitalization for hearing diseases	D.AUDITIVA
Hospitalization for circulatory diseases	D.CICULATÓRIA
Hospitalization for respiratory diseases	D.RESPIRATÓRIA
Hospitalization for digestive diseases	D.DIGESTIVA
Diseases due to skin diseases	D.PELE
Hospitalization for Musculoskeletal Diseases	D.OSTEOMUSC
Hospitalization for diseases of the urinary tract	D.AP.URINÁRIO
Hospitalizations due to pregnancy complications	COMPGRAVIDEZ
Perinatal infections hospitalizations	INF-PERINATAL
Congenital malformations	MALFORMCONGENITA
Injuries due to external causes	LESCAUEXTER
General death	**ÓBITOGERAL**
Infant death	**ÓB-INF**
Infant death by residence	OB-INF-RES
Death from External Causes	**ÓB-CAUS-EXTER**
Hypertension	HIPERTENSOS
Type 1 diabetes	DIABETESTIPO 1
Type 2 diabetes	Diabetes tipo 2
Hypertension and diabetes	HIPERT-DIAB
Hansen’s disease	AG-HANS
Diseases of Leishmaniasis	AGRAV-LEISH
Tuberculosis Disease	AG.TB
grievance for Dengue	AG.DENG
Born alive	NV
Preventive number (+)	NPRV (+)

Table 3: Statement of the 54 epidemiological variables, captured from the databases, of the 52 municipalities in the State of Rondônia.

Source: Adapted from DATASUS and RIPSA.

Service coverage indicators	Acronym
Percentage of nurse/1000/hab	%ENF/1000HAB
Doctor number/1000/inhab	NMÉDICO/1000HAB
Born alive without prenatal care	NVS/PRÉ-NATAL
Born alive with 1-3 prenatal consultations	NVC/1-3CONSPN
Born alive with 4-6 prenatal consultations	NVC/4-6CONSPN
Cervical examination	EXPREVENTIVO
Prenatal consultation of 7 and/+	NVC/7CONSPN
Vaccination coverage	COBERTVACINAL
Average visit number	MÉDIAVISDOM
Accompanied family number/Average	NFAMILAC/ACS

Table 4: Statement of health service coverage variables, captured from the databases, of the 52 municipalities in the State of Rondônia.

Source: Adapted from CNES and DATASUS.

Method

Description of the research 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 is classified as quantitative and qualitative in view of the presence of information in the form of data and the representative components of each variable, however it is an exploratory research in the sense that the intention is only to explore the breadth statistics for each variable in the Rondônia State scenario.

The data were treated statistically using the Factor Analysis (AF) technique and the Principal Component Analysis Method (PCA). The procedure at first was to apply the technique and method to the 121 grouped variables and after applying only to the group of variables with explanation above 95%, divided into two scenarios: Scenario 1: 121 variables; epidemiological, socioeconomic and health service coverage in the State of Rondônia; and Scenario 2: 42 variables; epidemiological, socioeconomic and health service coverage with explanation above 95% in the State of Rondônia.

Factor analysis (AF)

According to Mingoti [16] in the analysis of main components, the factorial analysis has as 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 is 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 = I, 2,..., p are grouped into subsets of new mutually unrelated variables, and the factorial analysis would aim to find these clustering factors.

Thus, in cases with a large number of variables measured and correlated with each other, it would be possible, based on the factor analysis, to identify a smaller number of new alternative variables, not correlated and that in some way summarize the main information of the original variables [16]. These new alternative variables are called latent factors or variables.

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 analyzes that involve other statistical techniques, such as regression analysis or analysis of variance [16].

Although factor analysis can be applied to the original variables contained in the vector X, to facilitate understanding we prefer to introduce the main concepts of this technique using the original variables $\boldsymbol{X_i}$ standardized by the respective mean and standard deviation.

Factor analysis model via correlation matrix

Let X_{pxi} be 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}. Let $Z_{i=}$ [$(X_i - \mu_i)/\sigma_i$] be the original standardized variables, where μ_i e σ_i, respectively represent the mean and standard deviation of the variable X_i, i = 1,2...p. In this case, the P_{pxp} matrix is the covariance matrix of the random vector Z = $(Z_1\ Z_2 \ldots Z_p)$'. The factor analysis model built from the P_{pxp} theoretical correlation matrix is a model that linearly relates the standardized variables and the common factors that, at first, are unknown. The model equations are given by:

$$Z_1 = l_{11} F_1 + l_{12} F_2 + \ldots + l_{1m} F_m + \varepsilon_1$$

$$Z_2 = l_{21} F_1 + l_{22} F_2 + \ldots + l_{2m} F_m + \varepsilon_2$$

$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots \quad \text{-----------------------}(4.1)$$

$$Z_p = l_{p1} F_1 + l_{p2} F_2 + \ldots + l_{pm} F_m + \varepsilon_p$$

$$Z_1 = I_{ii} F_i + I_{12} F_2 + \ldots + I_{im} F_m + \delta_i$$

$$Z_2 = I_{2i} F_1 + I_{22} F_2 + \ldots + I_{2m} F_m + \delta_2$$

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

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

Where:

$$(X - \mu)_{px1} = \begin{bmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \\ \vdots \\ X_p - \mu_p \end{bmatrix} \qquad \varepsilon_{px1} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_p \end{bmatrix} \qquad F_{mx1} = \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_m \end{bmatrix} \qquad L_{pxm} = \begin{bmatrix} l_{11} & l_{12} & \cdots & l_{1m} \\ l_{21} & l_{22} & \cdots & l_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ l_{p1} & l_{p2} & \cdots & l_{pm} \end{bmatrix}$$

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

In this model, based on Mingoti [16], F_{mx1} is a random vector containing m factors, also called latent variables, that describe the elements of the study population and are not observable, $1 \leq m \leq p$, that is, they cannot be measured a priori. Therefore, the factor analysis model assumes that the variables Z_i are linearly related to new random variables F_j, j = 1, 2,..., m, which will need to be identified in some way. The vector ε_{px1} is a vector of random errors and corresponds to the measurement errors and the Z_i variation, which is not explained by the common factors F_j, j = 1,2,...m, included in the model. The l_{ij}, coefficient, commonly called loading, is the coefficient of the i-th standardized variable Z_i in the j-th factor F_j and represents the degree of linear relationship between Z_i and F_j, j = 1,2,...m.

Orthogonal factor model

Some assumptions are necessary in order to be able to operationalize the model estimation in (4.1). Let's assume that.

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

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

That is, all factors F_i they are uncorrelated factors and have variances equal to 1.

E $[\boldsymbol{\varepsilon}_{px1}] = 0$, which implies that E $[\varepsilon_j] = 0$, j = 1,2,...,p, that is, all errors have averages equal to zero.

$$\text{(iv) Var}\ [\boldsymbol{\varepsilon}_{pxp}] = \boldsymbol{\psi}_{pxp} = \begin{bmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & \vdots \\ \vdots & \cdots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_p \end{bmatrix}$$

i.e., Var $[\varepsilon_j] = \boldsymbol{\psi}_j$ and Cov $(\boldsymbol{\varepsilon}_i, \boldsymbol{\varepsilon}_j) = 0$, "i ≠ j which means that the errors are not correlated with each other and do not necessarily have the same variance.

The vectors ε_{px1} e F_{mi1} are independent. Therefore, Cov ($\varepsilon_{px1}, F_{mx1}$) = E (ε F`) = 0.

Assumption (v) implies that the vectors ***ε and*** F represent two distinct sources of variation, related to 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 the 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 reparametrized in the form:

P_{pxp} ***= LL´ + ψ-------- (4.3)***

This comes from the fact that:

P_{pxp} = Var(Z) = Var (LF + ε)

P_{pxp} = Var(LF) + Var(***ε) =*** LIL` + ***ψ =*** LL´ + ***ψ***

Where I is the dimension identity matrix pxp.

The objective of factor analysis according to Wíchern; Johnson [17] is to find the matrices L_{pxm} e 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} that cannot be decomposed in the form LL´+ ***Ψ for a value of*** m less than p.

Johnson RA, Wichern DW (2002) Applied multivariate statistical analysis (5th ed). Pearson Education International, Upper Saddle River, USA Johnson RA, Wichern DW (2002) Applied multivariate statistical analysis (5th ed). Pearson Education International, Upper Saddle River, USA.

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.

$$P_{pxp} = \begin{bmatrix} \sum_{j=1}^{m} l_{1j}^2 & \sum_{j=1}^{m} l_{1j} l_{j2} & \cdots & \sum_{j=1}^{m} l_{1j} l_{p} \\ \sum_{j=1}^{m} l_{2j} l_{j1} & \sum_{j=1}^{m} l_{2j}^2 & \cdots & \sum_{j=1}^{m} l_{2j} l_{p} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{j=1}^{m} l_{p} l_{j1} & \cdot & \cdots & \sum_{j=1}^{m} l_{p}^2 \end{bmatrix} + \begin{bmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & \cdot \\ 0 & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ 0 & \cdot & 0 & \psi_p \end{bmatrix}$$

-----------(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 Z_i, is decomposed into two parts. The first, denoted by h_i^2, is the variability of Z_i explained by the m factors included in the factorial model. This part of the variability is called “commonality”, a name given due to the fact that the factors F_j, j = 1, 2, ..., m appear in all equations in the model (4.1) and the variables Z_i have a common source of variation. The second, denoted by ψ_i, is the part of the variability of Z_i associated only with the 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} e, portanto, Cov $(Z_i F_j)$ = Corr $(Z_i F_j) = l_j$, i = 1, 2, ..., p; j = 1,2,...,m.

This comes from the fact that

Cov (Z, F) = Cov $(E + \varepsilon, F)$ = Cov (E, F) + Cov (ε, F) = Cov (E, F) = L

So, you can use the matrix L_{pxm} in the search for understanding and interpretation of the factors F_j, j = 1, 2, ..., m. In order to operationalize factor analysis in practice, we first have to use mechanisms to estimate the value of m. From the estimated value of m we can then estimate the matrices L_{pxm} e ψ_{pxp}. (i_4) In relation to the total variance, the proportion explained by the factor F_j is given by:

$$PVTE_{Fj} = \frac{\sum_{i=1}^{p} l_{ij}^2}{p} \text{ ---------------- (4.5)}$$

The most representative factors in the model are those with the highest values of (4.5). It is common to express the values in (4.5) as a percentage.

Estimating the number of factors

The first step in conducting factor analysis is to estimate the theoretical correlation matrix P_{pxp} through the sample correlation matrix R_{pxp}, as was done in the principal component analysis. For the estimation of m, it will be enough to extract the eigenvalues from the R_{pxp} matrix and order them in descending order. It is observed, then, which eigenvalues are the most important in terms of numerical magnitude, using the following criteria: Criterion 1: the analysis of the proportion of the total variance related to each eigenvalue $\hat{\lambda}_i$, given by $\hat{\lambda}_i$ /p, i = 1, 2, ..., p. Those eigenvalues that represent greater 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 $\hat{\lambda}_i$ with the value 1, i = 1, 2, ..., p. The value of m will be equal to the number of eigenvalues $\hat{\lambda}_i$ greater or equal a 1. The basic idea of this criterion is to maintain new dimensions in the system 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 chart [18], which displays the values of $\hat{\lambda}_i$ sorted in descending order. By this criterion, a "jump point" is sought in the graph, which will represent 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 $\hat{\lambda}_1$ =2,24, $\hat{\lambda}_2$ =1,38, $\hat{\lambda}_3$ = 1,21, $\hat{\lambda}_4$ =0,63, $\hat{\lambda}_5$ =0,41, $\hat{\lambda}_6$ =0,13.

In this case, by Criterion 2, m would be estimated to be equal to 3. 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, that is, 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.

Matrix estimation methods L_{pxm} e ψ_{pxp}

Having chosen the value of m, it is possible to estimate the matrices L_{pxm} e ψ_{pxp}. Inicialmente, será introduzido o método de componentes principais, comumente utilizado como uma análise exploratória dos dados, em termos dos fatores subjacentes, por não exigir informações ou suposições sobre a distribuição de probabilidades do vetor aleatório Z. Initially, the principal component method will be introduced, commonly used as an exploratory analysis of the data, in terms of the underlying factors, as it does not require information or assumptions about the probability distribution of the random vector Z has multivariate normal distribution.

Principal component method for matrix estimation L_{pxm} e ψ_{pxp}

The principal component method works as follows: for each eigenvalue $\hat{\lambda}_i$, i = 1,2..., m retained in the estimation of the value of m, as discussed in section 3, there is the corresponding normalized eigenvector $\hat{e}_i$ Where $\hat{e}_i$ = The L_{pxm} and ψ_{pxp} matrices will be estimated respectively by:

$$\hat{L}_{pxm} = \left[\sqrt{\hat{\lambda}_1}\,\hat{e}_1 \; \sqrt{\hat{\lambda}_2}\,\hat{e}_2 \cdots \sqrt{\hat{\lambda}_m}\,\hat{e}_m\right] \text{----------(4.6)}$$

$$\hat{\psi}_{pxm} = \text{diag}\left(R_{pxp} - \hat{L}_{pxm}\hat{L}'_{mxp}\right) \text{------------(4.7)}$$

Where Diag (.) denotes the diagonal matrix. Thus, the ψ_{pxp} matrix has the main diagonal equal to the elements of the main diagonal of the matrix $(R - \hat{L}\hat{L}')$.

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 R_{pxp} matrix. For a fixed m value we have that:

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

Thus, an approximation for the LL` matrix will be given by:

$$\hat{L}\hat{L}' = \sum_{i=1}^{m} \hat{\lambda}_i \hat{e}_i \hat{e}_i' = \left[\sqrt{\hat{\lambda}_1}\,\hat{e}_1 \; \sqrt{\hat{\lambda}_2}\,\hat{e}_2 \cdots \sqrt{\hat{\lambda}_m}\,\hat{e}_m\right]\left[\sqrt{\hat{\lambda}_1}\,\hat{e}_1 \; \sqrt{\hat{\lambda}_2}\,\hat{e}_2 \cdots \sqrt{\hat{\lambda}_m}\,\hat{e}_m\right]'$$

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

$$\sum_{i=m+1}^{p} \hat{\lambda}_i \hat{e}_i \hat{e}_i' = R_{pxp} - \hat{L}_{pxm}\hat{L}_{mxp} \text{-------------(4.9)}$$

As the matrix in (4.9) is not diagonal, it cannot be used completely for the estimation of ψ_{pxp}. However, its diagonal can be considered. In this way, the matrix of specific variances is estimated as given in (4.7).

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

$$R \approx \hat{L}\hat{L}' + \hat{\psi} \text{-------------(4.10)}$$

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

MRES = R - $\left(\hat{L}\hat{L}' + \hat{\psi}\right)$-------------(4.11)

The residual matrix can serve as a criterion for assessing the quality of fit of the factorial model. Ideally, their 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.

Method of the main factors for the estimation of the matrices L_{pxm}

From Mingoti [16], it can be seen that another method can be used to estimate the matrices of specific loadings and variances introduced by Thompson (1934) and is called the method of main factors or iterative main 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 with a refinement of the L_{pxm} and ψ_{pxp} Lpxm estimates and generated by the principal component method. Considering the model P = LL '+ ψ , where P is the theoretical correlation matrix of the random vector of interest X. So, we have:

$$L` = P - \psi = \begin{bmatrix} h_1^2 & \rho_2 & \rho_3 & \cdots & \rho_{1p} \\ \rho_2 & h_2^2 & \rho_3 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \rho_{p3} & \cdots & h_p^2 \end{bmatrix} \text{-----------(4.12)}$$

Where $h_i^2 = 1 - \psi_i$, i = 1, 2. ..., p are the communalities. Suppose we estimate the matrix LL 'by R * given by:

$$R^*_{pxp} = \begin{bmatrix} h_1^{*2} & r_2 & r_3 & \cdots & r_{1p} \\ r_2 & h_2^{*2} & r_3 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & r_{p3} & \cdots & h_p^{*2} \end{bmatrix} \cong L^* L^{*'} \text{-----------(4.13)}$$

Where $\left(h_1^{*2}, h_2^{*2} \cdots h_p^{*2}\right)$ are initial estimates of communalities $\left(h_1^2, h_2^2 \cdots h_p^2\right)$. Using the principal component method, we have that:

$$L^* = \left[\sqrt{\hat{\lambda}_1^*}\,\hat{e}_1^* \; \sqrt{\hat{\lambda}_2^*}\,\hat{e}_2^* \cdots \sqrt{\hat{\lambda}_m^*}\,\hat{e}_m^*\right] \text{---------(4.14)*}$$

Where $\hat{\lambda}_1^* \geq \hat{\lambda}_2^* \geq \cdots \geq \hat{\lambda}_p^*$ are eigenvalues of R* and $\hat{e}_1^*, \hat{e}_2^*, \cdots \hat{e}_m^*$ are the respective standardized eigenvectors. From the L" matrix, therefore, we have new estimates of the communalities $\left(h_1^2, h_2^2 \cdots h_p^2\right)$, which are then placed on the main diagonal of the matrix in (4.13), and the procedure for estimating the L* matrix, using the principal component method, is repeated again. Proceed with the algorithm until the differences between the communalities of two successive interactions are negligible. Problems can occur during the execution of the algorithm, making it difficult to converge the iterative process.

For example, eventually at some stage of the procedure, some eigenvalues of the R* matrix can be negative, as these depend on the initial estimates of the communalities. This generates an inconsistency with the fact that a definite positive matrix is being estimated, as is the case with the correlation matrix.

There may also be, according to Rencher [19], the Heywood problem, caused by the fact that at some stage of the iterative process some estimate h_i^2 may be greater than 1, which generates a negative estimate ψ_1 of, and therefore inconsistent with, the definition of variance.

Principal component analysis (ACP) or (PCA)

The use of PCA can, in 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 highlight the technique and method applied to the statistical data analysis process, and in view of the research objectives, in order to leave it as a proposal for a real contribution to the Health System in Rondônia - more precisely for the planning area. From the task of explaining the method applied, the demonstration of the statistical way of how to obtain statistically results in a given sample follows.

Mingoti [16] proposes

Be $X=(X_1, X_2 \ldots X_p)'$ a random vector with measurements vector $\mu = (\mu_1\ \mu_2 \ldots \mu_p)'$ and covariance matrix Σ_{pxp}.

Be $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$ the eigenvalues of the matrix Σ_{pxp}, with the respective standardized eigenvectors $e_1, e_2, \ldots, e_p$, that is, eigenvectors and i satisfy the following conditions: (i) $e_i'\, e_j = 0$ for all i ≠ j ; (ii) $e_i'\, e_i = 1$ for all i = 1,2, ..., p; $\Sigma_{pxp}\, e_i = \lambda_i\, e_i$, for all i = 1,2, ..., p the eigenvector being denoted by $e_i = (e_{il}\, e_{i2} \ldots, e_{ip})'$. Consider the random vector Y = O' X, onde O_{pxp} is the orthogonal matrix of dimension pxp, consisting of the normalized eigenvectors of the matrix $\Sigma_{pxp,}$ that is

$$O_{pxp} = \begin{bmatrix} e_1 & e_1 & \cdot & \cdot & \cdot & e_{p1} \\ e_2 & e_2 & \cdot & \cdot & \cdot & e_{p2} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ e_{1p} & e_{2p} & \cdot & \cdot & \cdot & e_p \end{bmatrix} = \lfloor e_1 e_2 \cdots e_p \rfloor \text{----- (3.1)}$$

The vector Y is composed of p linear combinations of the random variables of the vector X, has vector of averages equal and O' μ and covariance matrix Λ_{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 make up the vector Y are not correlated with each other. Thus, the idea arises of using linear combinations in Y, as an alternative way of representing the covariance structure of the vector X, trying to obtain a reduction of the space of the variables, passing 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, with sector Y having the advantage of being composed of non-correlated random variables, thus facilitating their joint interpretation.

The following are some important definitions.

Definition 1

The j-th main component of the matrix Σpxp, 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 Y_j component are, respectively, equal to:

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

$Var\,[Y_j] = e_j{}'\, \Sigma pxp\, e_j = \lambda_j$

Where Cov $[Y_j, Y_k] = 0$, $j \neq k$. Each eigenvalue λ_j represents the variance of a main component Y_j. As 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 main component is defined at:

$$\frac{Var[Y_j]}{Variância\ Total\ d\ \ X} = \frac{\lambda_j}{Traço\left(\sum_{pxp}\right)} = \frac{\lambda_j}{\sum_{i=1}^{p} \lambda i} \text{------------(3.3)}$$

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 once

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

$$\left|\sum_{pxp}\right| = \prod_{i=1}^{p} \lambda_i = |\Lambda_{pxp}|$$

So, 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 in percentage. It is evident that the first main component has the highest proportion of explanation for the total variance of X.

Definition 3

The proportion of the total variance that is explained by the first k components

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

If the first k main components explain a large part of the total variance of the vector X, it may be possible to restrict the focus of attention only to the random vector $(Y_1\ Y_2\ ...\ Y_k)'$. In this way, a k-dimensional set of random variables can be examined, instead of a p-dimensional set, without losing a lot of information about the original variance and covariance structure of vector X. By the spectral decomposition theorem, by restricting the focus of attention to only the first k main components, the covariance matrix Σ_{pxp} will be approximated by the formula in (3.5):

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

Each portion of the sum in (3.5) involves an pxp dimension matrix corresponding only to the information of the j-th main component, j = 1,2,...,k. Thus, the original variability system of vector X will be approximated by the sum of k matrices, each representing the system of variability related to a component. When k=p, we have that the covariance matrix Σ_{pxp} is accurately reproduced by the sum of matrices related to the main components, that is,

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

Definition 4

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

$Y_i = a'_i X = a_i, X_1 + a_{i2} X_2 + ... + a_{ip} X_p$, i = 1,2,...,p

In this way, we have that:

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

$Cov(Y_i, Y_j) = a'_i \Sigma_{pxp} a_{j,}$ $i \neq j$, i, j = 1,2,...,p

Suppose that we want to find the values of the coefficients a_{ij} such that $a'_i a_j = 1$, so that the linear combinations $Y_1, Y_2, ..., Y_p$ were not correlated with each other and had maximum variance. So, it can be demonstrated by Wichern; Johnson [17] that the maximum variance of Y1 = a'1 X, under the constraint $a'_1 a_1 = 1$, is equal to λ_1 and is obtained when $a_1 = e_1$, that is, the normalized eigenvector corresponding to λ_1, where Y_1 called the first major 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, with Y_3 being called the third main component. Following the procedure, the maximum variance of $Y_i = a'_i X$, under the restrictions $a'_i a_i = 1$ and e $cov(Y_i, Y_j) = 0$ j <, is obtained when $a_i = e_i$, that is, the normalized eigenvector corresponding to the eigenvalue λ_i, Y_i being called the i-th main component. In this way, the p main components are constructed, being unique, except for the change of signal of all its coefficients. So, for example, if Y_i is a major component, $-Y_i$ will also be a major component of order i.

Due to the form of construction itself, the first main component is always the most representative in terms of total variance and the p-th is always the least representative.

Estimation of the main components: covariance matrix

In practice, the matrix Σ_{pxp} is unknown and needs to be estimated using the sampled data collected. In general, the matrix Σ_{pxp} is estimated by the sample covariance matrix S_{pxp}.

Be $\lambda_1, \lambda_2, ..., \lambda_p$ the eigenvalues of the matrix S_{pxp}, and be $\hat{e}_1, \hat{e}_2, ..., \hat{e}_p$, the corresponding standardized eigenvectors. So the j-ésima estimated main sample 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-------(3.6)

Some properties of the main sample 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 to say that these components are uncorrelated.

Property 3

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

$$\frac{Var\left[\hat{Y}_j\right]}{Variância\ Total\ Estimada\ de\ X} = \frac{\hat{\lambda}_j}{traço(S_{pxp})} = \frac{\hat{\lambda}_j}{\sum_{i=1}^{p} \hat{\lambda}_i}$$

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_i is the sample variance of the random variable X_i.

Property 5

By the spectral decomposition theorem, the matrix of $\boldsymbol{S_{pxp}}$ 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*** main sample components are used.

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

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

Results and Discussion

Scenario 1 = 121 epidemiological, socioeconomic and coverage variables of health services in the state of rondônia

The exploration of the data started from the principle that the ignorance of a certain situation is only eased with the confrontation of reality. In the application of the Factor Analysis (PA) technique and the Principal Component Analysis (PCA) method, the findings are in line 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 process of decision-making when distributing technical, financial and human resources. In the application of AP and PCA in the variables, with 121 variables grouped and considered from table 5-7, the statistical inferences explain the following findings with an explanation above 70%.

Factor 2 called socioeconomic indicators highlighted 4 variables with an explanation above 70%, which is described in table 8.

The factor (3) health service coverage indicators stood out only 2 (two) variables with an explanation above 70%, being described in table 9.

46 epidemiological variables with AF and ACP over 70%	
Hospitalization for infectious and parasitic diseases	Hospitalization in 5-year-old children
Hospitalization for dehydration	Hospitalization for leprosy
Tuberculosis hospitalization	Hepatitis hospitalization
HIV hospitalization	Hospitalization for mycoses
Hospitalization for malaria	Hospitalization for leishmaniasis
Hospitalization for câncer	Hospitalization diabetes mellitus
Hospitalization for mental illness	Hospitalization for neurological diseases
Hospitalization for ophthalmic diseases	Hospitalization for circulatory diseases
Hospitalization for respiratory diseases	Hospitalization for digestive diseases
Diseases due to skin diseases	Hospitalization for musculoskeletal disorders
Hospitalization for urinary tract disease	Hospitalizations due to pregnancy complications
Perinatal infections hospitalizations	Congenital malformations
Injuries due to external causes	Infant death by residence
Infant death	Death from external causes
General death	Type 1 diabetes
Type 2 diabetes	Hypertension
Hypertension and diabetes	Hansen’s disease
Disease due to tuberculosis	Dengue disease

Table 5: Demonstration of the epidemiological variables extracted from the application of PA and ACP, above 70%.

Source: Based on secondary data.

17 socioeconomic variables with AF and ACP over 70%	
Percentage of people in households with a computer	Percentage of people in households with a telephone
Women aged 15 and over	Women 25 years of age and over
Population up to 1 year of age	Population 10 to 14 years of age
Population aged 15 years and over	Population 25 years of age and over
5-year-old population	Population 65 years of age and over
Total population	Urban population
Annual average study of people aged 25 or over	Percentage of people aged 25 and/+ with less than 8 years of study
Percentage of people aged 25 and/+ with less than 8 years of study	Human - school development index (HDI-E) Literacy rate

Table 6: Demonstration of socioeconomic variables extracted from the application of PA and ACP, above 70%.

Source: Based on secondary data.

6 variables health services coverage with AF and ACP over 70%	
Born alive without prenatal care	Born alive with 1-3 prenatal consultations
Born alive with 4-6 prenatal consultations Prenatal consultation of 7 and/+, average number of visits	Cervical examination Accompanied/average family number

Table 7: Demonstration of health service coverage variables extracted from the application of PA and ACP, above 70%.

Source: Based on secondary data.

4 socioeconomic variables with Af and Acp over 70%	
Percentage of children aged 10 to 14 years with more than 1 year of school delay Percentage of people in households with running water	Percentage of children aged 10 to 14 with less than 4 years of schooling Percentage of children aged 7 to 14 years out of school

Table 8: Demonstration of socioeconomic variables extracted from the application of PA and ACP, above 70%.

Source: Based on secondary data.

2 health services coverage variables with AF and ACP over 70%	
Percentage of government income	Percentage of people with more than 50% of income from government agencies

Table 9: Demonstration of health service coverage variables extracted from the application of PA and ACP, above 70%.

Source: Based on secondary data.

Statistically in the context of the factorial exploration it was found that the factor 1 epidemiological indicators presented an explanation of variance of 69.90%, the factor 2 socioeconomic indicators of 14.47% and factor 3 indicators of coverage of health services 6.46%. The set of factors accumulated eigenvalues of 92.84% of explanation extracted from the Principal Component Analysis (PCA). Considering what Mingoti points out [16] “eigenvalues are ordered in descending order, the first component is the one with the greatest variability and the worst 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 health service coverage indicators.

Considering also the notes of Mingoti [16], the first main component is placed in the multivariate analysis in this study as of greater representativeness, and by the inserted context of greater importance in view of the object of study. Table 10 shows the statistical findings.

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

Table 10: Demonstration of eigenvalues, in the application of ACP in the 121 variables, in the year of study.

Source: Based on secondary data.

Another finding that draws attention is the coordination of factors with cases. There are in factor (1) 13 cases with negative inferences, in factor (2) 27 cases and in factor (3) 24 cases. Of these, the health regions headquartered by the following Municipalities are noteworthy: Porto Velho, Ariquemes, Ji-Paraná, Cacoal, Rolim de Moura and Vilhena, noting that only Porto Velho (capital) is negatively presented in a decreasing situation in the three factors, shown briefly the health regions in table 11.

On the other hand, table 12, which deals with the contribution of cases based on the correlations of the variables, shows that the largest contribution is from the Health Region of Porto Velho with 72.84% and 17.64% of the concentration of diseases in the factor 1 and 2 consecutively.

Porto Velho, specifically, corroborates the notes of Mingoti [16], because in this case it may be associated with the quantitative component of the diseases, justified even by the characteristic of this Health Region, which tends to concentrate the procedures as it is the only reference of the State for medium and high complexity procedures.

Scenario 2 = 42 epidemiological, socioeconomic and coverage variables of health services with explanation above 95%. in the state of rondônia

From the exploration of the 121 variables, there was a need to consider the notes of Hair., *et al.* [20], which proposes, when using Factor Analysis (PA), 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 is 95%.

County	F (1) IND. epidemiological	F (2) IND. socioeconomic	F (3) IND. coverage of health service
5 Ariquemes	-7,8181	4,2101	2,92478
9 Cacoal	-8,8542	5,4915	2,88012
24 Ji-Paraná	-11,4610	4,6260	2,01097
37 Porto Velho	-47,8678	-11,4336	-2,23120
41 Rolim de Moura	-4,9203	6,8001	0,65169
52 Vilhena	-9,0195	8,5311	1,33950

Table 11: Demonstration of negative and positive cases (city where the health region is based) based on correlations, extracted from the Principal Component Analysis (PCA), with 121 variables.

Source: Based on secondary data.

Regions of health	F (1) IND.epidemiological	F (2) IND. socioeconomic	F (3) IND. coverage of health service
Porto Velho	72,84	17,64	1,71
Ariquemes	1,94	2,39	2,94
Ji-Paraná	4,18	2,89	1,39
Cacoal	2,49	4,07	2,85
Rolim de Moura	0,77	6,24	0,15
Vilhena	2,59	9,82	0,62
All others	15,19	56,96	90,34
Grand total	100,00	100,00	100,00

Table 12: Demonstration of contribution of cases, by factor, based on a correlation of 121 variables, Rondônia.

Source: Based on secondary data.

According to Hair., *et al.* [20], if in the interpretation of PA it is found that:

- **Factorial matrix**: It 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 the variation in the data as a whole more 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;

- **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;
- **The factorial loads as cargas fatoriais**: Is the correlation of each variable with each factor. The loads indicate the degree of correspondence between the variable and the factor, with higher loads making the variable representative of the factor.

Observing these contributions by Hair., *et al.* [20] and as proposed in the research procedure, Scenario 1 represents the findings of the application of PA and ACP in the 121 variables. Scenario 2, in turn, represents the findings of the application of PA and ACP in variables with an explanation above 95% in the State of Rondônia. It should be made clear that Scenario 2 is extracted from Scenario 1, with the objective of obtaining greater clarity in the exploration of data and envisioning a greater possibility of contributing to the situational health diagnosis in the planning area.

The variables with an explanation above 95% are shown in table 13-15 below.

In this scenario of exploratory analysis, the application of PCA obtained that, of the 42 variables, 34 had an explanation above 70%. Of these, the expectations of high explanation were confirmed in factor 1 epidemiological indicators with 24 variables with an explanation above 95%. Table 16 shows the concentration of variables by factor.

25 socioeconomic variables with exp. over 95%	
% children 10 to 14 years old with more than 1 year of school delay; CUMATES	% children from 10 to 14 years old with less than 4 years of study; CQUATES
% of child 7-14 years old with 1 year of school delay; CÇ7-14C1AATRASADO	% children from 7 to 14 years old in elementary school; CENFUN
% children in households with per capita income less than R $ 75.50; CDVIQMS	% child 7-14 years out of school; CÇ7-14AFORAESC
% of people with more than 50% of income from government sources; PRMGOV	% of government income; RENDAGOV
% of people in households with electricity and TV; PDENERG	% of people in households with running water; PDAGENC
Women 25 years of age or older; MULAC25ANOS	Women aged 15 and over; NMAC15ANOS
Population between 10 and 14 years of age; POP-10-14ANOS	Population up to 1 year of age; POP1ANO
Population aged 25 or over; POPAC25ANOS	Population aged 15 or over; POPAC15ANOS
Population 65 years of age and over; POPAC65ANOS	5-year-old population; POP<5ANOS
Urban population; POPTOTAL	Total population; POPRURAL
Probability of survival up to 60 years; PROBSOBA-C65ANOS	Probability of survival up to 40 years; PROBSOBAC40A-NOS
Municipal Human Development Index Longevity; IDHM-LONG	Municipal Human Development Index; IDHM Life expectancy at birth; ESPVIDAONASC

Table 13: Demonstration of socioeconomic variables with explanation above 95% in the application of PA and ACP for the State of Rondônia.

Source: Based on secondary data.

14 epidemiological variables with exp. above 95%	
Hospitalization of infectious and parasitic diseases; DOENINFEC-PARASIT	Hospitalization for digestive diseases; D.DIGESTIVA
Hospitalizations due to pregnancy complications; COMPGRAVIDEZ General death; ÓBITOGERAL Death from External Causes; OB-CAUS-EXTER	Injuries due to external causes; LESCAUEXTER Infant death; ÓB-INF Infant death by residence; OB-INF-RES

Table 14: Demonstration of epidemiological variables with an explanation above 95% in the application of PA and ACP for the State of Rondônia.

Source: Based on secondary data.

3 health services coverage variables with exp over 95%	
Cervical examination; EXPREVENTIVO	Prenatal consultation of 7 and/+; NVC/7CONSPN
Average visit number; MÉDIAVISDOM	

Table 15: Demonstration of epidemiological variables with an explanation above 95% in the application of PA and ACP for the State of Rondônia.

Source: Based on secondary data.

Factors	Explanation	Variables
F (1) epidemiological	95% - 99%	27
F (2) socioeconomic	70% - 82%	08
F (3) health service coverage	70% - 80%	03

Table 16: Demonstration of the concentration of variables by factor and percentage of explanation.

Source: Based on secondary data.

Mingoti [16] states that the variance serves to measure the degree of linear relationship between two variables. Thus, in the application of ACP 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 to the 8th place, 99.99% of explanation is obtained, according to table 17. It can also be highlighted the variability associated with random error ε_i which is specific to each variable.

An interesting confirmation is the correlation between factors 1, 2 and 3, which shows 66.53% for epidemiological indicators, 18.72% for socioeconomic indicators and 7.53% for health coverage indicators. services. The finding corroborates what Hair., *et al.* [20] says "the first factor can be seen as the best summary of linear relationships shown in the data".

According to Reis [21] and Mingoti [16], commonality is "the total amount of variance that an original variable shares with all other variables included in the analysis" or "the variables have a common source of variation". In Scenario 1 the set of 121 variables (epidemiological, socioeconomic and health service coverage) showed 61 variables with communalities above 70%. In Scenario 2, the set of 42 variables (epidemiological, socioeconomic and health service coverage with explanation above 95%) presented 16 variables with communalities above 95%.

Conclusion

There is an almost absolute predominance of epidemiological variables in factor 1, which accounts for more than 70% of the explanation, with a total of 54 variables, 46 of which account for 66.90% of explanation. This fact, external, as well as the affinity of the object of the work in the face of being represented by the

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 17: Demonstration of matrix covariance eigenvalues (factors).

Source: Based on secondary data.

Variables	F (1) IND. epidemiological	F (2) IND. Socioeconomic	F (3) IND. coverage of health service	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 18: Demonstration of commonality in variables above 95% of explanation.

Source: Based on secondary data.

health problems, mainly the causes of hospitalizations and deaths. As for the factor 2 of the 57 socioeconomic variables, there was a predominance of 04 variables that tend to express the quality of the education system in Rondônia due to the presence of indicators that measure access to basic education. The factor 3 - of the health service coverage variables - was not significant 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%) that can be justified by the presence of conditions common to the others municipalities, for example, dengue, tuberculosis, leprosy, injuries from external causes, among others, here only observing numerical values. 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 municipality in concentrating services of medium and high complexity in health and being references for all municipalities in 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), urgent and pediatric emergency (Hospital Infantil Cosme e Damião) and in infectious-contagious disease (Hospital Centro de Medicina Tropical de Rondônia). In this case, some variables such as hospitalization and death can directly influence the municipality's contribution.

Scenario 2 aimed to stratify the result of Scenario 1, in order to ensure better clarity to the study. In the application of AF and ACP in the 42 variables with explanation above 95%, it 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 serve the group of access to basic education, income, non-health public goods and services. It can be said that the commonality statistical attribute directly influences the research result, since it was highly present in the first Scenario 1 with 61 variables and in the second

Scenario 2 with 16 variables. Among the variables that showed commonality for this study, one can consider the most important ones: number of preventive cancer exams, pregnancy complications and number of deaths in the period corresponding to 4 years of studies.

Even in an empirical way, health managers in the state of Rondônia are not acting far from the health needs of the population. As for the agreed indicators, 07 (seven) are part of the service coverage indicators and 26 of the epidemiological indicators. However, it is possible to attribute a criticism to the non-agreement of socioeconomic indicators, which shows a low power of articulation in the field of intersectoral policy between the segments of education, public security, environment and others that infer reach in the area of public health.

Bibliography

1. Brasil. "Constituição da república federativa do brasil de". Promulgada em 05 de outubro de 1988. Versão eletrônica, revista e atualizada (1988).

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 (1990).

3. Brasil. Lei nº 8.142, de 28 de dezembro de. 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 (1990).

4. Brasil. MINISTÉRIO DA SAÚDE. Norma Operacional da Assistência à Saúde - NOAS-SUS 01/2001 (2019).

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.

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). Relatório de Gestão. Departamento de Economia da Saúde (2005).

9. Brasil MS SE. "Aspectos Distributivos dos Gastos Públicos na Saúde". Projeto de Reforço à Reorganização do SUS – Reforsus. Série C. Projetos Programas e Relatórios. Brasília: Ministério da Saúde (2006).

10. Brasil MS SE. "Departamento de apoio à descentralização. coordenação-geral de apoio à gestão descentralizada". Diretrizes operacionais dos Pactos pela Vida, em Defesa do SUS e de Gestão (Série A. Normas e Manuais Técnicos). Brasília: Ministério da Saúde 1 (2006).

11. Piola Sérgio Francisco and Biasoto Júnior G. "Financiamento do SUS nos Anos 90". In: NEGRI, B. e DI GIOVANNI, Geraldo. (Orgs.). Brasil: Radiografia da Saúde. Campinas: UNICAMP: Instituto de Economia (2001).

12. Lucchese Patrícia Tavares Ribeiro. "Processo Alocativo e Redução da Desigualdade Regionais. Equidade na Gestão do Sistema Único de Saúde". *Ciência and Saúde Coletiva* Relatório Final, ENSP/FIOCRUZ, Rio de Janeiro (2003).

13. Escoda Maria do Socorro Quirino. "Iniquidade em Saúde". *Revista: Espaço para Saúde* 5.2 (2004).

14. UGÁ Maria Alicia Domínguez and SANTOS Isabela Soares. "Uma Análise da Equidade do Financiamento do Sistema de Saúde Brasileiro". *Cadernos de Saúde Pública* Rio de Janeiro ago. 22.8 (2006).

15. Porto Silvia Marta. "Metodologia de alocação equitativa de recursos". Rio de Janeiro: FIOCRUZ (2003).

16. Mingoti Sueli Aparecida. "Análise de Dados através de Métodos de Estatística Multivariada: Uma abordagem aplicada". Belo Horizonte: Editora UFMG (2005).

17. Wichern D W and Johnson R A. "Applied Mutauariate Statistical Analysis". Londres: 3=" Ed, Englewood (N.J,): Prentice-Hall, Inc (1998).

18. Cattell R B. "The Scree Plot Test for the Number of Factors". *Multivariate Behavioral Research* 1 (1966): 140-161.

19. 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.

20. Hair Jr., *et al*. “Análise Multivariada de Dados”. 5ª ed., Trad. Adonai Schlup Sant´Anna e Anselmo Chaves Neto. Porto Alegre: Bookman (2005).

21. Reis Elizabeth. “Estatística Multivariada Aplicada”. 2ª ed. Revista e corrigida. Lisboa: Edições Silabo (2001).