https://dx.doi.org/10.22161/ijaers.710.16 ISSN: 2349-6495(P) | 2456-1908(O

Logistic Regression Models and Classification Tree for Deaths and Recovered Patients Records of Covid-19 in the State of Minas Gerais, Brazil

Fabrício Pelizer Almeida[1], Moisés Keniel Guilherme de Lima[2], Demóstenes Coutinho Gomes[3], Esther Ferreira de Souza[4]

[1,2,3]Professor of the Department of Environmental and Civil Engineering, University of Uberaba, UNIUBE, Uberlandia-MG, Brazil.
[4]Undergraduate Student, Department of Environmental and Civil Engineering, University of Uberaba, UNIUBE, Uberlandia-MG, Brazil.

***Abstract*— *The challenges for the construction of a pandemic confrontation agenda by COVID-19 in Brazil come up against social inequalities, as a reflection of the segregation of access to comprehensive basic sanitation services and public health assistance programs. The objective of this work is to analyze the profile of deaths and recoveries by COVID-19 in the state of Minas Gerais, based on socio-environmental predictors, using a logistic regression model and classification tree (CHAID). Data on recovered individuals and confirmed deaths for COVID-19 were obtained from the Minas Gerais State Department of Health, containing records of age, sex, race, comorbidity and municipality of residence. The data regarding municipal basic sanitation were obtained from Instituto Trata Brasil. The Minitab and SPSS software were used in the elaboration of the logistic regression models and classification tree, respectively. The probability of death from COVID-19 in the state is significantly higher in males, over the age of 60 years old, with some comorbidity, declared black and brown, living in municipalities located in the poorest macro-regions of the state, where classes prevail inadequate or inadequate basic sanitation. The classification tree for deaths by COVID-19, differentiates young blacks and browns without comorbidity, and the elderly with comorbidity not assisted by a comprehensive basic sanitation network. It is concluded that the worsening of the pandemic in the state is related to aspects of social vulnerability, and that the implementation of inclusive public policies is urgent.*

***Keywords*— Pandemic, Basic sanitation, Racism, Statistical modeling.**

I. INTRODUCTION

The main discussions in the global political scenario, in view of the COVID-19 pandemic, are being directed towards the construction of an agenda capable of enabling structural and long-term solutions, and which considers equality, cohesion and social justice, through integrated and coordinated public policies. Therefore, studies on the evolution stages of the disease scenario by COVID-19, consider patterns of dissemination according to social, environmental, economic and political factors [39].

In more recent analyzes, combining data from records of those infected by severe infectious respiratory pandemics globally, they provide evidence that the increased risk in a population is largely driven by social vulnerability factors, and aggravated by disproportionate access to healthcare. health, especially basic sanitation services [21, 32]. Qualitative aspects and the recurrence of treated water supply and sewage services in peripheral global communities are put to the test as major factors in the transmission, severity and prevalence of diseases similar to COVID-19 [8, 43-46].

The complexity of the socio-environmental impacts of COVID-19 is extensive considering the developments and social strata, including the dynamics of risk factors associated with certain population groups. According to Men over 50 and with a history of pre-existing illnesses form the most vulnerable group. For the authors, viral infection can progress rapidly to cases of fatal respiratory diseases or acute respiratory failure [8].

The unprotected health of the elderly population associated with the prevalence of severe comorbidities (such as diabetes, obesity, hypertension), which are so prevalent in urban and peripheral conglomerates in poor and developing countries, is an important condition for

https://dx.doi.org/10.22161/ijaers.710.16 ISSN: 2349-6495(P) | 2456-1908(O

deaths due to COVID-19 [16, 25]. In Brazil, hypertension and diabetes are among the prognostic comorbidities most related to deaths caused by the new coronavirus [4].

The worsening of the spread of the disease in social strata in vulnerable situations in the field of health and care, finds support in the aspects of racism and ageism present in Brazilian society. A significant part of black elderly people lives in small and medium-sized cities in houses with few rooms, with intergenerational family arrangements and lacking material resources and, in many cases, with a lack of complete information about the disease and its severity [35]. The black and elderly population with immunosuppressive and hematopoietic diseases are at additional risk to the COVID-19 pandemic [42].

Even with specific objectives that ensure the state's protective strategic actions, outlined in the National Health Policy for the Elderly, and in the National Policy for Integral Health for the Black Population, the elderly black population maintains its invisibility in facing this pandemic, due to lack of transparency in the dissemination of epidemiological bulletins, and underreporting in the incidence and mortality rates due to COVID-19 [5-6, 31, 35].

Specifically, ethnoracial aspects were not eligible for analysis of the epidemiological situation of COVID-19 in the first epidemiological bulletins and, therefore, represented little public policy strategies and formulations at the beginning of the pandemic, in the different states of the federation. There is evidence of underreporting by COVID-19 in the country, given the frequent incompleteness of the race item in the different disease notification forms. also considering that the black population in the country as a whole, has less access to health services, and represents the largest proportion among vulnerable populations, who secularly experience the absence of public power in their territories [15, 34].

Associated with ethno racial factors, the country's socioeconomic scenario measured between the years 2012 and 2018, according to the Synthesis of Social Indicators (SIS), of the Brazilian Institute of Geography and Statistics (IBGE), further exposes the fragility of social groups and the worsening of extreme poverty in Brazil, in view of the evolution of the pandemic. The study points out that 13.5 million people in Brazil live with per capita monthly income below U$S 1.9 per day, representing 6.5% of the country's population. Poverty mainly affects the black and brown population, who represent 72.7% of the poor (38.1 million people). And, black or brown women make up the largest contingent, about 27.2 million people below the poverty line [17].

The same study points out that 56.2% of the total population below the poverty line (or 29.5 million people) do not have access to sanitation in their homes; 25.8% (or 13.5 million people) are not served by a water supply; and 21.1% (11.1 million people) do not have garbage collection. The proportions are higher among blacks and browns than among whites, both in relation to inadequate housing conditions and the lack of provision of sanitation services. Among blacks and browns, 42.8% (49.7 million people) are not served with sewage collection; 17.9% (20.7 million people) do not have water supply through a network; and 12.5% (14.5 million people) do not have access to garbage collection [17].

Therefore, it is very important to understand the implications of these imminent social, economic and environmental challenges in Brazil, in the context of the COVID-19 pandemic, considering the peculiarities of the states of the federation, with special attention to vulnerable populations, and for the estimation of possible impacts of the pandemic in the public health system [10]. This is a key research topic in this study. Therefore, the aim of this work is to analyze the profile of deaths and recoveries by COVID-19 in the state of Minas Gerais, based on socio-environmental predictors, using a logistic regression model and classification tree (CHAID).

II. METHODOLOGY

2.1. Database and Sociodemographic Characterization of the State of Minas Gerais

The study is supported by the survey of official notifications of cases of recovered individuals and confirmed deaths for COVID-19 at the website of the State Department of Health of Minas Gerais (SES/MG), from March 4 to September 10, 2020, and without any description that identifies the patient by name. The records of patients "In Monitoring", "Canceled" and "Ignored" were excluded (Table 1).

The secondary data considered in this research also include the records of information of individuals such as sex, age, race, comorbidity and municipality of residence. We also opted for the arbitrary exclusion of those incomplete records of any information, fields missing or not correctly described about the individuals, so that there is no incorrect or incomplete classification of the response variable in relation to the components of the model. Table 1 shows the variables used in the explanatory models, according to the type, composition and description of the

https://dx.doi.org/10.22161/ijaers.710.16 ISSN: 2349-6495(P) | 2456-1908(O

entry, paying attention to those that have undergone adaptations or some conversion.

The population and the total territorial area of the state of Minas Gerais, is 21.3 million inhabitants and 586.521,123 km² respectively, being the second most populous in Brazil, behind only the state of São Paulo [17]. The estimated rate of urbanization in the state is 85.3%, configuring a predominantly urban population [26].

Table 1. Description, types and composition of the response and predictor variables in the Logistic Regression (LR) and Decision Tree (DT) models.

Variable	Description	Type[1]	Models	Composition
'E.Covid-19'	Evolution of the COVID-19 cases	CaR	LR/DT	Death; Recovered.
'A'	Age	CoP	LR	Informed by the patient.
		CaP	DT	<1 year; 1 to 9 years old; 10 to 19 y.old; 20 to 29 y.old; 30 to 39 y.old; 40 to 49 y.old; 50 to 59 y.old; 60 to 69 y.old; 70 to 79 y.old; 80 to 89 y.old; 90 or more.
'R'	Race	CoP	LR	(1) Yellow; (2) Indigenous; (3) White; (4) Brown; (5) Black[2]
		CaP	DT	Yellow; Indigenous; White; Brown; Black
'C'	Comorbidity	CoP	LR	(1) Not Informed; (2) No; (3) Yes.
		CaP	DT	Not Informed; No; Yes.
'S'	Sex	CaP	LR/DT	Women (Female). Men (Male).
'PR'	Planning Regions	CaP	LR/DT	Grouping of municipalities corresponding to the 10 (ten) regions of the state of MG.[3]
'CSS'	Class of Sanitation Service	CaP	LR/DT	Inadequate (the Service Index without Collection and without Treatment prevails); Little Adequate (the Service Index with Collection and without Treatment prevails); Adequate (the Service Index with Collection and Treatment prevails).

[1]CaR: Categorical Response; CoP: Continuous Predictor; CaP: Categorical Predictor. [2]Based on studies [15, 36]. [3]João Pinheiro Foundation, an autarchy linked to the State Secretariat of Planning and Management of the State of Minas Gerais (SEPLAG).

The municipalities were grouped according to the division into Planning Regions state of Minas Gerais adopted by the State Secretariat for Planning and Management of the State of Minas Gerais (SEPLAG), and not individualized (Fig. 1). This condition allows to classify and compare regions of the state according to parameters of the state government itself [27].

The IDHM - Municipal Human Development Index - obtained through the geometric average of 3 sub-indices (income, education and longevity), was 0.769 for the state, while in Brazil it was 0.761 in 2015 [26]. Approximately 39.7% of the population declares itself to be white in the state, while browns and blacks represent 48.2% and 11.8%, respectively [17].

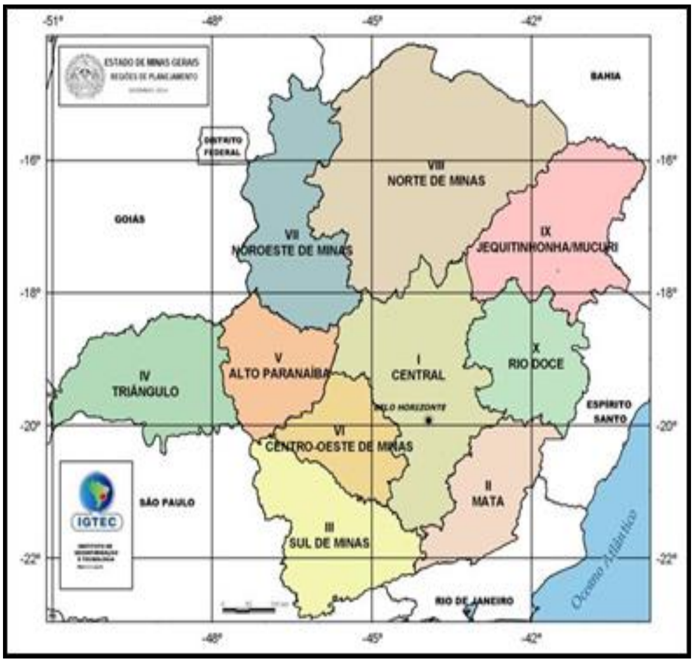

Fig. 1. Planning regions of the territory of the state of Minas Gerais adopted by the State Secretariat of Planning and Management of the State of Minas Gerais (SEPLAG).

Source: [27].

https://dx.doi.org/10.22161/ijaers.710.16 ISSN: 2349-6495(P) | 2456-1908(O

The data referring to basic sanitation in the municipalities of Minas Gerais were obtained from the Sanitation Panel Brazil, next to the website of the Instituto Trata Brasil - ITB (2020), which is based on data from the National Sanitation Information System (SNIS) - base year 2018. Minas Gerais was the second state of the federation in absolute investments in sanitation between 2014 and 2018, about R$ 6.49 billion, but in investments per capita, the state occupies the twelfth position, with investment of R$ 62.6 per inhabitant*year^{-1} [19].

2.2. Statistical Models Used in Research

The study was developed using the qualitative response model (evolution of COVID-19 cases), in which the response variable is binary, i.e., it assumes two possible events, deaths and recovered. A linear probability model, using the logit link function, simplified in eq. 1, with only two parameters is given by:

$$logit_i = g(p_i) = \beta_0 + \beta_1 x_i \qquad (\mathbf{1})$$

or, better described in eq. 2, the link function that relates the mean of the response in the ith observation to a linear predictor:

$$g(\mu_i) = X_i\beta \qquad (\mathbf{2})$$

since, μ_i mean response of the nth line, $g(\mu_i)$, the link function, X is the vector of the predictor variables and β, the vector of the coefficients associated with the predictors of the model.

It is reasonable to implement a link function that specifies a non-linear transformation to model responses, in which the dependent variable is related to the explanatory variables in a non-linear way [1]. In these terms, the regression model that uses logistic distribution as a link function describes the response groups, as described in eq. 3, for the set of predictor variables.

$$p = \Pr[Y = 1|X = x] = \Lambda(\beta' x) = \frac{e^{\beta' x}}{1+e^{\beta' x}} = \frac{1}{1+e^{-\beta x}} \qquad (\mathbf{3})$$

For the composition of the logistic regression model, continuous variables (Age, Race and Comorbidity) and categorical variables (Sex, Class of Sanitation Service and Planning Regions) were adopted. The outputs of results from the Binary Logistic Regression (LR) model were obtained using **MINITAB v. 19** [28].

Still as an analytical scope of this study, a decision tree was proposed for the evolution of COVID-19 cases, using the CHAID (*Chi-Square Automatic Interaction Detection*) method. This tree model is well used when segmentation is defined in terms of demographic characteristics or categorical variables with predictive power [24]. The CHAID method, which admits a binary categorical dependent variable in the model composition, maximizes the significance of the chi-square statistic in each partition.

The CHAID method is described based on the maximum (chi-square statistics) using $T_{(j)}^{(*)}$ the stepwise procedure evaluating the input of each independent categorical variable in the model, and checking whether its contribution is significant or not among the predictor variables in the construction of nodes and sub-nodes [20].

The validation method was obtained from an automatic test set, with the selection of a fraction of 30% of the lines and a basis for generating random data. The simulations for the construction of the decision tree were carried out with the support of **IBM SPSS v. 25** [18].

III. RESULTS

3.1. Logit model for COVID-19 evolution data

The binary logistic regression model (Table 2) for the composition of the response variable (Y'), evaluated the coefficients of the continuous predictors (age, race and comorbidity) and the categorical groups that identify the patient's gender (male/female), the class of sanitation service (inadequate, little adequate, adequate), and the Jequitinhonha-Mucuri region as significant at 1% ($p < 0.001$). The regions of Rio Doce, Sul de Minas and Centro-Oeste of the state of Minas Gerais, which also make up the predictive variable 'Regions of Planning', were significant at 5% ($p < 0.05$) for the evolution picture of individuals diagnosed with COVID-19.

Table 2: Output of results from the Binary Logistic Regression model in the composition of the variable (Y') of evolution of COVID-19 (event: death) according to continuous and categorical predictors for the state of Minas Gerais, Brazil.

Continuous Predictors	Coef.	SE Coef.	Z	*p-value*	VIF
Intercept	-8,884	0,211	-42,08	0,000 ***	
'Age' – 'A'	0,0624	0,00151	41,25	0,000 ***	1,06
'Race' – 'R'	0,2290	0,0317	7,23	0,000 ***	1,07
'Comorbidity' – 'C'	1,3033	0,0372	35,00	0,000 ***	1,06
Categorical Predictors	**Coef.**	**SE Coef.**	**Z**	***p-value***	**VIF**
'Sex' – 'S'					
Male	0,3315	0,0456	-7,27	0,000 ***	1,02
'Class of Sanitation Service' – 'CSS'					
Inadequate	0,2842	0,0568	5,00	0,000 ***	1,57
Little Adequate	0,3186	0,0706	4,52	0,000 ***	1,20
'Planning Regions ' – 'RP'					
Central	-0,120	0,130	-0,92	0,358 ns	7,89
Centro-Oeste	-0,480	0,173	-2,78	0,005 **	2,02
Jequitinhonha-Mucuri	0,780	0,183	4,27	0,000 ***	1,81
Noroeste	0,279	0,215	1,30	0,195 ns	1,51
Norte	-0,273	0,175	-1,56	0,119 ns	1,98
Rio Doce	-0,274	0,141	-1,95	0,051 **	4,20
Sul de Minas	-0,266	0,144	-1,85	0,064 **	3,61
Triângulo	-0,113	0,138	-0,82	0,410 ns	4,86
Zona da Mata	-0,231	0,146	-1,58	0,114 ns	3,73

P(death) = exp(Y')/(1 + exp(Y')). Training Count (death; recovered): (3856; 16865). Test Count (death; recovered): (1691; 7190). Lines used (29602). Test set fraction: 30%. Area under the ROC curve (89.77%) and Test área under the ROC curve (89.71%). * significant coefficient at 10%, ** at 5% and *** at 1%.

The results of the logit model indicated that the probability of occurrence of the event (death), responds positively to the gradual increase of the patient's age, to the ethnorracial aspects and to the association with some comorbidity. The chance of death in patients with COVID-19 increases 1.06 times with the advancing age of the patient, 3.68 times in the sense of the declarants as black and brown (in relation to whites, yellow and indigenous) and, 1.25 times when associated with a pre-existing disease (in comparison with those without comorbidity or not reported), as shown in Table 3.

Individuals in the male group were more likely to compose the event profile (death), given the positive marginal effect (0.3315). The chance of deaths from COVID-19 among men is 1.39 times more likely than among women (Table 3). According to Table 4, which has chi-square statistics for the contrasts between the groups that make up the study variables, the contrast between the groups - male *vs* female - regarding the contribution to the occurrence of deaths by COVID-19 is quite high (94.497) and significant (p <0.001).

The significant impact of the marginal effect of different classes of sanitation services on the response variable is also an important aspect of the research. The chances of obituaries in the municipalities where the basic sanitation class attendance rate is inadequate or little adequate is about 1.3 times higher than in those municipalities where the attendance rate with water and sewage collection and treatment prevails.

As for the contribution margin of effects between the sanitation classes (Table 4), from the chi-square statistic, a high amplitude was observed between the groups Adequate *vs* Inadequate and Adequate *vs* Little Adequate (p <0.001), in comparison with the two classes of service Inadequate *vs* Little Adequate (p=0.05). There is, therefore, a huge and contributory difference for the response variable [P (death)], on the order of 11 to 15 times, between the inadequate and inadequate models of basic sanitation in relation to the adequate standard of sanitation in the state of Minas Gerais.

Among the planning regions of the state of Minas Gerais, only the Jequitinhonha-Mucuri region showed a positive and significant coefficient (p <0.001) for the evolution of the disease. The social vulnerability of the Jequitinhonha-Mucuri and Noroeste regions in the context of the COVID-19 pandemic in the state of Minas Gerais is evident in the results shown in Table 3, noting that the probability of death in these two regions are higher than the order of 2.0 times in relation to the other regions.

https://dx.doi.org/10.22161/ijaers.710.16 ISSN: 2349-6495(P) | 2456-1908(O)

Table 3: Odds ratio of continuous and categorical predictors (for Level A in relation to Level B) in the composition of the variable (Y') of evolution of COVID-19 (event: death) for the state of Minas Gerais, Brazil.

Continuous Predictors		Odds Ratio	CI 95%
'Age' – 'A'		1,0644	(1,0612; 1,0676)
'Race' – 'R'		3,6813	(3,4222; 3,9600)
'Comorbidity' – 'C'		1,2573	(1,1816; 1,3378)
Categorical Predictors		**Odds Ratio**	**CI 95%**
Level A	**Level B**		
Sex – 'S'			
Male	Feminino	1,3931	(1,2739; 1,5234)
'Class of Sanitation Service' – 'CSS'			
Inadequate	Adequate	1,3287	(1,1886; 1,4853)
Little Adequate	Adequate	1,3753	(1,1976; 1,5792)
Little Adequate	Inadequate	1,0350	(0,8925; 1,2003)
'Planning Regions ' – 'RP'			
Central	Alto Paranaiba	0,8872	(0,6874; 1,1451)
Centro-Oeste	Alto Paranaiba	0,6186	(0,4410; 0,8679)
Jequitinhonha-Mucuri	Alto Paranaiba	2,1822	(1,5248; 3,1231)
Noroeste	Alto Paranaiba	1,3213	(0,8668; 2,0140)
Norte	Alto Paranaiba	0,7615	(0,5407; 1,0724)
Rio Doce	Alto Paranaiba	0,7603	(0,5771; 1,0017)
Sul de Minas	Alto Paranaiba	0,7664	(0,5784; 1,0154)
Triângulo	Alto Paranaiba	0,8928	(0,6817; 1,1694)
Zona da Mata	Alto Paranaiba	0,7940	(0,5964; 1,0572)
Centro-Oeste	Central	0,6973	(0,5434; 0,8948)
Jequitinhonha-Mucuri	Central	2,4596	(1,8713; 3,2329)
Noroeste	Central	1,4892	(1,0478; 2,1166)
Norte	Central	0,8583	(0,6682; 1,1023)
Rio Doce	Central	0,8570	(0,7351; 0,9991)
Sul de Minas	Central	0,8638	(0,7317; 1,0198)
Triângulo	Central	1,0063	(0,8781; 1,1532)
Zona da Mata	Central	0,8949	(0,7508; 1,0667)
Jequitinhonha-Mucuri	Centro-Oeste	3,5275	(2,4749; 5,0276)
Noroeste	Centro-Oeste	2,1358	(1,4018; 3,2541)
Norte	Centro-Oeste	1,2309	(0,8772; 1,7271)
Rio Doce	Centro-Oeste	1,2291	(0,9406; 1,6060)
Sul de Minas	Centro-Oeste	1,2388	(0,9417; 1,6297)
Triângulo	Centro-Oeste	1,4432	(1,1053; 1,8845)
Zona da Mata	Centro-Oeste	1,2835	(0,9742; 1,6910)
Noroeste	Jequitinhonha-Mucuri	0,6055	(0,3932; 0,9325)
Norte	Jequitinhonha-Mucuri	0,3489	(0,2444; 0,4982)
Rio Doce	Jequitinhonha-Mucuri	0,3484	(0,2595; 0,4678)
Sul de Minas	Jequitinhonha-Mucuri	0,3512	(0,2598; 0,4748)
Triângulo	Jequitinhonha-Mucuri	0,4091	(0,3070; 0,5452)
Zona da Mata	Jequitinhonha-Mucuri	0,3639	(0,2678; 0,4943)
Norte	Noroeste	0,5763	(0,3792; 0,8758)
Rio Doce	Noroeste	0,5755	(0,3971; 0,8338)
Sul de Minas	Noroeste	0,5800	(0,3983; 0,8446)
Triângulo	Noroeste	0,6757	(0,4714; 0,9686)
Zona da Mata	Noroeste	0,6009	(0,4098; 0,8813)
Rio Doce	Norte	0,9985	(0,7586; 1,3144)
Sul de Minas	Norte	1,0065	(0,7586; 1,3352)
Triângulo	Norte	1,1725	(0,8988; 1,5295)
Zona da Mata	Norte	1,0427	(0,7816; 1,3912)
Sul de Minas	Rio Doce	1,0079	(0,8314; 1,2220)
Triângulo	Rio Doce	1,1742	(0,9799; 1,4072)
Zona da Mata	Rio Doce	1,0443	(0,8621; 1,2649)
Triângulo	Sul de Minas	1,1650	(0,9650; 1,4065)
Zona da Mata	Sul de Minas	1,0361	(0,8448; 1,2707)
Zona da Mata	Triângulo	0,8893	(0,7268; 1,0882)

In Table 4, the high and significant contrast of the chi-square statistic comparing the regions of the state depicts the social, environmental and economic discrepancy in Minas Gerais. The Jequitinhonha-Mucuri region, in contrast to all other regions of the state, contributes significantly to the response variable [P (death)] of the model. Contrasts with less inflated chi-square values were observed between the regions of Alto Paranaíba, Central, Centro-Oeste, Triângulo, Zona da Mata and Sul de Minas.

Table 4: Chi-Square parameters (χ^2) for continuous variables (Wald's criterion) and comparison of categorical variables in the composition of the variable (Y') of COVID-19 evolution for the state of Minas Gerais, Brazil.

Contrast	Chi-Square (χ^2)	Pr>χ^2
'Age' – 'A'	2476,274	< 0,0001 ***
'Race' – 'R'	69,949	< 0,0001 ***
'Comorbidity' – 'C'	1783,299	< 0,0001 ***
'Sex' – 'S'		
Male *vs* Female	94,497	< 0,0001 ***
'Class of Sanitation Service' – 'CSS'		
Adequate *vs* Inadequate	40,303	< 0,0001 ***
Adequate *vs* Little Adequate	50,854	< 0,0001 ***
Inadequate *vs* Little Adequate	3,664	0,056 **
'Planning Regions' – 'RP'		
Alto Paranaíba *vs* Central	0,377	0,539 ns
Alto Paranaíba *vs* Centro-Oeste	6,574	0,010 ***
Alto Paranaíba *vs* Jequitinhonha-Mucuri	34,539	< 0,0001 ***
Alto Paranaíba *vs* Noroeste	0,723	0,395 ns
Alto Paranaíba *vs* Norte	2,331	0,127 ns
Alto Paranaíba *vs* Rio Doce	1,755	0,185 ns
Alto Paranaíba *vs* Sul de Minas	4,197	0,040 **
Alto Paranaíba *vs* Triângulo	0,033	0,855 ns
Alto Paranaíba *vs* Zona da Mata	1,106	0,293 ns
Central *vs* Centro-Oeste	8,043	0,005 ***
Central *vs* Jequitinhonha-Mucuri	67,357	< 0,0001 ***
Central *vs* Noroeste	2,087	0,149 ns
Central *vs* Norte	2,126	0,145 ns
Central *vs* Rio Doce	1,799	0,180 ns
Central *vs* Sul de Minas	6,257	0,012 **
Central *vs* Triângulo	0,607	0,436 ns
Central *vs* Zona da Mata	0,661	0,416 ns
Centro-Oeste *vs* Jequitinhonha-Mucuri	69,727	< 0,0001 ***
Centro-Oeste *vs* Noroeste	8,256	0,004 ***
Centro-Oeste *vs* Norte	1,048	0,306 ns
Centro-Oeste *vs* Rio Doce	3,492	0,062 *
Centro-Oeste *vs* Sul de Minas	1,130	0,288 ns
Centro-Oeste *vs* Triângulo	9,342	0,002 ***
Centro-Oeste *vs* Zona da Mata	4,179	0,041 **
Jequitinhonha-Mucuri *vs* Noroeste	15,823	< 0,0001 ***
Jequitinhonha-Mucuri *vs* Norte	54,177	< 0,0001 ***
Jequitinhonha-Mucuri *vs* Rio Doce	69,362	< 0,0001 ***
Jequitinhonha-Mucuri *vs* Sul de Minas	77,951	< 0,0001 ***
Jequitinhonha-Mucuri *vs* Triângulo	55,663	< 0,0001 ***
Jequitinhonha-Mucuri *vs* Zona da Mata	60,906	< 0,0001 ***
Noroeste *vs* Norte	4,329	0,037 **
Noroeste *vs* Rio Doce	3,695	0,055 **
Noroeste *vs* Sul de Minas	5,977	0,014 ***
Noroeste *vs* Triângulo	1,257	0,262 ns
Noroeste *vs* Zona da Mata	2,886	0,089 *
Norte *vs* Rio Doce	0,321	0,571 ns
Norte *vs* Sul de Minas	0,037	0,848 ns
Norte *vs* Triângulo	3,154	0,076 *
Norte *vs* Zona da Mata	0,580	0,446 ns
Rio Doce *vs* Sul de Minas	1,169	0,280 ns
Rio Doce *vs* Triângulo	2,999	0,083 *
Rio Doce *vs* Zona da Mata	0,109	0,741 ns
Sul de Minas *vs* Triângulo	7,730	0,005 ***

* significant coefficient at 10%, ** at 5% and *** at 1%.

3.2. Decision Tree Model for COVID-19 evolution data

The classification tree (Fig. 2) for the binary dependent variable shows in the first node (p-value: 0.000), the classification of obituaries in the sense of the age groups between 60 and 69 years (31.1%; n = 1193); 70 and 79 years (47.4% of deaths; n = 1446), 80 and 89 years (61.2%; n = 1375) and 90 years or more (68.1%; n = 370). The classification of recovered patients is quite expressive in those nodes that comprise the age groups between 50 and 59 years (85.1%; n = 3793), 40 and 49 years (93.7%; n = 4676), 30 and 39 years (97.6%; n = 5344) and 20 to 29 years, 10 to 19 years, 1 to 9 years and less than 1 year (98.9%; n = 4947).

The comorbidity variable classifies almost all nodes in a second depth layer (p-value: 0.000), with the exception of the node aged 90 years or older. This in turn, extends to a terminal node (p-value: 0.001) for the gender variable, with a narrow propensity of deaths in men (76%; n = 171) in relation to women (62.6%; n = 199).

In all these age groups, the percentage difference is significant between deaths and recoveries, considering only the nodes with and without reports of comorbidity, ranging between 5.8%, between 70 and 79 years and 12.8%, between 40 and 49 years. Therefore, even in those age groups that responded better to recovery, the report of pre-existing disease was significant for the classification of deaths by COVID-19 in the dependent variable. The "Not Informed" classification ends at terminal nodes for comorbidity (Fig. 2).

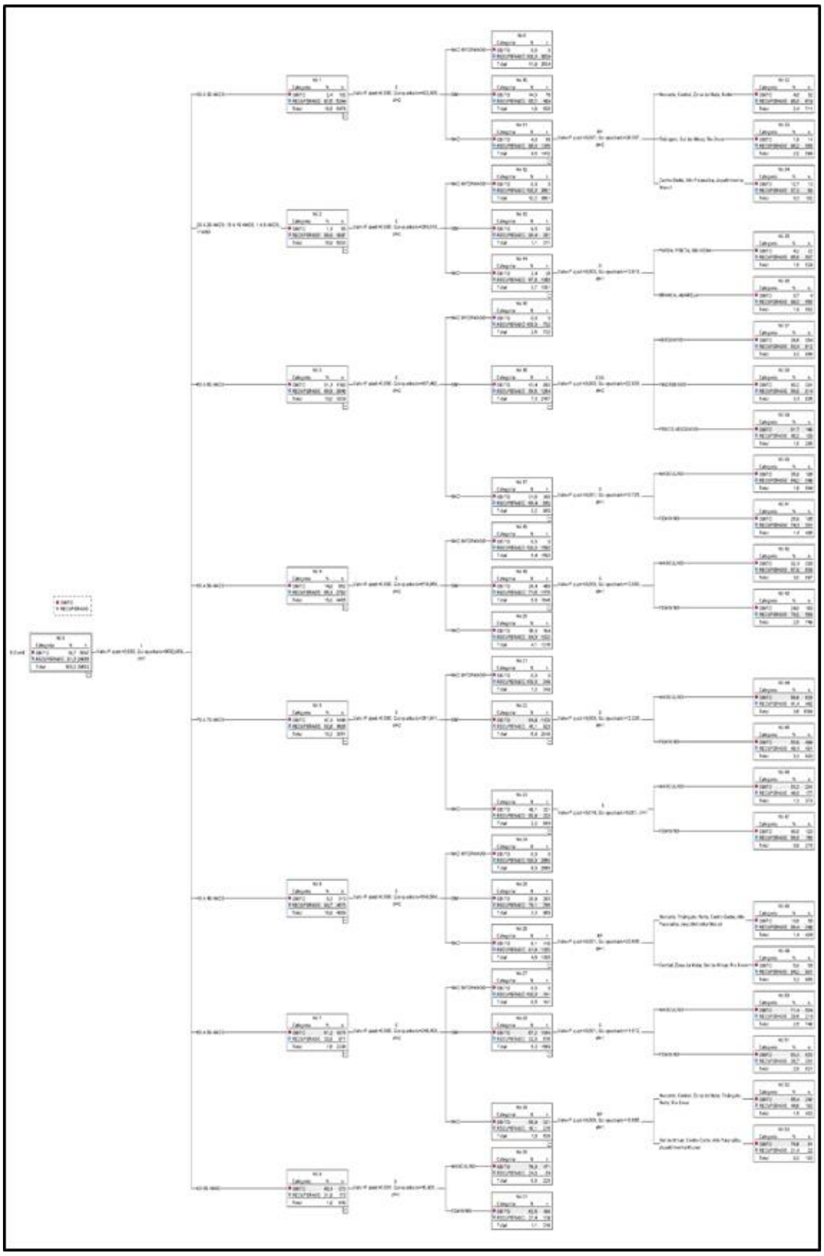

Fig. 2. Decision Tree Diagram (CHAID model) with 53 terminal nodes.

The third depth layer of the classification tree is quite diverse in the formation of the last nodes, the sex variable being the one that classifies most of these nodes. The proportional difference between men and women regarding recorded deaths is relevant in the five classified nodes, ranging from 7.7% (age range between 70 and 79 years, with comorbidity) and 9.9% (age range between 60 and 69 years, without comorbidity).

The ethno racial aspect (p-value: 0.003) is quite evident for the classification of deaths in the group of brown, black and indigenous individuals (4.2%; n = 22) in relation to white and yellow individuals (0.7%; n = 4), within the node of associated non-comorbidity and age groups of 20 to 29 years, 10 to 19 years, 1 to 9 years and less than 1 year (Fig. 2).

https://dx.doi.org/10.22161/ijaers.710.16 ISSN: 2349-6495(P) | 2456-1908(O

The class of basic sanitation service prevalent in the municipality of residence of the individual, classifies the age group between 60 and 69 years old, with some pre-existing disease, differentiating the appropriate, little adequate and inadequate services (p-value: 0.000). The death records in the municipalities with a prevalence of inadequate service (43.3%; n = 391) and little adequate (51.7%; n = 148) are proportionally higher than the obituaries in the municipalities classified as adequate (36.6 %; n = 354).

The variable that frames the individual's municipality of residence among the state's planning regions classifies three nodes, all of which have no associated comorbidity. In summary, the inclusion of the municipalities of Jequitinhonha-Mucuri and Noroeste in the terminal nodes greatly expands the expression of the percentage of deaths in relation to that of patients recovered by COVID-19, in the age groups between 30 and 39 years old (10.9%), 40 and 49 years old (7.8%) and 80 and 89 years old (23.2%). Similarities, were also observed in the composition of the planning regions for the nodes of the classification tree.

IV. DISCUSSION

The results of this research agree with the conclusions of recent studies on the evolution of COVID-19 in social groups in several Brazilian regions. The prevalence of death records in the population strata of adult and elderly individuals, between 60 and 69 years, 70 and 79 years, 80 and 89 years and 90 years or more, male and with some pre-existing disease, matches the profile of greater severity and susceptibility to COVID-19 [7, 30].

In this context, social distance and policies to contain the transmission of the disease must consider the age composition in the regional context, as well as intergenerational interactions. Protective public health measures against COVID-19 and with significant reach for individuals with chronic diseases such as hypertension, diabetes, obesity, cardiovascular and respiratory diseases, are becoming increasingly urgent in the state of Minas Gerais [11, 33, 37].

There is great concern in the elderly classes (aged between 60 and 69 years old) and with some comorbidity, living in regions where the basic sanitation service that is inadequate (Index of Service with Collection and without Treatment) and inadequate (Index of Service without Collection) prevails and without Treatment). The probability of occurrence of deaths by COVID-19 in these groups, to the detriment of recovery from the disease, exposes the critical and unequal socio-environmental context of some mining regions, greatly increasing the risk and prevalence of the disease in the state.

In this sense, the results of this work reinforce the discrepancy in the quality and reach of basic sanitation services and programs of prophylactic measures by the government, as a risk factor for the disease and its spread in the poorest regions of the state and among the population. elderly, resulting from the phenomenon of internalization of the disease [2, 12]. Such social, environmental and demographic factors are challenging in the country, including the lack of knowledge about the transmission characteristics of COVID-19, in the context of historical social and racial vulnerability. And, precisely in the regions of Jequitinhonha-Mucuri and Noroeste, which comprise the poorest municipalities in the state of Minas Gerais, there was a higher odds ratio for the registration of deaths by COVID-19 [3, 13].

Younger and classified individuals without any correlated comorbidity, in general, do not correspond to the highest probability of the event 'death' due to the disease. However, there is an aggravation of deaths due to COVID-19 among the brown, black and indigenous populations classified precisely among the youngest individuals, between the age groups of 20 to 29 years old, 10 to 19 years old, 1 to 9 years old and less than 1-year-old, with no record of comorbidity.

This is also a relevant aspect in this research, which reinforces observations that the pandemic of COVID-19 exposes unsuccessful social and political fractures in the peripheries of urban centers and fragmented in small and medium-sized municipalities, with racialized and discriminatory responses, exposing the fragility of young blacks and browns facing the disease [9, 34].

In this regard, the results are in line with some research that claims about racial disparities associated with health result from a historical process of racial segregation of households and occupation of urban areas, such as inequalities in employment opportunities, income and access to health care. Education [13, 41]. The structural and institutional racism present in the country, and so well explained before the numbers and records of the COVID-19 pandemic, reinforces the distance between the Brazilian State, the unequal condition of access of the black population to basic sanitation and public health services, and finally, the significant decrease in the life expectancy of these young people, in view of the glaring conditions of vulnerability of these communities [22].

And in this sense, the character of ethno racial vulnerability takes on alarming proportions, considering that public policies for the containment of the COVID-19 pandemic do not reach historically marginalized social groups, recognizing the limited access of these communities to health care and decent conditions. work and employment [9, 15]. And even when they are able to access health service spaces, some authors question the quality of the treatment provided as a result of racial discrimination in the health context [23, 29, 38].

It is also necessary to pay attention to the circulation aspect of the youngest who make up part of the economically active population. In general, they are users of public transport and remain working outside home-office standards or are looking for a replacement in the job market. This situation imposes a practically constant condition for the circulation of the virus and the spread of the disease among these social groups and among the individuals closest to the family and the community.

In summary, the research offers important subsidies for the composition of risk factors in the pandemic scenario of COVID-19 in view of the social, environmental and demographic strata of the spread of the disease in the state of Minas Gerais. And therefore, the actions of the public authorities need to ensure that the fragile population has safe conditions for prophylaxis, access to health services and a guarantee of minimum income.

V. CONCLUSION

The probability of occurrence of deaths due to COVID-19 in the state of Minas Gerais, in the data coverage period of this research, is significantly higher in males, aged over 60 years, with some comorbidity, declared black and brown, residents in municipalities located in the poorest macro-regions of the state, where inadequate or inadequate classes of basic sanitation prevail.

The classifications in the decision tree that admit the racial profile of deaths among the younger groups without associated comorbidity, and among the elderly with comorbidity, not assisted by an adequate basic sanitation network, reiterate the historical social asymmetries experienced by marginalized populations and alert to the urgency of protective actions by the state and municipal public authorities, far beyond the COVID-19 pandemic scenario.

The study and monitoring of the evolution of COVID-19, which is still being disseminated in the country, should cover more specific data and records on the progress of the disease in social, environmental, demographic and economic strata and groups. Based on regression models and learning machine techniques, it is possible to provide the necessary depth for reflection and the formulation of inclusive policies in the face of the worsening of the environmental crisis in the country characterized by the pandemic of COVID-19.

REFERENCES

[1] Aldrich, J. H.; Nelson, F. D. Linear probability, logit, and probit models. London: Sage, 1984. 94p.
[2] Amaral, L. S. do *et al*. Interiorização do Covid-19: uma análise da evolução dos casos/10 mil habitantes em municípios da microrregião de Garanhuns no estado de Pernambuco, através de modelos de regressão não linear. Research, Society and Development, [s.l.], v. 9, n. 9, p. 1-19, 30 ago. 2020.
[3] Barreto, M. L. *et al*. O que é urgente e necessário para subsidiar as políticas de enfrentamento da pandemia de COVID-19 no Brasil? Revista Brasileira de Epidemiologia, [s.l.], v. 23, p. 1-4, 2020.
[4] Brasil. Ministério da Saúde. Agência Saúde. Diabetes, hipertensão e obesidade avançam entre os brasileiros. 2020. Disponível em: https://www.saude.gov.br/noticias/agencia-saude/46982-diabetes-hipertensao-e-obesidade-avancam-entre-os-brasileiros-3. Acesso em: 09 de jun. 2020.
[5] Brasil. Constituição (2006). Portaria nº 2528, de 19 de outubro de 2006. Aprova a Política Nacional de Saúde da Pessoa Idosa. Brasília, DF, 19 out. 2006. Disponível em: https://bvsms.saude.gov.br/bvs/saudelegis/gm/2006/prt2528_19_10_2006.html. Acesso em: 30 set. 2020.
[6] Brasil. Secretaria de Gestão Estratégica e Participativa. Ministério da Saúde. Política Nacional de Saúde Integral da População Negra: uma política para o SUS. 2013. Departamento de Apoio à Gestão Participativa. Editora do Ministério da Saúde. 36p. Disponível em: http://bvsms.saude.gov.br/bvs/publicacoes/politica_nacional_saude_integral_populacao.pdf. Acesso em: 30 set. 2020.
[7] Brito, V. P. de *et al*. Associação da Diabetes Mellitus com a gravidade da COVID-19 e seus potenciais fatores mediadores: uma revisão sistemática. Revista Thema, [s.l.], v. 18, p. 204-217, 25 ago. 2020.
[8] Chen, N. et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. The Lancet, [s.l.], v. 395, n. 10223, p. 507-513, fev. 2020.
[9] Devakumar, D. et al. Racism and discrimination in COVID-19 responses. The Lancet, [s.l.], v. 395, n. 10231, p. 1194, abr. 2020.
[10] Dias, F. L. T. et al. Doenças respiratórias no Triângulo Mineiro: análise epidemiológica e projetiva com a pandemia de Covid-19. Journal of Health & Biological Sciences, [s.l.], v. 8, n. 1, p. 1-6, 24 abr. 2020.
[11] Dowd, J. B. *et al*. Demographic science aids in understanding the spread and fatality rates of COVID-19. Proceedings of The National Academy of Sciences, [s.l.], v. 117, n. 18, p. 9696-9698, 16 abr. 2020.
[12] Faria, R. *et al*. Difusão Espacial e Interiorização da Covid-19 no Estado do Rio Grande do Sul. Revista da Casa da

Geografia de Sobral (RCGS), v. 22, n. 2, p. 26-43, 5 set. 2020.

[13] Gaia, R da S P. Subcidadania, raça e isolamento social nas periferias brasileiras: reflexões em tempos de COVID-19. Revista Thema, v.18, número especial. p. 92-110. 2020.

[14] Goes, E F. *et al*. Vulnerabilidade racial e barreiras individuais de mulheres em busca do primeiro atendimento pós-aborto. Cadernos de Saúde Pública, [s.l.], v. 36, n. 1, p. 1-13, 2020.

[15] Goes, E. F.; Ramos, D. de O.; Ferreira, A. J. F. Desigualdades raciais em saúde e a pandemia da Covid-19. Trabalho, Educação e Saúde, [s.l.], v. 18, n. 3, p. 1-7, 2020.

[16] Guan, W-J. *et al*. Clinical Characteristics of Coronavirus Disease 2019 in China. New England Journal of Medicine, [s.l.], v. 382, n. 18, p. 1708-1720, 30 abr. 2020.

[17] Instituto Brasileiro de Geografia e Estatística (IBGE). Coordenação de População e Indicadores Sociais. Síntese de indicadores sociais: uma análise das condições de vida da população brasileira: 2018. Rio de Janeiro: IBGE, 2018. 151p. (39).

[18] IBM Corp. IBM SPSS® Statistics for Windows, Versão 25.0 IBM. 2012. Software. Armonk, NY: IBM Corp.

[19] Instituto Trata Brasil. Painel Saneamento Brasil, Ranking do Saneamento, 2019. Disponível em: http://www.tratabrasil.org.br/. Acesso em: 03 de jul. de 2020.

[20] Kass, G. V. An Exploratory Technique for Investigating Large Quantiles of Categorical Date. University of the Witwatersand. Appl. Statist. 29, n.2, p. 119-127, 1980.

[21] Lowcock, E. C. *et al*. The Social Determinants of Health and Pandemic H1N1 2009 Influenza Severity. American Journal of Public Health, [S.L.], v. 102, n. 8, p. 51-58, ago. 2012.

[22] Macedo, Y. Miguel; Ornellas, J. L.; Bomfim, H. F. COVID – 19 no BRASIL: o que se espera para população subalternizada? Revista Encantar - Educação, Cultura e Sociedade. Bom Jesus da Lapa, v. 2, p. 01-10, jan./dez. 2020.

[23] Macinko, J. *et al*. Who experiences discrimination in Brazil? Evidence from a large metropolitan region. International Journal for Equity in Health, [s.l.], v. 11, n. 1, p. 80-91, 2012.

[24] Magidson, J. SPSS® for Windows™ CHAID Release 6.0. SPSS Inc. Chicago, 1993.

[25] Melo, R. R. P. B. de *et al*. Desafios no acesso à água e saneamento básico no Brasil e o controle da COVID-19. Revista Augustus, Rio de Janeiro, v.25, n. 51, p. 281-293, 2020.

[26] Minas Gerais. Fundação João Pinheiro. Secretaria Estadual de Planejamento e Gestão do Estado de Minas Gerais (SEPLAG). Desenvolvimento Humano: Evolução do Índice de Desenvolvimento Humano Municipal (IDHM). Evolução do Índice de Desenvolvimento Humano Municipal (IDHM). 2015. FJP. Disponível em: http://minasedados.fjp.mg.gov.br/#dados-dh. Acesso em: 29 set. 2020.

[27] Minas Gerais. Fundação João Pinheiro. Secretaria Estadual de Planejamento e Gestão do Estado de Minas Gerais (SEPLAG). Regiões de Planejamento do Estado de Minas Gerais. 1995. FJP. Disponível em: https://www.mg.gov.br/conteudo/conheca-minas/geografia/regioes-de-planejamento. Acesso em: 03 jun. 2020.

[28] Minitab Inc. Statistical Software Data Analysis Software. Version 19, 2019.

[29] Miranda, R. de S. *et al*. Racismo no contexto da saúde: um estudo psicossociológico. 2015. 194 f. Tese (Doutorado) - Curso de Psicologia Social, Psicologia, Universidade Federal da Paraíba, Paraíba, 2015. Disponível em: https://repositorio.ufpb.br/jspui/handle/tede/7688. Acesso em: 30 set. 2020.

[30] Niquini, R. P. *et al*. SRAG por COVID-19 no Brasil: descrição e comparação de características demográficas e comorbidades com SRAG por influenza e com a população geral. Cadernos de Saúde Pública, [s.l.], v. 36, n. 7, p. 1-12, 2020.

[31] Oliveira, R. G. de *et al*. Desigualdades raciais e a morte como horizonte: considerações sobre a covid-19 e o racismo estrutural. Cadernos de Saúde Pública, [s.l.], v. 36, n. 9, p. 1-10, 2020.

[32] Pires, L. N. *et al*. Pandemic of Inequality. Public Policy Brief. Levy Economics Institute of Bard College. n. 149, 2020. 16p.

[33] Rezer, F.; Faustino, W. R.; Maia, C. S. Incidence of COVID-19 in the mesoregions of the state of Mato Grosso: confirmed and notified cases. Rev Pre Infec e Saúde [Internet]. v.6. n.10317. 2020.

[34] Santos, M. P. A. dos *et al*. População negra e Covid-19: reflexões sobre racismo e saúde. Estudos Avançados, [s.l.], v. 34, n. 99, p. 225-244, ago. 2020.

[35] Silva, A. da *et al*. Iniquidades raciais e envelhecimento: análise da coorte 2010 do Estudo Saúde, Bem-Estar e Envelhecimento (SABE). Revista Brasileira de Epidemiologia, v.21, n. suppl 2, 4 fev. 2019.

[36] Simões, A. *et al*. Panorama nacional e internacional da produção de indicadores sociais: grupos populacionais específicos e uso do tempo. 6. ed. Rio de Janeiro: IBGE, Coordenação de População e Indicadores Sociais, 2018. 352p.

[37] Souza, L. G.; Randow, R.; Siviero, P. C. L. Reflexões em tempos de COVID-19: diferenciais por sexo e idade. Com. Ciências Saúde, v.31, suppl 1. p. 75-83. 2020.

[38] Tavares, N. O. *et al*. A percepção dos psicólogos sobre o racismo institucional na saúde pública. Saúde em Debate, [s.l.], v. 37, n. 99, p. 580-587, dez. 2013.

[39] Ventura, D. de F. L. *et al*. Desafios da pandemia de COVID-19: por uma agenda brasileira de pesquisa em saúde global e sustentabilidade. Cadernos de Saúde Pública, [s.l.], v. 36, n. 4, p. 1-5, 2020.

[40] Wang, W. *et al*. Detection of SARS-CoV-2 in Different Types of Clinical Specimens. Jama, [s.l.], p. 1-2, 11 mar. 2020.

[41] Williams, D.; Collins, C. Racial residential segregation: a fundamental cause of racial disparities in health. Public Health Reports, v. 116, n. 5, p. 404-416, 2001.
[42] World Health Organization (WHO). Coronavirus disease (Covid-19) advice for the public. 2020. Disponível em: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public. Acesso em: 17 mai. 2020.
[43] Xiao, H. et al. The Effects of Social Support on Sleep Quality of Medical Staff Treating Patients with Coronavirus Disease 2019 (COVID-19) in January and February 2020 in China. Med Sci Monit, n.26: e923549-1–e923549-8. 2020.
[44] Xu, B. et al. Epidemiological data from the COVID-19 outbreak, real-time case information. Scientific Data. v. 7, n. 106. 2020.
[45] Yeo, C.; Kaushal, S.; Yeo, D. Enteric involvement of coronaviruses: is fecal oral transmission of SARS-COV-2 possible? The Lancet Gastroenterology & Hepatology, [s.l.], v. 5, n. 4, p. 335-337, abr. 2020.
[46] Zhao, S. et al. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: a data-driven analysis in the early phase of the outbreak. International Journal of Infectious Diseases, [s.l.], v. 92, p. 214-217, mar. 2020.