



## UTILIZAÇÃO DA ANÁLISE MULTIVARIADA PARA DADOS DE HEPATITES VIRAIS B E C: PARTE 1

### USE OF MULTIVARIATE ANALYSIS FOR DATA OF VIRAL HEPATITIS B AND C: PART 1

### UTILIZACIÓN DEL ANÁLISIS MULTIVARIADO PARA DATOS DE HEPATITIS VIRALES B Y C: PARTE 1

Sidnei Anastácio Sampaio<sup>1</sup>, Kuang Hongyu<sup>2</sup>, José Martins Pinto Neto<sup>3</sup>, Vando Edésio Soares<sup>4</sup>, Ricardo Scarparo Navarro<sup>5</sup>

#### RESUMO

**Objetivo:** analisar banco de dados de hepatites B e C, por meio da análise de componentes principais. **Método:** estudo descritivo e retrospectivo, de abordagem quantitativa. Os dados foram coletados no banco de dados do SINAN. **Resultados:** no banco de dados de hepatite B, o componente principal 1 (CP 1) explicou maior variabilidade dos dados (26,00%) entre todos os CPs. Existe um contraste entre as variáveis “idade” (coeficiente - 0,76738590) e “escolaridade do paciente” (coeficiente + 0,77306034). Para o banco de dados do VHC, o CP 1 (24,64%) obteve os maiores coeficientes. Existe um contraste entre as variáveis “escolaridade” (coeficiente - 0,68840788) e “bairro do paciente” (coeficiente + 0,64556072). **Conclusão:** este estudo conclui que os bancos de dados avaliados para as hepatites B e C mostraram maior frequência na população adulta, com relativa baixa e média escolaridade e predomínio do gênero feminino para a hepatite C. **Descritores:** Hepatite; Análise Multivariada; Análise de Componente Principal.

#### ABSTRACT

**Objective:** to analyze the hepatitis B and C database, through main component analysis. **Method:** descriptive and retrospective, quantitative approach. The data were collected in the SINAN database. **Results:** in the hepatitis B database, the main component 1 (MC 1) explained greater variability of the data (26.00%) among all MCs. There is a contrast between variables, “age” (coefficient - 0.76738590) and “schooling of the patient” (coefficient + 0.77306034). For the HCV database, the MC 1 (24.64%) had the highest coefficients. There is a contrast between the variables, “schooling” (coefficient - 0.68840788) and “patient’s neighborhood” (coefficient + 0.64556072). **Conclusion:** this study concludes that the databases evaluated for hepatitis B and C showed higher frequency in the adult population, with relatively low and medium schooling and predominance of the female genus for hepatitis C. **Descriptors:** Hepatitis; Multivariate Analysis; Principal Component Analysis.

#### RESUMEN

**Objetivo:** analizar la base de datos de hepatitis B y C, mediante el análisis de componentes principales. **Método:** estudio descriptivo y retrospectivo, de abordaje cuantitativo. Los datos se recopilaron en la base de datos del SINAN. **Resultados:** en la base de datos de hepatitis B, el componente principal 1 (CP 1) explicó mayor variabilidad de los datos (26,00%) entre todos los CPs. Existe un contraste entre las variables, “la edad” (coeficiente - 0,76738590) y “la escolaridad del paciente” (coeficiente + 0,77306034). Para el banco de datos del VHC, el CP 1 (24,64%) obtuvo los mayores coeficientes. Existe un contraste entre las variables, “escolaridad” (coeficiente - 0,68840788) y “barrio del paciente” (coeficiente + 0,64556072). **Conclusión:** este estudio concluye que los bancos de datos evaluados para las hepatitis B y C mostraron una mayor frecuencia en la población adulta, con relativa baja y media escolaridad y predominio del género femenino para la hepatitis C. **Descritores:** Hepatitis; Análisis Multivariante; Análisis de Componente Principal.

<sup>1</sup>Enfermeiro, Professor Mestre, Graduação em Enfermagem/Enf/ICEN/CUR, Universidade Federal do Mato Grosso/UFMT. Rondonópolis (MT), Brasil. E-mail: [sidnei.sampaio1@gmail.com](mailto:sidnei.sampaio1@gmail.com); <sup>2</sup>Matemático, Professor Doutor, Graduação em Estatística/Est/ICET, Universidade Federal do Mato Grosso/UFMT. Cuiabá (MT), Brasil. E-mail: [prof.kuang@gmail.com](mailto:prof.kuang@gmail.com); <sup>3</sup>Enfermeiro, Professor Doutor, Graduação Enfermagem/Medicina, Universidade Camilo Castelo Branco/UNICASTELO. Fernandópolis (SP), Brasil. E-mail: [jmpintoneto@gmail.com](mailto:jmpintoneto@gmail.com); <sup>4</sup>Zootecnista, Professor Doutor, Graduação em Zootecnia/Programa de Pós-Graduação em Engenharia Biomédica/Instituto de Pesquisa e Extensão- Universidade Camilo Castelo Branco/UNICASTELO. São Paulo (SP), Brasil.: [soaresvando@gmail.com](mailto:soaresvando@gmail.com); <sup>5</sup>Cirurgião-dentista, Professor Doutor, Graduação em Odontologia/Programa de Pós-Graduação em Engenharia Biomédica/Instituto de Pesquisa e Extensão- Universidade Camilo Castelo Branco/UNICASTELO. São Paulo (SP), Brasil. E-mail: [ricardosnavarro@gmail.com](mailto:ricardosnavarro@gmail.com)

## INTRODUÇÃO

Estima-se que 240 milhões de pessoas estejam cronicamente infectadas em todo o mundo pelo vírus da hepatite B (VHB).<sup>1</sup> Cerca de 800 mil pessoas morrem a cada ano a partir desta doença e menos de 650 mil, de cirrose e câncer de fígado, devido à infecção crônica da hepatite B (HB), e outras 130 mil, de hepatite B aguda, principalmente, por complicações de cirrose hepática ou por câncer primário do fígado (carcinoma hepatocelular, CHC).<sup>2</sup> Para efeitos de comparação, o vírus da hepatite B é de 50 a 100 vezes mais infeccioso que o Vírus da Imunodeficiência Humana (HIV).<sup>3</sup>

Em comparação à infecção pelo Vírus da Hepatite C (VHC), existem aproximadamente de 130 a 150 milhões de pessoas portadoras do vírus. Um número significativo de pessoas que sofrem de infecção crônica desenvolverá cirrose hepática ou câncer de fígado CHC. Estima-se que o número de pessoas que morrem anualmente de doenças hepáticas relacionadas ao vírus da hepatite C a cada ano esteja entre 350 a 500 mil.<sup>4</sup>

No Brasil, segundo dados sujeitos à revisão do Sistema de Informação de Agravos de Notificação (SINAN), em 2014, foram notificados 13.187 casos confirmados de hepatite B, seguidos de 51 casos de associação entre hepatite B e A (B+A), 333 para B+C, 226 de hepatite B+D. Nesse mesmo ano, para a hepatite C, houve 12.934 casos confirmados, seguidos de 26 casos de hepatite C+A. O montante geral de casos notificados no Brasil foi de 34.637 casos confirmados em todo o período de 2014, para todas as classificações etiológicas das hepatites virais.<sup>5</sup>

Nos residentes do Estado de São Paulo, em 2014, foi notificado um total de 2.229 casos confirmados de hepatite B, seguidos de associação das hepatites virais B+C, com 72 casos; B+A, com quatro casos. Em relação à hepatite C, foram notificados 3.192 casos confirmados, seguidos de quatro casos de C+A.<sup>5</sup>

A Análise de Componentes Principais (ACP) é uma técnica multivariada de modelagem da estrutura de covariância. A técnica foi inicialmente descrita por Pearson, em 1901.<sup>6-7</sup> Uma descrição de métodos computacionais práticos veio muito mais tarde, com Hotelling, em 1933 e 1936,<sup>8-9</sup> quem a usou com o propósito determinado de analisar as estruturas de correlação.<sup>7</sup>

A análise dos componentes principais é uma das mais importantes ferramentas estatísticas da análise multivariada, por constituir a base onde se fundamenta a maioria dos outros métodos multivariados de análise de dados,

sendo uma ferramenta de análise exploratória de dados que permite revelar a existência ou não de amostras diferentes, de relações entre as variáveis medidas e de relações ou agrupamentos entre amostras, visando à sua redução, mantendo apenas um número pequeno de Componentes Principais (CPs) que já respondem por uma parcela significativa da informação total contida na estrutura de dados.<sup>10</sup>

Como doenças transmitidas por sangue e fluidos corpóreos, as infecções pelos VHB ou VHC podem ser adquiridas em estabelecimentos onde existe exposição a materiais perfurocortantes, como agulhas, tesouras e lâminas. Este estudo foi realizado no município de Fernandópolis, Estado de São Paulo, onde existem vários estabelecimentos de saúde e de outras atividades que propiciam aumento do risco de infecções por hepatites virais B e C por ferimentos com inoculação de material contaminado, como nos casos dos que fazem tatuagem, manicures e pedicures, cabeleireiros, centros de estética, consultórios odontológicos, hospitais, unidades de saúde, farmácias, casas de profissionais do sexo, entre outros.

A doença parece ser bem conhecida pela população e pelos profissionais de saúde, mas é possível que muitos não sigam as orientações sobre o uso do material individual ou descartável, o que leva à vulnerabilidade à infecção/doença.

Portanto, com base no exposto, torna-se evidente a gravidade e problemática das hepatites virais B e C, o que justifica a realização deste estudo. Pela importância de ambas as infecções para a saúde humana, os estudos de prevalência ou incidência das mesmas se reveste de relevância, assim como o esforço para otimizar a análise dos bancos de dados formados por esta pesquisa. Este estudo se fundamenta na análise das variáveis de um banco de dados sobre hepatites B e C em Fernandópolis (SP), com o intuito de identificar as principais variáveis que emanaram da análise.

## OBJETIVOS

- Identificar quais variáveis são as mais importantes no banco de dados de hepatites viral B, de hepatite viral C e da junção dos bancos de dados das hepatites B e C.
- Reduzir o número de variáveis dos bancos de dados de hepatite viral B, de hepatite viral C e da junção dos bancos de dados das hepatites virais B e C.

## MÉTODO

O banco de dados deste estudo foi extraído do Programa de Pós-Graduação em Engenharia Biomédica/UNICASTELO, da dissertação de mestrado intitulada “Incidência de Hepatites Virais B e C em Fernandópolis SP”.

Estudo descritivo e retrospectivo, de abordagem quantitativa. Foram utilizados dados obtidos a partir de uma fonte secundária do Sistema de Informação de Agravos de Notificação (SINAN) disponível no Serviço de Vigilância Epidemiológica da Secretaria Municipal de Saúde do município de Fernandópolis (SP), referente aos casos notificados e confirmados de hepatites virais B e C, no período de 2007-2011, de pacientes residentes naquele município.

O Projeto foi aprovado pelo Comitê de Ética e Pesquisa da Universidade Camilo Castelo Branco (CEP UNICASTELO CAEE nº 01475612.7.0000.5494). Os dados foram analisados pelo método de análise multivariada, por meio de análise de componentes principais, ao nível de significância de 5%, pelo programa estatístico plataforma R.

O banco de dados bruto continha 46 variáveis, com 158 observações de casos confirmados de hepatites B e C, sendo 51 casos de infecção pelo VHB; 101, de infecção pelo VHC e oito casos de associação entre os dois agentes VHB e VHC. Foram retiradas, do banco de dados bruto, todas as variáveis cuja informação era de resposta “Não” e quando a “maioria dos dados como Não, Ignorado e em Branco”, sendo elas (Gestante; Zona de residência; Ocupação; Suspeita de hepatite; Tomou vacina; Agravos associados; Contato com paciente portador de VHB e VHC; O paciente foi Submetido ou Exposto a medicamento injetável, drogas inaláveis ou *crack*, drogas injetáveis, água/alimento contaminado, três ou mais parceiro sexuais, transplante, tatuagem/*piercing*, acupuntura, hemodiálise, acidente com material biológico, transfusão de sangue/derivados; Resultado da sorologia banco de sangue; Genótipo VHC; Classificação Final; Forma Clínica; Classificação Etiológica e Provável fonte de infecção). Foram retiradas também variáveis com resposta única como (Município de Residência, Regional de Saúde; Unidade de saúde; País; Unidade Federativa) e variáveis não pertinentes, ou seja, sem sentido para fazer análise, por exemplo, datas (dia, mês). Foram excluídas também todas as frequências (observações) em branco das variáveis que permaneceram no estudo para análise. O total de variáveis utilizadas no trabalho foi de oito

para cada banco de dados de HB, HC e junção de banco de dados de HB/HC. Foram 20 casos confirmados de HB, 30 casos de HC e 50 casos na junção de bancos de dados de HB e HC.

O principal local de coleta de dados da referida dissertação de mestrado, que serviu para a análise desta pesquisa, foi o município de Fernandópolis (SP).

### ◆ Análise Multivariada

A técnica de análise multivariada refere-se a todas as técnicas estatísticas que simultaneamente analisam múltiplas medidas sobre indivíduos ou objetos sob investigação. Assim, qualquer análise simultânea, de mais do que duas variáveis, pode ser considerada, a princípio, como multivariada, possibilitando que órgãos sanitários e organizações adquiram conhecimento e melhorarem, assim, suas tomadas de decisões.<sup>11</sup>

### ◆ Análise de Componentes Principais

A análise de componentes principais tem como objetivo reduzir a dimensionalidade do conjunto de variáveis e facilitar a interpretação da independência entre elas.<sup>7,11</sup> Para isso, são obtidas combinações lineares das variáveis originais que, geometricamente, representam a seleção de novos sistemas de coordenadas obtidos pela rotação do sistema original que tem  $p$  variáveis aleatórias como eixos das coordenadas. Esses novos eixos ortogonais (novas variáveis) são chamados de componentes principais e os valores das novas variáveis são chamados de escores dos componentes principais ou coordenadas principais.<sup>7</sup>

Em geral, escolhe-se o componente principal de maior importância (o primeiro componente principal), como sendo aquele de maior variância, que explique o máximo de variabilidade dos dados; o segundo componente de maior importância, o que apresenta a segunda maior variância, e assim sucessivamente, até o componente principal de menor importância.<sup>7,12-3</sup> Por outro lado, os últimos componentes principais serão responsáveis por direções que não estão associadas a muita variabilidade. Em outras palavras, esses últimos componentes principais identificarão relações lineares entre as variáveis originais próximas de constante.<sup>7,14-6</sup>

A obtenção dos componentes principais é realizada por meio da diagonalização de matrizes simétricas positivas semidefinidas. Então, podem-se calcular os componentes principais facilmente e utilizá-los em diferentes aplicações nas mais variadas áreas científicas. Esta facilidade é em função da existência de inúmeros programas capazes de

realizar cálculos matriciais para diagonalizar uma matriz simétrica positiva semidefinida.<sup>7</sup>

Muitos pesquisadores têm utilizado a análise de componentes principais para resolver problemas como da multicolinearidade em regressão linear, para estimar fatores, que representam outra técnica multivariada de modelagem da matriz de covariâncias, para realizar a modelagem da interação entre fatores em experimentos sem repetição, estudos de divergência e agrupamento entre genótipos em estudo de genética e melhoramento de plantas e animais, entre outras possibilidades.<sup>7,14</sup>

Para obter os componentes principais de uma forma geral, seja um conjunto de  $p$  variáveis  $X_1, X_2, \dots, X_p$ , com médias  $\mu_1, \mu_2, \dots, \mu_p$  e variância  $\sigma^2_1, \sigma^2_2, \dots, \sigma^2_p$ , respectivamente, estas variáveis não são independentes e, portanto, possuem covariância entre a  $i$ -ésima e  $k$ -ésima variável, definida por  $\sigma_{ik}$ , para  $i \neq k = 1, 2, \dots, p$ . Então, as  $p$  variáveis podem ser expressas na forma vetorial por:  $X = [X_1, X_2, \dots, X_p]^T$ , com vetor de médias  $\mu = [\mu_1, \mu_2, \dots, \mu_p]^T$  e matriz de covariância  $\Sigma$ . Encontram-se os pares de autovalores e autovetores  $(\lambda_1, e_1), (\lambda_2, e_2), \dots, (\lambda_p, e_p)$ , em que  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ , associados a  $\Sigma$  e então o  $i$ -ésimo componente principal é definido por:<sup>7,14</sup>

$$Z_i = e_i^T X = e_{i1}X_1 + e_{i2}X_2 + \dots + e_{ip}X_p$$

(1) em que  $i = 1, 2, \dots, p$ .

## RESULTADOS

Para o banco de dados de hepatite B, que possui 20 observações de casos e oito variáveis, o componente principal 1 (CP 1), ou seja, autovetor 1, obteve a maior porcentagem (26,00%) (Figura 1.2) entre todos os CPs. Ocorreu um contraste entre as variáveis. Uma aflora com sinal positivo para o coeficiente (autovetores) e a outra com sinal negativo para outro coeficiente. De modo que isto significa que, ao levar em consideração o sinal para a interpretação do resultado em relação às variáveis, elas possuem valores contrários em comparação uma com a outra. Têm a maior variação dos dados as variáveis idade do paciente (coeficiente - 0,76738590) e escolaridade do paciente (coeficiente + 0,77306034), (Figura 1). A Figura 1.2 destacou o indivíduo nove do banco de dados, sendo que no CP1 foi o que obteve a maior variação dos dados, com as variáveis idade e escolaridade. Essas variáveis, em suas categorias, mostraram que o indivíduo nove tem 22 anos e ensino médio completo (antigo colegial ou 2º grau), seguido do indivíduo três, que tem idade de 28 anos e escolaridade superior completa. Em seguida, o indivíduo

dez, que tem idade de 25 anos e escolaridade ensino médio incompleto (antigo colegial ou 2º grau), o que corresponde com o banco de dados observado. Todas essas variáveis do CP1 foram as que mais explicaram a variação dos dados de hepatite B, para este banco de dados.

Ao se generalizar o resultado do CP1 para os indivíduos observados, pode-se dizer que os indivíduos que aparecem na Figura 1.2 (gráfico), DOI no eixo X, a partir dos valores zero a três positivos, ou seja, quadrantes positivos (parte direita superior e inferior) do eixo X, representando o CP1, são os indivíduos que possuem menos idade e nível alto de escolaridade. Já na mesma Figura 1.2, pode-se dizer que os indivíduos que aparecem no eixo X próximo dos valores menos dois e próximo de zero, negativos, ou seja, quadrantes negativos (parte esquerda superior e inferior) do eixo X, representando o CP1, são os indivíduos que possuem mais idade e nível baixo de escolaridade, o que corresponde com o banco de dados observado. Esta mesma interpretação serve para as demais variáveis dos outros bancos de dados, quando acontece contraste.

O CP 2 do banco de dados de hepatite B explicou 19,87% (Figura 1.2) da variação dos dados. Existe um contraste entre as variáveis, sendo as variáveis mais importantes deste CP2 “exposição a tratamento cirúrgico” (coeficiente + 0,77329024) e “exposição a tratamento dentário” (coeficiente - 0,49345879) (Figura 1). Este contraste é observado pelos sinais contrários dos coeficientes: positivo para uma variável e negativo para outra variável. A figura 1.2 destacou o indivíduo 11 do banco de dados, sendo que, no CP2, foi o que obteve maior coeficiente, com as variáveis “exposição a tratamento cirúrgico” (sim, há mais de seis meses) e “exposição a tratamento dentário” (sim, há menos de seis meses). O total de explicação da variabilidade dos dados entre CP1 e CP2 é de aproximadamente 46% para o banco de dados de hepatite B (Figura 1.1).

Ao se acrescentar o CP 3 (Figura 1), destaca-se a variável “sexo (gênero)” com maior coeficiente (- 0.6829707) e eleva o poder de explicação da variabilidade dos dados no banco de dados de hepatite B em aproximadamente 61% (Figura 1.1), conforme mostra a proporção acumulada entre os três e mais importantes componentes principais (CP1, CP2 e CP3). Por outro lado, acrescentando mais um componente principal na sequência, neste caso, o CP4 (Figura 1), destaca-se a variável “raça (etnia)”, com coeficiente (- 0.66271869). A explicação

total do banco de dados de hepatite B, utilizando os quatro e mais importantes

componentes principais (CP1, CP2, CP3 e CP4), é de aproximadamente 74% (Figura 1.1).

Variáveis	CP1	CP2	CP3	CP4
ANO	- 0.47231924	+ 0.43805208	+ 0.4678939	-0.20556668
IDADE_ano	- 0.76738590	- 0.16221765	+ 0.2677251	+ 0.37747495
SEXO.Freq.	+ 0.19422132	+ 0.47380797	- 0.6829707	- 0.01982262
RACA.Freq.	+ 0.44380612	- 0.35494897	+ 0.3432241	- 0.66271869
ESCOLARIDADE.classe.	+ 0.77306034	+ 0.00835573	+ 0.3455104	+ 0.23508984
BAIRRO.Freq.	+ 0.56980275	+ 0.42367445	+ 0.1045699	+ 0.25760660
CIRÚRGICO.Freq.	+ 0.07733158	+ 0.77329024	+ 0.4025833	+ 0.06878936
DENTÁRIO.Freq	+ 0.32373183	- 0.49345879	+ 0.1512031	+ 0.48404653

Figura 1. Quadro com os valores dos autovetores (coeficientes) do banco de dados de hepatite B. Fernandópolis (SP), Brasil, 2016.

Fonte: Sistema de Informação de Agravo de Notificação - (SINAN) do município de Fernandópolis (SP).

CPs	Autovalores	Proporção (%)	Proporção acumulada (%)
CP1	2.0797323	25.996654	25.99665
CP2	1.5897362	19.871702	45.86836
CP3	1.1901013	14.876266	60.74462
CP4	0.9849954	12.312443	73.05706

Figure 1.1 - Table with the eigenvalues showing the proportion explained by the first four and most important MCs, database HB. 2016.

Source: Notification Aggravation Information System (SINAN) of the city of Fernandópolis (SP).

**Individuals factor map (PCA)**

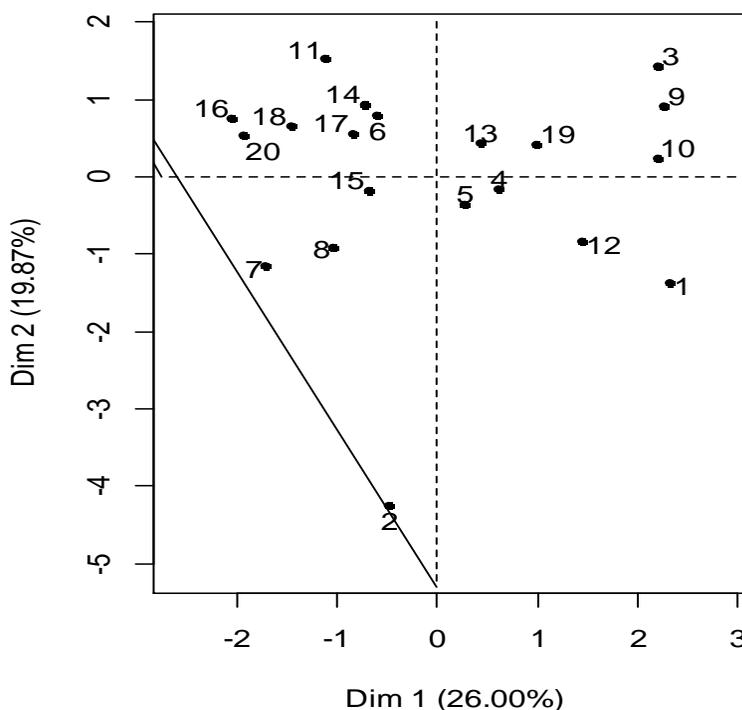


Figura 1.2 - Biplot do banco de dados dos infectados pelo VHB, no período de 2007 a 2011, gerando os CP1 e CP2. Fernandópolis (SP), Brasil, 2016.

Fonte: Sistema de Informação de Agravo de Notificação - (SINAN) do município de Fernandópolis. Fernandópolis (SP), Brasil, 2016.

Por sua vez, os resultados para o banco de dados da hepatite C, que possui 30 observações de casos e oito variáveis, revelou que o CP 1 (24,64%) (Figura 2.) possui os maiores coeficientes de variações deste banco de dados, sendo as variáveis escolaridade (coeficiente - 0,68840788) e bairro do paciente (coeficiente + 0,64556072) as mais importantes. Na Figura 2, destacou-se o

indivíduo seis do banco de dados, revelando que no CP1 foi o que obteve maior coeficiente, com as variáveis “escolaridade” resultando “analfabeto” e “bairro de residência” resultando “bairro Redentor”. Foi seguido do indivíduo 25, que teve escolaridade “1ª a 4ª série do ensino fundamental incompleta” (antigo primário ou 1º grau) e reside no “bairro Vila Regina”,

ambas as categorias das variáveis que tiveram as maiores variabilidades do banco de dados da hepatite C.

Em relação ao CP 2 (17,68%) (Figura 2.), os maiores coeficientes mostram as variáveis mais importantes deste banco de dados, sendo “sexo” (gênero) (coeficiente + 0,78498043) e “ano de notificação do caso” do paciente (coeficiente - 0,69346448). Na Figura 2, destacou-se o indivíduo um do banco de dados onde, no CP2, com a variável “gênero”, na categoria “feminino”, e o caso com notificação em “2007”. Seguido do indivíduo quatro, que é do gênero feminino, com

notificação do caso no ano de “2008”. O total de explicação da variabilidade dos dados entre CP1 e CP2 é de 42,32% para o banco de dados da hepatite C.

No CP 3, destacou-se a variável “exposição a tratamento dentário” (coeficiente + 0.93486633). No CP4, destacou-se a variável “raça (etnia)” (coeficiente + 0.50650838). O total de explicação da proporção acumulada entre os CP1, CP2 e CP3 é de 58%. Com os quatro componentes principais, o total de explicação da proporção acumulada entre os CP1, CP2 e CP3 e CP4, para o banco de dados de hepatite C, é de aproximadamente 69%.

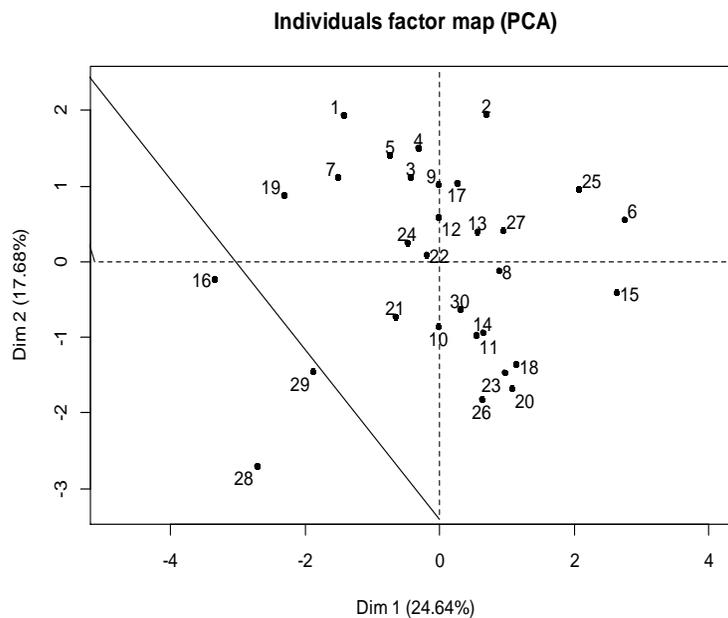


Figura 2 - Biplot do banco de dados dos infectados pelo VHC, no período de 2007 a 2011, gerando os CP1 e CP2. Fernandópolis (SP). 2016.

Fonte: Sistema de Informação de Agravo de Notificação - (SINAN) do município de Fernandópolis (SP).

Por fim, este estudo mostra o resultado da junção dos dados de hepatite B com hepatite C que têm 50 observações de casos e oito variáveis. O CP 1 (19,70%) (Figura 3) obteve os maiores coeficientes, mostrando as variáveis mais importantes desta junção de banco de dados, destacando-se “bairro” (coeficiente + 0,70469080) e “idade do paciente” (coeficiente - 0,69058785). Na Figura 3, o CP 1 destaca o indivíduo 25 do banco de dados, apresentando as variáveis “bairro onde reside”, das “Paineiras”, e “idade”, na categoria 25 anos. Foi seguido do indivíduo dois, que reside no bairro Jardim Paulista e tem idade de 35 anos.

Para o CP 2 (15,84%) (Figura 3), os maiores coeficientes são observados para as variáveis “escolaridade” (coeficiente + 0.725505640) e “exposição a tratamento dentário” (coeficiente + 0.592724503). Na Figura 3, destacou-se o indivíduo cinco do banco de dados. No CP2, ele obteve maior coeficiente, com as variáveis “escolaridade”, “superior

completo”, e “exposição a tratamento dentário”, “sim, há mais de seis meses”. Foi seguido do indivíduo 12, com escolaridade “superior completa” e exposição a tratamento dentário “sim, há mais de seis meses”. O total de explicação da variabilidade dos dados entre CP1 e CP2 é de aproximadamente 36% em relação ao banco de dados junção unificado entre as infecções pelo VHB ou VHC.

O CP 3 destacou a variável “sexo” (coeficiente + 0.81172170) e o CP 4 destacou a variável “raça (etnia)” (coeficiente - 0.709506482). O total de explicação da proporção acumulada entre os CP1, CP2 e CP3 é de aproximadamente de 50%. Com os quatro componentes principais, o total de explicação da proporção acumulada entre os CP1, CP2 e CP3 e CP4, para o banco de dados junção unificada entre as infecções pelo VHB ou VHC, é de aproximadamente 63%.

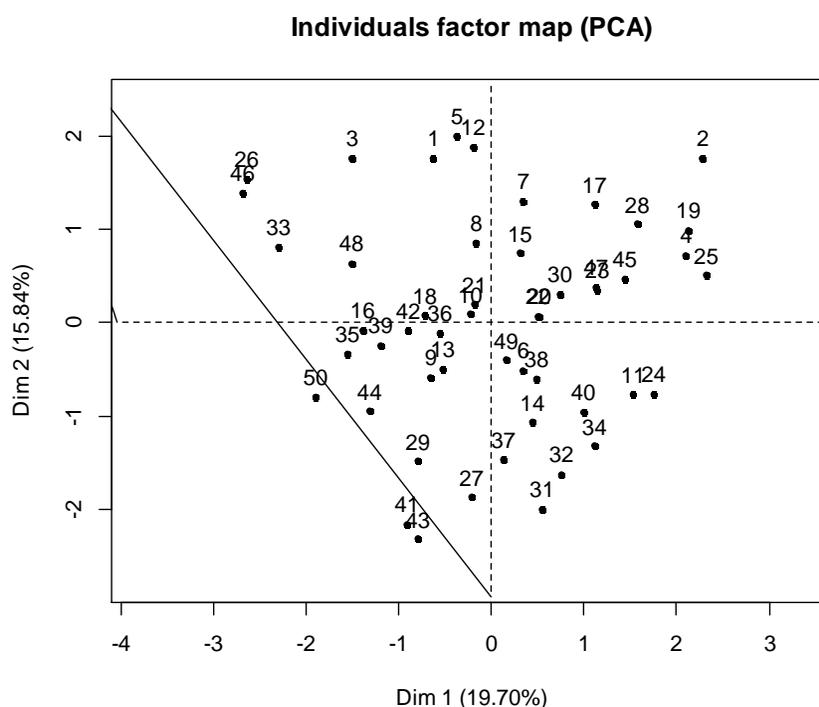


Figura 3. - Biplot do banco de dados dos infectados pelo VHB e VHC, no período de 2007 a 2011, gerando os CP1 e CP2. Fernandópolis (SP). 2016.

Fonte: Sistema de Informação de Agravos de Notificação - (SINAN) do município de Fernandópolis (SP).

## DISCUSSÃO

O resultado deste estudo, em relação à idade do paciente, condiz com estudo encontrado de literatura.<sup>17</sup> Por outro lado, apresenta sentido contrário ao resultado encontrado em outro estudo, sendo em população indígena onde predomina idade mais elevada entre os pacientes com hepatite B (HB). Normalmente, foram acima de 40 anos.<sup>18</sup> Para a hepatite C, evidências de literatura sugerem a maioria dos casos nos brasileiros com idade média superior aos 50 anos.<sup>17,19</sup>

É importante o trabalho de prevenção com as pessoas nesta idade ou entre a faixa etária dos 20 aos 30 anos e acima dos 50 anos, pois a vacina para esta doença - hepatite B - está disponível no SUS há muitos anos.<sup>20</sup>

Quanto à escolaridade para HB, visto que houve uma variabilidade no resultado, mas com um relativo predomínio em indivíduos de média escolaridade, entre ensino médio completo e ensino médio incompleto, e para hepatite C, baixa escolaridade, sendo analfabeto e fundamental incompleto. Para a junção do banco de dados para as hepatites B e C, a variável “escolaridade” foi superior completo. Semelhante a esta última é a pesquisa realizada em São Paulo-SP (2004-2007) entre servidores públicos, onde a faixa de escolaridade mais frequente foi acima de 12 anos, com 21,4% para a HB e 18,3% para HC, situação que revelou uma amostra significativa registrada como ignorado.<sup>21</sup>

Observando a informação “ignorado” desta pesquisa, encontrada na literatura, há dedução de uma possível mudança no resultado para escolaridade para hepatite B e C nesta pesquisa, se os dados de “ignorado” fossem preenchidos.

Um estudo de revisão integrativa na literatura revelou que a baixa escolaridade está associada à infecção pelo VHB em pessoas do sistema prisional.<sup>22</sup>

A baixa escolaridade e o não acesso à informação favorecem mecanismos de transmissão associados às condições socioeconômicas deficientes, tais como: promiscuidade, aglomeração, baixos padrões de higiene e dificuldade de acesso a uma assistência médica de qualidade.<sup>23</sup>

Em relação à “exposição a tratamento dentário” e “exposição a tratamento cirúrgico”, esses resultados são semelhantes aos encontrados em outro estudo onde foi evidenciado que, nas mulheres, o segundo maior percentual para a hepatite C (HC) é a exposição via tratamento odontológico e cirúrgico.<sup>24</sup>

Em relação aos bairros, a variável “lugar” ou “geográfica” é muito importante para a comparação de áreas ou comunidades onde o evento mostra diferentes incidências. É uma alternativa para sugerir ideias, identificar possíveis fatores etiológicos, em geral, do meio ambiente, ou confirmar as associações causais suspeitadas.<sup>25</sup>

Estudo recente sugere que houve queda na prevalência de hepatite B no Brasil, mas, em

Sampaio SA, Hongyu K, Pinto Neto J Martins et al.

localidades isoladas e distantes, ainda há baixas taxas de cobertura vacinal, o que reforça a necessidade de intensificar as estratégias de vacinação para jovens e adultos em regiões específicas, com persistência maior de prevalência da infecção pela HB.<sup>26</sup> Quanto à hepatite C, estudo prévio mostrou queda significativa no Brasil a partir de 2004, sugerindo este estudo que poderia ser consequência da criação do Programa Nacional de Hepatites Virais, testes sorológicos de hepatite nos Centros de Testagem e Aconselhamento e Programas de Redução de Danos.<sup>27</sup>

Com base no exposto, é importante que os serviços de saúde levem em consideração esta variável no planejamento das ações em saúde, lembrando que cada lugar tem suas especificidades, sejam elas socioeconômica, física e ambiental, cultural e organizacional.

Em relação ao gênero feminino, esses dados condizem com o encontrado em outro estudo onde, em 51,5%, predominou o gênero feminino para a HC.<sup>21</sup> É importante destacar, como hipótese, os fatores ambientais, comportamentais e a organização social, ter uma ligação e influência na distribuição por região das hepatites virais em relação ao gênero, devendo outros estudos serem realizados para mais esclarecimentos.

Por outro lado, é encontrado, na literatura, estudo de HC, onde a maioria dos infectados era do gênero masculino e encontrava-se na faixa etária acima de 40 anos.<sup>28</sup> Os autores descrevem que os dados encontrados sugerem que os indivíduos do gênero masculino foram infectados pelo HC antes da inclusão de exames sorológicos obrigatórios no Brasil e, por se tratar de uma infecção latente e muitas vezes silenciosa, passaram a apresentar alterações do quadro clínico mais tardiamente.

Outra explicação encontrada na literatura de predomínio de HC no sexo masculino aborda vários fatores de risco relacionados a hábitos sexuais que parecem contribuir para maior transmissão da HC, como: fatores culturais, falta de ações educativas de sexualidade, maior número de parceiros sexuais, baixa adesão ao uso de preservativos, experiências sexuais traumáticas, presença de outras doenças sexualmente transmissíveis associadas como sífilis, tricomoníase, HIV/AIDS e infecção por clamídia.<sup>19</sup> Além de abordarem que a transmissão homem-mulher parece ocorrer mais facilmente do que a transmissão mulher-homem mas que, apesar dessas evidências, estudos envolvendo casais monogâmicos demonstraram baixo risco de transmissão sexual. Contudo, a possibilidade

Utilização da análise multivariada para dados de...

de transmissão intrafamiliar, por compartilhamento de material de higiene pessoal ou eventual exposição a sangue contaminado, interfere na interpretação dos estudos que avaliaram a transmissão sexual da HC.

Um fator a ser destacado é que o SUS, por meio dos serviços de saúde, deve planejar e executar estratégias do Programa de Controle de HB e HC, considerando a variável “gênero” na operacionalização de suas ações, desde o horário do expediente da unidade, até os tipos de estratégias de educação em saúde.

Intervenções como ações educativas, de diagnóstico e tratamento, dentro de um programa Nacional de Atenção à Saúde do Homem, são fundamentais, pensando nas Doenças Sexualmente Transmissíveis (DSTs), câncer de próstata, doenças cardiovasculares, diabetes e adoção de práticas saudáveis, como as realizadas em programas de Atenção à Saúde da Mulher e Gestantes. Outra maneira de combater esta doença é intensificar a socialização de informações sobre sua transmissão, para fins de autoproteção.

Com base nos dados encontrados, fica evidente a importância do grau de instrução por meio da educação, o que leva o indivíduo a assimilar melhor as informações para a prevenção das infecções de hepatites virais.

As unidades de saúde devem levar em consideração esta variável no planejamento, organização e no desenvolvimento de suas atividades diárias, pois, em se tratando de uma clientela com relativa baixa e média escolaridade, as ações educativas devem ser desenvolvidas numa linguagem adequada à capacidade cognitiva dos usuários.

Este estudo faz a observação de que a utilização de banco de dados de hepatites virais, provinda do sistema de informação SINAN, para a análise multivariada, utilizando a técnica de componentes principais, para a pesquisa descritiva deste estudo, mostrou-se com pouca qualidade, no que se refere ao preenchimento de vários campos nas fichas de notificação compulsória, como variáveis com informações incompletas. Como consequência na metodologia deste estudo, houve a exclusão dessas várias variáveis, com frequências zero ou em branco no banco de dados, mascarando o resultado final do que já seria longe da realidade do município pesquisado. Na hipótese de análise dos resultados, mesmo com as variáveis completas, seriam apenas dos que procuraram os serviços de saúde do município no período. Portanto, deve-se ter o cuidado de não generalizar o resultado para todo o município.

## CONCLUSÃO

Com as variáveis e observações que este estudo pode trabalhar, a análise multivariada, por meio da técnica de componentes principais, mostrou que o resultado da análise do banco de dados de hepatites virais, provindo do sistema de informação SINAN, mostrou maior frequência na população adulta, com relativa baixa e média escolaridade e predomínio do gênero feminino para a hepatite C.

Todos os profissionais de saúde e agentes comunitários de saúde têm importante papel na realização de ações educativas, preventivas e de diagnóstico, favorecendo o tratamento precoce dessas patologias junto à população. Os resultados deste estudo também esperam colaborar na didática e interpretação de análise multivariada, por meio de componentes principais, no que se refere ao banco de dados secundários.

## AGRADECIMENTOS

Ao professor Dr. Francisco José Dutra Souto, pela revisão crítica intelectual do trabalho.

Ao professor Dr. Neuber José Segri, do departamento de Estatística do ICET/UFMT.

## REFERÊNCIAS

1. World Health Organization. Hepatitis B [Internet]. Geneva: WHO; 2015 [cited 2016 July 19]. Available from: <http://www.who.int/mediacentre/factsheets/fs204/en/>
2. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012 Dec;380(9859):2095-128. doi: 10.1016/S0140-6736(12)61728-0.
3. World Health Organization. Hepatitis B [Internet]. Geneva: WHO; 2002 [cited 2016 July 19]. Available from: [http://www.who.int/csr/disease/hepatitis/HepatitisB\\_whocdscsrlyo2002\\_2.pdf](http://www.who.int/csr/disease/hepatitis/HepatitisB_whocdscsrlyo2002_2.pdf)
4. World Health Organization. Hepatitis C [Internet]. Geneva: WHO; 2015 [cited 2016 July 19]. Available from: <http://www.who.int/mediacentre/factsheets/fs164/en/>
5. Ministério da Saúde (BR), DATASUS - Tecnologia da Informação a Serviço do SUS. Hepatites virais, casos notificados no sistema de informação de agravos de notificação - Brasil [Internet]. Brasília: Ministério da

Saúde; 2015 [cited 2016 July 19]. Available from:

<http://tabnet.datasus.gov.br/cgi/tabcgi.exe?sinannet/cnv/hepabr.def>

6. Pearson K. On lines and planes of closest fit to systems of points in space. *Philos Mag [Internet]*. 1991 [cited 2016 July 19];2:559-72. Available from:

7. <http://stat.smmu.edu.cn/history/pearson1901.pdf>

8. Hongyu K, Sandanielo VLM, Oliveira Junior GJ. Análise de Componentes Principais: resumo teórico, aplicação e interpretação. *Eng Sci [Internet]*. 2016 [cited 2016 July 19];1(5):83-90. Available from: <http://periodicoscientificos.ufmt.br/ojs/index.php/eng/article/view/3398/2744>

9. Hotelling H. Analysis of a complex of statistical variables into principal components. *The J Educ Psychol*. 1933;24(7):498-520. DOI: [10.1037/h0071325](https://doi.org/10.1037/h0071325)

10. Hotelling H. Simplified calculation of principal components. *Psychometrika [Internet]*. 1936 [cited 2016 July 19];1(3):27-35. Available from:

11. <https://pdfs.semanticscholar.org/07f2/665693e0903ddcf8e146da9e443f4709a057.pdf>

12. Lyra WS, Silva EC, Araújo MCU, Fragoso WD, Veras G. Classificação periódica: um exemplo didático para ensinar análise de componentes principais. *Quím Nova [Internet]*. 2010 [cited 2016 July 19];33(7):1594-97. Available from: <http://www.scielo.br/pdf/qn/v33n7/a30v33n7.pdf>

13. Hair JF, Anderson RE, Tatham RL, Black WC, BALBIN BJ. Análise multivariada de dados. 6th ed. Porto Alegre: Bookman; 2009.

14. Savegnago RP, Caetano SL, Ramos SB, Nascimento GB, Schmidt GS, Ledur MC, et al. Estimates of genetic parameters, and cluster and principal components analyses of breeding values related to egg production traits in a White Leghorn population. *Poult Sci [Internet]*. 2011 Oct [cited 2016 July 19];90(10):2174-88. Available from: <http://ps.oxfordjournals.org/content/90/10/2174.full.pdf+html>

15. Fraga AB, Silva FL, Hongyu K, Santos DS, Murphy TW, Lopes FB. Multivariate analysis to evaluate genetic groups and production traits of crossbred Holstein x Zebu cows. *Trop Anim Health Prod*. 2015 Mar;48(3):533-8. Doi: 10.1007/s11250-015-0985-2

16. Johnson RA, Wichern DW. Applied multivariate statistical analysis. Madison: Prentice Hall International; 1998.

Sampaio SA, Hongyu K, Pinto Neto J Martins et al.

Utilização da análise multivariada para dados de...

17. Anderson TW. An introduction to multivariate statistical analysis. 6th ed. New York: Wiley; 2003.
18. Ferreira DF. Estatística multivariada. Lavras: UFLA; 2011.
19. Aquino JÁ, Pegado KA, Barros LP, Machado LFA. Soro prevalência de infecções por vírus da hepatite B e vírus da hepatite C em indivíduos do Estado do Pará. Rev Soc Bras Med Trop [Internet]. 2008 July/Aug [cited 2016 July 19];41(4):334-7. Available from: <http://www.scielo.br/pdf/rsbmt/v41n4/a03v41n4.pdf>
20. Nunes HM, Monteiro MRCC, Soares MCP. Prevalência dos marcadores sorológicos dos vírus das hepatites B e D na área indígena Apyterewa, do grupo Parakanã, Pará, Brasil. Cad Saúde Pública [Internet]. 2007 Nov [cited 2016 July 19];23(11):2767-79. Available from: <http://www.scielo.br/pdf/csp/v23n11/22.pdf>
21. Martins T, Narciso-Schiavon JL, Schiavon LL. Epidemiologia da infecção pelo vírus da hepatite C. Rev Assoc Med Bras [Internet]. 2011 Jan/Feb [cited 2016 July 19];57(1):107-12. Available from: <http://www.scielo.br/pdf/ramb/v57n1/v57n1a24.pdf>
22. Ministério da Saúde (BR), Secretaria de Vigilância em Saúde. Guia de Vigilância em Saúde / Ministério da Saúde, Secretaria de Vigilância em Saúde. Brasília: Ministério da Saúde; 2014 [cited 2016 July 19]. Available from:
23. <http://portalsaude.saude.gov.br/images/pdf/2014/novembro/27/guia-vigilancia-saude-linkado-27-11-14.pdf>
24. Cruz CRB, Shirassu MM, Martins WP. Comparação do perfil epidemiológico das hepatites B e C em um serviço público de São Paulo. Arq Gastroenterol [Internet]. 2009 July/Sept [cited 2016 July 19];46(3):225-9. Available from: <http://www.scielo.br/pdf/ag/v46n3/16.pdf>
25. Silva AAS, Araújo TME. Fatores associado à hepatite B em população carcerária: revisão integrativa. J Nurs UFPE on line [Internet]. 2015 Sept [cited 2016 July 25];9(9):9276-84. Available from: [http://www.revista.ufpe.br/revistaenfermagem/index.php/revista/article/view/6922/pdf\\_8572](http://www.revista.ufpe.br/revistaenfermagem/index.php/revista/article/view/6922/pdf_8572)
26. Passos ADC, Gomes UA, Figueiredo JFC, Nascimento MMP, Oliveira JM, Gaspar AMC, et al. Influência da migração na prevalência de marcadores sorológicos de hepatite B em comunidade rural: 2 - Análise comparativa de algumas características das populações estudadas. Rev Saúde Pública [Internet]. 1993 [cited 2016 July 19];27(1):36-42. Available

from:

- <http://www.scielo.br/pdf/rsp/v27n1/06.pdf>
27. Acosta LMW, Fiorini M, Arruda KT, Carvalho VRS, Marques NS. A história dos dez anos da vigilância epidemiológica das hepatites virais do tipo B e C em Porto Alegre. Bol Epidemiol [Internet]. 2006 Nov [cited 2016 July 19];9(32):1-8. Available from: [http://lproweb.procempa.com.br/pmpa/pref\\_poa/sms/usu\\_doc/boletim32.pdf](http://lproweb.procempa.com.br/pmpa/pref_poa/sms/usu_doc/boletim32.pdf)
28. Pereira MG. Epidemiologia: teoria e prática. Rio de Janeiro: Guanabara Koogan; 2015.
29. Souto FJD. Distribution of hepatitis B infection in Brazil: the epidemiological situation at the beginning of the 21st century. Rev Soc Bras Med Trop [Internet]. 2016 Feb [cited 2016 jul 19];49(1):11-23. Available from: <http://www.scielo.br/pdf/rsbmt/2015nahead/0037-8682-rsbmt-1015900037868201762015.pdf>
30. Azevedo AO, Santos MM, Jerez-Roig J, Souza DLB. Incidência de hepatites virais no Brasil de 1997 a 2010. Rev Enferm UFPE on line [Internet]. 2015 Apr [cited 2016 July 25];9(4):7375-82. Available from:
31. [http://www.revista.ufpe.br/revistaenfermagem/index.php/revista/article/view/6044/pdf\\_7544](http://www.revista.ufpe.br/revistaenfermagem/index.php/revista/article/view/6044/pdf_7544)
32. Araújo AR, Almeida CM, Fraporti L, Garcia N, Lima TA, Maia LPV, et al. Caracterização do vírus da hepatite C em pacientes com hepatite crônica: genótipos no Estado do Amazonas, Brasil. Rev Soc Bras Med Trop [Internet]. 2011 Sept/Oct [cited 2016 July 19];44(5):638-40. Available from: <http://www.scielo.br/pdf/rsbmt/v44n5/24.pdf>

Submissão: 20/09/2016

Aceito: 12/07/2017

Publicado: 01/08/2017

### Correspondência

Sidnei Anastácio Sampaio  
 Graduação em Enfermagem  
 Instituto de Ciências Exatas e Naturais  
 Universidade Federal do Mato Grosso  
 Rua Marcio Bortolo, Quadra 2-F, Lote 13,  
 Bairro Paraíso,  
 CEP: 78734-020 – Rondonópolis (MT), Brasil