Journal of Nursing
Revista de Enfermagem
UFPE On Line
ISSN: 1981-8963

**ORIGINAL ARTICLE**

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

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

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

*Sidnei Anastácio Sampaio[1], Kuang Hongyu[2], José Martins Pinto Neto[3], Vando Edésio Soares[4], Ricardo Scarparo Navarro[5]*

**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.

**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.

**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. ***Descriptores:*** Hepatitis; Análisis Multivariante; Análisis de Componente Principal.

[1]Nurse, Master Professor, Graduation in Nursing / Enf / ICEN / CUR, Federal University of Mato Grosso / UFMT. Rondonópolis (MT), Brazil. E-mail: sidnei.sampaio1@gmail.com; [2]Mathematician, PhD Professor, Graduation in Statistics / Est / ICET, Federal University of Mato Grosso / UFMT. Cuiabá (MT), Brazil. E-mail: prof.kuang@gmail.com; [3]Nurse, Professor Doctor, Undergraduate Nursing / Medicine, Universidade Camilo Castelo Branco / UNICASTELO. Fernandópolis (SP), Brazil. E-mail: jmpintoneto@gmail.com; [4]Zootechnician, Ph.D. Professor, Graduation in Animal Science / Post-Graduate Program in Biomedical Engineering / Research and Extension Institute - Universidade Camilo Castelo Branco / UNICASTELO. São Paulo (SP), Brazil. E-mail: soaresvando@gmail.com; [5]Surgeon-dentist, PhD, Professor, Graduate in Dentistry / Post-Graduation Program in Biomedical Engineering / Research and Extension Institute - Universidade Camilo Castelo Branco / UNICASTELO. São Paulo (SP), Brazil. E-mail: ricardosnavarro@gmail.com

## INTRODUCTION

It is estimated that 240 million people are chronically infected worldwide with the hepatitis B virus (HBV) .[1] About 800,000 people die each year from this disease and less than 650,000, cirrhosis and liver cancer, chronic hepatitis B (HB), infection and another 130,000, acute hepatitis B, mainly, due to complications of liver cirrhosis or primary liver cancer (hepatocellular carcinoma, CHC). [2] For comparison purposes, hepatitis B virus is 50 to 100 times more infectious than the *Human Immunodeficiency Virus* (HIV).[3]

In comparison to infection with Hepatitis C virus (HCV), there are approximately 130 to 150 million people with the virus. A significant number of people suffering from chronic infection will develop liver cirrhosis or CHC liver cancer. It is estimated that the number of people dying annually of liver disease related to the hepatitis C virus each year is between 350 and 500 thousand.[4]

In Brazil, 13,187 confirmed cases of hepatitis B, followed by 51 cases of association between hepatitis B and A (B + A) were reported, according to data subject to revision of the Notification of Injury Information System (NIIS) reporting system in 2014, 333 For B + C, 226 hepatitis B + D. In that same year, for hepatitis C, there were 12,934 confirmed cases, followed by 26 cases of hepatitis C + A. The overall number of cases reported in Brazil was 34,637 confirmed cases throughout the 2014 period, for all the etiology classification of viral hepatitis.[5]

In the State of São Paulo, in 2014, a total of 2,229 confirmed cases of hepatitis B were reported, followed by association of viral hepatitis B + C, with 72 cases; B + A, with four cases. In relation to hepatitis C, 3,192 confirmed cases were reported, followed by four cases of C + A.[5]

Principal component analysis (PCA) is a multivariate model of covariance structure modeling. The technique was first described by Pearson, in 1901.[6-7] A description of practical computational methods came much later, with Hotelling, in 1933 and 1936,[8-9] who used it for the purpose of analyzing correlation structures.[7]

The analysis of the main components is one of the most important statistical tools of the multivariate analysis, since it is the basis of most of the other multivariate methods of data analysis, being an exploratory data analysis tool that reveals the existence or not of Different samples of relationships between measured variables and relations or groupings between samples, aiming at their reduction, maintaining only a small number of Main Components (MCs) that already account for a significant portion of the total information contained in the data structure.[10]

As diseases transmitted by blood and body fluids, HBV or HCV infections can be acquired in establishments where there is exposure to puncture-prone material such as needles, scissors and blades. This study was carried out in the city of Fernandópolis, state of São Paulo, where there are several health centers and other activities that increase the risk of infections with viral hepatitis B and C with contaminated material inoculation, as in the cases of which make tattoo, manicures and pedicures, hairdressers, cosmetic centers, dental offices, hospitals, health care units, pharmacies, sex workers, among others.

The disease appears to be well known to the population and health professionals, but it is possible that many do not follow guidelines on the use of individual or disposable material, which leads to vulnerability to infection/disease.

Therefore, based on the above, the severity and problematic of viral hepatitis B and C is evident, which justifies the accomplishment of this study. Due to the importance of both infections for human health, the prevalence or incidence studies are relevant, as well as the effort to optimize the analysis of the databases formed by this research. This study is based on the analysis of the variables of a database on hepatitis B and C in Fernandópolis (SP), in order to identify the main variables that emanated from the analysis.

## OBJECTIVES

- To identify which variables are the most important in the database for viral hepatitis B, viral hepatitis C and the junction of hepatitis B and C databases.
- To reduce the number of variables in the viral hepatitis B, viral hepatitis C, and viral hepatitis B and C databases.

## METHOD

The database of this study was extracted from the Graduate Program in Biomedical Engineering / UNICASTELO, master's thesis titled "Incidence of Viral Hepatitis B and C in Fernandópolis S.P.".

Descriptive and retrospective study, with a quantitative approach. It used data obtained from a secondary source of the Notification Injury Information System (NIIS) available at the Epidemiological Surveillance Service of the Municipal Health Department of the city

of Fernandópolis (SP), referring to the reported and confirmed cases of viral hepatitis B and C, in the 2007-2011 period, of patients residing in this municipality.

The Project was approved by the Ethics and Research Committee of Camilo Castelo Branco University (CEP UNICASTELO CAAE n ° 01475612.7.0000.5494). The data was analyzed by the multivariate analysis method, through analysis of main components, at a significance level of 5%, by the statistical program R platform.

The crude database contained 46 variables, with 158 confirmed cases of hepatitis B and C, 51 cases of HBV infection; 101, cases of HCV infection and eight cases of association between the two HBV and HCV agents. All the variables whose response data were "No" and when "most of the data, such as No, Ignored and Blank", were collected from the gross database, (Pregnant, Area of residence, Occupation, Suspected of hepatitis; exposed to injectable medication, inhaled drugs or crack, injectable drugs, contaminated water / food, three or more sex partners, transplantation, tattoo / piercing, acupuncture , hemodialysis, accident with biological material, blood transfusion / derivatives, blood bank serology, HCV genotype, final classification, clinical form, etiological classification and probable source of infection). Variables with a single response were also removed, such as (Municipality of Residence, Regional Health, Health Unit, Country, Federative Unit) and non pertinent variables, ie, meaningless to do the analysis, for example, dates (day, month). We also excluded all blank and ignored frequencies (observations) of the variables that remained in the study for analysis. The total variables used in the study were eight variables for each database of HB, HC and HB / HC database join. There were 20 confirmed cases of HB, 30 cases of HC and 50 cases of HB and HC database.

The main data collection site of the mentioned master's thesis, which was used for the analysis of this research was the municipality of Fernandópolis (SP).

**♦ Multivariate analysis**

The multivariate analysis technique refers to all statistical techniques that simultaneously analyze multiple measures on individuals or objects under investigation. Thus, any simultaneous analysis, of more than two variables, can be considered, in principle, as multivariate, enabling health agencies and organizations to acquire knowledge and improve, their decision making.[11]

**♦ Main component analysis**

The principal component analysis aims to reduce the dimensionality of the set of variables and to facilitate the interpretation of the independence between them.[7,11] For this, linear combinations of the original variables are obtained that, geometrically, represent the selection of new coordinate systems obtained by rotation of the original system that has p random variables as coordinate axes. These new orthogonal axes (new variables) are called the main components and the values of the new variables are called scores of the main components or main coordinates.

In general, the most important major component (the first main component), is chosen as the one of greater variance, which explains the maximum variability of the data; the second most important component, which presents the second largest variance, and And so on, to the major minor component.[7,12-3] On the other hand, the last major components will be responsible for directions that are not associated with much variability. In other words, these last major components will identify linear relationships between the original variables close to constant.[7,14-6]

The main components are obtained by means of the diagonalization of symmetrical positive semidefinite matrices. Then one can calculate the main components easily and use them in different applications in the most varied scientific areas. This facility is due to the existence of numerous programs capable of performing matrix calculations to diagonalize a semi-definite positive symmetric matrix.[7]

Many researchers have used main component analysis to solve problems such as multicollinearity in linear regression, to estimate factors, which represent another multivariate technique of modeling the covariance matrix, to perform the modeling of the interaction between factors in experiments without repetition, Divergence and grouping between genotypes in genetic study and breeding of plants and animals, among other possibilities.[7,14]

To obtain the main components in a general way, let p be a set of variables $X_1$, $X_2$, ... $X_p$, with averages $\mu_1$, $\mu_2$, ..., $\mu_p$ and variance $\sigma^2_1$, $\sigma^2_2$, ..., $\sigma^2_p$, respectively, these variables are not independent and therefore have covariance between *i*-th e *K*-th variable defined by $\sigma_{ik}$, for $i \neq k = 1, 2, ..., p$. Then the p variables can be expressed in vector form by: $X = [X_1, X_2, ..., X_p]^T$, with average vector $\mu = [\mu_1, \mu_2, ..., \mu_p]^T$ And covariance matrix Σ.

The pairs of eigenvalues and eigenvectors ($\lambda_1$, $e_1$), ($,\lambda_2$, $e_2$), ... , ($\lambda_p$, $e_p$), in which $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p$, associated with Σ and so the *i*-th main component is defined by:[7,14]

$$Zi = e^T_i X = e_{i1}X_1 + e_{i2}X_2 + ... + e_{ip}X_p \quad (1)$$ and that $i = 1, 2, ..., p$.

## RESULTS

For the hepatitis B database, which has 20 case observations and eight variables, the main component 1 (MC 1), that is, eigenvector 1, obtained the highest percentage (26.00%) (Figure 1.2) among all The MCs. There was a contrast between the variables. One emerges with a positive sign for the coefficient (eigenvectors) and the other with negative sign for another coefficient. So, this means that by taking into account the signal for interpretation of the result in relation to the variables, they have opposite values in comparison with each other. The patient's age (coefficient - 0.76738590) and patient's educational level (coefficient + 0.77306034), (Figure 1) have the greatest variation in the data. Figure 1.2 highlighted the individual nine of the database, and in MC1 it was the one that obtained the greatest variation of the data, with the variables age and schooling, these variables in their categories showed that individual nine is 22 years old and has completed high school (Former high school or high school), followed by individual three who is 28 years old and completes higher education, then individual ten, who is 25 years old and incomplete high school education (former high school or high school), which corresponds to the database. All of these MC1 variables were the ones that most explained the variation in hepatitis B data for this database.

When generalizing the result of MC1 for the individuals observed, it can be said that the individuals appearing in Figure 1.2 (graph), on the X axis, from zero to threeMC, are the individuals that have less age and high level of schooling. Already in the same Figure 1.2, it can be said that the individuals that appear in the X axis near the values - 2 (minus two) and near zero (0) negatives, that is, negative quadrants (upper and lower left) of the X axis representing MC1, are the individuals with the highest age and low level of schooling, which corresponds to the observed database. This same interpretation is used for the other variables of the other databases, when contrast occurs.

The MC 2 of the hepatitis B database, explained 19.87% (Figure 1.2) of the data variation. There is a contrast between the variables, being the most important variables of this MC2 "exposure to surgical treatment" (coefficient + 0.77329024) and "exposure to dental treatment" (coefficient - 0,49345879) (Figure 1). This contrast is observed by the opposite signs of the coefficients, positive for one variable and negative for another variable. Figure 1.2 highlighted the individual 11 of the database, and in MC2 it was the one that obtained the highest coefficient, with the variables “exposure to surgical treatment” (yes, more than six months ago) and exposure to dental treatment (yes, less than six months ago). The total explanation for the variability of data between MC1 and MC2 is approximately 46% for the hepatitis B database, (Figure 1.1).

When adding MC 3 (Figure 1), the variable “sex (gender)” with the highest coefficient (- 0.6829707) stands out and increases the power of explanation of data variability in the hepatitis B database by approximately 61% (Figure 1.1), shows the cumulative ratio of the three major components (MC 1, MC 2 and MC 3). On the other hand, adding another major component in the sequence, in this case, MC 4 (Figure 1), stands out the variable “race (ethnicity)” with coefficient (- **0.66271869**). The total explanation of the hepatitis B database, using the Four and most important major components (MC 1, MC 2, MC 3 and MC 4), is approximately 74% (Figure 1.1).

| Variables | MC1 | MC2 | MC3 | MC4 |
|---|---|---|---|---|
| YEAR | - 0.47231924 | + 0.43805208 | + 0.4678939 | -0.20556668 |
| AGE_year | - 0.76738590 | - 0.16221765 | + 0.2677251 | + 0.37747495 |
| SEX.Freq. | + 0.19422132 | + 0.47380797 | - 0.6829707 | - 0.01982262 |
| RACE.Freq. | + 0.44380612 | - 0.35494897 | + 0.3432241 | - 0.66271869 |
| EDUACATION.class. | + 0.77306034 | + 0.00835573 | + 0.3455104 | + 0.23508984 |
| NEIGHBORHOOD.Freq. | + 0.56980275 | + 0.42367445 | + 0.1045699 | + 0.25760660 |
| SURGICAL.Freq. | + 0.07733158 | + 0.77329024 | + 0.4025833 | + 0.06878936 |
| DENTAL.Freq | + 0.32373183 | - 0.49345879 | + 0.1512031 | + 0.48404653 |

Figure 1. Table with the eigenvector values (coefficients) of the hepatitis B database. Fernandópolis (SP), Brazil, 2016.
Source: Notification Aggravation Information System (SINAN) of the city of Fernandópolis (SP).

| MCs | Eigenvalues | Proportion(%) | Acumulated proportion(%) |
|---|---|---|---|
| MC1 | 2.0797323 | 25.996654 | 25.99665 |
| MC2 | 1.5897362 | 19.871702 | 45.86836 |
| MC3 | 1.1901013 | 14.876266 | 60.74462 |
| MC4 | 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).

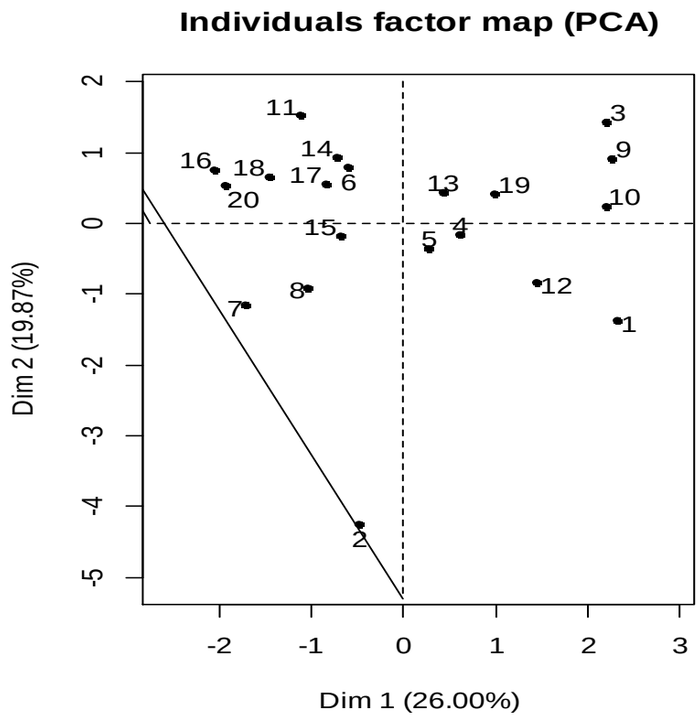


Figure 1.2 - Biplot of the HBV infected database from 2007 to 2011, generating MC1 and MC2. Fernandópolis (SP), Brazil, 2016.
Source: Notification of Invalidity Information System (SINAN) of the municipality of Fernandópolis. Fernandópolis (SP), Brazil, 2016.

In turn, the results for the hepatitis C database, which has 30 case observations and eight variables, revealed that MC 1 (24.64%) (Figure 2) has the highest coefficients of variation of this database. (Coefficient - 0.68840788) and neighborhood of the patient (coefficient + 0.64556072) were the most important variables. In Figure 2, individual six of the database was highlighted, revealing that in MC1 it was obtained a higher coefficient, with the variables “education” resulting in "illiterate" and “neighborhood of residence” resulting in "Redentor neighborhood". It was followed by the individual 25, who had "1st to 4th grade incomplete elementary school" (former primary or 1st grade) and resides in "Regina Village neighborhood", both categories of variables that had the greatest variabilities in the Hepatitis C.

In relation to MC 2 (17.68%) (Figure 2), the largest coefficients show the most important variables in this database, being "sex" (gender) (coefficient + 0.78498043) and "notification year Case "of the patient (coefficient - 0.69346448). In Figure 2, individual 1 of the database was highlighted, where in MC2, with the variables “gender” in the "female" category and the case with notification, "2007". Followed by individual four, who is female, with notification of the case in the year of "2008". The total explanation for the variability of data between MC1 and MC2 is 42.32% for the hepatitis C database.

At COP 3, the variable “exposure to dental treatment” (coefficient + 0.93486633) was highlighted. In MC4, the variable “race (ethnicity)” (coefficient + 0.50650838) was highlighted. The total explanation of the cumulative ratio between MC1, MC2 and MC3 is 58%. With the four main components, the total explanation of the cumulative ratio between MC1, MC2 and MC3 and MC4, for the hepatitis C database, is approximately 69%.

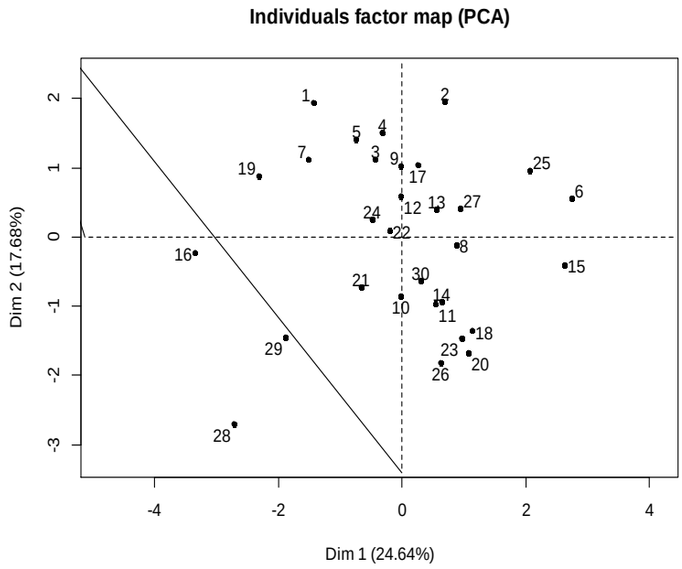


Figure 2 - Biplot of the HCV infected database from 2007 to 2011, generating MC1 and MC2. Fernandópolis (SP). 2016.

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

Finally, this study shows the results of the combination of hepatitis B and hepatitis C data, which has 50 case observations and eight variables. The MC 1 (19.70%) (Figure 3) showed the highest coefficients, showing the most important variables of this database junction, being the “neighborhood” (coefficient + 0.70469080) and “patient age” (coefficient - 0, 69058785). In Figure 3, MC 1 highlights the individual 25 from the database, showing the variables “neighborhood where he resides” in the "Paineiras", and age in the category 25 years. It was followed by individual two, who lives in the Jardim Paulista neighborhood and is 35 years old.

For MC 2 (15.84%) (Figure 3), the highest coefficients are observed for the variables "schooling" (coefficient + 0.725505640) and "exposure to dental treatment" (coefficient + 0.592724503). In Figure 3, individual five of the database was highlighted. In which in MC2, he obtained a higher coefficient, with the variables “schooling” to "complete upper" and “exposure dental treatment”, "yes, more than six months ago". It was followed by individual 12, with "complete upper" schooling and exposure to "yes to more than six months" dental treatment. The total explanation for the variability of data between MC1 and MC2 is approximately 36% relative to the unified junction database between HBV or HCV infections.

MC 3 highlighted the “sex” variable (coefficient + 0.81172170) and in MC 4 the “race (ethnicity)” variable (coefficient - 0.709506482) was highlighted. The total explanation of the cumulative ratio between MC1, MC2 and MC3 is approximately 50%. With the four main components, the total explanation of the cumulative ratio between MC1, MC2 and MC3 and MC4, for the unified junctional database between HBV or HCV infections, is approximately 63%.

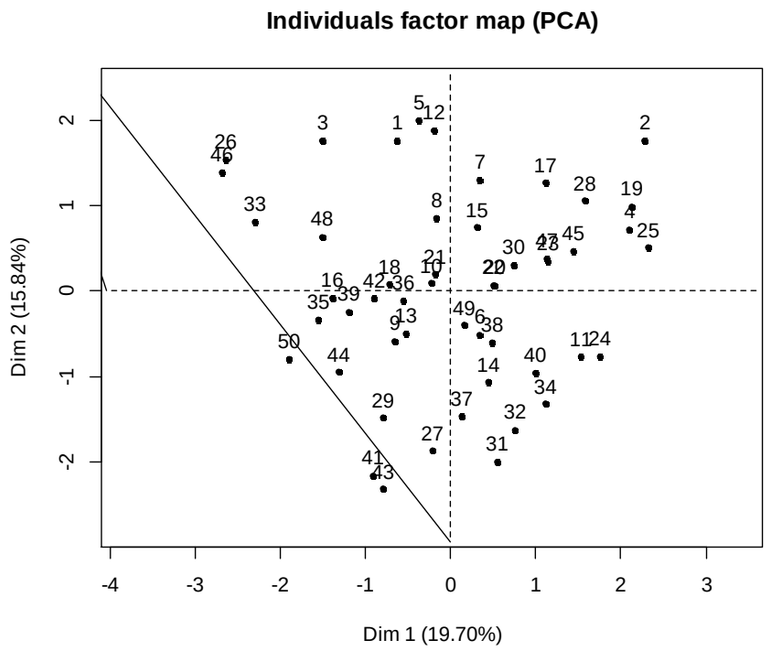


Figure 3 - Biplot of the database of those infected by HBV and HCV, in the period from 2007 to 2011, generating MC1 and MC2. Fernandópolis (SP). 2016.

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

## DISCUSSION

The result of this study, in relation to the patient's age, agrees with a study found in the literature.[17] On the other hand, it presents a contrary sense to the result found in another study, being in indigenous population, where the highest age predominates among patients with Hepatitis B (HB). Usually, they were older than 40 years.[18] For hepatitis C, literature evidence suggests the majority of cases in Brazilians with a mean age above 50 years.[17,19]

It is important to work with people at this age or between the age of 20 and 30 and over 50, because the vaccine for this disease - hepatitis B - has been available in the UHS for many years.[20]

As for schooling for HB, since there was a variability in the result, but with a relative predominance in individuals of average schooling, between high school and incomplete high school and for hepatitis C, low schooling being illiterate and fundamentally incomplete. For the combination of the database for hepatitis B and C, the "schooling" variable was complete higher education. Similar to the latter, the survey carried out in São Paulo-SP, (2004-2007) among public servants, the most frequent schooling was above 12 years, with 21.4% for HB and 18.3% for HC, a situation that revealed a significant sample recorded as ignored.[21] Observing the information that was "ignored" from this research, found in the literature, there is deduction of a possible change in the result for schooling for hepatitis B and C in this study if the "ignored" data was completed.

In an integrative review in the literature, low schooling was associated with HBV infection in prisoners.[22]

Low schooling and lack of access to information favor transmission mechanisms associated with poor socioeconomic conditions, such as: promiscuity, agglomeration, low hygiene standards and difficulty accessing quality health care.[23]

Regarding "exposure to dental treatment" and "exposure to surgical treatment". These results are similar to those found in another study where it was shown that in women, the second highest percentage for hepatitis C (HC) is exposure through dental and surgical treatment.[24]

In relation to the neighborhoods, the "place" or "geographic" variable is very important for comparing areas or communities, where the event shows different incidences. It is an alternative to suggest ideas, to identify possible aetiological factors, in general, of the environment, or to confirm the Suspicious causal associations.[25]

A recent study suggests that there was a decrease in the prevalence of hepatitis B in Brazil, but in isolated and distant localities there are still low rates of vaccination coverage, which reinforces the need to

intensify vaccination strategies for youth and adults in specific regions with greater persistence of Prevalence of HB infection.[26] Regarding hepatitis C, a previous study showed a significant decrease in Brazil from 2004, suggesting this study that could be a consequence of the creation of the National Program of Viral Hepatitis, serological tests of hepatitis in Test Centers and Counseling and Harm Reduction Programs.[27]

Based on the above, it is important that health services take into account this variable in planning health actions, noting that each place has its specific characteristics, be it socioeconomic, physical and environmental, cultural and organizational.

Regarding the female gender, these data are in agreement with another study, where 51.5%, predominated the female gender for HC.[21] It is important to highlight as hypothesis the environmental, behavioral and social organization factors, to have a connection and influence In the distribution by region of viral hepatitis in relation to the genus, and other studies should be performed for further clarification.

On the other hand, it is found, in the literature a study of HC, where the majority of the infected were male, in which they were in the age group over 40 years.[28] The authors describe that the data found suggest that individuals of the genus Were infected by HC prior to the inclusion of obligatory serological tests in Brazil and, because it was a latent and often silent infection, they began to present clinical alterations later.

Another explanation found in the literature on the prevalence of HC in males addresses several risk factors related to sexual habits that seem to contribute to a greater transmission of HC, such as: cultural factors, lack of educational sexuality actions, increased sexual partners, low adherence The use of condoms, traumatic sexual experiences, the presence of other associated sexually transmitted diseases such as syphilis, trichomoniasis, HIV / AIDS and chlamydia infection.[19] In addition, men-to-woman transmission seems to occur more easily than female-to-male transmission, but despite these evidences, studies involving monogamous couples demonstrated a low risk of sexual transmission. However, the possibility of intrafamily transmission through sharing of personal hygiene material or possible exposure to contaminated blood interferes with the interpretation of the studies that evaluated the sexual transmission of HC.

One factor to be highlighted is that UHS, through the health services, should plan and execute strategies of the HB and HC Control Program, considering the “gender” variable in the operationalization of its actions, from the unit's working hours to the types of health education strategies.

Interventions such as educational, diagnostic and treatment actions within a National Human Health Care program are critical, thinking about Sexually Transmitted Diseases (STDs), prostate cancer, cardiovascular diseases, diabetes and the adoption of healthy practices such as programs Of Women's Health Care and Pregnant Women. Another way to combat this disease is to intensify the socialization of information about its transmission, for purposes of self-protection.

Based on the data found, the importance of education through education is evident, which leads the individual to better assimilate information for the prevention of viral hepatitis infections.

Health units should take this variable into account in planning, organizing and developing their daily activities, since, in the case of a clientele with relatively low and medium schooling, educational actions should be developed in a language appropriate to the cognitive capacity of Users.

This study makes the observation that the use of a viral hepatitis database from the NIIS information system for multivariate analysis using the principal component technique for descriptive research of this study was shown to be poor in filling of several fields in the compulsory notification sheets, as variables with incomplete information, resulting in the methodology of this study, the exclusion of these several variables with zero or blank frequencies in the database, masking the final result, which would already be away from the reality of the municipality surveyed. For, in the hypothesis of analysis of the results, even with the complete variables, they would be only those who sought the health services of the municipality in the period. Care should be taken not to generalize the result to any municipality.

## CONCLUSION

With the variables and observations that the present study can work. The multivariate analysis using the principal components technique showed that the results of the analysis of the viral hepatitis database from the NIIS information system showed a higher frequency in the adult population, with a

relatively low and medium level of schooling and a predominance of the female genus for hepatitis C.

All health professionals and community health agents have an important role in carrying out educational, preventive and diagnostic actions, favoring the early treatment of these pathologies with the population. The results of thisstudy also hope to collaborate in didactics and interpretation of multivariate analysis by means of main components, as far as secondary database is concerned.

## AKNOWLEDGEMENTS

To Professor Dr. Francisco José Dutra Souto, for the intellectual critical revision of the work.

To Professor Dr. Neuber José Segri, of the Statistics Department of ICET / UFMT.

## REFERENCES

1. Word 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. Word 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

4. Word 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

10. Hotelling H. Simplifield 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.

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

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/prefpoa/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-10159000378682017620 15.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

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

Submission: 2016/09/20
Accepted: 2017/07/12
Publishing: 2017/08/01

**Corresponding Address**

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 Paraiso,
CEP: 78734-020 – Rondonópolis (MT), Brazil