



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

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

Sidnei Anastácio Sampaio¹, Kuang Hongyu², José Martins Pinto Neto³, Vando Edésio Soares⁴, Ricardo Scarparo Navarro⁵

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.

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.

¹Enfermeiro, Professor Mestre, Graduação em Enfermagem/Enf/ICEN/CUR, Universidade Federal do Mato Grosso/UFMT. Rondonópolis (MT), Brasil. E-mail: sidnei.sampaio1@gmail.com; ²Matemático, Professor Doutor, Graduação em Estatística/Est/ICET, Universidade Federal do Mato Grosso/UFMT. Cuiabá (MT), Brasil. E-mail: prof.kuang@gmail.com; ³Enfermeiro, Professor Doutor, Graduação Enfermagem/Medicina, Universidade Camilo Castelo Branco/UNICASTELO. Fernandópolis (SP), Brasil. E-mail: jmpintoneto@gmail.com; ⁴Zootecnista, 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; ⁵Cirurgião-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

INTRODUCCIÓN

Se estima que 240 millones de personas estén crónicamente infectadas en todo el mundo por el virus de la hepatitis B (VHB).¹ Cerca de 800 mil personas mueren cada año a partir de esta enfermedad y menos de 650 mil de cirrosis y cáncer de hígado debido a infección crónica de la hepatitis B (HB) y otras 130 mil de hepatitis B aguda, principalmente por complicaciones de cirrosis hepática o por cáncer primario del hígado (carcinoma hepatocelular, CHC).² Para efectos de comparación, el virus de la hepatitis B es de 50 a 100 veces más infeccioso que el Virus de la Inmunodeficiencia Humana (VIH).³

En comparación a la infección por el Virus de la Hepatitis C (VHC), existen aproximadamente de 130 a 150 millones de personas portadores del virus. Un número significativo de personas que sufren de infección crónica desarrollará cirrosis hepática o cáncer de hígado CHC. Se estima que el número de personas que mueren anualmente de enfermedades hepáticas relacionadas al virus de la hepatitis C a cada año esté entre 350 a 500 mil.⁴

En Brasil, según datos sujetos a revisión del Sistema de Información de Agravios de Notificación (SINAN), en 2014, fueron notificados 13.187 casos confirmados de hepatitis B, seguidos de 51 casos de asociación entre hepatitis B y A (B+A), 333 para B+C, 226 de hepatitis B+D. En ese mismo año, para la hepatitis C, hubo 12.934 casos confirmados, seguido de 26 casos de hepatitis C+A. El montante general de casos notificados en Brasil fue de 34.637 casos confirmados en todo el período de 2014, para todas las clasificaciones de etiología de las hepatitis virales.⁵

En los residentes del estado de São Paulo, en 2014, fueron notificados un total de 2.229 casos confirmados de hepatitis B, seguido de asociación de las hepatitis virales B+C con 72 casos, B+A con 4 casos. En relación a la hepatitis C, fueron notificados 3.192 casos confirmados, seguidos de 4 casos de C+A.⁵

El análisis de componentes principales (ACP) es una técnica multivariada de modelaje de la estructura de covariancia. La técnica fue inicialmente descrita por Pearson 1901.⁶⁻⁷ Una descripción de métodos computacionales prácticos vino mucho más tarde, con Hotelling en 1933 y 1936,⁸⁻⁹ quien la usó con el propósito determinado de analizar las estructuras de correlación.⁷

El análisis de los componentes principales es una de las más importantes herramientas

estadísticas del análisis multivariado, por constituir la base donde se fundamentan la mayoría de los otros métodos multivariados de análisis de datos, siendo una herramienta de análisis exploratorio de datos que permite revelar la existencia o no de muestras diferentes, de relaciones entre las variables medidas y de relaciones o agrupamientos entre muestras, visando su reducción, manteniendo apenas un número pequeño de Componentes Principales (CPs) que ya responden por una parcela significativa de la información total contenida en la estructura de datos.¹⁰

Como enfermedades transmitidas por sangre y fluidos corpóreales, infecciones por los VHB o VHC pueden ser adquiridas en establecimientos en que existe exposición a material cortante, como agujas, tijeras y láminas. El presente estudio fue realizado en el municipio de Fernandópolis, estado de São Paulo, donde existen varios establecimientos de salud y de otras actividades que propician aumento del riesgo de infecciones por hepatitis virales B y C de heridas con inoculación de material contaminado, como en los casos de los que hacen tatuajes, manicures y pedicuros, peluqueros, centros de estética, consultorios odontológicos, hospitales, unidades de salud, farmacias, casas de profesionales de sexo, entre otros.

La enfermedad parece ser bien conocida por la población y por los profesionales de salud, pero es posible que muchos no sigan las orientaciones sobre el uso del material individual o descartable, lo que lleva a vulnerabilidad para la infección/enfermedad.

Por lo tanto, con base en lo expuesto, se torna evidente la gravedad y problemática de las hepatitis virales B y C, lo que justifica la realización de este estudio. Por la importancia de ambas infecciones para la salud humana, los estudios de prevalencia o incidencia de las mismas se reviste de relevancia, así como el esfuerzo para optimizar el análisis de los bancos de datos formados por esa investigación. El presente estudio se fundamenta en el análisis de las variables de un banco de datos sobre hepatitis B y C en Fernandópolis (SP), con el intuito de identificar las principales variables que emanaron del análisis.

OBJETIVOS

- Identificar cuales variables son las más importantes en el banco de datos de hepatitis viral B, de hepatitis viral C y de la unión de los bancos de datos de las hepatitis B y C.
- Reducir el número de variables de los bancos de datos de hepatitis viral B, de

hepatitis viral C y de la unión de los bancos de datos de las hepatitis virales B y C.

MÉTODO

El banco de datos del presente estudio fue extraído del Programa de Post-Graduación en Ingeniería Biomédica/UNICASTELO, disertación de maestría intitulada “Incidencia de Hepatitis Virales B y C e Fernandópolis S.P.”.

Estudio descriptivo y retrospectivo, de enfoque cuantitativo. Utilizó datos obtenidos a partir de una fuente secundaria del Sistema de Información de Agravios de Notificación (SINAN) disponible en el Servicio de Vigilancia Epidemiológica de la Secretaría Municipal de Salud del municipio de Fernandópolis (SP), referente a los casos notificados y confirmados de hepatitis virales B y C, en el período de 2007-2011, residentes en este municipio.

El Proyecto fue aprobado por el Comité de Ética e Investigación de la Universidad Camilo Castelo Branco (CEP UNICASTELO CAAE nº01475612.7.0000.5494). Los datos fueron analizados por el método de análisis multivariado, por medio de análisis de componentes principales, a nivel de significancia de 5%, por el programa estadístico plataforma R.

El banco de datos bruto contenía 46 variables con 158 observaciones de casos confirmados de hepatitis B y C, siendo 51 casos de infección por el VHB, 101 de infección por el VHC y ocho casos de asociación entre los dos agentes VHB y VHC. Fueron retirados del banco de datos bruto todas las variables cuya información era de respuesta “No” y cuando la “mayoría de los datos como No, Ignorado y en Blanco”, siendo ellas (Gestante; Zona de residencia; Ocupación; Sospecha de hepatitis; Tomó vacuna; Agravios asociados; Contacto con paciente portador de VHB y VHC; El paciente fue Sometido o Expuesto: medicamento inyectable, drogas inhalables o crack, drogas inyectables, agua/alimento contaminado, tres o más compañeros sexuales, trasplante, tatuaje/*piercing*, acupuntura, hemodiálisis, accidente con material biológico, transfusión de sangre/derivados; Resultado de la serología banco de sangre; Genotipo VHC; Clasificación Final; Forma Clínica; Clasificación Etiológica y Probable fuente de infección). Fueron retiradas también variables con respuesta única como (Municipio de Residencia, Regional de Salud; Unidad de salud; País; Unidad Federativa) y variables no pertinentes, o sea, sin sentido para hacer análisis como por ejemplo fechas (día, mes). Fueron excluidas

también todas las frecuencias (observaciones) en blanco e ignorado de las variables que permanecieron en el estudio para análisis. El total de variables utilizadas en el trabajo fue de ocho variables para cada banco de datos de HB, HC y unión de banco de datos de HB/HC. Fueron 20 casos confirmados de HB, 30 casos de HC y 50 casos unión de bancos de datos de HB y HC.

El principal local de recolección de datos de la referida disertación de maestría, el cual sirvió para el análisis de la presente investigación fue el municipio de Fernandópolis (SP).

◆ Análisis Multivariado

La técnica de análisis multivariado se refiere a todas las técnicas estadísticas que simultáneamente analizan múltiples medidas sobre individuos u objetos sobre investigación. Siendo así, cualquier análisis simultánea de más de que dos variables pueden ser considerados, a principio, como multivariado, posibilitando que órganos sanitarios y organizaciones adquieran conocimiento y mejoren así sus tomadas de decisiones.¹¹

Análisis de Componentes Principales

El análisis de componentes principales tiene como objetivo reducir la dimensión del conjunto de variables y facilitar la interpretación de la independencia entre ellas.^{7,11} Para eso, son obtenidas combinaciones lineales de las variables originales que, geométricamente representan la selección de nuevos sistemas de coordenadas obtenidos por la rotación del sistema original que tiene *p* variables aleatorias como ejes de las coordenadas. Esos nuevos ejes ortogonales (nuevas variables) son llamados de componentes principales y los valores de las nuevas variables son llamado de puntos de los componentes principales o coordenadas principales.⁷

En general, se elige el componente principal de mayor importancia (el primer componente principal) como siendo aquel de mayor variancia, que explique el máximo de variabilidad de los datos, el segundo componente de mayor importancia, lo que presenta la segunda mayor variancia, y así sucesivamente, hasta el componente principal de menor importancia.^{7,12-3} Por otro lado, los últimos componentes principales serán responsables por direcciones que no están asociadas a mucha variabilidad. En otras palabras, esos últimos componentes principales identificarán relaciones lineares entre las variables originales próximo del constante.^{7,14-6}

La obtención de los componentes principales es realizada por medio de la diagonalización de matrices simétricas positivas semi-definidas. Entonces, se puede calcular los componentes principales fácilmente e utilizarlos en diferentes aplicaciones en las más variadas áreas científicas. Esta facilidad es en función de la existencia de inúmeros programas capaces de realizar cálculos matriciales para diagonalizar una matriz simétrica positiva semi-definida.⁷

Muchos investigadores han utilizado el análisis de componentes principales para resolver problemas como de la multicolinearidad en regresión linear, para estimar factores, que representan otra técnica multivariada de modelaje de la matriz de covariancias, para realizar el modelaje de la interacción entre factores en experimentos sin repetición, estudios de divergencia y agrupamiento entre genotipos en estudio de genética y mejora de plantas y animales, entre otras posibilidades.^{7,14}

Para obtener los componentes principales de una forma general, sea un conjunto de p variables $X_1, X_2, \dots, X_p$, con medias $\mu_1, \mu_2, \dots, \mu_p$ y variancia $\sigma^2_1, \sigma^2_2, \dots, \sigma^2_p$, respectivamente, estas variables no son independientes y por lo tanto, poseen covariancia entre la i -ésima y k -ésima variables definida por σ_{ik} , para $i \neq k = 1, 2, \dots, p$. Entonces las p variables pueden ser expresas en forma vectorial por: $X = [X_1, X_2, \dots, X_p]^T$, con vector de medias $\mu = [\mu_1, \mu_2, \dots, \mu_p]^T$ y matriz de covariancia Σ . Se encuentran los pares de auto-valores y auto-vectores $(\lambda_1, e_1), (\lambda_2, e_2), \dots, (\lambda_p, e_p)$, en que $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$, asociados a Σ y entonces el i -ésimo componente principal es definido por:^{7,14}

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

(1)

en que $i = 1, 2, \dots, p$.

RESULTADOS

Para el banco de datos de hepatitis B, que posee 20 observaciones de casos y 8 variables, el componente principal 1 (CP 1), o sea, auto-vector 1, obtuvo el mayor porcentaje (26,00%) (Figura 1.2) entre todos los CPs. Hubo un contraste entre las variables. Una aflora con señal positivo para el coeficiente (auto-vectores) y la otra con señal negativo para otro coeficiente. De modo que, esto significa que a llevar en consideración el señal para interpretación del resultado en relación a las variables, ellas poseen valores contrarios en comparación una con la otra. Tiene la mayor variación de los datos las variables edad del

paciente (coeficiente - 0,76738590) y escolaridad del paciente (coeficiente + 0,77306034), (Figura 1). La Figura 1.2 destacó el individuo 9 del banco de datos, siendo que en el CP1 fue el que obtuvo la mayor variación de los datos, con las variables edad y escolaridad, esas variables en sus categorías mostraron que el individuo 9 tiene 22 años y enseñanza media completa (antiguo colegial o 2º grado), seguido del individuo 3 que tiene edad de 28 años y escolaridad superior completa, en seguida el individuo 10 que tiene edad de 25 años y escolaridad de enseñanza media incompleta (antiguo colegial o 2º grado), el que corresponde con el banco de datos observado. Todas esas variables del CP1 fueron las que más explicaron la variación de los datos de hepatitis B, para este banco de datos.

Generalizando el resultado del CP1 para los individuos observados, se puede decir que los individuos que aparecen en la Figura 1.2 (gráfico) en el eje X a partir de los valores 0 (cero) a 3 (tres) positivos, o sea, cuadrantes positivos (parte derecha superior e inferior) del eje X representando el CP1, son los individuos que poseen menos edad y nivel alto de escolaridad. Ya en la misma Figura 1.2 se puede decir que los individuos que aparecen en el eje X próximo de los valores - 2 (menos dos) y próximo de cero (0) negativos, o sea, cuadrantes negativos (parte izquierda superior e inferior) del eje X representando el CP1, son los individuos que poseen más edad y nivel bajo de escolaridad, lo que corresponde con el banco de datos observado. Esta misma interpretación sirve para las demás variables de los otros bancos de datos, cuando acontece contraste.

El CP 2 del banco de datos de hepatitis B, explicó 19,87% (Figura 1.2) de la variación de los datos. Existe un contraste entre las variables, siendo las variables más importantes de este CP2 “exposición a tratamiento quirúrgico” (coeficiente + 0,77329024) y “exposición a tratamiento dentario” (coeficiente - 0,49345879) (Figura 1), este contraste es observado por los señales contrarios de los coeficientes, positivo para una variable y negativo para otra variable. La figura 1.2 destacó al individuo 11 del banco de datos, siendo que en el CP2 fue el que obtuvo mayor coeficiente, con las variables exposición a tratamiento quirúrgico (sí, hace más de seis meses) y exposición al tratamiento dentario (sí hace menos de seis meses). El total de explicación de la variabilidad de los datos entre CP1 y CP2 es de aproximadamente 46% para el banco de datos de hepatitis B, (Figura 1.1).

Acrecentando el CP 3 (Figura 1) se destaca la variable sexo (género) con mayor coeficiente (- **0.6829707**) y eleva el poder de explicación de la variabilidad de los datos en el banco de datos de hepatitis B en aproximadamente 61% (Figura 1.1) conforme muestra la proporción acumulada entre los tres y más importantes componentes principales (CP1, CP2 e CP3). Por otro lado,

acrecentando más un componente principal en la secuencia, en este caso, el CP4 (Figura 1) se destaca la variable raza (etnia) con coeficiente (- **0.66271869**), la explicación total del banco de datos de hepatitis B, utilizando los cuatro y más importantes componentes principales (CP1, CP2, CP3 y CP4) es de aproximadamente 74% (Figura 1.1).

Variables	CP1	CP2	CP3	CP4
AÑO	- 0.47231924	+ 0.43805208	+ 0.4678939	-0.20556668
EDADE_año	- 0.76738590	- 0.16221765	+ 0.2677251	+ 0.37747495
SEXO.Frec.	+ 0.19422132	+ 0.47380797	- 0.6829707	- 0.01982262
RAZA.Frec.	+ 0.44380612	- 0.35494897	+ 0.3432241	- 0.66271869
ESCOLARIDAD.clase.	+ 0.77306034	+ 0.00835573	+ 0.3455104	+ 0.23508984
BARRIO.Frec.	+ 0.56980275	+ 0.42367445	+ 0.1045699	+ 0.25760660
QUIRURGICO.Frec.	+ 0.07733158	+ 0.77329024	+ 0.4025833	+ 0.06878936
DENTARIO.Frec	+ 0.32373183	- 0.49345879	+ 0.1512031	+ 0.48404653

Figura 1. Cuadro con los valores de los auto-vectores (coeficientes) del banco de datos de hepatitis B. Fernandópolis (SP), Brasil, 2016. Fuente: Sistema de Información de Agravio de Notificación - (SINAN) del municipio de Fernandópolis (SP).

CPs	Auto-valores	Proporción(%)	Proporción 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

Figura 1.1: Cuadro con los auto-valores mostrando la proporción explicada por los cuatro primeros y más importantes CPs, banco de datos HB. 2016. Fuente: Sistema de Información de Agravio de Notificación - (SINAN) del municipio de Fernandópolis (SP).

MAPA DE FACTOR INDIVIDUALES

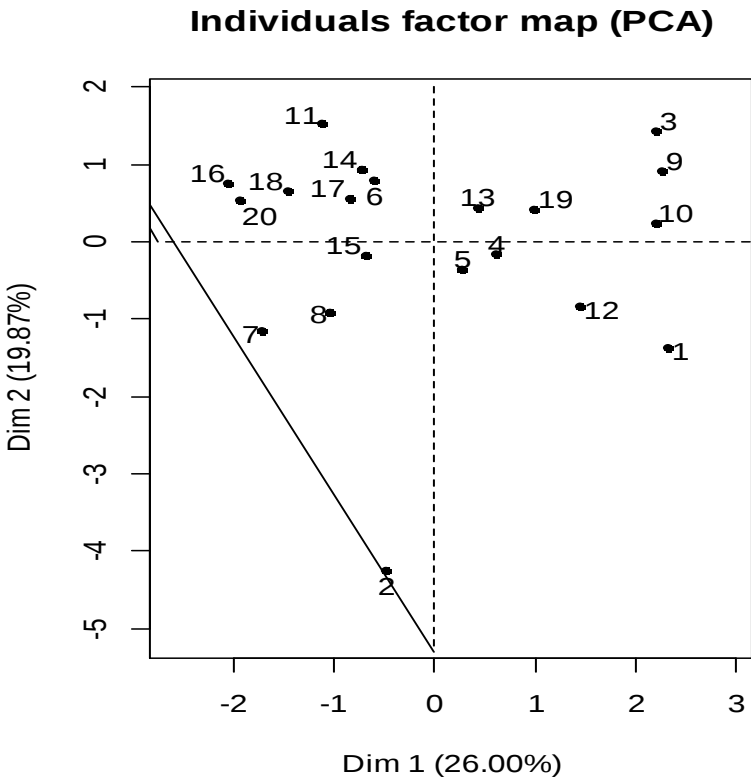


Figure 1.2: Biplot del banco de datos dos infectados por el VHB en el período de 2007 a 2011, generando los CP1 y CP2. Fernandópolis (SP), Brasil, 2016. Fuente: Sistema de Información de Agravio de Notificación - (SINAN) del municipio de Fernandópolis. Fernandópolis (SP), Brasil, 2016.

Por su vez, los resultados para el banco de datos de la hepatitis C, que posee 30 observaciones de casos y 8 variables, reveló que el CP 1 (24,64%) (Figura 2.) posee los mayores coeficientes de variaciones de este banco de datos, siendo las variables escolaridad (coeficiente - 0,68840788) y barrio del paciente (coeficiente + 0,64556072) las más importantes. En la Figura 2, se destacó el individuo 6 del banco de datos, revelando que en el CP1 fue el que obtuvo mayor coeficiente, con las variables escolaridad resultando “analfabeto” y barrio de residencia resultando “barrio Redentor”. Fue seguido del individuo 25, que tuvo escolaridad “1ª a 4ª serie de enseñanza fundamental incompleta” (antiguo primario o 1º grado) y reside en el “barrio vila Regina”, ambas categorías de las variables que tuvieron las mayores variabilidades del banco de datos de la hepatitis C.

En relación al CP 2 (17,68%) (Figura 2.), los mayores coeficientes muestran las variables más importante de este banco de datos, siendo “sexo” (género) (coeficiente +

0,78498043) y “año de notificación del caso” del paciente (coeficiente - 0,69346448). En la Figura 2, se destacó el individuo 1 del banco de datos, donde en el CP2, con las variables género en la categoría “femenino” y el caso con notificación, “2007”. Seguido del individuo 4, que es del género femenino, con notificación del caso en el año de “2008”. El total de explicación de la variabilidad de los datos entre CP1 y CP2 es de 42,32% para el banco de datos de la hepatitis C.

En el CP 3, se destacó la variable exposición al tratamiento dentario (coeficiente + 0.93486633). En el CP4, se destacó la variable raza (etnia) (coeficiente + 0.50650838). El total de explicación de la proporción acumulada entre los CP1, CP2 y CP3 es de 58%. Con los cuatro componentes principales, el total de explicación de la proporción acumulada entre los CP1, CP2 y CP3 y CP4 para el banco de datos de hepatitis C es de aproximadamente 69%.

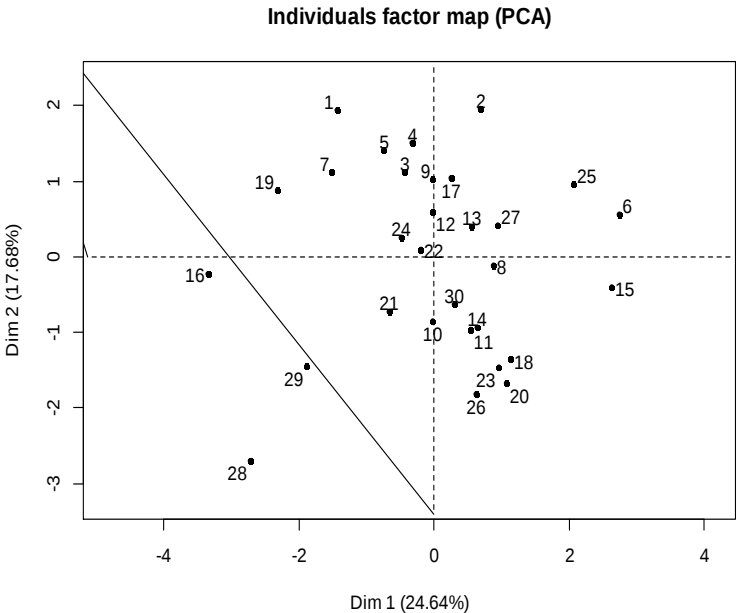


Figure 2: Biplot del banco de datos de los infectados por el VHC en el período de 2007 a 2011 generando los CP1 y CP2. Fernandópolis (SP). 2016.
Fuente: Sistema de Información de Agravio de Notificación - (SINAN) del municipio de Fernandópolis (SP).

Por fin, el presente estudio muestra el resultado de la junción de los datos de hepatitis B con hepatitis C que tienen 50 observaciones de casos y 8 variables. El CP 1 (19,70%) (Figura 3) obtuvo los mayores coeficiente, mostrando las variables más importantes de esta junción de banco de datos, destacándose barrio (coeficiente + 0,70469080) y edad del paciente (coeficiente - 0,69058785). En la Figura 3, el CP 1 destaca el individuo 25 del banco de datos, presentando las variables barrio donde reside en las

“Paineiras”, y edad en la categoría 25 años. Fue seguido del individuo 2, que reside en el barrio Jardim Paulista y tiene edad de 35 años.

Para el CP 2 (15,84%) (Figura 3), los mayores coeficiente, son observados para las variables “escolaridad” (coeficiente + 0.725505640) y “exposición al tratamiento dentario” (coeficiente + 0.592724503). En la Figura 3, se destacó el individuo 5 del banco de datos, en el cual en el CP2, obtuvo mayor coeficiente, con las variables respuestas a la

escolaridad “superior completa” y, exposición al tratamiento dentario “sí hace más de seis meses”. Fue seguido del individuo 12, con escolaridad “superior completa” y exposición al tratamiento dentario “sí hace más de seis meses”. El total de explicación de la variabilidad de los datos entre CP1 y CP2 es de aproximadamente 36% en relación al banco de datos junción unificado entre las infecciones por el VHB o VHC.

En el CP3 se destacó la variable sexo (coeficiente + 0.81172170) y en el CP 4 se

destacó la variable raza (etnia) (coeficiente - 0.709506482). El total de explicación de la proporción acumulada entre los CP1, CP2 y CP3 es de aproximadamente de 50%. Con los cuatro componentes principales, el total de explicación de la proporción acumulada entre los CP1, CP2 y CP3 y CP4 para el banco de datos junción unificada entre las infecciones por el VHB o VHC es de aproximadamente 63%.

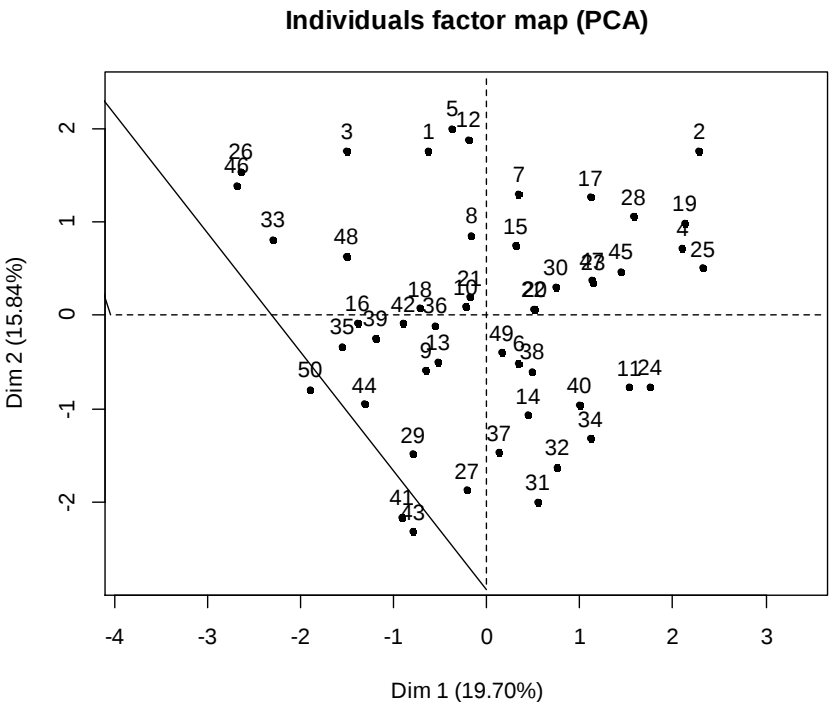


Figure 3. Biplot del banco de datos de los infectados por el VHB y VHC en el período de 2007 a 2011 generando los CP1 y CP2. Fernandópolis (SP). 2016.
Fuente: Sistema de Información de Agravio de Notificación - (SINAN) del municipio de Fernandópolis (SP).

DISCUSIÓN

El resultado del presente estudio, en relación a la edad del paciente, condice con un estudio encontrado de literatura.¹⁷ Por otro lado, presenta sentido contrario al del resultado encontrado en otro estudio, siendo en población indígena, donde predomina edad más elevada entre los pacientes con hepatitis B (HB). Normalmente, fueron con más de 40 años.¹⁸ Para la hepatitis C, evidencias de literatura sugieren la mayoría de los casos en los brasileños con edad media superior a los 50 años.^{17,19}

Es importante el trabajo de prevención con las personas en esta edad o entre este grupo de edad de los 20 a los 30 años y con más de 50 años, pues vacuna para esta enfermedad de hepatitis B está disponible en el SUS hace muchos años.²⁰

Sobre la escolaridad para HB, ya que hubo una variabilidad en el resultado, pero con un relativo predominio en individuos de media

escolaridad, entre enseñanza media completa y enseñanza media incompleta y para hepatitis C baja escolaridad siendo analfabeto y fundamental incompleto. Para la junción del banco de datos para las hepatitis B y C la variable escolaridad fue superior completo, semejante a este último es la investigación realizada en São Paulo-SP, 2004-2007 entre servidores públicos, el grupo de escolaridad más frecuente fue con más de 12 años, con 21,4% para la HB y 18,3% para HC, situación que reveló una muestra significativa registrada como ignorado.²¹ Observando la información que hubo ignorados de esta investigación encontrado en la literatura, se deduce un posible cambio en el resultado para escolaridad para hepatitis B y C en esta investigación si los datos de ignorado fuesen rellenados.

En un estudio de revisión integradora en la literatura reveló que la baja escolaridad está asociada a infección por el VHB en personas del sistema prisional.²²

La baja escolaridad y el no acceso a la información favorecen mecanismos de transmisión asociados a las condiciones socioeconómicas deficientes, sean promiscuidad, aglomeración, bajos padrones de higiene y dificultad de acceso a una asistencia médica de calidad.²³

En relación a la exposición al tratamiento dentario y exposición al tratamiento quirúrgico, esos resultados son semejantes a los encontrados en otro estudio donde fue evidenciado que, en las mujeres, el segundo mayor porcentaje para la hepatitis C (HC) es la exposición vía tratamiento odontológico y quirúrgico.²⁴

En relación a los barrios, la variable lugar o geográfica es muy importante para comparación de áreas o comunidades, en que el evento muestra diferentes incidencias, es una alternativa para sugerir ideas, identificar posibles factores etiológicos, en general, del medio ambiente, o confirmar las asociaciones causales sospechadas.²⁵

Estudio reciente sugiere que hubo una caída en la prevalencia de hepatitis B en Brasil, pero en localidades aisladas y distantes todavía hay bajas tasas de cobertura de vacunas, lo que refuerza la necesidad de intensificar las estrategias de vacunación para jóvenes y adultos en regiones específicas con persistencia mayor de prevalencia de la infección por la HB.²⁶ Cuanto a la hepatitis C, estudio previo mostró caída significativa en Brasil a partir de 2004, sugiriendo este estudio que podría ser consecuencia de la creación del Programa Nacional de Hepatitis Virales, testes serológicos de hepatitis en los Centros de Test y Consejo y Programas de Reducción de Daños.²⁷

Con base en lo expuesto, es importante que los servicios de salud lleven en consideración esta variable en el planeamiento de las acciones en salud, recordando que cada lugar tiene sus especificidades, sea socioeconómica, física y ambiental, cultural y organizacional.

En relación al género femenino, esos datos condicen con el encontrado en otro estudio, donde 51,5% predominó el género femenino para la HC.²¹ Es importante destacar como hipótesis los factores ambientales, comportamentales y la organización social, tener una ligación e influencia en la distribución por región de las hepatitis virales en relación al género, debiendo otros estudios ser realizados para mayores aclaraciones.

Por otro lado, es encontrado en la literatura estudio de HC, donde la mayoría de los infectados eran del género masculino, en el cual se encontraban en el grupo de edad de

más de 40 años.²⁸ Los autores describen que los datos encontrados sugieren que los individuos del género masculino fueron infectados por la HC antes de la inclusión de exámenes serológicos obligatorios en Brasil y, por tratarse de una infección latente y muchas veces silenciosa, pasaron a presentar alteraciones del cuadro clínico pero tardíamente.

Otra explicación encontrada en la literatura de predominio de HC en el sexo masculino enfocan varios factores de riesgo relacionados a hábitos sexuales que parecen contribuir para mayor transmisión de la HC, como: factores culturales, falta de acciones educativas de sexualidad, mayor número de compañeros sexuales, baja adherencia al uso de preservativos, experiencias sexuales traumáticas, presencia de otras enfermedades sexualmente transmisibles asociadas como sífilis, tricomoniasis, VIH/SIDA e infección por clamidia.¹⁹ Además de enfoque en que la transmisión hombre-mujer parece darse más fácilmente de que la transmisión mujer-hombre, pero que a pesar de esas evidencias, estudios envolviendo parejas monogámicos demostraron bajo riesgo de transmisión sexual. Con todo, la posibilidad de transmisión intrafamiliar por compartimento de material de higiene personal o eventual exposición a sangre contaminada interfiere en la interpretación de los estudios que evaluaron la transmisión sexual de la HC.

Un factor a ser destacado es que el SUS, a través de los servicios de salud, debe planear y ejecutar estrategias del Programa de Control de HB y HC, considerando la variable género en la operación de sus acciones, desde el horario del expediente de la unidad hasta los tipos de estrategias de educación en salud.

Intervenciones como acciones educativas, de diagnóstico y tratamiento dentro de un programa Nacional de Atención a la Salud del Hombre son fundamentales, pensando en las enfermedades sexualmente transmisibles (DSTs), cáncer de próstata, enfermedades cardiovasculares, diabetes y adopción de prácticas sanas, como realizado en programas de Atención a la Salud de la Mujer y Gestantes. Otra manera de combatir esta enfermedad es intensificar la socialización de informaciones sobre su transmisión, para fines de autoprotección.

Con base en los datos encontrados, es evidente la importancia del grado de instrucción por medio de la educación, lo que lleva al individuo a asimilar mejor las informaciones para prevención de las infecciones de hepatitis virales.

Las unidades de salud deben llevar en consideración esta variable en el planeamiento, organización y en el desarrollo de sus actividades diarias, pues, tratándose de una clientela con relativa baja y media escolaridad, las acciones educativas deben ser desarrolladas en un lenguaje adecuado a la capacidad cognitiva de los usuarios.

El presente estudio observa que la utilización de banco de datos de hepatitis virales viniendo del sistema de información SINAN para análisis multivariado utilizando la técnica de componentes principales para investigación descriptiva del presente estudio, se mostró con poca calidad en lo que se refiere a completar varios campos en las fichas de notificación compulsoria, como variables con informaciones incompletas, teniendo como consecuencia en la metodología de este estudio, la exclusión de esas varias variables con frecuencias cero o en blanco en el banco de datos, enmascarando el resultado final, del que, ya sería lejos de la realidad del municipio investigado. Pues, en la hipótesis de análisis de los resultados, mismo con las variables completas, serían apenas de los que buscaron los servicios de salud del municipio en el período. Se debe tener el cuidado de no generalizar el resultado para todo el municipio.

CONCLUSIÒN

Con las variables y observaciones que este estudio puede trabajar. El análisis multivariado por medio de la técnica de componentes principales mostró que el resultado del análisis del banco de datos de hepatitis virales proviniendo del sistema de información SINAN, mostró mayor frecuencia en la población adulta, con relativa baja y media escolaridad y predominio del género femenino para hepatitis C.

Todos los profesionales de salud y agentes comunitarios de salud tienen importante papel en la realización de acciones educativas, preventivas y de diagnóstico, favoreciendo el tratamiento precoz de esas patologías junto a la población. Los resultados del presente estudio también esperan colaborar en la didáctica e interpretación de análisis multivariado por medio de componentes principales, en lo que se refiere a banco de datos secundarios.

AGRADECIMIENTOS

Al profesor Dr. Francisco José Dutra Souto por la revisión crítica intelectual del trabajo.
Al profesor Dr. Neuber José Segri del departamento de Estadística de ICET/UFMT.

REFERENCIAS

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

7. 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: <https://pdfs.semanticscholar.org/07f2/665693e0903dddcf8e146da9e443f4709a057.pdf>

11. Lyra WS, Silva EC, Araújo MCU, Frago 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.

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

Utilización del análisis multivariado para datos...

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: <http://portalsaude.saude.gov.br/images/pdf/2014/novembro/27/guia-vigilancia-saude-linkado-27-11-14.pdf>
23. 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>
24. 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
25. 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>
26. 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
27. Pereira MG. Epidemiologia: teoria e prática. Rio de Janeiro: Guanabara Koogan; 2015.
28. 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>
29. 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: http://www.revista.ufpe.br/revistaenfermagem/index.php/revista/article/view/6044/pdf_7544
30. 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>

Sumisión: 20/09/2016

Acepto: 12/07/2017

Publicado: 01/08/2017

Correspondencia

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